Biology

for the IB DIPLOMA

C J Clegg

Biology

for the IB DIPLOMA

Biology

for the IB DIPLOMA

C J Clegg

HODDER EDUCATION
PART OF HACHETTE LIVRE UK

The cover image shows a green anole (*Anolis carolinensis*), a tree-dwelling reptile found in parts of North and South America that may grow to a length of 20 cm. Here it is perched momentarily upon a pitcher plant. These active, agile carnivores have pads on their feet that enable them to cling, climb and run on virtually any surface, as they hunt for crickets, cockroaches, spiders and moths. Prey is caught in the wide jaws, with the aid of a long, forked tongue. Their skin is characteristically scaly, and the anole, typically green coloured, can change its colour, as the related chameleon lizards do.

Pitcher plants are examples of 'carnivorous' plants; they grow and compete well on soils typically deficient in essential ions, such as nitrates. These photosynthetic organisms trap insects and other tiny animals in their tall-sided pitchers, lined internally with smooth, down-pointing scales. When the 'prey' dies it decays by bacterial action, aided by enzymes secreted in the pitcher. The ions released are actively absorbed into the plant's cells, enabling essential metabolic processes such as protein synthesis.

Orders: please contact Bookpoint Ltd, 130 Milton Park, Abingdon, Oxon OX14 4SB. Telephone: (44) 01235 827720. Fax: (44) 01235 400454. Lines are open 9.00–5.00, Monday to Saturday, with a 24-hour message answering service. Visit our website at www.hoddereducation.co.uk.

First published in 2007 by
Hodder Murray, an imprint of Hodder Education,
part of Hachette Livre UK,
338 Euston Road
London NW1 3BH

Impression number 5 4 3 2
Year 2010 2009 2008

Cover photo © David Aubrey/Science Photo Library
Illustrations by Oxford Designers & Illustrators, Barking Dog Art
Typeset in 10/12pt Goudy and Frutiger families by Eric Drewery
Printed in Italy

A catalogue record for this title is available from the British Library

ISBN: 978 0340 92652 9

Contents

Section 2: Additional higher level

Section 3: Options

this section appears in this ebook following the answers and index

Option A

Option B

Introduction

The International Baccalaureate Diploma programme, a pre-university course for 16–19-year-olds, is designed to develop not only a breadth of knowledge, skills and understanding, but well-rounded individuals and engaged world citizens. One of the Diploma's key requirements is concurrent study in six academic areas, at least one of which is an experimental science. Of these, biology, whether taken at standard or higher level, is the choice of many students. **This book is designed to serve them.**

Within the IB Diploma programme, the theory content for biology is organised into compulsory core topics and options. The organisation of this book follows that syllabus sequence:

- **Section 1** is the **common core material for standard and higher-level students**: Chapters 1–7;
- **Section 2** is the **additional higher level material for higher-level students**: Chapters 8–12;
- **Section 3** consists of the **options,** of which three are available to standard students only (Chapters 13–15), four are for standard and higher-level students (Chapters 16–19), and one is available to higher-level students only (Chapter 20).

The syllabus is presented as topics and options, and most are the subject of a single chapter in *Biology for the IB Diploma* (see pages xi and xii). The exceptions are Topics 3 and 4, which are split between two chapters in each case, in order to facilitate design and delivery of individual teaching programmes. The topic on statistical analysis is presented in Chapter 21.

Special features of the chapters of *Biology for the IB Diploma*:

- each begins with 'Starting points' that summarise the essential concepts on which the chapter is based; where the issues to be addressed have their genesis in earlier chapters, these are identified
- the text is written in straightforward language, uncluttered by phrases or idioms that might confuse **students for whom English is a second language**; the depth of treatment of topics carefully reflects the objectives and action verbs in which the syllabus assessment statements are phrased
- photographs, electron micrographs and full-colour **illustrations are linked to support the relevant text**, with annotations included to elaborate the context, function or applications
- the main sections within chapters specify the **IB syllabus subtopics paragraph numbers** being addressed, so links between text and syllabus are self-evident
- explanations of structure are linked to function and behaviour in living things, and the habitat and environment of organisms are identified where appropriate; application of biology in industry, and the economic, environmental and ethical consequences of developments are highlighted, where appropriate
- **processes of science** (science methods) and something of the history of developments are introduced selectively to aid appreciation of the possibilities and limitations of science
- self-assessment questions (**SAQs**) are phrased so as to assist comprehension and recall, but also help familiarise students with the assessment implications of the action verbs; at the end of chapters, typical examination questions are given
- links to the interdisciplinary **Theory of Knowledge** (TOK link) element of the IB course are made at appropriate points in most chapters.

Using this book

The sequence of chapters in *Biology for the IB Diploma* follows the sequence of the syllabus contents. However, the IB Diploma biology syllabus is not designed as a teaching syllabus, and the order in which the syllabus content is presented is not necessarily the order in which it should be taught. Different schools and colleges need to design a course delivery model based on individual circumstances. How this may be tackled is discussed in **Chapter 22**.

In addition to the study of theory issues on which this book focuses, IB science students are also involved in practical investigations and project work. Investigations are ultimately presented for the **internal assessment**, based on given assessment criteria. How all these components are integrated is also the subject of Chapter 22. This has been written by **guest author Gary Seston**, an experienced and enthusiastic teacher of IB Diploma biology, who, importantly, also has examiner experience. Prior to his present post at Sotogrande International School, Cadiz, Gary taught at the United World College of South East Asia in Singapore. This chapter is an excellent guide of interest to both teachers and students, though possibly at different stages of their experience of the course.

Author's acknowledgements

I am indebted to IB teachers who have welcomed me into their Departments, and updated me on the delivery of the Diploma programme, including:

- Chris Hall, Head of Science, Malvern College, Worcester, UK
- Carolyn Halliday, Head of Biology, and Keith Allen, St Clare's, Oxford, UK
- John Cherverton, Head of Biology, Oakham School, Rutland, UK.

These colleagues also read early drafts of chapters of the manuscript, and their advice guided me to the best approach, following helpful discussion with Katie Mackenzie Stuart, Science Publisher, of Hodder Murray.

To Gary Seston, of Sotogrande International School, Cadiz, I owe a special debt. His perceptive comments on the appropriateness of the presentation to students for whom English is a second language, together with his advice on organisms familiar to students across other continents, and especially his observations on content and approach to syllabus issues of all chapters, were invaluable.

On particular aspects of the IB Diploma biology syllabus, I had the advantage of insights from experts:

- Dr Robin Cook, The Marine Laboratory, Aberdeen, Scotland, concerning the conservation of marine fish stocks
- Professor Robert Turner, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College, London, concerning fMRI scanning of the brain
- Dr Richard Johnson, Aberdeen University, Scotland, concerning electron microscopy
- Dr Kevin McKenzie, School of Medical Science, Aberdeen University, concerning electron microscopy
- Dr Neil Millar, Head of Biology, Heckmondwike Grammar School, Kirklees, West Yorkshire, UK, for permission to introduce Merlin, his unique statistical package for biology students, which is made freely available for educational and non-profit use.

Finally, I am indebted to the Science publishing team of Phillipa Allum and Helen Townson at Hodder Murray, and freelances Penelope Lyons and Gina Walker, all of whose skill and patience brought together text and illustration as I have wished, and I am most grateful to them.

Dr Chris Clegg
Salisbury
Wiltshire
UK
May 2007

Matching chapters with the IB Diploma biology syllabus

Colour key for syllabus levels

Standard and higher levels

Higher level only

Standard level only

Chapter	IB Diploma biology syllabus topic
Chapter 1: Cells – the building blocks	Topic 2: Cells 2.1 Cell theory 2.2 Prokaryotic cells 2.3 Eukaryotic cells 2.4 Membranes 2.5 Cell division
Chapter 2: The chemistry of life	Topic 3: The chemistry of life (part a) 3.1 Chemical elements and water 3.2 Carbohydrates, lipids and proteins 3.3 DNA structure 3.4 DNA replication 3.5 Transcription and translation 3.6 Enzymes
Chapter 3: Energy transfer in cells	Topic 3: The chemistry of life (part b) 3.7 Cell respiration 3.8 Photosynthesis
Chapter 4: Genetics	Topic 4: Genetics (part a) 4.1 Chromosomes, genes, alleles and mutations 4.2 Meiosis 4.3 Theoretic genetics
Chapter 5: Genetic engineering and biotechnology	Topic 4: Genetics (part b) 4.4 Genetic engineering and biotechnology
Chapter 6: Ecology, evolution and biodiversity	Topic 5: Ecology and evolution 5.1 Communities and ecosystems 5.2 The greenhouse effect 5.3 Populations 5.4 Evolution 5.5 Classification
Chapter 7: Human physiology, health and reproduction	Topic 6: Human health and physiology 6.1 Digestion 6.2 The transport system 6.3 Defence against infectious disease 6.4 Gas exchange 6.5 Nerves, hormones and homeostasis 6.6 Reproduction
Chapter 8: Nucleic acids and proteins	Topic 7: Nucleic acids and proteins 7.1 DNA structure 7.2 DNA replication 7.3 Transcription 7.4 Translation 7.5 Proteins 7.6 Enzymes
Chapter 9: Energy transfer in cells II	Topic 8: Cell respiration and photosynthesis 8.1 Cell respiration 8.2 Photosynthesis
Chapter 10: Plant science	Topic 9: Plant science 9.1 Plant structure and growth 9.2 Transport in angiospermophytes 9.3 Reproduction in angiospermophytes

1 Cells – the building blocks

STARTING POINTS

- The **cell** is the basic unit of life; living things are made of cells.
- Cells are small, too small to be seen with the naked eye. They must be viewed using a **microscope**.
- Plant and animal cells have common features, including the **nucleus**, **cytoplasm** and cell membrane (**plasma membrane**).
- Many living things consist of a single cell (**unicellular organisms**), others are built of many cells (**multicellular organisms**).
- In most multicellular organisms, the cells perform specialised functions.

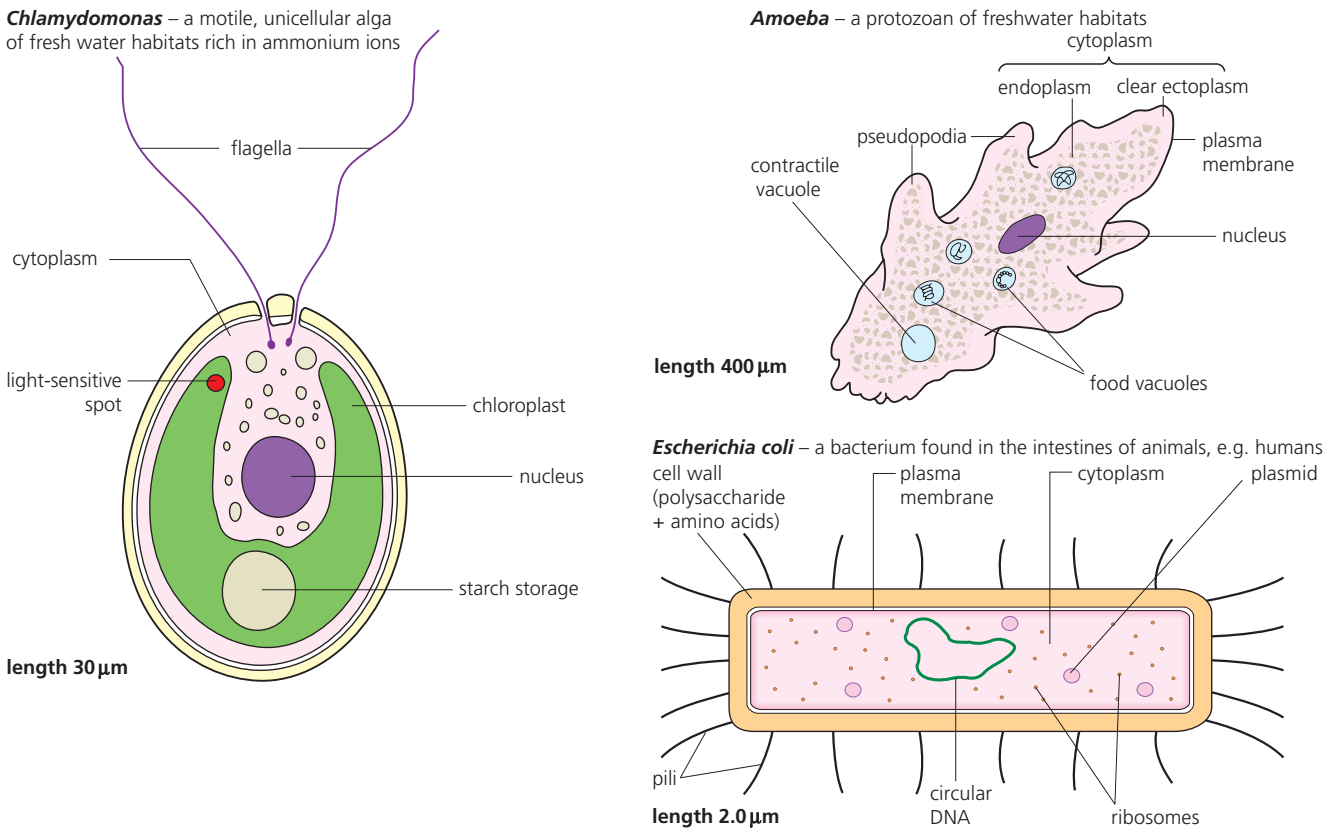
The cell is the basic unit of living matter – the smallest part of an organism which we can say is alive. It is cells that carry out the essential processes of life. We think of them as self-contained units of structure and function.

Introducing cells

2.1.1–2.1.10, 2.2.1–2.2.4, 2.3.1–2.3.6

Some organisms are made of a single cell, and are known as unicellular. Examples of **unicellular** organisms are introduced in Figure 1.1. In fact, there are vast numbers of different unicellular organisms in the living world, many with a very long evolutionary history.

Figure 1.1 Introducing unicellular organisation



Other organisms are made of many cells, and are known as **multicellular** organisms. Examples of multicellular organisms are the mammals and flowering plants. Much of the biology in this book is about multicellular organisms, including humans, and the processes that go on in these organisms. But remember, single-celled organisms carry out all the essential functions of life too, occurring within the confines of a single cell.

1 **State** the essential processes characteristic of living things.

Cell size

Cells are extremely small – most are only visible as distinct structures when we use a **microscope** (although a few types of cell are just large enough to be seen by the naked eye).

Observations of cells were first reported over 300 years ago, following the early development of microscopes (Figure 1.2). Today we use a compound light microscope to investigate cell structure – perhaps you are already familiar with the light microscope as a piece of laboratory equipment. You may have used one to view living cells, such as the single-celled animal, *Amoeba*, shown in Figure 1.1.

Since cells are so small, we need appropriate units to measure them. The **metre** (symbol **m**) is the standard unit of length used in science (it is an internationally agreed unit, or **SI unit**). Look at Table 1.1 below, showing the subdivisions of the metre used to measure cells and their contents. These units are listed in descending order of size. You will see that each subdivision is one thousandth of the unit above it. The smallest units are probably quite new to you; they may take some getting used to.

1 metre (m)	= 1000 millimetres (mm)
1 mm	= 1000 micrometres (μm) (or microns)
1 μm	= 1000 nanometres (nm)

Table 1.1 Units of length used in microscopy

So, the dimensions of cells are expressed in the unit called a **micrometre** or micron (**μm**). Notice this unit is one thousandth (10^{-3}) of a millimetre. This gives us a clear idea about how small cells are when compared to the millimetre, which you can see on a standard ruler.

2 Calculate:
- **a** how many cells of 100 μm diameter will fit side by side along a millimetre
- **b** the magnification of the image of *Escherichia coli* in Figure 1.1.

Bacteria are really small, typically 0.5–10 μm in size, whereas the cells of plants and animals are often in the range 50–150 μm or larger. In fact, the lengths of the unicells shown in Figure 1.1 are approximately:

- *Chlamydomonas* 30 μm
- *Escherichia coli* 2 μm
- *Amoeba* 400 μm (but its shape, and therefore length, varies greatly).

The features of cells

A cell consists of a **nucleus** surrounded by cytoplasm, contained within the cell membrane. The nucleus is the structure that controls and directs the activities of the cell. The **cytoplasm** is the site of the chemical reactions of life, which we call 'metabolism'. The cell membrane, known as the **plasma membrane**, is the barrier controlling entry to and exit from the cytoplasm.

Newly formed cells grow and enlarge. A growing cell can normally divide into two cells. Cell division is very often restricted to unspecialised cells, before they become modified for a particular task.

Cells may develop and specialise in their structure and in the functions they carry out. A common outcome of this is that many fully specialised cells are no longer able to divide, for example. But as a consequence of specialisation, **cells show great variety in shape and structure**. This variety in structure reflects the evolutionary adaptations of cells to different environments, and to different specialised functions – for example, within multicellular organisms.

Cell theory – a summary statement

The **cell theory** – the statement that cells are the unit of structure and function in living things – contains three very basic ideas:

- cells are the building blocks of structure in living things;
- cells are the smallest unit of life;
- cells are derived from other cells (pre-existing cells) by division.

Today we can confidently add two concepts:

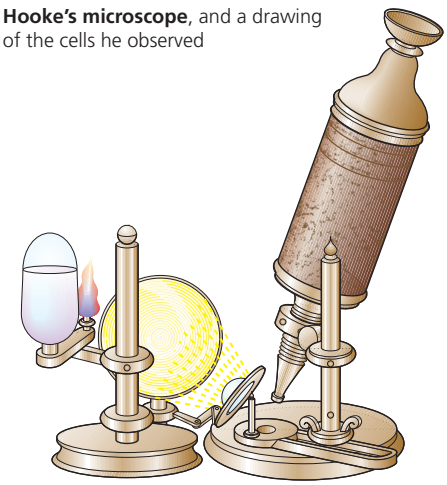
- cells contain a blueprint (information) for their growth, development and behaviour;
- cells are the site of all the chemical reactions of life (metabolism).

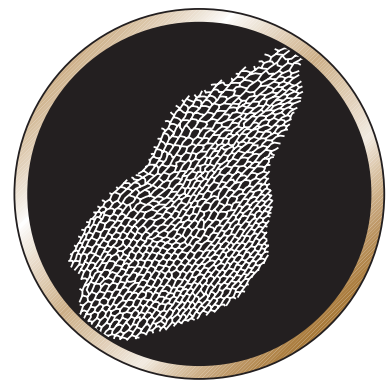
The origins of cell theory

Many biologists contributed to the development of the cell theory. This concept evolved gradually in western Europe during the nineteenth century as a result of a steadily accelerating pace of developments in **microscopy** and **biochemistry**. You can see a summary of the earliest steps in Figure 1.2.

Figure 1.2 Early steps in the development of the cell theory

Hooke's microscope, and a drawing of the cells he observed





Robert Hooke (1662), an expert mechanic and one of the founders of the Royal Society in London, was fascinated by microscopy. He devised a compound microscope, and used it to observe the structure of cork. He described and drew cork cells, and also measured them. He was the first to use the term 'cells'.

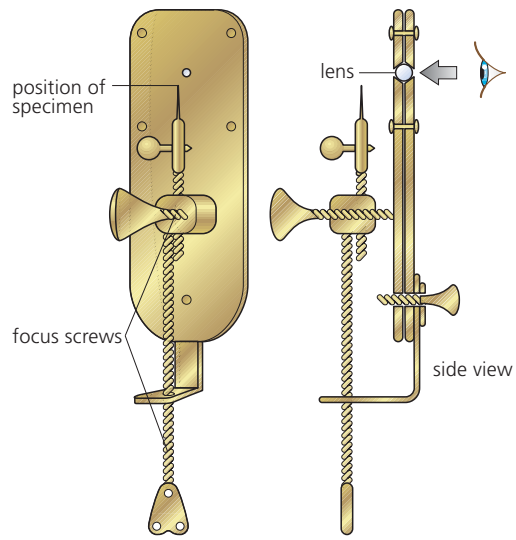
Anthony van Leeuwenhoek (1680) was born in Delft. Despite no formal training in science, he developed a hobby of making lenses, which he mounted in metal plates to form simple microscopes. Magnifications of ×240 were achieved, and he observed blood cells, sperms, protozoa with cilia, and even bacteria (among many other types of cells). His results were reported to the Royal Society, and he was elected a fellow.

Robert Brown (1831), a Scottish botanist, observed and named the cell nucleus. He also observed the random movements of tiny particles (pollen grains, in his case) when suspended in water (Brownian movement).

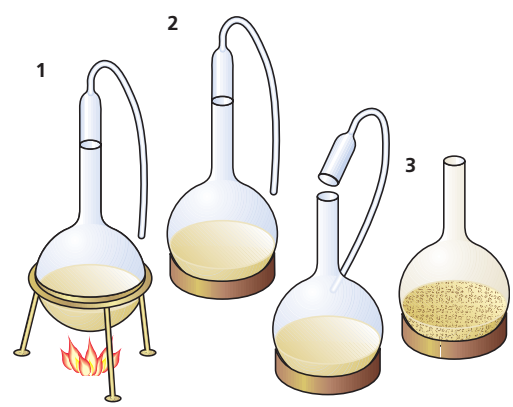
Matthias Schleiden (1838) and Theodor Schwann (1839), German biologists,established cells as the natural unit of form and function in living things: 'Cells are organisms, and entire animals and plants are aggregates of these organisms arranged to definite laws.'

Rudolf Virchow (1856), a German pathologist, established the idea that cells arise only by division of existing cells.

Louis Pasteur (1862), a brilliant French microbiologist, established that life does not spontaneously generate. The bacteria that 'appear' in broth are microbes freely circulating in the air, which contaminate exposed matter.



Leeuwenhoek's microscope



Pasteur's experiment, in which broth was sterilised (**1**), and then either exposed to air (**3**) or protected from air-borne spores in a swan-necked flask (**2**). Only the broth in **3** became contaminated with bacteria.

What evidence do we have in total, to support the cell theory?

Although cells were observed as long ago as the seventeenth century, it was not until the nineteenth century that biologists were confident enough to state what we now call the 'cell theory', categorically (Figure 1.2). Subsequent developments in many aspects of biological investigation have supported this concept. Cell theory has not been challenged, although it took some time for the true cause of apparent 'spontaneous generation of life' – the sudden appearance of living things in water that had previously seemed devoid of life, for example – to be understood.

Today, the study of the structure of cells, called **cytology**, is integrated within a larger, major branch of biology called **cell biology**. We know that cells contain the hereditary material **deoxyribonucleic acid (DNA)** in their nucleus (page 63). The structure and role of the DNA double helix in the control of metabolism and of growth via protein synthesis are understood. Now, developments in cell biology are closely linked to developments in genetics, biochemistry, biophysics, electron microscopy, biotechnology, enzymology, genetic engineering, and medicine – including oncology (the study of cancer), as we shall see. Table 1.2 is a summary of the evidence for the cell theory, and identifies where it is discussed further.

Concept	Evidence
living things are made of cells	**observations by microscopists** (light microscopy and electron microscopy) on the structure of unicellular and multicellular organisms, especially on the structure of tissues and organs (histology, page 7)
cells are the smallest units of life	**discovery of viruses** as particles that are 'crystalline' (non-cellular) when outside a host cell, and that can only reproduce themselves at the expense of their host cell's metabolic machinery (page 19)
	biochemical investigations of organelles (tiny substructures in cells, e.g. chloroplasts) showing their ability to function outside of a cell, under laboratory-controlled conditions, for a limited time (page 281)
cells come from pre-existing cells	**Pasteur's observations** on the origins of microbes in fermenter vessels (Figure 1.2), and related discoveries that cases of apparent 'spontaneous generations' of microorganisms in pond or puddle waters were due to the presence of (unnoticed) pre-existing cells
	discovery that the life cycles of many microorganisms include a resistant spore phase, and the linked discovery that spores of many microorganisms and unicellular protoctista are ubiquitous, but that they become active only in favourable growing conditions
	observations on the behaviour of cells at division (mitosis, page 31 and meiosis, page 94) and during reproduction (cytology)
cells contain a blueprint for growth, development and behaviour	**observations on the behaviour of chromosomes** and the establishment of the nature and role of genes / DNA in the day-to-day control of cells (page 66), and in the process of heredity (page 91)
	experimental evidence of the effects on cells of the deliberate transfer of genes between organisms (genetic engineering, page 117)
cells are the site of the chemical reactions of life	**discovery of enzymes** and the enzyme machinery of cellular processes such as cell aerobic respiration and fermentation (page 80)
	discovery of biochemical events in cells, such as the formation of proteins from amino acids (page 69)
	discovery of cell ultrastructure, of the presence of discrete organelles and of the biochemical events located in particular organelles (electron microscopy and biochemistry, page 14)

Table 1.2 The evidence for the cell theory – a review

Maintaining cell growth

The materials required for growth and maintenance of a cell enter through the outermost layer of the cytoplasm, a membrane called the **plasma membrane**. Similarly, waste products must leave the cell through the plasma membrane.

The rates at which materials can enter and leave a cell depend on the surface area of that cell, but the rates at which materials are used and waste products are produced depend upon the amount of cytoplasm present within the cell. Similarly, heat transfer between the cytoplasm and environment of the cell is determined by surface area.

Cell surface:volume ratios and cell size

As the cell grows and increases in size, an important difference develops between the surface area available for exchange and the volume of the cytoplasm in which the chemical reactions of life occur. The volume increases faster than the surface area; the **surface area:volume ratio** falls (Figure 1.3). So, with increasing size of a cell, less and less of the cytoplasm has access to the cell surface for exchange of gases, supply of nutrients, and loss of waste products.

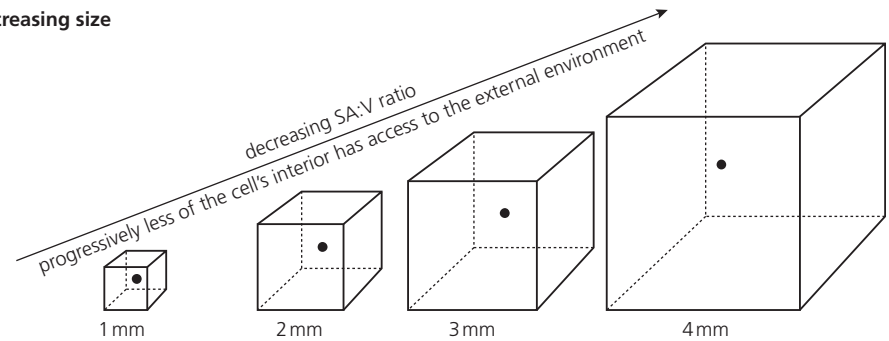


Figure 1.3 The effect of increasing size on the surface area:volume ratio

dimensions/mm	1 × 1 × 1	2 × 2 × 2	3 × 3 × 3	4 × 4 × 4
surface area/mm^2	6	24	54	96
volume/mm^3	1	8	27	64
surface area: volume ratio	6:1 = $^6/_1$ = 6	24:8 = $^{24}/_8$ = 3	54:27 = $^{54}/_{27}$ = 2	96:64 = $^{96}/_{64}$ = 1.5

3 For imaginary cubic 'cells' with sides 1, 2, 4 and 6 mm:

- **a** **calculate** the volume, surface area and ratio of surface area to volume for each
- **b** **state** the effect on the SA:V ratio of a cell as it increases in size
- **c** **explain** the effect of increasing cell size on the efficiency of diffusion in the removal of waste products from cell cytoplasm.

Put another way, we can say that the smaller the cell is, the more quickly and easily can materials be exchanged between its cytoplasm and environment. One consequence of this is that cells cannot continue growing larger, indefinitely. When a maximum size is reached, cell growth stops. The cell may then divide. The process of cell division is discussed later (page 30).

Metabolism and cell size

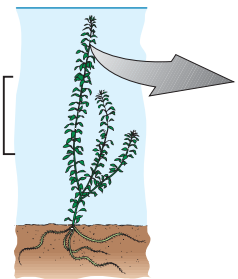
The extent of chemical reactions that make up metabolism in a cell (the subject of later chapters) is not directly related to the surface area of the cell, but it does relate to the amount of cytoplasm, expressed as the cell mass. In summary, we can say that the rate of metabolism of a cell is a function of its mass, whereas the rate of exchange of materials and heat energy that metabolism generates is a function of the cell's surface area.

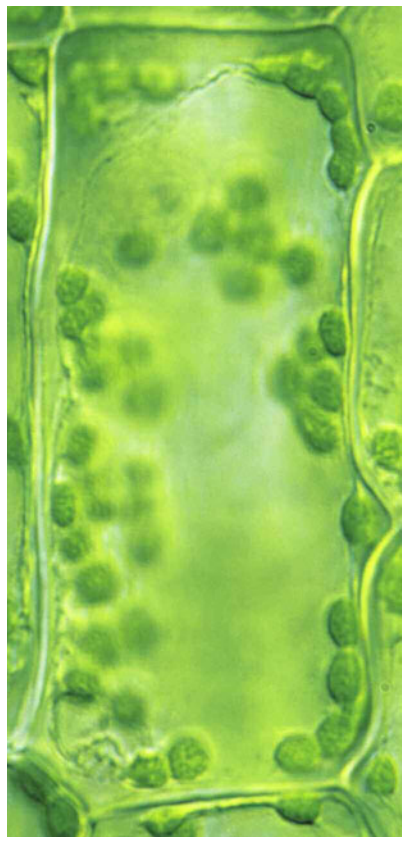
Introducing animal and plant cells

No 'typical' cell exists – there is a very great deal of variety among cells. However, we shall see that most cells have features in common. Viewed using a compound microscope, the initial appearance of a cell is of a simple sac of fluid material, bound by a membrane, and containing a nucleus. Look at the cells in Figure 1.4.

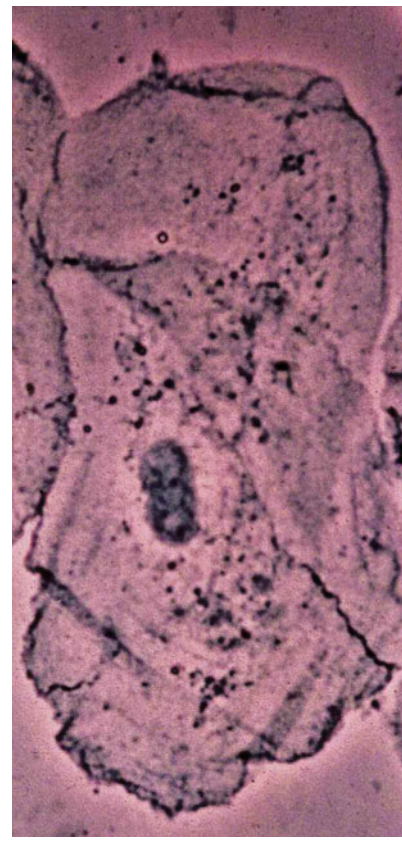
Figure 1.4 Animal and plant cells from multicellular organisms

Canadian pondweed (*Elodea*) grows submerged in fresh water



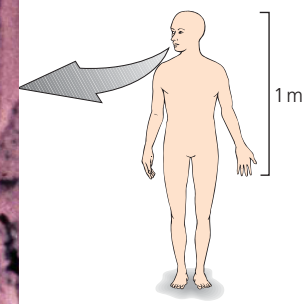


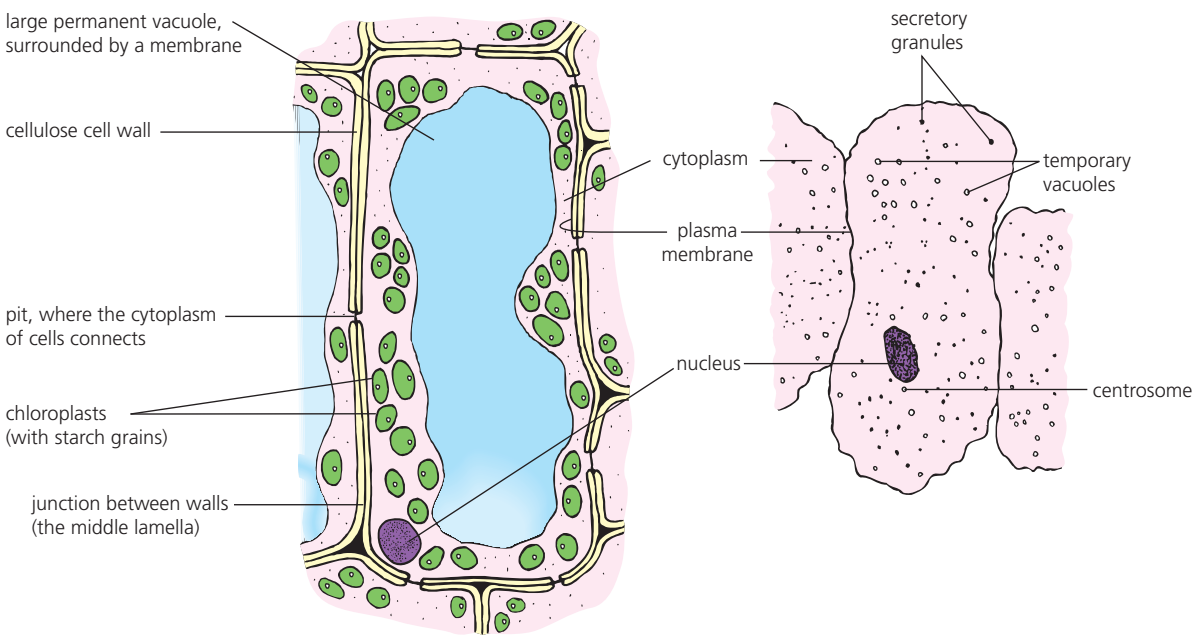
photomicrograph of a leaf cell of *Elodea* (×400)



photomicrograph of a human cheek cell (×800)

human





Animal and plant cells have at least three structures in common. These are their **cytoplasm** with its **nucleus**, surrounded by a **plasma membrane**. In addition, there are many tiny structures in the cytoplasm, called **organelles**, most of them common to both animal and plant cells. An organelle is a discrete structure within a cell, having a specific function. Organelles are all too small to be seen at this magnification. We have learned about the structure of organelles using the electron microscope (page 14).

There are some important basic differences between plant and animal cells (Table 1.3). For example, there is a tough, slightly elastic **cell wall**, made largely of cellulose, present around plant cells (page 16). Cell walls are absent from animal cells.

A **vacuole** is a fluid-filled space within the cytoplasm, surrounded by a single membrane. Plant cells frequently have a large, permanent vacuole present. By contrast, animal cells may have small vacuoles, but these are mostly temporary.

Green plant cells also contain organelles called **chloroplasts** in their cytoplasm. These are not found in animal cells. The chloroplasts are the sites where green plant cells manufacture elaborated food molecules by a process known as photosynthesis.

The **centrosome**, an organelle that lies close to the nucleus in animal cells (Figure 1.4), is not present in plants. This tiny organelle is involved in nuclear division in animal cells.

Finally, the **storage carbohydrate** (energy store) differs, too. Animal cells may store glycogen (page 47); plant cells normally store starch.

Plant cells	Feature	Animal cells
cellulose cell wall present	**cell wall**	no cellulose cell walls
many cells contain chloroplasts; site of photosynthesis	**chloroplasts**	no chloroplasts; animal cells cannot photosynthesise
large, fluid-filled vacuole typically present	**permanent vacuole**	no large permanent vacuoles
no centrosome	**centrosome**	a centrosome present outside the nucleus
starch	**carbohydrate storage product**	glycogen

Table 1.3 Differences between plant and animal cells

Multicellular organisms – specialisation and division of labour

Unicellular organisms are structurally simple in that they perform all the functions and activities of life within a single cell. The cell feeds, respires, excretes, is sensitive to internal and external conditions (and may respond to them), may move, and eventually divides or reproduces. You have seen examples of unicellular organisation in Figure 1.1.

By contrast, the majority of multicellular organisms are like the mammals (and flowering plants) in that they are made of cells, most of which are highly **specialised** to perform particular functions. Specialised cells occur organised into **tissues** and **organs**. A tissue is a group of similar cells specialised to perform a particular function, such as heart muscle tissue of a mammal. An organ is a collection of different tissues which performs a specialised function, such as the heart of a mammal. So the tissues and organs of multicellular organisms consist of specialised cells.

Control of cell specialisation

We have noted that the nucleus of each cell is the structure that controls and directs the activities of the cell, and that the information required for this exists in the form of a nucleic acid, DNA. The nucleus of a cell contains the DNA in thread-like **chromosomes**, which are linear sequences of **genes** (page 92). Genes control the development of each cell within the mature organism. We can define a gene in different ways, including:

- a specific region of a chromosome which is capable of determining the development of a specific characteristic of an organism;
- a specific length of the DNA double helix (hundreds or thousands of base pairs long) which codes for a protein.

So, when a cell is becoming specialised – we say the cell is **differentiating** – some of its genes are being activated and expressed. These genes determine how the cell develops. What happens during gene expression, and the mechanism by which a cell's chemical reactions are controlled, are explored in the next chapter.

For the moment we can just note that the nucleus of each cell contains all the information required to make each type of cell present within the whole organism, only a selected part of which is needed in any one cell and tissue. The potential of each cell to specialise in any number of different ways is called **totipotency**. Which genes are activated and how a cell specialises are controlled by the immediate environment of the differentiating cell, and its position in the developing organism.

The cost of specialisation

Specialised cells are efficient at carrying out their particular function, such as transport, or support, or protection. We say the resulting differences between cells are due to **division of labour**. By specialisation, increased efficiency is achieved, but at a price. The specialised cells are now totally dependent on the activities of other cells. For example, in animals, nerve cells are adapted for the transport of nerve impulses, but dependent on blood cells for oxygen, and on heart muscle cells to pump the blood. This modification of cell structure to support differing functions is another reason why no 'typical cell' really exists.

Extension: Non-cellular organisation – an exceptional condition

In addition to the familiar unicellular and multicellular organisation of living things, there are a few examples of multinucleate organs and organisms, without divisions into separate cells. This type of organisation is called acellular. An example of an **acellular organism** is the pin mould *Rhizopus*, where the 'plant' body consists of fine, thread-like structures called hyphae. An example of an **acellular organ** is the striped muscle fibres that make up the skeletal muscles of mammals (Figure 1.5).

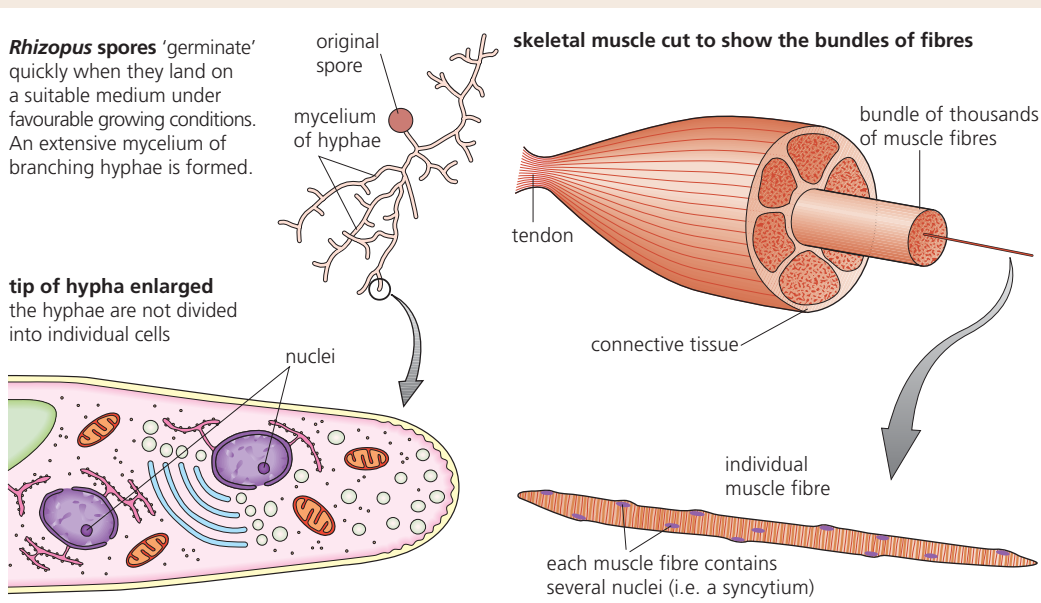


Figure 1.5 Acellular organisation in *Rhizopus* and in skeletal muscle fibres

Emergent properties of the multicellular organism

It has been said that a multicellular organism has properties that in total are greater than the sum of the individual parts – that is, of all the cells, tissues and organs of which it is made up. Such properties are described as **emergent**.

4 **Suggest** emergent properties of a human, as an example of a multicellular organism.

Examining cells, and recording structure and size

We use microscopes to magnify the cells of biological specimens in order to view them. Figure 1.6 shows two types of light microscope.

In the simple microscope (**hand lens**), a single biconvex lens is held in a supporting frame so that the instrument can be held very close to the eye. Today, a hand lens is mostly used to observe external structure, although some of the earliest detailed observations of living cells were made with single-lens instruments.

In the **compound microscope**, light rays are focused by the **condenser** on to a specimen on a microscope slide on the stage of the microscope. Light transmitted through the specimen is then focused by two sets of lenses (hence the name compound microscope). The **objective lens** forms an image (in the microscope tube) which is then further magnified by the **eyepiece lens**, producing a greatly enlarged image.

Biological material to be examined by compound microscopy must be sufficiently transparent for light rays to pass through. When bulky tissues and parts of organs are to be examined, thin sections are cut. Thin sections are largely colourless.

TOK Link

Researching the development of the microscope and the early history of The Royal Society may support your study of the methods of science. If so, the following websites provide useful links:

- The Royal Society
  www.royalsoc.ac.uk
- Royal Microscopical Society
  www.rms.org.uk

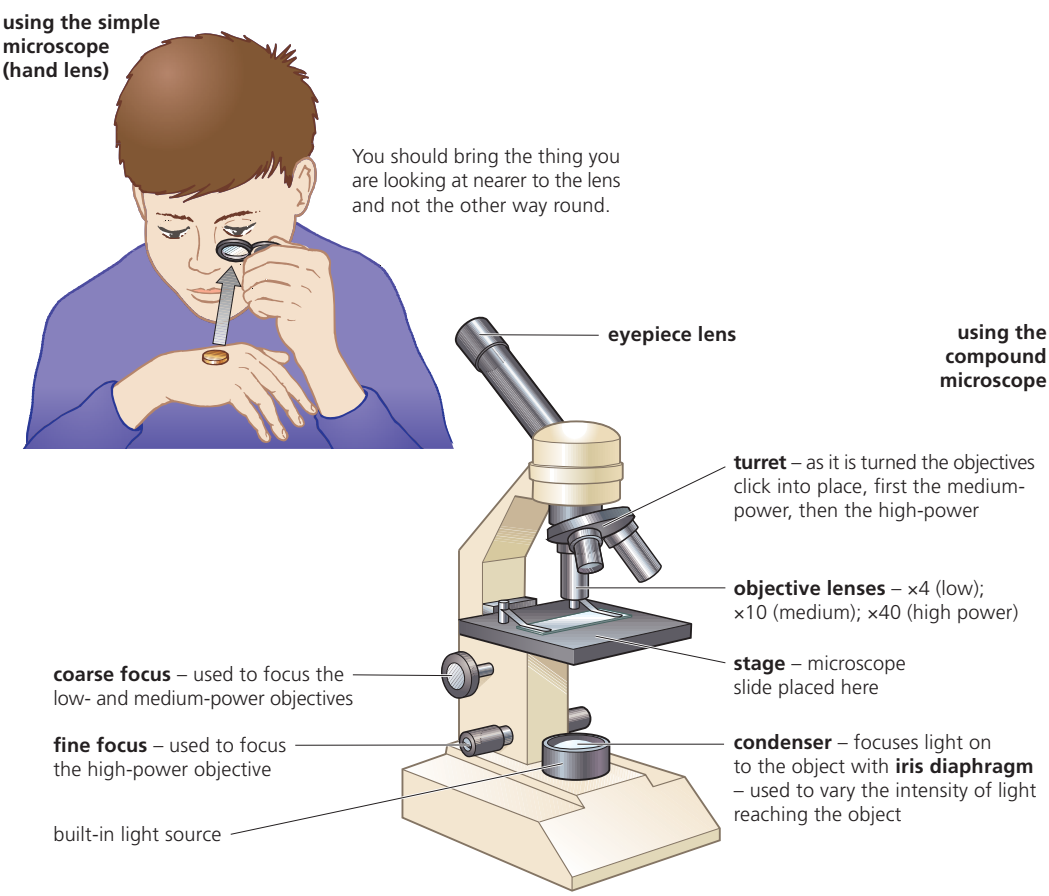


Figure 1.6 Light microscopy

Recording observations

Images of cells and tissues viewed may be further magnified, displayed or projected and saved for printing out by the technique of **digital microscopy** (Figure 1.7). A digital microscope is used, or alternatively an appropriate video camera is connected by microscope coupler or eyepiece adaptor that replaces the standard microscope eyepiece. Images are displayed via video recorder, TV monitor, or computer, and may be printed out by the latter.

Alternatively, a record of what you see via the compound microscope may be recorded by drawings of various types (Figure 1.8). For a clear, simple drawing:

- use a sharp HB pencil and a clean eraser;
- use unlined paper and a separate sheet for each specimen you record;
- draw clear, sharp outlines, avoiding shading or colouring (density of structures may be represented by degrees of stippling);
- label each sheet or drawing with the species, conditions (living or stained; if stained, which stains), TS or LS, and so forth;
- label your drawing fully, with labels well clear of the structures shown, remembering that label lines should not cross;
- annotate (add notes about function, role, development) if appropriate;
- include a statement of the magnification under which the specimen has been observed.

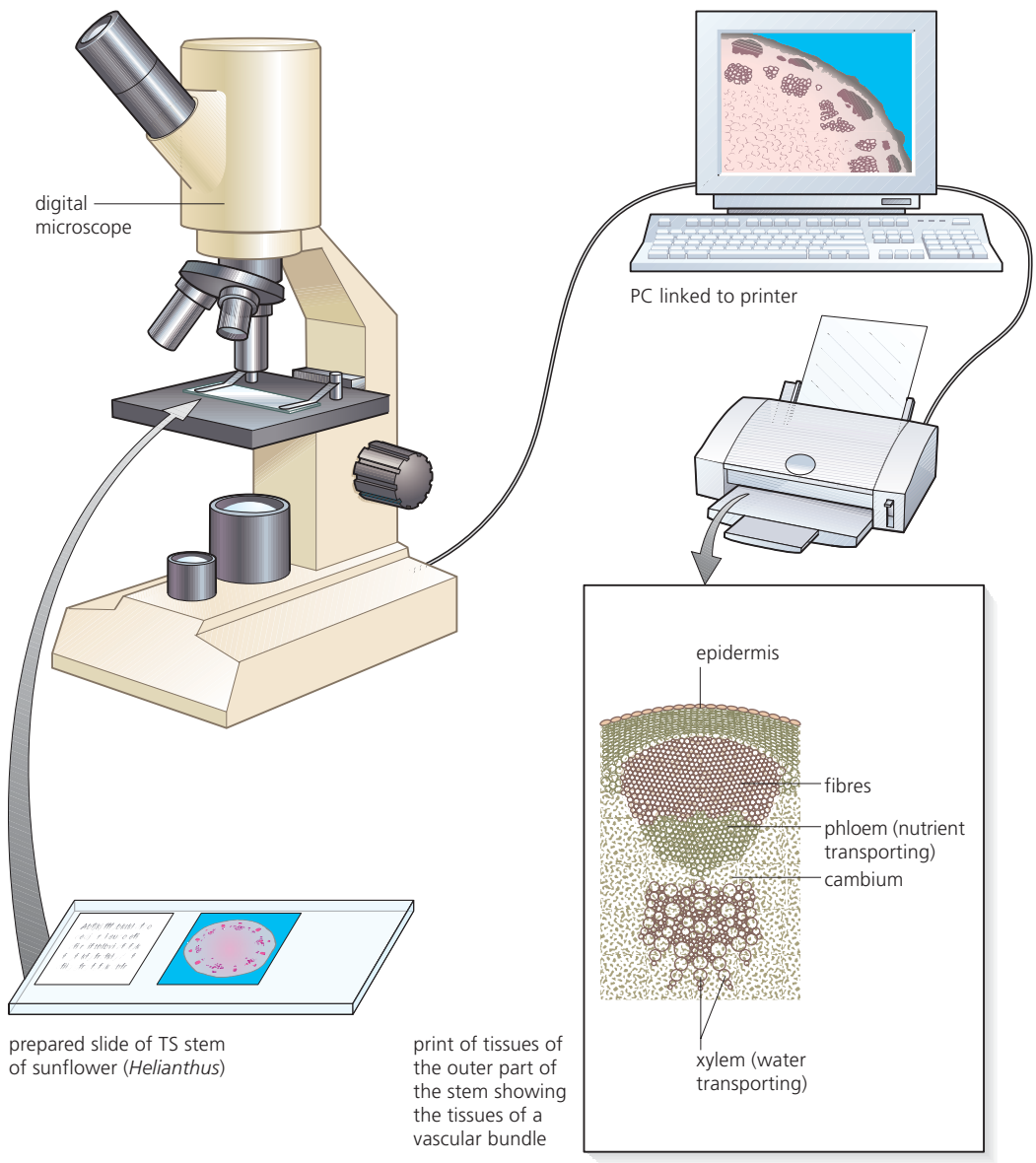
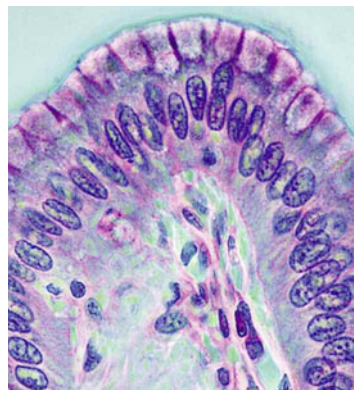


Figure 1.7 Digital microscopy in action.

Figure 1.8 Recording cell structure by drawing



view (phase contrast) of the layer of cells (epithelium) lining the stomach wall

The lining of the stomach consists of columnar epithelium. All cells secrete mucus copiously.

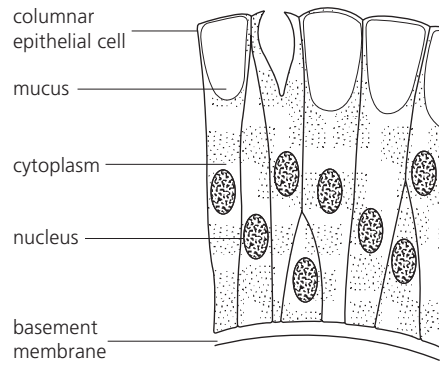
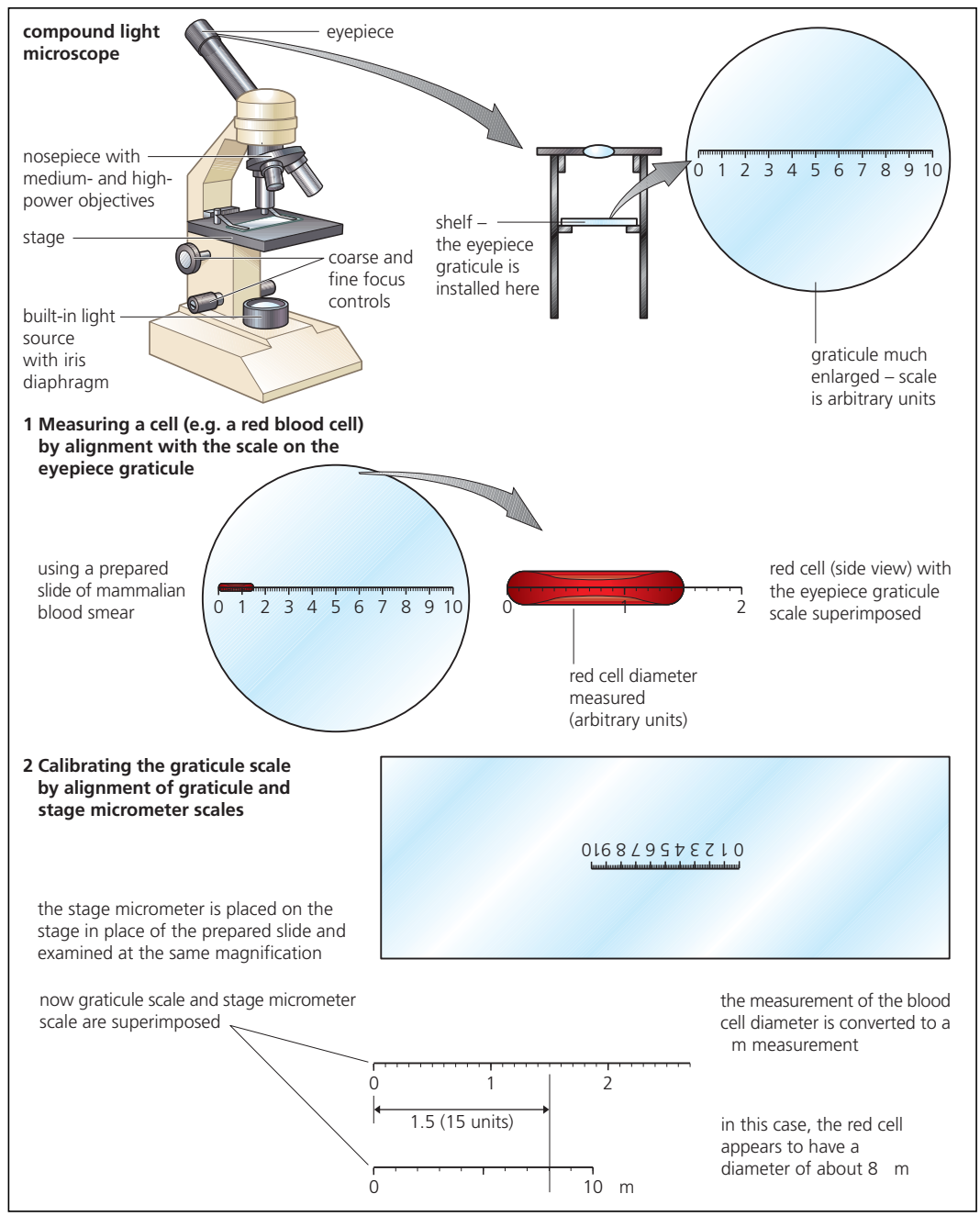


Figure 1.9 Measuring the size of cells

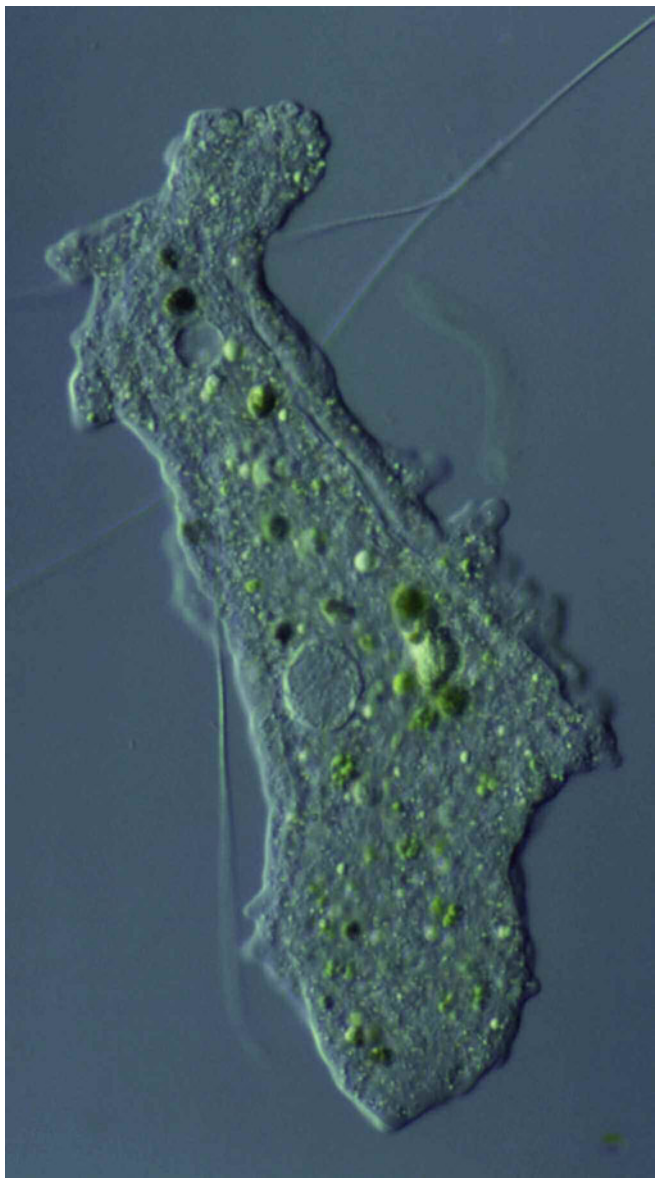


Measuring microscopic objects

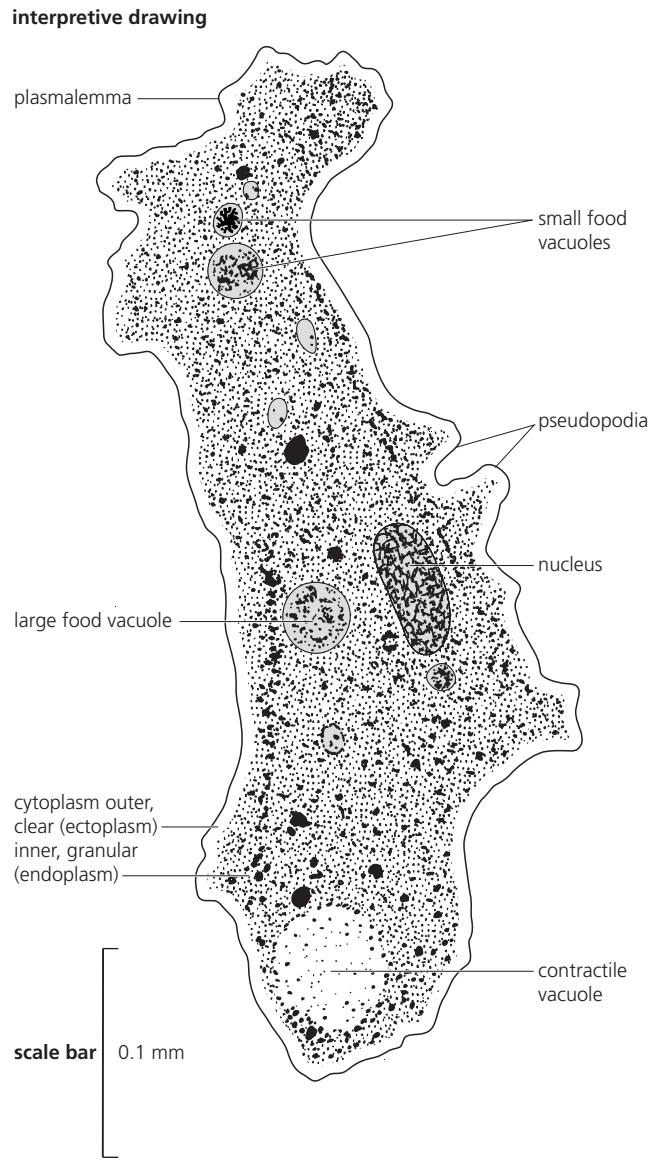
The size of a cell can be measured under the microscope. A transparent scale called a **graticule** is mounted in the eyepiece at the focal plane (there is a ledge for it to rest on). In this position, when the object under observation is in focus, so too is the scale. The size (e.g. length, diameter) of the object may then be recorded in arbitrary units. Next, the graticule scale is calibrated using a **stage micrometer** – in effect, a tiny, transparent ruler, which is placed on the microscope stage in place of the slide and then observed. With the eyepiece and stage micrometer scales superimposed, the true dimensions of the object can be estimated in microns. Figure 1.9 shows how this is done.

Once the size of a cell has been measured, a scale bar line may be added to a micrograph or drawing to record the actual size of the structure, as illustrated in the photomicrograph in Figure 1.10.

Figure 1.10 Recording size by means of scale bars



photomicrograph of *Amoeba proteus* (living specimen) – phase contrast microscopy



5 Using the scale bar given in Figure 1.10, **calculate** the maximum observed length of the *Amoeba* cell.

Magnification and resolution of an image

Magnification is the number of times larger an image is than the specimen. The magnification obtained with a compound microscope depends on which of the lenses you use. For example, using a ×10 eyepiece and a ×10 objective lens (medium power), the image is magnified ×100 (10 × 10). When you switch to the ×40 objective (high power) with the same eyepiece lens, then the magnification becomes ×400 (10 × 40). These are the most likely orders of magnification used in your laboratory work.

Actually, there is **no limit to magnification**. For example, if a magnified image is photographed, then further enlargement can be made photographically. This is what may happen with photomicrographs shown in books and articles. Magnification is given by the formula:

$$\text{magnification} = \frac{\text{size of image}}{\text{size of specimen}}$$

So, for a particular plant cell of 150 μm diameter, photographed with the microscope and with the image enlarged photographically, the magnification in a print showing the cell at 15 cm diameter (150 000 μm) is:

$$\frac{150\,000}{150} = 1000$$

If a further enlargement is made, to show the same cell at 30 cm diameter (300 000 μm), then the magnification is:

$$\frac{300\,000}{150} = 2000$$

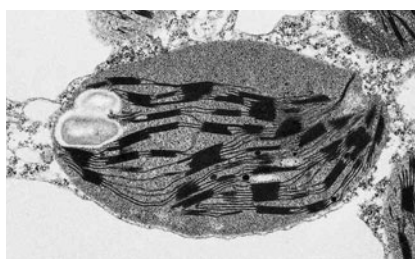
6 **Calculate** what magnification occurs with a ×6 eyepiece and a ×10 objective.

In this case, the image size has been doubled, but **the detail will be no greater**. You will not be able to see, for example, details of cell membrane structure, however much the image is enlarged. This is because the layers making up a cell's membrane are too thin to be seen as separate structures using the light microscope (Figure 1.11).

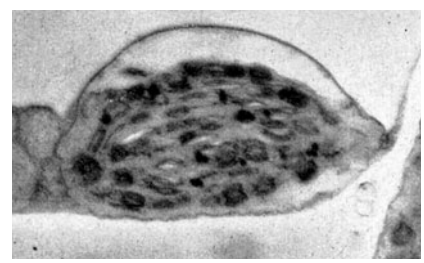
Figure 1.11 Magnification without resolution

chloroplast enlarged (×6000)

a) from a transmission electron micrograph



b) from a photomicrograph obtained by light microscopy



The **resolution** (resolving power) of the microscope is its ability to separate small objects which are very close together. If two separate objects cannot be resolved they will be seen as one object. Merely enlarging them will not separate them. Resolution is a property of lenses quite different from their magnification – and is more important.

Resolution is determined by the wavelength of light. Light is composed of relatively long wavelengths, whereas shorter wavelengths give the better resolution. For the light microscope, the limit of resolution is about 0.2 μm. This means two objects less than 0.2 μm apart may be seen as one object.

Electron microscopy

In the **electron microscope**, an electron beam is used to make a magnified image. Because an electron beam has a much shorter wavelength than light, resolution is much greater in the electron microscope than it is in the light microscope. For the electron microscope used with biological materials, the limit of resolution is about 5 nm (the size of nanometres is given in Table 1.1, page 2).

In **transmission electron microscopy** (**TEM**), the electron beam is passed through an extremely thin section of material. Membranes and other structures present are stained with heavy metal ions, making them electron-opaque so they stand out as dark areas in the image. We cannot see electrons, so the electron beam is focused onto a fluorescent screen for viewing, or onto a photographic plate for permanent recording (Figure 1.12).

In **scanning electron microscopy** (**SEM**) a narrow electron beam is scanned back and forth across the surface of the specimen. Electrons that are reflected or emitted from this surface are detected and converted into a three-dimensional image (Figures 10.17, 12.1).

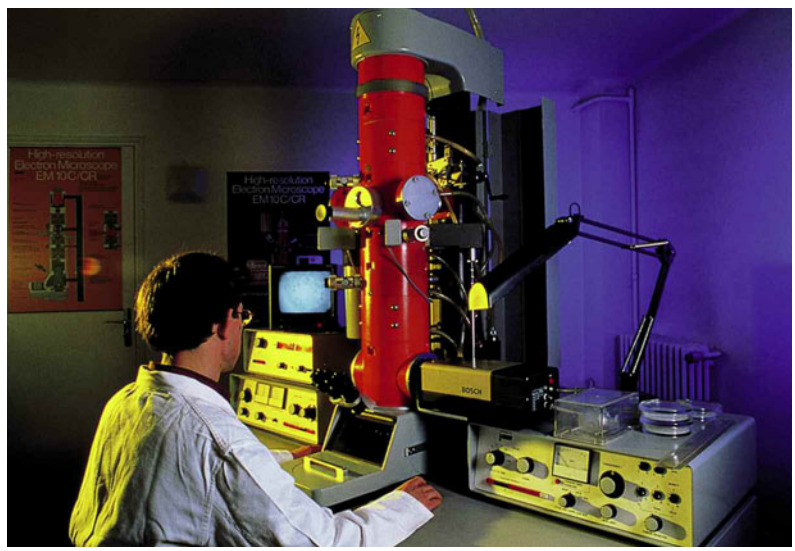


Figure 1.12 A transmission electron microscope in use

The impact of electron microscopy on cell biology

7 **Distinguish** between resolution and magnification.

The nucleus is the largest substructure (organelle) of a cell, and may be observed with the light microscope. Most organelles cannot be viewed by light microscopy, and none is large enough for internal details to be seen. It is the electron microscope that we use to learn about the fine details of cell structure (Figure 1.13).

Prokaryotic and eukaryotic organisation

Living things have traditionally been divided into two major groupings: animals and plants. However, the range of biological organisation is more diverse than this. The use of the electron microscope in biology has led to the discovery of two types of cellular organisation, based on the presence or absence of a nucleus.

Cells of plants, animals, fungi and protoctista have cells with a large, obvious nucleus. The surrounding cytoplasm contains many different membranous organelles. These types of cells are called **eukaryotic cells** (meaning a 'good nucleus').

On the other hand, bacteria contain no true nucleus and their cytoplasm does not have the organelles of eukaryotes. These are called **prokaryotic cells** (meaning 'before the nucleus').

This distinction between prokaryotic and eukaryotic cells is a fundamental division and is more significant than the differences between plants and animals. We will shortly return to examine the detailed structure of the prokaryotic cell, choosing a bacterium as our example (page 16). First, we need to look into the main organelles in eukaryotic cells.

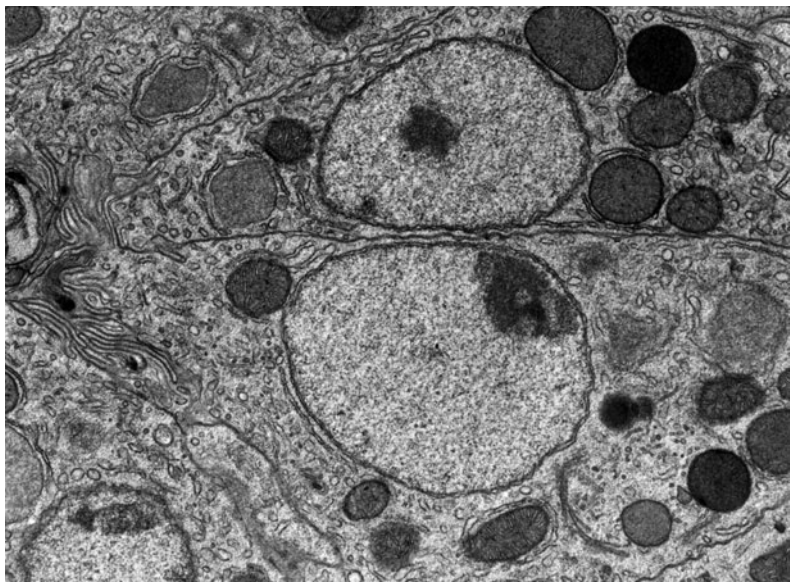
The ultrastructure of the eukaryotic cell

Our knowledge of the fine structure (known as **ultrastructure**) of cells has been built up by the examination of numerous TEMs. In effect, the eukaryotic cell is a bag of organelles suspended in a fluid matrix, contained within a special membrane, the plasma membrane.

The ultrastructure of a mammalian liver cell is shown in Figure 1.13. The functions of each of the six types of organelle shown (**nucleus**, **mitochondria**, **ribosomes**, **rough endoplasmic reticulum** (**rER**), **lysosomes**, and **Golgi apparatus**) are included on their labels.

Figure 1.13
Ultrastructure of a eukaryotic animal cell

TEM of liver cells

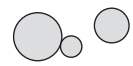


mitochondria – site of pathway of aerobic respiration and ATP formation

ribosomes – site of synthesis of proteins that remain in the cell

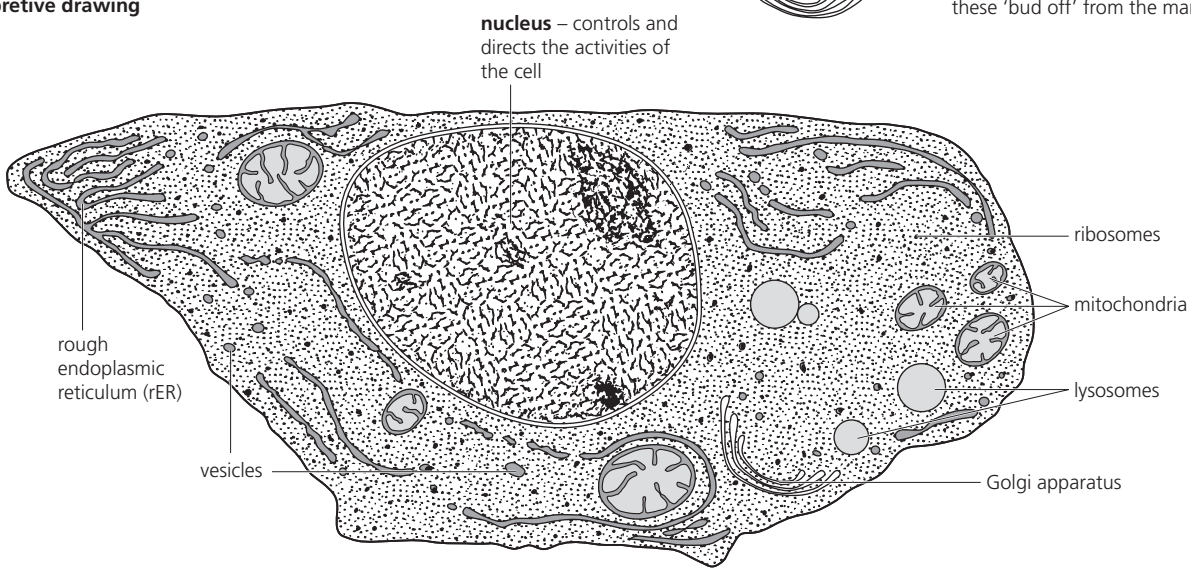
rough endoplamic reticulum (rER) – site of synthesis of proteins that will be exported from cells

lysosomes – membrane-bound vesicles containing enzymes



Golgi apparatus – site of synthesis of chemicals required by the cell, which are packaged into vesicles before these 'bud off' from the margins

interpretive drawing



Cells may have extracellular components

We have noted that the contents of cells are contained within the plasma membrane. However, cells may secrete material outside the plasma membrane; for example, plant cells have an external wall, and many animal cells secrete glycoproteins.

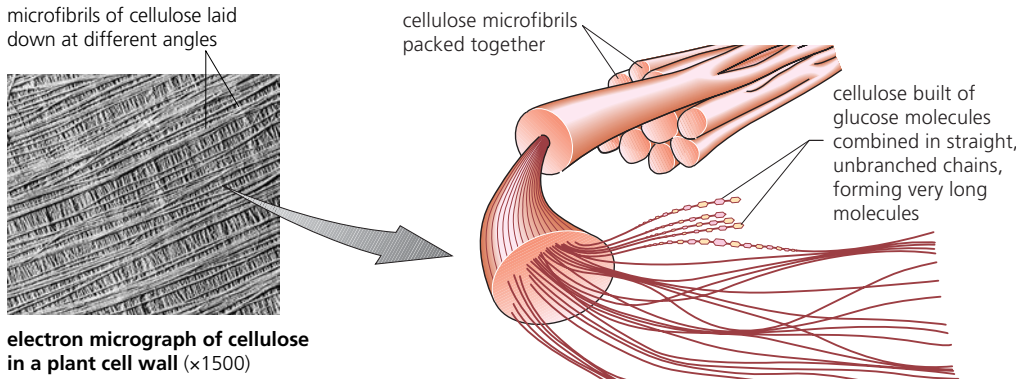
The plant cell and its wall

The plant cell differs from an animal cell in that it is surrounded by a wall. This wall is completely external to the cell. The plant cell wall is not an organelle. Plant cell walls are primarily constructed of cellulose – a polysaccharide and an extremely strong material. Cellulose molecules are very long, and are arranged in bundles called microfibrils (Figure 1.14).

Cell walls make the boundaries of plant cells easy to see when plant tissues are examined by microscopy. The presence of this strong structure allows the plant cell to develop high internal pressure due to water uptake, without danger of the cell bursting. This is a major difference between the cell water relations of plants and animals.

8 **Outline** how the electron microscope has increased our knowledge of cell structure.

Figure 1.14 The cellulose of a plant cell wall



electron micrograph of cellulose in a plant cell wall (×1500)

The effect of hydrated cell contents pressing against a cell wall (which is distinctly non-elastic) creates an internal pressure which supports the herbaceous (non-woody) plant, its stem, leaves and roots. For example, try cutting a 1 cm × 1 cm × 10 cm potato 'chip' from a potato tuber, and standing it in tap water for an hour or two. Then remove and gently bend it. The now turgid tissue may audibly snap, so stretched are the potato cells!

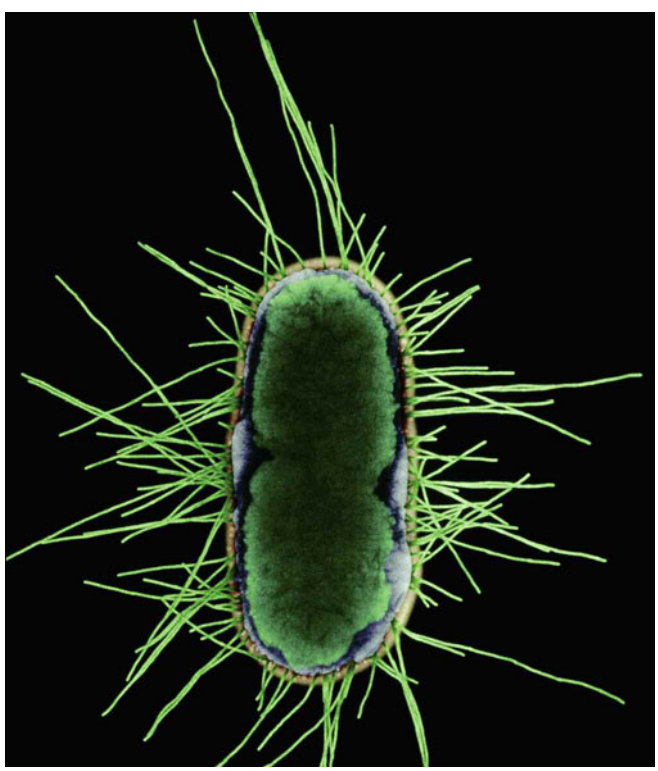
Extracellular glycoproteins around animal cells

Many animal cells have the ability to adhere to each other. We see the importance of this property in the ways in which cells may form compact tissues and organs. Other animal cells occur in simple sheets or layers, attached to a basement membrane below them. These cases of adhesion are brought about by glycoproteins that the cells have secreted. Glycoproteins are large molecules of protein to which quite large sugar molecules (called oligosaccharides) are attached. Glycoproteins play a part in the support and movement of cells.

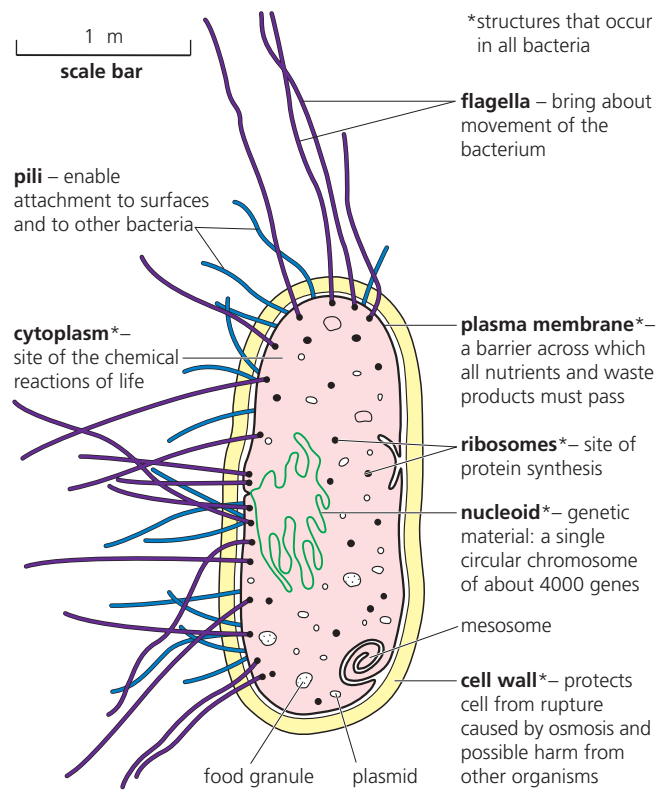
The structure of the prokaryotic cell

Figure 1.15 The structure of *Escherichia coli*

Figure 1.15 shows the structure of a prokaryotic cell called *Escherichia coli*. The bacteria and cyanobacteria are prokaryotes.



electron micrograph of *Escherichia coli*



E. coli is a common bacterium of the human gut – it occurs in huge numbers in the lower intestine of humans and other endothermic (once known as 'warm-blooded') vertebrates, such as the mammals, and it is a major component of the faeces of these animals. This tiny organism was named by a bacteriologist, Professor T. Escherich, in 1885. Notice the scale bar in Figure 1.15. This bacterium is typically about 1 × 3 μm in length – about the size of a mitochondrion in a eukaryotic cell. The cytoplasm lacks the range of organelles found in eukaryotic cells and there is no large nucleus surrounded by a double membrane.

In Figure 1.15, the ultrastructure of *E. coli* is shown. The functions of each of the structures (cell wall, plasma membrane, cytoplasm, **pili**, **flagella**, **ribosomes** and **nucleoid**) are included on their labels.

9 **Calculate** the approximate magnification of the image of *E. coli* in Figure 1.15.

Finally, we can note that all prokaryote cells are capable of extremely rapid growth when conditions are favourable for them. In such environments, prokaryote cells frequently divide into two cells (known as 'binary fission'). New cells formed then grow to full size and divide again.

Prokaryotic and eukaryotic cells compared

By contrasting Figures 1.13 and 1.15 we can see that there are fundamental differences, both in cell size and cell complexity. In Table 1.4, prokaryotic and eukaryotic cells are compared.

Prokaryotes (e.g. bacteria, cyanobacteria)	Feature	Eukaryotes (e.g. animals, plants, fungi)
cells are extremely small, typically 5–10 μm	**size**	cells are larger, typically 50–150 μm
nucleus absent; circular strand of DNA helix in the cytoplasm, not supported by histone protein, and called a 'nucleoid'	**genetic material**	nucleus has distinct nuclear membrane (with pores), and chromosomes of linear DNA helix supported by histone protein
cell wall present (not of cellulose)	**cell wall**	cell wall present in plants and fungi
few organelles; membranous organelles absent or very simple	**organelles**	many organelles bounded by double membrane (e.g. mitochondria, nucleus) or single membrane (e.g. Golgi apparatus, lysosome, vacuole, rough endoplasmic reticulum)
proteins synthesised in small ribosomes (70S)	**protein synthesis**	proteins synthesised in large ribosomes (80S)
some cells have simple flagella, 20 nm in diameter	**motile organelles**	some cells have cilia or flagella with internal structures, 200 nm in diameter

Table 1.4 Prokaryotic and eukaryotic cells compared

Extension: A possible origin for mitochondria and chloroplasts

Present-day prokaryotes are similar to fossil prokaryotes, some of which are 3500 million years old. By comparison, the earliest eukaryote cells date back only 1000 million years. Thus eukaryotes must have evolved surrounded by prokaryotes that were long-established organisms. It is possible that, in the evolution of the eukaryotic cell, prokaryotic cells that were taken up into food vacuoles came to survive as organelles inside the 'host' cell, rather than being digested as food items. If so, they have become integrated into the biochemistry of their 'host' cell, with time.

If this hypothesis is correct, it would explain why mitochondria (and chloroplasts) contain a ring of DNA double helix, just like a bacterial cell. They also contain small ribosomes like those of prokaryotes. These features have caused some evolutionary biologists to suggest that these organelles may be descendants of free-living prokaryotic organisms that came to inhabit larger cells. It may seem a fanciful idea, but not impossible.

The life history of the cell and the nature of stem cells

Multicellular organisms begin life as a single cell, which grows and divides, forming very many cells, and these eventually form the adult organism. So cells arise by division of existing cells. The time between one cell division and the next is known as the **cell cycle** (Figure 1.16).

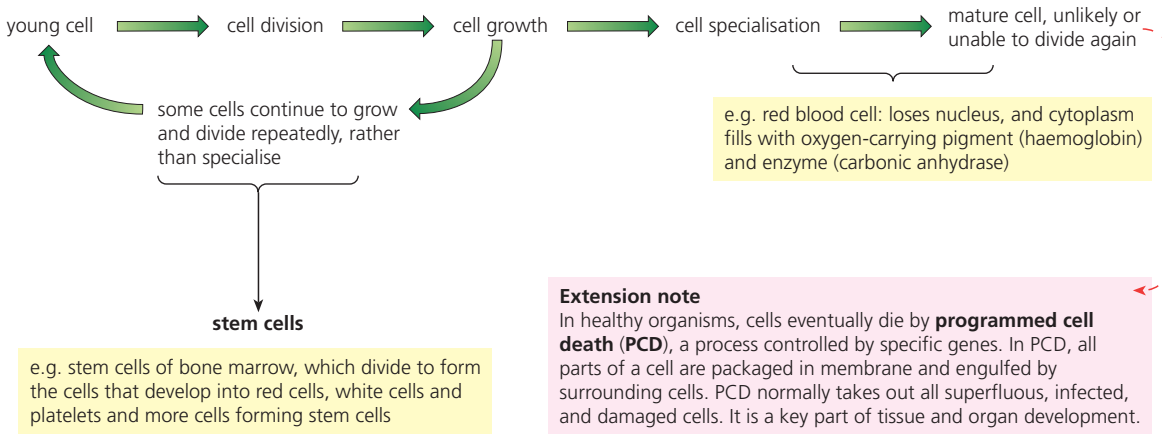


Figure 1.16 The life history of a cell and the role of stem cells

We have noted that most of the cells of a multicellular organism like ourselves have become highly specialised for a particular function. Many specialised cells are then unable to divide again, or can only do so if a special need arises such as when they are damaged (e.g. mammalian liver cells). However, a few cells are able to continue to divide and do so frequently throughout the life of the organism. These are the body's **stem cells**. Adult stem cells can divide an unlimited number of times, producing a new stem cell and a new body tissue cell each time. For example, blood stem cells, present in our bone marrow, produce the full range of different types of blood cell. In other positions in the body, stem cells are capable of producing other body tissue cells. In fact, stem cells may be able to grow into any of the 300 different kinds of cell in the human body.

Human stem cell research seeks to use human embryonic stem (ES) cells, obtained from embryos a few days old. ES cells are more flexible in that they may be coaxed to grow into any type of mature cell. ES cells may be extracted from human embryos that have been discarded during fertility treatments (page 230). Alternatively, therapeutic cloning is the creation of human embryos for the sole purpose of producing of ES cells (rather than cloning with the aim of producing a new human).

The hope is the ES cells, grown up in the laboratory, may eventually be successfully implanted in patients to treat diseases like Alzheimer's, Parkinson's or Type I diabetes. Perhaps genetically engineered stem cells could eventually be available to treat the genetic fault underlying sickle cell disease (page 101). Similarly, people with cystic fibrosis (page 576) might be treated with their own stem cells, removed and genetically engineered with the cystic fibrosis gene. Such cells would then be planted back in the patient in a way that might lead to the formation of healthy cells lining the airways of their lungs. This would eliminate the problem of tissue rejection that occurs in traditional transplant surgery.

Because these techniques are controversial and experimental, international scientists have agreed a consensus of principles guiding their work. New developments and challenges arise all the time around this therapy for human diseases. You can keep in touch with developments in this (and other) aspects of modern biology by reference to journals such as *New Scientist* (www.newscientist.com), and *Scientific American* (www.sciam.com). Other organisations and sources, including the BioNews website (www.bionews.org.uk), may be accessed using an internet search engine.

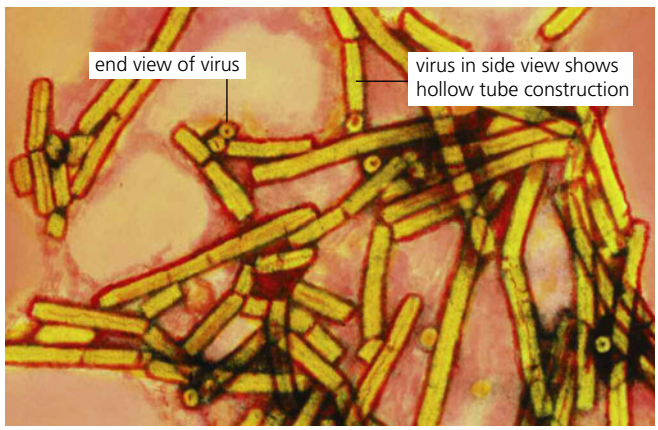
Introducing the viruses

Viruses are disease-causing agents, rather than organisms. They are not made of cells. A virus consists of a **core of nucleic acid**, either of DNA or RNA. This is surrounded by a **protein coat** called a capsid (Figure 1.17). In some viruses there is an additional external envelope of membrane made of lipids and proteins (e.g. human immunodeficiency virus – HIV).

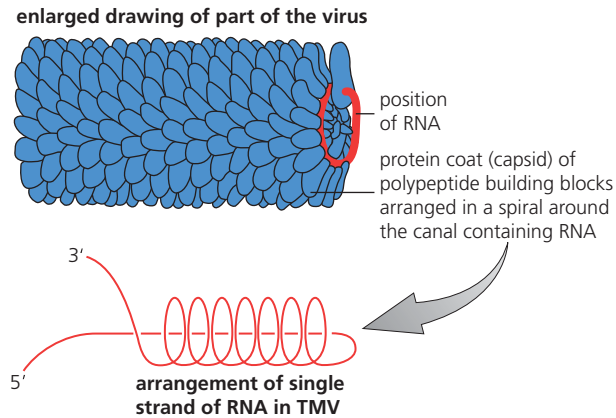
All viruses are all extremely small, most are in the size range 20–400 nm (0.02–0.4 μm). Consequently, they are visible only by means of electron microscopy.

Viruses reproduce inside specific living cells only. Here they function as parasites in their host organism. However, they have to be transported in some way between hosts. Viruses are mostly highly specific to particular host species, some to plant species, some to animal species and some to bacteria. An example of a plant virus is shown in Figure 1.17.

Figure 1.17 Tobacco mosaic virus (TMV) – an example of a virus



transmission electron micrograph of TMV (×40 000) negatively stained



10 **State** which characteristics of living things are not shown by viruses.

Relative sizes of molecules, macromolecules, viruses and organisms

There are great disparities of size among organisms, viruses, and the molecules they are built from. We measure size using the SI unit of length, the metre, and the accepted subdivisions (Table 1.1, page 2). In Figure 1.18, size relationships of biological and chemical levels of organisation are compared. The scale is logarithmic to accommodate the diversities in size in the space available, so each division is ten times larger than the division immediately below it. In science, 'powers to ten' are used to avoid writing long strings of zeros.

Remember, although sizes are expressed by a single length or diameter, all cells and organisms are three-dimensional structures, with length, breadth and depth.

11 **Distinguish** between the following pairs of terms:

- **a** cell wall and plasma membrane
- **b** nucleus and nucleoid
- **c** flagella and pili.

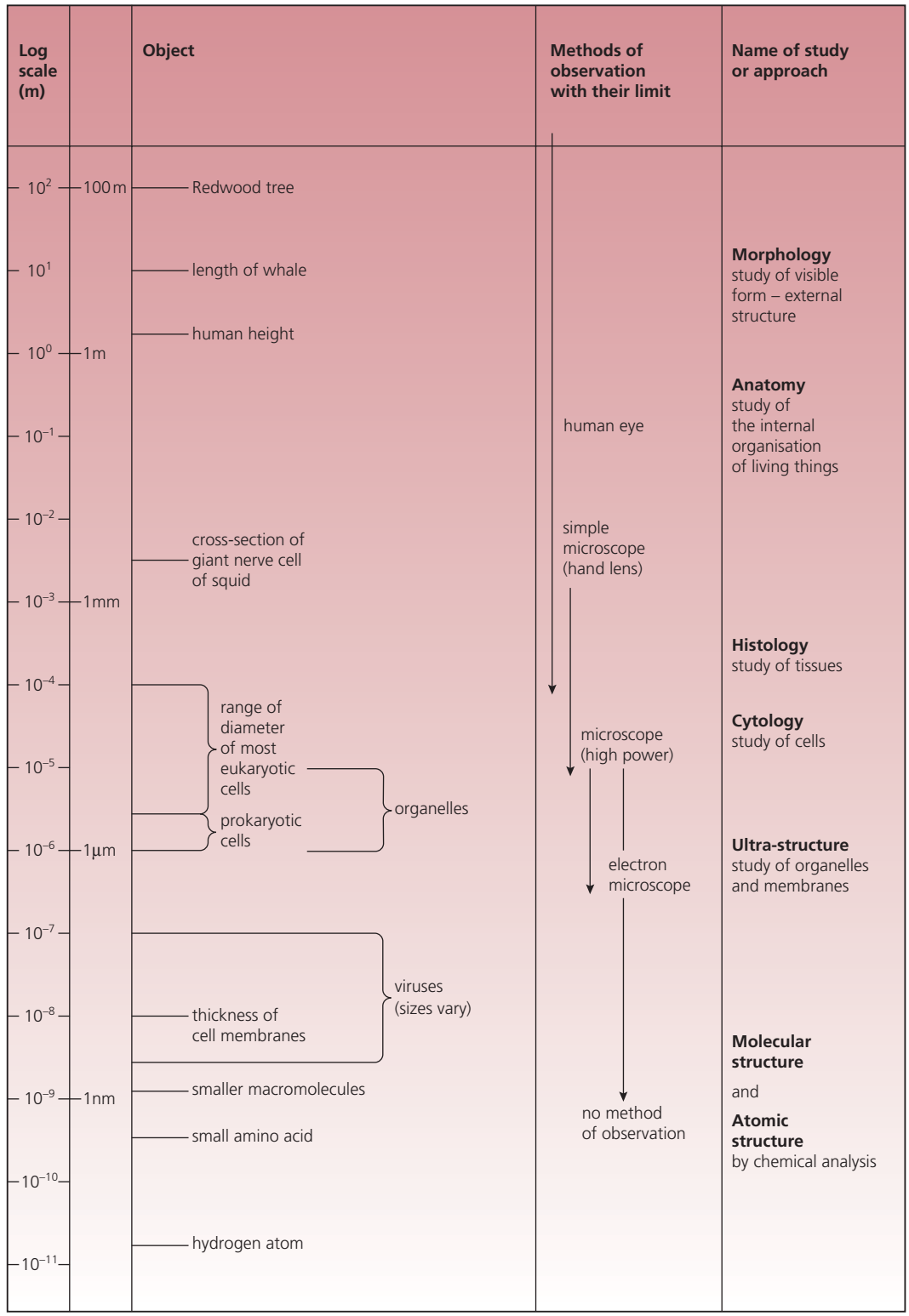


Figure 1.18 Size relationships on a logarithmic scale

Membranes

2.4.1–2.4.8

We have seen that a plasma membrane is a structure common to eukaryotic and prokaryotic cells. The plasma membrane maintains the integrity of the cell (it holds the cells contents together). Also, it is a barrier across which all substances entering and leaving the cell pass.

The structure of the plasma membrane

The plasma membrane is made almost entirely of protein and lipid, together with a small and variable amount of carbohydrate. How are these components assembled into the plasma membrane? Figure 1.19 shows the molecular structure of the plasma membrane, which is known as the **fluid mosaic model**. The plasma membrane is described as *fluid* because the components (lipids and proteins) are on the move, and *mosaic* because the proteins are scattered about in this pattern.

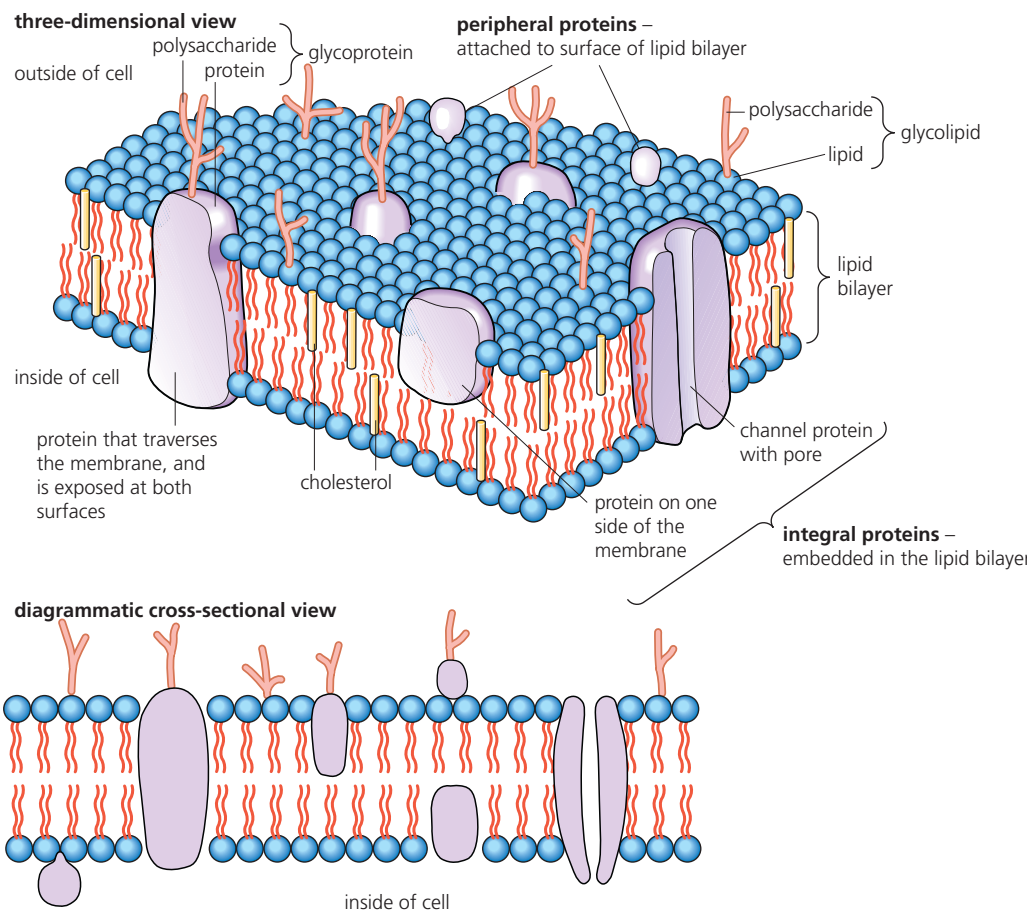


Figure 1.19 The fluid mosaic model of the plasma membrane

12 **Draw** a copy of the diagrammatic cross-section of the fluid mosaic membrane (part of Figure 1.19). **Label** it correctly, using the terms: phospholipid bilayer, cholesterol, glycoprotein, integral protein, peripheral protein.

The lipid of membranes is **phospholipid**. The chemical structure of phospholipid is shown in Figure 2.13, page 50.

Take a look at its structure, now.

You see that phospholipid has a 'head' composed of a glycerol group to which is attached one ionised phosphate group. This latter part of the molecule has **hydrophilic properties** (water-loving). For example, hydrogen bonds readily form between the phosphate head and water molecules.

The remainder of the phospholipid consists of two long, fatty acid residues consisting of hydrocarbon chains. These 'tails' have **hydrophobic properties** (water-hating).

So phospholipid is unusual in being partly hydrophilic and partly hydrophobic.

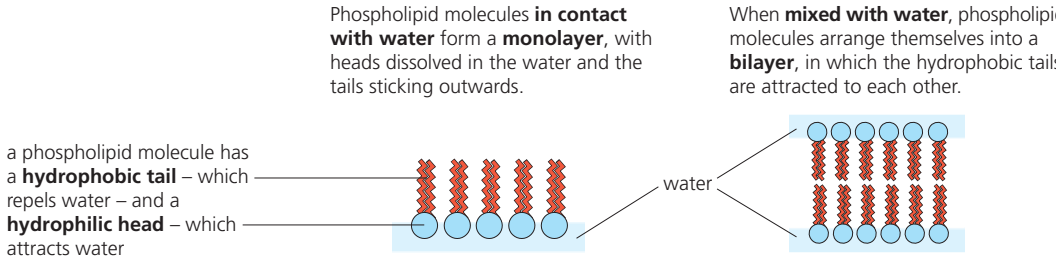


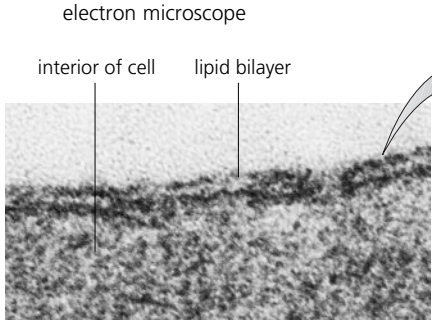
Figure 1.20 Phospholipid molecules and water – the formation of monolayers and bilayers

What are the consequences of this dual nature of phospholipid?

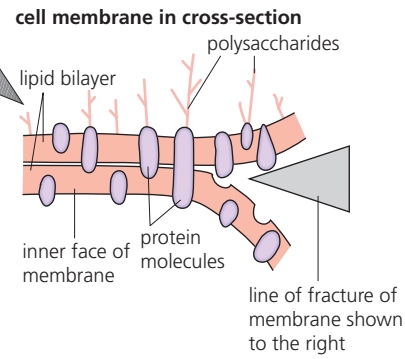
A small quantity of phospholipid in contact with water will float with the hydrocarbon tails exposed above the water, forming a monolayer of phospholipid molecules (Figure 1.20). When more phospholipid is available, the molecules arrange themselves as a bilayer, with the hydrocarbon tails facing together. This is the situation in the plasma membrane (Figure 1.21).

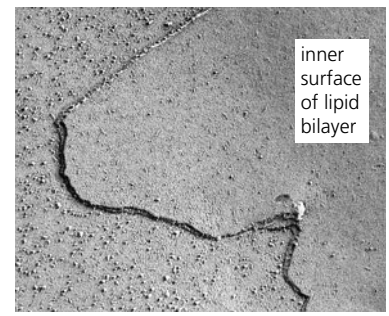
In the lipid bilayer, attractions between the hydrophobic hydrocarbon tails on the inside, and between the hydrophilic glycerol/phosphate heads and the surrounding water on the outside make a stable, strong barrier.

Figure 1.21 Plasma membrane structure; evidence from the electron microscope



TEM of the cell membrane of a red cell (×700 000)





electron micrograph of the cell membrane (freeze-etched)

The proteins of plasma membranes are **globular proteins** (page 257). Some of these proteins occur partially or fully buried in the lipid bilayer, and are described as integral proteins. Others are superficially attached on either surface of the lipid bilayer and are known as peripheral proteins. The functions of the membrane proteins are various; some are channels for transport of metabolites, some are enzymes or carriers, and some receptors or antigens (Figure 1.22).

The **carbohydrate** molecules of the membrane are relatively short-chain polysaccharides, some attached to the proteins (**glycoproteins**) and some to the lipids (**glycolipids**). This glycocalyx, as it is known, occurs only on the outer surface of the plasma membrane. Its various functions involve, for example, cell–cell recognition, acting as receptor sites for chemical signals, and the binding of cells into tissues.

13 **State** the difference between a lipid bilayer and the double membrane of many organelles.

Movement across the plasma membrane

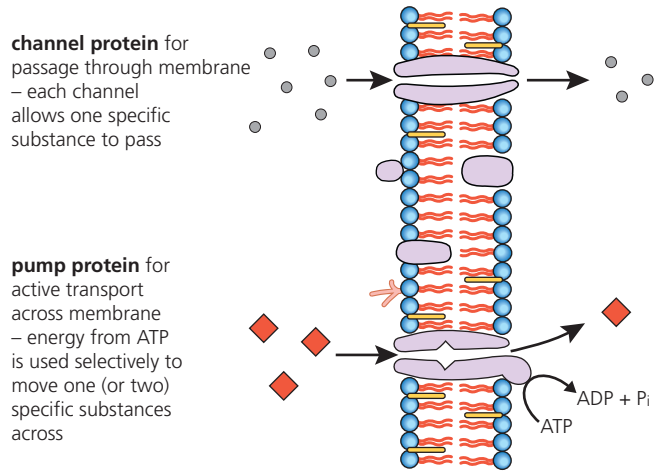
Movement of molecules across the plasma membrane of living cells is continuous and heavy. Into and out of cells pass **water**, **respiratory gases** (oxygen and carbon dioxide), **nutrients** such as glucose, **essential ions**, and **excretory products**.

Cells may secrete substances such as **hormones** and **enzymes**, and they may receive **growth substances** and certain hormones. Plants secrete the chemicals that make up their walls through their cell membranes, and assemble and maintain the wall outside the membrane. Certain mammalian cells secrete **structural proteins** such as collagen, in a form that can be assembled outside the cells.

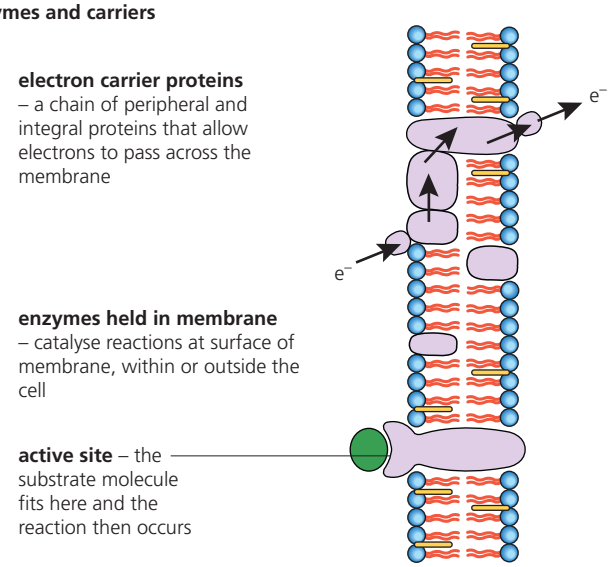
In addition, the plasma membrane is where the cell is identified by surrounding cells and organisms. For example, **protein receptor sites** are recognised by hormones, by neurotransmitter substances from nerve cells, and by other chemicals sent from other cells. Figure 1.23 is a summary of this movement, and Figure 1.24 summarises the mechanisms of transport across membranes, into which we need to look further.

Figure 1.22 The functions of membrane proteins

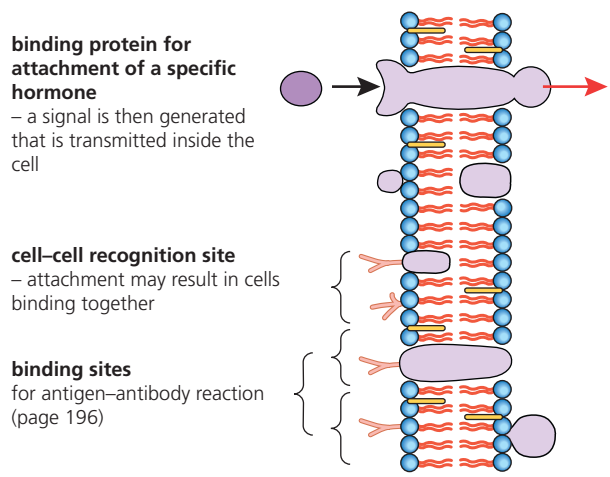
1 channels for transport of metabolites or water



2 enzymes and carriers



3 receptors, antigens, cell–cell recognition and cell binding sites



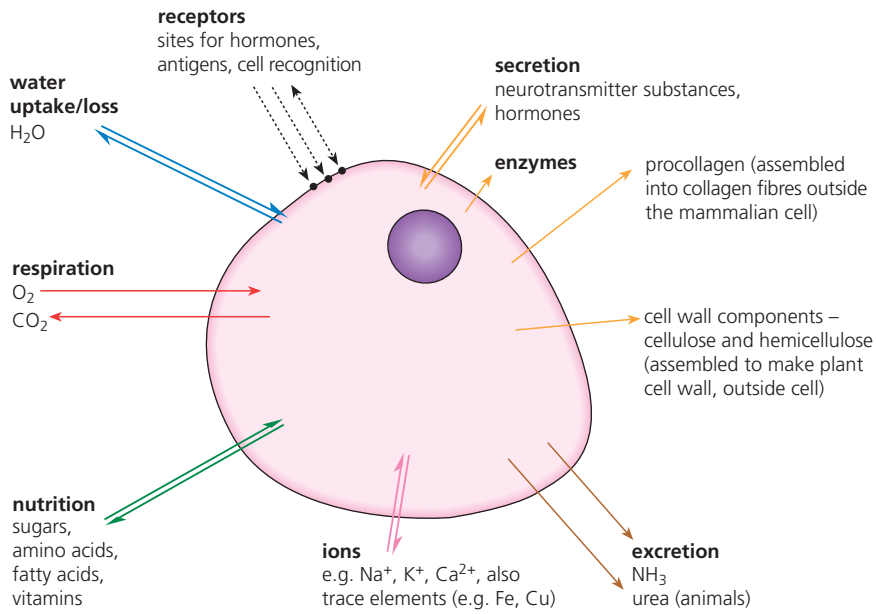


Figure 1.23 Movements across the plasma membrane

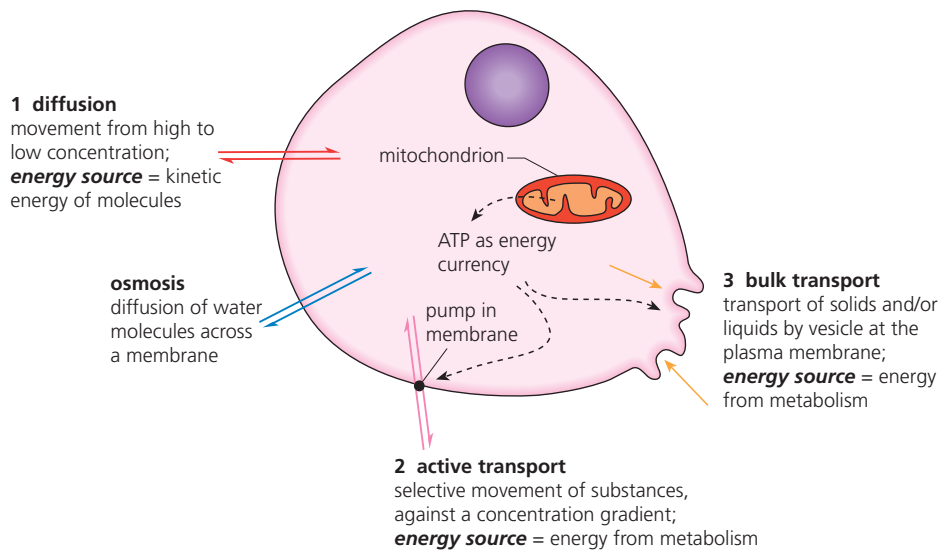


Figure 1.24 Mechanisms of movement across membranes

Movement by diffusion

The atoms, molecules and ions of liquids and gases undergo continuous random movements. These movements result in the even distribution of the components of a gas mixture and of the atoms, molecules and ions in a solution. So, for example, from a solution we are able to take a tiny random sample and analyse it to find the concentration of dissolved substances in the whole solution – because any sample has the same composition as the whole. Similarly, every breath we take has the same amount of oxygen, nitrogen and carbon dioxide as the atmosphere as a whole.

Continuous random movements of all molecules ensures complete mixing and even distribution, given time, in solutions and gases.

Diffusion is the free passage of molecules (and atoms and ions) from a region of their high concentration to a region of low concentration.

Where a difference in concentration has arisen in a gas or liquid, random movements carry molecules from a region of high concentration to a region of low concentration. As a result, the particles become evenly dispersed. The energy for diffusion comes from the **kinetic energy** of molecules. *Kinetic* means that a particle has this energy because it is in continuous motion.

Diffusion in a liquid can be illustrated by adding a crystal of a coloured mineral to distilled water. Even without stirring, the ions become evenly distributed throughout the water (Figure 1.25). The process takes time, especially as the solid has first to dissolve.

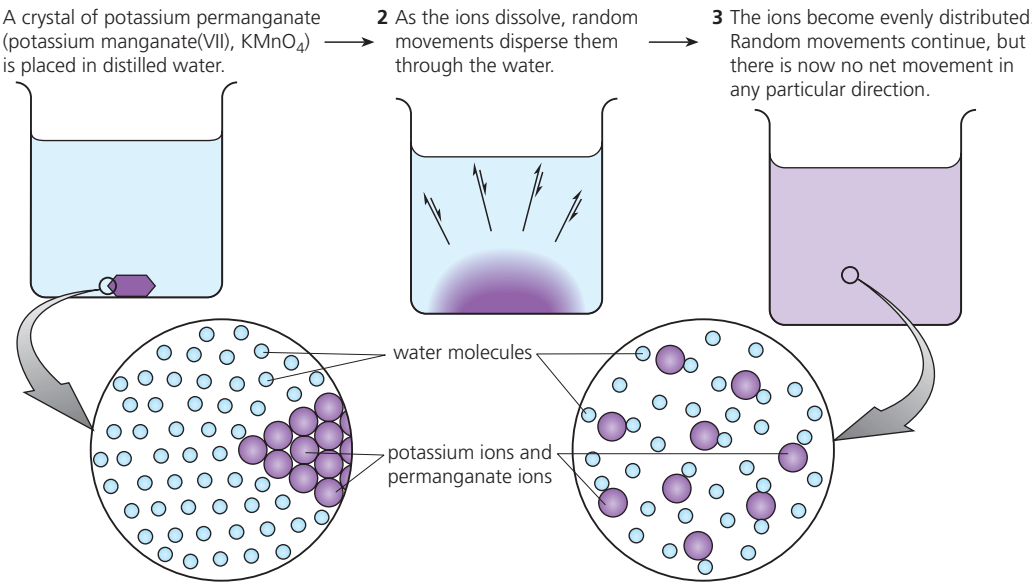


Figure 1.25 Diffusion in a liquid

Diffusion in cells

Diffusion across the cell membrane occurs where:

- The plasma membrane is fully permeable to the solute. The lipid bilayer of the plasma membrane is permeable to non-polar substances, including steroids and glycerol, and also oxygen and carbon dioxide in solution, all of which diffuse quickly via this route.
- The pores in the membrane are large enough for a solute to pass through (Figure 1.26). Water diffusing across the plasma membrane passes via the protein-lined pores of the membrane, and via tiny spaces between the phospholipid molecules. This latter occurs easily where the fluid mosaic membrane contains phospholipids with unsaturated hydrocarbon tails, for here these hydrocarbon tails are spaced more widely. The membrane is consequently especially 'leaky' to water, for example.

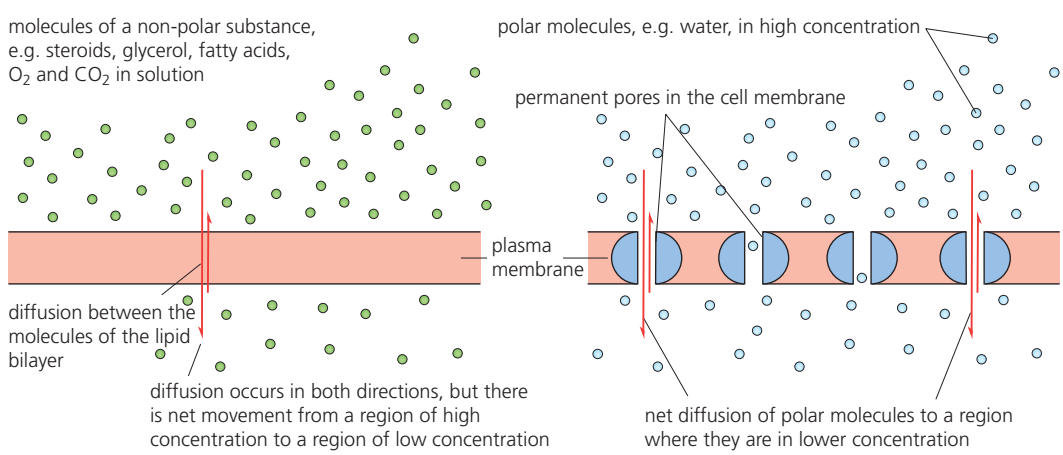


Figure 1.26 Diffusion across the plasma membrane

14 Students were provided with cubes of slightly alkaline gelatine of different dimensions, containing an acid–alkali indicator that is red in alkali but yellow in acid. The cubes were placed in dilute acid solution, and the time taken for the colour in the gelatine to change from red to yellow was measured.

Dimensions/mm	Surface area/mm^2	Volume/mm^3	Time/minutes
10 × 10 × 10	600.0	1000.0	12.0
5 × 5 × 5	150.0	125.0	4.5
4 × 4 × 4	96.0	64.0	4.2
2.5 × 2.5 × 2.5	37.5	15.6	4.0

a For each block, **calculate** the ratio of surface area to volume (SA:V).

b Plot a graph of the time taken for the colour change (vertical or *y* axis) against the SA:V ratio (horizontal or *x* axis).

c **Explain** why the colour changes more quickly in some blocks than others.

Facilitated diffusion

In facilitated diffusion, a substance that otherwise is unable to diffuse across the plasma membrane does so as a result of its effect on particular molecules present in the membrane. In the presence of the substance, these membrane molecules, made of globular protein, form into pores large enough to allow diffusion; they close up again when the substance is no longer present (Figure 1.27). In facilitated diffusion, the energy comes from the kinetic energy of the molecules involved, as is the case in all forms of diffusion. Energy from metabolism is not required. Important examples of facilitated diffusion are the movement of ADP into mitochondria and the exit of ATP from mitochondria (page 80).

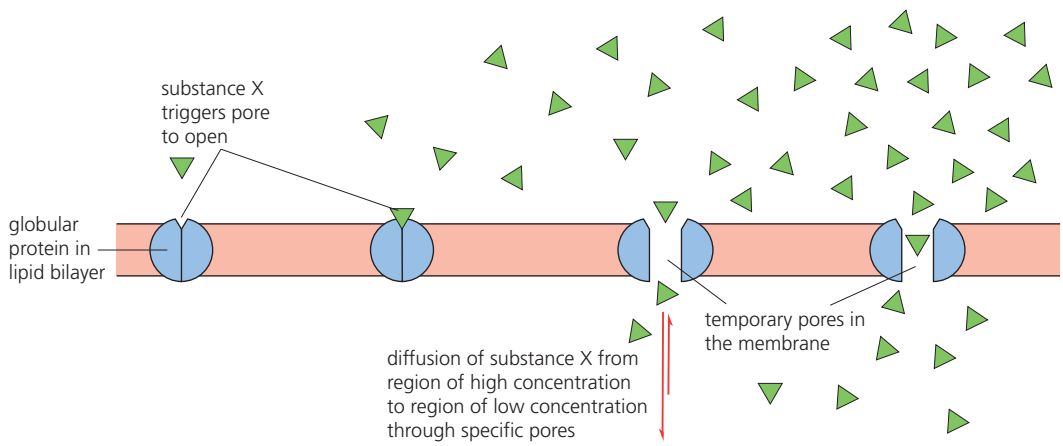


Figure 1.27 Facilitated diffusion

15 **Distinguish** between diffusion and facilitated diffusion.

Osmosis – a special case of diffusion

Osmosis is a special case of diffusion. It is the diffusion of water molecules across a membrane which is permeable to water (**partially permeable**). Since water makes up 70–90% of living cells and cell membranes are partially permeable membranes, osmosis is very important in biology.

Why does osmosis happen?

Dissolved substances attract a group of polar water molecules (page 38) around them. The forces holding the water molecules in this way are weak chemical bonds, including **hydrogen bonds**. Consequently, the tendency for random movement by the dissolved substances and their surrounding water molecules is restricted. Organic substances like sugars, amino acids, polypeptides and proteins, and inorganic ions like Na^+, K^+, Cl^- and NO_3^-, have this effect on the water molecules around them.

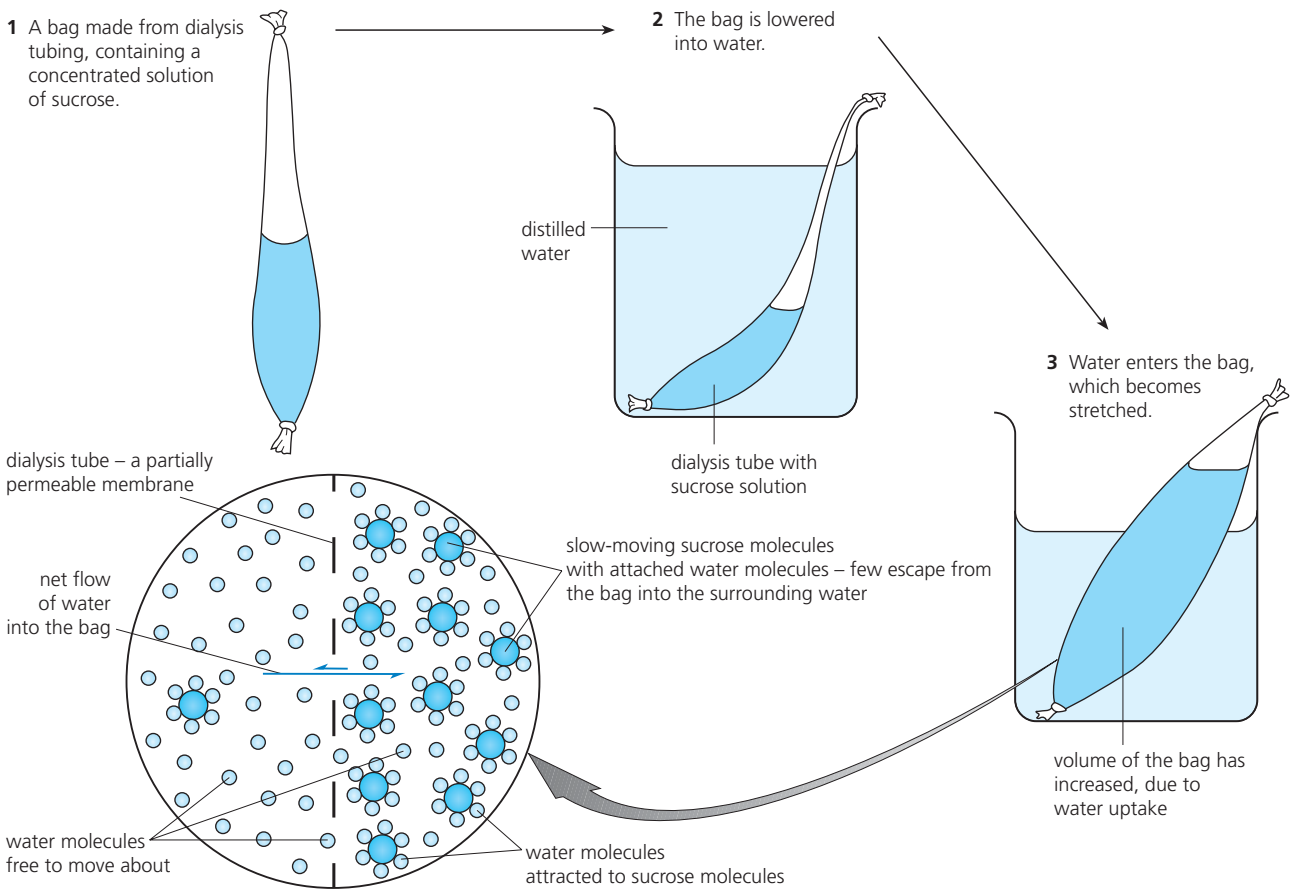
The stronger the solution (i.e. the more solute dissolved per volume of water) the greater the number of water molecules that are slowed up and held almost stationary. So, in a very concentrated solution, very many more of the water molecules have restricted movement than in a dilute solution. In pure water, all of the water molecules are free to move about randomly, and do so.

When a solution is separated from water (or a more dilute solution) by a membrane permeable to water molecules (such as the plasma membrane), water molecules free to move tend to diffuse, while dissolved molecules and their group of water molecules move very much less, if at all. So there is a net flow of water, from a more dilute solution into a more concentrated solution, across the membrane. The membrane is described as partially permeable.

Osmosis is the net movement of water molecules (solvent), from a region of high concentration of water molecules to a region of lower concentration of water molecules, across a selectively permeable membrane (Figure 1.28). Alternatively, we can state the following.

Osmosis is the passive movement of water molecules, across a partially permeable membrane, from a region of lower solute concentration to a region of higher solute concentration.

Figure 1.28 Osmosis



16 When a concentrated solution of glucose is separated from a dilute solution of glucose by a partially permeable membrane, **determine** which solution will show a net gain of water molecules.

Movement by active transport

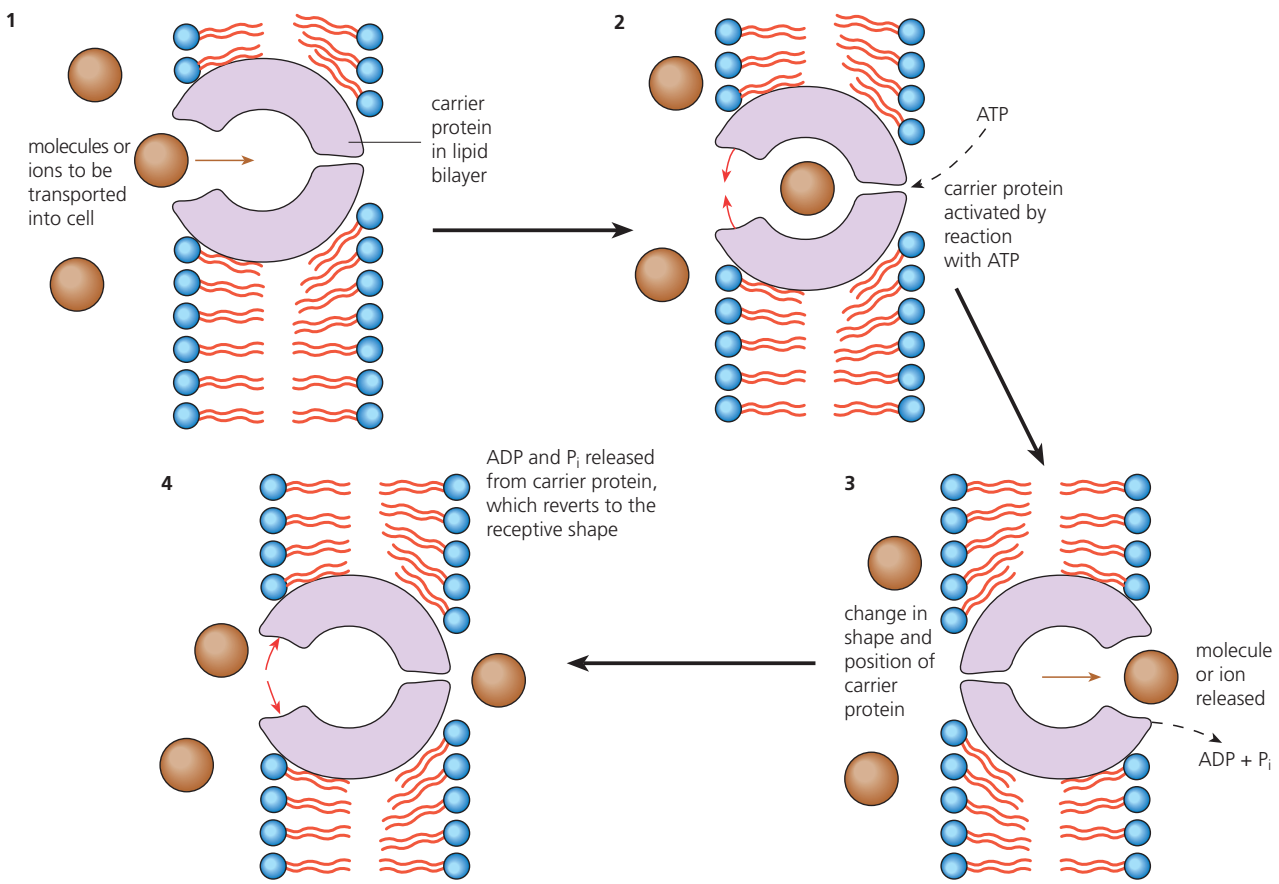
We have seen that diffusion is due to random movements of molecules, and occurs spontaneously, from a high to a low concentration. However, many of the substances required by cells have to be absorbed from a weak external concentration and taken up into cells that contain a higher concentration. Uptake against a concentration gradient cannot occur by diffusion. Instead, it requires a source of energy to drive it. This type of uptake is known as **active transport**.

In active transport, metabolic energy produced by the cell, held as ATP, is used to drive the transport of molecules and ions across cell membranes. Active transport has characteristic features distinctly different from those of movement by diffusion.

1 **Active transport occurs against a concentration gradient**
It occurs from a region of low concentration to a region of higher concentration. The cytoplasm of a cell normally holds some reserves of molecules and ions valuable in metabolism, like nitrate ions in plant cells, or calcium ions in muscle fibres. The reserves of useful molecules and ions do not escape; the cell membrane retains them inside the cell. Yet when more of these or other useful molecules or ions become available for uptake, they too are actively absorbed into the cells. This happens even though the concentration outside the cell is lower than that inside.

2 **Active uptake is highly selective**
For example, in a situation where potassium chloride (K^+ and Cl^- ions) is available to an animal cell, K^+ ions are more likely to be absorbed, since they are needed by the cell. Where sodium nitrate (Na^+ and NO_3^- ions) is available to a plant cell, it is likely that more of the NO_3^- ions are absorbed than the Na^+, since this reflects the needs of plant cells.

Figure 1.29 Active transport of a single substance



3 **Active transport involves special molecules of the membrane, called pump molecules**
The pump molecules pick up particular molecules or ions and transport them to the other side of the membrane, where they are then released. The pump molecules are globular proteins, sometimes also called carrier proteins, that span the lipid bilayers (Figure 1.29). Movements by the pump molecules require reaction with ATP; this reaction supplies metabolic energy to the process. Most membrane pumps are specific to particular molecules or ions and this is the way selective transport is brought about. If the pump molecule for a particular substance is not present, the substance will not be transported.

Active transport is a feature of most living cells. We meet examples of active transport in the gut where absorption occurs (Chapter 7), in the active uptake of ions by plant roots (Chapter 20), in the kidney tubules where urine is formed, and in nerve fibres where an impulse is propagated (Chapter 13).

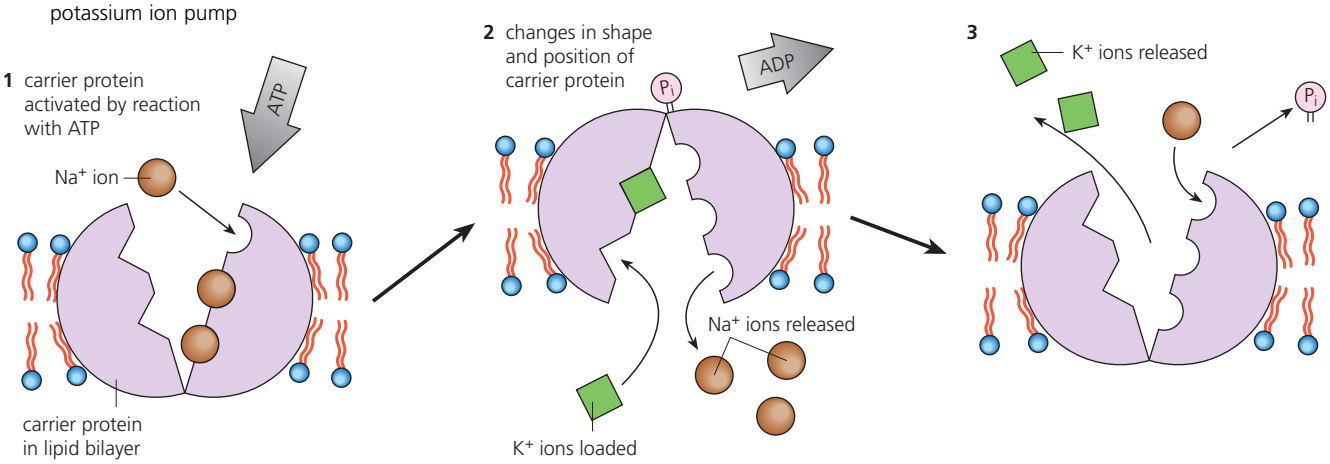
The protein pumps of plasma membranes are of different types. Some transport a particular molecule or ion in one direction (Figure 1.29), while others transport two substances (like Na^+ and K^+) in opposite directions (Figure 1.30). Occasionally, two substances are transported in the same direction; for example, Na^+ and glucose (Figure 20.13, page 654).

17 Samples of five plant tissue discs were incubated in dilute sodium chloride solution at different temperatures. After 24 hours, it was found that the uptake of ions from the solutions was as follows (arbitrary units).

	Sodium ions	Chloride ions
Tissue at 5 °C	80	40
Tissue at 25 °C	160	80

Comment on how absorption of sodium chloride occurs, giving your reasons.

Figure 1.30 The sodium/potassium ion pump



Movement by bulk transport

Another mechanism of transport across the plasma membrane is known as **bulk transport**. It occurs by movements of **vesicles** of matter (solids or liquids) across the membrane by processes known generally as **cytosis**. Uptake is called **endocytosis** and export is **exocytosis**.

The strength and flexibility of the fluid mosaic membrane makes this activity possible. Energy from metabolism (ATP) is also required to bring it about. For example, when solid matter is being taken in (**phagocytosis**), part of the plasma membrane at the point where the vesicle forms is pulled inwards and the surrounding plasma membrane and cytoplasm bulge out. The matter thus becomes enclosed in a small vesicle.

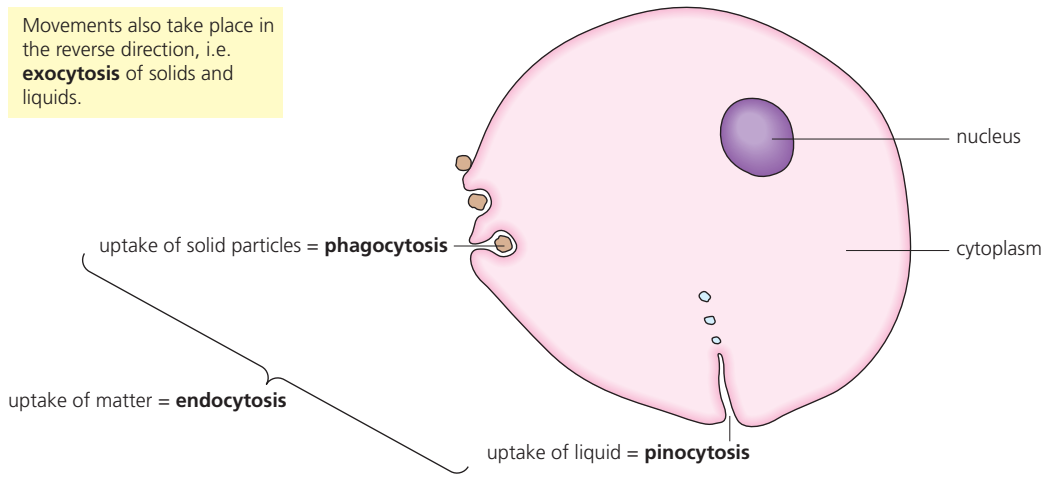
Vesicles are used to transport materials within cells, for example, between the rough endoplasmic reticulum (rER) and the Golgi apparatus, and on to the plasma membrane (Figure 1.13, page 15).

In the human body, there is a huge number of phagocytic cells (phagocytosis means 'cell eating'). These are called the **macrophages**. The macrophages engulf the debris of damaged or dying cells and dispose of it. For example, we break down about 2×10^{11} red cells each day. This number are ingested and disposed of by macrophages, every 24 hours.

Bulk transport of fluids is referred to as pinocytosis (Figure 1.31).

Figure 1.31 Transport by cytosis

Movements also take place in the reverse direction, i.e. **exocytosis** of solids and liquids.



18 **Distinguish** between the following pairs:

- a proteins and lipids in cell membranes
- b active transport and bulk transport
- c endocytosis and exocytosis.

Cell division

2.5.1–2.5.6

New cells arise by division of existing cells. Unicellular organisms grow quickly under favourable conditions, then divide into two cells. The growth and division cycle is repeated while conditions are suitable. Multicellular organisms begin life as a single cell which grows and divides, forming many cells which eventually make up the adult organism. In these situations, it is important that division of a nucleus produces two new cells (daughter cells) containing identical sets of chromosomes; this is achieved by **mitosis**.

The cell division cycle

The cycle of growth and division of a cell is called the cell division cycle. The **cell division cycle** has three main stages:

- **interphase**;
- division of the nucleus by **mitosis**;
- **cytokinesis** (cell division).

In each stage of the cell cycle particular events occur. These events are summarised in Figure 1.32, and they are discussed below.

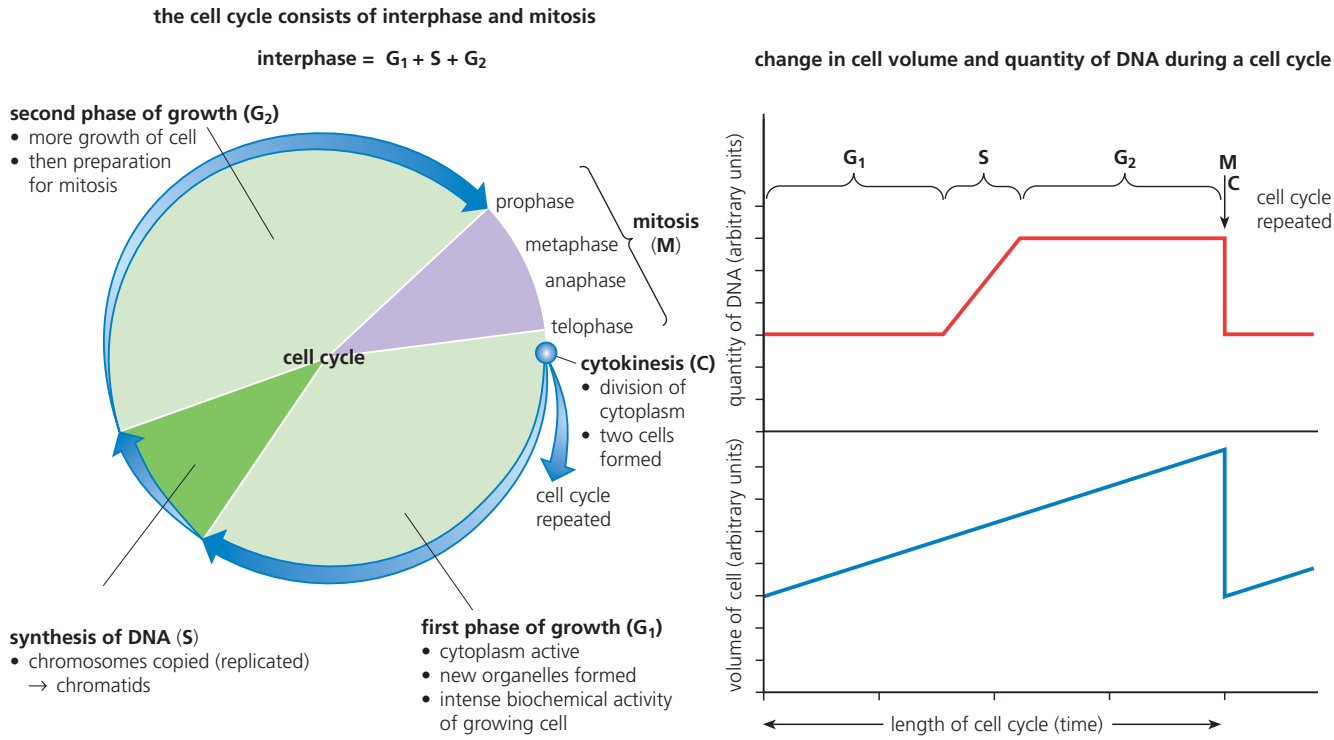


Figure 1.32 The stages of the cell cycle

Interphase

Interphase is always the longest part of the cell cycle, but it is of extremely variable length. When growth is fast, as in a developing human embryo, and the growing point of a young stem, interphase may last about 24 hours or less. On the other hand, in mature cells that infrequently divide it lasts a very long period – possibly indefinitely. For example, some cells, once they have differentiated, rarely or never divide again. Here the nucleus remains at interphase permanently.

What happens to the nucleus in interphase?

At first glance the nucleus appears to be resting, but this is not the case. The chromosomes cease to be visible as thread-like structures at interphase, becoming dispersed as chromatin. Now they are actively involved in protein synthesis. From the chromosomes, copies of the information in particular genes or groups of genes are taken for use in the cytoplasm (page 69). In organelles of the cytoplasm called ribosomes, proteins are assembled by combining amino acids in sequences dictated by the information from the gene.

During interphase, the synthesis of new organelles takes place in the cytoplasm. There is intense biochemical activity in the cytoplasm and the organelles, and there is an accumulation of stored energy before nuclear division occurs again.

Also in this period, between nuclear divisions, each chromosome **replicates** (makes a copy of itself). The two identical structures formed are called **chromatids**. The chromatids remain attached until they divide during mitosis.

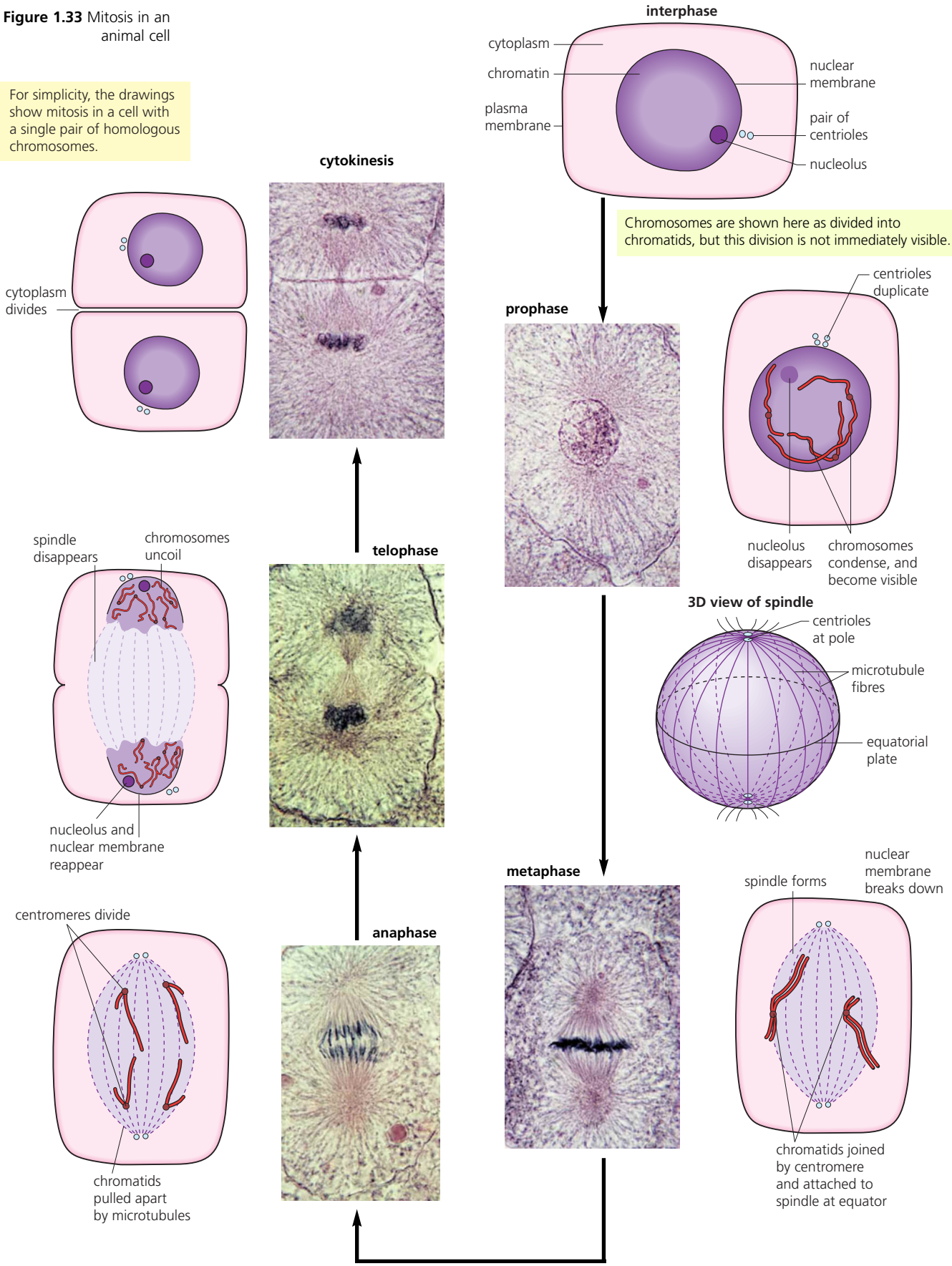
Mitosis

When cell division occurs, the nucleus divides first. In mitosis, the chromosomes, present as the chromatids formed during interphase, are separated, and accurately and precisely distributed to two daughter nuclei (Figure 1.33). In the following description, mitosis is presented and explained as a process in four phases, but remember this is for convenience of description only. Mitosis is a continuous process with no breaks between the phases.

In **prophase**, the chromosomes become visible as long thin threads. They increasingly shorten and thicken by a process of supercoiling. Only at the end of prophase is it possible to see that they consist of two chromatids held together at the **centromere**. At the same time, the nucleolus gradually disappears and the nuclear membrane breaks down.

Figure 1.33 Mitosis in an animal cell

For simplicity, the drawings show mitosis in a cell with a single pair of homologous chromosomes.



In **metaphase**, the centrioles move to opposite ends of the cell. Microtubules in the cytoplasm start to form into a spindle, radiating out from the centrioles (Figure 1.33, page 32). Microtubules attach to the centromeres of each pair of chromatids, and these are arranged at the equator of the spindle. (Note: in plant cells, a spindle of exactly the same structure is formed, but without the presence of the centrioles.)

In **anaphase**, the centromeres divide, the spindle fibres shorten, and the chromatids are pulled by their centromeres to opposite poles. Once separated, the chromatids are referred to as chromosomes.

In **telophase**, a nuclear membrane reforms around both groups of chromosomes at opposite ends of the cell. The chromosomes decondense by uncoiling, becoming chromatin again. The nucleolus reforms in each nucleus. Interphase follows division of the cytoplasm.

Cytokinesis

Division of the cytoplasm, known as **cytokinesis**, follows telophase. During division, cell organelles such as mitochondria and chloroplasts become distributed evenly between the cells. In animal cells, division is by in-tucking of the plasma membrane at the equator of the spindle, 'pinching' the cytoplasm in half (Figure 1.33).

In plant cells, the Golgi apparatus forms vesicles of new cell wall materials which collect along the line of the equator of the spindle, known as the cell plate. Here the vesicles coalesce forming the new plasma membranes and cell walls between the two cells (Figure 1.34).

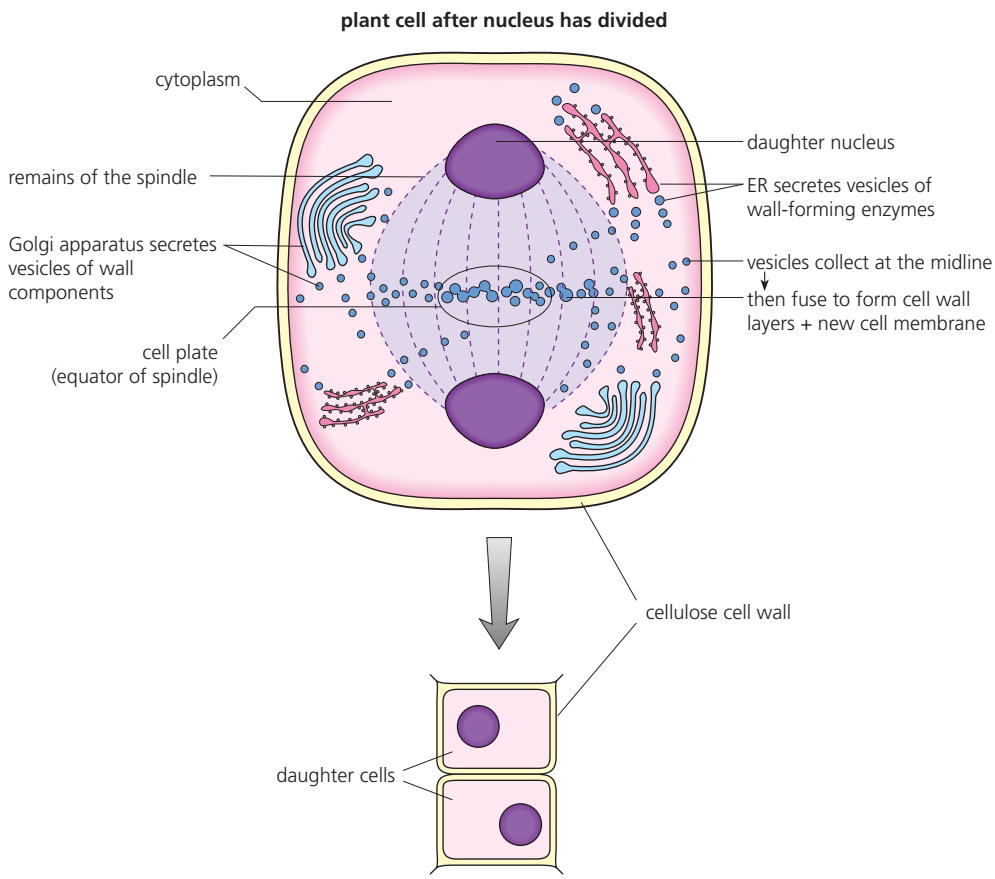


Figure 1.34 Cytokinesis in a plant cell

Observing chromosomes during mitosis

Actively dividing cells, such as those at the growing points of the root tips of plants, include many cells undergoing mitosis. This tissue can be isolated, stained with an orcein ethanoic (acetic orcein) stain, squashed, and then examined under the high-power lens of a microscope. Nuclei at interphase appear red–purple with almost colourless cytoplasm, but the chromosomes in cells undergoing mitosis will be visible, rather as they appear in the photomicrographs in Figure 1.33. The procedure is summarised in the flow diagram in Figure 1.35.

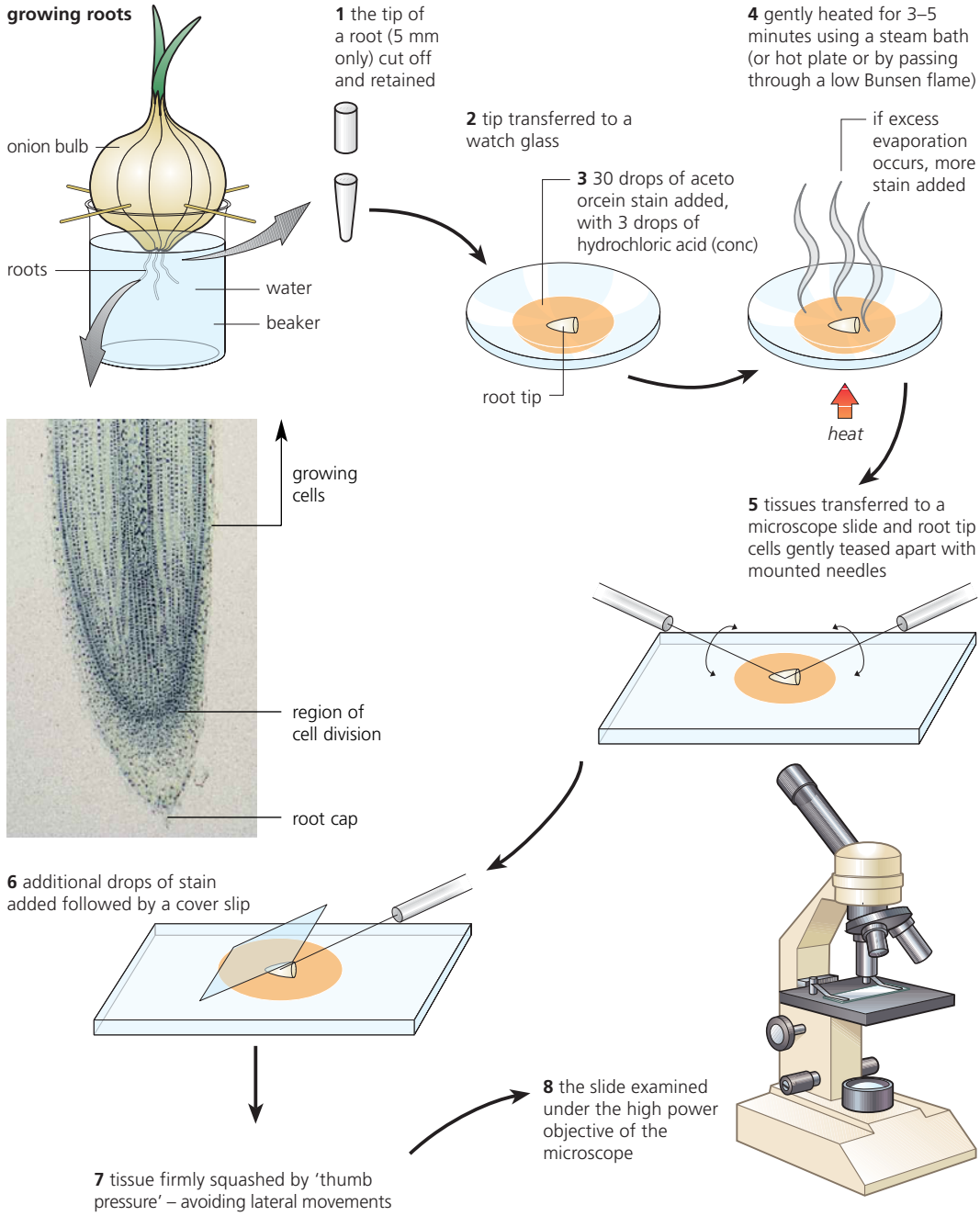


Figure 1.35 The orcein ethanoic stain of an onion root tip squash

19 Using slides they had prepared to observe chromosomes during mitosis in a plant root tip (Figure 1.35), five students observed and recorded the number of nuclei at each stage in mitosis in 100 cells as shown in the table below.

	Number of nuclei counted by				
Stage of mitosis	**student 1**	**student 2**	**student 3**	**student 4**	**student 5**
prophase	64	70	75	68	73
metaphase	13	10	7	11	9
anaphase	5	5	2	8	5
telophase	18	15	16	13	13

a **Calculate** the mean % of dividing cells at each stage of mitosis and present your results as a pie chart.
b Assuming that mitosis takes about 60 minutes to complete in this species of plant, **deduce** what these results imply about the lengths of the four steps.

The significance of mitosis

Daughter cells produced by mitosis have a set of chromosomes identical to each other and to the parent cell from which they were formed. This occurs because:

- an **exact copy** of each chromosome is made by accurate replication (page 66) during interphase, when two chromatids are formed;
- **chromatids remain attached** by their centromeres during metaphase of mitosis, when each becomes attached to a spindle fibre at the equator of the spindle;
- centromeres then divide during anaphase and the chromatids of each pair are pulled apart to **opposite poles** of the spindle; thus, one copy of each chromosome moves to each pole of the spindle;
- the chromosomes at the poles **form the new nuclei** – two to a cell at this point;
- two cells are then formed by division of the cytoplasm at the midpoint of the cell, **each with an exact copy of the original nucleus**.

In **growth and development of an embryo**, it is important that all cells carry the same genetic information as the existing cells from which they are formed, and which they share with surrounding cells or tissues. Similarly, when **repair of damaged or worn out cells** occurs, they are exact copies of what they replace. In fact, this is essential for growth, development and repair, because otherwise different parts of our body might start working to conflicting blueprints. The results would be chaos.

Mitotic cell division is also the basis of all forms of **asexual reproduction** (page 133), in which the offspring produced are identical to the parent.

Cancer – disease caused by uncontrolled cell division

There are very many different forms of cancer, affecting different tissues of the body. Cancer is not thought of as a single disease. However, in cancer, cells divide by mitosis repeatedly, without control or regulation, forming an irregular mass of cells, called a tumour. Sometimes tumour cells break away and are carried to other parts of the body, where they form a secondary tumour. Unchecked, cancerous cells ultimately take over the body at the expense of the surrounding, healthy cells, leading to malfunction and death.

Cancer is caused by damage to DNA of chromosomes. 'Mistakes' of different types build up in the DNA of the body cells. The accumulation of mistakes with time explains why the majority of cancers arise in older people. Different types of mistake can occur, and this explains why cancer is not one single disease. Another cause of cancer is damage to the gene that codes for a protein known as p53 which stops the copying of damaged DNA. This p53 protein and the gene that codes for it are called 'the guardians of the genome'.

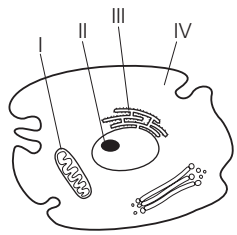
Damage to DNA has many potential causes, including the effects of ionising radiation (X-rays, gamma rays and others), certain chemicals (such as components of the tar in tobacco smoke) and some virus infections. Another set of factors is inherited, so that the members of some families are more likely to suffer from certain cancers than others.

20 **Describe** how the behaviour of cancerous cells differs from that of normal cells.

Examination questions – a selection

Questions 1–5 are taken from past IB Diploma biology papers.

Q1 The diagram represents an animal cell. Which processes occur in the locations labelled?



	Transcription	Translation	Glycolysis
A	II	III	I
B	III	II	I
C	II	III	IV
D	III	II	IV

Standard Level Paper 1, May 06, Q3

Q2 Which of the following is a characteristic of organelles?

A They are only found in eukaryotic cells.
B They are only found in prokaryotic cells.
C They are sub-cellular structures.
D They are all membrane-bound.

Standard Level Paper 1, May 04, Q2

Q3 **a** List **two** functions of membrane proteins. (1)
b Oxygen (O_2) moves across the membrane by diffusion. Define the term diffusion. (1)
c Potassium can move across the membrane by passive or active transport. Distinguish between active and facilitated diffusion of ions. (2)
d The hormone insulin leaves the cell by exocytosis. Describe the process of exocytosis. (2)

Standard Level Paper 2, May 06, Q3

Q4 The width of a human hair is 0.1 mm. What is the width in μm?

A 10 μm **B** 100 μm **C** 1000 μm **D** 10 000 μm

Higher Level Paper 1, May 06, Q2

Q5 **a** State **two** processes which involve mitosis. (2)
b Explain the importance of the surface area to volume ratio as a factor limiting cell size. (3)
c State **one** difference between the proteins produced by free ribosomes and those produced by ribosomes attached to the endoplasmic reticulum. (1)

Higher Level Paper 2, May 05, Q2

Questions 6–10 cover other syllabus issues in this chapter.

Q6 List the features animal and plant cells have in common. Construct a table of the differences in structure between typical plant and animal cells. (6)

Q7 The stages of the cell cycle are interphase, mitosis and cytokinesis. Four distinct phases make up the process of mitosis itself.
a Outline the events of interphase which establish that this is not a 'resting' stage in the cell cycle. (4)
b List the major changes that occur to the chromosomes of a nucleus undergoing mitosis during:
i prophase (the first phase)
ii anaphase (the third phase). (6)
c Explain how mitosis produces two genetically identical nuclei. (4)
d Distinguish between stem cells and cancer cells. (2)

Q8 The structure of a plasma membrane is represented by the fluid mosaic model. It is a barrier that has to be crossed by material entering or leaving a cell.
a Draw and label a diagram to illustrate the structure of the plasma membrane as seen in cross-section. (6)
b Explain how the properties of phospholipids help maintain the structure of the plasma membrane. (3)
c List the essential conditions necessary for diffusion. (4)
d Explain the process by which water moves across the plasma membrane. (4)

Q9 **a** Draw and label a diagram of a generalised prokaryotic cell. Annotate your diagram with the functions of each named structure. (6)
b Explain how the size of a prokaryotic cell relates to the size of some organelles typically found in animal cells. (2)
c Define binary fission. (2)

Q10 **a** List three ideas contained within the cell theory. (3)
b Outline the differences between the processes typically observed in a eukaryotic unicellular organism and a **named** highly differentiated animal cell. (4)

2 Chemistry of life

STARTING POINTS

- **Elements** are the basic units of pure substances. An **atom** is the smallest part of an element that can take part in a chemical change. At the centre of an atom is a nucleus of **protons** (positively charged particles), and usually also **neutrons** (uncharged particles). Around the nucleus are tiny particles called **electrons** (negatively charged). A **molecule** is a group of like or different atoms held together by chemical forces. A **compound** contains two or more elements chemically combined together.
- The element **carbon** has a branch of chemistry to itself, known as **organic chemistry**. Compounds built from carbon are called **organic compounds**.
- A vast number of organic compounds make up living things, but most of them fall into one of **four groups** of compounds, each with distinctive structures and properties. They are: **carbohydrates**, **lipids** (e.g. fats and oils), **proteins**, and **nucleic acids**.
- Most cells are at least **80% water** by mass. The properties of water are vital to life.

Chemical **elements** are the units of pure substance that make up our world. The Earth is composed of about 92 stable elements in all, in varying quantities, and living things are built from some of them. In Table 2.1 there is a comparison of the most common elements in the Earth's crust and in us. You can see that the bulk of the Earth is composed of the elements oxygen, silicon, aluminium and iron. Of these, only oxygen is a major component of cells.

	Earth's crust		Human body	
1	oxygen	47%	hydrogen	63%
2	silicon	28%	oxygen	25.5%
3	aluminium	7.9%	carbon	9.5%
4	iron	4.5%	nitrogen	1.4%
5	calcium	3.5%	calcium	0.31%
6	sodium	2.5%	phosphorus	0.22%

Table 2.1 Most common elements

Chemical elements and water

3.1.1–3.1.6

About 16 elements are required by cells, and are therefore essential for life. Consequently, the full list of essential elements is a relatively short one. Furthermore, about 99% of living matter consists of just four elements: **carbon**, **hydrogen**, **oxygen** and **nitrogen**.

Why do these four elements predominate in living things?

The elements carbon, hydrogen and oxygen predominate because living things contain large quantities of **water**, and also because most other molecules present in cells and organisms are compounds of carbon combined with hydrogen and oxygen, including the **carbohydrates** and **lipids**. The element nitrogen is combined with carbon, hydrogen and oxygen in compounds called amino acids from which **proteins** are constructed.

The roles of a selection of the other **essential elements** are listed in Table 2.2. Note that several of these elements are required in tiny, 'trace' amounts only; in fact, more of some would actually be very harmful to the body.

Element	In plants	In animals	In prokaryotes
sulphur	■ in some amino acids and proteins ■ in some vitamins	■ in some amino acids and proteins ■ in some vitamins	■ in some amino acids and proteins ■ in some vitamins
calcium	■ cell wall formation between dividing plant cells ■ co-factor for certain enzymes	■ constituent of bones ■ reacts in muscle fibre contraction, blood clotting and synapses ■ co-factor for certain enzymes	■ co-factor for certain enzymes ■ contributes to heat resistance of bacterial endospores
phosphorus	■ synthesis of nucleotides ■ ATP	■ synthesis of nucleotides ■ ATP ■ constituent of bones	■ synthesis of nucleotides ■ ATP
iron	■ constituent of electron transport molecules (e.g. cytochromes) ■ chlorophyll synthesis	■ constituent of electron transport molecules (e.g. cytochromes and haem – part of haemoglobin)	■ constituent of electron transport molecules ■ chlorophyll synthesis in photosynthetic prokaryotes
sodium	■ involved with **potassium** in membrane function	■ involved with **potassium** in membrane function and nerve impulse transport	■ involved with **potassium** in membrane function

Other essential elements include:
- **magnesium** (in chlorophyll – in green plants and certain prokaryotes);
- **manganese**, **copper**, **cobalt**, **zinc**, **molybdenum** (all in trace amounts, in certain enzymes).

Table 2.2 Roles of certain other elements in living organisms

The difference between atoms, molecules and ions

The fundamental unit of chemical structure is the atom.

An atom is the smallest part of an element that can take part in a chemical change.

Atoms group together to form molecules, and molecules are the smallest part of certain elements and compounds that can exist alone under normal conditions. For example, both oxygen and nitrogen naturally combine with another atom of the same type to form a molecule (O_2 and N_2).

1 Distinguish between the terms 'atom' and 'ion'.

Alternatively, if an atom gains or loses an electron an ion is formed.

Ions are formed when atoms gain or lose electrons to form positively or negatively charged ions.

Water

Living things are typically solid, substantial objects, yet water forms the bulk of their structures – between 65% and 95% by mass of most multicellular plants and animals (about 80% of a human cell consists of water). Despite this, and the fact that water has some unusual properties, water is a substance that is often taken for granted.

Water is composed of atoms of the elements hydrogen and oxygen. One atom of oxygen and two atoms of hydrogen combine by sharing electrons (**covalent bonding**). However, the water molecule is triangular rather than linear, and the nucleus of the oxygen atom draws electrons (negatively charged) away from the hydrogen nuclei (positively charged) – with an interesting consequence. Although overall the water molecule is electrically neutral, there is a net negative charge on the oxygen atom and a net positive charge on the hydrogen atoms. In other words, the water molecule carries an unequal distribution of electrical charge within it. This arrangement is known as a **polar molecule** (Figure 2.1).

2 Distinguish between ionic and covalent bonding.

Hydrogen bonds

With water molecules, the positively charged hydrogen atoms of one molecule are attracted to negatively charged oxygen atoms of nearby water molecules by forces called **hydrogen bonds**. These are weak bonds compared to covalent bonds, yet they are strong enough to hold water molecules together and attract water molecules to charged particles or a charged surface. In fact, hydrogen bonds largely account for the unique properties of water. We examine these properties next.

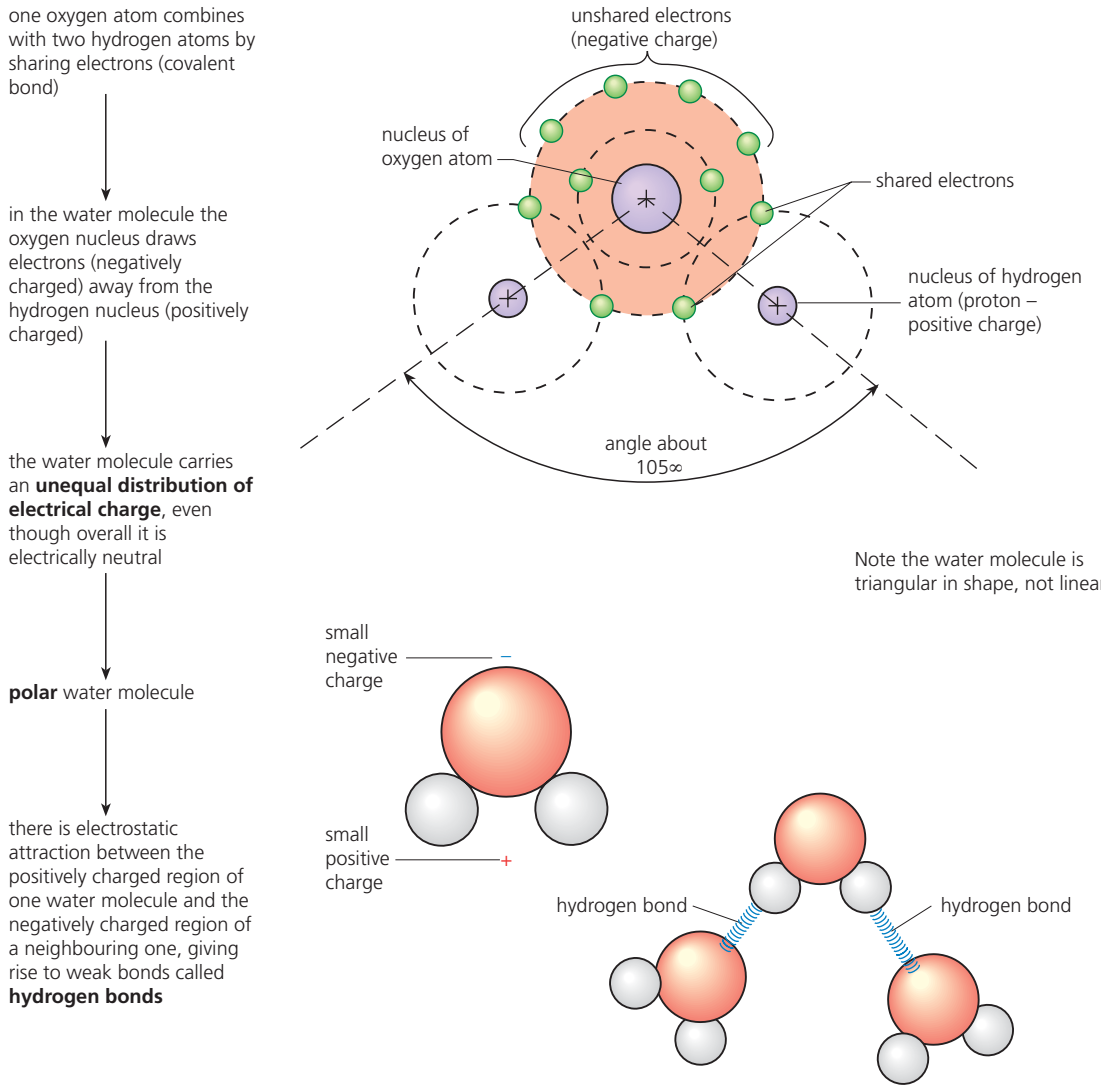


Figure 2.1 The water molecule and the hydrogen bonds it forms

Thermal properties of water

Heat energy and the temperature of water

A lot of heat energy is required to raise the temperature of water. This is because much energy is needed to break the hydrogen bonds that restrict the movements of water molecules. This property of water is its **specific heat capacity**. The specific heat capacity of water is the highest of any known substance. Consequently, aquatic environments like streams and rivers, ponds, lakes and seas are very slow to change temperature when the surrounding air temperature changes. Aquatic environments have much more stable temperatures than terrestrial (land) environments do.

Another consequence is that cells and the bodies of organism do not change temperature readily. Bulky organisms, particularly, tend to have a stable temperature in the face of a fluctuating surrounding temperature, whether in extremes of heat or cold.

Evaporation and heat loss

The hydrogen bonds between water molecules make it difficult for them to be separated and vaporised (evaporated). This means that much energy is needed to turn liquid water into water vapour (gas). This amount of energy is the **latent heat of vaporisation**, and for water it is very high. Consequently, the evaporation of water in sweat on the skin, or in transpiration from green leaves, causes marked cooling. The escaping molecules take a lot of energy with them. You experience this when you stand in a draught after a shower. And since a great deal of heat is lost with the evaporation of a small amount of water, cooling by evaporation of water is economical on water, too.

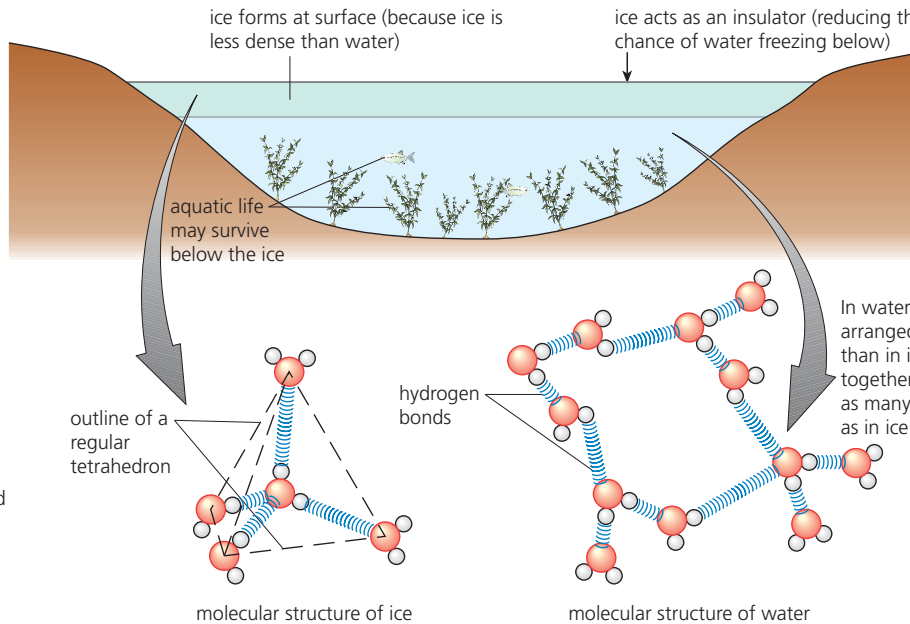
Heat energy and freezing

The amount of heat energy that must be removed from water to turn it to ice is very great, as is that needed to melt ice. This amount of energy is the **latent heat of fusion**, and is very high for water. As a result, both the contents of cells and the water in the environment are always slow to freeze in extreme cold.

The density of ice

Most liquids contract on cooling, reaching maximum density at their freezing point. Water is unusual in reaching its maximum density at 4 °C. So as water freezes, the ice formed is less dense than the cold water around it. As a consequence, ice floats on top of very cold water (Figure 2.2). The floating layer of ice insulates the water below. The consequence is that lakes rarely freeze solid; aquatic life can generally survive a freeze-up.

Figure 2.2 Ice forms on the surface of water



Because ice freezes from below, it is difficult to judge its depth. It can therefore be very dangerous to walk on!

The angle between the covalent bonds in the water molecule is very close to the angles of a perfect tetrahedron. In ice the molecules form a regular tetrahedral arrangement, and are spaced more widely apart than they are in liquid water.

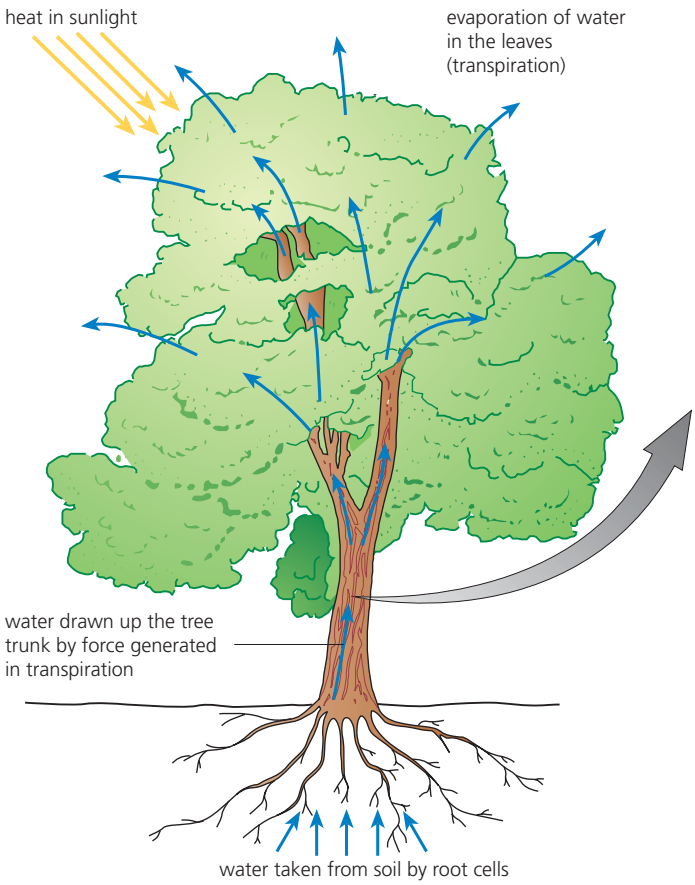
Cohesive properties of water

Cohesion is the force by which individual molecules stick together. Water molecules stick together as a result of hydrogen bonding. These bonds continually break and reform with other, surrounding water molecules, but at any one moment, a large number are held together by their hydrogen bonds. **Adhesion** is the force by which individual molecules cling to surrounding material and surfaces. Materials with an affinity for water are described as **hydrophilic** (see 'Solvent properties of water' on page 42). Water adheres strongly to most surfaces and can be drawn up long columns, through narrow tubes like the xylem vessels of plant stems, without danger of the water column breaking (Figure 2.3). Compared with other liquids, water has extremely strong adhesive and cohesive properties that prevent it breaking under tension.

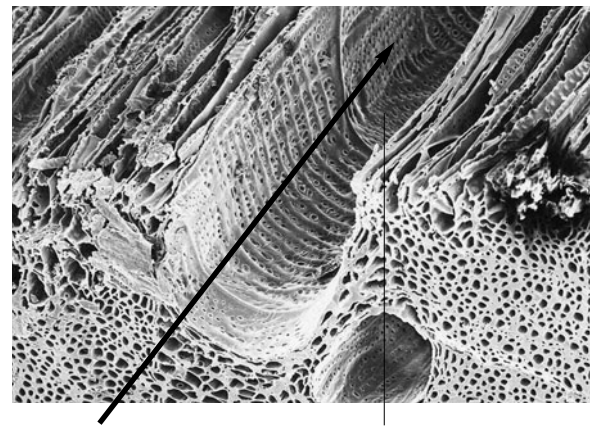
Related to the property of cohesion is the property of **surface tension**. The outermost molecules of water form hydrogen bonds with water molecules below them. This gives a very high surface tension to water – higher than any other liquid except mercury. The surface tension of water is exploited by insects that 'surface skate' (Figure 2.4). The insect's waxy cuticle prevents wetting of its body, and the mass of the insect is not great enough to break the surface tension.

Below the surface, water molecules slide past each other very easily. This property is described as low **viscosity**. Consequently, water flows readily through narrow capillaries and tiny gaps and pores.

Figure 2.3 Water is drawn up a tree trunk



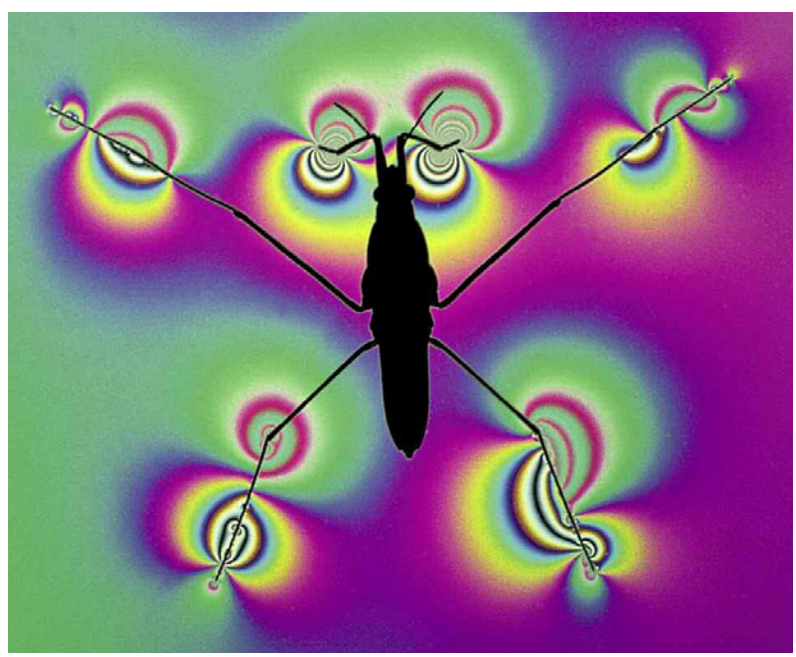
water can be drawn up to a great height without the column breaking or pulling apart



the column of water coheres (does not break), adheres to the walls of the xylem vessels and flows smoothly through them (because its viscosity is low)

xylem vessels run from roots to leaves, as continuous narrow tubes

Figure 2.4 A pond skater moving over the water surface



the photograph shows the interference patterns created by the water's surface tension

Solvent properties of water

Water is a powerful solvent for polar substances. These include the following:

- Ionic substances like sodium chloride (Na^+ and Cl^-, page 695). All cations (positively charged ions) and anions (negatively charged ions) become surrounded by a shell of orientated water molecules (Figure 2.5).
- Carbon-containing (organic) molecules with ionised groups (such as the carboxyl group $-COO^-$, and amino group $-NH_3^+$). Soluble organic molecules like sugars dissolve in water due to the formation of hydrogen bonds with their slightly charged hydroxyl groups (–OH).

Once they have dissolved, molecules (the **solute**) are free to move around in water (the **solvent**), and as a result, are more chemically reactive than when in the undissolved solid.

On the other hand, non-polar substances are repelled by water, as in the case of oil on the surface of water. Non-polar substances are **hydrophobic** (water-hating).

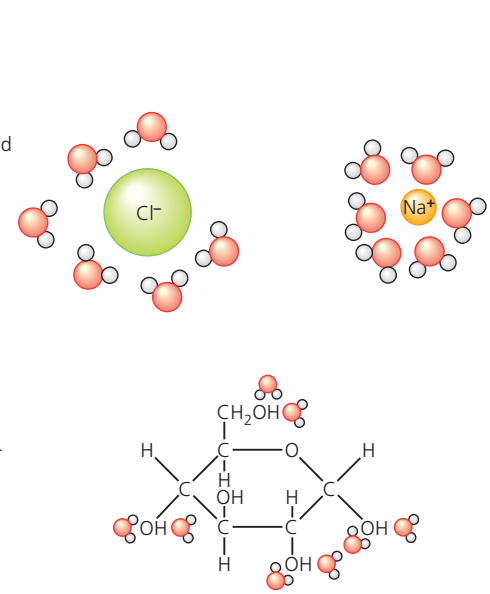


Figure 2.5 Water as universal solvent

3 In an aqueous solution of glucose, **state** which component is the solvent and which the solute.

Extension: Solvent properties arise because water is not a gas

Water has a relative molecular mass of only 18, yet it is a liquid at room temperature. This is surprising; it contrasts with other small molecules, which are **gases**. For example, methane (CH_4) = 16; ammonia (NH_3) = 17; and carbon dioxide (CO_2) = 44. But **none of these latter molecules contains hydrogen bonds**.

In gases, molecules are widely spaced and free to move about independently. In liquids, molecules are closer together. In the case of water, hydrogen bonds pull the molecules very close to each other, which is why water is a liquid at the temperatures and pressure that exist over much of the Earth's surface. As a result, we have a liquid medium which life exploits (Table 2.3).

Water is transparent

Because light penetrates water, aquatic photosynthetic plants can live at some depth (Figure 2.6). These plants support aquatic food chains, and therefore animal life. Light aids sight in aquatic animals.

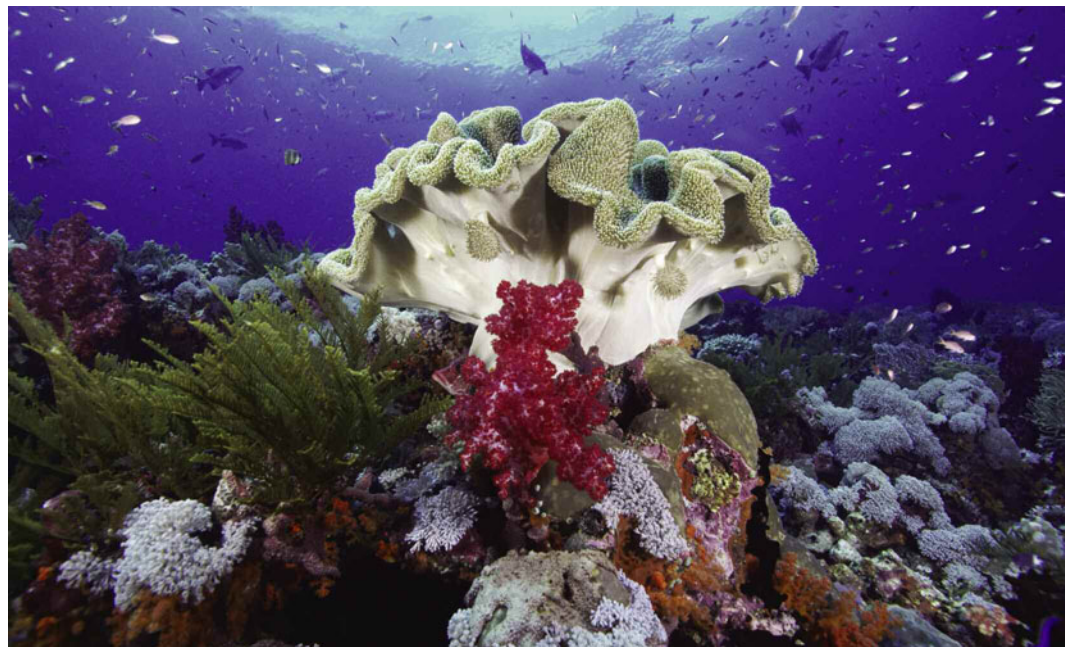


Figure 2.6 Seaweed can survive at depth

In warm, shallow seas, the plant life includes attached seaweeds and an abundance of tiny, unicellular, floating phytoplankton. These support animal life, including shoals of fish, and many non-vertebrates such as sessile corals like the large Toadstool Leather Coral seen here, in the Banda Sea, Indonesia.

Property	Benefit to life
1 a liquid at room temperature, water dissolves more substances than any other common liquid	liquid medium for living things and for the chemistry of life
2 much heat energy needed to raise the temperature of water	aquatic environment slow to change temperature bulky organisms have stable temperatures
3 evaporation requires a great deal of heat	evaporation causes marked cooling (much heat is lost by evaporation of a small quantity of water)
4 much heat has to be removed before freezing occurs	cell contents and water in aquatic environments are slow to freeze in cold weather
5 ice is at maximum density at 4 °C	ice forms on the surface of water, insulating the water below and allowing much aquatic life to survive freezing
6 surface water molecules orientate with hydrogen bonds formed inwards	water forms droplets and rolls off surfaces certain animals exploit surface tension to move over water surface
7 water molecules slide past each other easily (low viscosity)	water flows easily through narrow capillaries, and through tiny spaces (e.g. in soils, spaces in cell walls)
8 water molecules adhere to surfaces	water adheres to walls of xylem vessels as it is drawn up the stem to the leaves, from the roots
9 water column does not break or pull apart under tension	water can be lifted by forces applied at the top, and so can be drawn up xylem vessels of tree trunks by forces generated in the leaves
10 water is transparent	aquatic plants can photosynthesise at some depth in water

Table 2.3 Water and life – summary

4 Explain the significance to organisms (plants and animals) of water as a coolant and transport medium, in terms of it properties.

Carbohydrates, lipids and proteins

3.2.1–3.2.7

Carbohydrates, lipids and proteins are examples of compounds of carbon. Carbon is a relatively uncommon element of the Earth's crust, but in cells and organisms it is the second most abundant element by mass, after oxygen.

Compounds containing carbon that are found in living organisms are called organic compounds. The exceptions are the gas carbon dioxide (CO_2), hydrogen carbonates (the product of CO_2 when dissolved in water), and mineral calcium carbonate ($CaCO_3$), which are classified as non-organic carbon.

Carbon has remarkable properties. It is a relatively small atom, and is able to form four strong, stable, covalent bonds. Also, carbon atoms are able to react with each other to form extended chains. These carbon 'skeletons' may be straight chains, branched chains, or rings. Carbon also bonds covalently with other atoms, such as oxygen, hydrogen, nitrogen and sulphur, forming different groups of organic molecules with distinctive properties. So, there are vast numbers of organic compounds – more than the total of known compounds made from other elements, in fact. Biologists think the diversity of organic compounds has made possible the diversity of life.

5 Suggest where non-organic forms of carbon exist in the biosphere.

Fortunately, very many of the organic chemicals of living things fall into one of four discrete groups or 'families' of chemicals with many common properties. These families are the carbohydrates, lipids, proteins, and nucleic acids.

Carbohydrates

Carbohydrates are the largest group of organic compounds found in living things. They include sugars, starch, glycogen and cellulose. Carbohydrates contain only three elements: carbon, hydrogen and oxygen, with hydrogen and oxygen always present in the ratio 2:1 (as they are in water – H_2O), so they can be represented by the general formula $C_x(H_2O)_y$. Table 2.4 is a summary of the three types of carbohydrates commonly found in living things.

Carbohydrates – general formula $C_x(H_2O)_y$

Monosaccharides	Disaccharides	Polysaccharides
■ simple sugars (e.g. glucose and fructose – 6 carbon atoms; ribose – 5 carbon atoms)	■ two simple sugars condensed together (e.g. sucrose, lactose, maltose)	■ very many simple sugars condensed together (e.g. starch, glycogen, cellulose)

Table 2.4 Carbohydrates of cells and organisms

Monosaccharides – the simple sugars

Monosaccharides are carbohydrates with relatively small molecules. They taste sweet and are soluble in water. In biology, **glucose** is an especially important monosaccharide because:

- all green leaves manufacture glucose using light energy;
- our bodies transport glucose in the blood;
- all cells use glucose in respiration – we call it one of the respiratory substrates;
- in cells and organisms, glucose is the building block for very many larger molecules.

The structure of glucose

Glucose has the chemical or **molecular formula $C_6H_{12}O_6$**. This type of formula tells us what the component atoms are, and the numbers of each in the molecule. For example, glucose is a 6-carbon sugar, or **hexose**. But the molecular formula does not tell us the structure of the molecule.

Glucose can be written on paper as a linear molecule but it cannot exist in this form (because each carbon arranges its four bonds into a tetrahedron, so the molecule cannot be 'flat'). Rather, glucose is folded, taking a ring or cyclic form. Figure 2.7 shows the structural formula of glucose.

6 Suggest why simple sugars like glucose are not commonly found as a storage form of carbohydrate in cells or tissues.

The carbon atoms of an organic molecule may be numbered. This allows us to identify which atoms are affected when the molecule reacts and changes shape. For example, as the glucose ring forms, the oxygen on carbon-5 attaches itself to carbon-1. Note that the glucose ring contains five carbon atoms and an oxygen atom (Figure 2.7).

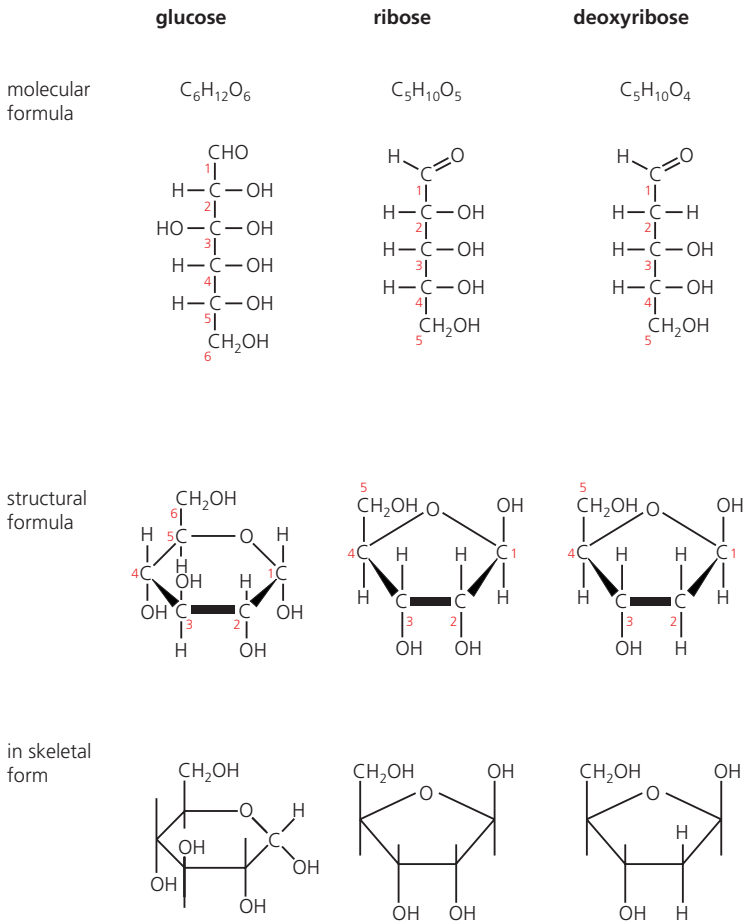


Figure 2.7 Structural formulae of some monosaccharides

Other monosaccharides of importance in living cells

Glucose, fructose and galactose are examples of hexose sugars commonly occurring in cells and organisms.

Other monosaccharide sugars produced by cells and used in metabolism include:

- 3-carbon sugars (**trioses**), early products in photosynthesis (page 284);
- 5-carbon sugars (**pentoses**), namely ribose and deoxyribose.

The pentoses ribose and deoxyribose are components of the nucleic acids (page 63). The structures of both these pentoses are also shown in Figure 2.7.

Disaccharides

Disaccharides are carbohydrates made of two monosaccharides combined together. For example, sucrose is formed from a molecule of glucose and a molecule of fructose chemically combined together.

Condensation and hydrolysis reactions

When two monosaccharide molecules are combined to form a disaccharide, a molecule of water is also formed as a product, and so this type of reaction is known as a **condensation reaction**. The linkage between monosaccharide residues, after the removal of H—O—H between them, is called a **glycosidic linkage** (Figure 2.8). This comprises strong, covalent bonds. The condensation reaction is brought about by an enzyme.

In the reverse process, disaccharides are 'digested' to their component monosaccharides in a hydrolysis reaction. This reaction involves adding a molecule of water (hydro-) as splitting (-lysis) of the glycosidic linkage occurs. It is catalysed by an enzyme, too, but it is a different enzyme from the one that brings about the condensation reaction.

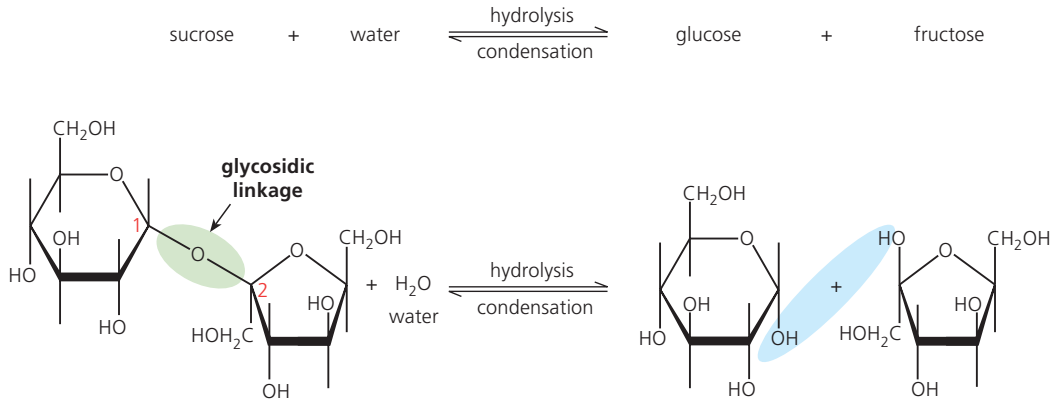


Figure 2.8 Disaccharides, and the monosaccharides that form them

This structural formula shows us how the glycosidic linkage forms/breaks, but the structural formulae of disaccharides should not be memorised.

maltose + water ⇌ glucose + glucose (hydrolysis / condensation)

lactose + water ⇌ galactose + glucose (hydrolysis / condensation)

Apart from sucrose, other disaccharide sugars produced by cells and used in metabolism include:

- **maltose**, formed by condensation reaction of two molecules of glucose;
- **lactose**, formed by condensation reaction of galactose and glucose.

Polysaccharides

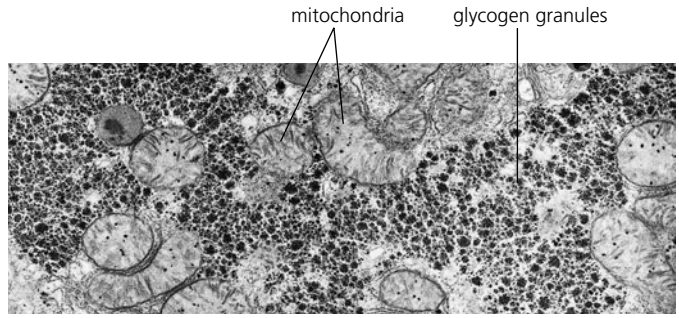
Polysaccharides are built from very many monosaccharide residues condensed together, all linked by glycosidic linkages. Poly- means 'many', and in fact thousands of monosaccharide residues make up a polysaccharide. They are examples of molecules known as polymers – that is, huge molecules normally made of only one type of repeated monomer residue.

Examples of polysaccharides produced by cells and used in metabolism include:

- **starch**, which is a polymer of glucose, and is in fact a mixture of two polysaccharides: amylose (an unbranched chain of glucose residues) and amylopectin (which has branches at points along its chain) – each chain in starch takes the form of a helix, and the whole molecule is stabilised by countless hydrogen bonds;
- **glycogen**, which is a polymer of glucose chemically similar to amylopectin, although larger and more highly branched (Figure 2.9);
- **cellulose**, which is a giant polymer also made of repeating glucose units, arranged in straight, unbranched, parallel chains held together in fibres by numerous hydrogen bonds (Figure 1.14, page 16).

Examples of some important carbohydrates are given in Table 2.5.

7 Define the terms 'monomer' and 'polymer', giving examples from the carbohydrates.



TEM of a liver cell (×7000)

Figure 2.9 Glycogen granules in a liver cell

Animals		**Plants**	
Monosaccharides		**Monosaccharides**	
glucose	■ transported to cells in the blood plasma ■ used as a respiratory substrate for cellular respiration or converted to glycogen (storage carbohydrate, see below)	glucose	■ a first product of photosynthesis
galactose	■ used in the production of lactose (milk sugar)	fructose	■ produced in cellular respiration as an intermediate of glucose breakdown ■ used in the production of sucrose
Disaccharides		**Disaccharides**	
lactose	■ produced in mammary glands and secreted into the milk as an important component in the diet of very young mammals	sucrose	■ produced in green leaves from glucose and fructose ■ transported in plant in solution, in the vascular bundles
		maltose	■ breakdown product in the hydrolysis of starch
Polysaccharides		**Polysaccharides**	
glycogen	■ storage carbohydrate formed from glucose in the liver and other cells (but not in brain cells) when glucose is not immediately required for cellular respiration (Figure 2.9)	cellulose	■ manufactured in cells and laid down externally, in bundles of fibres, as the main component of the cell walls
		starch	■ storage carbohydrate

Table 2.5 Examples of carbohydrates important in animals and plants

Lipids

Lipids occur in living things as animal **fats** and plant **oils**, and also as the **phospholipids** of cell membranes, for example. Fats and oils seem rather different substances, but their only difference is that at about 20 °C (room temperature) oils are liquid and fats are solid.

Lipids contain the elements carbon, hydrogen and oxygen, as do carbohydrates, but in lipids the proportion of oxygen is much less.

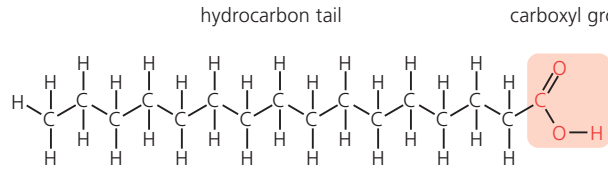
Lipids are insoluble in water. In fact they generally behave as 'water-hating' molecules, a property described as **hydrophobic**. However, lipids can be dissolved in organic solvents such as alcohol (e.g. ethanol) and propanone (acetone).

Fats and oils

Fats and oils are compounds called **triglycerides**. They are formed by reactions in which water is removed (another case of a condensation reaction) between fatty acids and an alcohol called glycerol. Here the bond formed is known as an **ester linkage**. The structure of a fatty acid commonly found in cells, and of glycerol, is shown in Figure 2.10, and the steps of triglyceride formation in Figure 2.11. In cells, enzymes catalyse the formation of triglycerides, and also the breakdown of glycerides by hydrolysis.

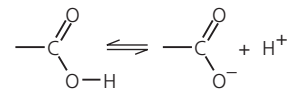
Figure 2.10 Fatty acids and glycerol, the building blocks

Fatty acid



this is palmitic acid with 16 carbon atoms

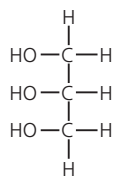
the carboxyl group ionises to form hydrogen ions, i.e. it is a weak acid



molecular formula of palmitic acid

$CH_3(CH_2)_{14}COOH$

Glycerol



molecular formula of glycerol

$C_3H_5(OH)_3$

The fatty acids combined in fats and oils have long hydrocarbon 'tails', typically of about 16–18 carbon atoms, but may be anything between 14 and 22. The hydrophobic properties of triglycerides are due to these hydrocarbon tails. A molecule of triglyceride is quite large, but relatively small when compared to polymer macromolecules such as starch and cellulose. It is because of their hydrophobic properties that triglyceride molecules clump together (aggregate) into huge globules, making them appear to be macromolecules.

8 Distinguish between condensation and hydrolysis reactions. Give an example of each.

The roles of fats and oils in living things

1 Energy store and metabolic water source

When triglycerides are oxidised in respiration a lot of energy is transferred (and used to make, for example, ATP – page 77). Mass for mass, fats and oils transfer more than twice as much energy as carbohydrates do, when they are respired (Table 2.6). This is because fats are comparatively low in oxygen atoms (the carbon of lipids is more reduced than that of carbohydrates), so more of the oxygen in the respiration of fats comes from the atmosphere. In the oxidation of carbohydrate, more oxygen is present in the carbohydrate molecule itself. Fat and oils, therefore, form a concentrated energy store.

They are insoluble, too, so the presence of fat or oil in cells does not cause osmotic water uptake there. A fat store is especially typical of animals that endure long unfavourable seasons in which they survive on reserves of food stored in the body. Oils are often a major energy store in plants, their seeds and fruits, and it is common for fruits and seeds – including maize, olives and sunflower seeds – to be used commercially as a source of edible oils for humans.

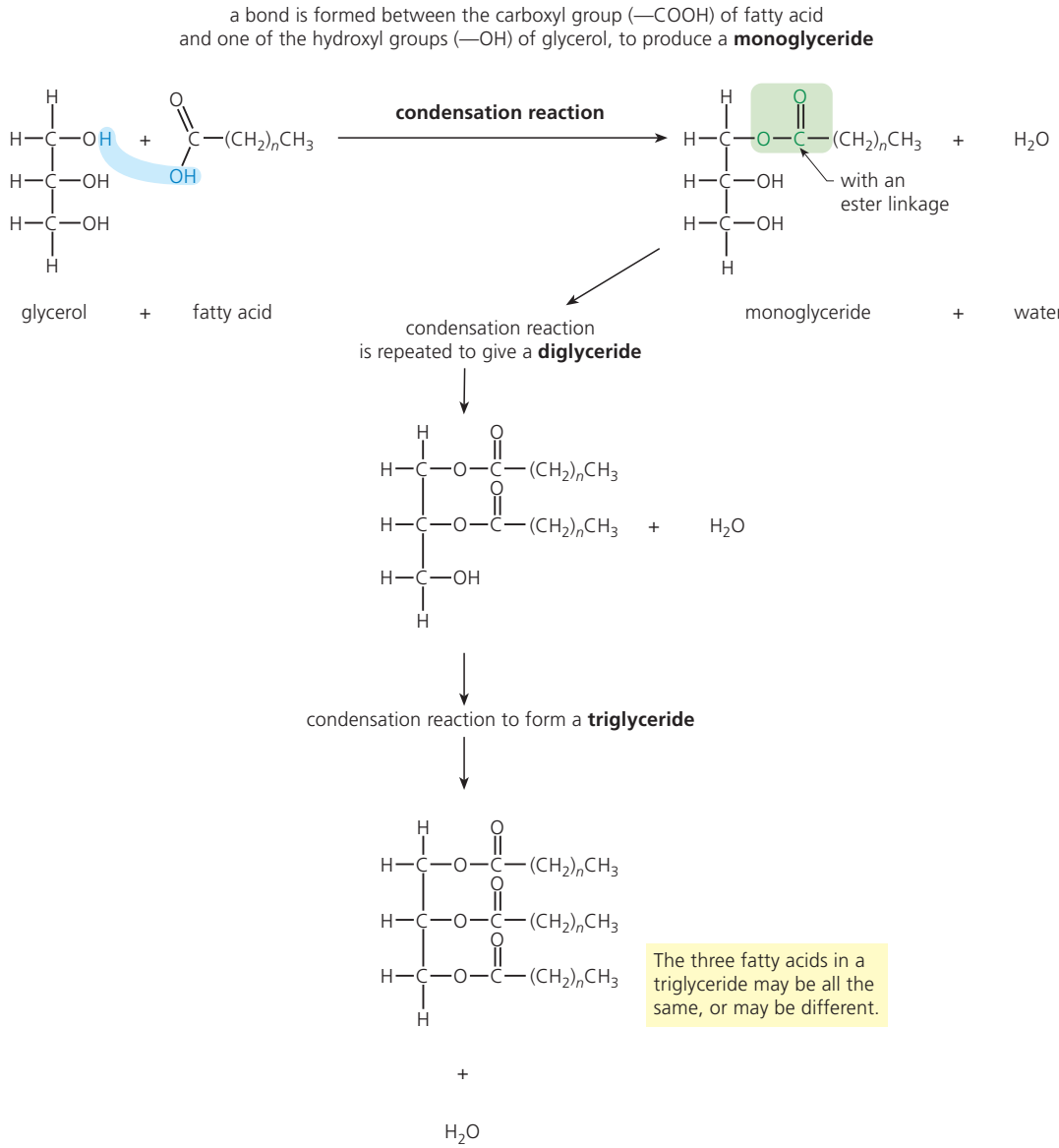


Figure 2.11 Formation of triglyceride

Complete oxidation of fats and oils produces a large amount of water, far more than when the same mass of carbohydrate is respired. Desert animals like the camel and the desert rat retain much of this metabolic water within the body, helping them survive when there is no liquid water for drinking. The development of the embryos of birds and reptiles, while in their shells, also benefits from metabolic water formed by the oxidation of the stored fat in their egg's yolk.

Lipids	Role	Carbohydrates
more energy per gram than from carbohydrates	**energy store**	less energy per gram than from lipids
much metabolic water is produced on oxidation	**metabolic water source**	less metabolic water is produced on oxidation
insoluble, so osmotic water uptake is not caused	**solubility**	sugars are highly soluble in water, causing osmotic water uptake
not quickly 'digested'	**ease of breakdown**	more easily hydrolysed – energy transferred quickly

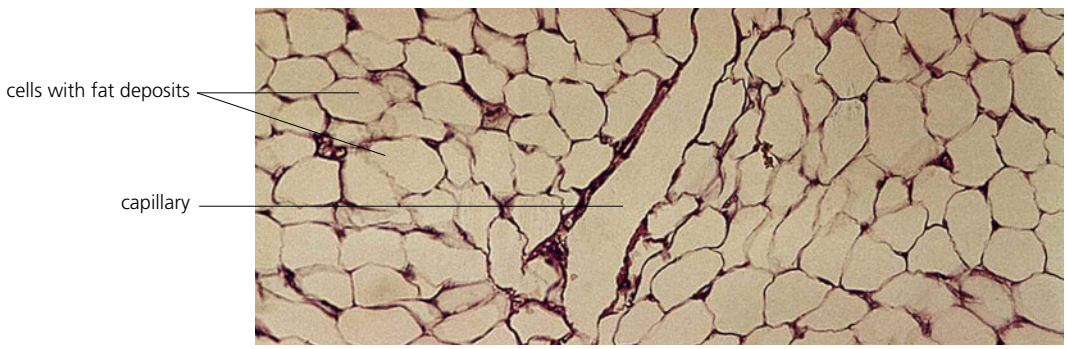
Table 2.6 Lipids and carbohydrates as energy stores – a comparison

2 Subcutaneous fat as a buoyancy aid (and thermal insulation)

Fat is stored in animals as adipose tissue, typically under the skin, where it is known as subcutaneous fat (Figure 2.12). In aquatic diving mammals, which have a great deal of subcutaneous fat, it is identified as blubber. Blubber undoubtedly gives buoyancy to the body since fat is not as dense as muscle or bone.

If fat reserves like these have a restricted blood supply (as is normally the case), then the heat of the body is not especially distributed to the fat under the skin. In these circumstances the subcutaneous fat also functions as a heat insulation layer.

Figure 2.12 Adipose tissue



3 Water-proofing of hair and feathers

Oily secretions of the sebaceous glands, found in the skin of mammals, act as a water repellent, preventing fur and hair from becoming waterlogged when wet. Birds have a preen gland that fulfils the same function for feathers.

4 Electrical insulation

Myelin lipid in the membranes of Schwann cells, forming the sheaths around the long fibres of nerve cells (page 211), electrically isolates the cell plasma membrane and facilitates the conduction of the nerve impulse there.

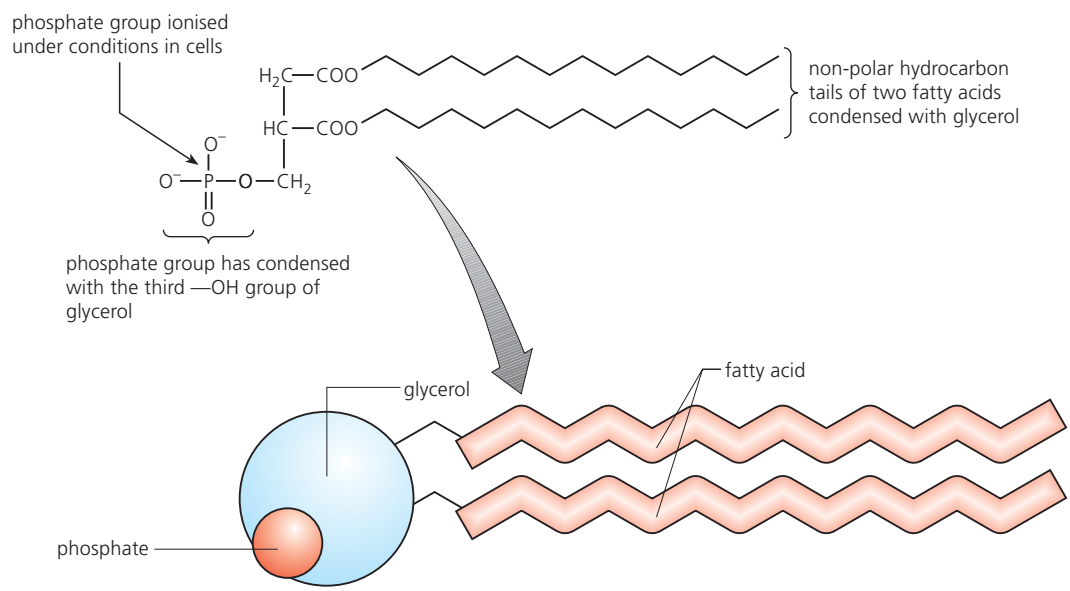
Phospholipid

9 Describe the property given to a lipid when it combines with a phosphate group.

A phospholipid has a similar chemical structure to triglyceride; here, one of the fatty acid groups is replaced by phosphate (Figure 2.13). This phosphate is ionised and therefore water soluble. So phospholipids combine the hydrophobic properties of the hydrocarbon tails with hydrophilic properties of the phosphate.

Phospholipid molecules form monolayers and bilayers in water (Figure 1.20, page 22). A phospholipid bilayer is a major component of the plasma membrane of cells.

Figure 2.13 Phospholipid



Amino acids, peptides and proteins

Proteins make up about two-thirds of the total dry mass of a cell. They differ from carbohydrates and lipids in that they contain the element nitrogen, and usually the element sulphur, as well as carbon, hydrogen and oxygen.

Amino acids are the molecules from which peptides and proteins are built – typically several hundred or even thousands of amino acid molecules are combined together to form a protein. The terms 'polypeptide' and 'protein' can be used interchangeably, but when a polypeptide is about 50 amino acid residues long it is generally agreed to have become a protein.

Once the chain is constructed, a protein takes up a specific shape. Shape matters with proteins – their shape is closely related to their function. This is especially the case in proteins that are enzymes, as we shall shortly see.

Amino acids – the building blocks of peptides

Figure 2.14 shows the structure of an amino acid. As their name implies, amino acids carry two groups:

- an **amino group** (—NH_2);
- an **organic acid group** (carboxyl group —COOH).

These groups are attached to the same carbon atom in the amino acids which get built up into proteins. Also attached here is a side-chain part of the molecule, called an **R group**.

Proteins of living things are built from just **20 different amino acids**, in differing proportions. All we need to note is that the R groups of these amino acids are all very different and consequently amino acids (and proteins containing them) have different chemical characteristics.

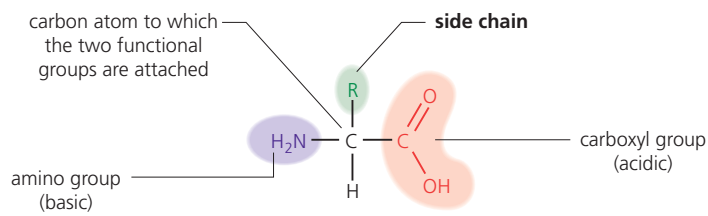
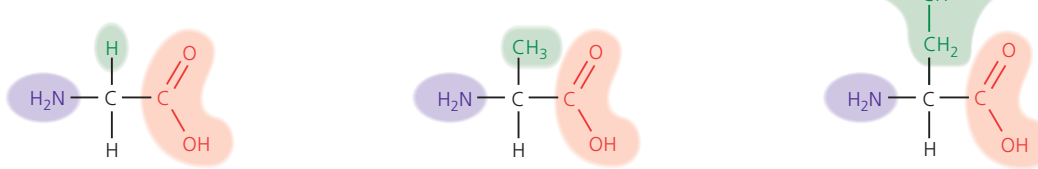


Figure 2.14 The structure of amino acids

The 20 different amino acids that make up proteins in cells and organisms differ in their side chains. Below are three illustrations but *details of R groups are not required*.



Some amino acids have an additional —COOH group in their side chain (= acidic amino acids).
Some amino acids have an additional —NH_2 group in their side chain (= basic amino acids).

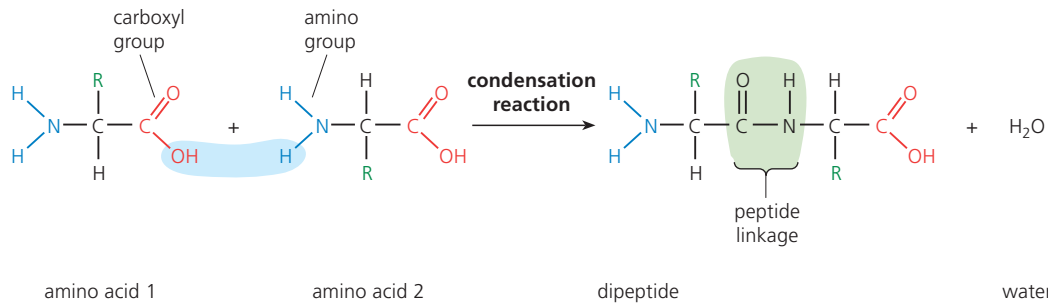
Peptide linkages

Two amino acids combine together with the loss of water to form a **dipeptide**. This is one more example of a condensation reaction. The amino group of one amino acid reacts with the carboxyl group of the other, forming a **peptide linkage** (Figure 2.15).

A further condensation reaction between the dipeptide and another amino acid results in a tripeptide. In this way, long strings of **amino acid residues**, joined by peptide linkages, are formed. Thus, peptides or protein chains are assembled, one amino acid at a time, in the presence of a specific enzyme (page 72).

Figure 2.15 Peptide linkage formation

amino acids combine together, the amino group of one with the carboxyl group of the other



When a further amino acid residue is attached by condensation reaction, a tripeptide is formed. In this way, long strings of amino acid residues are assembled to form polypeptides and proteins.

10 The possible number of different polypeptides, P, that can be assembled is given by $P = A^n$ where A = the number of different types of amino acids available, and n = the number of amino acid residues in the polypeptide molecule.

Given the naturally occurring pool of 20 different amino acids, **calculate** how many different polypeptides are possible if constructed from 5, 25 and 50 amino acid residues, respectively.

Enzymes – biological catalysts

3.6.1–3.6.5

Most chemical reactions do not occur spontaneously. In a laboratory or in an industrial process, chemical reactions may be made to occur by applying high temperatures, high pressures, extremes of pH, and by maintaining high concentrations of the reacting molecules. If these drastic conditions were not applied, very little of the chemical product would be formed. On the other hand, in cells and organisms, many chemical reactions occur simultaneously, at extremely low concentrations, at normal temperatures, and under the very mild, almost neutral, aqueous conditions we find in cells.

How is this brought about?

It is the presence of enzymes in cells and organisms that enables these reactions to occur at incredible speeds, in an orderly manner, yielding products that the organism requires, when they are needed. Sometimes, reactions happen even though the reacting molecules are present in very low concentrations. Enzymes are truly remarkable molecules.

Enzymes are biological catalysts made of protein.

By 'catalyst' we mean a substance that **speeds up the rate of a chemical reaction**. In general, catalysts:

- are **effective in small amounts**;
- remain **unchanged at the end of the reaction**.

Another important property of enzymes and catalysts is that they speed up the rate at which an equilibrium position is reached. In a reversible reaction:

$$A + B \rightleftharpoons C + D$$

A and B react to form C and D. This is a reversible reaction as shown by the sign $\rightleftharpoons$. As soon as C and D start to accumulate, some will react to re-form A and B. The reversible reaction reaches an equilibrium point when the rate of the forward reaction equals the rate of the reverse reaction. Most enzyme-catalysed reactions are reversible; the presence of the enzyme for these reactions means the equilibrium position is reached quickly.

The enzyme active site

In a reaction catalysed by an enzyme, the starting substance is called the **substrate**, and what it is converted to is the **product**. An enzyme works by binding to the substrate molecule at a specially formed pocket in the enzyme. This binding point is called the **active site**.

An active site is a region of an enzyme molecule where the substrate molecule binds.

As the enzyme and substrate form a complex, the substrate is raised to a transition state and instantly breaks down to products which are released, together with unchanged enzyme (Figure 2.16). The enzyme is available for use again.

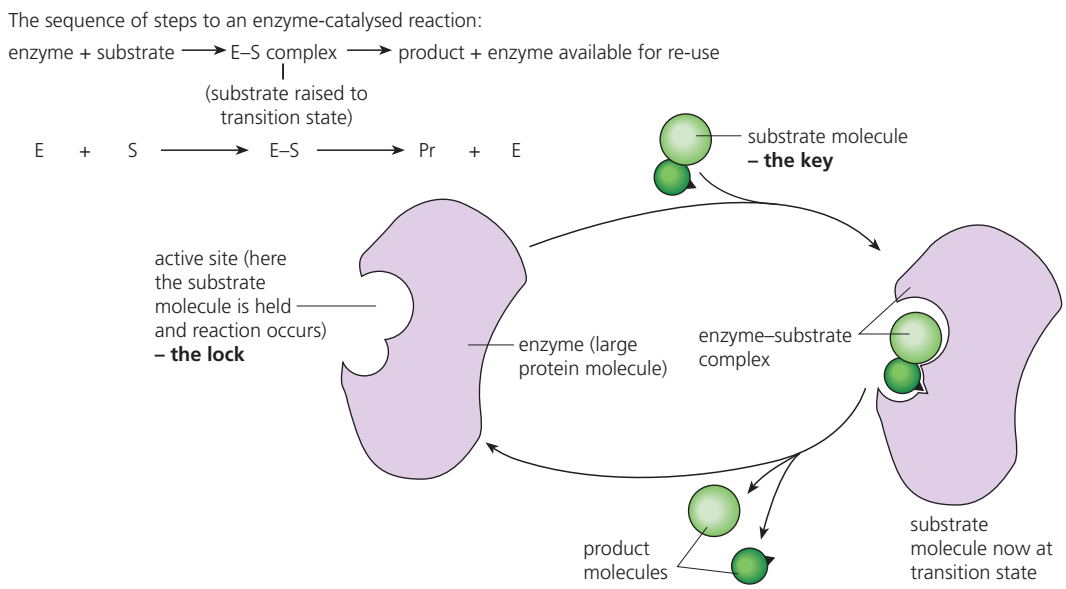


Figure 2.16 The enzyme–substrate complex and active site

Enzymes are typically large globular protein molecules. Most substrate molecules are quite small molecules by comparison. Even when the substrate molecules are very large, such as certain macromolecules like the polysaccharides, only one bond in the substrate is in contact with the enzyme active site. The active site takes up a relatively small part of the total volume of the enzyme.

The active site and enzyme specificity

Enzymes are highly specific in their action – they catalyse only one type of reaction or only a very small group of highly similar reactions. This means that enzymes 'recognise' a very small group of substrate molecules or even only a single type of molecule. This is because the active site where the substrate molecule binds has a **precise shape** and **distinctive chemical properties** (meaning the presence of particular chemical groups and bonds). Only particular substrate molecules are attracted to a particular active site and can fit there. All other substrate molecules are unable to fit and so cannot bind.

To emphasise this, the binding of enzyme and substrate is referred to as the **lock and key hypothesis** of enzyme action. The enzyme is the 'lock', the substrate is a 'key' that fits it, and the point of link-up is the active site on the enzyme, as shown in Figure 2.16.

Enzymes control metabolism

There is a huge array of chemical reactions that go on in cells and organisms – collectively the chemical reactions of life are called **metabolism**.

Since the reactions of metabolism can only occur in the presence of a specific enzyme, we know that if an enzyme is not present then the reaction it catalyses cannot occur. Many enzymes are always present in cells and organisms, but some enzymes are produced only under particular conditions or at certain stages. By making some enzymes and not others, cells can control what chemical reactions happen in the cytoplasm. Sometimes it is the presence of the substrate molecule that triggers the synthesis machinery by which the enzyme is produced. Very many of the proteins made by cells are enzymes.

Later in this chapter we shall see how protein synthesis (and therefore enzyme production) is controlled by the cell nucleus.

Where do enzymes operate?

Some enzymes are exported from cells, such as the digestive enzymes (page 180). Enzymes like these, that are parcelled up and secreted and work externally, are called **extracellular enzymes**. However, very many enzymes remain within cells and work there. These are the **intracellular enzymes**. They are found inside organelles, and in the membranes of organelles, in the fluid medium around the organelles, known as the cytosol, and in the plasma membrane.

11 **Explain** why the shape of globular proteins that are enzymes is important in enzyme action.

Factors affecting the rate of enzyme-catalysed reactions

Reactions catalysed by enzymes are typically extremely fast. However, the rate of an enzyme-catalysed reaction is sensitive to environmental conditions. Many factors within cells affect enzymes and therefore alter the rate of the reaction being catalysed. In extreme cases, proteins become **denatured**.

Denaturation of protein

Many of the properties of proteins depend on the three-dimensional shape of the molecule. This is true of enzymes, which are large, globular proteins where a small part of the surface is an active site. Here the precise chemical structure and physical configuration is critical. Provided the active site is unchanged, substrate molecules can attach and reactions can be catalysed.

Denaturation is a structural change in a protein that alters its three-dimensional shape and causes the loss of its biological properties.

Denaturation occurs when the bonds within the globular protein, formed between different amino acid residues, break. Possibly different bonds are formed. Temperature rises and changes in pH of the medium may cause denaturation of the protein of enzymes.

How may denaturation be brought about?

Exposure to heat causes atoms to vibrate violently and this disrupts bonds within globular proteins. Protein molecules change chemical characteristics. We see this most dramatically when a hen's egg is cooked. The translucent egg 'white' is a globular protein called albumen, which becomes irreversibly opaque and insoluble. Heat has triggered irreversible denaturation of this globular protein.

Small changes in pH of the medium similarly alter the shape of globular proteins. The structure of an enzyme may spontaneously re-form when the optimum pH is restored, but exposure to strong acids or alkalis is usually found to irreversibly denature enzymes.

It was investigation of the effect of change in factors such as temperature, pH and concentration of substrate that particularly helped our understanding of how enzymes work.

Temperature

Examine the investigation of the effects of temperature on the hydrolysis of starch by the enzyme amylase, shown in Figure 2.17. When starch is hydrolysed by the enzyme amylase, the product is maltose, a disaccharide (page 46). Starch gives a blue–black colour when mixed with iodine solution (iodine in potassium iodide solution), but maltose gives a red colour. The first step in this experiment is to bring samples of the enzyme and the substrate (the starch solution) to the temperature of the water bath before being mixed – a step called pre-incubation.

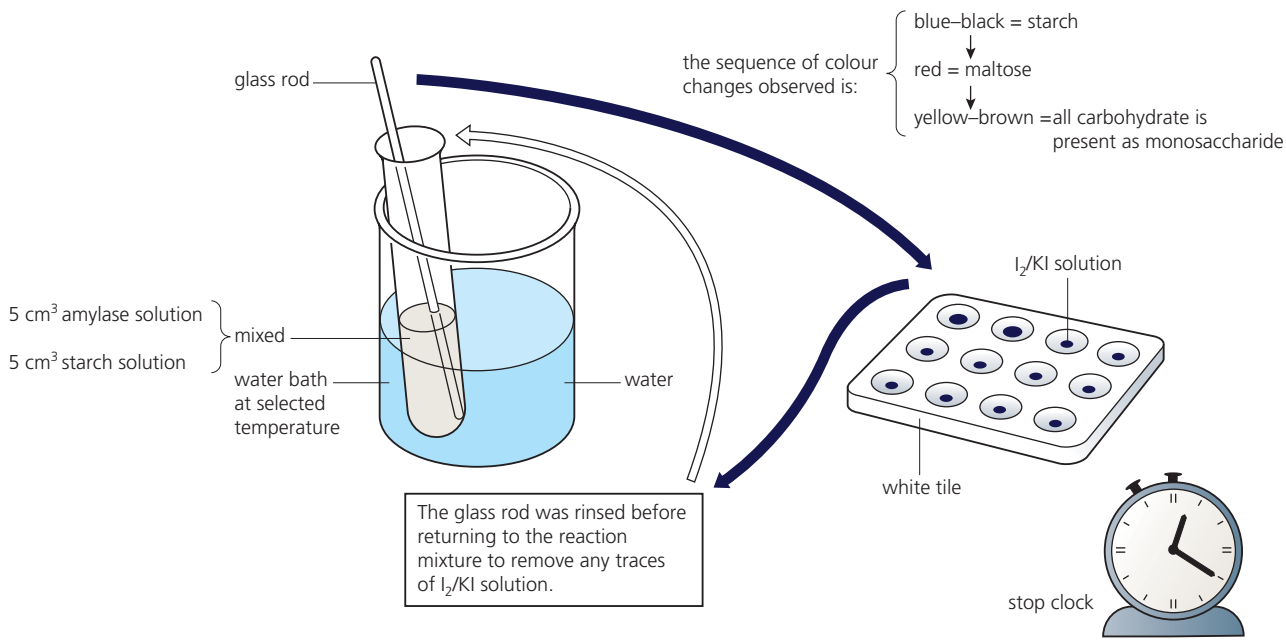


Figure 2.17 The effects of temperature on hydrolysis of starch by amylase

The experiment is **repeated at a range of temperatures,** such as at 10, 20, 30, 40, 50 and 60 °C.

A **control tube** of 5 cm^3 of starch solution + 5 cm^3 of distilled water (in place of the enzyme) should be included and tested for the presence/absence of starch, at each temperature used.

The progress of the hydrolysis reaction is then followed by taking samples of a drop of the mixture on the end of the glass rod, at half-minute intervals. These are tested with iodine solution on a white tile. Initially, a strong blue–black colour is seen confirming the presence of starch. Later, as maltose accumulates, a red colour predominates. The endpoint of the reaction is when all the starch colour has disappeared from the test spot. Using fresh reaction mixture each time, the investigation is repeated at a series of different temperatures, say at 10, 20, 30, 40, 50 and 60 °C. The times taken for complete hydrolysis at each temperature are recorded and the rate of hydrolysis in unit time is plotted on a graph (Figure 2.18).

A characteristic curve is the result – although the optimum temperature varies from reaction to reaction and with different enzymes.

How is the graph interpreted?

As temperature is increased, molecules have increased active energy, and reactions between them go faster. The molecules are moving more rapidly and are more likely to collide and react. In chemical reactions, for every 10 °C rise in temperature the rate of the reaction approximately doubles. This property is known as the **temperature coefficient** (Q_{10}) of a chemical reaction.

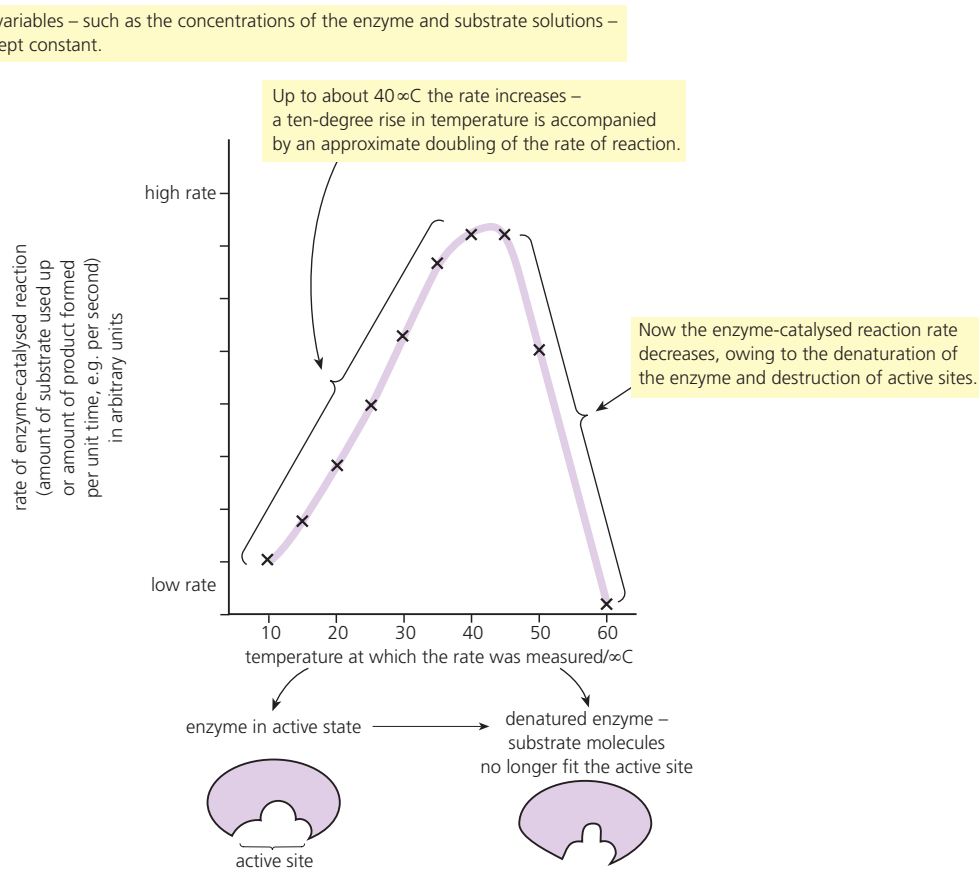


Figure 2.18 Temperature and the rate of an enzyme-catalysed reaction

However, in enzyme-catalysed reactions the effect of temperature is more complex, for proteins are denatured by heat. The rate of denaturation increases at higher temperatures, too. So as the temperature rises the amount of active enzyme progressively decreases, and the rate is slowed. As a result of these two effects of heat on enzyme-catalysed reactions, there is an apparent optimum temperature for an enzyme.

Not all enzymes have the same optimum temperature. For example, the bacteria in hot thermal springs have enzymes with optima between 80 °C and 100 °C or higher, whereas seaweeds of northern seas and the plants of the tundra have optima closer to 0 °C. Humans have enzymes with optima at or about normal body temperature. This feature of enzymes is often exploited in the commercial and industrial uses of enzymes (page 61).

12 In studies of the effect of temperature on enzyme-catalysed reactions, **suggest** why the enzyme and substrate solutions are pre-incubated to a particular temperature before they are mixed.

pH

Change in pH can have a dramatic effect on the rate of an enzyme-catalysed reaction. Each enzyme has a range of pH in which it functions efficiently. This is often at or close to neutrality point (pH 7.0). This effect of pH is because the structure of a protein (and therefore the shape of the active site) is maintained by various bonds within the three-dimensional structure of the protein. A change in pH from the optimum values alters the bonding patterns, progressively changing the shape of the molecule. The active site may quickly be rendered inactive. However, unlike temperature changes, the effects of pH on the active site are normally reversible – provided, that is, the change in surrounding acidity or alkalinity is not too extreme. As the pH reverts to the optimum for the enzyme, the active site may reappear.

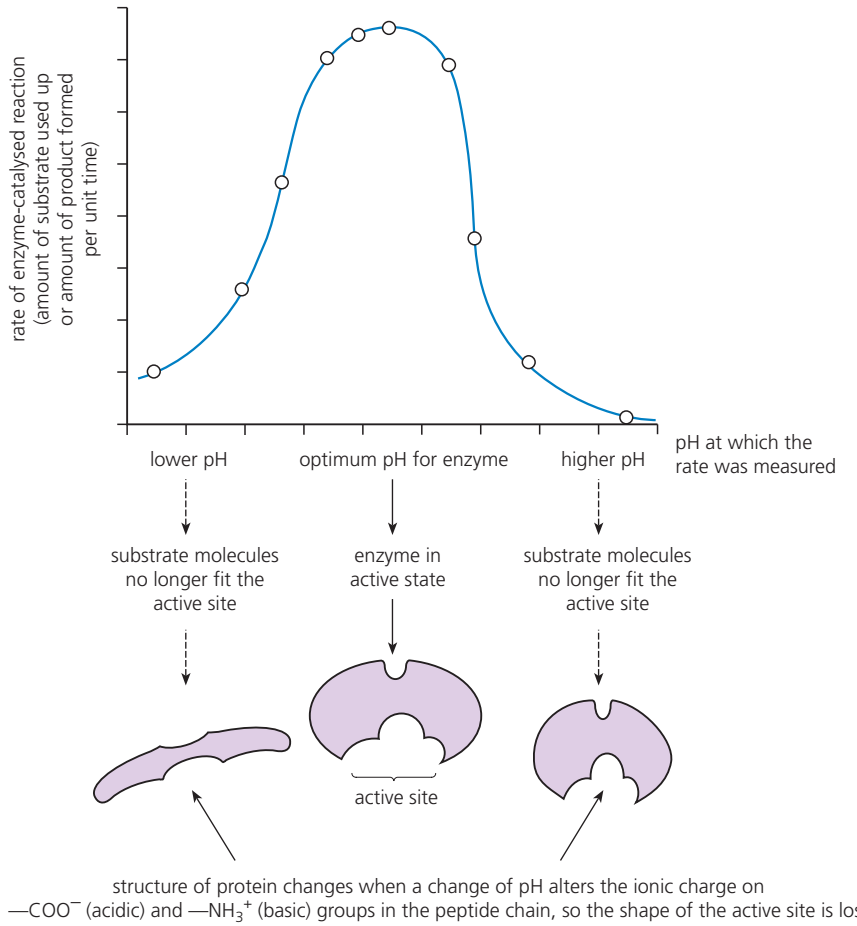
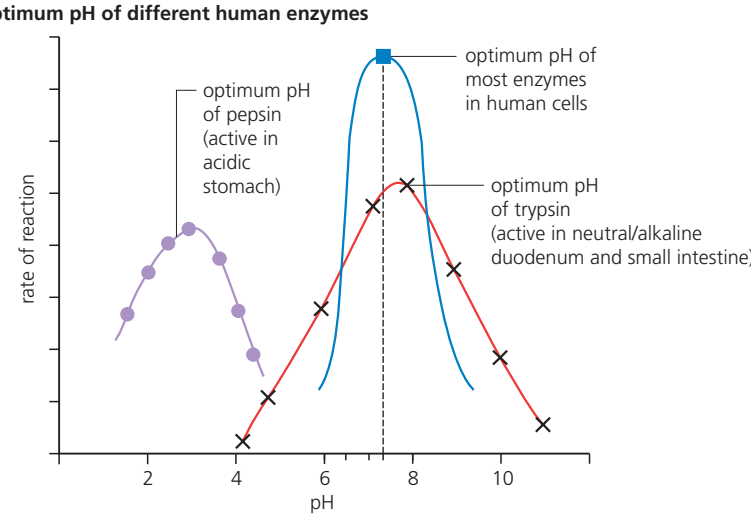


Figure 2.19 pH effect on enzyme shape and activity

the optimum pH of different human enzymes



13 **Explain** what a buffer solution is and why such solutions are often used in enzyme experiments.

Some of the digestive enzymes of the gut have different optimum pH values from the majority of other enzymes. For example, those adapted to operate in the stomach, where there is a high concentration of acid during digestion, have an optimum pH which is close to pH 2.0 (Figure 2.19).

Substrate concentration

The effect of different concentrations of substrate on the rate of an enzyme-catalysed reaction can be shown using an enzyme called catalase. This enzyme catalyses the breakdown of hydrogen peroxide:

$$2H_2O_2 \xrightarrow{\text{catalase}} 2H_2O + O_2$$

Catalase occurs very widely in living things; it functions as a protective mechanism for the delicate biochemical machinery of cells. This is because hydrogen peroxide is a common by-product of reactions of metabolism, but it is also a very toxic substance (a very powerful oxidising agent, page 270). Catalase inactivates hydrogen peroxide as it forms, before damage can occur.

You can demonstrate the presence of catalase in fresh liver tissue by dropping a small piece into dilute hydrogen peroxide solution (Figure 2.20). Compare the result obtained with that from a similar piece of liver that has been boiled in water (we have seen that high temperature denatures enzymes). If you do not wish to use animal tissues, then you can use potato or soaked and crushed dried peas, instead.

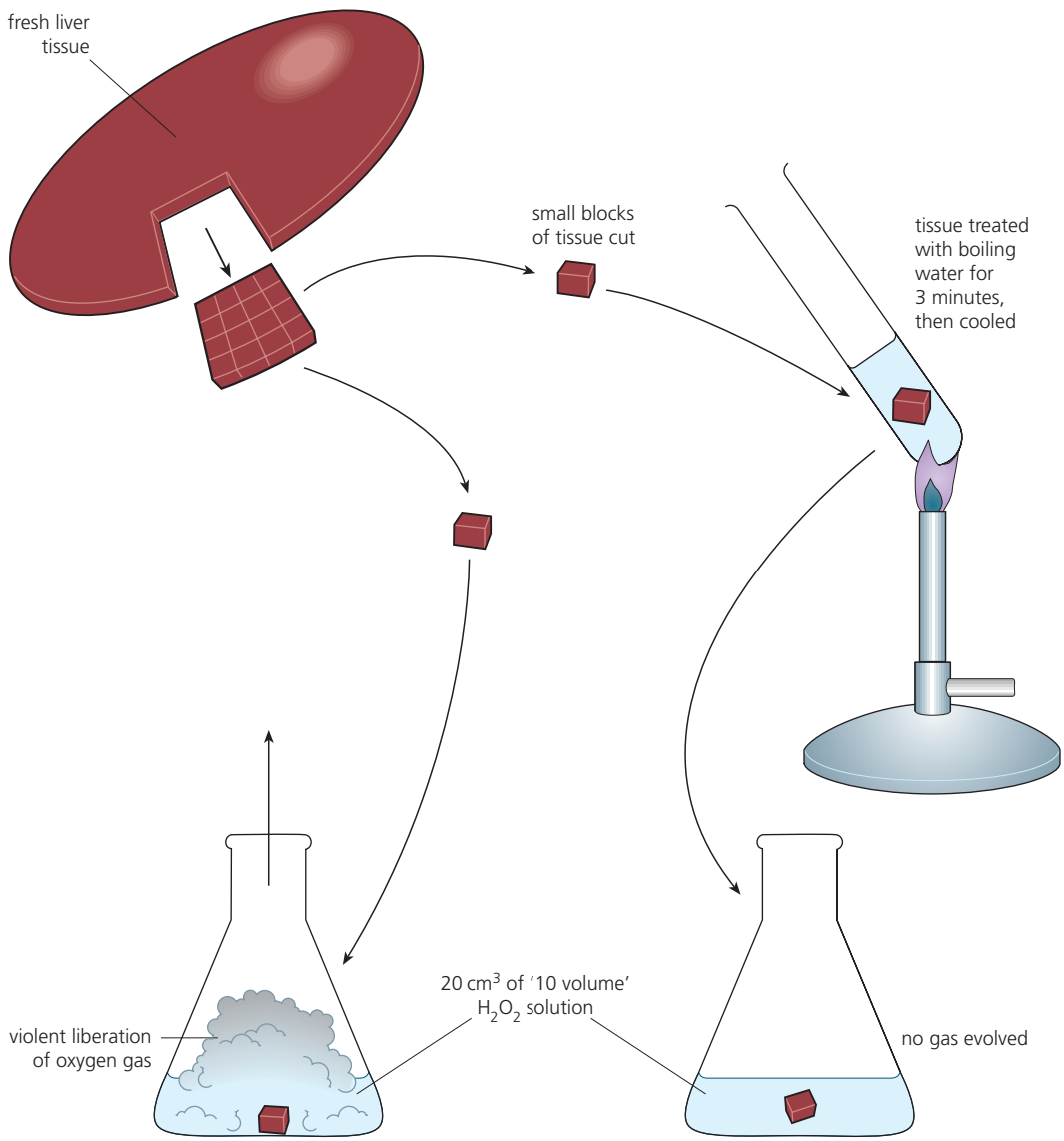


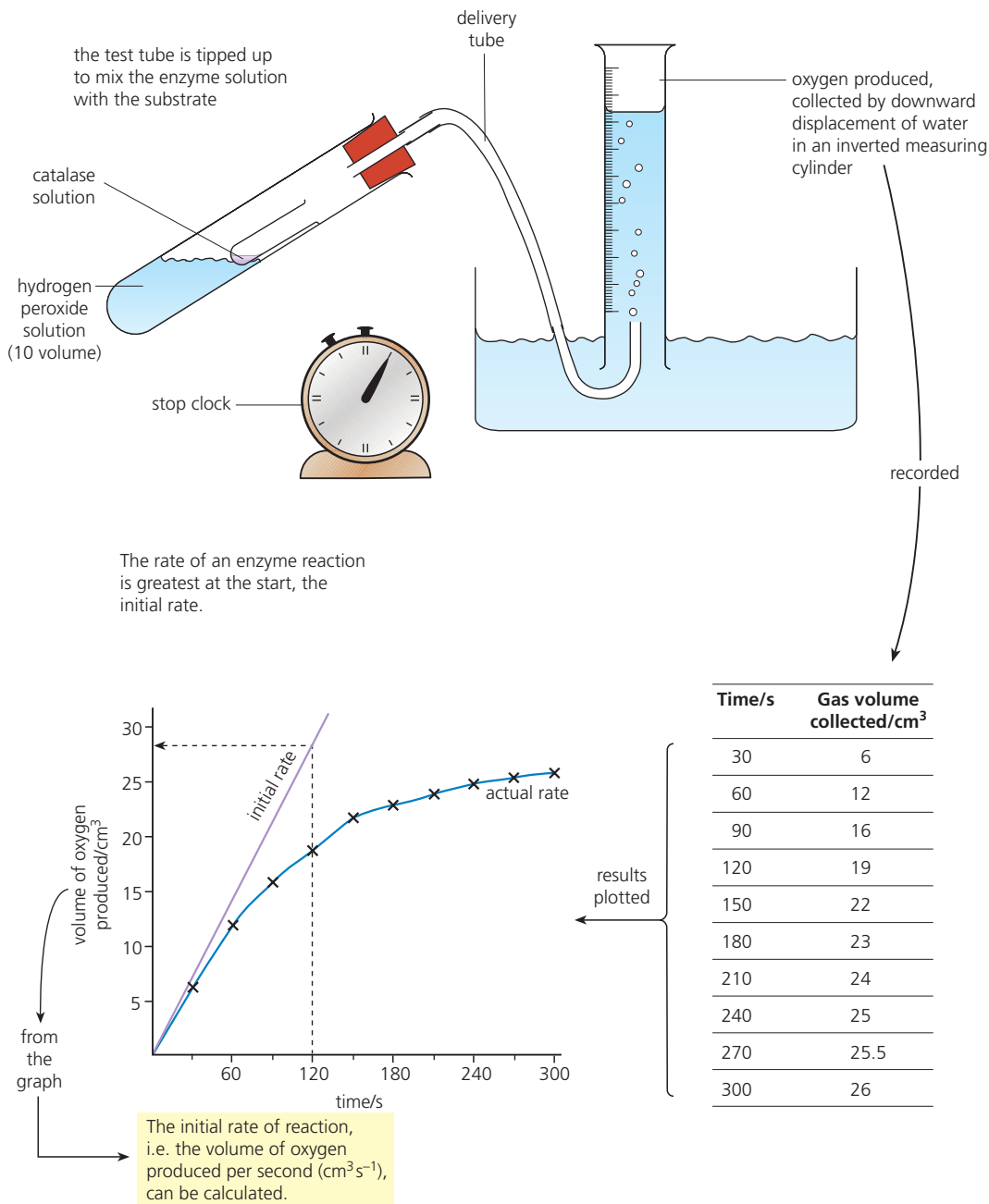
Figure 2.20 Liver tissue in dilute hydrogen peroxide solution, a demonstration

When measuring the rate of enzyme-catalysed reactions, we measure the amount of substrate that has disappeared from a reaction mixture, or the amount of product that has accumulated, in a unit of time. For example, in Figure 2.17 it is the rate at which the substrate starch disappears from a reaction mixture that is measured.

Working with catalase, it is convenient to measure the rate at which the product (oxygen) accumulates – the volume of oxygen that has accumulated at 30-second intervals is recorded (Figure 2.21).

Over a period of time, the initial rate of reaction is not maintained, but falls off quite sharply. This is typical of enzyme actions studied outside their location in the cell. The fall-off can be due to a number of reasons, but most commonly it is because the concentration of the substrate in the reaction mixture has fallen. Consequently, it is the initial rate of reaction that is measured. This is the slope of the tangent to the curve in the initial stage of reaction.

14 **Calculate** the initial rate of reaction shown in the graph in Figure 2.21.



Time/s	Gas volume collected/cm^3
30	6
60	12
90	16
120	19
150	22
180	23
210	24
240	25
270	25.5
300	26

Figure 2.21 Measuring the rate of reaction using catalase

To investigate the effects of substrate concentration on the rate of an enzyme-catalysed reaction, the experiment shown in Figure 2.21 is repeated at different concentrations of substrate, and the initial rate of reaction plotted in each case. Other variables such as temperature and enzyme concentration are kept constant.

When the initial rates of reaction are plotted against the substrate concentration, the curve shows two phases. At lower concentrations, the rate increases in direct proportion to the substrate concentration – but at higher substrate concentrations, the rate of reaction becomes constant, and shows no increase.

Now we can see that the enzyme catalase works by forming a short-lived enzyme–substrate complex. At low concentration of substrate, all molecules can find an active site without delay. Effectively, there is an excess of enzyme present. Here the rate of reaction is set by how much substrate is present – as more substrate is made available the rate of reaction increases.

However, at higher substrate concentrations, there comes a point when there is more substrate than enzyme. Now, in effect, substrate molecules have to 'queue up' for access to an active site. Adding more substrate merely increases the number of molecules awaiting contact with an enzyme molecule, so there is now no increase in the rate of reaction (Figure 2.22).

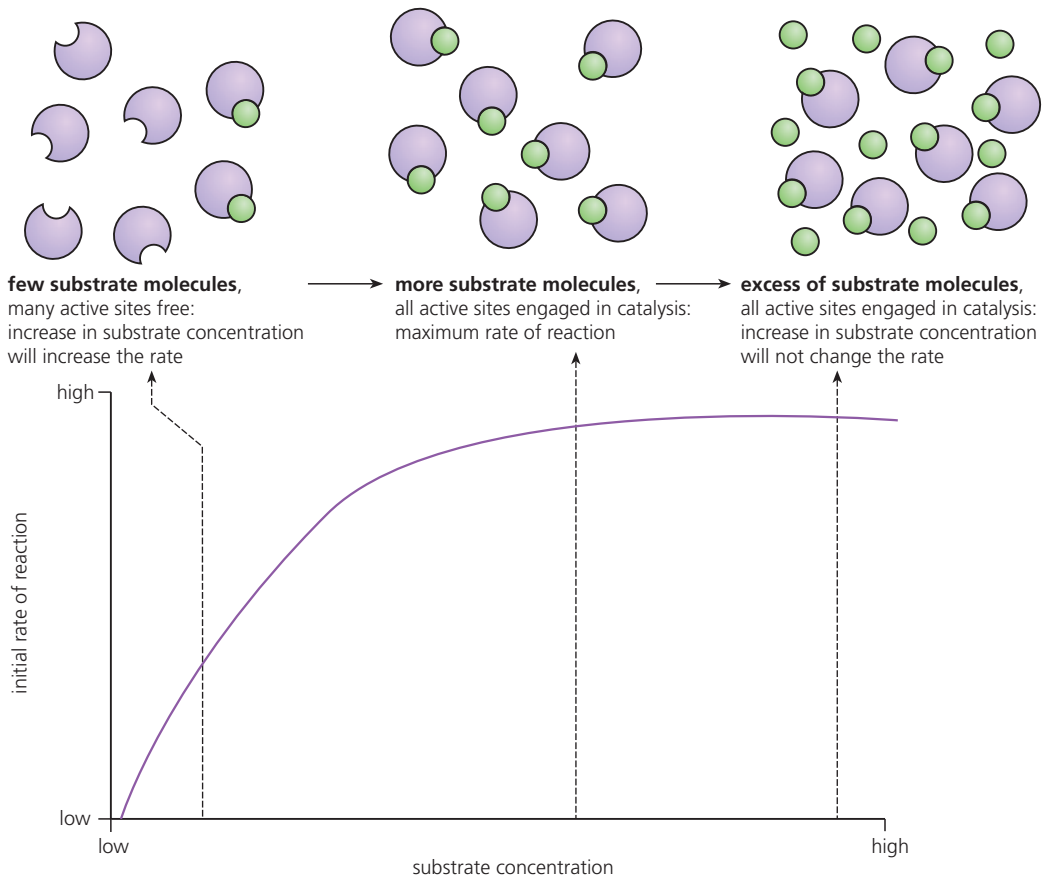


Figure 2.22 The effect of substrate concentration

15 When there is an excess of substrate present in an enzyme-catalysed reaction, **explain** the effects on the rate of reaction of increasing the concentration of:

a the substrate
b the enzyme.

Enzymes and a link with homeostasis

The requirements of enzymes for specific conditions – for example, of temperature and pH – have led to the need to control the internal environment of the body, a phenomenon known as **homeostasis** (page 217).

Industrial uses of enzymes

There is a tradition of using enzymes as biological catalysts in some long-established industries, all over the world. Here are some examples.

- In **leather tanning**, hides have been softened and hair removed using proteases naturally present in the faeces of carnivorous animals.
- In **brewing**, enzymes in barley grains at germination are used to convert the starch stored in the grain to sugars. These sugars are converted to ethanol by the fermentation enzymes of yeast.
- In **cheese manufacture**, the proteins in milk are coagulated (made insoluble), using the enzyme rennet (rennin, page 181), originally from young mammals.

Today, there are many new uses for enzymes as industrial catalysts, a part of the current development of **biotechnology**. Biotechnology is the industrial and commercial applications of biology, particularly of microorganisms, enzymology and genetic engineering (page 117). The use of enzymes in industrial processes is known as enzyme technology.

Advantages of enzymes as industrial catalysts

Enzymes are produced by living cells but often they may be separated from cells and continue to function outside the cell. They can be used in test tubes or industrial reactor vessels (known as *in vitro* conditions). Enzymes have advantages over inorganic catalysts in industrial processes because they are:

- **specific**, typically catalysing one particular compound or one type of bond only;
- **efficient**, requiring only a tiny quantity to catalyse the production of a huge quantity of product;
- able to **work at normal temperatures and pressures**, and so require much less input of energy than a chemical catalyst normally requires.

16 **Define** what we mean by *'in vitro'* and *'in vivo'*.

Sources of industrial enzymes

Useful enzymes are obtained from plant and animal sources; for example, enzymes for use in medicine in the treatment of thromboses are extracted from pineapple juice. However, the majority of enzymes for enzyme technology are from microorganisms, mainly from bacteria and fungi. Features of microorganisms making them valuable as a source of enzymes include the following.

- Microorganisms may be cultured economically and at high growth rates, in bulk fermenters, often using relatively low-grade nutrients.
- Some microorganisms occur naturally under extreme environmental conditions, and consequently may contain enzymes adapted to function efficiently under extremes of temperature and pH – for example, enzymes from hot springs' bacteria are used in biological washing powders.

Lactose-free milk – an example of industrial use of enzymes

People unable to produce lactase in their pancreatic juice or on the surface of the villi of the small intestine (page 183) fail to digest milk sugar. Lactose passes on to their large intestine without being hydrolysed to its constituent monosaccharides (page 44). As a result, bacteria in the large intestine feed on the lactose, producing fatty acids and methane, causing diarrhoea and flatulence. Such people are said to be lactose intolerant.

The enzyme lactase is produced in the gut of all human babies while they are dependent on milk. However, it is only found in adults from Northern Europe (and their descendents, wherever they now live) and from a few African tribes. On the other hand, people from the Orient, Arabia and India, most African people and those from the Mediterranean typically produce little or no lactase as adults.

Such people may be prescribed **lactose-free milk**; this product can be supplied by the application of enzyme technology, using lactase obtained from bacteria.

How can bacterial lactase be used safely in this context?

Initially, enzymes were used in industrial processes as **whole-cell preparations**. This meant applying a culture of particular microorganisms, usually some species of bacteria. This is not efficient, and may be highly inappropriate for a food product such as liquid milk. **Cell-free preparations of enzymes** have tended to replace the use of whole cells, especially where the enzymes are secreted by the microorganism into the culture medium, and so are relatively easily isolated for use. However, isolated enzymes, once applied, cannot be used over and over again, and removal of the enzyme may be expensive.

The favoured, alternative arrangement is the use of **immobilised enzymes**. Figure 2.23 illustrates the ways in which enzymes may be effectively immobilised. The advantages of using an immobilised enzyme in industrial productions are:

- it permits re-use of the enzyme preparation;
- the product is obtained enzyme free;
- the enzyme may be much more stable and long lasting, due to protection by the inert matrix.

Today, lactose-free milk is produced by passing milk through a column containing immobilised lactase. The enzyme is obtained from bacteria, purified, and enclosed in capsules (Figure 2.24).

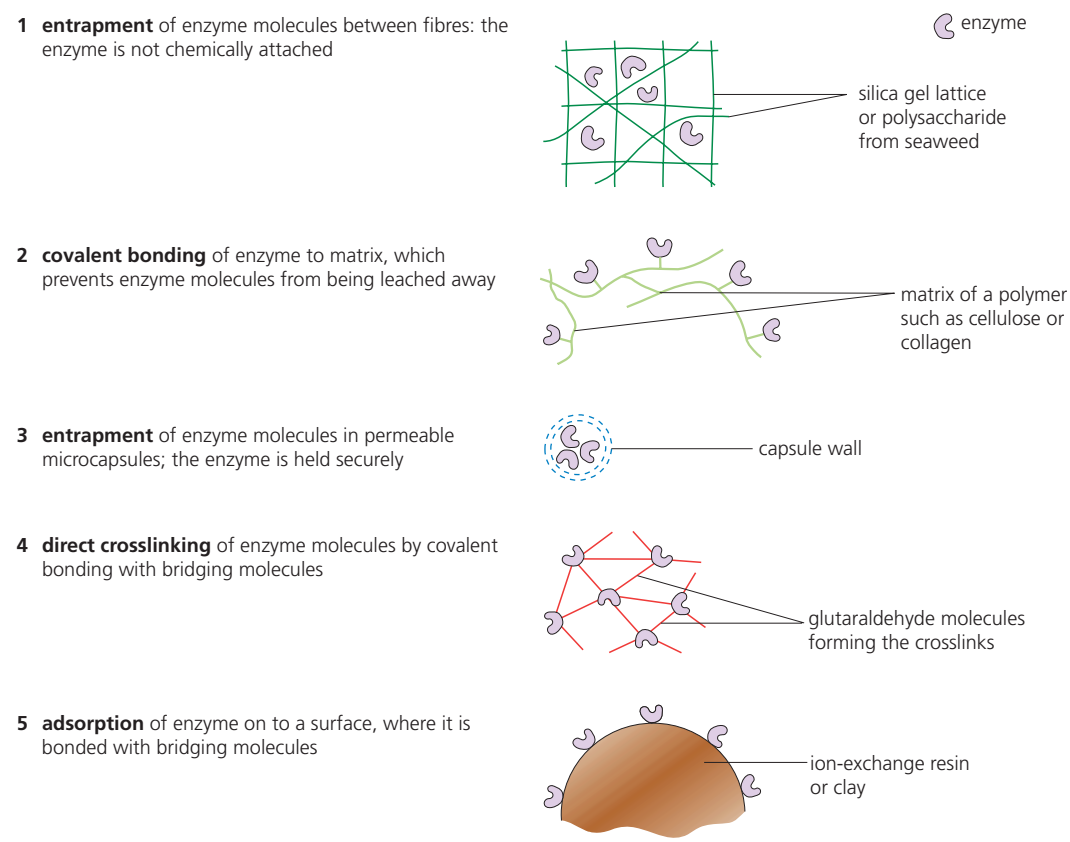


Figure 2.23 Methods of immobilising enzymes

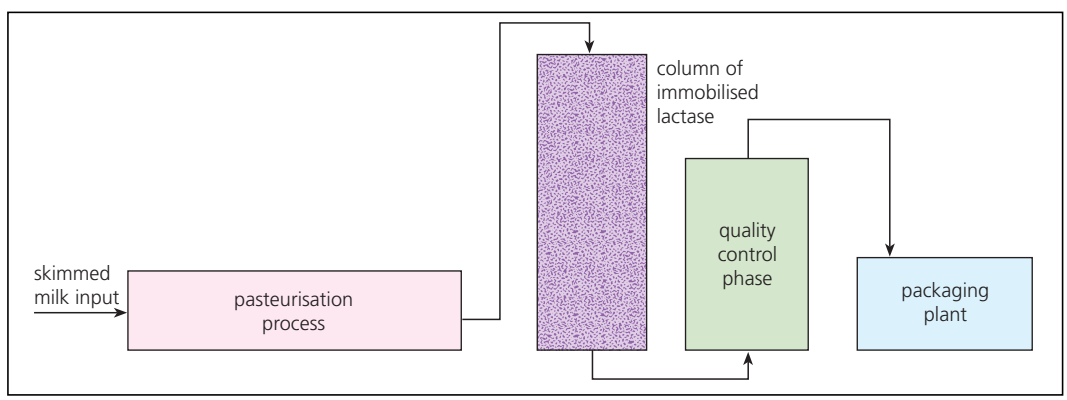


Figure 2.24 Production of lactose-free milk

17 **Define** the term 'catalyst'. **List** two differences between inorganic catalysts and enzymes.

Nucleic acids

3.1.1–3.3.5, 3.4.1–3.4.3

Nucleic acids are the **information molecules** of cells found throughout the living world. This is because the code containing the information in nucleic acids, known as the **genetic code**, is a universal one. That means that it makes sense in all organisms. It is not specific to a few organisms or to just one group, such as bacteria.

There are two types of nucleic acid found in living cells: **deoxyribonucleic acid (DNA)**, and **ribonucleic acid (RNA)**. DNA is the genetic material and occurs in the chromosomes of the nucleus. But while some RNA also occurs in the nucleus, most is found in the cytoplasm – particularly in the ribosomes. Both DNA and RNA have roles in the day-to-day control of cells and organisms, as we shall shortly see. First, we will look into the structure of nucleotides and the way they are built up to form the unique **DNA double helix**.

Nucleotides

A nucleotide consists of three substances combined together. These are:

- a **nitrogenous base** – the four bases of DNA are cytosine (C), guanine (G), adenine (A), and thymine (T);
- a **pentose sugar** – deoxyribose occurs in DNA and ribose in RNA;
- **phosphoric acid**.

These components are combined by condensation reaction to form a nucleotide with the formation of two molecules of water. Since any one of the four bases can be selected, four different types of nucleotide are formed to make DNA. How these components are combined together is shown in Figure 2.25, together with the diagrammatic way the components are represented to illustrate their spatial arrangement. Simple shapes are used rather than complex structural formulae, and these shapes are all that are required here.

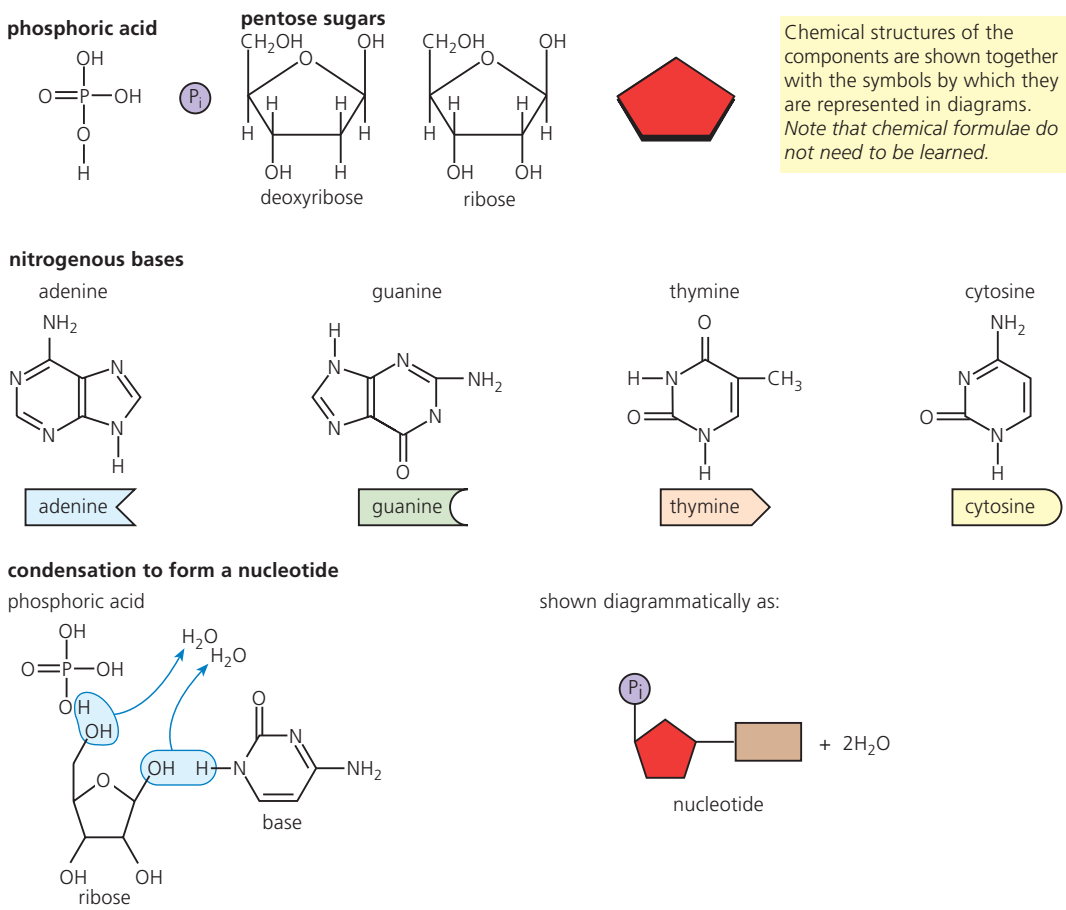
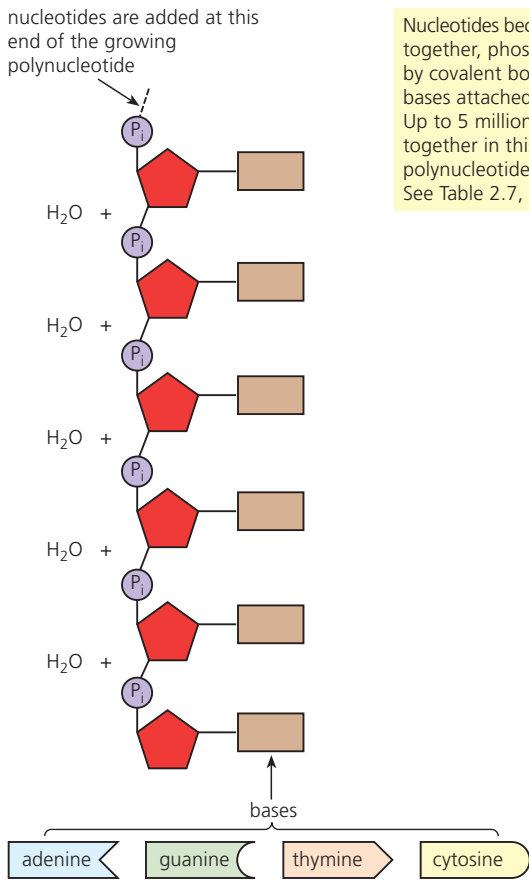


Figure 2.25 The components of nucleotides

Nucleotides become nucleic acid

Nucleotides themselves then condense together, one nucleotide at a time, to form huge molecules called nucleic acids or **polynucleotides** (Figure 2.26). So, nucleic acids are very long, thread-like macromolecules with alternating sugar and phosphate molecules forming the 'backbone'. This part of the nucleic acid molecule is uniform and unvarying. However, also attached to each of the sugar molecules along the strand is one of the bases, and these project sideways. Since the bases vary, they represent a unique sequence that carries the coded information held by the nucleic acid.

Figure 2.26 How nucleotides make up nucleic acid



Nucleotides become chemically combined together, phosphate to pentose sugar, by covalent bonds, with a sequence of bases attached to the sugar residues. Up to 5 million nucleotides condense together in this way, forming a polynucleotide (nucleic acid).
See Table 2.7, page 69.

18 **Distinguish** between a nitrogenous base and a base found in inorganic chemistry.

The DNA double helix

The DNA molecule consists of two polynucleotide strands, paired together, and held by hydrogen bonds. The two strands take the shape of a double helix (Figure 2.27). The pairing of bases is between **adenine (A)** and **thymine (T)**, and between **cytosine (C)** and **guanine (G)**, simply because these are the only combinations that fit together along the helix. This pairing, known as **complementary base pairing**, also makes possible the very precise way that DNA is copied in a process called **replication**. The existence of the DNA double helix was discovered, and the way DNA holds information was suggested, by Francis Crick and James Watson in 1953 – they received a Nobel Prize for this work (Figure 2.28).

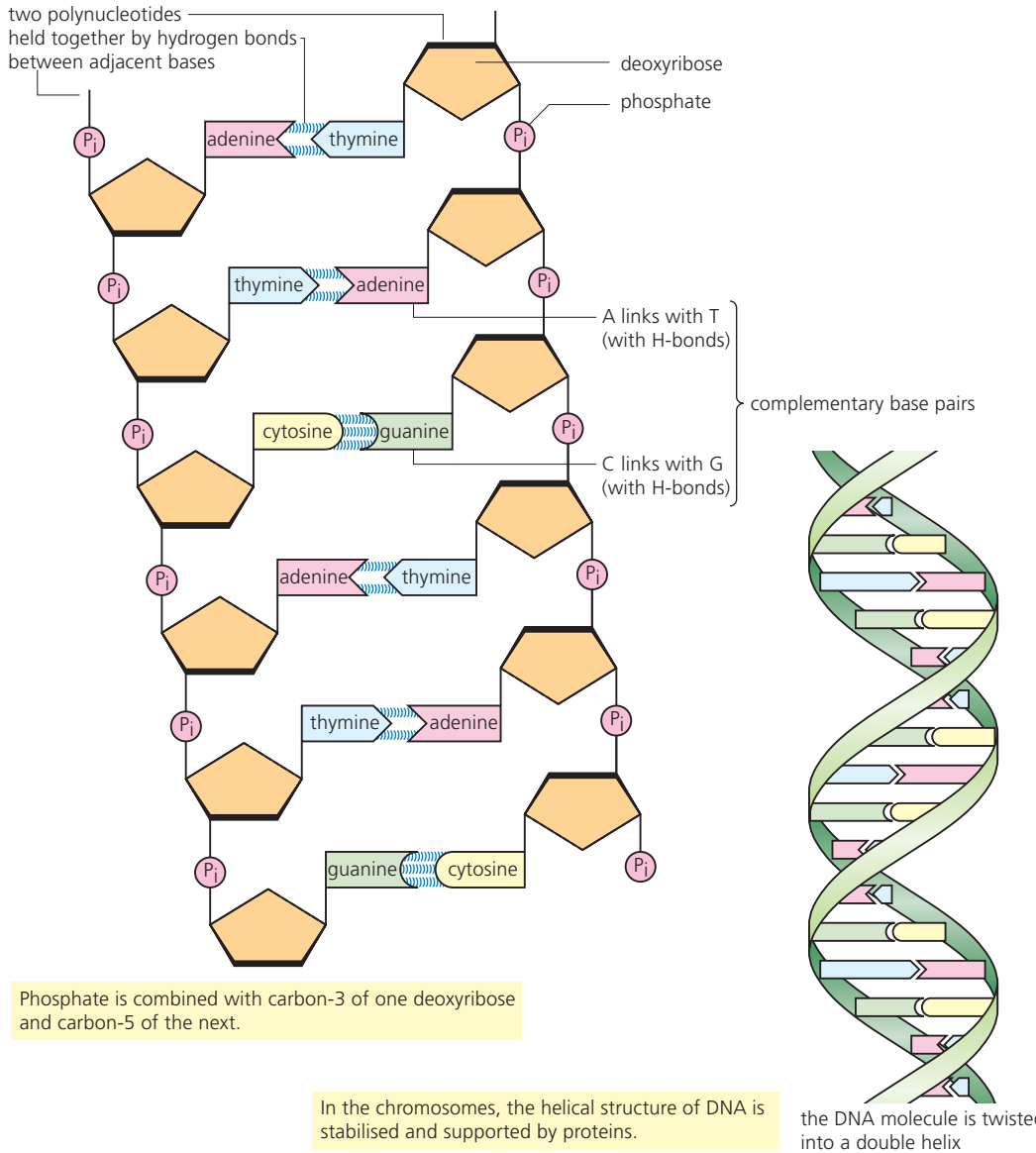


Figure 2.27 The DNA double helix

TOK Link

Crick and Watson had a distinctive method of working, including re-interpreting already published data and developing studies they had access to. Research this for your study of the **methods of science**. Try these resources.

1. Crick and Watson's original, short 'letter' to *Nature*, published 25 April 1953, if your school or college can make a copy available for you. For a one-off fee, the paper can be obtained from: www.nature.com/nature/archive
2. Their own accounts in:
   James D. Watson (1965), *The Double Helix*, Weidenfeld & Nicolson
   Francis Crick (1988), *What Mad Pursuits*, Penguin
3. Find out about the contribution of Rosalind Franklin:
   Brenda Maddox (2003), *The Dark Lady of DNA*, Harper Collins

Francis Crick (1916–2004) and **James Watson** (1928–) laid the foundations of a new branch of biology – **cell biology** – and achieved this while still young men. Within two years of their meeting in the Cavendish Laboratory, Cambridge (1951), Crick and Watson had achieved their understanding of the nature of the gene in chemical terms.

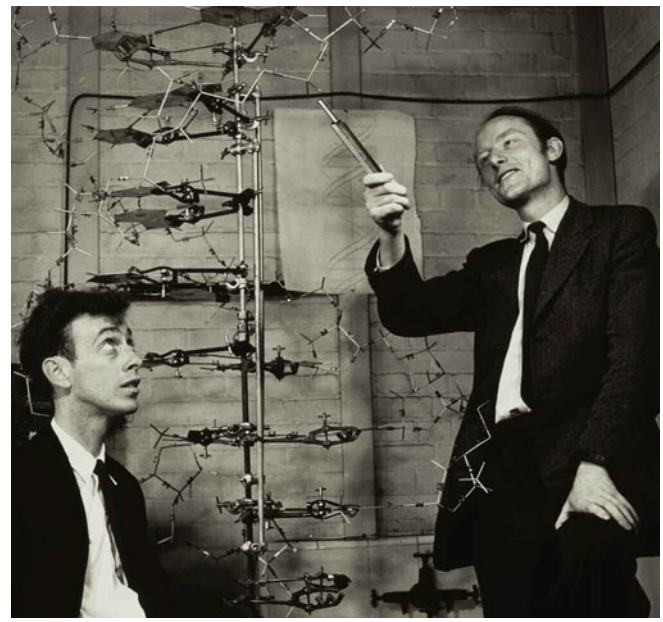
Crick and Watson brought together the experimental results of many other workers, and from this evidence they deduced the likely structure of the DNA molecule.

- **Edwin Chergaff** measured the exact amount of the four organic bases in samples of DNA, and found the ratio of A:T and of C:G was always close to 1.

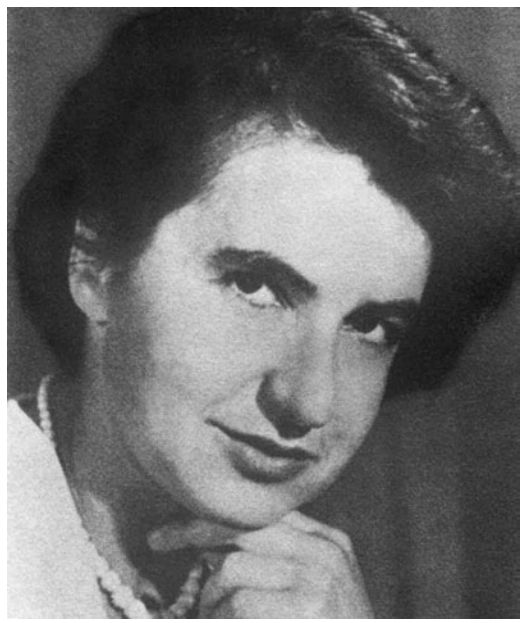
Chergaff's results suggest consistent base pairing in DNA from different organisms.

Organism	Ratio of bases in DNA samples	
	Adenine : Thymine	Guanine : Cytosine
Cow	1.04	1.00
Human	1.00	1.00
Salmon	1.02	1.02
Escherichia coli	1.09	0.99

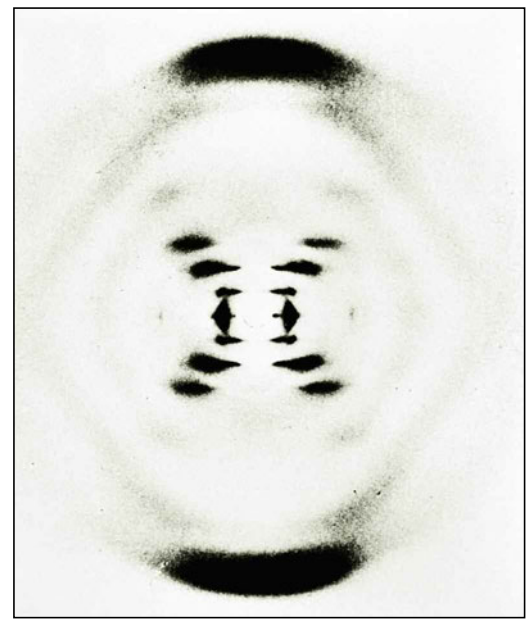
- **Rosalind Franklin** and **Maurice Wilkins** produced X-ray diffraction patterns by bombarding crystalline DNA with X-rays.



Watson and Crick with their demonstration model of DNA



Rosalind Franklin produced the key X-ray diffraction pattern of DNA at Kings College, London



X-ray diffraction pattern of DNA

Watson and Crick concluded that DNA is a double helix consisting of:

- two polynucleotide strands with nitrogenous bases stacked on the inside of the helix (like rungs on a twisted ladder)
- parallel strands are held together by hydrogen bonds between the paired bases (A–T, C–G)
- ten base pairs occur per turn of the helix
- the two strands of the double helix are antiparallel.

They built a model. See Figure 2.27 (page 65) for a simplified model of the DNA double helix.

Figure 2.28 The discovery of the role of DNA

Replication – how DNA copies itself

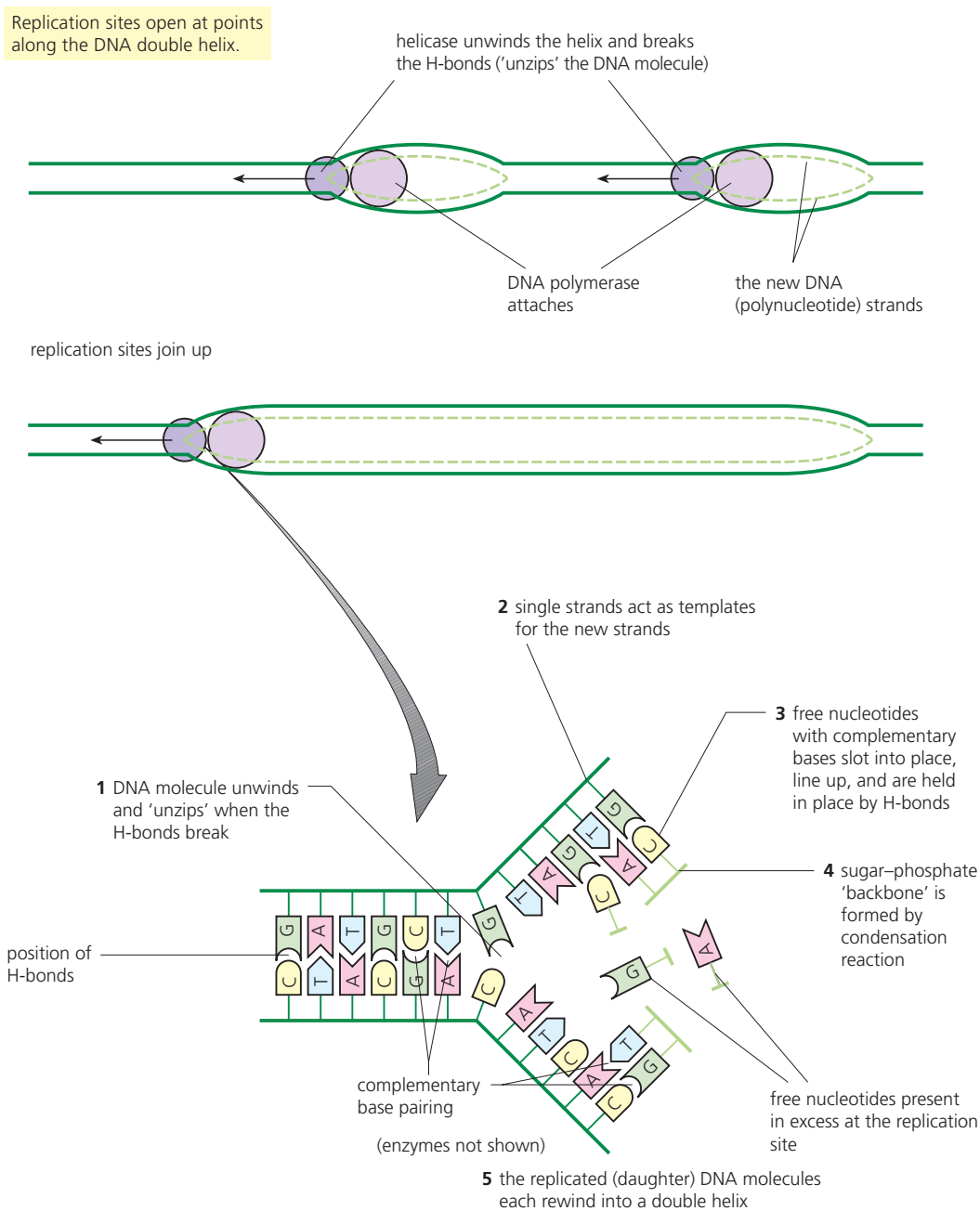
In order that a copy of each chromosome can pass into the daughter cells at cell division, the DNA of the chromosomes must first be copied (replicated). As the DNA carries the genetic message, replication must be extremely accurate. Remember, it is a process that takes place in the interphase nucleus, well before the events of nuclear division (page 31).

In replication, the DNA double helix must unwind, and the hydrogen bonds holding the strands together be broken. This allows the two strands of the helix to separate. An enzyme, **helicase**, is involved in these steps and holds the strands apart while replication occurs.

Both strands of DNA act as templates in replication. New nucleotides with the appropriate complementary bases line up opposite the bases of the exposed strands. Adenine pairs with thymine, cytosine with guanine. Because of this process of base pairing, the sequence of bases in one strand exactly determines the sequence of bases in the other strand. The two strands are said to be complementary. The bases of the two strands fit together only if the sugar molecules they are attached to point in opposite directions. The strands are said to be antiparallel.

Hydrogen bonds then form between the complementary bases, holding them in place. Finally, the sugar and phosphate groups of adjacent nucleotides of the new strand condense together. This reaction is catalysed by an enzyme called **DNA polymerase**. DNA replication is summarised in Figure 2.29.

Figure 2.29 DNA replication



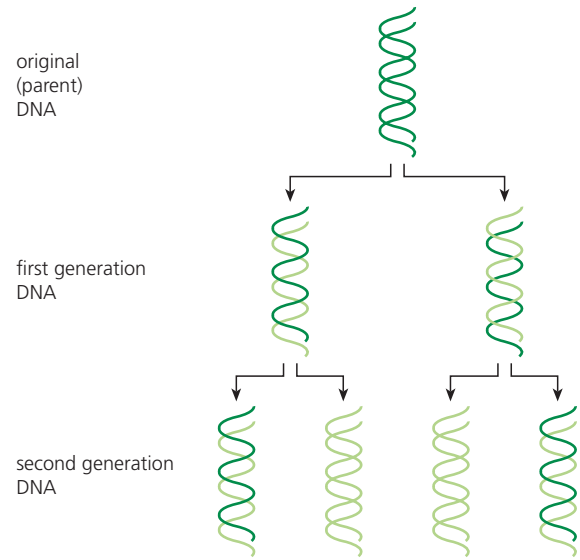
Each new pair of double strands winds up into a double helix. One strand of each new double helix came from the original chromosome and one is a newly synthesised strand. This arrangement is known as **semi-conservative replication** because half the original molecule is kept the same (Figure 2.30).

DNA polymerase also has a role in 'proof-reading' the new strands. Any mistakes that start to happen (such as the wrong bases attempting to pair up) are corrected. Each new DNA double helix is an exact copy of the original.

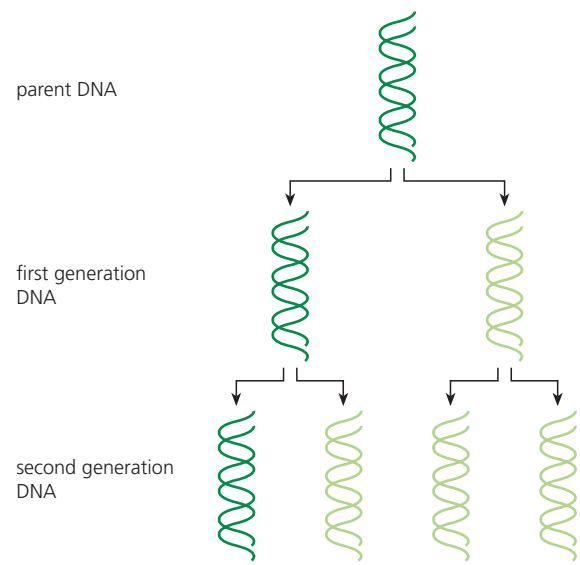
Figure 2.30 Semi-conservative versus conservative replication

Crick and Watson suggested replication of DNA would be 'semi-conservative', and this has since been shown experimentally, using DNA of bacteria 'labelled' with a 'heavy' nitrogen isotope.

In **semi-conservative replication** one strand of each new double helix comes from the parent chromosome and one is a newly synthesised strand (i.e. half the original molecule is conserved)



If an entirely new double helix were formed alongside the original, then one DNA double helix molecule would be conserved without unzipping, in the next generation (i.e. **conservative replication**).



Protein synthesis

3.5.1–3.5.5

Proteins are linear series of amino acids condensed together – most proteins contain several hundred amino acid residues. There are only 20 different amino acids used in protein synthesis; all cell proteins are built from these. The unique properties of a protein lie in:

- **which amino acids** are involved in its construction;
- **the sequence** in which these amino acids are condensed together.

The sequence of bases in DNA dictates the order in which specific amino acids are assembled and combined together. The huge length of the DNA molecule in a single chromosome codes for a very large number of proteins. Within this extremely long molecule, the relatively short length of DNA that codes for a single protein is called a **gene**. Proteins are very variable in size and, therefore, so are genes. A very few genes are as short as 75–100 nucleotides. Most are at least 1000 nucleotides long, and some are more.

Proteins are formed at ribosomes in the cytoplasm, so a copy of the gene's information is first required, and transported to the site of protein synthesis. That copy is made of RNA, and is called messenger RNA. So both DNA and RNA are involved in protein synthesis.

Introducing RNA

RNA molecules are relatively short in length when compared with DNA. In fact, RNAs tend to be from a hundred to thousands of nucleotides long, depending on the particular role they have. The RNA molecule is a single strand of polynucleotide in which the sugar is ribose. The bases found in RNA are cytosine, guanine, adenine, and **uracil** (which replaces thymine of DNA). There are three functional types of RNA, known as **messenger RNA (mRNA)**, **transfer RNA (tRNA)**, and **ribosomal RNA** (Table 2.7). While mRNA is formed in the nucleus and passes out to ribosomes in the cytoplasm, tRNA and ribosomal RNA occur only in the cytoplasm.

DNA	Feature	RNA
very long strands, several million nucleotides	**length**	relatively short strands, 100 to several thousand nucleotides
contains deoxyribose	**sugar**	contains ribose
contains bases C, G, A and T (not U)	**bases**	contains C, G, A and U (not T)
consists of two polynucleotide strands of complementary base pairs (C with G and A with T) held by H-bonds in the form of a double helix	**forms**	consists of single strands, and in three functional forms: messenger RNA (mRNA) transfer RNA (tRNA) ribosomal RNA

Table 2.7 The differences between DNA and RNA

Transcription – the first step in protein synthesis

Stage 1 of protein synthesis occurs in the nucleus where a complementary copy of the code is made by the building of a molecule of mRNA. This process is called **transcription** (Figure 2.31), and the enzyme is **RNA polymerase**. The DNA double helix unwinds and the hydrogen bonds are broken at the site of the gene being transcribed. At this site there is a pool of free nucleotides. Then, one strand of the DNA, the coding strand, is used as the template for transcription by complementary base pairing. Of course, in RNA synthesis it is uracil which pairs with adenine. Once the mRNA strand is formed, it leaves the nucleus through pores in the nuclear membrane and passes to ribosomes in the cytoplasm, where the information can be 'read' and used in the synthesis of a protein.

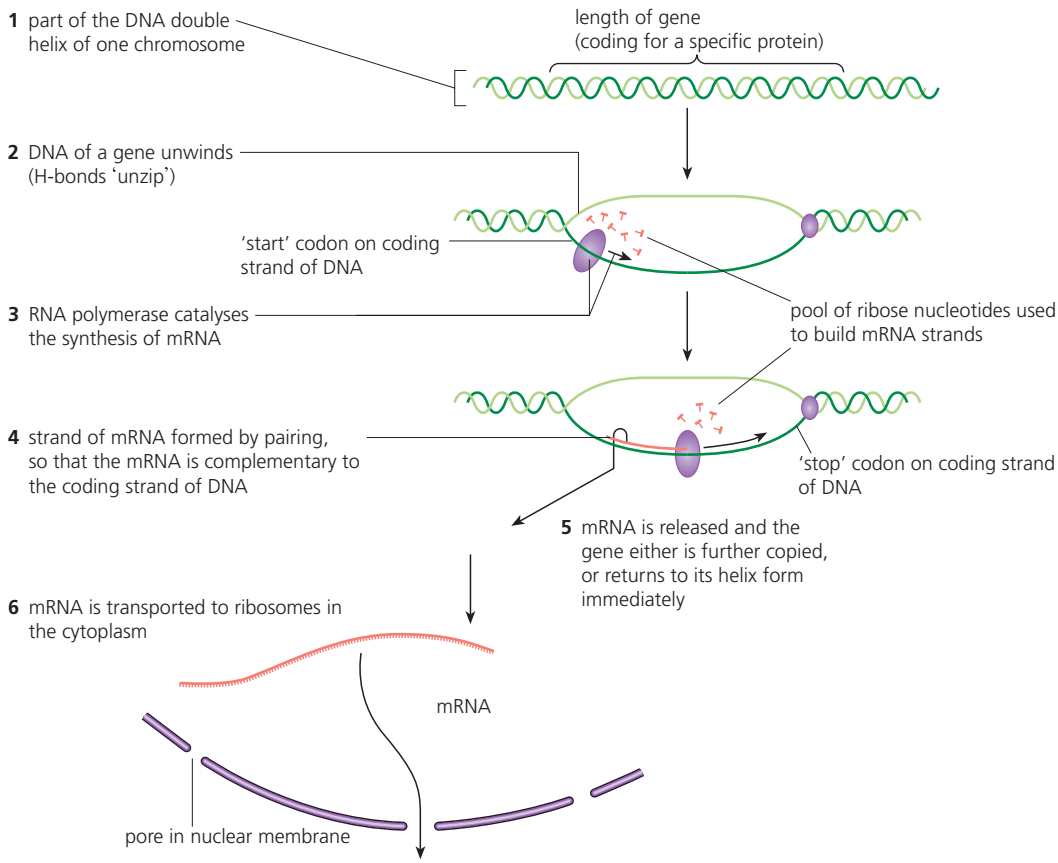


Figure 2.31 Transcription

The genetic code

The sequence of bases in nucleic acid dictates the order in which specific amino acids are to be assembled and combined to make proteins. The code is a three-letter or **triplet code**, meaning that each sequence of three bases stands for one of the 20 amino acids, and is called a **codon**.

In fact, with the four-letter alphabet of DNA (C, G, A, T) there are 64 possible different triplet combinations (4 × 4 × 4). In other words, the genetic code has many more codons than there are amino acids – since only 20 amino acids require coding. This type of code is described as **degenerate** since most amino acids have two or three similar codons that code for them. A degenerate code has more than one base triplet (codon) to code for one amino acid.

Of course, some of the codons represent the 'punctuations' of the code, meaning there are 'start' and 'stop' triplets. Thus the enzyme machinery involved in protein synthesis is instructed where to begin and to end a particular protein sequence.

Amino acid activation – the second step in protein synthesis

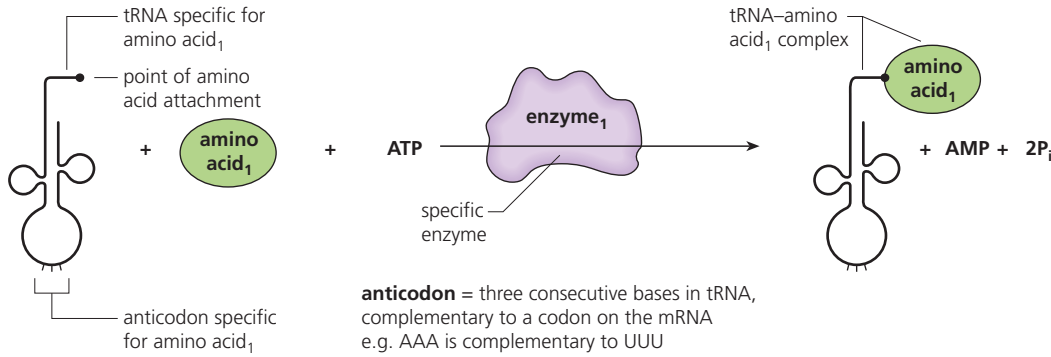
In Stage 2 of protein synthesis, the amino acids of the pool available for incorporation into protein are activated by combining with short lengths of a different sort of RNA, tRNA. This activation occurs in the cytoplasm. The special significance of tRNA is that translates a three-base sequence into an amino acid sequence.

How does this occur?

All the tRNAs have a cloverleaf shape, but there is a different tRNA for each of the 20 amino acids involved in protein synthesis. At one end of each tRNA molecule is a site where one particular amino acid of the 20 involved can be joined. At the other end, there is a sequence of three bases called an **anticodon**. This anticodon is complementary to the codon of mRNA that codes for the specific amino acid (Figure 2.32).

Figure 2.32 Amino acid activation

Each amino acid is linked to a specific transfer RNA (tRNA) before it can be used in protein synthesis. This is the process of amino acid activation. It takes place in the cytoplasm.



The amino acid becomes attached to its tRNA by an enzyme in a reaction that also requires ATP. These enzymes are specific to the particular amino acids (and types of tRNA) to be used in protein synthesis. The specificity of the enzymes is a way of ensuring the correct amino acids are used in the right sequence.

Translation – the last step in protein synthesis

In Stage 3, the final stage, a protein chain is assembled, one amino acid residue at a time, in tiny organelles called ribosomes. These move along the messenger RNA 'reading' the codons from the 'start' codon. In the ribosome, for each mRNA codon, the complementary anticodon of the tRNA–amino acid complex slots into place and is temporarily held in position by hydrogen bonds. While held there, the amino acids of neighbour tRNA–amino acid complexes are joined by a peptide linkage. This frees the first tRNA which moves back into the cytoplasm for re-use. Once this is done, the ribosome moves on to the next mRNA codon. The process continues until a ribosome meets a 'stop' codon (Figure 2.33).

19 **Outline** the different forms of RNA involved in transcription and translation, and state the roles of each.

One gene – one enzyme?

So far we have assumed that each gene works by coding for the synthesis of one particular protein, such as an enzyme.

Do we know this is the case? If so, how?

The answer is that evidence came from two American scientists, G. W. Beadle and E. L. Tatum, who worked together in America in 1940. Their work is summarised below, and in Figure 2.34.

TOK Link

While the details of Beadle and Tatum's experiments are not required, comparison of their **experimental approach** with that of Watson and Crick may support your discussion of the methods of science.

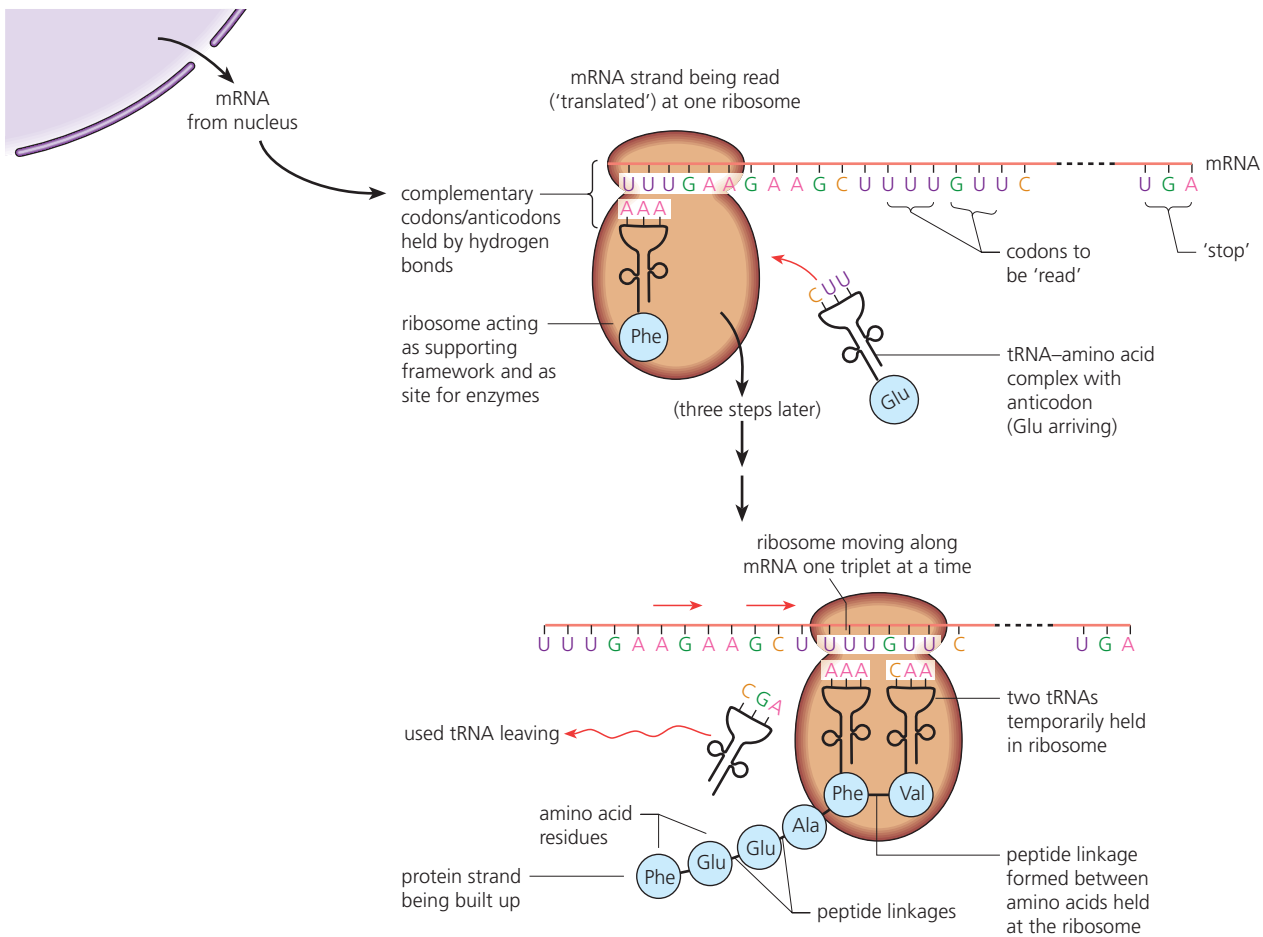
Beadle and Tatum worked with a fungus called *Neurospora* that was a pest organism in bakeries, causing bread to turn mouldy. This simple organism was easily cultured on laboratory agar plates containing a **minimal medium** of sugar, mineral ions, the vitamin biotin, and a source of nitrogen such as nitrate or ammonium ions. The fungus naturally manufactures all other organic compounds required, including all the amino acids, from this medium – hence the term 'minimal'.

They exposed some of their cultures of *Neurospora* to X-rays that caused the formation of mutants. A **mutant** organism has altered genetic material. Interestingly, some of the mutant forms produced by chance were unable to grow on the minimal medium, but could still grow normally on media enriched with amino acids used in protein synthesis.

Further investigations with these mutants showed that most needed only one particular amino acid added to minimal medium in order to be able to grow and reproduce. These mutants had lost the ability to synthesise one particular enzyme.

In a whole series of experiments, many different mutants were produced and cultured – each unable to synthesise a different enzyme. They also conducted breeding experiments between mutants and unaltered *Neurospora* cultures, and found that the new characteristics of mutant *Neurospora* were inherited just as other, naturally occurring characteristics were (Figure 2.34).

Figure 2.33 Translation



Several ribosomes may move along the mRNA at one time. This structure (mRNA, ribosomes plus growing protein chains) is called a **polysome**.

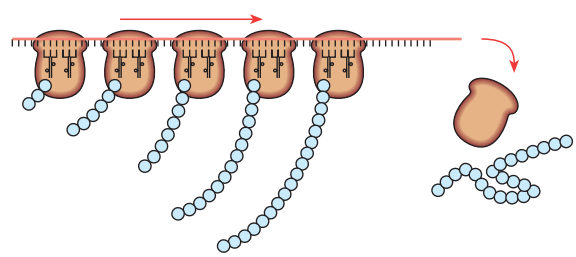
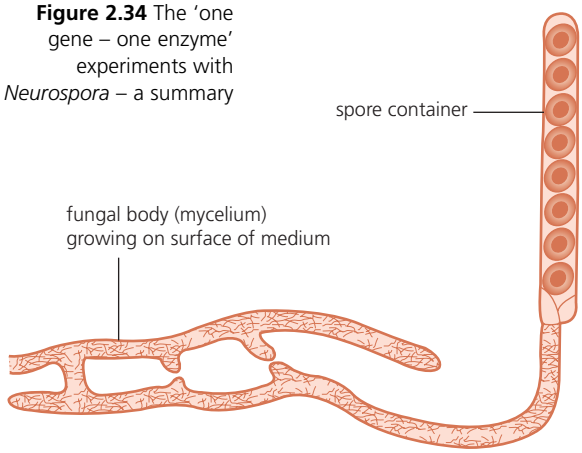
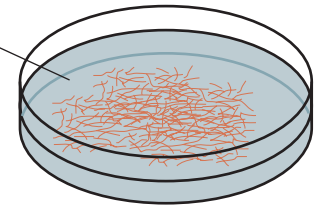
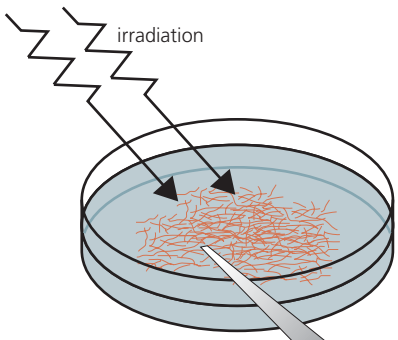


Figure 2.34 The 'one gene – one enzyme' experiments with *Neurospora* – a summary



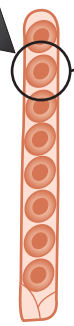
1 culturing *Neurospora*
grown in a Petri dish on the surface of a sterilised agar medium containing sucrose, inorganic salts (e.g. NO_3^-, PO_4^{3-} etc.) and a growth factor (biotin)
= **minimal medium**



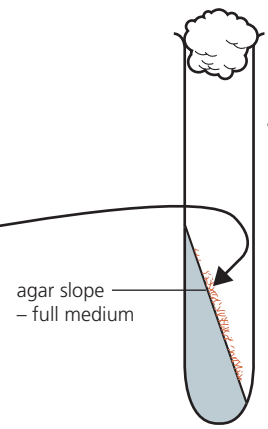


2 induction of mutations
Mutations are modifications to the chemistry of the base pairs of DNA.
Mutagens (which cause mutations) include ionising radiation (α, β or γ radiation from radioactive sources, or X-rays), or non-ionising sources (UV light or some chemicals).

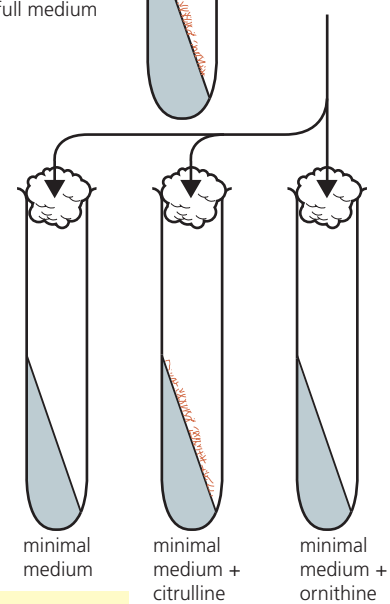
3 isolating spores
Groups of eight spores are lifted from the surface of the medium and **separately plated** on medium in a sterile tube, i.e. each spore gives rise to a separate reference culture grown on full medium.



4 reference culture, one from each spore (kept, used for sub-culturing)



5 screening spores for mutations
Spores are plated out on a full medium, containing most of the elaborate nutrients required for growth (called the **reference culture**).
Part of the reference culture is **sub-cultured** on minimal medium. Any mutant unable to manufacture any of the essential amino acids will start to die.
Once a mutant fungus is detected, it is experimentally sub-cultured on a range of media providing different amino acids to find out which one the mutant requires (i.e. which genes it has lost through mutation).



Results and discovery

- A range of mutants were produced, each deficient in one of the enzymes needed to produce arginine.
- Arginine production in *Neurospora* starts with glutamic acid, and involves two intermediates and three enzymes:

glutamic acid $\xrightarrow{\textit{enzyme 1}}$ ornithine $\xrightarrow{\textit{enzyme 2}}$ citrulline $\xrightarrow{\textit{enzyme 3}}$ arginine

- In different mutants, enzymes 1, or 2, or 3 were no longer produced, e.g. one mutant lacked enzyme 2, and could not produce citrulline (only if citrulline was added to the minimal medium would growth occur).

Beadle and Tatum concluded that in an organism **each gene coded for a particular enzyme**. This was known as the **one gene – one enzyme hypothesis**.

This work was carried out in the middle of the last century, well before Crick and Watson's work on the chemical nature of the gene. Beadle and Tatum's contribution earned them a Nobel Prize. Interestingly, the same conclusions about the role of a gene had been proposed by a doctor studying inherited human diseases of metabolism several years earlier, but his work was ignored, and so was unknown to the American scientists at the time of their study.

Subsequently, as protein chemistry became better understood, several proteins were found to consist of two (or more) polypeptide chains. Also, of course, not all proteins are enzymes, so the hypothesis was modified to **one gene – one polypeptide**.

Another development has been the appreciation that in different parts of an organism (or at different times), selected genes are switched on or switched off. The existence of **regulator genes** which activate or suppress the action of neighbouring genes has been demonstrated in, for example, prokaryotes (page 244).

Human haemoglobin consists of two sorts of polypeptide chain, α (**alpha**) **chains** and β (**beta**) **chains**. In the fetus, the β chains are replaced by γ (**gamma**) **chains**, the effect of which gives fetal haemoglobin a higher affinity for oxygen (page 667). Control of haemoglobin production may involve:

- a gene that codes for α chains – always switched on;
- a gene that codes for γ chains – switched on in the fetal stage only;
- a gene that codes for β chains – switched off at the fetal stage but switched on at birth;
- other genes that switch on and off the β gene and the γ gene.

Consequently, the picture has emerged of genes with a range of roles, sometimes activated, but possibly inactivated at other times. The combined effect of gene action, however, is the realisation of the characteristics of an organism by the interactions of gene products, not all of which are enzymes.

20 Not all proteins formed in cells are enzymes. **State** the other important roles proteins may have.

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 Which property of water is most important to plants living below the surface of water?

A cohesion
B oxygen solubility
C surface tension
D transparency

Higher Level Paper 1, May 06, Q8

Q2 **a** Define the term *active site* of an enzyme. (1)
b Outline how enzymes catalyse biochemical reactions. (2)
c Explain the effect of pH on enzyme activity. (3)
d State three functions of lipids. (2)

Higher Level Paper 1, May 03, Q3

Q3 Where do transcription and translation occur in eukaryotic cells?

	Transcription	Translation
A	cytoplasm	cytoplasm
B	cytoplasm	mitochondria
C	nucleus	cytoplasm
D	nucleus	nucleus

Standard Level Paper 2, May 05, Q8

Q4 Which of the following elements is most common in living organisms?

A sodium
B oxygen
C iron
D nitrogen

Standard Level Paper 1, May 04, Q6

Questions 5–8 cover other syllabus issues in this chapter.

Q5 Carbohydrates, lipids and proteins make up the bulk of the organic compounds found in organisms.

a Identify a monosaccharide commonly found in both animal and plant cells. (1)
b State the polysaccharide with a role as energy store found in skeletal muscle, and what monomer it is constructed from. (2)
c Draw a generalised structural formula for:
i an amino acid
ii a fatty acid. (4)
d Define the type of chemical reaction involved in the formation of a peptide linkage between two amino acids, and state the other product, in addition to a dipeptide. (2)
e List three advantageous features of lipids as energy stores for terrestrial animals. (3)

Q6 **a** State at which period in a cell cycle the replication process takes place. (1)
b For the process of DNA replication, draw a diagrammatic representation of the structure of a 'replication fork' in which two important enzymes essential to replication are shown *in situ*. Annotate your diagram to make clear the roles of these enzymes. (6)
c Explain the part played by complementary base pairing in replication. (2)
d Explain why the replication process is described as semi-conservative. (2)

Q7 **a** Outline the key features of biological macromolecules. (1)
b Glycogen and protein are both polymers. Explain why there is only one type of glycogen but many different types of protein. (2)

Q8 **a** Nucleic acids are the information molecules of cells, and structurally consist of nucleotides combined together. State what three substances combine together to form a nucleotide. (3)
b List the structural differences between RNA and DNA. (2)
c Describe the types of RNA that occur in cells, what roles each type has, and where each type is commonly found in growing cells. (3)
d Outline the difference in roles of DNA polymerase and RNA polymerase. (2)

3 Energy transfer in cells

STARTING POINTS

- **Organisms require energy** to maintain cells and carry out their activities and functions.
- **Cellular respiration** occurs in every living cell, making energy available.
- Green plants manufacture glucose by **photosynthesis**, using energy from sunlight.
- **Chloroplasts** contain chlorophyll and are the site of photosynthesis.

Living things require **energy** to build and repair body structures, and to maintain all the activities of life, such as movement, reproduction, nutrition, excretion and sensitivity. The need for energy for protein synthesis, and for the active transport of molecules and ions across membranes by membrane pumps, has already been discussed (Chapter 1).

Energy is transferred in cells by the breakdown of nutrients, principally of carbohydrates like glucose. The process by which energy is made available from nutrients in cells is called **cellular respiration**; this process is investigated first in this chapter.

Organisms require a supply of nutrients to sustain respiration, but they obtain their nutrients in very different ways. Green plants synthesise sugars from simpler substances by **photosynthesis**. Photosynthesis is discussed later in this chapter. On the other hand, animals and other heterotrophic organisms take nutrients into the body tissues following the digestion of food (Chapter 7, page 178).

Cellular respiration – the controlled release of energy

3.7.1–3.7.4

We can measure the total amount of energy that can be released from the nutrient glucose by means of a simple **calorimeter** (Figure 3.1). Here, a known amount of glucose is placed in the crucible in a closed environment, and burnt in oxygen. The energy, released as heat, is transferred to the surrounding jacket of water, and the rise in temperature of the water is measured. Then the energy value of the glucose sample may be calculated, based on the fact that it takes 4.2 joules (J) of heat energy to raise 1 g of water by 1 °C. One well known outcome of this technique is the energy value labelling of manufactured foods that we can read on the packaging – and the publicity that 'low-calorie' items may receive in slimming diets.

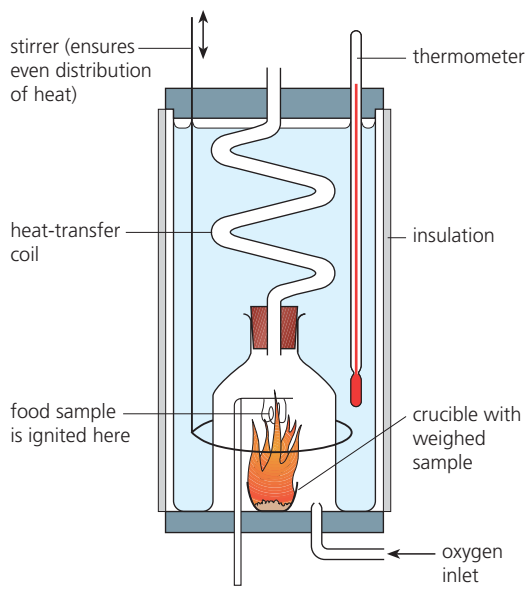


Figure 3.1 A calorimeter for measuring energy value

Samples are completely oxidised in air.

The energy values of foods are published in tables, and those of manufactured and packaged foods may be recorded on the wrapping.

People sometimes talk about 'burning up food' in respiration. In fact, likening respiration to combustion is unhelpful. In combustion, energy in fuel is released in a one-step reaction, as heat (Figure 3.2). Such a violent change would be disastrous for body tissues. In cellular respiration, a large number of small steps occur, each catalysed by a specific enzyme. Because energy in respiration is transferred in small quantities, much of the energy is made available to the cells, and may be trapped in the energy currency molecule **adenosine triphosphate (ATP)**. However, some energy is still lost as heat in each step – we notice how warm we become with strenuous physical activity!

Cellular respiration is the controlled release of energy, in the form of ATP, from organic compounds in cells.

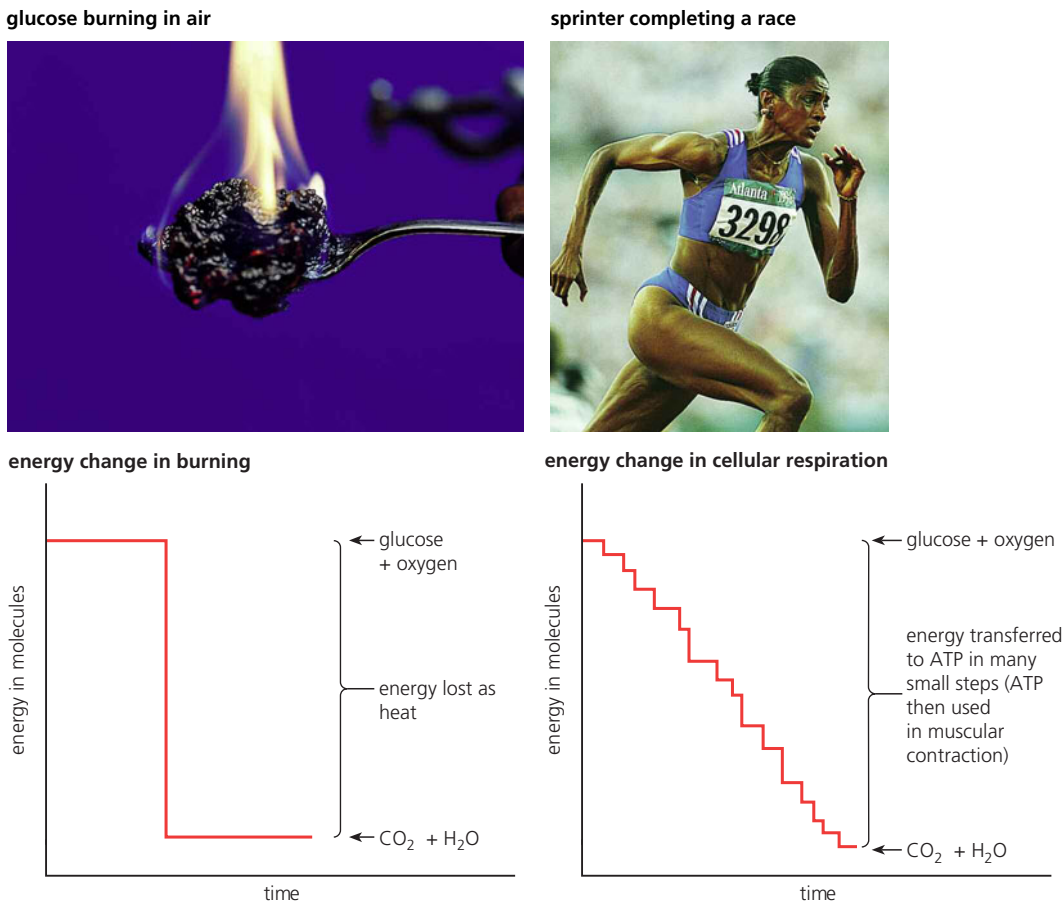


Figure 3.2 Combustion and respiration compared

ATP – the universal energy currency

Energy made available within the cytoplasm is transferred to a molecule called adenosine triphosphate (ATP). ATP is referred to as **energy currency** because, like money, it can be used in different contexts, and it is constantly recycled.

ATP is a **nucleotide** (page 63) with an unusual feature. It carries three phosphate groups linked together in a linear sequence (Figure 3.3). ATP may lose both of the outer phosphate groups, but usually only one at a time is lost. ATP is a relatively small, soluble organic molecule. It occurs in cells at a concentration of 0.5–2.5 mg cm^{-3}.

Like many organic molecules of its size, ATP contains a good deal of chemical energy locked up in its structure. What makes ATP special as a reservoir of stored chemical energy, is its role as a common intermediate between energy-yielding reactions and energy-requiring reactions and processes. Energy-yielding reactions include many of the individual steps in respiration. Energy-requiring reactions include the synthesis of cellulose from glucose, the synthesis of proteins from amino acids, and the contraction of muscle fibres.

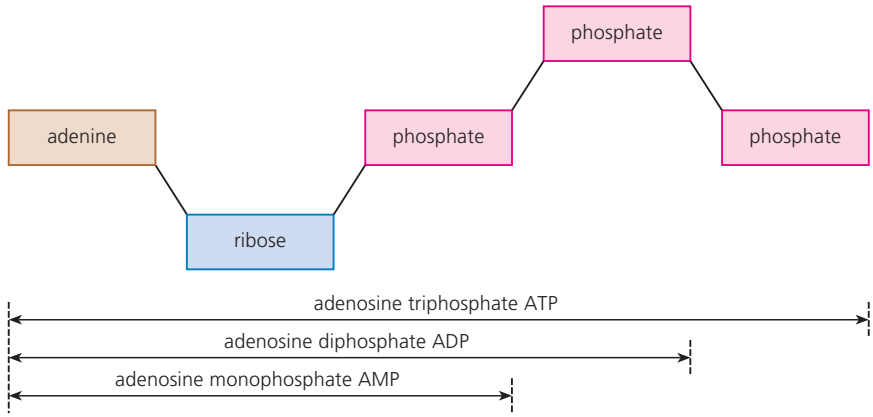


Figure 3.3 ATP, ADP and AMP

In summary, ATP is a molecule universal to all living things; it is the source of energy for chemical change in cells, tissues and organisms. The important features of ATP are that it is:

- a substance that moves easily within cells and organisms – by facilitated diffusion;
- able to take part in many steps of cellular respiration and in very many reactions of metabolism;
- able to deliver energy in relatively small amounts, sufficient to drive individual reactions.

The ATP–ADP cycle and metabolism

In cells, ATP is formed from **adenosine diphosphate (ADP)** and **phosphate ion (P_i)** using energy from respiration (Figure 3.4). Then, in the presence of enzymes, ATP participates in energy-requiring reactions. The free energy available in ATP is approximately 30 – 34 kJ mol^{-1}. Some of this energy is lost as heat in a reaction, but much free energy is made available to do useful work, more than sufficient to drive a typical energy-requiring reaction of metabolism (Figure 3.4).

Sometimes ATP reacts with water (a hydrolysis reaction) and is converted to ADP and P_i. For example, direct hydrolysis of the terminal phosphate groups happens in muscle contraction.

1 Outline why ATP is an efficient energy currency molecule.

Mostly, ATP reacts with other metabolites and forms phosphorylated intermediates, making them more reactive in the process. The phosphate groups are released later, so both ADP and P_i become available for re-use as metabolism continues.

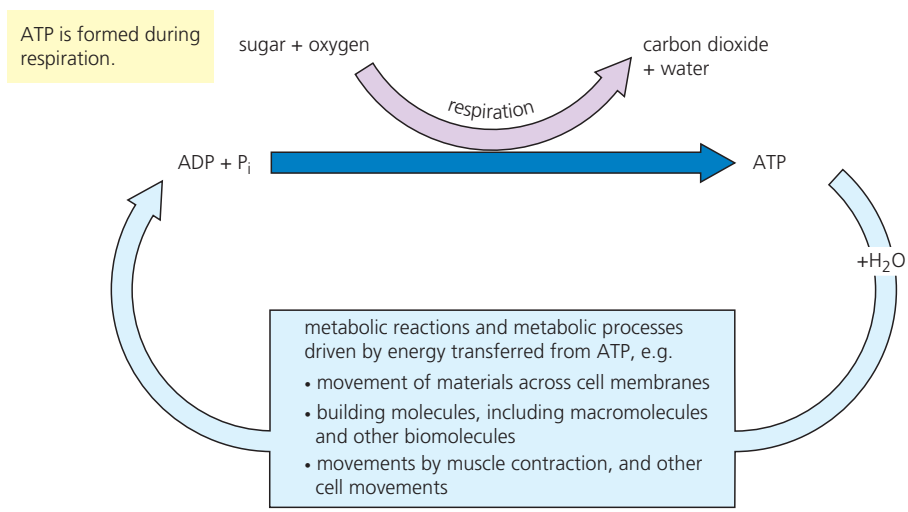


Figure 3.4 The ATP–ADP + P_i cycle

The steps of cellular respiration

In the first steps of cellular respiration, the glucose molecule (a 6-carbon sugar) is split into two 3-carbon molecules. Then the products are converted to an organic acid called pyruvic acid (also a 3-carbon compound). Under conditions in the cytoplasm, organic acids are weakly ionised, and therefore pyruvic acid exists as the **pyruvate** ion.

Obviously, two molecules of pyruvate are formed from each molecule of glucose. In addition, there is a small amount of ATP formed, using a little of the energy that had been locked up in the glucose molecule. No molecular oxygen is required for these first steps of cellular respiration.

Because glucose has been split into smaller molecules, these steps are known as **glycolysis** (glyco + lysis). The enzymes that catalyse these reactions are found in the cell cytoplasm generally, but not inside an organelle. Throughout cellular respiration, a series of oxidation–reduction reactions occur. (What oxidation and reduction consist of is explained on page 270.)

In summary:

$$\text{glucose} \xrightarrow{\text{enzymes in the cytoplasm}} \text{pyruvate + small amount of ATP}$$

Aerobic and anaerobic cellular respiration

While no oxygen is required for the formation of pyruvate by cells in the early steps of cellular respiration, most animals and plants and very many microorganisms do require oxygen for cell respiration, in total. We say that they respire aerobically.

In **aerobic cellular respiration**, sugar is completely oxidised to carbon dioxide and water and much energy is made available. The steps of aerobic respiration can be summarised by a single equation:

$$\text{glucose + oxygen} \longrightarrow \text{carbon dioxide + water + ENERGY}$$

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O + \text{ENERGY}$$

This equation is a balance sheet of the inputs (raw materials) and the outputs (products). It tells us nothing about the separate steps, each catalysed by a specific enzyme, by which cellular respiration occurs. It does not mention pyruvate, for example.

What happens to pyruvate in aerobic respiration?

If oxygen is available to cells and tissues, the pyruvate is completely oxidised to carbon dioxide, water and a large quantity of ATPs. Before these reactions take place, the pyruvate first passes into mitochondria by facilitated diffusion. This is because it is only in mitochondria that the required enzymes are found (Figure 3.5).

In summary:

$$\text{pyruvate} \xrightarrow{\text{enzymes in the mitochondria}} \text{carbon dioxide + water + large amount of ATP}$$

In this phase of cellular respiration, the pyruvate is oxidised by:

- removal of hydrogen atoms by hydrogen acceptors (oxidising agents);
- addition of oxygen to the carbon atoms to form carbon dioxide.

These reactions occur one at a time, each catalysed by a different enzyme.

Also, it is in the mitochondria that the reduced hydrogen acceptor molecules (these carry H) react with oxygen to form water. The reduced hydrogen acceptor is re-oxidised (loses its H), and is available for re-use in the production of more pyruvate. The majority of ATP molecules (the key product of respiration for the cell) are generated in this step, also.

The sites of cellular respiration

The enzymes of cellular respiration occur partly in the cytoplasm (the enzymes of glycolysis) and partly in the mitochondria (the enzymes of pyruvate oxidation and most ATP formation). After formation, ATP passes to all parts of the cell. Both ADP and ATP pass through the mitochondrial membranes by facilitated diffusion. The locations of the different stages of aerobic respiration are shown in Figure 3.5.

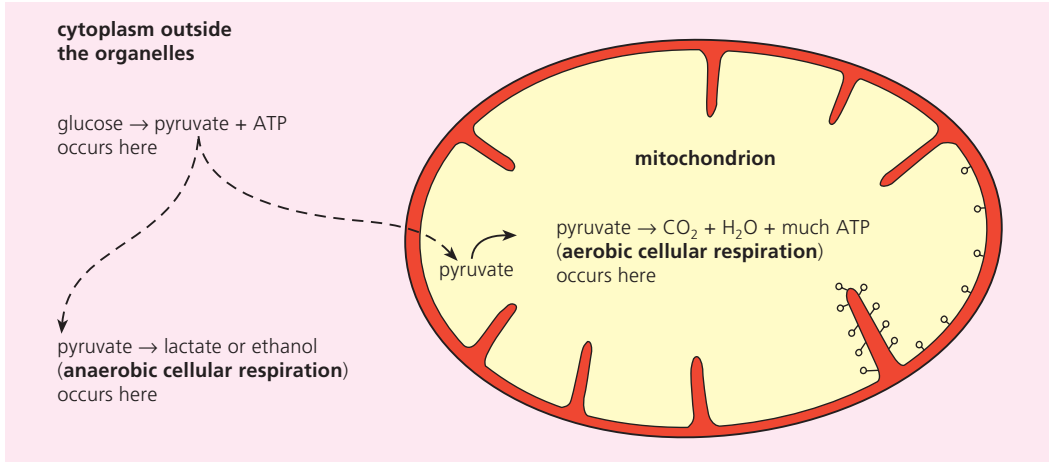


Figure 3.5 The sites of cellular respiration in cells

2 State which steps in cellular respiration occur whether or not oxygen is available to cells.

Anaerobic respiration – fermentation

In the absence of oxygen, many organisms (and sometimes tissues in organisms, when these have become deprived of oxygen) will continue to respire pyruvate by different pathways, known as fermentation or **anaerobic respiration**, at least for a short time.

Many species of yeast (*Saccharomyces*) respire anaerobically, even in the presence of oxygen. The products are ethanol and carbon dioxide. You will already be aware that **alcoholic fermentation** of yeast has been exploited by humans in wine and beer production for many thousands of years:

glucose ⟶ ethanol + carbon dioxide + ENERGY

Vertebrate muscle tissue can respire anaerobically, too, but in this case it involves the formation of lactic acid rather than ethanol. Once again, under conditions in the cytoplasm, lactic acid is weakly ionised, and therefore exists as the **lactate** ion.

Lactic acid fermentation occurs in muscle fibres, but only when the demand for energy for contractions is very great, and cannot be fully met by aerobic respiration. In lactic acid fermentation the sole waste product is lactate.

glucose ⟶ lactate + ENERGY

Extension:

Out of general interest we can also note that just a few organisms respire only in the absence of oxygen. One such is the tetanus bacillus *Clostridium tetani* which causes tetanus, also called lock-jaw. This bacterium is active in oxygen-free environments rich in nutrients, such as occur in deep body wounds. Otherwise the bacteria merely survive as resistant spores in the soil or among the faeces of the lower gut of many animals. *Clostridium tetani* (during growth, but not as a dormant spore) produces toxins which, when released inside wounds, are carried around the body in the blood circulation, and cause agonising muscular spasms. Today, tetanus may be prevented by vaccination, supported by periodic booster injections.

What happens to pyruvate in anaerobic respiration?

In human skeletal muscle tissue, when oxygen is not available, the pyruvate remains in the cytoplasm and is converted to lactate. In yeast, whether or not oxygen is available, the pyruvate is converted to the alcohol called ethanol.

How is the supply of pyruvate maintained in cells in the absence of oxygen?

This is the key issue in the breakdown of pyruvate in the absence of oxygen. Remember, in pyruvate formation, hydrogen acceptor molecules are reduced (take up hydrogen atoms). Without using oxygen, these must be re-oxidised (lose their H) if production of pyruvate is to continue.

The answer is that in anaerobic cellular respiration the reduced hydrogen acceptor molecules donate their hydrogen to form lactate or ethanol from pyruvate. This is how they are re-oxidised. In this way the acceptor molecules are available for further pyruvate synthesis. Pyruvate formation is able to continue.

3 Identify two products of anaerobic respiration in muscle.

We can summarise this as:

Anaerobic respiration is 'wasteful'

Anaerobic respiration is 'wasteful' of respiratory substrate. This is the case because the total energy yield per molecule of glucose respired, in terms of ATP generated, in both alcoholic and lactic acid fermentation is limited to the net two molecules of ATP produced in pyruvate formation. No additional energy is transferred in the latter steps and made available in cells.

So we can think of both lactate and ethanol as energy-rich molecules. (That is correct; ethanol is sometimes used as a fuel in cars, for example.) The energy locked up in these molecules may be used later, however. For example, in humans, lactate is transported to the liver and later metabolised aerobically. Energy yields of cellular respiration are compared in Table 3.1.

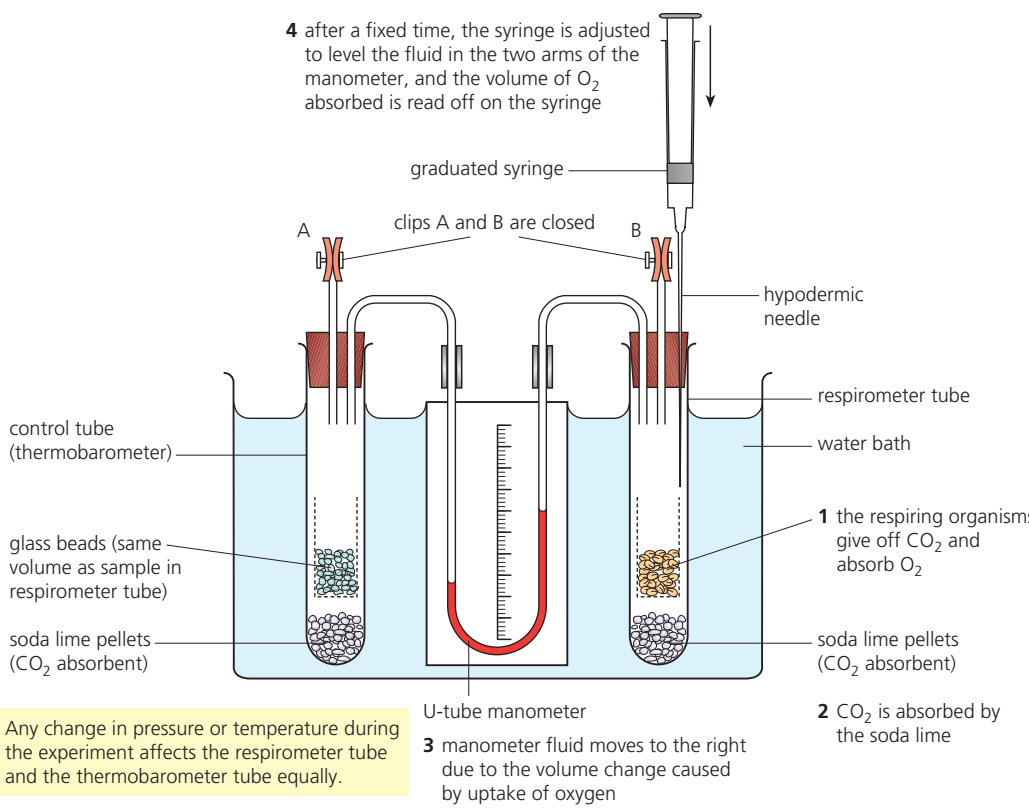
Table 3.1 Energy yield of aerobic and anaerobic cellular respiration compared

Yield from each molecule of glucose respired		
Aerobic respiration		**Anaerobic respiration**
2 ATPs	**glycolysis**	2 ATPs
up to 36 ATPs	**fates of pyruvate**	nil
38 ATPs	**Totals**	2 ATPs

Investigating respiration

The rate of respiration of an organism is an indication of its demand for energy. Respiration rate, the uptake of oxygen per unit time, may be measured by means of a **respirometer**. This apparatus, a form of manometer, detects change in pressure or volume of a gas. Respiration by tiny organisms (germinating seeds or fly maggots are ideal) trapped in the chamber of the respirometer alters the composition of the gas there, once the screw clip has been closed (Figure 3.6).

Figure 3.6 A respirometer to measure respiration rate



Any change in pressure or temperature during the experiment affects the respirometer tube and the thermobarometer tube equally.

Soda lime removes carbon dioxide gas as it is released by the respiring organism. Consequently, only oxygen uptake causes a change in volume. The bubble of coloured liquid in the attached capillary tube will move in response. The change in the volume of air in the respirometer tube, due to oxygen uptake, can be estimated from measurements of the movement of the manometric fluid during the experiment. In this apparatus, the volume of gas absorbed per unit time is given by readings on the syringe – the volume of air from the syringe that must be added to the respirometer tube to make the manometric fluid level in the two arms is equal to the volume of oxygen taken up by the respiring organisms.

4 In the respirometer (Figure 3.6), **explain** how changes in temperature or pressure in the external environment are prevented from interfering with measurement of oxygen uptake by respiring organisms in the apparatus.

Photosynthesis

3.8.1–3.8.8

Green plants use the energy of sunlight to produce **sugars** from the inorganic raw materials **carbon dioxide** and **water**, by a process called **photosynthesis**. The waste product is oxygen. Photosynthesis occurs in plant cells that contain the organelles called **chloroplasts** (page 6) – such as many of the cells of the leaves of green plants. Here energy of light is trapped by the green pigment **chlorophyll**, and becomes the chemical energy in molecules such as **glucose** and in **ATP**. Note that we now say that **light energy is transferred to organic compounds** in photosynthesis, rather than talking of the 'conversion of energy' – a term that was once widely used.

Sugar formed in photosynthesis may temporarily be stored as starch, but sooner or later most is used in metabolism. For example, plants manufacture other carbohydrates, together with the lipids, proteins, growth factors, and all other metabolites they require. For this they additionally need certain mineral ions, which are absorbed from the soil solution. Figure 3.7 is a summary of photosynthesis and its place in plant metabolism.

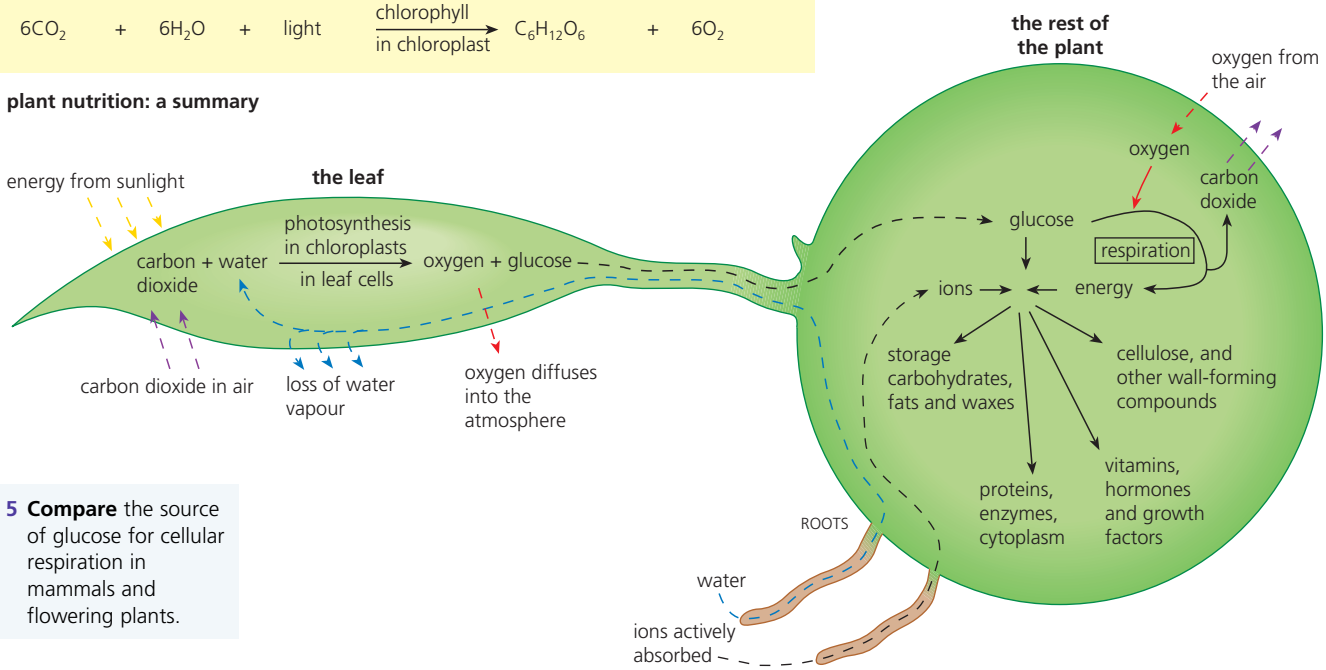
Figure 3.7 Photosynthesis and its place in plant nutrition

photosynthesis: a summary

The process in the chloroplast can be summarised by the equation:

carbon dioxide	+	water	+	LIGHT ENERGY	chlorophyll → in chloroplast	organic compounds, e.g. sugars	+	oxygen
raw materials				*energy source*		*products*		*waste product*
$6CO_2$	+	$6H_2O$	+	light	chlorophyll → in chloroplast	$C_6H_{12}O_6$	+	$6O_2$

plant nutrition: a summary



5 **Compare** the source of glucose for cellular respiration in mammals and flowering plants.

What is light?

Light is a form of the electromagnetic radiation produced by the Sun. Visible light forms only a part of the total magnetic radiation reaching the Earth. When this visible 'white' light is projected through a prism, we see a continuous spectrum of light – a rainbow of colours, from red to violet. Different colours have different wavelengths. The wavelengths of electromagnetic radiation and of the components of light are shown in Figure 3.8.

The significance of the spectrum of light in photosynthesis is that not all the colours of the spectrum present in white light are absorbed equally by chlorophyll. Some are even transmitted (or reflected), rather than being absorbed.

Which of these individual colours does chlorophyll absorb and then exploit as the energy source for photosynthesis?

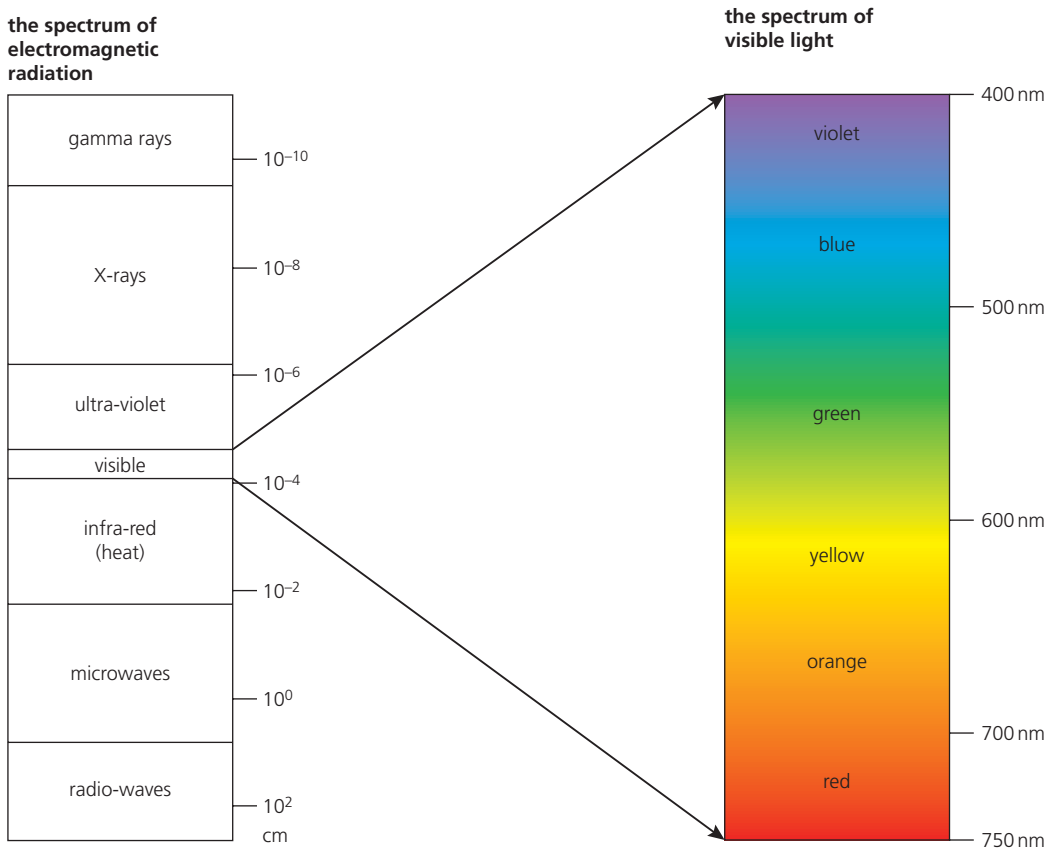


Figure 3.8 Electromagnetic radiation and the spectrum of visible light

Investigating chlorophyll

Some plant pigments are soluble in water, but chlorophyll is not. While we cannot extract chlorophyll from leaves with water, it can be extracted by dissolving in an organic solvent like propanone (acetone). With this solvent, extraction of chlorophyll is straightforward. Figure 3.9 shows how it may be done.

With chlorophyll in solution, the colours of light it absorbs can be investigated. We have seen that white light consists of a roughly equal mixture of violet, blue, green, yellow, orange and red light (Figure 3.8). When these different colours are projected through the chlorophyll solution in turn, we see that the greatest absorption occurs in the blue and red parts of the spectrum, whereas green light is transmitted or reflected (Figure 3.10).

The chemical structure of the chlorophyll molecule allows absorption of the energy of blue and red light. When extracted from the chloroplasts and dissolved in organic solvent, chlorophyll still absorbs light, but it cannot now use the energy gained to cause photosynthesis to take place. This is because, in the extraction process, the chlorophyll has been separated from the membrane systems and enzyme systems that surround it in chloroplasts, and which are essential for carrying out the biochemical steps of photosynthesis.

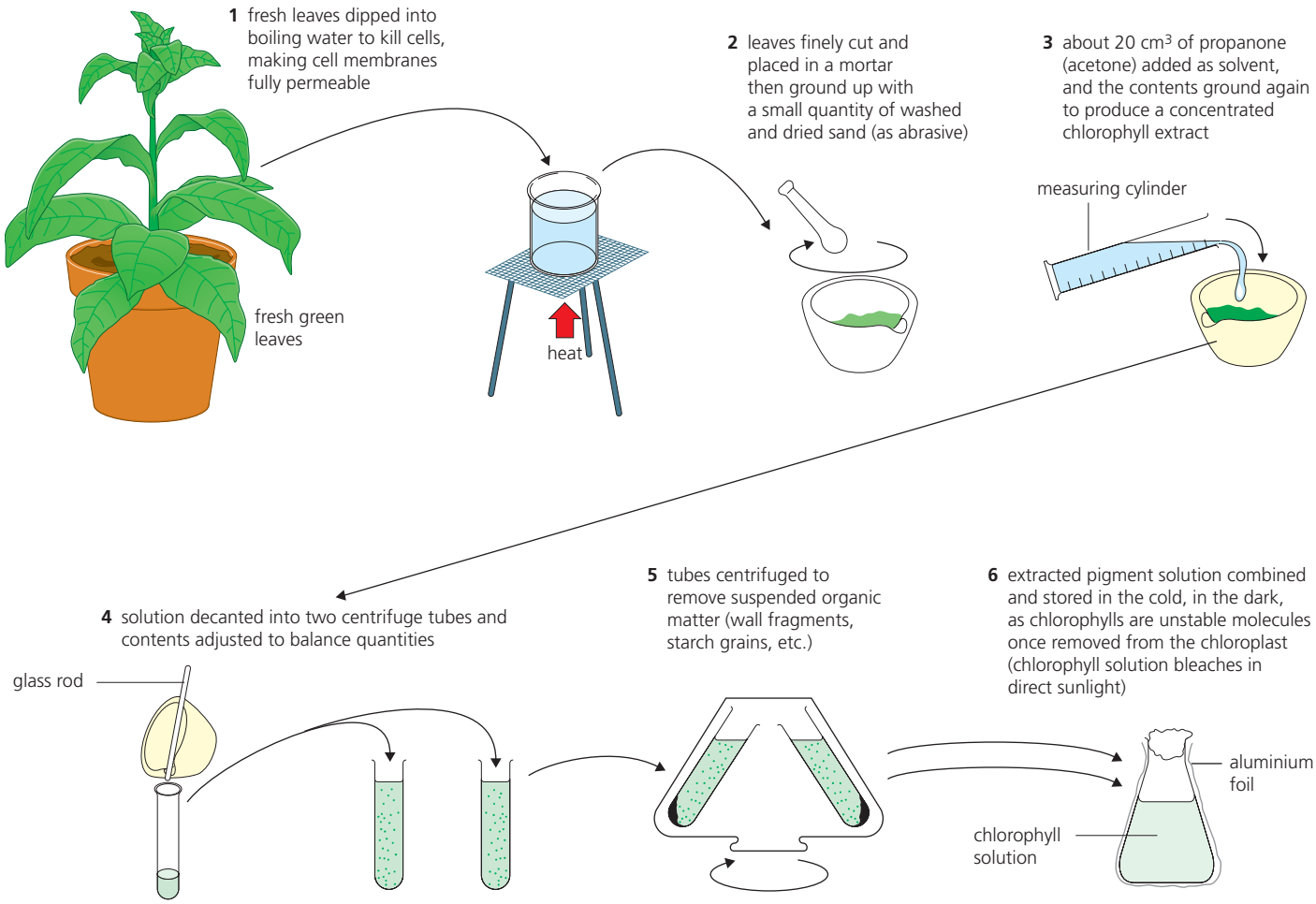


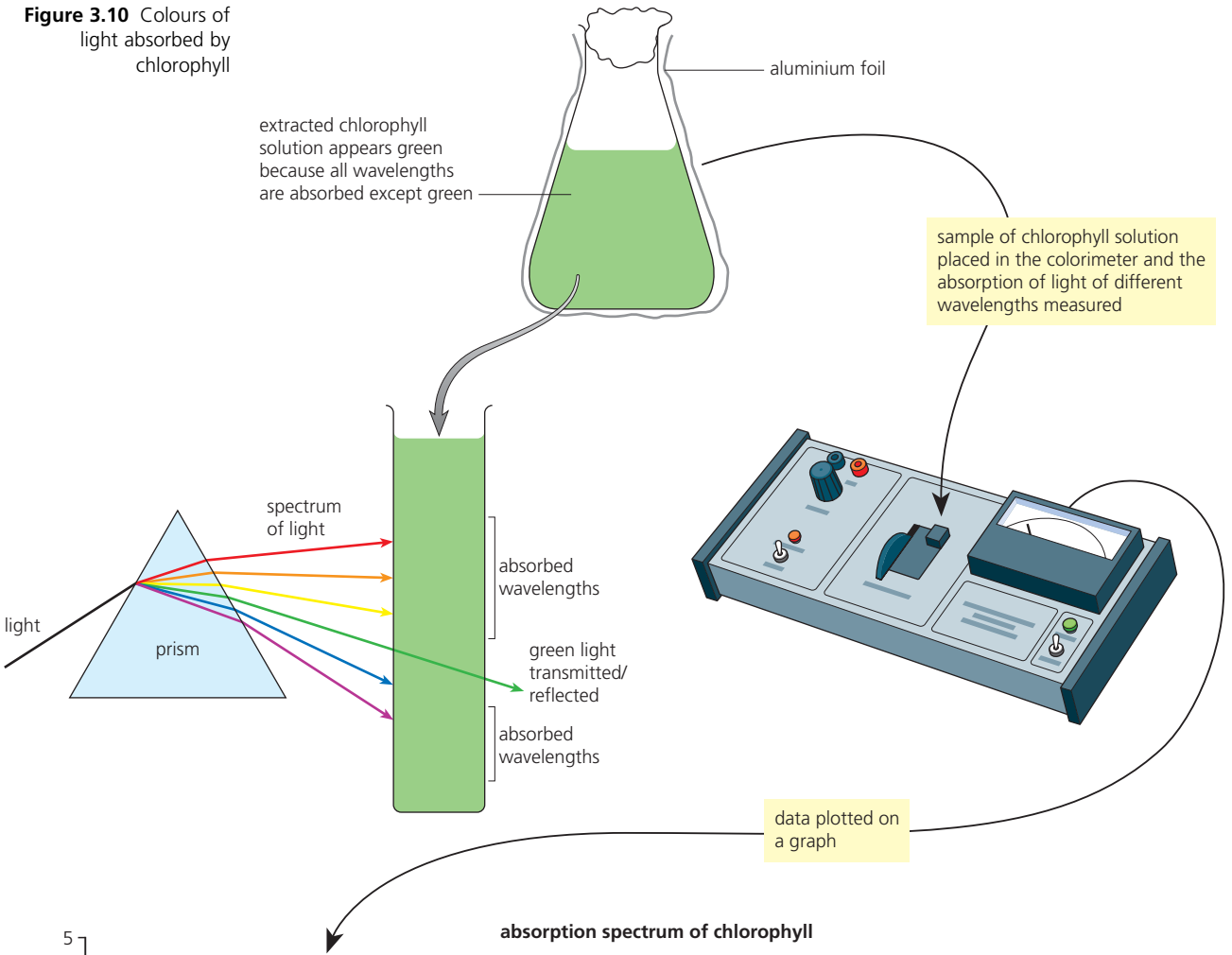
Figure 3.9 Steps in the extraction of chlorophyll

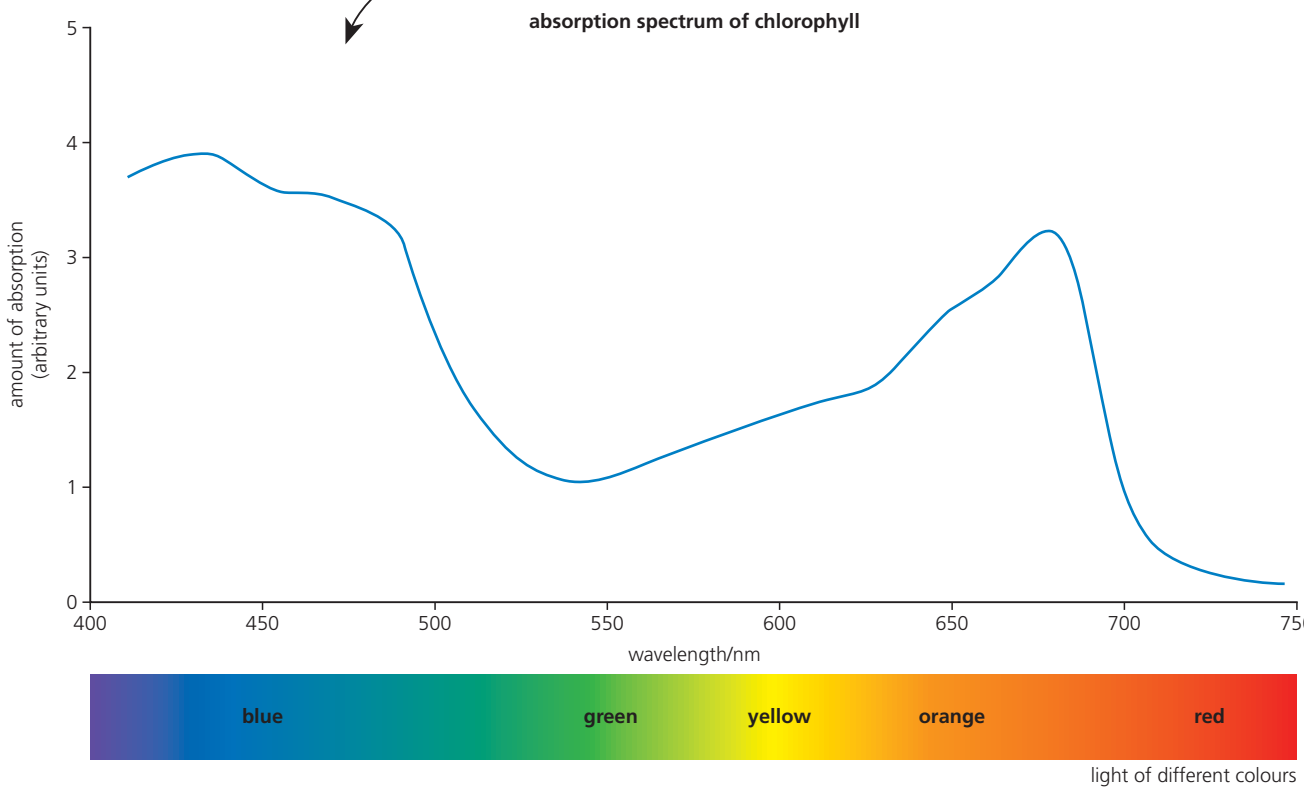
Green plants suitable to use for chlorophyll extraction

- spinach
- bougainvillea
- hibiscus
- green grass
- in Asia, the leaf vegetables kai lan or kang kong

6 **Evaluate** the essential safety precautions required when chlorophyll is extracted as shown in Figure 3.9.

Figure 3.10 Colours of light absorbed by chlorophyll

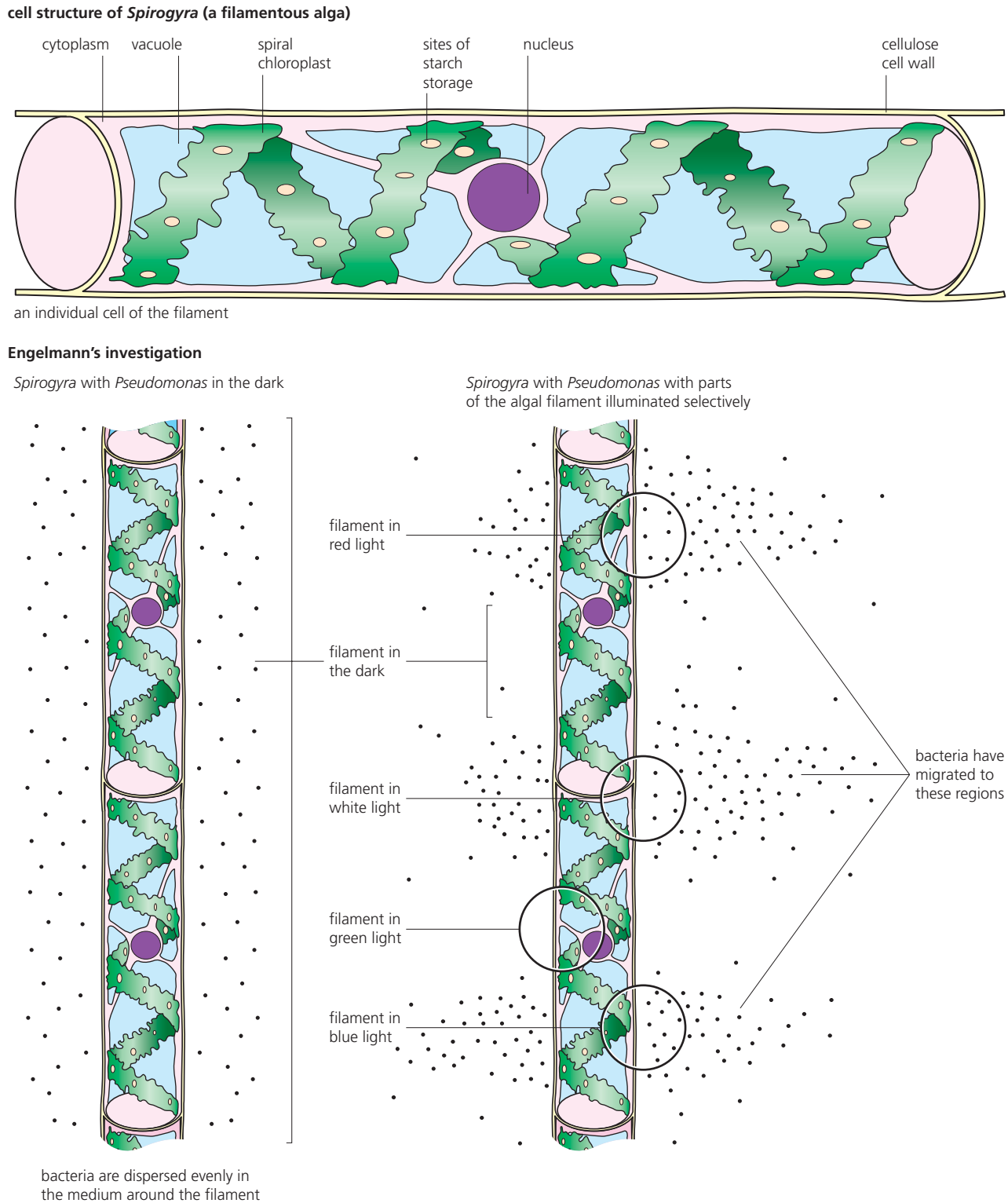




Extension: Which wavelengths are used in photosynthesis?

We can see that chlorophyll absorbs light energy most strongly in the blue and red ranges of the spectrum. How can we tell that this light energy, once absorbed, is being used to drive photosynthesis in chloroplasts?

Figure 3.11 Engelmann's investigation of light and photosynthesis



This question was first resolved by a German biologist, T. W. Engelmann, in 1882. Using filamentous algae such as *Spirogyra* (an aquatic, photosynthetic organism), he shone fine beams of white light and of light of different wavelengths onto different parts of a filament of the alga – so different parts were exposed to blue, green or red light, or to full (white) light. The other parts of the filament were in darkness.

Present in the water around the filament was a population of a motile bacterium (*Pseudomonas* sp.). This is a strongly aerobic species – the bacterium requires a supply of oxygen to survive, and migrates away from regions where oxygen is scarce.

Where algal cells are carrying out photosynthesis, oxygen diffuses out and is dissolved in the water immediately around the cell. Engelmann found that the bacteria migrated to only certain of the illuminated regions, and were visible there when the filaments were examined microscopically. They occurred around the parts illuminated with red and blue light only (Figure 3.11). He concluded that it is these wavelengths that are used in photosynthesis.

What happens in photosynthesis?

Photosynthesis is a set of many reactions occurring in chloroplasts in the light. However, these can be conveniently divided into two main steps:

1 **Light energy is used to split water (photolysis)**
This releases the waste product of photosynthesis, oxygen, and allows the hydrogen atoms released to be retained on hydrogen acceptor molecules. The hydrogen is one requirement of step 2. At the same time, ATP is generated from ADP and phosphate, also using energy from light.

2 **Sugars are built up from carbon dioxide**
We say that carbon dioxide is fixed to make organic molecules. To do this, both energy of ATP and hydrogen atoms from the reduced hydrogen acceptor molecules are required.

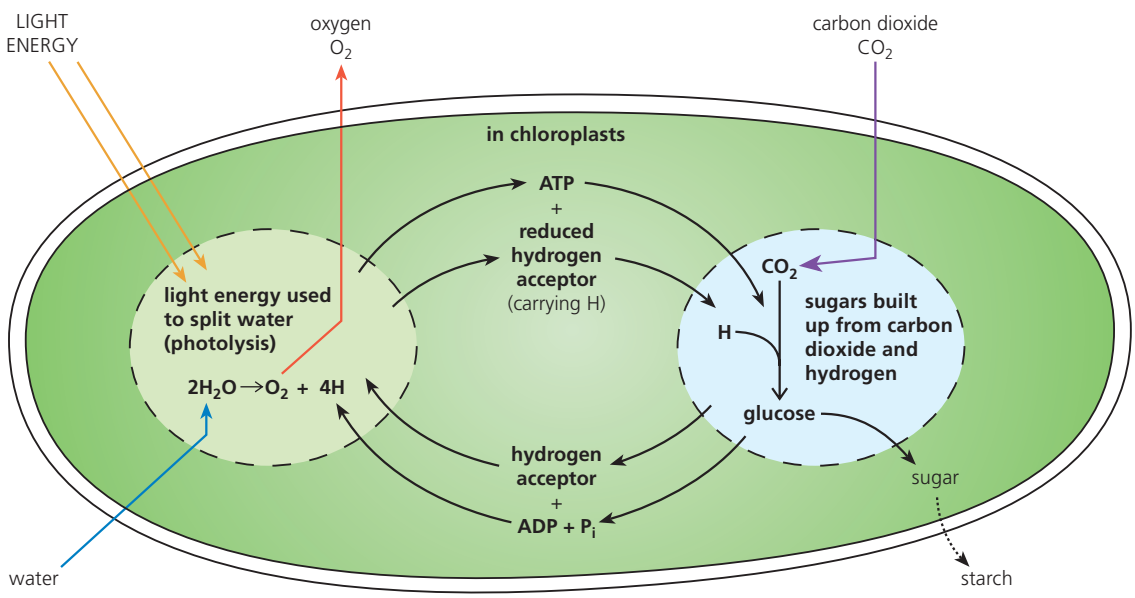


Figure 3.12 Two steps of photosynthesis

7 **State** from where a green plant obtains **a** carbon and **b** hydrogen used to build up glucose molecules.

Measuring the rate of photosynthesis

Measuring the rate of photosynthesis is a straightforward task in the laboratory. The rate can be estimated indirectly, by measuring the **increase in biomass** in samples of illuminated leaves, since most of the sugar produced in the light is immediately stored as starch in the cells. This technique is illustrated in Figure 3.13. It is best to destarch the leaves of the plant beforehand by keeping the intact plant in the dark for 48 hours first. The result is a more accurate estimate of gain in mass due to photosynthesis.

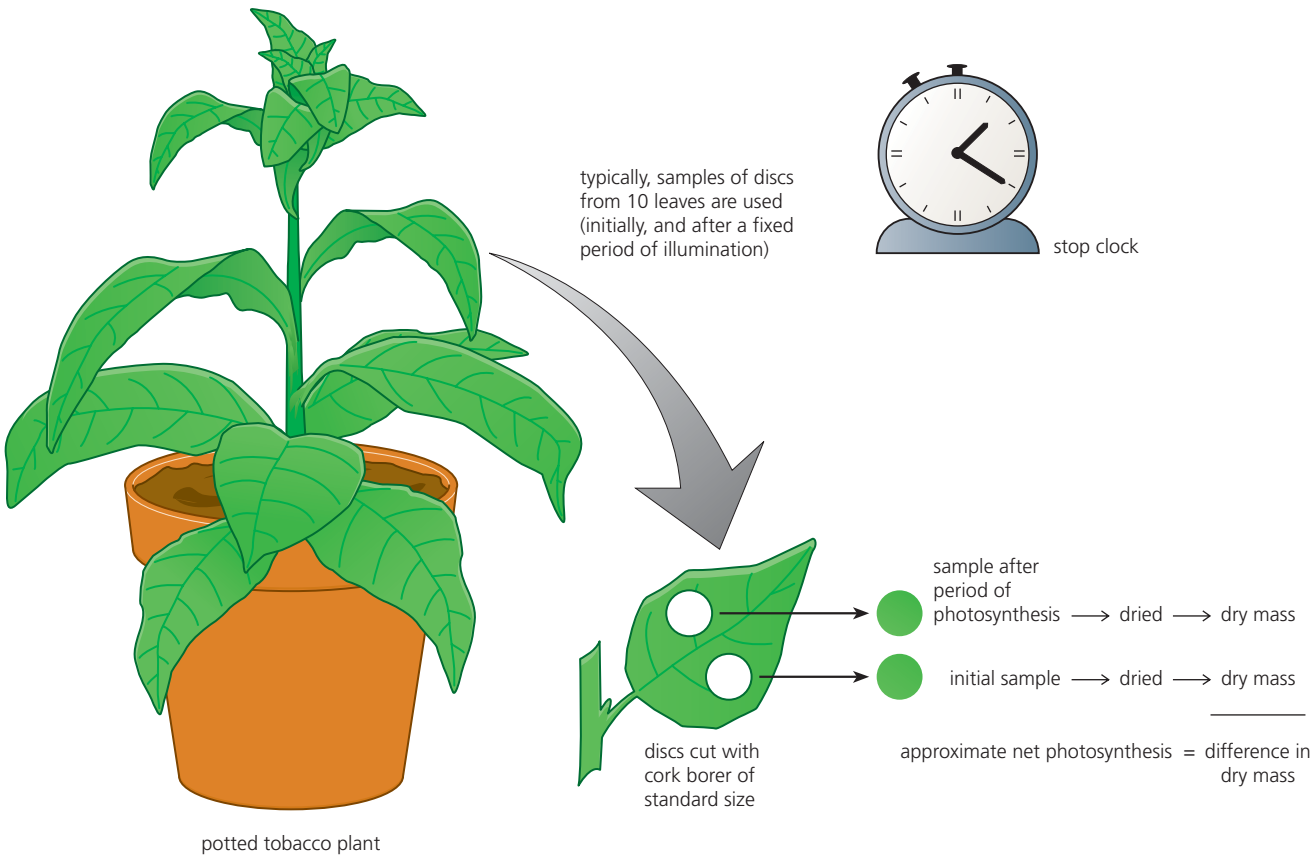


Figure 3.13 Rate of photosynthesis estimated by dry mass increase in leaf discs

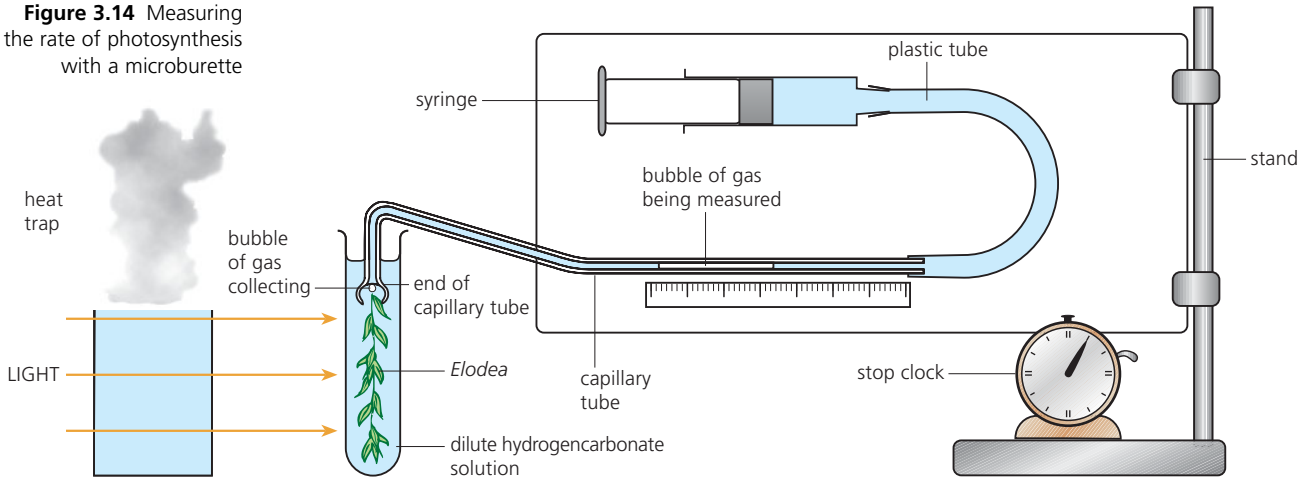
8 Suggest the likely fate of starch stored in a green leaf, during the destarching process.

The rate of photosynthesis can also be estimated using an oxygen sensor probe connected to a data-logging device or, using a microburette, measured as the **volume of oxygen given out in the light** by an aquatic plant (Figure 3.14). The experiment requires a freshly cut shoot of a pondweed which, when inverted, produces a vigorous stream of gas bubbles from the base. The bubbles tell us the pondweed is actively photosynthesising. The pondweed is placed in a very dilute solution of sodium hydrogencarbonate, which supplies the carbon dioxide (as —HCO_3 ions) required by the plant for photosynthesis. The quantity of gas evolved in a given time, say in 30 minutes, is measured by drawing the gas bubble that collects into the capillary tube, and measuring its length. This length is then converted to a volume.

Alternatively, the **uptake of carbon dioxide in the light** can be measured. Using an aquatic plant, the dissolved carbon dioxide absorbed from the water will cause the pH to rise and this may be monitored with a pH meter connected to a data-logging monitor.

Using this apparatus the effects of external conditions such as light intensity, carbon dioxide concentration and temperature on the rate of photosynthesis have been investigated.

Figure 3.14 Measuring the rate of photosynthesis with a microburette



9 A thermometer is not shown in the apparatus in Figure 3.14. **Predict** why one is required, and **state** where.

External factors and the rate of photosynthesis

The effect of the **concentration of carbon dioxide** on the rate of photosynthesis is shown in Figure 3.15. Look at this figure now – note the shape of the curve.

In this experiment, when the concentration of carbon dioxide is at zero there is no photosynthesis, of course. As the concentration is steadily increased, the rate of photosynthesis rises, and the rate of that rise is positively correlated with the increasing carbon dioxide concentration. However, at much higher concentrations of carbon dioxide, the rate of photosynthesis reaches a plateau – now there is no increase in rate with rising carbon dioxide concentration.

The effect of **light intensity** on the rate is shown in Figure 3.16. Look at this graph now – the shape of the curve is familiar.

In this experiment, when light intensity is low, the rate of photosynthesis is directly proportional to light intensity. However, at much higher light intensities, there is no increase in rate with rising light intensity – this rate had reached a plateau, too.

Figure 3.15 The effect of carbon dioxide concentration on photosynthesis

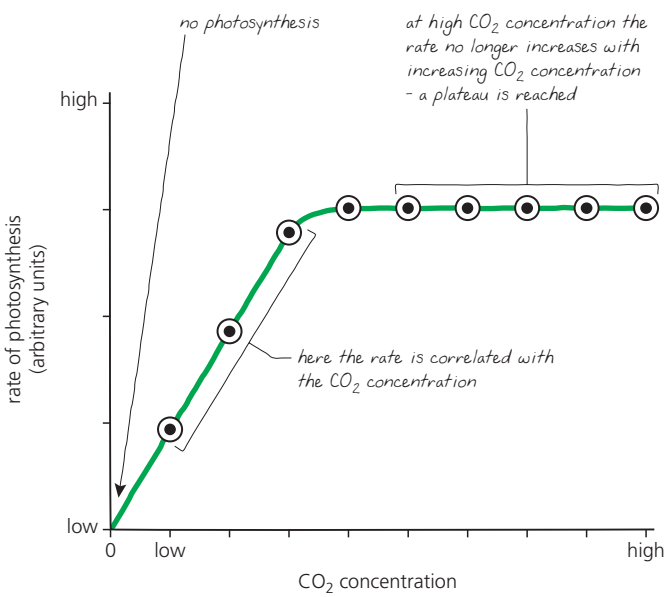
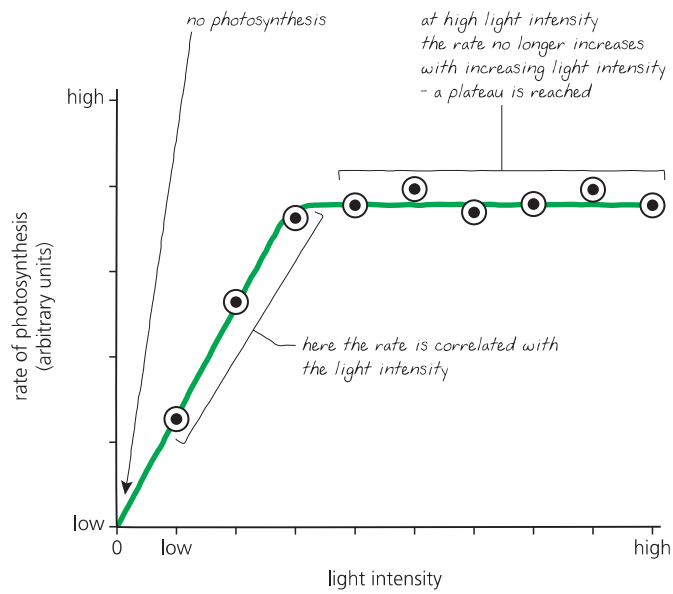


Figure 3.16 The effect of light intensity on photosynthesis



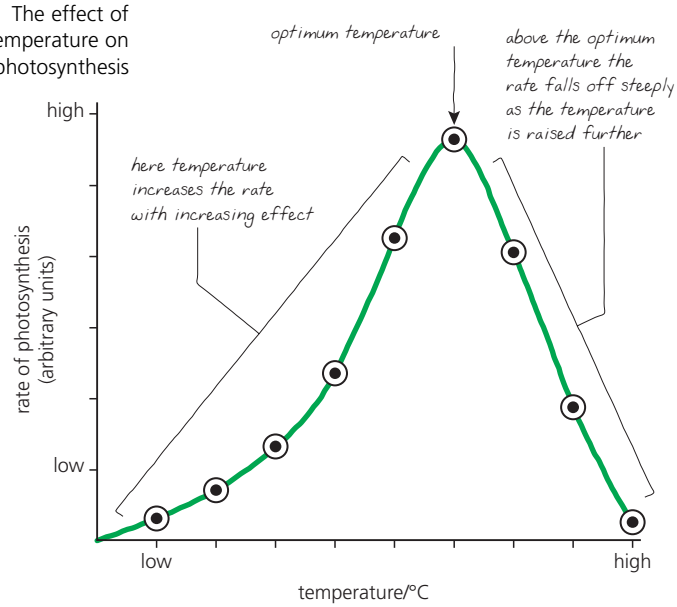


Figure 3.17 The effect of temperature on photosynthesis

The effect of **temperature** on the rate of photosynthesis is shown in Figure 3.17. Here the curve of the graph is an entirely different shape. At relatively low temperatures, as the temperature increases, the rate of photosynthesis increases more and more steeply. However, at higher temperatures, the rate of photosynthesis abruptly stops rising and actually falls steeply. The result is a clear optimum temperature for photosynthesis.

10 Suggest why we should expect that high external temperatures cause the rate of photosynthesis to fall off very rapidly.

Examination questions – a selection

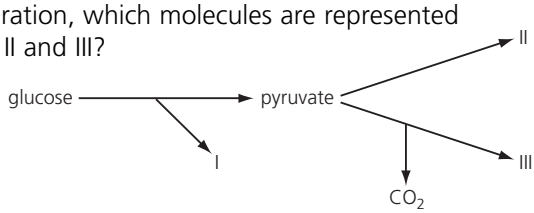
Questions 1–3 are taken from past IB Diploma biology papers.

Q1 In which process of photosynthesis is light directly involved?

A conversion of ATP to ADP
B the fixing of carbon
C the splitting of water
D the conversion of pyruvate to ethanol

Standard Level Paper 1, May 06, Q11

Q2 In the following generalised diagram of anaerobic respiration, which molecules are represented by I, II and III?



	I	II	III
A	ATP	ethanol	lactate
B	ethanol	ATP	lactate
C	lactate	ethanol	ATP
D	ATP	lactate	ethanol

Higher Level Paper 1, May 05, Q5

Q3 **a** State the main photosynthesis pigment in plants. (1)
b State the **two** materials used to convert carbon dioxide to organic molecules in plants. (2)
c Explain **two** ways in which the rate of photosynthesis can be measured. (4)

Standard Level Paper 2, May 04, Q4

Questions 4–8 cover other syllabus issues in this chapter.

Q4 Define cell respiration. Include the essential raw materials, *and* the products. (2)
State where respiration occurs in a living organism. (2)

Q5 Explain the composition of ATP and the role of this molecule in cell metabolism. (2)

Q6 Construct a flow diagram to show the steps in aerobic cellular respiration. Demonstrate which steps are restricted to the cytoplasm, and which to a mitochondrion, and the connection between these steps. (6)

Q7 Photosynthesis involves the transfer of light energy to chemical energy of glucose and other organic molecules.

a State why photosynthesis is restricted to green plants in the light. (1)
b State the waste product of photosynthesis. (1)
c Define the term 'photolysis'. (1)
d Photosynthesis leads to an increase in biomass of a green plant. Suggest the metabolic process mostly responsible for reducing the amount of biomass present in the plant. (1)

Q8 Design an experiment to test whether temperature has an effect on photosynthesis. Predict what outcome you would expect. (6)

4 Genetics

STARTING POINTS

- The **nucleus** is the largest organelle in a cell. It contains the **chromosomes** on which **genes**, the blueprints (coded instructions) for the cell, occur.
- When cells divide, the **nucleus divides first**, and each new (daughter) cell receives a nucleus.
- The nucleus **controls and directs the activities** of the cell throughout life; it **contains the hereditary material** which is passed from generation to generation during reproduction.

Genetics is the study of inheritance and of variation in the inherited characteristics that **chromosomes** control. The nucleus contains the chromosomes, as we have already seen. The chromosomes hold the blueprint (coded instructions) for the organisation and activities of cells and for the whole organism, in the form of **genes**. Genes are involved both in the continuous control of cell activity and in reproduction, carrying instructions from one generation to the next.

Chromosomes, genes, mutations and meiosis

4.1.1–4.1.4, 4.2.1–4.2.7

Chromosomes can be seen in cells by staining them with certain dyes (Figure 1.33, page 34). At the time a nucleus divides, the chromosomes become compact, much-coiled structures, and are clearly visible. At all other times, the chromosomes are very long thin, uncoiled threads. In this condition, they give the stained nucleus a granular appearance. The granules are called chromatin.

Each chromosome consists of **DNA**, a huge molecule consisting of **two paired strands in the form of a double helix** (Figure 2.27, page 64). The DNA molecule runs the full length of the chromosome, and is supported by protein. About 50% of a chromosome is built of protein. While some of these proteins are enzymes that are involved in copying and repair reactions of DNA, the bulk of chromosome protein has a support and packaging role for DNA.

There are five features of the chromosomes of organisms which it is helpful to note now.

1 **The number of chromosomes per species is fixed**
The number of chromosomes in the cells of different species varies, but in any one species the number of chromosomes per cell is normally constant. For example, the mouse has 40 chromosomes per cell, the onion has 16, humans have 46, and the sunflower has 34. Each species has a characteristic chromosome number. Note, these are all even numbers.

2 **The shape of a chromosome is characteristic**
Chromosomes are long thin structures of a particular, fixed length. Somewhere along the length of the chromosome is a narrow region called the **centromere**. Centromeres may occur anywhere along the chromosome, but they are always in the same position on any given chromosome. The position of the centromere, as well as the length of a chromosome, is how they are identified in photomicrographs (photographs taken down a light microscope, see page 10).

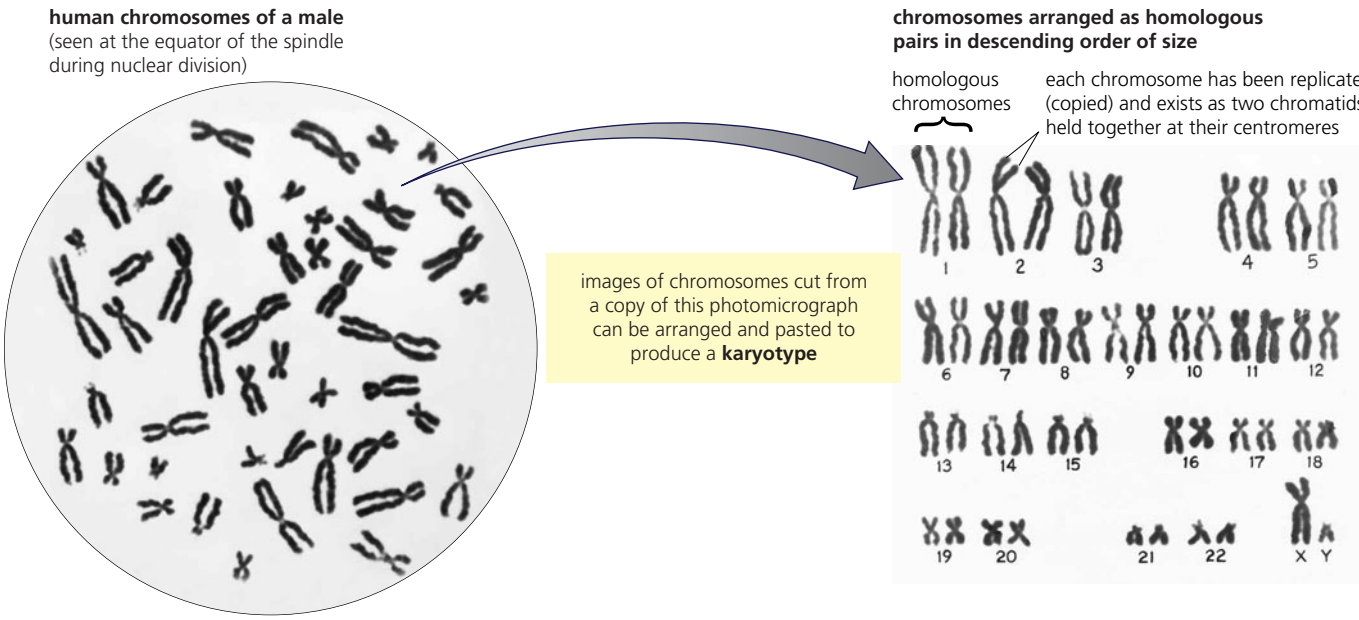
3 **The chromosomes of a cell occur in pairs called homologous pairs**
The word 'homologous' just means 'similar in structure'. One of each pair came originally from one parent, and the other from the other parent. So, for example, the human has 46 chromosomes, 23 coming originally from each parent in the process of sexual reproduction. This is why chromosomes occur in homologous pairs.

Homologous chromosomes resemble each other in structure. They occur in a diploid cell, contain the same sequence of genes, but have come from different parents.

You can see this from the photomicrographs in Figure 4.1. Here, chromosomes of a human male are shown at a stage of the nuclear division called **mitosis** (page 31). They are cut out of a photomicrograph and arranged in numbered descending order of size. This type of enlarged photographic image of the mitotic chromosomes of a cell, arranged in pairs in order of size, is called a **karyotype**.

The procedure of karyotyping is used by genetic counsellors, in conjunction with other techniques, to detect the presence of (rare) abnormalities, such as Down's syndrome (page 98), in a patient's chromosome set. This is discussed later in this chapter.

Figure 4.1 Chromosomes as homologous pairs, as seen during nuclear division



4 Chromosomes hold the hereditary factors – genes

The chromosomes are effectively a linear series of genes (Figure 4.2). We can define 'gene' in different ways. For example, a gene is a:

a specific region of a chromosome which is capable of determining the development of a specific characteristic of an organism;
b specific length of the DNA double helix, hundreds or (more typically) thousands of base pairs long, which codes for a protein;
c unit of inheritance.

A particular gene always occurs on the same chromosome in the same position.

The position of a gene is called a **locus** (plural, **loci**).

Each gene has two or more forms, called **alleles**. The word 'allele' just means 'alternative form'.

The total of all the genetic information – all the genes present on an organism's chromosomes – is called the **genome** of the organism.

5 Chromosomes copy themselves

Between nuclear divisions, while the chromosomes are uncoiled and cannot be seen, each chromosome makes a copy of itself. It is said to **replicate**.

Replication occurs in the cell cycle, during interphase (page 31). The two identical structures formed are called **chromatids** (Figure 4.3). The chromatids remain attached by their centromeres until they are divided during nuclear division. After division of the centromeres, the chromatids are recognised as chromosomes again.

Of course, when chromosomes copy themselves, the critical event is the copying of the DNA double helix that runs the length of the chromosome. Replication occurs in a very precise way, brought about by specific enzymes, as we have already discussed (page 66).

1 **Explain** why chromosomes occur in homologous pairs in cells.

Figure 4.2 Genes and alleles of a homologous pair of chromosomes

The **loci** are the positions along the chromosomes where genes occur, so alleles of the same gene occupy the same locus.

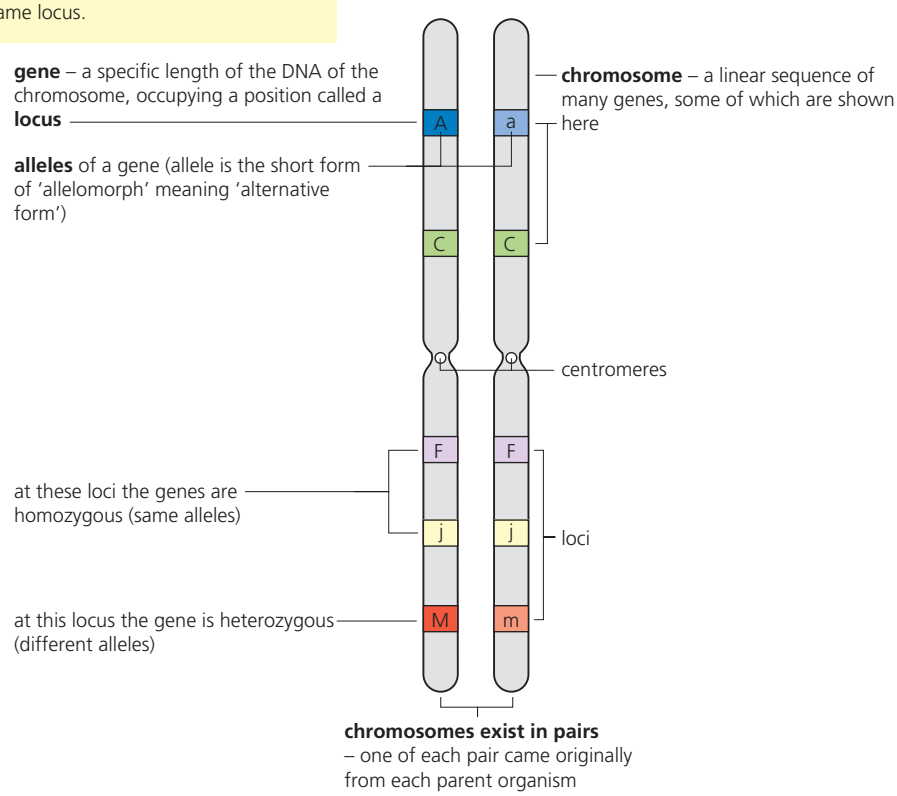
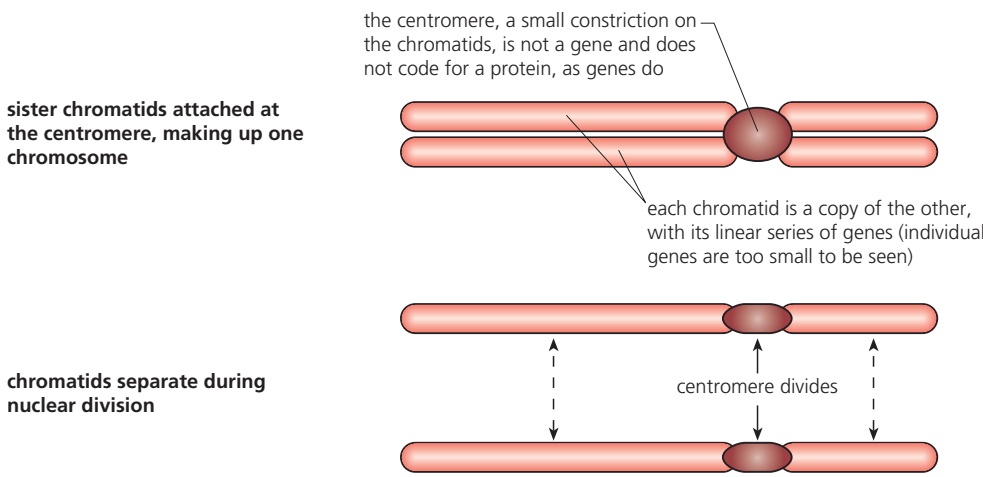


Figure 4.3 One chromosome as two chromatids



Nuclear divisions and the fate of chromosomes

Divisions of the nucleus are very precise processes, ensuring the correct distribution of chromosomes between the new cells (daughter cells). There are two types of nuclear division, known as mitosis and meiosis.

In **mitosis**, the daughter cells produced have the same number of chromosomes as the parent cell, typically two of each type, known as the **diploid** (**2*n***) state. Mitosis is the nuclear division that occurs when cells grow, old cells are replaced, and when an organism reproduces asexually. The way mitosis works is explained in Chapter 1, page 32.

In **meiosis**, the daughter cells contain half the number of chromosomes of the parent cell. That is, one chromosome of each type is present in the nuclei formed; this is known as the **haploid (*n*)** state. Meiosis is the nuclear division that occurs when sexual reproduction occurs, normally during the formation of the gametes.

The differences between mitosis and meiosis are summarised in Figure 4.4.

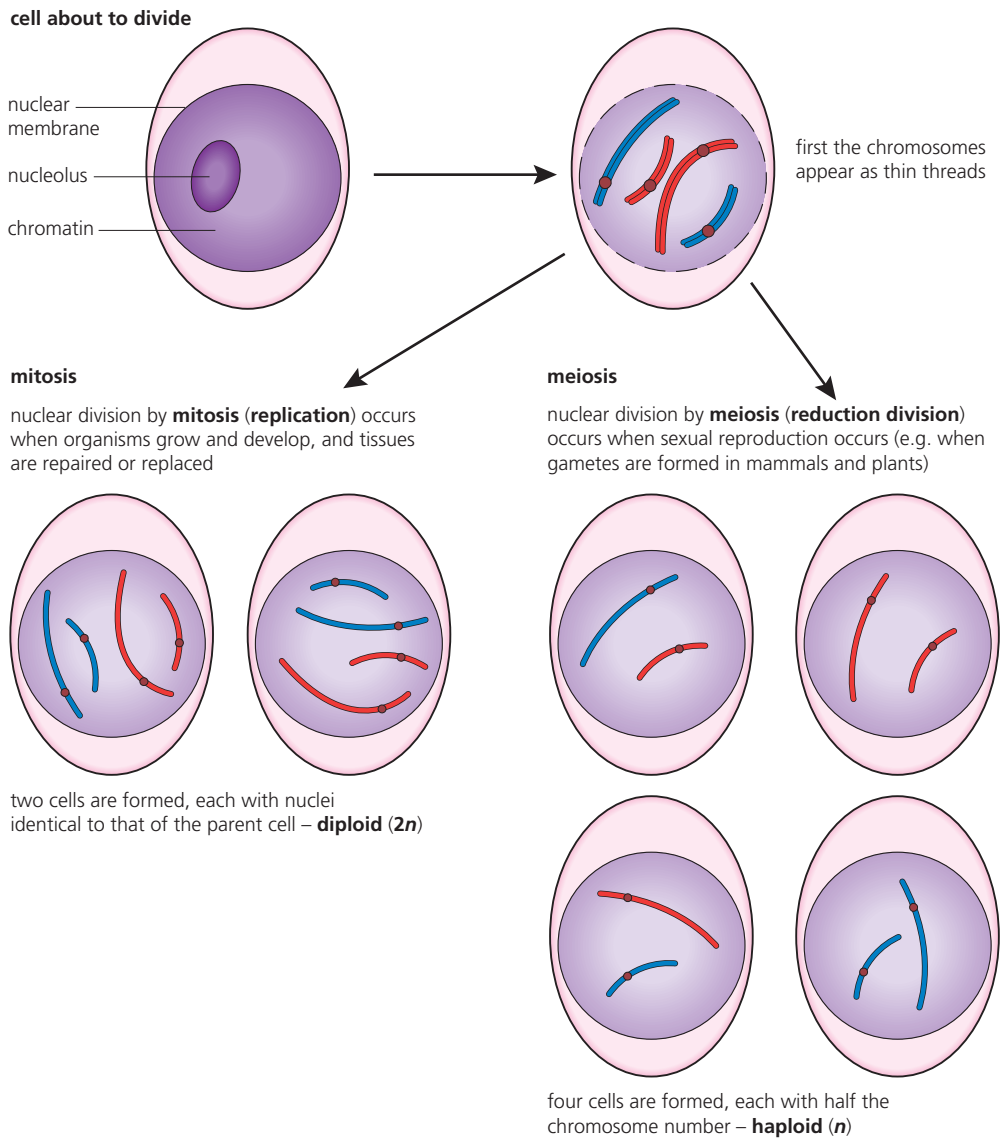


Figure 4.4 Mitosis and meiosis, the significant differences

2 **Suggest** why it is essential that nuclear division is a precise process.

Meiosis, the reduction division

Meiosis is part of the life-cycle of every organism that reproduces sexually. In meiosis, four daughter cells are produced – each with half the number of chromosomes of the parent cell. Halving of the chromosome number of gametes is essential because at fertilisation the number is doubled.

How does meiosis work?

Meiosis involves **two divisions** of the nucleus, known as **meiosis I** and **meiosis II**, both of which superficially resemble mitosis. As in mitosis, chromosomes replicate to form chromatids during interphase, before meiosis occurs. Then, early in meiosis I, **homologous chromosomes pair up**. By the end of meiosis I, homologous chromosomes have separated again, but the chromatids they consist of do not separate until meiosis II. Thus, meiosis consists of two nuclear divisions but only **one replication of the chromosomes**.

The process of meiosis

The key events of meiosis are summarised in Figure 4.5.

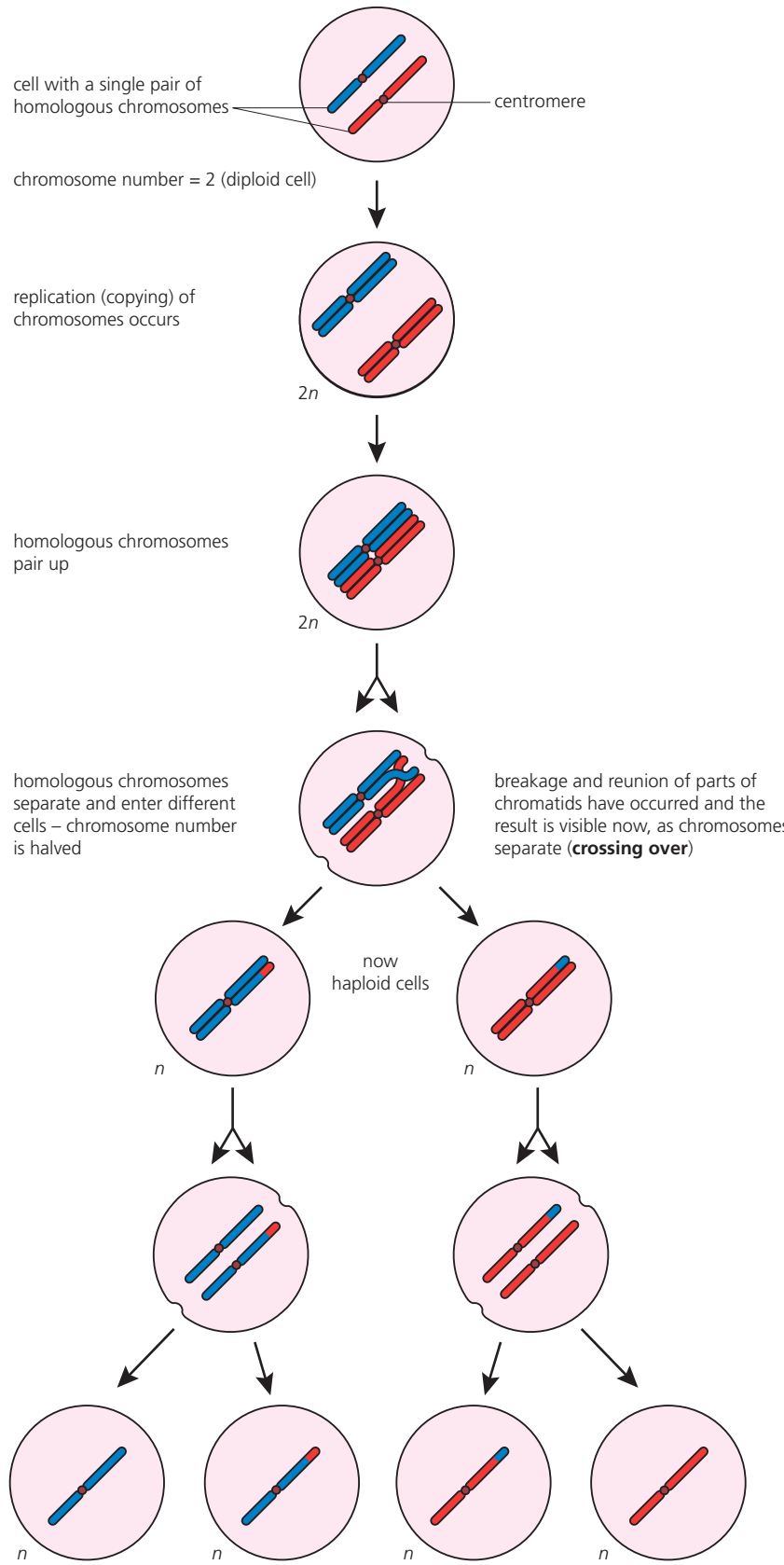


Figure 4.5 What happens to chromosomes in meiosis

In the interphase that precedes meiosis, the chromosomes are replicated as chromatids, but between meiosis I and II there is no further interphase, so no replication of the chromosomes occurs during meiosis.

As meiosis begins, the chromosomes become visible. At the same time, homologous chromosomes pair up. (Remember, in a diploid cell each chromosome has a partner that is the same length and shape and with the same linear sequence of genes. It is these partner chromosomes that pair.)

When the homologous chromosomes have paired up closely, each pair is called a **bivalent**. Members of the bivalent continue to shorten.

During the coiling and shortening process within the bivalent, the **chromatids frequently break**. Broken ends rejoin more or less immediately. When non-sister chromatids from homologous chromosomes break and rejoin they do so at exactly corresponding sites, so that a cross-shaped structure called a chiasma is formed at one or more places along a bivalent. The event is known as a **crossing over** because lengths of genes have been exchanged between chromatids.

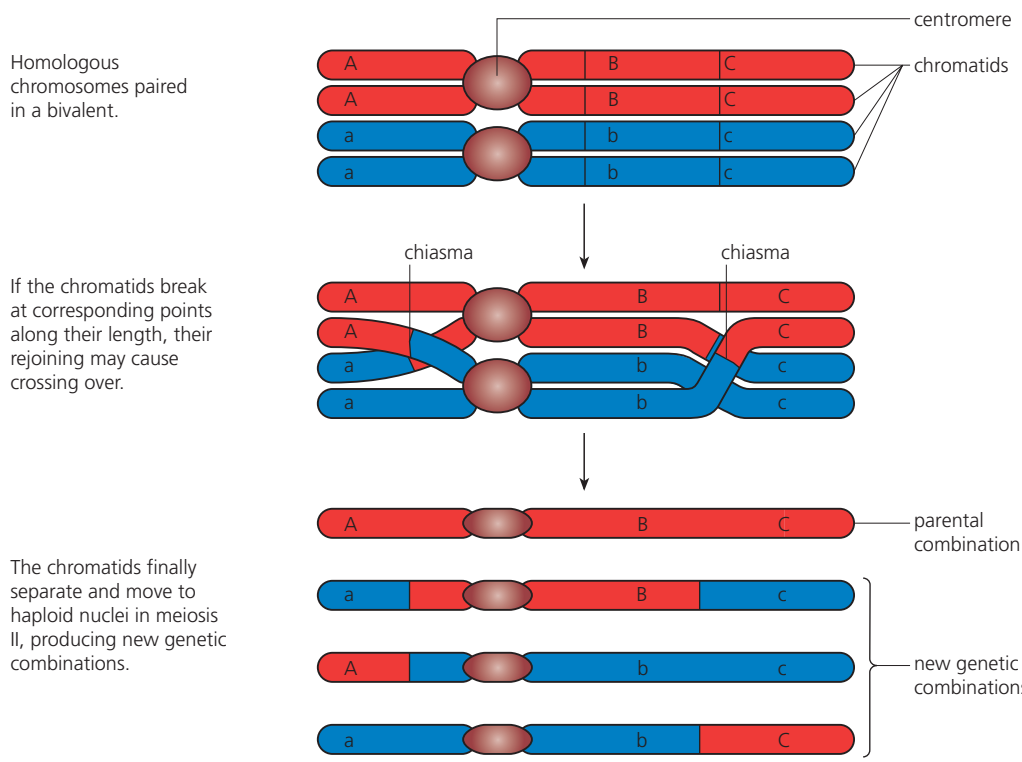
Then, when members of the bivalents start to repel each other and separate, the bivalents are (initially) held together by one or more chiasmata. This temporarily gives an unusual shape to the bivalent. So, crossing over is an important mechanical event (as well as a genetic event).

Next, the spindle forms. Members of the bivalents become attached by their centromeres to the fibres of the spindle at the equatorial plate of the cell. Spindle fibres pull the homologous chromosomes apart, to opposite poles, but the **individual chromatids remain attached** by their centromeres.

Meiosis I ends with two cells each containing a single set of chromosomes each made of two chromatids. These cells do not go into interphase, but rather continue smoothly into meiosis II. This takes place at right angles to meiosis I, but is exactly like mitosis. Centromeres of the chromosomes divide and **individual chromatids now move to opposite poles**. Following division of the cytoplasm, there are four cells – each with half the chromosome number of the original parent cell (haploid).

Figure 4.6 Genetic variation due to crossing over between non-sister chromatids

The effects of genetic variation are shown in one pair of homologous chromosomes. Typically, two, three or more chiasmata form between the chromatids of each bivalent at prophase I.



Meiosis and genetic variation

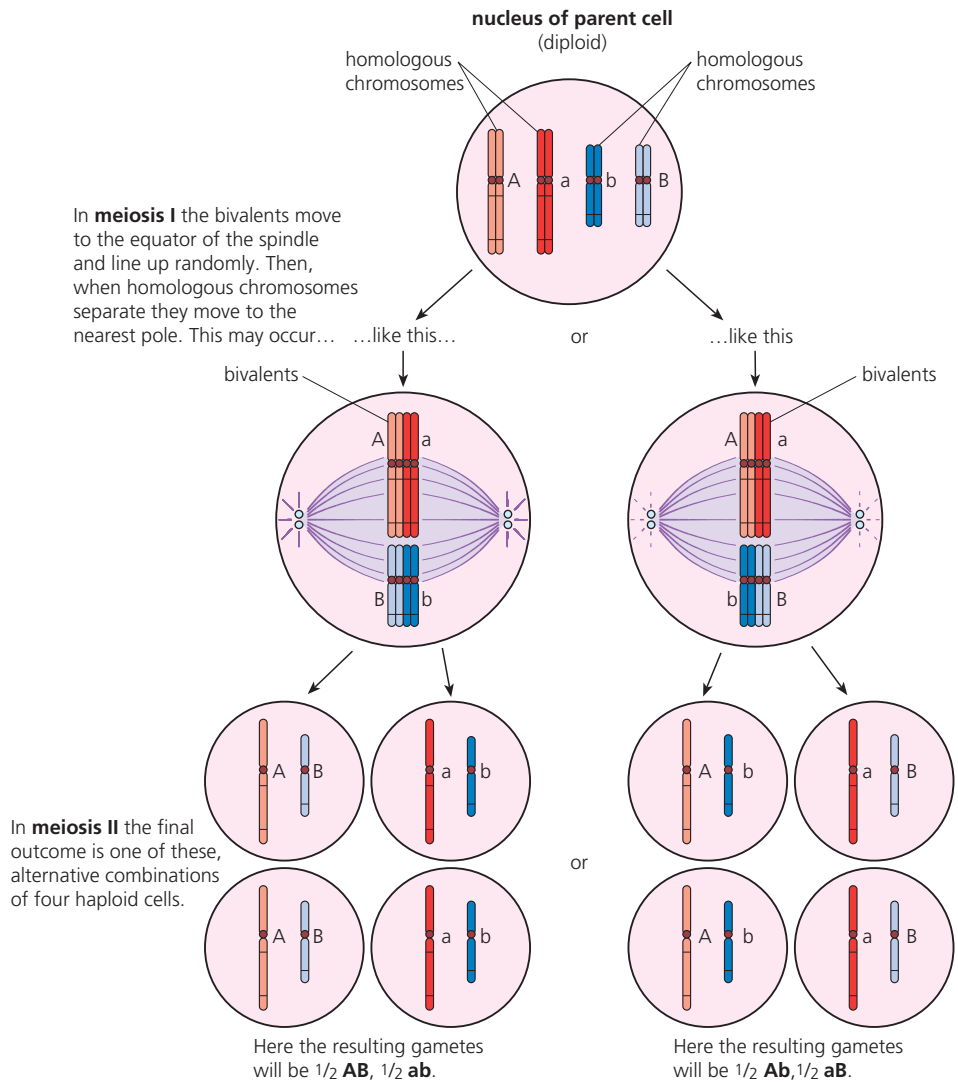
The four haploid cells produced by meiosis differ genetically from each other for two reasons.

- There is **independent assortment** of maternal and paternal homologous chromosomes. This happens because the way the bivalents line up at the equator of the spindle in meiosis I is entirely random. Which chromosome of a given pair goes to which pole is unaffected by (independent of) the behaviour of the chromosomes in other pairs.
- There is **crossing over** of segments of individual maternal and paternal homologous chromosomes. These events result in new combinations of genes on the chromosomes of the haploid cells produced (Figure 4.6).

Independent assortment is illustrated in Figure 4.7, in a parent cell with a diploid number of 4 chromosomes. In human cells, the number of pairs of chromosomes is 23; the number of possible combinations of chromosomes that can be formed by random orientation during meiosis is 2^{23}, which is over 8 million. We see that independent assortment alone, generates a huge amount of variation in the coded information carried by different gametes into the fertilisation stage.

Figure 4.7 Genetic variation due to independent assortment

Independent assortment is illustrated in a parent cell with two pairs of homologous chromosomes (four bivalents). The more bivalents there are, the more variation is possible. In humans, for example, there are 23 pairs of chromosomes giving over 8 million combinations.



Abrupt change in hereditary information – mutations

An abrupt change in the structure, arrangement or amount of DNA of chromosomes may result in a change in the characteristics of an organism or an individual cell in which they occur. This is because the change, called a **mutation**, results in the alteration or non-production of a cell protein (and of the **mRNA** which codes for it, page 69). Mutations can occur spontaneously as a result of errors in normal cell processes such as DNA replication, although this is rare. Alternatively, they can be induced by certain chemicals or types of radiations that cells are sometimes exposed to.

Mutations occurring in body cells of multicellular organisms are called **somatic mutations**. They are only passed on to the immediate descendents of that cell. However, we have already noted that somatic mutations can also be the cause of a cancer (page 35). Somatic mutations do occur with increasing frequency as an organism ages; cancers are more common in elderly people than in the young.

On the other hand, mutations occurring in the cells of the gonads (called **germ line mutations**) that give rise to gametes with an altered genome can be inherited by the offspring and so be the cause of genetic changes in future generations.

A mutation is a change in the amount or chemical structure of DNA which may result in a change in the characteristics of an organism.

Chromosome mutations

In chromosome mutations, there is a change in the number or the sequence of genes, brought about in a number of different ways. Additional sets of chromosomes may occur; for example, the cultivated potato has double the number of chromosomes of the smaller, wild potato (polyploidy, page 467).

Alternatively, there may be an alteration to part of the chromosome set. Sometimes, chromosomes that should separate and move to opposite poles during the nuclear division of gamete formation do not separate but instead move to the same pole.

Non-disjunction is the term for the failure of a pair of chromatids to separate and go to opposite poles during division of the nucleus. In meiosis, this results in gametes with more than and less than the haploid number of chromosomes.

For example, people with Down's syndrome have an extra chromosome 21, giving them a total of 47 chromosomes. How this case of non-disjunction arises is illustrated in Figure 4.8. The symptoms of Down's syndrome are variable, but when severe they include congenital heart defects, defects in the eyes, subsequent learning difficulties and immune system deficiencies.

How are chromosome mutations detected?

A chromosome mutation is exposed by karyotyping the chromosome set of a fetus, after fetal cells have been obtained by amniocentesis or chorionic villus sampling (Figure 4.9).

Gene mutations

Gene mutations are changes in the sequence of bases in the DNA of a gene. They may occur spontaneously as a result of errors in DNA replication although this is extremely rare since DNA polymerase has a built-in checking mechanism as it operates (page 67).

Alternatively, gene mutations can be caused by environmental agents we call **mutagens**. These can include ionising radiation in the form of X-rays, cosmic rays, and radiation from radioactive isotopes (α, β and γ rays). Any of these can cause the break-up of the DNA molecule. Non-ionising mutagens include UV light and various chemicals, including carcinogens in tobacco smoke (tar compounds). These act by modifying the chemistry of the base pairs of DNA.

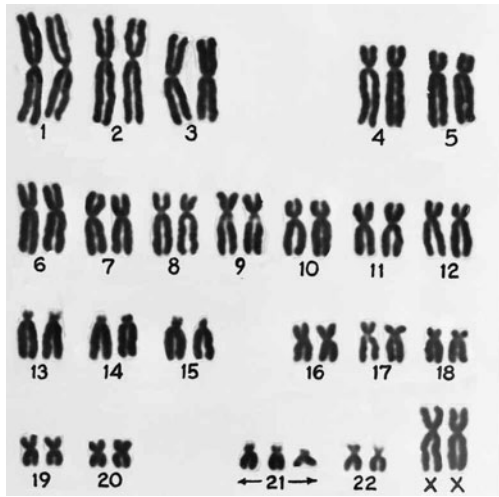
Base substitution mutation and sickle cell condition

The smallest form of gene mutation change occurs when one base is replaced by another – this is called base substitution. An example of this type of gene mutation is the cause of the human condition known as sickle cell anaemia.

Figure 4.8 Down's syndrome, an example of a chromosome mutation

An extra chromosome causes Down's syndrome. The extra one comes from a meiosis error. The two chromatids of chromosome 21 fail to separate, and both go into the daughter cell that forms the secondary oocyte.

karyotype of a person with Down's syndrome



an extra chromosome 21

Steps of non-disjunction in meiosis
(illustrated in nucleus with only two pairs of homologous chromosomes – for clarity)

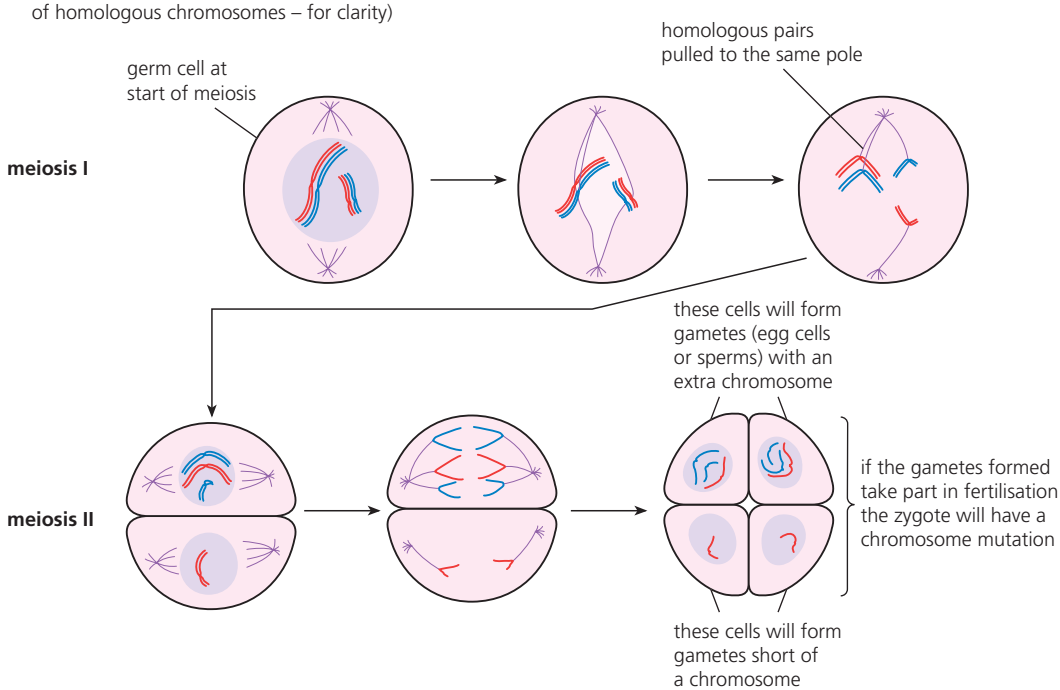
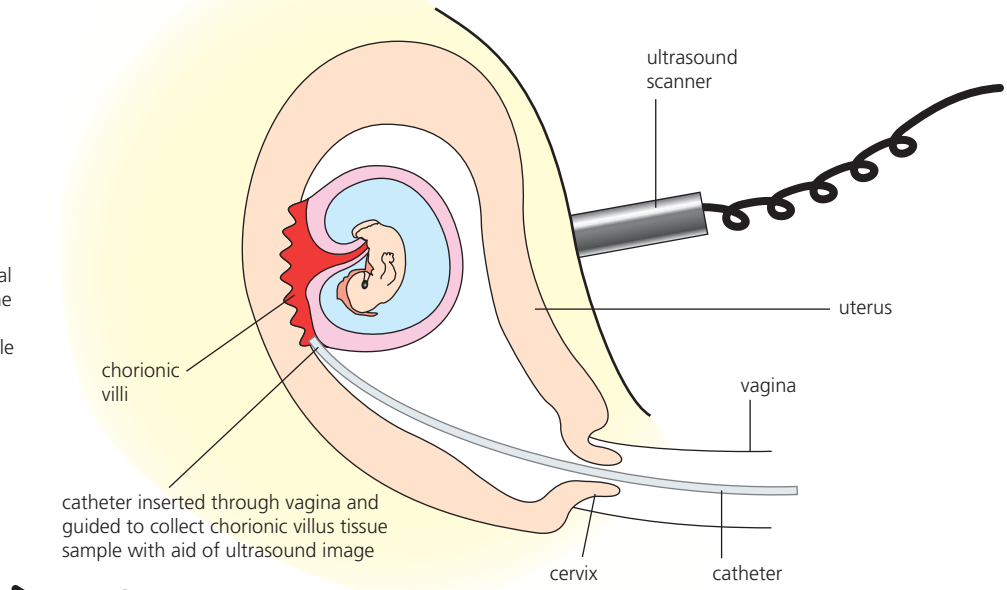


Figure 4.9 Screening of a fetus in the uterus

chorionic villus sampling – withdrawal of a sample of the fetal tissue part-buried in the wall of the uterus in the period 8–10 weeks into the pregnancy; the tiny sample is of cells that are actively dividing and can be analysed quickly.



amniocentesis – withdrawal of a sample of amniotic fluid in the period 16–30 weeks of gestation; the fluid contains cells from the surface of the embryo

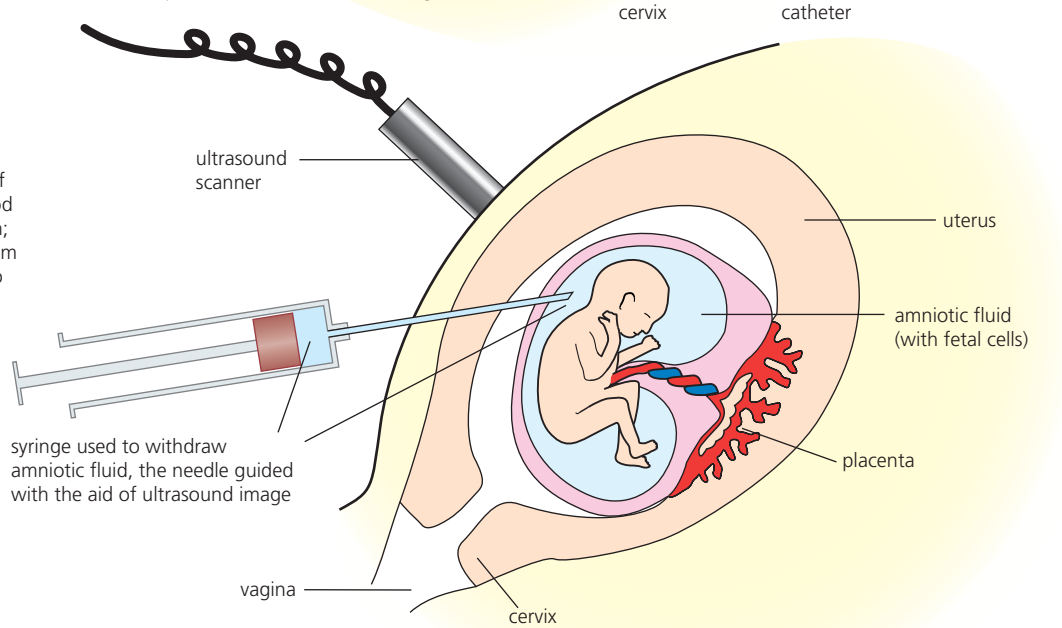
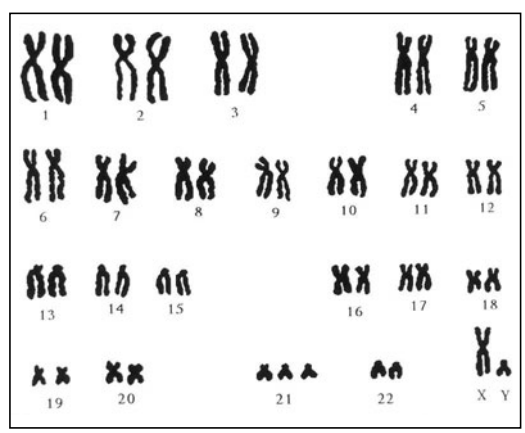


Figure 4.10 Human karyotype used in a genetic screening for counselling



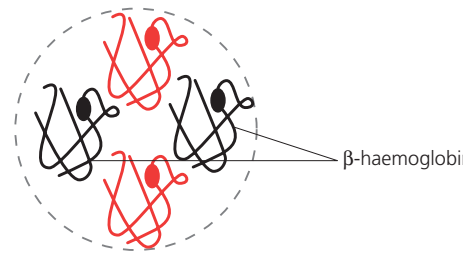
3 **Analyse** the human karyotype in Figure 4.10 to determine the gender of the patient and whether non-disjunction has occurred.

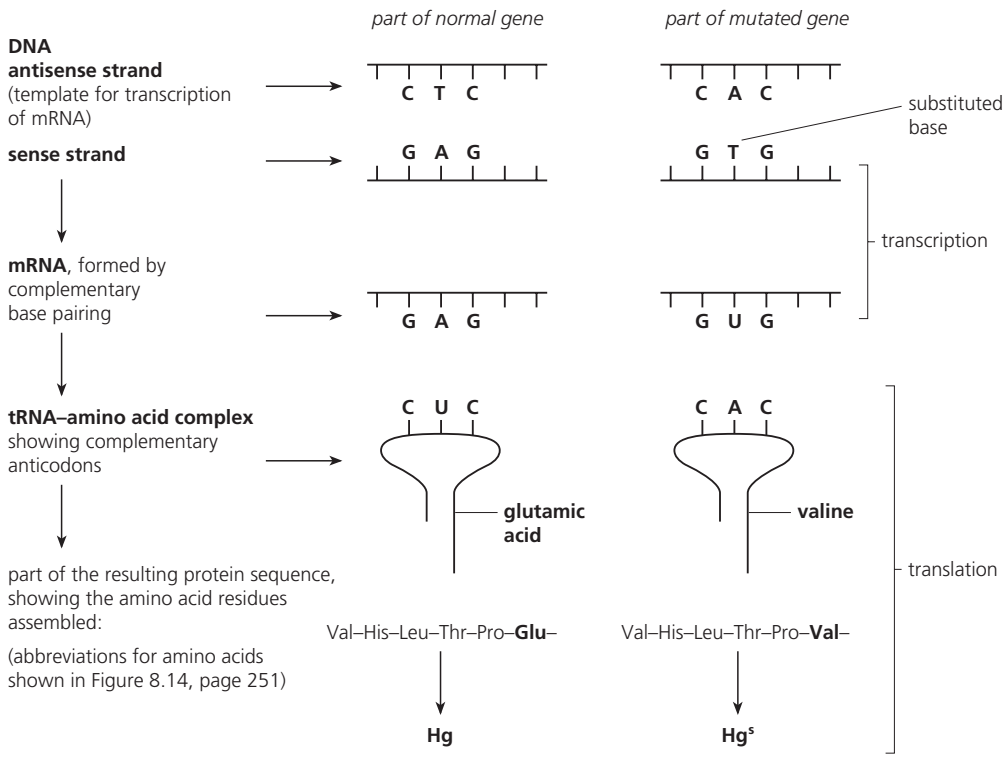
Figure 4.11 Sickle cell anaemia; an example of a gene mutation

Anaemia is a disease typically due to a deficiency in healthy red cells in the blood.

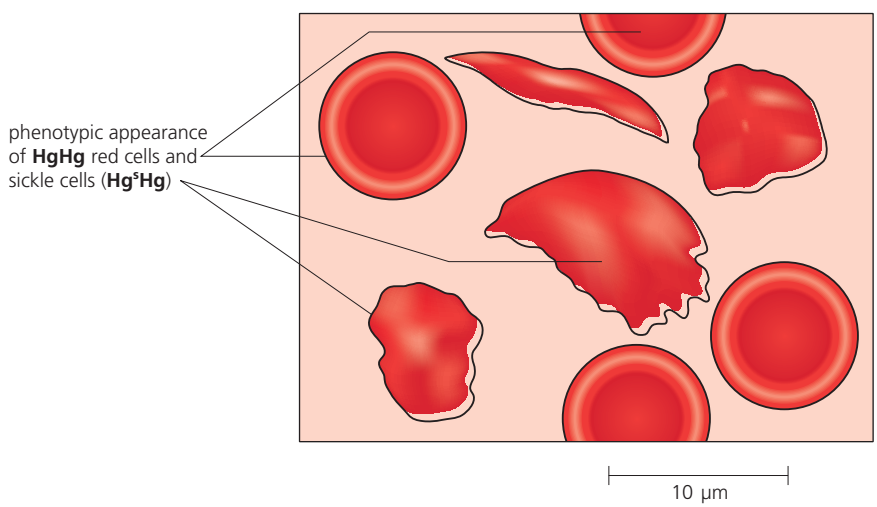
Haemoglobin occurs in red cells – each contains about 280 million molecules of haemoglobin. A molecule consists of two α-haemoglobin and two β-haemoglobin subunits, interlocked to form a compact molecule.

The **mutation** that produces sickle cell haemoglobin (**Hg^S**) is in the gene for β-haemoglobin. It results from the substitution of a single base in the sequence of bases that make up all the codons for β-haemoglobin.





drawing based on a photomicrograph of a blood smear, showing blood of a patient with sickle cells present among healthy red cells



Haemoglobin, the oxygen-transporting pigment of red cells, is made of four protein molecules that interlock to form a compact molecule. Two of the haemoglobin molecules are known as α haemoglobin and two are β haemoglobin. The gene that codes for the amino acid sequence of β haemoglobin occurs on chromosome 11, and is prone to a substitution of the base A to T in a codon for the amino acid glutamic acid. As a consequence, the amino acid valine appears at that point, instead, as illustrated in Figure 4.11. Note the consequence of this substitution at the transcription and translation stages.

The effects on the properties of the haemoglobin molecule of this simple change of one amino acid residue for another within the whole β haemoglobin molecule are dramatic. The molecules with this unusual haemoglobin, known as **haemoglobin S (Hb^s)**, tend to clump together and form long fibres that distort the red cells into sickle shapes. In this condition, they transport little oxygen and the sickle cells may even block smaller vessels.

People who are heterozygous for haemoglobin S (**Hb Hb^s**) have less than 50% haemoglobin S. Such a person is said to have **sickle cell trait**, and they are only mildly anaemic. However, people who are homozygous for haemoglobin S (**Hb^s Hb^s**) have a serious problem, and are described as having **sickle cell anaemia**. Heart and kidney failure problems are common in people affected with sickle cell anaemia.

An advantage in having sickle cell trait?

Malaria is the most important of all insect-borne diseases. It is in Africa south of the Sahara that about 80% of the world's malaria cases are found, and here that most fatalities due to the disease occur. It is estimated that some 400 million people are infected, of whom 1.5 million (mostly children under 5 years) die each year.

Malaria is caused by *Plasmodium*, a protozoan, which is transmitted from an infected person to another person by blood-sucking mosquitoes of the genus *Anopheles*. Only the female mosquito is the vector (the male mosquito feeds on plant juices).

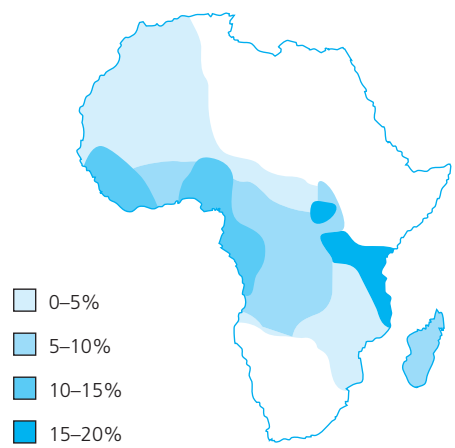
Plasmodium completes its life cycle in red cells, but it cannot do so in red cells containing haemoglobin S. People with sickle cell trait are protected to a significant extent. Where malaria is endemic in Africa, possession of one mutant gene (having sickle cell trait, not full anaemia) is advantageous. This means that 'survival of the fittest' ensures this allele is selected for. Fewer of the alleles for normal haemoglobin are carried into the next generation (Figure 4.12).

Figure 4.12 How sickle cell trait may confer an advantage

distribution of haemoglobin S is virtually the same as that of malaria

distribution of malaria caused by *Plasmodium falciparum* or *P. vivax* (the forms of malaria that are most frequently fatal, especially in childhood)

distribution of sickle cell gene in the population



Sickle cell anaemia is a genetically transmitted disease of the blood caused by an abnormal form of haemoglobin. Red cells with 'sickle' haemoglobin do not carry oxygen and can cause blockage of arterioles. Normally, sickle cell anaemia confers a disadvantage, but the malarial parasite (page 595) is unable to complete its life cycle in red cells with haemoglobin S, so the gene for sickle cell is selected for in regions of the world where malaria occurs.

4 **Explain** how knowledge of the cause of sickle cell trait and sickle cell anaemia supports the concept of a gene as a linear sequence of bases.

Chromosomes and the mechanism of inheritance

4.3.1–4.3.12

The mechanism of inheritance was successfully investigated before chromosomes had been observed or genes were known about. It was **Gregor Mendel** who made the first discovery of the fundamental laws of heredity (Figure 4.13).

Figure 4.13 Gregor Mendel – founder of modern genetics

Gregor Mendel was born in Moravia in 1822, the son of a peasant farmer. As a young boy, he worked to support himself through schooling, but at the age of 21 he was offered a place in the monastery at Bruno (now in the Czech Republic). The monastery was a centre of research in natural sciences and agriculture, as well as in the humanities. Mendel was successful there. Later, he became Abbot.

Mendel discovered the principles of heredity by studying **the inheritance of seven contrasting characteristics of the garden pea plant**. These did not 'blend' on crossing, but retained their identities, and were inherited in fixed mathematical ratios.

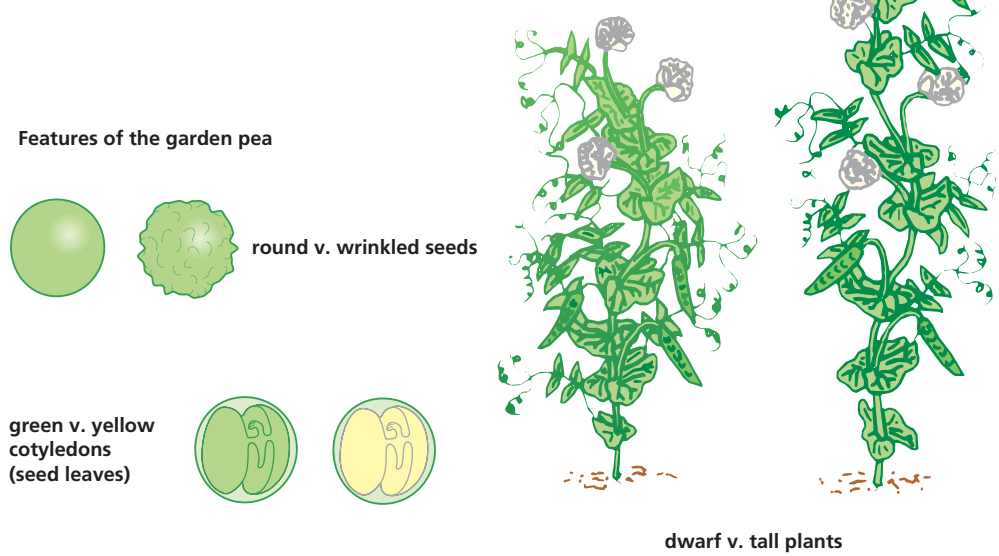
He concluded that hereditary factors (we now call them genes) determine these characteristics, that these factors occur in duplicate in parents, and that the two copies of the factors segregate from each other in the formation of gametes.

Today, we often refer to Mendel's laws of heredity, but Mendel's results were **not presented as laws** – which may help to explain the difficulty others had in seeing the significance of his work at the time.

Mendel was successful because:

- his experiments were carefully planned, and used large samples
- he carefully recorded the numbers of plants of each type but expressed his results as ratios
- in the pea, contrasting characteristics are easily recognised
- by chance, each of these characteristics was controlled by a single factor (gene)* rather than by many genes, as most human characteristics are
- pairs of contrasting characters that he worked on were controlled by factors (genes) on separate chromosomes*
- in interpreting results, Mendel made use of the mathematics he had learnt.

* Genes and chromosomes were not known then.



TOK Link

At the time of Mendel, it was thought that the characteristics of parents 'blended' in their offspring. What features of Mendel's methods enabled him to avoid this error?

Mendel's investigation of the inheritance of a single contrasting characteristic is known as a **monohybrid cross**. Mendel had noticed that the garden pea plant was either tall or dwarf. How was this contrasting characteristic controlled?

Figure 4.14 summarises the steps of his investigation of the inheritance of height in the pea plants. Note that his experiment began with plants that always 'bred true'; that is, the tall plants produced progeny that were all tall and the dwarf plants produced progeny that were all dwarf, when each was allowed to self-fertilise. Self-fertilisation is the normal condition in the garden pea plant.

In Mendel's interpretation of the monohybrid cross he argued that because the dwarf characteristic had apparently disappeared in the F_1 generation and reappeared in the F_2 generation, there must be a factor controlling dwarfness that remained intact from one generation to another. However, this factor did not express itself in the presence of a similar factor for tallness. In other words, as characteristics, tallness is dominant and dwarfness is recessive. Logically there must be two independent factors for height, one from one parent and the other factor from the other parent in the cells of an organism. A sex cell (gamete) must contain only one of these factors.

Mendel saw that a 3:1 ratio could be the product of randomly combining two pairs of unlike factors (**T** and **t**, for example). This can be shown using a grid, now known as a Punnett grid, after the mathematician who first used it (Figure 4.15).

Mendel's conclusions from the monohybrid cross were that:

- within an organism there are breeding factors controlling characteristics such as 'tall' and 'dwarf';
- there are two factors in each cell;
- one factor comes from each parent;
- the factors separate in reproduction, and either can be passed on to an offspring;
- the factor for 'tall' is an alternative form of the factor for 'dwarf';
- the factor for 'tall' is dominant over the factor for 'dwarf'.

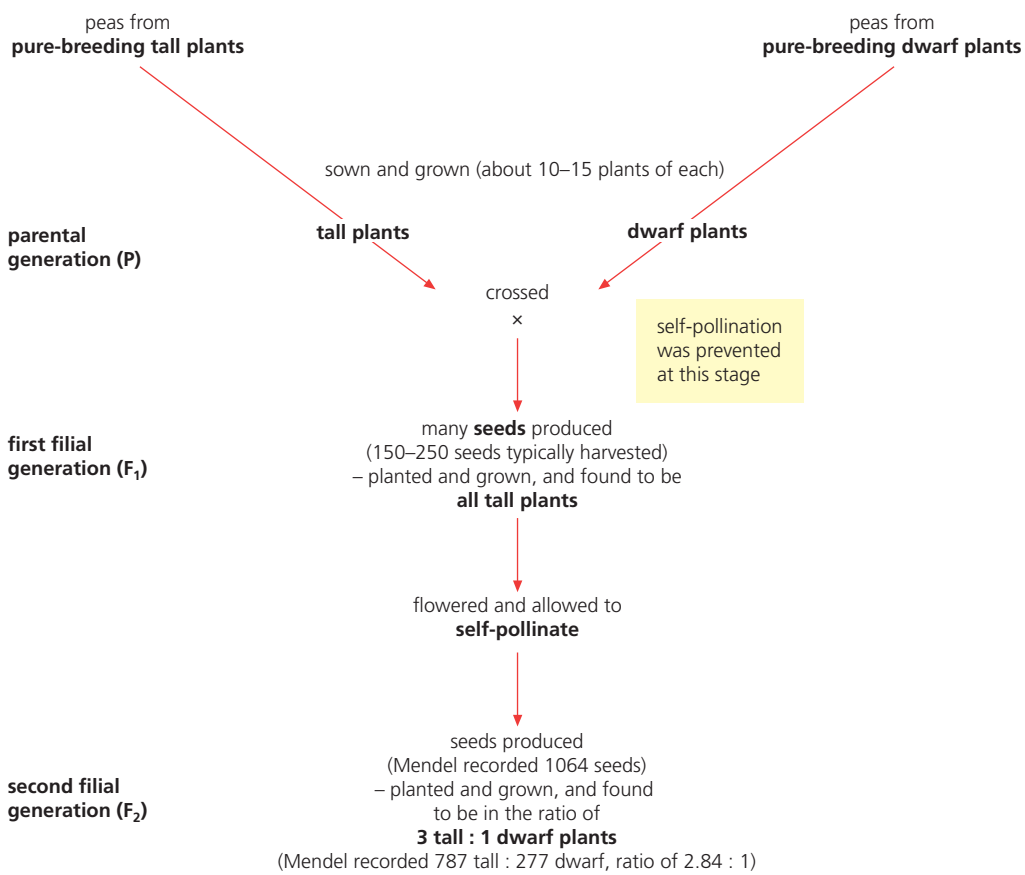


Figure 4.14 The steps of Mendel's monohybrid cross

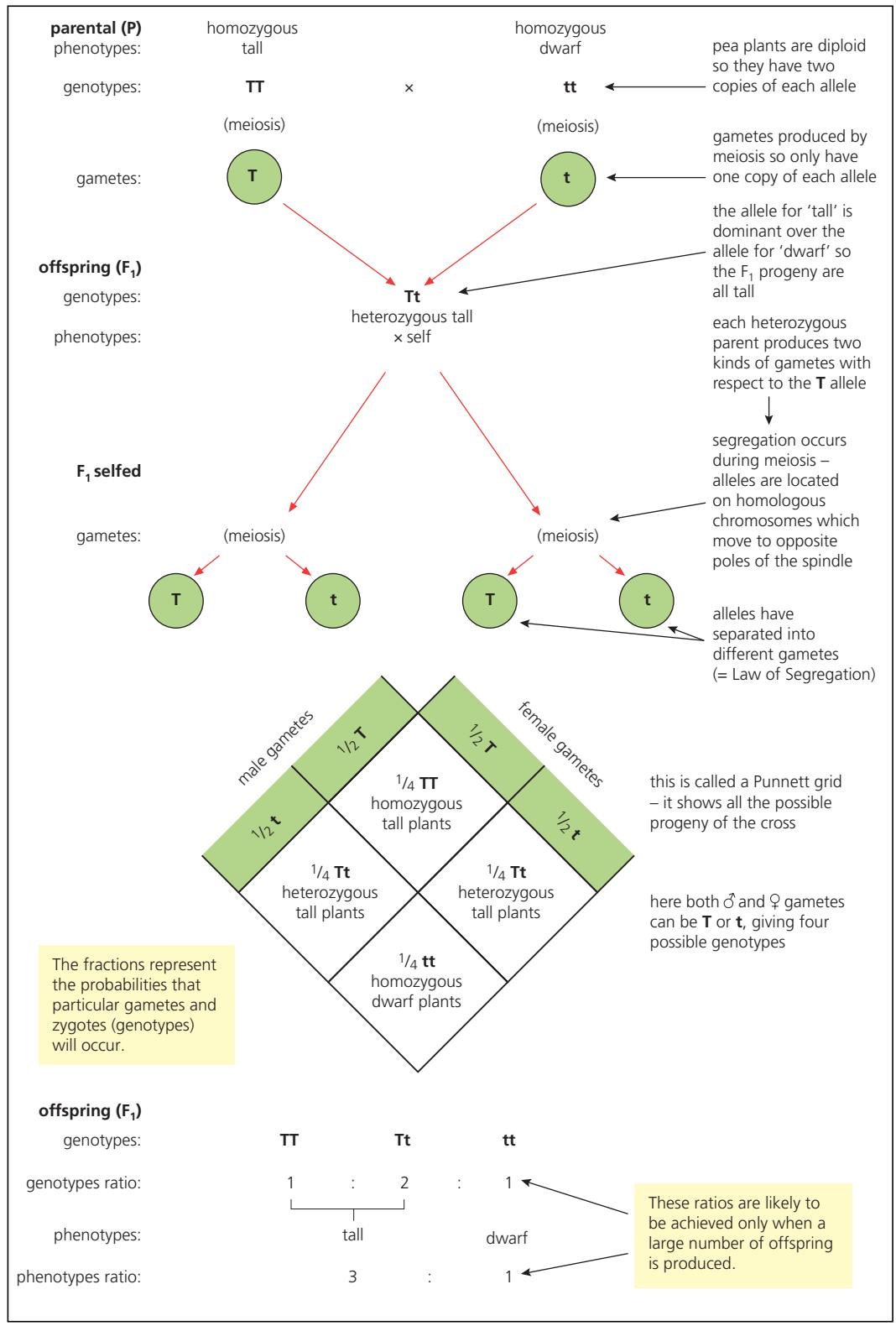


Figure 4.15 Genetic diagram showing the behaviour of alleles in Mendel's monohybrid cross.

5 From the Punnett grid (Figure 4.15), a ratio of 3 tall to 1 dwarf pea plants was predicted. In fact, a ratio of 2.84:1 occurred (Figure 4.14). **Suggest** what chance events may influence the actual ratios of offspring obtained in breeding experiments like these.

Mendel never stated his discoveries as laws, but it might have helped others understand his work had he done so. For example, he might have said, 'Each characteristic of an organism is determined by a pair of factors of which only one can be present in each gamete'. Today we call a similar statement Mendel's First Law, the **Law of Segregation: 'The characteristics of an organism are controlled by pairs of alleles which separate in equal numbers into different gametes as a result of meiosis.'**

Genotype, phenotype and the test cross

The alleles that an organism carries (present in every cell) make up the genotype of that organism. A genotype in which the two alleles of a gene are the same is said to 'breed true' or, more scientifically, to be **homozygous** for that gene. In Figure 4.15, the parent pea plants (P generation) were either homozygous tall or homozygous dwarf.

If the alleles are different, the organism is **heterozygous** for that gene. In Figure 4.15, the progeny (F_1 generation) were heterozygous tall.

So the **genotype** is the genetic constitution of an organism. Alleles interact in various ways and with environmental factors. The outcome is the phenotype. The **phenotype** is the way in which the genotype of the organism is expressed – including the appearance of the organism. Here the heights of the plants were their phenotypes.

If an organism shows a recessive characteristic in its phenotype (like the dwarf pea) it must have a homozygous genotype (**tt**). But if it shows the dominant characteristic (like the tall pea) then it may be either homozygous for a dominant allele (**TT**) or heterozygous for the dominant allele (**Tt**). **In other words, TT and Tt look alike**. They have the same phenotype but different genotypes.

You could tell plants like these apart only by the offspring they produced in a particular cross. When the tall heterozygous plants (**Tt**) are crossed with the homozygous recessive plants (**tt**), the cross yields 50% tall and 50% dwarf plants (Figure 4.16). This type of cross has become known as a **test cross**. If the offspring produced had all been tall then we would have known that the tall plants under test were homozygous plants (**TT**). Of course, sufficient plants have to be used to obtain these distinctive ratios.

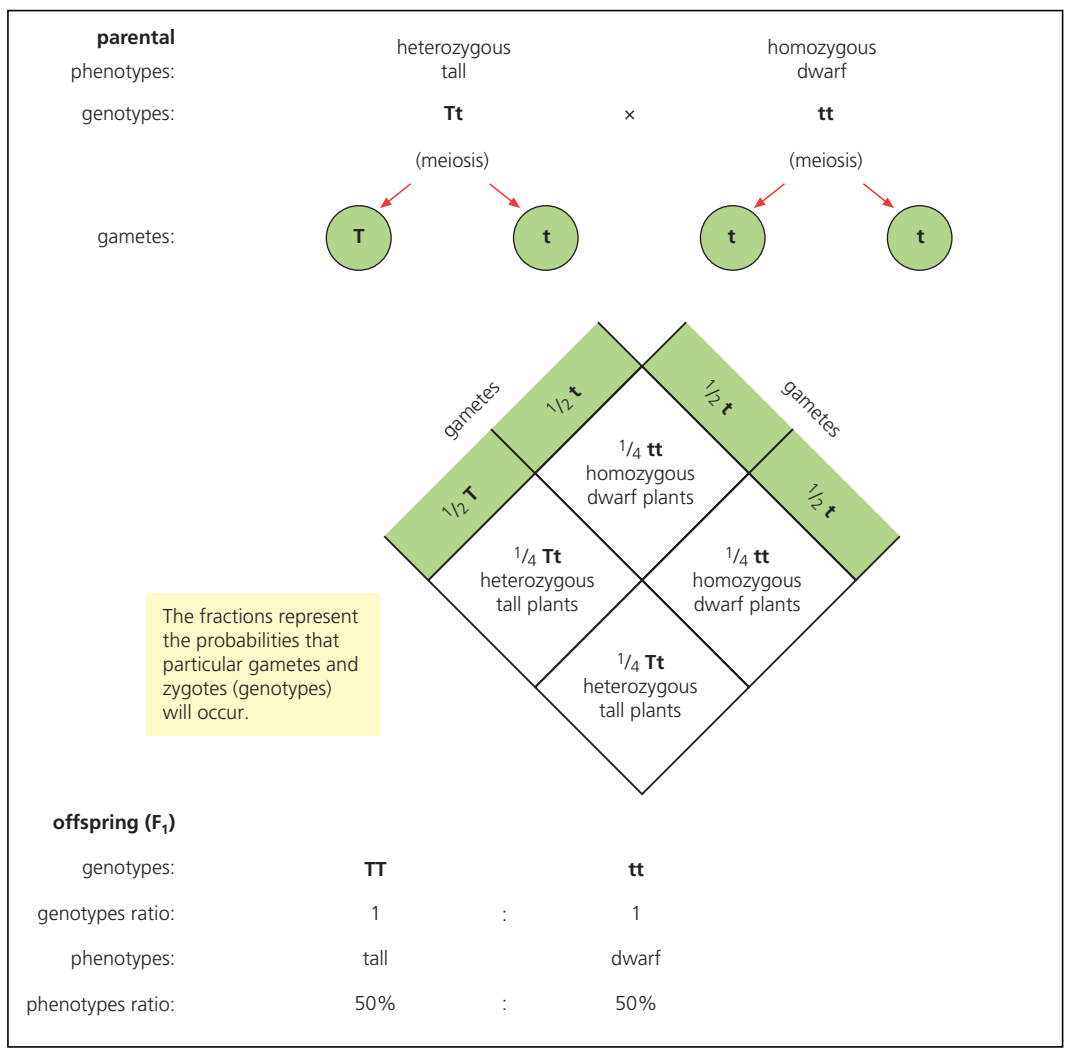


Figure 4.16 Genetic diagram of Mendel's test cross

6 **Construct** a table to explain the relationship between Mendel's Law of Segregation and meiosis.

Modification of the 3:1 monohybrid ratio

In certain types of monohybrid cross the 3:1 ratio is not obtained. Two of these situations are illustrated next.

Codominance – when both alleles are expressed

In the case of some genes, both alleles may be expressed simultaneously, rather than one being dominant and the other recessive in the phenotype, as has been illustrated up to now. For example, in the common garden flower *Antirrhinum*, when red-flowered plants are crossed with white-flowered plants, the F_1 plants have pink flowers. When pink-flowered *Antirrhinum* plants are crossed, the F_2 offspring are found to be red, pink and white in the ratio 1:2:1 respectively.

Pink coloration of the petals occurs because both alleles have been expressed in the heterozygote and two pigment systems are present, rather than that of a dominant allele only. Red and white are **codominant alleles**. In genetic diagrams, codominant alleles are represented by a capital letter for the gene, and different superscript capital letters for the two alleles, in recognition of their equal influence (Figure 4.17).

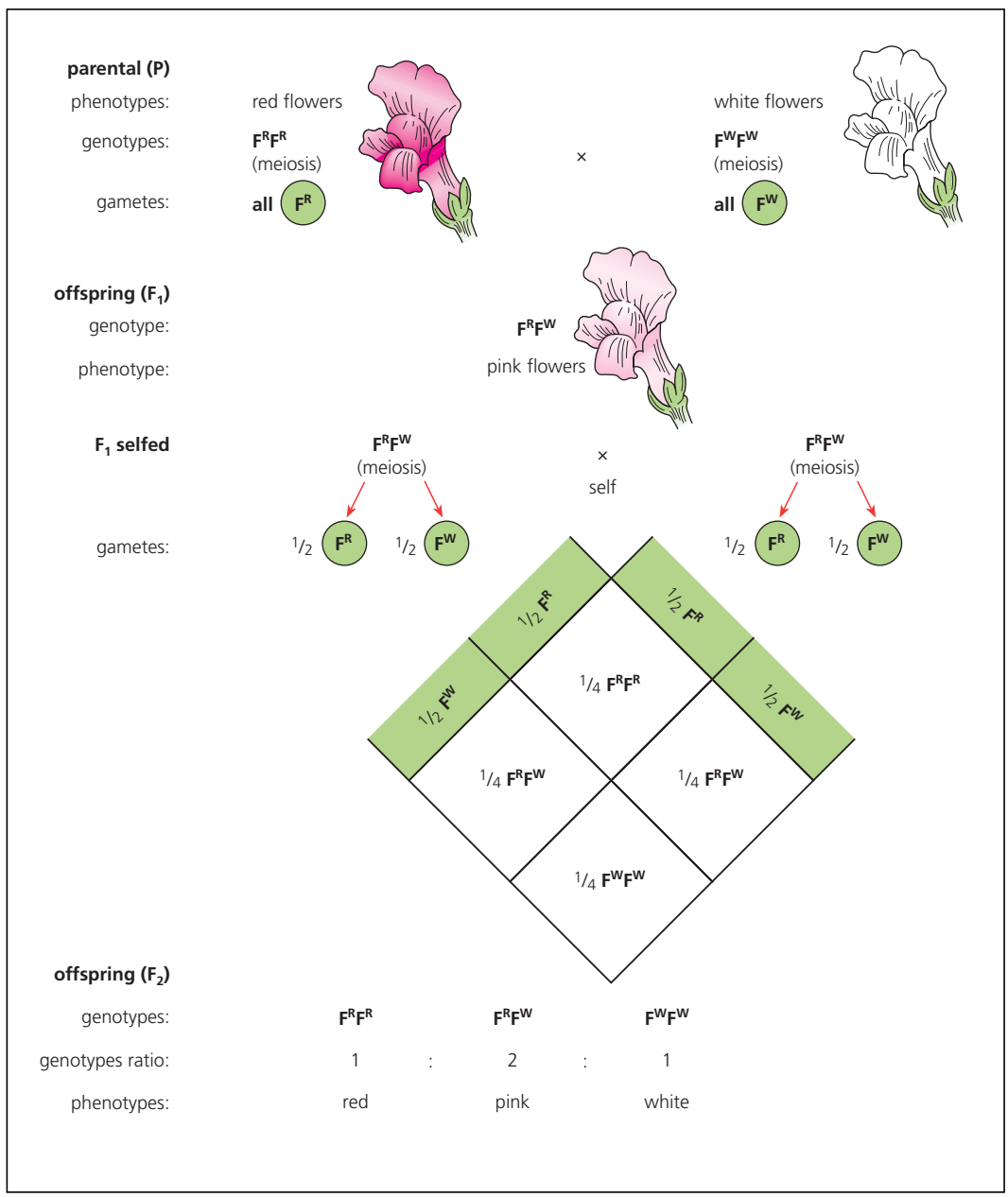


Figure 4.17 Codominance in the garden flower, *Antirrhinum*

7 **Construct** for yourself (pencil and paper) a monohybrid cross between cattle of a variety that has a gene for coat colour with codominant alleles, for red and white coats. Homozygous parents produce roan offspring (red and white hairs together). **Predict** what offspring you would expect and in what proportions, when a sibling cross (equivalent to selfing in plants) occurs between roan offspring.

More than two alleles exist for a particular locus

The genes introduced so far exist in two forms (two alleles), like the height gene of the garden pea, existing as tall or dwarf alleles. That means that in genetic diagrams we can represent alleles by a single letter (here, **T** or **t**) according to whether they are dominant or recessive.

For simplicity we began by considering inheritance of a gene for which there are just two alleles. However, we now know that not all genes are like this. In fact, most genes have more than two alleles, and these are cases of **multiple alleles**.

With multiple alleles, we choose a single capital letter to represent the locus at which the alleles may occur, and the individual alleles are then represented by an additional single letter (usually capital) in a superscript position (Table 4.1) – as with codominant alleles. An excellent example of multiple alleles is found in those controlling the ABO blood group system of humans. Human blood belongs to one of four groups: A, B, AB or O. Table 4.1 lists the possible phenotypes and the genotypes that may be responsible for each.

Table 4.1 The ABO blood groups – phenotypes and genotypes

Phenotype	Genotypes
A	I^AI^A or I^Ai
B	I^BI^B or I^Bi
AB	I^AI^B
O	ii

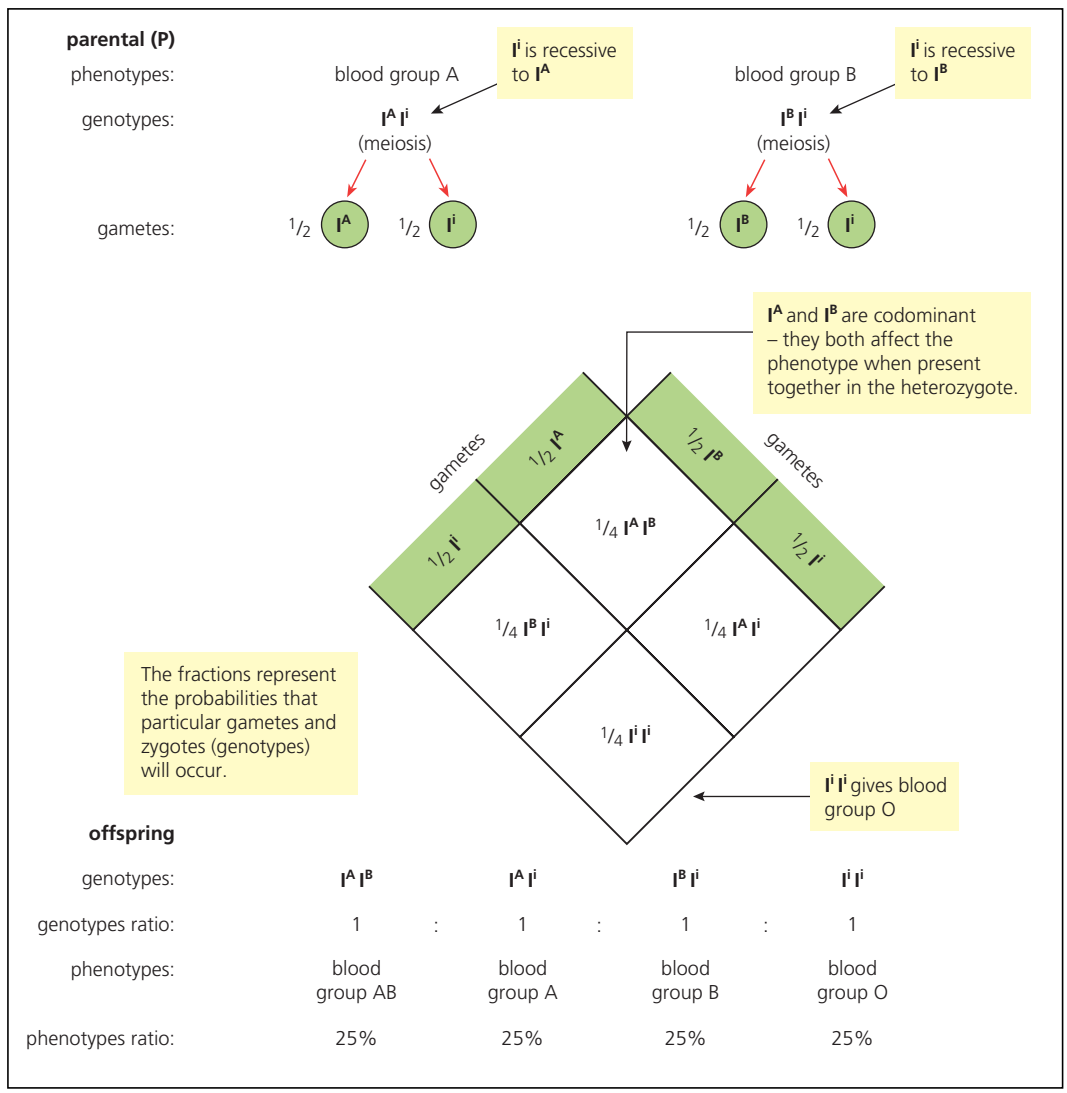


Figure 4.18 Inheritance of blood groupings A, B, AB, and O

So, the ABO blood group system is determined by combinations of alternative alleles. In each individual, only two of the three alleles exist, but they are inherited as if they were alternative alleles of a pair. However, I^A and I^B are codominant alleles and both I^A and I^B are dominant to the recessive I^i. In Figure 4.18 the way the alternative blood groups may be inherited is shown.

8 One busy night in an understaffed maternity unit, four children were born about the same time. Then the babies were muddled up by mistake; it was not certain which child belonged to which family. Fortunately the children had different blood groups: A, B, AB and O.

The parents' blood groups were also known:

Mr and Mrs Jones	A × B	Mr and Mrs Lee	B × O
Mr and Mrs Gerber	O × O	Mr and Mrs Santiago	AB × O

The nurses were able to decide which child belonged to which family. **Deduce** how this was done.

Reviewing the genetic terms learned so far

Table 4.2 lists these essential genetic terms and defines each, succinctly.

Cover the right-hand column with a piece of paper, temporarily, and see if you can define each correctly, in your own words.

Gene	the basic unit of inheritance by which inherited characteristics are transferred from parents to offspring, consisting of a length of DNA on a chromosome
Allele	one alternative form of a gene, occupying a specific position (locus) on a chromosome
Loci	the positions along the chromosomes where genes occur, so alleles of the same gene occupy the same locus (singular)
Genotype	the genetic constitution of an organism
Phenotype	the characteristics displayed by the organism – the way in which the genotype is expressed (appearance of an organism)
Dominant allele	an allele that affects the phenotype of the organism, whether present in the heterozygous or homozygous condition
Recessive allele	an allele that affects the phenotype of the organism only when the dominant allele is absent (in homozygous recessive individuals)
Homozygous	a diploid organism that has inherited the same allele (for any particular gene) from both parents
Heterozygous	a diploid organism that has inherited different alleles from each parent
Test cross	testing a suspected heterozygote by crossing it with a known homozygous recessive
Codominant alleles	pairs of alleles that both affect the phenotype when present in a heterozygote
Carrier	an individual with a recessive allele of a gene that does not affect their phenotype – especially one responsible for a genetic disorder such as colour blindness or haemophilia (page 114)

Table 4.2 Essential genetic terms

Human inheritance investigated by pedigree chart

Studying human inheritance by experimental crosses (with selected parents, sibling crosses, and the production of large numbers of progeny) is out of the question. Instead, we may investigate the pattern of inheritance of a particular characteristic by researching a family pedigree, where appropriate records of the ancestors exist. A human pedigree chart uses a set of rules. These are identified in Figure 4.19.

9 In the human pedigree chart in Figure 4.19 (page 110) **state**:

a who are the female grandchildren of Richard and Judith
b who are Alan's (**i**) grandparents; (**ii**) uncles
c how many people in the chart have parents unknown to us
d the names of two offspring who are cousins.

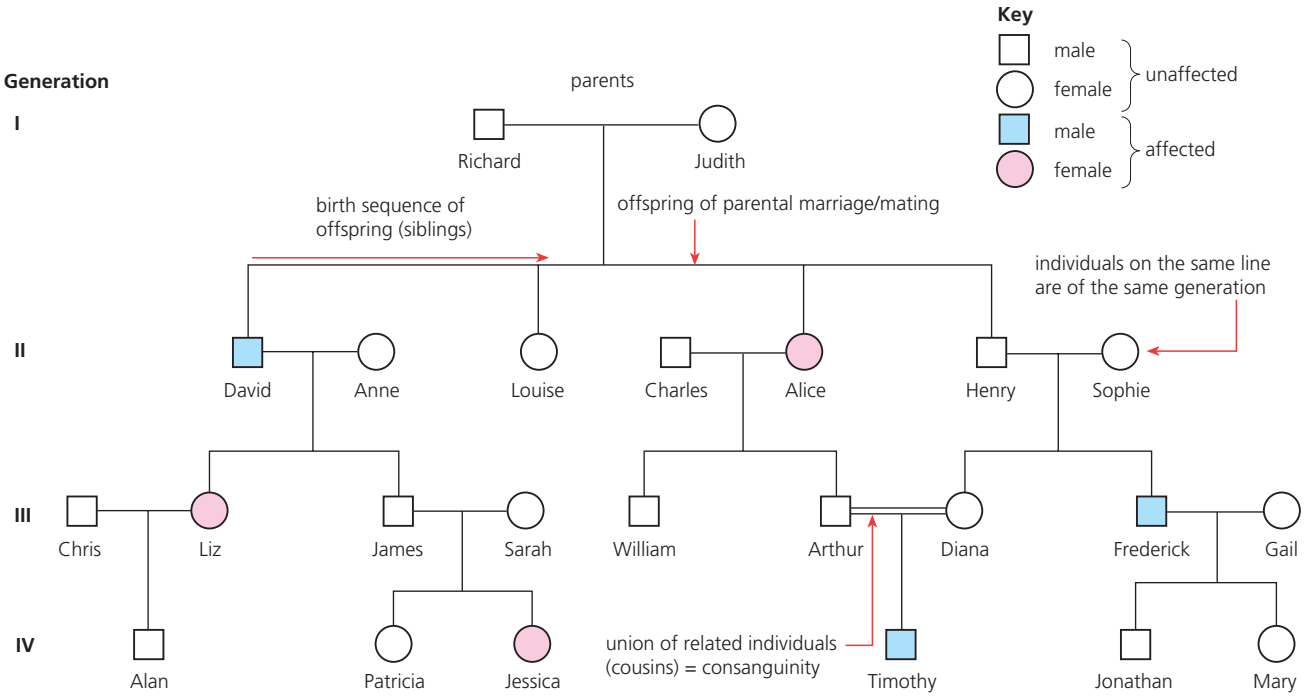


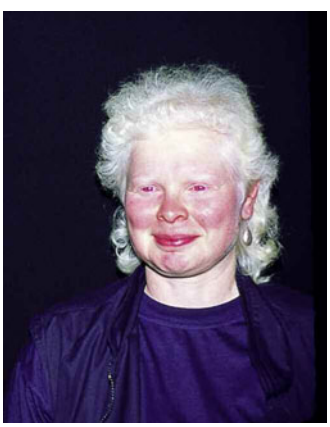
Figure 4.19 An example of a human pedigree chart

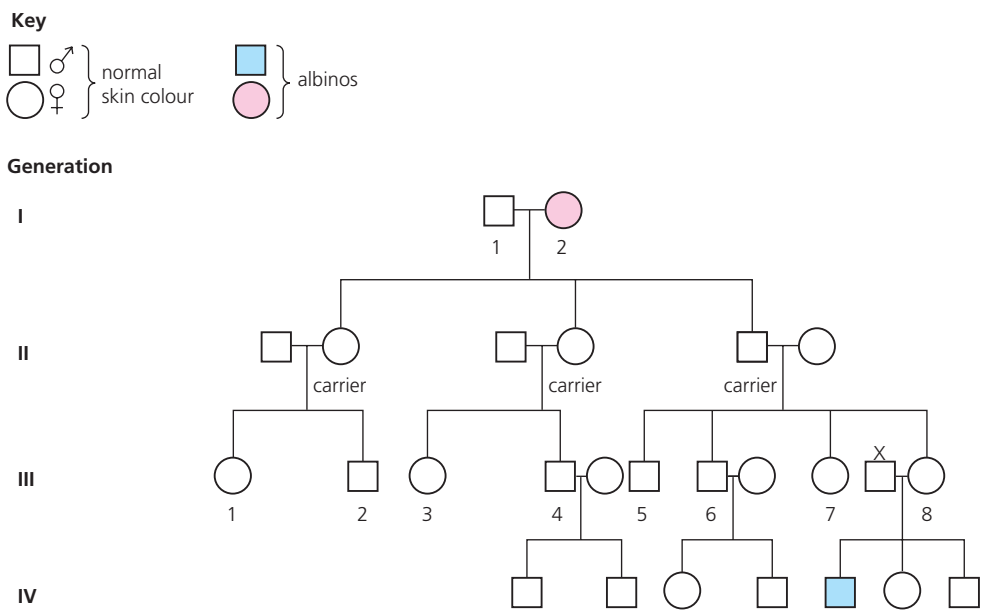
We can use a pedigree chart to detect conditions likely to be due to dominant and recessive alleles. In the case of a characteristic due to a dominant allele, the characteristic tends to occur in one or more members of the family in every generation (Figure 4.19). On the other hand, a recessive characteristic is seen infrequently, often skipping many generations.

For example, albinism is a rare inherited condition of humans (and other mammals) in which the individual has a block in the biochemical pathway by which the pigment melanin is formed. Albinos have white hair, very light-coloured skin and pink eyes. Albinism shows a pattern of recessive monohybrid inheritance in humans (Figure 4.20). In the chart shown of a family with albino members, albinism occurs infrequently, skipping two generations altogether.

Figure 4.20 Pedigree chart of a family with albino members

Albino people must be homozygous for the recessive albino allele (**pp**). People with normal skin pigmentation may be homozygous normal (**PP**) or carriers (**Pp**).



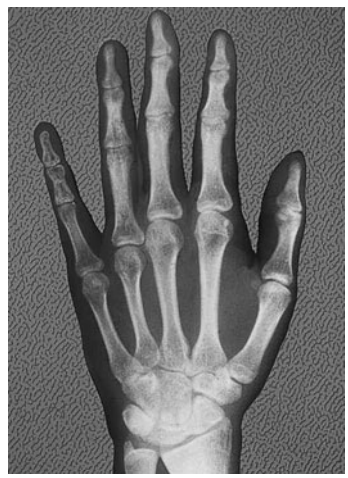


This is a typical family tree for inheritance of a characteristic controlled by a recessive allele.

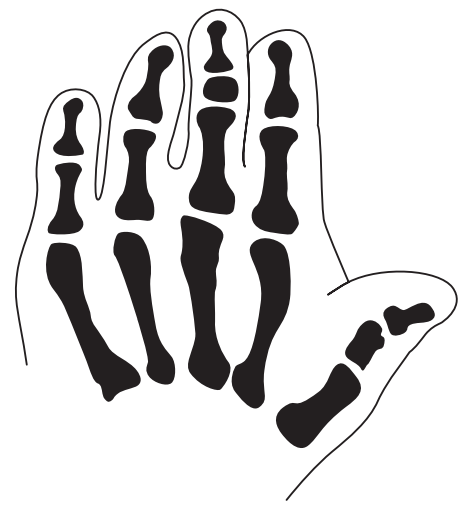
Brachydactyly is a rare condition of humans in which the fingers are very short. Brachydactyly is due to a mutation in the gene for finger length. Unusually, the mutant allele is dominant, so brachydactyly shows a pattern of dominant monohybrid inheritance among members of a family in which it occurs; that is, it tends to occur in every generation (Figure 4.21). Indeed, the frequencies of brachydactylous and normal handed people are similar to those obtained in a test cross (page 106).

Figure 4.21 Brachydactyly, and pedigree chart of a family with brachydactylous genes

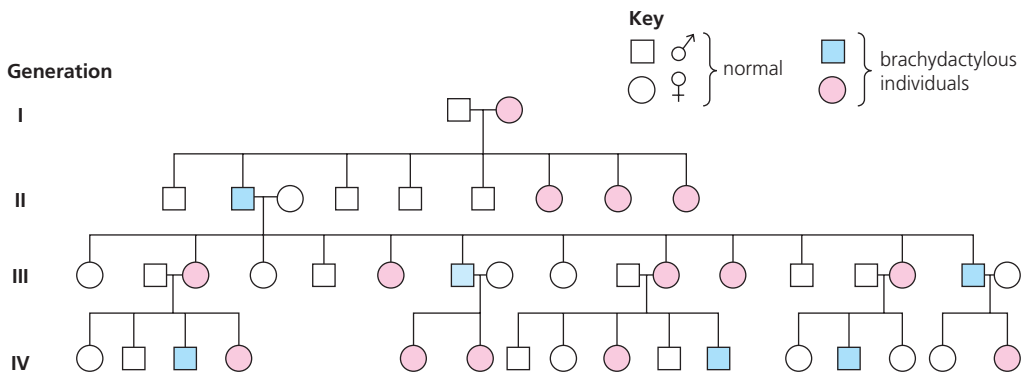
X-ray of bones of hand of normal length



Drawing of brachydactylous hand



Pedigree chart of family with brachydactylous alleles



In **generation I** the parents are assumed to be normal male (**nn**) and brachydactylous female (heterozygous **Nn**), as the ratio of offspring is similar to that of a test cross.

In **each subsequent generation** about half the offspring are brachydactylous (i.e. **Nn** or **NN**) and half are normal (**nn**).

10 If a homozygous normal-handed parent (**nn**) were crossed with a heterozygous brachydactylous parent (**Nn**) **calculate** the probability of an offspring with brachydactylous hands, using a Punnett grid. **Construct** a genetic diagram to show your workings.

Sex chromosomes and gender

In humans, gender is determined by specific chromosomes known as the **sex chromosomes**. Each of us has one pair of sex chromosomes (either XX or XY chromosomes) along with the 22 other pairs (known as **autosomal chromosomes**).

Egg cells produced by meiosis all carry an X chromosome, but 50% of sperms carry an X chromosome and 50% carry a Y chromosome. At fertilisation, an egg cell may fuse with a sperm carrying an X chromosome leading to a female offspring. Alternatively, the egg cell may fuse with a sperm carrying a Y chromosome, leading to a male offspring. So, the gender of offspring in humans (and all mammals) is determined by the male partner. Also, we would expect equal numbers of male and female offspring to be produced by a breeding population, over time (Figure 4.22).

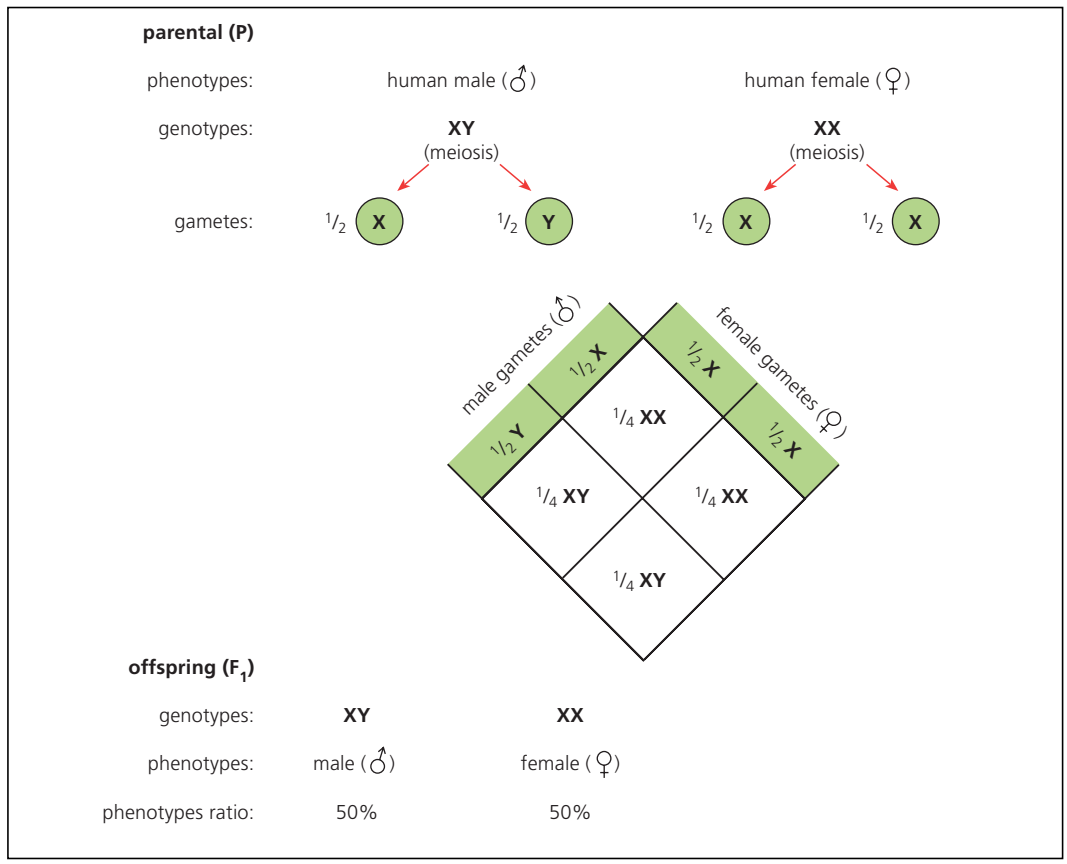


Figure 4.22 X and Y chromosomes and the determination of sex

Human X and Y chromosomes and the control of gender

Initially, male and female embryos develop identically in the uterus. At the seventh week of pregnancy, however, a cascade of developmental events is triggered, leading to the growth of male genitalia if a Y chromosome is present in the embryonic cells.

On the Y chromosome is the **prime male-determining gene**. This gene codes for a protein – the **testis-determining factor (TDF)**. TDF functions as a molecular switch; on reaching the embryonic gonad tissues, TDF initiates the production of a relatively low level of testosterone. The effect of this hormone at this stage is to inhibit the development of female genitalia, and to cause the embryonic genital tissues to form testes, scrotum and penis.

In the absence of a Y chromosome, the embryonic gonad tissue forms an ovary. Then, partly under the influence of hormone from the ovary, the female reproductive structures develop.

Extension: Non-disjunction and the sex chromosomes

We have noted already that non-disjunction in meiosis results in gametes with more and less than the haploid number of chromosomes (page 98). Table 4.3 identifies cases that occur in the determination of gender, and how frequently they arise.

Genotype	Condition	Incidence at birth
XO ♀	**Turner's syndrome:** common at conception, but normally fatal (99% naturally aborted); infertile adult	1 in 2500
XXY ♂	**Klinefelter's syndrome:** male with feminine features; often infertile; typically mildly mentally disabled	1 in 1000
XXX ♀	normal individual	1 in 1000
XYY ♂	almost normal individual	1 in 1000

Table 4.3 Non-disjunction of X and Y chromosomes

Pairing of X and Y chromosomes in meiosis

We know that homologous chromosomes pair up early in meiosis (Figure 4.5, page 95). Pairing is an essential step in the mechanism of meiosis. However, only a very small part of the X and Y chromosomes of humans have complementary alleles and can pair up during meiosis. The bulk of the X and Y chromosomes contain genes that have no corresponding alleles on the other (Figure 4.23).

The short Y chromosome carries genes specific for male sex determination and sperm production, including the gene that codes for TDF, which switches development of embryonic gonad tissue to testes early in embryonic development.

The X chromosome carries an assortment of genes, very few of which are concerned with sex determination.

We shall examine the effects of some of the alleles of these genes next.

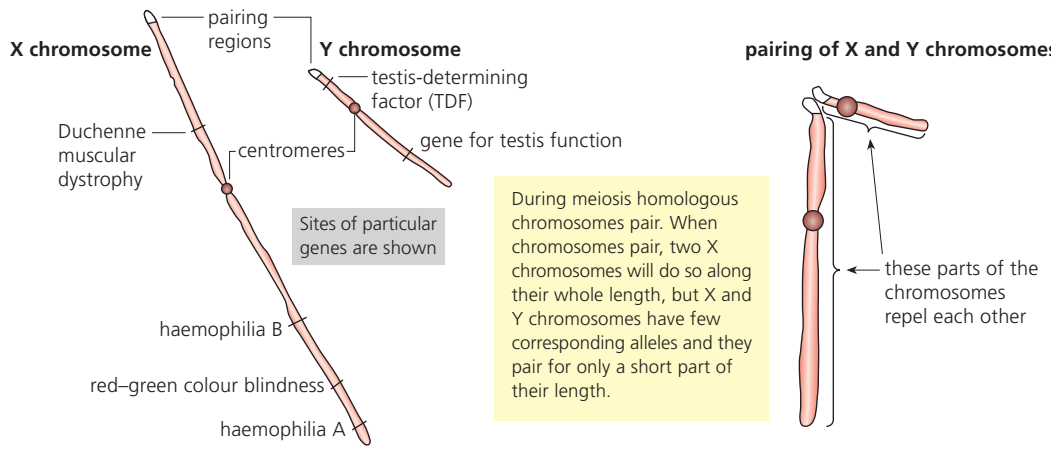


Figure 4.23 The pairing of X and Y chromosomes in meiosis

Sex linkage

Genes present on the sex chromosomes are inherited with the sex of the individual. They are said to be sex-linked characteristics.

Sex linkage is a special case of linkage occurring when a gene is located on a sex chromosome (usually the X chromosome).

The inheritance of these sex-linked genes is different from the inheritance of genes on the autosomal chromosomes. This is because the X chromosome is much longer than the Y chromosome (because many of the genes on the X chromosome are absent from the Y chromosome). In a male (XY), an allele present on the X chromosome is most likely to be on its own and will be apparent in the phenotype *even if it is recessive*.

Meanwhile, in a female, a single recessive gene is often masked by a dominant allele on the other X chromosome, and in these cases the recessive allele is not expressed. A human female can be homozygous or heterozygous with respect to sex-linked characteristics whereas males will have only one allele.

An individual with a recessive allele of a gene that does not have an effect on their phenotype (i.e. is heterozygous for that allele) is known as a **carrier**. They carry the allele but it is not expressed. So, female carriers are heterozygous for sex-linked recessive characteristics. Of course, the unpaired alleles of the Y chromosome are all expressed in the male. However, the alleles on the (short) Y chromosome are mostly concerned with male structures and male functions.

Examples of recessive conditions controlled by genes on the X chromosome are: Duchenne muscular dystrophy, red–green colour blindness, and haemophilia. In the case of these **genetically controlled diseases**, if a single recessive allele is present in a male human, the disease-triggering allele will be expressed. However, a female must be homozygous recessive for a sex-linked characteristic for the allele to be expressed.

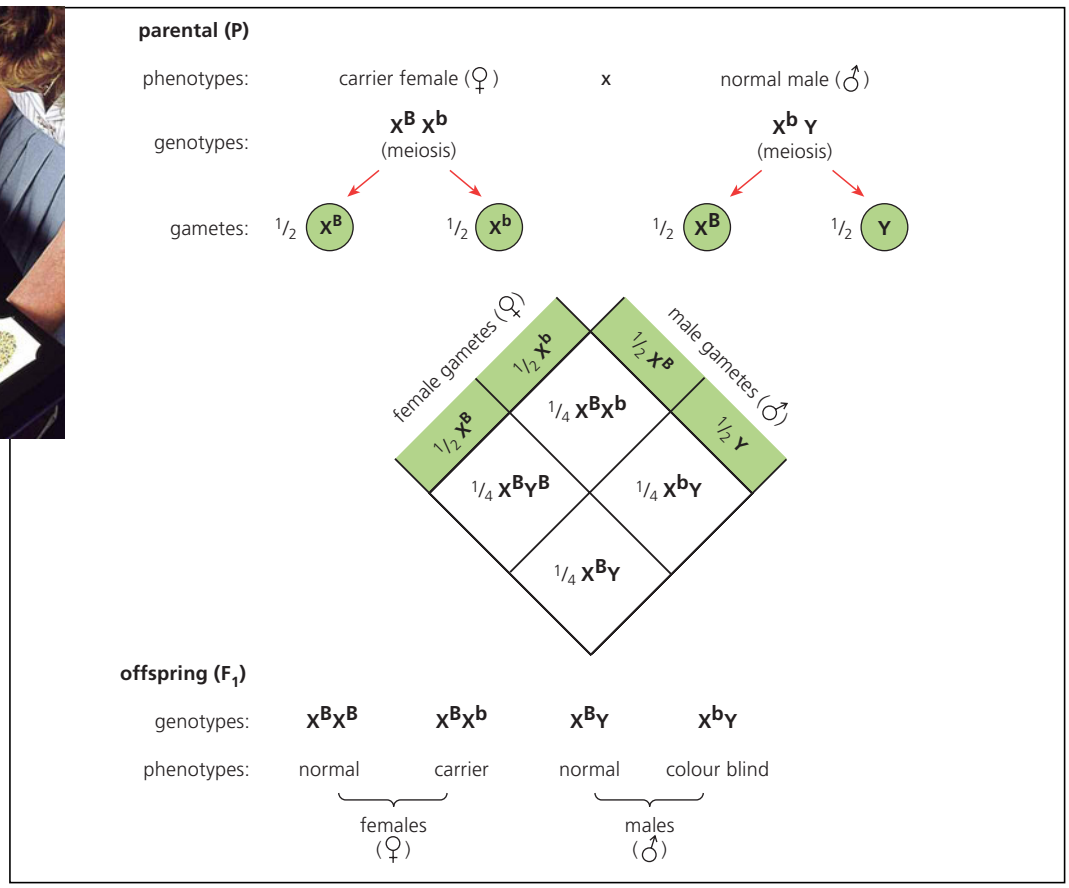
Red–green colour blindness

A red–green colour blind person sees green, yellow, orange and red as all the same colour. The condition afflicts about 8% of males, but only 0.4% of females in the human population. This is because a female with normal colour vision may by homozygous for the normal colour vision allele (X^BX^B) or she may be heterozygous for normal colour vision (X^BX^b). For a female to be red–green colour blind, she must be homozygous recessive for this allele (X^bX^b), and this occurs extremely rarely. On the other hand a male with a single recessive allele for red–green colour vision (X^bY) will be affected.

The inheritance of red–green colour blindness is illustrated in Figure 4.24. It is helpful for those who are red–green colour blind to recognise their inherited condition. Red–green colour blindness is detected by the use of multicoloured test cards.

Colour blindness is detected by multicoloured test cards. A mosaic of dots is arranged on the cards so that those with normal vision see a pattern that is not visible to those with colour blindness.

Figure 4.24 Detection and inheritance of red–green colour blindness



11 **State** how the genetic constitution of a female who is red–green colour blind is represented. **Explain** why it is impossible to have a 'carrier' male.

Haemophilia

If a break occurs in the circulatory system of a mammal, there is a risk of uncontrolled bleeding. This is normally overcome by the blood clotting mechanism that causes any gap to be plugged (page 349). Haemophilia is a rare, genetically determined, condition in which the blood will not clot normally. The result is frequent, excessive bleeding.

There are two forms of haemophilia, known as haemophilia A and haemophilia B. They are due to a failure to produce adequate amounts of particular blood proteins essential to the complex blood clotting mechanism. Today, haemophilia is effectively treated by the administration of the clotting factor the patient lacks.

Haemophilia is a sex-linked condition because the genes controlling production of the blood proteins concerned are located on the X chromosome. Haemophilia is caused by a recessive allele. As a result, haemophilia is largely a disease of the male since a single X chromosome carrying the defective allele (X^hY) will result in disease. For a female to have the disease, she must be homozygous for the recessive gene (X^hX^h), but this condition is usually fatal in the uterus, typically resulting in a natural abortion.

A female with only one X chromosome with the recessive allele (X^HX^h) is described as a carrier. She has normal blood clotting. When a carrier is partnered by a normal male, there is a 50% chance of the daughters being carriers and a 50% chance of the sons having haemophilia (Figure 4.25).

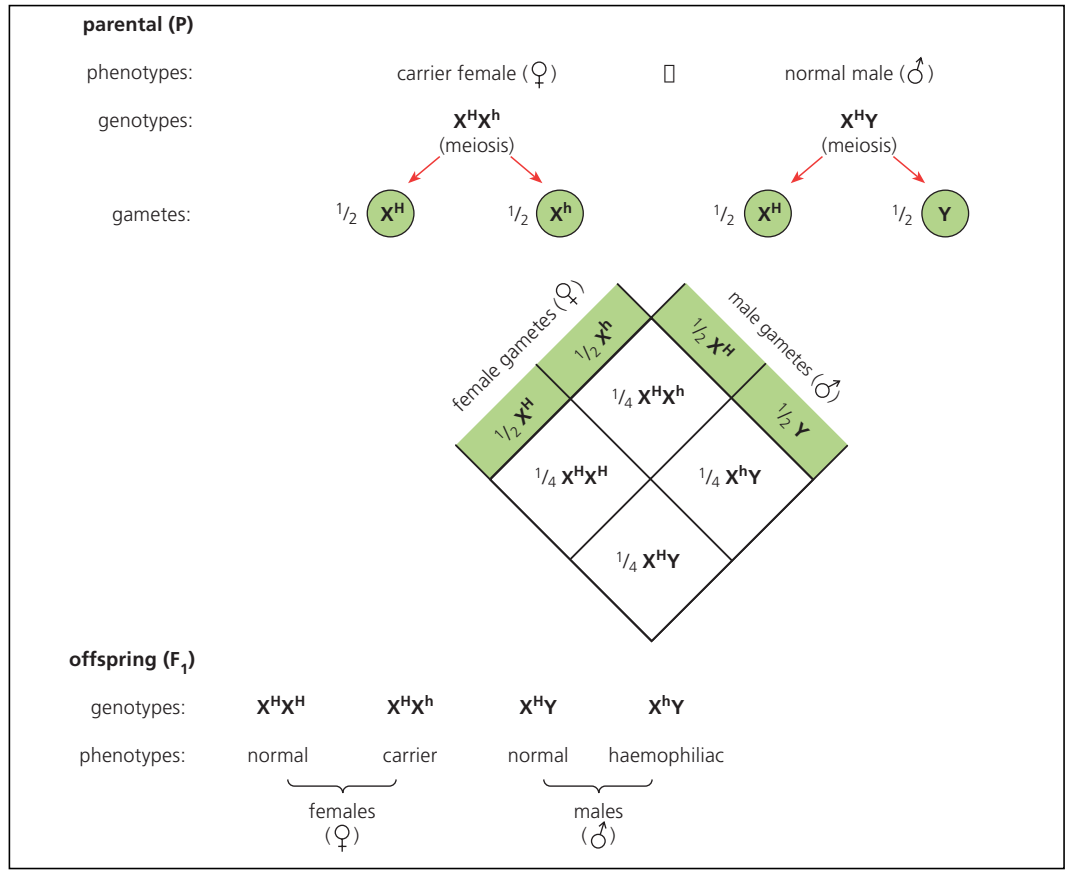


Figure 4.25 The inheritance of haemophilia

12 Haemophilia results from a sex-linked gene. The disease is most common in males, but the haemophilia allele is on the X chromosome. **Explain** this apparent anomaly.

Examination questions – a selection

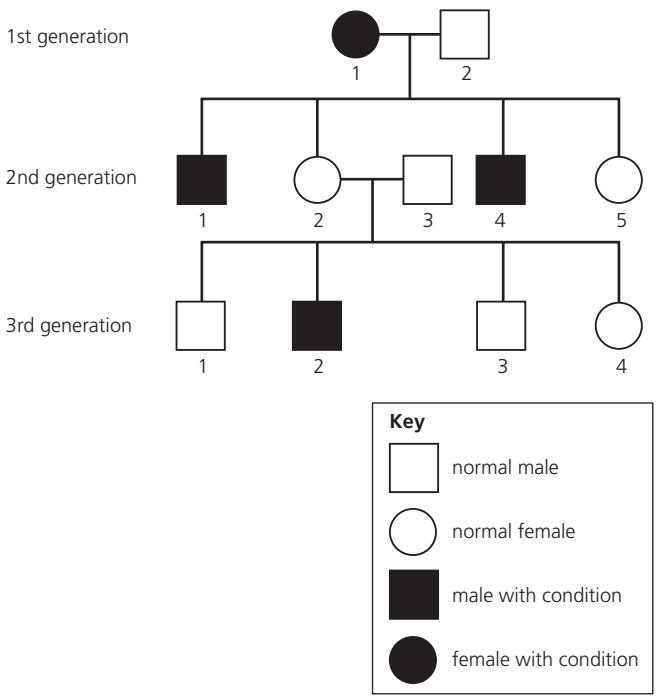
Questions 1–4 are taken from past IB Diploma biology papers.

Q1 Which response summarises meiosis?

	Pairing of chromosomes	Number of divisions	Result
A	no	one	two diploid cells
B	no	two	four diploid cells
C	yes	one	two haploid cells
D	yes	two	four haploid cells

Higher Level Paper 1, May 02, Q8

Q2 The diagram below shows the pedigree of a family with red–green colour blindness, a sex-linked condition.



a Define the term *sex linkage*. (1)

b Deduce, with a reason, whether the allele producing the condition is dominant or recessive. (2)

c **i** Determine all the possible genotypes of the individual (2nd generation-1) using appropriate symbols. (1)

ii Determine all the possible genotypes of the individual (3rd generation-4) using appropriate symbols. (1)

Standard Level Paper 2, May 03, Q2

Q3 What is the genetic cross called between an individual of unknown genotype and an individual who is homozygous recessive for a particular trait?

A test cross
B hybrid cross
C dihybrid cross
D F_1 cross

Higher Level Paper 1, May 06, Q11

Q4 When red shorthorn cattle are crossed with white shorthorn cattle, the offspring are roan, a colour that has both red and white hairs. What does this cross illustrate?

A codominance
B multiple alleles
C sex linkage
D mutation

Standard Level Paper 1, May 03, Q15

Questions 5 and 6 cover other syllabus issues in this chapter.

Q5 Mendel carried out a breeding experiment with garden pea plants in which pure-breeding pea plants grown from seeds with a smooth coat were crossed with plants grown from seeds with a wrinkled coat. All the seeds produced (called the F_1 generation) were found to have a smooth coat. When plants were grown from these seeds and allowed to self-pollinate, the second generation of seeds (the F_2 generation) included both smooth and wrinkled seeds in the ratio of 3:1.

a In Mendel's explanation of his results he used the term 'hereditary factor'. Identify our term for 'factors' and state where in the cell they occur. (2)

b The parent plants have diploid cells, while the gametes (sex cells) are haploid. Define what we mean by *haploid* and *diploid*. (4)

c By reference to the above experiment, define the following terms and give an example of each:

i *homozygous* and *heterozygous* (4)

ii *dominant* and *recessive* (4)

iii *genotype* and *phenotype*. (4)

d Using appropriate symbols for the alleles for smooth and wrinkled coat, construct a genetic diagram (including a Punnett grid) to show the behaviour of the alleles in this experiment. (6)

Q6 **a** Define the term *mutation*. (2)

b By means of specific examples, explain the difference between chromosome mutation and gene mutation. (4)

5 Genetic engineering and biotechnology

STARTING POINTS

- Chromosomes are a linear sequence of **genes**. They consist of **a length of DNA double helix**, hundreds or thousands of base pairs long.
- The sequence of bases in nucleic acid forms the **genetic code**. The code dictates the order in which specific amino acids are assembled and combined to make **proteins**.
- The **code is universal** to prokaryotes and eukaryotes.
- Many of the proteins coded for by genes and produced in cells function as **enzymes**. Reactions of metabolism are catalysed by specific enzymes.
- **Bacteria are prokaryotes** and have a distinctive cell structure. They lack chromosomes enclosed in a nucleus, but have a **nucleoid** – a single, circular chromosome of DNA in the cytoplasm, attached to the plasma membrane.

The practical skill of animal and plant breeding made possible the first great revolution in human history. This is known as the **Neolithic Revolution**. It commenced after the last Ice Age, about 10 000 years ago. At this time, groups of *Homo sapiens* stopped their nomadic existence and became farmers in settled communities.

The greatest achievement of these Neolithic (new stone age) peoples was the breeding of **domesticated animals** and the cultivation of many **crop plants**, all from the wild stock around them. This technology, now called **artificial selection**, was developed without theoretical knowledge of heredity. Many of the plant and animal species of modern farming were first bred at this time.

Today a new type of genetic modification is also in use. Genes from one organism are transferred to the set of genes (the genome) of another, unrelated organism. The process is called recombinant DNA technology or **genetic engineering**. The outcomes are new varieties of organisms, mostly but not exclusively of microorganisms such as bacteria. Genetic material can be transferred because the **genetic code is universal** (page 251). This means that any particular triplet code represents the same amino acid whether it is being read in a bacterium, cyanobacterium, green plant, fungus or animal. Thus the amino acid sequence of the protein coded for by the gene is the same wherever that gene occurs or is placed.

Gene technology has important applications in, for example, the pharmaceuticals industry, medicine, agriculture, horticulture and forensic science. It is a branch of **biotechnology** – the industrial and commercial application of biology. Another aspect of biotechnology is the industrial exploitation of enzymes (page 61).

1 **Define** the terms 'genome' and 'gene'.

Genetic engineering and its applications 4.4.1–4.4.13

The steps of genetic engineering

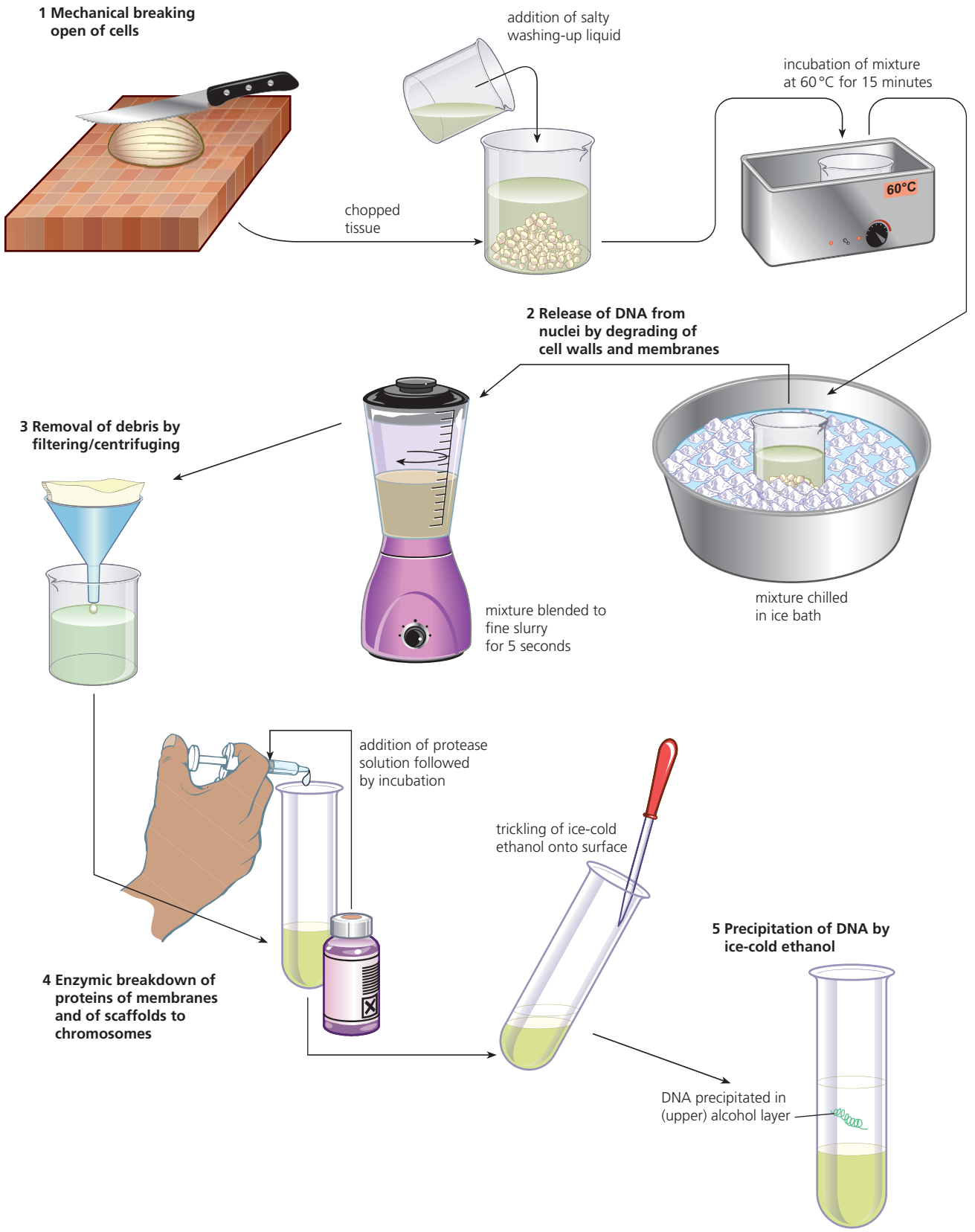
The five steps of genetic engineering are discussed, each in turn, below.

Isolating the gene from the source genome

The long double strand of DNA of a chromosome is functionally divided into genes, the units of inheritance. A chromosome holds a linear sequence of very many genes. For example, there is an average of 2200 genes on each human chromosome. Genes are variable in their length, but usually consist of hundreds or thousands of bases. Each gene codes for a particular protein. For genetic engineering to take place, individual genes may have to be isolated (and eventually identified).

How can one gene be found and isolated among so many?

Figure 5.1 The extraction of DNA – the procedure or protocol



DNA may be extracted from tissue samples by mechanically breaking up the cells, filtering off the debris, and breaking down cell membranes by treatment with detergents. After this, the protein framework of the chromosomes, often called a scaffold, is removed by incubation with a protein-digesting enzyme (protease). As a result the DNA is now present as long naked threads and can be isolated from the mixture of chemicals by precipitation with ethanol; it is thus 'cleaned' (Figure 5.1). The DNA strands are then re-suspended in aqueous, pH-buffered medium and are ready for 'slicing' into fragments.

2 **State** the products of digestion of proteins by protease enzymes.

Cutting up of DNA

Genetic engineering really began with the discovery of restriction endonucleases (**restriction enzymes**). These enzymes occur naturally in bacteria. Here they protect against the activity of viruses by cutting up viral DNA that enters the bacterium, and thereby completely inactivating it. Viral DNA might otherwise take over the host cell. The viruses that specifically parasitise bacteria have a special name. They are called bacteriophage or phage. Restriction enzymes were so named because they *restrict* the multiplication of phage viruses.

Many different restriction enzymes have been discovered and purified, and today they are used widely in genetic engineering experiments. Restriction enzymes cut at particular base sequences, and are of two types, forming either **blunt ends** or **sticky ends** to the cut fragments. Sticky ends are single-stranded extensions formed in the double-stranded DNA after 'digestion' with a restriction enzyme that cuts in a staggered fashion (Figure 5.2).

In experiments, a selected restriction enzyme is used to cut the DNA extracted from cells of an organism known to carry a required gene.

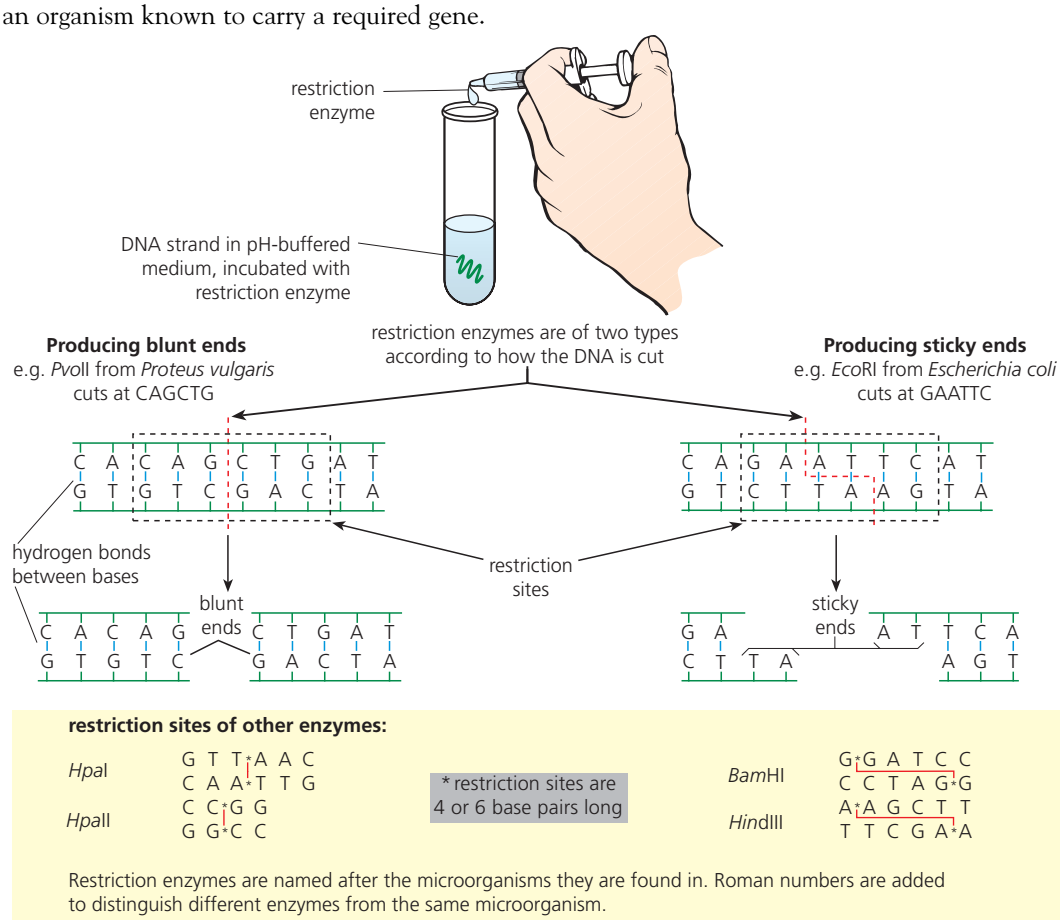


Figure 5.2 The role of restriction endonucleases (restriction enzymes)

3 **Outline** why we describe DNA as double stranded.

Isolating DNA fragments – electrophoresis

Electrophoresis is a process used to separate particles, including biologically important molecules such as DNA, RNA, proteins and amino acids. Electrophoresis is typically carried out on an agarose gel (a very pure form of agar) or on polyacrylamide gel (PAG). Both these substances contain tiny pores which allow them to act like a **molecular sieve**. Small particles can move quite quickly through these gels whereas larger molecules move much more slowly.

Biological molecules separated by electrophoresis also carry an **electrical charge**. In the case of DNA, it is the phosphate groups in DNA fragments that give them a net negative charge. Consequently, when these molecules are placed in an electric field they migrate towards the positive pole.

So, in electrophoresis, separation occurs according to the size of the molecule and the charge carried. This is the double principle of electrophoretic separations. Separation of DNA fragments produced by the actions of restriction enzyme is shown in Figure 5.3.

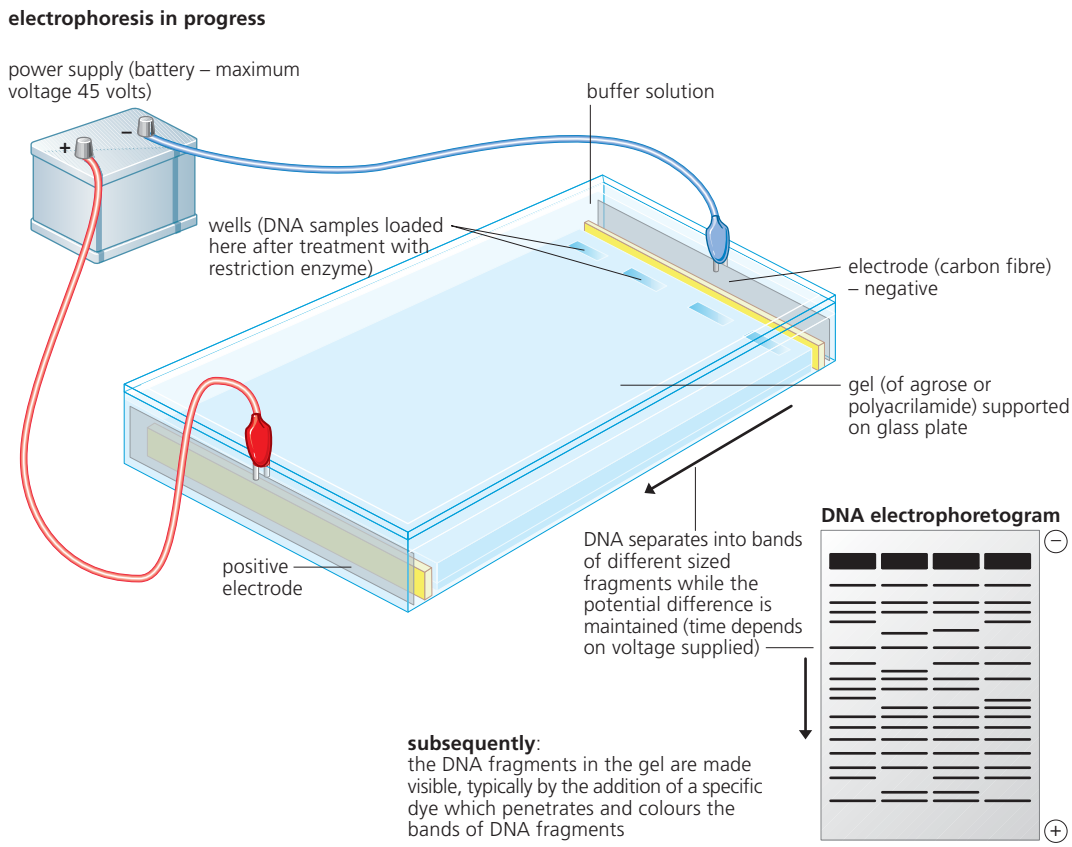


Figure 5.3 Electrophoretic separation of DNA fragments

Vectors for cloning

Electrophoresis is used to sort the DNA fragments by size. The required gene will occur in fragments of a particular size (Figure 5.4). The gene fragment required can be identified eventually, after it has been copied (a process known as **cloning**). For these steps to be carried out, the gene has to be added to a **vector** by which it is transferred to a receiving organism. These steps and techniques of genetic engineering are outlined next.

In biology, a vector is an agent that transports between one organism and another. In genetic engineering, a vector is 'carrier DNA' into which a DNA fragment containing a particular gene can be inserted, and which is then used to introduce the gene into a new organism. One such vector is commonly found in bacteria.

A bacterium such as *E. coli* (page 16) contains two types of genetic material. One is the long, double strand of DNA in the form of a ring, which we can think of as a single, circular chromosome. This is called the **nucleoid**.

Figure 5.4 Locating the required gene

Restriction endonuclease enzyme cuts DNA strands into fragments of differing lengths at the restriction sites. One fragment will contain the **required gene.**

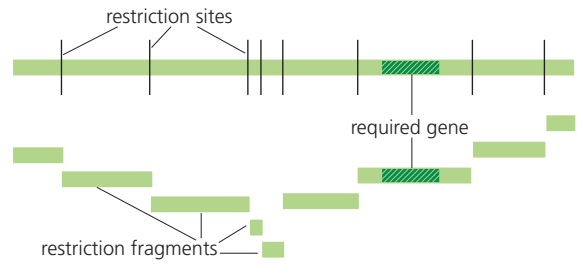
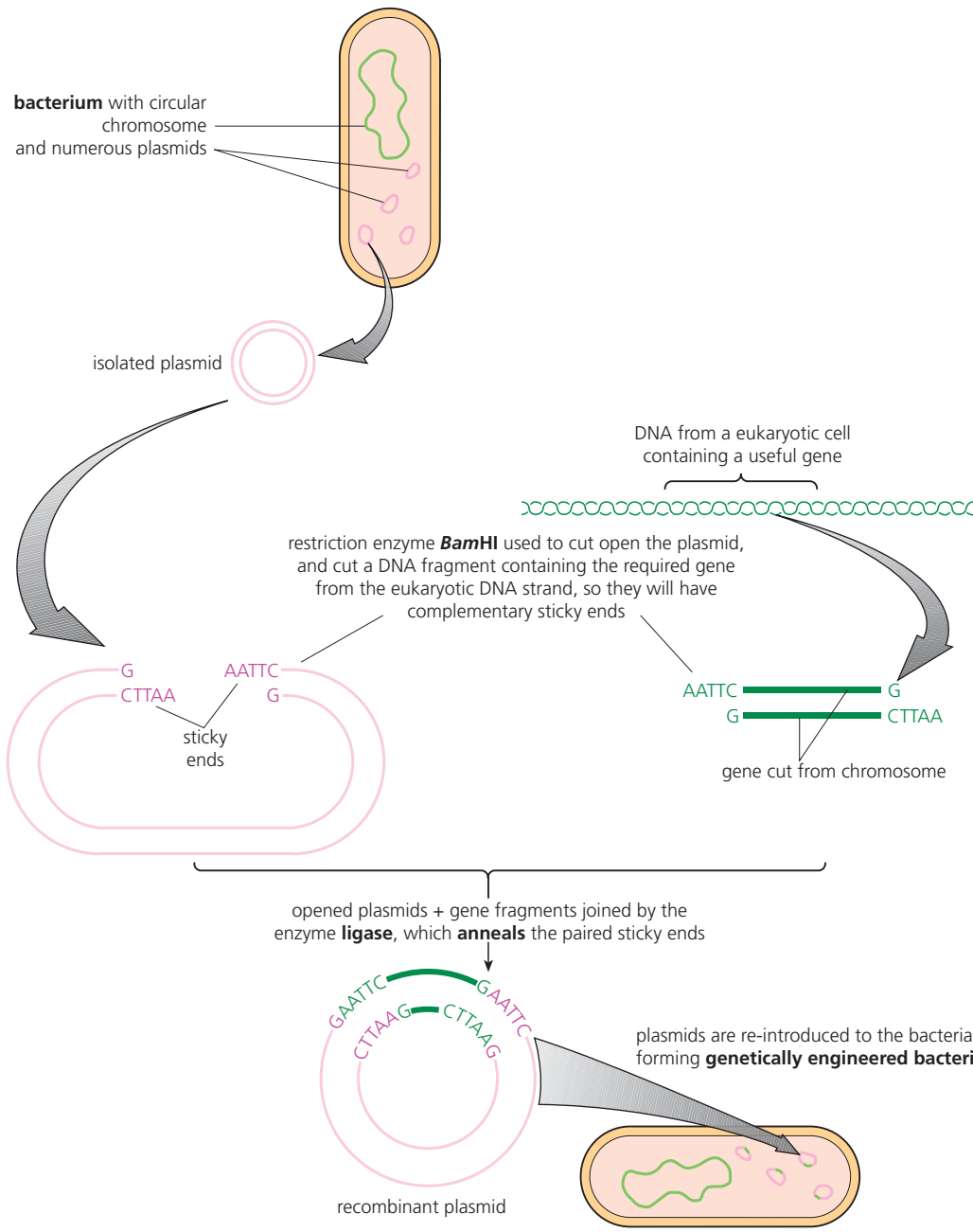


Figure 5.5 Plasmids as vectors



The other type of DNA in bacteria takes the form of numerous smaller rings, called **plasmids** (Figure 5.5). Plasmids are easily isolated from a bacterium and can be re-introduced to a bacterial cell, too. They are therefore extremely useful as vectors. In the bacterium, plasmids copy themselves (replicate) independently of the chromosome, so any gene added to a plasmid is also copied many times. This is how the isolated gene is cloned, once inside a bacterial cell.

Using plasmids as vectors

To prepare isolated plasmids as vectors, a restriction enzyme that forms sticky ends is used to cut open the plasmid. The *same* restriction enzyme is used to cut out the gene from longer strands of DNA, previously separated by electrophoresis. In this way, both will carry complementary sticky ends. Then the opened plasmids and the gene-containing fragments may be combined together by a second enzyme, **ligase**.

Ligase occurs naturally in the nuclei of organisms where it will repair DNA damaged in replication. Ligase catalyses the joining of sugar–phosphate backbones of adjacent DNA strands (a process called **annealing**), after their sticky ends have aligned by complementary base pairing (Figure 5.6).

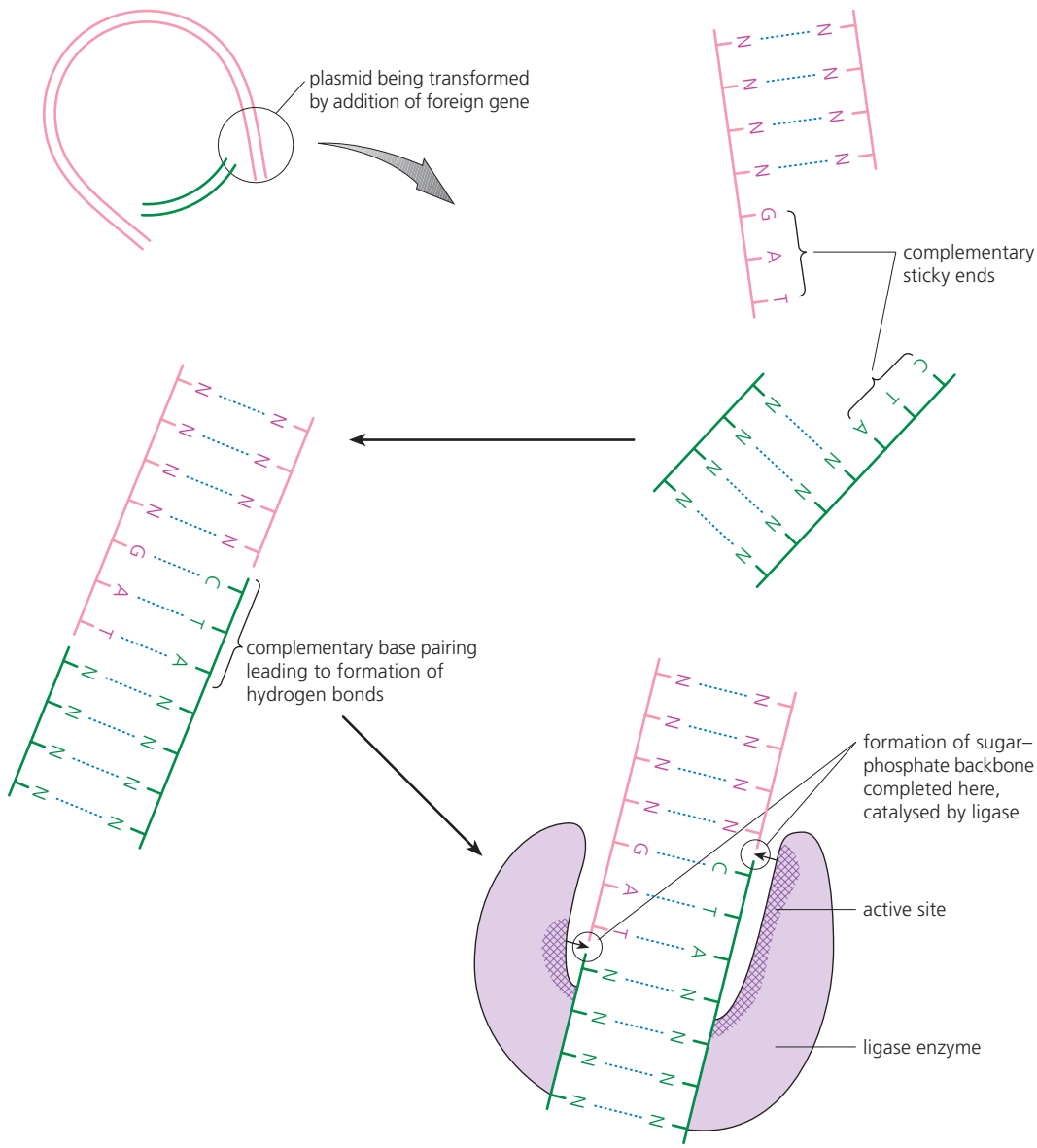


Figure 5.6 The action of ligase

4 **State** the property of DNA that makes complementary sticky ends hold together.

Finally, genetically engineered plasmids are re-introduced to the bacterium. This may occur by mixing bacterial cells and engineered plasmids in the presence of a salt (calcium chloride) which makes the bacterial cell walls especially permeable to plasmids. Some of the recombinant plasmids enter the cytoplasm of the bacteria. Because the genetic code is universal, the bacterium can translate the code and produce the gene's protein product.

Current uses of genetically modified organisms

Genetically modified (GM) bacteria, animals and plants all have roles in the world we live in today.

The production of transformed bacteria

One of the earliest applications of the techniques of genetic engineering was in the production of GM bacteria containing human genes, or genes taken from other eukaryotes. The manufacture of human growth hormone by the pharmaceutical industry is a good example.

Human growth hormone (hGH), also known as somatostatin, is one of a group of hormones normally synthesised in the pituitary gland and circulated around the body in the blood stream. It stimulates growth of body cells, particularly those of the skeleton and skeletal muscles. It works by enhancing cell protein synthesis, and the uptake of amino acids by the plasma membrane, from the plasma and tissue fluid. Also, it enhances conversion of glycogen to glucose in the liver, so tending to cause blood sugar level to rise.

Failure to secrete sufficient hGH in children and young people in their growth years leads to a condition called **pituitary dwarfism**. This can be overcome by a series of injections of hGH (Figure 5.7). It proved extremely difficult to extract a significant quantity of the equivalent hormone from, for example, sheep – and such hormone was largely not effective in humans. Hence the importance of a supply of hGH from GM bacteria.

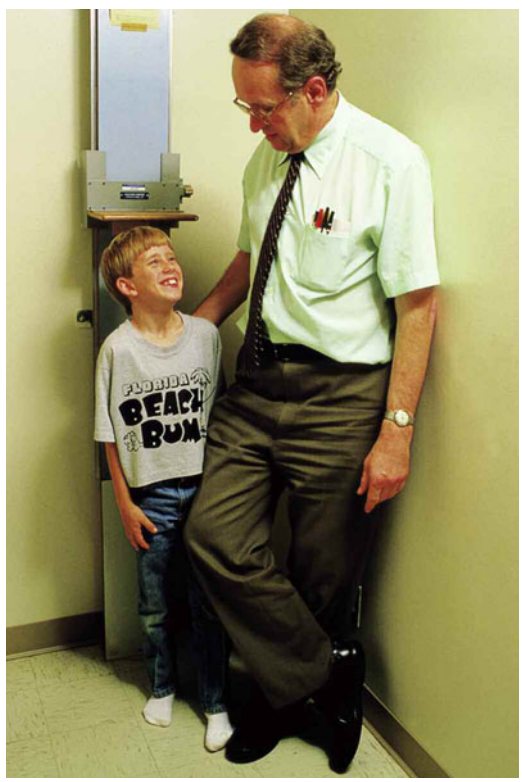


Figure 5.7 This 11-year-old boy required treatment with genetically engineered hGH to overcome the effects of childhood deficiency of the hormone, which resulted in dwarfism; regular injections were given and the effects monitored

This hormone is produced from cultures of the bacterium *E. coli* that have been engineered to carry the gene for hGH. This is a mutant form of *E. coli* – one that requires special laboratory conditions to survive at all. This is an obvious **safety precaution**, since otherwise genetically engineered bacteria might escape and exchange genes with a gut population of *E. coli*, with possibly unpredictable consequences.

Interestingly, hGH is a short protein (a polypeptide, in this case), only 14 amino acid residues in length. Consequently, it was found easier to manufacture an artificial gene coding for the hormone, than to isolate the real gene from human chromosomes. The artificial gene was inserted into plasmids and introduced into host bacteria. Transformed bacteria with the hGH gene 'switched on' are cultured in large vats called fermenters, where they produce hGH polypeptide in significant quantities. This is extracted and purified from the culture medium.

5 **Calculate** the *minimum* number of nucleotides an artificial gene (a polynucleotide) coding for hGH would need in length, given that in addition to codons for 14 amino acids, 'stop' and 'start' codons are also required, and the plasmid vector has to be 'cut open' with *Eco*RI restriction enzyme (Figure 5.2) to create sticky ends.

In eukaryotes

Manipulating genes in eukaryotes is a more difficult process than in prokaryotes. There are several reasons for this, including the following.

- Plasmids, the most useful vehicle for moving genes, do not occur in eukaryotes (except in yeasts) and, if introduced, do not always survive there to be replicated.
- Eukaryotes are diploid organisms, so two forms (alleles) for every gene are required to be engineered into the nucleus. By comparison, prokaryotes have a single, circular 'chromosome', so only one copy of the gene has to be engineered into their chromosome.

Producing transgenic animals

Transgenic organisms have genetic material introduced artificially from another organism. Despite the difficulties of engineering eukaryotic cells, several varieties of transgenic animal have been produced. For example, transgenic sheep have been successfully engineered to produce in their milk rare and expensive human proteins that may be useful as medicines.

One example is the production of a special human blood protein, known as AAT. Production of **AAT protein** in our bodies enables us to **maintain lung elasticity** which is essential in breathing movements. Patients with a rare genetic disease are unable to manufacture AAT protein at all, and they develop emphysema (destruction of the walls of the air sacs, so that the lungs remain full of air during expiration).

The chemical industry is unable to manufacture AAT protein in the laboratory on a practical scale. However, the human gene for AAT protein production has been identified and isolated, and it has been cloned into sheep, together with a promoter gene (a sheep's milk protein promoter) attached to it. Consequently the sheep's mammary glands produce the human protein and secrete it in their milk, during lactations (Figure 5.8). Thus, human AAT protein is made available for use with patients.

Another product important to medicine is the human blood clotting factor known as factor IX, now obtained from the milk of GM sheep for use with patients whose bodies are unable to synthesise this protein.

6 **Explain** why a gene that is inserted into the nucleus of a fertilised egg cell is also passed to the progeny of the animal that forms from the zygote.

Producing transgenic plants

Many commercially valuable plant species have been genetically engineered and field trials undertaken. The crops most involved are cotton, tobacco, oilseed rape, maize, potatoes, soya and tomatoes. The improvements achieved include:

- herbicide resistance in crop plants, so that application of weed killer will remove weeds without harm to the crop (Figure 5.10, page 128);
- rice varieties that contain β carotene (vitamin A precursor);
- tomato varieties able to ripen on the plant and develop full flavour, without rotting quickly when picked and marketed as red fruit (traditionally, tomatoes are picked green and induced to ripen but in this state they may lack flavour);
- resistance to insect pests – crop plant tissue contains a naturally occurring insecticide which is harmless to other animal species;
- oilseed rape varieties that make and store commercially useful oils;
- thale cress (*Arabidopsis thaliana*), engineered as a cheap, renewable source of the biodegradable plastic called polyhydrobutyrate (most plastics come from petrochemicals and are slow to degrade; reserves of petrochemicals are limited).

Transgenic flowering plants may be formed using tumour-forming *Agrobacterium*. This soil-inhabiting bacterium sometimes invades broad-leaved plants at the junction of stem and root, forming a huge growth called a tumour or crown gall. The gene for tumour formation occurs naturally in a plasmid in the bacterium, known as a Ti plasmid. Useful genes may be added to the Ti plasmid, using restriction enzyme and ligase, and the recombinant plasmid placed back into *Agrobacterium*. A host crop plant is then infected by the modified bacterium. The gall tissue that results may be cultured into independent plants, all of which also carry the useful gene. All the plants will make the particular useful gene product that has been introduced (Figure 5.9).

7 **Explain** why plants formed from the gall tissue will contain the recombinant gene that has been added to the Ti plasmid.

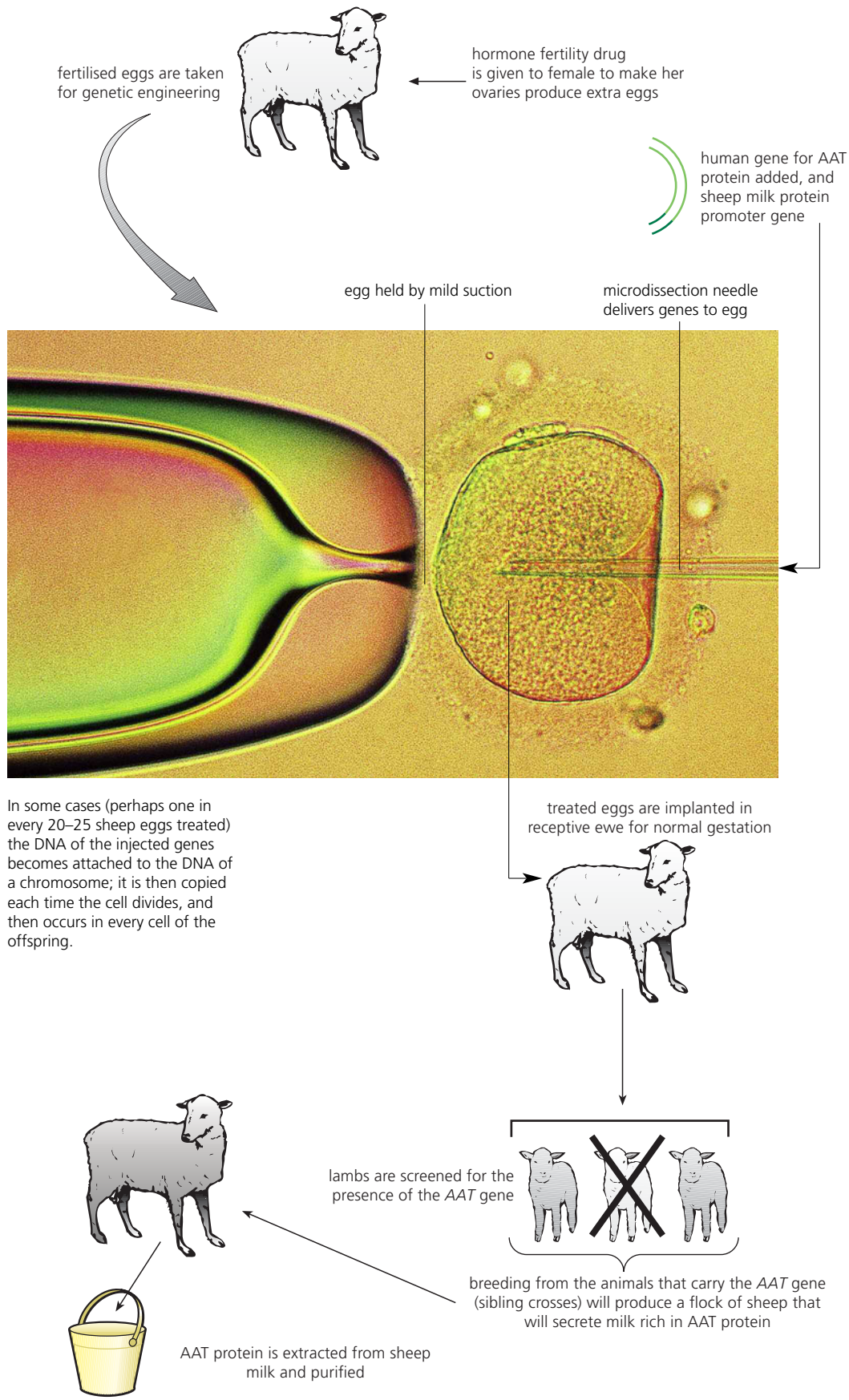
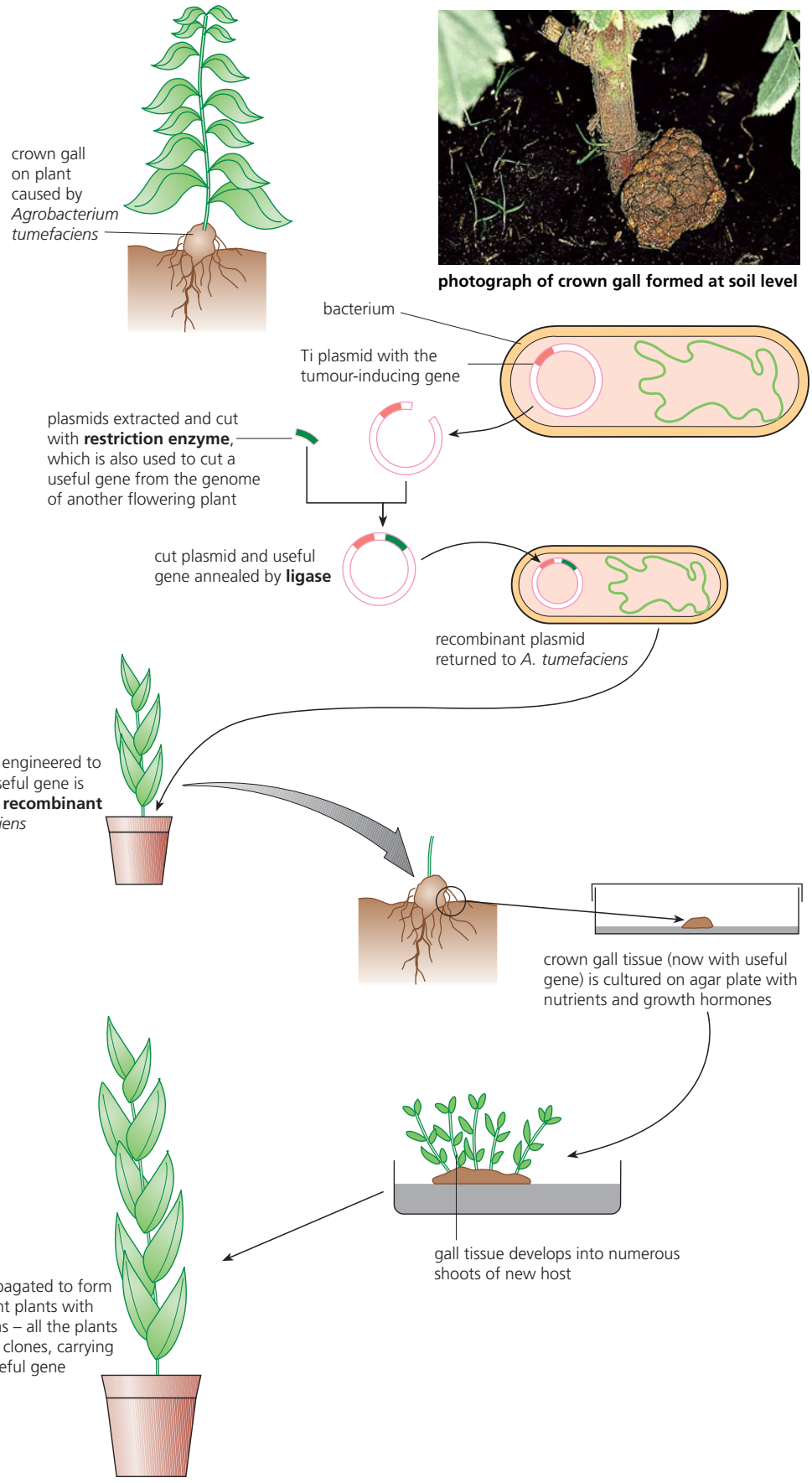


Figure 5.8 The cloning of transgenic sheep to secrete human AAT protein

Figure 5.9 Transgenic plants via the Ti plasmid of *Agrobacterium*



Issues raised by genetic engineering

There are many examples of genetic modification of organisms that are beneficial. However, geneticists are really producing new organisms when genes are transferred. Consequently, this work is potentially a source of hazards and it certainly generates concerns. These include the following.

- Will a gene, added to a genome, function in an unforeseen manner – perhaps, for example, triggering some disease in the recipient?
- Might an introduced gene for resistance to adverse conditions get transferred from a crop plant or farm animal into a weed species or to some predator?
- Is it possible that a harmless organism such as the human gut bacterium *E. coli* might, with recombinant DNA technology, be transformed into a harmful pathogen that escapes the laboratory and infects the population?
- Is there an important overriding principle that humans should not 'change nature' in a deliberate way?
- Genetic engineering is a costly technology, mostly beneficial to the health and life expectancy of people of developed nations. If the funds were made available for more basic problems of housing, health and nutrition in less developed countries instead, vastly more humans would benefit immediately.

An evaluation of the potential benefits and possible harmful effects of genetic modification of plants is presented in the following case study (Table 5.1 and Figure 5.10).

Case study: Consider the production of herbicide-tolerant crops, such as sugar beet modified to be tolerant of glyphosate herbicide applications. This herbicide normally kills plants it comes in contact with, but is inactivated in a 'tolerant' plant (Figure 5.10).

Benefits	Dangers
Use of herbicide eradicates weeds around sugar beet crops, where loss of yield due to competition with weeds is very high.	Fewer weed species on farmland, and longer periods without weeds growing, breaks the wildlife food chain.
Glyphosate herbicide is transported all around the weed plant, killing even the largest weeds with extensive roots.	Selection of a glyphosate-resistant crop plant ties the grower to one particular herbicide product – choice is lost.
Glyphosate inhibits an enzyme for the production of essential amino acids in plants. This enzyme is absent from animals, so glyphosate has very low toxicity.	Will GM plant material, when consumed by humans, release novel toxins or otherwise adversely affect enzyme systems in the human digestive system?
An average of five sprays of herbicides are applied to sugar beet in conventional cultivation, whereas with GM sugar beet a maximum of three spays (and normally only one or two) are needed because of the greater activity of glyphosate.	Genetically engineered genes are vectored in plasmids containing an antibiotic-resistant gene – to facilitate GM processes in the lab. This latter gene might be accidentally transferred to human gut bacteria, via food eaten.
Cereals normally precede spring-sown sugar beet. Cereal stubble is an important source of food for bird life in autumn and winter. With glyphosate herbicide applied to GM beet being so effective, there is no need for weed control in stubble the preceding autumn – so wild birds' food sources are available for longer.	'Superweeds' may develop by cross-pollination between herbicide-tolerant crop plants and compatible weed species (fat-hen, sea beet and weed beet are naturally occurring near relatives of sugar beet). Superweeds would be difficult to eradicate from crops.
Glyphosate herbicide provides good weed control, allowing cultivation by discs or tines (conservation tillage) rather than by ploughing. These systems have less harmful effects on soil organisms, and increase moisture, improve soil structure, and reduce tillage costs.	There is a possibility of cross-pollination between GM crops and conventional and organic crops. The maintenance of sufficient distance or the devising of effective barriers to prevent or reduce pollen transfers between crops may be difficult to achieve.
Glyphosate herbicide is applied as large droplets from coarse nozzles, rather than as a fine mist of tiny droplets that are prone to drift onto surrounding habitats.	If glyphosate herbicide droplets do reach hedgerow plants (or further), their size and chemical activity in plants means they are more likely to do damage.

Table 5.1 Benefits from and dangers of GM herbicide tolerance in a crop plant such as sugar beet

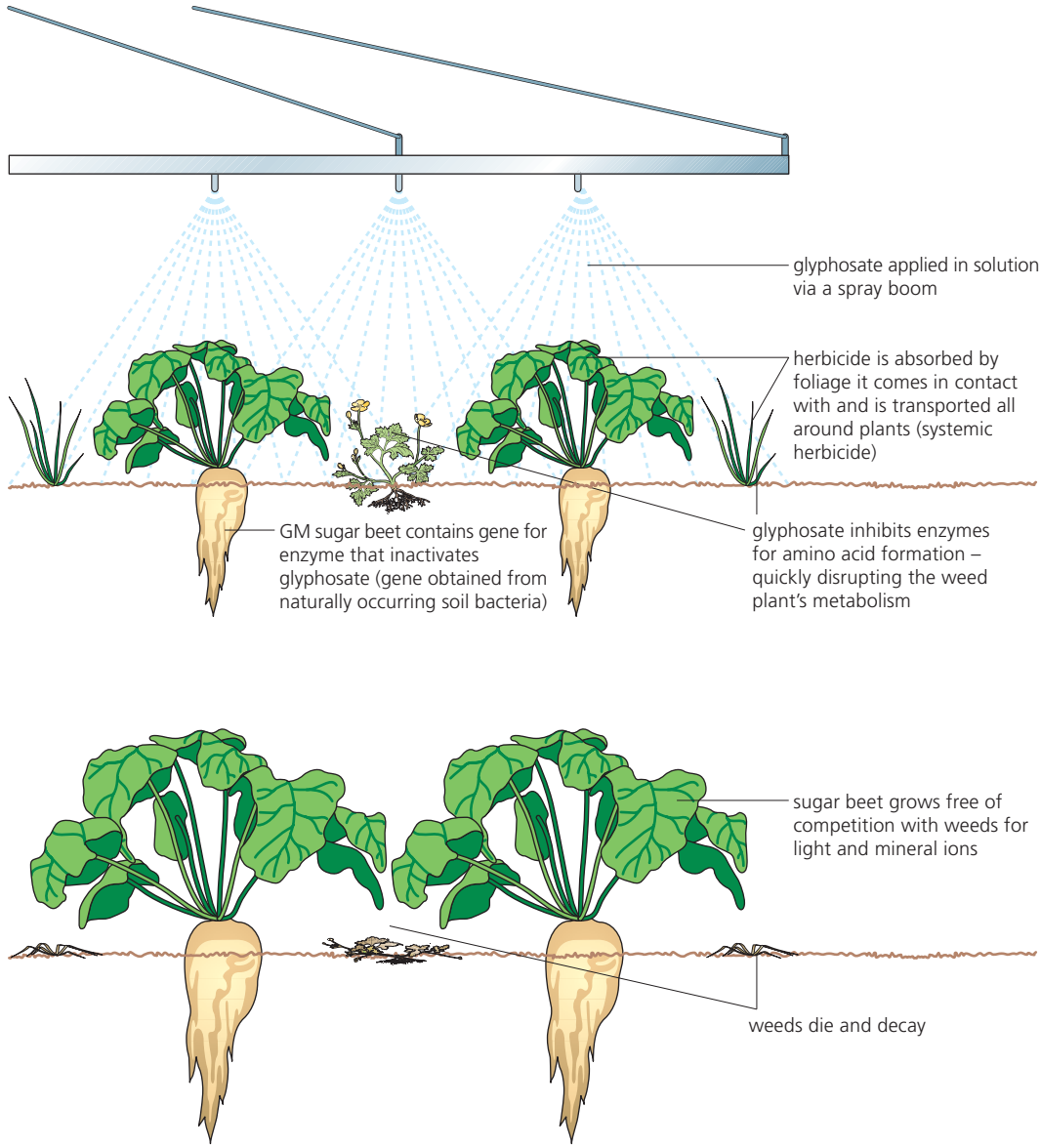


Figure 5.10 Herbicide-tolerant sugar beet and the action of glyphosate

8 **Suggest** what advantage would result from the eventual transfer of genes for nitrogen fixation from nodules in leguminous plants to cereals such as wheat.

TOK Link

'Frankenstein foods!' Occasional articles in the press make strong criticisms of genetic modification of food organisms, sometimes with alarmist headlines. Select such an article and prepare the following to use in a discussion with your peers:

1 a concise summary of the criticisms made, avoiding extremist language or unnecessary exaggeration

2 a list of balancing arguments in favour of genetic modification of food organisms

3 a concise statement of your own view on this issue.

DNA profiling

DNA profiling exploits the techniques of genetic engineering to identify a person or organism from a sample of their DNA. The DNA of our chromosomes, which is unique, includes that of our genes.

Actually, the bulk of our DNA does not code for proteins. In these regions, short sequences of bases are repeated many times. While some of these sequences are scattered throughout the length of the DNA molecule, many are joined together in major clusters. It is these major lengths of the non-coding, 'nonsense' DNA, often known as **satellite DNA**, that are used in genetic profiling. Between 10% and 25% of our total DNA is made up of short sequences of 5–10 nucleotides (**microsatellites**) repeated thousands of times. We inherit a distinctive combination of these apparently non-functional repeat regions, half from our mother and half from our father. Consequently, each of us has a unique sequence of nucleotides in our DNA (except for identical twins, who share the same pattern).

To produce a genetic 'fingerprint' a sample of DNA is cut with a restriction enzyme which recognises specific sequences in the DNA. Electrophoresis (Figure 5.3, page 120) is used to separate pieces according to length and size, and the result is a pattern of bands similar to a bar code (Figure 5.11). The steps to produce a DNA fingerprint are as follows.

1. A sample of cells is obtained from blood, semen, hair root, or body tissues, and the DNA is extracted. Where a tiny quantity of DNA is all that can be recovered, this is precisely copied (polymerase chain reaction, below) to obtain sufficient DNA to analyse.
2. The DNA is cut into small, double-stranded fragments using a particular restriction enzyme chosen because it 'cuts' close to, but not within, the satellite DNA.
3. The resulting DNA fragments are of varying lengths, and are separated by gel electrophoresis into bands (invisible, at this stage).
4. The gel is treated to split DNA into single strands, and then a copy is transferred to a membrane.
5. Selected, radioactively labelled DNA probes are added to the membrane to bind to particular bands of DNA, and then the excess probes are washed away.
6. The membrane is now overlaid with X-ray film which becomes selectively fogged by emission from the retained labelled probes.
7. The X-ray film is developed, showing up the positions of the bands (fragments) to which probes have attached. The result is a profile with the appearance of a bar code.

DNA profiling has applications in forensic investigations

DNA profiles are produced from samples taken from serious crime scenes, both from victims and suspects, as well as from others who have certainly not been involved in the crime, as a control. The greatest care has to be taken to ensure the authenticity of the sample. There must be no possibility of contamination if the outcome of subsequent testing is to be meaningful. DNA profiling helps eliminate innocent suspects, and to identify a person or people who may be responsible or related. It cannot prove with absolute certainty anyone's guilt or connection. An example is shown in Figure 5.11.

Crimes such as suspected murder and burglaries may be solved in cases where biological specimens are left at the scene of the crime, such as a few hair roots or a tiny drop of blood. Also, specimens may be collected from people suspected of being present at the crime. DNA fingerprinting may also help forensic scientists identify corpses otherwise too decomposed for recognition, or where only parts of the body remain, as may occur after bomb blasts or other violent incidents, including natural disasters.

Note: There are various circumstances where the amount of DNA available or which can be recovered (such as at a crime scene) is very small indeed – apparently too little for analysis, in fact. It is now possible to submit such minute samples to a process known as the **polymerase chain reaction** (**PCR**) in which the DNA is replicated in an entirely automated process, *in vitro*, to produce a large amount of the sequence. A single molecule is sufficient as the starting material, should this be all that is available. The product is exact copies in quantities sufficient for analysis.

Figure 5.11 DNA profiles used to investigate relatedness

1 DNA profiles used to establish family relationships

Is the male (F) the parent of both children?

Examine the DNA profiles shown to the right.

Look at the children's bands (C).

Discount all those bands that correspond to bands in the mother's profile (M).

The remaining bands match those of the biological father.

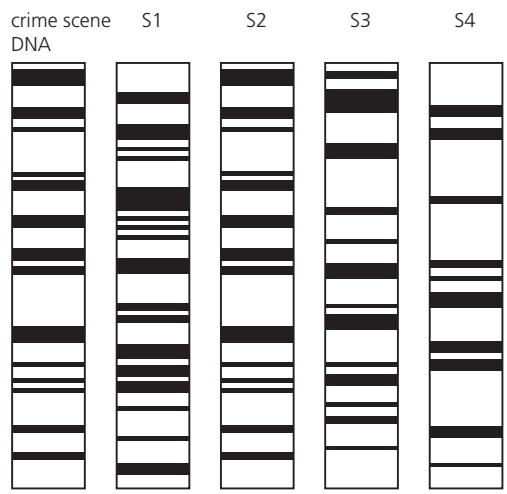
DNA fingerprinting has been widely applied in biology. In ornithology, for example, DNA profiling of nestlings has established a degree of 'promiscuity' in breeding pairs, the male of which was assumed to be the father of the whole brood. In birds, the production of a clutch of eggs is extended over a period of days, with copulation and fertilisation preceding the laying of each egg. This provides the opportunity for different males to fertilise the female.

2 DNA profiling in forensic investigation

Identification of criminals

At the scene of a crime (such as a murder or burglary), hairs – with hair root cells attached – or blood may be recovered. If so, the resulting DNA profiles may be compared with those of DNA obtained from suspects.

Examine the DNA profiles shown to the right, and suggest which suspect should be interviewed further.



Identification in a rape crime involves the taking of vaginal swabs. Here, DNA will be present from the victim and also from the rapist. The result of DNA analysis is a complex profile that requires careful comparison with the DNA profiles of the victim and of any suspects. A rapist can be identified with a high degree of certainty, and the innocence of others established.

Identification of a corpse which is otherwise unidentifiable is achieved by taking DNA samples from body tissues and comparing their profile with those of close relatives or with DNA obtained from cells recovered from personal effects, where these are available.

DNA profiling has applications in determining paternity

Another very important application of DNA profiling is in issues of parentage. A range of samples of DNA of the people who are possibly related are analysed side by side. The banding patterns are then compared (Figure 5.11). Because a child inherits half his DNA from his mother and half from his father, the bands in a child's DNA fingerprint that do not match his mother's must come from the child's father.

DNA profiling also has wide applications in studies of wild animals – for example, concerning breeding behaviour, and in the identification of unrelated animals as mates in captive breeding programmes of species in danger of extinction. Note the use of a DNA profile in Figure 5.13 (page 134), for example.

9 Distinguish between the following pairs:

a genotype and genome
b restriction endonuclease and ligase
c 'blunt ends' and 'sticky ends'
d bacterial chromosome and a plasmid.

Human Genome Project

The Human Genome Project (HGP), an initiative to map the entire human genome, is a publicly funded project that was launched in 1990.

The ultimate objectives of the HGP were to discover the location of each human gene and the base sequence within its DNA structure.

The work was shared among more than 200 laboratories around the world, avoiding duplication of effort. However, in 1998 the task became a race when a commercially funded company set out to achieve the same outcome in only three years, by different techniques. Because of a fear that a private company might succeed, patent the genome, and then sell access to it (rather than making the information freely available to all), the HGP teams accelerated their work.

In fact, both teams were successful well ahead of the project completion dates. On the 26th June 2000, a joint announcement made at the White House in Washington DC (USA) established that the sequencing of the human genome had been achieved. At the same time as the HGP had got under way, teams of scientists set about the sequencing of the DNA of other organisms, and more than 30 have been completed to date.

Outcomes of the HGP

Ultimately, most if not all aspects of biological investigation will be enhanced by the outcomes of this project. At this stage, we can illustrate the impact of the HGP with three examples.

How many individual genes do we have and how do they work?
The three billion bases that make up the human genome represent 30 000 to 45 000 genes – far fewer than the expected 100 000. This is only a few thousand more than the genes of a sea urchin. Even *Drosophila* (the fruit fly) has almost half our number. A rice plant has many more genes than a human – up to 55 000 genes.

How is the structural, physiological and behavioural complexity of humans delivered by so relatively few genes?

The secret of our complexity may lie not in the number of our genes but how we use them. Promoters and enhancers are associated with many genes and may determine which cells express genes, when they are expressed, and at what level.

Another issue is the effect of experience. The genetic code of DNA is not altered by the environment (except in cases of mutation) – information flows out of genes and not back into them. However, while our genes are immune to direct outside influence, can our experience regulate the expression of particular genes? Genes may be switched on and off by promoters, and this may occur in response to external factors. If so, our genes may be responding to our actions as well as causing them.

This field of investigation is in its infancy compared with its likely final impact.

Locating the cause of genetic disease

Another outcome of the HGP is the ability to locate genes responsible for human genetic diseases. Genetic diseases cause suffering in about 1–2% of the population.

More than half of all genetic diseases are due to a mutation of a single gene, and the mutant gene is commonly recessive. To prove a gene is associated with a disease it must be shown that patients have a mutation in that gene, but unaffected individuals do not. The outcome of the mutation in affected people is the loss of the ability to form the normal product of the gene. Common genetic diseases include cystic fibrosis (page 576), sickle cell disease (page 100), and haemophilia (page 114).

Development of gene therapy

The discovery of genes responsible for human genetic disease generates the possibility of a fundamental cure by gene therapy.

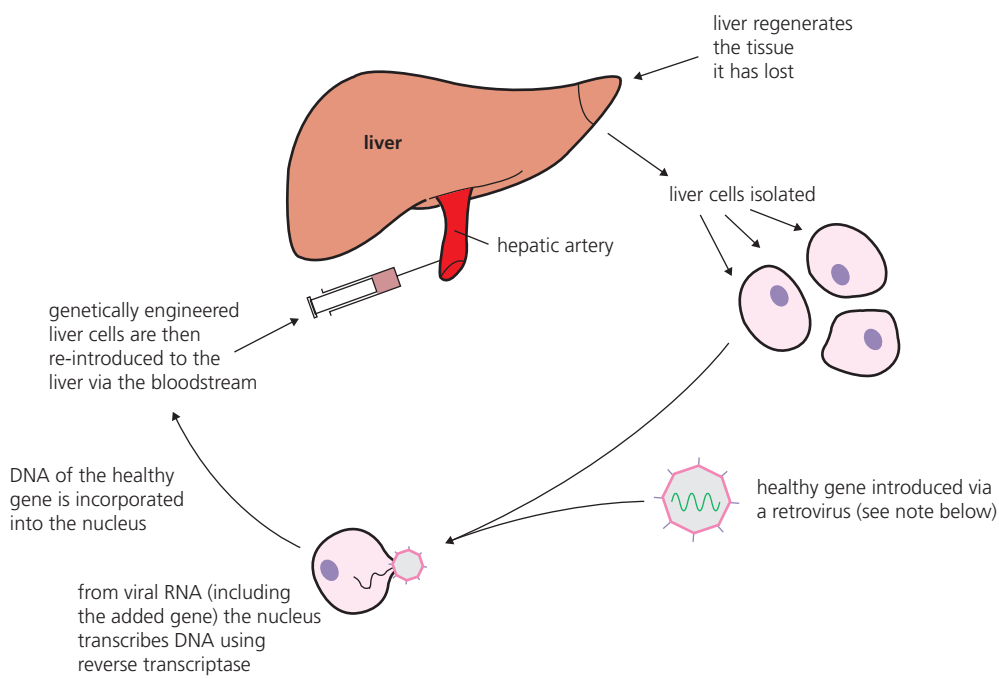
Gene therapy is the use of recombinant DNA technology to overcome genetic disease, where this is thought safe and ethically sound. Gene therapy is a very recent, and is a highly experimental science. One approach is to supply the body with the missing gene's product (and then periodically re-supply it). More difficult, but a permanent solution, would be to supply the missing gene to body cells in such a way that it remains permanently functional. This approach is illustrated here by reference to the genetic disease **familial hypercholesterolaemia (FH)**.

Lipids are transported around the body and across cell membranes as lipoproteins, such as low-density lipoproteins (LDLs). In the genetic disease FH, the membrane receptors for LDLs are defective, and LDLs accumulate in the blood, leading to atherosclerosis (page 664) and coronary heart disease.

Gene therapy for FH involves engineering a healthy human gene into a sample of liver cells so that genetically corrected cells may be re-established in the liver and restore correct functioning (Figure 5.12).

Figure 5.12 Familial hypercholesterolaemia (FH); correcting liver cell functions

Samples of liver tissue may be removed, the cells genetically modified, and then returned. The human body quickly regenerates missing liver tissue. If removed liver cells are returned to the body they may be rapidly incorporated into the liver.



Note: An *alternative vector* is used in this case because plasmids are not components of eukaryotic cells. **Retroviruses** are viruses that, on entry to the host cell, convert their RNA into DNA and add it to a host chromosome. They use a special enzyme for this, reverse transcriptase (page 573). Once the added genes are installed, they may be transcribed and translated in the usual way.

10 **Suggest** the factors that tend to maintain the levels of alleles for a genetic disease within a population, despite their unfavourable effects on quality and length of life of those affected.

The new discipline of bioinformatics

Bioinformatics is the storage, manipulation and analysis of biological information via computer science. At the centre of this development is the creation and maintenance of databases concerning nucleic acid sequences and the proteins derived from them. Already, the genomes of several prokaryotes and eukaryotes have been sequenced, as well as that of humans. This represents a deluge of data that requires organisation, storage and indexing to facilitate subsequent effective analysis involving applied mathematics, informatics, statistics and computer science.

From these database resources, potentially highly rewarding discoveries may follow, including:

- the location of specific genes in the DNA sequences of various organisms, as well as the sequences that regulate genes (remember, within a genome not all the nucleotides are genes);
- development of methods of predicting the structure and suggesting the function of newly discovered proteins and RNA sequences that the genes code for;
- establishment of evolutionary relationships between species (by comparing genes of different species); new phylogenetic trees would follow, based on molecular genetics – it would become possible to trace the evolutionary relationship of a large number of organisms by measuring changes in their DNA, and to estimate the point in geological time when related organisms diverged (page 499).

Cloning

Clones are identical things.

Genetic engineers clone genes when a single gene is introduced into a plasmid. Bacteria are then induced to take up the recombinant plasmid (page 122). Once this has occurred, the plasmid is replicated many times in the cytoplasm. In this way, very many identical copies of the gene are produced. This is an important step in recombinant DNA technology. This is one important aspect of cloning with which we are already familiar.

A clone is also defined as a group of **genetically identical individuals** or cells produced by **asexual reproduction** from a single parent cell or individual. There are numerous cases of asexual reproduction in plants, animals and protoctista. Cloning has been used widely in commercial plant propagation for many years.

Human clones occur naturally, too, in the persons of identical twins. In sexual reproduction, the fertilised egg cell divides to form a tiny hollow ball of cells called a blastocyst, as it passes down the oviduct. In the uterus, the blastocyst becomes embedded in the wall, and forms a group of cells, the inner cell mass, that will go on to form the embryo (page 385). If this cell mass (unusually) divides into two separated masses before it embeds, then both cell masses go on to form embryos, and these become genetically identical twins.

In recent years, animal clones have been produced by **nuclear transfer techniques**. Dolly the sheep was the first mammal to be cloned from non-embryonic cells. This was achieved at the Roslin Institute, Edinburgh in 1996. Dolly was produced from a fully differentiated udder cell taken from a six-year-old ewe. The isolated cell was induced to become 'dormant' or genetically quiescent, and then converted to the embryonic state by fusion with an egg cell from which the nucleus had been removed, taken from a different ewe. The process is summarised in Figure 5.13.

Human cloning?

Preliminary experiments have shown that it may be possible to clone humans, too. However, the ethical issues of such a move have stimulated opposition and comment from many people and organisations, and human cloning is currently banned in many countries.

In 2004, the UK's Human Fertilisation and Embryology Authority (HFEA) gave a Newcastle biomedical team permission to create human embryos that are clones of patients. The team is licensed to use the embryos to make embryonic stem cells to investigate diabetes, though their work could be relevant to diseases such as Alzheimer's and Parkinson's, also.

steps in the production of Dolly the clone

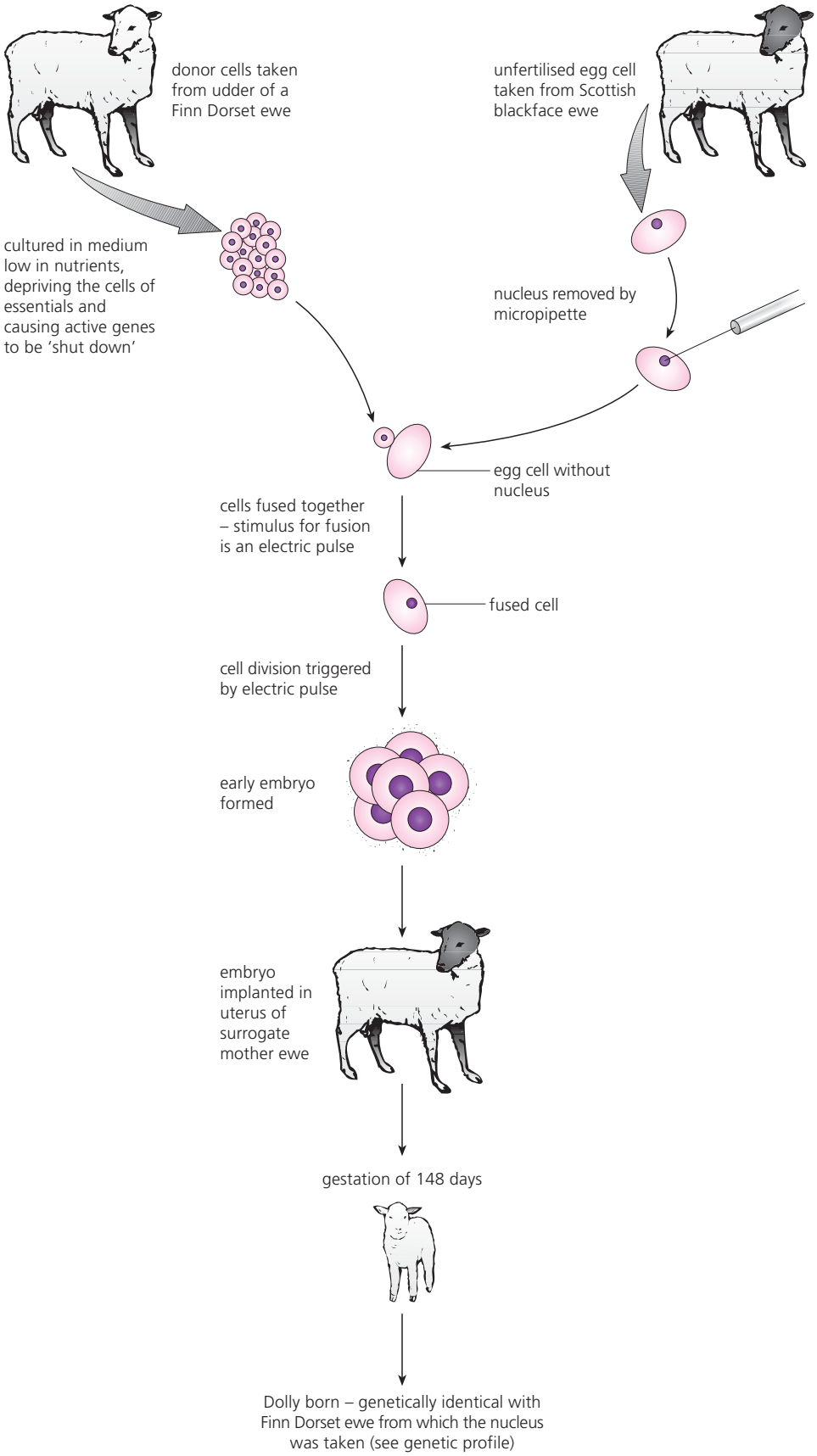
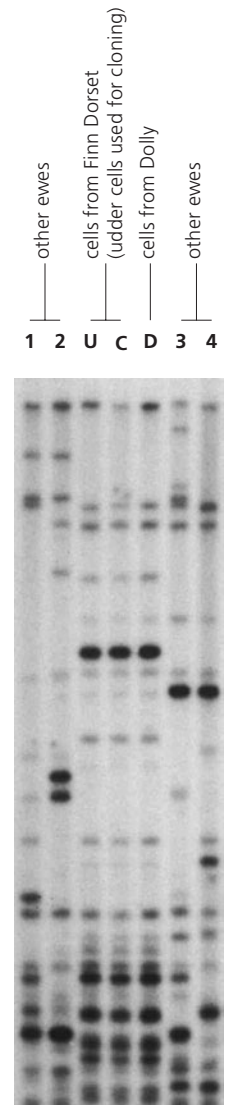


Figure 5.13 Creating Dolly the cloned sheep, using nuclear transfer techniques



genetic profile of Finn Dorset ewe cells, Dolly's cells and other ewes

Like the team who cloned Dolly the sheep, this team use nuclear transfer techniques – in this case, replacing the DNA of a human embryo cell with the DNA from a human skin cell. The embryos created in these Newcastle experiments will never be implanted in a uterus because reproductive cloning is illegal in the UK.

Instead, the team work on embryonic **stem cells** derived from the embryos. Stem cells have the ability to produce a variety of cells (page 18). It is thought these cells may have great potential in treating disease by replacing damaged tissue. Unlike an organ transplant, stem cells are not expected to be rejected by the patient's body, as the process ensures the engineered cells have the same DNA as the patient.

This is the first licence for cloning experiments in humans that the HFEA has issued. The Chair of the HFEA said:

'After careful consideration of all the scientific, ethical, legal and medical aspects of the project, the HFEA Licence Committee agreed to grant an initial one-year research licence to the Newcastle Centre for Life. This is an important area of research and a responsible use of technology. The HFEA is there to make sure any research involving human embryos is scrutinised and properly regulated.'

Since human clones can form naturally, and because experiments in human cloning are being planned, permitted and funded in several countries around the world, this controversial issue needs our careful, informed consideration.

Opponents of research cloning served an application for a judicial review on the HFEA over its decision to issue the first research licence granting permission to create cloned human embryos. At the time, UK 'pro-life' groups condemned the move and have mounted a legal challenge. This remains a controversial issue.

Arguments for and against human cloning are listed in Table 5.2.

What views do you hold?

Points against human cloning	Points in favour of human cloning
Human beings might be planned and produced with the sole intention of supplying 'spare parts' for a related human being with some passing health problem.	Parents at high risk of producing offspring with genetic disease would have the opportunity of healthy children. Infertile couples could have children of their own.
Cloning could facilitate 'improving' humans by designing and delivering a race of 'superior' people (like attempts to 'improve' by some arbitrary standards proposed in earlier, discredited, eugenics movements).	Cloning technologies are being developed to deliver organs for transplants that are entirely compatible and are not rejected by the recipient's immune system. New treatments for genetic diseases are planned and may shortly be achieved.
Cloning techniques are experimental and unreliable, resulting in the death of many embryos and newborns, since there are still so many unknown factors operating.	Cloning techniques are as safe and reliable as other comparable medical procedures, given this early stage in development and our limited experiences.
The traditional concept of 'family' is of a group of people with individuality, and with a clear sense of personal worth. Clones might have diminished rights and a lessened sense of individuality.	Clones are not true duplicates because environmental factors and personal experiences influence development and who we are. Identical twins have a strong sense of individuality and personal worth.
Some aspects of human life should exist above the values and standards of the laboratory. Many believers consider that cloning is against the will of their god. Human inventiveness must not tamper with 'nature', regarding human life issues.	All new developments take time for acceptance. Most improvements in medical technologies have received trenchant opposition which has receded with familiarity, and as the advantages are perceived.

Table 5.2 The cases for and against human cloning

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 What is a clone?

A a group of organisms which could interbreed and produce fertile offspring
B a group of cells descended from two parent cells
C a group of organisms of the same species living together and interbreeding
D a group of organisms with identical genotype

Higher Level Paper 1, May 02, Q10

Q2 **a** Outline how the process of meiosis can lead to Down's syndrome. (4)
b Discuss the advantages and disadvantages of genetic screening for chromosomal and genetic disorders. (8)
c Describe the technique for the transfer of the insulin gene using *E. coli*. (6)

Standard Level Paper 2, May 06, QB6

Q3 What is copied by the polymerase chain reaction (PCR)?

A polypeptides
B polysaccharides
C polynucleotides
D polyunsaturated fatty acids

Standard Level Paper 1, May 04, Q16

Q4 Which of the following conditions has been treated by gene therapy?

A emphysema
B SCID
C coronary heart disease
D colon cancer

Standard Level Paper 1, May 06, Q14

Questions 5–8 cover other syllabus issues in this chapter.

Q5 **a** By means of examples, explain what is meant by:
i therapeutic cloning
ii reproductive cloning. (6)
b Suggest three reasons why many people may object in principle to the cloning of human embryos as being unethical. (3)

Q6 DNA profiling uses the techniques of genetic engineering to identify an individual (typically a human) from a sample of their DNA.

a It is satellite DNA that is used. Explain what you understand by *satellite DNA*. (1)
b Restriction endonuclease is the enzyme that cuts DNA into short lengths. Identify at what specific points along the DNA molecule the restriction enzyme selected for use in DNA profiling cuts nucleic acid. (2)
c Outline the processes by which gel electrophoresis separates (sieves out) DNA fragments. (2)
d During forensic investigation, application of the polymerase chain reaction (PCR) is sometimes required prior to DNA profiling. Explain what the PCR does and in what circumstances it is required. (2)

Q7 Many human communities appear highly critical of attempts at the genetic modification of farm crops or animals. Choosing a specific example with which you are familiar, list the points you would make if you attempted to change their outlook to being supportive. (6)

Q8 **a** Outline what you understand by the term *biotechnology*. Give examples that show its relevance to human communities over time. (2)
b Suggest reasons why the application of enzymes is so highly significant in biotechnological processes. (4)

6 Ecology, evolution and biodiversity

STARTING POINTS

- **Ecology** is the study of living things in their environment.
- **Energy** reaching the green plants as light is transferred to the chemical energy of **nutrients** during **photosynthesis**.
- During **feeding**, energy is transferred from organism to organism, but is eventually returned to the environment as heat.
- **Cell respiration** is a process by which nutrients are broken down and much energy transferred to carry out the **activities and functions of cells and organisms**.
- **ATP** is the **energy currency** molecule of cells and organisms.

Ecology is the study of living things in their environment. It is an essential component of modern biology. Understanding the relationships between organisms and their environment is just as important as knowing about the structure and physiology of animals and plants.

Communities and ecosystems

5.1.1–5.1.14

One of the ideas that ecologists have introduced is that of the **ecosystem**. This is defined as a **community of organisms** and their surroundings, the **environment in which they live** and with which they interact. An ecosystem is a basic functional unit of ecology since the organisms that make up a community cannot realistically be considered apart from their physical environment.

Examples such as woodlands or a lake illustrate two important features of an ecosystem, namely that it is:

- a largely **self-contained** unit, since most organisms of the ecosystem spend their entire lives there and their essential nutrients will be endlessly recycled around and through them;
- an **interactive** system, in that the kinds of organism that live there are largely decided by the physical environment, and the physical environment is constantly altered by the organisms; the organisms (community) of an ecosystem are called the **biotic** component, and the physical environment is known as the **abiotic** component.

Ecology	The study of living things within their environment.
Ecosystem	A stable, settled unit of nature consisting of a community of organisms, interacting with each other and with their surrounding physical and chemical environment. Examples of ecosystems are ponds or lakes, woods or forests, sea shores or salt marshes, grassland, savannah or tundra. Ecosystems are necessarily very variable in size.
Habitat	The locality in which an organism occurs. It is where the organism is normally found. If the area is extremely small, we call it a microhabitat. The insects that inhabit the crevices in the bark of a tree are in their own microhabitat. Conditions in a microhabitat are likely to be very different from conditions in the surrounding habitat.
Population	All the living things of the same species in a habitat at any one time. The members of a population have the chance of interbreeding, assuming the species concerned reproduces sexually. The boundaries of populations are often hard to define, but those of aquatic organisms occurring in a small pond are clearly limited by the boundary of the pond.
Community	All the living things in a habitat or ecosystem, the total of all the populations. So, for example, the community of a well-stocked pond would include the populations of rooted, floating and submerged plants, the populations of bottom-living animals, the populations of fish and non-vertebrates of the open water, and the populations of surface-living organisms – typically a very large number of organisms.
Species	A group of individuals of common ancestry that closely resemble each other and that are normally capable of interbreeding to produce fertile offspring.

Table 6.1 An introduction to ecological terms

1 **Suggest** why the dominant plant species of a habitat may more or less determine the types of animal life present.

Environment is a term we commonly use to mean 'surroundings'. In biology, we talk about the environment of cells in an organism, or the environment of organisms in a habitat. It is also our term for the external conditions affecting the existence of organisms, so 'environment' is a rather general, unspecific term, but useful, nonetheless.

Alternatively, there are several other terms that ecologists have introduced and regularly use that have precise meaning. Table 6.1 gives definitions of these ecological terms; you can refer to it as the need arises.

Feeding relationships – producers, consumers and decomposers

Think of an ecosystem that you are very familiar with. Perhaps it is one near your home, school or college. It might be savannah, forest, a lake, woodland or meadow. Whatever you have in mind, it will certainly contain a community of plants, animals, and microorganisms, all engaged in their characteristic activities. Some of these organisms will be much easier to observe than others, possibly because of their size, or the times of day (or night) at which they feed, for example.

However, the essence of survival of organisms is their activity. To carry out these activities organisms need **energy**. We have already seen that the immediate source of energy in cells is the molecule **ATP** (page 77) produced by respiration. The energy ATP represents has been transferred from sugar and other organic molecules, the respiratory substrates. These organic molecules are obtained from nutrients as a result of the organism's mode of nutrition.

We know that green plants make their own organic nutrients from an external supply of inorganic molecules, using energy from sunlight in photosynthesis (page 82). The nutrition of the green plants is described as **autotrophic** (meaning 'self feeding') and in ecology, green plants are known as **producers**.

An autotroph is an organism that synthesises its organic molecules from simple inorganic substances.

In contrast, animals and most other types of organism use only existing nutrients, which they obtain by digestion and then absorb into their cells and tissues for use. Consequently, animal nutrition is dependent on plant nutrition either directly or indirectly. Animal nutrition is described as **heterotrophic** (meaning 'other nutrition') and in ecology, animals are known as **consumers**.

A heterotroph is an organism that obtains organic molecules from other organisms.
A consumer is an organism that ingests other organic matter that is living or recently killed.

Some of the consumers, known as herbivores, feed directly and exclusively on plants. **Herbivores are primary consumers.** Animals that feed exclusively on other animals are carnivores. **Carnivores** feeding on primary consumers are known as **secondary consumers**. Carnivores that feed on secondary consumers are called **tertiary consumers**, and so on.

Eventually all producers and consumers die and decay. Organisms that feed on dead plants and animals, and on the waste matter of animals, are described as **saprotrophs** (meaning 'putrid feeding'), and in ecology these feeders are known as **detritivores** or **decomposers**.

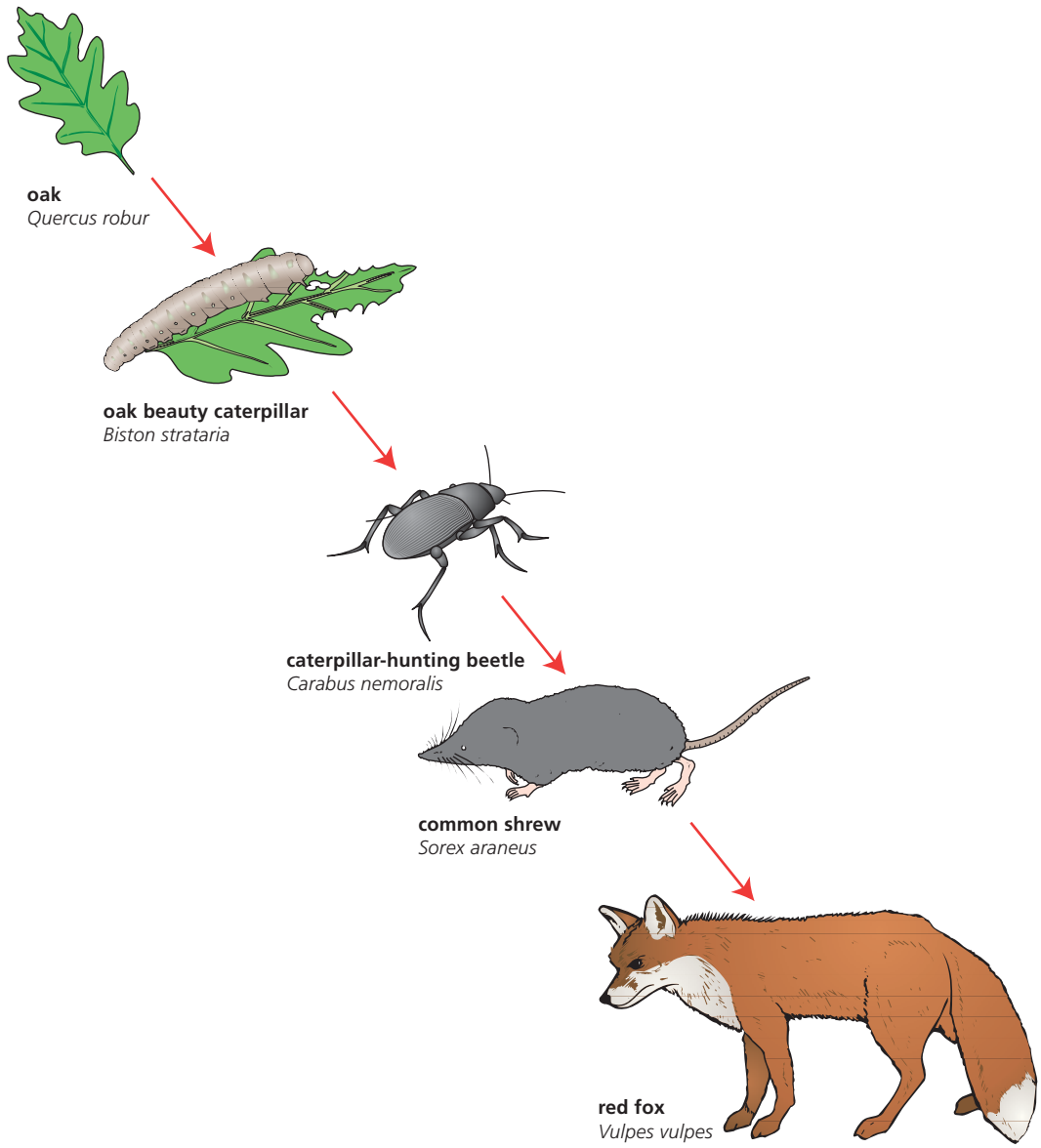
A saprotroph is an organism that lives on or in dead organic matter, secreting digestive enzymes into it and absorbing the products of digestion.
A detritivore or a decomposer is an organism that ingests dead organic matter.

Feeding by saprotrophs releases inorganic nutrients from the dead organic matter, including carbon dioxide, water, ammonia, and ions such as nitrates and phosphates. Sooner or later, these inorganic nutrients are absorbed by green plants and re-used. We will look in more detail at the cycling of nutrients in the biosphere later in this chapter.

Drawing up a food chain

A feeding relationship in which a carnivore eats a herbivore which itself has eaten plant matter is called a **food chain** (Figure 6.1). Here, light is the initial energy source, as it is with most other food chains. Note that in a food chain the arrows point to the consumers, and so indicate the direction of energy flow. Food chains from contrasting ecosystems are shown in Figure 6.2.

Figure 6.1 A food chain



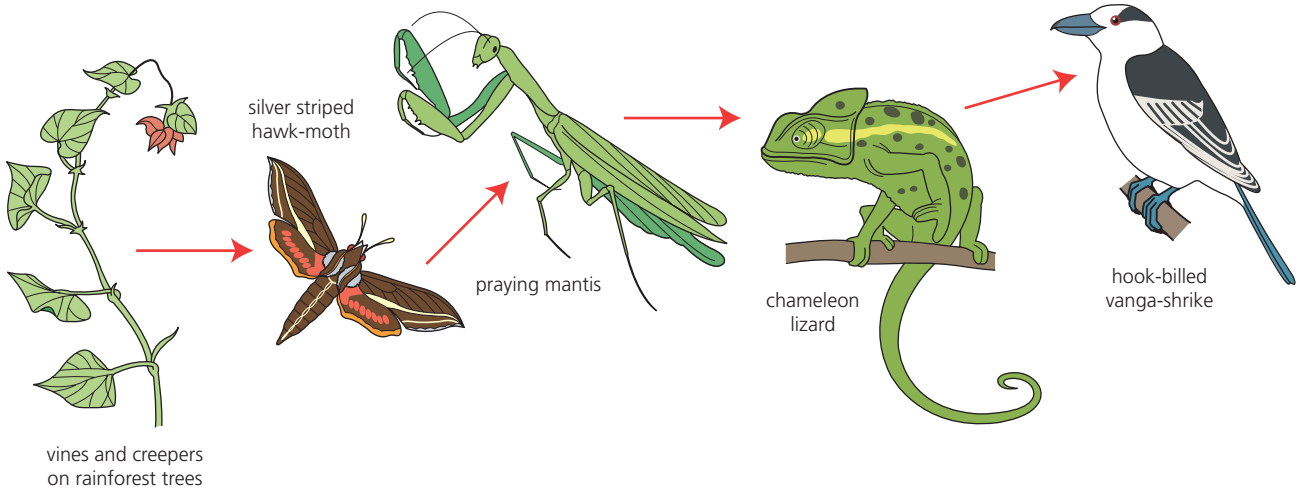
2 **Apply** one or more of the terms below to describe each of the listed features of a freshwater lake:

population **ecosystem** **habitat** **abiotic factor** **community** **biomass**

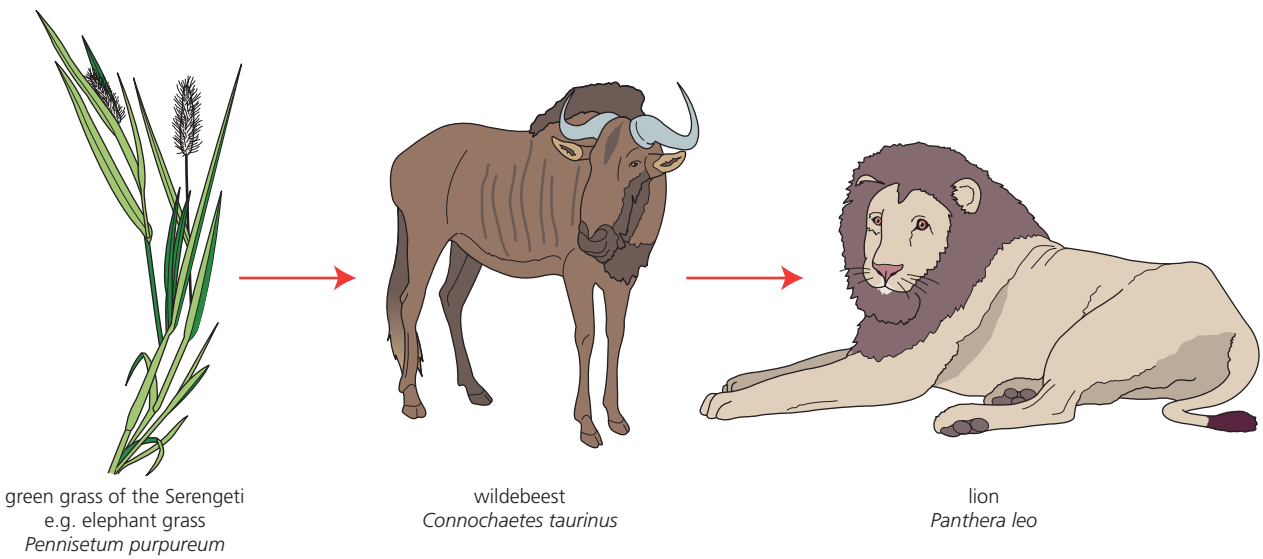
a the whole lake
b all the frogs of the lake
c the flow of water through the lake
d all the plants and animals present
e the total mass of vegetation growing in the lake
f the mud of the lake
g the temperature variations in the lake

Figure 6.2 Food chains from contrasting ecosystems

A in a tropical rainforest



B on the savannah of the Serengeti, East Africa



3 Using the information in the food web in Figure 6.3 B, **construct** two individual food chains from the marine ecosystem, each with at least three linkages (four organisms).

Identify each organism with its common name, and **state** whether each is a producer, primary consumer, secondary consumer, etc.

Food webs

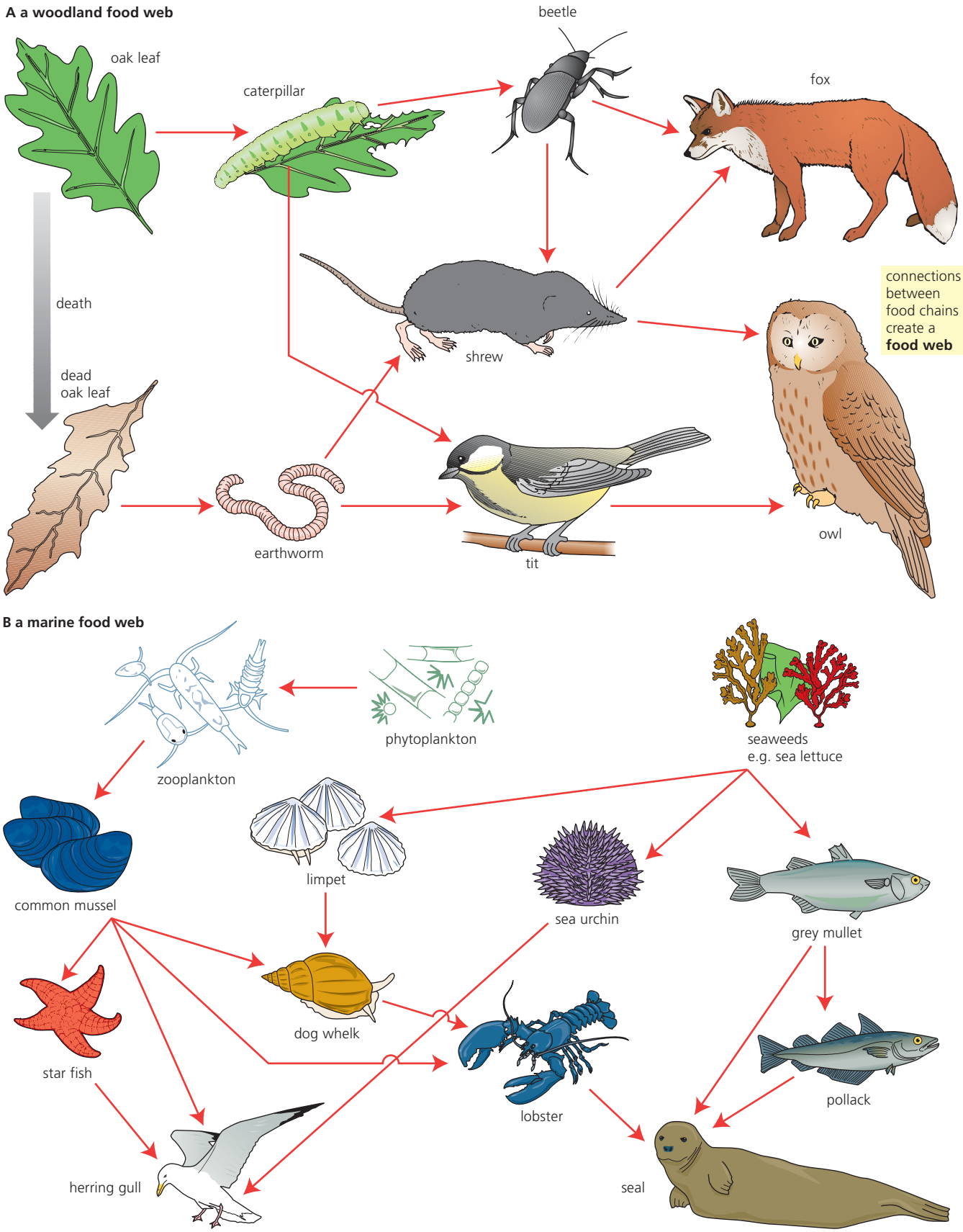
In an ecosystem the food chains are not isolated. Rather, they interconnect with other chains. This is because most prey species have more than one predator. Predators, as well as having preferences, need to exploit alternative food sources when any one source becomes scarce. They also take full advantage of gluts of food as particular prey populations become temporarily abundant.

Consequently, individual food chains may be temporary, and are interconnected so that they may form a **food web**. Two examples of food webs are shown in Figure 6.3, one from a woodland ecosystem and one from a marine ecosystem.

Examine these now.

Note that food chains tell us about the feeding relationships of organisms in an ecosystem, but they are entirely **qualitative** relationships (we know which organisms are present as prey and as predators) rather than providing **quantitative** data (we do not know the numbers of organisms at each level).

Figure 6.3 Examples of food webs



Trophic levels

The level at which an organism feeds in a food chain is called its **trophic level** (feeding level). In this way of classifying feeding relationships, the **producers** are designated as **trophic level 1**, because energy has been transferred once, from Sun to plant. All the herbivores make up **trophic level 2**, because here energy has been transferred twice, and so on. The trophic levels of organisms in some of the above food chains and food webs are classified in Table 6.2.

Trophic level	Woodland	Rainforest	Savannah
Producer Level 1	oak	vines and creepers on rainforest trees	grass
Primary consumer Level 2	caterpillar	silver striped hawk-moth	wildebeest
Secondary consumer Level 3	beetle	praying mantis	lion
Tertiary consumer Level 4	shrew	chameleon lizard	
Quaternary consumer Level 5	fox	hook-billed vanga-shrike	

Table 6.2 An analysis of trophic levels

4 **Identify** the initial energy source for almost all communities.

Note that there is not a fixed number of trophic levels to a food chain, but typically they have three, four or five levels only. There is an important reason why stable food chains remain quite short. We will come back to this point, shortly.

Sometimes it can be difficult to decide at which trophic level to place an organism. For example, an **omnivore** feeds on both plant matter (level 2 – primary consumer) and on other consumers (level 3 – secondary consumer, or higher). In the woodland food chain of Figure 6.1, the fox more commonly feeds on beetles than shrews because there are many more beetles about, and they are easier to catch.

5 **Suggest** what trophic levels humans occupy. Give examples.

Energy flow

Between each trophic level there is an energy transfer. At the base of the food chain, green plants, the producers, transfer light energy into the chemical energy of sugars in photosynthesis. Much of the light energy is not retained in the green leaf. Some is reflected, some transmitted, and some lost as heat energy. Heat is also a waste product of the reactions of respiration and of the plant's metabolism, as sugar is converted into lipids and amino acids, for example. Some of these metabolites are used in the growth and development of the plant, and by these reactions, energy is locked up in the organic molecules of the plant body.

Inevitably, energy is transferred every time the tissues of a green plant are consumed by herbivores.

Finally, on the death of the plant, the remaining energy passes to detritivores and saprotrophs when dead plant matter is broken down and decayed. The diverse routes of energy flow through a primary producer are summarised in Figure 6.4.

Consumers eat producers or primary or secondary consumers, according to where they occur in a food chain. In this way, energy is transferred to the consumer. The consumer in turn, transfers energy in muscular movements by which it hunts and feeds, and as it seeks to escape from predators. Some of the food eaten remains undigested, passing through the consumer unchanged, and is lost in the faeces. Also, heat energy, a waste product of the reactions of respiration and of the animal's metabolism, is continuously lost as the consumer grows and develops and forms body tissues. If the consumer is itself caught and consumed by another, larger consumer, energy is again transferred. Finally, on the death of the consumer, the remaining energy passes to detritivores and saprotrophs when dead matter is broken down and decayed.

The flow of energy through a consumer is summarised in Figure 6.5.

Figure 6.4 Energy flow through a producer

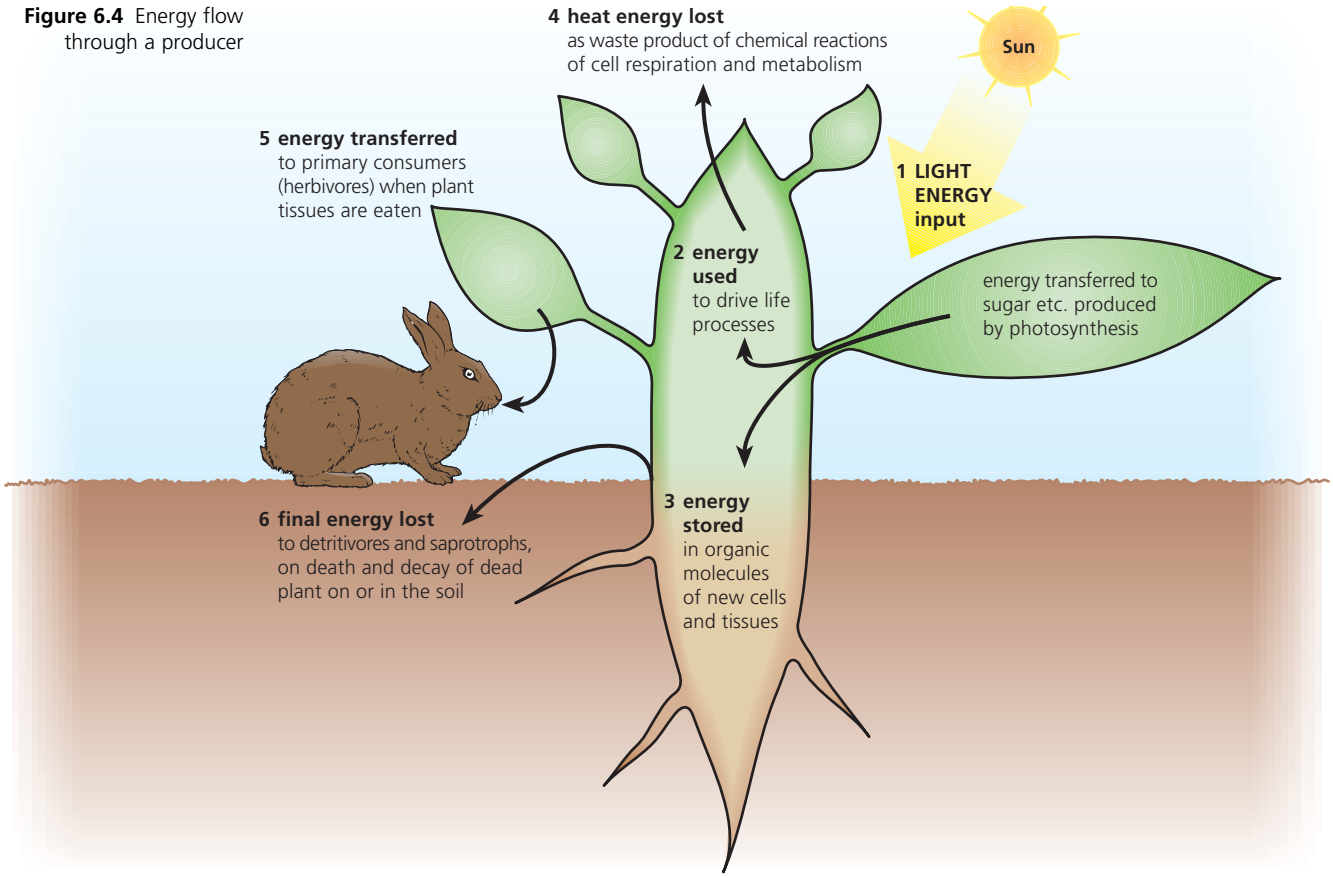
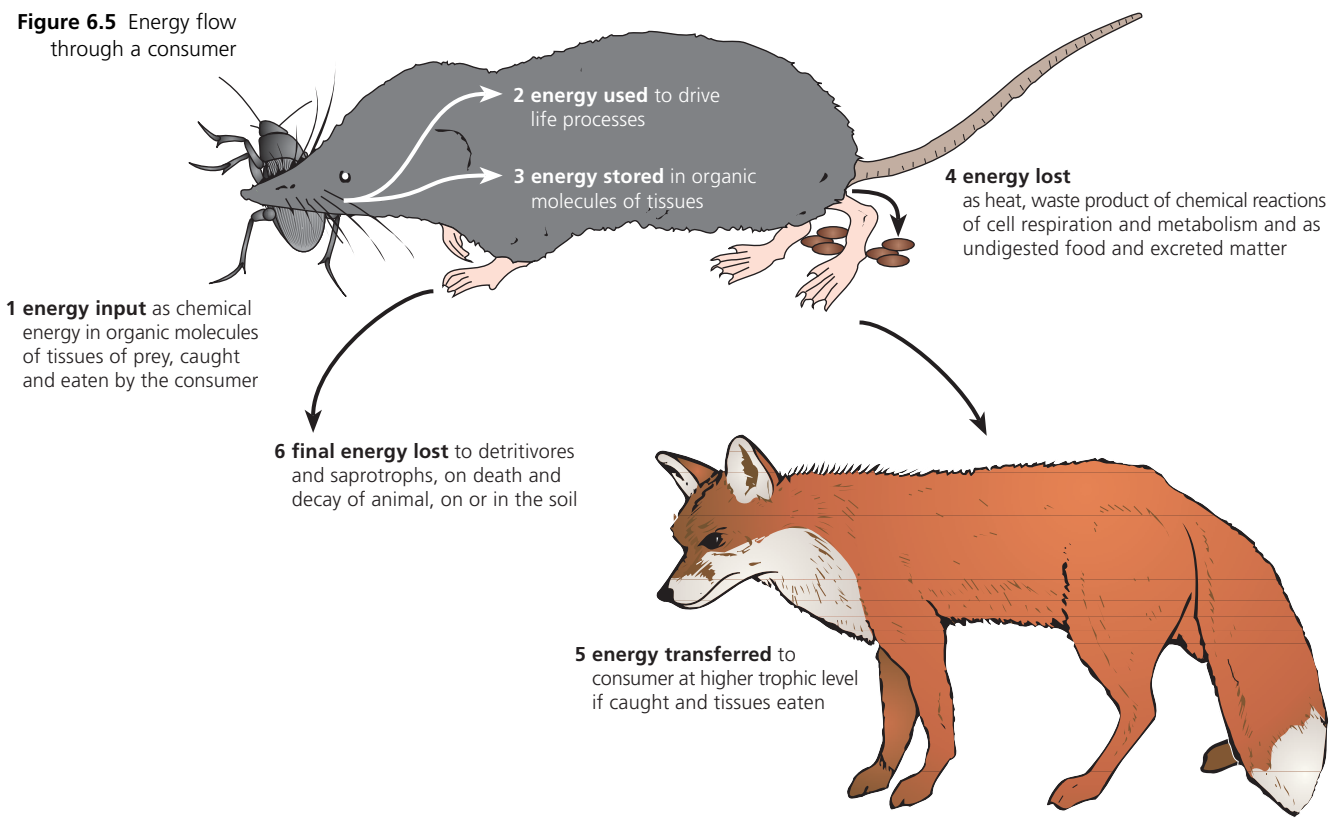


Figure 6.5 Energy flow through a consumer



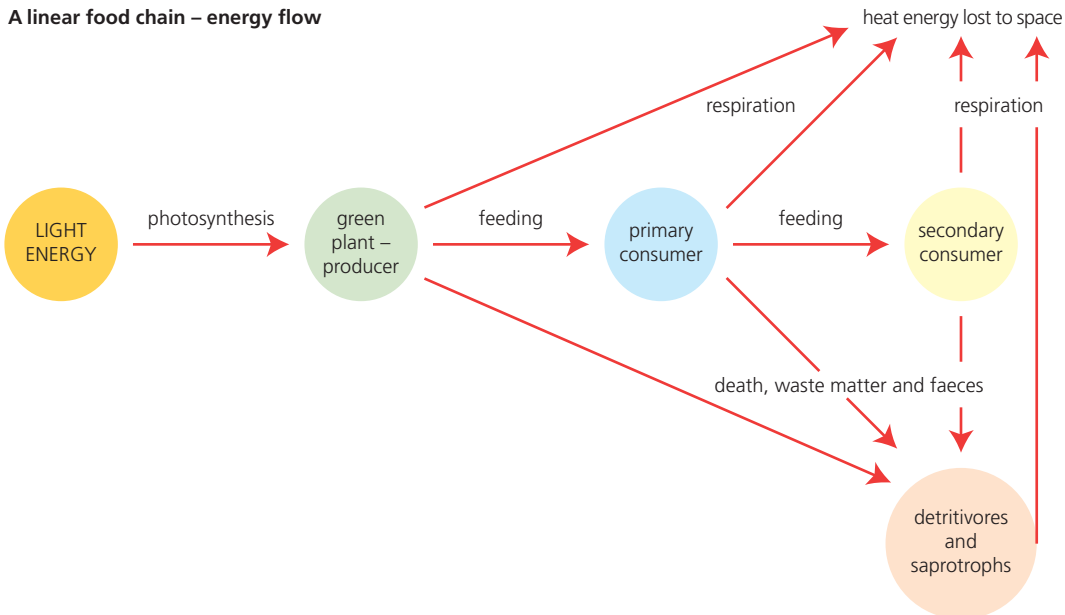
So, energy is transferred from one organism to another in a food chain, but only some of the energy transferred becomes available to the next organism in the food chain. In fact, only about 10% of what is eaten by a consumer is built into that organism's body, and so is potentially available to be transferred on in predation or browsing. There are two consequences of this.

- The energy loss at transfer between trophic levels is the reason why food chains are short. Few transfers can be sustained when so little of what is eaten by one consumer is potentially available to the next step in the food chain. Consequently, it is very uncommon for food chains to have more than four or five links between producer (green plant) and top carnivore, at most.
- Feeding relationships of a food chain may be structured like a pyramid. At the start of the chain is a very large amount of living matter (biomass) of green plants. This supports a smaller biomass of primary consumers, which in turn supports an even smaller biomass of secondary consumers. A generalised ecosystem pyramid diagram, representing the structure of an ecosystem in terms of the biomass of the organisms at each trophic level, is shown in Figure 6.6.

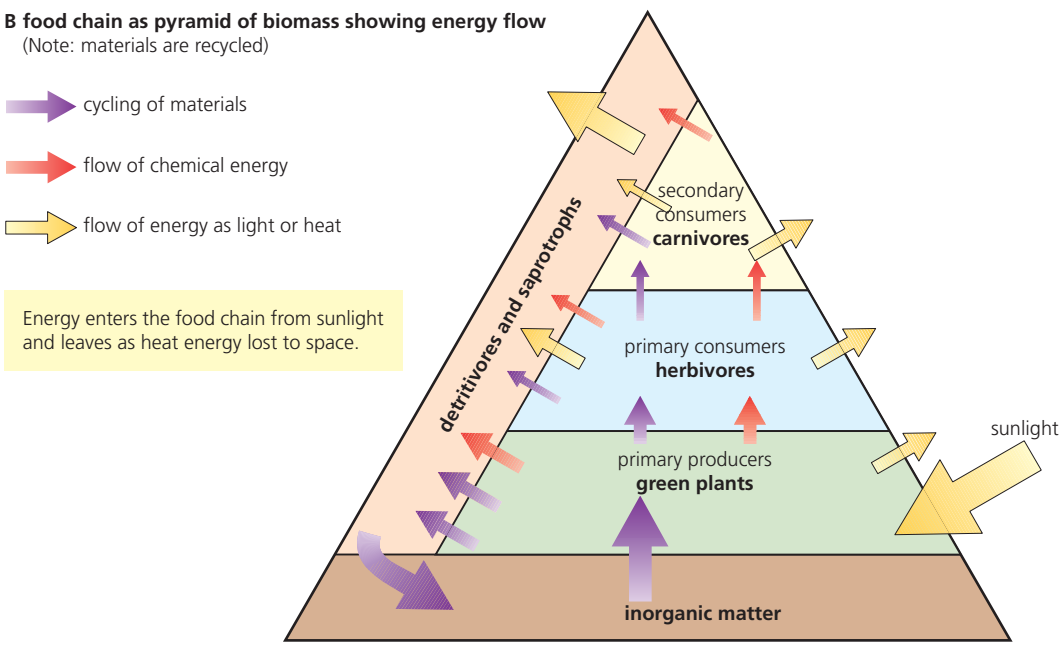
6 **Explain** why, in a food chain, a large amount of plant material supports a smaller mass of herbivores and an even smaller mass of carnivores.

Figure 6.6 Energy flow through a food chain – the pyramid of biomass

A linear food chain – energy flow



B food chain as pyramid of biomass showing energy flow
(Note: materials are recycled)



Pyramids of numbers, biomass and energy

Ecological pyramids are a practical means of analysing ecosystems. For example, they have been used to express seasonal changes in one ecosystem, or to compare different ecosystems at comparable times.

Initially, **pyramids of numbers** were produced. All the organisms at each trophic level within a given area were recognised and counted. The data were relatively easily obtained without destroying the organisms (given that the experimenter recognised all the organisms). The results were presented as rectangular blocks, stacked on top of each other, representing the numbers of organisms at each trophic level in a bar diagram.

In this approach, no allowance was made for differences in size of individual organisms; a giant oak tree and a single microscopic green alga counted the same. Consequently, the pyramids of numbers sometimes created strange shapes of limited meaning (Figure 6.7).

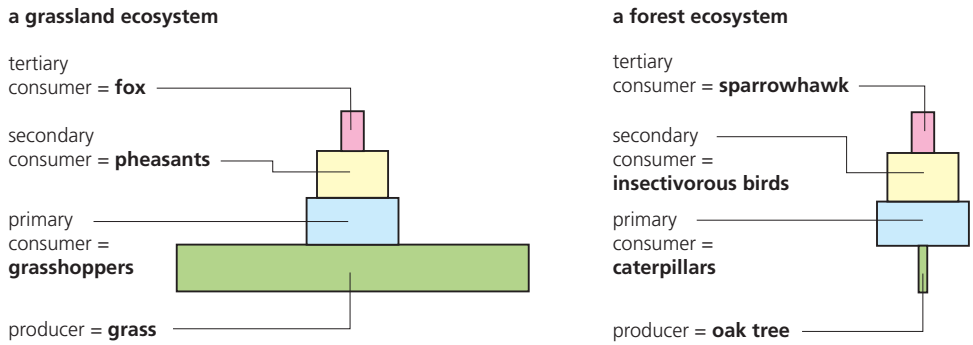


Figure 6.7 Examples of ecological pyramids of numbers

This problem of erratically shaped 'pyramids' was overcome by producing **pyramids of biomass**. To do this, the fieldwork involved estimating the numbers of organisms of each type at each trophic level, and then finding the biomass (dry weight) of a representative sample of each type of organism.

However, problems emerged with pyramids of biomass, too. A pyramid of biomass represents matter in different organisms at the moment the analysis was made, but the data do not indicate how fast matter is being added or consumed. Consequently, a pyramid based on the inflow of energy to each trophic level is preferred. This is called a **pyramid of energy**.

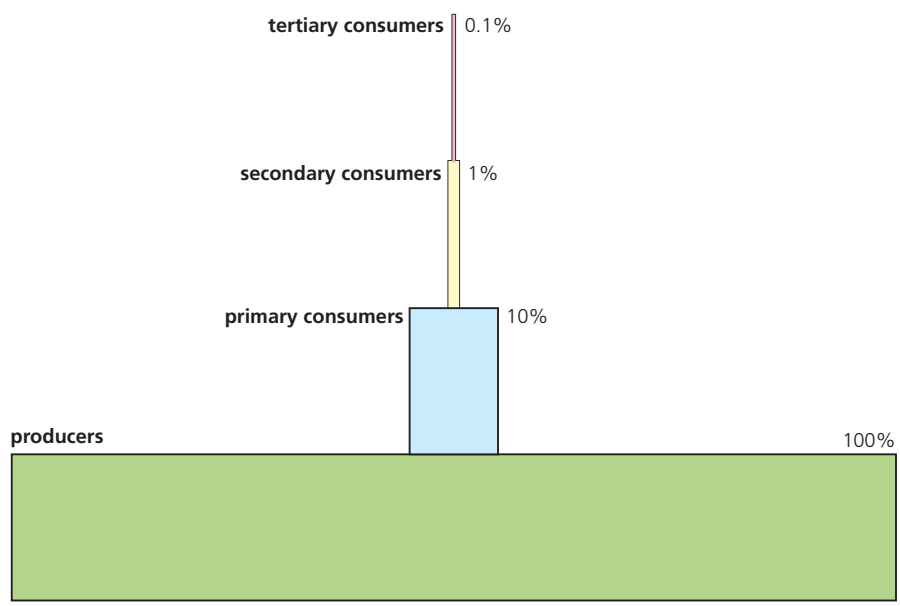
There are few examples of pyramids of energy because obtaining the data is difficult and time-consuming, and is a destructive activity. In practice, pyramids of biomass can be converted to pyramids of energy using published values of energy content, made in previous experiments in habitats that are comparable. The amounts of energy (kJ) measured are expressed per square metre of area occupied by the community per year (kJ m^{-2} $year^{-1}$).

Figure 6.8 A generalised pyramid of energy

Only energy taken in at one trophic level and then built in as chemical energy in the molecules making up the cells and tissues is available to the next trophic level. This is about 10% of the energy.

The reasons are:

- Much energy is used for cell respiration to provide energy for growth, movement, feeding, and all other essential life processes.
- Not all food eaten can be digested. Some passes out with the faeces. Indigestible matter includes bones, hair, feathers, and lignified fibres in plants.
- Not all organisms at each trophic level are eaten. Some escape predation.



A generalised pyramid of energy is shown in Figure 6.8. It is annotated to summarise the reasons why the shape is always pyramidal. Figure 6.9 shows a researched example – a study made over 50 years ago by an American ecologist working on a river system in Florida, USA. It is interesting to note that it is a quite sweeping generalisation to say that only about 10% of energy is available for the next trophic level. For example, in the Florida river analysis it was found that 17% of energy of the producers (green aquatic plants) was transferred to primary consumers. The reason in this case was that the plants concerned were almost entirely of highly digestible matter. They lacked any woody tissues common to most terrestrial plant matter.

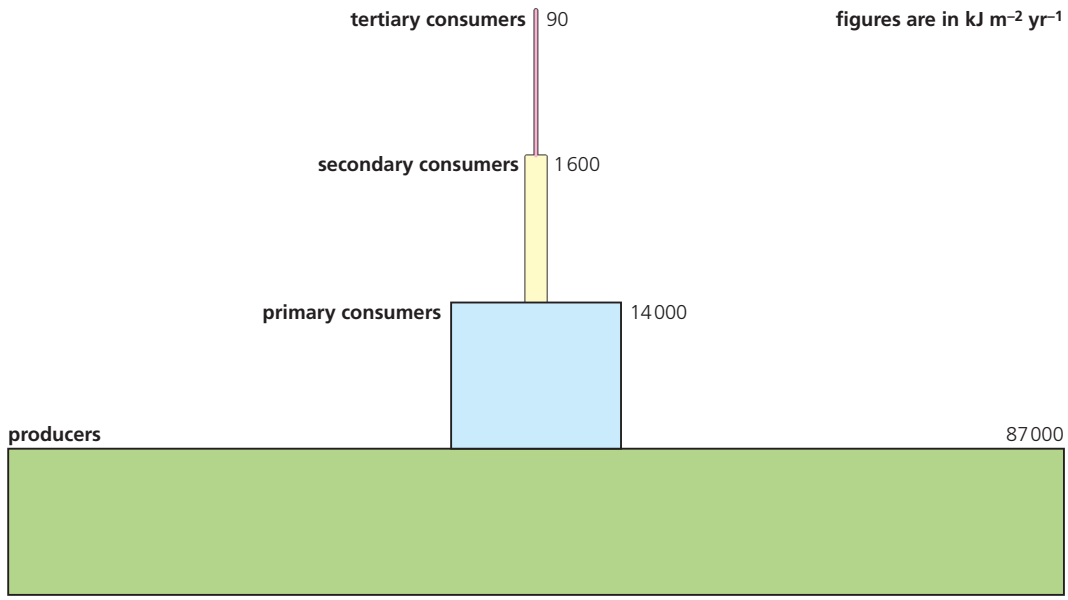


Figure 6.9 Pyramid of energy for a river ecosystem at Silver Springs, Florida

Cycling of nutrients

We have seen that energy enters ecosystems as light energy, is transferred to chemical energy in the nutrient molecules that sustain the growth of the tissues of producers and consumers, but ultimately is lost to space as heat. Energy continuously flows through ecosystems in this way. On the other hand, the **nutrients** of the ecosystem are not lost – they are **recycled and re-used**.

Nutrients provide the chemical elements that make up the biochemical molecules of cells and organisms. We have seen that all organisms are made of carbon, hydrogen and oxygen, together with mineral elements nitrogen, calcium, phosphorus, sulphur and potassium, and several others, in increasingly small amounts (Table 2.2, page 38). Plants obtain their essential nutrients as carbon dioxide and water, from which they manufacture sugar. With the addition of mineral elements absorbed as ions from the soil solution, they build up the complex organic molecules they require (Figure 3.7, page 82). Animals, on the other hand, obtain nutrients as complex organic molecules of food which they digest, absorb and assimilate into their own cells and tissues.

Recycling of nutrients is essential for the survival of living things, because the available resources of many elements are limited. When organisms die, their bodies are broken down and decomposed, mainly by **bacteria and fungi**, and the nutrients are released. Elements may become part of the soil solution, and some may react with chemicals of soil or rock particles, before becoming part of living things again by being reabsorbed by plants.

The cycling processes by which essential elements are released and re-used are called **biogeochemical cycles** because the cycle of changes involves both living things (the biota) and the non-living (abiotic) environment, consisting of atmosphere, hydrosphere (oceans, rivers and lakes) and the lithosphere (rocks and soil). All the essential elements take part in biogeochemical cycles – one example is the carbon cycle, which we will discuss shortly.

Roles of detritivores and saprotrophs in recycling

When organisms die, their bodies are broken down to simpler substances (for example, CO_2, H_2O, NH_3 and various ions) by a succession of organisms (mostly microorganisms), as illustrated in Figure 6.10. Scavenging actions of detritivores often begin the process, but saprotrophic bacteria and fungi always complete the breakdown processes.

Ultimately, both plants and animals depend on the activities of saprotrophic microorganisms to release matter from dead organisms for re-use. There is a limited supply of many of the chemical elements required as nutrients, so recycling is essential for life. Microorganisms are the main contributors in the vital recycling that goes on in the environment, largely unseen.

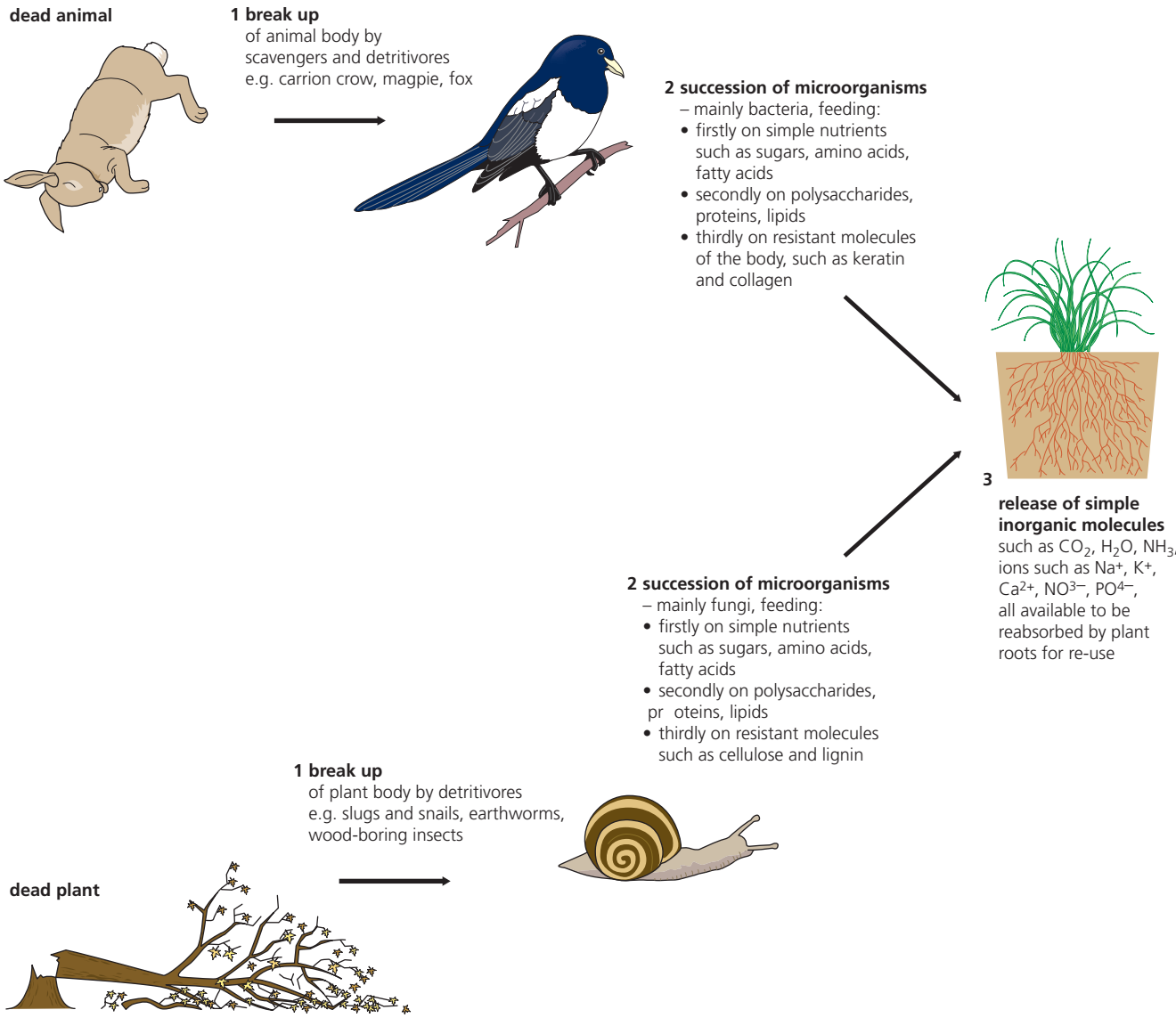


Figure 6.10 The sequence of organisms involved in decay

7 **Explain** how it is that animal life is dependent on the actions of saprotrophs.

Greenhouse effect

5.2.1–5.2.6

Carbon dioxide is present in the atmosphere at about 0.038% by volume (which represents 0.057% by mass). This amount of atmospheric carbon dioxide is maintained by a balance between the fixation of this gas during photosynthesis and release of carbon dioxide into the atmosphere by respiration, combustion, and decay by microorganisms – an interrelationship illustrated in the carbon cycle (Figure 6.11).

Photosynthesis withdraws about as much carbon dioxide during the daylight each day as is released into the air by all the other processes, day and night – or nearly so.

In fact, the level of atmospheric carbon dioxide is now rising for reasons we will examine shortly. First, we need to look at the effect of low levels of atmospheric carbon dioxide because this natural component of the atmosphere maintains **a favourable environmental temperature** on Earth – a phenomenon known as the **greenhouse effect**.

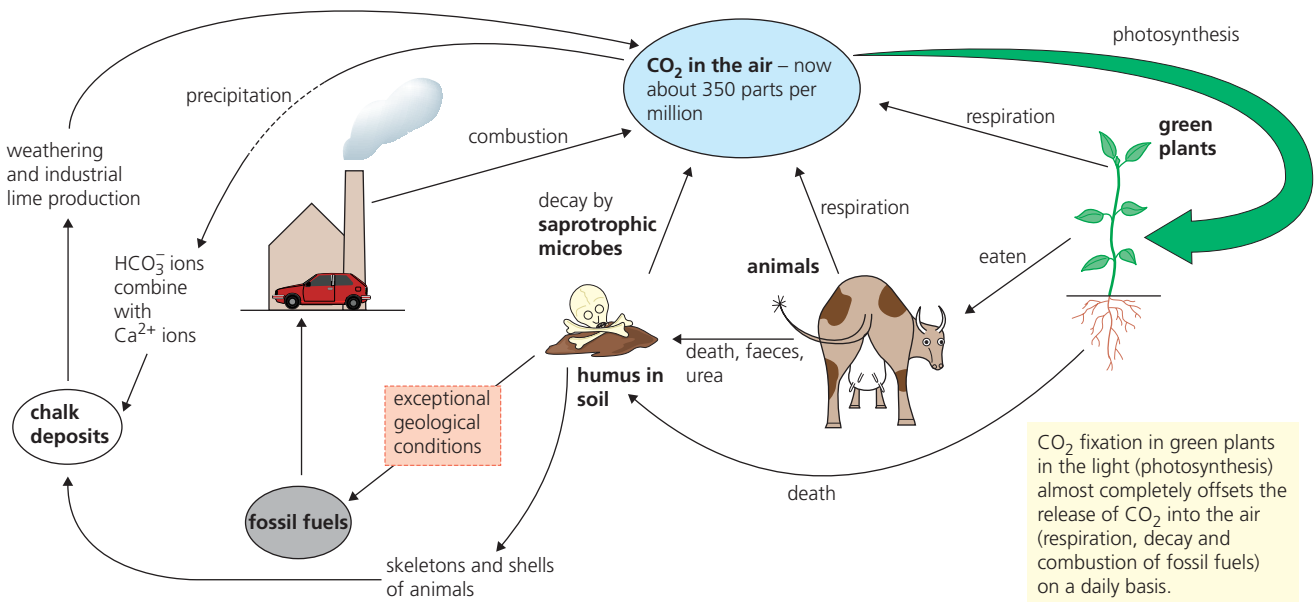


Figure 6.11 The carbon cycle

8 In the cycling of carbon in nature, **state** in what forms inorganic carbon can exist in:

a the atmosphere
b the hydrosphere
c the lithosphere.

The mechanism of the greenhouse effect

The radiant energy reaching the Earth from the Sun includes visible light (short wave radiation), and infra-red radiation (longer wave radiation – heat), which warms up the sea and the land. As it is warmed, the Earth radiates infra-red radiation back towards space. However, much of this heat does not escape from our atmosphere. Some is reflected back by clouds and much is absorbed by gases in the atmosphere which are warmed. In this respect, the atmosphere is working like the glass in a greenhouse, which is why the effect is called the greenhouse effect (Figure 6.12). We must stress that the greenhouse effect is very important to life on the Earth: without it surface temperatures would be too cold for life.

Gases in the atmosphere that absorb infra-red radiation are referred to as **greenhouse gases**. **Carbon dioxide** is not the only component of our atmosphere with this effect – both **water vapour** and **methane** are naturally occurring greenhouse gases. In addition, atmospheric pollutants such as the oxides of nitrogen (particularly **nitrous oxide**) and the **chlorofluorocarbons (CFCs)** also have greenhouse properties. Oxides of nitrogen are waste products of the combustion of fossil fuels (natural gas, oil and coal), and a particularly abundant source is the exhaust fumes of internal combustion engines. CFCs, on the other hand, are unreactive molecules that were deliberately manufactured by the chemical industry to use as propellants in aerosol cans and as the coolant in refrigerators. With the passage of time these gases have escaped into the atmosphere, and have slowly been carried up to the stratosphere.

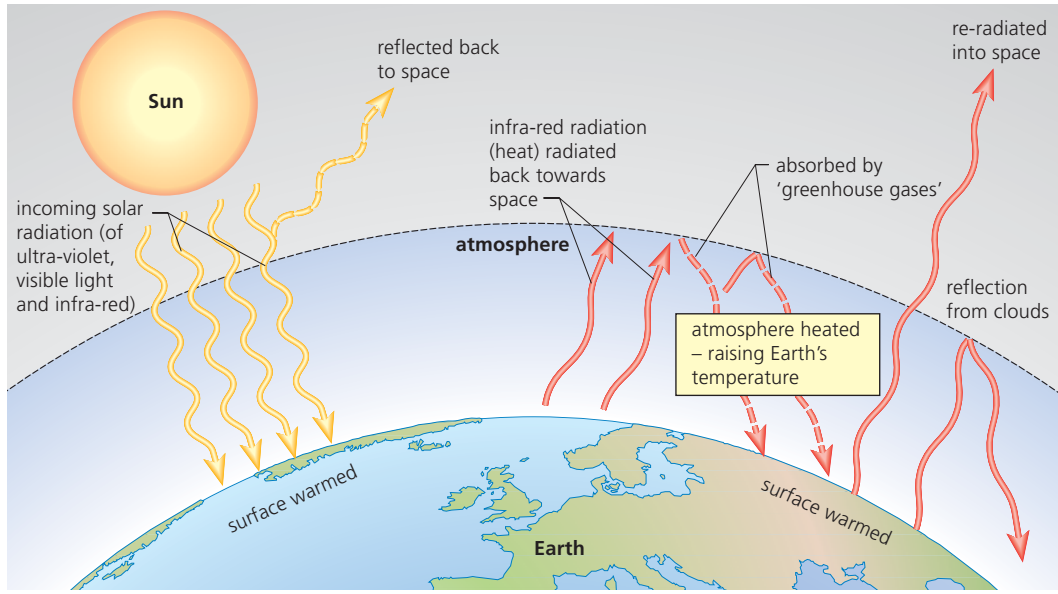


Figure 6.12 The greenhouse effect

An enhanced greenhouse effect leading to global warming?

The composition of the atmosphere has changed over time. That is beyond dispute.

How are current and historic levels of atmospheric carbon dioxide known?

The best **long-term records** of changing levels of greenhouse gases (and associated climate change) are based on evidence obtained from ice cores drilled in the Antarctic and Greenland ice sheets. For example, data from the Vostok ice core in East Antarctic and evidence from the composition of the bubbles of gas obtained from these cores, show us how methane and carbon dioxide levels have varied over no less than 400 000 years. Similarly, variations in the concentration of oxygen isotopes from the same source indicate how temperature has changed during the same period (Figure 6.13).

Clearly, the levels of greenhouse gas in the atmosphere can be closely correlated with global temperature. Environmental conditions on Earth have changed as a consequence, and we may anticipate further changes unless the current, rising levels of atmospheric carbon dioxide are reversed by internationally agreed action. We need to examine this issue next.

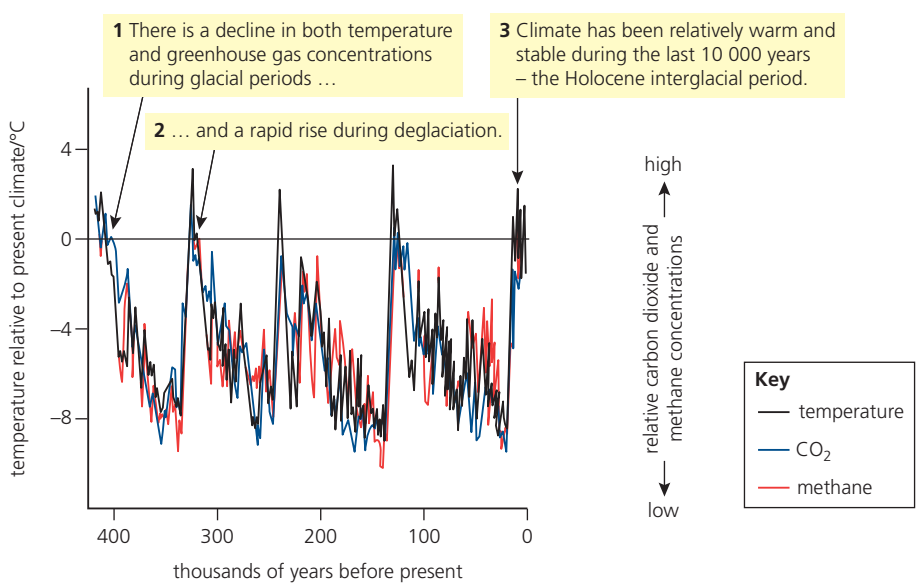


Figure 6.13 Three types of data recovered from the Vostok ice cores over 400 000 years of Earth history

Changes in atmospheric carbon dioxide levels since 1957

Appropriate measuring devises were established at the Mauna Low monitoring station on Hawaii, beginning in 1957 as part of the International Geophysical Year initiative, to monitor the global environment. The most recent data furnished by this on-going investigation are shown in Figure 6.14. Seasonal variations cause fluctuations in atmospheric carbon dioxide, but the trend is unmistakeable. Note that there are differences in the scales of the graphs in Figures 6.13 and 6.14, so the slopes of the curves appear profoundly different.

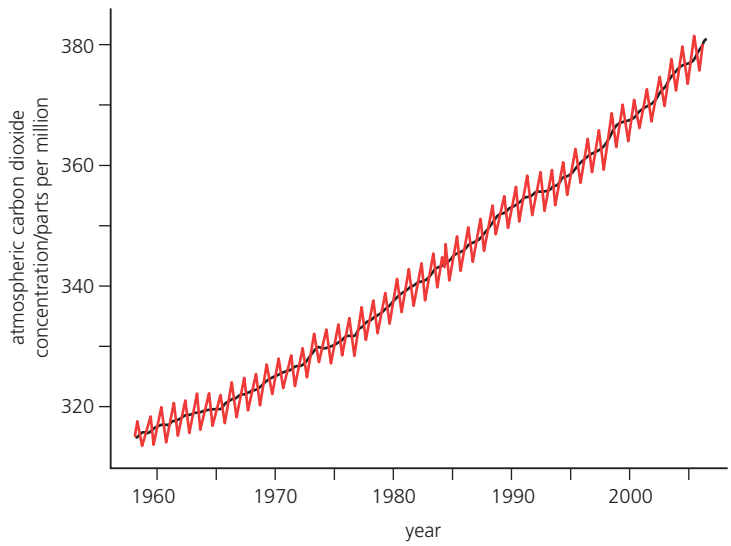


Figure 6.14 Atmospheric carbon dioxide levels recorded at Mauna Low, Hawaii, 1957–2005

9 From the data given in Figure 6.14:

a **calculate** the % change in mean atmospheric carbon dioxide between 1960 and 2000
b **explain** why it is that the atmospheric carbon dioxide varies between high and low values within each 12-month period of the graph.

The causes of raised levels of atmospheric carbon dioxide

From the graphs in Figures 6.13 and 6.14, we can see that while the levels of carbon dioxide have risen at times during the past 20 000 years, it is since the beginnings of the industrial revolutions in the developed countries of the world (the past 200 years or so) that levels have been consistently high and steadily rising.

We may assume that when the carbon dioxide levels were raised in earlier Earth history, processes like volcanic eruptions and the weathering of chalk and limestone rocks added additional carbon dioxide. Today, the release of carbon dioxide from volcanoes is estimated to contribute only about 1% of the amounts released by human activities.

The most recent sharp rise is attributed to the burning of fossil fuels (natural gas, coal and oil). Fossil fuels were mostly laid down in the Carboniferous Period, so we are now adding to our atmosphere carbon dioxide containing carbon that has been locked away for about 350 million years. This is an entirely new development in geological history. This enhanced global warming is beginning to generate problems that pose a major environmental threat to life as we know it.

Global warming and the Precautionary Principle

The Precautionary Principle is summed up by the expression 'Better safe than sorry'! Given that this idea is generally accepted once it is put to people in a particular context, it is surprising that the principle has not been clearly defined. However, no official definition has been shown to be other than vague, on careful consideration. The most common definition used is:

> **When an activity raises threats of harm, measures should be taken, even if a cause-and-effect relationship has not been established scientifically.**

To evaluate the relevance of this principle as a justification for strong action to combat the threats posed by enhanced global warming, we need to identify:

- the most likely consequences that may result from failure to slow down the rising level of atmospheric carbon dioxide and then reduce it to earlier values;
- the actions needed to deliver measures that enable us to achieve this.

In Table 6.3 these aspects are identified.

Likely consequences of unchecked enhanced global warming	Effective strong actions that may combat these threats
■ complete melt-down of the polar ice caps and the glaciers of mountain regions, releasing vast amounts of water into the seas, so raising sea levels	■ conserve fossil fuel stocks, using them only sparingly, and only where there are no apparent alternatives (such as oils from biofuel sources)
■ warming of sea waters, causing them to expand in volume, so further raising sea levels	■ develop nuclear power sources to supply electricity for industrial, commercial and domestic needs
■ permanent flooding of much lowland territory where currently vast numbers of the human population live and where the most agriculturally productive land occurs	■ develop so-called renewable sources of power, exploiting environmental energy sources, such as wave energy and wind power
■ raising of the sea water temperature to levels where surface waters (where algae may flourish) is no longer ion-enriched by natural mixing with deep, colder water, causing failure of algal photosynthesis (a major CO_2-sump) so further raising atmospheric CO_2 levels	■ develop biofuel sources of energy that exploit organic waste matter that will naturally decay anyway, and biofuel crops that are renewable sources of energy using current photosynthesis products
■ destruction of forests by global temperatures that, in many parts of the world, are too high to support the survival and natural growth of trees	■ reduce use of fuels for heating of homes (where necessary) to minimum levels by designing well-insulated houses
■ interruption of ocean current systems that distribute warm waters from the tropical regions of the globe northwards, thereby disrupting agricultural production levels that sustain the human food chain for many peoples	■ reduce use of fuels for transport systems
	■ terminate the destruction of forests in general and, in particular, of rainforest all around the tropical regions of the Earth, since these are major CO_2-sumps

Table 6.3 Evaluating the threats posed by global warming and appropriate strong actions to rectify or avoid the resulting environmental degradation

To be effective, strong actions taken in response to these challenges need to:

- be agreed internationally as acceptable to all nations and be acted on by everyone simultaneously;
- recognise that existing developed countries have previously experienced their industrial revolutions (which largely initiated these processes of environmental damage) and need to fully share the benefits with less-developed countries that would otherwise be required to forego some of the benefits of development.

Figure 6.15 summarises the problem.

In less-developed countries
natural resources such as rainforest may be the only product people can market to earn foreign currency to purchase goods and services, often at developed-world prices – and there is huge demand for their resources from developed and developing countries.

In developed countries
the destruction of natural resources like rainforest elsewhere in the world is seen as a threat to stable environmental conditions essential for an established, comfortable way of life. This way of life is often based on earlier exploitation of similar resources.

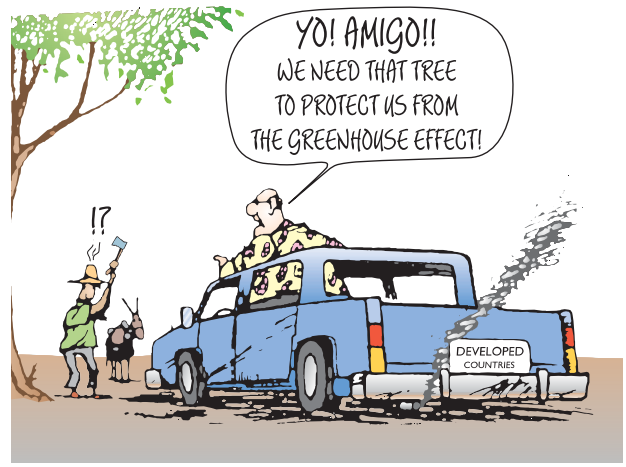


Figure 6.15 The environmental conundrum for today's world!

These contrasting needs cannot be met without huge sacrifices on both sides – yet it is easier for each country to recognise only what other countries need to do!

Is effective international agreement possible?

The need is for drastic cuts in global emissions of carbon dioxide and other greenhouse gases. An international agreement to limit release of greenhouse gases by all industrial countries and emerging industrial countries was first agreed at the **Earth Summit** in Rio de Janeiro in 1992. The initiative, known as **Agenda 21**, was launched. Subsequently, at the **Kyoto Conference** in Japan a first attempt was made to meet the pledges made at Rio. Carbon dioxide emission targets for the industrial nations were set for the period 2009–2012.

However, nations were allowed to offset emissions with devices such as carbon sinks – mechanisms by which atmospheric carbon dioxide is removed from the air either permanently, or on a long-term basis. For example, carbon dioxide may be absorbed by additional forest trees and become the carbon of wood that is not harvested and burnt. However, many of the arrangements are so complex as to be difficult to enforce. Real progress is extremely slow.

The Arctic and the consequences of global temperature rise – a case study

Arctic ecosystems consist of areas of permanent **ice desert** and huge areas of low-lying treeless land with an environment and vegetation regime described as **tundra** (Figure 6.16). The few plants present all through the year are sedges, lichens and sometimes dwarf willows – often described as low cushion plants. Trees do not grow because the growing season is too short for them, the winters too cold and dry, and what soil exists there is unable to provide support for the growth of significant root systems. Most of the year the temperatures are well below freezing point, so the moisture, soil and organic remains present are frozen hard as rock (permafrost). Dead organic matter accumulates, because the periods of the year when bacteria and fungi might decay this to humus in the usual way are far too short. Thus, the bulk of nutrients are locked away and few are released in growing periods. Nevertheless, during the brief growing season the surface of the substratum melts and plant growth is briefly spectacular, including flowering and seed dispersal, accompanied by a tremendous number of insects.

Figure 6.16 Arctic ecosystems

Some arctic ecosystems consist of **permanent ice desert**.

Other areas are of low-lying land with vegetation described as **tundra**.

The consequences of global warming for arctic ecosystems include:

- loss of the ice habitats, leading to extensive flooding of surrounding low-lands, at least temporarily;
- significant decay by microorganisms of the accumulated detritus, once released from its permafrost state, leading to huge releases into the atmosphere of methane and carbon dioxide, previously locked away in dead organic matter; this contributes to further global warming;
- an expansion in the range of habitats as significant quantities of soil rich in humus are formed;
- appearance and growth of conifers, forming boreal forests (also known as taiga); these areas absorb radiant heat energy from sunlight and contribute to further warming of the region, since they replace ice, snow and frozen tundra;
- appearance of insect-eating species, particularly of birds, able to take advantage of further increasing numbers of insects;
- a wider flora including annual plants typical of alpine meadows and grassland in summer seasons, where these appear;
- appearance of small mammals such as the alpine marmot, able to take advantage of the expanding range of plant biota and habitats (in winter periods such mammals need to be able to retreat into excavated dens for group hibernation);
- predators for the expanded vertebrate populations, mostly birds of prey (raptors) that can fly on when winter returns;
- increased presence of pathogens that parasitise the expanded range of animal and plant life the changing habitat supports.

Extension: The enhanced greenhouse effect – important sources

Evidence on the causes and effects of enhanced global warming continues to be sought and the data interpreted, not least so that steps and proposals for action to avoid an environmental disaster may be effective. However, all aspects of the issue are controversial to some extent. Our understanding is likely to continue to develop in what is a fast-changing situation. Consequently, there are very likely to be relevant developments within the time span of your International Baccalaureate programme, and you may need to consult other sources. In the table below are some references currently pertinent to this issue.

Environmental change due to global warming will be less extreme than predicted
www.newscientist.com/article/mg18524861.500.html 'Meet the global warming sceptics', 12 February 2005
W. Kininmonth (2004), *Climate change: A Natural Hazard,* Multi-Science Publishing Co. ISBN 09065 22269
Evidence of how global temperatures have regularly fluctuated in geological time
Search Google for 'Climate and the Carboniferous Period'. You may well find articles providing a graph of atmospheric CO_2 and average global temperature variation for the past 600 million years of Earth history.
Current and recent human activities as the cause of enhanced global warming, the disastrous consequences, and necessary action now
James Lovelock (2006) *The Revenge of Gaia: Why the Earth is Fighting Back – and How We Can Still Save Humanity*, Allen Lane (London) ISBN 07139 99144
Al Gore (2006) *An Inconvenient Truth: The Planetary Emergency of Global Warming and What We Can Do About It*, Bloomsbury Publishing ISBN 07475 89062

Table 6.4 Sources on the issues of enhanced global warming

Populations

5.3.1–5.3.4

Another practical approach to analysing what goes on in an ecosystem is to find out how the separate populations which make up the ecosystem grow and maintain themselves. Remember, a population is defined as a group of individuals of the same species that have the potential to breed with each other. In an ecosystem such as a wood or lake, the numbers of some species remain remarkably stable, but other species have times of rapid growth, and sometimes populations 'crash'. So, populations can vary greatly.

What are the factors in an ecosystem that control population size?

Population growth

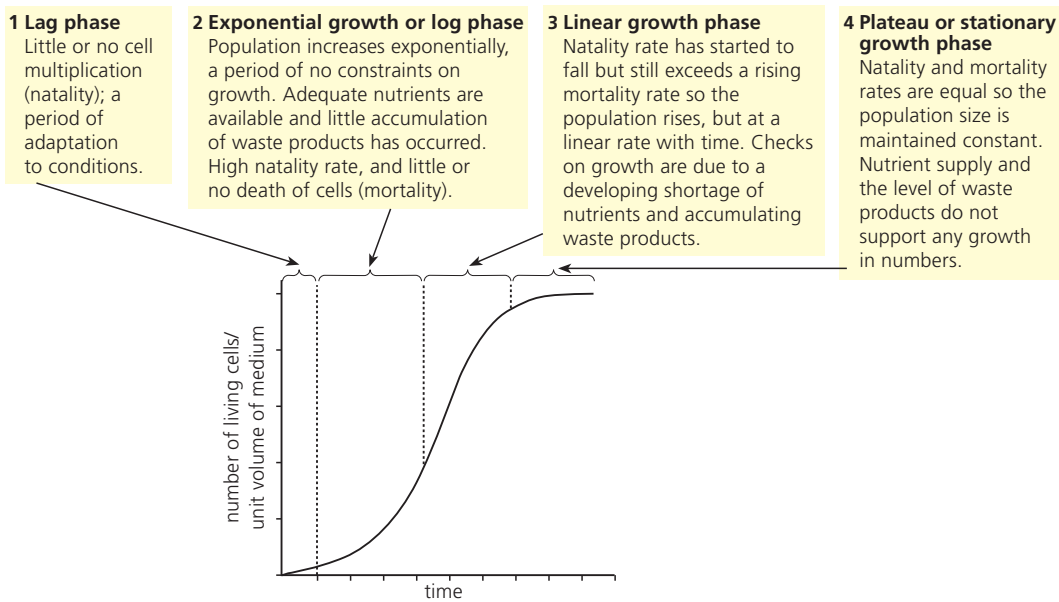
First, think of the situation where a small number of young rabbits enter a large, well-stocked meadow area from which other rabbits continue to be excluded. The rabbits initially familiarise themselves with the new territory, and establish burrows. Then, benefiting from the more or less unrestricted access to food supplies, breeding would commence.

The population of rabbits would increase very rapidly, but eventually, because of the very large numbers of rabbits produced, the vegetation would be used up faster than it grew. Further increases in population now stop, and the population enters a plateau or stationary phase. In this situation, the supply of food has become a limiting factor in the growth of the rabbit population.

If we had maintained a careful census of rabbit numbers throughout this experiment, a graph of number of individuals plotted against time would give us an S-shaped (sigmoid) curve (Figure 6.17).

Figure 6.17 The sigmoid curve of growth

The pattern of growth of a population of microorganisms after being inoculated into a fresh medium sample is shown. The distinct phases to growth are identified.



This situation of a population in the wild is similar to the change in population produced when a quantity of microbiological growth medium, called a broth, is inoculated by a few bacteria, under controlled laboratory conditions, and the growth of the bacterial population is measured. The bacteria reproduce by a process called binary fission in which fully grown bacteria each divide into two and these cells then grow to full size. The bacteria feed on the nutrients in the medium (such as glucose, amino acids, and growth factors), slowly using them up. Waste products start to accumulate.

The typical change in this population with time is shown by the graph in Figure 6.17. As in the case of the rabbit population described above, we see that the population numbers pass through distinct phases. Each phase is identified by a descriptive name:

- organisms become adapted to the medium (**lag phase**);
- growth proceeds at an exponential rate (exponential growth or **log phase**);
- growth proceeds at a more steady rate (**linear growth phase**);
- growth slows down and stops (stationary or **plateau phase**).

Exponential growth cannot be maintained

So, when a population has access to ideal conditions, there is an exponential period of growth, whether it is a population of mammals or microorganisms. In exponential growth, the population doubles in each generation: one individual becomes 2, then they become 4, 8, 16, 32, 64, 128, 256, and so on. This type of increase in a population is quick, but is not maintained because there are insufficient resources for it.

As a population increases, it begins to experience environmental resistance, because:

- space and resources are reduced;
- competition for space and resources increases.

The population tends to stabilise at a level which ecologists call the **carrying capacity** of the habitat.

If the numbers start to increase above carrying capacity, shortage of resources reduces the numbers of offspring produced, and the population regulates itself at the carrying capacity. If the population is reduced, say by heavy predation, the additional resources that then become available lead to an increase in reproductive rate, and the carrying capacity is again reached.

In other words, populations tend to be naturally self-regulating.

10 **Suggest** what factors might:

a lead to an increase in a population of a species of songbird in a wood or forest habitat near you

b eventually limit the growth of that population.

Why populations fluctuate in the wild

In practice, most populations in an ecosystem are already established, but do show significant fluctuations in numbers over a period of time. One example is seen in the study of the populations of phytoplankton (producers) and zooplankton (primary consumers) in a fresh water lake, studied for a 12-month period (Figure 6.18). We can see there is a significant spring 'bloom' in the phytoplankton population in the third and fourth months of the year, and that a corresponding surge in the zooplankton population occurred shortly afterwards. Note that the abiotic factors of essential nutrients, light and mean temperature were also measured and recorded.

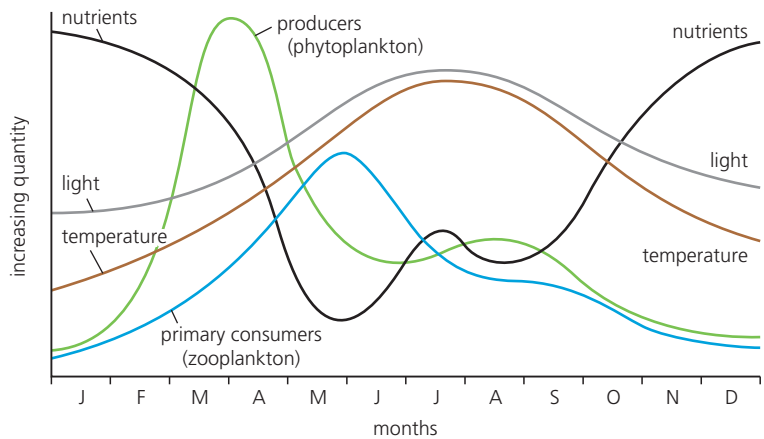


Figure 6.18 Plankton of a fresh water lake in Northern Europe, with data on abiotic factors

11 **Examine** the results of the ecological study shown in Figure 6.18, and then **suggest**:

- **a** two soluble nutrients in the lake water that would be taken up by phytoplankton and facilitate their rapid growth rate in March and April
- **b** why the numbers of phytoplankton decrease significantly by May and June, despite the favourable conditions of light and temperature
- **c** the most likely source of the increase in nutrients that begins in October.

So, in ecosystems the numbers in each population may fluctuate to varying degrees with time. In fact, population sizes vary for the following reasons:

- the birth rate (**natality**) varies;
- the death rate (**mortality**) varies;
- mobile members of the population may move away to new habitats, referred to as **emigration**;
- new members of species may arrive from other habitats, referred to as **immigration**.

In addition, it is possible that:

- A sudden, rapid change in one or more of the physical or chemical components of the environment may occur, reducing the size of the population or reducing the birth and death rates. This might be due to severe drought, or intense cold, for example. The effects of such change are unrelated to the density of the population. They are called **density-independent factors**.
- The effect of other members of the population and of members of other populations, in the form of competition for resources, predation and grazing, and parasitism, may all affect the population numbers adversely. The effects of these factors increase with increasing population numbers. They are known as **density-dependent factors**.

Evolution

5.4.1–5.4.8

By 'evolution' we mean 'the gradual development of life in geological time'. The word 'evolution' is used very widely, but in biology it specifically means 'the processes by which life has been changed from its earliest beginnings to the diversity of organisms we know about today, living and extinct'. It is the development of new types of living organisms from pre-existing types by the accumulation of genetic differences over long periods of time.

Evolution is the process of cumulative change in the heritable characteristics of a population.

Evidence for evolution

Evidence for evolution comes from many sources, including from the study of fossils, from artificial selection in the production of domesticated breeds, and from studies of the comparative anatomy of groups of related organisms.

Fossils

We learn about the history of life from the evidence of the fossils that have been found. Fossilisation is an extremely rare, chance event. This is because predators, scavengers and bacterial action normally break down dead plant and animal structures before they can be fossilised. Of the relatively few fossils formed, most remain buried, or if they do become exposed, are overlooked or are accidentally destroyed before discovery.

Nevertheless, numerous fossils have been found – and more continue to be discovered all the time. The various types of fossil and the steps of fossil formation by petrification are illustrated in Figure 6.19. Where it is the case that the fossil or the rock that surrounds it can be accurately dated (as is often the case, using radiometric dating techniques that exploit natural radioactivity – as in ^{14}carbon or the ratio of ^{40}potassium to ^{40}argon), we have good evidence of the history of life.

Figure 6.19 Fossilisation

Fossil forms

petrification – organic matter of the dead organism is replaced by mineral ions
mould – the organic matter decays, but the space left becomes a mould, filled by mineral matter
trace – an impression of a form, such as a leaf or a footprint, made in layers that then harden
preservation – of the intact whole organism; for example, in amber (resin exuded from a conifer, which then solidified) or in anaerobic, acidic peat

Steps of fossil formation by petrification

1. dead remains of organisms may fall into a lake or sea and become buried in silt or sand, in anaerobic, low-temperature conditions
2. hard parts of skeleton or lignified plant tissues may persist and become impregnated by silica or carbonate ions, hardening them
3. remains hardened in this way become compressed in layers of sedimentary rock
4. after millions of years, upthrust may bring rocks to the surface and erosion of these rocks commences
5. land movements may expose some fossils and a few are discovered by chance but, of the relatively few organisms fossilised, very few will ever be found by humans

Sedimentary rock layers (with a fault line) and the remains of extinct fossil species

12 Many fossils are preserved in sedimentary rocks. **Explain** why is this so, and how sedimentary rocks formed.

Figure 6.20 Charles Darwin's observation of pigeon breeding

From his breeding of pigeons, Darwin noted that there were more than a dozen varieties that, had they been presented to an ornithologist as wild birds, would have been classsified as separate species.

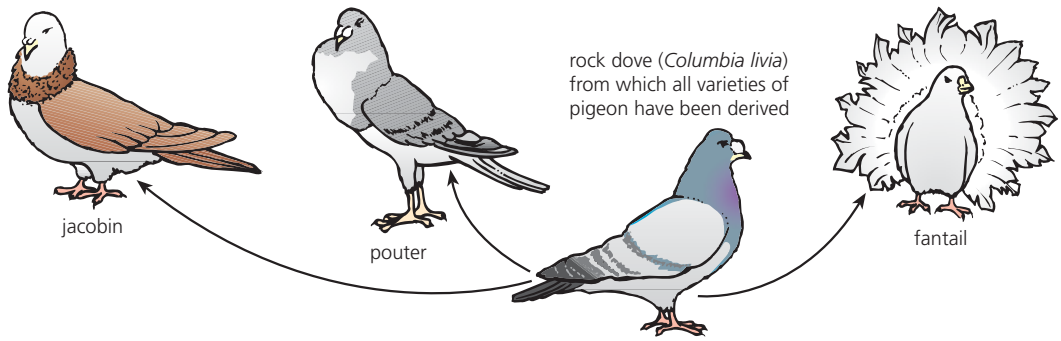
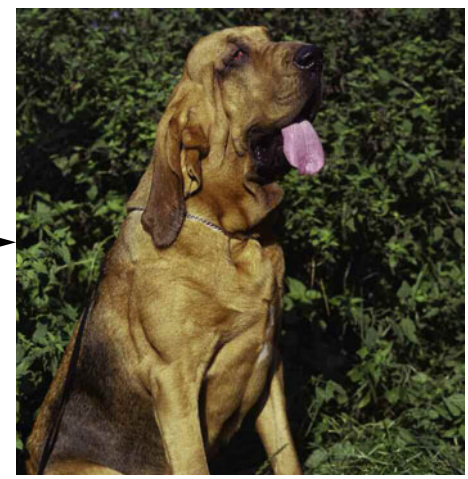


Figure 6.21 Dog breeding

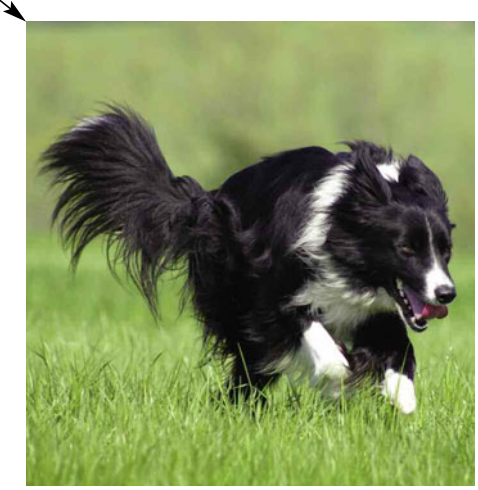
wolf (*Canis lupus*)
differences in size and other features of American, European and Asian wolves are reflected in the dog breeds that have been developed from them

The dog was domesticated from the wolf about 13 000 years ago. It was the first animal to be domesticated.



bloodhound
selectively bred for following air and ground scent in hunting

Irish setter
selectively bred for recreational hunting (with guns), along with pointers and retrievers



collie dog
selectively bred to herd and control flocks of domestic animals

13 Charles Darwin argued that the great wealth of varieties we have produced in domestication supports the concept of evolution. **Outline** how this is so.

Artificial selection

Artificial selection is the process by which all the plants and animals used by humans (in horticulture, agriculture, transport, companionship and leisure) have been derived from wild organisms. Artificial selection involves identifying the largest, the best or the most useful of the progeny for the intended purpose, and using them as the next generation of parents. The continuous culling out of progeny deficient in the desired features, generation by generation, leads to deliberate genetic change in the population; the genetic constitution of the population changes rapidly (Figures 6.20, 6.21).

14 Alsatians, Pekinese and Dachshunds are different in appearance yet are all classified as members of the same species. **Explain** how this is justified.

Comparative anatomy

Studies of the comparative anatomy of many groups of related organisms show that, although adapted to different habitats or life styles, often their underlying organisation is similar (that is, they have **homologous structures**). In such cases, a likely explanation is that they share a common ancestor, and show adaptive radiation from a basic plan (Figure 6.22). On the other hand, fundamentally different structures that have similar functions show only superficial resemblances. These are described as **analogous structures**.

Figure 6.22 Homologous and analogous structures

Homologous structures:

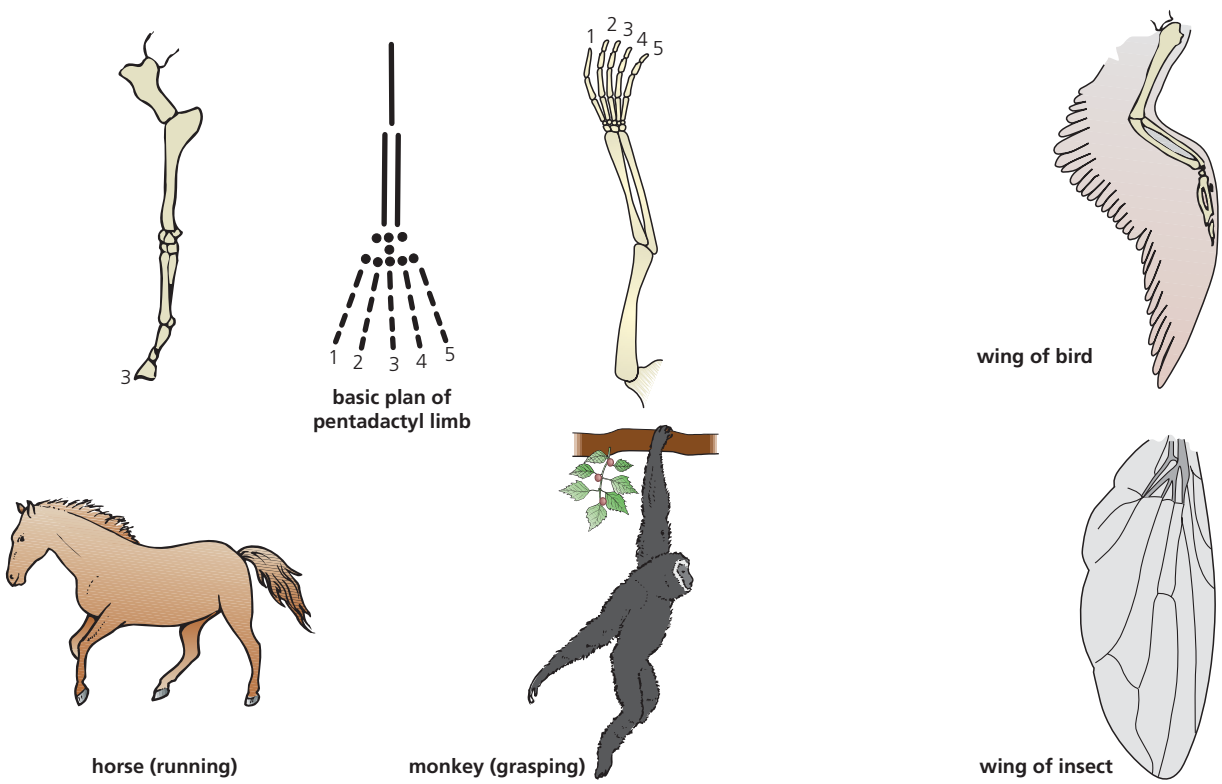
- are similar in fundamental structure
- are similar in position and development, but not necessarily in function
- are similar because of common ancestry

Examples of homologous structures are the **limbs of vertebrates**, all of which appear to be modifications on an ancestral five-fingered (pentadactyl) limb.

Analogous structures:

- resemble each other in function
- differ in their fundamental structure
- illustrate only superficial resemblances

Examples of analogous structures are the **wings of birds and insects**, which are similar only in their function as aerofoils.



Evolution by natural selection – the ideas and arguments

Charles Darwin (1809–82) was a careful observer and naturalist who made many discoveries in biology. After attempting to become a doctor (at Edinburgh University) and then a clergyman (at Cambridge University), he became the unpaid naturalist on an Admiralty-commissioned expedition to the southern hemisphere, on a ship called HMS *Beagle*. On this five-year expedition around the world and in his later investigations and reading, he developed the idea of **organic evolution by natural selection**.

Darwin remained very anxious (always) about how the idea of evolution might be received, and he made no moves to publish it until the same idea was presented to him in a letter by another biologist and traveller, **Alfred Russel Wallace**. Only then (1859) was *On the Origin of Species by Natural Selection* completed and published. The arguments and ideas in the *Origin of Species* are summarised in Table 6.5.

	Statement / deduction
S1	Organisms produce a far greater number of progeny than ever give rise to mature individuals.
S2	The number of individuals in species remain more or less constant.
D1	Therefore, there must be a high mortality rate.
S3	The individuals in a species are not all identical, but show variations in their characteristics.
D2	Therefore, some variants will succeed better than others in the competition for survival. So the parents for the next generation will be selected from those members of the species better adapted to the conditions of the environment.
S4	Hereditary resemblance between parents and offspring is a fact.
D3	Therefore, subsequent generations will maintain and improve in the degree of adaptation of their parents, by gradual change.

Table 6.5 Charles Darwin's ideas about the origin of species, summarised in four statements (S) and three deductions (D) from these statements

Neo-Darwinism

Charles Darwin (and nearly everyone else in the scientific community of his time) knew nothing of Mendel's work. Instead, biologists generally subscribed to the concept of 'blending inheritance' when mating occurred (which would reduce the genetic variation available for natural selection).

Neo-Darwinism is an essential restatement of the concepts of evolution by natural selection in terms of Mendelian and post-Mendelian genetics, as follows.

Organisms produce many more offspring than survive to be mature individuals

Darwin did not coin the phrase 'struggle for existence', but it does sum up the point that the over-production of offspring in the wild leads naturally to their competition for resources. Table 6.6 is a list of the normal rate of production of offspring in some common species.

Organism	No. of eggs / seeds / young per brood or season
rabbit	8–12
great tit	10
cod	2–20 million
honey bee (queen)	120 000
poppy	6 000

Table 6.6 Numbers of offspring produced

How many of these offspring survive to breed themselves?

In fact, in a stable population, a breeding pair may give rise to a single breeding pair of offspring, on average. All their other offspring are casualties of the 'struggle'; many organisms die before they can reproduce.

So, populations do not show rapidly increasing numbers in most habitats, or at least, not for long. Population size is naturally limited by restraints we call **environmental factors**. These include space, light, and the availability of food. The never-ending competition for resources results in the majority of organisms failing to survive and reproduce. In effect, the environment can only support a certain number of organisms, and the number of individuals in a species remains more or less constant over a period of time.

The individuals in a species are not all identical, but show variations in their characteristics

Today, modern genetics has shown us that there are several ways by which **genetic variations** arise in **gamete formation** during meiosis, and at **fertilisation**. We have seen that genetic variations arise via:

- **random assortment** of paternal and maternal chromosomes in meiosis (occurs in the process of gamete formation, page 224);
- **crossing over** of segments of individual maternal and paternal homologous chromosomes (results in new combinations of genes on the chromosomes of the haploid gametes produced by meiosis, page 96);
- the **random fusion** of male and female gametes in sexual reproduction (this source of variation was understood in Darwin's time).

Additionally, variation arises due to **mutations** – either chromosome mutations or gene mutations, page 98.

As a result of all these, the individual offspring of parents are not identical. Rather, they show variations in their characteristics.

Natural selection results in offspring with favourable characteristics

When genetic variation has arisen in organisms:

- the favourable characteristics are expressed in the phenotypes of some of the offspring;
- these offspring may be better able to survive and reproduce in a particular environment; others will be less able to compete successfully to survive and reproduce.

Thus, natural selection operates to determine the survivors and the genes that are perpetuated in future progeny. In time, this selection process may lead to new varieties and new species.

The operation of natural selection is sometimes summarised in the phrase **survival of the fittest**, although these were not words that Darwin used, at least not initially.

15 **Deduce** the importance of modern genetics to the theory of the origin of species by natural selection.

To avoid the criticism that 'survival of the fittest' is a circular phrase (how can fitness be judged except in terms of survival?), the term 'fittest' is understood in a particular context. For example, the fittest of the wildebeest of the African savannah (hunted herbivores) may be those with the acutest senses, quickest reflexes, and strongest leg muscles for efficient escape from predators. By natural selection for these characteristics, the health and survival of wildebeests is assured.

Environmental change and speciation

When environmental changes put selected individuals at a disadvantage, then natural selection is likely to operate on individuals and causes changes to gene pools. We have already noted that individuals possessing a particular allele, or combination of alleles, may be more likely to survive, breed, and pass on their alleles than other, less well-adapted individuals are. This process is also referred to as **differential mortality**; two examples follow.

1 **Multiple antibiotic resistance in bacteria**
This is an example of the situation in which, while the majority of an existing form of an organism may no longer be best suited to the environment, some unusual or abnormal forms of the population have a selective advantage.

Certain bacteria cause disease. Patients with bacterial infections are frequently treated with an antibiotic to help them overcome the infection. Antibiotics are very widely used. In a large population of a species of bacteria, some may carry a gene for resistance to the antibiotic in question. Sometimes, such a gene arises by spontaneous mutation; sometimes the gene is acquired in a form of sexual reproduction between bacteria of different populations.

The resistant bacteria in the population have no selective advantage *in the absence of the antibiotic*, and must compete for resources with non-resistant bacteria. But when the antibiotic is present, most bacteria of the population will be killed off. The resistant bacteria are very likely to survive and be the basis of the future population. In the new population, all now carry the gene for resistance to the antibiotic; the genome has been changed abruptly (Figure 6.23).

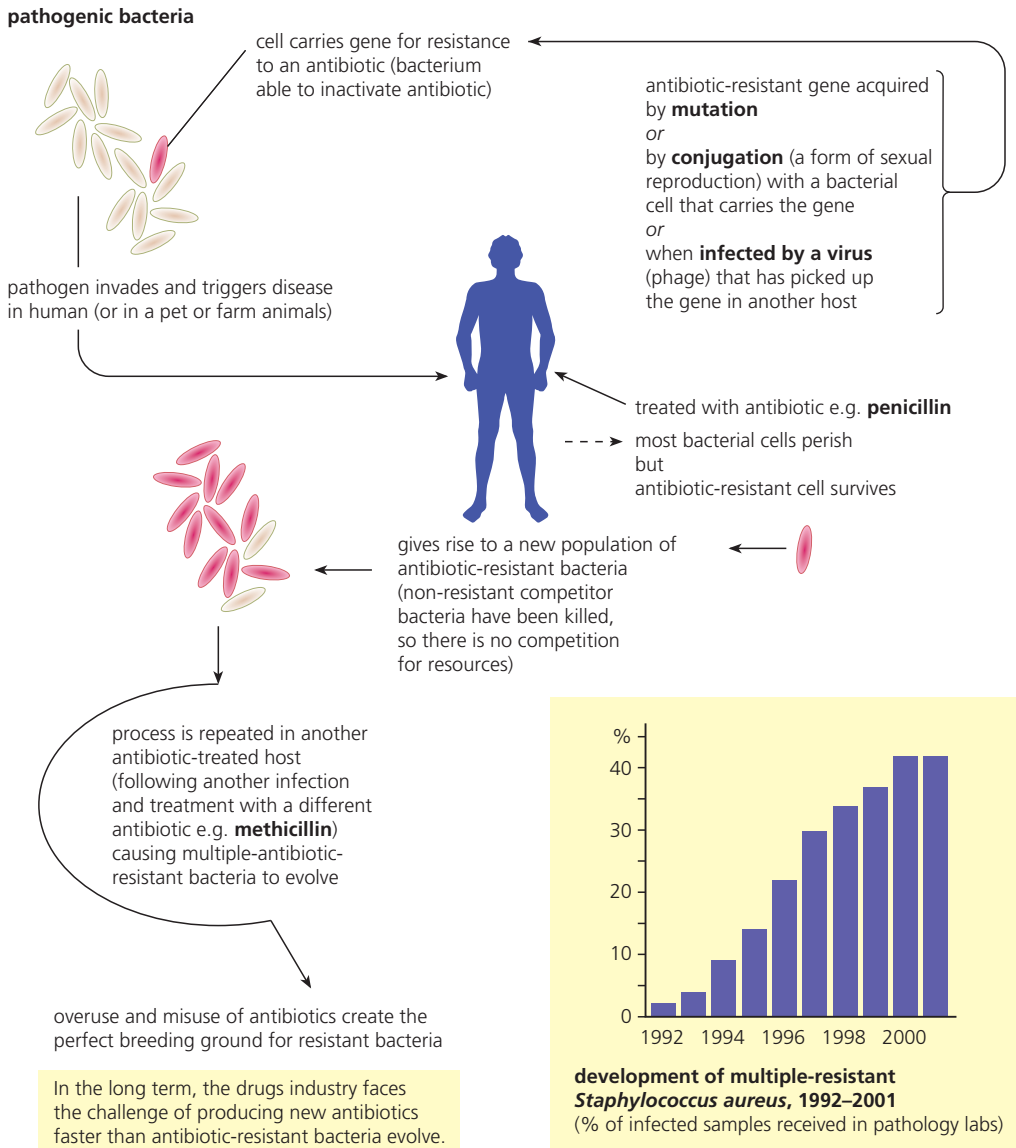


Figure 6.23 Multiple antibiotic resistance in bacteria

16 Explain:

a why doctors ask patients to ensure that they complete their course of antibiotics fully

b why the medical profession tries to combat resistance by regularly alternating the type of antibiotic used against an infection.

2 Heavy metal tolerance in plants

This is a phenomenon associated with those plants able to survive and even flourish on the otherwise bare waste tips and spoil heaps commonly found at mining sites. Here, heavy metals such as zinc, copper, lead, and nickel may be present as ions dissolved in soil moisture at concentrations that generate toxic conditions for plants normally present on the surrounding unpolluted soils. We have seen that several heavy metal ions are essential for normal plant growth when present in trace amounts (page 38). In mining spoils, these levels are frequently exceeded, and spoil heaps from eighteenth- and nineteenth-century mining activities in several countries around the world remain largely bare of plant cover, even when surrounding, unpolluted soils have dense vegetation cover. Seeds from these plants regularly fall on spoil heap soil, but plants fail to establish themselves.

However, careful observations of spoil heaps at many locations have disclosed the presence of local populations of plants that have evolved tolerance. One example is the grass *Agrostis tenuis* (Bent grass), populations of which are tolerant of otherwise toxic concentrations of copper (Figure 6.24).

A variety of biochemical and physiological mechanisms have evolved in tolerant species, including:

a the selective ability to avoid uptake of heavy metal ions;
b the accumulation of ions that enter in insoluble compounds in cell walls by formation of stable complexes with wall polysaccharides;
c transport of toxic ions into the vacuoles of cells, the membranes of which are unable to pump them out again, so avoiding interactions with cell enzymes.

The evolution of this form of tolerance has been demonstrated in several species of terrestrial plants, and also in species of seaweeds now tolerant of copper-based antifouling paints frequently applied to the hulls of ships.

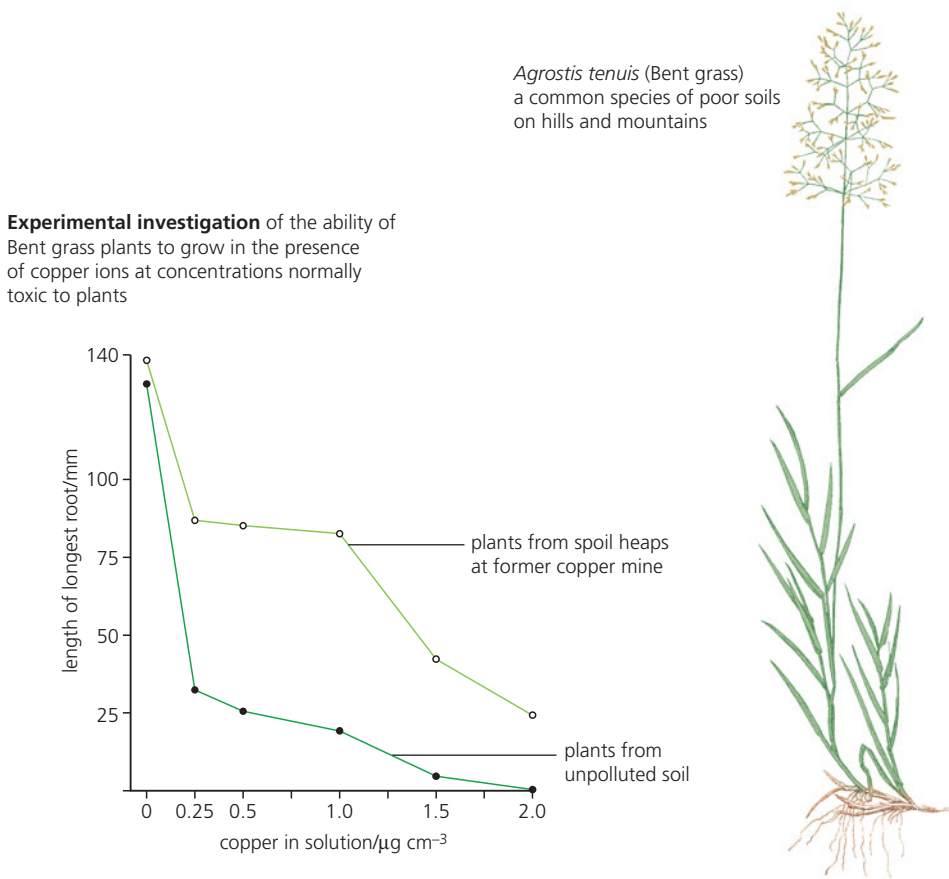


Figure 6.24 Copper ion tolerance in populations of Bent grass

Extension: The development of the ideas of evolution

The idea of biological evolution is closely linked to the name of **Charles Darwin**, but in fact evolution was discussed by **Erasmus Darwin** (1731–1802), Charles's grandfather, and by other biologists and geologists long before the publication of Charles Darwin's controversial theory shocked Victorian Britain.

The earliest, significant contribution was the realisation that the **Earth was extremely old**. In Western culture, the biblical account of creation was generally accepted as authoritative until the eighteenth century, at least. Furthermore, the chronology detailed in the Bible suggested that life had appeared on Earth a mere few thousand years ago. This timescale, which meant that the Earth was only 5000–6000 years old, was widely accepted in Europe until well into the nineteenth century.

James Hutton (1726–97), a doctor, farmer and experimental scientist, realised that the sedimentary rocks of many existing mountain ranges had once been the beds of lakes and seas and, before that, had been the rocks of even older mountains. He made no estimate of the age of the Earth, but he realised that, in contrast to biblical estimates, the **Earth's timescale** virtually had **no beginning and no end**. Nowadays, geologists estimate the age of the Earth to be **4500 million years**, and **life to have originated 3500 million years ago**. With these timescales it became possible to imagine organic evolution by gradual change.

17 **Explain** the significance for evolution theory of the realisation by early geologists that the Earth was more than a few thousand years old.

18 **Suggest** what sort of events might, in the past, have caused violent and speedy habitat change over a substantial part of the surface of the Earth.

TOK Link

The word 'theory' comes from a Greek word meaning 'seeing'. Today, a scientific theory normally explains something, usually via some mental model. However, there is diversity of scientific theories. They are not always precise and mathematical, like those of Isaac Newton (1642–1727), which seized people's imaginations and enabled them to see science as extremely precise and exact. This remains a very commonly held view.

Dip into the *The Origin of Species* – say Chapter 7. Find the paragraph on the Greenland whale and the evolution of the baleen plates. How would you describe this reasoning? What sort of 'theory' is presented here?

Classification

5.5.1–5.5.5

Classification is essential to biology because there are far too many different living things to sort out and compare unless they are organised into manageable categories. With an effective classification system in use, it is easier to organise our ideas about organisms and make generalisations. The scheme of classification has to be flexible, allowing newly discovered living organisms to be added into the scheme where they fit best. It should also include fossils, since we believe living and extinct species are related.

The process of **classification** involves:

- giving every organism an **agreed name**;
- arrangement of organisms into **groupings of apparently related organisms**, as far as these relationships are understood, so imposing an overall scheme on the diversity of living things.

The binomial system of naming

Many organisms have local names, but these often differ from locality to locality around the world, so they do not allow observers to be confident they are talking about the same thing. For example, in America the name 'robin' refers to a bird of the size of the European blackbird and altogether different from the European robin (Figure 6.25). To avoid confusion, scientists use an international approach called the **binomial system** (meaning 'a two-part name'). By this system everyone, anywhere knows exactly which organism is being referred to.

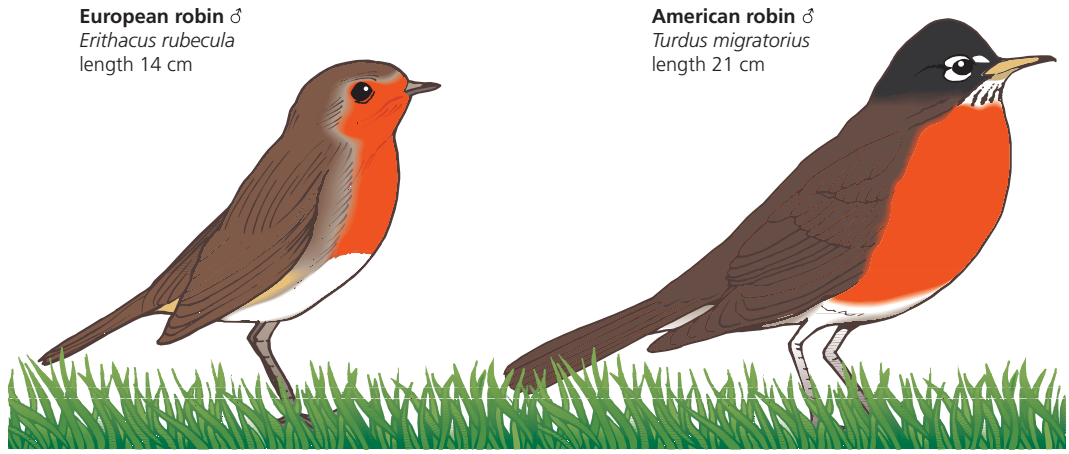


Figure 6.25 Two robins of the northern hemisphere

In the binomial system, each organism is given a scientific name consisting of two Latin words. The first (a noun) designates the **genus**, the second (an adjective) names the **species**. The generic name begins with a capital letter, and is followed by the specific name, which begins with a lowercase letter. Conventionally, this name is written in *italics* when in print (or is underlined in handwriting). As shown in Figure 6.26, closely related organisms have the same generic name; only their species names differ. You will see that when organisms are frequently referred to in a piece of writing, the full name is given initially and thereafter the generic name is shortened to the first (capital) letter. Thus, in continuing references to humans in an article or scientific paper, *Homo sapiens* becomes *H. sapiens*.

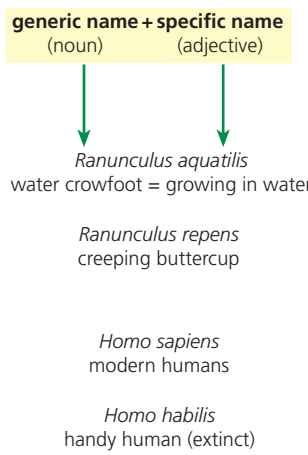


Figure 6.26 Naming organisms by the binomial system

19 Scientific names of organisms are often difficult to pronounce or remember. **State** why they are used.

The scheme of classification

The science of classification is called taxonomy. The word comes from taxa (singular, taxon), which is the general name for groups or categories within a classification system. The taxa used in taxonomy are given in Figure 6.27.

Biological classification schemes are the invention of biologists, based on the best available evidence at the time. In classification, the aim is to use as many characteristics as possible in placing similar organisms together and dissimilar ones apart. Just as similar **species** are grouped together into the same **genus** (plural, **genera**), so similar genera are grouped together into **families**. This approach is extended from families to **orders**, then **classes**, **phyla** and **kingdoms**. This is a hierarchical scheme of classification, each successive group containing more and more different kinds of organism.

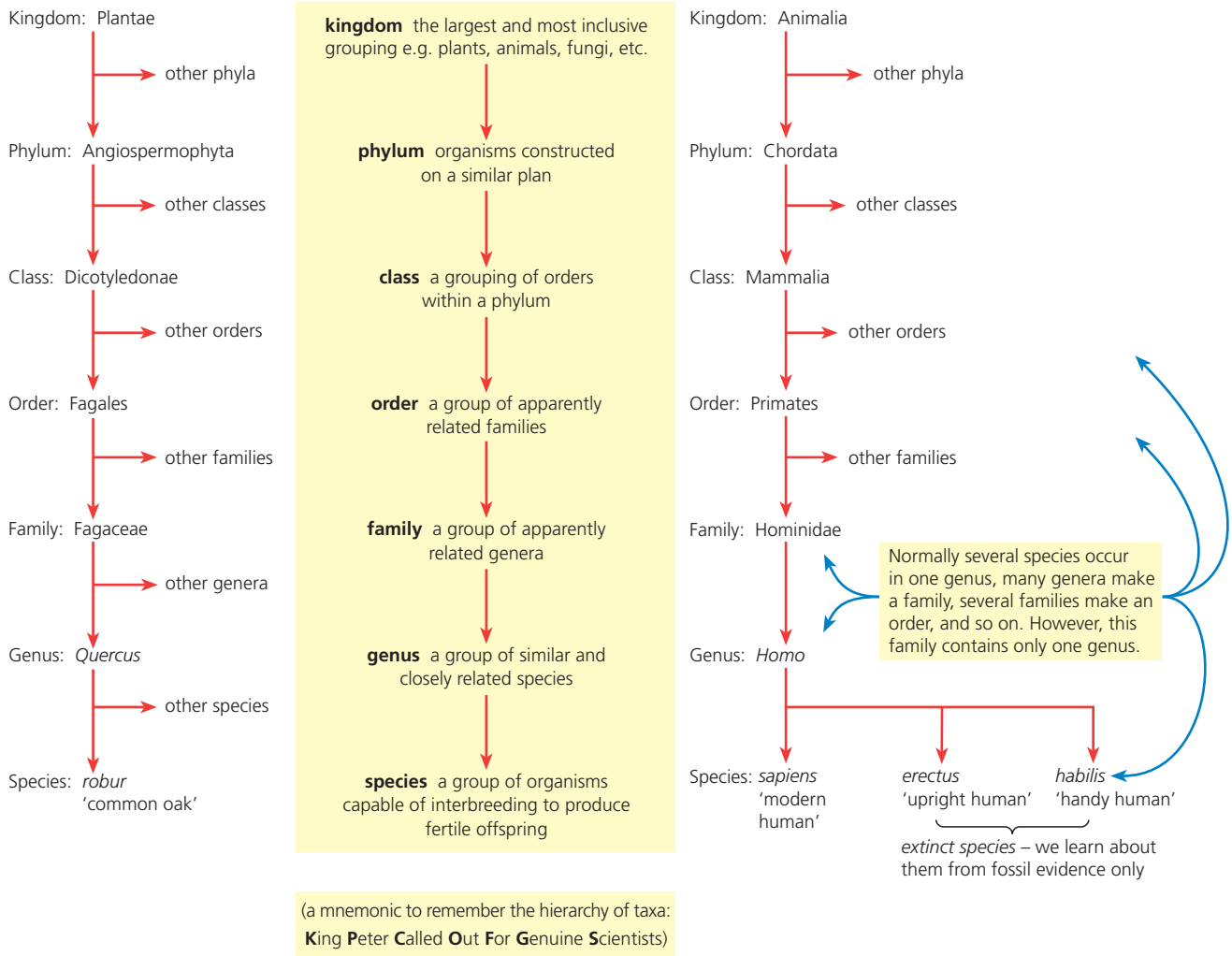


Figure 6.27 Taxa used in taxonomy, applied to genera from two different kingdoms

20 **Design** a method of classifying animals and plants commonly found in gardens that might be useful to an enthusiastic gardener.

Classification of the plant kingdom

The green plants are terrestrial organisms, adapted to life on land, although some do occur in aquatic habitats. They are eukaryotic organisms, with a wall containing cellulose around each cell. In their nutrition, green plants are autotrophic organisms, manufacturing sugar by photosynthesis in their chloroplasts. The sugar is then stored or used immediately to sustain the whole of their metabolism.

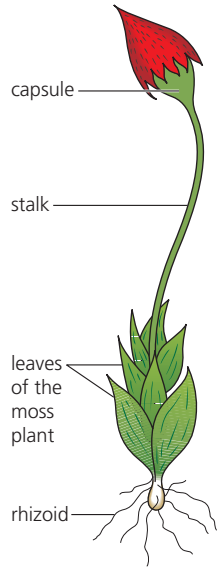
A distinctive feature of green plants is their rather **complex life cycle**: there are two stages or generations, a gametophyte generation that produces gametes, and a sporophyte generation that forms spores. *The details of this feature do not concern us here, but it does account for some otherwise puzzling aspects of green plant structures and life cycles.*

The phyla that make up the green plants

Green plants range from simple, tiny mosses to huge trees – incidentally, trees include some of the largest and oldest living things. The four main phyla of green plants are the mosses (Phylum Bryophyta); ferns (Phylum Filicinophyta); conifers (Phylum Coniferophyta) and flowering plants (Phylum Angiospermophyta). Figures 6.28–31 present the **simple recognition features of the four phyla**, in turn.

Figure 6.28 The mosses – Phylum Bryophyta

the moss *Funaria* is an early coloniser of land after a fire has killed larger plants in woods or heathland

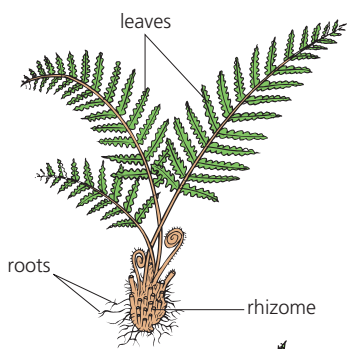


Introducing the Bryophyta

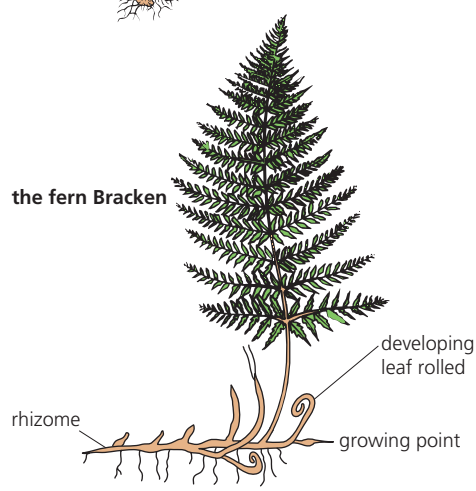
- All are land plants, yet poorly adapted to terrestrial conditions – typically restricted to damp environments.
- Plant constructed of a tiny stem and radially arranged leaves. Roots absent, but stem anchored by hair-like rhizoids.
- Leaves not protected from water loss by a waxy cuticle, and stem contains no water-conducting cells or supporting 'fibres'.
- Spore-containing capsule grows on the main (haploid) plant on a long stalk with its 'foot' in the moss stem.
- Spore capsule may have an elaborate spore-dispersing valve mechanism.
- Alternatively, some Bryophyta exist as flat leaf-like structures on the soil surface, and are known as liverworts.

Figure 6.29 The ferns – Phylum Filicinophyta

the fern *Dryopteris*



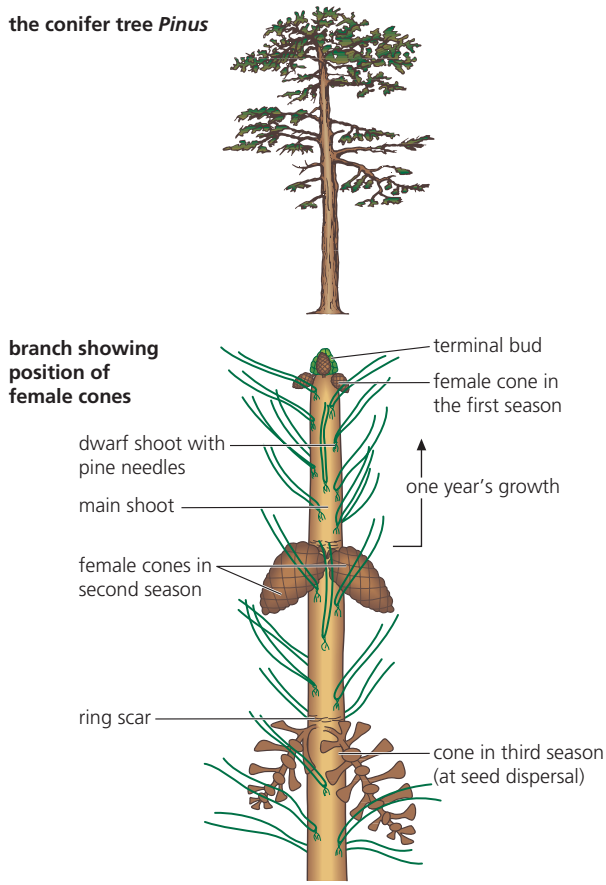
the fern Bracken



Introducing the Filicinophyta

- Ferns are green plants with stems, leaves and roots, and are well-adapted to terrestrial conditions. (Stems growing just below ground are called rhizomes.)
- Within stem, leaves and roots is vascular tissue for conducting water and nutrients around the plant.
- The leaves are elaborate structures that form tightly coiled up, and uncoil in early growth.
- Leaves are covered by a waxy cuticle that protects against water loss by evaporation.
- Spore-producing structures (sporangia) occur in clusters on the under-surface of the leaves, protected below a flap of tissue.
- Spores are released explosively, thrown some distance and then germinate to produce a tiny, independent leaf-like plant.
- This tiny gamete-forming plant (haploid) is where the zygote is formed that then grows into a new fern plant.
- Present day ferns are relatively small survivors of an ancient group, dominant within the Carboniferous period (about 355 million years ago). Huge forests grew across a swampy land, with tree-like ferns – now present as today's fossil fuels.

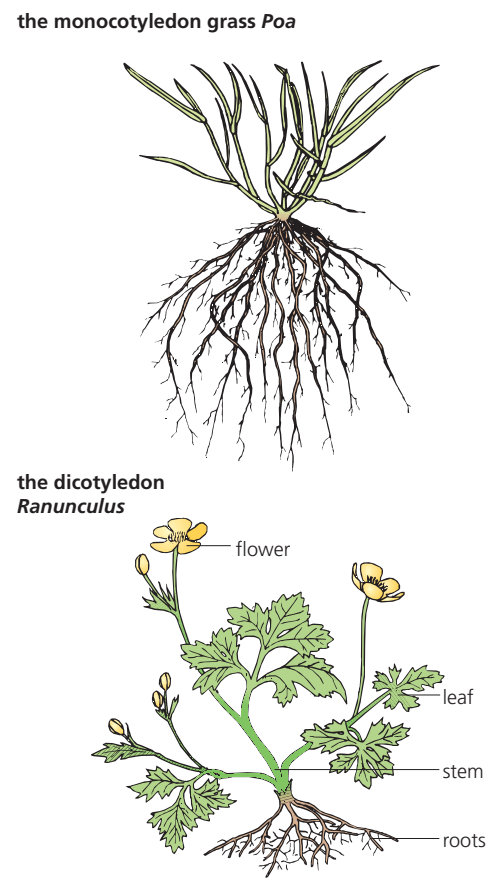
Figure 6.30 The conifers – Phylum Coniferophyta



Introducing the Coniferophyta

- Cone-bearing trees, usually large with strong stem (trunk). They grow well, even on poor soils, and are the dominant plants in northern (boreal) forests, for example.
- The main trunk continues to grow straight, and side branches are typically formed in whorls, below, giving the trees a simple cone-shaped outline, at least as young trees.
- Leaves are usually waxy and needle-shaped. They are mostly evergreen, the leaves able to resist damage in low temperatures and heavy snow fall, when water supplies are locked away as ice.
- Seed-bearing plants, the seeds formed in female cones. Male and female cones often occur on the same tree. Typically, the seed takes 2–3 years to form from arrival of pollen (from male cone) to release and dispersal of seeds.
- Survival on poor soils is aided by their mutually advantageous relationship with soil fungi in modified roots (mycorrhiza) in which fungal threads (hyphae) 'trade' ions from the soil for excess sugar that forms in the tree.
- Since confers mostly grow fast and straight, they form economically important crops of timbers known as 'soft woods'.

Figure 6.31 The flowering plants – Phylum Angiospermophyta



Introducing the Angiospermophyta

- The Angiospermophyta are the dominant group of land plants. Many are herbaceous (non-woody) plants, others are trees (hard woods) or shrubs.
- The stem, leaves and roots contain vascular tissue (xylem and phloem) that delivers water and nutrients all around the plant.
- Leaves are elaborate structures with a waxy, waterproof covering and pores (stomata) in the surface.
- Flowers are unique to the Angiospermophyta, and from them seeds are formed. Seeds are enclosed in an ovary, and after fertilisation, the ovary develops into a fruit.
- With development of the flower have come complex mechanisms of pollen transfer and seed dispersal, often involving insects, birds, mammals or wind and water.
- The Angiospermophyta are divided into the monocotyledons and dicotyledons.
- The monocotyledons (e.g. the grasses) mostly have parallel veins in their leaves, and have a single seed-leaf in the embryo in the seed.
- The dicotyledons (the broad-leaved plants) have net veins in their leaves and two seed leaves in the embryo.

Classification of the animal kingdom

Animals are multicellular, eukaryotic organisms with heterotrophic nutrition. Typically, the body of an animal has cells highly specialised by their structure and physiology to perform particular functions, such as muscle cells for movement. Specialised tissues often occur together to form organs, which perform a particular function in the body as a whole. Most animals have some form of nervous system to coordinate their body actions and responses. These features of animal structure and function are mostly explored in relation to the human mammal, in Chapter 7.

A point of contrast with plants is the simplicity of the life cycle of animals (although some parasitic animals are an exception to this). Their life cycle is diploid, with the adult producing haploid gametes (sperms and ova) by meiosis. After fertilisation, the zygote divides to produce an embryo, which early in development becomes a characteristic hollow ball of cells, called a **blastula** (page 384).

An animal's life style may be reflected in its body plan – many animals are in more or less constant movement, often in search of food, for example. The symmetry of the body of a motile organism is typically **bilateral**, meaning that there is only one plane that cuts the body into two equal halves. Also, with motility comes a compact body, elongated in the direction of movement, so shaped to offer least resistance to the surrounding medium, air or water. Since the anterior (front) end experiences the changing environment first, sense organs become located there. The result is the evolution of a head distinct from the rest of the body, a developmental process called **cephalisation**.

However, the body of an animal with a non-motile or sessile life style may show radial symmetry. The body is approximately cylindrically organised, with body parts arranged around a central axis. There are many planes by which the body can be cut into equal halves. No head is formed.

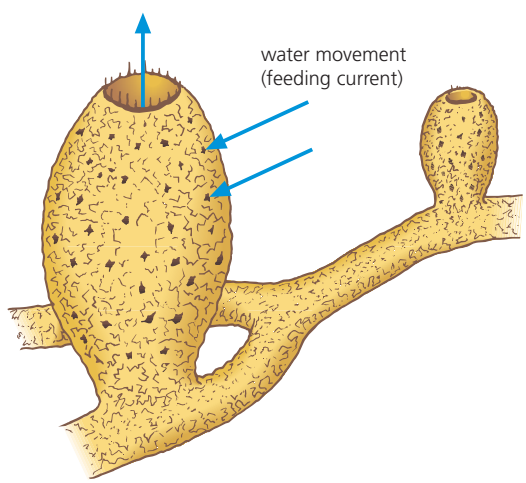
Some non-vertebrate phyla of the animal kingdom

The phyla of animals introduced here are a selection of non-vertebrate groups only, namely the sponges (Phylum Porifera); jellyfishes and sea anemones (Phylum Cnidaria); flatworms (Phylum Platyhelminthes), roundworms (Phylum Nematoda), the molluscs (Phylum Mollusca), and the jointed-limbed animals (Phylum Arthropoda).

Figures 6.32–6.37 present the **simple recognition features of these six phyla**, in turn.

Figure 6.32 The sponges – Phylum Porifera

the sponge *Leucosolenia*



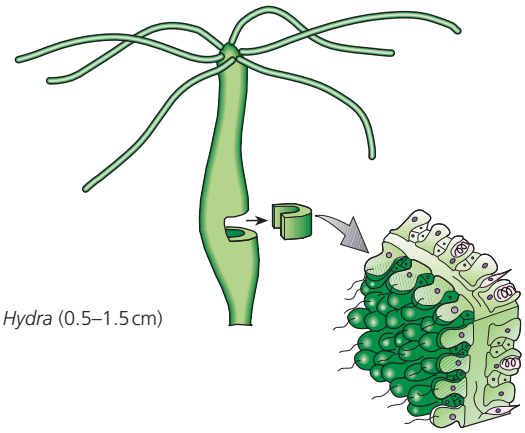
Leucosolenia (1–2 cm wide)

Introducing the Porifera

- The Porifera are the simplest multicellular animals, structurally little more than colonies of cells. They are aquatic and mostly marine animals (once thought to be plants).
- Sponges are formed into simple sac-like structures of cells in two layers, arranged around a central (gastric) cavity.
- Cells of the walls of the sponge specialise for feeding, structural support or reproduction. They form the only multicellular animals to entirely lack any nervous system.
- Feeding is on tiny suspended particles – plankton – which are drawn in through pores in the walls, and are taken up by individual cells and digested.
- Sponges may reproduce asexually by budding, and also sexually – forming a free-swimming larva which is the dispersal stage for this sedentary animal.

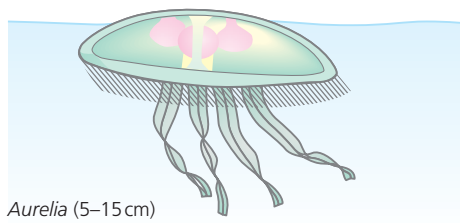
Figure 6.33 The jellyfishes and sea anemones – Phylum Cnidaria

Hydra
a fresh water hydroid cnidarian



Hydra (0.5–1.5 cm)

Aurelia
a marine jellyfish where the medusoid form is the dominant stage



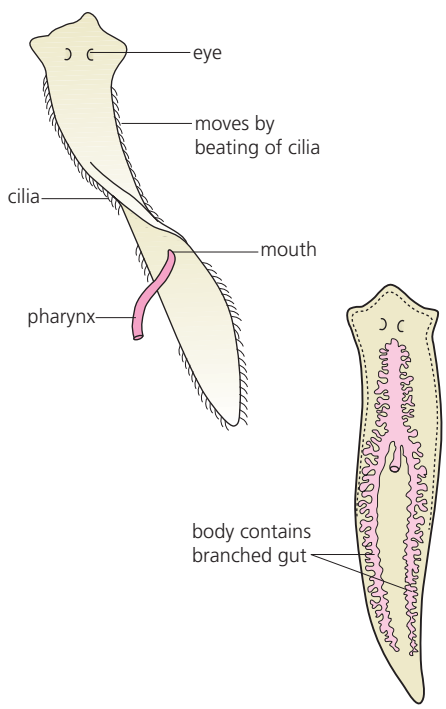
Aurelia (5–15 cm)

Introducing the Cnidaria

- The Cnidaria or coelenterates (meaning 'hollow gut') are aquatic animals, mostly marine, with radially symmetrical body plans.
- The body cavity is a gut with a single opening for ingestion of food and egestion of waste matter.
- The body wall is of two layers of cells – an outer **ectoderm** and an inner **endoderm** – separated by a layer of jelly, the **mesoglea**.
- The ectoderm includes **stinging cells** which may be triggered by passing prey. With these cells, prey is poisoned, paralysed and held until it is pushed into the gut for digestion.
- The stinging cells are found especially on the tentacles.
- Behaviour of the body is coordinated by a nerve net in the mesoglea, in contact with the bases of all wall cells.
- There are alternative body forms in the Cnidaria – either a sessile **hydroid** form (illustrated by *Hydra*) or a floating **medusa** form (illustrated by the jellyfish). These body forms typically alternate in the life cycle.
- The Cnidaria also include the sea anemones and corals.

Figure 6.34 The flatworms – Phylum Platyhelminthes

Planaria
a free-living flatworm

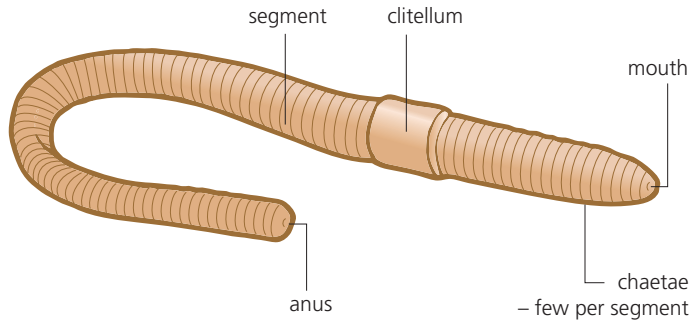


Introducing the Platyhelminthes

- The Platyhelminthes are flat, unsegmented animals with a body built from three cell layers (**triploblastic organisation**).
- There is no cavity in the middle layer. They have a mouth and a gut with numerous branches, but there is no anus. Feeding is by scavenging or predating on other small animals.
- There is no circulatory system in the platyhelminth body, but the generally small, thin, often flat body means that oxygen can diffuse easily to most cells.
- There are 'flame cells' present for excretion and regulation of water and ions in the body.
- Platyhelminthes often have both male and female reproductive organs in one individual (hermaphrodite organisation), but the chances of self-fertilisation are minimal.
- Some are free-living flatworms, but others are parasitic flukes or tapeworms. The phylum contains many important parasites.

Figure 6.35 The segmented worms – Phylum Annelida

Lumbricus
the earthworm

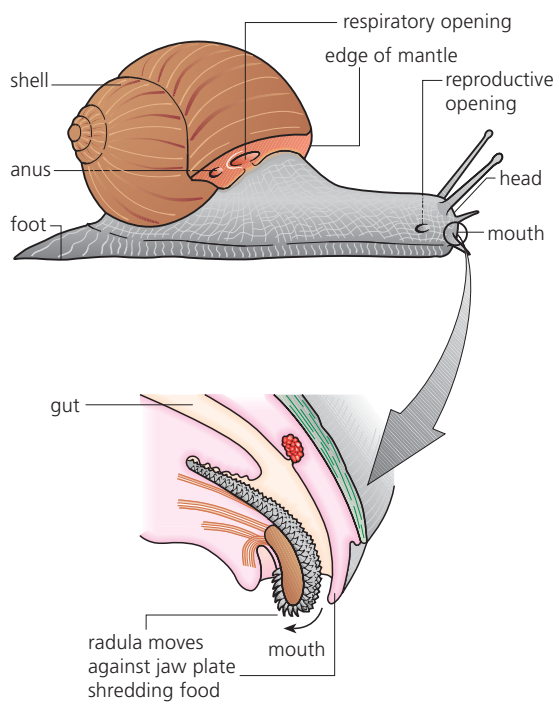


Introducing the Annelida

- Worm-like animals with a soft body (without a rigid skeleton), built of a fixed number of **similar segments** (visible externally as rings). Each segment contains the same pattern of nerves, blood vessels and excretory organs. This arrangement is known as **metameric segmentation**.
- Internally, the segments are separated by septa that divide the **body cavity** (**coelom**) into compartments, each surrounded by the muscular body wall. The coelum is fluid-filled and muscles of the body wall bring about movements by working against the pressure of the fluid in the coelum compartments. This forms a flexible, **hydrostatic skelton**.
- Gaseous exchange occurs through the whole body surface, which is kept moist. The blood contains an oxygen-transporting pigment.
- The collection of sense organs and the feeding structures at the anterior end, known as **cephalisation**, modifies the segmentation pattern somewhat. A solid (non tubular) ventral nerve cord runs the length of the body.

Figure 6.36 The molluscs – Phylum Mollusca

Helix
the snail



Introducing the Mollusca

- The molluscs are the slugs, snails, limpets, mussels and octopuses. They are a huge and diverse group of organisms, the second largest phylum in terms of numbers of species. Most are aquatic and found in fresh water or marine habitats, but a few are terrestrial.
- They are animals with generally soft, flexible bodies which show little or no evidence of segmentation. The body is divided into a head, a flattened muscular foot and a hump or visceral mass often covered by a shell, secreted by a layer of tissue called the mantle.
- The compact body shape of molluscs means that diffusion is not effective for the transport of nutrients, and molluscs have gills or occasionally lungs for gaseous exchange, as well as a well-developed blood circulation.
- Most molluscs have a rasping, tongue-like radula used for feeding.

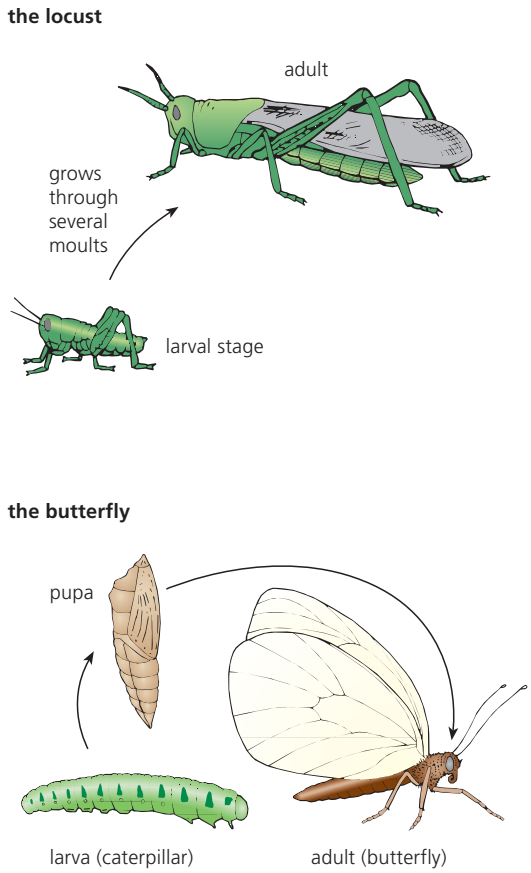


Figure 6.37 The jointed-limbed animals – Phylum Arthropoda

Introducing the Arthropoda

- The Arthropoda are the most numerically successful of all animals, and are divided into five distinct groups: the **crustaceans**, the **arachnids**, the **centipedes**, the **millipedes**, and the **insects**.
- These animals have segmented bodies, covered by a hard external skeleton made of chitin, with jointed limbs (after which they are named). Typically, there is one pair of legs per segment, but this pattern has been lost from some arthropods.
- The exoskeleton cannot grow with the animal, so it has to be shed periodically (moulting), and replaced with a larger one into which the enlarging animal grows.
- The blood circulation is open; the blood is in a haemocoel cavity surrounding all the organs of the body. A tubular heart pumps blood into the haemocoel.
- The functioning of the body is co-ordinated by a ventral nerve cord with nerves running to each segment. There is a concentration of nerves at the front of the body.
- The insects are by far the most numerous and have:
  - a body divided into head, thorax and abdomen
  - three pairs of legs and two pairs of wings attached to the thorax
  - a pair of compound eyes and a pair of antennae on the head
  - head with mouthparts that are modified paired limbs
  - air piped to the tissues by a system of tubes, called tracheae.

The construction of dichotomous keys

The process of naming unknown organisms we come across – for example, in ecological field work – is important but time-consuming. We often attempt this by comparisons, using identification books that are illustrated with drawings and photographs, and that provide information on habitat and habits to give us clues to the identity of organisms.

Alternatively, the use of **keys** may assist us in the identification of unknown organisms. The advantage of using keys is that it requires careful observation. We learn a great deal about the structural features of organisms and get some understanding of how different organisms may be related.

The steps in the construction of a dichotomous key are illustrated first.

Follow the steps, then put them into practice yourself.

Steps in key construction

The steps in key construction are illustrated using eight different tree leaves as shown in Figure 6.38. In selecting the leaves, care must be taken to ensure that each is entirely representative of the majority of leaves the tree has.

First, each leaf is carefully examined and the most significant features of structure are listed in a matrix where their presence (or absence) is recorded against each specimen, as shown in Figure 6.39.

Then, from the matrix, a characteristic shown by half (or thereabouts) of the leaves is selected. This divides the specimens into two groups. A dichotomous flow diagram is constructed, progressively dividing the specimens into smaller groups. Each division point is labelled with the critical diagnostic feature(s), as shown in Figure 6.40.

Finally, a dichotomous key is constructed, reducing the dichotomy points in the flow diagram to alternative statements to which the answer is either 'yes' or 'no'. To each alternative is given a number to which the reader must refer to carry on the identification, until all eight leaves have been identified (Figure 6.41).

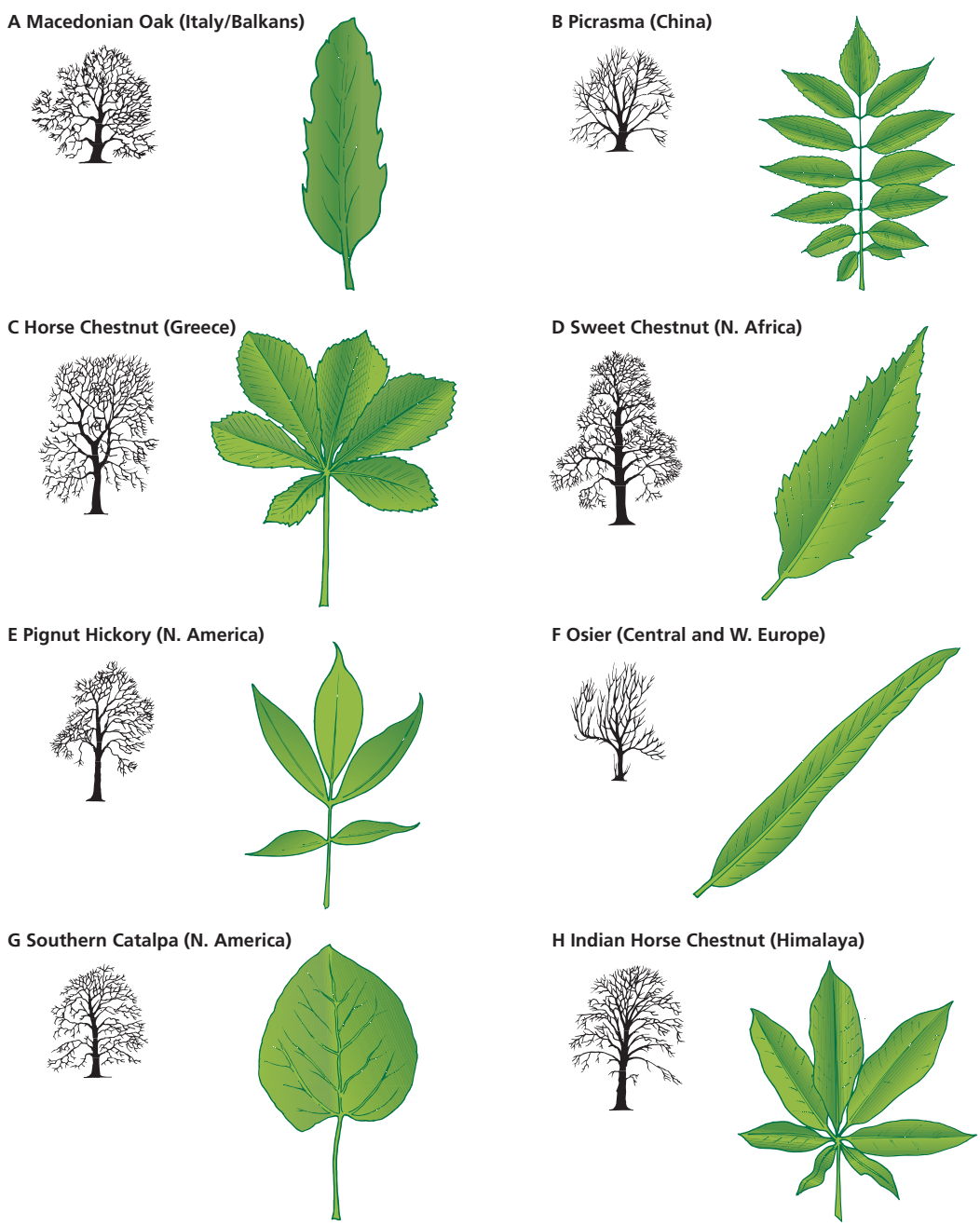


Figure 6.38 Collection of leaves for the construction of a dichotomous key

Figure 6.39 Matrix of characteristics shown by one or more of the sample

Feature of leaf: present (✓) or absent (–)	Tree leaves, identified by number							
	A Macedonian Oak	B Picrasma	C Horse Chestnut	D Sweet Chestnut	E Pignut Hickory	F Osier	G Southern Catalpa	H Indian Horse Chestnut
leaf not divided into leaflets (leaf entire)	✓	–	–	✓	–	✓	✓	–
leaf consists of leaflets	–	✓	✓	–	✓	–	–	✓
leaf blade narrow, with almost parallel sides	✓	–	–	–	–	✓	–	–
leaf blade broad	–	–	–	✓	–	–	✓	–
leaflets radiate from one point (palmate)	–	–	✓	–	–	–	–	✓
leaflets arranged in two rows along stalk	–	✓	–	–	✓	–	–	–
leaf / leaflet margin smooth	–	–	–	–	✓	✓	✓	✓
leaf / leaflet margin toothed, like a saw	✓	✓	✓	✓	–	–	–	–
leaf blade heart-shaped	–	–	–	–	–	–	✓	–
leaf blade boat-shaped (widest in middle)	–	–	–	✓	–	–	✓	–
leaflet paddle-shaped, widest near one end	–	–	✓	–	–	–	–	–
leaflets 5 in number or less	–	–	–	–	✓	–	–	–
leaflets between 6 and 10 in number	–	–	✓	–	–	–	–	✓
leaflets more than 10 in number	–	✓	–	–	–	–	–	–

Figure 6.40 Dichotomous flow diagram of leaf characteristics

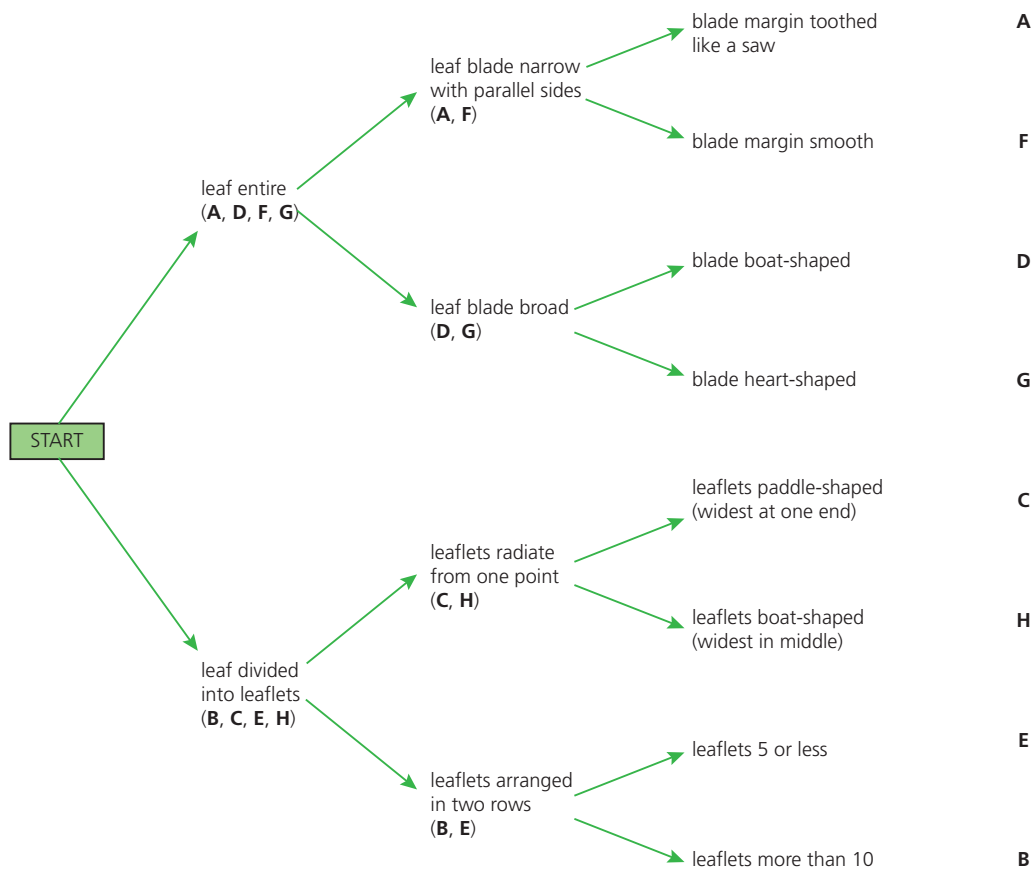


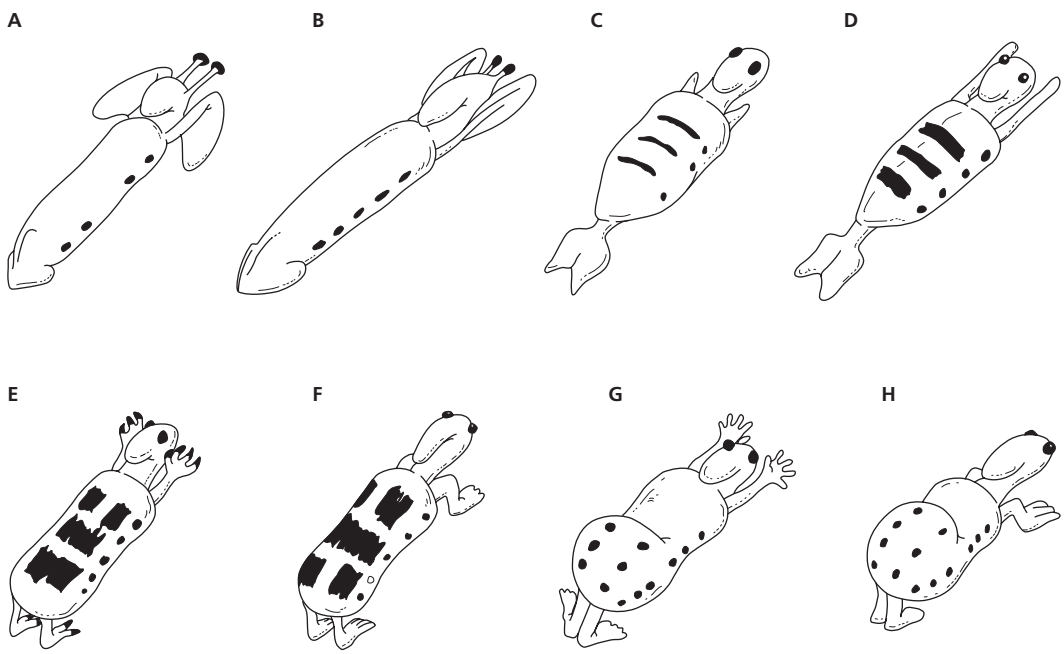
Figure 6.41 Dichotomous key to the sample of eight tree leaves

		Go to ...
1	Leaves entire, not divided into leaflets	**2**
	Leaf blade divided into leaflets	**5**
2	Leaf blade narrow, with almost parallel sides	**3**
	Leaf blade broad rather than narrow	**4**
3	Blade margin toothed like a saw	**Macedonian Oak**
	Blade margin smooth	**Osier**
4	Blade boat-shaped	**Sweet Chestnut**
	Blade heart-shaped	**Southern Catalpa**
5	Leaflets radiate from one point	**6**
	Leaflets arranged in two rows	**7**
6	Leaflets paddle-shaped – widest at one end	**Horse Chestnut**
	Leaflets boat-shaped – widest in the middle	**Indian Horse Chestnut**
7	Leaflets 5 (or less) in number	**Pignut Hickory**
	Leaflets more than 10 in number	**Picrasma**

Practising key construction

Joseph Camin, a biology teacher at Kansas University, designed a family of 'animals' in order to introduce the principles of classification in a practical way. Some of his 'caminacules' are shown in Figure 6.42. Use a copy of these to construct a key, following the steps detailed above. Alternatively, use a suitable selection of eight living biological specimens.

Figure 6.42 A selection of Camin's 'caminacules'



Extension: The five kingdoms

At one time the living world seemed to divide naturally into two kingdoms:

Plants	Animals
photosynthetic (autotrophic nutrition)	ingestion of complex food (heterotrophic nutrition)
mostly rooted (therefore, stationary) organisms	typically mobile organisms

These two kingdoms grew from the original disciplines of biology, namely **botany**, the study of plants, and **zoology**, the study of animals. Fungi and microorganisms were 'added' to botany.

Initially there was only one problem: fungi possessed the typically 'animal' heterotrophic nutrition but were 'plant-like' in structure. Later, with the use of the **electron microscope**, came the discovery of the two types of cell structure, namely **prokaryotic** and **eukaryotic** (page 14). As a result, the bacteria with their prokaryotic cells could no longer be 'plants' since plants have eukaryotic cells. The divisions of living things into kingdoms needed overhauling.

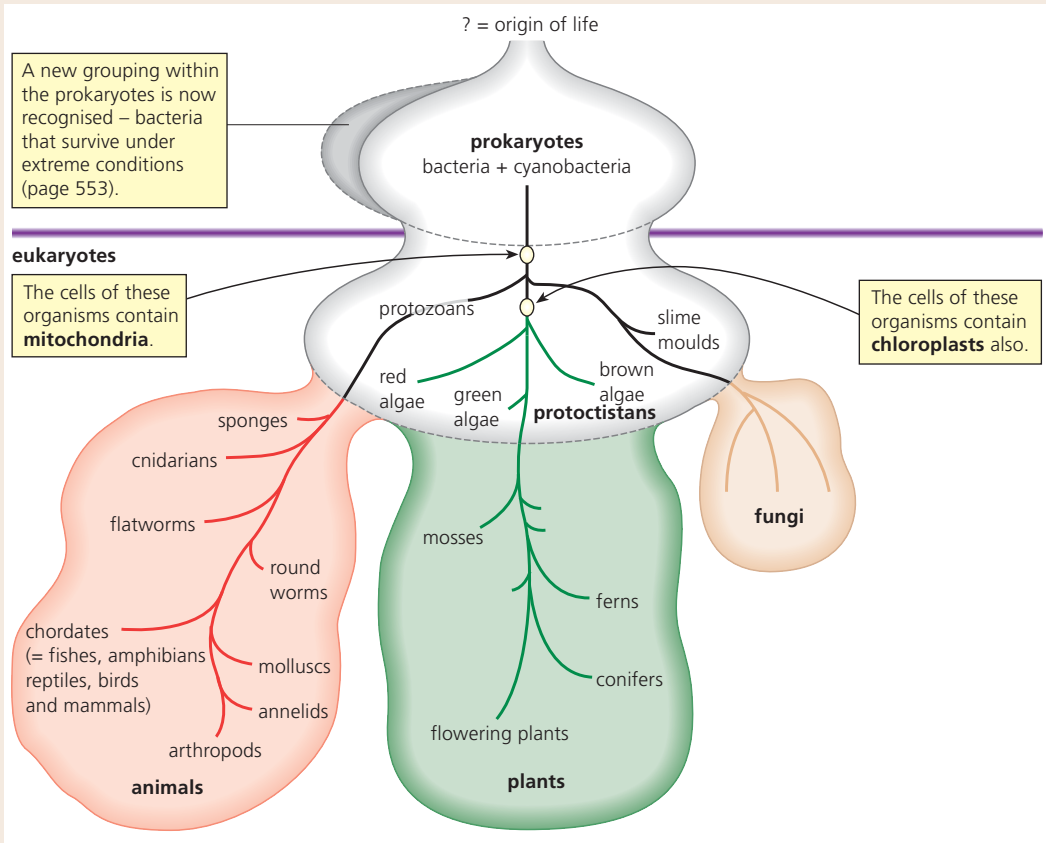


Figure 6.43 Possible evolutionary relationship of the five kingdoms

Today, the agreed division of living things is into five kingdoms (Figure 6.43):

- **Prokaryotae** – the prokaryote kingdom, the bacteria and cyanobacteria (photosynthetic bacteria), predominantly unicellular organisms;
- **Protoctista** – the protoctistan kingdom (eukaryotes), predominantly unicellular, and seen as resembling the ancestors of the fungi, plants and animals;
- **Fungi** – the fungal kingdom (eukaryotes), predominantly multicellular organisms, non-motile, and with heterotrophic nutrition;
- **Plantae** – the plant kingdom (eukaryotes), multicellular organisms, non-motile, with autotrophic nutrition;
- **Animalia** – the animal kingdom (eukaryotes), multicellular organisms, motile, with heterotrophic nutrition.

Examination questions – a selection

Questions 1–5 are taken from past IB Diploma Biology papers:

Q1 Which characteristics apply to all evolving populations?

I overproduction of offspring
II there are differing genotypes in the population
III different chances of survival

A I and II only **C** II and III only
B I and III only **D** I, II and III

Standard Level Paper 1, May 02, Q21

Q2 Which of the following represents a kingdom?

A eukaryotes **C** protoctista
B viruses **D** mammals

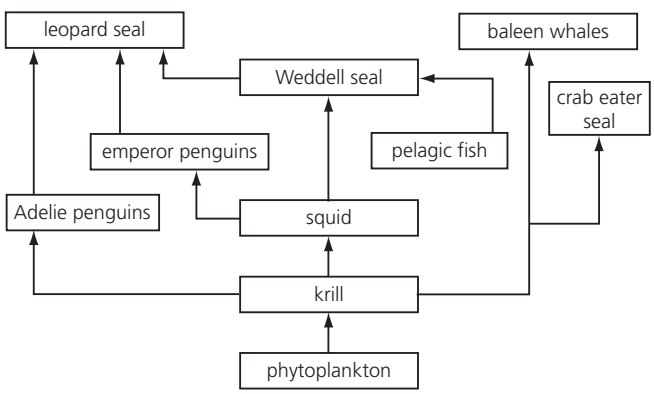
Higher Level Paper 1, May 06, Q18

Q3 Which of the following correctly identifies the role(s) of bacteria in an ecosystem?

A autotroph
B autotroph and detritivore
C detritivore and heterotroph
D autotroph, detritivore and heterotroph

Standard Level Paper 1, May 05, Q21

The following diagram refers to questions 4 and 5. It shows part of the food web of the community that inhabits Antarctica.



Q4 What trophic level do squid belong to?

A tertiary consumers
B secondary consumers
C primary consumers
D producers

Standard Level Paper 1, May 04, Q18

Q5 Which human activity would have the greatest impact on the food web?

A overfishing the krill
B killing the crab eater seals for their skins
C a continued ban on hunting whales
D building an airstrip through the emperor penguin nesting colony

Standard Level Paper 1, May 04, Q19

Questions 6–10 cover other syllabus issues in this chapter.

Q6 By means of concise definition and examples, distinguish between the following terms:

a *ecosystem* and *habitat* (4)
b *population* and *community* (4)
c *species* and *genus*. (4)

Q7 **a** Carbon is present in all organic compounds and in carbon dioxide gas in the atmosphere. Suggest where else inorganic carbon is commonly found in the biosphere, and in what forms it occurs. (4)

b Describe by what processes the carbon of organic compounds will be converted to carbon of carbon dioxide. (2)

c State where conversion of carbon dioxide to organic compounds occurs and what source of energy is involved. (2)

d Explain the fact that the average concentration of atmospheric carbon dioxide is almost stable on a daily basis. (2)

e Identify the part played by the oceans, and organisms in them, in the carbon cycle. (3)

Q8 The individuals in a species are not identical, but show variations in their characteristics. Explain how variations may arise, and what part variations may play in the evolution of species by natural selection. (6)

Q9 **a** Draw a graph of the growth of an animal population after introduction to a new, highly favourable environment. Annotate the stages of the curve to establish the factors influencing change in population size at each stage. (6)

b In a natural environment the numbers of a population in the community of organisms present normally fluctuate unpredictably. Identify the natural influences that change population numbers in these circumstances. (6)

Q10 Use a table as shown to state one external feature characteristic of each group. (7)

Group	Characteristic external feature
cnidarians	
chordates	
mosses	
arthropods	
ferns	
annelids	
flowering plants	

7 Human physiology, health and reproduction

STARTING POINTS

- The bulk of the **carbon compounds** that make up the cytoplasm of cells consist of **carbohydrates**, **lipids** and **proteins**.
- In cells, **energy** is transferred by the breakdown of **nutrients**, principally carbohydrates such as glucose, by the process of **cell respiration**.
- Animals obtain nutrients by **digestion** of food substances, followed by absorption of the products of digestion – described as **heterotrophic nutrition**.
- **Enzymes** are biological catalysts made of **protein**, produced by cells to bring about **chemical changes** inside cells, such as cell respiration, or outside cells, such as **digestion**.
- **Entry into cells** occurs by **diffusion** or **active transport**. Water enters by **osmosis**, a special case of diffusion.
- Microorganisms are a diverse group, some **prokaryotes**, some **eukaryotes**, and some **protoctista**. **Viruses** are not living organisms.
- Reproduction is a characteristic of living things. In sexual reproduction two sex cells (**gametes**) fuse (**fertilisation**) to form a **zygote**.
- As the chromosome number is doubled at fertilisation, the process of gamete formation involves **meiosis** (reductive nuclear division).

Physiology is about how and why the parts of the body function the way they do. Linked with physiology is **anatomy** – the study of structure. Structure gives vital clues to function, as when the structures of our arteries and veins are compared, for example.

Physiology is also an experimental science. For example, physiological experiments have established that cells of organisms function best in an internal environment that stays fairly constant, held and maintained within quite narrow limits. We talk about a **constant internal environment** – which is certainly the case in mammals such as the human, on which we focus here.

In this chapter, we first examine how the body **obtains nutrients**, breathes and **exchanges gases** with the air, and maintains an efficient **internal transport system**.

Pathogens, many of them microorganisms, may bring disease to healthy organisms. Accordingly, cells of our blood circulation also have roles in **defence against communicable diseases**. Finally, the roles of **nerves** and **hormones**, the processes of **homeostasis**, and aspects of **sexual reproduction** are discussed.

Nutrition

6.1.1–6.1.7

An animal takes in food, which is complex organic matter, and digests it in the **alimentary canal** or gut, producing molecules suitable for absorption into the body cells via the blood circulation system. This is known as **holozoic nutrition** (meaning 'feeding like an animal'). Holozoic nutrition is just one of the forms of **heterotrophic nutrition** (meaning 'feeding on complex, ready-made foods') to obtain the required nutrients. The important ecological point to remember about all heterotrophs is their dependence, directly or indirectly, on organisms that manufacture their own elaborated foods. These organisms are the photosynthetic green plants, known as autotrophs. This dependence is demonstrated in food chains (page 139).

1 **State** two forms of heterotrophic nutrition, other than holozoic.

Digestion, where and why

The food an animal requires has to be searched for or possibly hunted. Some mammals eat only plant material (**herbivores**), some eat only other animals (**carnivores**), and others eat both animal and plant material (**omnivores**). Whatever is eaten, a balance of the essential nutrients is required by the body. The sum total of the nutrients eaten is the diet.

The mammalian gut is a long, hollow muscular tube connecting mouth to anus. Along the gut are several glands, and the whole structure is specialised for the movement and digestion of food and for the absorption of the useful products of digestion. The regions of the human gut are shown in Figure 7.1, and the steps of nutrition in Table 7.1.

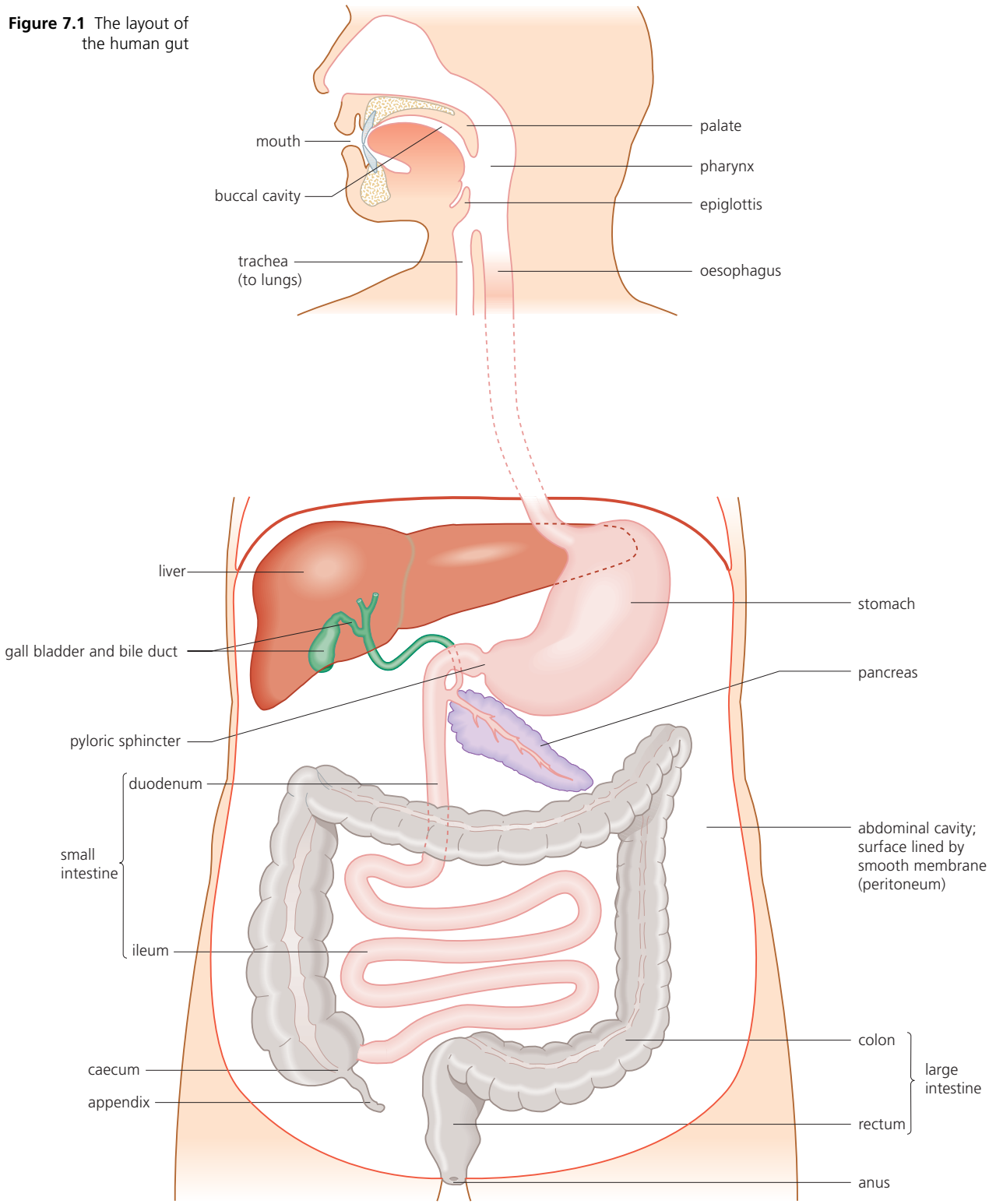
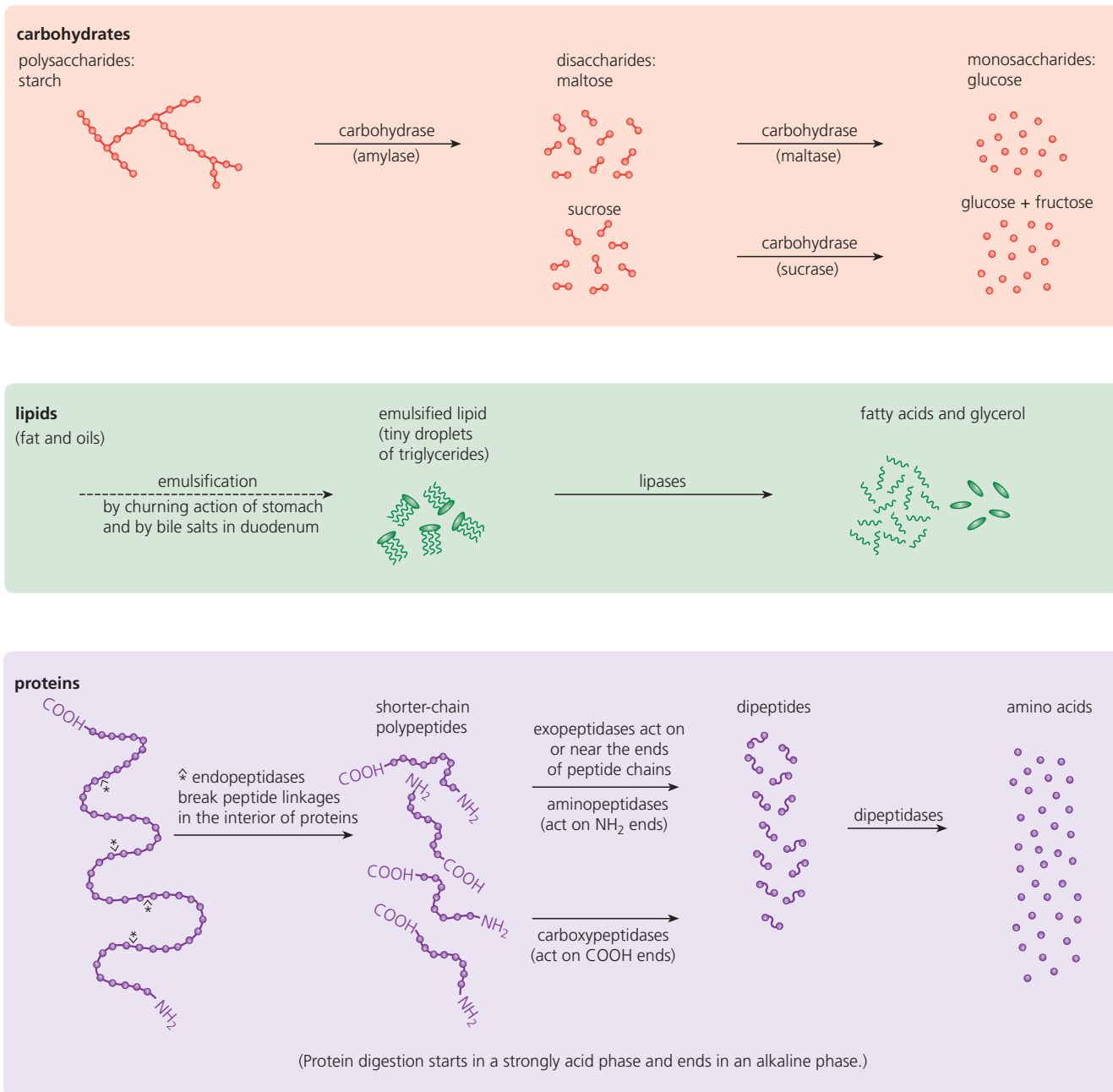


Figure 7.1 The layout of the human gut

Step	Process
ingestion	food taken into mouth for processing in the gut
digestion	■ mechanical digestion by the action of teeth and the muscular walls of the gut ■ chemical digestion by enzymes, mainly in the stomach and intestine
absorption	soluble products of digestion absorbed into blood circulation system (into lymphatic system if fat droplets)
assimilation	products of digestion absorbed from blood into body cells (such as liver and muscle cells) and used or stored
egestion	undigested food and dead cells from the lining of the gut, together with bacteria from the gut flora, expelled from the body as faeces

Table 7.1 The five steps of holozoic nutrition

Figure 7.2 Chemical digestion – the steps



(Protein digestion starts in a strongly acid phase and ends in an alkaline phase.)

Digestion is an essential step because the bulk of the food taken in consists of insoluble molecules that are far too large to cross the gut wall and enter the blood stream. The bulk of our diet consists of carbohydrates, lipids and proteins. These must be hydrolysed to monosaccharides, fatty acids and glycerol, and free amino acids, before they can be absorbed and, later, built up into the carbohydrates, lipids and proteins required by our bodies. The chemical structure of carbohydrates, lipids and proteins is discussed in Chapter 2. The chemical changes that occur to large, insoluble carbohydrates, lipids and proteins during digestion are summarised in Figure 7.2. Food cannot be said to have truly entered the body until it has been digested and the products absorbed across the gut wall.

The first stage in the breakdown of large, insoluble food molecules, then, is **mechanical digestion**. This occurs by the action of the jaws and teeth in the mouth, and then later, through the churning action of the muscular walls as the food is moved along the gut, particularly in the stomach. Throughout the gut, waves of contraction and relaxation of the circular and longitudinal muscles of the wall propel food along. This process is known as **peristalsis**. (Gut muscles are described as involuntary muscles because they are not under conscious control.)

2 a **Explain** why digestion is an essential stage in nutrition.
b **State** two essential foods that do not require digestion.
c **Identify** a polysaccharide the human gut cannot digest.

The **chemical digestion** that follows is brought about by enzymes. Digestive enzymes are protein catalysts produced in specialised cells in glands associated with the gut. Several different enzymes are secreted onto the food as it passes along the gut. The action of these enzymes very greatly speeds up the breakdown of insoluble food substances. They work efficiently at the relatively low temperature the body is maintained at. To complete the digestion processes, enzymes secreted onto the food work together with those held in the plasma membranes of cells of the gut lining. Examples of digestive enzymes are listed in Table 7.2

Table 7.2 Examples of digestive enzymes

Enzyme	Source	Substrate	Product	Optimum pH
amylase	salivary glands	starch	maltose	6.5–7.5
pepsin	gastric glands	protein	polypeptides	2.0
lipase	pancreas	triglyceride	fatty acids and glycerol	7.0

Digestion in the stomach

The human stomach is a J-shaped muscular bag located high in the abdominal cavity, below the diaphragm and liver. The function of this organ is to retain a meal while enzymic digestion of protein begins.

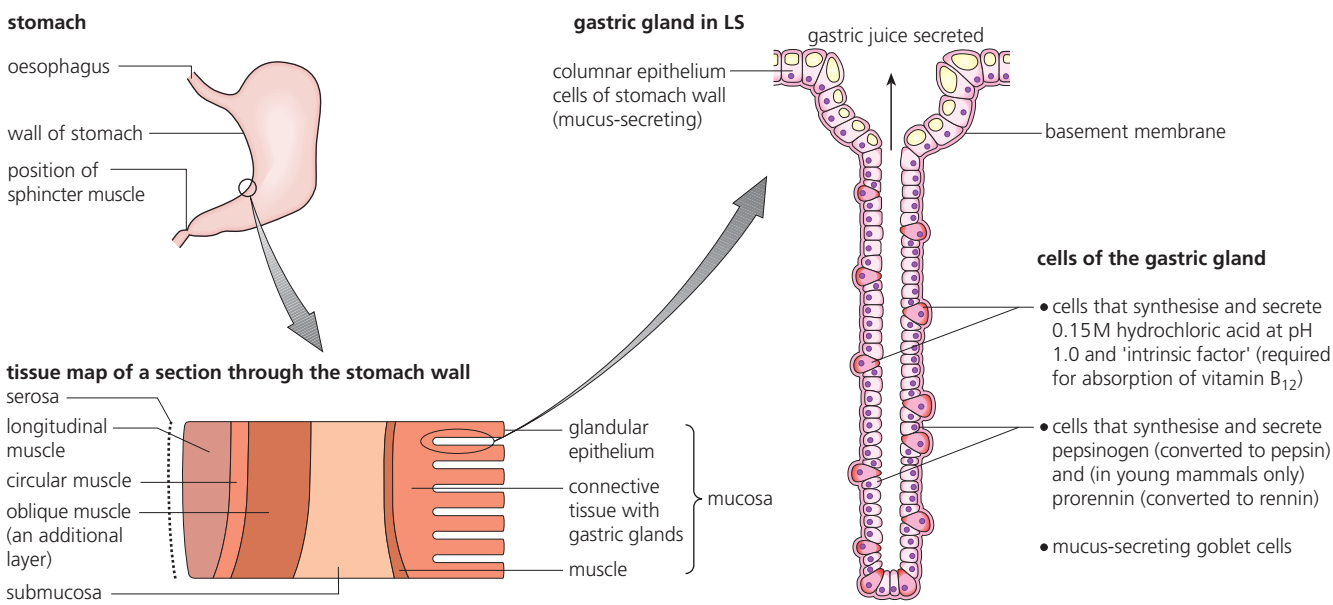


Figure 7.3 Gastric glands, structure and function

Present in the wall of the stomach are millions of tiny pits called **gastric glands** which secrete the components of **gastric juice** (Figure 7.3). This juice includes **hydrochloric acid** – sufficiently acidic to create an environment of pH 1.5–2.0, which is the optimum pH for protein digestion by the **protease enzymes** of the gastric juice. These proteases, of which **pepsin** is one (Table 7.2), are formed in cells of the gastric glands and secreted in an inactive state. The hydrochloric acid then activates them, and it also kills many bacteria present in the incoming food.

The whole stomach lining is supplied with **goblet cells** that secrete mucus. Mucus bathes the interior lining of the stomach, forming an effective barrier to both the hydrochloric acid and the proteases of the gastric juices, preventing **autolysis** (self-digestion) of the stomach wall.

3 **Predict** why it is essential that the protease enzymes of digestion are secreted in inactive forms.

As the food is mixed with gastric juice and churned by muscle action it becomes a semi-liquid called **chyme**. The churning action of the stomach is an important part of the mechanical digestion process. A typical meal may spend up to four hours in the stomach.

The roles of the small intestine

It is in the small intestine that completion of digestion of carbohydrates, lipids and proteins occurs, and the useful products of digestion are absorbed. In humans, the small intestine is about five metres long. Throughout its length, the innermost layer of the wall is formed into vast numbers of finger-like or leaf-like projections called villi.

Digestion in the small intestine

Food enters the first part of the small intestine (known as the **duodenum**) a little at a time. Here the chyme meets bile from the bile duct, and the pancreatic juice from the **pancreas**. Bile is strongly alkaline and neutralises the acidity of the chyme. It also lowers the surface tension of large fat globules, causing them to break into tiny droplets, a process called **emulsification**. This speeds digestion by the enzyme lipase later on. Bile itself contains no enzymes.

Pancreatic juice contains several enzymes, including an **amylase** that catalyses the hydrolysis of starch to maltose, a **lipase** that catalyses the hydrolysis of fats to fatty acids and glycerol, and **proteases** that hydrolyse proteins to polypeptides, and shorter-length peptides to free amino acids (Table 7.3).

All these enzymes act as the chyme, bile and pancreatic juice are mixed together by a churning action (a form of peristalsis) called segmentation (Figure 7.4).

Figure 7.4 Chyme meets bile and pancreatic juice

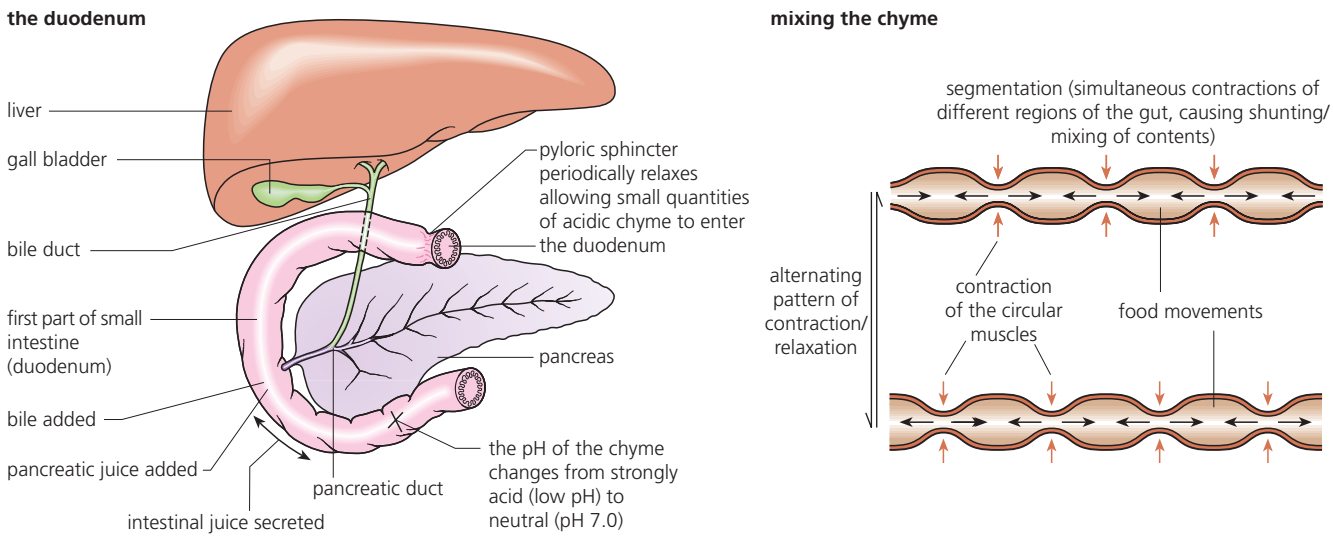


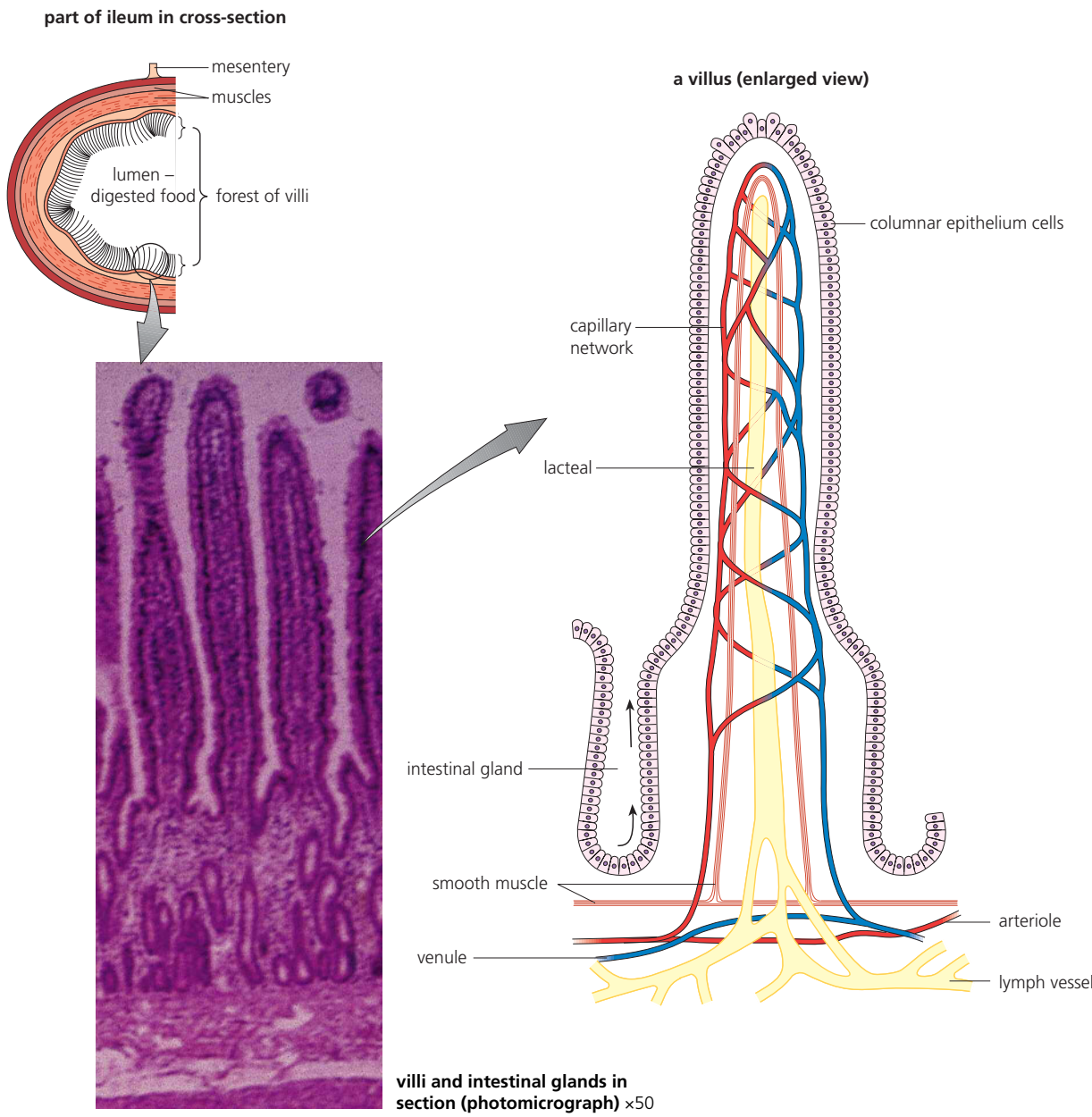
Table 7.3 The groups of hydrolytic enzymes operating in the small intestine

Enzyme	Substrate	Products
amylases	starch	maltose
lipases	lipids	fatty acids and glycerol
proteases	proteins + polypeptides	polypeptides + amino acids

Absorption in the small intestine

The products of digestion, mostly monosaccharide sugars, amino acids, fatty acids and glycerol, together with various vitamins and mineral ions, are absorbed as they make contact with the epithelial cells of the villi of the small intestine (Table 7.4). The process is efficient because the small intestine has a huge surface area, due to the vast number of **villi** (Figure 7.5). Epithelial cells expend energy in the **active transport** process by which most of the products of digestion are taken into the cells. Transport involves protein pump molecules in the plasma membrane, activated by reaction with ATP.

Figure 7.5 The small intestine – the absorption surface



Feature	Description/function
villi	provide a huge surface area for absorption
epithelium cells	single layer of small cells, packed with mitochondria – the source of ATP (metabolic energy) for active uptake across the plasma membrane
pump proteins in the plasma membrane of epithelial cells	actively transport nutrients across the plasma membrane into the villi
network of capillaries	large surface area for uptake of amino acids, monosaccharides, and fatty acids and glycerol into blood circulation
lacteal	branch of the lymphatic system into which triglycerides (combined with protein) pass for transport to body cells
mucus from goblet cells in epithelium	lubricates movement of digested food among the villi and protects plasma membrane of epithelial cells

Table 7.4 Functions of the absorption surface of the small intestine

Assimilation follows absorption

The fate of absorbed nutrients is termed **assimilation**. In the first stage of assimilation, absorbed nutrients are transported from the intestine. In the villi, **sugars** are passed into the capillary network, and from here they are transported to the liver. The liver maintains a constant level of blood sugar (page 222). The **amino acids** are also passed into the capillary network and transported to the liver. Here they contribute to the pool or reserves of amino acids from which new proteins are made in cells and tissues all over the body.

4 Distinguish between absorption and assimilation.

Finally, **lipids** are absorbed as fatty acids and glycerol and are largely absorbed into the lacteal vessels. From there they are carried by the lymphatic system to the blood circulation outside the heart.

Role of the large intestine

The large intestine wall has no villi, but the surface area for absorption is increased somewhat by numerous folds of the inner lining. At this point in the gut, most of the useful products of digestion have been absorbed. What remains is the undigested matter (such as plant fibre), mucus, dead intestinal cells, bacteria, some mineral ions and water. Water is an important component of our diet, and many litres of water are also secreted onto the chyme in the form of digestive juices.

In the colon, water and mineral salts (such as Na^+ and Cl^- ions) are absorbed. What remains of the meal is now referred to as faeces. Bacteria compose about 50% of the faeces, and some of these microorganisms produce gases and odoriferous substances. Bile pigments (excretory products formed from the routine breakdown of red cells), which were added in the duodenum, uniformly colour the faeces.

The rectum is a short muscular tube which terminates at the anus. Discharge of faeces from the body at the anus is controlled by sphincter muscles.

The transport system

6.2.1–6.2.7

Living cells require a supply of water and nutrients such as glucose and amino acids, and most need oxygen. The waste products of cellular metabolism have to be removed. In single-celled organisms and very small organisms, internal distances are small, so movements of nutrients can occur efficiently by **diffusion**, although some substances require to be transported across membranes by **active transport**. In larger organisms, these mechanisms alone are insufficient – an internal transport system is required to service the needs of the cells.

Internal transport systems at work are examples of **mass flow**. The more active an organism is, the more nutrients are likely to be required by cells, and so the greater is the need for an efficient system for internal transport. Larger animals have a **blood circulatory system** that links the parts of the body and makes resources available where they are required.

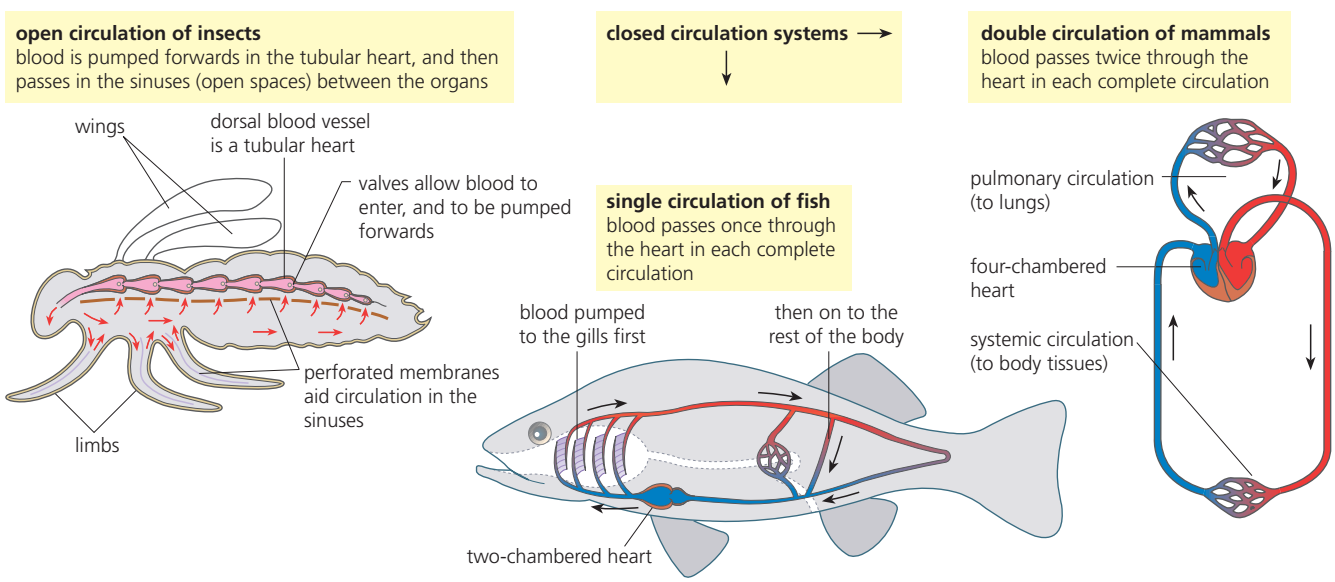
Transport in mammals

Mammals have a **closed circulation** in which blood is pumped by a powerful, muscular heart and circulated in a continuous system of tubes – the **arteries**, **veins** and **capillaries** – under pressure. The heart has four chambers, and is divided into right and left sides. Blood flows from the right side of the heart to the lungs, then back to the left side of the heart. From here it is pumped around the rest of the body and back to the right side of the heart. As the blood passes twice through the heart in every single circulation of the body, the system is called a **double circulation**.

The circulatory system of mammals is shown in Figure 7.6, alongside alternative systems. Looking at these other systems helps us to understand the features of the mammalian circulation. It becomes clear that the major advantages of the mammalian circulation are:

- simultaneous high-pressure delivery of oxygenated blood to all regions of the body;
- oxygenated blood reaches the respiring tissues, undiluted by deoxygenated blood.

Figure 7.6 Open and closed circulations



5 In an open circulation there is 'little control over circulation'. **Suggest** what this means.

The transport medium – the blood

Blood is a special tissue consisting of a liquid medium called **plasma** in which are suspended red cells or **erythrocytes**, white cells or **leucocytes**, and the platelets (Figures 7.7 and 7.8). The plasma is the medium for exchange of substances between cells and tissues, the erythrocytes are involved in transport of respiratory gases, and the leucocytes are adapted to combat infection. The roles of the components of blood are summarised in Table 7.5.

Figure 7.7 The composition of the blood

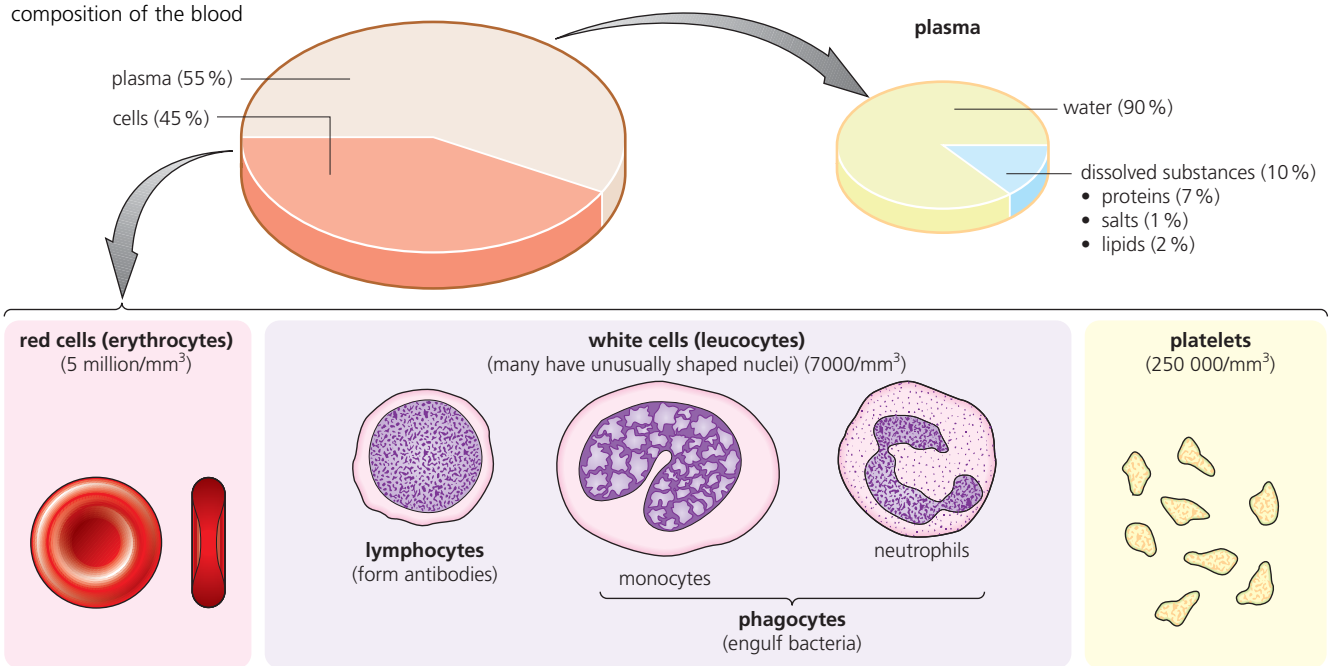


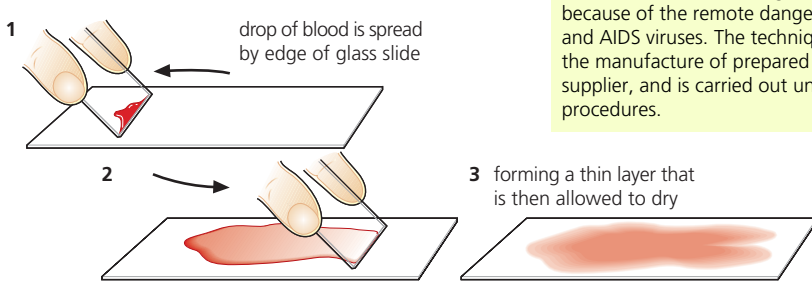
Figure 7.8 Examining a blood smear

preparing a blood smear for staining

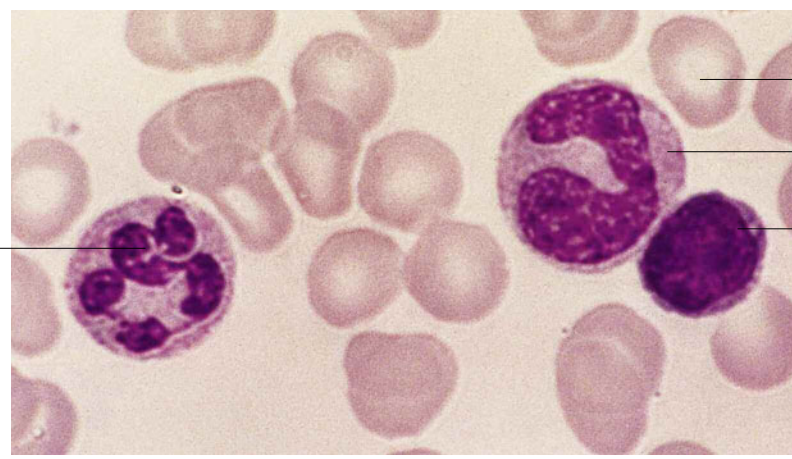
Once the smear has dried, it is usually double-stained.

The first stain is selected to dye the cytoplasm of white cells. The second stain will dye the nuclei. When coloured in this way, the few white cells in a blood sample are easy to find.

This technique might be used in a hospital pathology laboratory to search for abnormalities in the blood cells.



The taking of blood samples and the handling of human blood in school and college laboratories are **forbidden** because of the remote danger of infection from hepatitis and AIDS viruses. The technique shown here is used in the manufacture of prepared slides by a laboratory supplier, and is carried out under proper sterile procedures.



human blood smear stained to show the few white cells present (×4500)

Table 7.5 The components of the blood and their roles

Component	Role
plasma	transport of: ■ **nutrients** from gut or liver to all the cells ■ excretory products such as **urea** from the liver to the kidneys ■ **hormones** from the endocrine glands to all tissues and organs ■ **dissolved proteins** which have roles including regulating the osmotic concentration (water potential) of the blood ■ dissolved proteins which are **antibodies** ■ **heat** distribution to all tissues
erythrocytes	transport of: ■ **oxygen** from lungs to respiring cells ■ **carbon dioxide** from respiring cells to lungs (also carried in plasma)
lymphocytes	important in the **immune system** because they form antibodies
phagocytes	**ingest bacteria** or cell fragments
platelets	play a part in the **blood clotting mechanism**

6 Outline where in the body phagocytic leucocytes function.

The plumbing of the circulation system – arteries, veins and capillaries

There are three types of vessel in the circulation system:

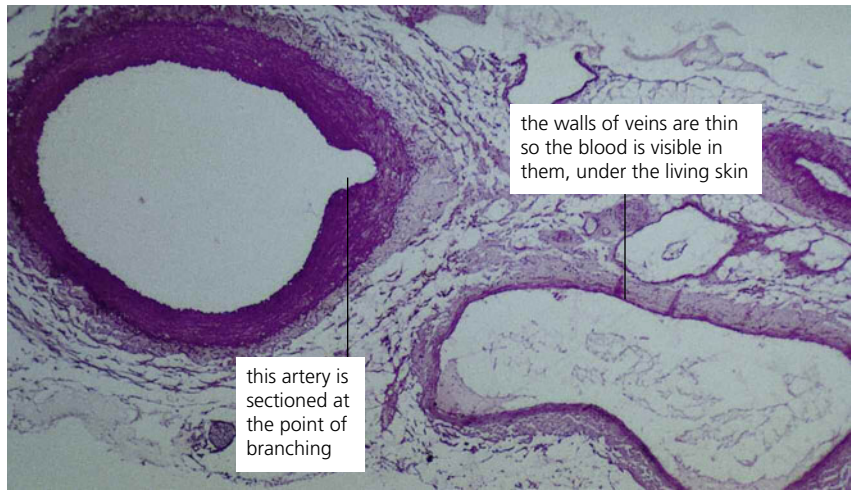
- **arteries**, which carry blood away from the heart;
- **veins**, which carry blood back to the heart;
- **capillaries**, which are fine networks of tiny tubes linking arteries and veins.

Both arteries and veins have strong, elastic walls, but the walls of the arteries are very much thicker and stronger than those of the veins. The strength of the walls comes from the collagen fibres present, and the elasticity is due to the elastic and involuntary (smooth) muscle fibres. The walls of the capillaries, on the other hand, consist of endothelium only (endothelium is the innermost lining layer of arteries and veins). Capillaries branch profusely and bring the blood circulation close to cells – no cell is far from a capillary.

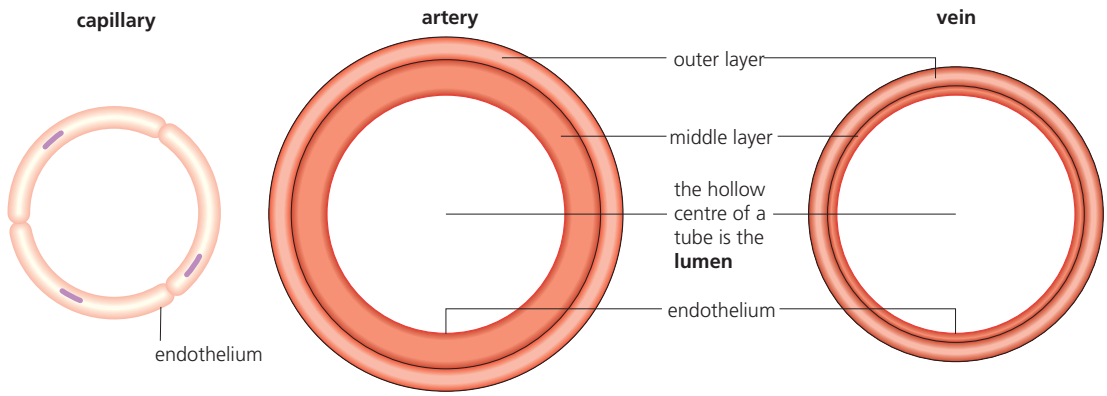
Blood leaving the heart is under high pressure, and travels in waves or **pulses**, following each heart beat. By the time the blood has reached the capillaries it is under very much lower pressure, without a pulse. This difference in blood pressure accounts for the differences in the walls of arteries and veins (Table 7.6). Figure 7.9 shows an artery, vein and capillary vessel in section, and details the wall structure of these three vessels. Because of the low pressure in veins there is a possibility of backflow here. Veins have **valves** at intervals that prevent this (Figure 7.10).

Figure 7.9 The structure of the walls of arteries, veins and capillaries

TS artery and vein, LP (×20)



In sectioned material (as here), veins are more likely to appear squashed, whereas arteries are circular in section.



	Capillary – site of exchange between blood and body tissues	Artery – carries blood under high pressure away from the heart	Vein – carries blood under low pressure back to the heart
Outer layer (collagen fibres)	absent	present	present
Middle layer (elastic fibres and involuntary muscle fibres)	absent	thick layer	thin layer
Inner layer or endothelium (pavement epithelium)	present	present	present
Valves	absent	absent	present

Table 7.6 Structural differences between arteries, veins and capillaries

Valves are especially common in the veins of the limbs.

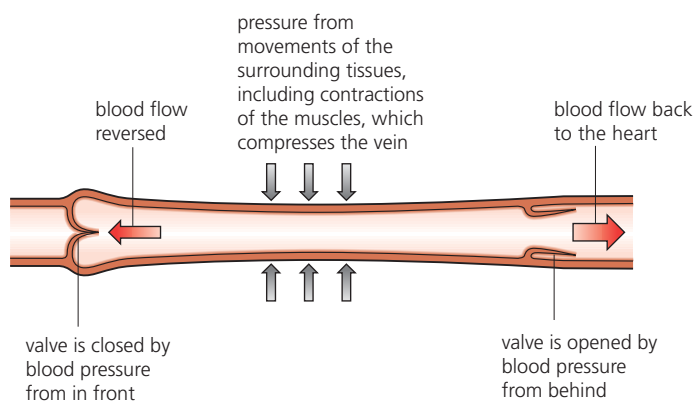


Figure 7.10 The valves in veins

The arrangement of arteries and veins

We have already noted that mammals have a double circulation. It is the role of the right side of the heart to pump deoxygenated blood to the lungs, and the arteries, veins and capillaries serving the lungs are known as the **pulmonary circulation**. The left side of the heart pumps oxygenated blood to the rest of the body, and the arteries, veins and capillaries serving the body are known as the **systemic circulation**.

In the systemic circulation, organs are supplied with blood by an artery branching from the main **aorta**. Within the organ, the artery branches into numerous arterioles (smaller arteries), and the smallest arterioles supply the capillary networks. Capillaries drain into venules (smaller veins), and venules join to form veins. The veins join the **vena cava** carrying blood back to the heart. The branching sequence in the circulation is, therefore:

aorta → artery → arteriole → capillary → venule → vein → vena cava

Arteries and veins are often named after the organs they serve (Figure 7.11). The blood supply to the liver is via the hepatic artery, but the liver also receives blood directly from the small intestine, via a vein called the **hepatic portal vein**. This brings much of the products of digestion, after they have been absorbed into the capillaries of the villi.

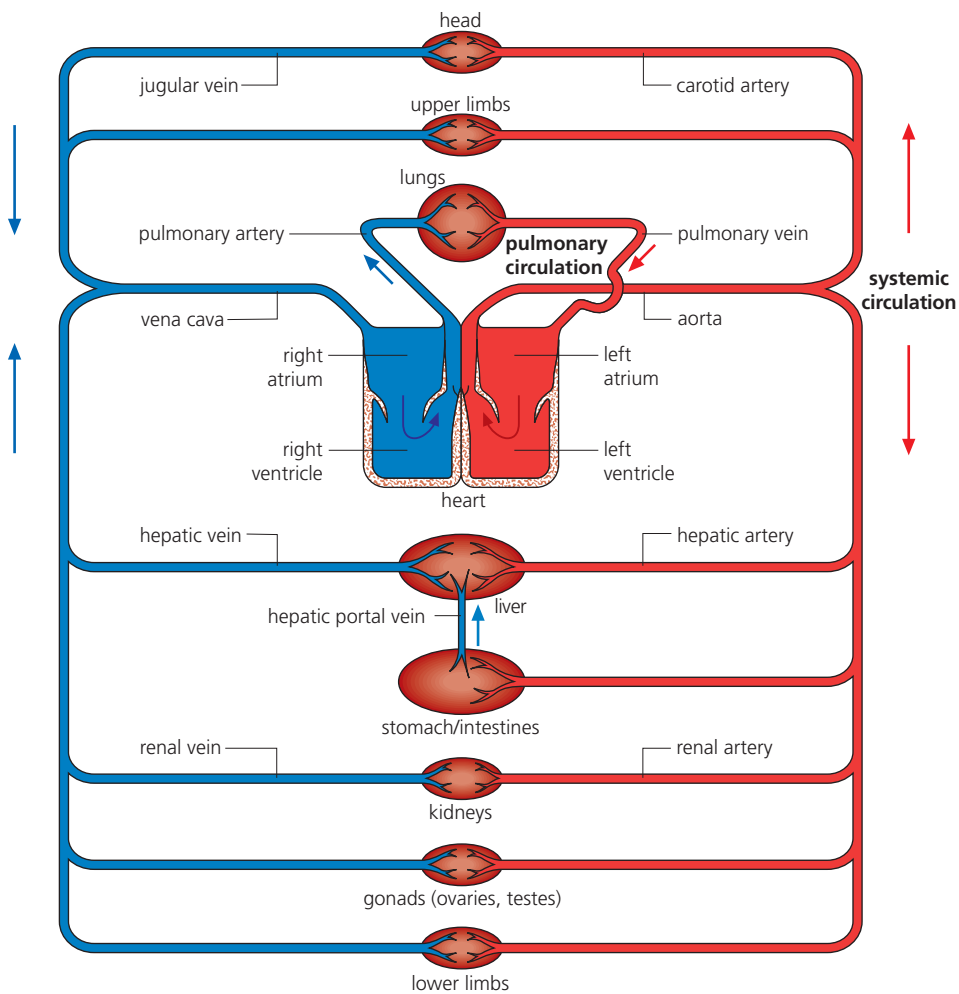


Figure 7.11 The layout of the human circulation

The circulatory system as shown here is simplified, e.g. limbs, lungs, kidneys and gonads are paired structures in the body.

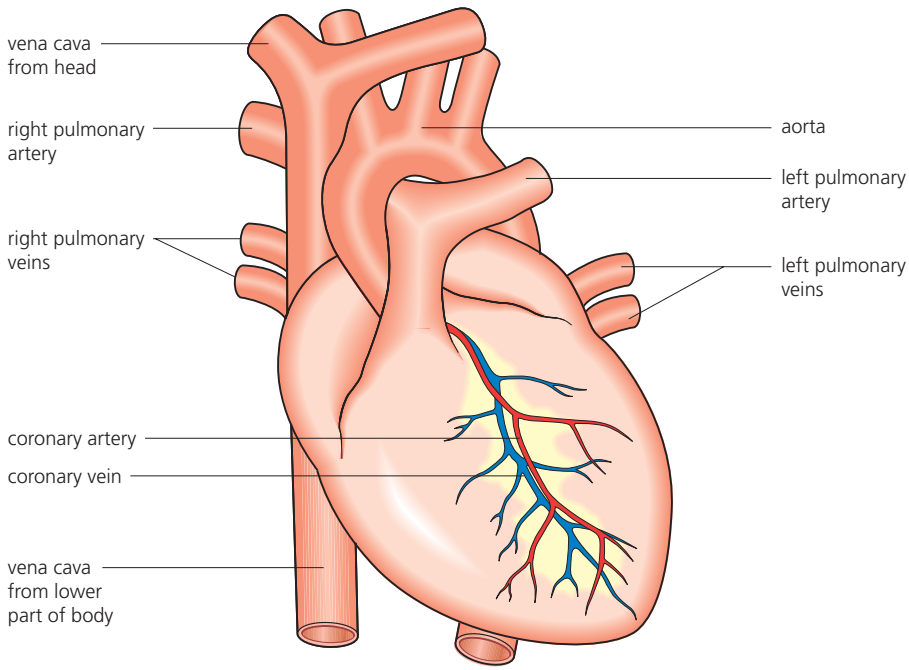
The heart as a pump

The human heart is the size of a clenched fist. It is found in the thorax between the lungs and beneath the breast bone (sternum). The heart is a hollow organ with a muscular wall, and is contained in a tightly fitting membrane, the pericardium – a strong, non-elastic sac that anchors the heart within the thorax.

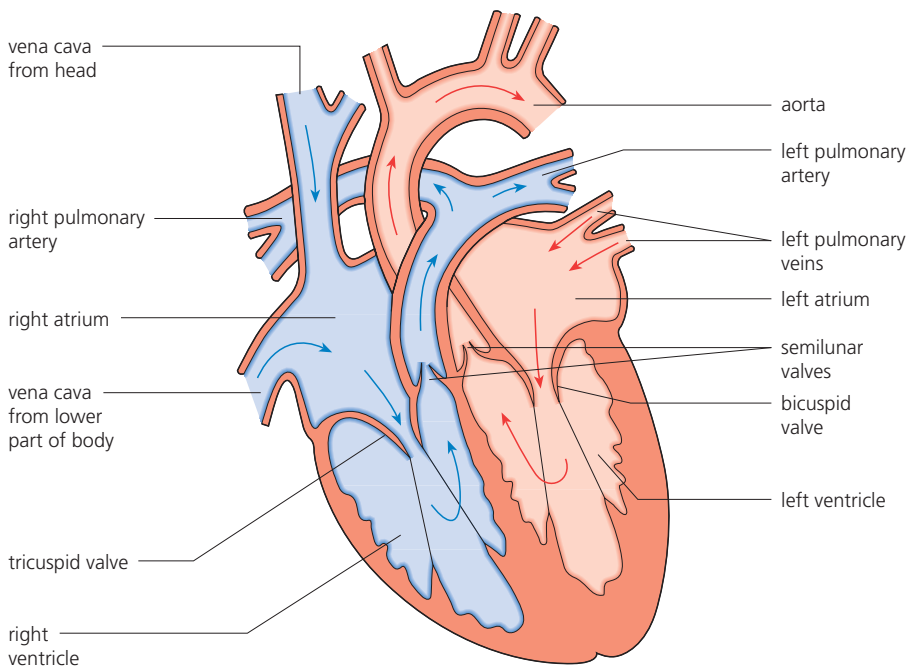
The cavity of the heart is divided into four chambers, with those on the right side of the heart completely separate from those on the left. The two upper chambers are thin-walled **atria** (singular, **atrium**). These receive blood into the heart. The two lower chambers are thick-walled

Figure 7.12 The structure of the heart

heart viewed from the front of the body with pericardium removed



heart in LS



ventricles, with the muscular wall of the left ventricle much thicker than that of the right ventricle. However, the volumes of the right and left sides (the quantities of blood they contain) are identical. The ventricles pump blood out of the heart.

Note that the walls of the heart (the heart muscle) are supplied with oxygenated blood via **coronary arteries** (Figure 7.12). These arteries, and the capillaries they serve, deliver to the muscle fibres of the heart the oxygen and nutrients essential for the pumping action.

The **valves** of the heart prevent backflow of the blood, thereby maintaining the direction of flow through the heart. You can see these valves in action, in the lower drawing in Figure 7.12. The **atrio-ventricular valves** are large valves, positioned to prevent backflow from ventricles to atria. The edges of these valves are supported by tendons anchored to the muscle walls of the ventricles below, and which prevent the valves from folding back due to the (huge) pressure that develops here with each heart beat. Actually, these atrio-ventricular valves are individually named: on the right side is the **tricuspid valve**, on the left is the **bicuspid** or mitral valve.

A different type of valve separates the ventricles from pulmonary artery (right side) and aorta (left side). These are pocket-like structures called **semilunar valves**, rather similar to the valves seen in veins. These cut out backflow from aorta and pulmonary artery into the ventricles as the ventricles relax between heart beats.

The action of the heart

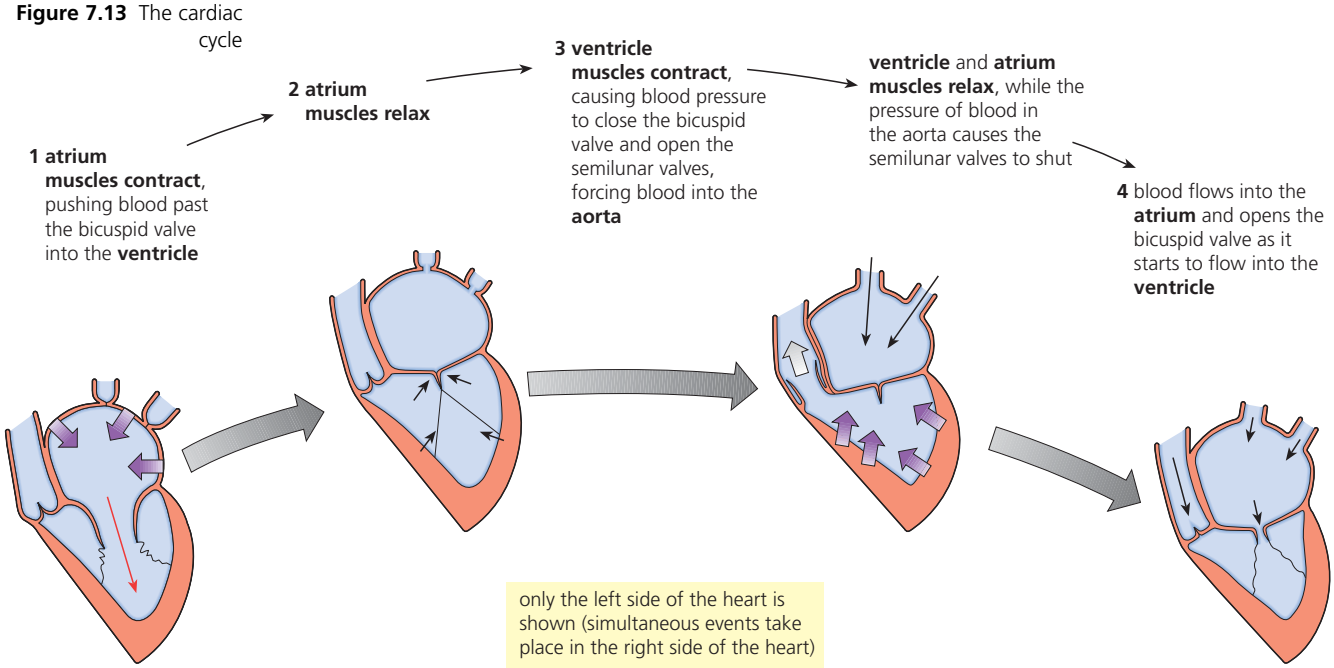
The heart normally beats about 75 times per minute – approximately 0.8 seconds per beat. In each beat, the heart muscle contracts strongly, followed by a period of relaxation. As the muscular walls of a chamber of the heart contract, the volume of that chamber decreases. This increases the pressure on the blood contained there, forcing the blood to a region where pressure is lower. Since the valves prevent blood flowing backwards, blood consistently flows on through the heart.

Look at the steps of contraction and relaxation in Figure 7.13, illustrated in the left side of the heart. (Both sides function together, in exactly the same way, of course.)

We start at the point where the **atrium contracts**. Blood is pushed into the ventricles (where the contents are under low pressure) by contraction of the walls of the atrium. This contraction also prevents backflow by blocking off the veins which brought the blood to the heart.

The **atrium now relaxes**.

Figure 7.13 The cardiac cycle



Next the **ventricle contracts**, and contraction of the ventricle is very forceful indeed. The high pressure this generates slams shut the atrio-ventricular valve and opens the semilunar valves, forcing blood into the aorta. A pulse, detectable in arteries all over the body, is generated.

This is followed by **relaxation of the ventricles**. Each contraction of cardiac muscle is followed by relaxation and elastic recoil.

7 The edges of the atrio-ventricular valves have non-elastic strands attached, which are anchored to the ventricle walls (Figure 7.12). **Suggest** the role of these strands.

Origin of the heart beat

The heart beats rhythmically throughout life, without rest, apart from the momentary relaxation between beats. Even more remarkably, the origin of each beat is within the heart itself – we say the heart beat is **myogenic** in origin.

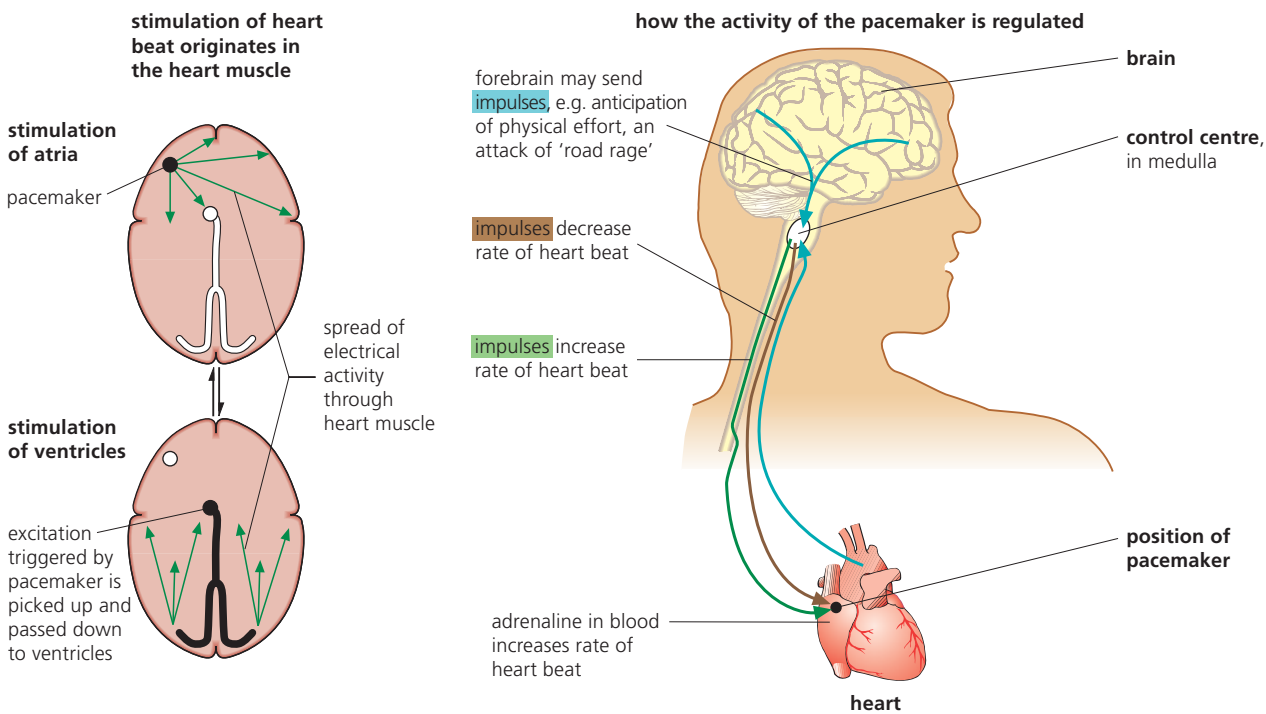
Beats originate in a structure in the muscle of the wall of the right atrium, called the **pacemaker** (Figure 7.14). Special muscle fibres radiate out from the pacemaker, conducting the impulse to the muscles of both atria, triggering contraction there. Then a second structure picks up the excitation and passes it to the ventricles. Now ventricular contraction is triggered. Then, after every contraction, cardiac muscle has a period of insensitivity to stimulation (in effect, a period of enforced non-contraction), when the heart refills with blood. This period is a relatively long one in heart muscle, and is undoubtedly important in enabling the heart to beat throughout life.

The heart's natural rhythm, set by the pacemaker, is about 50 beats per minute, but we are well aware that conditions in our body override this basic rate and increase heart performance.

How is the pacemaker regulated?

Nervous control of the heart is by reflex action – it is an involuntary response, not under our conscious control. The heart receives impulses from a **control centre** in the hindbrain (medulla), via two nerves. One nerve, when stimulated, triggers speeding up of the heart rate, and the other nerve triggers a slowing down of the heart (Figure 7.14). Since these nerves have opposite effects we say they are **antagonistic**.

Figure 7.14 The pacemaker and control of heart rate



Nerves supplying the control centre in the hindbrain bring impulses from strategically placed stretch receptors located in the walls of the aorta and various arteries, and indeed from the heart wall itself. As a result, when blood pressure is high in the arteries, the rate of heart beat is lowered by impulses from the control centre. When blood pressure is low, the rate of heart beat is increased. The rate of heart beat is also influenced by impulses from the higher centres of the brain – for example, emotion, stress and anticipation of events can all cause impulses that speed up heart rate.

8 Suggest likely conditions or situations in which the body is likely to secrete adrenaline.

In addition, the hormone **adrenaline**, which is secreted by the adrenal glands and carried in the blood, causes the pacemaker to increase the heart rate.

The roles of the blood circulation system

The blood circulation has roles in the body's defence against diseases (see below) as well as being the all-important transport system of the body (Table 7.7). Nutrients from digestion, oxygen and carbon dioxide, urea, hormones and antibodies are all transported. In the tissues of the body, exchange between the blood and cells of the tissues occurs from the capillaries, the walls of which are permeable and highly 'leaky'.

Function	Transport of
tissue respiration	**O_2** from lungs to all tissues **CO_2** back to the lungs
hydration	**water** to all the tissues
nutrition	**nutrients** (sugars, amino acids, lipids, vitamins) and inorganic ions to all cells
excretion	waste product **urea** to kidneys
development and co-ordination	**hormones** from endocrine glands to target organs
temperature regulation	distribution of **heat**
defence against disease	**antibodies** are circulated in the blood stream

Table 7.7 Transport roles of the blood circulation

Defence against infectious disease

6.3.1–6.3.8

Infectious disease is caused when another organism or virus invades the body and lives there parasitically. The invader is known as a **pathogen** and the infected organism – a human in this case – is the **host**.

A pathogen is an organism or virus that causes a disease. Most, but not all, are microorganisms.

Pathogens may pass from diseased host to healthy organisms, so these diseases are known as infectious or **communicable diseases**. Disease in general is defined as an 'unhealthy condition of the body', and this rather broad definition includes some distinctly different forms of ill-health. For example, ill-health may be caused by unfavourable environmental conditions. Diseases of this type are non-infectious or **non-communicable** diseases, and they include conditions such as cardiovascular disease, malnutrition and cancer.

Good health is more than the absence of harmful effects of a disease. This point is emphasised by the **World Health Organization** (**WHO**) which identifies health as 'a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity'.

Pathogens and disease

The range of **disease-causing organisms** that may infect humans includes not only microorganisms such as certain **bacteria** and **fungi**, but also **viruses**, some **protozoa** (single-celled animals) and certain non-vertebrate animals in the phyla of **flatworms** and of **roundworms** (pages 170–1).

9 **State** the differences in structure between a bacterial cell and a virus.

Not all bacteria or fungi are parasitic and pathogenic – only relatively few species are. On the other hand, no virus can function outside a host organism, so we can say that all viruses are parasitic. Remember, a virus consists of a nucleic acid core surrounded by a protein coat (page 19). Viruses, once introduced into a host cell, take over the machinery of protein and nucleic acid synthesis, and coerce their host cells to manufacture more virus components and assemble them.

Antibiotics against infection

Many bacterial diseases of humans and other animals can be successfully treated with antibiotics. **Antibiotics** are naturally occurring substances that slow down or kill microorganisms. They are obtained mainly from fungi or bacteria, and are substances which such organisms manufacture in their natural habitats. An antibiotic, when present in low concentrations, may inhibit the growth of other microorganisms.

Since their original discovery, over 4000 different antibiotics have been isolated, but only about 50 have proved to be safe to use as drugs. The antibiotics which are effective over a wide range of pathogenic organisms are called **broad spectrum antibiotics**. Others are effective with just a few pathogens. Many antibiotics in use today have been synthesised.

How do antibiotics work?

Most antibiotics are so extremely effective at disrupting bacterial metabolism that whole populations are quickly suppressed. Figure 7.15 shows an actively dividing bacterial cell. A key process here is the formation and laying down of the new dividing wall. Bacterial cells have a rigid wall containing giant, complex molecules of amino-sugars and peptide units (polymers) rather than the polysaccharides of plant cell walls. Some bacteria have additional wall layers.

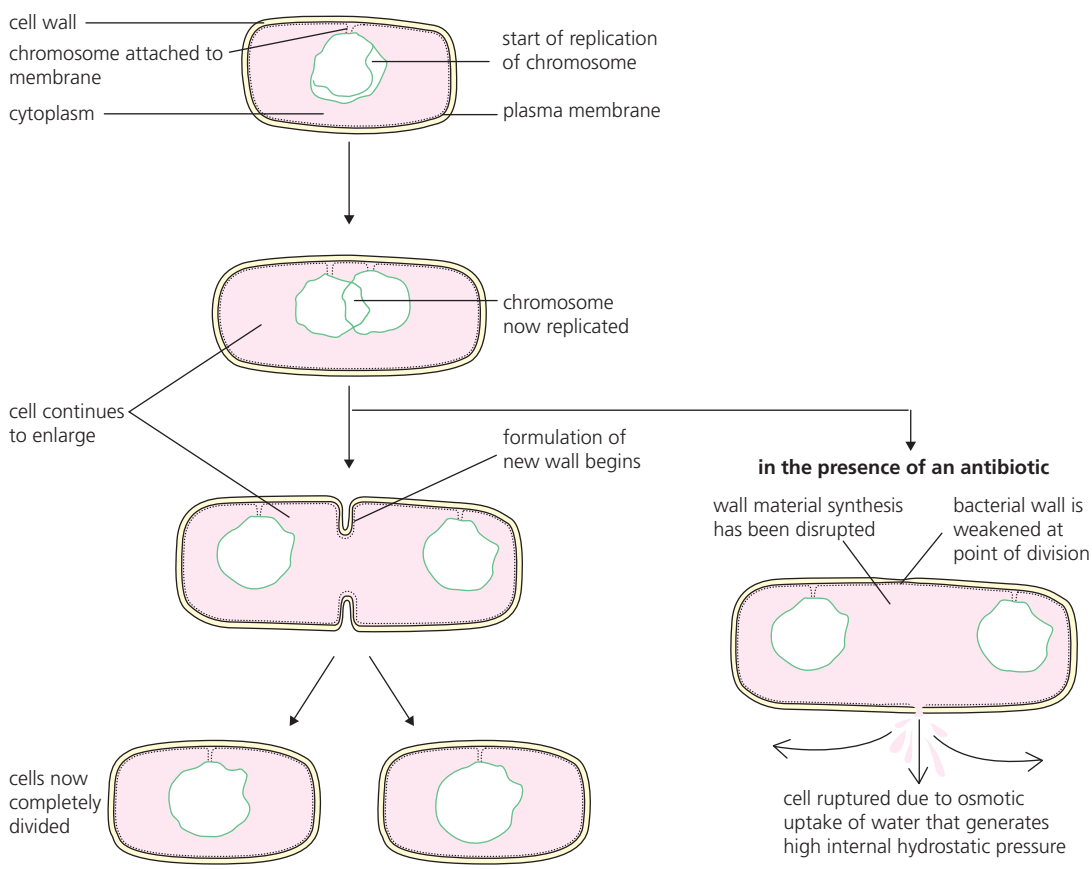


Figure 7.15 How antibiotics work

In division and growth, bacteria are vulnerable to antibiotic action because the antibiotic molecules enter the bacterial cells and disrupt the metabolic reactions of wall formation and growth.

10 Explain why antibiotics are effective against bacteria but not viruses.

Other bacteria have dormant spores whose growth is stopped by antibiotics, by similar mechanisms.

Viruses, on the other hand, are not living cells and have no metabolism of their own to be interfered with. Viruses reproduce using metabolic pathways in their host cell that are not affected by antibiotics. Antibiotics cannot be used to prevent viral diseases.

The body's defence against infectious disease

The **first line of defence** is the skin. Both our skin and the internal linings of lungs, trachea and gut are potential ports for entry to the body's tissues and organs by pathogens. Not surprisingly, protective measures have evolved at these surfaces.

The **external skin** is covered by keratinised protein of the dead cells of the epidermis (Figure 7.43, page 221). This is a tough and impervious layer, and an effective barrier to most organisms unless the surface is broken, cut or deeply scratched. However, folds or creases in the

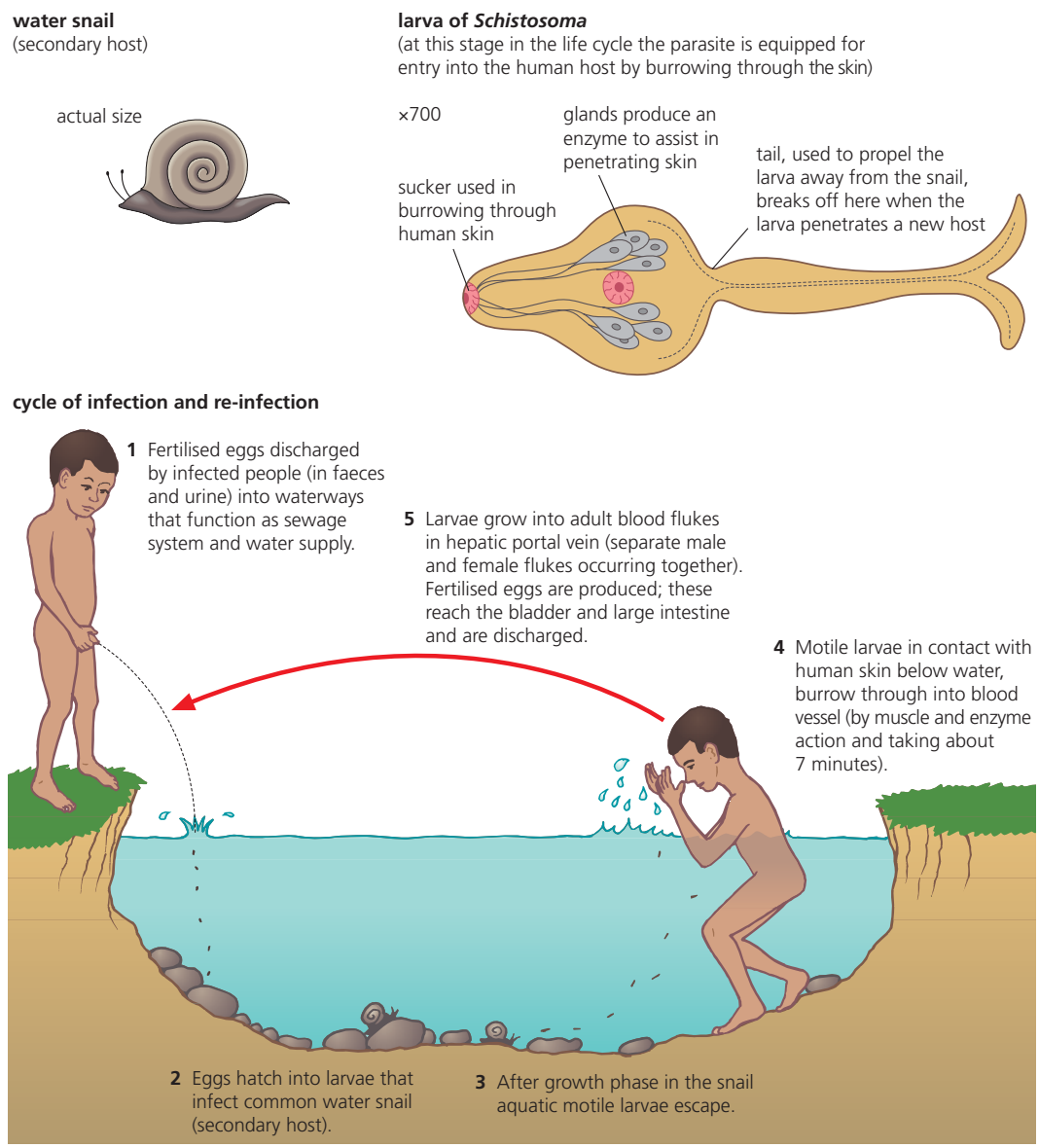


Figure 7.16 The life cycle of *Schistosoma* and the spread of schistosomiasis

skin that are permanently moist can harbour microorganisms that degrade the barrier and cause infection, as in athlete's foot. Also, the aquatic larvae of the pathogenic flatworm *Schistosoma*, known as a blood worm, may burrow through, when people bathe in infected water (Figure 7.16). Clearly, the body requires additional lines of defence.

11 Explain how the skin acts as a barrier against many pathogens, yet is often traversed by *Schistosoma*.

The **internal surfaces** of our breathing apparatus (the trachea, bronchi and the bronchioles) and of the gut are all lined by moist epithelial cells. These vulnerable internal barriers are protected by the secretion of copious quantities of **mucus**, and by the actions of **cilia** removing the mucus.

12 Suggest how mucus secreted by the lungs may protect lung tissue.

Cilia are organelles that project from the surface of certain cells. Cilia occur in large numbers on the lining (epithelium) of the air tubes (bronchi) serving the lungs where they sweep the fluid mucus across the epithelial surface, away from the delicate air sacs of the lungs.

In the gut, digestive enzymes provide some protection, as does the strong **acid secreted in the stomach** on arrival of food (page 181).

However, all these internal barriers may be crossed by certain pathogens – and often are. It is fortunate there are internal lines of defence too.

First, the body responds to localised damage by **inflammation**. Inflammation is the initial, rapid, localised response the tissues make to damage (cuts, scratches, bruising, or deep wounds). The site becomes swollen, warm and painful. The volume of blood is increased at the damaged site. **Leucocytes (white cells)** escape from the enlarged capillaries with plasma, and function in the tissue fluid, especially at the sites of inflammation. These cells originate in the bone marrow and are initially distributed by the blood circulation all over the body.

The leucocytes of the body and blood circulation play a complex part in the resistance to infection – their roles are detailed below.

Meanwhile, if a blood vessel is ruptured when the skin is damaged, then the **blood clotting mechanism** is activated, so that the vessel is quickly sealed off.

Phagocytic leucocytes

Some of the leucocytes have the role of engulfing foreign material including invading bacterial cells. Certain of these leucocytes are short-lived cells of the plasma. Others are the long-lived, rubbish-collecting cells found throughout the body tissues. Both types of cell take up material into their cytoplasm, much as the protozoan *Amoeba* is observed to feed, by a mechanism known as **phagocytosis** (Figure 7.17). Once inside the cell, the material taken up is destroyed in a controlled way by the activity of lysosomes.

Immunity: lymphocytes and the antigen–antibody reaction

Other leucocytes are responsible for the body's **antibody reaction** to infection or invasion of foreign matter – this is the basis of our immune system.

Immunity is based on the body's ability to recognise 'self' (body cells and specific proteins, for example) and to distinguish 'self' from 'non-self' substances (produced by an invading organism, for example). Any 'non-self' substance is called an **antigen**. This ability to recognise antigens and to take steps to overcome them is the property of special white cells called **lymphocytes**. Each type of lymphocyte recognises one specific antigen. In the presence of an antigen, these cells divide rapidly, producing many cells – known as a clone. These cloned lymphocytes then secrete the antibody specific to that antigen.

An **antibody** is a special kind of protein constructed in the shape of a Y, whose top contains an antigen-binding site (Figure 7.18). Millions of different types of antibodies are produced – as many as there are types of foreign matter (antigens) invading the body. The amino acid sequence of the antigen-binding site differs according to the chemistry of the antibody it binds with. It is the antigen-binding site that gives each antibody its specificity.

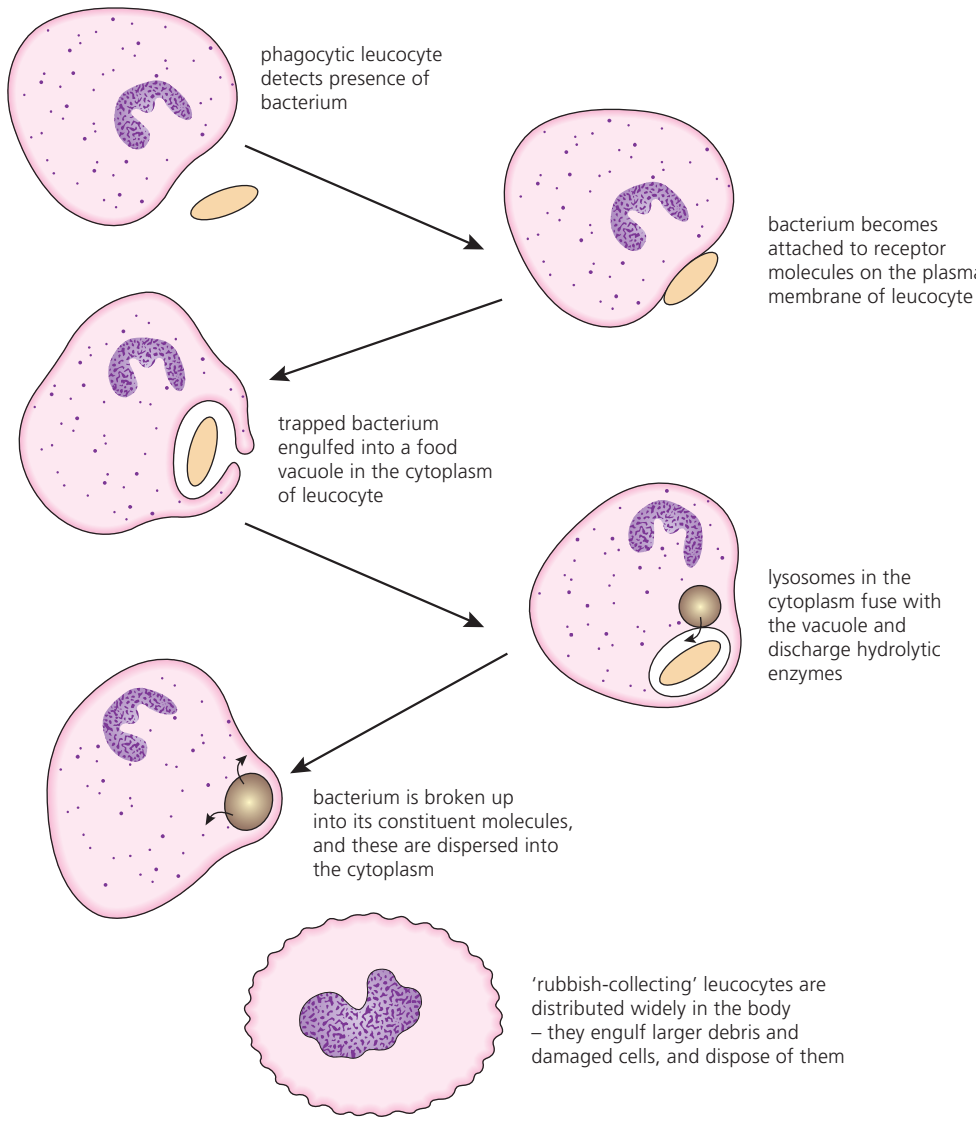


Figure 7.17 Phagocytosis of a bacterium

So a huge range of **different antibody-secreting lymphocytes** exists, each type recognising one specific antigen. The more antigens we encounter, the more antibodies we are able to form, should they be required. In the presence of an invasion by an antigen, the immune system responds: the appropriate antibody-secreting lymphocytes rapidly divide, and sufficient antibodies necessary to overcome the antigen are secreted. However, this type of lymphocyte also has a short life span. Once the harmful effects of the invading antigen are neutralised, the respective lymphocytes disappear from the blood circulation. But 'knowledge' of the antigen is not lost, because lymphocytes we call **memory cells** remain behind, stored in the lymph nodes. With the aid of our memory cells, our body can respond rapidly, if the same antigen re-invades. We say we have **immunity** to that antigen.

Antibodies destroy antigens in different ways. Antigens that are toxins may be inactivated by reaction with the antibody, and bacterial cells may be clumped together so they 'precipitate' and can be engulfed by phagocytic cells. Antibodies may attach to foreign matter, ensuring its recognition by phagocytic cells. Antibodies may also act by destroying bacterial cell walls, causing lysis of the bacterium.

13 Deduce which synthetic machinery of the lymphocytes is active in the production of the components of their antibodies.

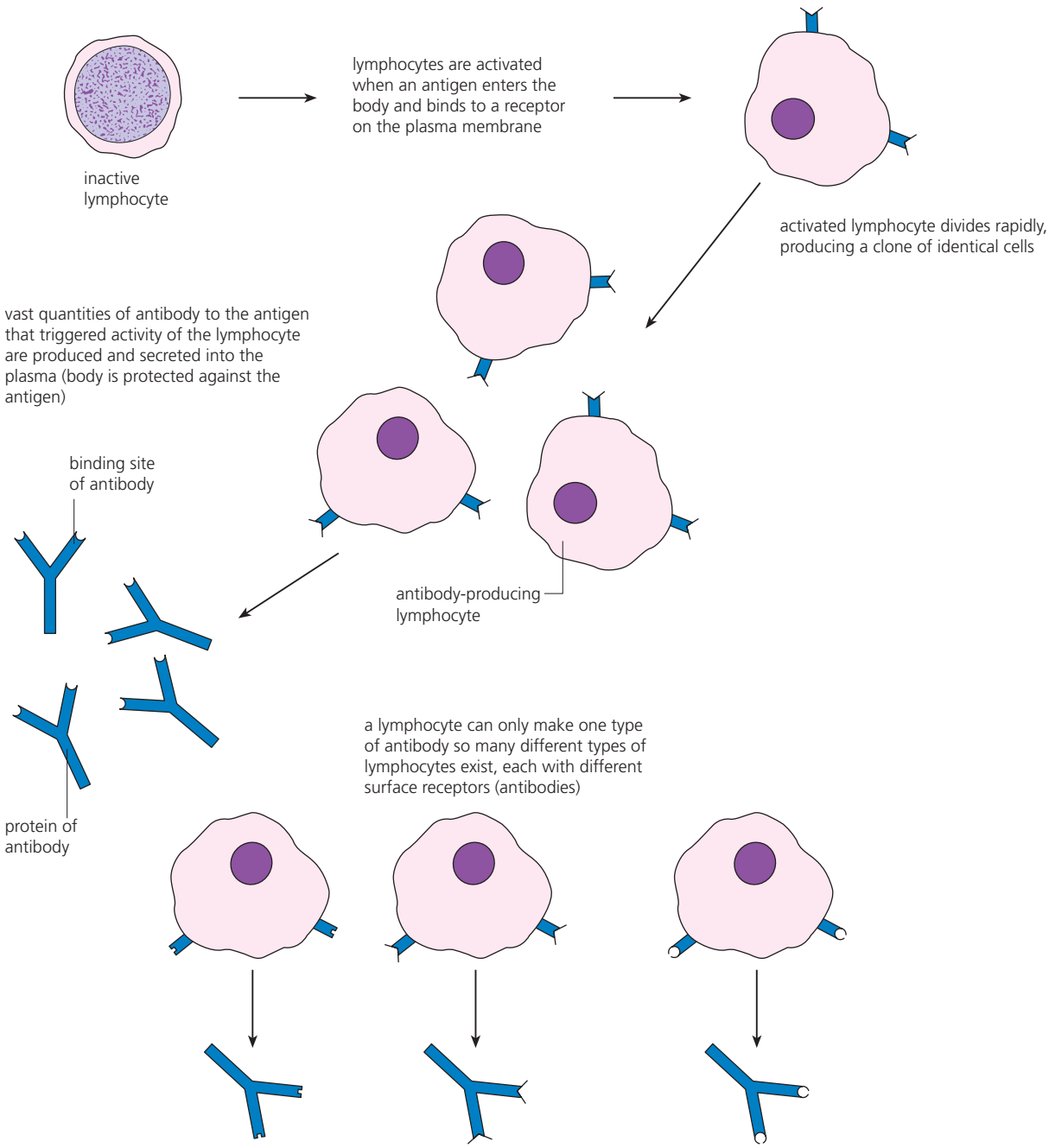


Figure 7.18 Formation and function of an antibody

Human immunodeficiency virus (HIV) and AIDS

Human immunodeficiency virus (**HIV**) was first identified in 1983 as the cause of a disease of the human immune system known as **acquired immune deficiency syndrome** (**AIDS**).

HIV is a tiny virus, less than 0.1 μm in diameter (Figure 7.19). It consists of two single strands of RNA which, together with enzymes, are enclosed by a protein coat. A membrane derived from the human host cell in which the virus was formed encapsulates each new virus particle leaving the host cell.

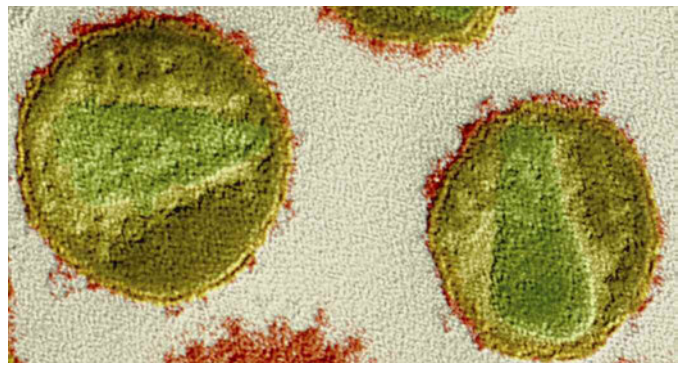
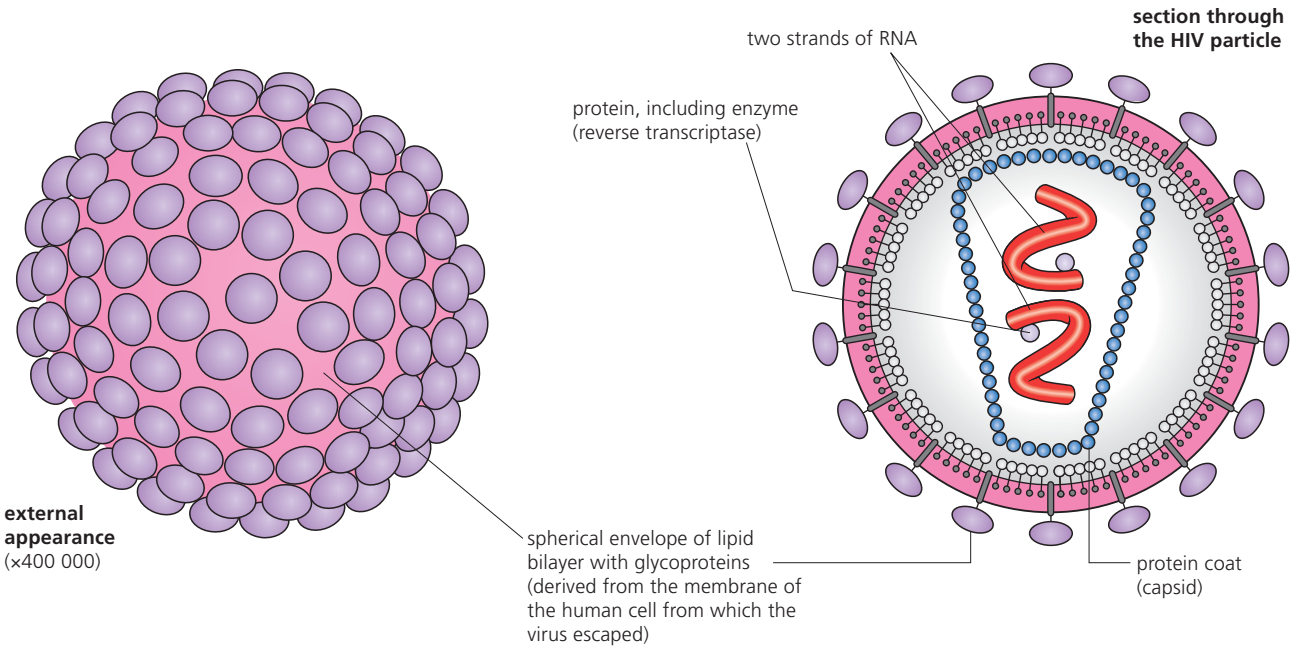


Figure 7.19 The human immunodeficiency virus (HIV)

TEM of HIV (×300 000)



HIV was first detected in central Africa in the 1950s, perhaps as a mutation of a similar virus present in African green monkeys (but other theories about the origin of HIV exist). From Africa, HIV was spread to the Caribbean and later to the USA and Europe. Now, HIV and AIDS occur worldwide. AIDS is probably already the greatest current threat to public health because it kills people in the most productive stage of their lives.

HIV is a retrovirus

The word 'retrovirus' is not self-explanatory; the term needs explaining. A **retrovirus** reverses the normal flow of genetic information, which is from the DNA of genes to messenger RNA in the cytoplasm (page 69). The idea that information always flows in this direction in cells is called the **central dogma of cell biology**. However, in retroviruses the information in RNA in the cytoplasm is translated into DNA within a host cell, and then becomes attached to the DNA of a chromosome in the host's nucleus.

How does a retrovirus work?

Taking HIV as an example, the virus binds to a host cell (lymphocyte) membrane (Figure 7.20), and the core of the virus passes inside. Inside the host cell, the RNA and virus enzymes are released. One enzyme from the virus is called **reverse transcriptase**. This enzyme catalyses the copying of the genetic code of each of the virus's RNA strands into a DNA double helix. This DNA then enters the host nucleus and is 'spliced' into the host's DNA of a chromosome (Figure 7.21). Here it may be replicated with the host's genes every time the host cell divides. The viral genes remain **latent**, giving no sign of their presence in the host cells at this stage.

Figure 7.20 HIV infection of a white cell

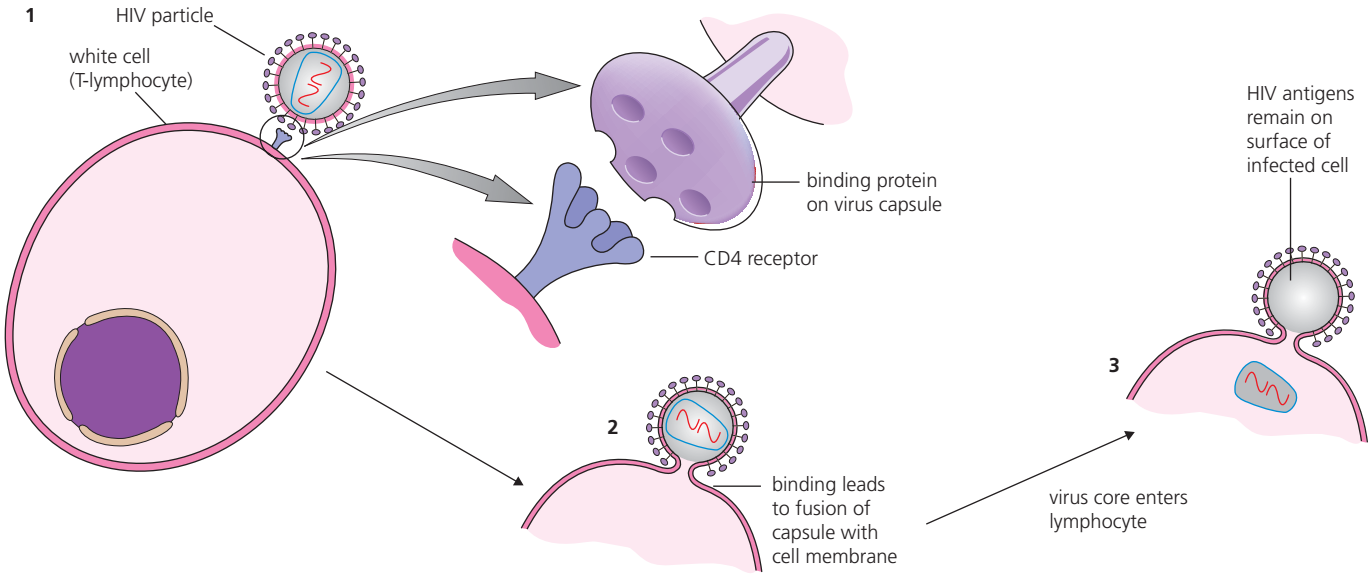
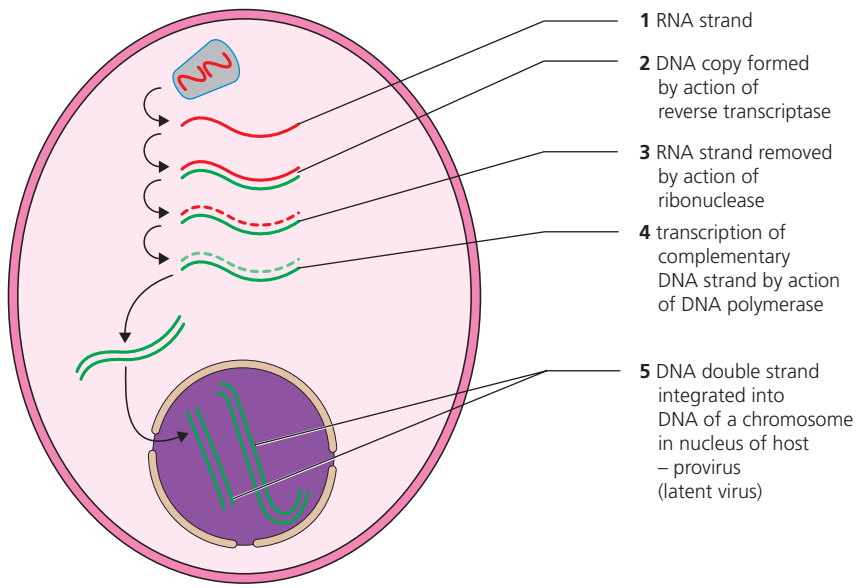


Figure 7.21 How HIV becomes part of the white cell's genome



The onset of AIDS

At a later stage, some event in the patient's body activates the HIV genes and the outcome is AIDS. The average interval between HIV infection and the onset of AIDS is about 8–10 years. The result of activation is the synthesis of viral messenger RNA. This then passes out into the cytoplasm, and there codes for viral proteins (enzymes and protein coat) at the ribosomes. Viral RNA (single-stranded), enzymes and coat protein are formed into viral cores. These move against the cell membrane and 'bud-off' new viruses (Figure 7.22). These infect more lymphocytes and the cycle is rapidly repeated. Without treatment, this process causes the body's reserve of lymphocytes (T-cells) to decrease very quickly. Eventually, no infection, however trivial, can be resisted; death follows (Figure 7.23).

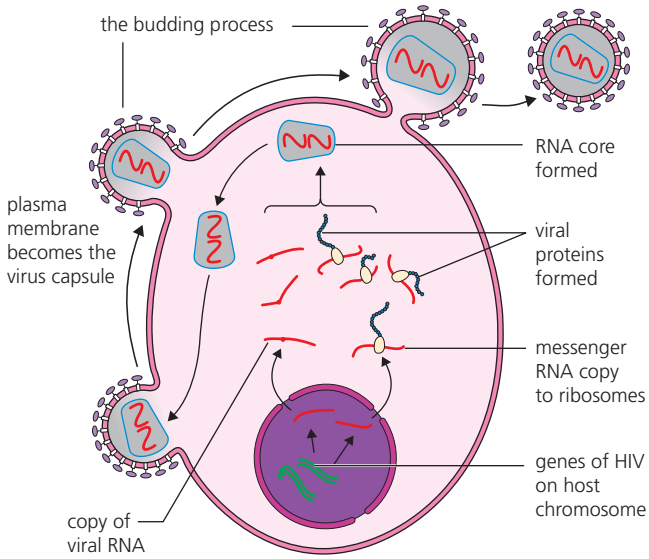


Figure 7.22 Activation of the HIV genome and production of new HIV

viral proteins and RNA being synthesised and new HIV cores assembled, then budding off with part of the host plasma membrane as each capsule

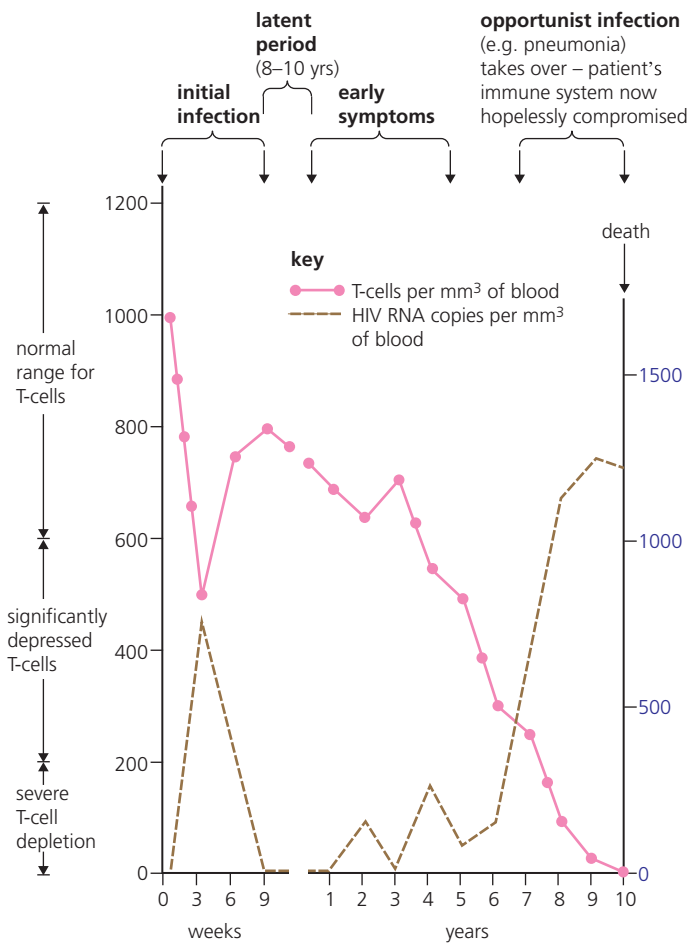


Figure 7.23 Profile of an AIDS infection

Treatment, prevention and the social consequences of AIDS

Infection with HIV is possible through contact with blood or body fluids of infected people, such as may occur during sexual intercourse, sharing of hypodermic needles by intravenous drug users, and breast feeding of a newborn baby. Also, blood transfusions and organ transplants will transmit HIV, but donors are now screened for HIV infection in most countries. HIV is *not* transferred by contact with saliva on a drinking glass, or by sharing a towel, for example. Nor does the female mosquito transmit HIV when feeding on human blood (page 595).

The spread of HIV and the eventual onset of AIDS in patients are outpacing the current efforts of scientists and doctors to prevent them. The WHO charts the spread of this pandemic (Figure 7.24).

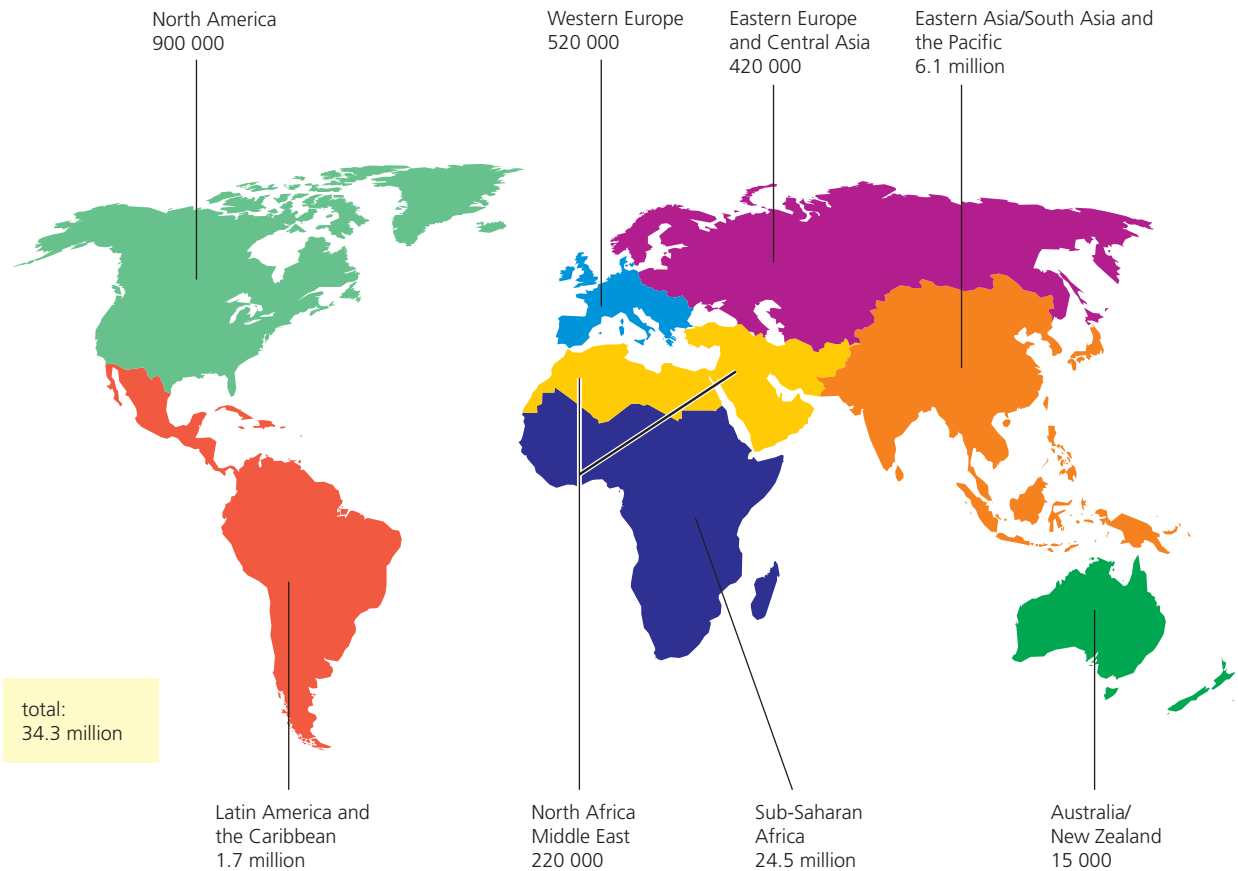


Figure 7.24 Worldwide incidence of HIV infection (WHO, 2000)

HIV infection and AIDS are difficult to treat. We have already noted that viruses are not controlled by antibiotics. Once the presence of HIV has been confirmed, patients may be offered drugs that slow down the progress of the infection – a combination of drugs that reduces the number of HIV-infected cells, at least temporarily. The three most popular drugs (AZT and two protease inhibitors) interrupt the steps of nucleic acid reverse transcription. A combination of drugs is used to prevent HIV from rapidly developing resistance to any one drug, and to avoid dangerous or unpleasant side-effects that some patients experience.

Ideally, a vaccine against HIV would be the best solution – one designed to wipe out infected lymphocytes and HIV particles in the patient's blood stream. The work of several laboratories is dedicated to this solution. The problem is that in the latent state of the infection, the infected (T4) lymphocyte cells frequently change their membrane marker proteins because of the presence of the HIV genome within the cell. Effectively, HIV can hide from the body's immune response by changing its identity.

Current strategies for preventing the spread of HIV are:

- practising 'safe sex' by using condoms to prevent transmission of the virus through infected blood or semen (condoms need to be freely available to the sexually active population);
- use of sterile needles by intravenous drug users (sterile needles need to be freely available);
- effective education programmes so that the vulnerable understand the cause and effects of HIV infection, and the best steps to remain healthy, whether or not they are literate.

TOK Link

Sexually transmitted diseases are rampant among teenagers in developed countries. Worldwide, there is a relentless rise in cases of AIDS. What are the ideal practical responses to the risk of infection by:

- vulnerable individual citizens?
- national governments?

Why may current responses be proving ineffective?

Extension

Recent studies have shown that the prevalence of AIDS is significantly higher in uncircumcised males. Circumcised males were seven times less likely to transmit HIV to, or receive HIV from, their partner. The reason for this is the mucous membrane of the inner surface of the foreskin has cells with receptors that HIV can exploit. Consequently, it is possible that HIV infection of future generations might be reduced if male circumcision were practised more widely. But health workers point out that if this were to encourage males to participate in unsafe sex with numerous partners, it would defeat the object.

In fact, AIDS is just one of several **sexually transmitted diseases (STDs)** currently on the increase in various parts of the world. But the **social implications of the AIDS pandemic** are devastating the economies and social life of communities in many less-developed countries and threatening to do the same in many others. The social consequences are summarised in Table 7.8.

Table 7.8 The social consequences of AIDS

	Social consequence
Psychological	Patient, family and friends suffer grief, loneliness and guilt.
Economic	Where the producer or wage-earner is incapacitated, poverty results or is increased. Infected people may be stigmatised, and deprived of housing or employment.
Child development and education	Children may become orphans at early age, older ones having to rear siblings with little support, and unable to proceed with education or training.
Child health	Offspring may be infected at birth or during breast feeding, leading to early ill-health and death.
Medical services provision	Meagre resources are diverted to treat long-term symptoms, without hope of cure. Other diseases not treated properly.
National factors	Reduction in numbers of the economically productive members of the community and incapacity of many of the more sexually active members cramp aspirations and achievements for a nation.

14 **Describe** why AIDS patients typically die from common infectious diseases that are not normally fatal.

15 **Distinguish** between the following pairs:

a antigens and antibodies;
b antibiotics and vaccines;
c vector and host.

Effects of HIV on the immune system

We have seen that HIV specifically attaches to antibody-secreting lymphocytes and takes over their nucleic acid and protein synthesis machinery. As the virus becomes active after a dormant period, the number of these lymphocytes drops profoundly, leaving the patient vulnerable to attack by opportunistic infections such as pneumonia and meningitis (Figure 7.23, page 201). This is because, with the reduction in the number of active lymphocytes, there is a dramatic loss in the ability to produce antibodies to a wide range of common infections.

Gaseous exchange

6.4.1–6.4.5

All living things respire.

Cellular respiration is the controlled release of energy in the form of ATP from organic compounds in cells. It is a continuous process in all cells.

To support aerobic cellular respiration, cells take in oxygen from their environment and give out carbon dioxide, by a process called gaseous exchange (Figure 7.25).

Gaseous exchange is the exchange of gases between an organism and its surroundings, including the uptake of oxygen and the release of carbon dioxide in animals and plants.

The exchange of gases between the individual cell and its environment takes place by **diffusion**. For example, in cells respiring aerobically there is a higher concentration of oxygen outside the cells than inside, and so there will be a continuous net inward diffusion of oxygen.

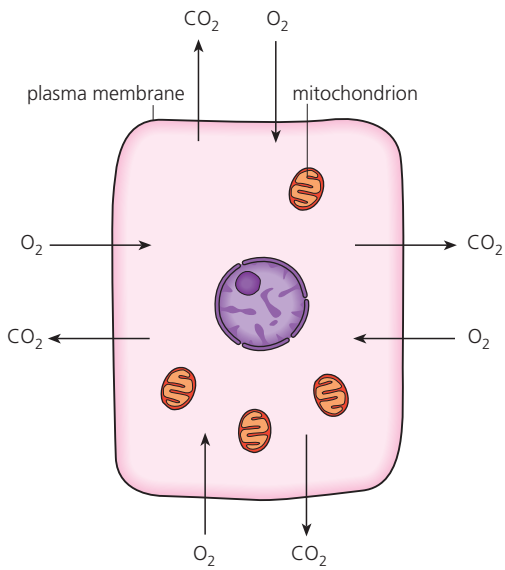


Figure 7.25 Gaseous exchange in an animal cell

What speeds up diffusion? In living things there are three factors which effectively determine the rate of diffusion in practice.

- The size of the **surface area** available for gaseous exchange (the respiratory surface). The greater this surface area, the greater the rate of diffusion. Of course, in a single cell, the respiratory surface is the whole plasma membrane.
- The **difference in concentration**. A rapidly respiring organism will have a very much lower concentration of oxygen in the cells and a higher than normal concentration of carbon dioxide. The greater the gradient in concentration across the respiratory surface, the greater the rate of diffusion.
- The **length of the diffusion path**. The shorter the diffusion path, the greater the rate of diffusion, so the respiratory surface must be as thin as possible.

Gaseous exchange in animals

We can see that the surface area of singled-celled animal is large in relation to the amount of cytoplasm it contains – here the surface of the cell is sufficient for efficient gaseous exchange. On the other hand, large, multicellular animals have very many of their cells too far from the body surface to receive enough oxygen by diffusion alone.

In addition, animals often develop an external surface of tough or hardened skin that provides protection to the body, but which is not suitable for gaseous exchange. These organisms require an alternative respiratory surface.

Active organisms have an increased metabolic rate, and the demand for oxygen in their cells is higher than in sluggish and inactive organisms. So for many reasons, large active animals such as **mammals** have specialised organs for gaseous exchange. In mammals, the respiratory surface consists of **lungs**. Lungs provide a large, thin surface area, suitable for gaseous exchange. However, the lungs are in a protected position inside the thorax (chest), so air has to be brought to the respiratory surface there. The lungs must be ventilated.

A ventilation system is a pumping mechanism that moves air into and out of the lungs efficiently, thereby maintaining the concentration gradient for diffusion.

In addition, in mammals the conditions for diffusion at the respiratory surface are improved by:

- a blood circulation system, which rapidly moves oxygen to the body cells as soon as it has crossed the respiratory surface, thereby maintaining the concentration gradient;
- a respiratory pigment, which increases the oxygen-carrying ability of the blood. This is the haemoglobin of the red cells which are by far the most numerous of the cells in our blood circulation.

16 **List** three characteristics of an efficient respiratory surface and explain how each influences diffusion.

The working lungs of mammals

The structure of the human thorax is shown in Figure 7.26. Lungs are housed in the **thorax**, an airtight chamber formed by the **rib-cage** and its muscles (**intercostal muscles**), with a domed floor, the **diaphragm**. The diaphragm is a sheet of muscle attached to the body wall at the base of the rib-cage, separating thorax from abdomen. The internal surfaces of the thorax are lined by the **pleural membrane**, which secretes and maintains pleural fluid. Pleural fluid is a lubricating liquid derived from blood plasma that protects the lungs from friction during breathing movements.

Lungs connect with the pharynx at the rear of the mouth by the trachea. Air reaches the trachea from the mouth and nostrils, passing through the larynx (voice box). Entry into the larynx is via a slit-like opening, the glottis. Above is a cartilaginous flap, the **epiglottis**. Glottis and epiglottis work to prevent the entry of food into the trachea. The trachea initially runs beside the oesophagus. Incomplete rings of cartilage in the trachea wall prevent collapse under pressure from a large bolus of food passing down the oesophagus.

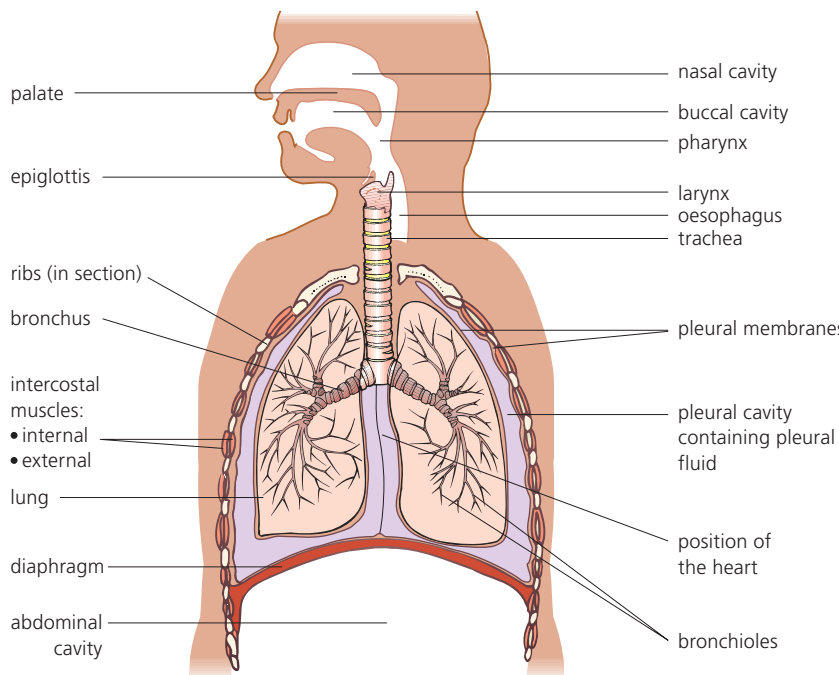


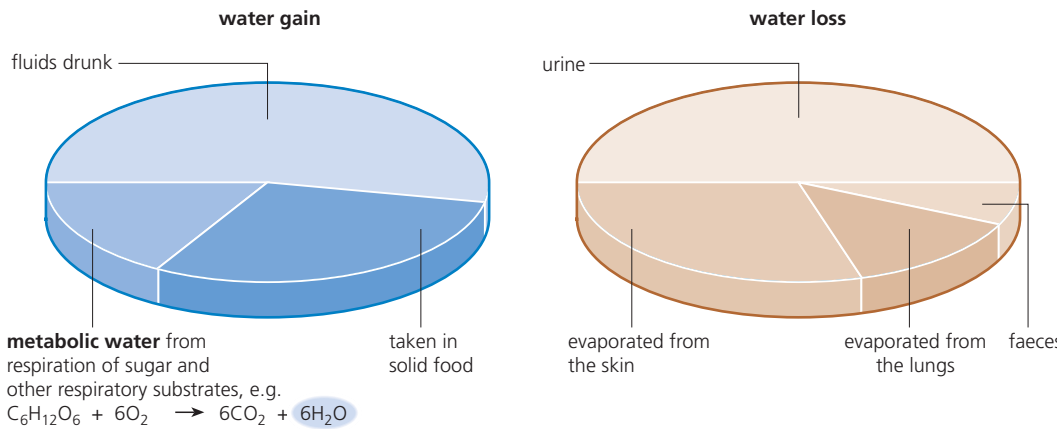
Figure 7.26 The structure of the human thorax

Figure 7.27 Water loss in gaseous exchange in humans

Exhaled air is saturated with water vapour, an invisible component except in freezing weather or when breathed on to a very cold surface.

The water balance of the body:

- about 3 dm^3 is taken in and lost daily
- about 10–15% of this total is lost from the lungs.



The trachea then divides into two **bronchi**, one to each lung. Within the lungs the bronchi divide into smaller bronchioles. The finest bronchioles end in air sacs (**alveoli**). The walls of bronchi and larger bronchioles contain smooth muscle, and are also supported by rings or tiny plates of cartilage, preventing collapse that might be triggered by a sudden reduction in pressure that occurs with powerful inspirations of air.

Lungs are extremely efficient, but of course they cannot prevent some water loss during breathing – an issue for most terrestrial organisms (Figure 7.27).

Ventilation of the lungs

Air is drawn into the alveoli when the air pressure in the lungs is lower than atmospheric pressure, and it is forced out when pressure is higher than atmospheric pressure. Since the thorax is an airtight chamber, pressure changes in the lungs occur when the volume of the thorax changes (Table 7.9 and Figure 7.28).

The volume of the thorax is increased when the ribs are moved upwards and outwards, and the diaphragm dome is lowered. These movements are brought about by contraction of the diaphragm and external intercostal muscles.

The volume of the thorax is decreased by the diaphragm muscles relaxing and the diaphragm becoming more dome-shaped by pressure from below (natural elasticity of the stomach and liver, displaced and stretched at inspiration). Also, the internal intercostal muscles contract. The ribs now move down and inwards.

17 **Compare** roles of the internal and external intercostal muscles during ventilation of the lungs (Figure 7.28 and Table 7.9).

Table 7.9 The mechanism of lung ventilation – a summary

Inspiration (inhalation)	Structure/outcome	Expiration (exhalation)
muscles contract, flattening the diaphragm and pushing down on contents of abdomen	**diaphragm**	muscles relax, pressure from abdominal contents pushes diaphragm into a dome shape
contract, moving rib-cage up and out	**external intercostal muscles**	relax
relax	**internal intercostal muscles**	contract, moving rib-cage down and in
increases	**volume of thorax cavity**	decreases
falls below atmospheric pressure	**air pressure of thorax**	rises above atmospheric pressure
in	**air flow**	out

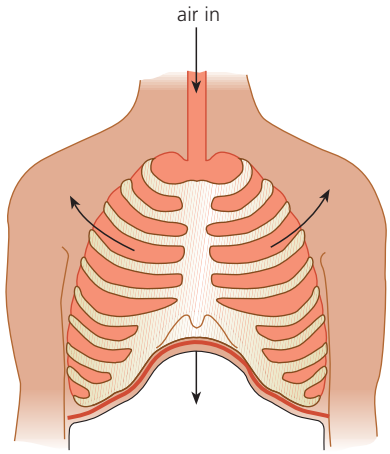
Figure 7.28 The ventilation mechanism of the lungs

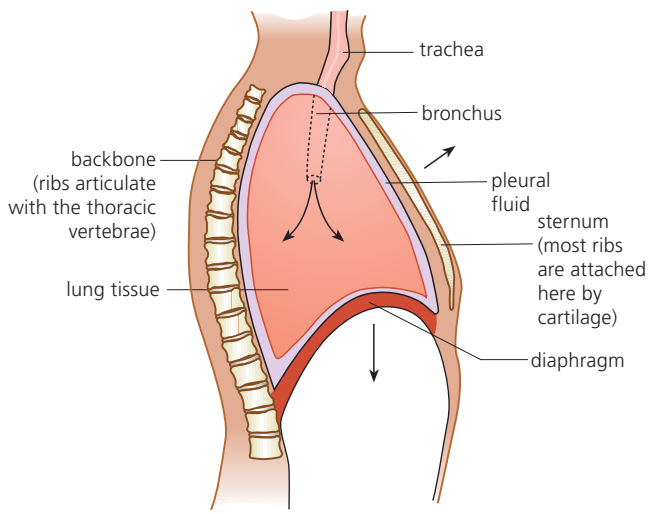
inspiration:
- external intercostal muscles contract
- internal intercostal muscles relax

} ribs moved upwards and outwards, and the diaphragm down

- diaphragm muscles contract

air in





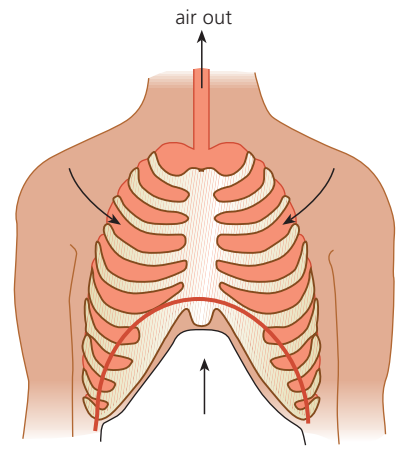
volume of the thorax (and therefore of the lungs) increases; pressure is reduced below atmospheric pressure and air flows in

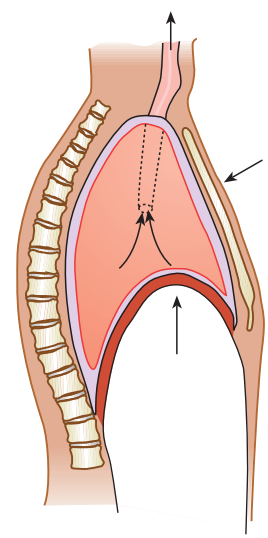
expiration:
- external intercostal muscles relax
- internal intercostal muscles contract

} ribs moved downwards and inwards, and the diaphragm up

- diaphragm muscles relax

air out



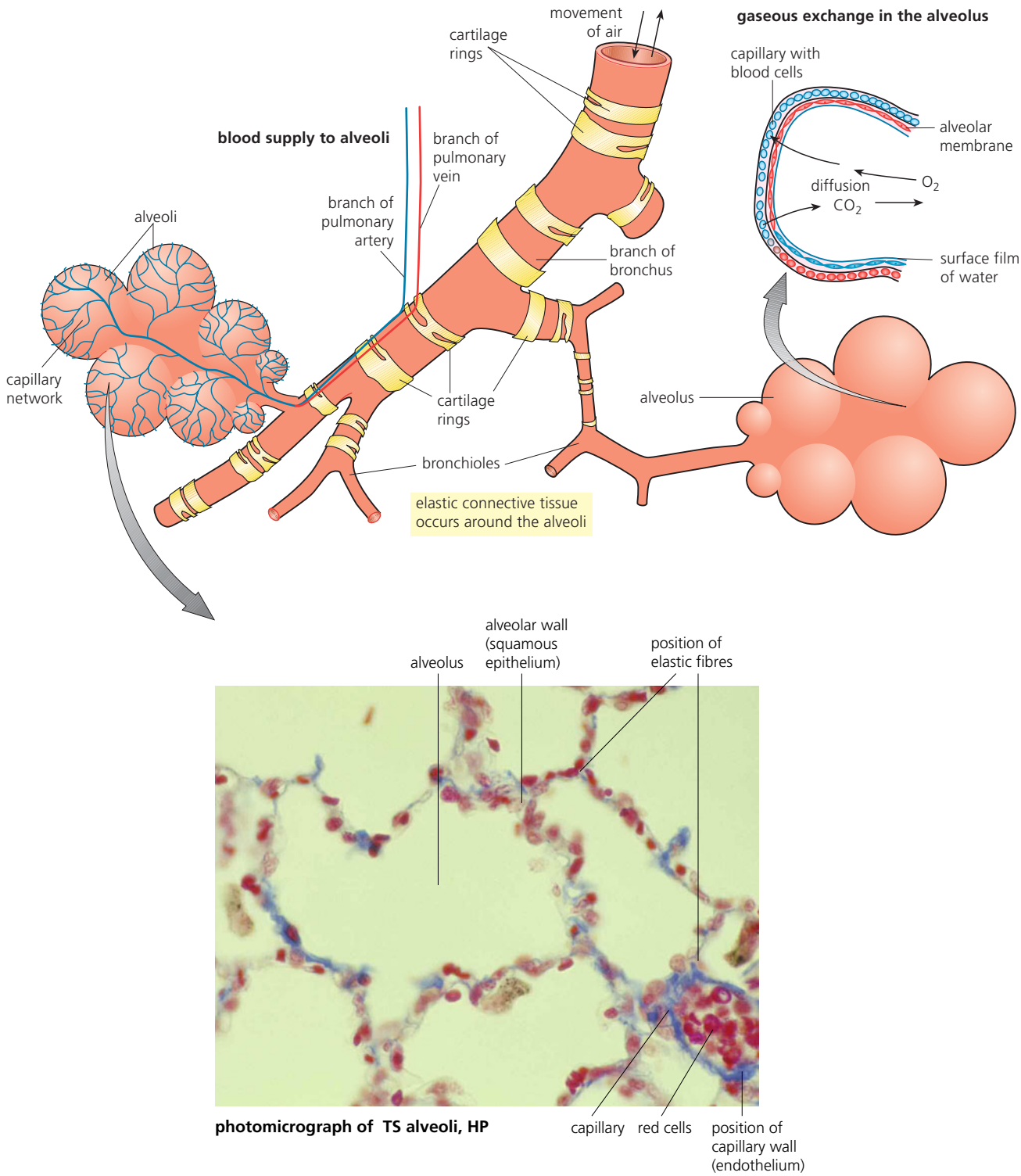


volume of the thorax (and therefore of the lungs) decreases; pressure is increased above atmospheric pressure and air flows out

Alveolar structure and gaseous exchange

The lung tissue consists of the alveoli, arranged in clusters, each served by a tiny bronchiole. Alveoli have elastic connective tissue as an integral part of their walls. A capillary system wraps around the clusters of alveoli (Figure 7.29). Each capillary is connected to a branch of the pulmonary artery and is drained by a branch of the pulmonary vein. The pulmonary circulation is supplied with deoxygenated blood from the right side of the heart, and returns oxygenated blood to the left side of the heart to be pumped to the rest of the body.

Figure 7.29 Gaseous exchange in the alveoli



There are some 700 million alveoli in our lungs, providing a surface area of about 70 m^2 in total. This is an area 30–40 times greater than that of the body's external skin. The wall of an alveolus is one cell thick, and is formed by pavement epithelium. Lying very close is a capillary, its wall also composed of a single layer of flattened (endothelium) cells. The combined thickness of walls separating air and blood is typically 5 μm. The capillaries are extremely narrow, just wide enough for red cells to squeeze through, so red cells are close to or in contact with the capillary walls.

Blood arriving in the lungs is low in oxygen but high in carbon dioxide. As blood flows past the alveoli, gaseous exchange occurs by diffusion. Oxygen dissolves in the surface film of water, diffuses across into the blood plasma and into the red cells where it combines with haemoglobin to form oxyhaemoglobin. At the same time, carbon dioxide diffuses from the blood into the alveoli (Table 7.10).

Component	Inspired air/%	Alveolar air/%	Expired air/%
oxygen	20	14	16
carbon dioxide	0.04	5.5	4.0
nitrogen	79	81	79
water vapour	variable	saturated	saturated

Table 7.10 The composition of air in the lungs

18 Explain why, if the concentration of carbon dioxide was to build up in the blood of a mammal, this would be harmful.

Efficiency of lungs as organs of gaseous exchange

The lungs of mammals are just one evolutionary response to the need for efficient organs of gaseous exchange in compact multicellular animals. Alternative arrangements include the system of tubes that pipe air directly to respiring cells of insects, the internal gills of fish, and the lungs of birds in which air travels completely through, to and from the air sacs beyond.

How effective are mammalian lungs?

Air flow in the lungs of mammals is tidal, in that air enters and leaves by the same route. Consequently there is a residual volume of air that cannot be expelled. Incoming air mixes with and dilutes the residual air, rather than replaces it. The effect of this is that air in the alveoli contains significantly less oxygen than the atmosphere outside (Table 7.10).

Nevertheless, the lungs are efficient organs. Their success is due to numerous features of the alveoli that adapt them to gaseous exchange. These are listed in Table 7.11.

Feature	Effects and consequences
surface area of alveoli	a huge surface area for gaseous exchange (50 m^2 = area of doubles tennis court)
wall of alveoli	very thin, flattened (squamous) epithelium (5 μm) – diffusion pathway is short
capillary supply to alveoli	network of capillaries around each alveolus supplied with deoxygenated blood from pulmonary artery and draining into pulmonary veins – maintains the concentration gradient of O_2 and CO_2
surface film of moisture	O_2 dissolves in water lining the alveoli; O_2 diffuses into the blood in solution

Table 7.11 Features of alveoli that adapt them to efficient gaseous exchange

19 Explain the difference between gaseous exchange and cellular respiration.

20 Explain why a significant amount of our total daily water loss occurs from the lungs (Figure 7.27).

Nerves, hormones and homeostasis

6.5.1–6.5.12

Control and communication within the body involves both the **nervous system** and **hormones from the endocrine glands**. We now look at these systems, in turn.

Introducing the nervous system

The nervous system is built from nerve cells called **neurones**. A neurone has a **cell body** containing the **nucleus** and the bulk of the cytoplasm. From the cell body run fine cytoplasmic **fibres**. Most fibres are very long indeed.

Neurones are specialised for the transmission of information in the form of impulses. An **impulse** is a momentary reversal in the electrical potential difference in the membrane of a neurone. The transmission of an impulse along a fibre occurs at speeds between 30 and 120 metres per second in mammals, so nervous coordination is extremely fast, and responses are virtually immediate. Impulses travel to particular points in the body, served by the fibres. Consequently, the effects of impulses are localised rather than diffuse.

Neurones are grouped together to form the **central nervous system**, which consists of the **brain** and **spinal cord**. To and from the central nervous system run nerves of the **peripheral nervous system**. Communication between the central nervous system and all parts of the body occurs via these nerves (Figure 7.30).

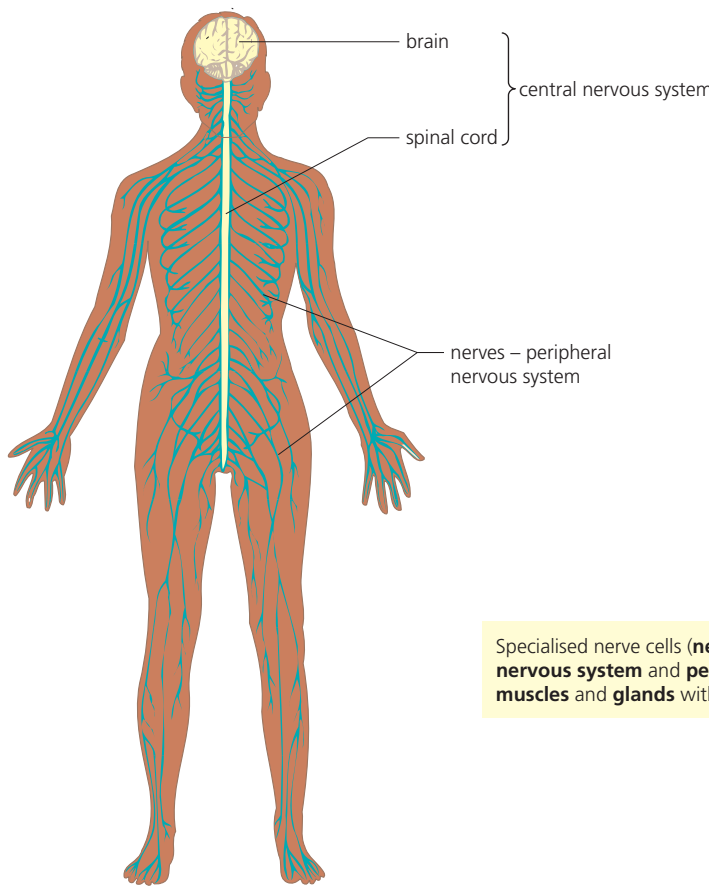


Figure 7.30 The organisation of the mammalian nervous system

Specialised nerve cells (**neurones**) are organised into **central nervous system** and **peripheral nerves** linking **sense organs**, **muscles** and **glands** with the **brain** or **spinal cord**.

Looking at neurone structure

Three types of neurone make up the nervous system (sensory neurones, relay neurones and motor neurones). The structure of a motor neurone is shown in Figure 7.31.

The motor neurones have many fine **dendrites** which bring impulses towards the cell body, and a single long **axon** which carries impulses away from the cell body. The function of the motor neurone is to carry impulses from the central nervous system to a muscle or gland (known as an **effector**).

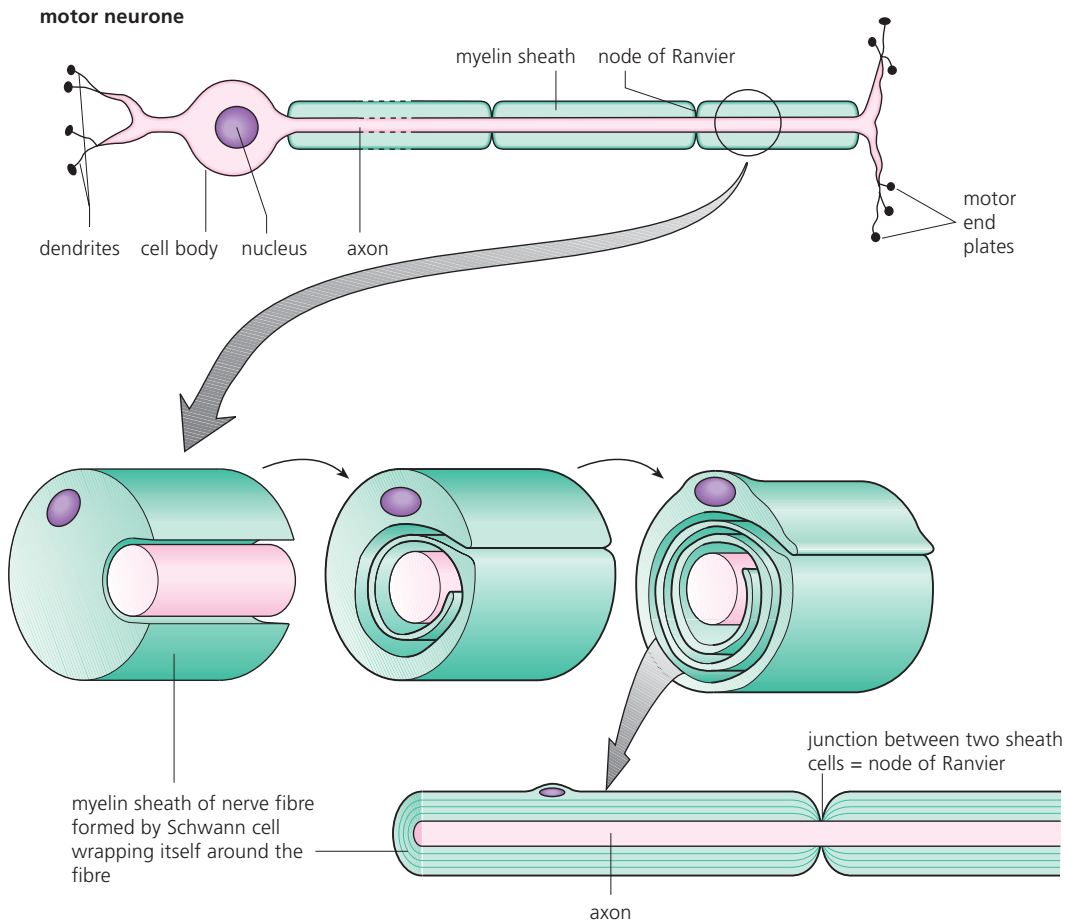


Figure 7.31 Motor neurone structure

A neurone is surrounded by many supporting cells, one type of which, Schwann cells, become wrapped around the axons of motor neurones, forming a structure called a **myelin sheath**. Myelin consists largely of lipid, and has high electrical resistance. Frequent junctions occur along a myelin sheath, between the individual Schwann cells. The junctions are called nodes of Ranvier.

Neurones and the transmission of an impulse

Neurones transmit information in the form of impulses. An impulse is transmitted along nerve fibres, but it is not an electrical current that flows along the 'wires' of the nerves. As stated above, an impulse is a momentary reversal in electrical potential difference in the membrane – a change in the position of charged ions between the inside and outside of the membrane of the nerve fibres. This reversal flows from one end of the neurone to the other in a fraction of a second.

Between conduction of one impulse and the next the neurone is sometimes said to be resting, but this not the case. The 'resting' neurone membrane actively sets up and maintains the electrical potential difference between the inside and the outside of the fibre.

How is this done?

The resting potential

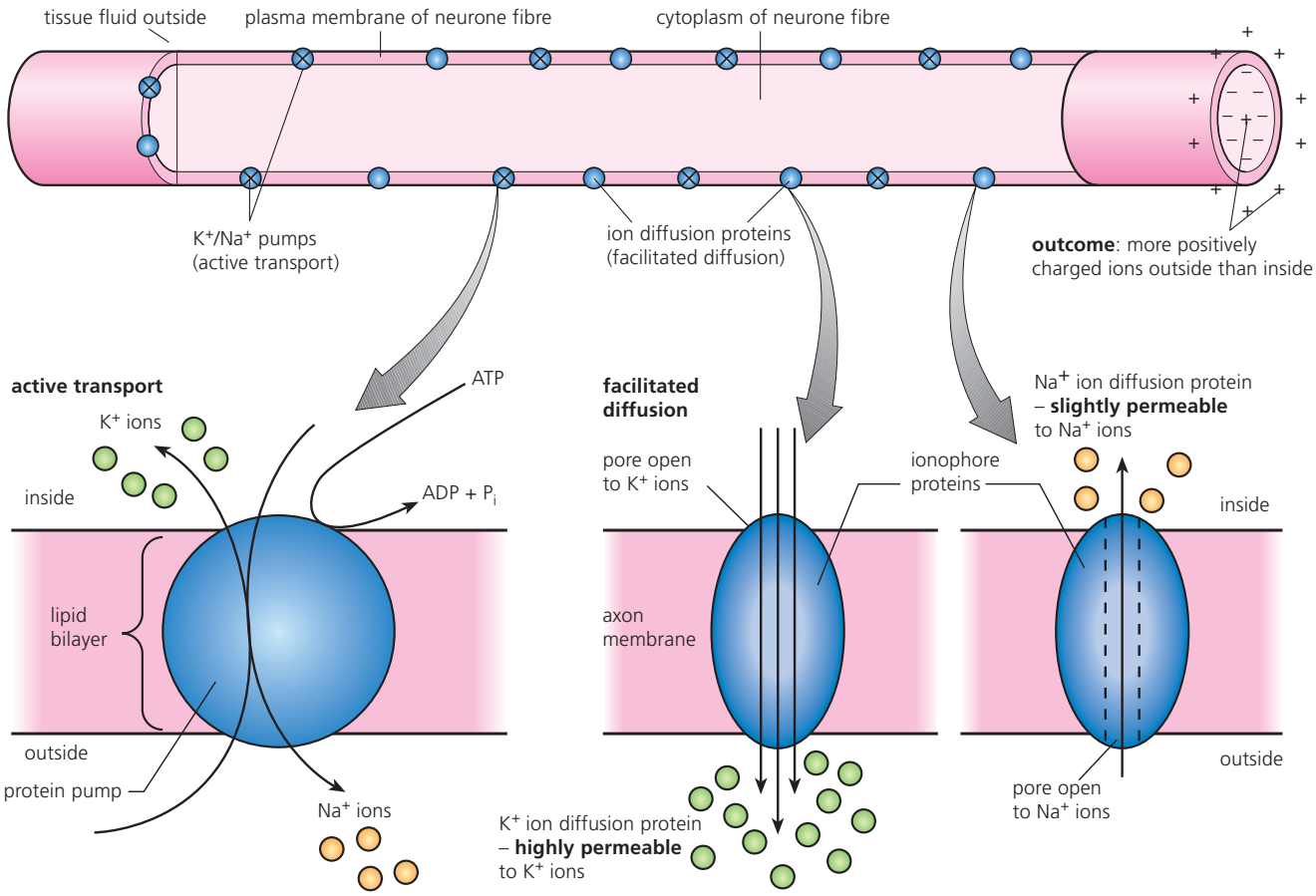
The resting potential is the potential difference across a nerve cell membrane when it is not being stimulated. It is normally about –70 millivolts (mV).

The resting potential difference is re-established across the neurone membrane after a nerve impulse has been transmitted. We say the nerve fibre has been **repolarised**.

The resting potential is the product of two processes:

1 **The active transport of potassium ions (K^+) in across the membrane, and sodium ions (Na^+) out across the membrane.** This occurs by a K^+/Na^+ pump, using energy from ATP (Figure 1.30, page 29). The concentration of potassium and sodium ions on opposite sides of the membrane is built up, but this in itself makes no change to the potential difference across the membrane.
2 **Facilitated diffusion of K^+ ions out, and Na^+ ions back in.** The important point here is that the membrane is far more permeable to K^+ ions flowing out than to Na^+ ions returning. This causes the tissue fluid outside the neurone to contain many more positive ions than are present in the tiny amount of cytoplasm inside. As a result, a **negative charge** is developed inside compared to outside and the resting neurone is said to be **polarised**. The difference in charge or potential difference (about –70 mV), is known as the **resting potential** (Figure 7.32).

Figure 7.32 The establishment of the resting potential



The next event, sooner or later, is the passage of an impulse, known as an **action potential**.

The action potential

The action potential is the potential difference produced across the plasma membrane of the nerve cell when stimulated, reversing the resting potential from about –70 mV to about +40 mV.

An action potential is triggered by a stimulus received at a receptor cell or sensitive nerve ending. The energy of the stimulus causes a temporary and local reversal of the resting potential. The result is that the membrane is briefly **depolarised**.

This change in potential across the membrane occurs because of pores in the membrane called **ion channels**. These special channels are globular proteins that span the membrane. They have a central pore with a **gate** that can open and close. One type of channel is permeable to sodium ions, and another to potassium ions. During a resting potential these channels are all closed.

The transfer of energy of the stimulus first opens the gates of the **sodium channels** in the plasma membrane and sodium ions diffuse in, down their **electrochemical gradient**.

What causes an electrochemical gradient?

The electrochemical gradient of an ion is due to its electrical and chemical properties.

- **Electrical properties** are due to the charge on the ion (an ion is attracted to an opposite charge).
- **Chemical properties** are due to concentration in solution (an ion tends to move from a high to a low concentration).

With sodium channels opened, the cytoplasm of the neurone fibre (the interior) quickly becomes progressively more positive with respect to the outside. When the charge has been **reversed** from −70 mV to +40 mV (due to the electrochemical gradient), an **action potential has been created** in the neurone fibre (Figure 7.33).

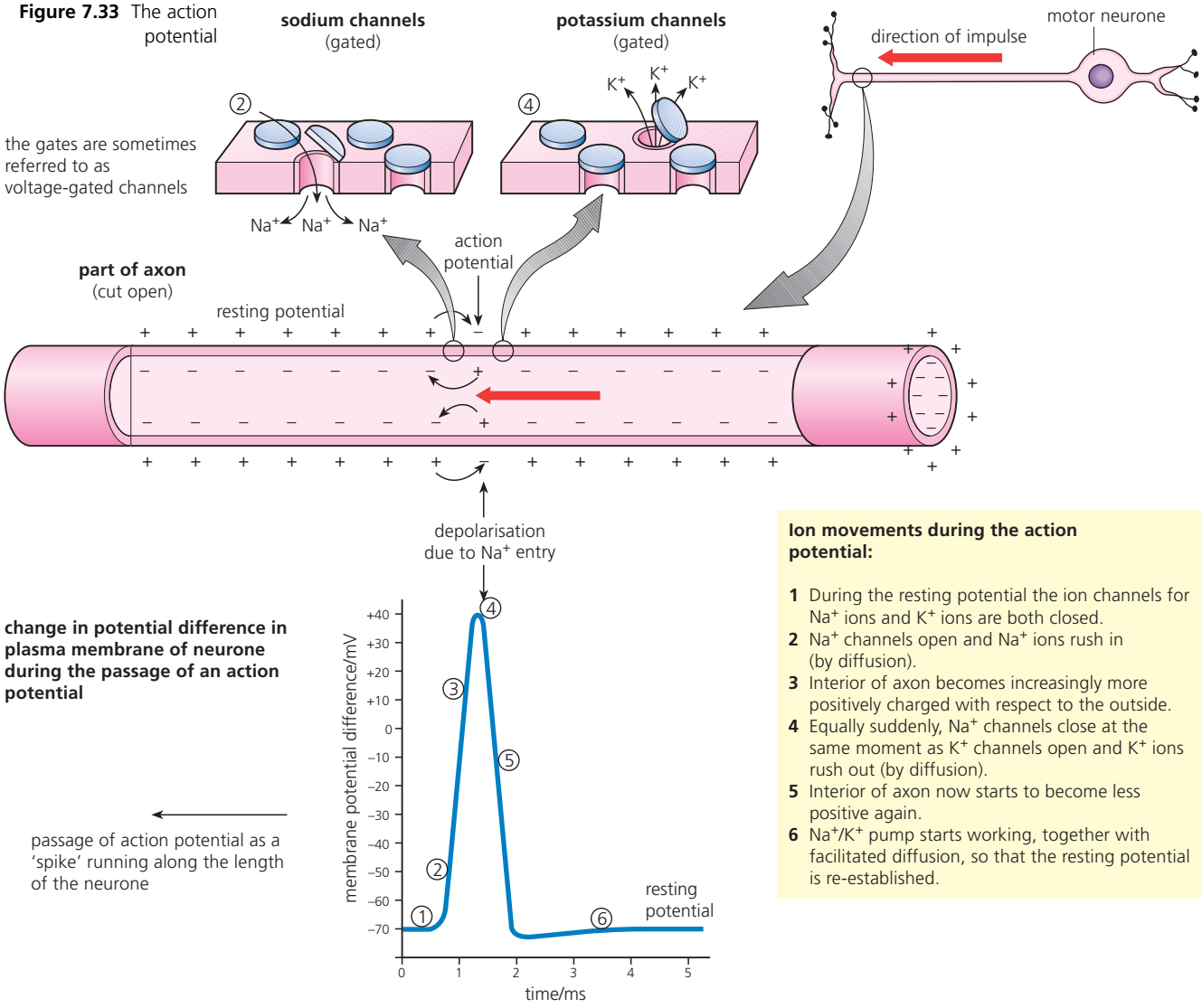
The action potential then runs the length of the neurone fibre. At any one point it exists for only two thousandths of a second (2 milliseconds), before the resting potential starts to be re-established. So action potential transmission is exceedingly quick.

Almost immediately an action potential has passed, the sodium channels close and the **potassium channels** open. Now, potassium ions can exit the cell, again down an electrochemical gradient, into the tissue fluid outside. The interior of the neurone fibre starts to become less positive again. Then the potassium channels also close. Finally, the resting potential is re-established by the action of the sodium/potassium pump and the process of facilitated diffusion.

21 Deduce the source of energy used to:

a establish the resting potential
b power an action potential.

Figure 7.33 The action potential



Ion movements during the action potential:

1. During the resting potential the ion channels for Na^+ ions and K^+ ions are both closed.
2. Na^+ channels open and Na^+ ions rush in (by diffusion).
3. Interior of axon becomes increasingly more positively charged with respect to the outside.
4. Equally suddenly, Na^+ channels close at the same moment as K^+ channels open and K^+ ions rush out (by diffusion).
5. Interior of axon now starts to become less positive again.
6. Na^+/K^+ pump starts working, together with facilitated diffusion, so that the resting potential is re-established.

The refractory period

For a brief period, following the passage of an action potential, the neurone fibre is no longer excitable. This is the **refractory period**, and it lasts only 5–10 milliseconds in total. The neurone fibre is not excitable during the refractory period because there is a large excess of sodium ions inside the fibre and further influx is impossible. Subsequently, as the resting potential is progressively restored, it becomes increasingly possible for an action potential to be generated again. Because of the refractory period, the maximum frequency of impulses is between 500 and 1000 per second.

The all-or-nothing principle

Obviously stimuli are of widely different strengths: for example, a light touch and the pain of a finger hit by a hammer. A stimulus must be at or above a minimum intensity, known as the **threshold of stimulation**, in order to initiate an action potential. Either the depolarisation is sufficient to fully reverse the potential difference in the cytoplasm (from −70 mV to +40 mV), or it is not. If not, no action potential arises. With all subthreshold stimuli, the influx of sodium ions is quickly reversed, and the full resting potential is re-established.

However, as the **intensity of the stimulus increases**, the **frequency** at which the **action potentials pass** along the fibre **increases** (the individual action potentials are all of standard strength). For example, with a very persistent stimulus, action potentials pass along a fibre at an accelerated rate, up to the maximum possible permitted by the refractory period. This means the effector (or the brain) is able to recognise the intensity of a stimulus from the frequency of action potentials (Figure 7.34).

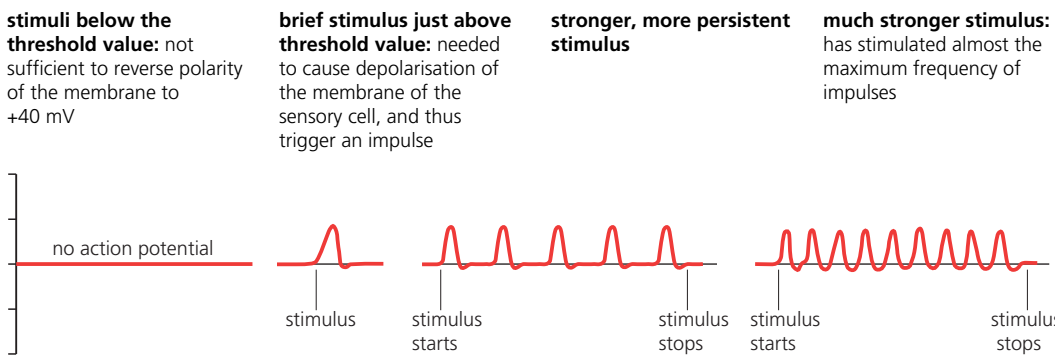


Figure 7.34 Weak and strong stimuli and the threshold value

Junctions between neurones

The synapse is the link point between neurones. A synapse consists of the swollen tip (synaptic knob) of the axon of one neurone (**pre-synaptic neurone**) and the dendrite or cell body of another neurone (**post-synaptic neurone**). At the synapse, the neurones are extremely close but they have **no direct contact**. Instead there is a tiny gap, called a **synaptic cleft**, about 20 nm wide (Figure 7.35).

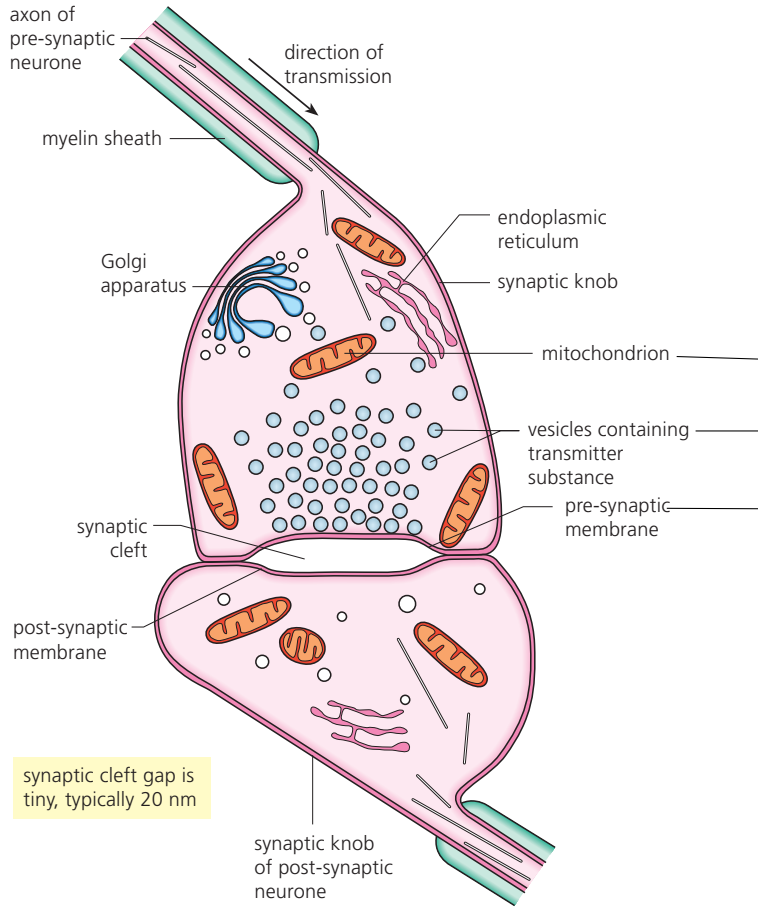
The practical effect of the synaptic cleft is that an action potential can only cross it via specific chemicals, known as **transmitter substances**. Transmitter substances are all relatively small, diffusible molecules. They are produced in the Golgi apparatus in the synaptic knob, and held in tiny vesicles before use.

Acetylcholine (ACh) is a commonly occurring transmitter substance (the neurones that release acetylcholine are known as cholinergic neurones). Another common transmitter substance is **noradrenalin** (from adrenergic neurones). In the brain, the commonly occurring transmitters are glutamic acid and dopamine.

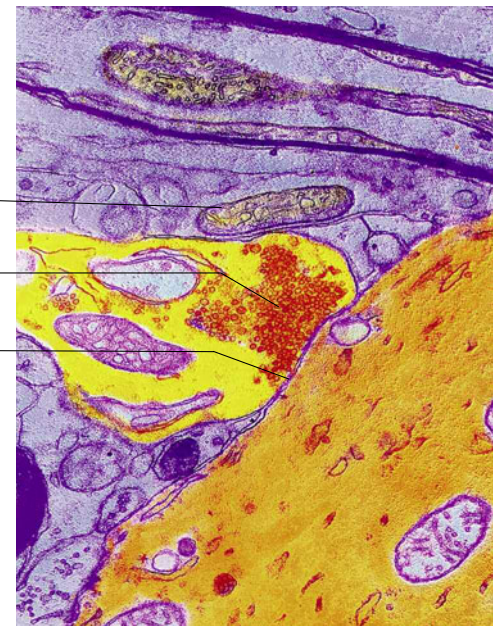
Steps of synapse transmission

You may find it helpful to follow each step in Figure 7.36, the diagram of this event.

Figure 7.35 A synapse in section



TEM of a synapse (×80 000)



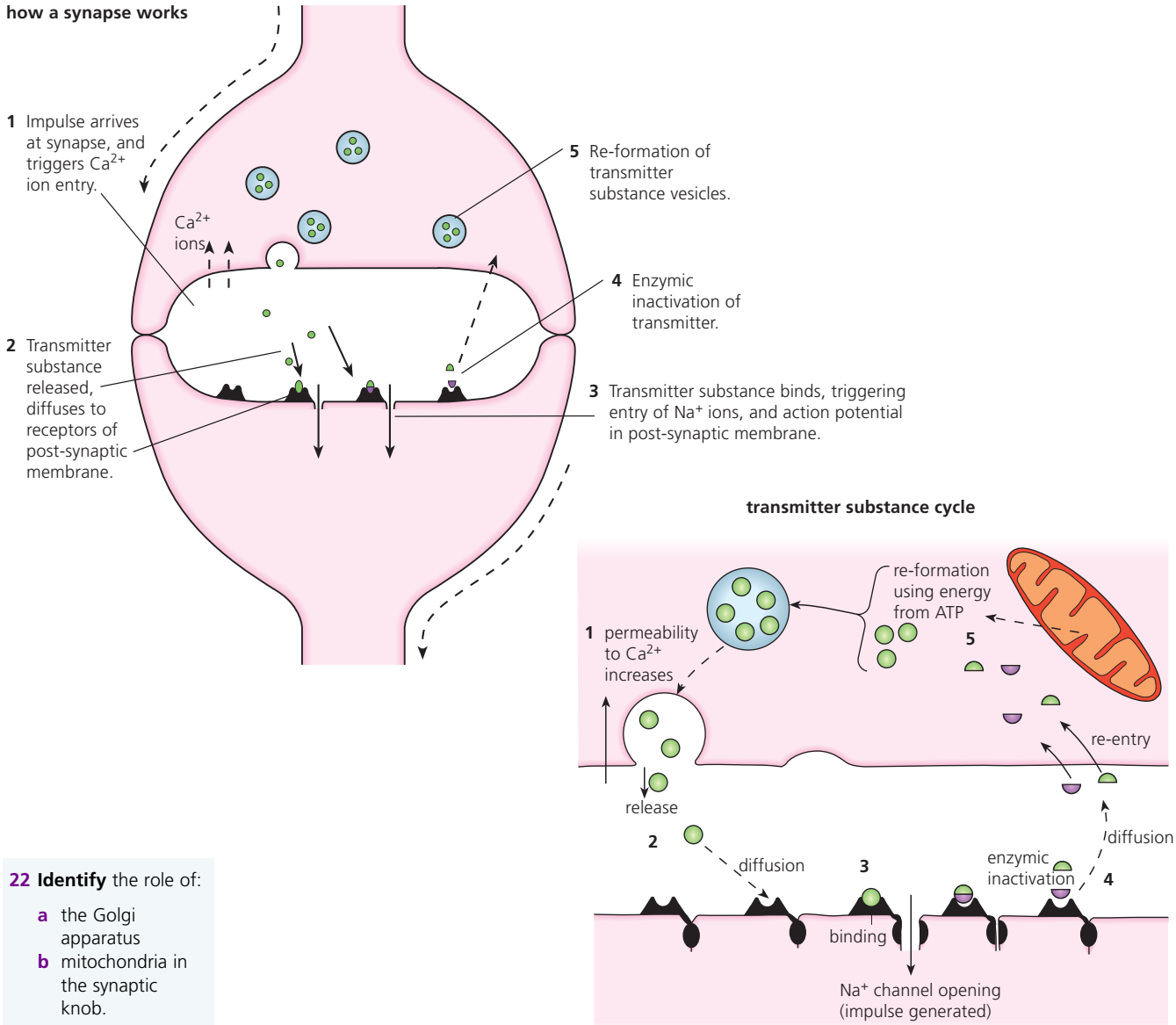
1 The arrival of an action potential at the synaptic knob opens **calcium ion channels in the pre-synaptic membrane**, and calcium ions flow in from the synaptic cleft.
2 The calcium ions cause **vesicles of transmitter substance** to fuse with the pre-synaptic membrane and release transmitter substance into the synaptic cleft.
3 The transmitter substance diffuses across the synaptic cleft and binds with a **receptor protein**.

In the **post-synaptic membrane** there are specific receptor sites for each transmitter substance. Each of these receptors also acts as a channel in the membrane which allows a specific ion (e.g. Na^+, or Cl^- or some other ion) to pass. The attachment of a transmitter molecule to its receptor instantly **opens the ion channel**.

When a molecule of ACh attaches to its receptor site, a Na^+ channel opens. As the sodium ions rush into the cytoplasm of the post-synaptic neurone, **depolarisation** of the post-synaptic membrane occurs. As more and more molecules of ACh bind, it becomes increasingly likely that depolarisation will reach the **threshold level**. When it does, an **action potential is generated** in the post-synaptic neurone. This process of build-up to an action potential in post-synaptic membranes is called **facilitation**.

4 The transmitter substance on the receptors is immediately **inactivated** by enzyme action. For example, the enzyme cholinesterase hydrolyses ACh to choline and ethanoic acid, which are inactive as transmitters. This causes the ion channel of the receptor protein to close, and so allows the resting potential in the post-synaptic neurone to be re-established.
5 The inactivated products from the transmitter re-enter the pre-synaptic knob, are **resynthesised** into transmitter substance, and packaged for re-use.

Figure 7.36 Chemical transmission at the synapse



22 Identify the role of:

- **a** the Golgi apparatus
- **b** mitochondria in the synaptic knob.

Introducing the endocrine system

Hormones are chemical substances produced and secreted from the cells of the ductless or **endocrine glands**. In effect, hormones carry messages about the body – but in a totally different way from the nervous system.

Hormones are transported indiscriminately in the blood stream, but they act only at specific sites, called **target organs**. Although present in small quantities, hormones are extremely effective messengers, helping to control and co-ordinate body activities. Once released, hormones typically cause changes to specific metabolic reactions of their target organs, but a hormone circulates in the blood stream only briefly. In the liver, they are broken down and the breakdown products are excreted in the kidneys. So, long-acting hormones must be secreted continuously to be effective.

An example of a hormone to be discussed is insulin, released from endocrine cells in the pancreas. Insulin regulates blood glucose (page 222). The positions of endocrine glands of the body are shown in Figure 7.37.

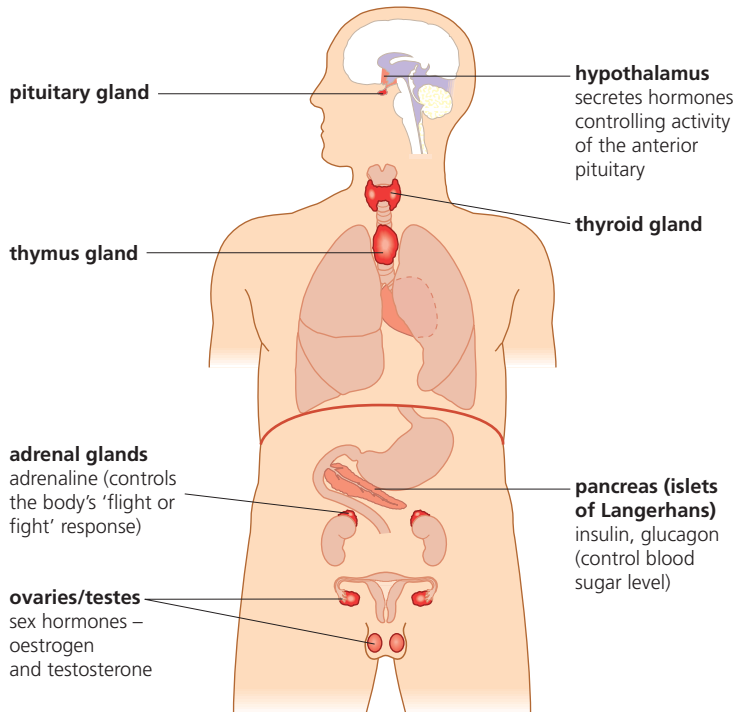


Figure 7.37 The human endocrine system

Maintaining a constant internal environment – homeostasis

Living things face changing and sometimes hostile environments; some external conditions change slowly, others dramatically. For example, temperature changes quickly on land exposed to direct sunlight, but the temperature of water exposed to sunlight changes very slowly (page 39).

How do organisms respond to environmental changes?

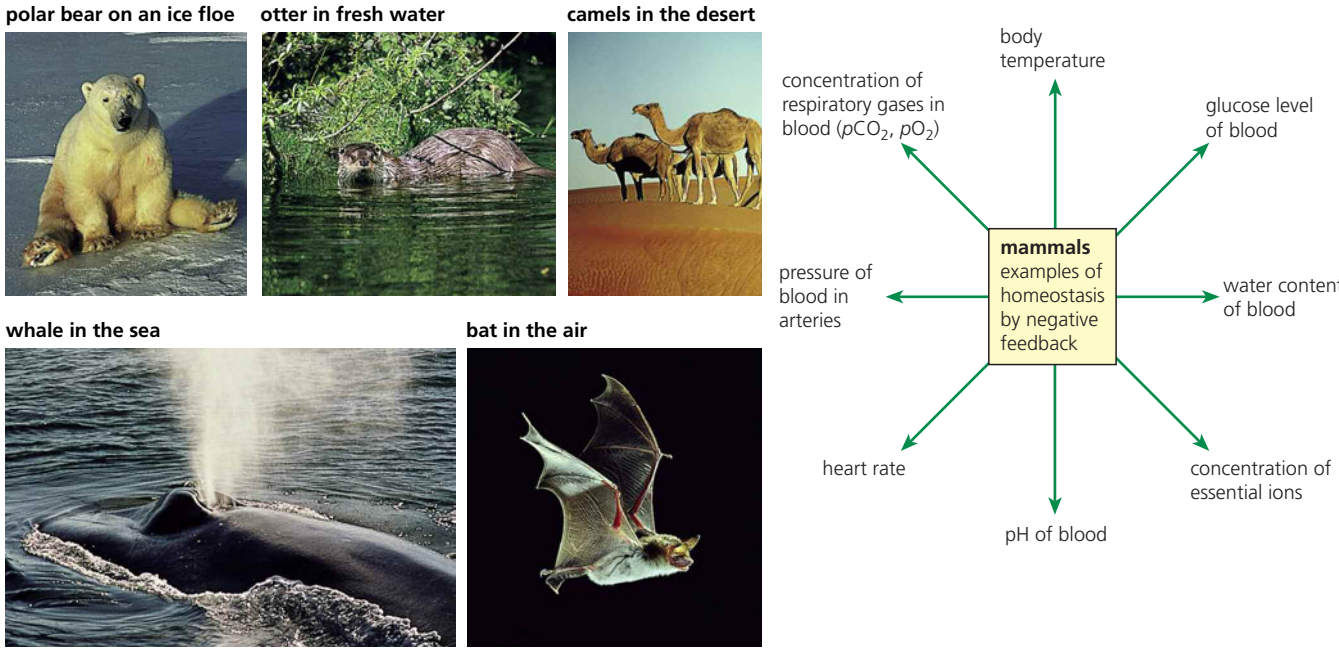
An animal that is able to maintain a constant internal environment, enabling it to continue normal activities, more or less whatever the external conditions, is known as a **regulator**. For example, mammals and birds maintain a high and almost constant body temperature over a very wide range of external temperatures. Their bodies are at or about the optimum temperature for the majority of the enzymes that drive their metabolism. Their muscles contract efficiently, and the nervous system co-ordinates responses precisely, even when external conditions are unfavourable. They are often able to avoid danger, and perhaps they may also benefit from the vulnerability of prey organisms which happen to be **non-regulators**. So regulators may have greater freedom in choosing where to live. They can exploit more habitats with differing conditions than non-regulators.

Homeostasis is the name for this ability to maintain a constant internal environment. Homeostasis means 'staying the same'. The internal environment consists of the blood circulating in the body and the fluid circulating among cells (tissue fluid) that forms from it, delivering nutrients and removing waste products while bathing the cells. Mammals are excellent examples of animals that hold internal conditions remarkably constant. They successfully regulate and maintain their blood pH, oxygen and carbon dioxide concentrations, blood glucose, body temperature and water balance at constant levels or within narrow limits (Figure 7.38).

How is homeostasis achieved?

Figure 7.38 Homeostasis in mammals

Mammals are a comparatively recent group in terms of their evolutionary history, yet they have successfully settled in significant numbers in virtually every type of habitat on Earth. This success is directly linked to their ability to control their internal environment by homeostasis.



Negative feedback – the mechanism of homeostasis

Negative feedback is the type of control in which conditions being regulated are brought back to a set value as soon as it is detected that they have deviated from it. We see this type of mechanism at work in a system to maintain the temperature of a laboratory water bath. Analysis of this familiar example will show us the components of a negative feedback system (Figure 7.39).

A negative feedback system requires a **detector** device that measures the value of the variable (the water temperature in the water bath) and transmits this information to a **control unit**. The control unit compares data from the detector with a pre-set value (the desired water temperature of the water bath). When the value from the detector is below the required value, the control unit activates an **effector** device (a water heater in the water bath) so that the temperature starts to be raised. Once data from the detector register in the control box that the water has reached the set temperature, the control box switches off the response (the water heater). How precisely the variable is maintained depends on the sensitivity of the detector, but negative feedback control typically involves some degree of 'overshoot'.

In mammals, regulation of body temperature, blood sugar level, and the amounts of water and ions in blood and tissue fluid (**osmoregulation**) are regulated by negative feedback. The detectors are specialised cells in either the brain or other organs, such as the pancreas. The effectors are organs such as the skin, liver and kidneys. Information passes between them via the nerves of the nervous system or via hormones (the endocrine system), or both. The outcome is an incredibly precisely regulated internal environment.

Homeostasis in action

Control of body temperature

The regulation of body temperature, known as **thermoregulation**, involves controlling the amount of heat lost and heat gained across the body surface. Heat may be transferred between an animal and the environment by convection, radiation and conduction, and the body loses heat by evaporation (Figure 7.40).

Figure 7.39 Negative feedback, the mechanism

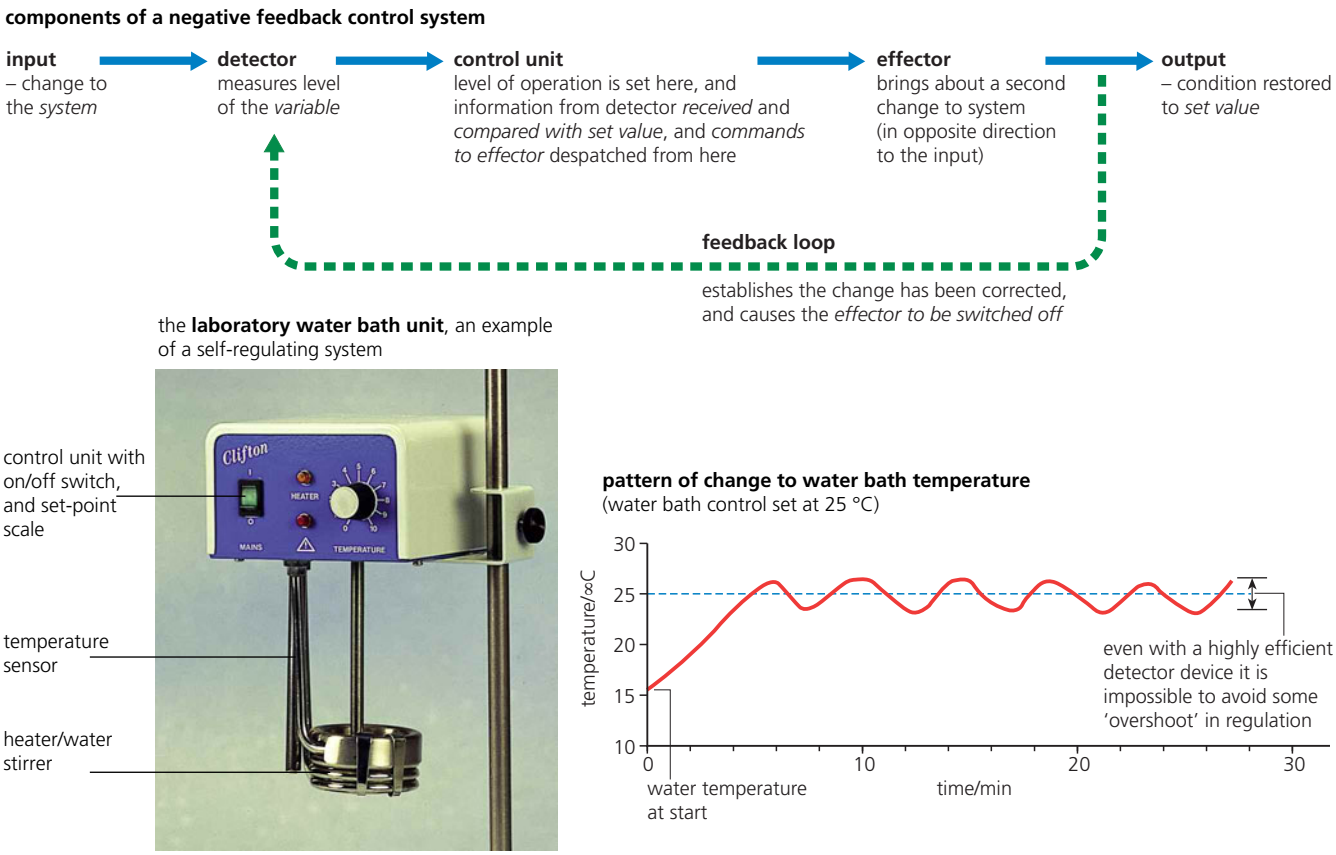
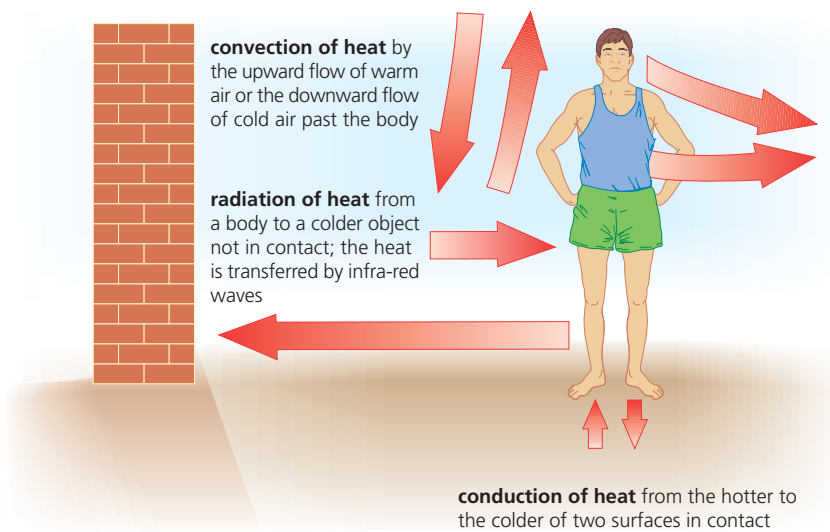


Figure 7.40 How heat is transferred between organism and environment



Mammals maintain a high and relatively constant body temperature, using the heat energy generated by metabolism within their bodies (or by generating additional heat in the muscles when cold) and carefully controlling the loss of heat through the skin. An animal with this form of thermoregulation is called an **endotherm**, meaning 'inside heat'. Birds too have perfected this mechanism. Humans hold their inner body temperature (core temperature) just below 37 °C. In fact, human core temperature only varies between about 35.5 and 37.0 °C within a 24-hour period, when we are in good health. Under low external temperature, however, only the temperature of the trunk is held constant; there is a progressive fall in temperature along the limbs (Figure 7.41).

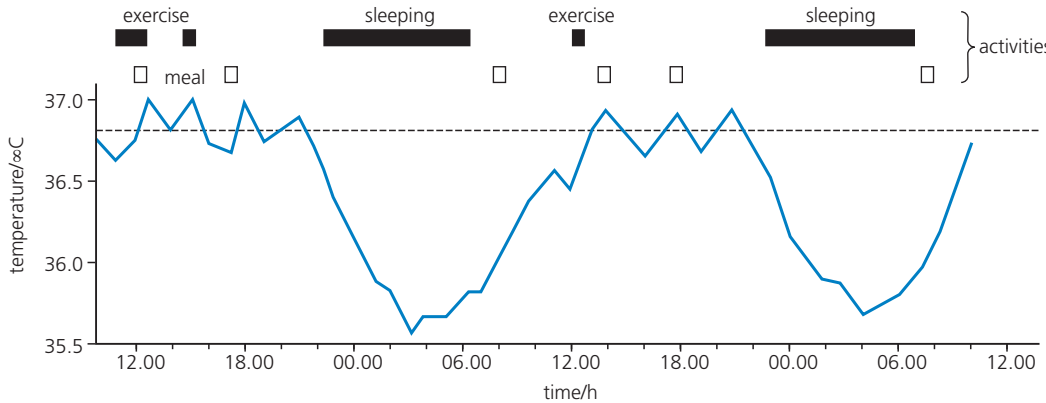
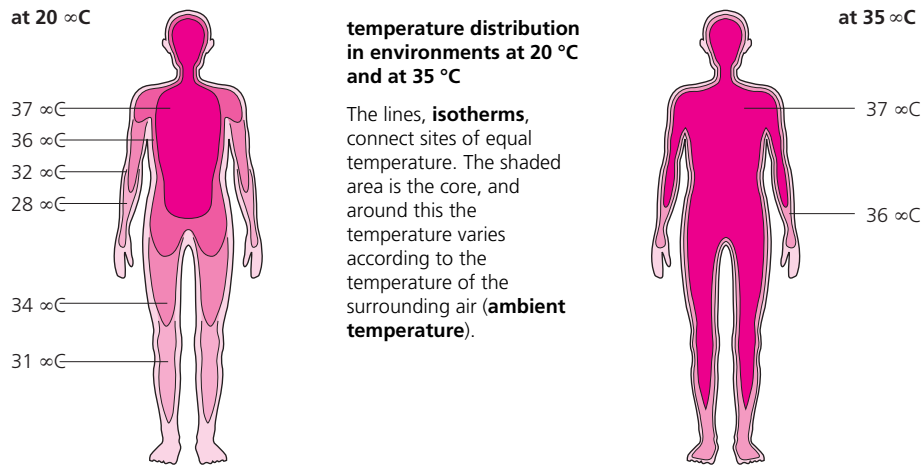


Figure 7.41 Body temperature of a human

body temperature over a 48-hour period

The body temperatures shown were taken with the thermometer under the tongue. Although this is a region close to the body 'core', temperatures here may be altered by eating/drinking, and by the breathing in through the mouth of cold air, for example. More accurate values are obtained by taking the rectal temperature.



temperature distribution in environments at 20 °C and at 35 °C

The lines, **isotherms**, connect sites of equal temperature. The shaded area is the core, and around this the temperature varies according to the temperature of the surrounding air (**ambient temperature**).

23 Deduce the times and conditions under which body temperature typically varies from the normal in a 24-hour period (Figure 7.41).

Heat production in the human body

The major sources of heat are the biochemical reactions of metabolism that generate heat as a waste product. Heat is then distributed by the blood circulation – we might say that heat is transferred by the blood circulation. The organs of the body vary greatly in the amount of heat they yield. For example, the liver is extremely active metabolically, but most of its metabolic reactions require an input of energy (they are **endergonic reactions**) and little energy is lost as heat. Mostly, the liver is thermally neutral.

The bulk of our body heat (over 70%) comes from other organs, mainly from the heart and kidneys, but also from the lungs and brain (which, like a computer central processing unit, needs to be kept cool). While the body is at rest, the skeleton, muscles and skin, which make up over 90% of the body mass, produce less than 30% of the body heat (Figure 7.42). Of course, in times of intense physical activity, the skeletal muscles generate a great deal of heat as a waste product of respiration and contraction.

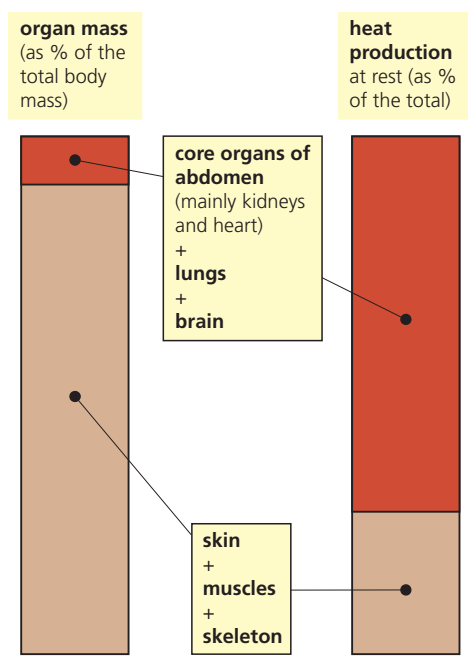


Figure 7.42 Heat production in the body at rest

24 State what it means to say that most reactions in liver cells are endergonic.

The roles of the skin in thermoregulation

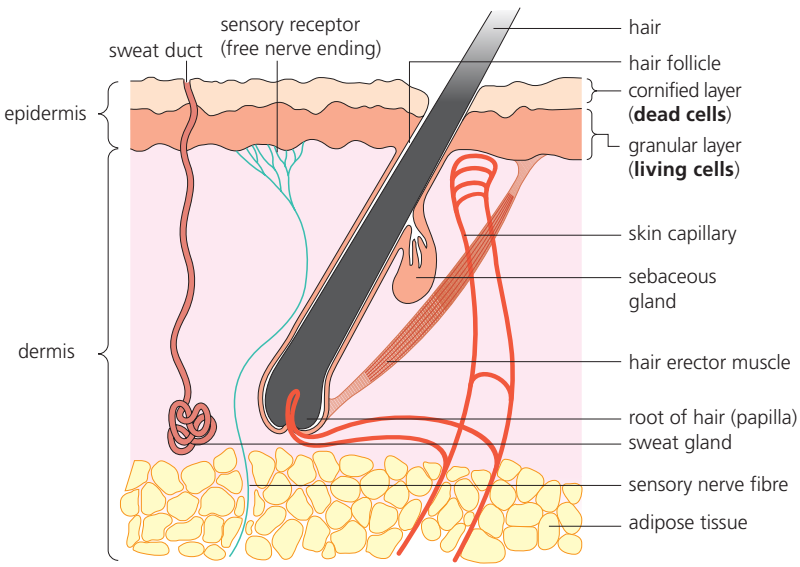
Heat exchanges occur at the skin. The structure of the skin is shown in Figure 7.43. Heat loss at the skin may be varied by:

- arterioles supplying **capillary networks** being dilated (**vasodilation**) when the body needs to lose heat, but constricted (**vasoconstriction**) when the body needs to retain heat;
- **hair erector muscles**, which contract when heat must be lost but relax when heat loss must be increased;
- **sweat glands**, which produce sweat when the body needs to lose heat, but do not when the body needs to retain heat.

Figure 7.43 The skin and temperature regulation

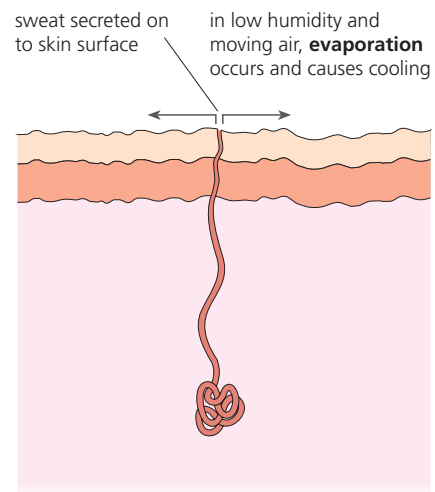
The operation of these mechanisms is shown in Figure 7.43.

structure of the skin



role of the sweat glands in regulating heat loss through the skin

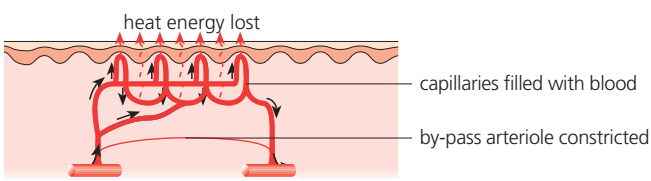
warm conditions



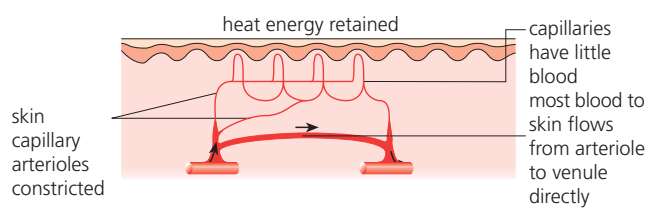
role of capillaries in regulating heat loss through the skin

In skin that is especially exposed (e.g. outer ear, nose, extremities of the limbs) the capillary network is extensive, and the arterioles supplying it can be dilated or constricted.

warm conditions

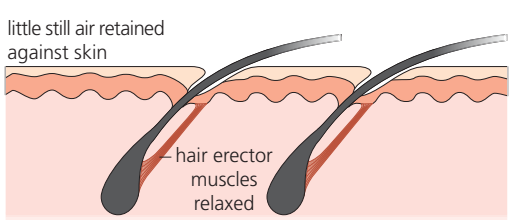


cold conditions



role of the hair in regulating heat loss through the skin

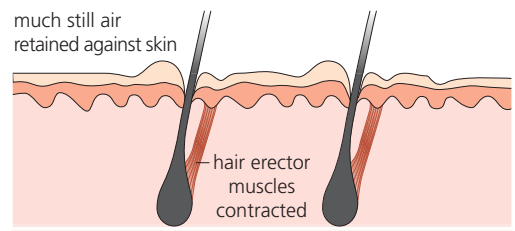
warm conditions



The hair erector muscles may be contracted or relaxed.

still air is a poor conductor of heat

cold conditions



Other mechanisms in thermoregulation

If the body experiences persistent cold then heat production is increased. Under chilly conditions, heat output from the body muscles is raised by contraction of skeletal muscles that is not co-ordinated and is known as **shivering**. This raises muscle heat production about five times above basal rate.

Regulation of blood glucose

Transport of glucose to all cells is a key function of the blood circulation. In humans, the **normal level of blood glucose** is about 90 mg of glucose in every 100 cm^3 of blood, but it can vary. For example, during an extended period without food, or after prolonged and heavy physical activity, blood glucose may fall to as low as 70 mg. After a meal rich in carbohydrate has been digested, blood glucose may rise to 150 mg.

The maintenance of a constant level of this monosaccharide in the blood plasma is important for two reasons.

1 Respiration is a continuous process in all living cells. To maintain their metabolism, cells need a regular supply of glucose, which can be quickly absorbed across the cell membrane. Glucose is the principal respiratory substrate for many tissues. Most cells (including muscle cells) hold reserves in the form of glycogen which is quickly converted to glucose during prolonged physical activity. However, glycogen reserves may be used up quickly. In the brain, glucose is the only substrate the cells can use and there is no glycogen store held in reserve.

   If our blood glucose falls below 60 mg per 100 cm^3, we have a condition called **hypoglycaemia**. If this is not quickly reversed, we may faint. If the body and brain continue to be deprived of adequate glucose levels, convulsions and coma follow.

2 An abnormally high concentration of blood glucose, known as **hyperglycaemia**, is also a problem. Since high concentration of any soluble metabolite lowers the water potential of the blood plasma, water is drawn from the cells and tissue fluid by osmosis, back into the blood. As the volume of blood increases, water is excreted by the kidney to maintain the correct concentration of blood. As a result, the body tends to become dehydrated, and the circulatory system is deprived of fluid. Ultimately, blood pressure cannot be maintained.

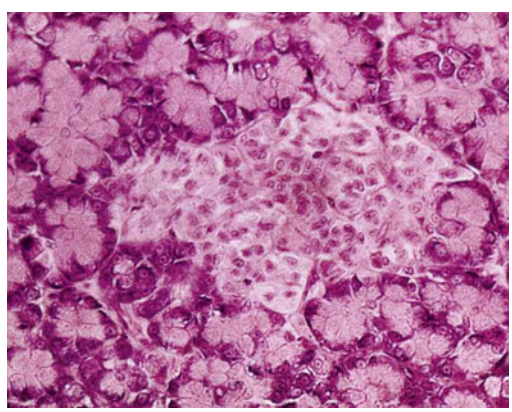
For these reasons, it is critically important the blood glucose is held within set limits.

Mechanism for regulation of blood glucose

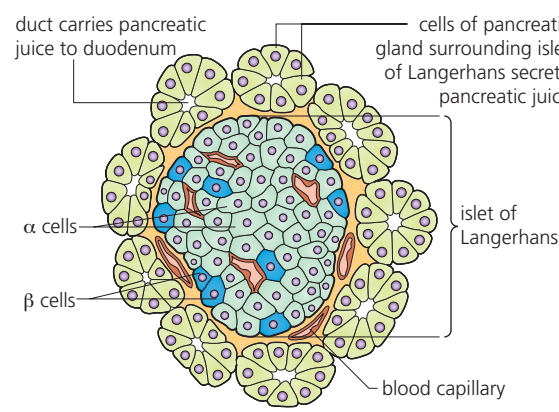
After the digestion of carbohydrates in the gut, glucose is absorbed across the epithelial cells of the villi (Figure 7.5, page 183) into the hepatic portal vein (Figure 7.11, page 189). The blood carrying the glucose reaches the liver first. If the glucose level is too high, glucose is withdrawn from the blood and stored as glycogen. But even so, blood circulating in the body immediately after a meal has a raised level of glucose. At the pancreas, the presence of an excess of blood glucose is detected in patches of cells known as the islets of Langerhans (Figure 7.44). These islets are hormone-secreting glands (endocrine glands); they have a rich capillary network, but no ducts that would carry secretions away. Instead, their hormones are transported all over the body by the blood. The islets of Langerhans contain two types of cell, **α cells** and **β cells**.

Figure 7.44 Islet of Langerhans in the pancreas

TS of pancreatic gland showing an islet of Langerhans



drawing of part of pancreatic gland



25 **Explain** what liver cells receive from blood from the hepatic artery that is not present in blood from the hepatic portal vein.

26 **Predict** what type of organelle you would expect to see most frequently when a liver cell is examined by EM.

In the presence of a **raised blood glucose** level, the **β cells** are stimulated. They secrete the hormone **insulin** into the capillary network. Insulin stimulates the uptake of glucose into cells all over the body, but especially by the liver and the skeletal muscle fibres. It also increases the rate at which glucose is used in respiration, in preference to alternative substrates (such as fat). Another effect of insulin is to trigger conversion of glucose to glycogen in cells (**glycogenesis**), and of glucose to fatty acids and fats, and finally the deposition of fat around the body.

As the blood glucose level reverts to normal this is detected in the islets of Langerhans, and the β cells respond by stopping insulin secretion. Meanwhile the hormone is excreted by the kidney tubules and the blood insulin level falls.

When the **blood glucose level falls below normal**, the **α cells** of the pancreas are stimulated. These secrete a hormone called **glucagon**. This hormone activates the enzymes that convert glycogen and amino acids to glucose (**gluconeogenesis**). Glucagon also reduces the rate of respiration (Figures 7.45 and 7.46)

As the blood glucose level reverts to normal, glucagon production ceases, and this hormone in turn is removed from the blood in the kidney tubules.

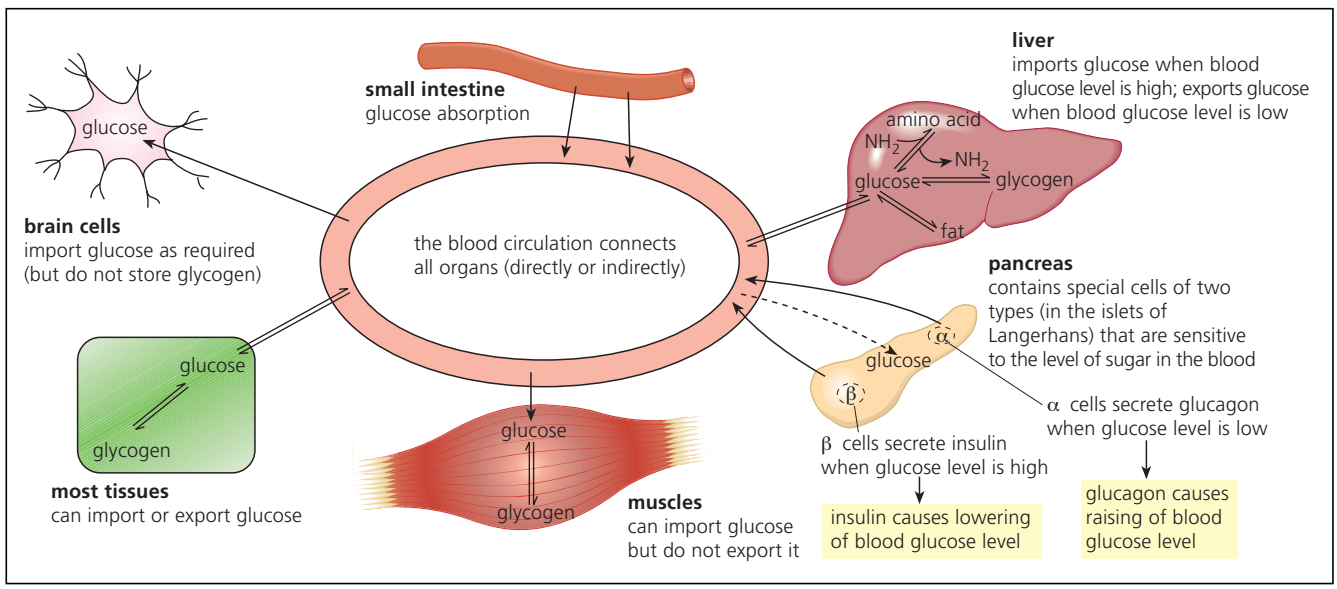


Figure 7.45 The sites of blood glucose regulation

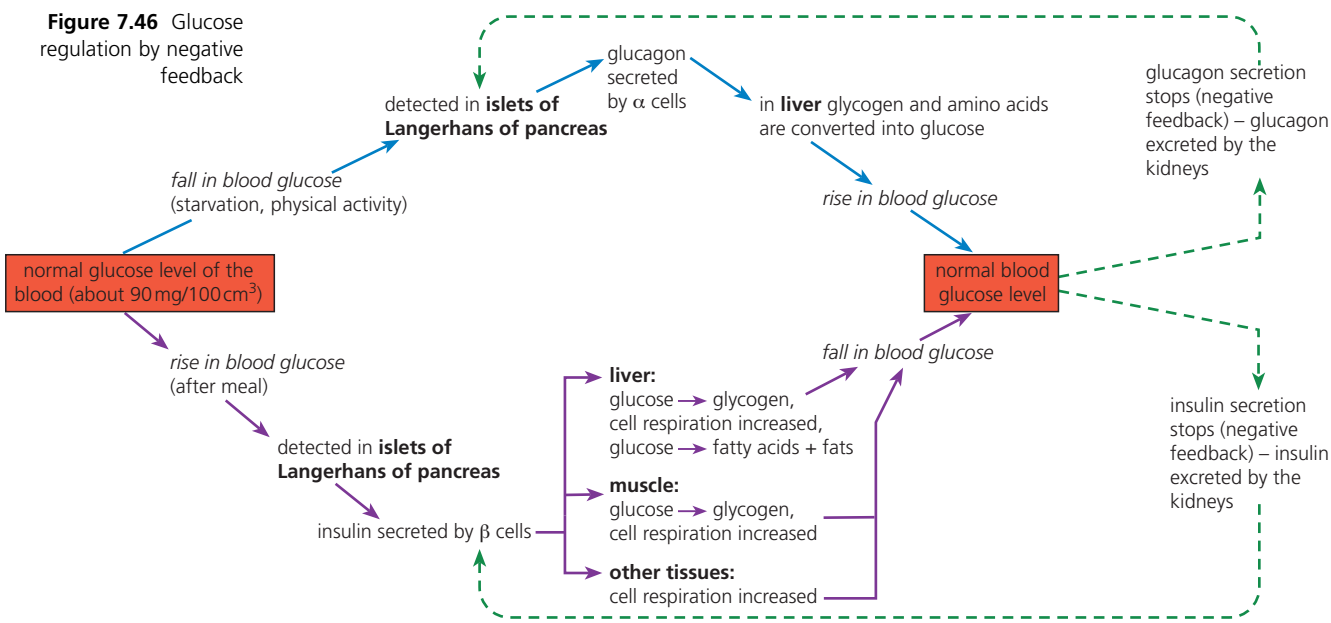


Figure 7.46 Glucose regulation by negative feedback

The disease diabetes

Diabetes is the name for a group of diseases in which the body fails to regulate blood glucose levels. **Type I diabetes** results from a failure of insulin production by the β cells. **Type II diabetes** (diabetes mellitus) is a failure of the insulin receptor proteins on the cell membranes of target cells (Figure 7.47). As a consequence, blood glucose is more erratic and generally, permanently raised. Glucose is also regularly excreted in the urine. If this condition is not diagnosed and treated, it carries an increased risk of circulatory disorders, renal failure, blindness, strokes or heart attacks.

Figure 7.47 Diabetes, cause and treatment

type I diabetes, 'early onset diabetes'
affects young people, below the age of 20 years

due to the destruction of the β cells of the islets of Langerhans by the body's own immune system

symptoms:
- constant thirst
- undiminished hunger
- excessive urination

treatment:
- injection of insulin into the blood stream daily
- regular measurement of blood glucose level

patient injecting with insulin, obtained by genetic engineering

type II diabetes, 'late onset diabetes'
the common form (90% of all cases of diabetes are of this type)

common in people over 40 years, especially if overweight, but this form of diabetes is having an increasing effect on human societies around the world, including young people and even children in developed countries, seemingly because of poor diet

symptoms:
mild – sufferers usually have sufficient blood insulin, but insulin receptors on cells have become defective

treatment:
largely by diet alone

Reproduction

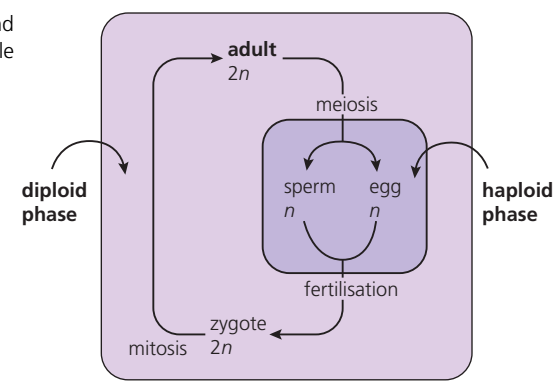
6.6.1–6.6.6

Reproduction is the production of new individuals by an existing member or members of the same species. It is a fundamental characteristic of living things; the ability to self-replicate in this way sets the living world apart from the non-living. In reproduction, a parent generation effectively passes on a copy of itself in the form of the genetic material, to another generation, the offspring. The genetic material of an organism consists of its chromosomes, made of nucleic acid (page 63).

Organisms reproduce either asexually or sexually and many reproduce by both these methods. However, mammals reproduce by **sexual reproduction** only.

In sexual reproduction, two **gametes** (specialised sex cells) fuse to form a **zygote** which then grows into a new individual. Fusion of gametes is called **fertilisation**. In the process of gamete formation, a nuclear division by **meiosis** (page 94) halves the normal chromosome number. That is, gametes are **haploid**, and fertilisation restores the **diploid** number of chromosomes (Figure 7.48). Without the reductive nuclear division in the process of sexual reproduction, the chromosome number would double in each generation. Remember, the offspring produced by sexual reproduction are unique, in complete contrast with offspring formed by asexual reproduction.

Figure 7.48 Meiosis and the diploid life cycle



Sexual reproduction in mammals

In mammals, like most vertebrates, the sexes are separate (unisexual individuals). In the body, the reproductive and urinary (excretory) systems are closely bound together, especially in the male, so biologists refer to these as the urinogenital system. Here, just the reproductive systems are considered (Figures 7.49 and 7.50).

Male and female gametes are produced in paired glands called **gonads**, the testes and ovaries.

The male reproductive system

- Two **testes** (singular, **testis**) are situated in the scrotal sac, hanging outside the main body cavity; this allows the testes to be at the optimum temperature for sperm production, a temperature 2–3 °C lower than the normal body temperature. As well as producing the male gametes, **spermatozoa** (singular, **spermatozoon**) or sperms, the testes also produce the male sex hormone, **testosterone**; the testes are, therefore, also endocrine glands (page 216).
- Ducts which store the sperms and carry them in a fluid, called **seminal fluid**, to the outside of the body during a process called an **ejaculation**.
- Ducted or exocrine glands which secrete the nutritive seminal fluid in which the sperms are transported; these include the **seminal vesicles** and **prostate gland**.
- The **penis**, through which the urethra runs. This duct carries semen during an ejaculation (and urine during urination) to the outside. The penis also contains spongy erectile tissue that can fill with blood when the male is sexually stimulated. This causes the penis to enlarge, lengthen and become rigid, in a condition known as an **erection**. The erect penis penetrates the vagina in sexual intercourse.

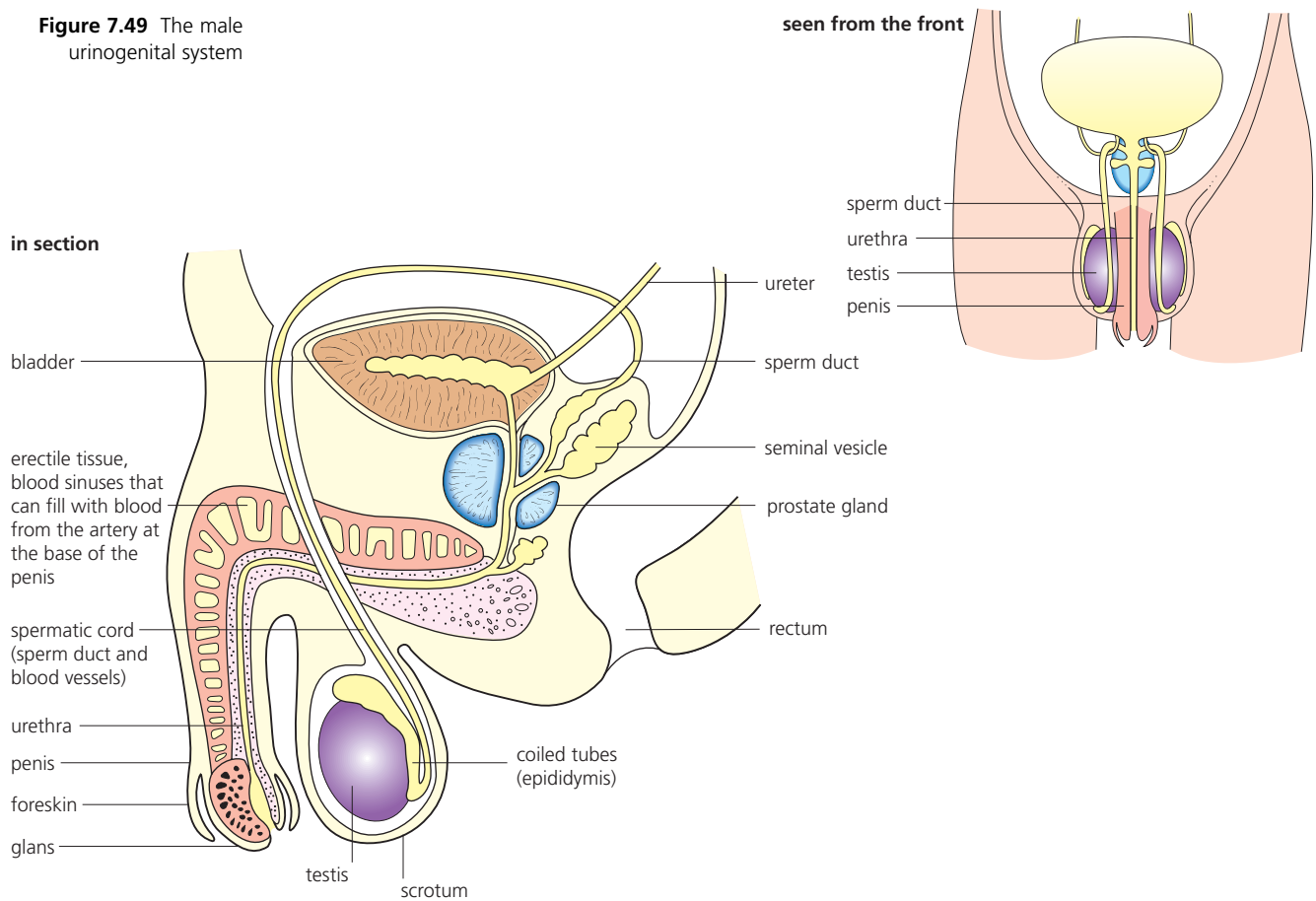


Figure 7.49 The male urinogenital system

The female reproductive system

- The **ovaries** are held near the base of the abdominal cavity. As well as producing the female gametes, ova or **egg cells**, the ovaries are also endocrine glands, secreting the female sex hormones **oestrogen** and **progesterone**.
- A pair of **oviducts** extend from the uterus and open as funnels close to the ovaries. The oviducts transport egg cells, and are the site of fertilisation.
- The **uterus**, which is about the size and shape of an inverted pear, has a thick muscular wall and an inner lining of mucous membrane richly supplied with arterioles. This lining, called the **endometrium**, undergoes regular change in a 28-day cycle. The lining is built up each month in preparation for implantation and early nutrition of a developing embryo, should fertilisation occur. If it does not occur, the endometrium disintegrates and menstruation starts.
- The **vagina**, a muscular tube that can enlarge to allow entry of the penis, and exit of a baby at birth. The vagina is connected to the uterus at the **cervix**, and it opens to the exterior at the **vulva**.

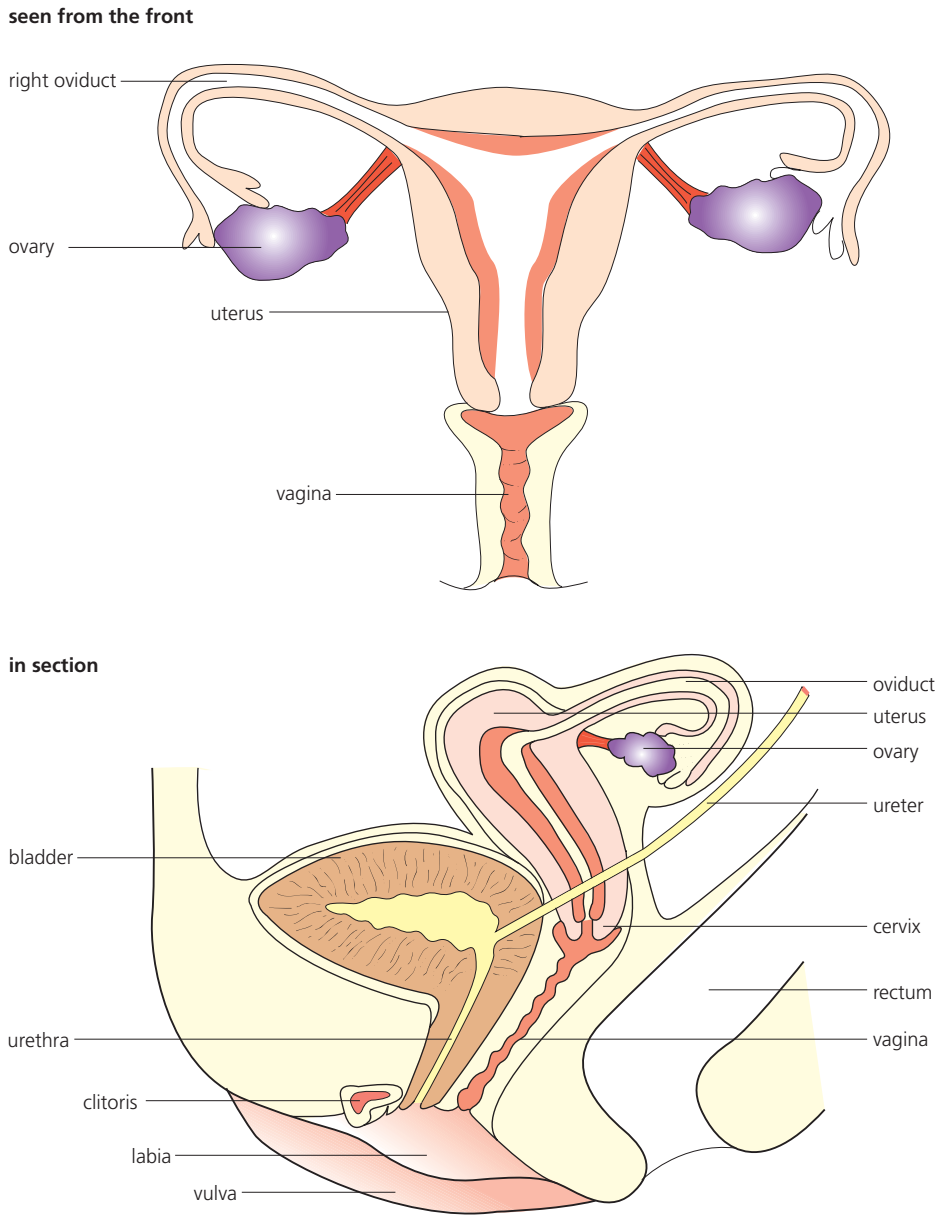


Figure 7.50 The female urinogenital system

Primary and secondary sexual characteristics

The primary sexual characteristics of a male or a female are possession of their respective reproductive organs. We have seen that in humans, gender is primarily determined by the sex chromosomes, XX or XY (Figure 4.22, page 112).

The **secondary sexual characteristics** are those that develop in males and females at **puberty**, the time in growth and development at the beginning of sexual maturation. At this time there is a significant increase in the production of the **sex hormones** by the gonads, testosterone in cells of the testes, and oestrogen and progesterone in cells of the ovaries. These hormones are chemically very similar molecules; they are manufactured from the steroid **cholesterol** that has been synthesised in the liver and absorbed as part of the diet. (Steroids are a form of lipid.)

Perhaps the most noticeable effect of the increased secretion of the sex hormones is the stimulation they cause in muscle protein formation and bone growth. Because of this effect, testosterone, oestrogen and progesterone are known as **anabolic steroids** (anabolic means 'build up'). The effects of the female sex hormones are less marked, in this respect, than those of testosterone in the male, and the onset of puberty occurs, on average, about two years earlier in girls (Figure 7.51).

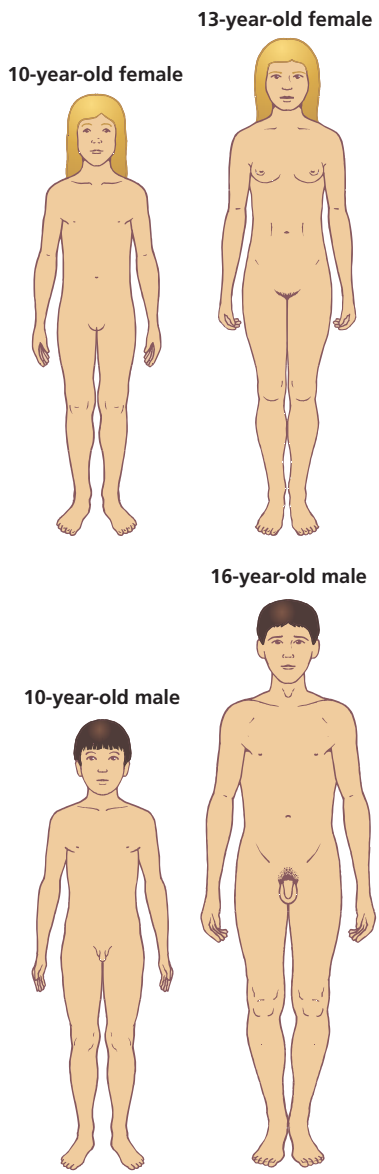


Figure 7.51 The onset of the secondary sexual characteristics

Secondary sexual characteristics of a human female:

- maturation of the ovaries, and enlargement of the vagina and uterus
- development of the breasts
- widening of the pelvis
- deposition of fat under the skin of the buttocks and thighs
- growth of pubic hair and hair under the arm pits
- monthly ovulation and menstruation
- changes in behaviour associated with a sex drive.

Secondary sexual characteristics of a human male:

- development and enlargement of the testes, scrotum, penis and glands of the reproductive tract
- increased skeletal muscle development
- enlargement of the larynx, deepening the voice
- growth of pubic hair, underarm hair, and body hair
- continuous production of sperms, and in the absence of sexual intercourse, occasional erections and the discharge of seminal fluid
- changes in behaviour associated with a sex drive.

Roles of hormones in the control of reproduction

The onset of puberty is triggered by a part of the brain called the hypothalamus. Here, production and secretion of a releasing hormone causes the nearby pituitary gland (the 'master' endocrine gland) to produce and release into the blood circulation two hormones, **follicle-stimulating hormone (FSH)** and **luteinising hormone (LH)**. They are so named because their roles in sexual development were discovered in the female, although they do operate in both sexes. Their first effects are to enhance secretions of the sex hormones. Then, in the presence of FSH, LH and the respective sex hormone, there follows the development of the secondary sexual characteristics, and the preparation of the body for its role in sexual reproduction.

In females

In the female reproductive system, the secretion of oestrogen and progesterone is cyclical, rather than at a steady rate. Together with FSH and LH, the **changing concentrations of all four hormones** bring about a repeating cycle of changes that we call the **menstrual cycle**.

The menstrual cycle consists of **two cycles**, one in the **ovaries** and one in the **uterus lining**. The ovarian cycle is concerned with the monthly preparation and shedding of an egg cell from an ovary, and the uterus cycle, with the build-up of the lining of the uterus. 'Menstrual' means 'monthly'; the combined cycles take 28 days (Figure 7.52).

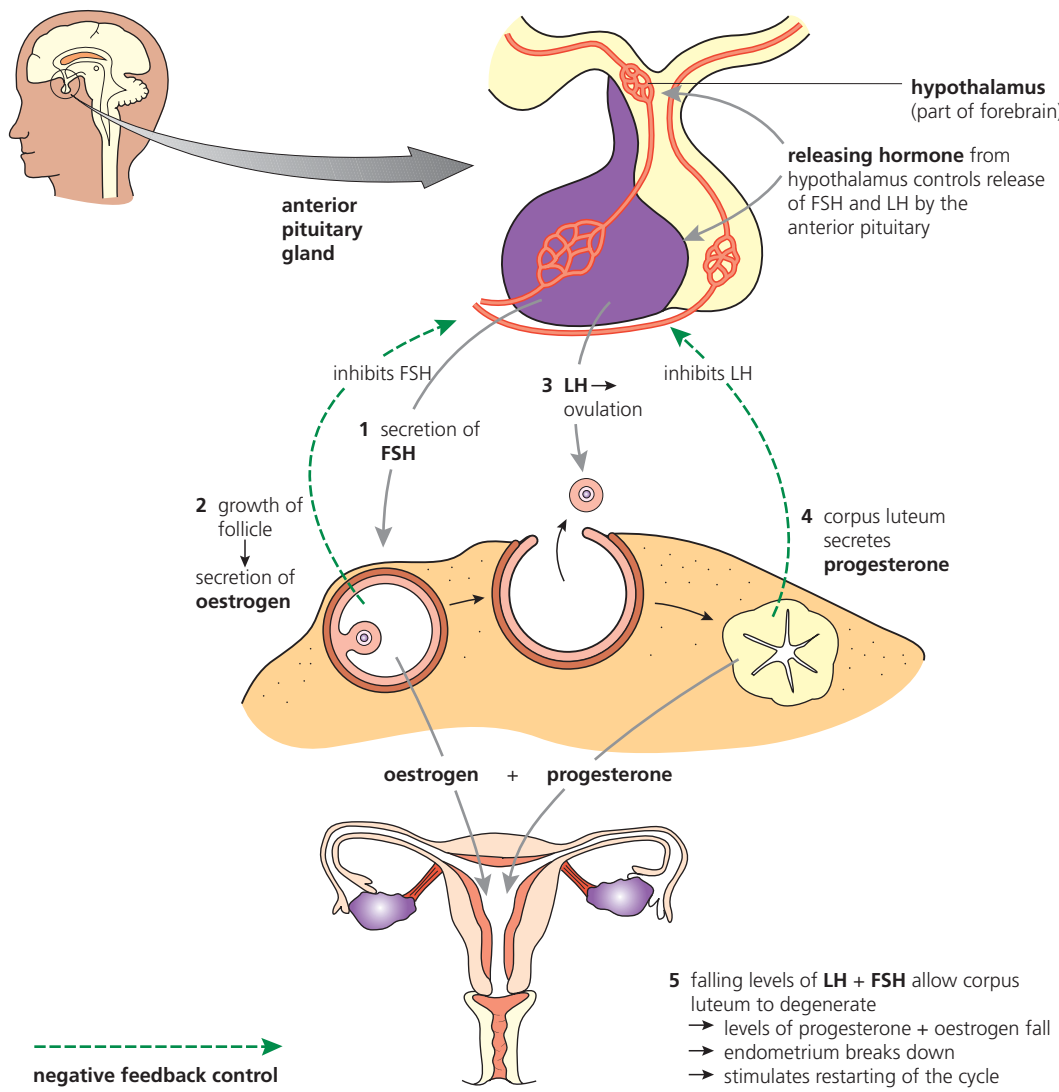


Figure 7.52 Hormone regulation of the menstrual cycle

By convention the start of the cycle is taken as the first day of menstruation (bleeding), which is the shedding of the endometrium lining of the uterus. The steps, also summarised in Figure 7.53, are as follows.

1 **FSH** is secreted by the pituitary gland, and stimulates development of several immature egg cells (in primary follicles) in the ovary. Only one will complete development into a mature egg cell (now in the ovarian follicle).
2 The developing follicle then secretes **oestrogen**. Oestrogen has two targets:
   a in the uterus, it stimulates the build-up of the endometrium, the lining for a possible implantation of an embryo, should fertilisation take place;

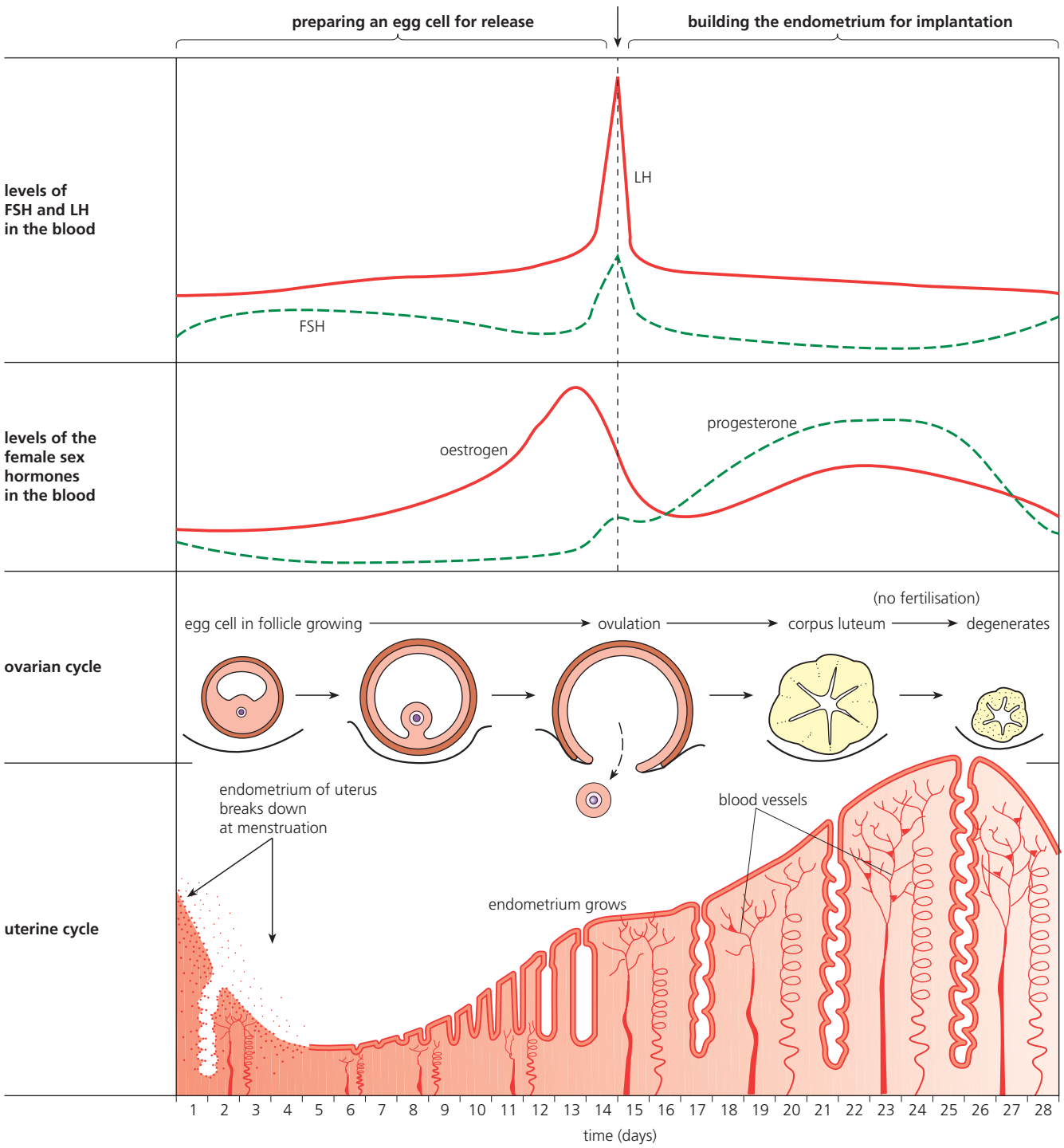


Figure 7.53 Changing levels of hormones in the menstrual cycle

b in the pituitary gland, oestrogen inhibits the further secretion of FSH, which prevents the possibility of further follicles being stimulated to develop, and is an example of **negative feedback control** (Figure 7.53).

3 The concentration of oestrogen continues to increase to a peak value just before the mid-point of the cycle. This high and rising level of oestrogen suddenly stimulates the secretion of **LH** and, to a slightly lesser extent, FSH, by the pituitary gland. LH stimulates **ovulation** (the shedding of the mature egg cell from the ovarian follicle and its escape from the ovary).

As soon as the ovarian follicle has discharged its egg, LH also stimulates the conversion of the vacant follicle into an additional, temporary gland, called a **corpus luteum**.

4 The corpus luteum secretes **progesterone** and, to a lesser extent, oestrogen. Progesterone has two targets:

a in the uterus, it continues the build-up of the endometrium, further preparing for a possible implantation of an embryo, should fertilisation take place;

b in the pituitary gland, it inhibits further secretion of LH, and also of FSH; this is a second example of **negative feedback control**.

5 The levels of FSH and LH in the blood stream now rapidly decrease. Low levels of FSH and LH allow the corpus luteum to degenerate. As a consequence, the level of progesterone and oestrogen also fall. Soon the level of these hormones is so low that the extra lining of the uterus is no longer maintained. The **endometrium breaks down** and is lost through the vagina in the first five days or so of the new cycle. Falling levels of progesterone again cause the secretion of FSH by the pituitary.

A new cycle is under way.

6 **If the egg is fertilised** (the start of a pregnancy), then the developing embryo itself immediately becomes an endocrine gland, secreting a hormone that circulates in the blood and maintains the corpus luteum as an endocrine gland for at least 16 weeks of pregnancy. When eventually the corpus luteum does break down, the **placenta** takes over as an endocrine gland, secreting oestrogen and progesterone. These hormones continue to prevent ovulation, and maintain the endometrium.

27 **Identify** the critical hormone changes that respectively trigger ovulation, and cause degeneration of the corpus luteum.

In males

In the male, the secretion of sex hormone commences in pre-natal development, and is a continuous process, rather than cyclic. Testosterone secretion has the following roles:

- it initiates the pre-natal development of male genitalia;
- it triggers and regulates the development of secondary sexual characteristics;
- it maintains the sex drive (libido) in the adult.

Infertility, and *in vitro* fertilisation

In mammals, fertilisation – the fusion of male and female gametes to form a **zygote** – is internal. It occurs in the upper part of the oviduct. As the zygote is transported down the oviduct, mitosis and cell division commence. By the time the embryo has reached the uterus, it is a solid ball of tiny cells. Division continues and the cells organise themselves into a fluid-filled ball, the **blastocyst**, which becomes embedded in the endometrium of the uterus, a process known as **implantation**. In humans, implantation takes from day 7 to day 14 approximately (Figure 12.35, page 384).

Not all partners are fertile in this way. Either the male or female or both may be infertile, due to a number of different causes (Table 7.13).

In some cases, a couple's infertility may be overcome by the process of fertilisation of eggs outside the body (***in vitro* fertilisation**, **IVF**). The key step in IVF is the successful removal of sufficient eggs from the ovaries. To achieve this, normal menstrual activity is temporarily suspended with hormone-based drugs.

Then the ovaries are induced to produce a large number of eggs simultaneously, at a time controlled by the doctors. In this way, the correct moment to collect the eggs can be known accurately.

Extension: Contraception

People who make a study of human population growth (demographers) have predicted that by 2050 the Earth may have 10 billion human inhabitants – that is, very many more than we currently have. The projections suggest that the world population will stabilise at this level. Can such a total human population be sustained by the Earth's resources?

If we think not, an immediate task would be to reduce the **birth rate**. The mechanisms for birth control exist. **Contraception** is any procedure which prevents conception. There are many different methods, some easier and more reliable than others. It is a convention to divide them into **mechanical**, **chemical** and **behavioural** (Table 7.12).

Mechanical	barrier between sperm and egg	■ condom – sheath fitted to erect penis ■ femidom** – sheath fitted to the vagina ■ diaphragm or cap** – 'cup' fitted over the cervix
	prevention of implantation	■ loop or coil** – fitted inside the uterus
Chemical – prevention of ovulation by use of hormones	combined pill	■ oestrogen and progesterone** – daily for 21 days
	mini-pill	■ progesterone** – daily at the same time each day
Behavioural	rhythm method – calculating a time when fertile eggs are not present and sperms cannot survive until ovulation	■ if ovulation occurs at days 13–15, egg may survive till day 18 and sperms introduced on or after day 8 may survive to ovulation; so intercourse not practised on days 8–18; ovulation detected by body temperature changes

** methods that can be used without involvement of the male partner, thereby giving the female partner control of contraception

Table 7.12 Alternative methods of contraception

Extension: Sexually transmitted diseases (STDs) and contraception

STDs cause enormous suffering worldwide. However, apart from AIDS, they are seldom mentioned or discussed.

Earlier in the last century, the common STDs were the bacterial infections **gonorrhoea** and **syphilis**, both becoming particularly prevalent in communities after both the First and Second World Wars.

Today, STDs due to *Chlamydia* (a bacterium), to herpes simplex virus, and to human papilloma viruses (genital warts) are adding to patient numbers. AIDS and the special factors in HIV transmission have already been discussed (page 198).

STDs are mostly transmitted between infected people who are sexually active. The more sexual partners a person has, the greater their risk of infection. In many cases, 'safe sex' using **barrier contraception** (such as the condom) **reduces the risks**, but now that oral contraception may be preferred, risks have increased enormously. Another source of infection comes at birth for newborn babies of infected mothers.

In males	In females
failure to achieve or maintain an erect penis	conditions in cervix cause death of sperms
structurally abnormal sperms sperms with poor mobility short-lived sperms too few sperms	conditions in uterus prevent implantation of blastocyst
blocked sperm duct preventing semen from containing sperms	oviducts blocked or damaged, preventing egg from reaching sperms
	eggs fail to mature or be released

Table 7.13 Causes of infertility

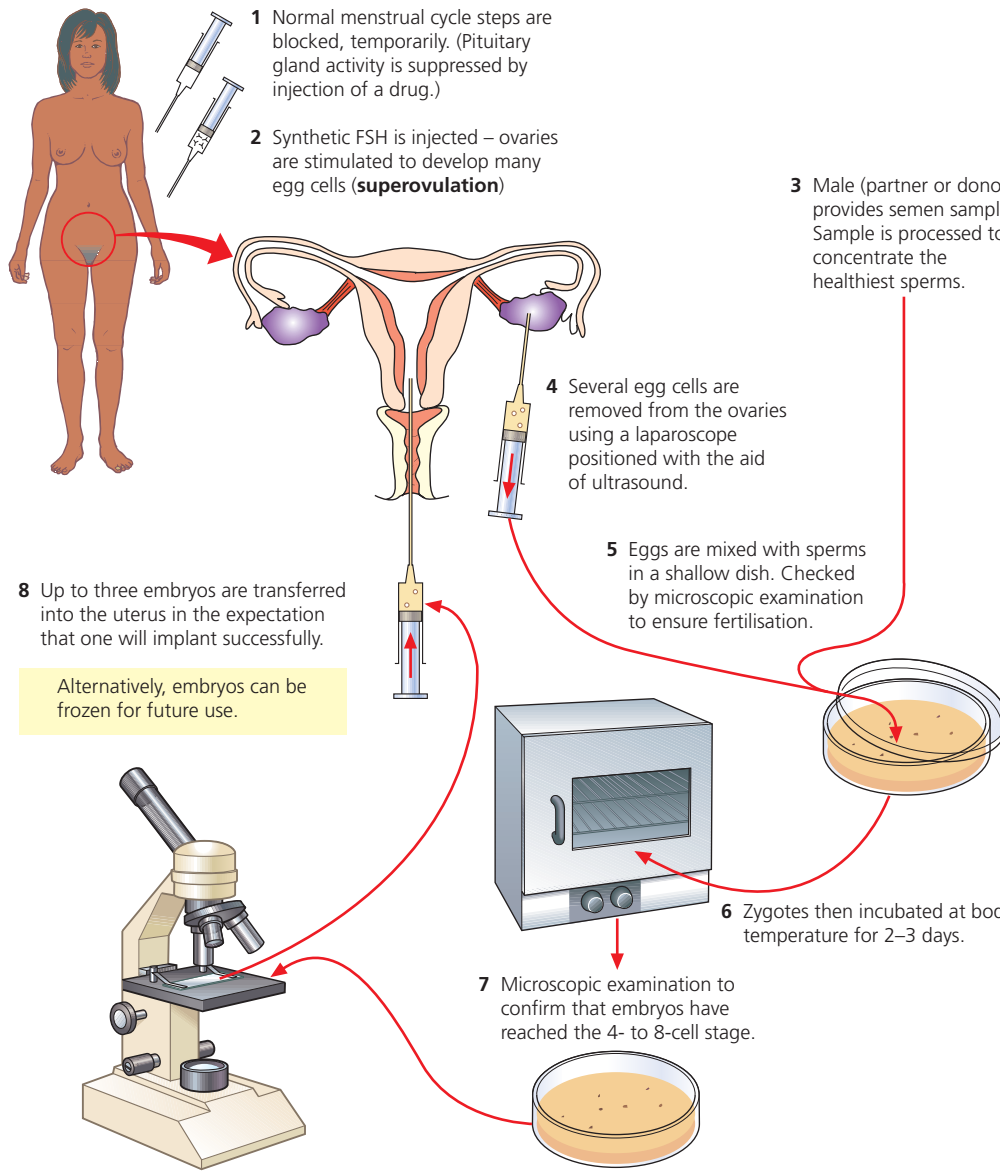


Figure 7.54 *In vitro* fertilisation – the process

Egg cells are then isolated from surrounding follicle cells, and mixed with sperms. If fertilisation occurs, the fertilised egg cells are incubated so that embryos at the eight-cell stage may be placed in the uterus. If one (or more) imbed there, then a normal pregnancy may follow.

The first 'test-tube baby' was born in 1978. The steps of IVF are illustrated in Figure 7.54. Today, the procedure is regarded as a routine one.

Ethical issues raised by IVF

IVF raises many issues for society. Infertility is a personal problem that generates stress and unhappiness in those affected. Many of these couples may seek assistance. Medical treatments involved are expensive, and their deployment inevitably deflects finite resources away from other needs. Success rates are very low.

In the face of the world's booming human population and the large numbers of parentless children seeking adoption, some argue against these sorts of development in reproduction technology.

Another controversial issue is that current fertility treatments tend to create several embryos, all with the potential to become new people. However, the embryos that are not selected for implantation ultimately must be destroyed – a stressful task in itself.

Ethical issues are listed in Table 7.14.

Favourable arguments	Critical arguments
For some otherwise childless couples, desired parenthood may be achieved.	Allows infertility due to inherited defects to be passed on (unwittingly) to the offspring who may then experience the same problem in adulthood.
Allows men and women surviving cancer treatments the possibility of having children later, using gametes harvested prior to radiation treatment or chemotherapy.	Excess embryos are produced to ensure success, and so an embryologist has to select some new embryo(s) to live, and to allow the later destruction of other potential human lives.
Permits screening and selection of embryos before implantation stage to avoid an inherited disease.	Multiple pregnancies have been a common outcome, sometimes producing triplets, quads or sextuplets. This can lead to increased risk to the mother's health, risks of premature birth, and put the babies at risk of conditions such as cerebral palsy.
If IVF treatment were to be banned, the state may be interfering in the lives of individuals with medical problems that otherwise could be cured.	Infertility is not always strictly a health problem; it may have arisen in older parents who chose to delay having a family (a life-style issue).
Offspring produced by IVF are much longed-for children who will more certainly be loved and cared for.	There is an excess of unwanted children, cared for in orphanages or in foster homes. These children may have benefited from adoption by couples, childless or otherwise, keen to be caring parents.

Table 7.14 Ethical issues raised by IVF

28 Outline key points that might be put by a genetic counsellor who felt that some alternative way to establish a family was more appropriate for a particular childless couple seeking IVF treatment.

Examination questions – a selection

Questions 1–8 are taken from past IB Diploma biology papers.

Q1 **a** Draw a diagram of the heart showing the chambers, valves and associated blood vessels. (4)

b Outline the control of the heart beat. (6)

c Discuss the structure and function of the blood vessels that are found in a human. (8)

Standard Level Paper 2, May 06, QB5

Q2 Which of the following is correct for a pathogen?

	Can be a virus	Can cause antibody response	Is an antigen
A	✔	✘	✘
B	✔	✔	✘
C	✘	✔	✔
D	✔	✔	✔

Higher Level Paper 1, May 05, Q18

Q3 Which of the following correctly explains the functions of parts of the digestive system?

	Stomach	Small intestine	Large intestine
A	digests proteins	absorbs vitamin K	absorbs water
B	absorbs water	digests carbohydrates	digests proteins
C	digests lipids	digests proteins	absorbs water
D	digests proteins	absorbs glucose	absorbs water

Standard Level Paper 1, May 05, Q25

Q4 **a** Explain how the skin and mucous membranes prevent entry of pathogens into the body. (3)

b Explain why antibiotics are used to treat bacterial but not viral diseases. (2)

Standard Level Paper 2, May 03, Q5

Q5 Which hormone affects the heart beat?

A glucagon
B insulin
C adrenaline
D oxytocin

Higher Level Paper 1, May 04, Q19

Q6 **a** Draw a labelled diagram of an adult male reproductive system. (4)

b Outline the process of *in vitro* fertilisation (IVF). (6)

c Discuss the advantages and disadvantages of genetic engineering. (8)

Standard Level Paper 2, May 04, QB5

Q7 **a** Draw a diagram of the human digestive system. (4)

b Describe the role of enzymes in the process of digestion of proteins, carbohydrates and lipids in humans. (6)

c Explain how blood glucose concentration is controlled in humans. (8)

Standard Level Paper 2, May 03, QB8

Q8 Where are the chemoreceptors that detect the changes in blood pH and levels of glucose found?

	Changes in blood pH	Changes in blood glucose
A	brain stem	small intestine
B	carotid vein	liver
C	carotid artery	pancreas
D	venae cavae	liver

Higher Level Paper 1, May 02, Q17

Questions 9–12 cover other syllabus issues in this chapter.

Q9 **a** In the absorption of digested food molecules in the small intestine, explain where each of the following structures occurs and how each contributes to the absorption process:

i villi (2)
ii capillary networks (2)
iii microvilli (2)
iv protein pumps. (2)

b Explain the term *assimilation*, and state where it occurs in the body. (3)

Q10 Blood has roles in transport and in the body's defences against disease. By means of a table, identify the specific roles of each component of the blood (plasma, red cells, white cells, platelets) in these functions. (8)

Q11 **a** The human immune deficiency virus (HIV) is a retrovirus that is responsible for the disease called AIDS. Explain what *retrovirus* means. (3)

b With the onset of AIDS, the immune system is said to become 'hopelessly compromised'. Explain this, and outline how HIV brings this about. (3)

Q12 **a** In gaseous exchange, oxygen from the air reaches mitochondria in respiring cells of the body. Identify the precise pathway that oxygen takes and the processes by which it is moved at each step in this journey into the innermost human body cells. (6)

b The regulation of the concentrations of oxygen and carbon dioxide in the blood is by negative feedback. Explain the principle and outline the mechanism of negative feedback in homeostasis. (4)

8 Nucleic acids and proteins

STARTING POINTS

- The **chromosomes** of eukaryotes are found in the nucleus and consist of **DNA** and **protein**.
- DNA exists as a **double helix**, the strands held by **complementary base pairing** and **hydrogen bonds**.
- **Replication of DNA** occurs during interphase of **mitosis** by unwinding of the double helix and formation of new complementary strands.
- In **protein synthesis**, the **genetic code** of the chromosome is **transcribed** into a single strand of **messenger RNA** (**mRNA**) which passes out to ribosomes in the cytoplasm.
- In a **ribosome** the sequence of triplets of bases in mRNA is **translated** into the sequence of amino acids that condense together to form **proteins**.
- This chapter extends the study of nucleic acids, proteins and enzymes begun in Chapter 2, pages 37–75.

We have seen that **nucleic acids** are very long, thread-like macromolecules of alternating sugar and phosphate molecules (forming a 'backbone') with a nitrogenous base – either cytosine (C), guanine (G), adenine (A), thymine (T), or uracil (U) – attached to each sugar molecule along the strand. Furthermore, nucleic acids are described as the information molecules of cells, and they fulfil this role throughout the living world – the genetic code is a universal one.

The structures of the two types of nucleic acid, **deoxyribonucleic acid** (**DNA**), and **ribonucleic acid** (**RNA**), are compared in Table 2.7, page 69. We have also noted that in eukaryotes, while DNA occurs only in the nucleus, RNA is found in both the cytoplasm and the nucleus.

1 **Distinguish** between an organic base, a nucleoside, a nucleotide, and a nucleic acid.

In this chapter, the way **DNA is structured** and **packaged** in chromosomes is discussed, prior to looking again at the **replication of DNA**, the **transcription** of the genetic code, and its **translation** into the sequence of amino acids selected to build specific polypeptides. Finally, the structure and roles of **cell proteins**, particularly those that function as **enzymes**, are examined.

Chromosome structure and the packaging of DNA

7.7.1–7.7.5

DNA occurs in the chromosomes in the nucleus, along with protein. Actually, more than 50% of a chromosome is built of **protein**. While some of the proteins of the chromosome are enzymes involved in copying and repair reactions of DNA, the bulk of chromosome protein has a **support and packaging role** for DNA.

Why is packaging necessary?

Well, take the case of human DNA. In the nucleus, the total length of the DNA of the chromosomes is over 2 metres. We know this is shared out between 46 chromosomes, and that each chromosome contains one, very long DNA molecule. Chromosomes are different lengths (Figure 4.1, page 92), but we can estimate that within a typical chromosome of 5 μm length, there is a DNA molecule approximately 5 cm long. This means that about 50 000 μm of DNA is packed into 5 μm of chromosome.

This phenomenal packaging is achieved by the much-coiled DNA double helix being looped around protein beads called **nucleosomes**, as illustrated in Figure 8.1.

The packaging protein of the nucleosome, called **histone**, is a basic (positively charged) protein containing a high concentration of amino acid residues with additional base groups ($-NH_2$), such as lysine and arginine (Figure 8.17, page 254). In nucleosomes, eight histone molecules combine to make a single bead. Around each bead, the DNA double helix is wrapped in a double loop.

Figure 8.1 The structure of a nucleosome

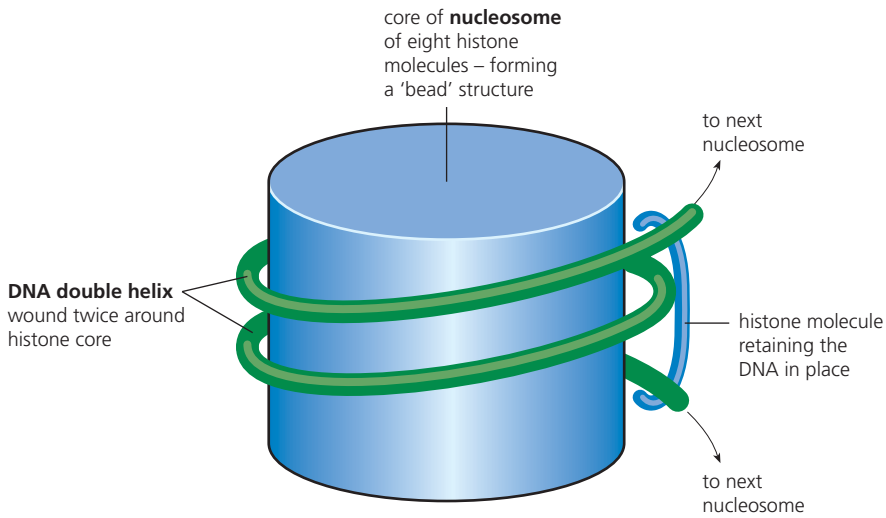
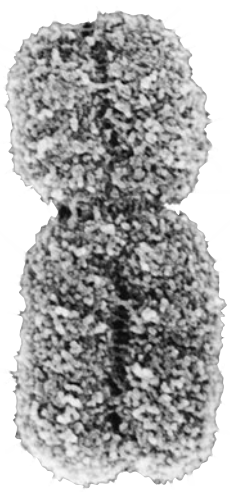
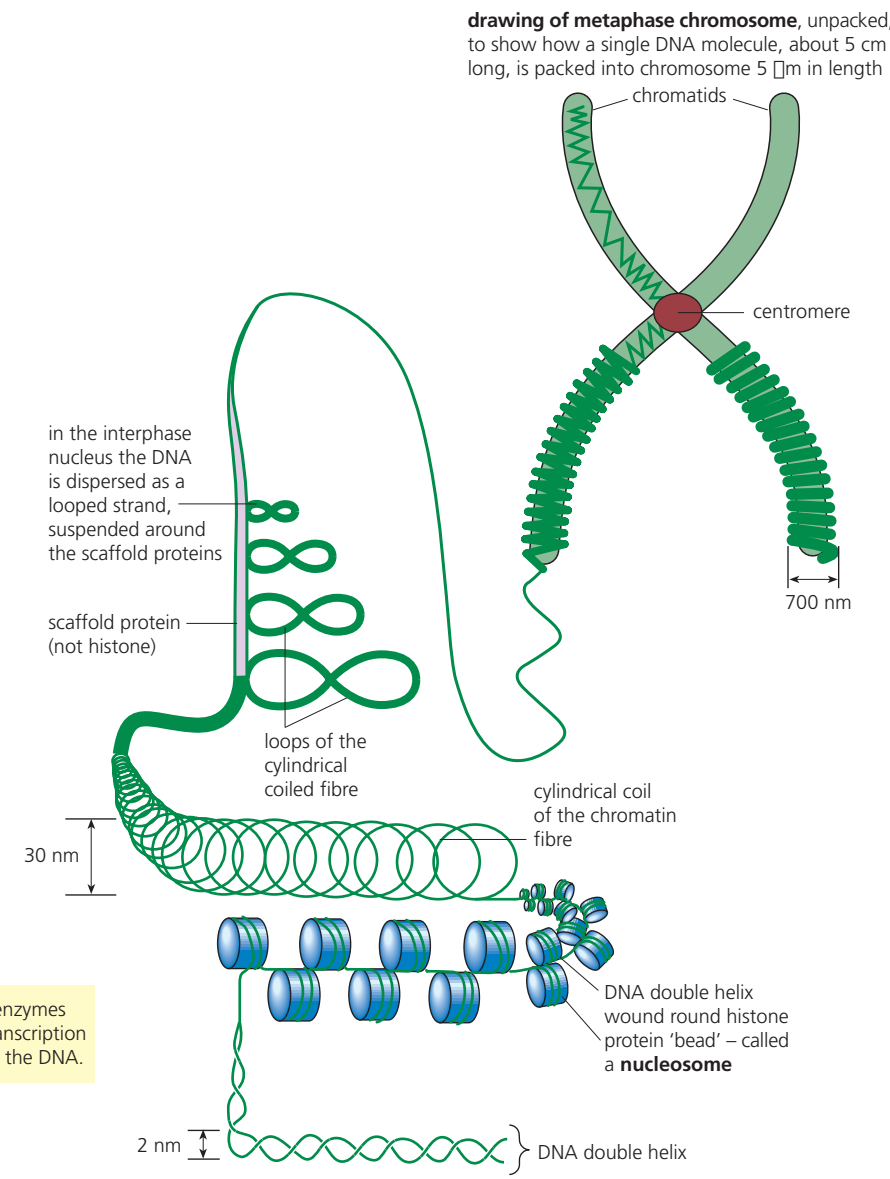


Figure 8.2 The packaging of DNA in the chromosome



electron micrograph of metaphase chromosome (×40 000) – at this stage the chromosome is at maximum condensed state (supercoiled)



Also present are enzymes involved in the transcription and replication of the DNA.

2 Suggest a main advantage of chromosomes being 'supercoiled' in metaphase of mitosis.

The whole beaded thread is itself coiled up, forming the chromatin fibre. The chromatin fibre is again coiled, and the coils are looped around a 'scaffold' protein fibre, made of a **non-histone protein**. This whole structure is folded (supercoiled) into the much-condensed metaphase chromosome, as shown in Figure 8.2. Clearly, the nucleosomes are the key structures that facilitate supercoiling of these phenomenal lengths of DNA that are packed in the nuclei. Furthermore, they facilitate access to selected lengths of the DNA (particular genes) during transcription – a process we will discuss shortly.

Discovery of the role of DNA as information molecule

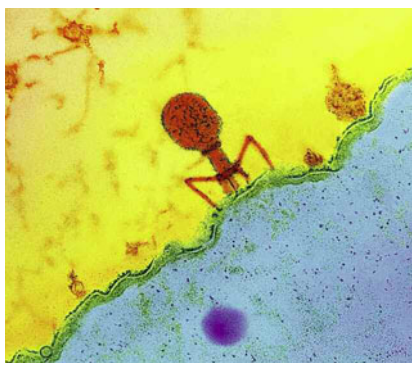
Since about 50% of a chromosome consists of protein, it is no wonder that people once speculated that protein of the chromosomes might be the information substance of the cell. For example, there is more chemical 'variety' within protein's structure than in nucleic acid. However, this idea proved incorrect. We now know that the DNA of the chromosomes holds the information that codes for the sequence of amino acids from which the proteins of the cell cytoplasm are built.

How was this established?

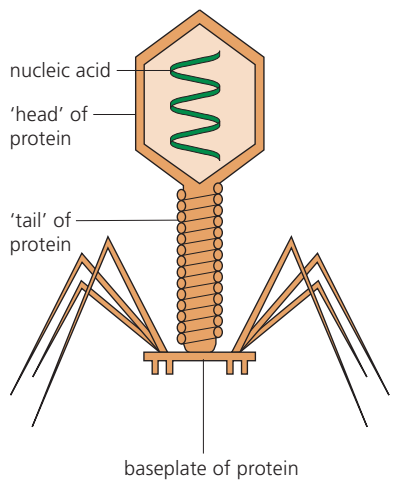
The unique importance of DNA was shown by an experiment with a bacteriophage virus. A **bacteriophage** (or **phage**, as it is known) is a virus that parasitises a bacterium. We know that viruses consist of a protein coat surrounding a nucleic acid core (page 19). Once a virus has gained entry to a host cell it may take over the cell's metabolism, switching it to the production of new viruses. Eventually, the remains of the host cell break down (lysis) and the virus particle escapes, usually to repeat the infection with new host cells. The life cycle of a bacteriophage, a virus with a complex 'head' and 'tail' structure, is shown in Figure 8.3.

Figure 8.3 The life cycle of a bacteriophage

TEM of bacteriophage infecting a bacterium

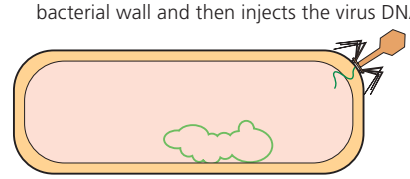


structure of the phage



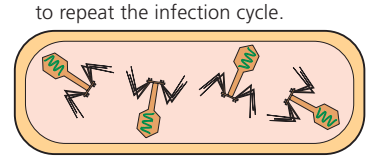
steps to replication of the phage

1 The phage attaches to the bacterial wall and then injects the virus DNA.



2 Virus DNA takes over the host's synthesis machinery.

3 New viruses are assembled and then escape to repeat the infection cycle.



Two experimental scientists, Martha Chase and Alfred Hershey, used a bacteriophage that parasitises the bacterium *Escherichia coli* (page 16) to answer the question of whether genetic information lies in the protein (coat) or the DNA (core) (Figure 8.4).

Two batches of the bacteriophage were produced, one with radioactive phosphorus (^{32}P) built into the DNA core (hence the DNA was labelled) and one with radioactive sulphur (^{35}S) built into the protein coat (hence the protein was labelled). Sulphur occurs in protein, but there is no sulphur in DNA. Likewise, phosphorus occurs in DNA, but there is no phosphorus in protein. Thus the radioactive labels were specific.

Two identical cultures of *E. coli* were infected, one with the ^{32}P-labelled virus and one with the ^{35}S-labelled virus. Subsequently, radioactively labelled viruses were obtained only from the bacteria infected with virus labelled with ^{32}P. In fact, the ^{35}S label did not enter the host cell at all. Their experiment clearly demonstrated that it is the DNA part of the virus which enters the host cell and carries the genetic information for the production of new viruses.

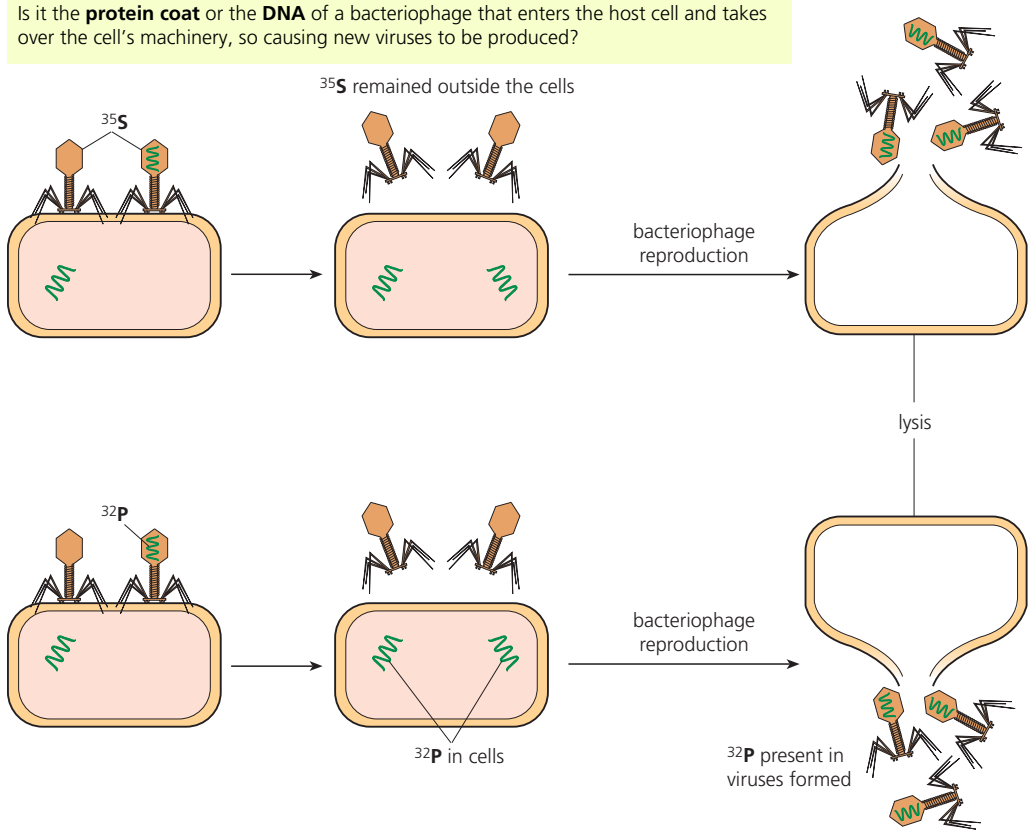
Figure 8.4 The Hershey–Chase experiment

Is it the **protein coat** or the **DNA** of a bacteriophage that enters the host cell and takes over the cell's machinery, so causing new viruses to be produced?

The question was answered using:

1 phage labelled with radioactive sulphur (^{35}**S**) – sulphur is a component of protein *but not of DNA*

2 phage labelled with radioactive phosphorus (^{32}**P**) – phosphorus is a component of DNA *but not of protein*



Only the DNA part of the virus got into the host cell (and radioactively labelled DNA was present in the new viruses formed). It was the virus DNA that controlled the formation of new viruses in the host, so Hershey and Chase concluded that **DNA carries the genetic message**.

3 Deduce what would have been the outcome of the Hershey–Chase experiment if protein had been the carrier of genetic information.

Base pairing and 'direction' in the DNA molecule

The structure of the DNA molecule as a double helix, proposed by Watson and Crick in 1953, is illustrated in Figure 2.27 (page 65). In terms of its function, the key feature of DNA is **base pairing**.

Discovery of the principle of base pairing by Watson and Crick was achieved by interpretation of the work of Edwin Chargaff. In 1935, Chargaff had analysed the composition of DNA from a range of organisms, and found rather remarkable patterns. The significance of these patterns was not immediately obvious, though. His discoveries were:

- the numbers of purine bases (adenine and guanine) always equalled the number of pyrimidine bases (cytosine and thymine);
- the number of adenine bases equalled the number of thymine bases, and the number of guanine bases equalled the number of cytosine bases.

What does this mean?

The organic bases found in DNA are of two distinct types with contrasting shapes:

- cytosine and thymine are pyrimidines or **single-ring** bases;
- adenine and guanine are purines or **double-ring** bases.

Only a **purine will fit with a pyrimidine** between the sugar–phosphate backbones, when base pairing occurs (Figure 8.5).

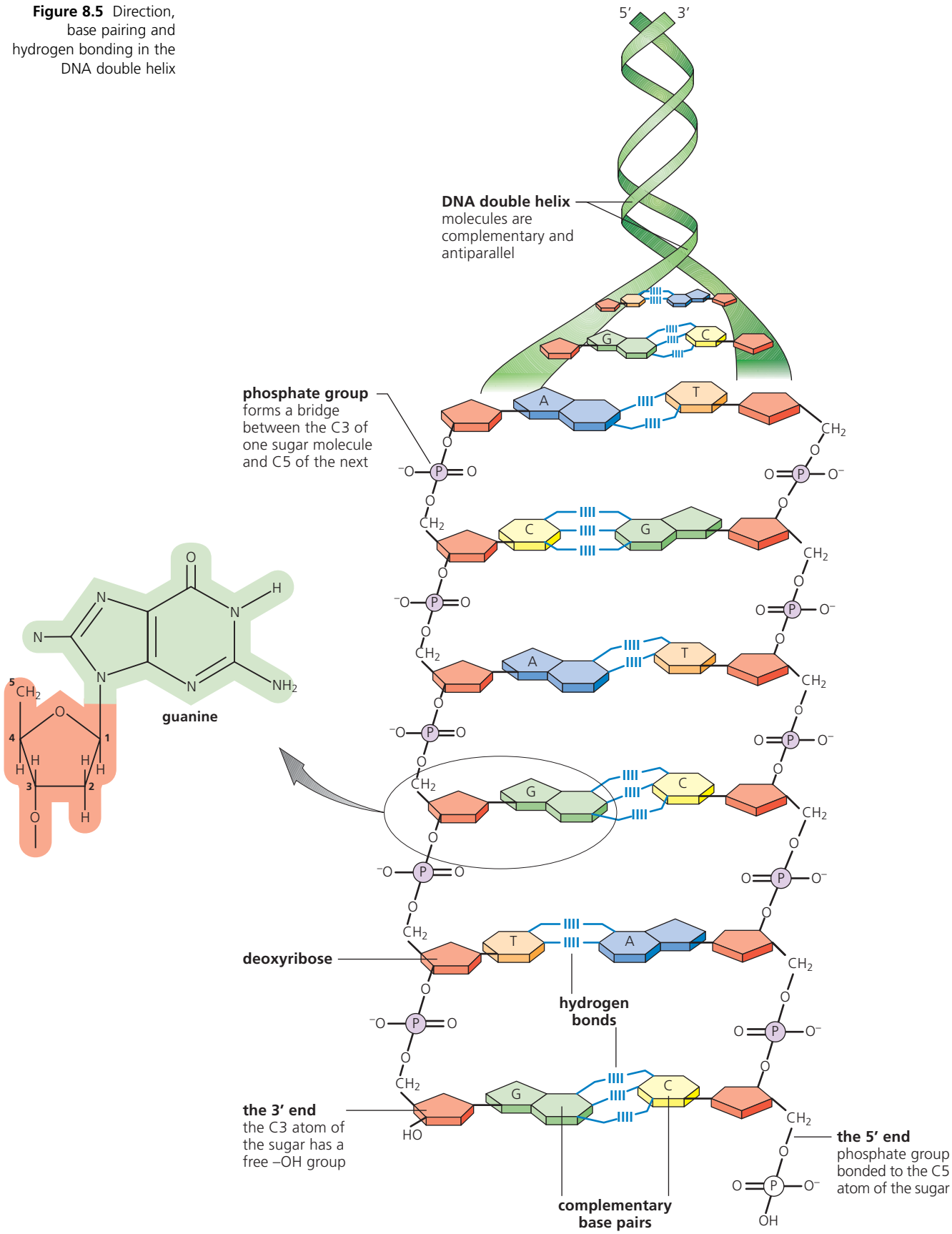


Figure 8.5 Direction, base pairing and hydrogen bonding in the DNA double helix

From the model of DNA in Figure 8.5 we can also see that when A pairs with T, they are held together by two hydrogen bonds; when C pairs with G, they are held by three hydrogen bonds. **Only these pairs can form hydrogen bonds.**

Because of base pairing and the formation of specific hydrogen bonds, the sequence of bases in one strand of the helix determines the sequence of bases in the other – a principle we know as complementary base pairing.

4 In base pairing, organic bases are held together (A–T, C–G) by hydrogen bonds. **Describe** what a hydrogen bond is.

Note that we can identify **direction** in the DNA double helix. The phosphate groups along each strand are bridges between carbon-3 of one sugar molecule and carbon-5 of the next, and one chain runs from 5' to 3' while the other runs from 3' to 5'. That is, the two chains of DNA are **antiparallel**, as illustrated in Figure 8.5. The existence of direction in DNA strands becomes important in replication and when the genetic code is transcribed into mRNA.

Chromosomes, genes and 'nonsense' DNA sequences

We know more about the structure of the human genome than about any other vertebrate genome. The **Human Genome Project**, the initiative to map the entire human genome, has established that the three million bases of our chromosomes represent far fewer genes than was expected (page 131). In fact, only a relatively small part of the total DNA in the nucleus occurs in genes. It is a fact that the bulk of our DNA does not code for proteins at all.

So, what are the different types of DNA present?

Well, about 30% of the human genome is made up of **genes** and base sequences relating to genes. The latter category includes the exons and introns of genes, of which more shortly (page 245). Also included are regions that contain promoters and regulatory sequences, of which there will again be further discussion shortly (page 244).

That leaves 70% of our DNA in the category of non-coding or **'nonsense' DNA**. Of this:

- about 80% of sequences are unique or low-copy-number sequences of bases;
- about 20% are moderately and highly repetitive DNA, and are divided as follows:
  - sequences in which base sequences are repeated from a few times up to 10^5 times per genome (moderately repetitive);
  - sequences in which between 5 and 300 base pairs per repeat occur, duplicated more than 10^5 times per genome (highly repetitive); it is these sequences that make up the **satellite DNA**, which is successfully exploited in **DNA profiling** (page 129).

We will return to the issue of nonsense DNA briefly, when we discuss the maturation of mRNA prior to it leaving the nucleus for the ribosomes.

Replication – DNA copying itself

7.2.1–7.2.3

The way that DNA of the chromosome is replicated was introduced in Chapter 2 (page 66). The key features are:

- replication must be an extremely accurate process since DNA carries the genetic message;
- replication is quite separate from cell division; replication of DNA takes place in the **interphase** nucleus, well before the events of nuclear division;
- strands of the DNA double helix are built up individually from free **nucleotides** (the structure of a nucleotide is shown in Figure 2.25, page 63);
- before nucleotides can be condensed together, the DNA double helix has to **unwind**, and the hydrogen bonds holding the strands together must be broken, allowing the two strands of the helix to separate;
- the enzyme **helicase** brings about the unwinding process and holds the strands apart for replication to occur;
- **both strands act as templates**; nucleotides with the appropriate complementary bases line up opposite the bases of the exposed strands (A with T, C with G);

- **hydrogen bonds form** between complementary bases, holding the nucleotides in place;
- finally, the sugar and phosphate groups of adjacent nucleotides of the new strand condense together, catalysed by the enzyme **DNA polymerase**;
- the two strands of each DNA molecule wind up into a double helix;
- one strand of each new double helix came from the parent chromosome and one is a newly synthesised strand, an arrangement known as **semi-conservative replication** because half the original molecule is conserved;
- DNA polymerase also has a role in '**proof-reading**' the new strands; any mistakes that start to happen (such as the wrong bases attempting to pair up) are immediately corrected and each new DNA double helix is exactly like the original.

5 Predict the experimental results you would expect to see if the Meselson–Stahl experiment (Figure 8.6) were carried on for three generations.

Extension: The evidence for DNA replication

Experimental evidence that DNA is replicated semi-conservatively, and not by some other mechanism, came from growing a culture of the bacterium *E. coli* (page 16) in a medium (food source) where the available nitrogen contained only the heavy nitrogen isotope, ^{15}N. Consequently, the DNA of the bacterium became entirely 'heavy'.

These bacteria were then transferred to a medium of the normal (light) isotope, ^{14}N. New DNA manufactured by the cells was now made of ^{14}N. The change in concentration of ^{15}N and ^{14}N in the DNA of succeeding generations was measured. Interestingly, the bacterial cell divisions in a culture of *E. coli* are naturally synchronised; every 60 minutes, they all divided again.

The DNA was extracted from samples of the bacteria from each succeeding generation and the DNA in each sample was separated. This was done by placing the sample on top of a salt solution of increasing density, in a centrifuge tube. On being centrifuged, the different DNA molecules were carried down to the level where the salt solution was of the same density. Thus, DNA with 'heavy' nitrogen ended up nearer the base of the tubes, whereas DNA with 'light' nitrogen stayed near the top of the tubes. Figure 8.6 shows the results that were obtained.

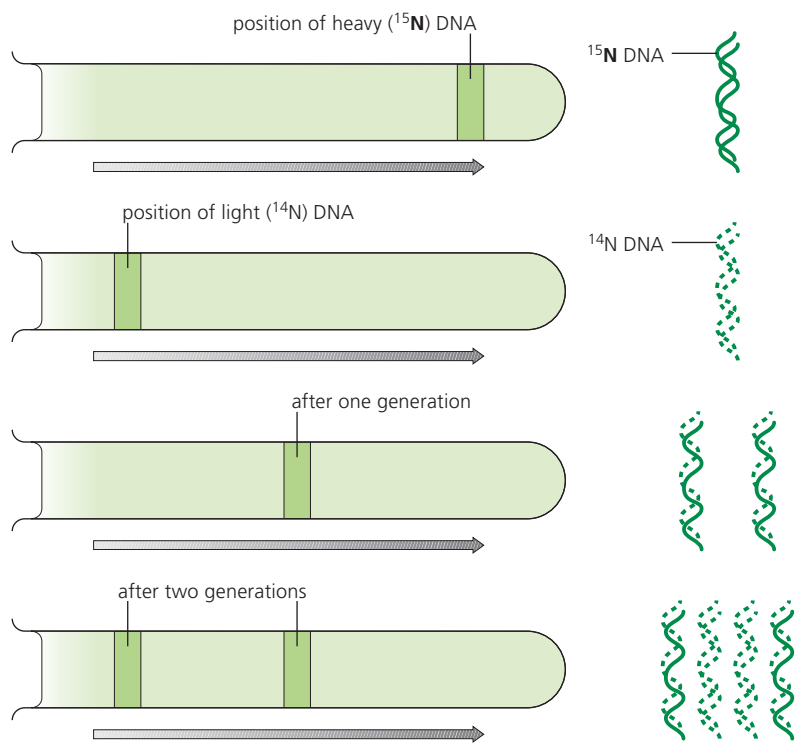
Figure 8.6 DNA replication is semi-conservative

1 Meselson and Stahl 'labelled' nucleic acid (i.e. DNA) of the bacterium *Escherichia coli* with 'heavy' nitrogen (^{15}**N**), by culturing in a medium where the only nitrogen available was as $^{15}NH_4^+$ ions, for several generations of bacteria.

2 When DNA from labelled cells was extracted and centrifuged in a density gradient (of different salt solutions) all the DNA was found to be 'heavy'.

3 In contrast, the DNA extracted from cells of the original culture (before treatment with ^{15}**N**) was 'light'.

4 Then a labelled culture of *E.coli* was switched back to a medium providing unlabelled nitrogen only, i.e. $^{14}NH_4^+$. Division in the cells was synchronised, and:
- after **one generation** all the DNA was of intermediate density (each of the daughter cells contained (i.e. *conserved*) one of the parental DNA strands containing ^{15}**N** alongside a newly synthesised strand containing DNA made from ^{14}N)
- after **two generations** 50% of the DNA was intermediate and 50% was 'light'. This too agreed with semi-conservative DNA replication, given that in only half the cells was labelled DNA present (one strand per cell).



Events at the replication fork

In eukaryotic chromosomes, replication is initiated at many points along the DNA double helix. At each advancing replication fork, DNA replication always occurs in the 5' → 3' direction. This is because the DNA polymerase that catalyses the condensation reaction (known as **DNA polymerase III**) can only work in this direction. So, it is the 5' end of a free DNA nucleotide which is added to 3' end of a chain of nucleotides.

The consequence of 5' → 3' replication is that, since both strands of DNA are replicated simultaneously, the details of the mechanism must vary in the strand that runs 3' → 5' (Figure 8.7). Here, after helicase has catalysed the uncoiling of the DNA double helix, causing exposure of the base sequence of the two template strands, the enzyme **RNA primase** adds a short RNA primer to the exposed 3' → 5' strand. To this, **DNA polymerase III** then adds succeeding free nucleotides by complementary base pairing, working in the 5' → 3' direction, away from the direction of the advancing replication fork.

Relatively short lengths of replicated DNA are formed by this process, known as **Okazaki fragments** after the biochemist who discovered this phenomenon. Later, the RNA primer between fragments is removed and replaced with DNA by **DNA polymerase I**, and then fragments are annealed into one continuous strand by ligase.

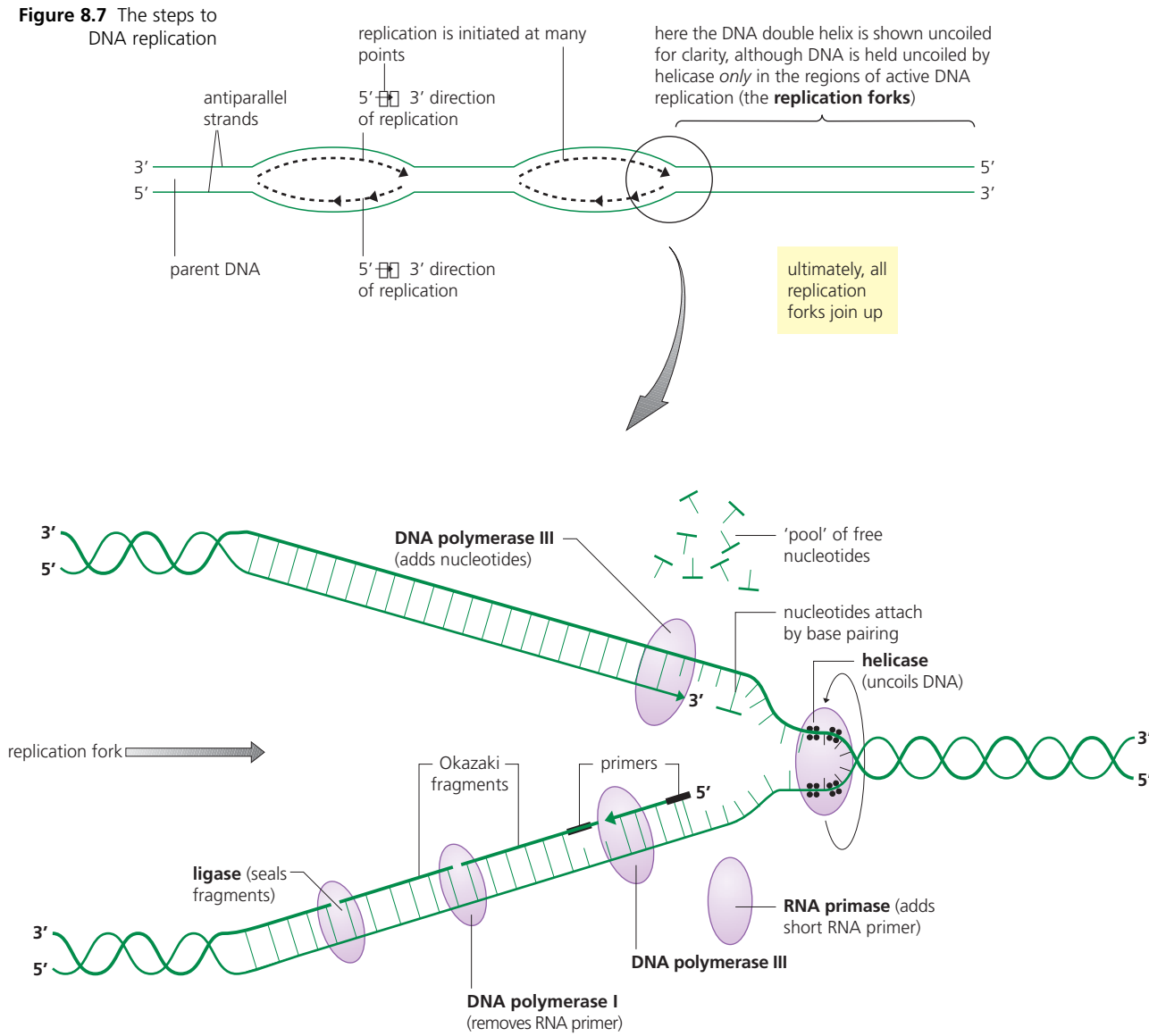


Figure 8.7 The steps to DNA replication

Meanwhile, on the exposed 5' → 3' strand, **DNA polymerase III** adds nucleotides by complementary base pairing to the free 3' end of the new strand, in the same direction as the replication fork. This process proceeds continuously, immediately behind the advancing helicase, as fresh template is exposed.

A note on 'nucleosides' and 'nucleotides'

The combination of a pentose sugar with a base forms a compound known as a **nucleoside**. Then, the combination of nucleoside with phosphate group forms a **nucleotide**. Table 8.1 lists the five nucleotides which are components of nucleic acids.

Base	+ pentose sugar	= Nucleoside	+ H_3PO_4	= Nucleotide
adenine		adenosine		adenosine phosphate
thymine		thymidine		thymidine phosphate
uracil		uridine		uridine phosphate
cytosine		cytidine		cytidine phosphate
guanine		guanosine		guanosine phosphate

Table 8.1 The common nucleosides and nucleotides found in cells

Nucleotides combined by the action of DNA polymerase are not in their monophosphate form, but are as **nucleoside triphosphates**: adenosine triphosphate (ATP), thymidine triphosphate, cytidine triphosphate, and guanosine triphosphate, as base pairing dictates. In each condensation reaction, two of the three phosphate groups are excluded.

So, it is chemically correct to refer to the nucleotides involved in DNA replication as nucleoside triphosphates.

DNA in protein synthesis – the genetic code

7.3.1–7.3.4, 7.4.1–7.4.7

The roles of nucleic acid in protein synthesis are introduced in Chapter 2, page 71. The key features are as follows.

- The DNA molecule in a single chromosome codes for a very large number of proteins. Within this extremely long molecule, the relatively short length of DNA that codes for a single protein is called a **gene**.
- Proteins are very variable in size and, therefore, so are genes. A few genes are as short as 75–100 nucleotides. Most are at least 1000 nucleotides long, and some are longer.
- Only 20 amino acids are used in protein synthesis; all cell proteins are built from them. The unique properties of each protein lie in:
  – which amino acids are involved in its construction;
  – the sequence in which these amino acids are joined.
- It is the sequence of the four bases (C, G, A and T) in a gene that dictates the order in which specific amino acids are assembled and combined.
- The code's sequence lies in one of the two strands of DNA – the coding strand. The other strand is complementary to it. As we shall shortly discover, the correct name of the coding strand is the **antisense strand** (see Table 8.2).
- The code is a three-letter or triplet code – a sequence of three of the four bases stands for one of the 20 amino acids, and is called a 'codon'. The genetic code has more codons than there are amino acids, and most amino acids have two or three similar codons that code for them (a degenerate code). Some codons represent 'punctuations' – that is, they are 'start' and 'stop' triplets.

There are three major steps in the process by which the information of the gene is used to determine how the protein molecule is constructed:

- Stage 1, **transcription** occurs in the nucleus.
  A copy of the code is made by building a molecule of **messenger RNA** (**mRNA**). This is catalysed by RNA polymerase, and involves complementary base pairing, with the antisense strand used as the template. The mRNA strand then leaves the nucleus through pores in the nuclear membrane. It passes to ribosomes in the cytoplasm where the information can be 'read' and is used.
- Stage 2 occurs in the cytoplasm.
  Amino acids are activated for protein synthesis by combining with short lengths of a different sort of RNA, called **transfer RNA** (**tRNA**). There is a different tRNA for each of the 20 amino acids involved in protein synthesis. One end of each tRNA molecule is a site where a particular amino acid can attach. At the other end, is a sequence of three bases called an **anticodon**. This anticodon is complementary to the codon of mRNA that codes for the specific amino acid.
- Stage 3, **translation** occurs in ribosomes in the cytoplasm.
  A protein is assembled, one amino acid residue at a time, as a ribosome moves along the messenger RNA 'reading' the codons. Complementary anticodons of the amino acid–tRNA slot into place and are temporarily held in position by hydrogen bonds. While held there, the amino acids of neighbour amino acid–tRNAs are joined by peptide linkages.

Regulation of gene expression

The cell nucleus contains many more genes than are required to be active at any one moment. For example, in a multicellular organism, every nucleus contains the coded information relating to the development and maintenance of all mature tissues and organs. Both during development and later, this genetic information is used selectively. Genes function (are **expressed**) only in cells they relate to, when they are needed. For example, a gene for the enzyme pepsin is present in all cells of our body but it is active only in cells of the gastric glands in the stomach. Similarly, the gene for insulin production is active only in cells of the islets of Langerhans in the pancreas.

How is gene expression regulated?

A mechanism for control of gene expression was first discovered in **prokaryotes**. However, it is now appreciated that this mechanism does not operate in eukaryotes, so the details do not concern us here.

In **eukaryotes**, genes are only transcribed if an RNA polymerase enzyme binds to a region of DNA situated close to the gene, known as a **promoter**. The special features of promoters in eukaryotic chromosomes are:

- some permit repeated, unrestricted binding of RNA polymerase, resulting in continuous expression of the gene;
- some require a **regulatory protein** to be present and bound, prior to binding of RNA polymerase and gene expression (regulatory proteins are the products of other genes, often on different chromosomes);
- some regulatory proteins must first be activated by reaction with a steroid hormone or some other metabolite, before binding of RNA polymerase and gene expression.

Other aspects of eukaryotic gene regulation concern:

- the change that may occur to the messenger RNA immediately after it has been formed by transcription and in which introns are removed (see 'Some eukaryotic genes are discontinuous', below);
- the changes that may occur to proteins formed by translation of the mRNA code, out in the cytoplasm (see 'Post-translational modification of protein', page 251).

Both of these processes are under the control of specific genes, although these are often found on other chromosomes; they too are part of the story of how gene action is brought about.

The sequence of events in transcription in eukaryote chromosomes

The steps to the expression of a gene begin when the genetic information in DNA is **transcribed** into a molecule of **mRNA** by complementary base pairing. mRNA is an intermediary molecule between the information in the gene and the formation of the protein it codes for, out in the cytoplasm. Remember, mRNA is single-stranded nucleic acid in which the base uracil pairs with adenine in place of thymine found in DNA. The events of transcription in eukaryote chromosomes are illustrated in Figure 8.8.

Look at the sequence of events described there, now.

The enzyme **RNA polymerase** binds to a **promoter region**, the 'start' signal for transcription, located immediately before the gene, and the new nucleic acid strand is formed in the 5' → 3' direction.

During transcription, only one strand of the DNA double helix serves as a **template** for synthesis of mRNA. This is called the **antisense strand**. The partner strand is called the sense strand. Sense and antisense strands are compared in Table 8.2 below.

Sense strand	Antisense strand
carries the promoter sequence of bases to which RNA polymerase binds and begins transcription	is the template sequence for transcription by complementary base pairing by RNA polymerase
has the same base sequence as the mRNA (the mRNA has U in place of T)	has the same base sequence as the tRNA (the tRNA has base U in place of T)
carries the terminator sequence of bases at the termination of each gene that cause RNA polymerase to stop transcription	is 'read' in the 3' → 5' direction, and mRNA synthesis occurs in the 5' → 3' direction (by addition of 5' end of a nucleotide to the 3' end of the RNA molecule already synthesised)

Table 8.2 Sense and antisense strands of DNA

RNA polymerase draws on the pool of **free nucleotides**. As with DNA replication, these nucleotides are present in the form of nucleoside triphosphates (although in RNA synthesis uridine triphosphate replaces thymidine triphosphate). As the RNA is formed, it falls away from the antisense strand that has acted as template, and hydrogen bonds re-form between the two DNA strands.

6 Explain what is meant by antiparallel strands.

The process continues until a base sequence known as the **terminator** is reached. At this signal, both RNA polymerase and the completed, new strand of mRNA are freed from the site of the gene.

Some eukaryotic genes are discontinuous

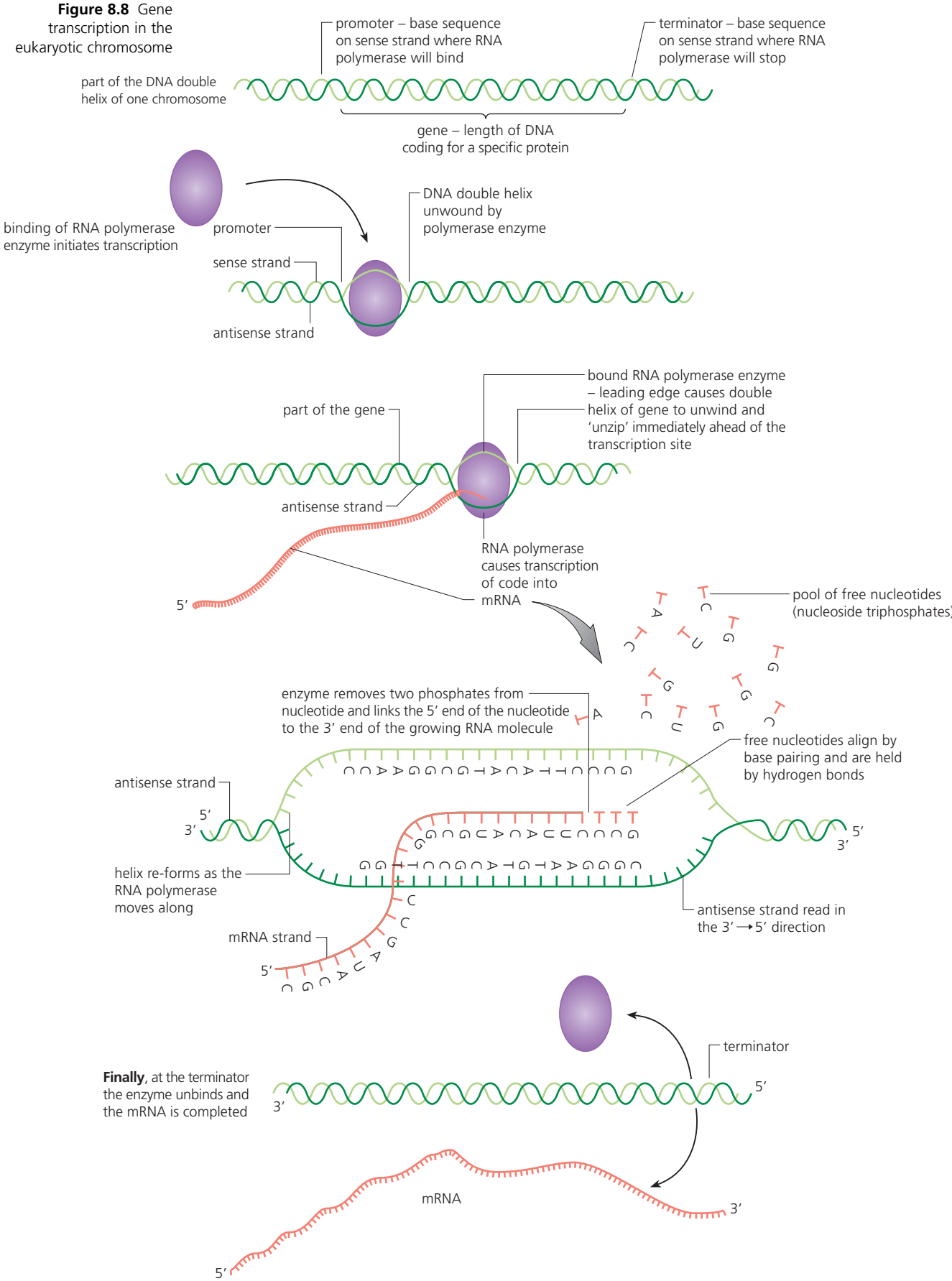
Where the exact sequence of bases in a particular gene has been determined, it is sometimes discovered that **non-coding sequences** are present. In other words, some eukaryotic genes have units of non-gene DNA *within their boundaries*.

The sections that carry meaningful information are called **exons**, whereas the intervening lengths of DNA (interruptions, in effect) are called **introns**. Genes split in this way are common in higher plants and animals, although others contain no introns at all.

When a split gene is transcribed into mRNA, the newly formed mRNA contains the sequence of introns and exons exactly as they occur in the DNA. The persistence of nonsense sequences would undoubtedly present problems in the subsequent protein-synthesis steps.

In fact, an enzyme-catalysed reaction known as **post-transcriptional modification**, removes the introns at this stage. The production of this enzyme is also under the control of a gene. The short lengths of nonsense transcribed into the RNA sequence of bases are disposed of. The resulting (shortened) length of mRNA is described as **mature**. It passes out into the cytoplasm, to ribosomes, where it is involved in protein synthesis (Figure 8.9).

Figure 8.8 Gene transcription in the eukaryotic chromosome



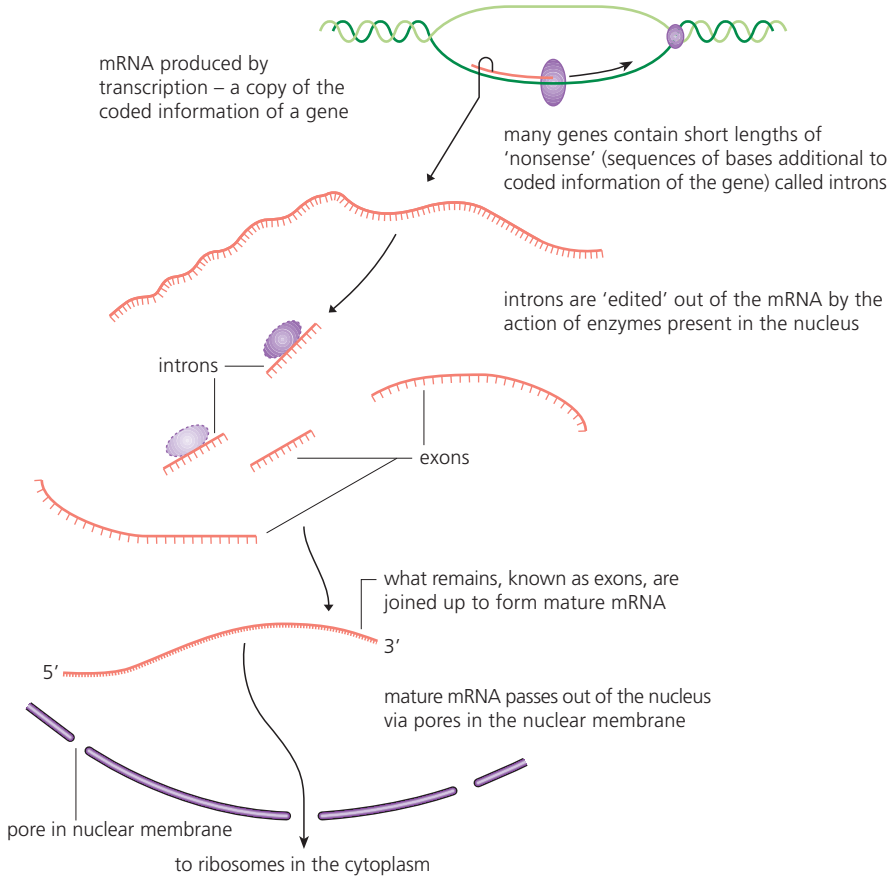


Figure 8.9 Post-transcriptional modification of mRNA

Translation – protein synthesis in ribosomes

The first step to protein synthesis in the **cytoplasm** involves the **activation of amino acids** by combining them with short lengths of a different sort of RNA, called **transfer RNA** (**tRNA**). This activation process occurs in the cytoplasm and requires ATP. It is this tRNA, once attached to its amino acid, which facilitates the translation of the three-base sequences of each codon of mRNA into a sequence of amino acids in a protein.

All tRNAs have a cloverleaf shape, but there is a different tRNA for each of the 20 amino acids involved in protein synthesis (Figure 8.10). At one end of the tRNA molecule is a site where **one particular amino acid** can be joined (and no other type). At the other end, there is a sequence of three bases called an **anticodon**. This anticodon is complementary to the codon of mRNA that codes for that specific amino acid. Of course, the amino acid is attached to its tRNA by enzyme action and these enzymes are specific to the particular amino acids (and types of tRNA) to be used in protein synthesis. Thus, the structure of the tRNA allows recognition by a specific tRNA-activating enzyme which attaches a specific amino acid to the tRNA. The specificity of the enzymes is a way of ensuring the correct amino acids are used in the right sequence. Some amino acids have more than one type of tRNA specific to them.

The next step, the translation process, occurs in the **ribosomes**. These tiny organelles consist of a large and a small subunit, both composed of RNA (known as rRNA) and protein, as illustrated in Figure 8.11. During translation, the mRNAs bind with the small subunit. Here occur the three sites where the tRNAs interact:

- at the **first site**, codons of the incoming tRNA bind to specific tRNA–amino acids through their anticodons (**complementary base pairing**);
- the **second site** is where the amino acid attached to its tRNA is condensed with the growing polypeptide chain by **formation of a peptide linkage**;
- the **third site** is where the **tRNA leaves** the ribosome following transfer of its amino acid to the growing protein chain.

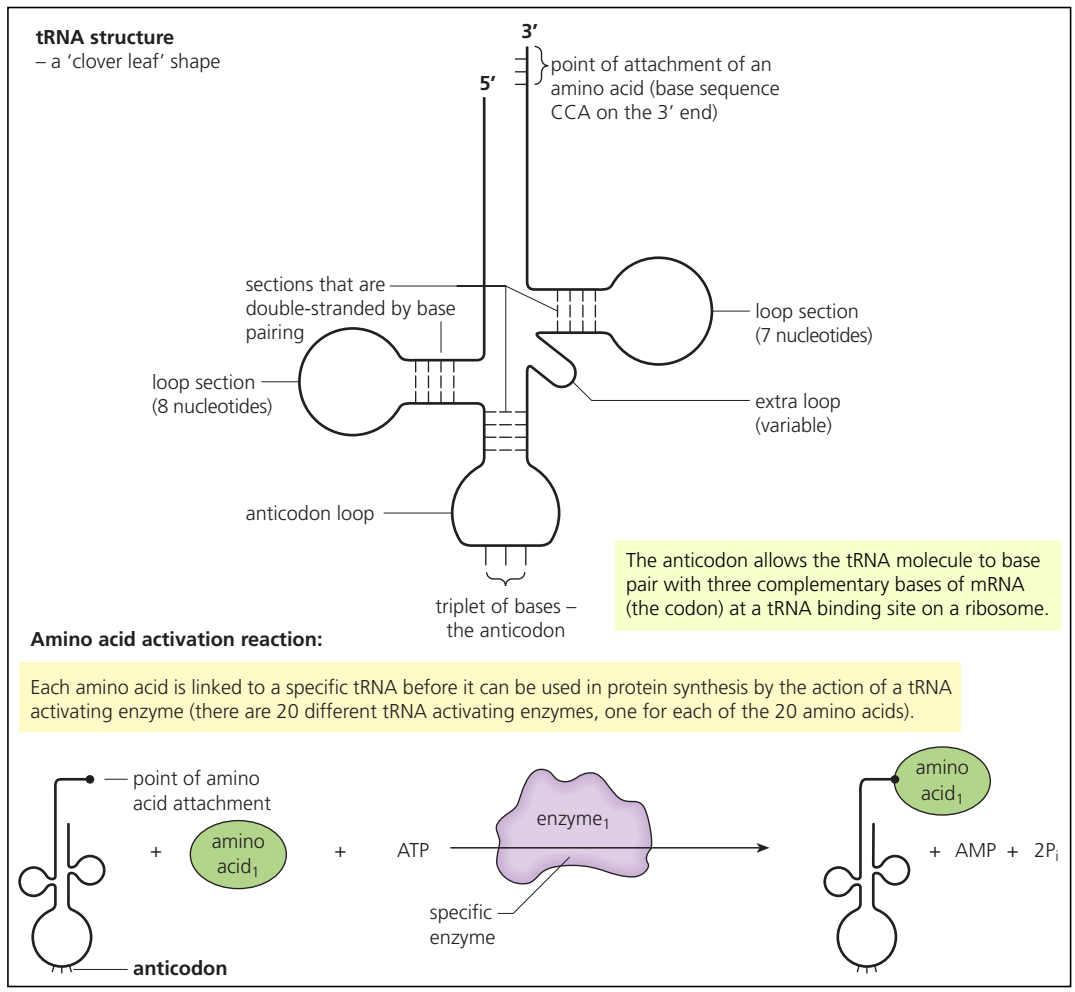


Figure 8.10 Transfer RNA (tRNA) and amino acid activation

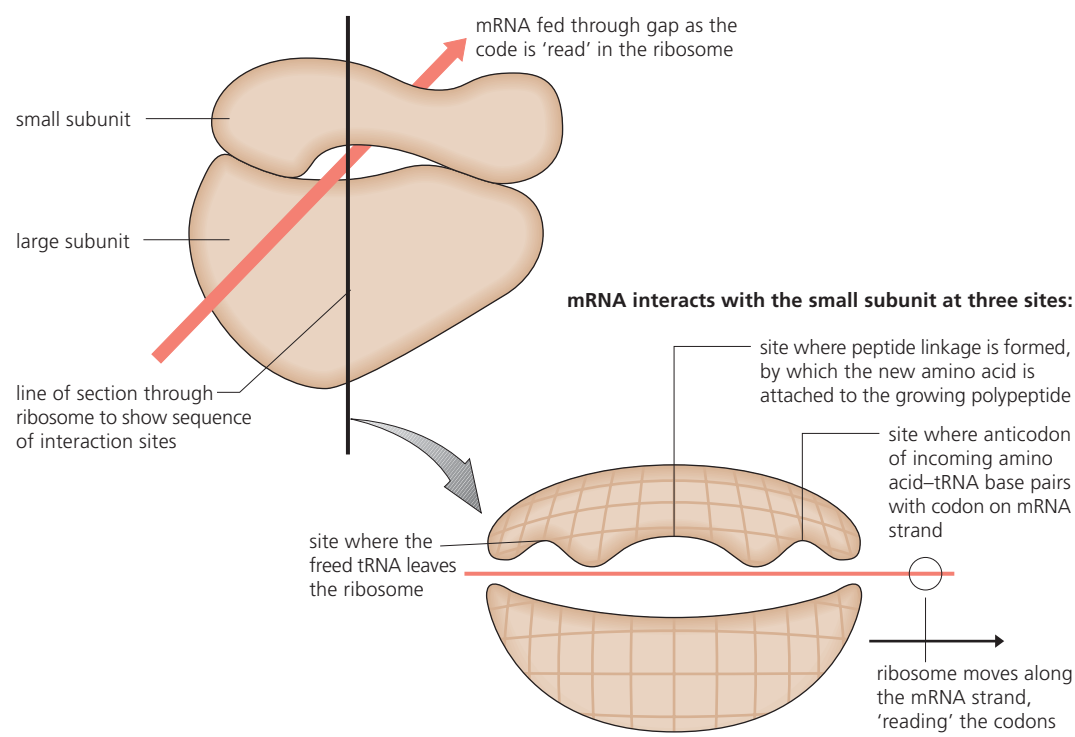


Figure 8.11 Ribosome structure and function

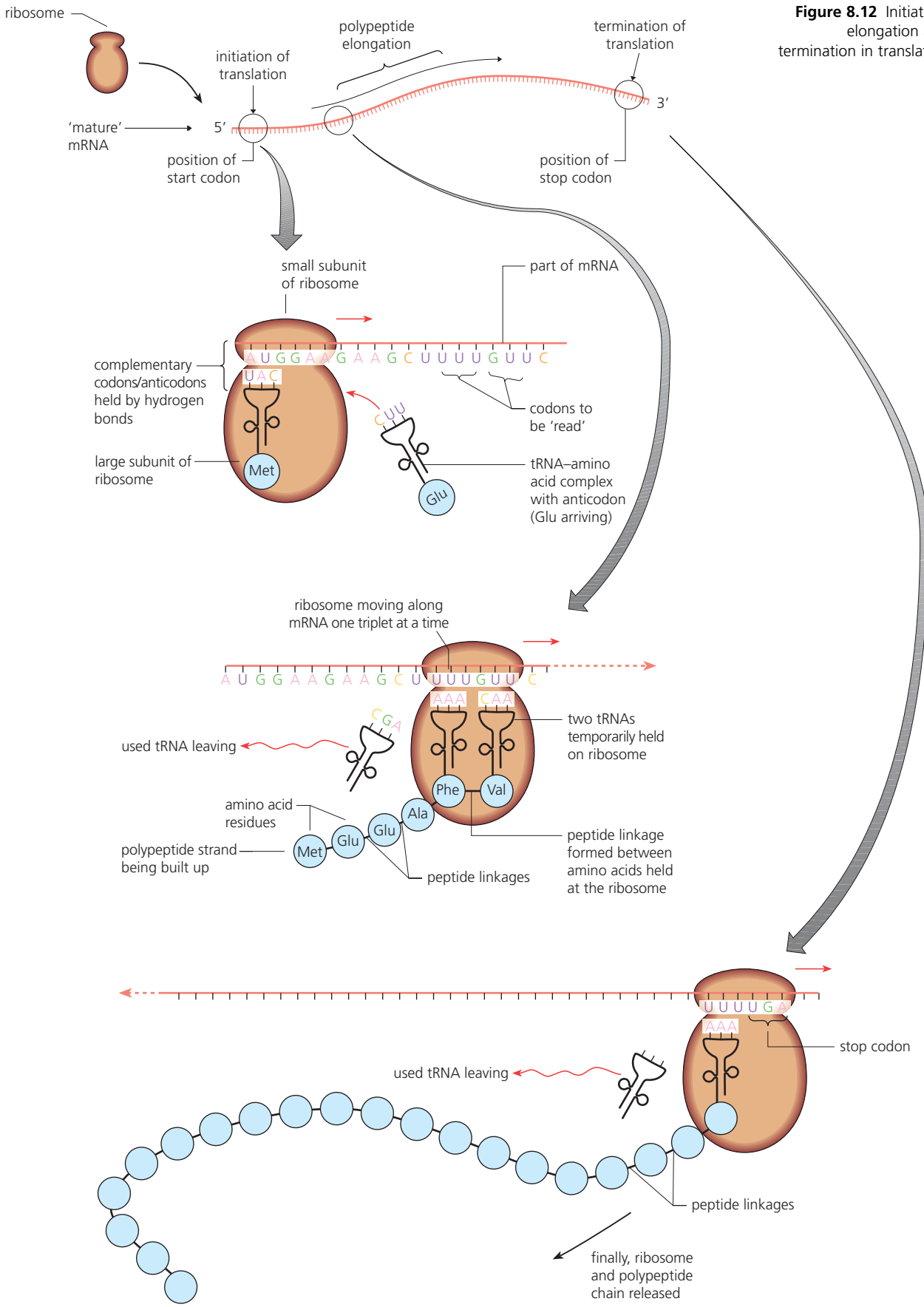


Figure 8.12 Initiation, elongation and termination in translation

In the ribosome a protein chain is assembled, one amino acid residue at a time. This occurs as the ribosome moves along the messenger RNA 'reading' the codons from the **'start' codon**.

Translation occurs in the 5' → 3' direction, the organelle moving along the mRNA from the 'start' codon towards the 3' end and the **'stop' codon**. As the ribosome moves along, complementary anticodons of the **specific amino acid–tRNA complex** slot into place. They are temporarily held in position by hydrogen bonds.

7 **Draw** and **label** the structure of a peptide linkage between two amino acids.

While held there, the amino acids of neighbour amino acid–tRNAs are joined by **peptide linkages** (Figure 2.15, page 52). This frees the first tRNA which moves back into the cytoplasm for re-use. Once this is done, the ribosome moves on to the next mRNA codon. The process continues until a 'stop' codon occurs (Figure 8.12).

Many ribosomes occur freely in the cytoplasm, and these are the site of synthesis of proteins that remain in the cell and fulfil particular roles there. It is common for several ribosomes to move along the mRNA at one time; the structure (mRNA, ribosomes and their growing protein chains) is called a **polysome** (Figure 8.13).

8 **State** the sequence of changes catalysed by RNA polymerase.

Other ribosomes are bound to the membranes of the rough endoplasmic reticulum (**rER**, page 15), and these are the site of synthesis of proteins that are subsequently secreted from cells or packaged in lysosomes there.

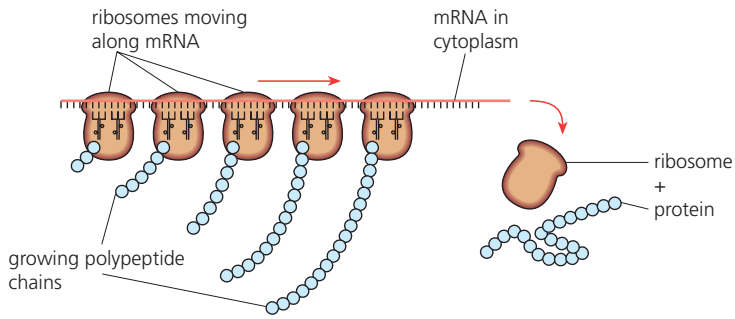


Figure 8.13 A polysome

Deciphering of the genetic code

The sequence of bases (A, C, G, and T) in DNA carries the genetic code, specifying the 20 amino acids from which cell proteins are built. It was clear to Watson and Crick from the outset that combinations of three bases would be most likely to code for the amino acids. This was because:

- a doublet code (AA, AG, AC, AT, GA, etc.) could specify only 16 amino acids ($4^2 = 16$);
- a four-base code could specify 256 amino acids ($4^4 = 256$), far more possibilities than are needed.

In a triplet code, there are still more possibilities than are required, and it was later found that the spare capacity is used rather than left unused (this is why the genetic code is described as a **degenerate** code).

The code was first 'cracked' by the preparation of lengths of nucleic acid (synthetic mRNA, in effect) in which one 'word' was repeated (UUU–UUU–UUU, CCC–CCC–CCC etc).

Subsequently, synthetic mRNAs with known, variable sequences of nucleotides were produced. These mRNAs were added in turn to cell-free protein-synthesising systems containing all 20 amino acids. Each polypeptide formed was analysed. In this way, the complete code was deciphered.

Two amino acids were found to be coded for by a single triplet, and a few other amino acids were each coded for by two codons. However, most amino acids were coded for by several triplets, up to a maximum of six. The genetic code (for mRNA) is given in Figure 8.14. Note that certain 'words' code for a 'full stop', rather than an amino acid. The codon for methionine is the 'start' codon.

Figure 8.14 The genetic code – a universal code

The 20 amino acids used in protein synthesis

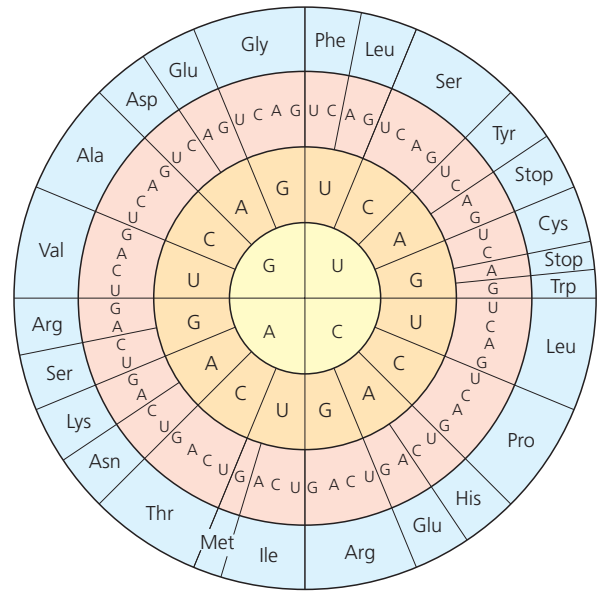
Amino acids	Abbreviations
alanine	Ala
arginine	Arg
asparagine	Asn
aspartic acid	Asp
cysteine	Cys
glutamine	Gln
glutamic acid	Glu
glycine	Gly
histidine	His
isoleucine	Ile
leucine	Leu
lysine	Lys
methionine	Met
phenylalanine	Phe
proline	Pro
serine	Ser
threonine	Thr
tryptophan	Trp
tyrosine	Tyr
valine	Val

The genetic code in circular form

The codons are those of messenger RNA (where uracil, U, replaces thymine, T).

Read the code from the centre of the circle outwards along a radius. For example, serine is coded by UCU, UCC, UCA or UCG, or by AGU or AGC.

In addition, some codons stand for 'stop', signalling the end of a peptide or protein chain.



9 The sequence of bases in a sample of mRNA was found to be:

GGU–AAU–CCU–UUU–GUU–ACU–CAU–UGU.

Predict:

a the sequence of amino acids coded for by this sequence of bases

b the sequence of bases in the antisense strand of DNA from which this mRNA was transcribed.

Post-translational modification of protein

When a protein exits from translation at a ribosome, it may take up its active, three-dimensional shape and be functional immediately. For example, it may be active as an enzyme in the cytoplasm in some essential, continuous biochemical pathway. On the other hand, many proteins are in the form of inactive precursors requiring processing steps. These steps occur after translation and so are known as **post-translational modifications** (Figure 8.15). In fact, it is often important for proteins to become active only at particular sites. For example, the protein-digesting enzyme trypsin is produced in the pancreas but in an inactive form (trypsinogen, page 649) which will not digest the proteins of the pancreas cells in which it is formed.

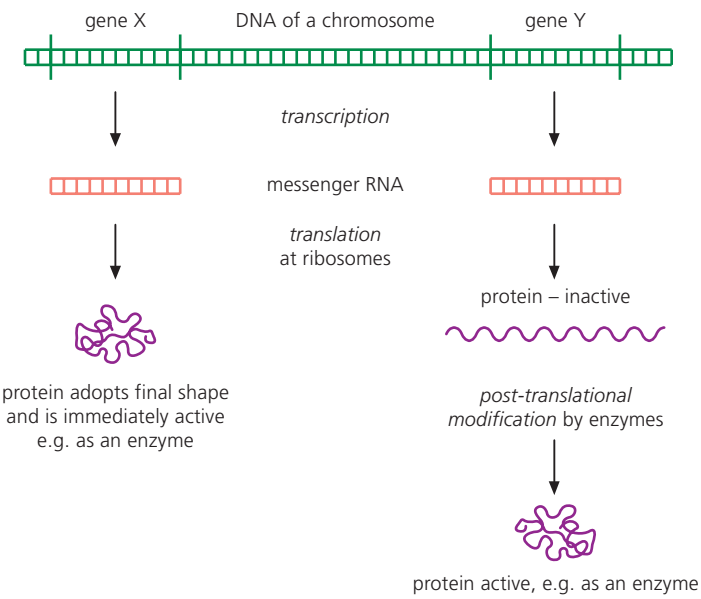


Figure 8.15 Protein synthesis and post-translational modification

10 Distinguish between the different forms of protein in the nucleus.

The central dogma

We have seen that the information of genes is copied by transcription into mRNA, and that these single strands of nucleic acid pass to the ribosomes in the cytoplasm during protein synthesis. This is a **one-way flow** of coded information from nucleus to cytoplasm and has been named the **central dogma of molecular biology**.

The term 'one-way flow' draws attention to the direction in which information travels – from the nucleus to the cytoplasm, *and never the other way round* (Figure 8.16).

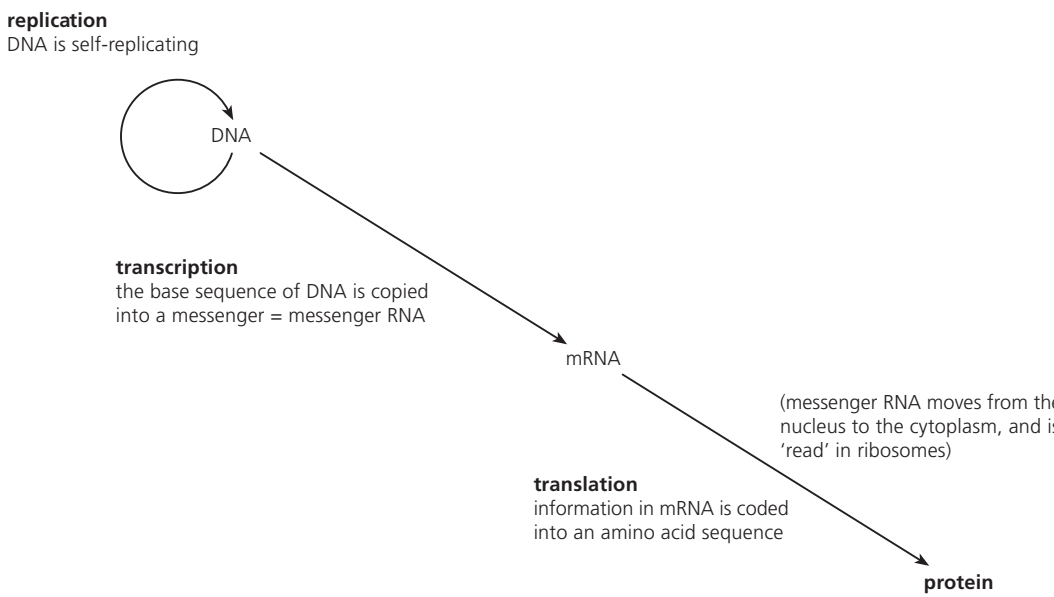


Figure 8.16 Steps to protein synthesis and the central dogma

Incidentally, if this principle did not apply, events in cells due to environmental changes might lead to changes in the characteristics that particular genes code for.

Since this principle of molecular biology was asserted by Crick and Watson, it has been discovered that occasional reversing of the flow of information does occur.

You will remember that human immunodeficiency virus (HIV) is an RNA virus (page 198). Once HIV nucleic acid is inside a host cell, an enzyme called **reverse transcriptase** catalyses the copying of the RNA strand to form a DNA double helix. Reverse transcriptase is an enzyme the virus introduces into the host cell. The DNA this enzyme synthesises then enters the host nucleus and becomes attached to a host chromosome. There it functions as other genes do, eventually being transcribed, leading to HIV replication and the death of the host cell.

TOK Link

What limitations are placed on advancement of knowledge if we hold dogmas dogmatically?

The genetic engineer's tool-kit

The enzyme reverse transcriptase may be extracted from retroviruses (of which there are many besides HIV) and it is used in genetic modification. In fact, reverse transcriptase is an important enzyme in the genetic engineer's tool-kit (Table 8.3). The steps to the reverse transcription reaction are illustrated in Figure 7.21 (page 200). Genetic engineers make use of purified enzymes extracted mainly from microorganisms or viruses to manipulate nucleic acids in very precise ways.

Enzyme	Natural source	Application in genetic modification
restriction enzymes (restriction endonuclease)	cytoplasm of bacteria (combats viral infection by breaking up viral DNA)	breaks DNA molecules into shorter lengths, at specific nucleotide sequences
DNA ligase	with nucleic acid in the nucleus of all organisms	joins together DNA molecules during replication of DNA
polymerase	with nucleic acid in the nucleus of all organisms	synthesises nucleic acid strands, guided by template strand of nucleic acid
reverse transcriptase	in retroviruses only	synthesis of a DNA strand complementary to an existing RNA strand

Table 8.3 The genetic engineer's tool-kit of enzymes

The special value of reverse transcriptase to the genetic engineer lies in the fact that mRNA is often present in large quantities in cells, compared to the original DNA (the gene) from which it has been synthesised. Moreover, mRNA is easily extracted from cells.

For example, the gene for insulin is present in human pancreatic tissue, but it is difficult to extract and contains introns as well as exons (thus providing additional problems for the genetic engineer). But the mRNA complementary to the gene for insulin is relatively easy to extract and can be used to synthesise a version of the gene that contains only exons.

There are many cases of the successful application of reverse transcriptase in the production of modified microorganisms leading to the commercial production of important biochemicals. Continuing from the insulin example above, reverse transcription is exploited in the production of human insulin in strains of the genetically modified bacterium *E. coli*, for the treatment of type I diabetes. The steps of this process are illustrated in Figure 18.21 (page 573).

11 Explain the value of the discovery of reverse transcriptase to molecular biologists.

Proteins

7.5.1–7.5.4

We know that proteins are built up by condensation reactions between **amino acids**. We have seen that amino acids are molecules that have a basic amino group ($-NH_2$) and an acidic carboxyl group (–COOH), attached to the same carbon atom (page 51). These are the reactive or **functional groups** of the molecule; they are the groups that participate in the condensation reactions that form peptide linkages.

The remainder of the amino acid molecule, the side chain or **–R part** as it is known, may be very variable. Something of the variety of amino acids is shown in Figure 8.17. We can see that, while amino acids have the same basic structure, they are all rather different in character, because of the different R groups they carry. Categories into which many of the amino acids found in cell proteins fit are:

- **acidic amino acids**, having additional carboxyl groups (e.g. aspartic acid);
- **basic amino acids**, having additional amino groups (e.g. lysine);
- **amino acids with hydrophilic properties** (water soluble) have polar R groups (e.g. serine);
- **amino acids with hydrophobic properties** (insoluble) have non-polar R groups (e.g. alanine).

The bringing together of differing amino acids in contrasting combinations produces proteins with very different properties. The amino acid residues of proteins affect their shape, where they occur in cells, and the functions they carry out. We will see examples of this when the functions of cell membranes and enzymes are discussed.

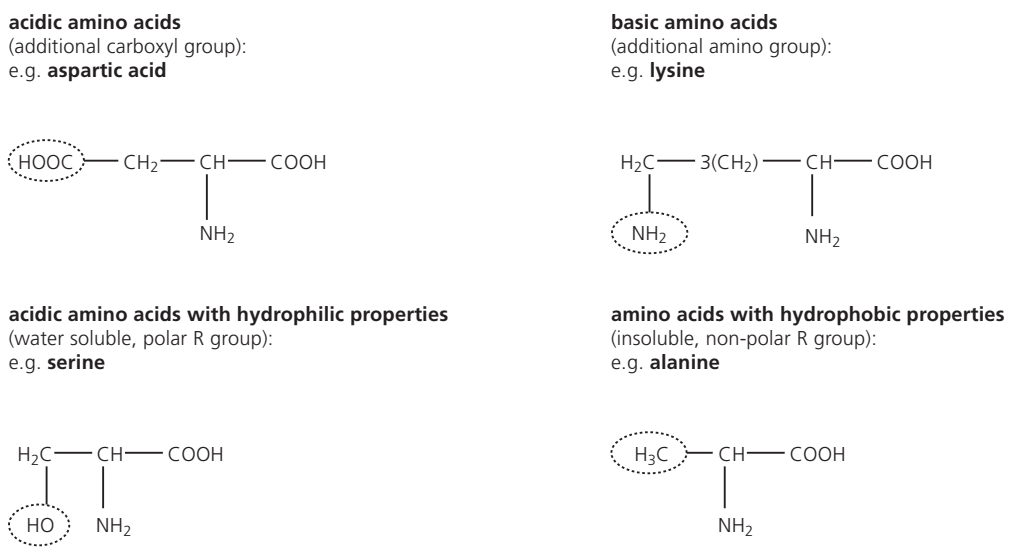


Figure 8.17 The range of amino acids used in protein synthesis

The structure of proteins

There are four levels of protein structure, each of significance in biology.

The **primary structure** of a protein is the sequence of amino acid residues attached by peptide linkages. Proteins differ in the variety, number and order of their constituent amino acids. We have seen how, in the living cell, the sequence of amino acids in the polypeptide chain is controlled by the coded instructions stored in the DNA of the chromosomes in the nucleus, mediated via mRNA. Changing just one amino acid in the sequence of a protein alters its properties, often quite drastically. This sort of mistake or mutation does happen (page 100).

The **secondary structure** of a protein develops when parts of the polypeptide chain take up a particular shape, immediately after formation at the ribosome. Parts of the chain become folded or twisted, or both, in various ways.

The most common shapes are formed either by coiling to produce an **α helix** or folding into **β sheets** (Figure 8.18). These shapes are permanent, held in place by hydrogen bonds (page 38).

Figure 8.18 The secondary structure of protein

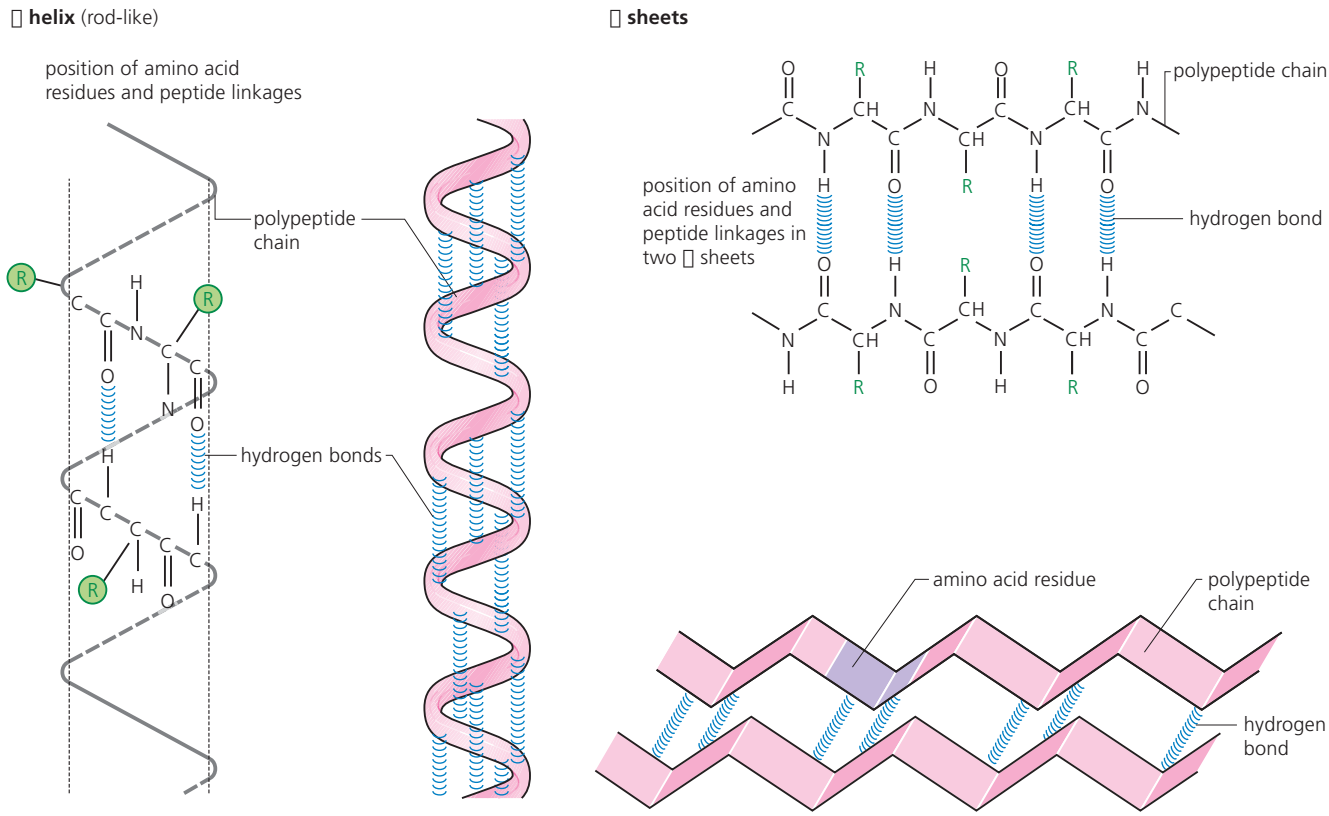
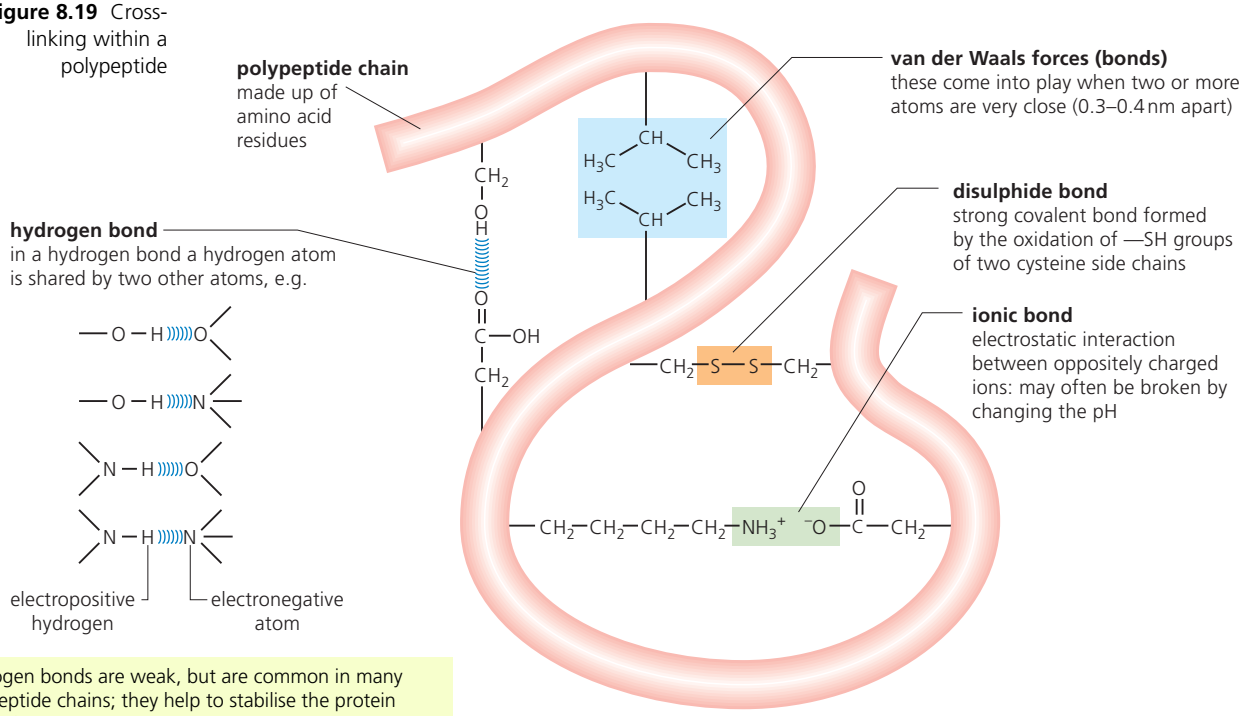


Figure 8.19 Cross-linking within a polypeptide



Hydrogen bonds are weak, but are common in many polypeptide chains; they help to stabilise the protein molecule.

Figure 8.20 Lysozyme – primary, secondary and tertiary structure

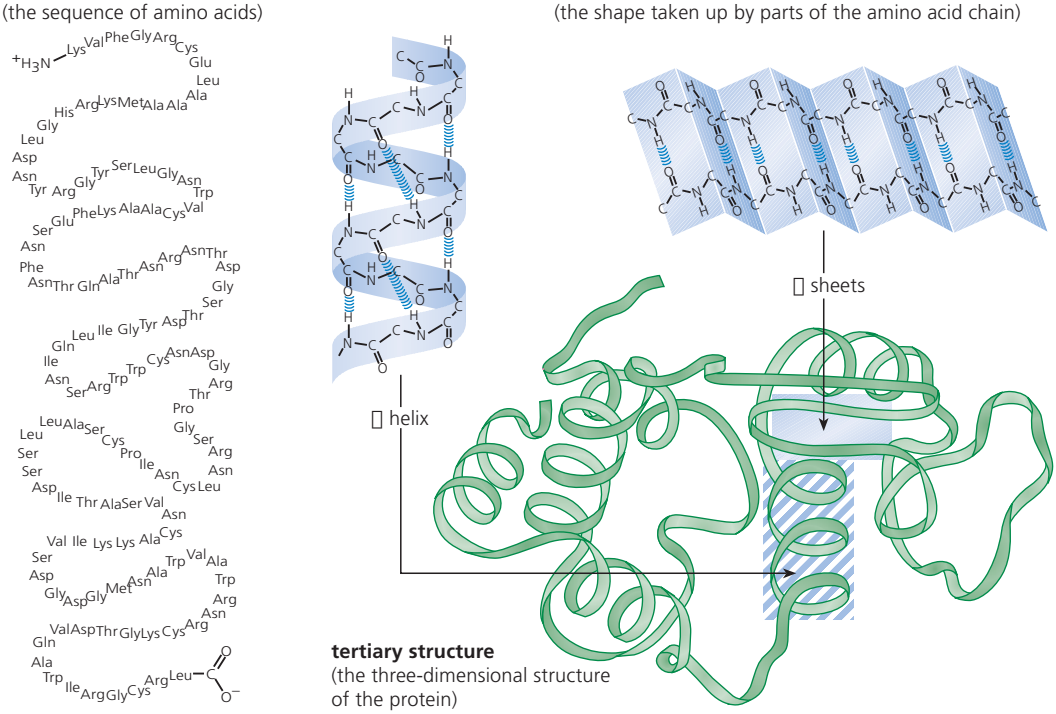
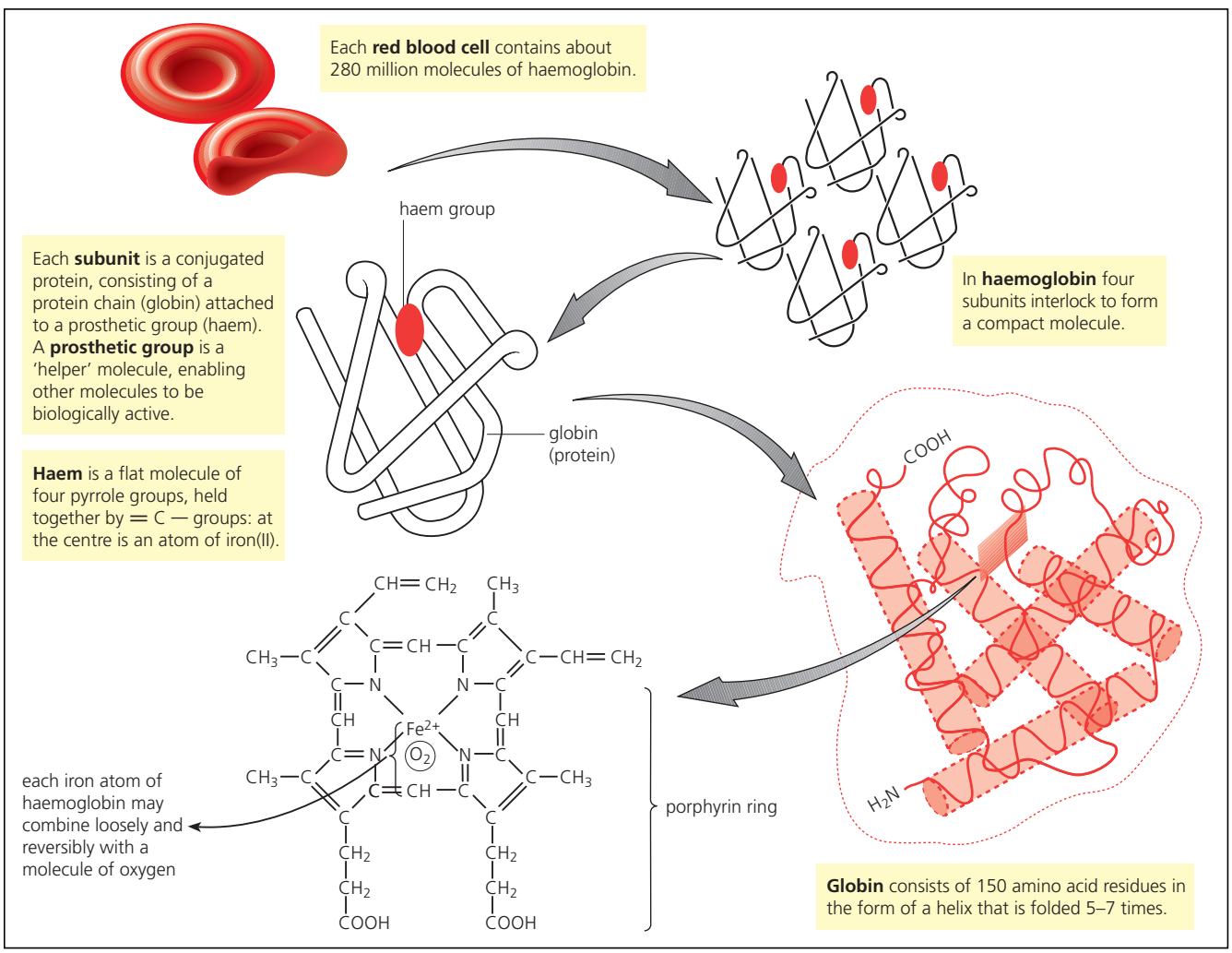


Figure 8.21 Haemoglobin – a quaternary protein of red cells



12 **Describe** three types of bond that contribute to a protein's secondary structure.

The **tertiary structure** of a protein is the precise, compact structure, unique to that protein, which arises when the molecule is further folded and held in a particular complex shape. This shape is made permanent by four different types of bonding, established between adjacent parts of the chain (Figure 8.19). The primary, secondary and tertiary structure of the protein lysozyme is shown in Figure 8.20.

The **quaternary structure** of protein arises when two or more proteins become held together, forming a complex, biologically active molecule. An example is haemoglobin, consisting of four polypeptide chains held around a non-protein haem group, in which an atom of iron occurs (Figure 8.21).

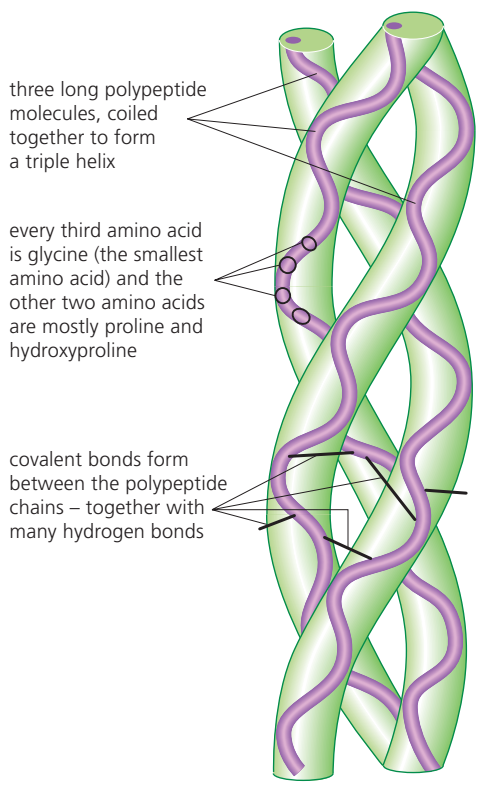
Fibrous and globular proteins – contrasting tertiary structures

Some proteins take up a tertiary structure that is a long, much-coiled chain, and are called fibrous proteins. They have long, narrow shapes. Examples of fibrous proteins are **collagen**, a component of bone and tendons (Figure 8.22), and **keratin**, found in hair, horn and nails. Fibrous proteins are often insoluble.

Figure 8.22 Collagen – example of a fibrous protein

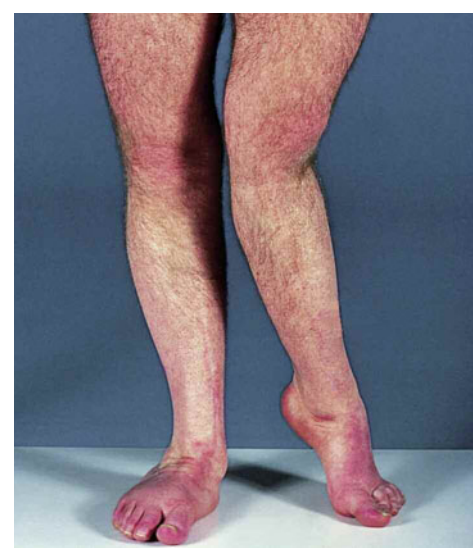
Collagen:

- is the most abundant structural protein in animals – skin is largely composed of collagen, forming a mat in the deeper layers; it is also a major component of bones, teeth and tendons
- chemically consists of three polypeptide chains, each about 1000 amino acid residues long, wound together as a triple helix – a unique arrangement in proteins – with an appearance of a twisted rope
- has a primary structure of glycine (the smallest amino acid) at every third point in the chain in which the repeating sequence is glycine–proline–hydroxyproline. The R groups of these amino acids are ring structures that prevent collagen from folding or forming pleated sheeets; collagen is forced to take up the triple helix shape
- is an insoluble protein that is assembled by cells as procollagen, a soluble molecule built to make collagen outside the cell, with cross-links formed between the polypeptide chains as they are assembled
- when assembled, binds with other components such as calcium phosphate in bone.



Patients with brittle bone disease (see below) may suffer multiple fractures of their bones.

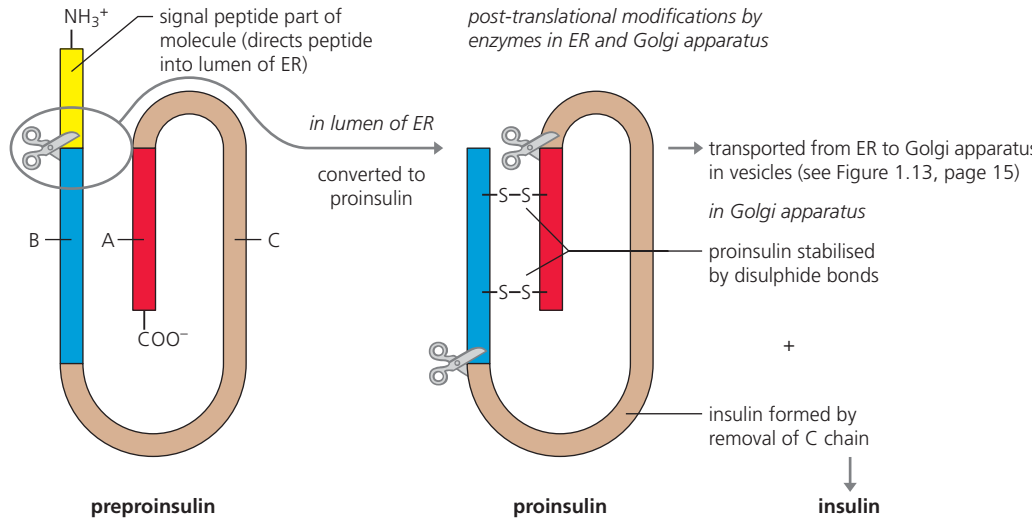
The disease arises because the patient's collagen does not form links to the mineral component of bone.

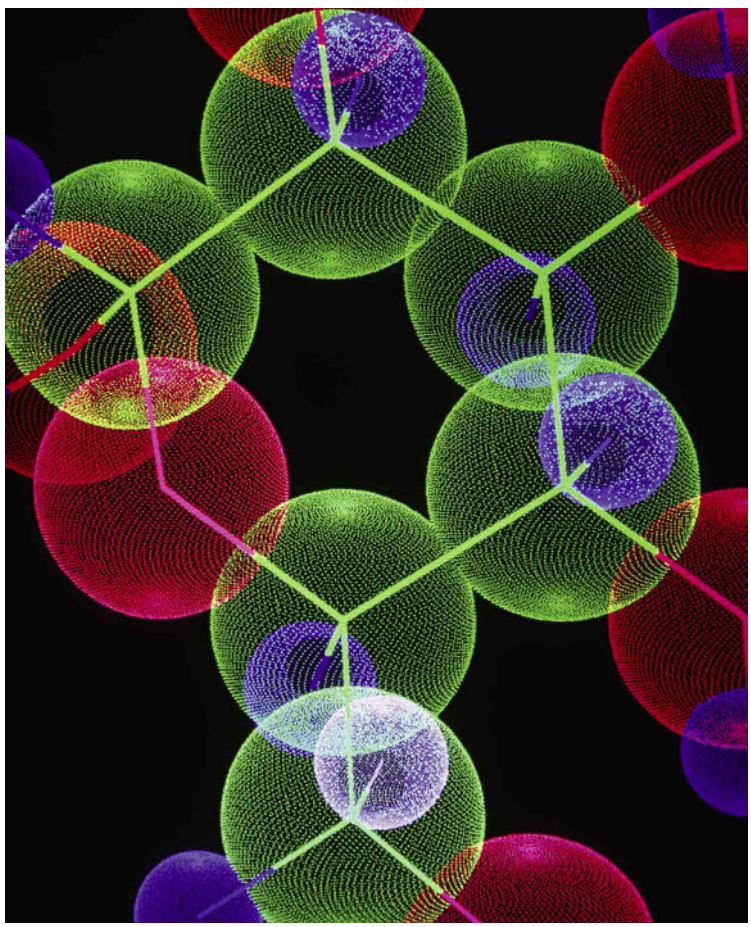


Other proteins take up a tertiary structure that is more spherical, and are called **globular proteins** (Figure 8.23). They are mostly highly soluble in water. Enzymes, such as **catalase**, are typical globular proteins. Some of our hormones are globular proteins, too, including **insulin** (page 222).

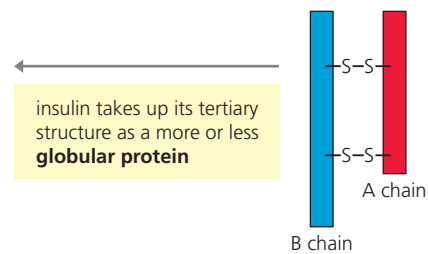
Figure 8.23 Insulin – example of a globular protein

Insulin is a hormone produced in the β cells of the islets of Langerhans in the pancreas by ribosomes of the rough endoplasmic reticulum (rER) as a polypeptide of 102 amino acid residues (preproinsulin).





computer-generated image of 3D shape of insulin



insulin takes up its tertiary structure as a more or less **globular protein**

Polar and non-polar amino acids in proteins

Amino acids with polar R groups have hydrophilic properties. When these amino acids are built into protein in prominent positions they may influence the properties and functioning of the proteins in cells. Similarly, amino acids with non-polar R groups have hydrophobic properties.

Examples of these outcomes are illustrated in Figures 8.24 (for cell membrane proteins) and 8.25 (for an enzyme that occurs in the cytoplasm).

Figure 8.24 Polar and non-polar amino acids in membrane proteins

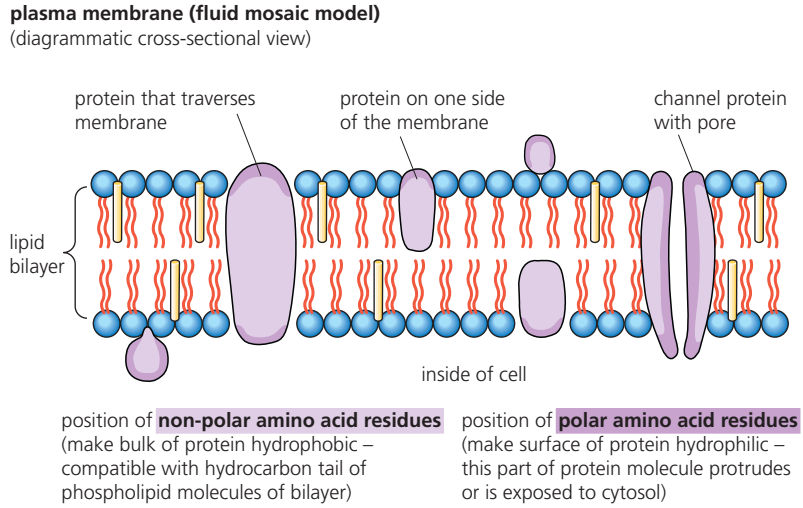
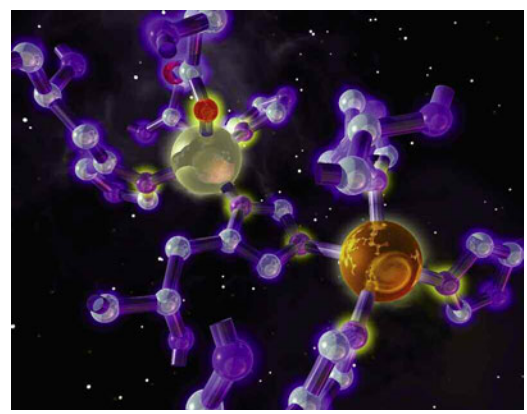


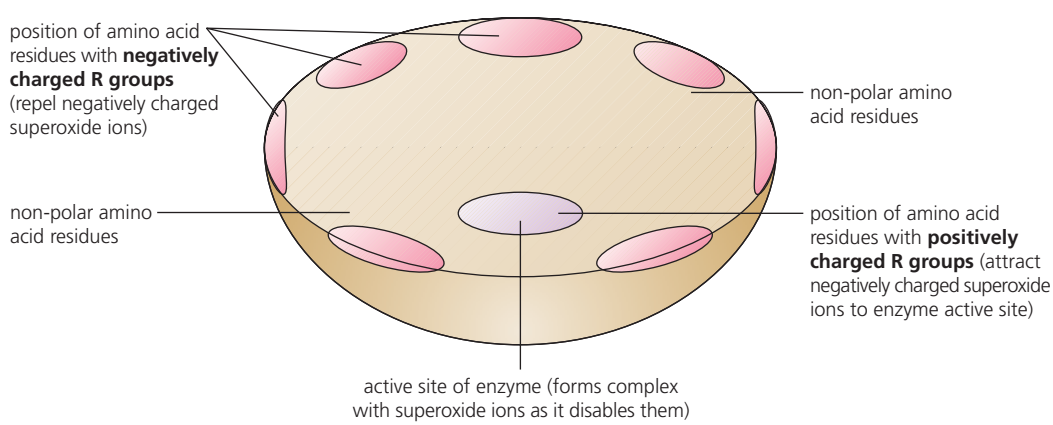
Figure 8.25 Polar and non-polar amino acids in superoxide dismutase

Superoxide dismutase is an enzyme common to all cells – breaking down superoxide ions as soon as they form as by-products of the reactions of metabolism.

The molecule is approximately saucer-shaped, with active site centrally placed.



computer simulation of tertiary structure of superoxide dismutase



Roles of proteins

Some proteins have a structural role in cells and organisms, others have biochemical or physiological roles. The roles of proteins are reviewed in Table 8.4.

Whatever their roles are in metabolism, cell proteins are in a continuous state of flux. They are continuously built up, used, and broken down again into their constituent amino acids, to be rebuilt or replaced by fresh proteins, according to the needs of the cells. This dynamic state of cell proteins is shown in Figure 8.26.

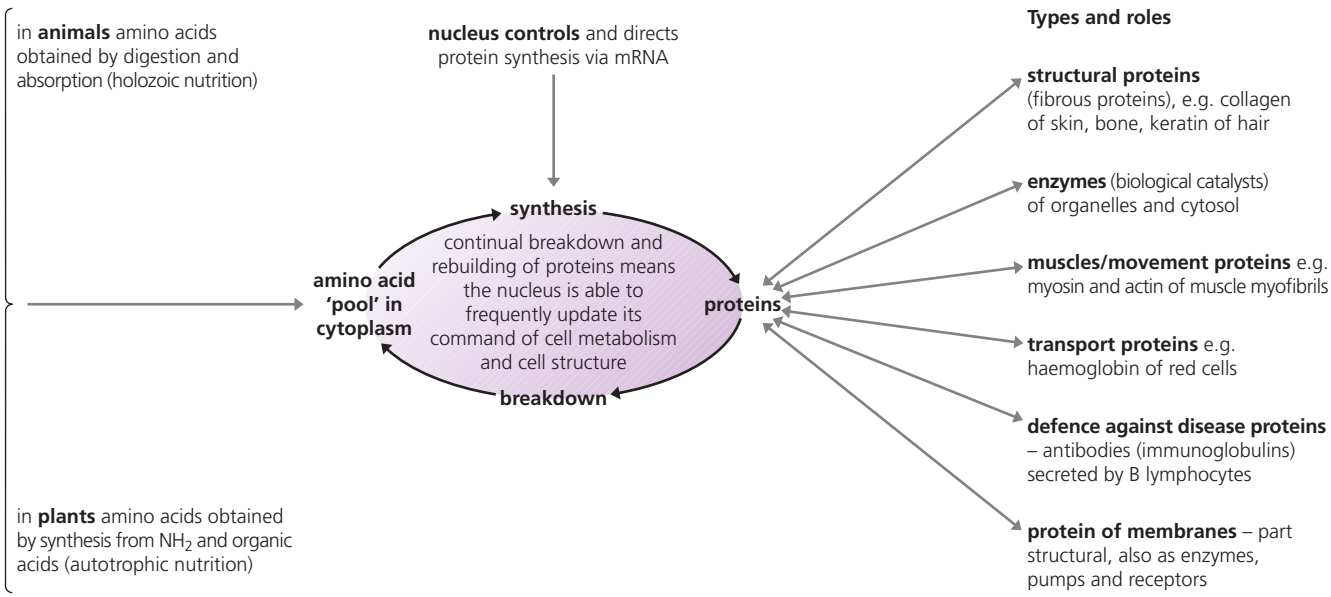


Figure 8.26 The dynamic state of cell proteins

Function	Role / examples
Biological catalysts – enzymes	Alter the speed of chemical reactions, making possible biochemical changes of organisms under the normal conditions of life (e.g. digestive enzyme pepsin and carbonic anhydrase of red cells).
Defence against disease	Antibodies secreted by B-lymphocytes (a type of white cell) in response to non-self substances (antigens) that may invade the body. Antibody proteins are known as immunoglobulins.
Transport of respiratory gases	Haemoglobin in red cells – a conjugated protein (non-protein haem part attached to globin protein) which combines with oxygen in the lungs and frees the oxygen (dissociates) in respiring tissues.
Muscle movement	Myosin and actin proteins of voluntary muscle myofibrils bring about muscle contraction by reaction with ATP, triggered by calcium ions (page 368).
Chemical messengers	Hormones are chemical messengers produced and secreted by cells of endocrine glands. Some are polypeptides or proteins (e.g. ADH, page 374, and insulin). Other hormones are steroids (lipids).
Structure and support	Fibrous proteins (e.g. collagen of bone, and keratin of hair).

Table 8.4 The functions of cellular proteins

Proteins also occur in membranes where they fulfil some of the above roles.

13 Outline three ways membrane proteins are important to the functioning of a cell.

Enzymes

7.6.1–7.6.5

There are literally many thousands of chemical reactions taking place within cells and organisms. **Metabolism** is the name we give to the totality of these enzyme-catalysed chemical reactions of life, and the molecules involved are collectively called metabolites. Many metabolites are made in organisms, but others are imported from the environment – for example, nutrients from food substances taken in, water, and the gases carbon dioxide or oxygen.

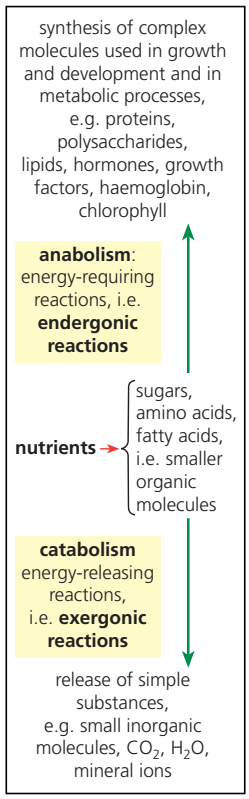
Metabolism actually consists of chains (linear sequences) and cycles of enzyme-catalysed reactions, such as we see in respiration (page 270), photosynthesis (page 284), protein synthesis (page 247), and very many other pathways. All these reactions may be classified as one of just two types, according to whether they involve the build-up or breakdown of organic molecules.

- In **anabolic reactions**, larger molecules are built up from smaller molecules (Figure 8.27). Examples of anabolism are the synthesis of proteins from amino acids and the synthesis of polysaccharides from simple sugars.
- In **catabolic reactions**, larger molecules are broken down (Figure 8.27). Examples of catabolism are the digestion of complex foods and the breakdown of sugar in respiration.

Overall:

metabolism = anabolism + catabolism

Figure 8.27 Metabolism, an overview



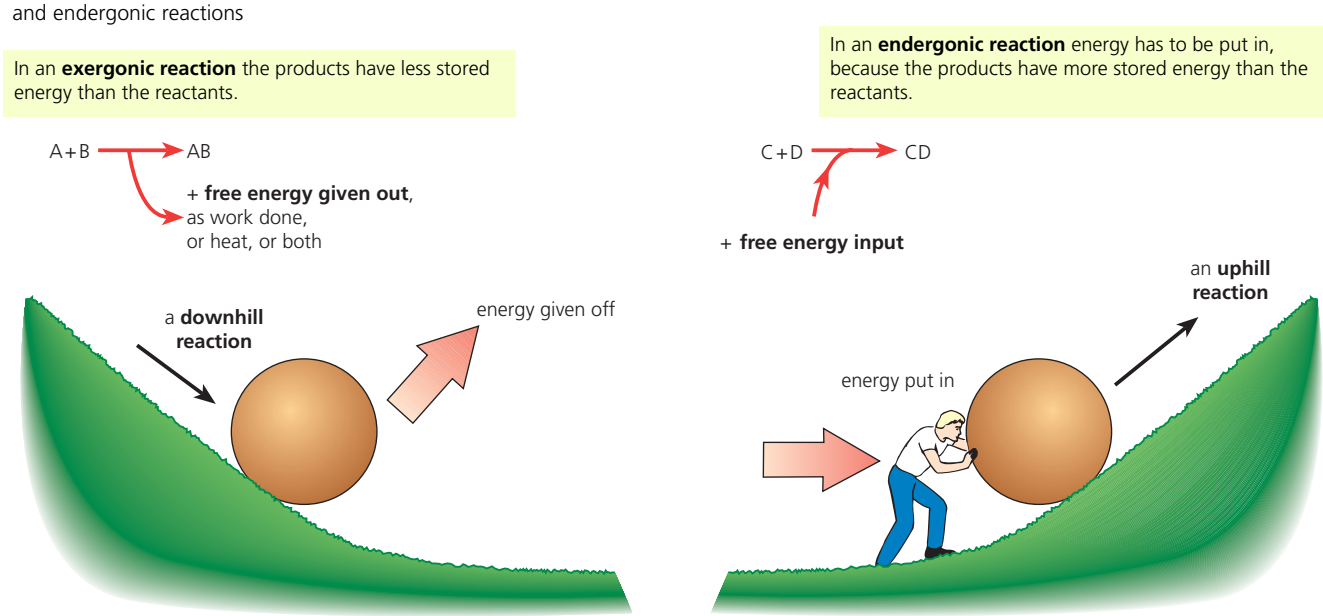
Metabolism and energy

Chemical energy exists in the structure of molecules. Every molecule contains a quantity of stored energy equal to the quantity of energy needed to synthesise it originally. Chemical energy does not exist in the chemical bonds of molecules, but rather in the structural arrangement of the molecule.

When glucose is oxidised to carbon dioxide and water in aerobic cell respiration, energy is transferred. This energy is no longer in store but is on the move; it is **active energy**. Actually, only part of the stored energy in a molecule is available, known as free energy, and can be used to do work. Reactions that release free energy are known as **exergonic reactions** (Figure 8.28). The oxidation of glucose is an example of an exergonic reaction.

On the other hand, reactions that require energy are called **endergonic reactions** (Figure 8.28). The synthesis of a protein from amino acids is an example of an endergonic reaction.

Figure 8.28 Exergonic and endergonic reactions



The many endergonic reactions that occur in metabolism are made possible by being coupled to exergonic reactions. Coupling occurs through **ATP**, which is referred to as the **energy currency** molecule in biology (Figure 3.4, page 78). Molecules of ATP work in metabolism by acting as common intermediates, linking energy-requiring and energy-yielding reactions. Metabolic processes mostly involve ATP, directly or indirectly.

Enzymes work by lowering the energy of activation

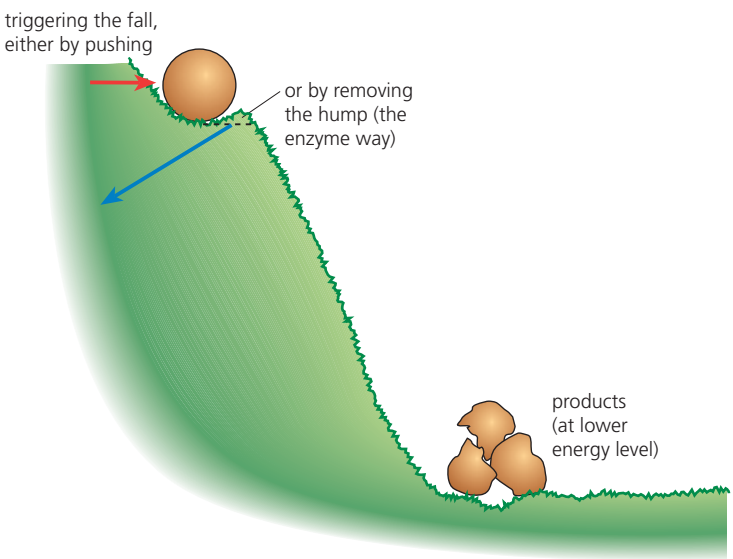
As molecules react they become unstable, high-energy intermediates, but only momentarily. We say they are in a transition state because the products are formed immediately. The products have a lower energy level than the substrate molecules. The minimum amount of energy needed to raise substrate molecules to their transition state is a called the **activation energy**. This is the energy barrier that has to be overcome before the reaction can happen.

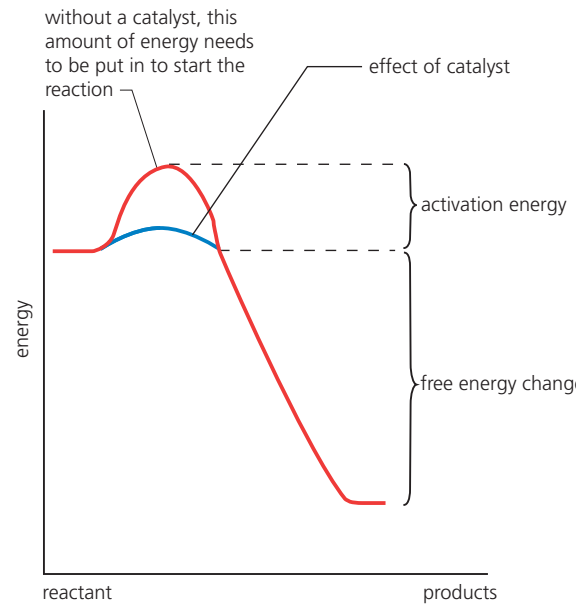
Enzymes work by lowering the amount of energy required to activate the reacting molecule.

Another model of what is going on is the boulder (substrate) perched on a slope, prevented from rolling down by a small hump (activation energy) in front of it. The boulder can be pushed over the hump, or the hump can be dug away (= lowering the activation energy), allowing the boulder to roll, and shatter at a lower level (products) (Figure 8.29).

Figure 8.29 Activation energy

'boulder on hillside' model of activation energy





Example:

sucrose + water $\xrightarrow{\text{sucrase}}$ glucose + fructose

Enzyme specificity and 'induced fit'

We have seen that enzymes are highly specific in their action, which makes them different from most inorganic catalysts. They are specific because of the way enzymes bind with their substrate at the active site, which is a pocket or crevice in the protein. To emphasise this, the binding of enzyme and substrate is sometimes referred to as the **lock and key hypothesis** of enzyme action (page 53). The necessity of binding the substrate at the active site of the enzyme before catalysis explained why changes in temperature and pH affected enzyme action since these changes often caused the shape of the enzyme molecules to change (Figures 2.18 and 2.19, pages 55–6).

However, the lock and key model does not fully account for the combined events of binding and simultaneous chemical change observed in most enzyme-catalysed reactions. At the active site, the arrangement of certain amino acid residues in the enzyme exactly matches certain groupings on the substrate molecule, enabling the enzyme–substrate complex to form. As the complex is formed, it seems an essential, critical **change of shape** is induced in the enzyme molecule. It is this shape change that is important in momentarily raising the substrate molecule to the transitional state in which it is able to react.

With a transitional state achieved, other amino acid residues of the active site bring about the breaking of particular bonds in the substrate molecule, at the point where it is temporarily held by the enzyme. Because different enzymes have different arrangements of amino acids in their active sites, each enzyme catalyses either a single chemical reaction or a group of closely related reactions (Figures 8.30 and 8.31). In other words, induced fit also accounts for the broad specificity of some enzymes (that is, the ability of some enzymes to bind to several substrates).

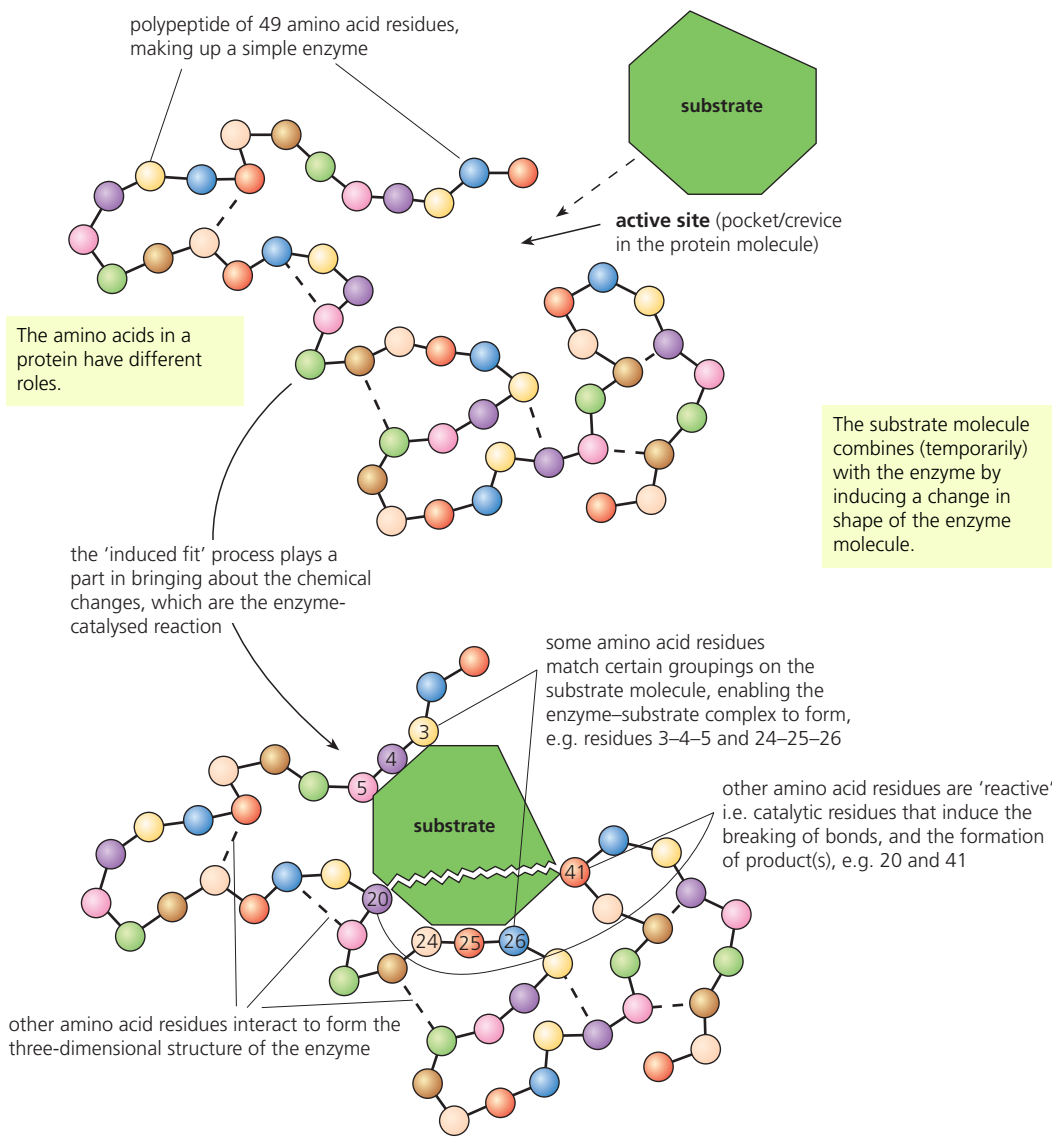


Figure 8.30 The induced fit model of enzyme action

Specificity:
- some amino acid residues allow a particular substrate molecule to 'fit'
- some amino acid residues bring about particular chemical changes.

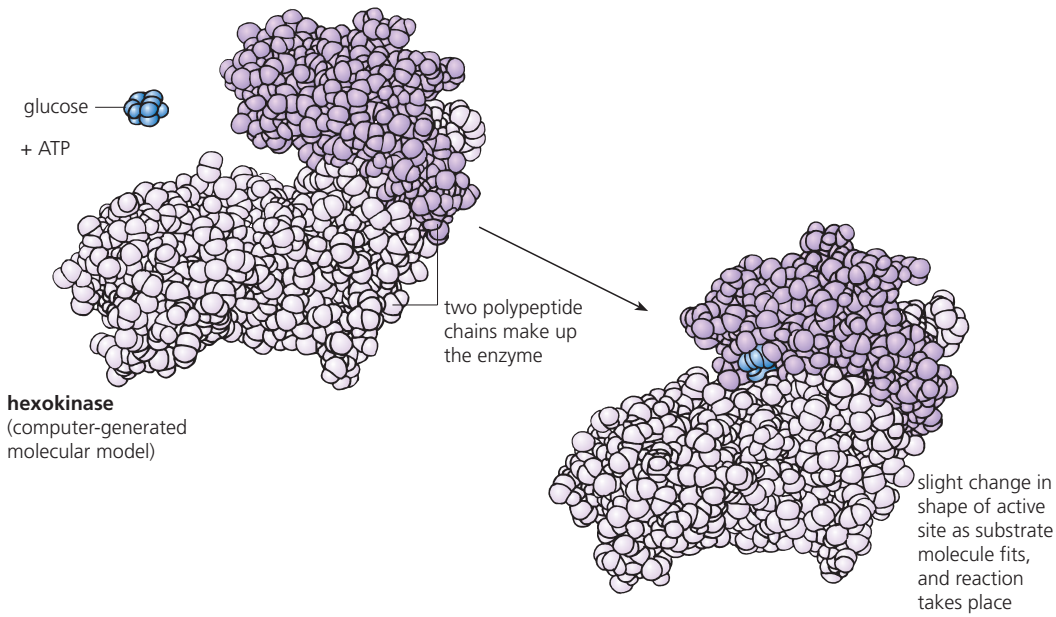


Figure 8.31 Computer-generated image of induced fit in action

14 Explain the differences between the 'lock and key' and 'induced fit' hypotheses of enzyme action.

Inhibitors of enzymes

Certain substances present in cells (and some which enter from the environment) may react with an enzyme, altering the rate of reaction. These substances are known as **inhibitors**, since their effects are generally to lower the rate of reaction. Studies of the effects of inhibitors have helped our understanding of:

- the chemistry of the active site of enzymes;
- the natural regulation of metabolism and which pathways operate;
- the ways certain commercial pesticides and some drugs work, namely by inhibiting specific enzymes and preventing particular reactions.

For example, molecules that sufficiently resemble the substrate in shape may compete to occupy the active site. They are known as **competitive inhibitors**. The enzyme that catalyses the reaction between carbon dioxide and the acceptor molecule in photosynthesis is known as ribulose bisphosphate carboxylase (Rubisco, page 285) and is competitively inhibited by oxygen in the chloroplasts.

Because these inhibitors are not acted on by the enzyme and turned into products, as normal substrate molecules are, they tend to remain attached. However, if the concentration of the substrate molecule is raised to a sufficiently high level, the inhibitor molecules are progressively displaced from the active sites.

Alternatively, an inhibitor may be unlike the substrate molecule, yet still combine with the enzyme. In these cases, the attachment occurs at some other part of the enzyme, probably quite close to the active site. Here the inhibitor either partly blocks access to the active site by substrate molecules, or it causes the active site to change shape and thus be unable to accept the substrate. These are called **non-competitive inhibitors**, since they do not compete for the active site. Adding excess substrate does not overcome their inhibiting effects (Figure 8.32). Cyanide ions combine with cytochrome oxidase but not at the active site. Cytochrome oxidase is a respiratory enzyme present in all cells, and is a component in a sequence of enzymes and carriers that oxidise the hydrogen removed from a respiratory substrate such as glucose, forming water.

Features and examples of competitive and non-competitive inhibition of enzymes are set out in Table 8.5.

Figure 8.32 Competitive and non-competitive inhibitors, the principles

When the initial rates of reaction of an enzyme are plotted against substrate concentration, the effects of competitive and non-competitive inhibitors are seen to be different.

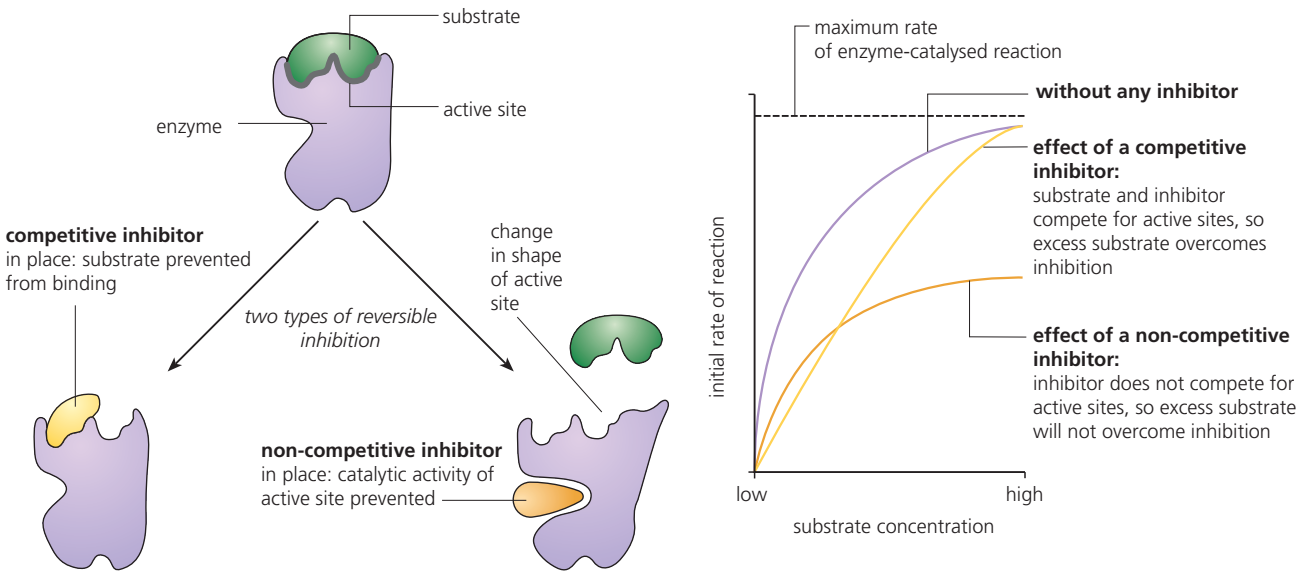


Table 8.5 Competitive and non-competitive inhibition of enzymes compared

Competitive inhibition	Non-competitive inhibition
inhibitor chemically resembles the substrate molecule and occupies (blocks) the active site	inhibitor chemically unlike the substrate molecule, but reacts with bulk of enzyme, reducing accessibility of active site
with a low concentration of inhibitor, increasing concentration of substrate eventually overcomes inhibition as substrate molecules displace inhibitor	with low concentration of inhibitor, increasing concentration of substrate cannot prevent binding – some inhibition remains at high substrate concentration
Examples: ■ O_2 competing with CO_2 for active site of Rubisco ■ malonate competing with succinate for active site of succinate dehydrogenase	Examples: ■ cyanide ions blocking cytochrome oxidase in terminal oxidation in cell aerobic respiration ■ nerve gas Sarin blocking acetyl cholinesterase in synapse transmission

The role of allostery in the control of metabolic pathways

Many metabolic pathways are switched off by a process known as **allosteric inhibition**. Allosteric enzymes have two sites. One is the active site of the enzyme and the other is an additional site on the molecule where another substance can lock in. Once such a substance is locked in, the whole enzyme is altered, resulting in inactivation of the active site.

Individual pathways in metabolism may be regulated and adjusted by means of allosteric inhibition. A case in point is end-product inhibition (Figure 8.34). Here, the **final product** inhibits the enzyme that catalyses the **first step** in the pathway.

So, in end-product inhibition, as the product molecules accumulate, the steps in their production are switched off. But these product molecules may now become the substrates in subsequent metabolic reactions. If so, the accumulated product molecules will be removed, and production of new product molecules will recommence.

Extension: Enzymes and the control of metabolism

The metabolic pathways of cells and tissues are numerous, and many interconnect. We have seen that the reactions of metabolism are made possible by the presence of specific enzymes. In addition, the enzyme machinery of cells plays a part in the **control** of the pathways of metabolism (Figure 8.33). How do enzymes achieve this?

- The **specificity** of enzymes means that a reaction occurs only in the presence of a specific enzyme. If the enzyme is not present, the reaction cannot occur.
- Because enzymes are **effective in very small amounts**, the production of a specific enzyme leads to immediate reactions and the formation of products.
- Where enzymes compete for the same substrate molecule, the amount of each enzyme present and **how readily each forms its enzyme–substrate complex** decides the extent to which pathways operate.
- Many enzymes are always present in cells, but some are **enzymes produced only in the presence of the substrate**. This is a case of the substrate molecule triggering the synthesis machinery by which the enzymes are produced.

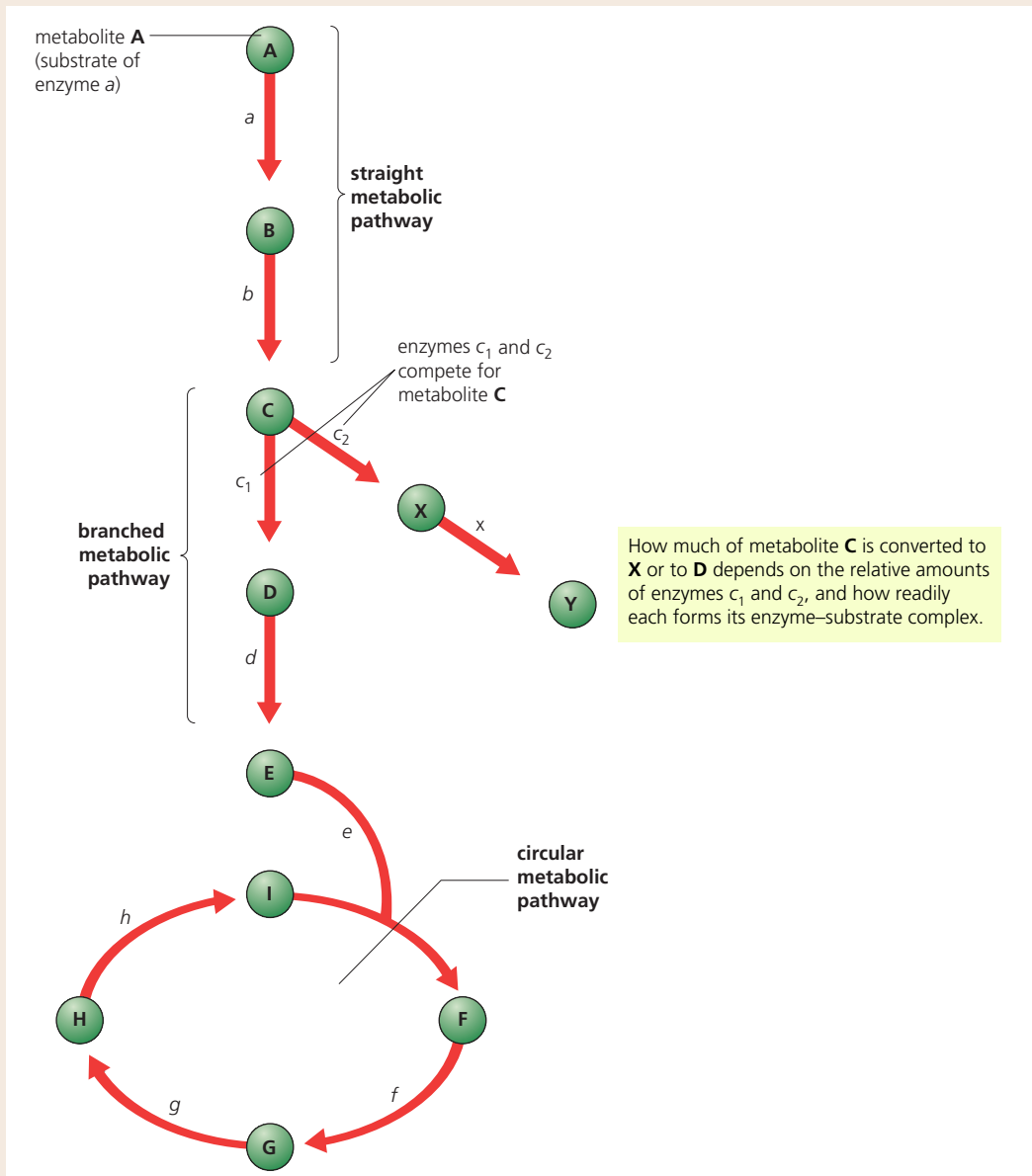
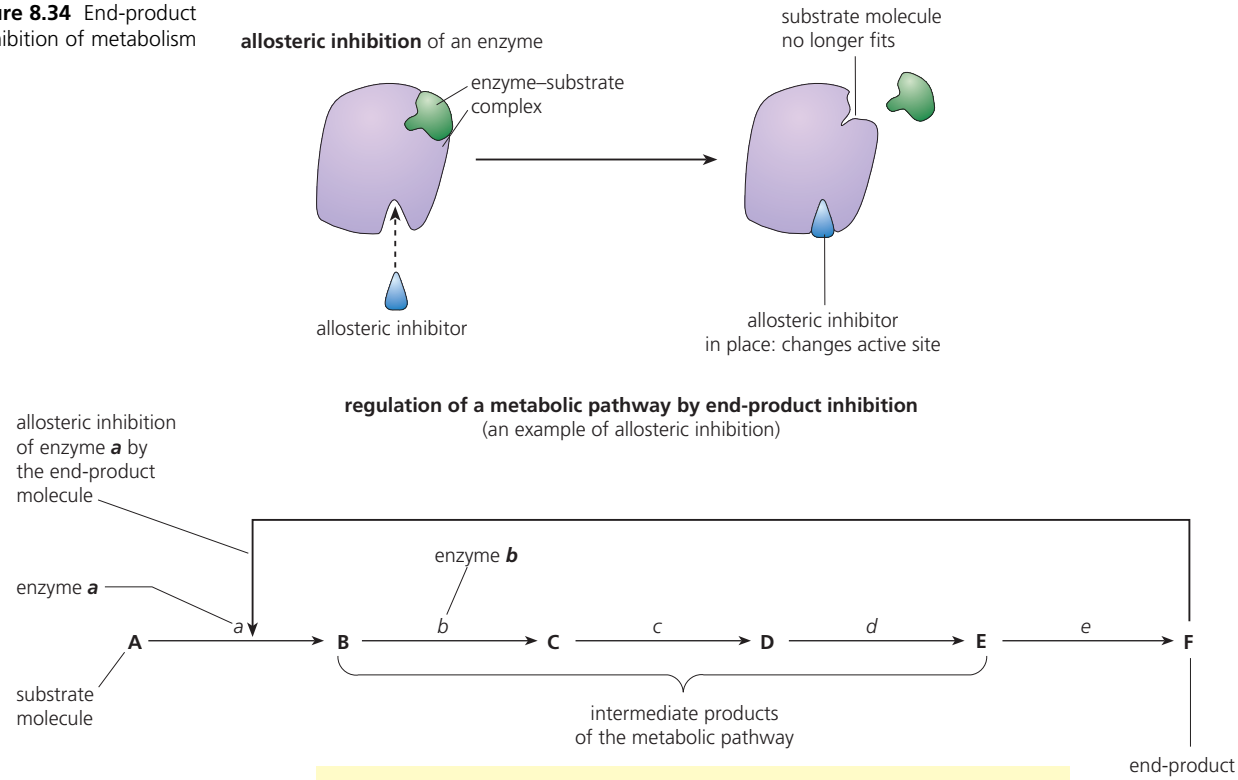


Figure 8.33 Some metabolic pathways and the roles of enzymes

Figure 8.34 End-product inhibition of metabolism



This is an example of the regulation of a metabolic pathway by **negative feedback**.

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 How do enzymes speed up biochemical reactions?

A By being used up in the reaction.
B By increasing activation energy.
C By changing the pH of the reaction.
D By lowering activation energy.

Higher Level Paper 1, May 05, Q24

Q2 Which molecule is involved in the process of transcription?

A DNA polymerase
B Helicase
C DNA ligase
D mRNA

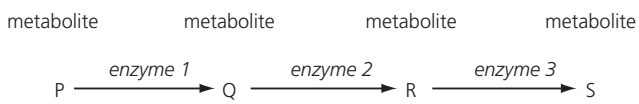
Higher Level Paper 1, May 03, Q12

Q3 Which process involves the removal of introns?

A the reaction catalysed by RNA polymerase
B the formation of mature eukaryotic mRNA
C the activation of reverse transcriptase
D the binding of ribosomes to mRNA

Higher Level Paper 2, May 02, Q21

Q4 The diagram below illustrates a simple metabolic pathway.



Which response shows allosteric feedback inhibition?

	Metabolite	Site of binding of metabolite
A	P	allosteric site of enzyme 1
B	S	allosteric site of enzyme 1
C	P	active site of enzyme 1
D	S	active site of enzyme 1

Higher Level Paper 2, May 02, Q23

Questions 5–8 cover other syllabus issues in this chapter.

Q5 **a** State the **two** types of organic molecule that form the bulk of chromosomes of eukaryotes, and their approximate proportions. (3)

b Suggest why the complex structural packaging of these molecules in chromosomes may be essential in the nucleus. (2)

c By means of an annotated diagram, explain the composition and structure of a nucleosome. (4)

Q6 **a** DNA replication is semi-conservative and occurs in the 5' → 3' direction. Explain the meaning of this statement. (6)

b By means of a fully annotated diagram, outline the steps to replication at a replication fork, making explicit the role of each of the essential enzymes. (6)

Q7 **a** Construct a concise flow diagram of the sequence of key events in transcription in the eukaryotic nucleus up to the point when mRNA is about to pass into the cytoplasm. (8)

b Identify to what extent translation in prokaryotes differs from this. (3)

c List the differences in composition between mRNA and the DNA of chromosomes. (3)

Q8 **a** Proteins are essential components of organisms. Describe the key property of the proteins specifically involved in:

i biological catalysis
ii movement driven by muscle fibres
iii structural support
iv the movement of molecules across membranes. (8)

b Outline the differences between the primary and tertiary structures of proteins. (4)

c Identify the types of bonds that maintain the secondary structure of protein. (5)

d Membrane proteins typically arrange themselves with component amino acids that are non-polar and others that are polar in contrasting positions. With the aid of an annotated diagram, explain the property of the molecules of the lipid bilayer of membranes that causes this. (8)

9 Energy transfer in cells II

STARTING POINTS

- **Respiration** occurs in every living cell, making energy available to maintain cells and to carry out activities and functions.
- Cellular respiration requiring oxygen, **aerobic respiration**, involves the oxidation of sugar to carbon dioxide and water.
- Green plants manufacture glucose by **photosynthesis**, using the raw materials carbon dioxide and water and energy from sunlight. Oxygen is produced as a waste product.
- **Chloroplasts** are the photosynthetic organelles. Chloroplasts contain chlorophyll needed for the transfer of light energy into chemicals like glucose during photosynthesis.
- This chapter extends discussion of cell respiration and photosynthesis begun in Chapter 3, pages 76–90.

Respiration and photosynthesis have traditionally been summarised in single equations:

Cellular respiration

glucose + oxygen → carbon dioxide + water + ENERGY

$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + ENERGY$

Photosynthesis

carbon dioxide + water + ENERGY → glucose + oxygen

$6CO_2 + 6H_2O + ENERGY \rightarrow C_6H_{12}O_6 + 6O_2$

At first sight it seems that the one process is simply the reverse of the other. In fact these equations are merely 'balance sheets' of the inputs and outputs. Photosynthesis and respiration occur by different pathways with many enzymes unique to one or the other.

In this chapter, the **biochemical events in respiration** and **photosynthesis** are considered in detail, beginning with aerobic respiration.

The steps of aerobic cellular respiration

8.1.1–8.1.6

In cellular respiration, the chemical energy of organic molecules such as glucose is made available for use in the living cell. Much of the energy transferred is lost in the form of heat energy, but cells are able to retain significant amounts of chemical energy in **adenosine triphosphate** (**ATP**, page 78). ATP, found in all cells, is the universal energy currency in living systems. ATP is a relatively small, soluble molecule. It is able to move, by facilitated diffusion, from the mitochondria where it is synthesised to all the very many sites where energy is required, such as in muscles for contraction movements (page 368), in membranes for active transport (page 23) and in ribosomes for protein synthesis (page 248).

How is energy transferred from respiratory substrates like glucose?

In summary, glucose is a relatively large molecule containing six carbon atoms all in a reduced state. During aerobic cellular respiration, glucose undergoes a series of enzyme-catalysed oxidation reactions (Figure 9.1). These reactions are grouped into three major phases:

- **glycolysis**, in which glucose is converted to pyruvate;
- the **Krebs cycle**, in which pyruvate is converted to carbon dioxide;
- the **electron-transport system**, in which hydrogen removed in oxidation reactions of glycolysis and the Krebs cycle is converted to water, and the bulk of the ATP is synthesised.

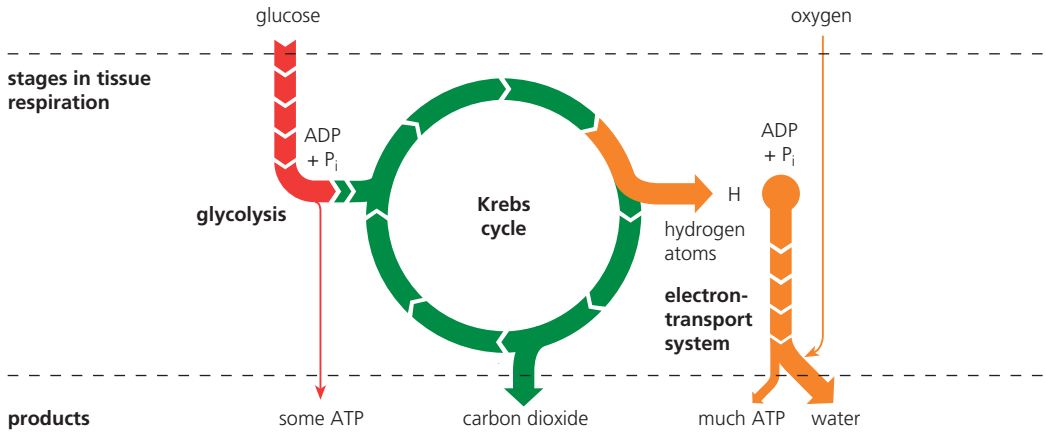


Figure 9.1 The three phases of aerobic cellular respiration

Respiration as a series of redox reactions

The terms '**reduction**' and '**oxidation**' recur frequently in respiration.

What do these terms mean?

In cellular respiration, glucose is oxidised to carbon dioxide, but at the same time, oxygen is reduced to water (Figure 9.2). In fact, tissue respiration is a series of oxidation–reduction reactions, so described because when one substance in a reaction is oxidised another is automatically reduced. The short-hand name for **red**uction–**ox**idation reactions is **redox reactions**.

In biological oxidation, oxygen atoms may be added to a compound, but alternatively, hydrogen atoms may be removed. In respiration, all the hydrogen atoms are gradually removed from glucose. They are added to hydrogen acceptors, which are themselves reduced.

Since a hydrogen atom consists of an electron and a proton, gaining hydrogen atom(s) (a case of reduction) involves gaining one or more electrons. In fact, the best definition of **oxidation is the loss of electrons**, and **reduction is the gain of electrons**. Remembering this definition has given countless people problems, so a mnemonic has been devised:

OIL RIG = **O**xidation **I**s **L**oss of electrons; **R**eduction **I**s **G**ain of electrons.

Redox reactions take place in biological systems because of the presence of a compound with a strong tendency to take electrons from another compound (an **oxidising agent**) or the presence of a compound with a strong tendency to donate electrons to another compound (a **reducing agent**).

Another feature of oxidation and reduction is **energy change**. When reduction occurs, energy is absorbed (an **endergonic reaction**, Figure 8.28, page 261). When oxidation occurs, energy is released (an **exergonic reaction**).

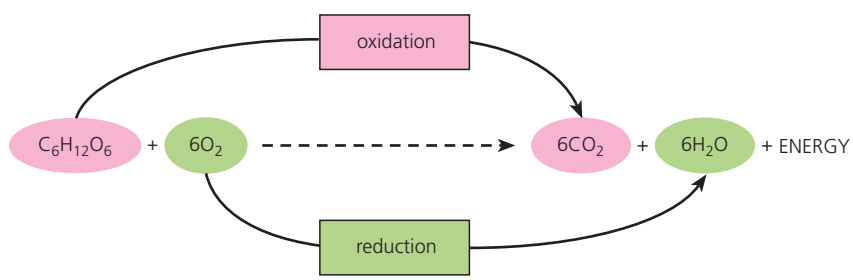


Figure 9.2 Respiration as a redox reaction

oxidation involves:
- addition of oxygen
- removal of hydrogen
- loss of electrons
- release of energy

reduction involves:
- removal of oxygen
- addition of hydrogen
- gain of electrons
- uptake of energy

An example of energy release in oxidation is the burning of a fuel in air. Here, energy is given out as heat. In fact, the amount of energy in a molecule depends on its degree of oxidation. An oxidised substance has less stored energy than a reduced substance, illustrated by the fuel molecule methane (CH_4) which has more stored chemical energy than carbon dioxide (CO_2).

Cellular respiration

Glycolysis

Glycolysis is a linear series of reactions in which a 6-carbon sugar molecule is broken down to two molecules of the 3-carbon pyruvate ion (Figure 9.3). The enzymes of glycolysis are located in the cytoplasm outside organelles (known as the cytosol), rather than in the mitochondria. Glycolysis occurs by four stages:

- **Phosphorylation** by reaction with ATP is the way glucose is first activated, forming glucose phosphate. Conversion to fructose phosphate follows, and a further phosphate group is then added at the expense of another molecule of ATP. So, two molecules of ATP are consumed per molecule of glucose respired, *at this stage of glycolysis*.
- **Lysis** (splitting) of the fructose bisphosphate now takes place, forming two molecules of 3-carbon sugar, called **triose phosphate**.
- **Oxidation** of the triose phosphate molecules occurs by removal of hydrogen. The enzyme for this reaction (a dehydrogenase) works with a coenzyme, **nicotinamide adenine dinucleotide (NAD)**. NAD is a molecule that can accept hydrogen ions (H^+) and electrons (e^-). In this reaction, the NAD is reduced to NADH and H^+ (reduced NAD):

  $NAD^+ + 2H^+ + 2e^- \rightarrow NADH + H^+$ (sometimes represented as $NADH_2$)

  (Reduced NAD can pass hydrogen ions and electrons on to other acceptor molecules (see below), and when it does, it becomes oxidised back to NAD.)

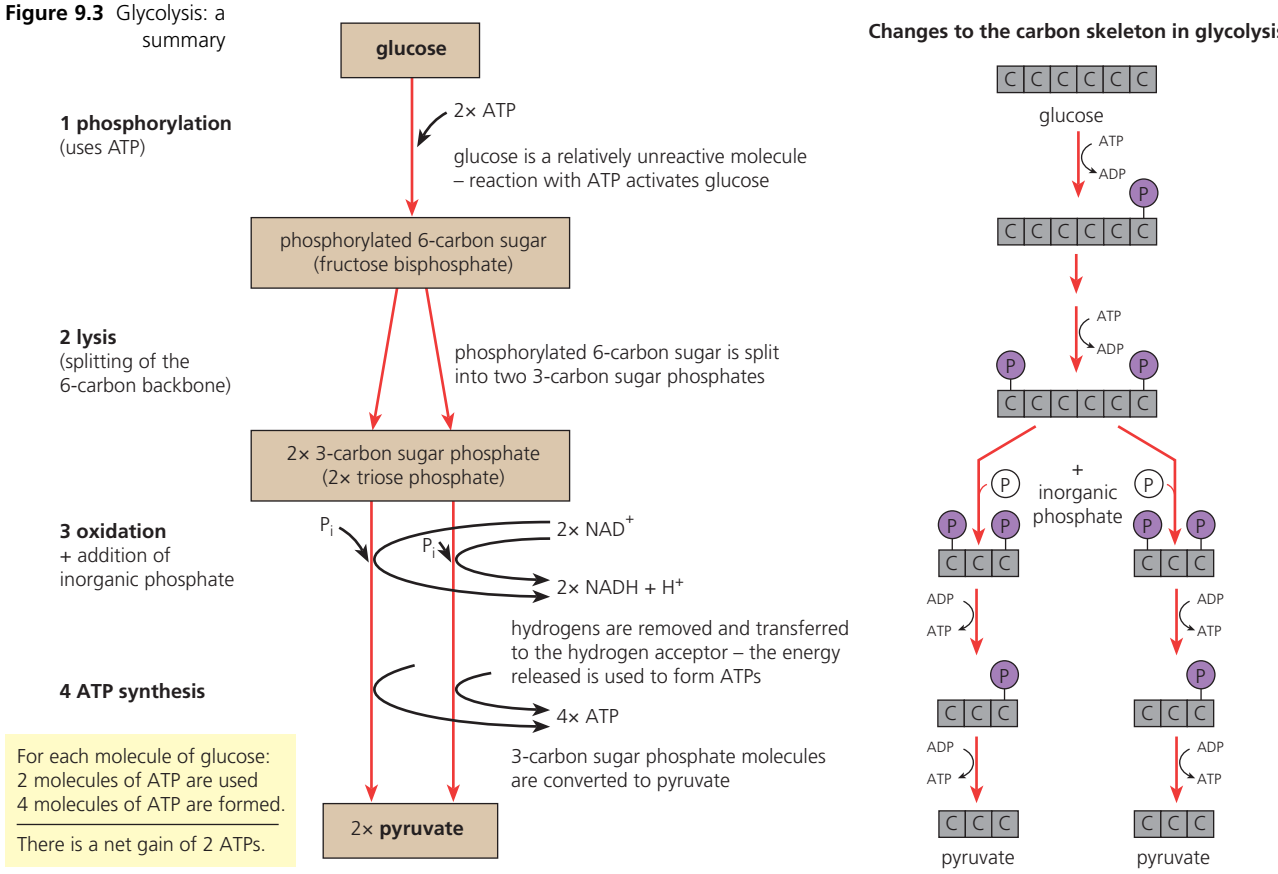


Figure 9.3 Glycolysis: a summary

- **ATP formation**, which occurs twice in the reactions by which each triose phosphate molecule is converted to pyruvate. This form of ATP synthesis is described as **at substrate level** in order to differentiate it from the bulk of ATP synthesis that occurs later in cell respiration, during operation of the electron-transport chain (see below). As two molecules of triose phosphate are converted to pyruvate, four molecules of ATP are synthesised *at this stage of glycolysis*. So in total, there is a **net gain of two ATPs in glycolysis**.

1 **State** which of the following are produced during glycolysis:

carbon dioxide	NADH	ATP	pyruvate
lactate	glycogen	NAD^+	glucose

The mitochondria and the steps of the Krebs cycle

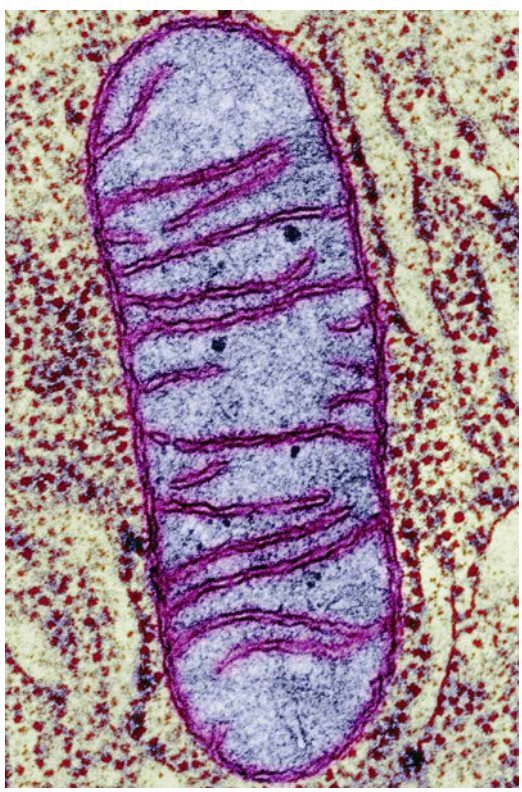
From this point on, the steps of aerobic cell respiration occur within organelles known as **mitochondria** (singular, **mitochondrion**). This is where the enzymes concerned with the Krebs cycle and electron transport are all located.

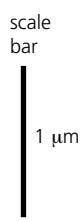
Mitochondria are found in the cytoplasm of all eukaryotic cells, usually in very large numbers. The structure of a mitochondrion is best investigated by transmission electron microscopy. The clarity of the resulting TEMs often makes possible a line drawing to show the mitochondrion's double membrane around the matrix, and the way the inner membrane intucks to form cristae (Figure 9.4). Of course, the mitochondrion is a three-dimensional structure.

2 Using the scale bar in Figure 9.4, **calculate** the length and width in microns of the mitochondrion shown in section in the TEM.

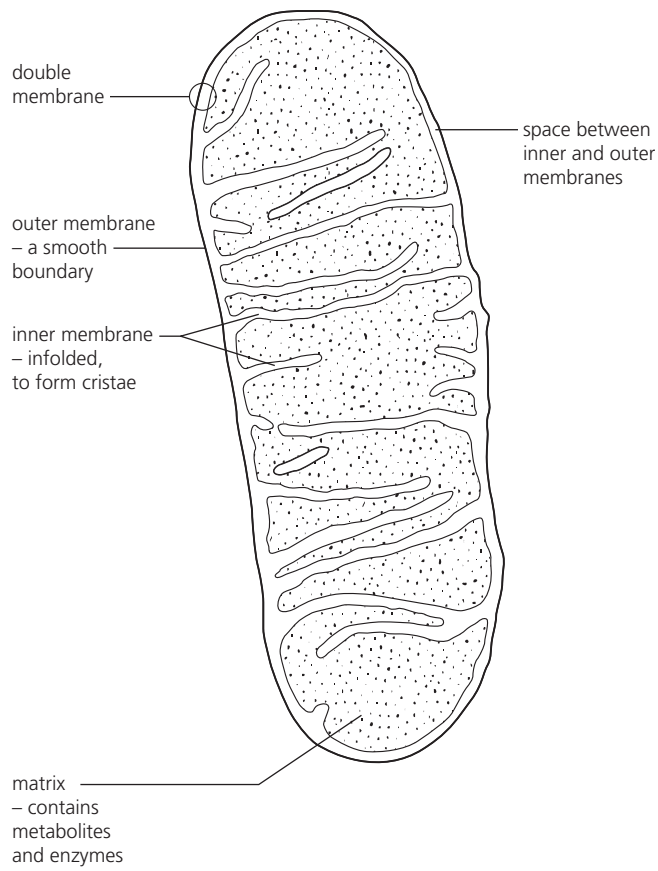
Figure 9.4 TEM of a mitochondrion, with an interpretive drawing

TEM of a mitochondrion. Mitochondria are the site of the aerobic stage of respiration.





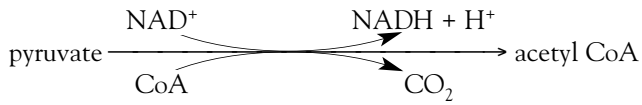
interpretive drawing



Link reaction

Pyruvate diffuses into the matrix of the mitochondrion as it forms, and is metabolised there.

First, the 3-carbon pyruvate is decarboxylated by removal of carbon dioxide and, at the same time, oxidised by removal of hydrogen. Reduced NAD is formed. The product of this **oxidative decarboxylation** reaction is an acetyl group – a 2-carbon fragment. This acetyl group is then combined with a coenzyme called coenzyme A (CoA), forming **acetyl coenzyme A (acetyl CoA)**. The production of acetyl coenzyme A from pyruvate is known as the link reaction because it connects glycolysis to reactions of the **Krebs cycle**, which now follow.



Krebs cycle

The Krebs cycle is named after Hans Krebs who discovered it, but it is also sometimes referred to as the citric acid cycle, after the first intermediate acid formed.

The acetyl coenzyme A enters the Krebs cycle by reacting with a **4-carbon organic acid** (oxaloacetate, OAA). The products of this reaction are a **6-carbon acid** (citrate) and coenzyme A which is released and re-used in the link reaction.

The citrate is then converted back to the 4-carbon acid (an acceptor molecule, in effect) by **the reactions of the Krebs cycle**. These involve the following changes:

- **two molecules of carbon dioxide are given off**, in separate decarboxylation reactions;
- **a molecule of ATP is formed**, as part one of the reactions of the cycle; as in glycolysis, this ATP synthesis is at substrate level;
- **three molecules of reduced NAD are formed**;
- **one molecule of another hydrogen accepter**, the coenzyme flavin adenine dinucleotide (FAD) is reduced (NAD is the chief hydrogen-carrying coenzyme of respiration but FAD has this role in the Krebs cycle).

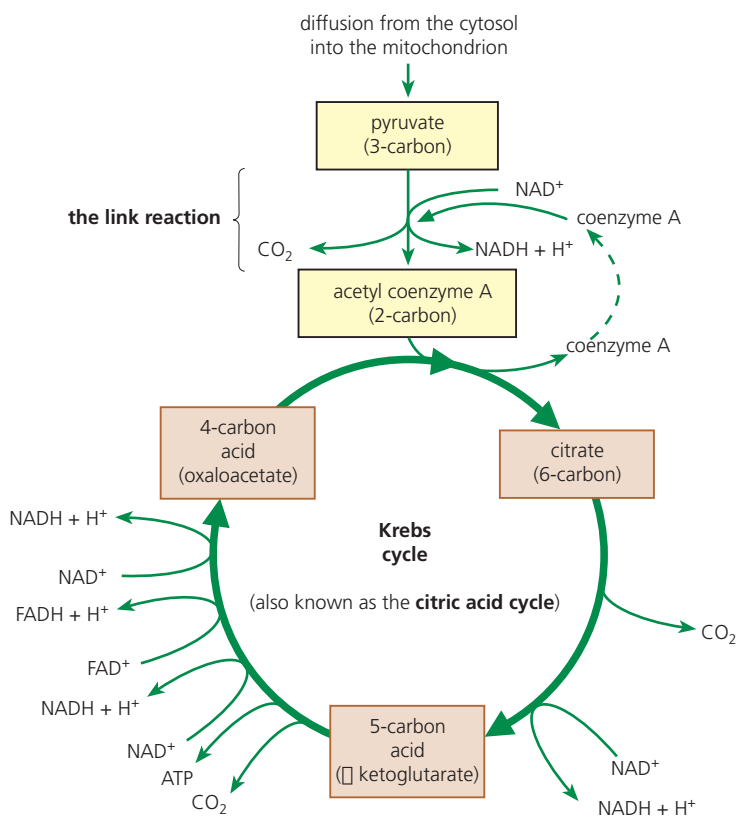


Figure 9.5 The Krebs cycle: a summary

There are several other organic acid intermediates in the cycle not shown here.

Because glucose is converted to two molecules of pyruvate in glycolysis, the whole Krebs cycle sequence of reactions 'turns' twice for every molecule of glucose that is metabolised by aerobic cellular respiration (Figure 9.5).

Now we are in a position to summarise the changes to the molecule of glucose that occurred in the reactions of glycolysis and the Krebs cycle. A 'budget' of the products of glycolysis and two turns of the Krebs cycle is shown in Table 9.1.

Table 9.1 Net products of aerobic respiration of glucose at the end of the Krebs cycle

	Product			
Step	**CO_2**	**ATP**	**reduced NAD**	**reduced FAD**
glycolysis	0	2	2	0
link reaction (pyruvate → acetyl CoA)	2	0	2	0
Krebs cycle	4	2	6	2
Totals	$6CO_2$	4 ATP	10 reduced NAD	2 reduced FAD

3 **Outline** the types of reaction catalysed by:

a dehydrogenases
b decarboxylases.

Extension: Fats can be respired

In addition to glucose, **fats** (lipids) are also commonly used as respiratory substrates – they are first broken down to **fatty acids** (and glycerol). Fatty acid is 'cut up' into 2-carbon fragments and fed into the Krebs cycle via **coenzyme A**. Vertebrate muscle is well adapted to the respiration of fatty acids in this way (as is our heart muscle), and they are just as likely as glucose to be the respiratory substrate.

Terminal oxidation and oxidative phosphorylation

The removal of pairs of hydrogen atoms from various intermediates of the respiratory pathway is a feature of several of the steps in glycolysis and the Krebs cycle. On most occasions, oxidised NAD is converted to reduced NAD, but once in the Krebs cycle an alternative hydrogen-acceptor coenzyme, FAD, is reduced.

In the final stage of aerobic respiration, the hydrogen atoms (or their electrons) are transported along a **series of carriers**, from the reduced NAD (or FAD), to be combined with oxygen to form water.

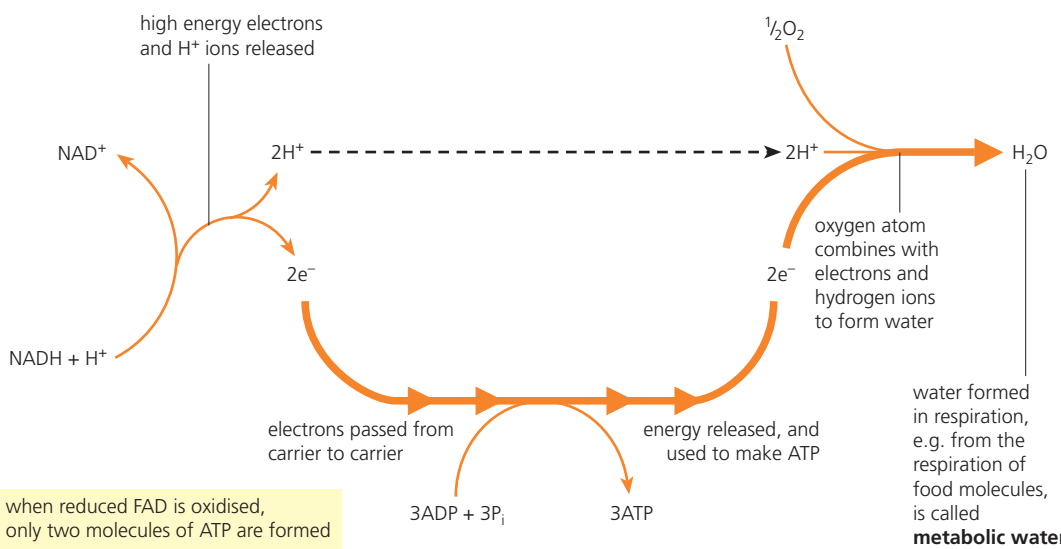


Figure 9.6 Terminal oxidation and the formation of ATPs

4 **Suggest** how the absence of oxygen in respiring tissue might 'switch off' both the Krebs cycle and terminal oxidation.

As electrons are passed between the carriers in the series, **energy is released**. Release of energy in this manner is controlled and can be used by the cell. The energy is transferred to ADP and P_i, to form ATP. Normally, for every molecule of reduced NAD which is oxidised (that is, for every pair of hydrogens) approximately three molecules of ATP are produced.

The process is summarised in Figure 9.6. The total yield from aerobic respiration is 38 ATPs per molecule of glucose respired.

Phosphorylation by chemiosmosis?

How could a mitochondrion use the energy made available in the flow of electrons between carrier molecules to drive the synthesis of ATP?

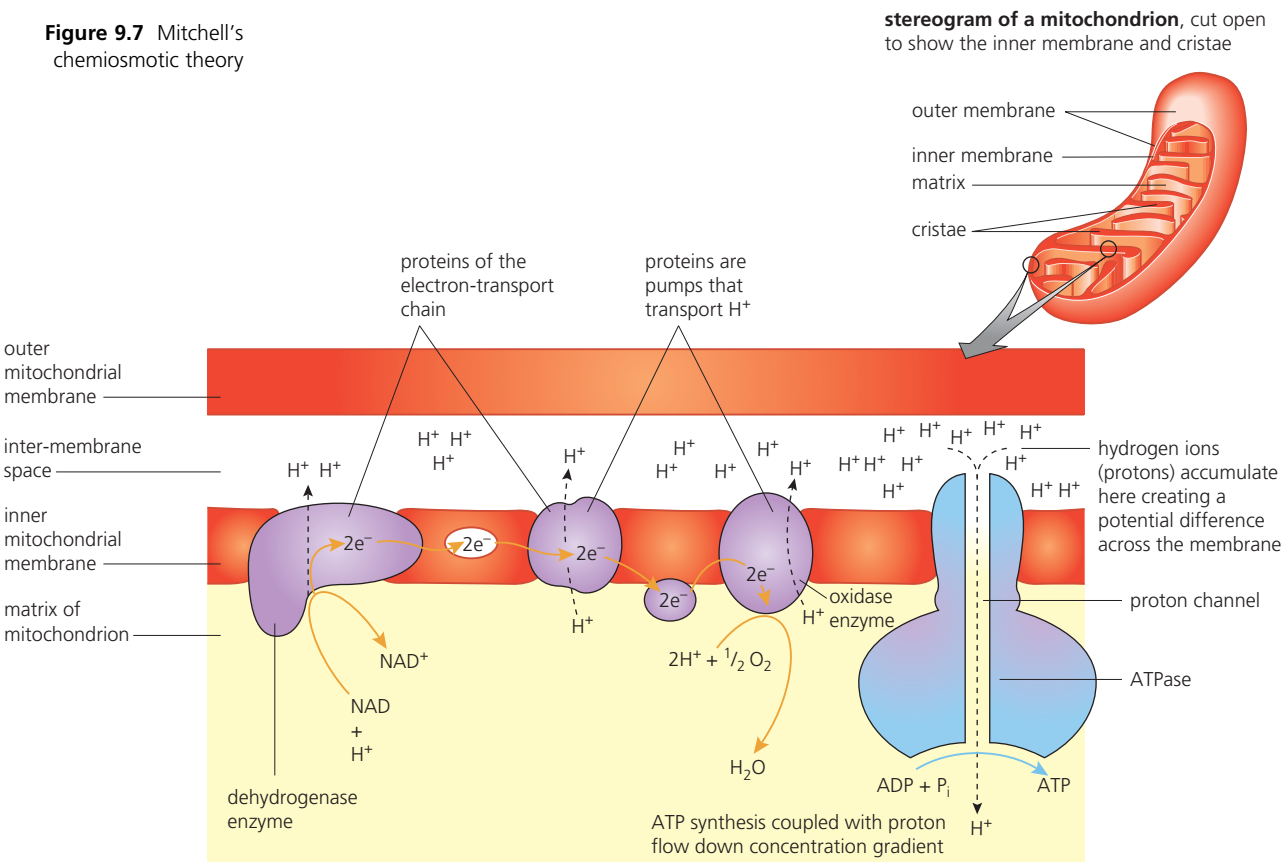
It was a biochemist, Peter Mitchell, who first suggested the chemiosmotic theory when he was at an independently funded research institute in Cornwall, UK. This was in 1961; at the time he was studying the metabolism of bacteria. His hypothesis was not generally accepted for many years, but about two decades later he was awarded a Nobel Prize for his discovery.

Chemiosmosis is a process by which the synthesis of ATP is coupled to electron transport via the movement of protons (Figure 9.7). The electron-carrier proteins are arranged in the inner mitochondrial wall in a highly ordered way. These carrier proteins oxidise the reduced coenzymes, and energy from the oxidation process is used to pump hydrogen ions (protons) from the matrix of the mitochondrion into the space between inner and outer mitochondrial membranes.

5 When ATP is synthesised in mitochondria, **explain** where the electrochemical gradient is set up, and in which direction protons move.

Here they accumulate – incidentally, causing the pH to drop. Because the inner membrane is largely impermeable to ions, a significant **gradient in hydrogen ion concentration** builds up across the inner membrane, generating a potential difference across the membrane. This represents a store of potential energy. Eventually, the protons do flow back into the matrix, but this occurs via the channels in **ATP synthetase** enzymes (ATPase), also found in the inner mitochondrial membrane. As the protons flow down their concentration gradient through the enzyme, the energy is transferred as ATP synthesis occurs.

Figure 9.7 Mitchell's chemiosmotic theory



Site of cellular respiration – structure in relation to function

The location of the stages of cellular respiration are summarised in Figure 9.8. Once pyruvate has been formed from glucose in the cytosol, the remainder of the pathway of aerobic cell respiration is located in the mitochondria. The relationship between structure and function in this organelle is summarised in Table 9.2.

Table 9.2 Mitochondrial structure in relation to function

Structure	Function / role
external double membrane	permeable to pyruvate, CO_2, O_2 and NAD/NADH + H^+
matrix	site of enzymes of link reaction and Krebs cycle
inner membrane	surface area greatly increased by intucking to form cristae – since the electron-transport chain and ATP synthetase enzymes are housed here, opportunities for ATP synthesis are enhanced
	impermeable to hydrogen ions (protons) – permitting the establishment of a potential difference between the inter-membrane space and the matrix
inter-membrane space	relatively tiny space – allowing the accumulation of hydrogen ions (protons) to generate a large concentration difference with matrix, facilitating phosphorylation

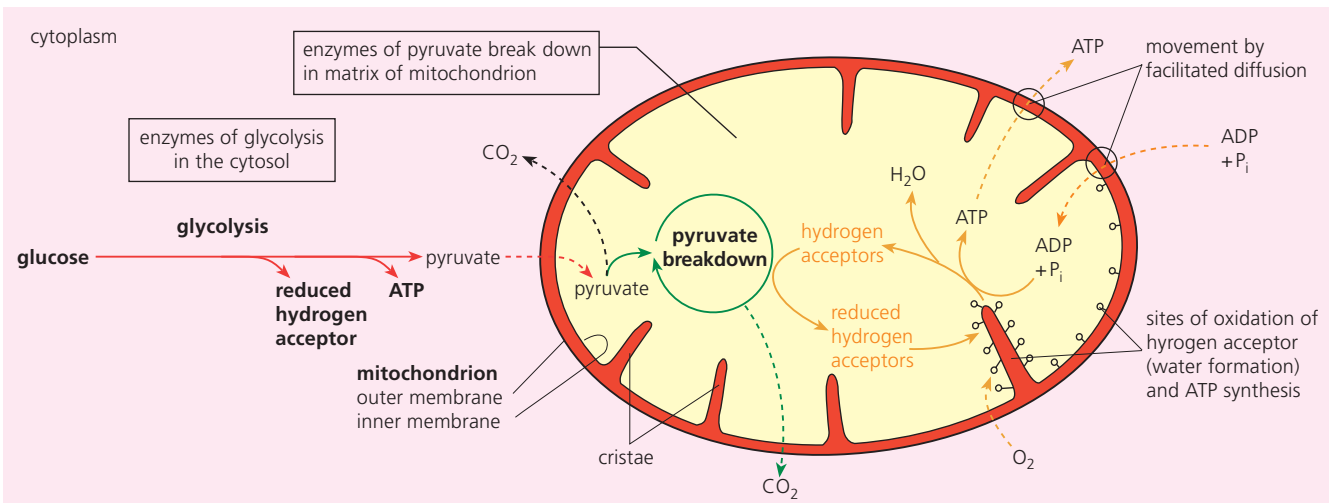


Figure 9.8 The site of respiration in the eukaryotic cell

Extension: Respiration as a source of intermediates

We have seen that the main role of respiration is to provide a pool of ATP, and that this is used to drive the endergonic reactions of synthesis. But the compounds of the glycolysis pathway and the Krebs cycle (**respiratory intermediates**) may also serve as the **starting points of synthesis of other metabolites** needed in the cells and tissues of the organism. These include polysaccharides like starch and cellulose, glycerol and fatty acids, amino acids, and many others.

6 **Distinguish** between the following pairs:

a substrate and intermediate
b glycolysis and the Krebs cycle
c oxidation and reduction

The steps of photosynthesis

8.2.1–8.2.6

Just as the mitochondria are the site of reactions of the Krebs cycle and respiratory ATP formation, so the **chloroplasts** are the organelles where the reactions of photosynthesis occur.

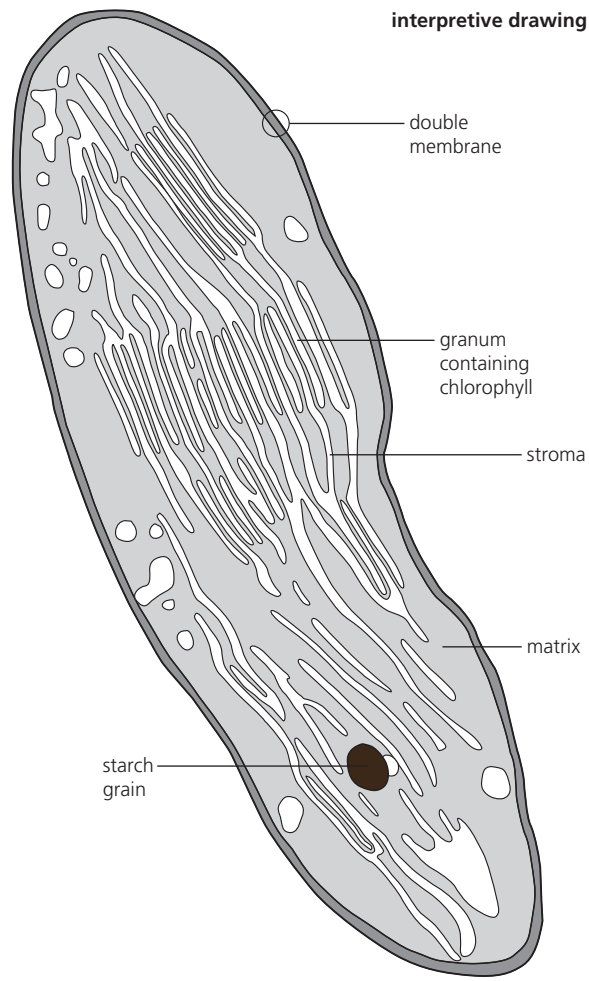
Chloroplasts are found in green plants. They contain the photosynthetic pigments along with the enzymes and electron-transport proteins for the reduction of carbon dioxide to sugars and for ATP formation, using light energy. The detailed structure of chloroplasts is investigated with the transmission electron microscope. TEMs of thin sections of a chloroplast show the arrangement of the membranes within this large organelle (Figure 9.9).

You can see there is a double membrane around the chloroplast, and that the inner of these membranes intucks at various points to form a system of branching membranes. Here, these membranes are called **thylakoids**. Thylakoid membranes are organised into flat, compact, circular piles called **grana** (singular, **granum**), almost like stacks of coins. Between the grana are loosely arranged tubular membranes suspended in a watery matrix, called the **stroma**.

Chlorophyll, the photosynthetic pigment that absorbs light energy, occurs in the grana. Also suspended in the matrix are starch grains, lipid droplets and ribosomes. We shall see that the structure, composition and arrangements of membranes are central to the biochemistry of photosynthesis, just as the mitochondrial membranes are the sites of many of the reactions of aerobic cell respiration.

Figure 9.9 TEM of a chloroplast, with an interpretive drawing

TEM of a thin section of a chloroplast (×15 000)



The reactions of photosynthesis

Photosynthesis is a complex set of many reactions that take place in illuminated chloroplasts. However, biochemical studies of photosynthesis by several teams of scientists have established that the many reactions by which light energy brings about the production of sugars, using the raw materials water and carbon dioxide, fall naturally into two inter-connected stages (Figure 9.10).

- **The light-dependent reactions**, in which light energy is used directly to split water (known as **photolysis** for obvious reasons). Hydrogen is then removed and retained by the photosynthesis-specific hydrogen acceptor, known as $NADP^+$. ($NADP^+$ is very similar to the coenzyme NAD^+ of respiration, but it carries an additional phosphate group, hence NADP.) At the same time, ATP is generated from ADP and phosphate, also using energy from light. This is known as photophosphorylation. Oxygen is given off as a waste product of the light-dependent reactions. This stage occurs in the grana of the chloroplasts.
- **The light-independent reactions**, in which sugars are built up using carbon dioxide. Of course, the products of the light-dependent reactions (ATP and reduced hydrogen acceptor NADPH + H^+) are used in sugar production. This stage occurs in the stroma of the chloroplast. It requires a continuous supply of the products of the light-dependent reactions, but does not directly involve light energy (hence the name). Names can be misleading, however, because this stage is an integral part of photosynthesis, and photosynthesis is a process that is powered by light energy.

We shall now consider both sets of reactions, in order to understand more about how these complex changes are brought about.

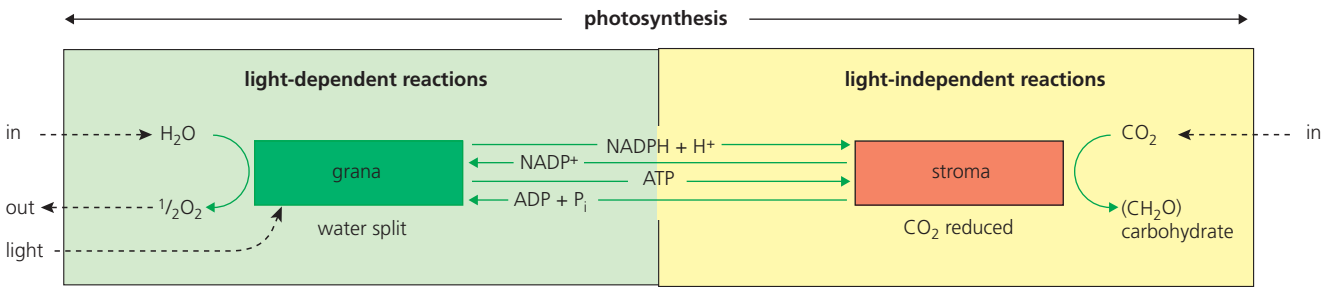


Figure 9.10 The two sets of reactions of photosynthesis, inputs and outputs

The light-dependent reactions

In the light-dependent stage, light energy is trapped by photosynthetic pigment, chlorophyll. Chlorophyll molecules do not occur haphazardly in the grana. Rather, they are grouped together in structures called **photosystems**, held in the thylakoid membranes of the **grana** (Figure 9.11).

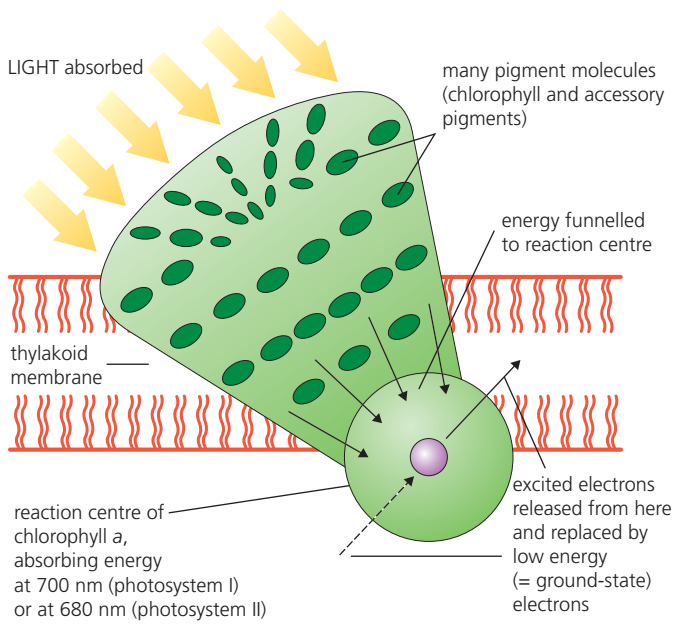


Figure 9.11 The structure of photosystems

In each photosystem, several hundred chlorophyll molecules plus accessory pigments (carotene and xanthophylls) are arranged. All these pigment molecules harvest light energy, and they funnel the energy to a single chlorophyll molecule of the photosystem, known as the **reaction centre**. The different pigments around the reaction centres absorb light energy of slightly different wavelengths.

There are **two types of photosystem** present in the thylakoid membranes of the grana, identified by the wavelength of light that the chlorophyll of the reaction centre absorbs:

- **photosystem I** has a reaction centre activated by light of wavelength 700 nm. This reaction centre is referred to as P700;
- **photosystem II** has a reaction centre activated by light of wavelength 680 nm. This reaction centre is referred to as P680.

Photosystems I and II have specific and differing roles, as we shall see shortly. However, they occur grouped together in the thylakoid membranes of the grana, along with specific proteins that function quite specifically. These consist of:

- **enzymes** catalysing
  – splitting of water into hydrogen ions, electrons and oxygen atoms;
  – formation of ATP from ADP and phosphate (P_i);
  – conversion of oxidised H-carrier ($NADP^+$) to reduced carrier ($NADPH + H^+$);
- **electron carrier molecules**.

7 **Construct** a table of the components of photosystems that identifies the role of each.

When light energy reaches the reaction centre, **ground-state electrons** of the key chlorophyll molecule are raised to an 'excited state' by the light energy received. As a result, **high-energy electrons** are released from this chlorophyll molecule, and these electrons bring about the biochemical changes of the light-dependent reactions. The spaces vacated by the high-energy (excited) electrons are continuously refilled by non-excited or ground-state electrons.

We will examine this sequence of reactions in the two photosystems next.

First, the excited electrons from photosystem II are picked up and passed along a chain of electron carriers. As these excited electrons pass, some of the energy causes the pumping of hydrogen ions (protons) from the chloroplast's matrix into the thylakoid spaces. Here they accumulate, causing the pH to drop. The result is a proton gradient that is created across the thylakoid membrane, and which sustains the synthesis of ATP. This is another example of chemiosmosis (see below).

As a result of these energy transfers, the excitation level of the electrons falls back to ground state and they come to fill the vacancies in the reaction centre of photosystem I. Thus, electrons have been transferred from photosystem II to photosystem I.

Meanwhile the 'holes' in the reaction centres of photosystem II are filled by electrons (in their ground state) from water molecules. In fact, the positively charged 'vacancies' in photosystem II are powerful enough to cause the splitting of water (photolysis), in the presence of a specific enzyme. The reaction this enzyme catalyses then triggers release of hydrogen ions and oxygen atoms, as well as ground-state electrons.

The oxygen atoms combine to form molecular oxygen, the waste product of photosynthesis. The hydrogen ions are used in the reduction of $NADP^+$ (see below).

Photophosphorylation

In the grana of the chloroplasts, the synthesis of ATP is coupled to electron transport via the movement of protons by chemiosmosis, as it was in mitochondria (Table 9.3). Here it is the hydrogen ions trapped within the thylakoid space which flow out via ATP synthetase enzymes, down their electrochemical gradient. At the same time, ATP is synthesised from ADP and P_i.

We have seen that the excited electrons which provided the energy for ATP synthesis originated from water, and move on to fill the vacancies in the reaction centre of photosystem II. They are subsequently moved on to the reaction centre in photosystem I, and finally are used to reduce $NADP^+$. Because the pathway of the electrons is linear, the photophosphorylation reaction in which they are involved is described as **non-cyclic photophosphorylation**.

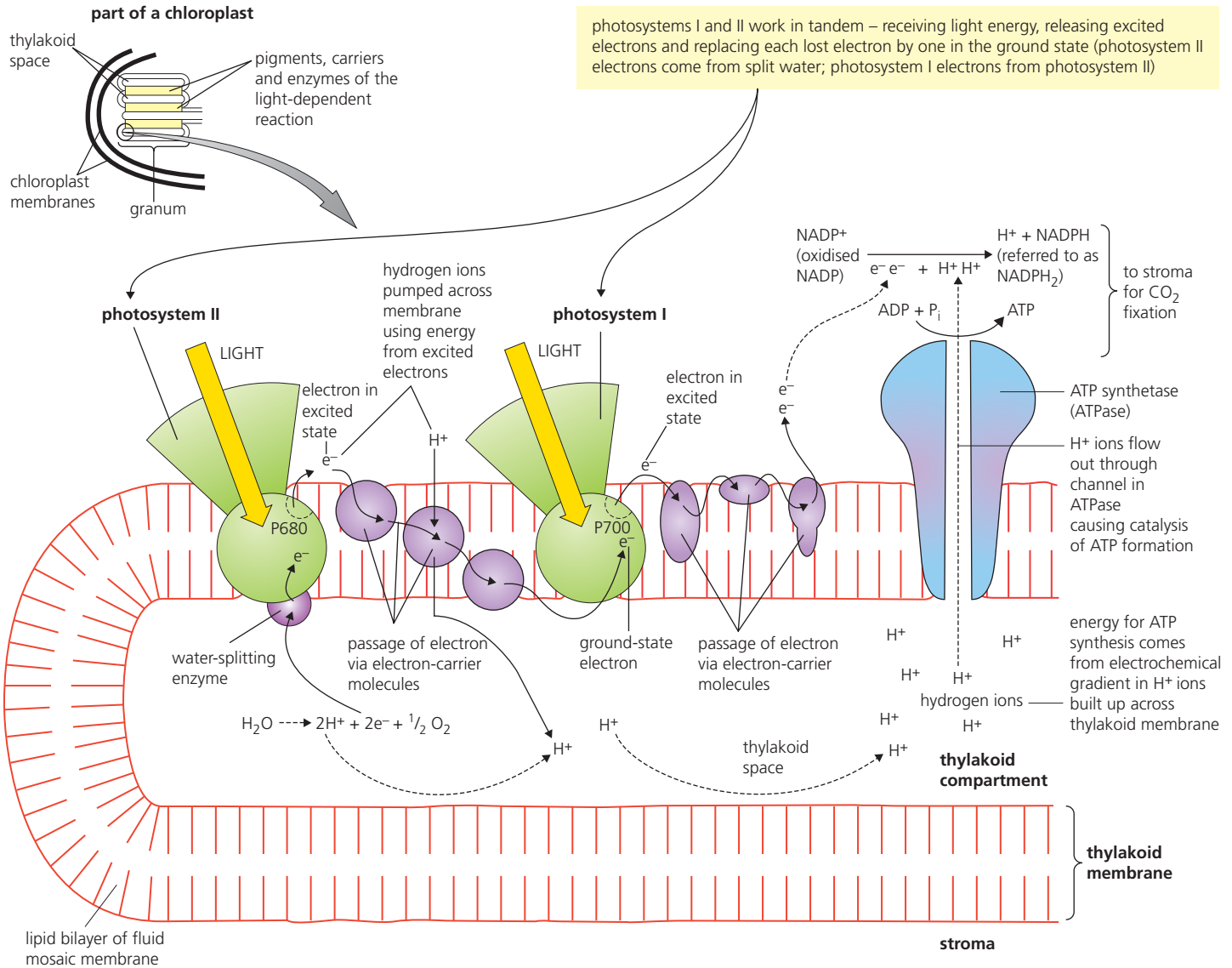


Figure 9.12 The light-dependent reaction

8 In non-cyclic photophosphorylation, **deduce** the ultimate fate of electrons displaced from the reaction centre of photosystem II.

The excited electrons from photosystem I are then picked up by a different electron acceptor. Two at a time, they are passed to $NADP^+$, which, with the addition of hydrogen ions from photolysis, is reduced to form $NADPH + H^+$.

By this sequence of reactions, repeated again and again at very great speed throughout every second of daylight, the products of the light-dependent reactions (**ATP + NADPH + H^+**) are formed (Figure 9.12).

ATP and reduced NADP do not normally accumulate, however, as they are immediately used in the fixation of carbon dioxide in the surrounding stroma (light-independent reactions, page 283). Then the ADP and $NADP^+$ diffuse back into the grana for re-use in the light-dependent reactions.

Photosynthesis in chloroplasts		Cell respiration in mitochondria
thylakoid space in grana	**site of proton (H^+) accumulation**	space between inner and outer membranes of mitochondria
from water molecules after photolysis has occurred	**origin of protons**	from reduced hydrogen acceptors (e.g. $NADH + H^+$)
sunlight	**energy source**	glucose and respiratory intermediates
diffuses to stroma and used to sustain reduction of carbon dioxide in light-independent reactions	**fate of ATP formed**	diffuses into matrix of mitochondria and to cytosol and mainly involved in anabolic reactions of metabolism

Table 9.3 Chemiosmosis in mitochondria and chloroplasts compared

Studying the light-dependent reactions with isolated chloroplasts

Chloroplasts can be isolated from green plant leaves, and suspended in buffer solution of the same concentration as the cytosol (using an isotonic buffer). Suspended in such a buffer, it has been found that the chloroplasts are undamaged, and function much as they do in the intact leaf. So, these isolated chloroplasts can be used to investigate the reactions of photosynthesis – for example, to show they evolve oxygen when illuminated. This occurs provided the natural electron acceptor enzymes and carrier molecules are present.

In the research laboratory, a sensitive piece of apparatus called an **oxygen electrode** is used to detect the oxygen given off by isolated chloroplasts.

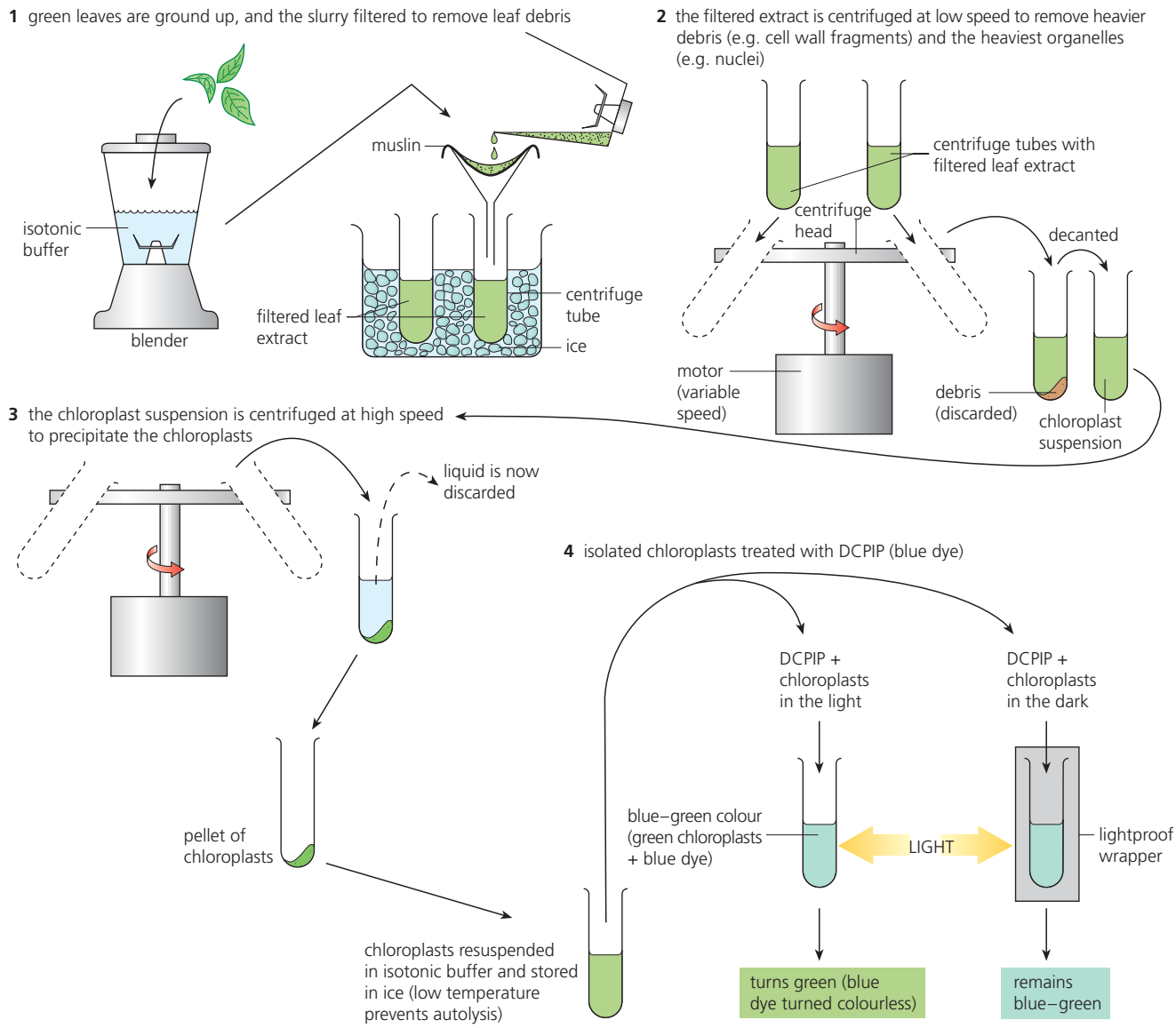
Alternatively, a **hydrogen-acceptor dye** that changes colour when it is reduced can be used. The dye known as DCPIP is an example. DCPIP does no harm when added to chloroplasts in a suitable buffer solution, but changes from blue to colourless when reduced. The splitting of water by light energy (photolysis) is the source of hydrogen that turns DCPIP colourless. The photolysis of water and the reduction of the dye is represented by the equation:

$$2\text{DCPIP} + 2H_2O \rightarrow 2\text{DCPIPH}_2 + O_2$$

In Figure 9.13 the steps of isolation of chloroplasts and the investigation of their reducing activity are shown.

9 In the experiment shown in Figure 9.13, isolated chloroplasts are retained in isotonic buffer, standing in an ice bath. **Explain** the significance of this step.

Figure 9.13 The reducing activity of isolated chloroplasts

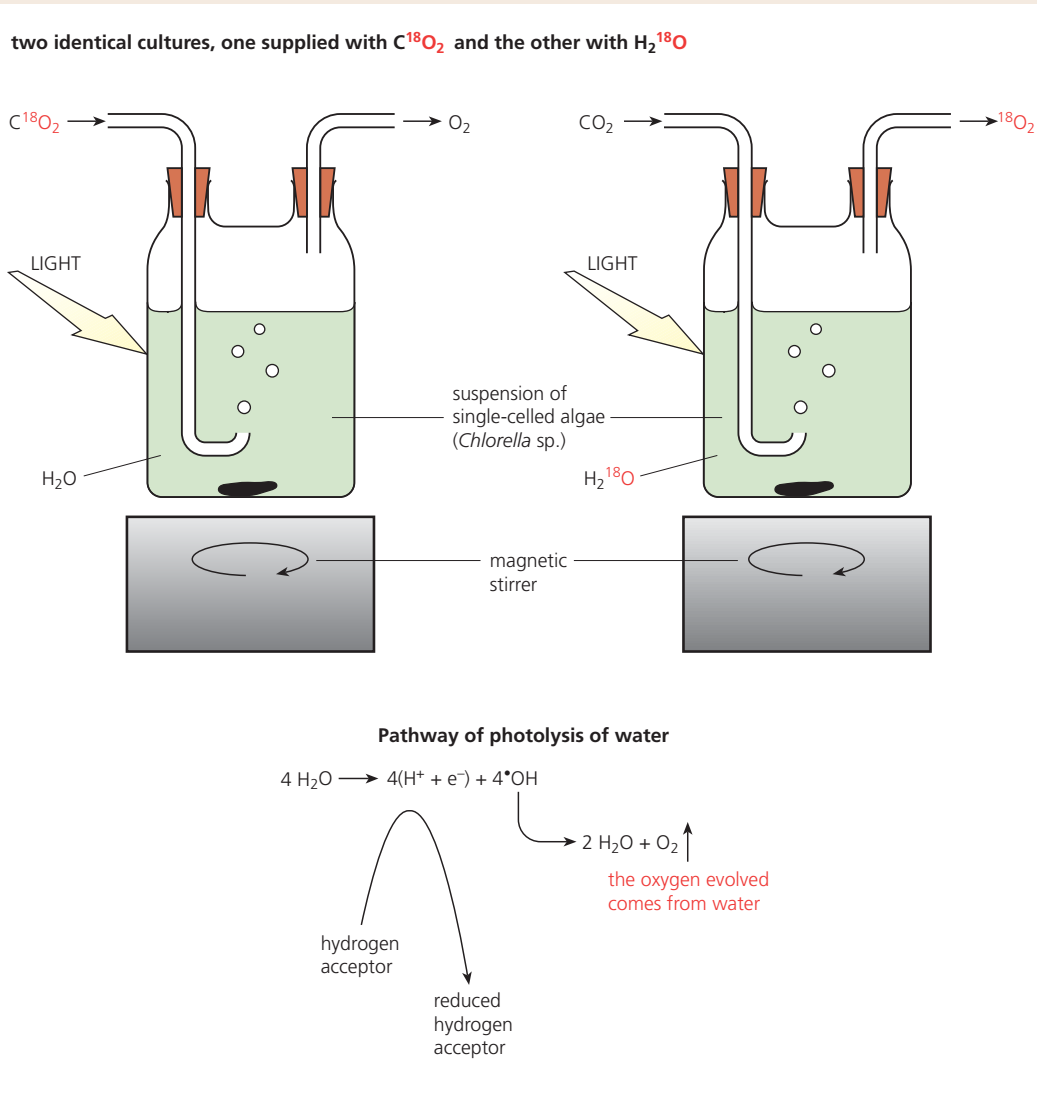


Extension: The light-dependent reactions and the source of oxygen

The explanation of the light-dependent reactions tells us that the oxygen given off in photosynthesis is derived exclusively from water (rather than carbon dioxide).

An experiment using water containing the isotope oxygen-18 confirmed this was correct. Illuminated suspensions of photosynthetic cells were provided with either carbon dioxide enriched with $C^{18}O_2$ or water enriched with $H_2{}^{18}O$. From both, the oxygen evolved was analysed to determine the oxygen-18 concentration, as shown in Figure 9.14.

Figure 9.14 Photolysis of water



Incidentally, this experiment established that the traditional balanced equation for photosynthesis

$$6CO_2 + 6H_2O + \text{LIGHT ENERGY} \rightarrow C_6H_{12}O_6 + 6O_2$$

is incorrect because it implies that some or all of the oxygen evolved comes from carbon dioxide. This is because it shows 12 oxygen atoms are produced, but the six water molecules on the left-hand side of the equation contain only six.

The summary equation for photosynthesis is **less misleading** when written as:

$$6CO_2 + 12H_2O + \text{LIGHT ENERGY} \rightarrow C_6H_{12}O_6 + 6O_2 + 6H_2O$$

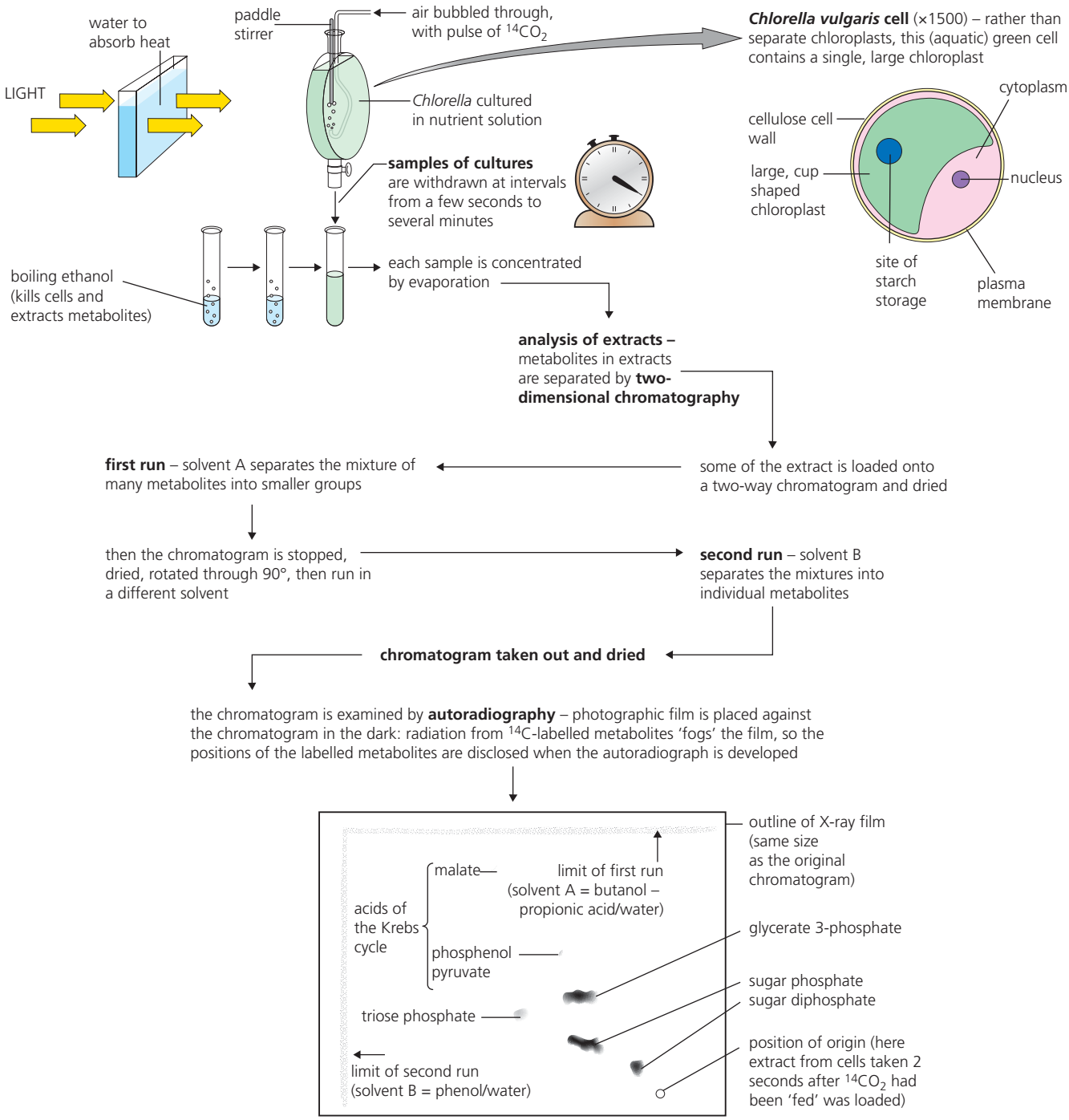
The light-independent reactions

In the light-independent reactions, carbon dioxide is converted to carbohydrate. These reactions occur in the **stroma** of the chloroplasts, surrounding the grana.

How is this brought about?

The pathway by which carbon dioxide is reduced to glucose was investigated by a method we now call **feeding experiments**. In these particular experiments, radioactively labelled carbon dioxide was fed to cells. $^{14}CO_2$ is taken up by the cells in exactly the same way as non-labelled carbon dioxide is, and is then fixed into the same products of photosynthesis.

Figure 9.15 Investigation of the light-independent reactions of photosynthesis



So, a brief pulse of labelled $^{14}CO_2$ was introduced into the otherwise continuous supply of $^{12}CO_2$ to photosynthesising cells in the light, and its progress monitored. Samples of the photosynthesising cells, taken at frequent intervals after the $^{14}CO_2$ had been fed, contained a sequence of radioactively labelled intermediates, and (later) products, of the photosynthetic pathway. These compounds were isolated by **chromatography** from the sampled cells and identified (Figure 9.15).

The experimenters chose for their photosynthesising cells a culture of *Chlorella*, a unicellular alga. They used these in place of mesophyll cells (page 296), since they have identical photosynthesis, but allowed much easier sampling.

The whole technique was pioneered by a team at the University of California, led by Melvin Calvin in the middle of the last century. He was awarded a Nobel Prize in 1961. The chromatography technique that the team exploited was then a relatively recent invention, and radioactive isotopes were only just becoming available for biochemical investigations.

10 Calvin used a suspension of aquatic, unicellular algal cells in his 'lollipop flask' (Figure 9.15) to investigate the fixation of carbon dioxide in photosynthesis. **Suggest** what advantages were obtained by this choice, compared to the use of cells of intact green leaves.

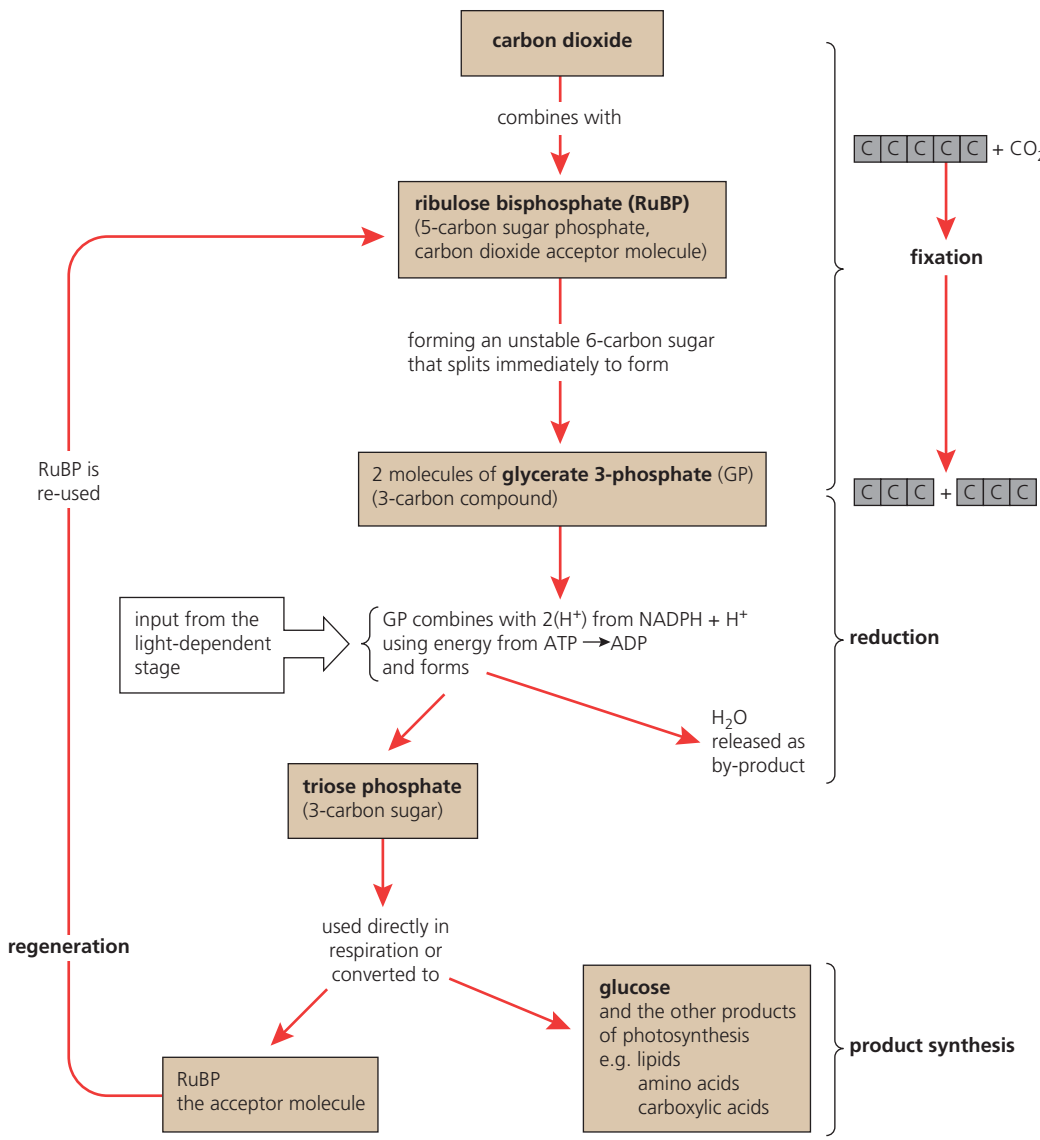


Figure 9.16 The path of carbon in photosynthesis

The steps of the light-independent reactions

The experiments of Calvin's team established the details of the path of **carbon from carbon dioxide to glucose** and other products (Figures 9.16 to 9.19). They showed that:

- The first product of the fixation of carbon dioxide is **glycerate 3-phosphate (GP)**. This is known as the **fixation step**.
- This initial product is immediately reduced to the 3-carbon sugar phosphate, **triose phosphate**, using **NADPH + H^+** and **ATP**. This is the **reduction step**.
- Then the triose phosphate is further metabolised to produce **carbohydrates** such as **sugars**, **sugar phosphates** and **starch**, and later **lipids**, amino acids such as **alanine**, and organic acids such as **malate**. This is the **product synthesis step**.
- Some of the triose phosphate is metabolised to produce the molecule that first reacts with carbon dioxide (the acceptor molecule). This is the **regeneration of acceptor step**. The reactions of this regeneration process are today known as the **Calvin cycle** (Figure 9.18).

This done, a big problem remained.

Which intermediate is the actual acceptor molecule?

At first, a 2-carbon acceptor molecule for carbon dioxide was sought, simply because the first product of carbon dioxide fixation was known to be a 3-carbon compound. None was found.

Eventually the acceptor molecule proved to be a 5-carbon acceptor (**ribulose bisphosphate**). When carbon dioxide has combined, the 6-carbon product immediately split into two 3-carbon GP molecules – hence the initial confusion.

The enzyme involved is called **ribulose bisphosphate carboxylase** (commonly shortened to **Rubisco**). Rubisco is by far the most common protein of green plant leaves, as you would expect.

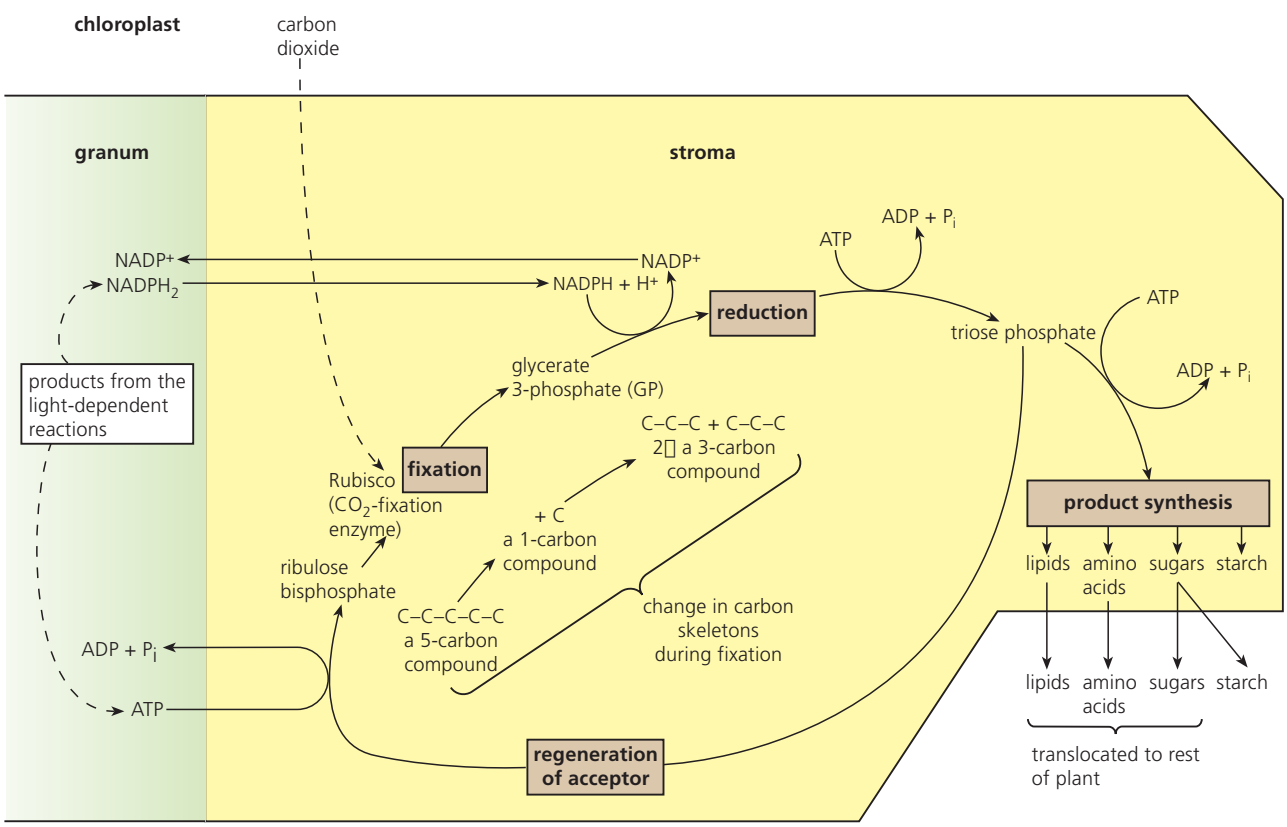


Figure 9.17 Summary of the light-independent reactions

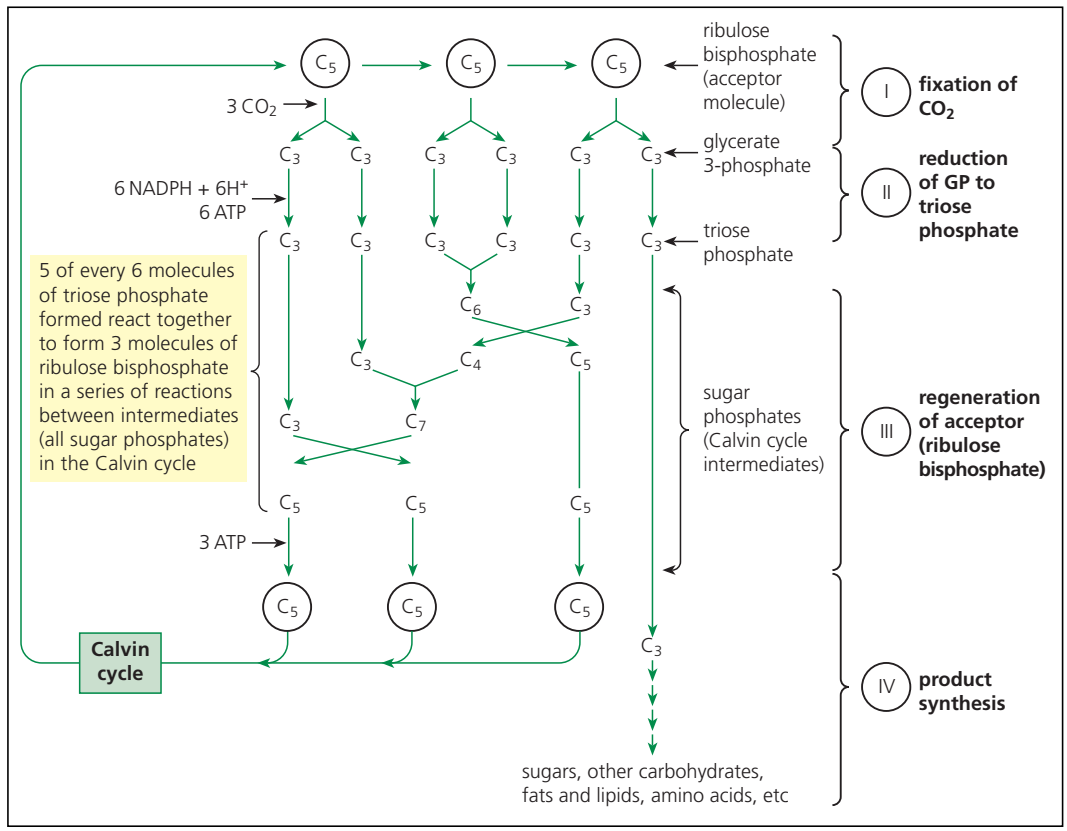


Figure 9.18 Regeneration of ribulose bisphosphate – Calvin cycle

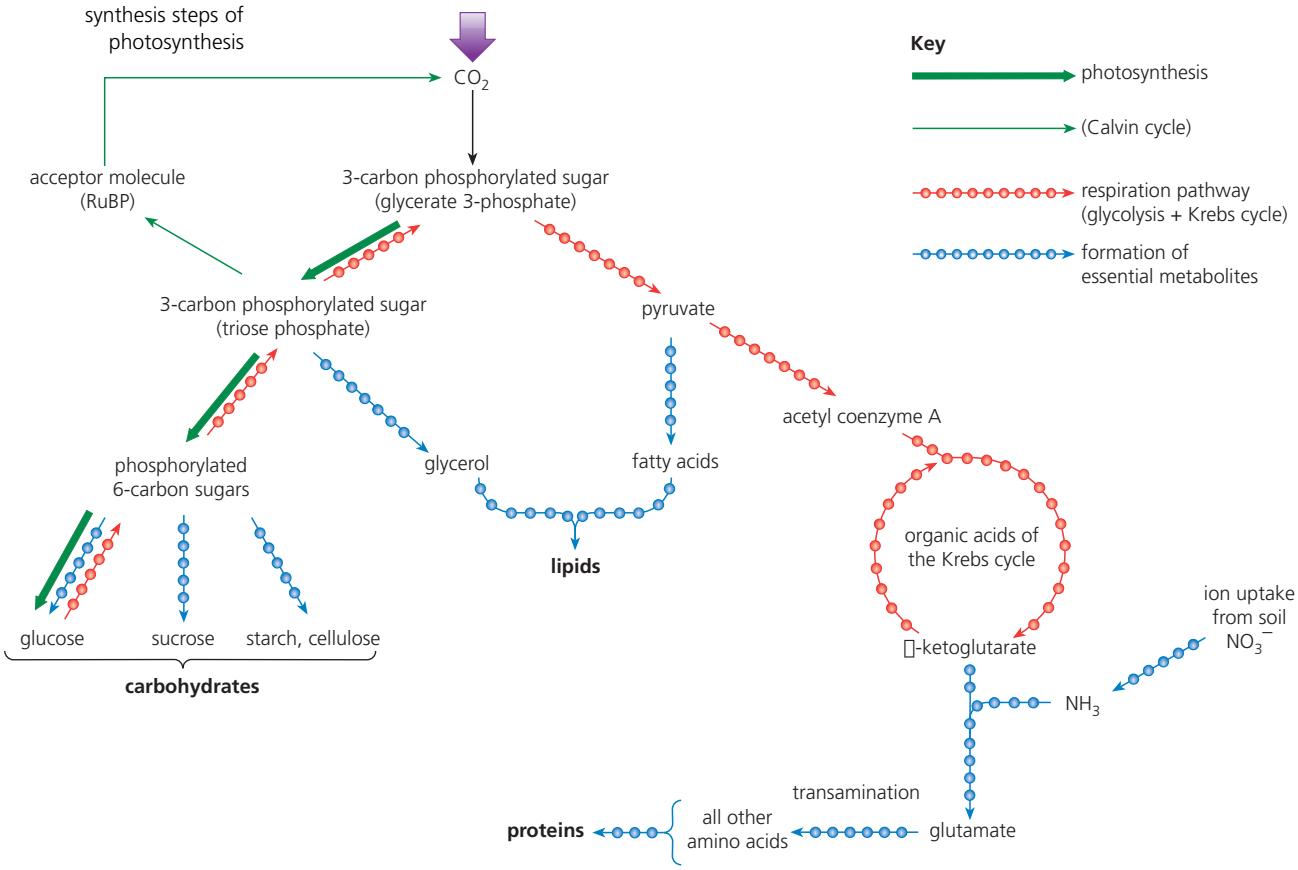


Figure 9.19 The product synthesis steps of photosynthesis

Chloroplast, grana and stroma – the venue for photosynthesis

We have seen that the grana are the site of the light-dependent reactions and the stroma the site of the light-independent reactions of photosynthesis. Table 9.4 identifies how the structure of the chloroplast facilitates function.

Structure of chloroplast	Function / role
double membrane bounding the chloroplast	contains the grana and stroma, and is permeable to CO_2, O_2, ATP, sugars and other products of photosynthesis
photosystems with chlorophyll pigments arranged on thylakoid membranes of grana	provide huge surface area for maximum light absorption
thylakoid spaces within grana	restricted regions for accumulation of protons and establishment of the gradient
fluid stroma with loosely arranged thylakoid membranes	site of all the enzymes of fixation, reduction and regeneration of acceptor steps of light-independent reactions, and many enzymes of the product synthesis steps

Table 9.4 Structure and function in chloroplasts

11 **Distinguish** between the following:

a light-dependent reactions and light-independent reactions
b photolysis and photophosphorylation.

12 **Deduce** the significant difference between the starting materials and the end-products of photosynthesis.

Light and photosynthesis

8.2.7–8.2.8

Visible light represents a small part of the spectrum of electromagnetic radiation reaching the Earth from the Sun (Figure 3.8, page 83). We have seen that the photosynthetic pigments that constitute chlorophyll absorb the various wavelengths of visible light to varying extents. The technique of chromatography has shown that chlorophyll present in green plants is a mixture of four pigments: chlorophyll α and β, and two carotenoids, carotene and xanthophylls. A **colorimeter** is the apparatus we use to find out the absorption spectrum of chlorophyll once this has been extracted from leaves (Figure 3.10, page 85).

The action spectrum for photosynthesis

Clearly, chlorophyll absorbs most strongly in the blue and red parts of the spectrum. But which wavelengths present in white light are most effective in bringing about the light-dependent reactions of photosynthesis?

The **action spectrum** for photosynthesis may be obtained most simply by using a photosynthometer, as illustrated in Figure 3.14, page 89. This apparatus allows estimation of the efficiency of photosynthesis in an aquatic plant such as *Elodea*. Here, the volume of oxygen produced in unit time is measured under particular conditions – for example, at different wavelengths. A light source that projects different wavelengths in turn is required. The amount of oxygen produced consistently at each wavelength is measured, and the table of results indicates the more effective wavelengths.

13 **Discuss** the likely limitations of using a simple photosynthometer (Figure 3.14, page 89) to obtain an action spectrum of photosynthesis.

More accurate data may be obtained when oxygen evolution from a suspension of algal cells or from isolated chloroplasts is measured, using an **oxygen electrode**. The oxygen electrode is an especially accurate instrument for measuring the concentration (partial pressure) of oxygen present in solution. The suspension is constantly stirred, and so changes in oxygen production – for example, as light intensity or carbon dioxide concentration is varied – are detected.

The **absorption spectrum** for chlorophyll pigments and the action spectrum for photosynthesis show that the wavelengths of light absorbed by chlorophyll pigments, namely the red and blue light, are very similar to the wavelengths that cause photosynthesis. The absorption and action spectra match quite well. So the wavelengths optimally absorbed are the ones that provide most energy for photosynthesis; both blue and red light are used by green plants as the energy source for photosynthesis (Figure 9.20).

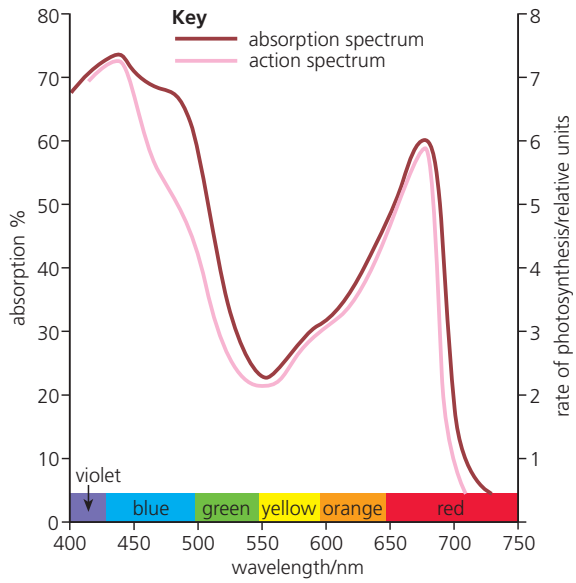


Figure 9.20 Absorption and action spectra compared

14 Distinguish between the following:

a chlorophyll and chloroplast
b stroma and grana
c absorption spectrum and action spectrum.

Incidentally there is one part of the two curves where they diverge somewhat – in the upper wavelengths of blue light. Light absorption at this point is largely by particular photosynthetic pigments (carotenes) and is not fully used in photosynthesis. Carotenes are ancillary pigments – meaning they are pigments that do not directly contribute light energy to a reaction centre, but rather pass the energy to the other photosynthetic pigments first. This energy transfer between pigments is clearly not 100% efficient.

Light as a limiting factor in photosynthesis

By definition, light is an essential for photosynthesis. Green plant cells in the dark are unable to photosynthesise at all; under this condition, cells of the plant take in oxygen from the atmosphere for aerobic cell respiration.

As light starts to reach green cells, at dawn for example, photosynthesis starts, and some oxygen is produced. Oxygen consumption in respiration continues, of course. Eventually the light intensity increases to the point where photosynthetic oxygen production is equal to oxygen consumption in respiration. Now the leaf is neither an oxygen importer nor exporter.

This point is known as the **compensation point** (Figure 9.21). Then, as the Sun rises higher in the sky and the light intensity increases further, the rate of photosynthesis also increases; the leaf becomes a net exporter of oxygen.

We have already noted the general effect of increasing light intensity on the rate of photosynthesis (Figure 3.16, page 90).

At low light intensities the rate of photosynthesis increases linearly with increasing light. Here the low light intensity (lack of sufficient light) is limiting the rate of photosynthesis. Light is a factor limiting the rate of photosynthesis under these conditions.

However, as light intensity is raised further, a point is reached where increasing light intensity produces no further effect on the rate of photosynthesis. Here, some factor other than light is limiting photosynthesis. What may now be rate-limiting?

The limiting factor at this point is disclosed when the investigation is repeated at higher carbon dioxide concentration (Figure 9.22). *Look at this graph now; follow the annotations.*

Figure 9.21 Light intensity and the compensation point

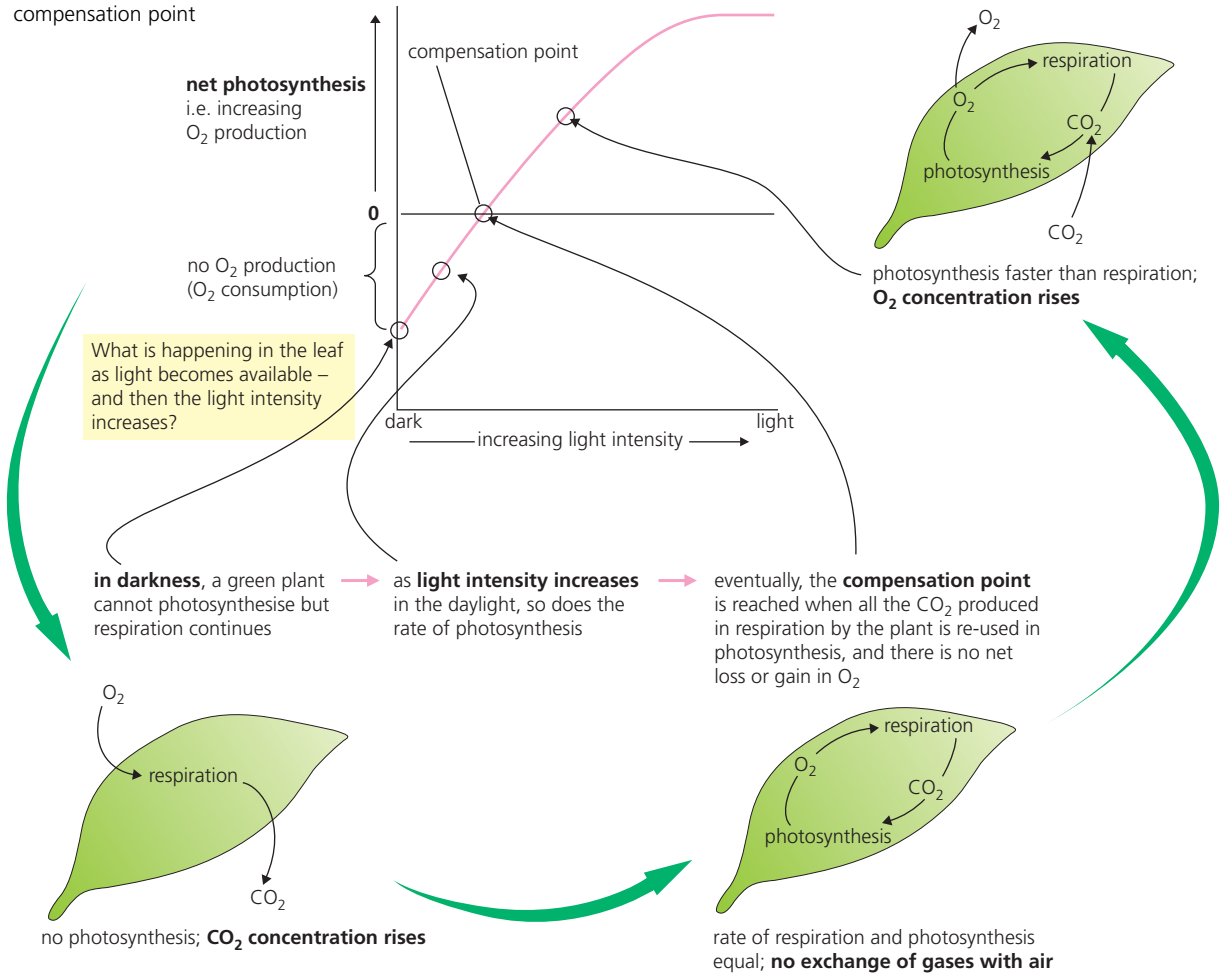
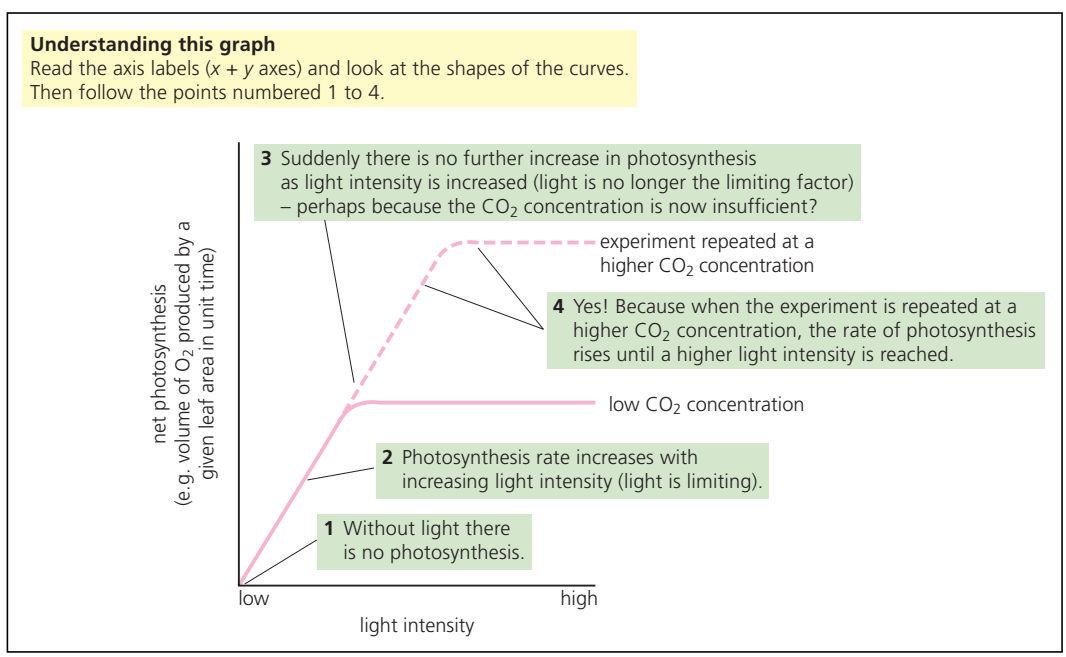


Figure 9.22 The effect of light intensity on the rate of photosynthesis



The results of these earlier, physiological investigations fit well with our current understanding. Photosynthesis is a biochemical process involving a series of interconnected reactions. All these reactions contribute to the overall rate; photosynthesis depends on several essential conditions being favourable. At any one time, the rate of photosynthesis will be limited by the slowest of these reactions – the overall rate will be limited by the factor that is in shortest supply. This factor, which ever one it is, is known as the **limiting factor**. Clearly, a limited supply of either light (Figure 9.22) or carbon dioxide (Figure 9.23) could limit the rate of photosynthesis, since both of these are essential. Clearly, both light intensity and carbon dioxide concentration can be limiting factors in photosynthesis.

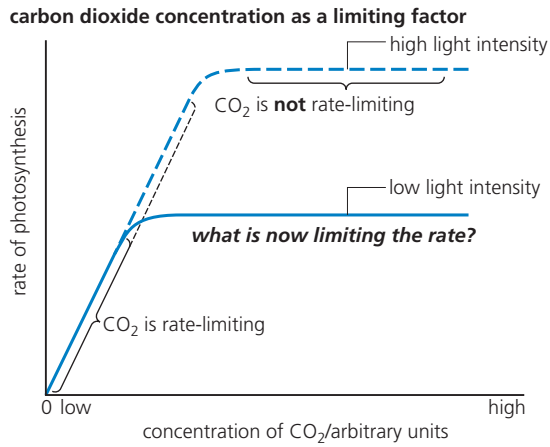


Figure 9.23 The effect of carbon dioxide concentration on photosynthesis

Carbon dioxide-limiting conditions and cyclic photophosphorylation

When the carbon dioxide concentration becomes the limiting factor, we know that the light-independent reactions may be slowed relative to the light-dependent reactions.

An outcome is that NADPH accumulates in the stroma and the concentration of $NADP^+$ reduces. Then, without an adequate supply of oxidised H-acceptor ($NADP^+$), photosystems II and I are unable to operate together, as they do in non-cyclic photophosphorylation (Figure 9.12, page 280). Now it has been found that another type of ATP formation occurs under these conditions, called **cyclic photophosphorylation** (Figure 9.24).

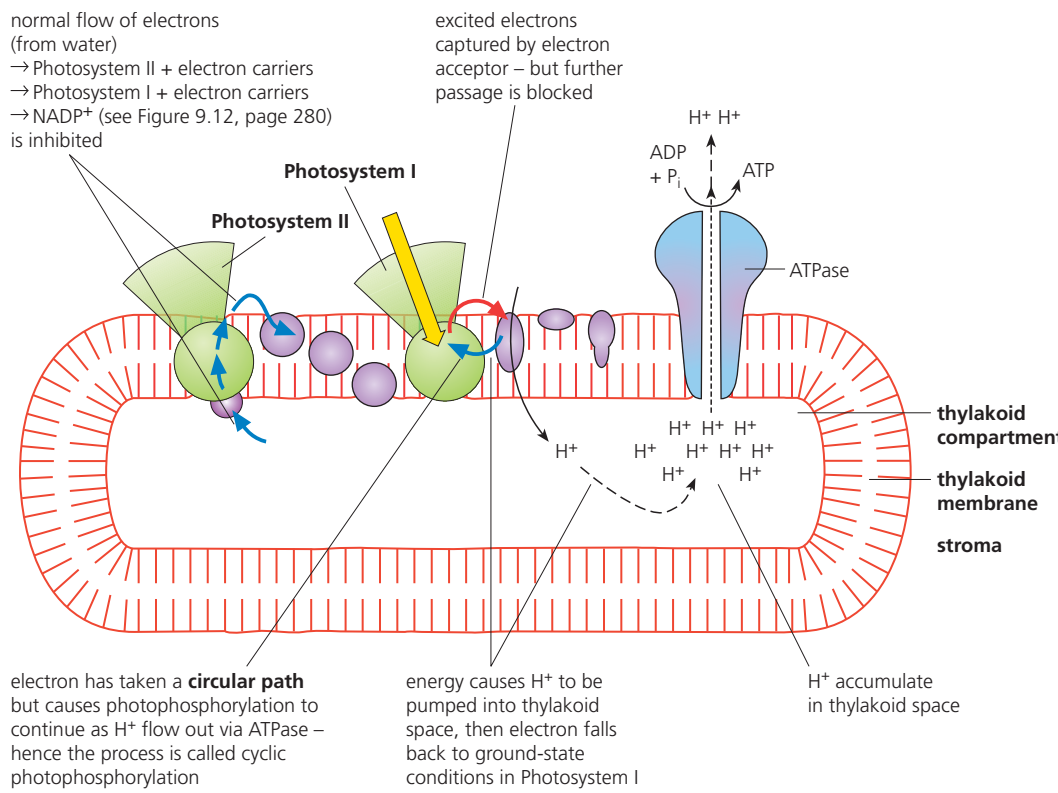


Figure 9.24 Cyclic photophosphorylation

Cyclic photophosphorylation occurs when the CO_2 concentration is the limiting factor (and with relatively high light intensity), so that the light-independent reactions are slowed relative to the light-dependent reactions. **NADPH accumulates in the stroma.**

In cyclic photophosphorylation, excited electrons from the reaction centre in photosystem I fall back to their point of origin. They do this via electron carriers that transfer energy from those electrons and pump H^+ into the thylakoid space. ATP synthesis by chemiosmosis follows. The Z-diagram illustrates this (Figure 9.25).

This type of photophosphorylation is called cyclic because the electrons have returned to the photosystem from which they originated (a cyclic path, rather than a linear one).

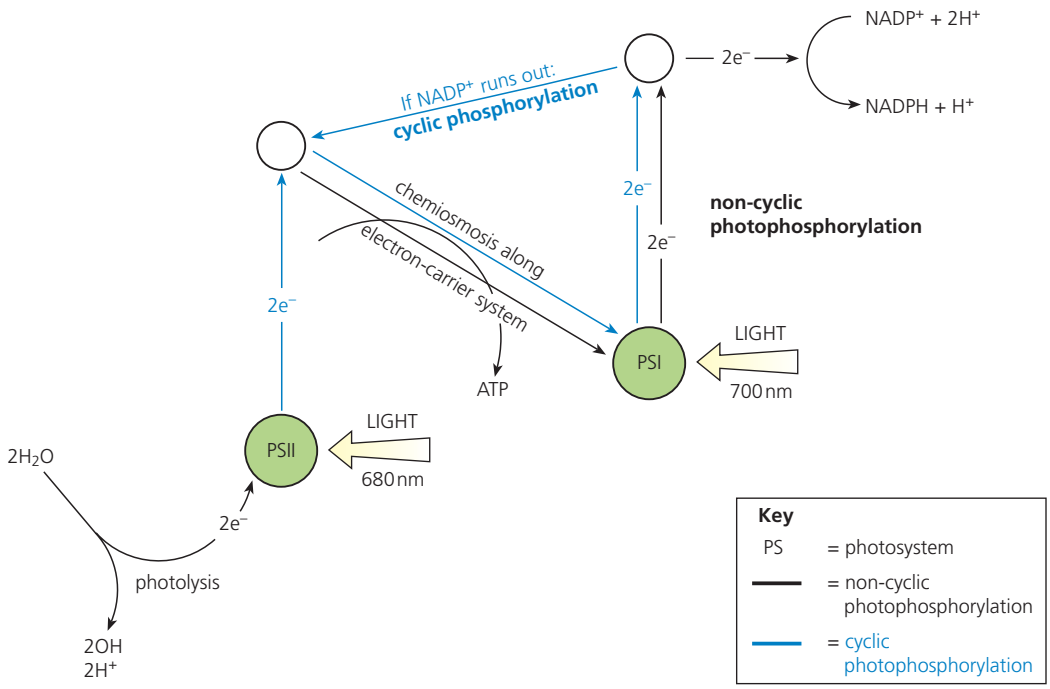


Figure 9.25 A comparison of cyclic- and non-cyclic photophosphorylation (the so-called Z diagram)

15 In cyclic photophosphorylation, **predict** which of the following occur:

photoactivation of photosystem I
reduction of $NADP^+$
production of ATP.

The effect of changing temperature on the rate of photosynthesis

The effect of temperature on the rate of photosynthesis, under controlled laboratory conditions, is shown in Figure 9.26. It can be seen that the effect of increasing temperature, say from about 15 °C to 25 °C, actually depends on the light intensity.

Under low light intensities, a rise in temperature has less effect than it does under higher intensities. Of course, at very much higher temperatures, the rate of all metabolic reactions falls away since all cell enzymes are denatured by excess heat energy (page 55).

The graph in Figure 9.26 has been interpreted as showing that photosynthesis is made up of the two sequential sets of reactions we have been studying: light-dependent reactions (which, like all photochemical events, are largely temperature indifferent) and light-independent reactions (which, like all biochemical steps catalysed by enzymes, are temperature sensitive).

Can you see why?

The arguments are as follows:

- If photosynthesis includes **photochemical reactions** – temperature-insensitive changes brought about by light energy – then under conditions of low light intensity (when these photochemical reactions are rate limiting), a rise in temperature should have little effect. This is what happens (Figure 9.26, curve B).
- If photosynthesis also includes **enzymic reactions** – temperature-sensitive changes involving enzymes – then under conditions of high light intensity (when enzymic reactions are rate limiting), a rise in temperature should have a significant effect. It should increase the rate of photosynthesis. This, too, is what happens (Figure 9.26, curve A).

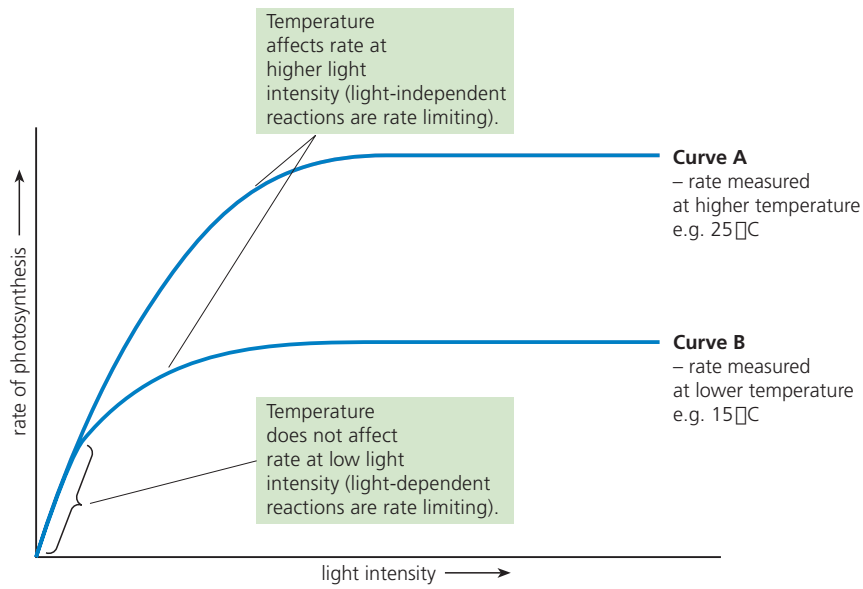


Figure 9.26 The effect of temperature on the rate of photosynthesis

This reasoning provided the first evidence that photosynthesis is made up of two sequential sets of reactions:

- **light-dependent reactions** – photochemical steps largely unaffected by temperature;
- **light-independent reactions** – biochemical reactions catalysed by enzymes and therefore temperature sensitive. Here a 10 °C rise in temperature has the effect of doubling the rate of reaction.

In fact, this understanding of the nature of photosynthesis was deduced from the above experimental evidence before it was confirmed by biochemical techniques, using chloroplasts isolated from the leaf.

Examination questions – a selection

Questions 1–5 are taken from past IB Diploma biology papers.

Q1 What is acetyl (ethanoyl) CoA?

I an intermediate in carbohydrate metabolism under aerobic conditions
II a product of the oxidation of fatty acids in lipid metabolism
III an intermediate in carbohydrate metabolism under anaerobic conditions

A I and II only
B I and III only
C II and III only
D I, II and III

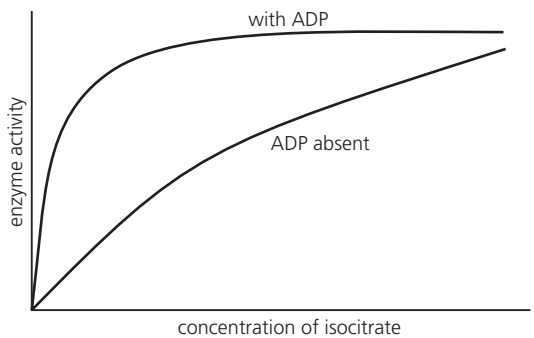
Higher Level Paper 1, May 03, Q15

Q2 Which of the following is produced during glycolysis?

A NADH
B CO_2
C glucose
D glycogen

Higher Level Paper 1, May 05, Q26

Q3 Isocitrate dehydrogenase is an enzyme of the Krebs cycle. Its activity in the presence and absence of ADP is shown below.



What effect will a high level of energy consumption have on the activity of this enzyme?

A The activity will increase.
B The activity will decrease.
C The enzyme will maintain a constant activity.
D The activity will fluctuate up and down.

Higher Level Paper 1, May 04, Q27

Q4 At which stage of photosynthesis is light involved most directly?

A reduction of $NADP^+$ to $NADPH_2$
B chemiosmosis
C the synthesis of chlorophyll
D the photoactivation of chlorophyll

Higher Level Paper 1, May 06, Q29

Q5 Which of the following statements about pyruvate is true?

A It contains less energy than glucose per molecule.
B Every molecule of glucose is converted to one molecule of pyruvate.
C Pyruvate is produced in the mitochondria.
D Under aerobic conditions, pyruvate is converted to lactate.

Higher Level Paper 1, May 06, Q28

Questions 6–10 cover other syllabus issues in this chapter.

Q6 Although oxidation may be illustrated by reference to the addition of oxygen to a substance, oxidation is defined precisely as 'loss of electrons'. Explain this definition fully. (4)

Q7 Chemiosmosis is a process by which the synthesis of ATP is coupled to electron transport via the movement of protons. Draw and annotate a diagram to show how the structure of the membranes of a mitochondrion (given the position of electron carriers and specific enzymes within it) allows a large proton gradient to be formed, and ATP synthesis to follow. (8)

Q8 Draw a diagram of a chloroplast to show its structure as disclosed by electron micrography. Annotate this drawing to link function to structures where possible. (6)

Q9 **a** Explain as concisely as you are able the distinctive nature of non-cyclic and cyclic photophosphorylation. (6)
b Describe the specific circumstances which cause cyclic photophosphorylation to occur. (2)

Q10 **a** Explain the concept of limiting factors in photosynthesis, with reference to light intensity and the concentration of carbon dioxide. (4)
b Draw a fully annotated graph of net photosynthesis rate (*y* axis) against light intensity (*x* axis) at low and high concentrations of carbon dioxide. (4)

10 Plant science

STARTING POINTS

- Green plants (**Plantae**) are **autotrophic** (self-feeding) organisms, manufacturing their own nutrients from simple, inorganic molecules.
- Green plants show wide diversity, including the mosses (**bryophytes**), the ferns (**filicinophytes**), the conifers (**coniferophytes**), and the flowering plants (**angiospermophytes**).
- **Photosynthesis** is the process by which green plants manufacture **carbohydrates** from carbon dioxide and water, using **energy from sunlight**; oxygen is the waste product. Photosynthesis occurs in specialised organelles called **chloroplasts**.
- Photosynthesis can be divided into two linked steps – the **light-dependent reactions** in the **grana** produce NADPH + H^+ and ATP, and the **light-independent reactions** in the **stroma** fix carbon dioxide to carbohydrate.
- Energy is transferred in cells by **aerobic cellular respiration**. **ATP** is the **universal energy currency** of cells. ATP is a reactant in energy-requiring reactions (e.g. protein synthesis) and processes (e.g. active transport mechanisms).
- Cells of organisms are surrounded by an aqueous solution, and **exchange of molecules** between cells and their environment occurs in solution across the plasma membrane (of **fluid mosaic** construction), either by **diffusion** (osmosis is a special case), or **active transport** (by protein pumps in the membrane), or by **bulk transport** (via the movements of vesicles, for example).
- In **sexual reproduction**, two sex cells (**gametes**) fuse (**fertilisation**) to form a **zygote**. The zygote develops into a new organism, similar but not identical to its parents. Gamete formation involves **meiosis** (a reduction nuclear division).

The green plants (**Plantae**) make up one of the five kingdoms of living things (Figure 6.43, page 176). The features that make the green plants distinctly different from other organisms are as follows.

- There is a wall around each cell, the chief component of which is **cellulose**. Cellulose is a polysaccharide, and is an extremely tough, protective material.
- Cell organelles called **chloroplasts** are the site of photosynthesis. By photosynthesis, the plant generates energy-rich nutrients from simple, inorganic substances (carbon dioxide, water and inorganic ions, such as nitrates and phosphates).

In the long history of life, green plants evolved about 500 million years ago, from aquatic, single-celled organisms called green algae (possibly very similar to *Chlorella*, Figure 9.15, page 283). The present-day diversity of green plants was introduced in Chapter 6, pages 137–77. Today, it is the angiospermophytes (flowering plants) that are the dominant terrestrial plants. The fossil record indicates they achieved this dominance early in their evolutionary history – about 100 million years ago.

In this chapter, **flowering plant structure** and **growth**, **internal transport mechanisms**, and **reproduction** are discussed.

Plant structure and growth

9.1.1–9.1.7

The flowering plants or angiospermophytes dominate plant life in almost every habitat across the world. Some are trees and shrubs with woody stems, but many are non-woody (herbaceous) plants. Whether woody or herbaceous, the plant consists of stem, leaves and root.

The **stem** supports the leaves in the sunlight, and transports organic materials (such as sugar and amino acids), ions and water between the roots and leaves. At the top of the stem is a **terminal bud** or terminal growing point, and in the axil of each leaf is an **axillary bud**. New cells are produced at these growing points.

A **leaf** consists of a leaf blade connected to the stem by a leaf stalk. The leaf is an organ specialised for photosynthesis.

The **root** anchors the plant and is the site of absorption of water and ions from the soil.

The structure of the sunflower, *Helianthus annus*, a herbaceous plant, illustrates these features (Figure 10.1).

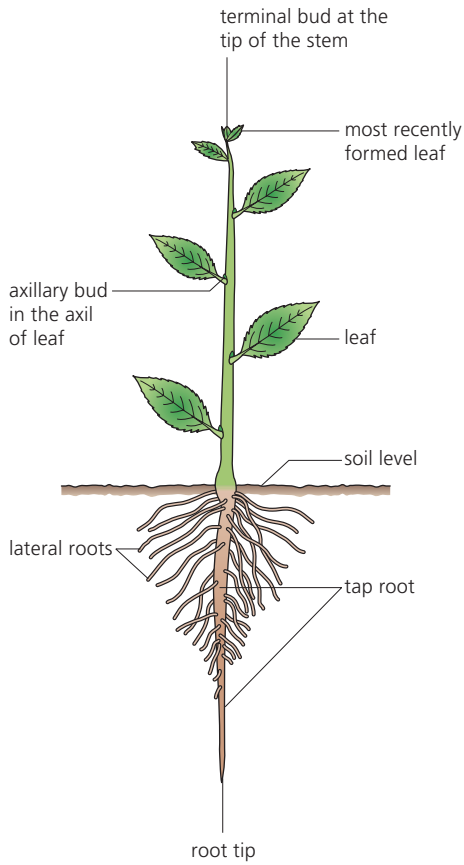


Figure 10.1 The structure of the sunflower plant, *Helianthus annus*

The distribution of tissues in the stem of the sunflower

A **tissue map** (sometimes called a low-power diagram) is a drawing that records the relative positions of structures within an organ or organism, as seen in section; it does not show individual cells.

From the tissue map in Figure 10.2, it can be seen that the stem is an organ surrounded or contained by a layer called the **epidermis**, and that it contains **vascular** tissue (**xylem** for water transport and **phloem** for transport of organic solutes) in a discrete system of veins or **vascular bundles**. In the stem, the vascular bundles are arranged in a ring, positioned towards the outside of the stem, rather like the steel girders of a ferro-concrete building.

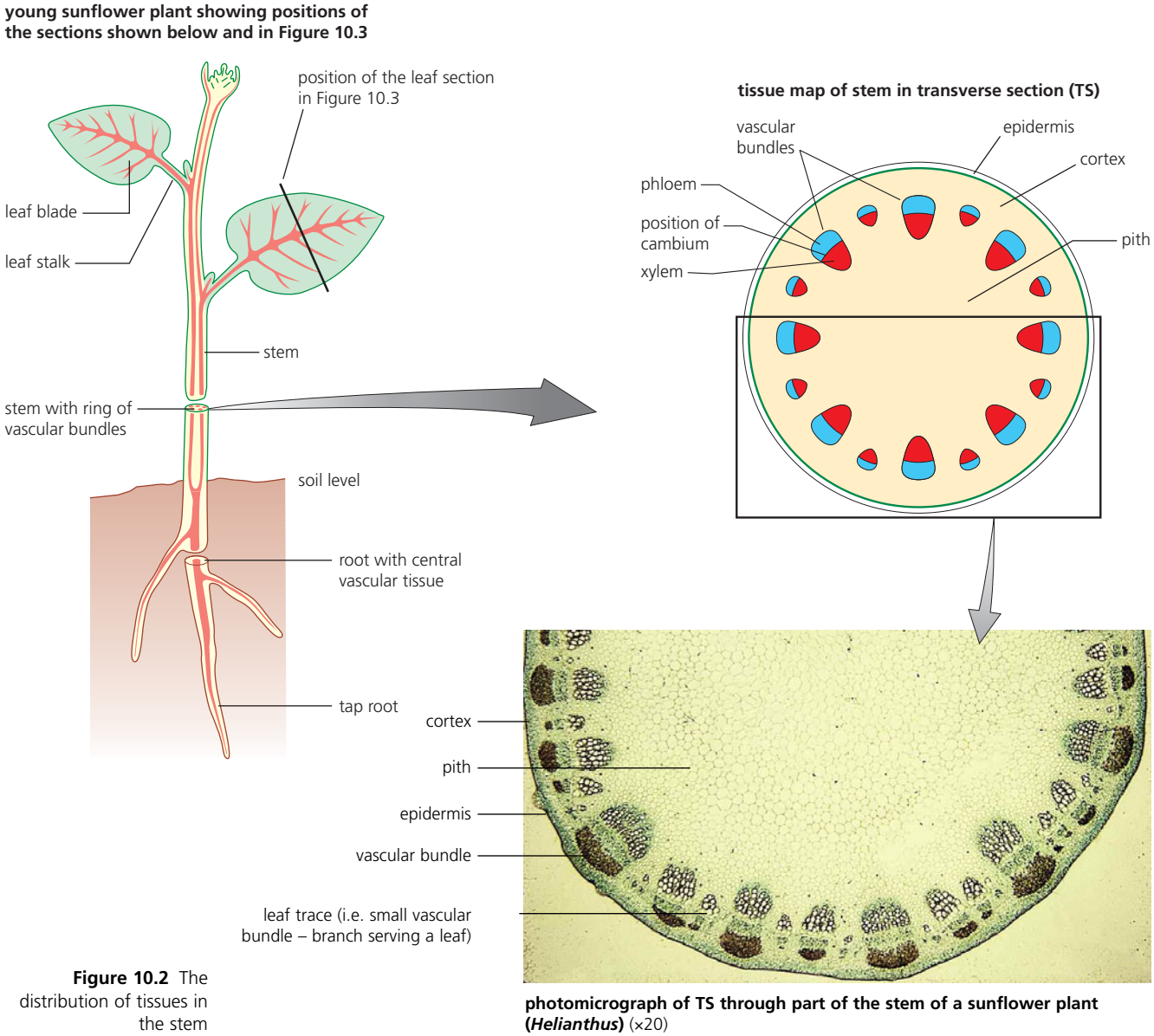


Figure 10.2 The distribution of tissues in the stem

The distribution of tissues in the leaf

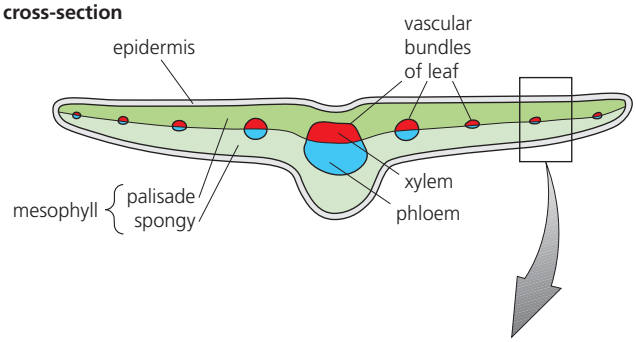
A tissue map showing the distribution of tissues in a leaf is shown in Figure 10.3. Like the stem, the leaf is contained by a single layer of cells, the **epidermis**, and also contains vascular tissue in a system of vascular bundles. The vascular bundles in leaves are often referred to as **veins**. The bulk of the leaf is taken up by a tissue called **mesophyll**, and the cells here are supported by veins arranged in a branching network.

Dicotyledonous and monocotyledonous plants

The flowering plants are divided into two groups with distinguishing structural features, and known as the **monocotyledons** and **dicotyledons**. The sunflower is a typical dicotyledonous plant (Figure 10.1) and the grasses are a major family of monocotyledonous plants (Figure 10.4). The defining difference between members of the two groups is the number of seed leaves (**cotyledons**) present in the embryo. (Cotyledons are structurally simpler leaves than the normal leaves of these plants that develop later.) The important common differences are identified in Table 10.1.

Figure 10.3 The distribution of tissues in a leaf

tissue map of leaf in cross-section



photomicrograph of part of a leaf in cross-section (×180)

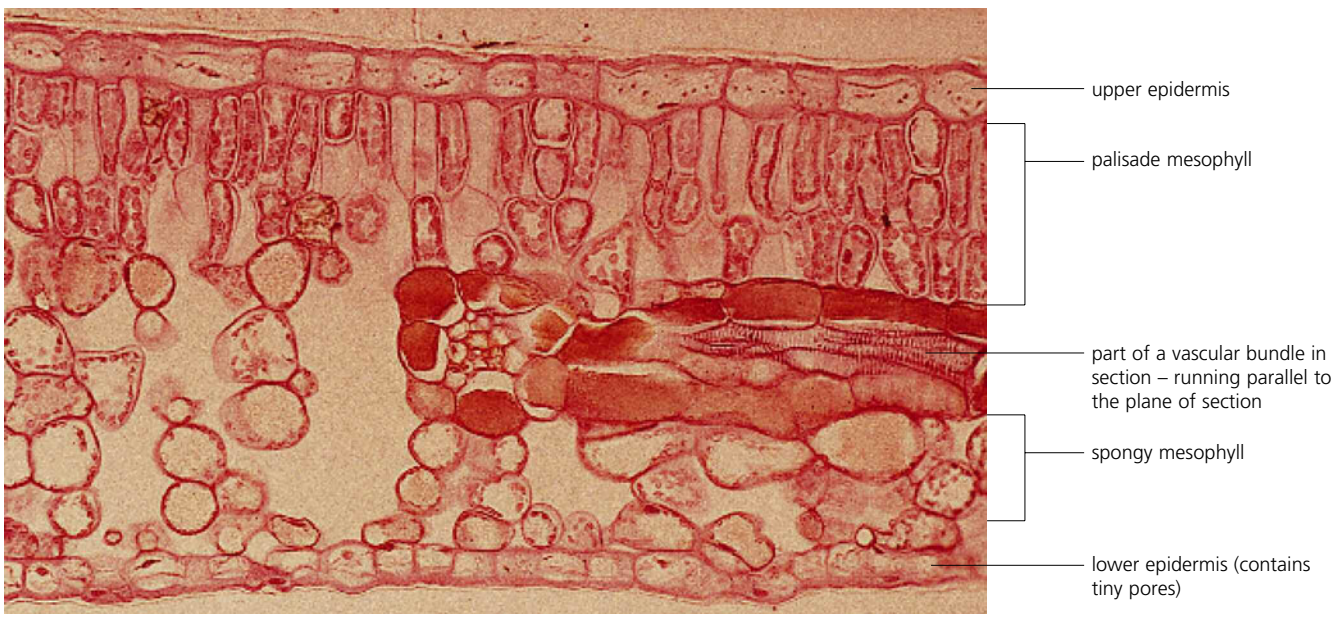


Figure 10.4 A plant of annual meadow grass (*Poa* sp.) showing monocotyledonous features

1 **Draw** and **label** a dicotyledonous plant and a monocotyledonous plant growing near your school or home, identifying their diagnostic features that can be observed with a simple microscope (hand lens).

Dicotyledons (e.g. sunflower)		Monocotyledons (e.g. meadow grass)
embryo in seed has two cotyledons	**cotyledons**	embryo in seed has one cotyledon
broad leaves with veins forming a network	**veins in leaves**	bayonet or strap-shaped leaves with parallel veins
vascular bundles of stem in a ring	**vascular bundles in stem**	vascular bundles of stem numerous and scattered
branched roots	**root growth**	unbranched roots
parts of the flowers (sepals, petals, etc.) in fours or fives	**flower parts (floral organs)**	parts of the flowers (sepals, petals, etc.) in threes.

Table 10.1 Monocotyledons and dicotyledons – the differences

Leaf structure and function

Leaves are the organs specialised for **photosynthesis**, in which light energy is used to build sugar from carbon dioxide obtained from the air, and water obtained from the soil.

The leaf is a thin structure with a large surface area in which **mesophyll cells** are spread out over a wide area, so maximising the amount of light absorbed. The mesophyll cells in the upper layer are called palisade cells. They are packed with **chloroplasts** – the organelles in which photosynthesis occurs.

The **epidermis** is a tough, transparent layer with an external waxy **cuticle**. The cuticle is impervious and effectively reduces water vapour loss from the leaf surface. (As a leaf warms in the light it is inevitably vulnerable to loss of water vapour by evaporation.) The epidermis has many **stomata** (singular, **stoma**), mainly in the lower surface. These are tiny pores that occur between specialised epidermal cells called **guard cells**. They are the sites of inward diffusion of carbon dioxide. Within the leaf's mesophyll tissue are continuous **air spaces** that enhance diffusion of carbon dioxide to the cells.

Between the mesophyll cells lies a complete **network of vascular bundles** which is so extensive that no mesophyll cell is more than a few cells away from a bundle. This is important because it is **xylem** in the vascular bundles that delivers water from the root system, and **phloem** that transports sugars, formed in the light, to sites of use and storage in the rest of the plant.

Another role of this network is mechanical **support**. Mesophyll tissue (rarely more than 10–15 cells thick) is supported by the turgidity of all its cells, contained within the non-elastic epidermis, and reinforced by the network of bundles. Leaves are relatively delicate structures, yet have to withstand destructive forces such as wind and rain, at times.

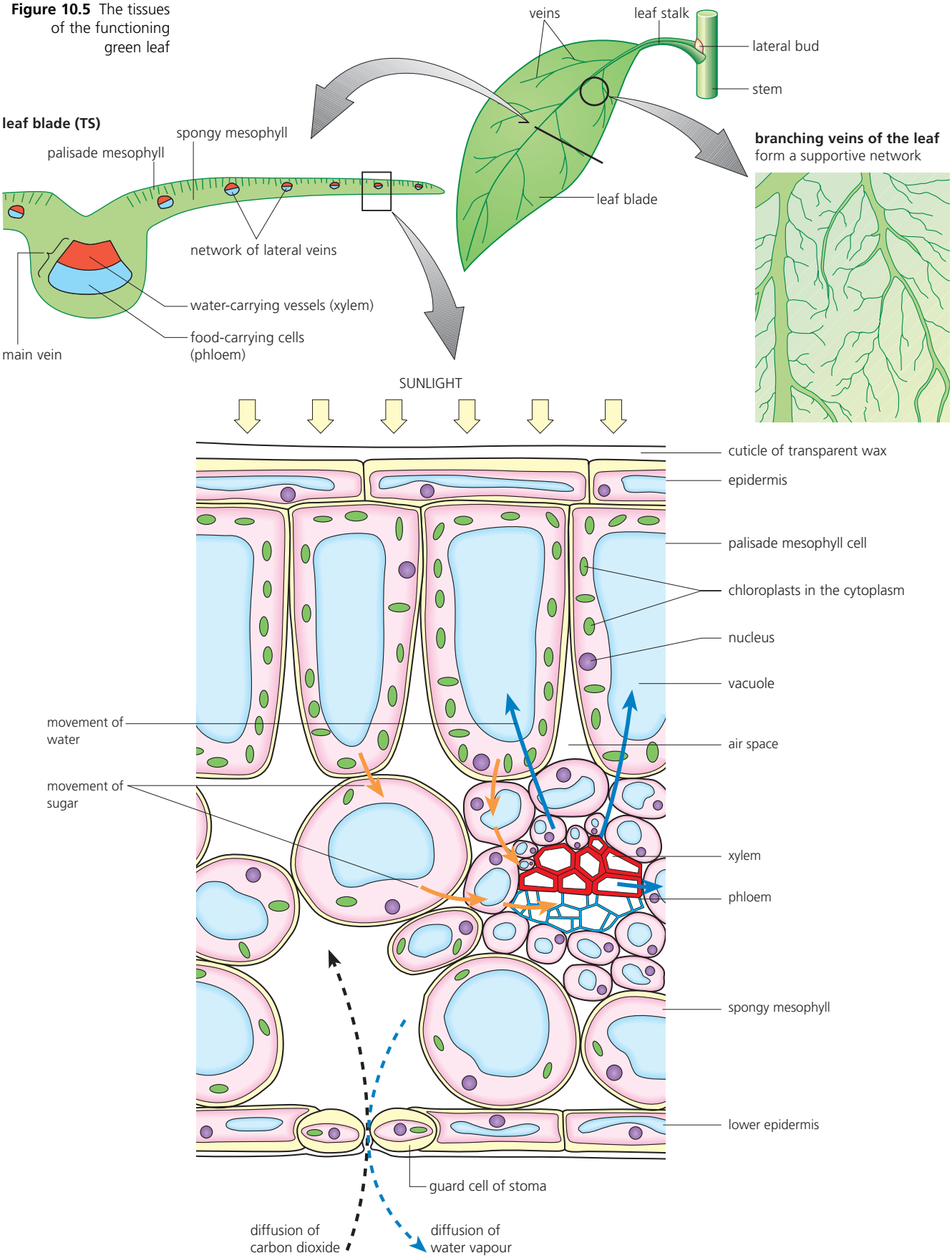
These features are illustrated in Figure 10.5, and summarised in Table 10.2.

2 **Outline** the properties of cellulose that make it an ideal material for plant cell walls.

Feature / structure	Role/function
thin structure of large surface area	maximises light absorption by chloroplasts in the (upper) palisade mesophyll cells
tough, transparent epidermis with waxy cuticle	contains and protects delicate mesophyll cells, particularly from excessive water loss
stomata	site of inward diffusion of carbon dioxide
continuous air spaces between mesophyll cells	pathway of diffusion between external air and surface of mesophyll cells
network of vascular bundles	supports leaf tissue and enables xylem and phloem to service mesophyll cells closely
xylem of vascular bundles	delivers water to leaf cells
phloem of vascular bundles	transports products of photosynthesis to rest of plant

Table 10.2 Functional adaptations of green leaves

Figure 10.5 The tissues of the functioning green leaf



Plant structures can be modified for different functions

We have seen that flowering plant organs (stem, leaf and root) have clearly defined roles in the functioning of a plant such as the sunflower. These same organs have evolved other (or additional), specialised functions in some plants, in particular circumstances.

For example, leaves, stems and roots have become modified for **storage of food reserves** by plants surviving an unfavourable season for growth (Figure 10.6). In bulbs such as the onion plant, it is leaves (or **leaf bases**, actually) that are the site of food reserves on which subsequent growth is initially supported. Similarly, **underground stems** can be the site of storage of food reserves (e.g. the potato), as indeed may be the **main root** (e.g. the carrot).

Not all flowering plants support themselves upright, as free-standing organisms. The habit of creeping growth (on walls and fences), or climbing (particularly over other, taller plants), is often observed. With this growth habit, materials normally invested in supporting the plant body can be used alternatively – for example, in extensive leaf growth – to obvious advantage. Climbing plants frequently form **tendrils** by which they secure their positions. Tendrils are typically forms of modified stems or leaves (Figure 10.7).

Figure 10.6 Plant organs modified for food storage

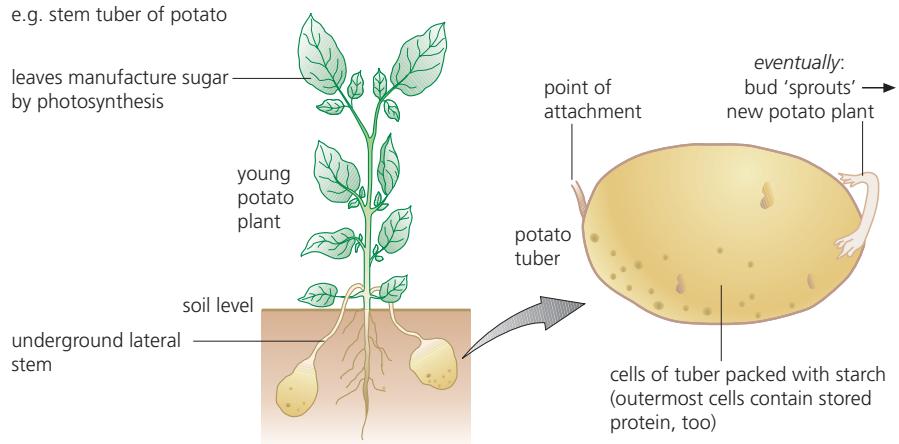
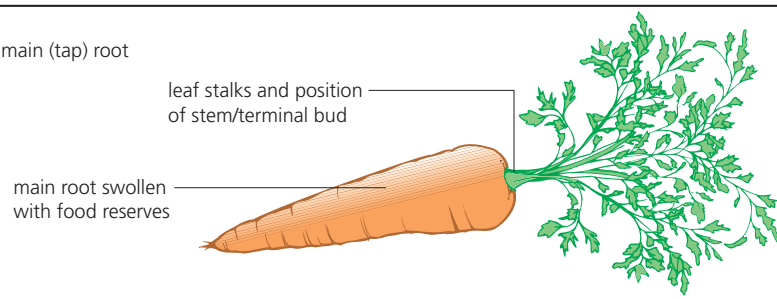
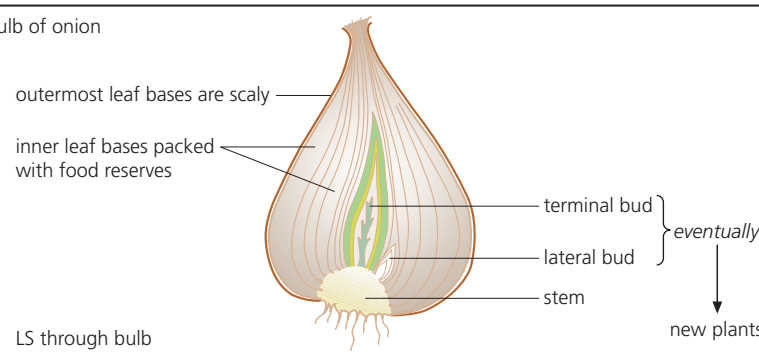
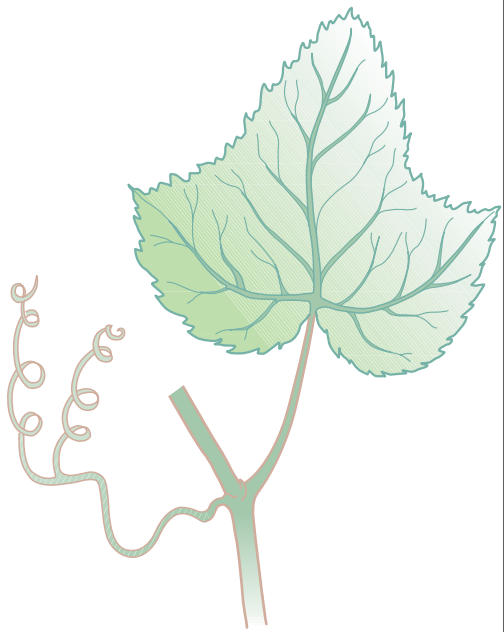
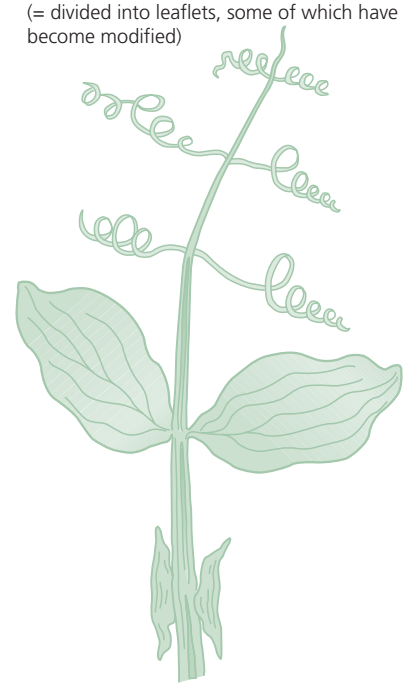
Plant organ	Example
stem	e.g. stem tuber of potato 
root	e.g. carrot, a modified main (tap) root 
leaf	e.g. bulb of onion 

Figure 10.7 Tendrils of climbing plants

Stem tissue modified as a tendril e.g. grape vine (*Vitis* sp.)	**Leaf tissue modified as a tendril** e.g. sweet pea (*Lathyrus* sp.)
tendril – a modified stem	**tendril – a modified part of a compound leaf** (= divided into leaflets, some of which have become modified)
	
Tendrils grow around other plant stems and any other available support, and flexibly support the relatively weak stem and leaves of the climbing plant in a favourable position.	

Plant growth – the role of meristems in dicotyledonous plants

Cells of a plant that are capable of dividing repeatedly are described as **meristematic** cells. A meristem is a group of cells that retain the ability to divide by mitosis. Once a plant has grown past the early embryo stage (an embryonic plant is present in the seed), subsequent growth of the plant occurs by cell division at the meristems (Table 10.3). In Figure 10.8, two types of meristem are identified, namely **apical meristems** and **lateral meristems**, and the growth they are responsible for is illustrated.

Apical meristems occur at the tips of the stem and root and are responsible for primary growth. This leads to an increase in the length of stem and root. First, the new cells formed by **division** rapidly increase in size. Then, this cell **enlargement** phase is followed by cell **differentiation** as the new, enlarging cells become specialised. For example, new cells of the ground tissue (Figure 10.16), contained within the external layer of cells known as the epidermis, form. So do cells of the vascular tissue. These contain water-carrying cells (xylem) and elaborated-food-carrying cells (phloem), and these are assembled as extensions to the existing vascular bundles. These are the primary tissues that make up stems (and roots), and so apical meristems are also called **primary meristems**. Between phloem and xylem of the bundles, a few meristematic cells remain after primary growth, and these form a meristematic tissue called **cambium**.

Lateral meristems form from the cambium cells in the centre of vascular bundles, between the outer phloem tissue and the inner xylem tissue. When the lateral meristem forms and grows, it causes the secondary growth of the plant. Secondary growth involves additions of vascular tissue (secondary phloem and secondary xylem), and results in an increase in the girth of the stem. The first stage in secondary growth occurs when the cambium in the vascular bundles grows into a complete cylinder around the stem. Growth of the lateral meristem increases the circumference of the stem, and also increases the strength of the stem.

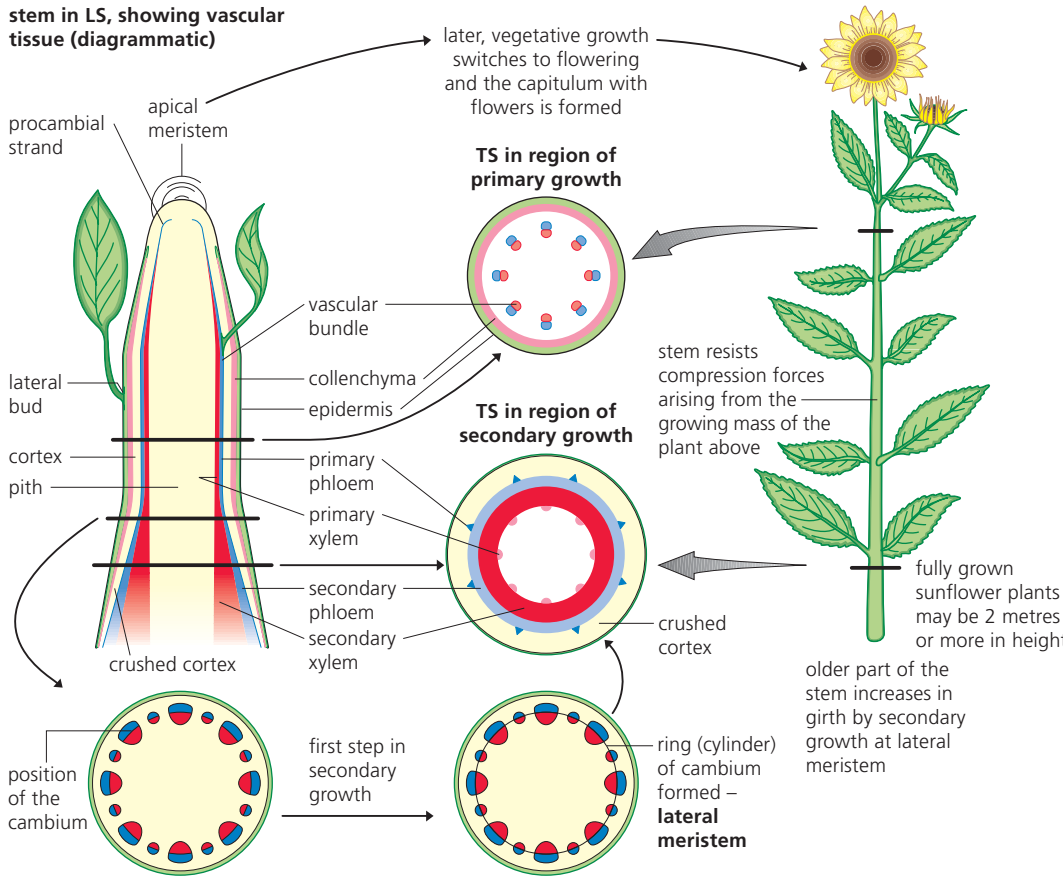


Figure 10.8 The roles of apical and lateral meristems in the growth of stems

Growth due to apical meristem		Growth due to lateral meristem
occurs at tip of stems and roots	**position of meristem**	occurs laterally, between primary phloem and primary xylem
product of embryonic cells	**origin**	cambium – meristematic cells left over from primary growth
produces initial tissues of actively growing plant from the outset	**timing of activity**	functions in older stems (and roots), and in woody plants from the outset
forms epidermis, ground tissues, and primary phloem and xylem	**cell products**	forms mainly secondary phloem and xylem (and often fibres)
produces growth in length and height of plant	**outcome for stem**	produces growth in girth of stem, plus strengthening of stem

Table 10.3 Growth due to apical and lateral meristems compared

The control of plant growth

Among the internal factors that play a part in control of plant growth and sensitivity are substances known as plant growth hormones or, better, as **plant growth substances**. In learning about these substances, bear the following points in mind:

- there are five major types of compound, naturally occurring in plants, that we classify as plant growth substances, one of which is known as **auxin**;
- these substances tend to interact with each other in the control of growth, rather than working in isolation;
- plant growth substances occur in low concentrations in plant tissues, making their extraction and investigation quite challenging.

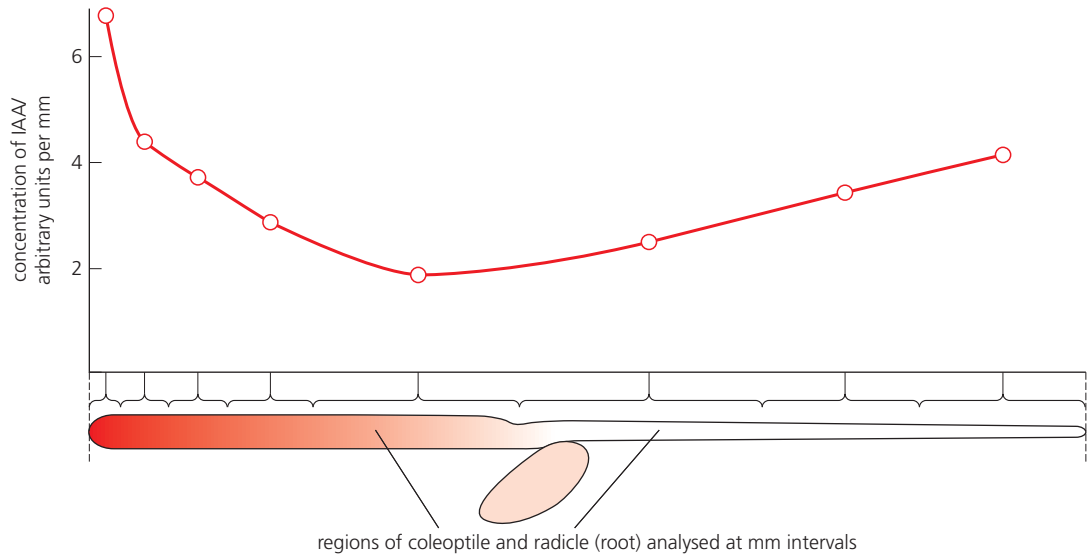


Figure 10.9 Distribution of auxin (IAA) in a young, growing plant

Auxin is manufactured by cells undergoing repeated cell division, such as those found at the stem and root tips. Consequently, the concentration of auxin is highest there (Figure 10.9). Auxin is then transported to the region of growth behind the tip where it causes cells to elongate. In the process, the auxin is used up and inactivated.

Plant stems and light

The response of green plants to light is **complex** and very interesting.

In the **dark**, plant stems grow thin and weak, and they bear tiny, undeveloped leaves, yellow in colour. The shoots of plants in the dark are said to be etiolated. You can demonstrate this by covering growing plants to exclude the light from them completely for a while. By contrast, shoots grown in **full light** are short, with sturdy stems. The leaves are fully expanded and dark green in colour.

We already know that sunlight is essential for photosynthesis, to sustain plant nutrition. But we cannot also say that light is essential for plant growth in length. Quite the contrary in fact – light inhibits plant stem growth – but light *is* essential for chlorophyll formation and leaf expansion.

The shoots grown in unilateral light confirm this. (Unilateral light is a beam of light coming from one direction.) Here, the stems grow towards the light.

We explain this on the basis that growth on the illuminated side of the stem is inhibited by light, but stem growth is unchecked on the dark side, so a growth curvature results. This confirms that light does inhibit plant stem growth in length.

To investigate this response, we can mark a growing stem at regular intervals, using a felt-tip marker pen. Some of the marked stems are then exposed to unilateral light and others to normal illumination. This reveals that the region where elongation of the stem occurs is also the region of growth curvature in unilateral light. So we conclude that the response of the plant stem to unilateral light is a **growth response**.

Growth movements of plant organs in response to an external stimulus in which the direction of the stimulus determines the direction of the response are called **tropic movements** or tropisms. For example, when the stem tip responds by growing towards the light, it is said to be **positively phototropic**.

3 **Construct** a list of the various effects of light on plant growth and development.

Figure 10.10 The coleoptile as an experimental organ

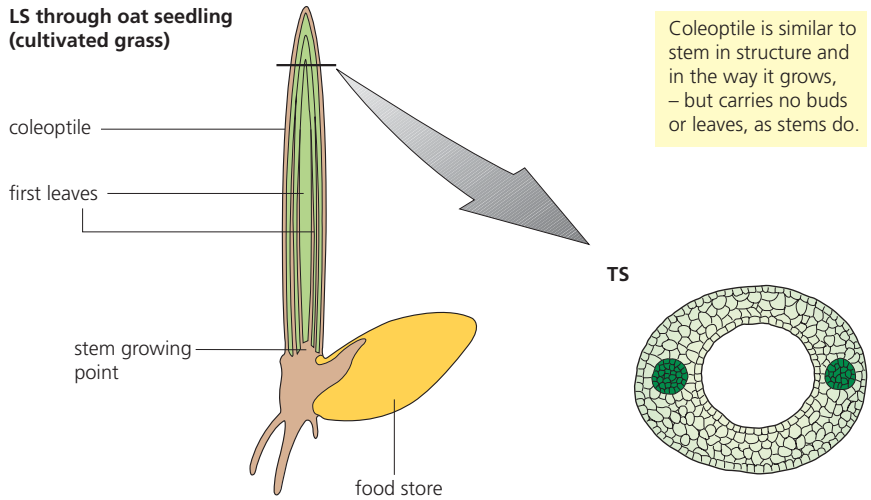
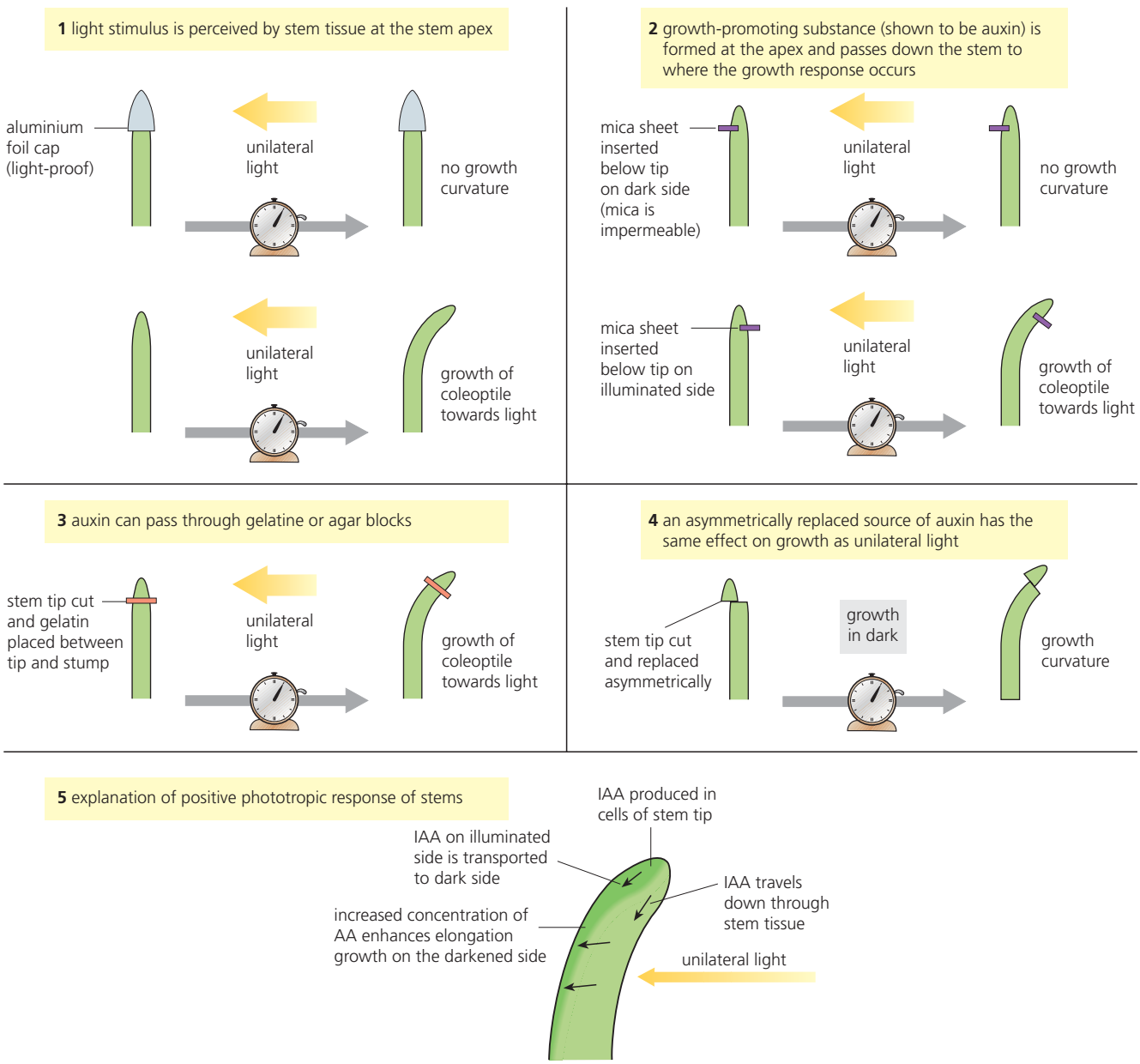


Figure 10.11 Auxin and positive phototropism – establishing the connection



Using coleoptiles to investigate phototropism

The **coleoptile** is a sheath of tissue, unique to the grass family, which encloses the shoot of a germinating grass seedling, but only as long as it grows up through the soil (Figure 10.10). The coleoptile grows rather like a stem does, but it is uncluttered by leaves or buds, so its growth is easily observed.

Experiments to investigate phototropism by examining the responses of oat coleoptiles (oat is a cultivated grass species) to unilateral light led to the discovery of auxin. Much later, auxin was isolated, analysed, and identified as **indoleacetic acid** (**IAA**).

If the tip of the stem or coleoptile is cut off, stood on a small gelatine (or agar) block for a short while, and the *block* then placed on a cut stump of stem or coleoptile, growth in length is continued. The explanation is that the plant growth substance auxin has passed from the coleoptile into the gelatine block, and when the gelatine took the place of a stem tip, the auxin then passed down into the tissue and stimulated elongation of the cells. This technique has been used to investigate auxin actions.

The effect of light on auxin distribution was discovered in experiments with coleoptiles exposed to unilateral light. Here it is the auxin passing down the coleoptile that is redistributed to the darkened side, causing differential growth and the curvature of the stem (Figure 10.11). (The degree of curvature is proportional to the amount of auxin, up to a certain concentration. Above this, additional auxin inhibits growth.)

Transport in angiosperms

9.2.1–9.2.11

Transport within flowering plants is of nutrients (carbon dioxide, water, and essential ions), oxygen for respiration, and elaborated foods (mainly sugar and amino acids). Plant growth substances are also transported, as illustrated by the movements of auxin, discussed above. Transport of water, ions and elaborated food substances involve the functioning vascular bundles, and it is this transport that is discussed now.

Root system, absorption and uptake

The root system provides a huge surface area in contact with soil, for plants have a system of branching roots that continually grow at each root tip, through the soil. This is important because it is the dilute solution that occurs around soil particles that the plant draws on for **essential ions** and the huge volume of **water** it requires.

4 **List** the features of root hairs that facilitate absorption from the soil.

Contact with the soil is vastly increased by the **region of root hairs** that occur just behind the growing tip of each root (Figure 10.12). Root hairs are extensions of individual epidermal cells, and are relatively short-lived. As root growth continues, fresh resources of soil solution are exploited.

Water uptake

Water uptake occurs from the soil solution in contact with the root hairs. Uptake is largely by mass-flow through the interconnecting 'free' spaces in the cellulose cell walls, but there are three possible routes of water movement through plant cells and tissues, in total, as illustrated in Figure 10.13.

- **Mass flow** through the interconnecting free spaces between the cellulose fibres of the plant cell walls. This free space in cellulose makes up about 50% of the wall volume. This route is a highly significant one in water movement about the plant. This pathway, entirely avoiding the living contents of cells, is called the apoplast. The **apoplast** also includes the water-filled spaces of dead cells (and the hollow xylem vessels – as we shall shortly see).
- **Diffusion** through the cytoplasm of cells, and via the cytoplasmic connections between cells (called plasmodesmata). This route is called the **symplast**. As the plant cells are packed with many organelles, these offer resistance to the flow of water and so this pathway is not the major one.

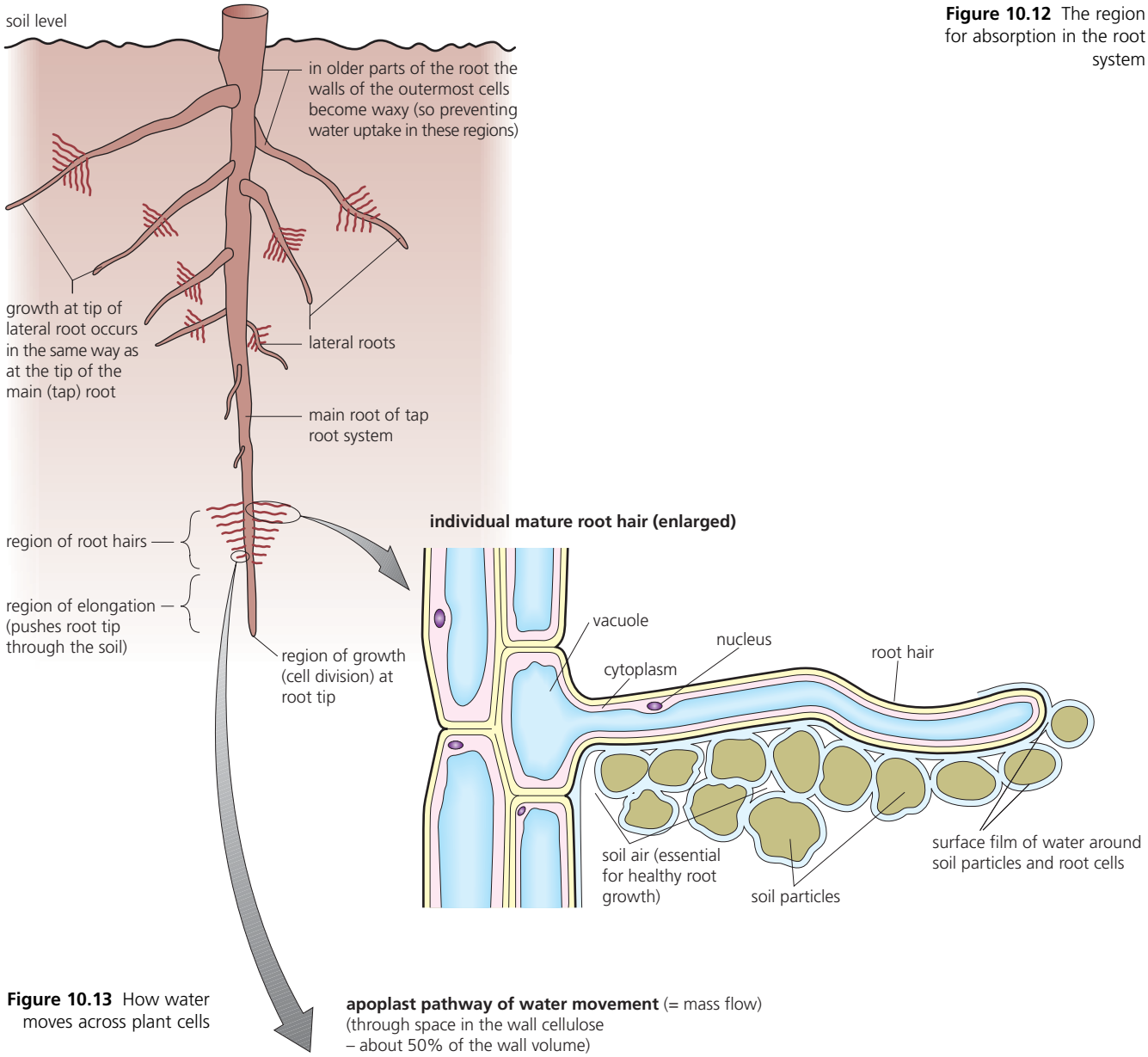


Figure 10.12 The region for absorption in the root system

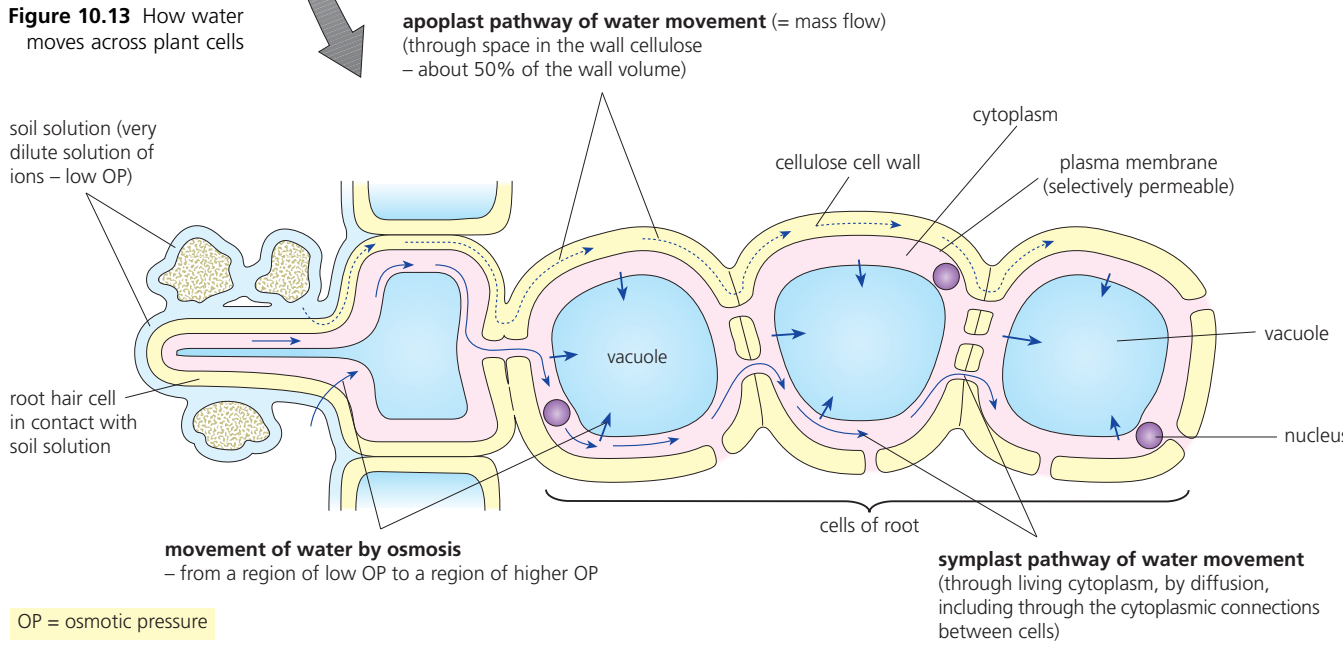


Figure 10.13 How water moves across plant cells

- **Osmosis** from vacuole to vacuole of the cells, driven by a gradient in osmotic pressure. This is not a significant pathway of water transport across the plant, but it is the means by which individual cells absorb water.

5 **Explain** the difference between the symplast and the apoplast.

While all three routes are open, the bulk of water crosses the root tissue to the xylem via the apoplast (Figures 10.13 and 10.18).

Ion uptake

Ion uptake by the roots from the surrounding soil solution is by active transport.

In active transport, metabolic energy is used to drive the transport of molecules and ions across cell membranes. Active transport has characteristic features distinctly different from those of movement by diffusion, for example.

- **Active transport may occur against a concentration gradient** – that is, from a region of low to a region of higher concentration. The cytosol of a cell normally holds some reserves of ions essential to metabolism, like nitrate ions in plant cells. These reserves of useful ions do not escape; the cell membranes retain them inside the cell. Indeed, when additional ions become available to root hair cells, they are actively absorbed into the cells, too. In fact, plant cells tend to hoard valuable ions like nitrates and calcium ions, even when they are already at higher concentration inside the cytoplasm than outside the cell.
- **Active uptake is a highly selective process.** For example, in a situation where sodium nitrate (Na^+ and NO_3^- ions) is available to the root hairs, it is likely that more of the NO_3^- ions are absorbed than the Na^+, since this reflects the needs of the whole plants.
- **Active transport involves special molecules of the membrane, called pumps.** The pump molecule picks up particular ions and transports them to the other side of the membrane, where they are then released. The pump molecules are globular proteins that traverse the lipid bilayer. Movements by these pump molecules require reaction with ATP; by this reaction, metabolic energy is supplied to the process. Most membrane pumps are specific to particular molecules or ions and this is the way selective transport is brought about. If the pump molecule for a particular substance is not present, the substance will not be transported.

It is the presence of numerous specific protein pumps in the plasma membranes of all the root hair cells that makes possible the efficient way that ions reaching the cell surface are absorbed (Figure 10.14). Roots are metabolically very active, and they require a supply of oxygen for aerobic cell respiration. By this process, the required supply of ATP for ion uptake is maintained.

Ions must reach the cell membranes

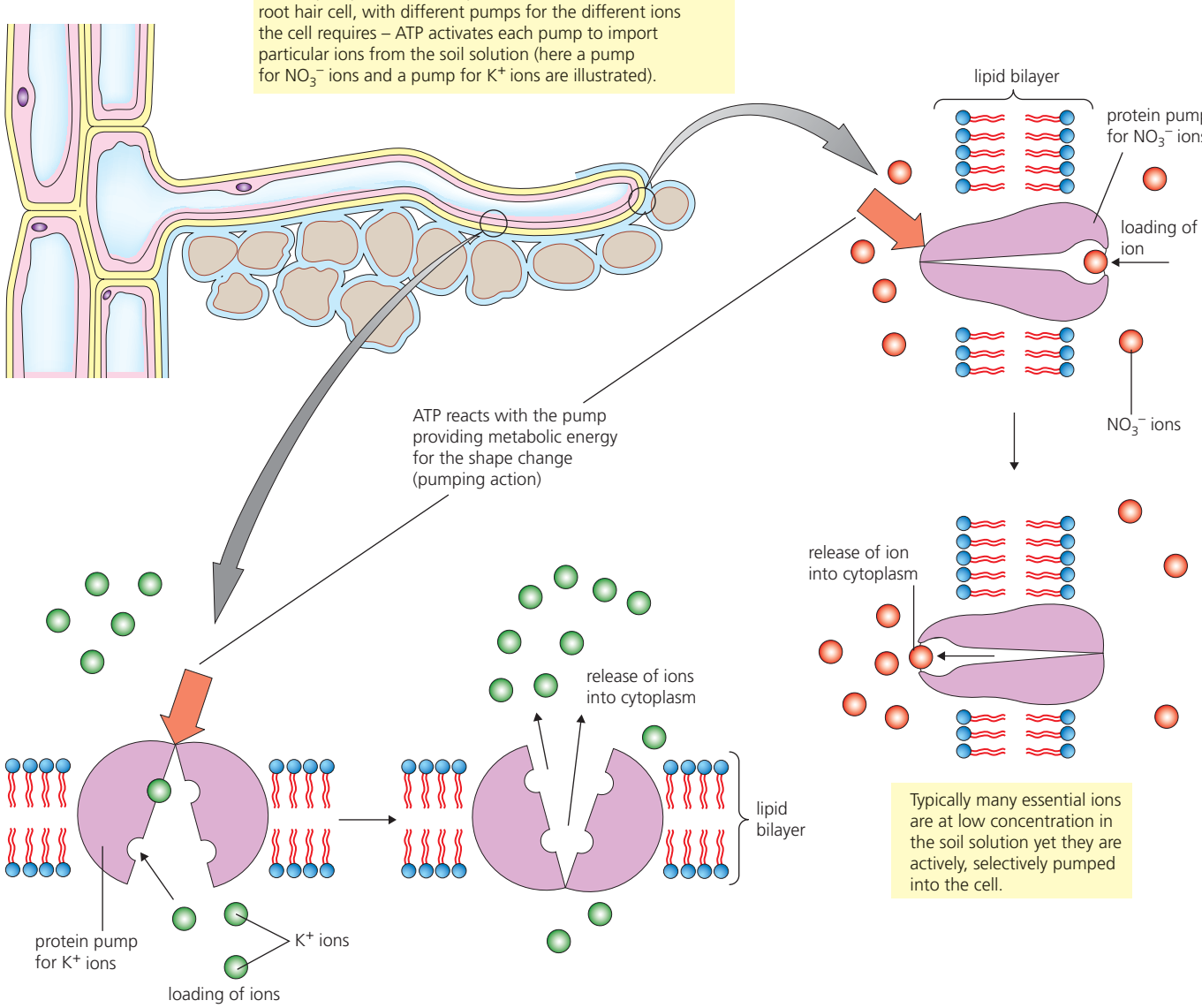
Soil solution contains ions at relatively low concentrations; so how are plasma membranes of the root hair cell supplied with them, in adequate quantities? Well, there are three mechanisms that maintain an adequate supply of ions.

1 The **mass flow of water through the free spaces in the cellulose walls** (the apoplast pathway of water movement – Figure 10.13) delivers fresh soil solution alongside the root hair plasma membranes, continuously.
2 The active uptake of valuable ions from the soil solution of the apoplast maintains a **concentration gradient**, so ions diffuse from higher concentrations outside the apoplast to the solution of lower concentration immediately adjacent to the protein pumps.
3 Many species of **plants live in a mutualistic relationship with species of soil-inhabiting fungi**. In this relationship, the fungal hyphae receive a supply of sugar from the plant root cells. Plants generally have an excess of sugar. In return, the fungal hyphae release to the root cells ions that have previously been taken up as and when they became available in the soil. The spread of fungal hyphae through the surrounding soil is very extensive. The periodic death and decay of other organisms on or in the soil, releases a supply of ions – very often some distance from the plant roots. Nevertheless, plant roots obtain many of these ions later, from the fungal hyphae, thereby establishing that the plant–fungus relationship is one of mutual benefit.

6 **Explain** the significance of root hair cells being able to take up nitrate ions from the soil solution even though their concentration in the cell is already higher than in the soil.

7 **Suggest** why plants often fail in soil which is persistently waterlogged.

Figure 10.14 The active uptake of ions by protein pumps in root hair cell membranes



The stem, support and water transport

The stem supports **leaves in the light**, and **transports water and ions** between the roots and the whole **aerial system** (leaves, flowers, and buds). It also transports **organic nutrients** (amino acids, sugars, etc.), as we shall see shortly.

The ground tissues consist of living cells that make up the bulk of the stem (and root) of herbaceous (non-woody) plants. These cells, known as parenchyma, show little structural adaptation. However, they have an important role in **support of the herbaceous stem**. This they achieve by means of turgidity, which exerts pressure on surrounding cells, and on the tough, continuous epidermis (as we saw in the leaf). How effective this is in maintaining the aerial system in an upright position is demonstrated when a herbaceous plant wilts (Figure 10.15). Incidentally, these ground tissue cells are also sites of starch storage in many plants.

The outermost layers of ground tissue cells of the stem often have additional layers of cellulose thickening laid down unevenly. These are known as collenchyma cells (Figure 10.16).

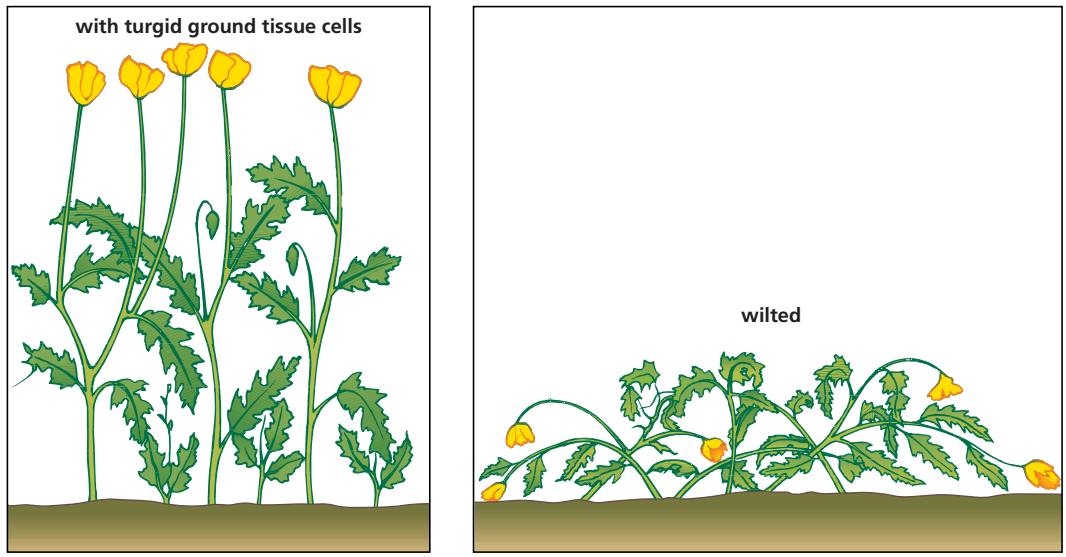


Figure 10.15 A herbaceous plant with and without adequate water

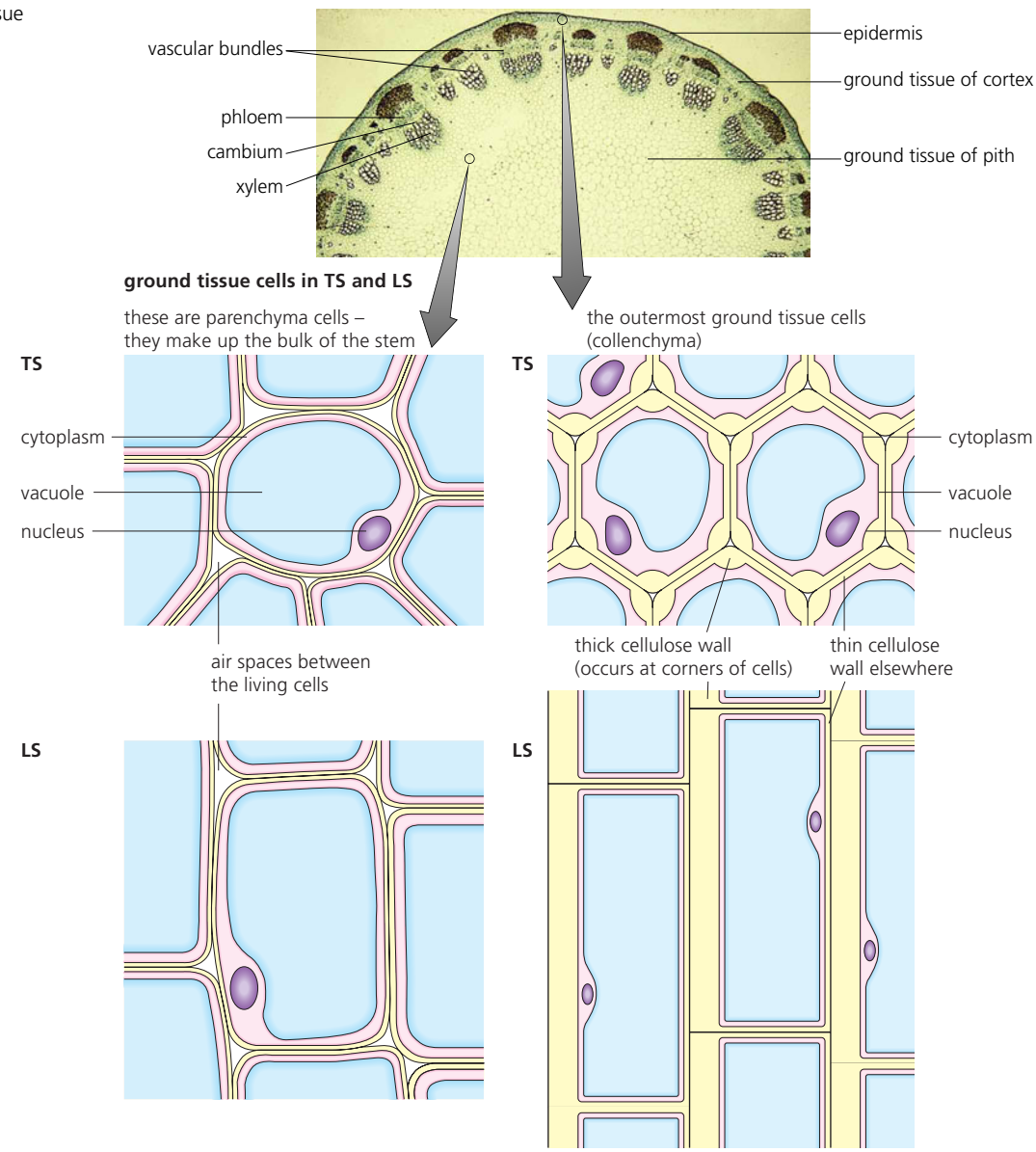


Figure 10.16 The structure of ground tissue

Additional support to the stem comes from the walls of the **xylem tissue** (water-carrying) which are strengthened with cellulose thickenings that are hardened with a chemical substance, **lignin** (Figure 10.17). In **woody plants**, it is lignified cells like the xylem that supply virtually all the support to the stem. The great strength a piece of wood has is evidence of how strong lignin makes a block of plant tissue, otherwise composed only of cellulose cell walls.

Transport of water through the plant

Transport of water through the plant occurs in the xylem tissue. In a root, the xylem is centrally placed; in the stem, xylem occurs in the ring of vascular bundles, as we saw in Figure 10.2. Nevertheless, xylem of root and stem are connected.

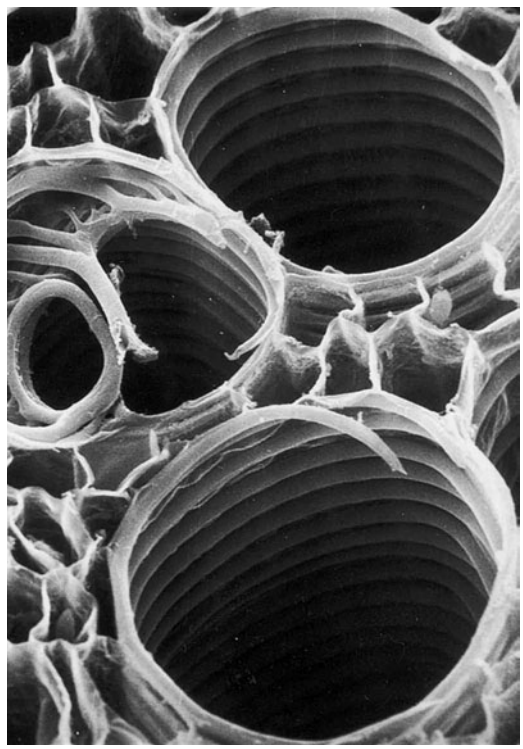


Figure 10.17 SEM of spiral xylem vessels

Xylem begins as elongated cells with cellulose walls and living contents, connected end to end. During development, the end walls are dissolved away so that mature xylem vessels are **long, hollow tubes**. The living contents of a developing xylem vessel are used up in the process of depositing cellulose thickening to the inside of the lateral walls of the vessel, and hardening this by the deposition of lignin. Consequently, xylem is extremely tough tissue and is strengthened internally; it is able to resist negative pressure (suction) without collapsing in on itself. Figure 10.17 is a scanning electron micrograph (SEM) of spirally thickened xylem vessels. Note that other vessels may have differently deposited thickening; many have rings of thickening, for example.

We have already seen that water uptake occurs from the soil solution, mainly at the root hairs (Figure 10.12). This occurs largely by mass flow through the interconnecting free spaces in the cellulose cell walls (apoplast – Figure 10.13).

A tissue map of a root tissue in transverse section shows that the centrally placed vascular tissue is contained by the **endodermis**. The endodermis is a layer of cells unique to the root. At the endodermis, a waxy strip in the radial walls blocks the passage of water by the apoplast route momentarily. This waxy strip is called the **Casparian strip**. Water passes through the endodermis by osmosis.

Meanwhile, in the leaves, evaporation of water occurs and water vapour diffuses out of the stomata, a process called **transpiration**.

Transpiration is the evaporation of water vapour through stomata of green plant leaves (and stems).

Water lost by leaf cells inevitably raises the osmotic pressure in these cells, causing water uptake from surrounding cells, and so from the xylem vessels of the network of veins there. Consequently, a stream of water is drawn up the xylem of the stem by a force generated by transpiration in the leaves, sometimes called the transpiration stream, or the **transpiration pull**.

The properties of water in relation to the chemistry of cell walls ensure the water column coheres to the lateral walls of the xylem and the water column does not break under tension (i.e. negative pressure or suction). This is known as the **cohesive property of water** (page 40). So, a continuous column of water is maintained, moving from root cell walls to leaf cell walls. Much of this water ultimately evaporates – it is lost from the plant as vapour (Figure 10.18).

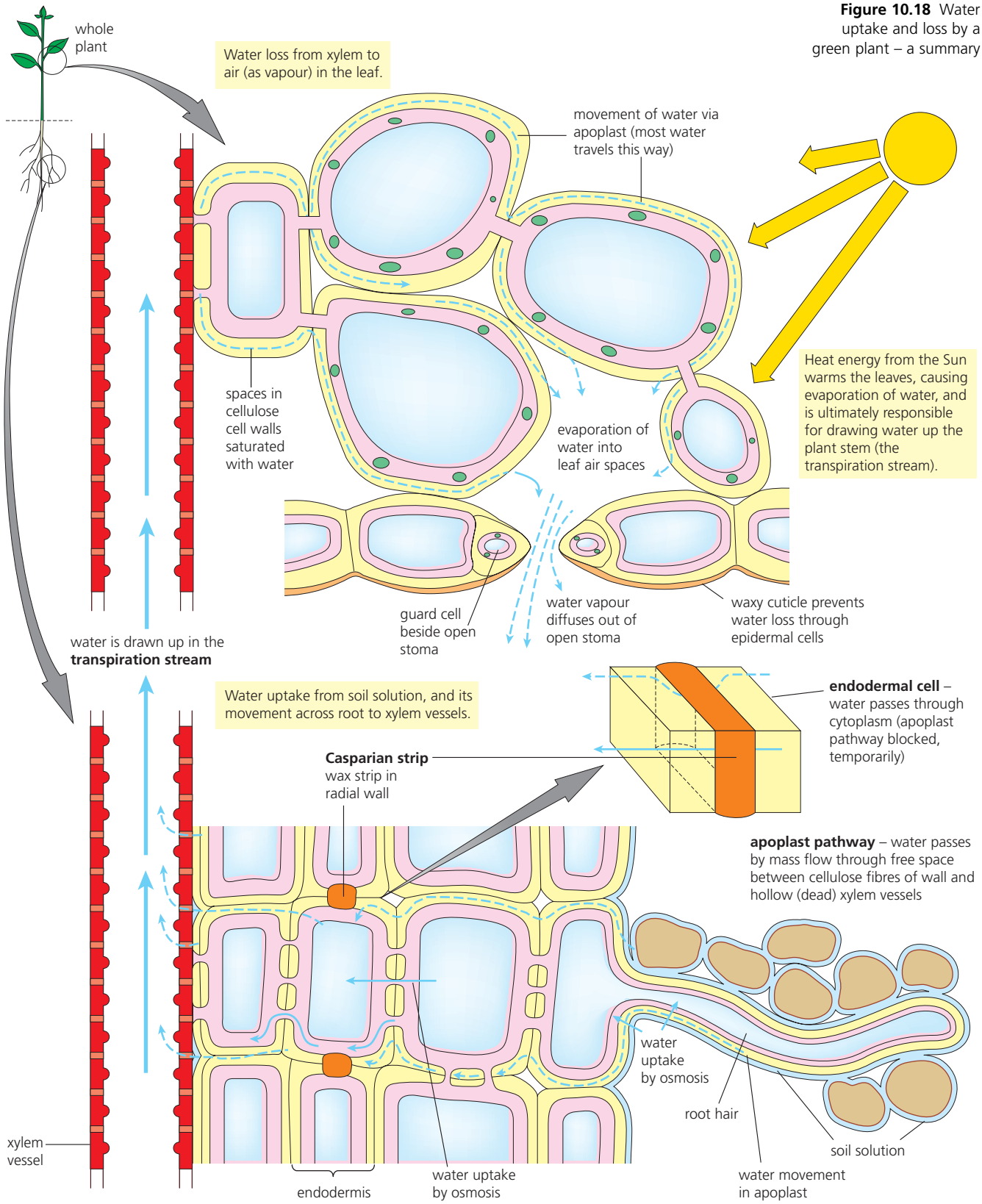


Figure 10.18 Water uptake and loss by a green plant – a summary

8 **Explain** the consequence of the Casparian strip for the apoplast pathway of water movement.

Stomata and transpiration

The tiny pores of the epidermis of leaves through which gas exchange can occur are known as stomata (Figure 10.19). Most stomata occur in the epidermis of leaves, but some do occur in stems. In the broad, flattened leaves typical of many dicotyledonous plants, stomata are concentrated in the lower epidermis. Each stoma consists of two elongated guard cells. These cells are attached to ordinary epidermal cells that surround them and are securely joined together at each end. However, guard cells are detached and free to separate along the length of their abutting sides. When they separate, a pore appears between them.

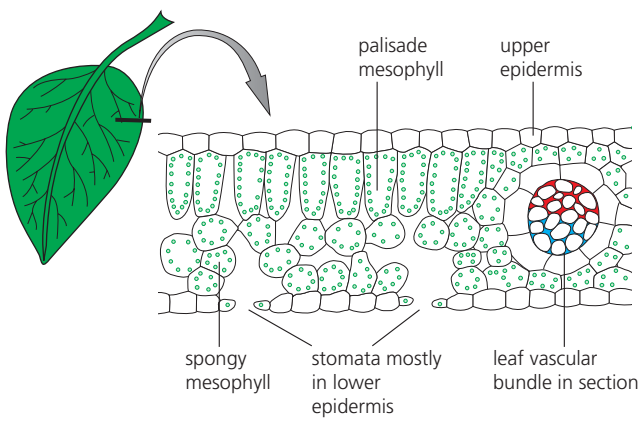
Stomata open and close due to change in turgor pressure of the guard cells. They open when water is absorbed by the guard cells from the surrounding epidermal cells. The guard cells then become fully turgid, and they each push into the epidermal cell beside them (because of the way cellulose is laid down in the walls, Figure 10.19). A pore develops between the guard cells. When water is lost and the guard cells become flaccid, the pore closes again. This has been demonstrated experimentally (Figure 10.20).

Stomata tend to open in daylight and be closed in the dark (but there are exceptions to this). This diurnal pattern is overridden, however, if and when the plant becomes short of water and starts to wilt. For example, in very dry conditions when there is an inadequate water supply, stomata inevitably close relatively early in the day (turgor cannot be maintained). This curtails water vapour loss by transpiration and halts further wilting. Adequate water reserves from the soil may be taken up subsequently, thereby allowing the opening of stomata again – for example, on the following day. The effect of this mechanism is that stomata regulate transpiration in that they prevent excessive water loss when this is threatened (Figure 10.21).

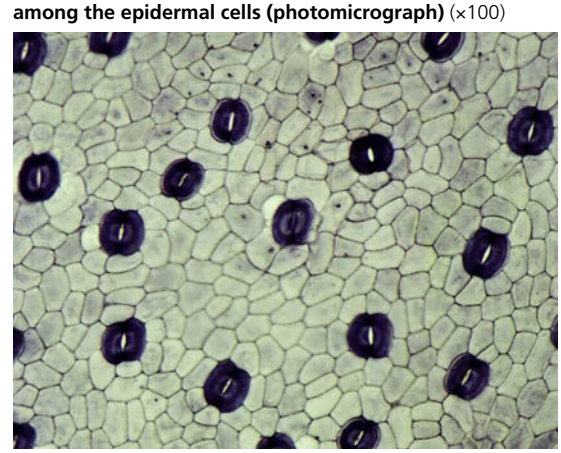
9 Suggest why changes in the turgor of guard cells cause opening of stomata.

Figure 10.19 The distribution and structure of stomata

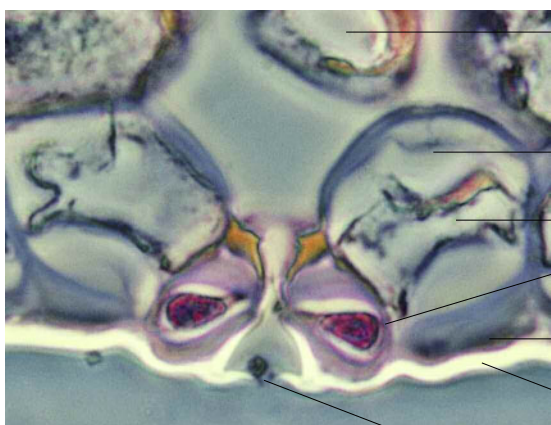
distribution of stomata in typical dicotyledonous leaf



lower surface of leaf – showing distribution of stomata among the epidermal cells (photomicrograph) (×100)



structure of individual stoma (photomicrograph) (×500)



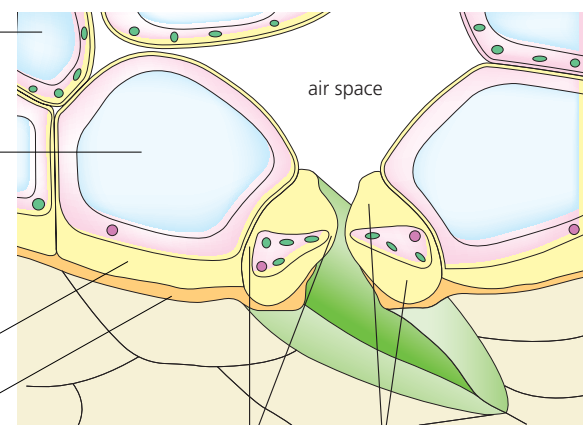
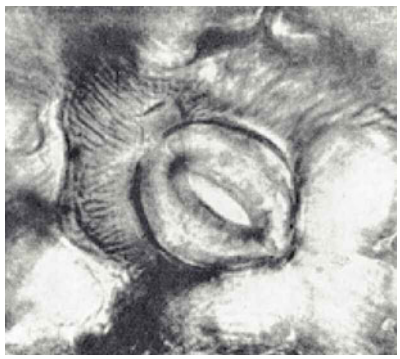


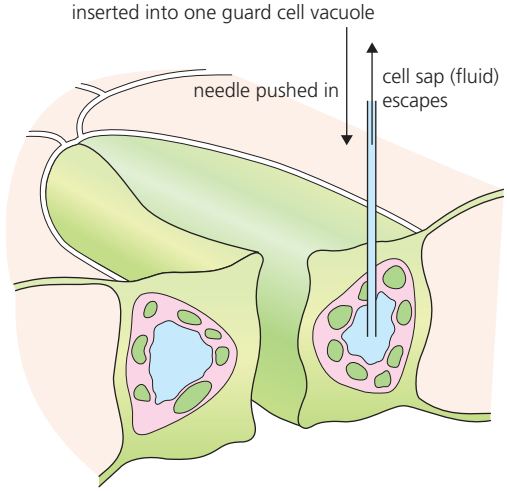
Figure 10.20 It's turgor pressure that does it!

In an experimental demonstration that turgor pressure of the guard cells causes the opening of the stomatal pore, a microdissection needle was inserted into a guard cell.

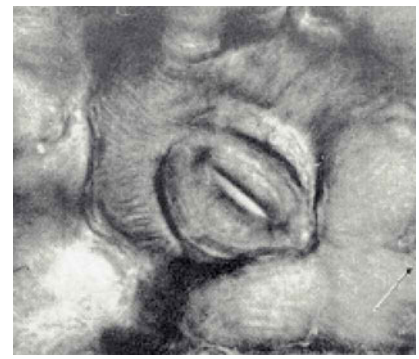
A fully open pore due to turgid guard cells



B microdissection needle (fine hollow tube) inserted into one guard cell vacuole



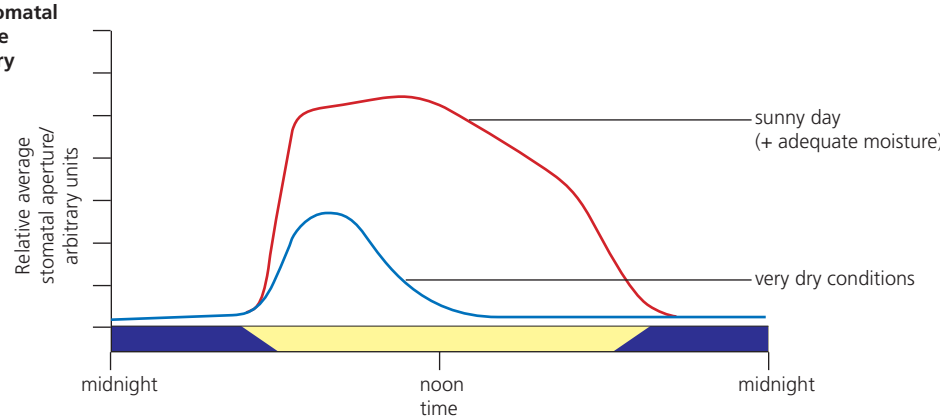
C half-open pore due to collapse of turgor



Observation: on release of the turgor pressure in one guard cell, the distinctive shape of the cell when the pore is open was lost. 'Half' of the pore disappeared.

Figure 10.21 Stomatal opening and environmental conditions

how stomatal aperture may vary



10 Examine Figure 10.21. **Suggest** why the stomatal apertures of the plant in very dry conditions differed in both maximum size and duration of opening from those of the plant with adequate moisture.

Abscisic acid (ABA) and water stress conditions

Abscisic acid is another plant growth regulator found to occur in stems, leaves and fruits. Many of the effects of ABA result from its interaction with other growth regulator substances, including its effects on seed germination, *but these do not concern us here*. ABA is important in leaves because its presence assists plant survival at times of physiological stress, such as prolonged drought. In these conditions, the level of ABA present is raised, and it maintains stomatal closure, reducing water loss.

The rate of transpiration may be investigated using a potometer (Figure 10.22) and is found to be dramatically affected by **environmental conditions** around the plant.

Why is this so?

Transpiration occurs because water molecules continuously evaporate from the cellulose walls of cells in the leaf which are saturated with water. This makes the air in the air spaces between mesophyll cells more or less saturated with water vapour. If the air outside the plant is less saturated (less humid – as it very often is) *and the stomata are open*, water vapour will diffuse out into the drier air outside.

In effect, transpiration is an unfortunate consequence of plant structure and nutrition that can only be slowed or stopped by closure of the stomata. In the light, plants tend to dry out the soil they grow in! Thus we can see how changes in environmental conditions will affect transpiration, and why.

For example:

1 **Light** affects transpiration because the stomata tend to be open in the light, and open stomata are essential for loss of water vapour from the leaf. There is also an indirect effect of light because it comes from the Sun and contains infra-red rays which warm the leaf and raise its temperature. Light is an essential factor for transpiration.
2 **Temperature** affects transpiration because it causes the evaporation of water molecules from the surfaces of the cells of the leaf. A rise in the concentration of water vapour within the air spaces increases the difference in concentration in water vapour between the leaf's interior and the air outside, and diffusion is enhanced. So an increase in temperature of the leaf raises the transpiration rate.
3 **Wind** sweeps away the water vapour molecules accumulating outside the stomata of the epidermis of the leaf surface, so enhancing the difference in concentration of water vapour between the leaf interior and the outside. Movements of air around the plant enhance transpiration.
4 **Humid air**, if it collects around a leaf, decreases the difference in concentration of water vapour between the interior and exterior of the leaf, so slowing diffusion of water vapour from the leaf. High humidity slows transpiration.

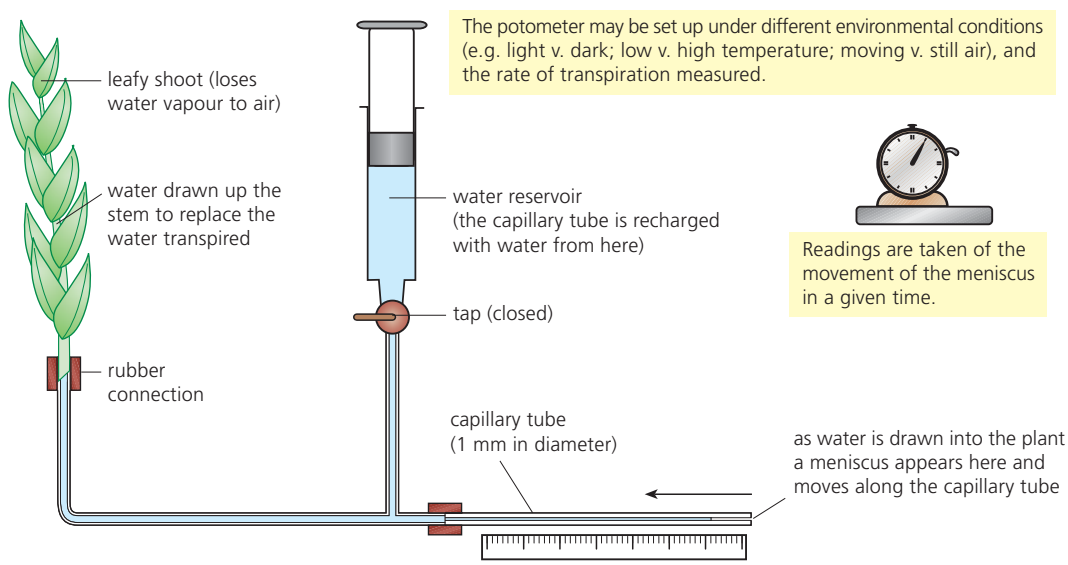
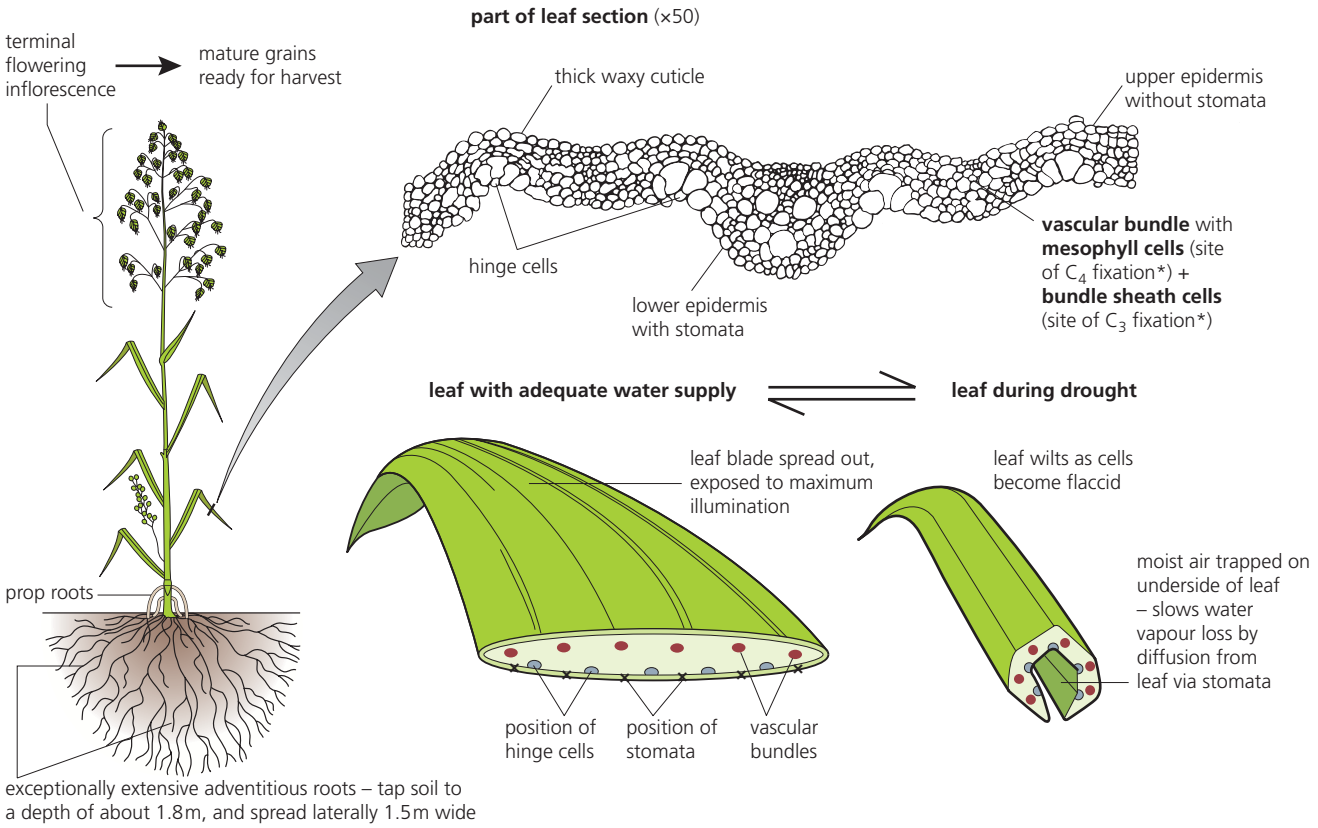
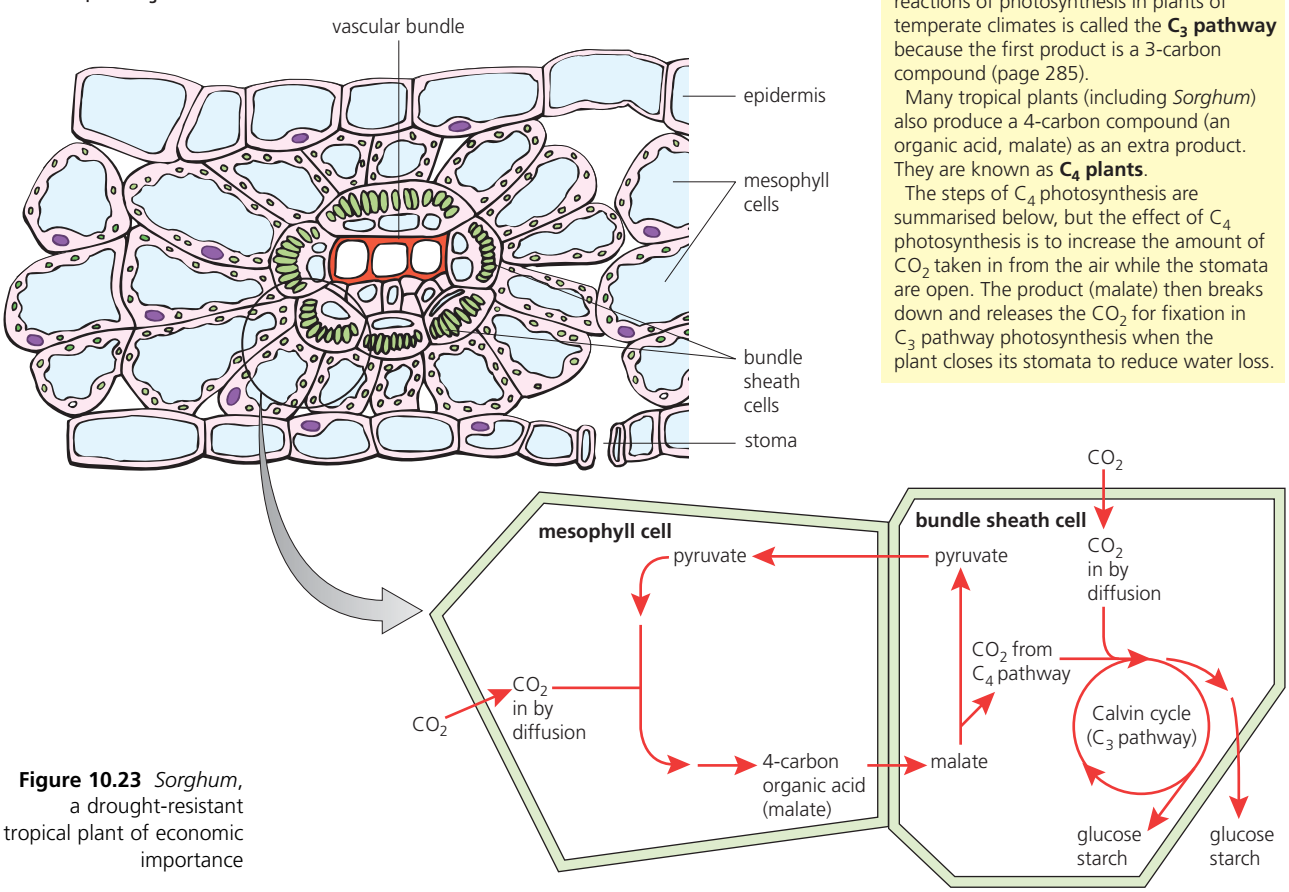


Figure 10.22 Investigating transpiration and the factors that influence it

Xerophytes – plants of permanently dry and arid conditions

Most native plants of temperate and tropical zones, and most of our crop plants, grow best in habitats with adequate rainfall, well-drained soils, and with their aerial system (stem and leaves) exposed to moderately dry air. Loss of water vapour from the leaves may be substantial in drier periods, particularly in the early part of the day, but excessive loss from the leaves is prevented by the responses of the stomata (Figure 10.21). Any deficit is normally made good by the water uptake that continues, day and night. It is the structure of these sorts of plants (known as mesophytes) that we have been considering in this chapter.





***Photosynthesis in C_4 plants**

The fixation of CO_2 in the light-independent reactions of photosynthesis in plants of temperate climates is called the **C_3 pathway** because the first product is a 3-carbon compound (page 285).

Many tropical plants (including *Sorghum*) also produce a 4-carbon compound (an organic acid, malate) as an extra product. They are known as **C_4 plants**.

The steps of C_4 photosynthesis are summarised below, but the effect of C_4 photosynthesis is to increase the amount of CO_2 taken in from the air while the stomata are open. The product (malate) then breaks down and releases the CO_2 for fixation in C_3 pathway photosynthesis when the plant closes its stomata to reduce water loss.

Figure 10.23 *Sorghum*, a drought-resistant tropical plant of economic importance

On the other hand, **xerophytes** are plants able to survive and often grow well in habitats where water is scarce. These plants show features that directly or indirectly help to minimise water loss, due to transpiration. Their adaptations are referred to as **xeromorphic features**, and these are summarised in Table 10.4.

Table 10.4 Xeromorphic features

Structural features	**Effect**
exceptionally thick cuticle to leaf (and stem) epidermis	prevents water loss through the external wall of the epidermal cells
layer of hairs on the epidermis	traps moist air over the leaf and reduces the diffusion
reduction in the number of stomata	reduces outlets through which moist air can diffuse
stomata in pits or groves	moist air trapped outside the stomata, reducing diffusion
leaf rolled or folded when short of water (cells flaccid)	reduces area from which transpiration can occur
superficial roots	exploit overnight condensation at soil surface
deep and extensive roots	exploit a deep water table in the soil
Biochemical features	**Effect**
C_4 photosynthesis pathway (Figure 10.23) operates in addition to normal photosynthesis (C_3)	enhances photosynthetic CO_2 fixation in daylight while stomata are open – allowing the stomata to be closed in dry conditions
CAM metabolism (Figure 10.25)	retains CO_2 from the air in organic acids in cells overnight (in the dark) while stomata are open, and then releases this CO_2 into leaf air space in the light (stomata now closed) for fixation in photosynthesis

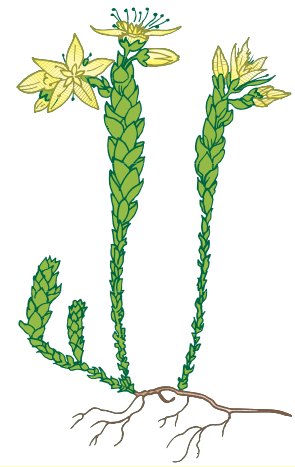
Figure 10.24 Marram grass (*Ammophila*), a pioneer plant of sand dunes

marram grass has the ability to grow in the extremely arid environment of sand dunes, accelerating the build-up of sand

Xerophytic plants typically show some of the adaptations listed below.

- **Sorghum** (*Sorghum*), a grain plant that thrives in extremely arid conditions, shows marked adaptations in its leaf structure and root system, and in its C_4 photosynthesis (Figure 10.23).
- **Marram grass** (*Ammophila*), a wild grass plant of sand dunes, has leaves that, at first glance, look more like stems (Figure 10.24).
- **Stonecrop** (*Sedum*), a plant of dry stone walls and rocky slopes, shows pronounced structural and biochemical adaptations (Figure 10.25).

Figure 10.25 Stonecrop (*Sedum*) – a plant of dry walls and rocky slopes



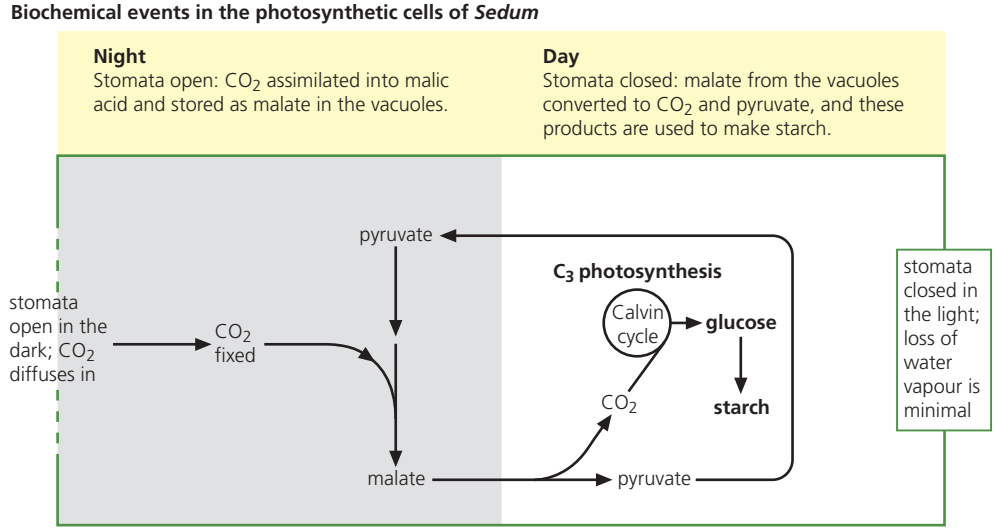
Sedum occurs in habitats where water is rarely available.

Sedum belongs to the angiospermophyte family Crassulaceae – succulent (fleshy) xerophytic plants. They show many xeromorphic features (Table 10.4), but their special feature is **crassulacean acid metabolism** (**CAM**).

In CAM plants, stomata are open in the night only (transpiration is at a minimum), and they fix CO_2 into the organic acid malate.

In the light, the stomata close, and the malate breaks down to release CO_2 into the air spaces of the plant. This CO_2 is fixed by C_3 photosynthesis (page 285), with minimal loss of water vapour from the plant.

Biochemical events in the photosynthetic cells of *Sedum*



11 **Suggest** why it is that, of all the environmental factors which affect plant growth, the issue of water supply is so critical.

12 **Explain** why we can describe the external epidermis of the leaf of marram grass as botanically equivalent to the lower epidermis of a mesophyte leaf.

The stem and organic solute transport

Translocation is the movement of manufactured food (sugars and amino acids, mainly) which occurs in the phloem tissue of the vascular bundles. Sugars are made in the leaves (in the light) by photosynthesis and transported as sucrose. The first-formed leaves, once established, transport sugars to sites of new growth (new stem, new leaves and new roots). In older plants, sucrose is increasingly transported to sites of storage, such as the cortex of roots or stems, and in seeds and fruits (see below).

Amino acids are mostly made in the root tips. Here, absorption of nitrates occurs which the plant uses in the synthesis of amino acids. After their manufacture, amino acids are transported to sites where protein synthesis is occurring. These are mostly in the buds, young leaves and young roots, and in developing fruits.

Translocation is not restricted to organic compounds manufactured within the plant. Chemicals that are applied to plants, for example by spraying, and are then absorbed by the leaves, may be carried all over the organism. Consequently, pesticides of this type are called systemic.

Phloem tissue consists of **sieve tubes** and **companion cells**. Sieve tubes are narrow, elongated elements, connected end to end to form tubes. The end walls, known as **sieve plates**, are perforated by pores. The cytoplasm of a mature sieve tube has no nucleus, nor many of the other organelles of a cell. However, each sieve tube is connected to a companion cell by strands of cytoplasm passing through gaps (called pits) in the walls. The companion cells are believed to service and maintain the cytoplasm of the sieve tube, which has lost its nucleus.

Phloem is a living tissue, and has a relatively high rate of aerobic respiration during transport. In fact, transport of manufactured food in the phloem is an active process, using energy from metabolism.

Phloem transport may occur in either direction in stem, leaves and roots, and is believed to occur by mass flow, as shown in Figure 10.26.

How does mass flow work?

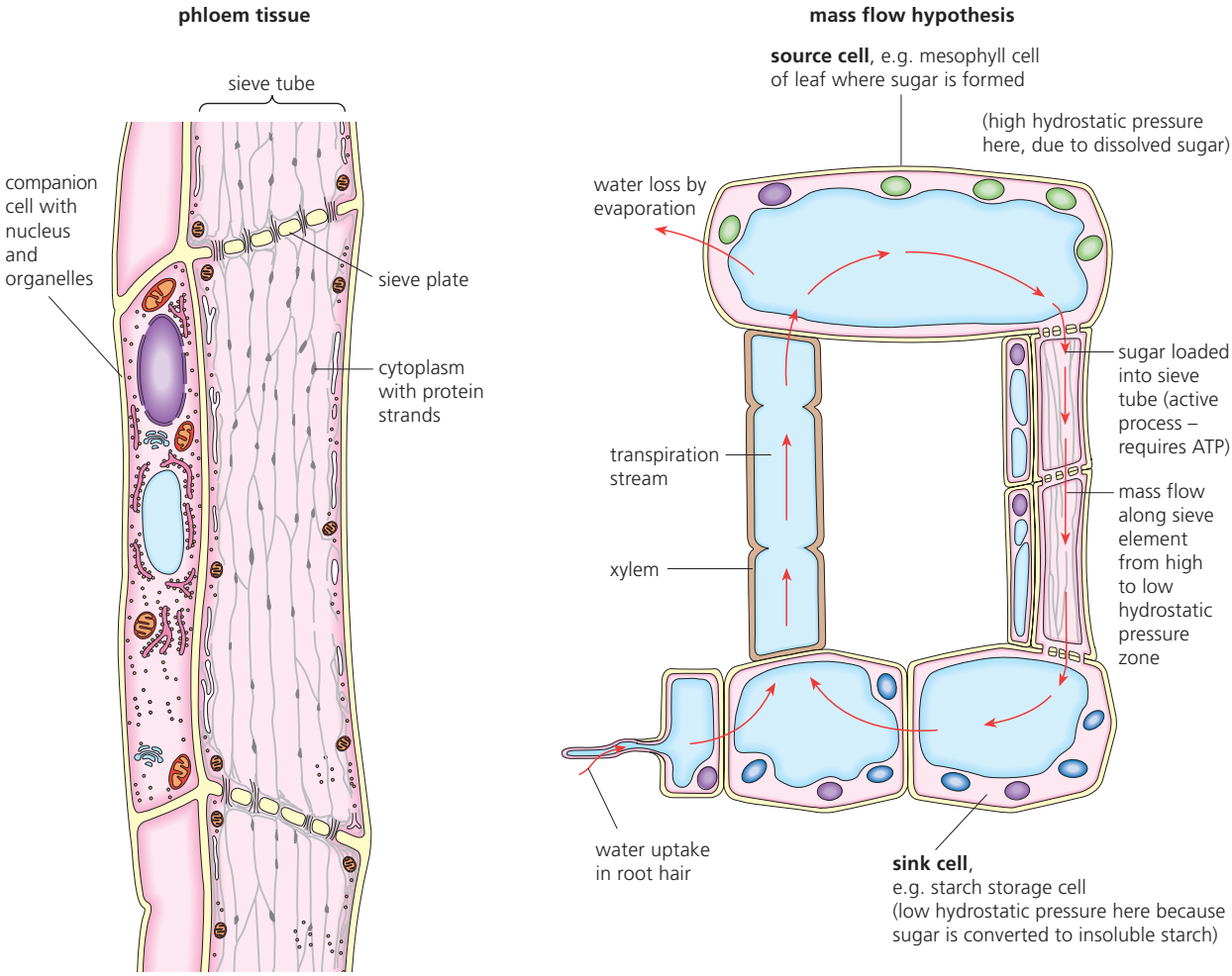


Figure 10.26 The structure of phloem and the mass flow hypothesis of phloem transport

Solutes are loaded into the phloem sieve tubes (a process requiring ATP), and then solutes flow through the phloem from a region of high hydrostatic pressure to a region of low hydrostatic pressure. Hydrostatic pressure is high in and around photosynthesising cells in the light (mesophyll cells of the leaf), and in the phloem sieve tubes nearby, because of the presence of sugar, which generates a high osmotic pressure. Consequently water flows in, raising the hydrostatic pressure still further. This is called a source area.

Hydrostatic pressure is low in cells where sugar is converted to starch and stored (for example, ground tissue cells of the stem and root) and in the phloem sieve tubes nearby. Here the removal of sugar lowers the osmotic pressure, and water flows away. Starch storage cells like these are called sink areas.

13 Describe the differences between transpiration and translocation.

Reproduction in flowering plants

9.3.1–9.3.6

Flowering plants contain their reproductive organs in the flower. Flowers are often hermaphrodite structures, carrying both male and female parts.

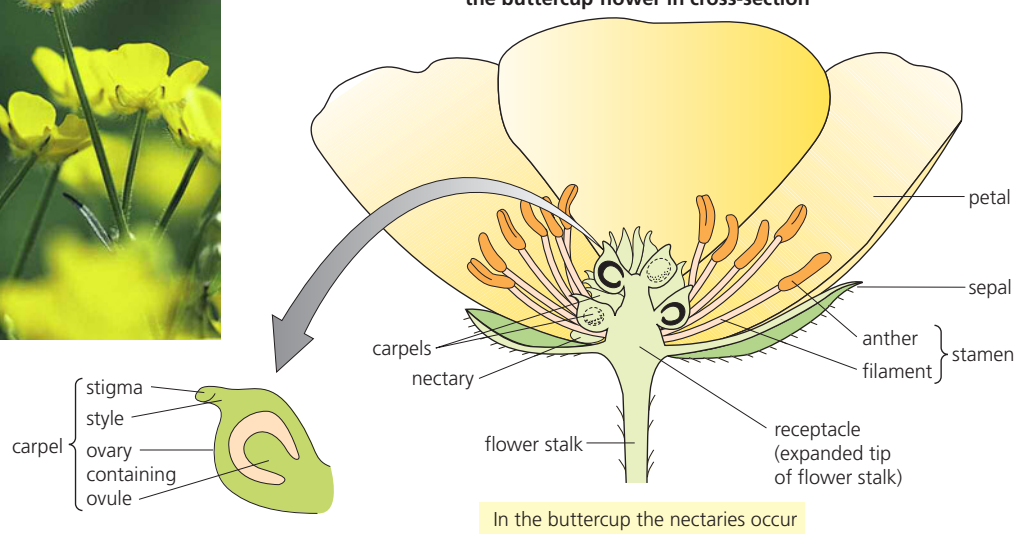
The parts of flowers occur in rings or whorls, attached to the swollen tip of the flower stalk, called the receptacle. The **sepals** (collectively, the calyx) enclose the flower in the bud, and are usually small, green and leaf-like. The **petals** (collectively, the corolla) are often coloured and conspicuous, and may attract insects or other small animals. The stamens are the male parts of the flower, and consist of **anthers** (housing pollen grains) and the **filament** (stalk). The carpels are the female part of the flower. There may be one or many, free-standing or fused together. Each carpel consists of an **ovary** (containing ovules), a **stigma** (surface receiving pollen), and a connecting style.

The buttercup flower is shown in Figure 10.27, and other flowers common in different parts of the world are shown in Figure 10.28.

Figure 10.27 The buttercup (*Ranunculus*) flower

the inflorescence of buttercup (*Ranunculus acris*)

the buttercup flower in cross-section



In the buttercup the nectaries occur one at the base of each petal.

Figure 10.28 Other animal-pollinated flowers common in parts of the world

Bougainvillea rosenka Native of tropical and sub-tropical South America. The flowers are small but surrounded by brightly coloured leaves (bracts).

Hibiscus syriacus Native of warm temperate, tropical and sub-tropical regions throughout the world. The large, trumpet-shaped flowers have five petals.

Pollination and fertilisation

Pollination is the transfer of pollen from a mature anther to a receptive stigma.

The pollen may come from the anthers of the same flower or flowers of the same plant, in which case, this is referred to as **self-pollination**. Alternatively, pollen may come from flowers on a different plant of the same species, which is referred to as **cross-pollination**.

Transfer of the pollen is usually by **insects** or by the **wind**, although in the flowers of certain species, running water or bird or bat visitors to the flowers may be the agent that carries out pollination. Insect-pollinated flowers typically produce a sugar solution, called nectar, which attracts insects to the flower.

Fertilisation in flowering plants can occur only after an appropriate pollen grain has landed on the stigma, and germinated there.

Fertilisation is the fusion of male and female gametes to form a zygote.

The pollen grain produces a pollen tube which grows down between the cells of the style, and into the ovule through the micropyle (Figure 10.29). Incidentally, the pollen tube delivers two male nuclei. One of these male nuclei then fuses with the egg nucleus in the embryo sac, forming a diploid zygote. The other fuses with another nucleus which triggers formation of the food store for the developing embryo. This 'double fertilisation' is unique to flowering plants.

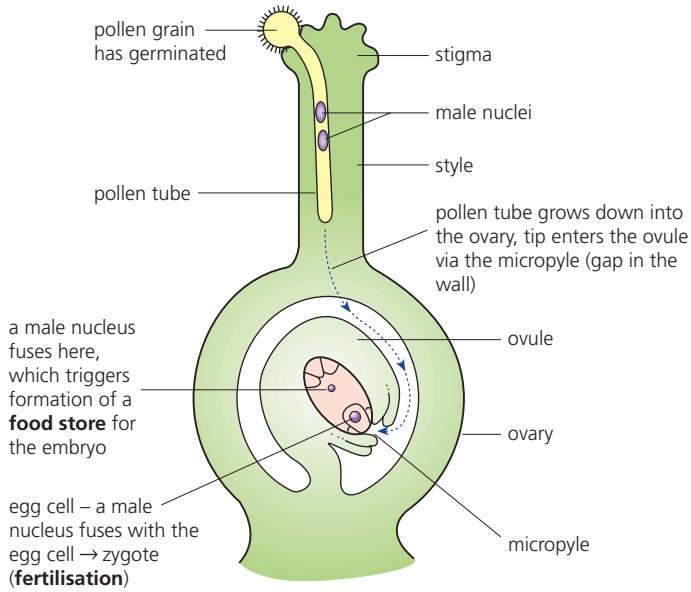


Figure 10.29 Fertilisation in a flowering plant

14 Explain the differences between pollination and fertilisation in the flowering plant.

Extension:

Because many flowers are hermaphrodite there is the potential for self-fertilisation that cannot arise in a unisexual organism. If self-fertilisation occurs, the offspring produced will show less variation from their parents than if fertilisation occurs between gametes from different individuals. So it is interesting to note that several mechanisms that prevent self-fertilisation are common in hermaphrodite flowers, such as the stamens and stigmas maturing at different times. This means that when the stigma is receptive the stamens are not releasing pollen grains, and vice versa.

Seed formation and dispersal

The seed develops from the fertilised ovule and contains an embryo plant and a food store. After fertilisation:

- The zygote grows by repeated mitotic division to produce cells that form an **embryonic plant**, consisting of an embryo root, an embryo stem, and either a single cotyledon (seed leaf) or two cotyledons. (Remember, the phylum Angiospermophyta is divided into two classes, according to the number of cotyledons present. The monocotyledons have a single seed leaf, the dicotyledons have two.)
- Formation of **stored food reserves** is triggered. In many seeds, the developing food store is absorbed into the cotyledons, rather than remaining as a separate store, packed round the embryonic plant. For example, this is the case in peas and beans (Figure 10.30). Note that formation of food reserves can only occur if fertilisation occurs – in the absence of fertilisation, food reserves are not moved into the unfertilised ovule.

As the seed matures, the outer layers of the ovule become the protective seed coat or **testa** and the whole ovary develops into the **fruit**. Next, the water content of the seed decreases, and the seed moves into a dormancy period. In a mature, fully dormant seed, water makes up only 10–15% of seed weight.

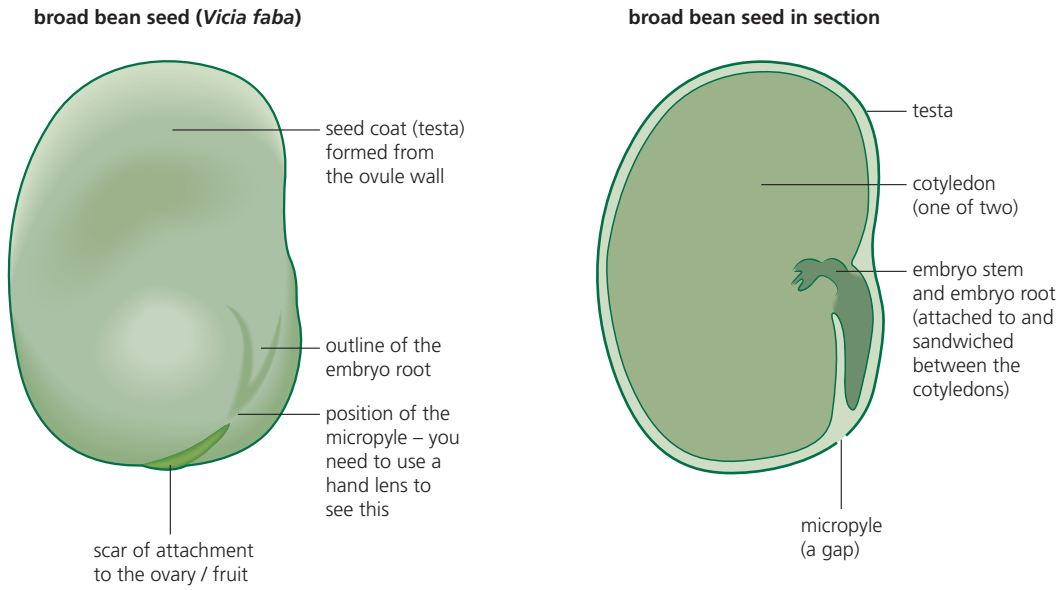


Figure 10.30 The structure of a dicotyledonous seed

15 **State** a fruit or vegetable we eat that originates from:

- **a** an ovary containing one seed
- **b** an ovary containing many seeds
- **c** several ovaries fused together, containing many seeds.

The **seed** is a form in which the flowering plant may be dispersed. If offspring seeds eventually germinate some distance apart, there is more likelihood they will not be competing for the same resources of space, water and light.

Seed dispersal is the carrying of the seed away from the vicinity of the parent plant.

The plant structures to aid dispersal that have evolved variously exploit air currents (wind), passing animals or flowing water to transport seeds. In a few plants, seeds are flung away from the ripening fruit by an explosive mechanism. All seeds are compact, nutritious, and relatively lightweight – in effect they are food packages to a hungry animal. In the process, many seeds taken in this way are dropped or lost in transit, and so these seeds, too, may have been successfully dispersed.

The physiology of seed germination

Many seeds do not germinate as soon as they are formed and dispersed. Such seeds are said to have a **dormant period** and germinate only when this has elapsed. Dormancy may be imposed within the seed, due to:

- **incomplete seed development** that causes the embryo to be immature, and which is overcome in time;
- the **presence of a plant growth regulator** – abscisic acid, for example – that inhibits development, and which only disappears from the seed tissues with time;
- an **impervious seed coat** that is eventually made permeable – for example, by abrasion with coarse soil or by the action of microorganisms;
- a **requirement for pre-chilling** under moist conditions, before the seed can germinate; some seeds need to be held at or below 5 °C for up to 50 days (possibly the equivalent of winter in temperate climates).

Once dormancy is overcome, germination occurs if the following essential external conditions are met (Table 10.5).

- **Water** uptake has occurred so that the seed is fully hydrated and the embryo is enabled to be physiologically active.
- **Oxygen** is present at a high enough partial pressure to sustain aerobic respiration. Growth demands a continuous supply of metabolic energy in the form of ATP that is best generated by aerobic cell respiration in all the cells.

- **A suitable temperature** exists, one that is close to the optimum temperature for the enzymes involved in the mobilisation of stored food reserves, the translocation of organic solutes in the phloem, and the synthesis of intermediates for cell growth and development. For example, wheat seeds germinate in the range 1–35 °C, and maize in the range 5–45 °C.

The **steps of germination** are summarised in Figure 10.31. Note that a plant growth regulator, **gibberellic acid (GA)**, is produced by the cells of the embryo. GA passes to the food stored in the cotyledons. Here protein reserves are converted to **hydrolytic enzymes** which hydrolyse the stored food reserves. The main event is the production of the enzyme **amylase** which hydrolyses starch to maltose. This disaccharide is then hydrolysed to glucose. The resulting soluble sugar (and other compounds) sustain **respiration** and also provide the **building blocks for synthesis** of the intermediates essential for new cells.

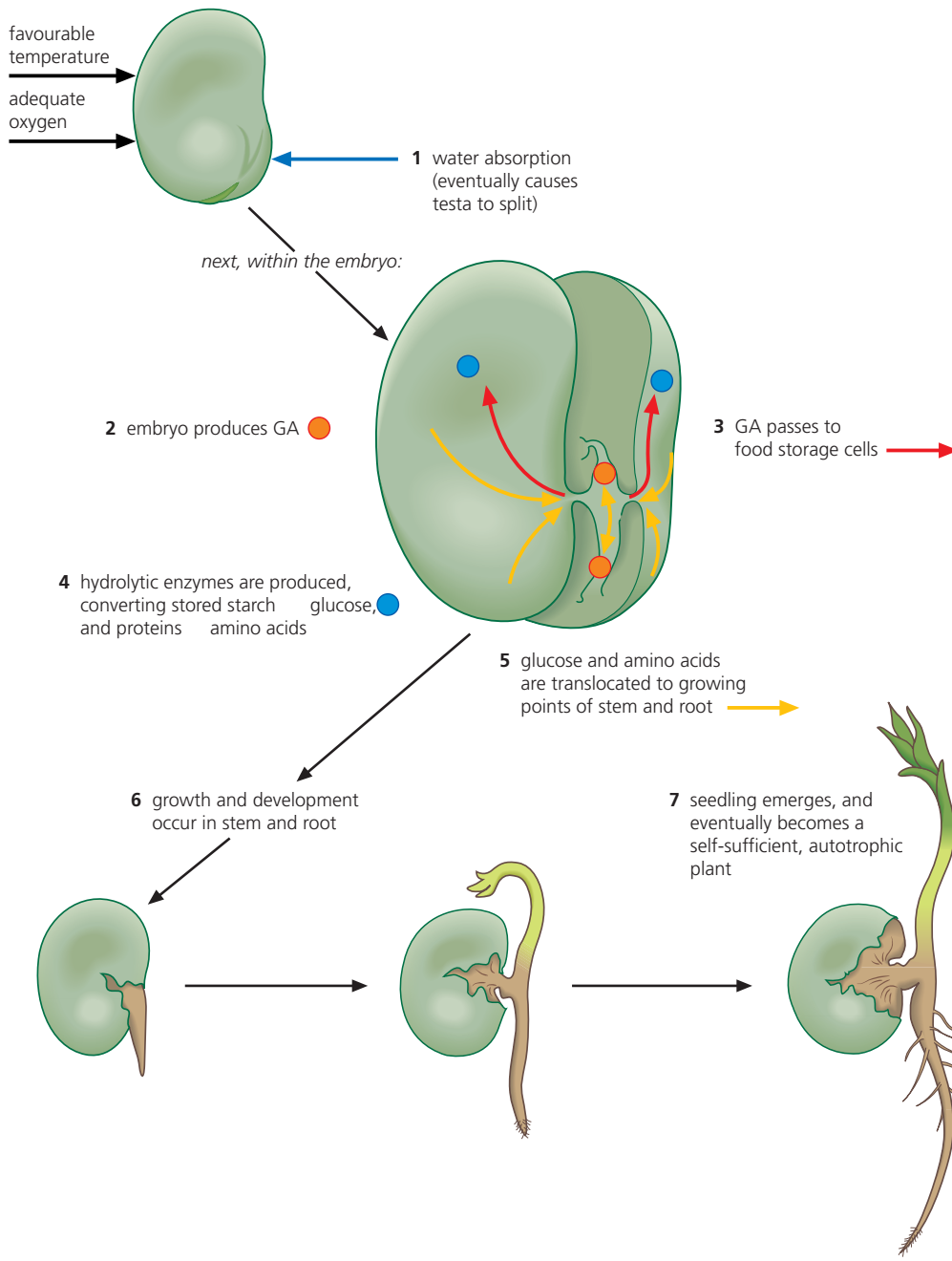


Figure 10.31 Metabolic events of germination in a starchy seed

External	Internal
water uptake – hydration of the cytoplasm of cells of embryo	overcoming of dormancy
ambient **temperature** – within optimum range for enzyme action	production of GA by embryo cells to initiate biochemical changes of germination, leading to production of hydrolytic enzymes for mobilisation of stored food
oxygen – to sustain aerobic cell respiration	

Table 10.5 Conditions for germination

The control of flowering

You will be well aware that plants flower at different times of the year; very many species have a precise season when flowers are produced. At other times, no flowers are formed on these plants.

How is flowering switched on by environmental conditions?

The answer is that day length provides important signals, and that a different plant pigment molecule is involved in the process.

Plant development and phytochrome

A blue–green pigment called **phytochrome** is present in green plants in very low concentrations. The amount of phytochrome is not sufficient to mask chlorophyll, and it has been a substance difficult to isolate and purify from plant tissue, although this has been done.

Phytochrome is a very large conjugated protein (protein molecule + pigment molecule, combined), and it is a highly reactive molecule. It is not a plant growth substance, but it is a photoreceptor pigment, able to absorb light of particular wavelength, and change its structure as a consequence. It is likely to react with different molecules around it, according to its structure.

Two forms of phytochrome

We know that phytochrome exists in two interconvertible forms.

One form, referred to as $\mathbf{P_R}$, is a blue pigment which absorbs mainly red light of wavelength 660 nm (this is what the $_R$ stands for).

The other form is $\mathbf{P_{FR}}$, a blue–green pigment which mainly absorbs far-red light of wavelength 730 nm.

When $\mathbf{P_R}$ is exposed to light (or red light on its own) it is converted to $\mathbf{P_{FR}}$. However, in the dark (or if exposed to far-red light alone) it is converted back to $\mathbf{P_R}$.

$$\mathbf{P_R} \underset{\text{darkness (slow)}}{\overset{\text{light (slow)}}{\rightleftharpoons}} \mathbf{P_{FR}} \qquad \begin{matrix}\text{(or red light) (fast)} \\ \text{(or far-red light) (fast)}\end{matrix}$$

Where plant growth and development are influenced by light, this is known as **photomorphogenesis**. Phytochrome is the pigment system involved in photomorphogenesis. We know this because the red–far-red absorption spectrum of phytochrome corresponds to the action spectrum of some specific effects of light on development (see pages 287–8 if you have forgotten the terms 'absorption spectrum' and 'action spectrum').

It appears that it is $\mathbf{P_{FR}}$ that is the active form of phytochrome, stimulating some effects in plant development and inhibiting others. In particular, $\mathbf{P_{FR}}$ controls of the onset of flowering.

Photoperiodism is the response of an organism to changing length of day. Many plants flower only at a particular time of the year, and plants where flowering is controlled by day length fall into two categories (Figure 10.32).

- **Short-day plants** – these are plants which flower only if the period of darkness is longer than a certain critical length. If darkness is interrupted by a brief flash of red light the plant will not flower, but this is reversed by a flash of far-red light.

Interpretation

Phytochrome in P_{FR} form inhibits flowering in short-day plants. The very long nights required by short-day plants allow the concentration of P_{FR} to fall to a low level, removing the inhibition. A flash of light in the darkness reverses this, but a flash of far-red light reverses the reversal, and flowering still takes place.

- **Long-day plants** – these are plants which flower only if the period of uninterrupted darkness is less than a certain critical length each day.

Interpretation

Phytochrome in P_{FR} form promotes flowering in long-day plants. The long period of daylight causes the accumulation of P_{FR}, because P_R is converted to P_{FR}.

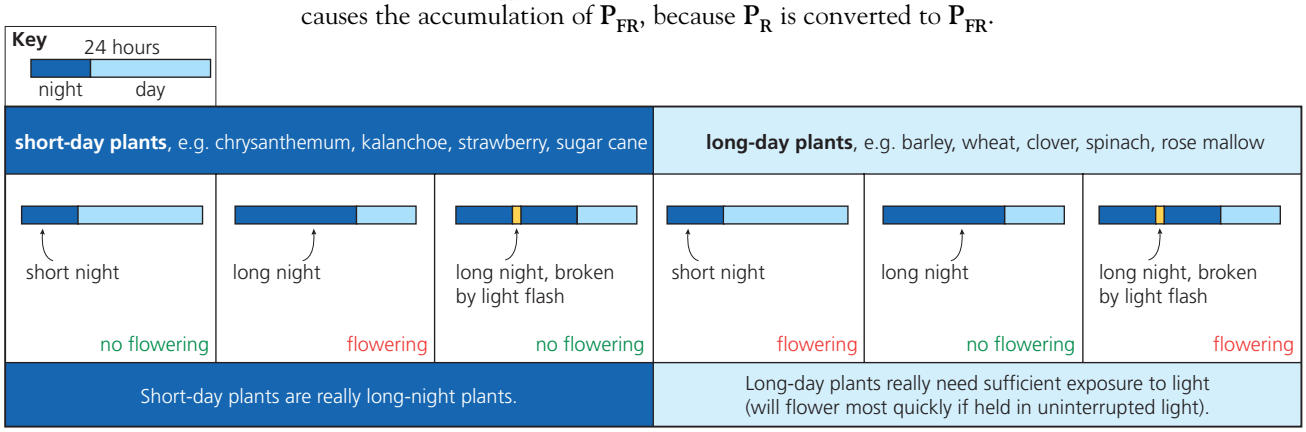


Figure 10.32 Flowering related to day length

Extension: The conversion of vegetative buds to flower buds

It is the leaves of plants that are sensitive to day length, yet the structural switch to flowering occurs in a stem apex. It has been assumed that a growth regulator substance is formed in leaves under the correct regime of light and dark, and is then transported to the stem apex where it causes the switch in development. For example, a leaf that has been exposed to the correct photoperiod, if immediately grafted onto another plant of the same type, will cause flowering there.

A hormone is believed to exist, and has been named **florigen**, but it has not been isolated. Since it is P_{FR} that is enzymically active, this substance might cause the formation of an enzyme which promotes the formation of florigen in long-day plants but inhibits its formation in short-day plants. Much remains to be discovered about the control of flowering (Figure 10.33).

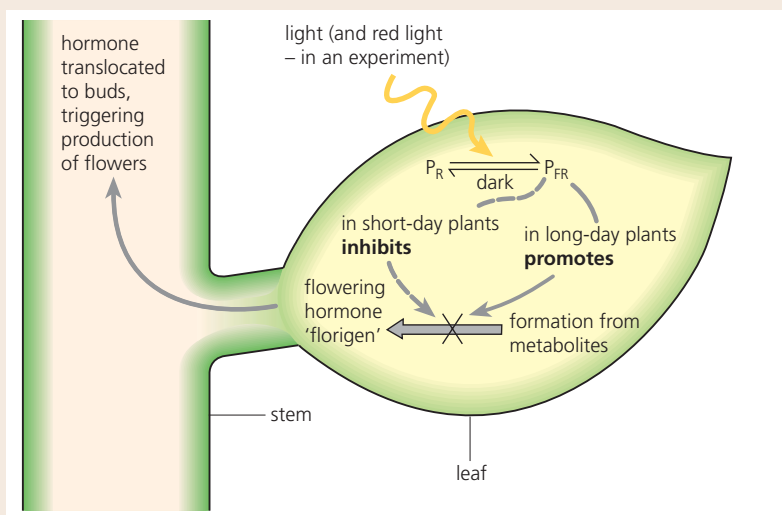


Figure 10.33 Phytochrome and flowering, a suggested hypothesis

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 Which of the following would be an adaptation made by a xerophyte plant?

A reduced root surface area
B increased air space
C increased number of stomata
D a thicker cuticle

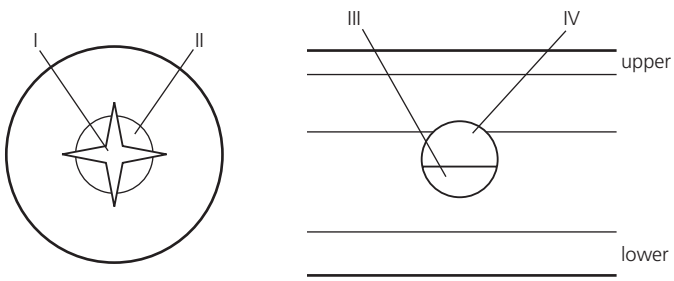
Higher Level Paper 1, May 05, Q39

Q2 What are the differences between the transport of minerals in the apoplast and the symplast of the root?

	Apoplast	Symplast
A	they travel through the cytoplasm of the root cells	they travel through the intercellular spaces
B	their movement is blocked by the endodermis	their movement continues across the endodermis
C	they enter by active transport	they enter by passive transport
D	they cross between cells by the plasmodesmata	they cannot travel via the plasmodesmata

Higher Level Paper 1, May 04, Q40

Q3 The diagrams show the distribution of tissues in the root and in the leaf of a dicotyledonous plant.



Which tissues are phloem?

A I and III only
B I and IV only
C II and III only
D II and IV only

Higher Level Paper 2, May 03, Q38

Q4 What conditions are needed for the germination of all seeds?

I light **II** sufficient water **III** oxygen

A I and II only
B I and III only
C II and III only
D I, II and III

Higher Level Paper 1, May 02, Q40

Questions 5–10 cover other syllabus issues in this chapter.

Q5 The leaf is the site of the bulk of photosynthesis in the flowering plant. Outline, by means of concise notes, **five** structural features of leaves that specifically favour photosynthesis and the ways they may enhance this process. (5 + 5)

Q6 *"Transpiration is an unfortunate consequence of plant structure and plant nutrition; it cannot be described as a purposeful process!"*
Suggest how you may support and justify this statement to a sceptical biologist. (6)

Q7 **a** Explain what *meristems* are. (1)
b Identify the key features of meristematic plant cells. (3)
c Explain the difference between an apical and a lateral meristem by reference to the growth they make possible. (4)

Q8 **a** Define *translocation* and *transpiration*. (4)
b Identify the sources of energy for movement of substances in:
i translocation
ii transpiration. (4)
c By means of a fully annotated drawing, describe the structure of phloem tissue. (6)
d List the structural features of phloem tissue that are not shown by xylem tissue. (3)
e Outline the essential features of the mass flow hypothesis of phloem transport. (6)

Q9 **a** Draw a fully labelled half-flower diagram of an insect-pollinated flower you have studied. (6)
b State what insects may visit this flower. (1)
c Identify the features of this flower that may attract insects. (3)
d Explain how pollination is brought about in this flower. (4)

Q10 **a** List the ways guard cells of stomata differ from the ordinary epidermal cells around them. (3)
b Explain why stomatal pores close as leaves wilt, irrespective of whether the leaf is in the light or dark. (2)
c Outline the ideal conditions for stomata to be fully open. (4)

11 Genetics II

STARTING POINTS

- In **diploid organisms** chromosomes occur in pairs, called **homologous chromosomes**, one coming originally from each parent. A gene occupies a position on a chromosome, its **locus**. Genes at a locus exist in two or more forms, called **alleles**.
- **Meiosis**, the nuclear division in which the new cells formed have **half the chromosome number** of the parent cell, is associated with sexual reproduction and the production of gametes.
- Gregor Mendel's **monohybrid cross** established that characteristics may be controlled by a pair of alleles that **segregate** in equal numbers into different gametes. Segregation occurs in meiosis.
- Genes on the same chromosome are **linked** and may be inherited together. Characteristics controlled by genes on the **sex chromosomes** are sex-linked characteristics.
- **Mutations** (abrupt changes) may arise in genes or chromosomes.
- This chapter extends aspects of the study of genetics begun in Chapter 4, pages 91–116.

The painstaking experimental work on inheritance by **Gregor Mendel** (Figure 4.13, page 103), although overlooked in his life-time, supplied the foundations for **modern genetics**, once it was rediscovered, read and understood.

Heredity is responsible for many of the similarities and differences (**variations**) between parents and offspring. We know that many variations that are inherited may be discrete or **discontinuous variations**, such as we saw in tall or dwarf pea plants (page 104). Alternatively, other characteristics of organisms show continuous variation over a range (such as height in humans). These are known as **continuous variations**.

In this chapter, **chromosome behaviour in meiosis** and the causes of **genetic variation in gametes** are considered first. Mendel's investigation of the **dihybrid cross** and the **Law of Independent Assortment** are explained, and their relation to meiosis established. Then, gene linkage, **crossing over** and the formation of **recombinants** are discussed. Finally, how continuous variation may also be genetically controlled in **polygenic inheritance** is demonstrated.

Meiosis and genetic variety in gametes

10.1.1–10.1.3

Meiosis is an essential event in life cycles that include sexual reproduction, because at fertilisation the chromosome number is doubled.

The events in meiosis have already been established in Chapter 4. Meiosis is a nuclear division that is slower and more complex than mitosis, for it involves two successive divisions of the nucleus (**meiosis I** and **meiosis II**), but both of these divisions superficially resemble mitosis.

We have already seen that in meiosis I the **homologous chromosomes separate**, and that in meiosis II the **chromatids separate**.

Take a look at these points, shown in Figure 4.5 on page 95, now.

Equally important is the fact that there is no interphase between meiosis I and II, so no replication of the chromosomes occurs during meiosis. (Remember, chromosomes replicate to form chromatids during interphase, well before nuclear divisions occur.)

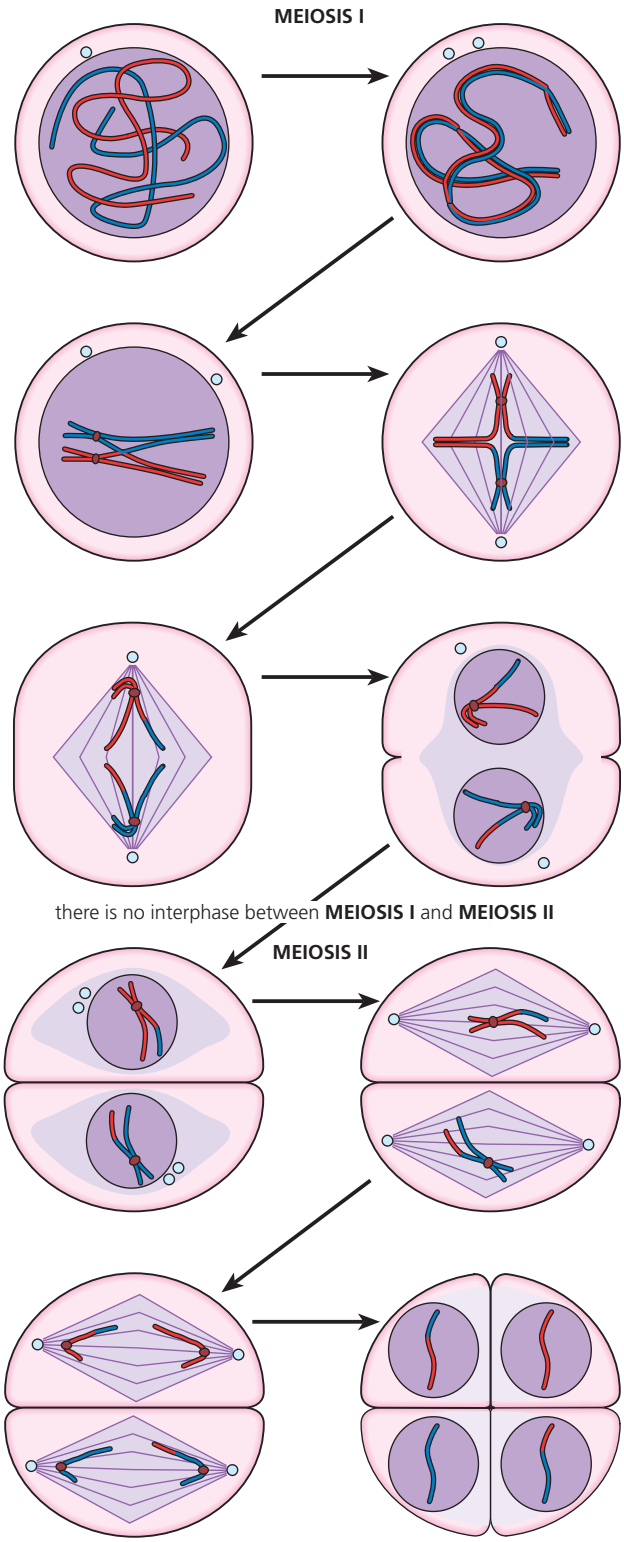
Meiosis consists of two nuclear divisions but only one replication of the chromosomes.

Figure 11.1 What happens in meiosis

prophase I (early)
During interphase the chromosomes replicate into chromatids held together by a centromere (the chromatids are not visible). Now the chromosomes condense (shorten and thicken) and become visible.

prophase I (late)
Homologous chromosomes repel each other. Chromosomes can now be seen to consist of chromatids. Sites where chromatids have broken and rejoined, causing crossing over, are visible as chiasmata.

anaphase I
Homologous chromosomes separate. Whole chromosomes are pulled towards opposite poles of the spindle, centromere first (dragging along the chromatids).



prophase I (mid)
Homologous chromosomes pair up (becoming **bivalents**) as they continue to shorten and thicken. Centrioles duplicate.

metaphase I
Nuclear membrane breaks down. Spindle forms. Bivalents line up at the equator, attached by centromeres.

telophase I
Nuclear membrane re-forms around the daughter nuclei. The chromosome number has been halved. The chromosomes start to decondense.

prophase II
The chromosomes condense and the centrioles duplicate.

anaphase II
The chromatids separate at their centromeres and are pulled to opposite poles of the spindle.

metaphase II
The nuclear membrane breaks down and the spindle forms. The chromosomes attach by their centromere to spindle fibres at the equator of the spindle.

telophase II
The chromatids (now called chromosomes) decondense. The nuclear membrane re-forms. The cells divide.

The process of meiosis

Once started, meiosis proceeds steadily as a continuous process of nuclear division. The steps of meiosis are explained in four distinct phases (**prophase**, **metaphase**, **anaphase** and **telophase**), but this is just for convenience of analysis and description – there are no breaks between the phases in nuclear divisions.

The behaviour of the chromosomes in the phases of meiosis is shown in Figure 11.1. For clarity, the drawings show a cell with a single pair of homologous chromosomes.

Meiosis I – prophase I

What happens to chromosomes during prophase I is especially complex. They appear in the light microscope as single threads with many tiny bead-like thickenings along their lengths. These thickenings represent an early stage in the process of shortening and thickening by coiling that continues throughout prophase. This packaging of DNA in the chromosome is shown in Figure 8.2, page 236. Of course, each chromosome is already replicated as two chromatids but individual chromatids are not visible as yet.

Formation of bivalents

As the chromosomes continue to thicken, homologous chromosomes are seen to come together in specific pairs, point by point, all along their length. The product of pairing is called a **bivalent**. Remember, in a diploid cell each chromosome has a partner that is the same length and shape and with the same linear sequence of alleles.

The homologous chromosomes of the bivalents continue to shorten and thicken. Later in prophase, the individual chromosomes can be seen to be double-stranded, as the sister chromatids of which each consists become visible.

Crossing over

Within the bivalent, during the coiling and shortening process, breakages of the chromatids occur frequently. Breakages are common in non-sister chromatids at the same points along the length of each. Broken ends rejoin more or less immediately, but where the rejoins are between non-sister chromatids, swapping of pieces of the chromatids has occurred, hence the term 'crossing over'.

The point of join between different chromatids is called a **chiasma** (plural, **chiasmata**). Virtually every pair of homologous chromosomes forms at least one chiasma at this time, and to have two or more chiasmata in the same bivalent is very common (Figures 11.2 and 11.3).

In the later stage of prophase I, the attraction and tight pairing of the homologous chromosomes ends, but the attraction between sister chromatids remains for the moment. This attraction of sister chromatids keeps the bivalents together. The chromatids are now at their shortest and thickest.

Later still, the centrioles present in animal cells (page 7) duplicate, and start to move apart as a prelude to spindle formation. Plant cells are without a centriole.

Finally, the disappearance of the nucleoli and nuclear membrane marks the end of prophase I.

Metaphase I

Next, the spindle forms and the bivalents become attached to individual microtubules of the spindle by their centromeres. The bivalents are now arranged at the equatorial plate of the spindle framework – we say that they line up at the centre of the cell. By the end of metaphase I, the members of the bivalents start to repel each other and separate. However, at this point they are held together by one or more chiasmata, and this gives temporary but unusual shapes to the bivalents.

Anaphase I

The homologous chromosomes of each bivalent now move to opposite poles of the spindle, but with the individual chromatids remaining attached by their centromeres. The attraction of sister chromatids has lapsed, and they separate slightly – both are clearly visible. However, they do not separate yet, but go to the same pole. Consequently, meiosis I has separated homologous pairs of chromosomes but not the sister chromatids of which each is composed.

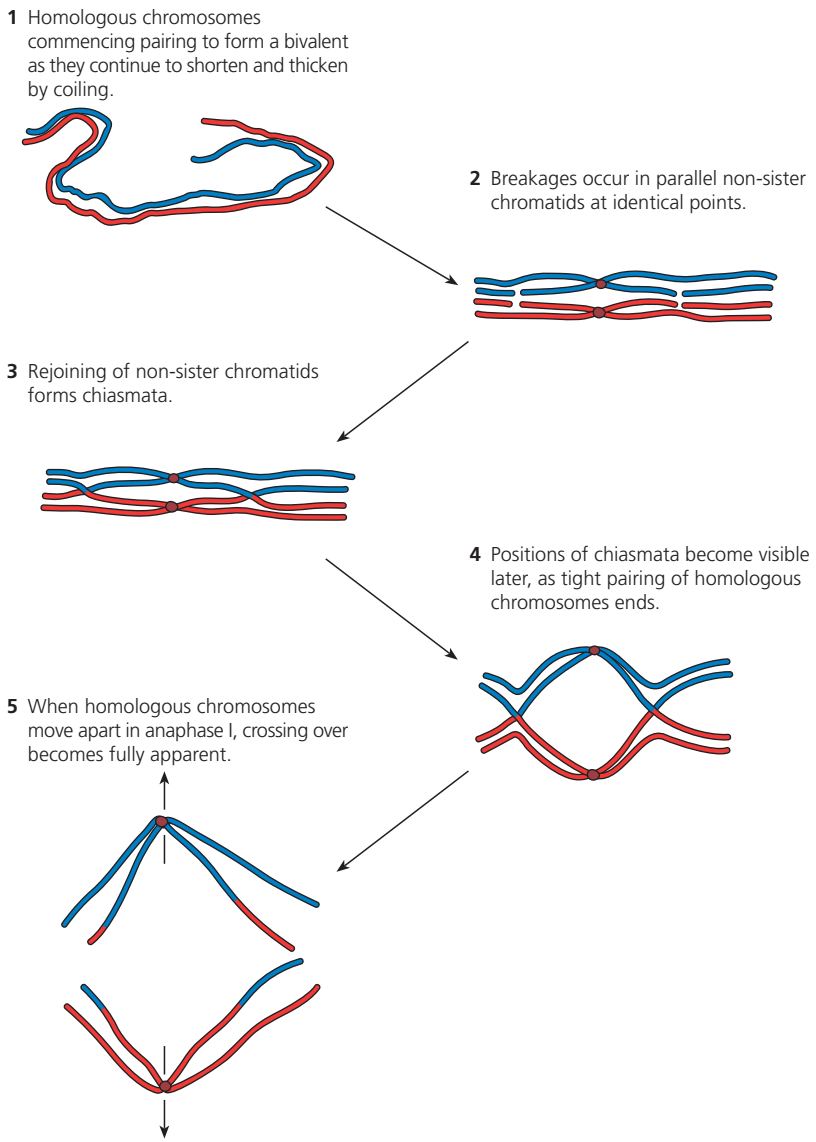


Figure 11.2 Formation of chiasmata

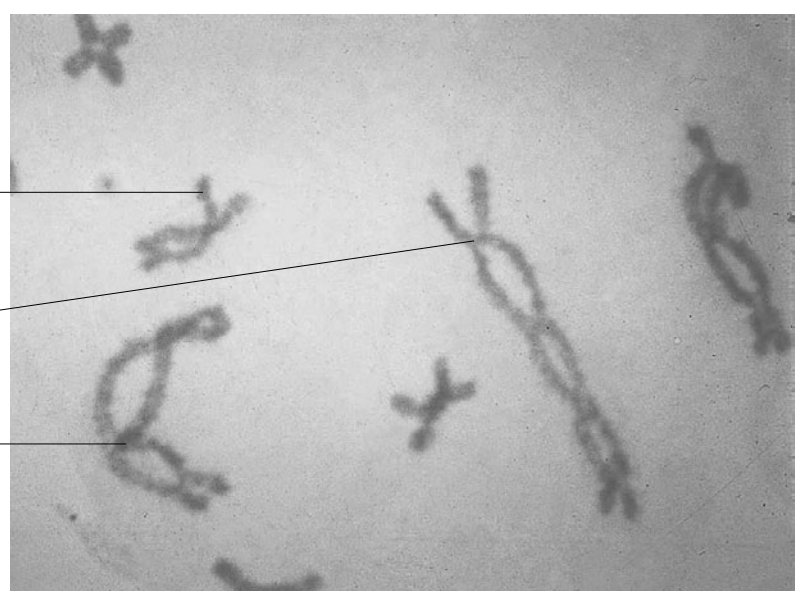


Figure 11.3 Photomicrograph of bivalents held together by chiasmata

Telophase I

The arrival of homologous chromosomes at opposite poles signals the end of meiosis I. The chromosomes tend to uncoil to some extent, and a nuclear membrane re-forms around both nuclei. The spindle breaks down. However, these two cells do not go into interphase, but rather continue into meiosis II, which takes place at right angles to meiosis I. Meiosis II is remarkably similar to mitosis.

Meiosis II – prophase II

The nuclear membranes break down again, and the chromosomes shorten and re-thicken by coiling. Centrioles, if present, move to opposite poles of the cell. By the end of prophase II the spindle apparatus has re-formed, but is present at right angles to the original spindle.

Metaphase II

The chromosomes line up at the equator of the spindle, attached by their centromeres.

Anaphase II

The centromeres divide and the chromatids are pulled to opposite poles of the spindle, centromeres first.

Telophase II

Nuclear membranes form around the four groups of chromatids, so that four nuclei are formed. Now there are four cells, each with half the chromosome number of the original parent cell. Finally, the chromatids – now recognised as chromosomes – uncoil and become apparently dispersed as chromatin. Nucleoli re-form.

The process of meiosis is now complete, and is followed by division of the cells (cytokinesis, page 33).

Meiosis and genetic variation

The variation in the genetic information carried by different gametes that arises in meiosis is highly significant for the organism, as we shall see. The four haploid cells produced by meiosis differ genetically from each other because of:

- **Independent assortment of maternal and paternal homologous chromosomes**
  The way the bivalents line up at the equator of the spindle in meiosis I is entirely random. Which chromosome of a given pair goes to which pole is unaffected by (independent of) the behaviour of the chromosomes in other pairs. This was introduced in Figure 4.7, but it is represented again here in terms of the critical steps in meiosis where it occurs (Figure 11.4). These illustrations show a parent cell with only four chromosomes, for clarity. Of course, the more bivalents there are in the nucleus, the more variation is possible. In humans, there are 23 pairs of chromosomes, so the number of possible combinations of chromosomes that can be formed as a result of independent assortment is 2^{23}. This is over 8 million.
- **Crossing over of segments of individual maternal and paternal homologous chromosomes**
  This results in new combinations of genes on the chromosomes of the haploid cells produced, as illustrated in Figure 11.5. Crossing over generates the possibility of an almost unimaginable degree of variation. For example, if we were to assume for sake of discussion that there are 30 000 individual genes on the human chromosome complement, all with at least two alternative alleles, and that crossing over was equally likely between all of these genes, then there would be $2^{30\,000}$ different combinations. Of course, all these assumptions would be inaccurate to varying extents, but the point that virtually unlimited recombinations are possible is established.

Figure 11.4 Genetic variation due to independent assortment

Independent assortment is illustrated in a cell with two pairs of chromosomes.

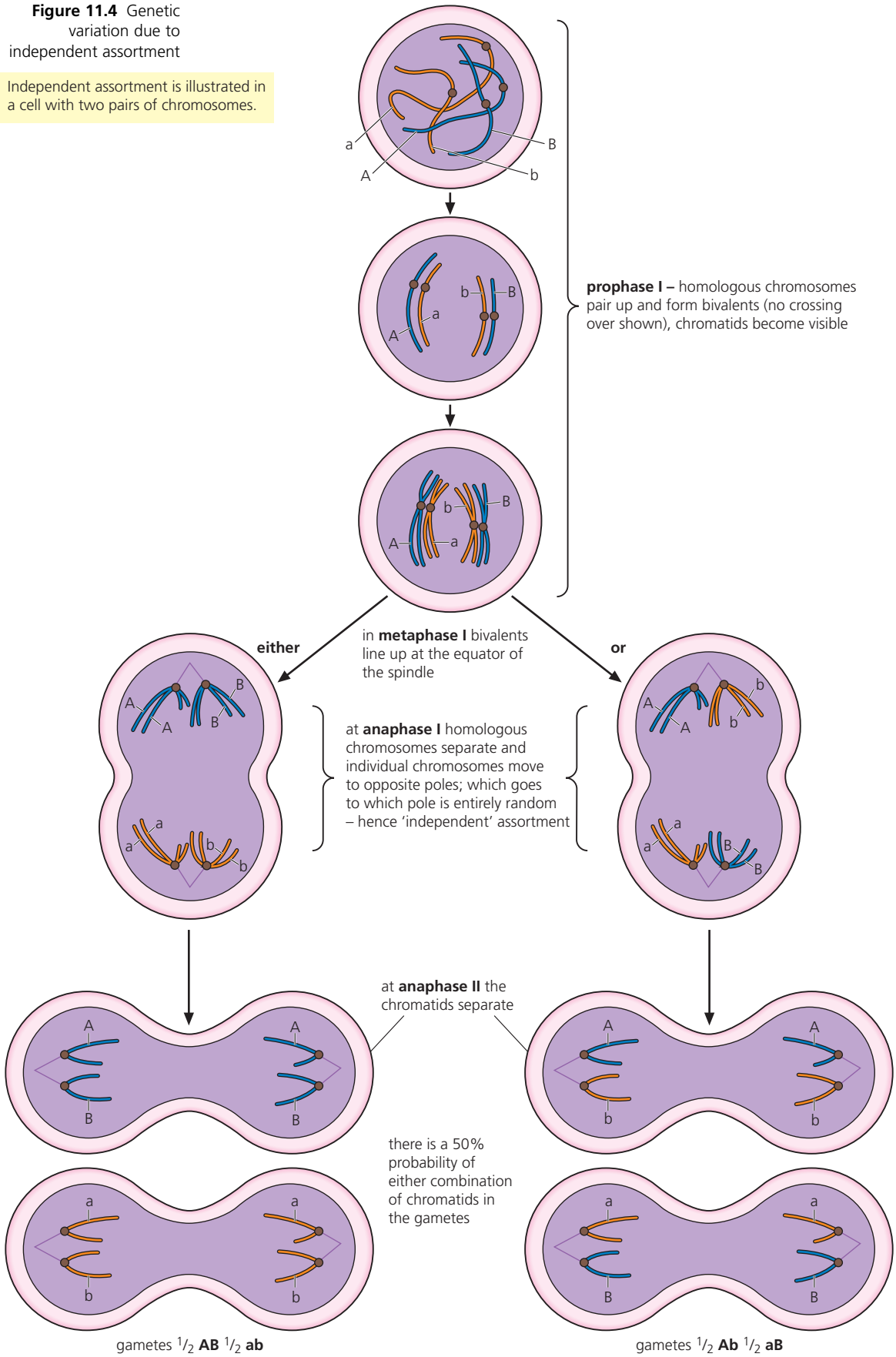
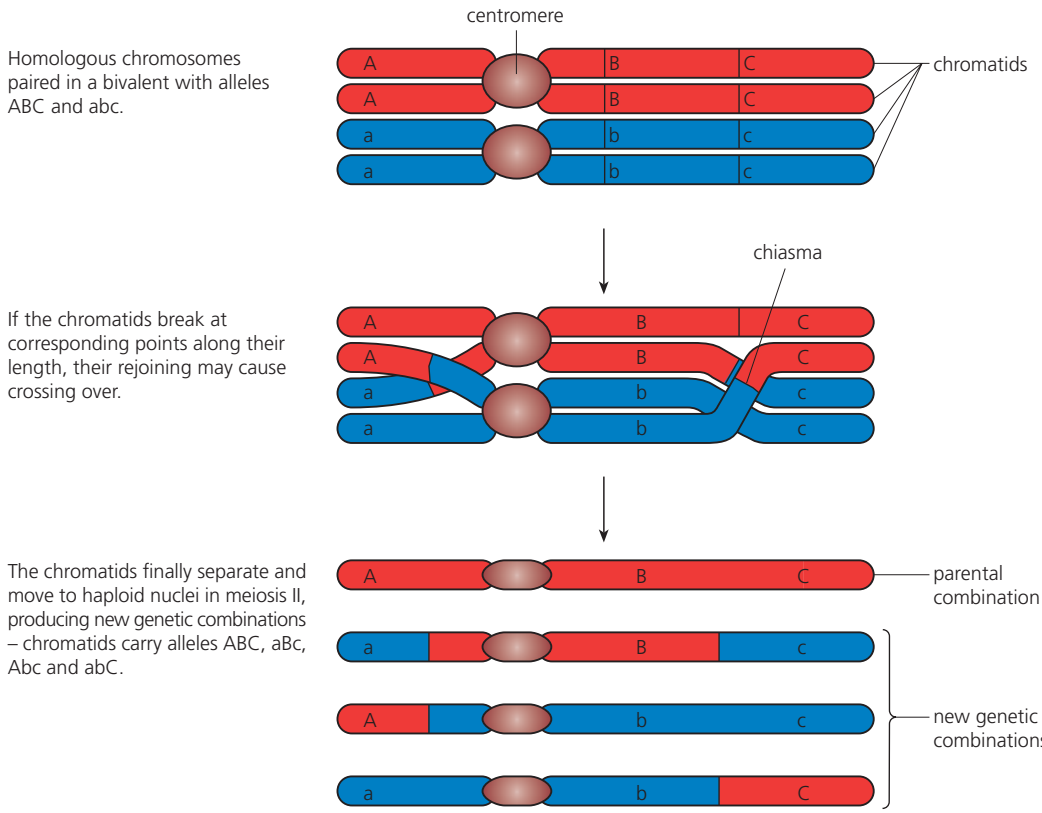


Figure 11.5 Genetic variation due to crossing over

The effects of chiasmata on genetic variation are illustrated in one pair of homologous chromosomes. Typically, two or three chiasmata form between the chromatids of a bivalent in prophase I.



A note on recombinants

Offspring with new combinations of characteristics different from those of their parents are called **recombinants**.

Recombination in genetics is the re-assortment of alleles or characters into different combinations from those of the parents. We have seen that recombination occurs for genes located on separate chromosomes (unlinked genes) by chromosome assortment in meiosis (Figure 11.4), and for genes on the same chromosomes (linked genes) by crossing over during meiosis (Figure 11.5).

1 **Distinguish** the essential differences between mitosis and meiosis.

Dihybrid crosses and gene linkage 10.1.4–10.1.5, 10.2.1–10.2.6

The contribution to our understanding of the mechanism of inheritance made by Gregor Mendel was introduced on page 103, by discussion of his investigation of the inheritance of a single pair of contrasting characteristics, namely of tall and dwarf pea plants. This was his so-called **monohybrid cross**.

Mendel also investigated the simultaneous inheritance of **two pairs of contrasting characters**, again using the garden pea plant. This he referred to as a **dihybrid cross**.

For example: Mendel crossed (P generation) pure-breeding pea plants from **round seeds** with **yellow cotyledons** (seed leaves) with pure-breeding plants from **wrinkled seeds** with **green cotyledons**. All the progeny (F_1 generation) were **round, yellow peas**.

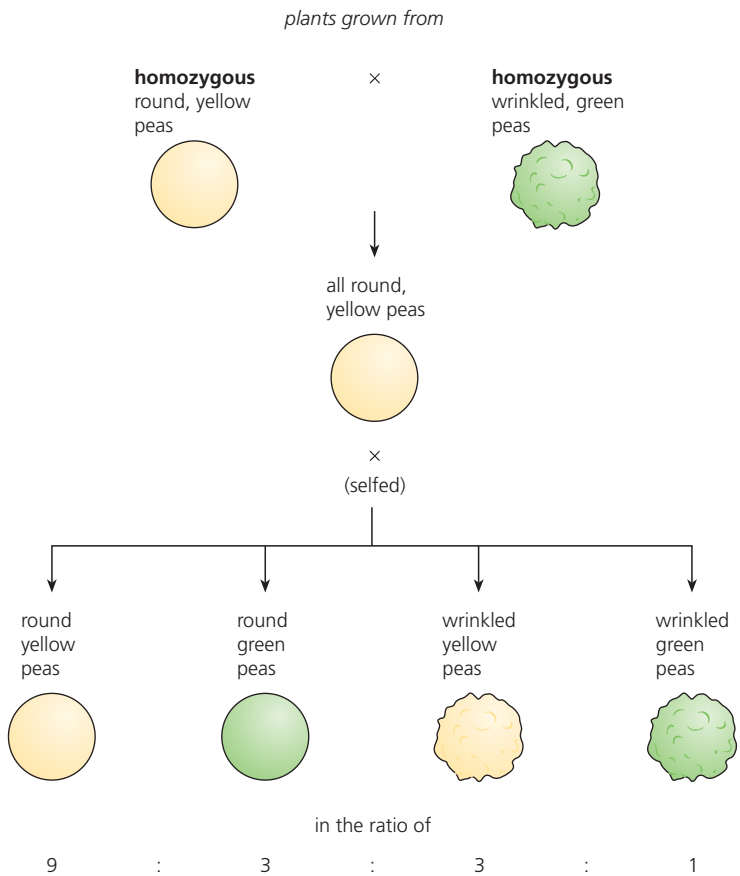
When plants grown from these seeds were allowed to self-fertilise the following season, the resulting seeds (F_2 generation) – of which there were more than 500 to be classified and counted – were of the following four phenotypes, and were present in the ratio shown below:

Phenotypes	round seed with yellow cotyledons	round seed with green cotyledons	wrinkled seed with yellow cotyledons	wrinkled seed with green cotyledons
Ratio	**9**	**3**	**3**	**1**

Mendel noticed that two new combinations, not represented in the parents (i.e. **recombinations**) appeared in the progeny: both round and wrinkled seeds can turn up with either green or yellow cotyledons.

Thus, the two pairs of factors were inherited independently. He had noticed that either one of a pair of contrasting characters could be passed to the next generation. This meant that a heterozygous plant must produce four types of gametes in equal numbers (Figures 11.6 and 11.7).

Figure 11.6 Mendel's dihybrid cross



2 In Figure 11.6, **identify** the progeny that are:

a heterozygous
b recombinants.

3 **Deduce** the positions of the genes for round/wrinkled and for yellow/green cotyledons within the nucleus of the pea plant. Explain the significance of their location (which was unknown to Mendel).

Mendel confirmed the results of his first dihybrid cross by carrying out experiments with peas showing other pairs of contrasting characteristics (and also with other species of plant).

He published his results through his local natural history and agricultural research society, and by regular correspondence with certain influential scientists of the time. However, we cannot be confident that anyone apart from Mendel himself saw the significance of his discoveries in his life-time.

parental (P)

phenotypes: homozygous round and yellow × homozygous wrinkled and green

genotypes: **RRYY** × **rryy** ← pea plants are diploid so they have two copies of each allele

(meiosis) (meiosis)

gametes: **RY** **ry** ← gametes produced by meiosis so only have one copy of each allele

offspring (F_1)

genotypes: **RrYy** ← F_1 progeny are heterozygous for both genes

phenotypes: heterozygous round and yellow ← the alleles for 'round' and 'yellow' are dominants; the F_1 progeny are all round yellow peas

F_1 selfed **RrYy** × **RrYy** ← genes are on separate chromosomes and random assortment occurs i.e. **R** may assort with **Y** or **y** **r** may assort with **Y** or **y** so four different types of gametes (both ♂ and ♀) are formed, in equal proportions (**independent assortment**)

(meiosis) (meiosis)

gametes: **RY** **Ry** **rY** **ry** | **RY** **Ry** **rY** **ry**

	¼**RY**	¼**Ry**	¼**rY**	¼**ry**
¼**RY**	1/16**RRYY** round yellow	1/16**RRYy** round yellow	1/16**RrYY** round yellow	1/16**RrYy** round yellow
¼**Ry**	1/16**RRYy** round yellow	1/16**RRyy** round green	1/16**RrYy** round yellow	1/16**Rryy** round green
¼**rY**	1/16**RrYY** round yellow	1/16**RrYy** round yellow	1/16**rrYY** wrinkled yellow	1/16**rrYy** wrinkled yellow
¼**ry**	1/16**RrYy** round yellow	1/16**Rryy** round green	1/16**rrYy** wrinkled yellow	1/16**rryy** wrinkled green

← random fertilisation between all four gametes produces 9 : 3 : 3 : 1 ratio of offspring given many crosses occur

offspring (F_2)

	round yellow	round green	wrinkled yellow	wrinkled green
phenotypes:				
phenotypes ratio:	9 :	3 :	3 :	1 :
genotypes:	1 **RRYY** 2 **RrYY** 2 **RRYy** 4 **RrYy**	1 **RRyy** 2 **Rryy**	1 **rrYY** 2 **rrYy**	1 **rryy**

Figure 11.7 Genetic diagram showing the behaviour of alleles in Mendel's dihybrid cross

Mendel did not express the outcome of the dihybrid cross as a succinct Second Law. However, today we call Mendel's Second Law the **Law of Independent Assortment**. It is stated as:

Two or more pairs of alleles segregate independently of each other as a result of meiosis, provided the genes concerned are not linked by being on the same chromosome.

The relationship between Mendel's Law of Independent Assortment and meiosis is detailed in Table 11.1.

Mendel's dihybrid cross	Feature of meiosis
Within an organism there are breeding factors controlling characteristics like round or wrinkled seeds, and yellow or green cotyledons.	Each chromosome holds a linear sequence of genes.
These factors remain intact from generation to generation.	A particular gene always occurs on the same chromosome in the same position after each nuclear division.
There are two factors for each characteristic in each cell.	The chromosomes of a cell occur in pairs, called homologous pairs.
One factor comes from each parent. (A recessive factor is not expressed in the presence of a dominant factor.)	One of each pair came originally from one parent and the other from the other parent.
Factors separate in reproduction; either can be passed to an offspring. Only one of the factors can be in any gamete.	At the end of meiosis, each cell (gamete) contains a single member of each of the homologous pairs of chromosomes present in the parent cell.
The factors for seed shape and seed colour segregate independently of each other as a result of meiosis.	The genes for seed shape and seed colour are on separate chromosomes.
The 9:3:3:1 ratio shows that all four types of gamete are equally common. The inheritance of the two characteristics is separate.	The arrangement of bivalents at the equatorial plate of the spindle is random; maternal and paternal homologous chromosomes are independently assorted. In a large number of matings, all possible combinations of chromosomes will occur in equal numbers.

Table 11.1 How the Law of Independent Assortment relates to meiosis

The dihybrid test cross

Look back to Chapter 4, page 106, to remind yourself of the issue that the **monohybrid test cross** sorts out.

Why is a test cross sometimes necessary?

In the case of dihybrid inheritance, too, while homozygous recessive genotypes such as wrinkled green peas (**rryy**) can be recognised in the phenotype, homozygous round yellow peas (**RRYY**) and heterozygous round yellow peas (**RrYy**) look exactly the same (Figure 11.7). They can only be distinguished by the progeny they produce, as illustrated in a dihybrid test cross (Figure 11.8).

Unlike the monohybrid test cross, where the outcome is a ratio of 1:1, the dihybrid test cross outcome is a ratio of 1:1:1:1. We will return to this point later.

TOK Link

Given the absence of understanding of chromosomes, genes and meiosis at the time, is it appropriate to describe Mendel's scientific contributions as 'nothing but trained and organised common sense'?

Drosophila and the dihybrid cross

Mendel died in 1884 – in relative obscurity, at least as far as the scientific community was concerned. Few seemed to know of his scientific work and those who were aware of it did not appear to understand his results. But his papers were rediscovered in 1900, as a result of careful literature searches by people keen to advance our understanding of inheritance. Mendel's results came to be confirmed and then extended by many others, using a range of species.

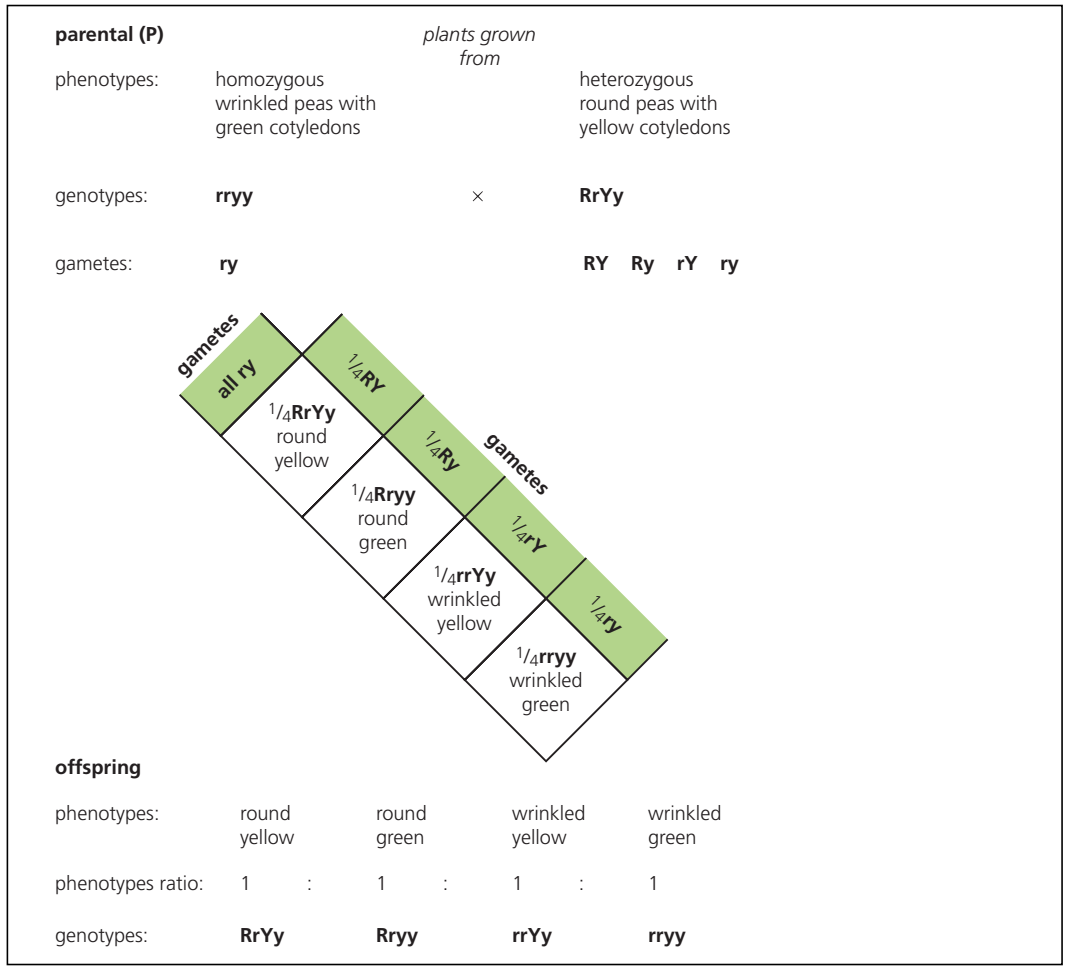


Figure 11.8 Genetic diagram of the dihybrid test cross

Drosophila melanogaster (the fruit fly) was first selected by an American geneticist called **Thomas Morgan** in 1908 as an experimental organism for investigation of Mendelian genetics in an animal. Morgan was awarded a Nobel Prize in 1933 because by his experiments he had:

- shown that Mendel's 'factors' are linear sequences of genes on chromosomes (what is now called the **Chromosome Theory of Inheritance**);
- discovered sex chromosomes and sex linkage;
- demonstrated crossing over, and the exchange of alleles between chromosomes, resulting from chiasmata formed during meiosis.

Drosophila commonly occurs around rotting vegetable material. This form is often called the wild type, simply to differentiate it from the various naturally occurring mutant forms.

Why has this insect become such a useful experimental animal in the study of genetics?

Drosophila has four pairs of chromosomes (Figure 11.9). From mating to emergence of adult flies (generation time) takes about 10 days at 25°C. One female produces hundreds of offspring. These flies are relatively easily handled, cultured on sterilised artificial medium in glass bottles, and they can be temporarily anaesthetised for setting up cultures and sorting progeny.

4 **Define** what is meant by the term 'mutant'.

A dihybrid cross can be shown in *Drosophila* – for example, by crossing normal flies (wild type) with flies homozygous for vestigial wing and ebony body (Figure 11.10). These characteristics are controlled by genes not on the sex chromosomes. All chromosomes other than the sex chromosomes are called **autosomal chromosomes** (see below).

The genes concerned in this *Drosophila* cross are on separate autosomal chromosomes, so they are unlinked genes. SAQ 5 on page 338 sets the task of determining the genotypes of offspring of the F_2 generation of this dihybrid cross in *Drosophila*, summarised in Figure 11.10.

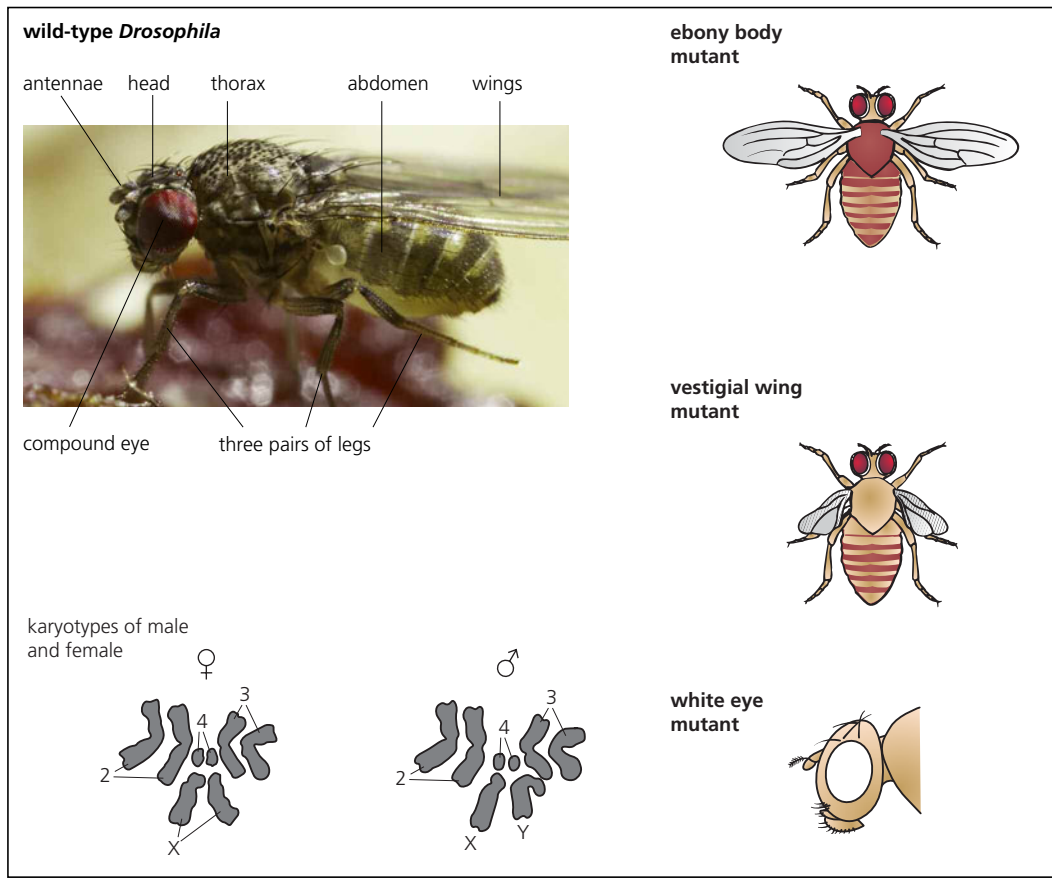


Figure 11.9 Wild-type *Drosophila* (photograph) and some common mutants (drawings)

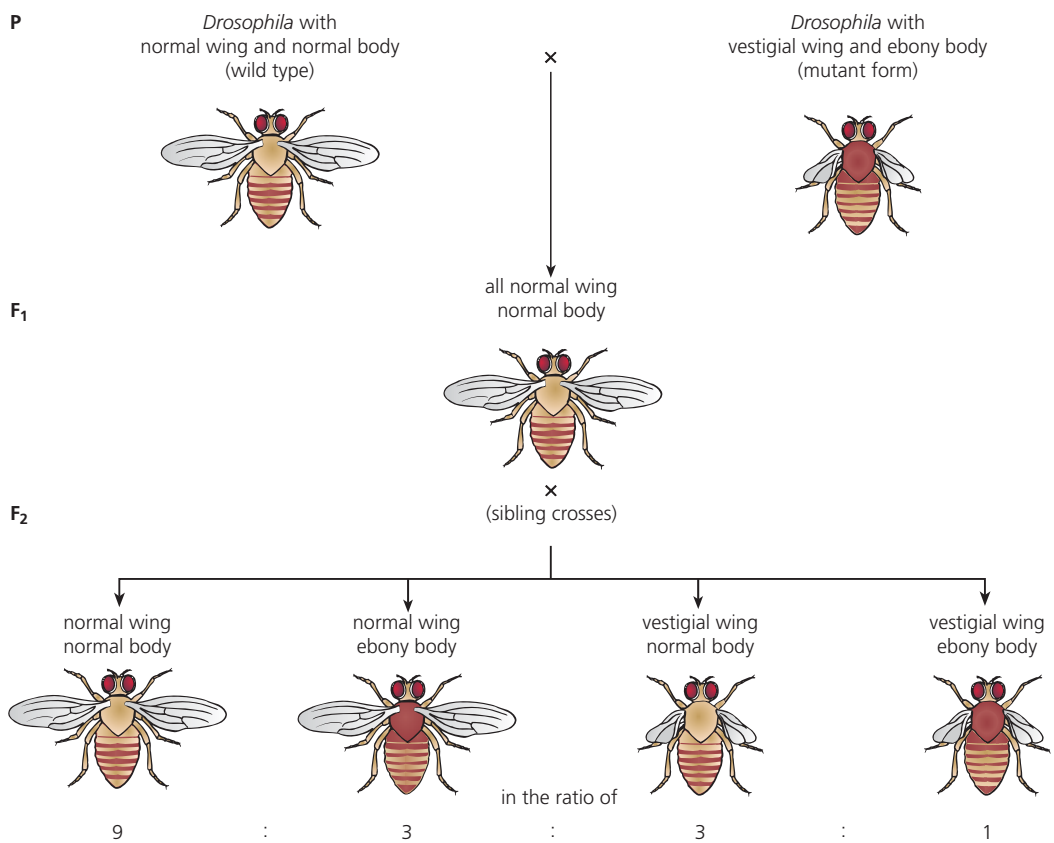


Figure 11.10 A dihybrid cross in *Drosophila* – a summary

5 **Construct** a genetic diagram for the dihybrid cross shown in Figure 11.10, using the layout as given in Figure 11.7. **Determine** the genotypes of the offspring of the F_2 generation.

Extension: Probability and chance in genetic crosses

We can see that there is an expected ratio of offspring of 9:3:3:1 when the dihybrid cross is carried out. Actually, the offspring produced in many dihybrid cross experiments do not exactly agree with the expected ratio. This is illustrated by the results of an experiment with mutant forms of *Drosophila* in Table 11.2.

Offspring in F_2 generation	normal wing, grey body = 315	normal wing, ebony body = 108	vestigial wing, grey body = 101	vestigial wing, ebony body = 32	total = 556
Predicted ratio	9	3	3	1	
Expected numbers of offspring	313	104	104	35	

Table 11.2 Observed and expected offspring

Clearly, these results are fairly close to the ratio 9:3:3:1, but they are not the predicted ratio, precisely.

What, if anything, went wrong?

Well, we can expect this ratio among the progeny only if three conditions are met:

- fertilisation is entirely random;
- there are equal opportunities for survival among the offspring;
- very large numbers of offspring are produced.

In the above experiment with *Drosophila*, the exact ratio may not be obtained because, for example:

- more male flies of one type may have succeeded in fertilising females than of the other type;
- more females of one type may have died before reaching egg laying condition than of the other type;
- fewer eggs of one type may have completed their development than of the other type.

Similarly, in breeding experiments with plants such as the pea plant, exact ratios may not be obtained because of parasite damage or by the action of browsing predators on the anthers or ovaries in some flowers, or because some pollen types fail to be transported by pollinating insects as successfully as others, perhaps.

Gene linkage

In humans, the nucleus contains 46 chromosomes, although we have seen that these are only clearly visible as such during the nuclear divisions, mitosis or meiosis. In fact, it is particularly in metaphases of these divisions, when the chromosomes are at their shortest, and are arranged at the plate of the spindle, that it is particularly easy to see them. It is from photomicrographs of this stage that **karyotypes** are constructed (Figure 4.1, page 92).

A human karyotype shows us that one pair of our chromosomes are the **sex chromosomes**. These carry the critical gene that determines sex. The remainder of our set of chromosomes carry genes that determine characteristics of the growth, development and functioning of all other aspects of the body. These chromosomes are known as the autosomes (Figure 11.11).

Linkage groups

There are many thousands of genes per cell in an organism, whereas the number of chromosomes is often less than 50 and very rarely exceeds 100. Each chromosome, then, consists of many genes; it may be thought of as a linear series of genes that are all linked together.

The genes carried on a particular chromosome are called a linkage group.

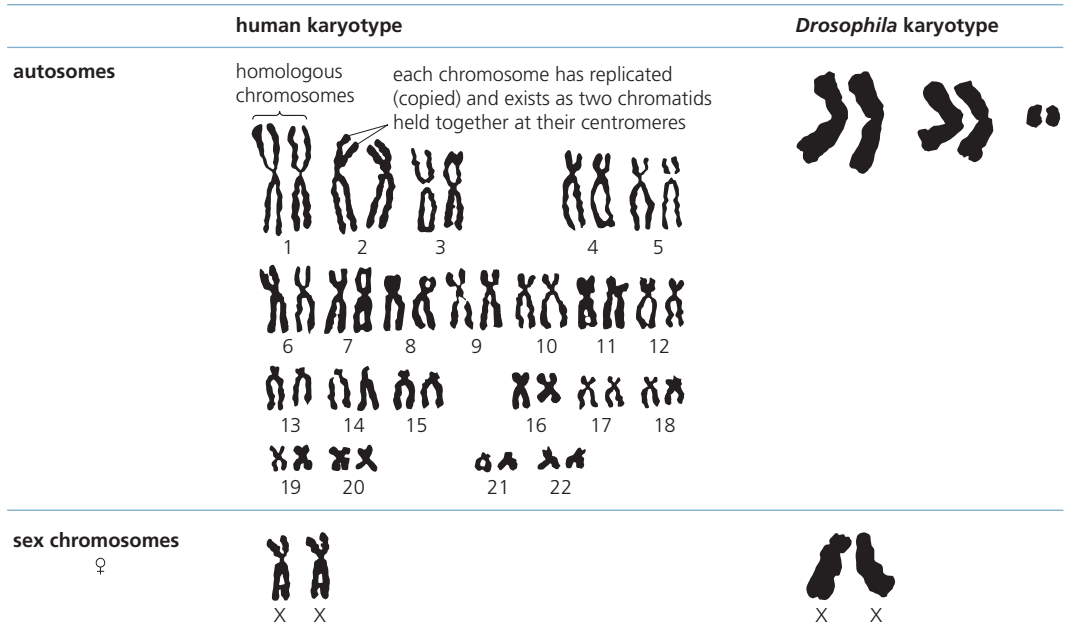


Figure 11.11 Chromosomes as autosomes and sex chromosomes

Obviously, genes of a linkage group will tend to be inherited together. And, in the case of a dihybrid cross involving linked genes, the Mendelian ratio we would expect from a cross with two pairs of contrasting characteristics in the F_2 generation (9:3:3:1) would not be obtained.

In fact, we might expect the ratio to be the same as that from crosses of one pair of contrasting characteristics in the F_2 generation, namely 3:1. This would certainly be the case if no chiasmata formed **between** the linked genes during prophase of meiosis, between non-sister chromatids.

The first example of this was discovered by experimental geneticists in 1906, very soon after Mendel's work was rediscovered. This experiment caused a mystery because linkage was not understood at that time. In studies of the inheritance in the **sweet pea plant**, plants homozygous for purple flowers and elongated pollen grains were crossed with plants homozygous for red flowers and rounded pollen grains. The offspring (the F_1 generation) had purple flowers and elongated pollen grains. Clearly, the alleles responsible for these characteristics were dominant.

When the F_1 generation plants were selfed, it was anticipated that the two pairs of alleles would segregate independently of each other, producing the four possible combinations (two parental types and two recombinant types) of a dihybrid cross in the expected ratio of 9:3:3:1.

This did not happen!

Instead, the progeny included a higher proportion of the parental characteristics (purple flowers and elongated pollen, and red flowers and round pollen) at the expense of the new combinations (purple flowers with round pollen, and red flowers with elongated pollen).

From Figure 11.12 it can be seen that this outcome is due to:

- gene linkage;
- where the chiasmata occur along the length of the chromosome.

6 The relative positions of linked genes have been extensively studied in organisms such as *Drosophila* by determining the proportions of recombinants in large samples of crosses. Recombination experiments have allowed the construction of gene maps of individual chromosomes, but this is not the case with humans. **Explain** why this is so, and **outline** how linkage is now investigated in human chromosomes.

parents homozygous purple flowers, elongated pollen × homozygous red flowers, round pollen

genotypes: F E / F E × f e / f e

gametes: F E × f e

offspring (F_1) heterozygous purple flowers, elongated pollen

F E / f e

× self

When the F_1 offspring are selfed, the F_2 offspring that result depend on whether a chiasma forms *between* these alleles *or elsewhere* along the chromosome during meiosis, in gamete formation.

Figure 11.12 Crossing over in the sweet pea plant

flowers of sweet pea (*Lathyrus odoratus*)

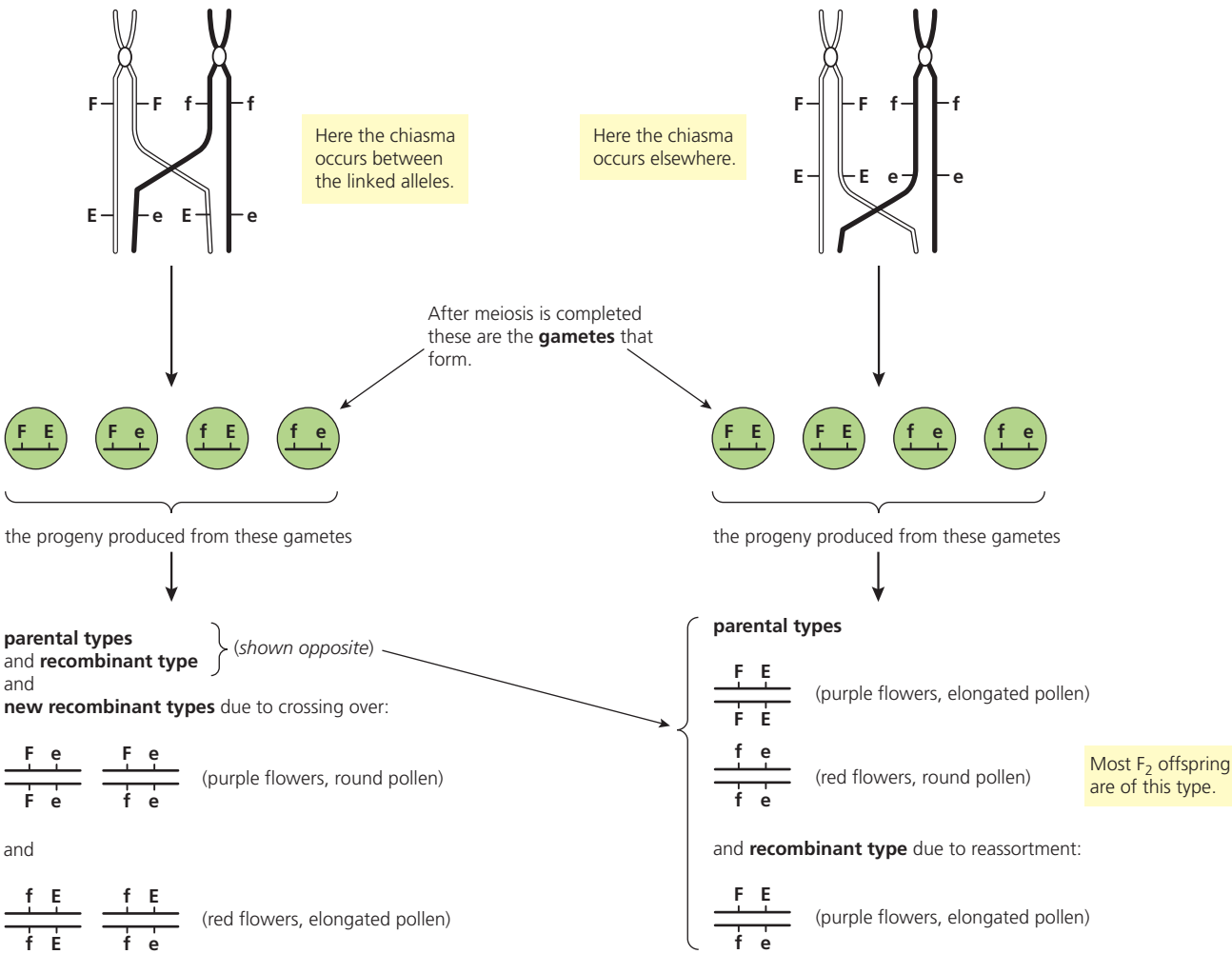
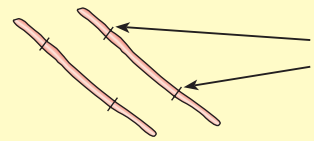


Figure 11.13 Linkages, crossing over and the test cross

One of the four pairs of chromosomes in *Drosophila* carries the alleles for the genes for both:
body colour – normal (**GG** or **Gg**) or ebony body (**gg**)
wing form – straight (**SS** or **Ss**) or curled wing (**ss**)



linked genes:
body colour
wing form

test cross

parental (**P**)

a (male) homozygous recessive fly with ebony body and curled wing

a (female) heterozygous fly with normal body and straight wing

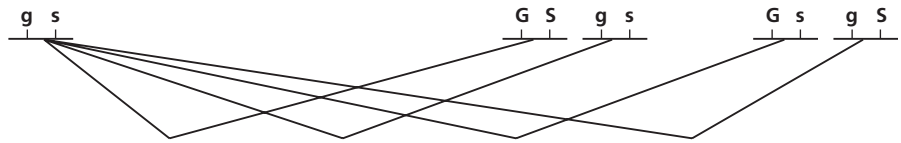
phenotypes:

×

genotypes:

gametes:

if no chiasma between genes : if chiasma occurs between genes



offspring (F_1)

genotypes:

phenotypes:

normal body straight wing | ebony body curled wing : normal body curled wing | ebony body straight wing

parental types | recombinants due to crossing-over

Note: If these genes were on separate chromosomes we would expect these offspring in the ratio 1:1:1:1 (**dihybrid test cross**), but these genes are linked, and if no crossing over occurred we would expect parental types only, in the ratio 1:1 (**monohybrid test cross**).

The outcome of this experiment was:

Offspring	Phenotypes	Genotypes	Numbers obtained
parental types	normal body straight wing	G S / g s	536
	ebony body curled wing	g s / g s	481
recombinants	normal body curled wing	G s / g s	101
	ebony body straight wing	g S / g s	152

The majority of the offspring were parental types, so more chiasmata in this chromosome occur somewhere other than between these two genes.

Gene linkage and the test cross

The role of the test cross is to differentiate between the homozygous dominant and the heterozygous dominant since their appearance (their phenotype) is identical (pages 106 and 336).

The possibility of crossing over complicates the outcome of a test cross involving two genes that are linked, as Figure 11.13 shows.

This illustration comes from an investigation of inheritance in *Drosophila*, where a fly homozygous for ebony body and curled wing was crossed with a fly heterozygous for normal body and straight wing. Note that the characteristic 'curled wing' is different from the characteristic 'vestigial wing' shown in Figure 11.10 (a cross concerned with two contrasting characteristics controlled by genes on separate chromosomes – a straightforward dihybrid cross).

7 Deduce what recombinants may be formed in a test cross involving the linked genes:

$$\frac{tb}{tb} \times \frac{TB}{tb}$$

Polygenes and continuous variation

10.3.1–10.3.2

We began the story of genetics in Chapter 4 with an investigation of the inheritance of height in the garden pea (page 104), where one gene with two alleles gave tall or dwarf plants. This clear-cut difference in an inherited characteristic is an example of **discontinuous variation** – there is no intermediate form, and no overlap between the two phenotypes.

In fact, **very few characteristics of organisms are controlled by a single gene**. Mostly, characteristics of organisms are controlled by a number of genes. Groups of genes which together determine a characteristic are called **polygenes**.

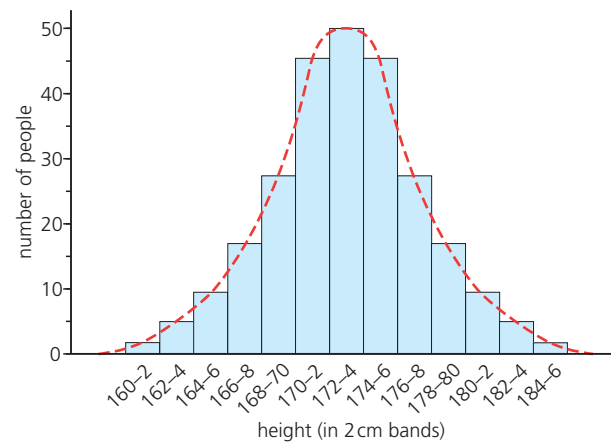
Polygenic inheritance is the inheritance of phenotypes that are determined by the collective effect of several genes.

The genes that make up a polygene are often (but not necessarily always) located on different chromosomes. Any one of these genes has a very small or insignificant effect on the phenotype, but the combined effect of all the genes of the polygene is to produce infinite variety among the offspring.

Many features of humans are controlled by polygenes, including body weight and height. The graph of the variation in the heights of a population of 400 people, in Figure 11.14, shows continuous variation in height between the shortest at 160 cm and the tallest at 186 cm and a mean height of 173 cm. Human height is controlled by several different genes.

Figure 11.14 Human height as a case of polygenic inheritance

Human height is determined genetically by interactions of the alleles of several genes, probably located at loci on different chromosomes.



Variation in the height of adult humans
The results cluster around a mean value and show a normal distribution. For the purpose of the graph, the heights are collected into arbitrary groups, each of a height range of 2 cm.

Human skin colour

The colour of human skin is due to the amount of the pigment called **melanin** that is produced in the skin. Melanin synthesis is genetically controlled. It seems that three, four or more separately inherited genes control melanin production. The outcome is an almost continuous distribution of skin colour from very pale (no alleles coding for melanin production) to very dark brown (all alleles for skin colour code for melanin production).

In Figure 11.15, polygenic inheritance of human skin colour involving only two independent genes is illustrated. This is because dealing with all four genes is unwieldy (Figure 11.16), and the principle can be demonstrated clearly enough using just two genes.

It should be noted, too, that both human height and skin colour are characteristics that may also be influenced by environmental factors.

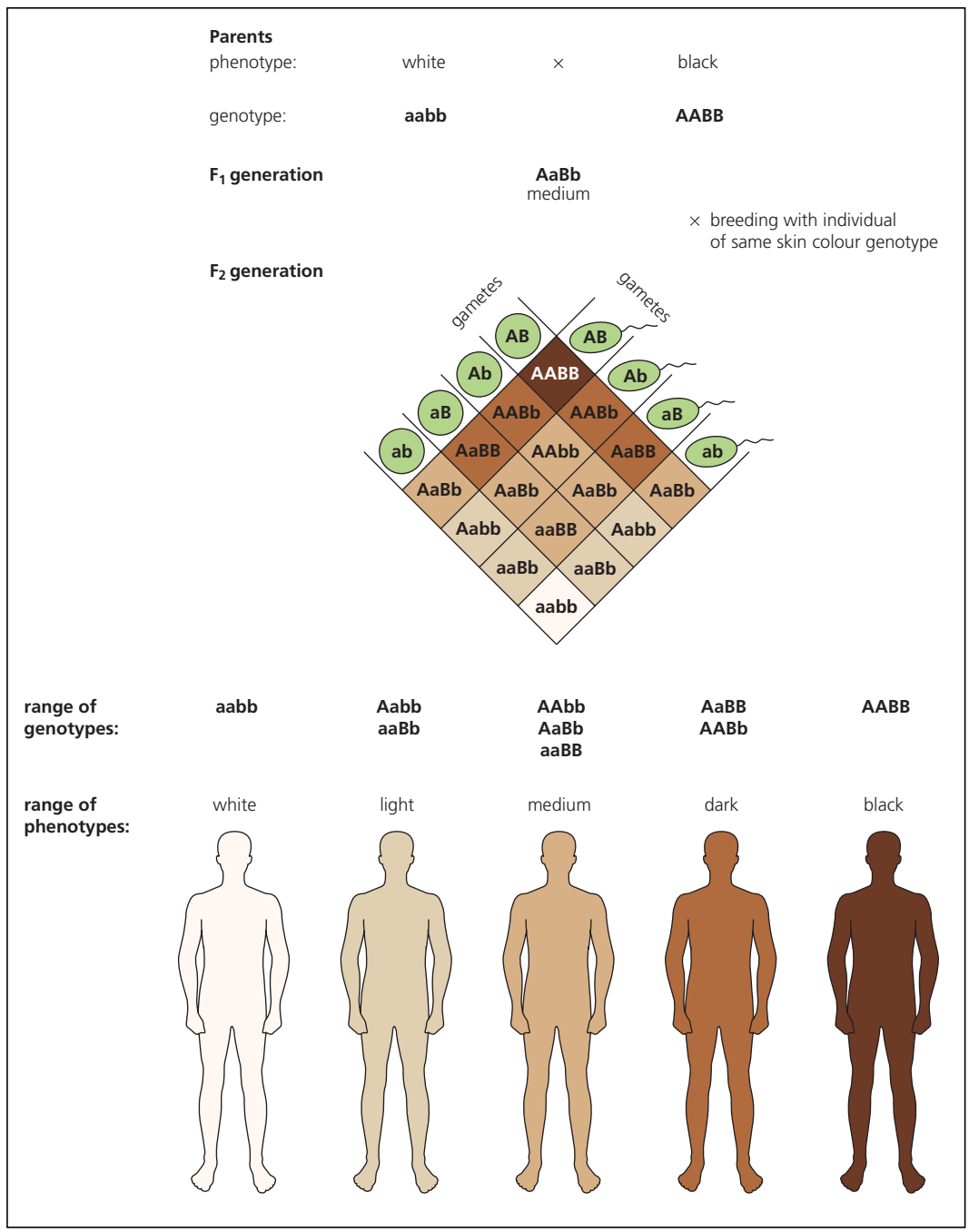


Figure 11.15 Human skin colour as a characteristic controlled by two independent genes – an illustration of polygenic inheritance

8 **Derive** the ratio of phenotypes produced in the F_2 generation shown in Figure 11.15.

Polygenic inheritance in a plant species

Many other examples of polygenic inheritance have been discovered. A good example, one in which only three genes interact together to control the phenotype, is the colour of the grain (a fruit) in certain varieties of wheat, *Triticum vulgare*.

In this case, for all three of these genes there are only two alleles. Despite this limitation, the full range of colours obtained varies from dark red (genotype $\mathbf{R^1R^1R^2R^2R^3R^3}$) to white (genotype $\mathbf{r^1r^1r^2r^2r^3r^3}$), with five intermediate shades of pink (Figure 11.16).

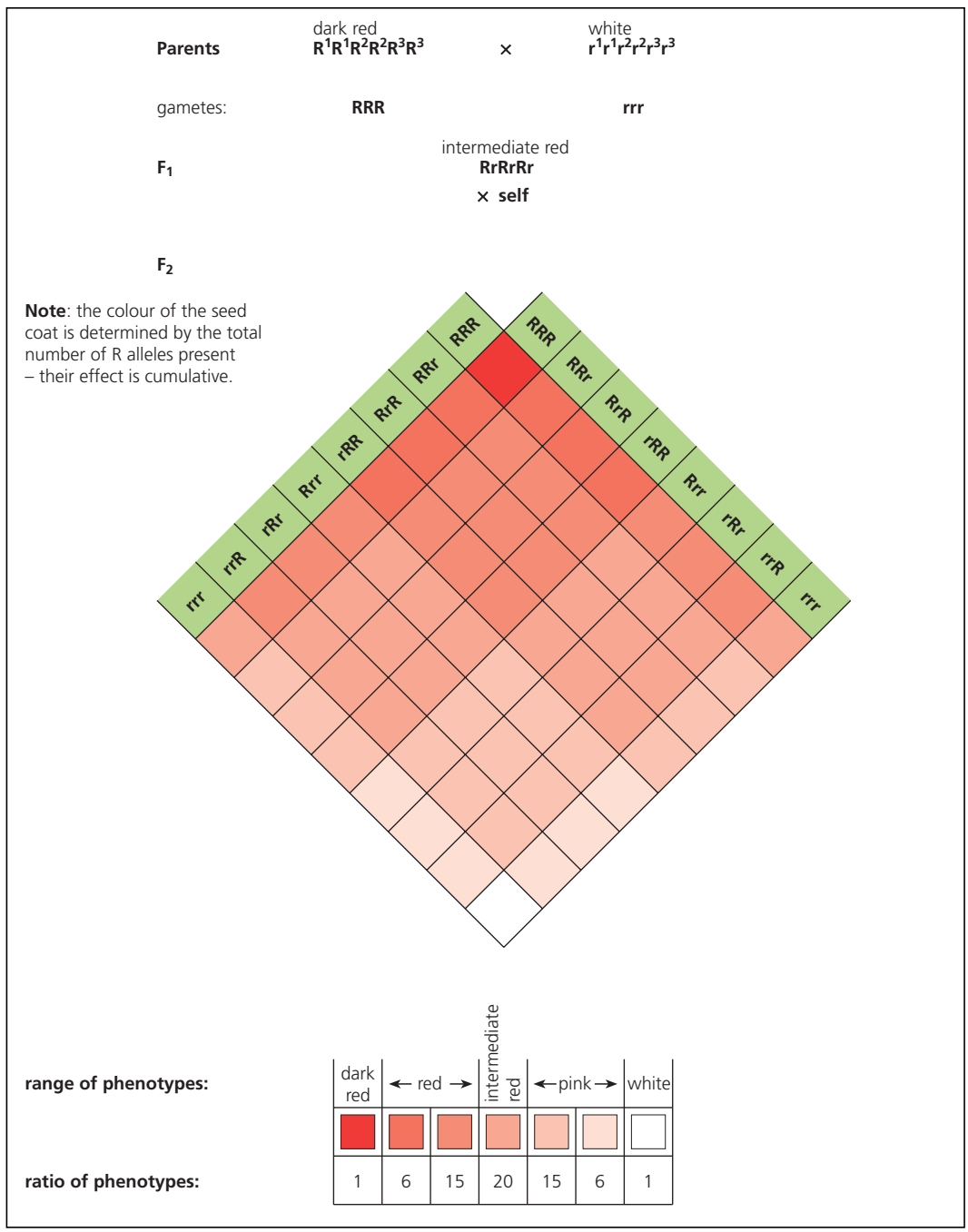


Figure 11.16 Inheritance of seed coat colour in *Triticum vulgare*, a case of polygenic inheritance

Continuous variation

We can see that the number of genes controlling a characteristic does not have to be large before the variation in a phenotype becomes more or less continuous, where a large group of offspring are produced. The outcome is that the characteristics controlled by polygenes tend to show **continuous variation**.

Nevertheless, the individual genes concerned are inherited in accordance with the principles established above. However, there are so many intermediate combinations of alleles that the discrete ratios are not observed. Even where only two genes constitute the polygene, significant variation results where the effect of the alleles is additive.

Extension: Other forms of gene interaction

We have seen that polygenic inheritance adds to the range of phenotypes that may exist. In fact, there are also other mechanisms by which genetic variety is controlled. It is helpful at this stage to be aware of the existence of some of these mechanisms, although knowing about how they operate is beyond the scope of this course.

Cases where characteristics are controlled by two or more genes, one or more of which **masks or modifies** the expression of the other(s) to various degrees, include:

- plumage pigmentation and comb form in the domestic fowl;
- agoutie coat colour in some mice;
- shell banding in the snail *Cepaea nemoralis*.

These and several other examples have similarly been much studied and widely reported on.

Another important factor that may affect the appearance of the phenotype is the effect of **environmental conditions** (already mentioned above). For example, a tall plant may appear almost dwarf if it has been consistently deprived of adequate essential mineral ions.

Similarly, the physique of humans may be greatly affected by the levels of nourishment received, particularly as children.

We must remember then, that the phenotype of an organism is the product of both its genotype and the **influences of the environment**.

Examination questions – a selection

Questions 1–6 are taken from past IB Diploma biology papers.

Q1 Which response describes the behaviour of chromosomes in metaphase I and anaphase II of meiosis?

	Metaphase I	Anaphase II
A	chromosomes line up at the equator	separation of homologous chromosomes
B	tetrads (bivalents) line up at the equator	separation of homologous chromosomes
C	chromosomes line up at the equator	separation of sister chromatids
D	tetrads (bivalents) line up at the equator	separation of sister chromatids

Higher Level Paper 1, May 03, Q24

Q2 In garden peas, the pairs of alleles coding for seed shape and seed colour are unlinked. The allele for smooth seeds (S) is dominant over the allele for wrinkled seeds (s). The allele for yellow seeds (Y) is dominant over the allele for green seeds (y).

If a plant of genotype Ssyy is crossed with a plant of genotype ssYy, which offspring are recombinants?

A SsYy and Ssyy
B SsYy and ssYy
C SsYy and ssyy
D Ssyy and ssYy

Higher Level Paper 1, May 03, Q26

Q3 A polygenic character is controlled by two genes each with two alleles. How many different possible genotypes are there for this character?

A 2
B 4
C 9
D 16

Higher Level Paper 2, May 04, Q30

Q4 If red (RR) is crossed with white (rr) and produces a pink flower (Rr), and tall (D) is dominant to dwarf (d), what is the phenotypic ratio from a cross of Rrdd and rrDd?

A 9:3:3:1
B 50% pink, 50% white and all tall
C 1:1:1:1, in which 50% are tall, 50% are dwarf, 50% are pink and 50% are white
D 3:1

Higher Level Paper 1, May 05, Q30

Q5 If the haploid number of an organism is 8, how many gametes are possible, not considering the effects of crossing over?

A 16
B 64
C 128
D 256

Higher Level Paper 1, May 06, Q40

Q6 **a** Define the term *gene linkage* and outline an example of a cross between two linked genes. (8)

b Describe the inheritance of ABO blood groups, including an example of the possible outcomes of a homozygous blood group A mother having a child with a blood group O father. (5)

c Outline sex linkage. (5)

Higher Level Paper 2, November 05, QB8

Questions 7–10 cover other syllabus issues in this chapter.

Q7 **a** Explain the essential differences between prophase of mitosis and prophase I of meiosis. (4)

b Identify the events in meiosis that result in gametes having different chromosome content. (4)

c Describe the steps in meiosis by which a chiasma is formed. (4)

Q8 **a** Mendel conducted many experiments with garden pea plants, but some later workers used the fruit fly *Drosophila* in experimental genetics investigations. Suggest **three** reasons why this insect was found to be useful. (3)

b When dihybrid crosses are carried out the progeny are rarely present in the exact proportions predicted. Explain why small deviations of this sort arise in dihybrid crosses with:

i garden pea plants (3)
ii *Drosophila*. (3)

Q9 In *Drosophila*, mutants with scarlet eyes and vestigial wings are recessive to flies with red eyes and normal wings. (These contrasting characters are controlled by single genes on different chromosomes, i.e. they are not linked on a single chromosome.) Explain the phenotypic ratio to be expected in the F_2 generation when normal flies are crossed with scarlet eyes/vestigial wing mutants and sibling crosses of the F_1 offspring are then conducted. Show your reasoning by means of a genetic cross diagram. (8)

Q10 Distinguish between the following pairs:

a *discontinuous variable* and *continuous variable* (4)
b *monohybrid cross* and *dihybrid cross* (4)
c *linkage* and *crossing over* (4)
d *autosomes* and *sex chromosomes* (4)
e *X chromosomes* and *Y chromosomes* (4)
f *multiple alleles* and *polygenes* (4)

12 Human physiology, health and reproduction II

STARTING POINTS

- The **blood circulation** has a role in the body's **defence against disease**.
- The **immune system** responds to the presence of foreign matter (**antigens**), including **pathogens**, by producing **antibodies** by special **leucocytes** (white cells).
- **Energy** is transferred in cells by cell respiration. **ATP**, the universal energy currency, is a reactant in **energy-requiring reactions** and **processes**.
- **Neurones** are the basic units of the **nervous system**, which consists of **receptors** (sense organs) linked to **effectors** (**muscles** or glands) by pathways of neurones called **reflex arcs**.
- **Excretion** is the removal of **waste products** of metabolism, including carbon dioxide, expelled at the lungs.
- **Homeostasis** involves maintaining the internal environment of the body between narrow limits. This includes **osmoregulation** in which the balance of **water** and **solutes** in body fluids is maintained at a constant level despite variations in intake.
- **Sex hormones** (testosterone, oestrogen and progesterone) have roles in the onset of **sexual maturity**, and the **production of sperm and egg cells**.
- Cell divisions by **mitosis** lead to the production of identical new cells. By **meiosis** the number of chromosomes per nucleus is halved – an essential step in **gamete production**.
- This chapter extends study of aspects of the structure and functioning (**physiology**) **of the mammal** begun in Chapter 7, pages 178–234.

We think of the **blood circulation** as the essential transport system for the body, but it also plays an intricate part in our resistance to harmful invasions. It is particularly the **white cells** that combat **infection**.

In the operation of the **nervous system**, changes that bring about responses, called **stimuli**, are detected by **receptors** (sense organs), but it is an **effector** (including our muscles) that brings about a response. Since in mammals, receptors and effectors are widely spaced, an efficient system of internal communication is essential.

Mammals are **regulators**, able to keep their internal environment in a steady state under a wide variety of external conditions. Negative feedback is the type of control mechanism that operates in **homeostasis**.

In mammals the **sexes are separate**. The **male reproductive system** includes the testes, in which sperms are produced. The **female reproductive system** includes the ovaries, in which oocytes are formed, and the uterus, where early development of the offspring occurs.

In this chapter, we extend study of the body's mechanisms of **defence against disease**. Then, the roles of the skeletal, muscular and nervous systems in bringing about **movement** are investigated, followed by discussion of the roles of the **kidneys** in excretion and osmoregulation – an aspect of homeostasis. **Human reproduction** is revisited in an investigation of the structure of the gonads and the process of gamete formation. Finally, the steps of fertilisation, the processes of pregnancy and birth, and the roles of certain hormones of the systems are considered.

Defence against infectious disease

11.1.1–11.1.7

Our intact skin is no easy barrier for pathogens to cross – mostly they are prevented from entry to the body. However, the defences of the skin are sometimes broken down, and so it is fortunate there are also internal lines of defence. The body responds to localised damage (cuts and abrasions, for example) by **inflammation**. Inflammation is the initial, rapid, localised response the tissue makes to damage. As a result, the volume of blood in the damaged area is increased. If a blood vessel has been ruptured, then the **blood clotting mechanism** is activated. In the blood and tissue fluid, the **immune system** operates. We will examine the blood clotting mechanism, first.

The blood clotting mechanism

In the event of a break in our closed blood circulation, the danger arises of a loss of blood and possibly a fall in blood pressure. It is by the clotting of blood that escapes are prevented, either at small haemorrhages, or at cuts and other wounds. In these circumstances, a clot both stops the outflow of blood and reduces invasion opportunities for pathogens. Subsequently, repair of the damaged tissues can get under way. Initial conditions at the wound trigger a **cascade of events** by which a blood clot is formed.

What is meant by 'a cascade of events'?

First, **platelets** collect at the site. These are components of the blood that are formed in the bone marrow along with the erythrocytes (red cells) and leucocytes (white cells), and are circulated throughout the body, suspended in the plasma, with the blood cells (Figure 7.7, page 186).

Platelets are actually cell fragments, disc-shaped and very small (only 2 μm in diameter) – too small to contain a nucleus. Each platelet consists of a sack of cytoplasm rich in vesicles containing enzymes, and is surrounded by a plasma membrane. Platelets stick to the damaged tissues and clump together there (at this point they change shape from sacks to flattened discs with tiny projections that interlock). This action alone seals off the smallest breaks.

The collecting platelets release a **clotting factor** (a protein called thromboplastin) which is also released by damaged tissues at the site. This clotting factor, along with vitamin K and calcium ions, always present in the plasma, causes a soluble plasma protein called **prothrombin** to be converted to the active, proteolytic enzyme **thrombin**. The action of this enzyme is to convert another soluble blood protein, **fibrinogen** to insoluble **fibrin** fibres at the site of the cut. Red cells are trapped within the mass of fibres, and the blood clot has formed (Figure 12.1).

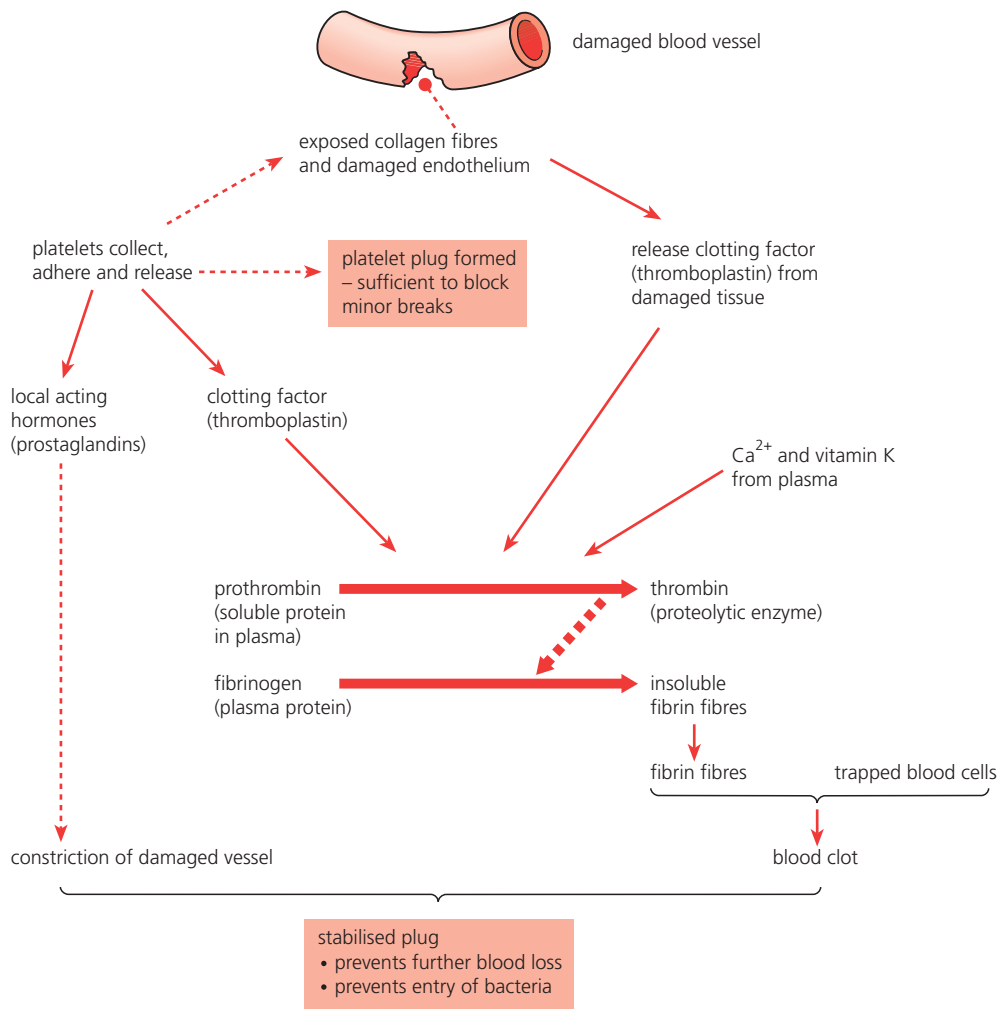
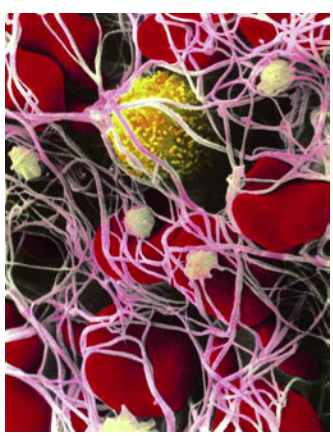


Figure 12.1 The blood clotting mechanism



SEM of blood clot showing meshwork of fibrin fibres and trapped blood cells

It is most fortunate that clot formation is not normally activated in the intact circulation; clotting is triggered by the abnormal conditions at the break. These complex steps of clotting may be seen as an essential **fail-safe mechanism**. This is necessary because casual formation of a blood clot within the intact circulation system immediately generates the risk of a dangerous and possibly fatal blockage in the lungs, heart muscle or brain.

1 Identify the correct sequence of the following events during blood clotting:

fibrin formation; clotting factor release; thrombin formation.

The immune system and the response to invasion

It is the leucocytes in the blood that provide our main defences **once invasion of the body by harmful microorganisms has occurred**. Among the millions of erythrocytes (red cells) of our blood circulation are the relatively few leucocytes (white cells).

Leucocytes are not restricted to the blood circulation. Not only do they continually circulate in the arteries, capillaries and veins, they move through the walls of these vessels into tissues and organs. Very many occur in lymph of the lymph nodes and vessels, and many are found in organs like the lungs (page 196) and in the liver, and elsewhere in the body.

It is special leucocytes called **lymphocytes** that are responsible for the immune response. Lymphocytes make up 20% of the leucocytes circulating in the blood plasma. Lymphocytes detect matter entering from outside our bodies (including foreign macromolecules and microorganisms) as different from 'self' (body cells and our own proteins). 'Non-self' invading macromolecules and microorganisms are known as **antigens**.

What recognition of 'self' entails

Cells are identified by specific molecules – markers that are lodged in the outer surface of the plasma membrane. These molecules, which identify a cell, are highly variable **glycoproteins** on the cell surface.

You can see glycoproteins on the plasma membrane in Figure 1.19, page 21.

The glycoproteins that identify cells are known as the **major histocompatability complex** antigens – but we can refer to them as MHC. There are genes on one of our chromosomes (chromosome 6) that code for MHC, so each individual's MHC is genetically determined and is a feature we inherit. As with all inherited characteristics that are products of sexual reproduction, variation occurs. Each of us has distinctive MHC antigens present on the plasma membrane of most of our body cells. Unless you have an identical twin, your MHC antigens are unique.

So, lymphocytes of our immune system have **antigen receptors** that recognise our own MHC antigens, and differentiate these from foreign antigens detected in the body.

The principle of challenge and response

Our lymphocytes all appear the same, but we have two distinct types, based on the ways they function:

- B-lymphocytes (**B-cells**) secrete antibodies (**humoral immunity**);
- **T-cells** assist B-cells and may attack infected cells (**cell-mediated responses**).

As B- and T-cells are formed (in the red bone marrow) and then mature, they develop **immunocompetence**. This means they acquire an ability to recognise one specific antigen. An immunocompetent cell has a specific receptor protein on its cell surface. Each cell has lots of these receptor molecules that are capable of binding to only **one type of antigen**. The receptors on B-cells are actually attached copies of the antibody that will be secreted later, if the antigen is encountered.

Immunocompetent lymphocytes are stored in the lymph nodes. Here all the different lymphocytes, each with multiple copies of their antigen receptor, are found.

If antigens invade the body, they reach the tissue fluids and lymph, and encounter a lymphocyte with the corresponding receptor, typically in a lymph node. The lymphocytes challenged in this way immediately respond by growing and dividing repeatedly, leading to vast numbers of cells producing antibodies. The antibodies eventually overcome and dispose of the source of the foreign antigens.

Clonal selection and a polyclonal response

The product of vast numbers of identical plasma cells, as described above, is known as **clonal selection**, for obvious reasons.

Invading microorganisms have a complex chemical exterior – their wall or plasma membrane. Here are exposed very many different types of antigen. Consequently, a typical pathogen activates many different types of B-cell, triggering the secretion of many different antibodies, all of which attach and attack that pathogen. The activation of many different cloning events in this way is called **polyclonal selection**.

'Memory' in the immune system

After each immune response event has overcome the source of antigens, the plasma cells that secreted the antibodies are disposed of. Antibody-secreting cells are relatively short-lived. However, some of the cells produced in the cloning episode become memory cell lymphocytes. Memory cells survive in the body, mostly in the lymph nodes, typically for years after the event.

If the antigen returns, the body is able to respond quickly. We say that we have immunity.

Types of immunity

At the start of life, a mammal acquires **passive immunity** from the antibodies in its mother's blood circulation. Only later, after birth and direct stimulation by antigens, is **active immunity** acquired.

> **Passive immunity is immunity due to the acquisition of antibodies from the mother (in whom active immunity has arisen), via the placenta while in the uterus, and then via the first-formed milk, received immediately after birth. This milk is known as colostrum. Alternatively, antibodies can be received by injection, and so confer passive immunity.**

Note that passive immunity fades away eventually because the offspring cannot also receive the necessary memory cells, and so is unable to make the antibodies itself.

> **Active immunity is immunity due to the production of antibodies by the organism, after the body's defences have been exposed to antigens.**

Note that the exposure to antigens which generates active immunity may occur due to infection by a pathogen, as already described (**natural acquired immunity**), or following inoculation with a vaccine (**artificial acquired immunity**).

2 Outline in tabulated form the parts played by the blood in the protection of the body.

Steps of antibody production

We have seen that lymphocytes are involved in repelling an invasion by the production of antibodies, and that lymphocytes consist of two types, T-lymphocytes and B-lymphocytes. Both types are formed by divisions of the **stem cells** in the bone marrow, and undergo a development process in preparation for their distinctive roles (Figure 12.2).

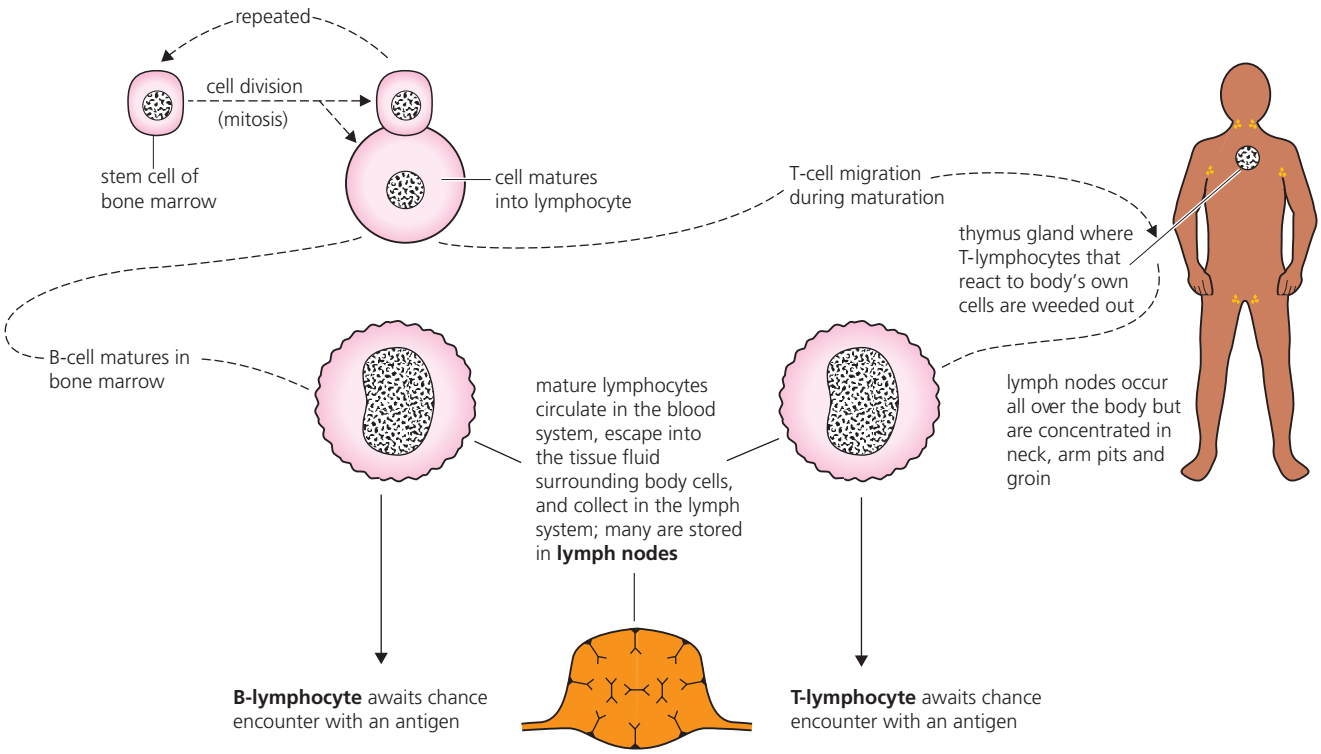
T-lymphocytes (T-cells) leave the bone marrow during development, and differentiate in the **thymus gland** before circulating and storage in lymph nodes. It is while they are present in the thymus gland that the body apparently selects out the lymphocytes that would otherwise react to the body's own cells.

The role of T-cells is **cell-mediated immunity** – an immune response not directly involving antibodies, although some have a role in the activation of B-cells, as we shall see shortly. Some T-cells are effective against pathogens located within host cells.

B-lymphocytes (B-cells) complete their maturation in the bone marrow, prior to circulating in the body and being stored in lymph nodes. The role of the majority of B-cells, after recognition and binding to a specific antigen, is to proliferate into cells (called **plasma cells**) that **secrete antibodies** into the blood system. This is known as **humoral immunity**.

Both T-cells and B-cells have molecules on the outer surface of their plasma membrane that enable them to recognise antigens, but each lymphocyte has only one type of surface receptor. Consequently, each lymphocyte can recognise only one type of antigen.

Figure 12.2 T-lymphocytes and B-lymphocytes



3 Explain the significance of the role of the thymus gland in destroying T-cells that would otherwise react to body proteins.

When an infection occurs the leucocyte population responds. Their numbers increase enormously, and many collect at the site of the invasion. A complex response to infection is begun. The special roles of T-cells and B-cells in this response are as follows, and are illustrated in Figure 12.4 (page 354).

1 On the arrival of a specific antigen in the body, B-cells with surface receptors (antibodies) that recognise that particular antigen, bind to it.

What exactly is an antibody?

An antibody is a special protein called an **immunoglobulin**, made of four polypeptide chains held together by disulphide bridges (—S—S—), forming a molecule in the shape of a Y. The arrangement of amino acid residues in the polypeptides that form the fork region in this molecule is totally unique to that antibody. It is this region that forms the highly specific binding site for the antigen (Figure 12.3). Antibodies initially occur attached to the plasma membrane of B-cells, but later are also mass-produced and secreted by cells derived from the B-cell by exocytosis, but only after that B-cell has undergone an **activation step** (step 5, below).

2 On binding to the B-cell, the antigen is taken into the cytoplasm by phagocytosis, before being expressed on the plasma membrane of the B-cell.

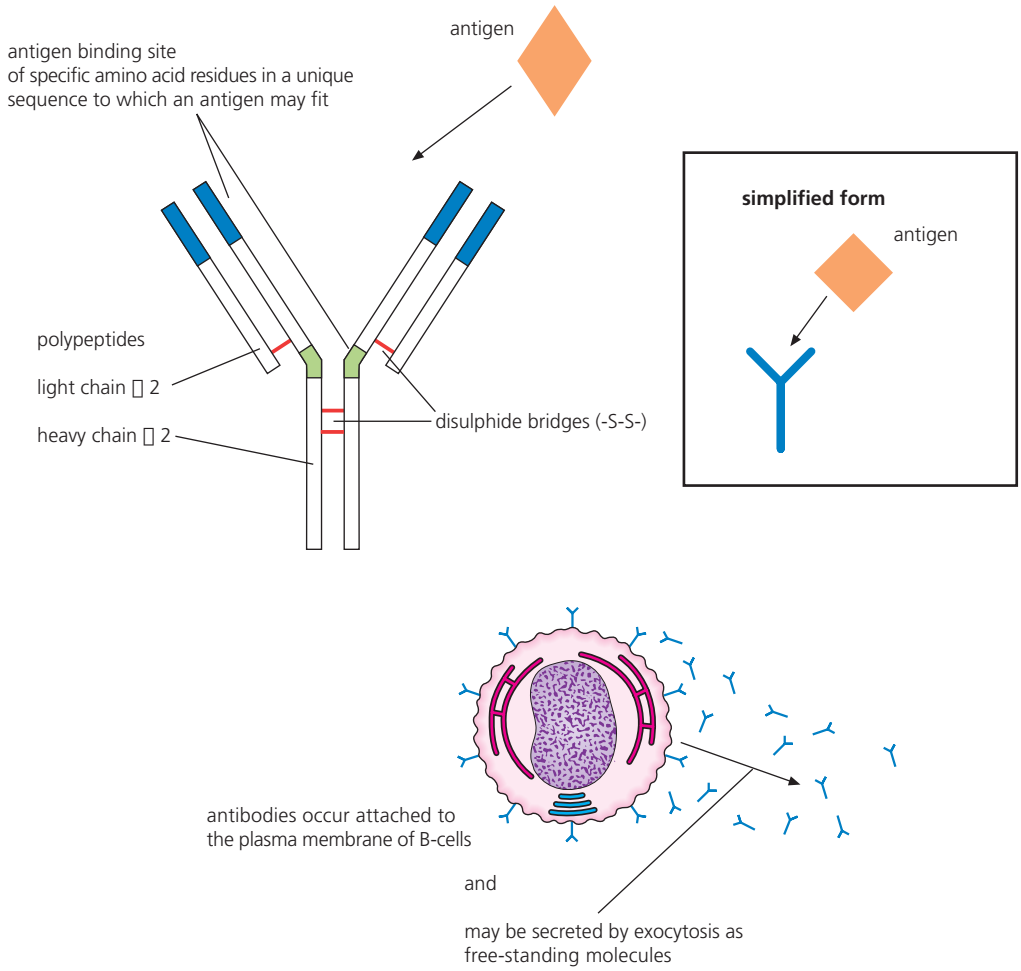


Figure 12.3 The structure of an antibody

3 Meanwhile, T-cells can only respond to antigens when presented on the surface of other cells. Phagocytic cells of the body, including **macrophages**, engulf antigens they encounter. This occurs in the plasma and lymph. Once these antigens are taken up, the macrophage presents them externally by attaching the antigen to their surface membrane protein, MHC antigens. This is called **antigen presentation** by a macrophage.

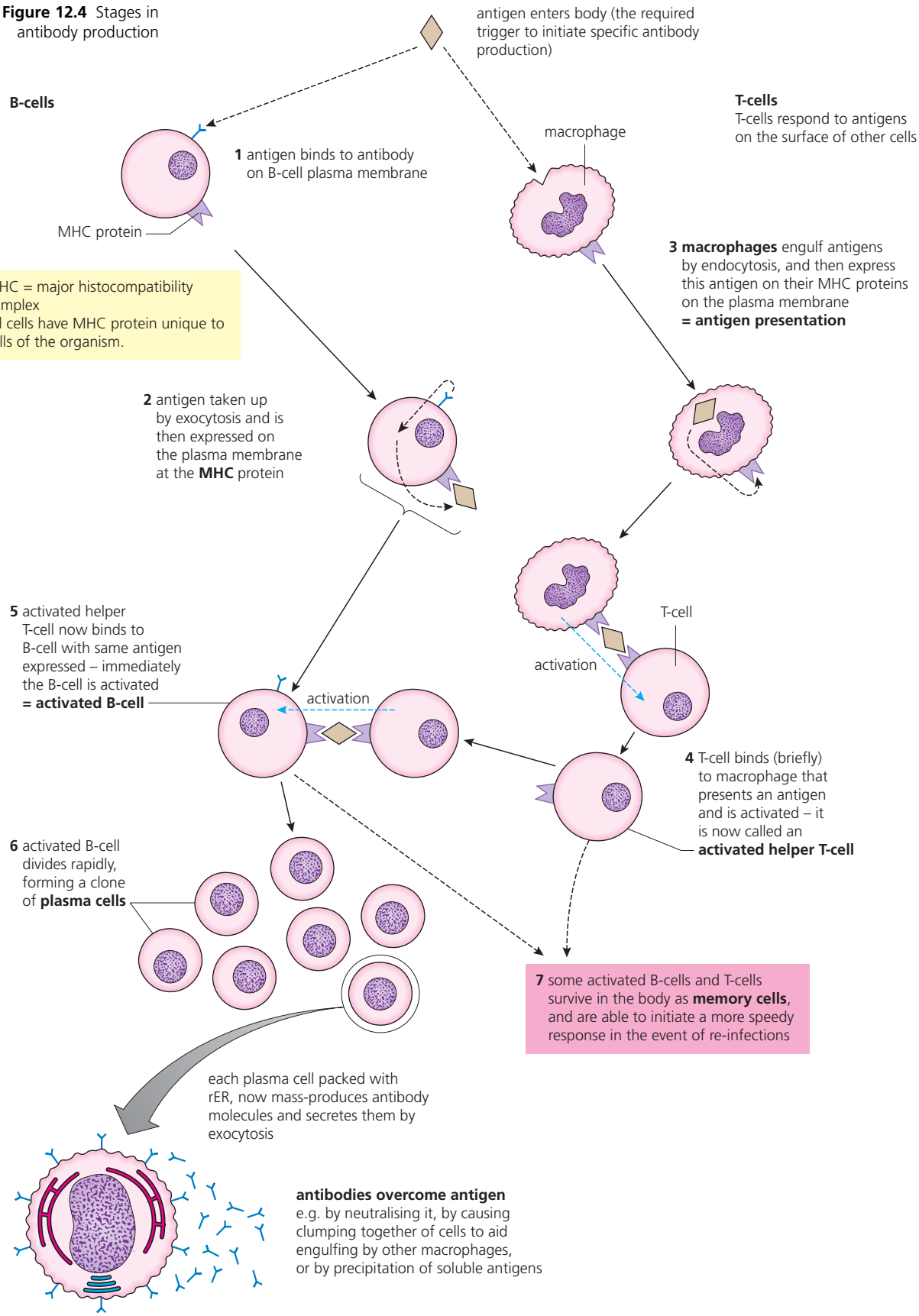
4 T-cells come in contact with these macrophages and briefly bind to them. The T-cell is immediately activated. They are armed or **activated helper T-cells**.

5 Activated helper T-cells now bind to B-cells with the same antigen expressed on their plasma membrane (step 2 above), and the activated T-cell sends a message to the B-cell, activating it. This is the activation step mentioned in step 1 above. It is now an armed or **activated B-cell**.

6 Activated B-cells immediately divide very rapidly by mitosis, forming a clone of cells called plasma cells. A TEM of plasma cells shows them to be packed with rough endoplasmic reticulum (rER). It is in these organelles that the antibody is mass-produced and exported from the B-cell by exocytosis. The antibodies are normally produced in such numbers that the antigen is overcome (Figure 12.4).

As noted above, the production of an activated B-cell, its rapid cell division to produce a clone of plasma cells, and the resulting production of antibodies that react with the antigen is called clonal selection. Sometimes several different antibodies react with the one antigen – polyclonal selection.

Figure 12.4 Stages in antibody production



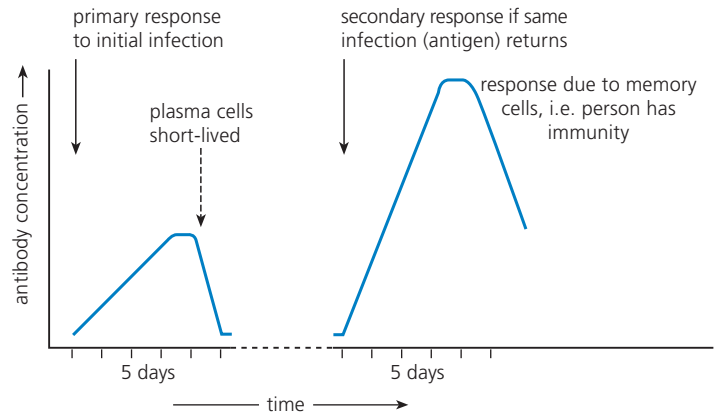


Figure 12.5 Profile of antibody production in infection and re-infection

Memory cells are retained in lymph nodes. They allow a quick and specific response if the same antigen reappears.

7 After these antibodies have tackled the foreign matter and the disease threat it introduced, they disappear from the blood and tissue fluid, along with the bulk of the specific B-cells and T-cells responsible for their formation. However, certain of these specifically activated B-cells and T-cells are retained in the body as memory cells. These are long-lived cells, in contrast to plasma cells and activated B-cells. Memory cells make possible an early and effective response in the event of a re-infection of the body by the same antigen (Figure 12.5). This is the basis of natural immunity (see below).

4 Identify where antigens and antibodies may be found in the body.

Extension: Additional roles for T-cells

We have noted that T-cells of the immune system identify antigens **when presented on the surface of cells** (as opposed to responding to free antigens in the plasma of the blood or in tissue fluid, as B-cells do).

A **virus** causes a disease only when it has entered a host cell and then has converted the host metabolic machinery to the mass-production of more viruses (page 237). However, a virus-infected cell also automatically 'advertises' the presence of the virus within, early in the invasion, by attaching antigens from the virus's protein coat onto the exterior of the plasma membrane.

The immune system, in the form of special T-cells, recognises these antigens. The killer T-cell attaches to the infected cell and then destroys it; the infection is overcome. Killer T-cells are known as **cytotoxic T-cells**.

T-cells and cancer

A similar response occurs with body cells that have become cancerous. This is because cancer cells develop **tumour-associated antigens (TAA)** on their surface. Chance encounters between T-cells and cancer cells with exposed TAA sensitise the T-cells (we say the T-cells have become armed). They then become active as **cytotoxic cells**. In this condition, the cancer cells may be destroyed.

Unfortunately, the immune system is generally less efficient at dealing with cancerous tumours than it is with infections. This may be because cancer cells are almost identical to healthy body cells from which they formed. Consequently, they also carry MHC proteins on their plasma membranes, marking them out as the body's own cells. We saw that T-cells pass through the thymus gland during formation, where T-cells that attack the body's own cells are removed. It is likely that the dual messages on cancer cells (TAA alongside MHC proteins) may disable cytotoxic T-cells to some extent.

It is now helpful to summarise the complex roles of B-cells and T-cells in the immune system (Figure 12.6).

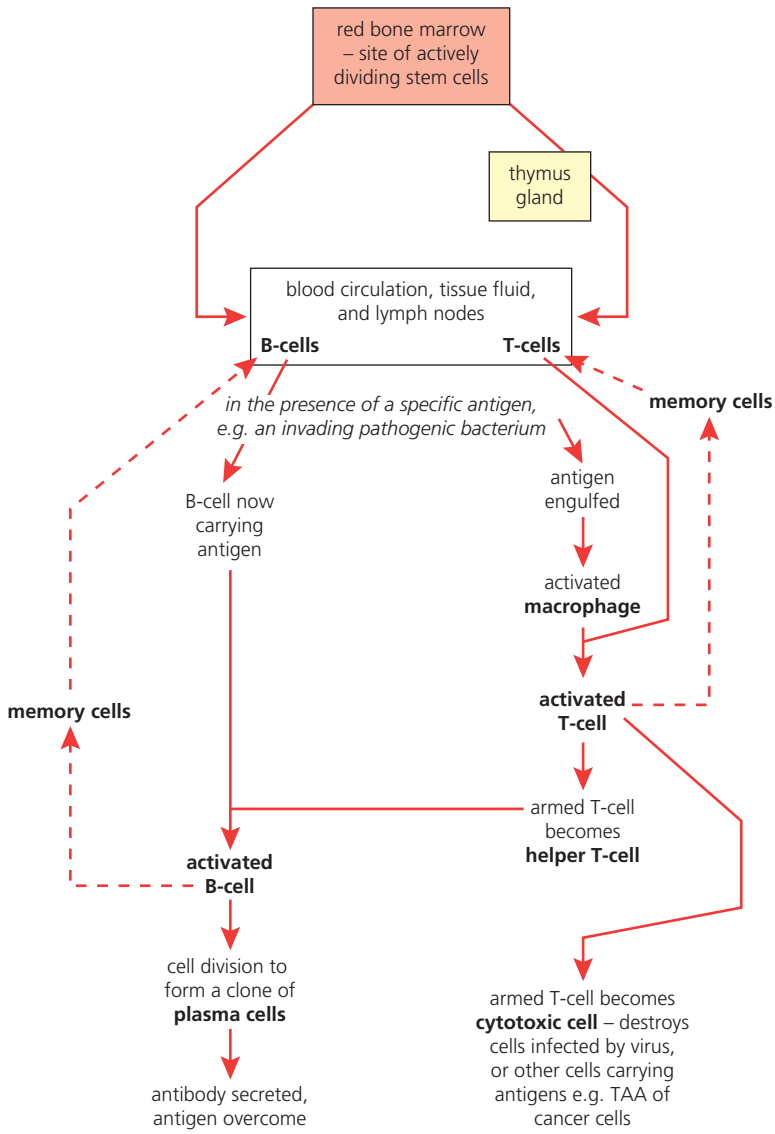


Figure 12.6 The roles of B-cells and T-cells in the immune system – a summary

5 Distinguish between a plasma cell and memory cells of the immune system.

Monoclonal antibody production and their uses

We have seen that antibodies are secreted by plasma B-cells and are effective in the destruction of antigens within the body.

Antibodies also have great potential in modern medicine, where they can be made available. The trouble has been that the plasma cells from which they are derived are short-lived. **Monoclonal antibodies** are a recent invention by which antibodies are made available in the long term, for applications in new circumstances. They are already used widely in medicine and research.

A monoclonal antibody is a single antibody that is stable and that can be used over a period of time. Each specific antibody is made by one particular type of B-cell. The problem of the normally brief existence of a plasma cell is overcome by fusing the specific lymphocyte with a cancer cell which, unlike other body cells, goes on dividing indefinitely. The resulting cell divides to form a clone of cells which persists, and which conveniently goes on secreting the antibody in significant quantities (Figure 12.7).

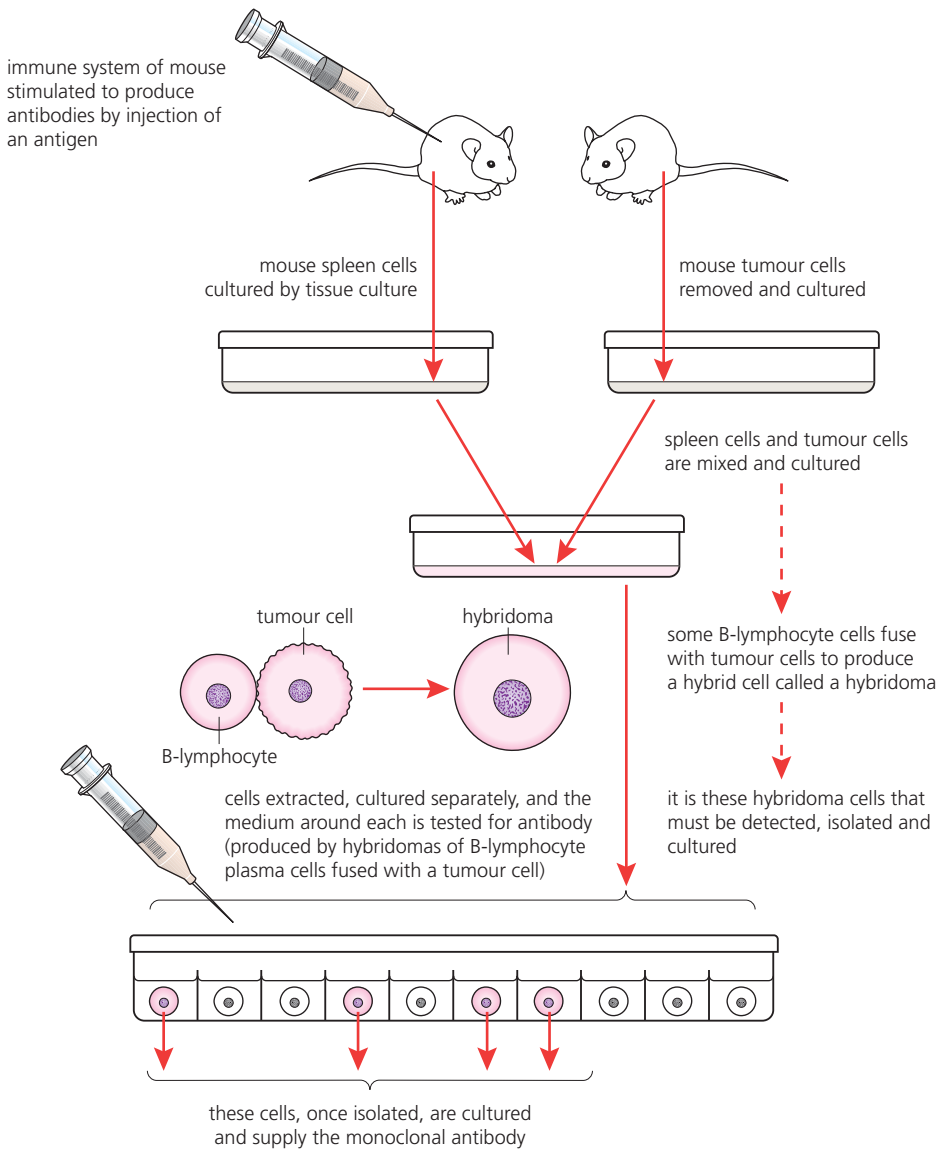


Figure 12.7 Formation of monoclonal antibodies

The use of monoclonal antibodies in medicine is already established, and is under further development in these and other areas. For example, in **diagnosis**, monoclonal antibodies are used in pregnancy testing.

A pregnant woman has a significant concentration of the hormone human chorionic gonadotrophin (HCG) (page 387) in her urine, whereas a non-pregnant woman has a negligible amount. Monoclonal antibodies to HCG have been engineered to also carry (become attached to) coloured granules, so that in a simple test kit the appearance of a coloured strip in one compartment provides immediate, visual confirmation of pregnancy. How this works is illustrated in Figure 12.8.

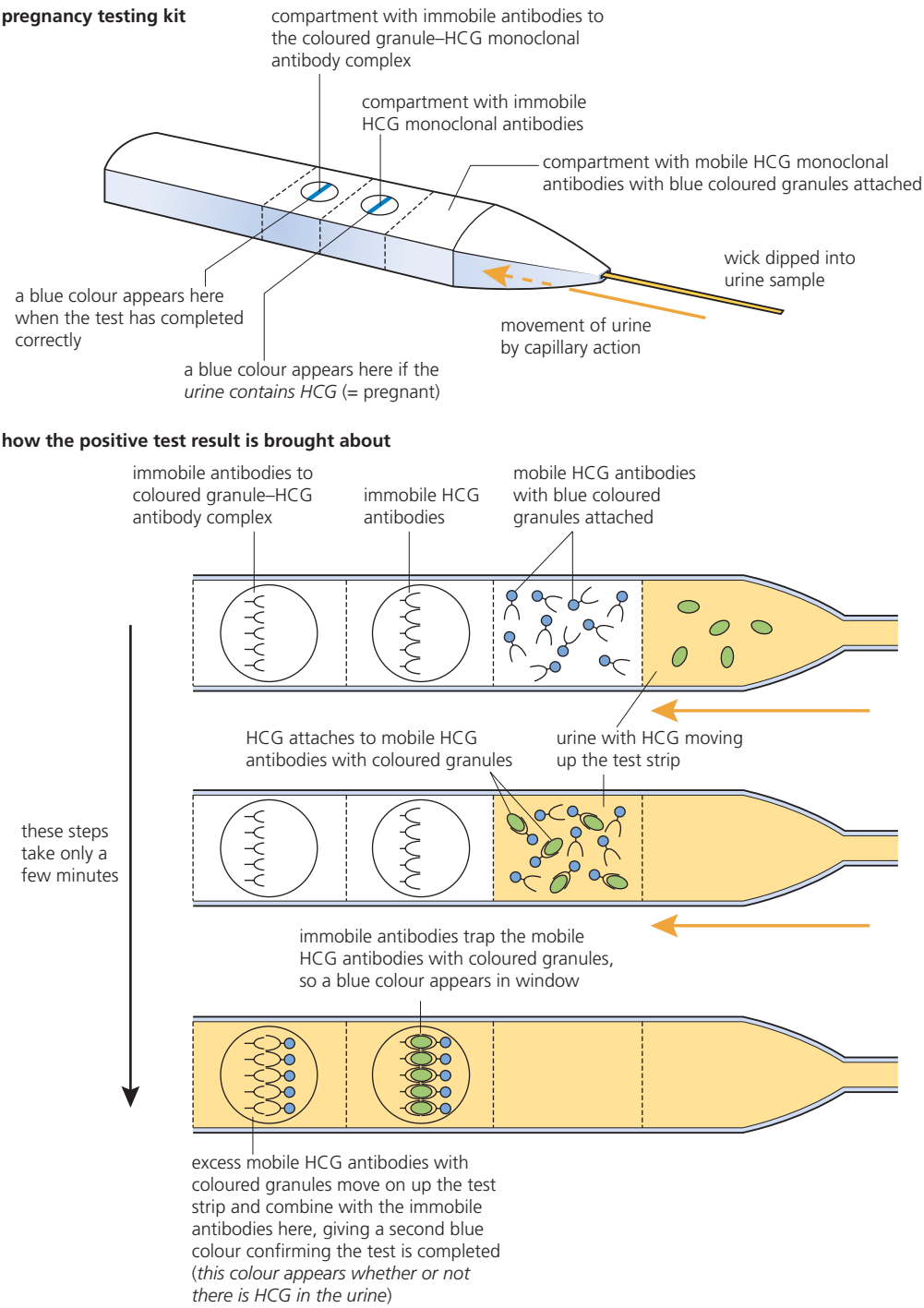


Figure 12.8 Detecting pregnancy using monoclonal antibodies

Monoclonal antibodies also have applications in the **treatment** of disease. For example, cancer cells carry specific tumour-associated antigens (TAA) on their plasma membrane. Monoclonal antibodies to TAA have been produced. Then, drugs to kill cells or inhibitors to block key tumour proteins have been attached, so cancer cells can be specifically targeted and killed. The advantage here is that many drugs and treatments effective against cancer are harmful to other, healthy cells. The specificity of antibodies avoids this problem.

The original monoclonal antibodies were developed from mouse cells (Figure 12.7), but patients have developed an adverse reaction to antibodies made this way, since they are foreign proteins to the patient's immune system. Genetically engineered antibodies that will be compatible with the patient's immune system are sought, to avoid triggering the immune response.

Vaccination – the process of immunisation

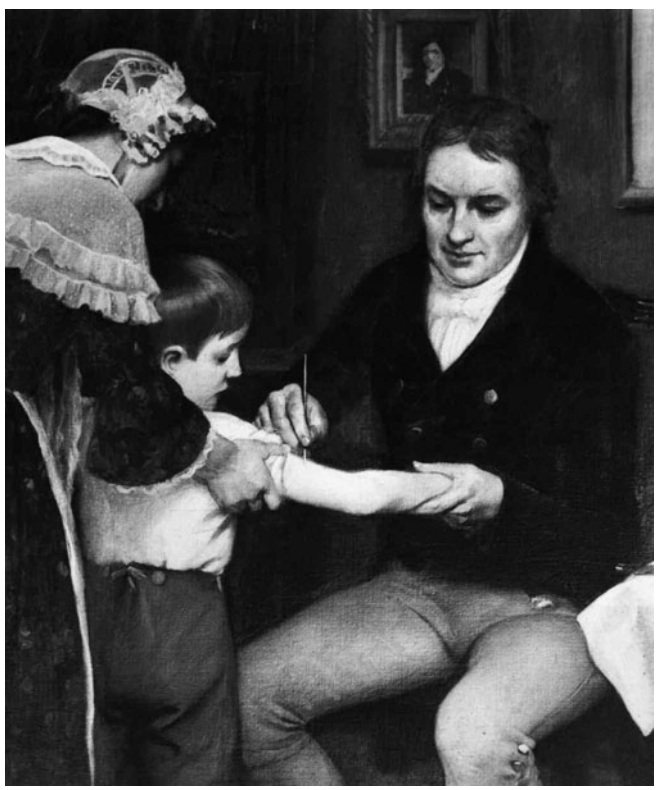
It has long been known that people who recovered from plague or smallpox rarely contracted those diseases again. The possibility of a patient acquiring immunity from smallpox was apparently appreciated in the East in the seventeenth century, if not earlier. An English ambassador's wife tried to introduce the practice of vaccination in Britain at the start of the eighteenth century. Vaccination was practised in Turkey at the time, and she had observed how successful it was there. Sadly, she had little success.

The first attempt in the West at immunisation was made by **Edward Jenner (1749–1823)**, a country doctor from Gloucestershire, UK. At the time, many people who got smallpox died of it, but *not* those who had earlier contracted cowpox (workers who handled cows typically did so at some stage). Jenner saw the significance of this protection the patients had acquired. He extracted fluid from a cowpox pustule on an infected milkmaid, and injected it into the arm of an eight-year-old boy (Figure 12.9). The child got a mild infection, but when exposed to smallpox, remained healthy. Jenner named this technique **vaccination**, after the cowpox *vaccinia* (a virus). Of course, he did not understand the cause of smallpox; the chemical nature of viruses was not reported until 1935.

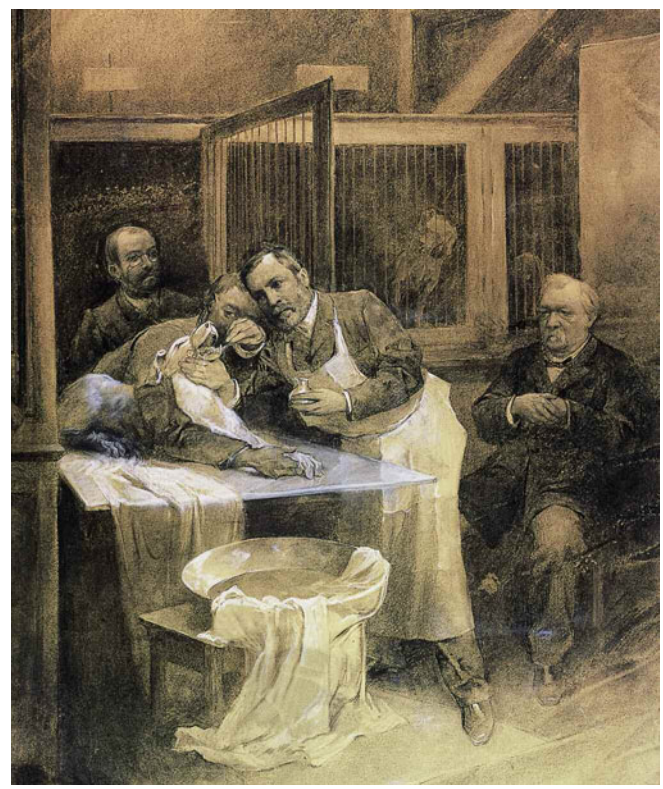
The French scientist **Louis Pasteur (1822–95)** discovered that cultures of chicken cholera bacterium, allowed to age for 2–3 months, produced only a mild infection of cholera when inoculated into chickens. The old cultures had become less pathogenic – today we say they had **attenuated**. Fresh, virulent strains of the chicken cholera bacterium failed to infect chickens previously exposed to the mild form of the disease.

Pasteur, one of the greatest experimentalists in microbiology, had notable successes in immunisation (Figure 12.9). However, he recognised the original contribution of Jenner to the discovery of immunity by using the name vaccine for injections of the attenuated organisms that he developed to prevent chicken cholera disease, and (later) anthrax (due to a bacterium) and rabies (due to a virus).

Figure 12.9 Edward Jenner and Louis Pasteur at work



Jenner inoculating James Phipps, from a drawing by William Thompson, c. 1880 (*by courtesy of the Wellcome Trustees*)



Pasteur made a major contribution to our understanding of diseases of humans and other animals; working with dogs, he showed that an injection of the attenuated rabies microorganism can produce immunity to the disease

Vaccines today

The deliberate administration of antigens that have been made harmless, after they are obtained from disease-causing organisms, in order to confer immunity in future, is a very important contribution to public health today.

Vaccines are administered either by injection or by mouth. They cause the body's immune system to briefly make antibodies against the disease (*without becoming infected*), and then to retain the appropriate memory cells. Active artificial immunity is established in this way. The profile of response in terms of antibody production caused by any later exposure to the antigen is exactly the same as if the immunity was acquired after the body overcame an earlier infection (Figure 12.5).

Vaccines are manufactured from dead or attenuated bacteria, or from inactivated viruses, purified polysaccharides from bacterial walls, toxoids, and even recombinant DNA produced by genetic engineering.

So successful has vaccination been *where vaccines are widely available and the take up is by about 85–90% of the relevant population* that, in many human communities, some formerly common and dangerous diseases have become very uncommon occurrences. As a result, the public in such communities have sometimes become casual about the threat such diseases still pose (see below). The recommended schedule of vaccinations for children brought up in one developed country can be accessed via the website www.doh.gov.uk

6 **Identify** the steps of plasma cell formation that the existence of memory cells avoids.

Controversy about vaccination

The issues that vaccination can raise with the general public are shown by cases such as the controversy that raged over the **combined measles, mumps and rubella (MMR) vaccine**.

MMR vaccine was first licensed for use in 1975. For MMR vaccine, attenuated viruses are extracted from separate cell cultures and freeze-dried. The dried measles, mumps and rubella strains are then mixed together. In this condition, they remain stable until added to distilled water, for injection. The preparation must then be used within 6 hours. Children under a year old are rarely vaccinated because they are protected by antibodies from their mother (passive immunity). Infants in the UK, for example, receive their first MMR injection aged 15–24 months, and they receive a booster injection between three and five years of age.

Dr Andrew Wakefield, a gastroenterologist employed at the Royal Free Hospital in London at the time (1998), studied 12 children who had all apparently developed autism within 14 days of being given MMR. He published his suggestion (in the medical journal *Lancet*) that for some children, inoculation with the MMR vaccine triggered an inflammatory bowel condition. In this state, the gut might become leaky, allowing 'rogue peptides' into the blood circulation, quickly leading to brain damage – in particular, to autism.

Autism is a life-long disorder in which children withdraw into themselves, reject human contact, and may become impossible to handle. To the families in which it occurs, autism can be most distressing. The film *Rain Man*, in which Dustin Hoffman starred, raised the profile of this disease, so it is more widely recognised now.

Dr Wakefield noted 'We did not prove an association', meaning this was a suggestion that needed following up (and it was). However, his ideas immediately received widespread publicity, and have lead to public anxiety sufficient to cause a marked fall in the take-up of MMR vaccine.

Why do some parents confidently blame MMR for their child's condition?

Autism, when it arises, does so at about 18 months. We have noted the MMR vaccine is first given when children are around 15 months old. When autism does develop soon after inoculation, it is inevitable that there will *appear* to be a connection, whether there is or not (page 683).

To date, no study has produced any evidence that MMR causes autism. Major investigations and follow-up studies have been undertaken with large samples of children in Australia, in Finland, and in Japan. In communities where the opinion is that cases of autism are on the increase, it is noted that the increase began before MMR was introduced, and the incidence of autism continued to rise even when the numbers of children receiving MMR reached a plateau. There is no doubt in the minds of health authorities that vaccination is the safest option for children and their parents. However, after the BSE (bovine spongiform encephalitis, page 596) episode in the UK, governments in the developed world have a major problem in convincing doubters.

Benefits and dangers arising from the general practice of vaccination are listed in Table 12.1.

TOK Link

The coincidence of autism onset and MMR administration may be described as circumstantial evidence. What part, if any, can it play in scientific discovery?

Benefits	Dangers
With the assistance of an effective vaccine, dangerous disease may be eradicated, as was smallpox in 1977. Poliomyelitis may also soon be eradicated; the Salk vaccine was developed in 1955, and the (oral) Sabin vaccine in 1962.	In the worst cases, vaccines can actually cause the disease they are designed to prevent. This is extremely rare.
Vaccines prevent diseases that otherwise may result in very unpleasant and sometimes life-threatening conditions; for example, measles is a major cause of infant death in many less-developed countries.	Vaccines can cause occasional adverse reactions; for example, MMR may trigger tenderness, slight fever and rashes in mild cases, and seizures, allergic reactions, and anaphylactic shock in serious cases (but none has been fatal). These reactions are rare. Whooping cough vaccine may cause brain damage.
Long-term disability from disease can be prevented. Disability for unborn children can be prevented since rubella infections in pregnant women can infect the unborn child, leading to deafness, blindness, brain damage and heart disease.	As the incidence of a disease begins to fall dramatically due to an immunisation programme, rare side-effects of the vaccine, appearing in very a small proportion of those vaccinated, can become unacceptable to the public, leading to serious loss of confidence in what are, in fact, favourable treatments.

Table 12.1 Benefits and dangers of vaccination, today

7 State what we mean by immunity.

Muscles and movement

11.2.1–11.2.8

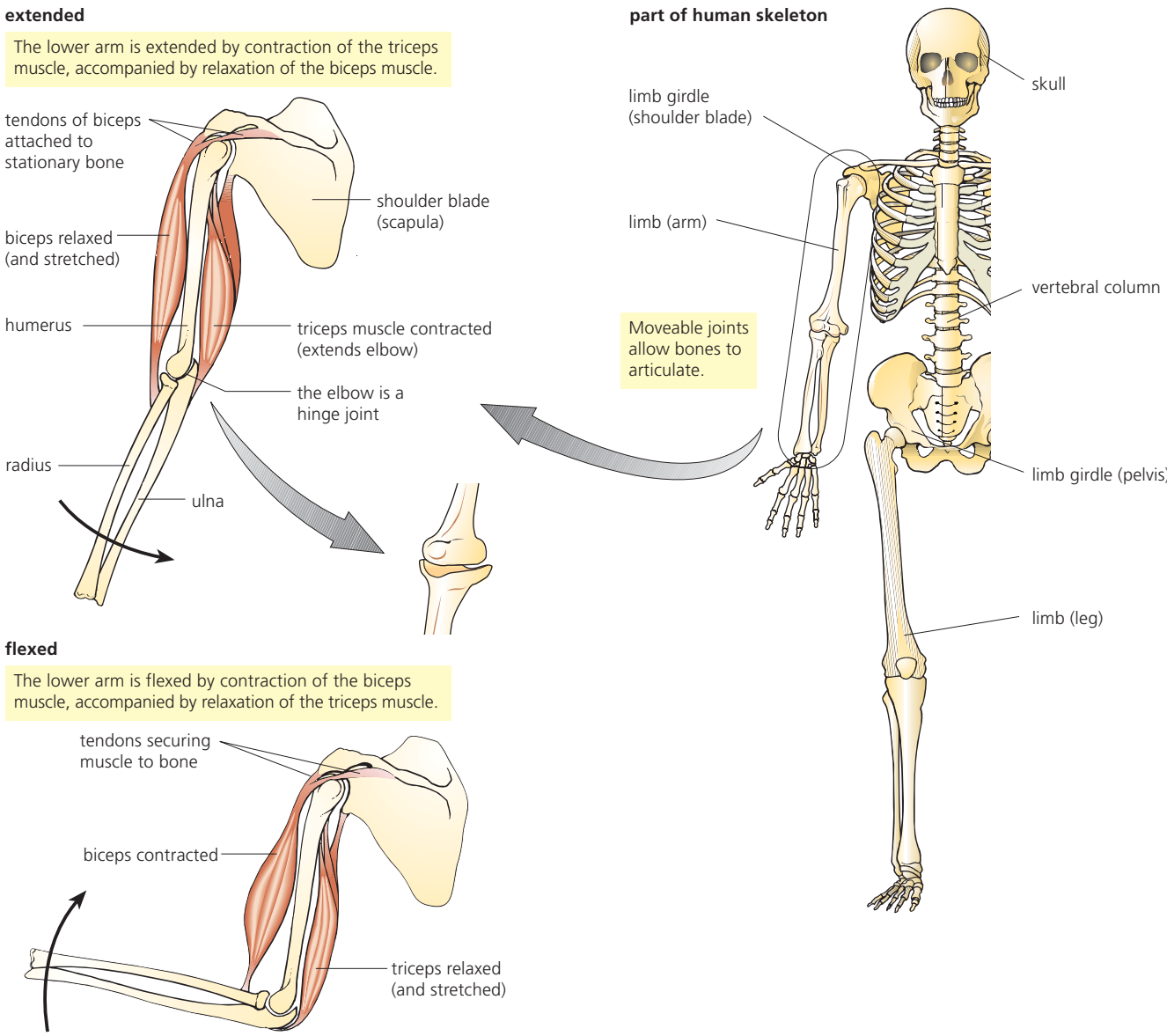
Movement is a characteristic of all living things. It occurs within cells (e.g. cytoplasmic streaming), within organisms (e.g. pumping action of the heart), and as movements of whole organisms, known as **locomotion**. It is the issue of **muscular locomotion** in humans that is examined here.

Roles of the components of our locomotory system

Locomotion is the result of the interactions of nervous, muscular and skeletal systems. The component parts and their roles in locomotion are as follows:

- **Bones** support and partially protect the body parts. Also, they articulate with other bones at **joints**, and they provide anchorage for the muscles. In mammals, the skeleton consists of the axial skeleton (skull and vertebral column) and the appendicular skeleton (limb girdles and limbs). In our study of exercise physiology, we shall be more concerned with the latter.
- **Ligaments** hold bones together, and form protective capsules around the moveable joints. Ligaments are made of fibres of strong but very slightly elastic connective tissue.
- **Muscles** cause movements by contraction. Skeletal muscle is one of three types of muscle in the mammal's body. Skeletal muscles occur in pairs, anchored to bones across joints. They are arranged so that when one of the pair contracts the other is stretched, a system known as antagonistic pairs (Figure 12.10). Contractions of skeletal muscle may either maintain the posture and position of the body, or go on to bring about movement at joints.
- **Tendons** attach muscles to bones at their points of anchorage. They are cords of dense connective tissue.

Figure 12.10 Muscles operate in antagonistic pairs



- **Nerves** are bundles of many nerve fibres of individual nerve cells (called neurones). They connect the central nervous system (brain and spinal cord) with other parts of the body, including the skeletal muscle. Nerve impulses are transmitted in a few milliseconds, travelling along individual nerve fibres to particular points in the body. Nervous control is therefore precise and specific as well as quick. Nerve impulses stimulate muscles to contract, and the nervous system as a whole co-ordinates movement.

How do these components work together to bring about movements at joints?

Moveable joints in the body are of different types, but they all permit controlled movements, and are all examples of **synovial joints** because a thick viscous fluid, the synovial fluid, is secreted and retained in the joint, to help lubricate it. We will consider first a hinge joint, as found in the human elbow.

The human elbow joint

Moveable joints are contained in fibrous **capsules** attached to the immediately surrounding bone. The capsule consists of tough connective tissue but is flexible enough to permit movement. Also present at the joint, but outside the capsule, are ligaments that hold the bones together, preventing dislocations.

At the elbow, the humerus of the upper arm articulates with the radius and ulna of the lower arm at a hinge joint (Figure 12.11 and Table 12.2). At the joint a fluid-filled space separates the articulating surfaces. This fluid, the **synovial fluid**, is secreted by the **synovial membrane**. The fluid nourishes the living cartilage layers that cover the articulating surfaces of the bones as well as serving as a lubricant.

Components	Functions
humerus, radius and ulna	the bones of the skeleton, together with the muscles attached across joints, function as a system of levers to maintain body posture and bring about actions, typically movements
biceps muscle	anchored to shoulder blade and attached to radius so contraction flexes the lower arm (and stretches triceps)
triceps muscle	anchored to shoulder blade and attached to ulna so contraction extends the lower arm (and stretches biceps)
ligaments	hold bones (humerus, radius and ulna) in correct positions at the joint (combats dislocation)
capsule	contains and protects the joint without restricting movement
synovial membrane	secretes synovial fluid
synovial fluid	lubricates the joint, nourishes the cartilage, and removes any (harmful) detritus from worn bone and cartilage surfaces
cartilage	firm, flexible material – a slippery covering that reduces friction

Table 12.2 Components of the elbow joint (a synovial joint) and their functions

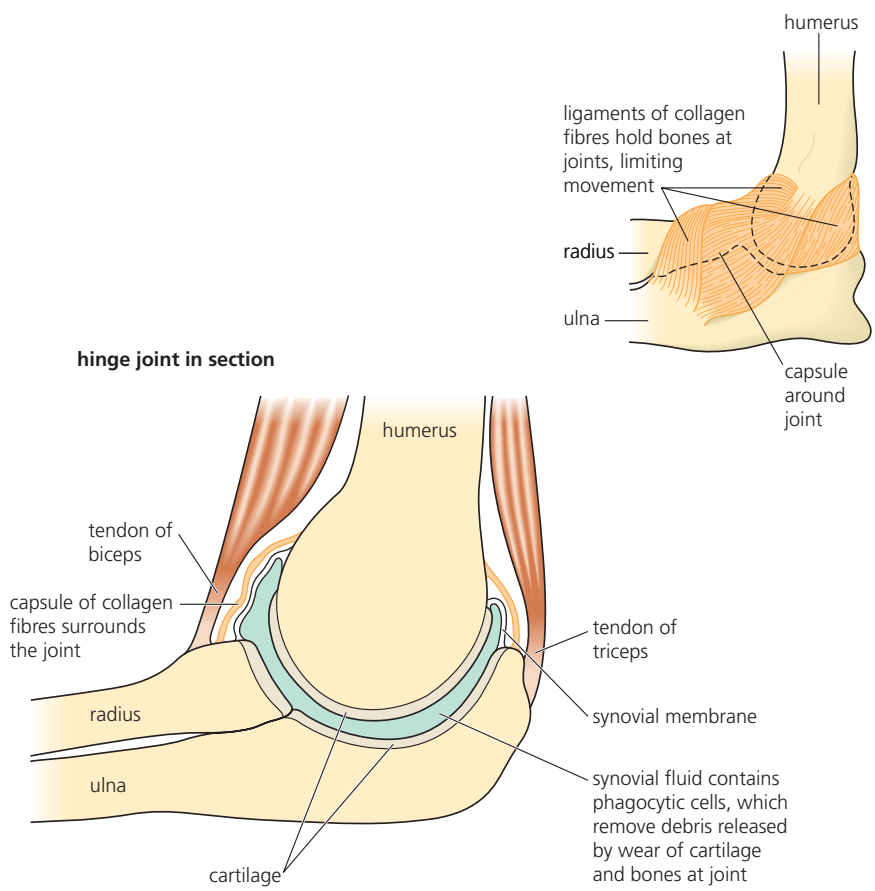


Figure 12.11 The hinge joint of the elbow

Comparative movements at joints

You will have noticed that we have contrasting degrees of movement at our knees and hips (and at shoulder and elbow). This is because of a fundamental difference in the types of joints involved (Figure 12.12 and Table 12.3).

At the hip is a **ball and socket joint**, which is also a synovial joint. In this type of joint, the ball-like surface of one bone (here, the femur of the upper leg) fits into a cup-like depression on another bone (here, the pelvic girdle or hip bone). In the hip joint, the socket is called the acetabulum. This type of joint permits movements in all three planes, a type of movement described as circumduction.

At the knee is a **hinge joint**, which like that of the elbow, restricts movement to one plane. This is because of the shape of the articulating surfaces, and also the ligaments that hold the bones together. Movements at the knee are described as flexions and extensions.

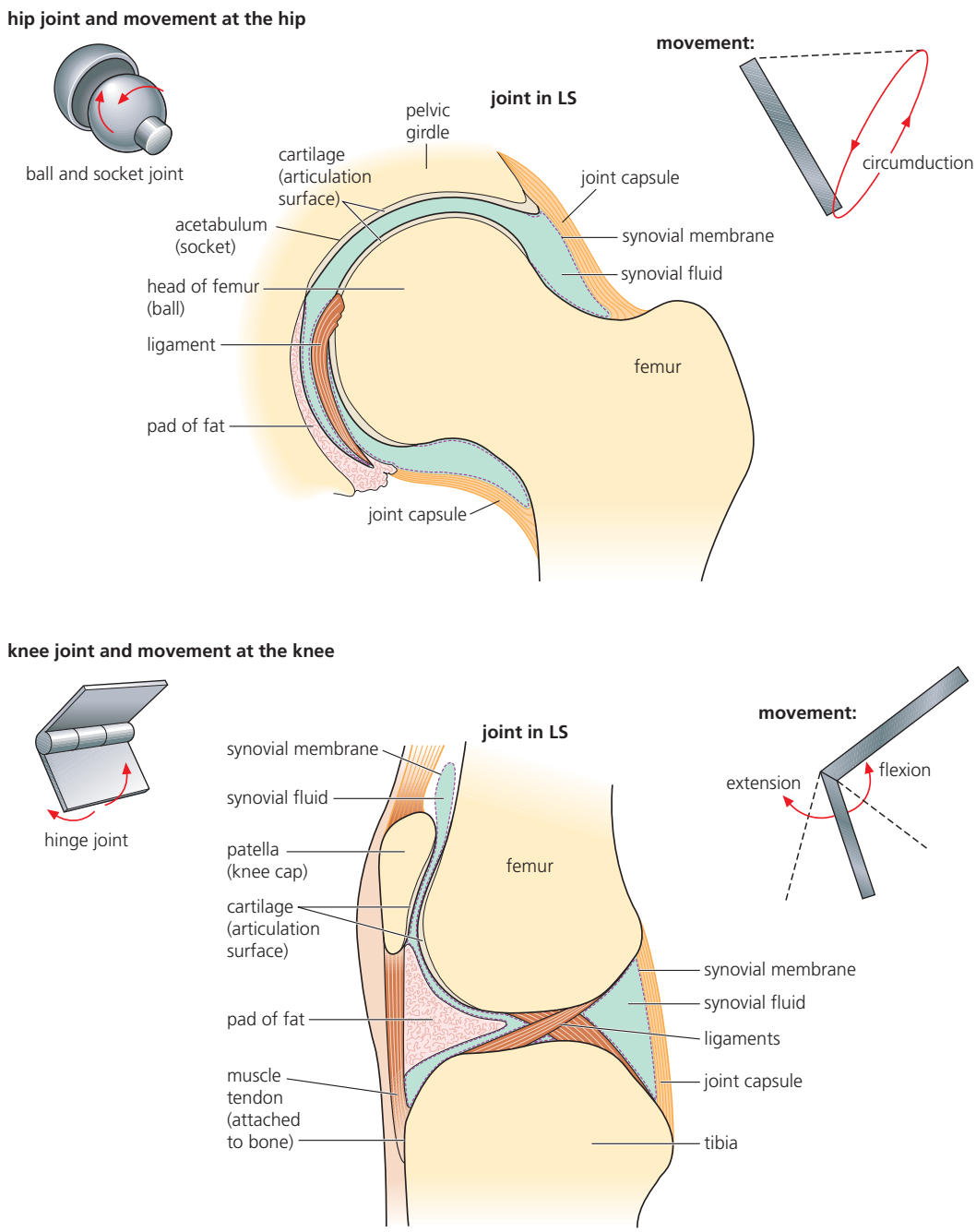


Figure 12.12 Movement at hip and knee joints

hip joint		knee joint
synovial – ball and socket	**type**	synovial – hinge
acetabulum of pelvic girdle and femur (head)	**articulating bones**	femur and tibia
none	**additional bones**	knee cap (patella)
between acetabulum and head of femur	**articulating surface(s)**	between femur and tibia and between femur and patella
circumduction	**permitted movement**	flexion and extension

Table 12.3 Movement of hip and knee joints compared

8 When the tendon that secures the patella is tapped, a knee jerk is observed. Sketch a diagram and **annotate** it to show the structures and the events involved in the knee jerk response.

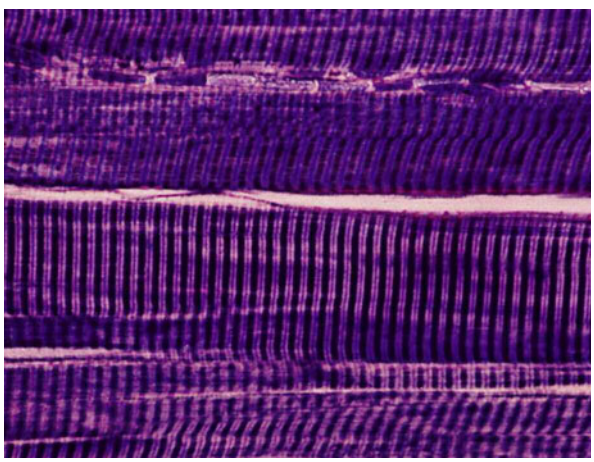
Muscle structure

We have seen that muscles that cause locomotion are attached to the moveable parts of skeletons and are known as **skeletal muscles**. Skeletal muscles are attached by **tendons** and work in antagonistic pairs, as illustrated in Figure 12.10.

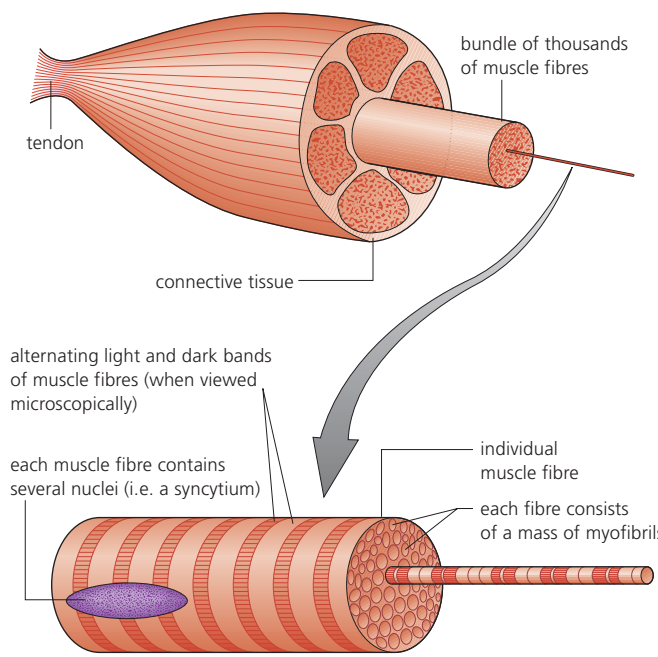
Skeletal muscle consists of bundles of muscle fibres (Figure 12.13). A muscle **fibre** is a long, multinucleate cell. The remarkable feature of a muscle fibre is the ability to shorten to half or even a third of the relaxed or resting length. Fibres appear striped under the light microscope, so skeletal muscle is also known as striated muscle. Actually, each fibre is itself composed of a mass of myofibrils, but we need the electron microscope to see this.

Figure 12.13 The structure of skeletal muscle

photomicrograph of LS voluntary muscle fibres, HP (×1500)



skeletal muscle cut to show bundles of fibres



The ultrastructure of skeletal muscle

By use of the electron microscope we can see that each muscle fibre consists of very many parallel **myofibrils** within a plasma membrane known as the **sarcolemma**, together with cytoplasm. The cytoplasm contains **mitochondria** packed between the myofibrils. The sarcolemma infolds to form a system of transverse tubular endoplasmic reticulum, known as **sarcoplasmic reticulum**. This is arranged as a network around individual myofibrils. The arrangements of myofibrils, sarcolemma and mitochondria, surrounded by the sarcoplasmic membrane, are shown in Figure 12.14.

Figure 12.14 The ultrastructure of a muscle fibre

electron micrograph of TS through part of a muscle fibre, HP (×36 000)

stereogram of part of a single muscle fibre

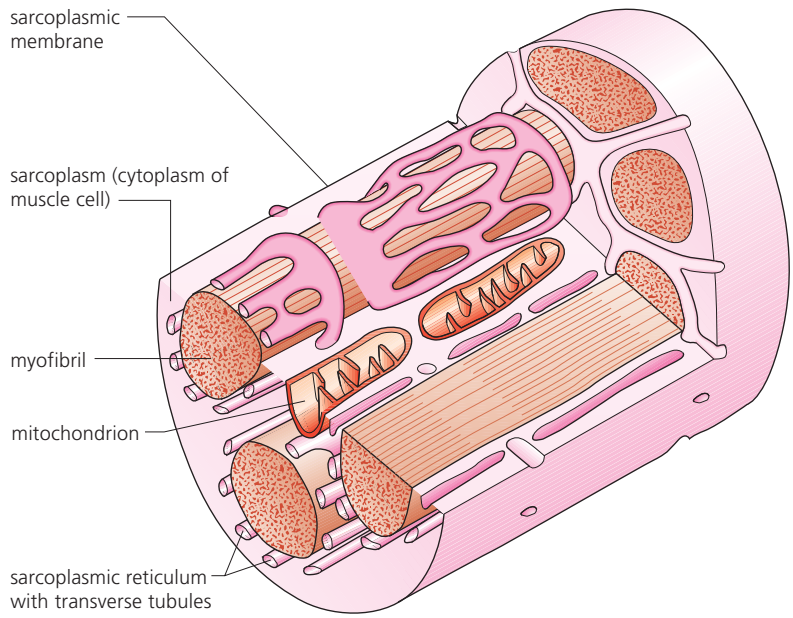
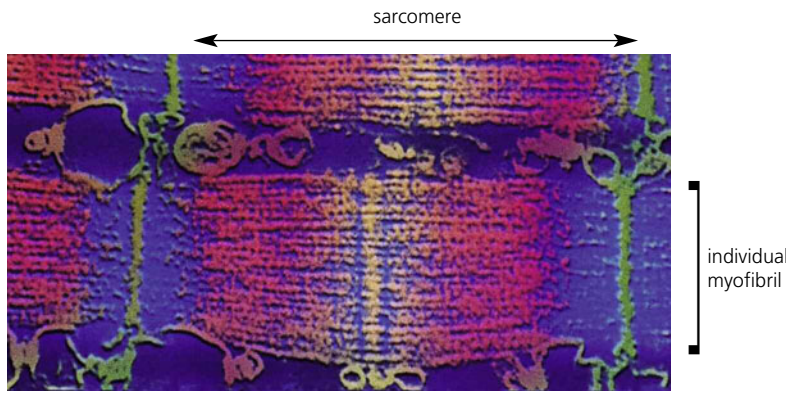
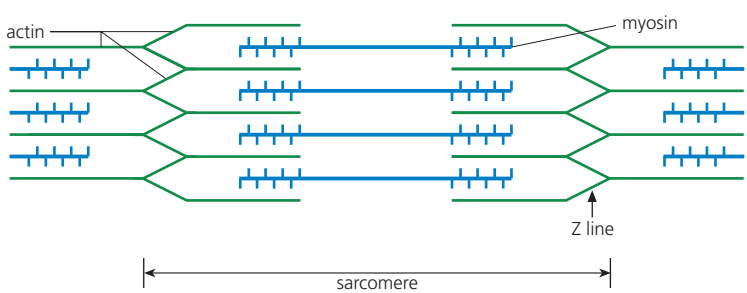


Figure 12.15 The ultrastructure of a myofibril

electron micrograph of LS through part of voluntary muscle fibre, HP (×36 000)



interpretive drawing of the thick filaments (myosin) and thin filaments (actin)



The striped appearance of skeletal muscle is due to an interlocking arrangement of two types of protein filaments, known respectively as **thick** and **thin filaments**, that make up the myofibrils. These protein filaments are aligned, giving the appearance of stripes (alternating **light and dark bands**). This is shown in the more highly magnified electron micrograph and interpretive drawing in Figure 12.15.

9 **Explain** the relationship to a muscle of:

a a muscle fibre
b a myofibril
c a myosin filament.

The **thick filaments** are made of a protein called **myosin**. They are about 15 nm in diameter. The longer, **thin filaments** are made of another protein, **actin**. Thin filaments are about 7 nm in diameter, and are held together by transverse bands, known as **Z lines**. Each repeating unit of the myofibril is, for convenience of description, referred to as a **sarcomere**. So we can think of a myofibril as consisting of **a series of sarcomeres attached end to end**.

Skeletal muscle contracts by sliding of the filaments

Thick filaments lie in the central part of each sarcomere, sandwiched between thin filaments. When skeletal muscle contracts, the **actin** and **myosin filaments slide past each other**, in response to nervous stimulation, causing shortening of the sarcomeres (Figure 12.16). This occurs in a series of steps, sometimes described as a ratchet mechanism. A great deal of **ATP** is used in the contraction process.

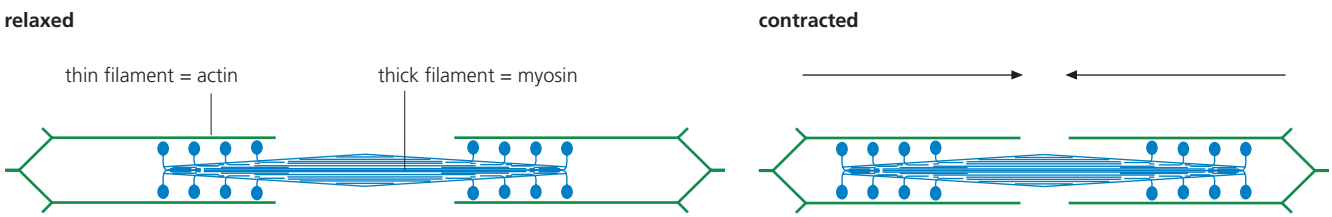
Shortening is possible because the thick filaments are composed of many myosin molecules, each with a **bulbous head** which protrudes from the length of the myosin filament. Along the actin filament are a complementary series of binding sites to which the bulbous heads fit. However, in muscle fibres at rest, the binding sites carry **blocking molecules** (a protein called **tropomyosin**), so binding and contraction are not possible. The contraction of a sarcomere is best described in the following four steps.

1 The myofibril is stimulated to contract by the arrival of an **action potential**. This triggers release of calcium ions from the sarcoplasmic reticulum, to surround the actin molecules. Calcium ions now react with an additional protein present (**troponin**) which, when so activated, triggers the removal of the blocking molecule, tropomyosin. The **binding sites are now exposed**.

Figure 12.16 Muscle contraction of a single sarcomere

change in a single sarcomere in relaxed and contracted myofibril

Muscles contract as the actin and myosin filaments slide between each other, shortening each sarcomere.



electron micrographs of muscle fibres, relaxed (left) and contracted (right)

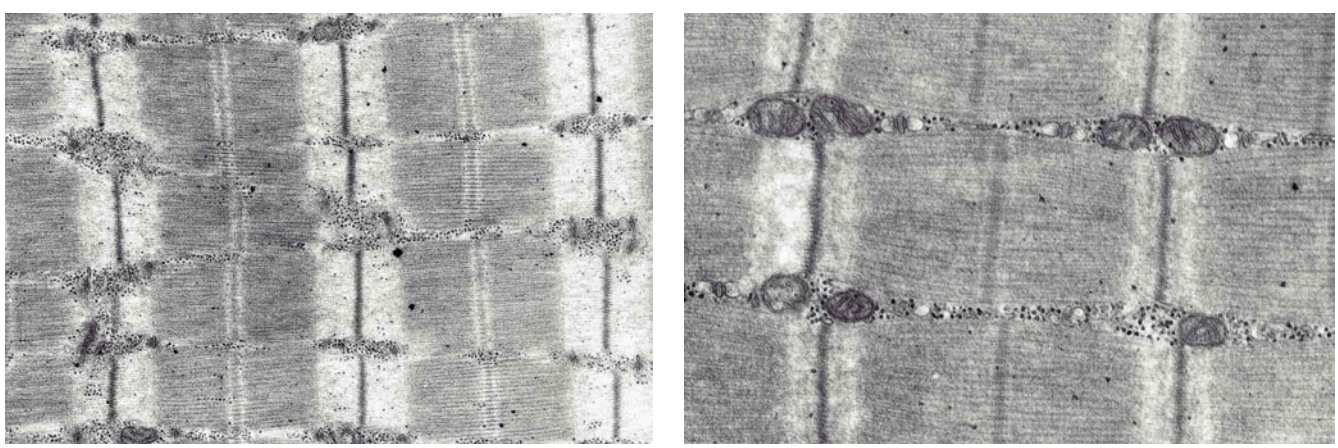
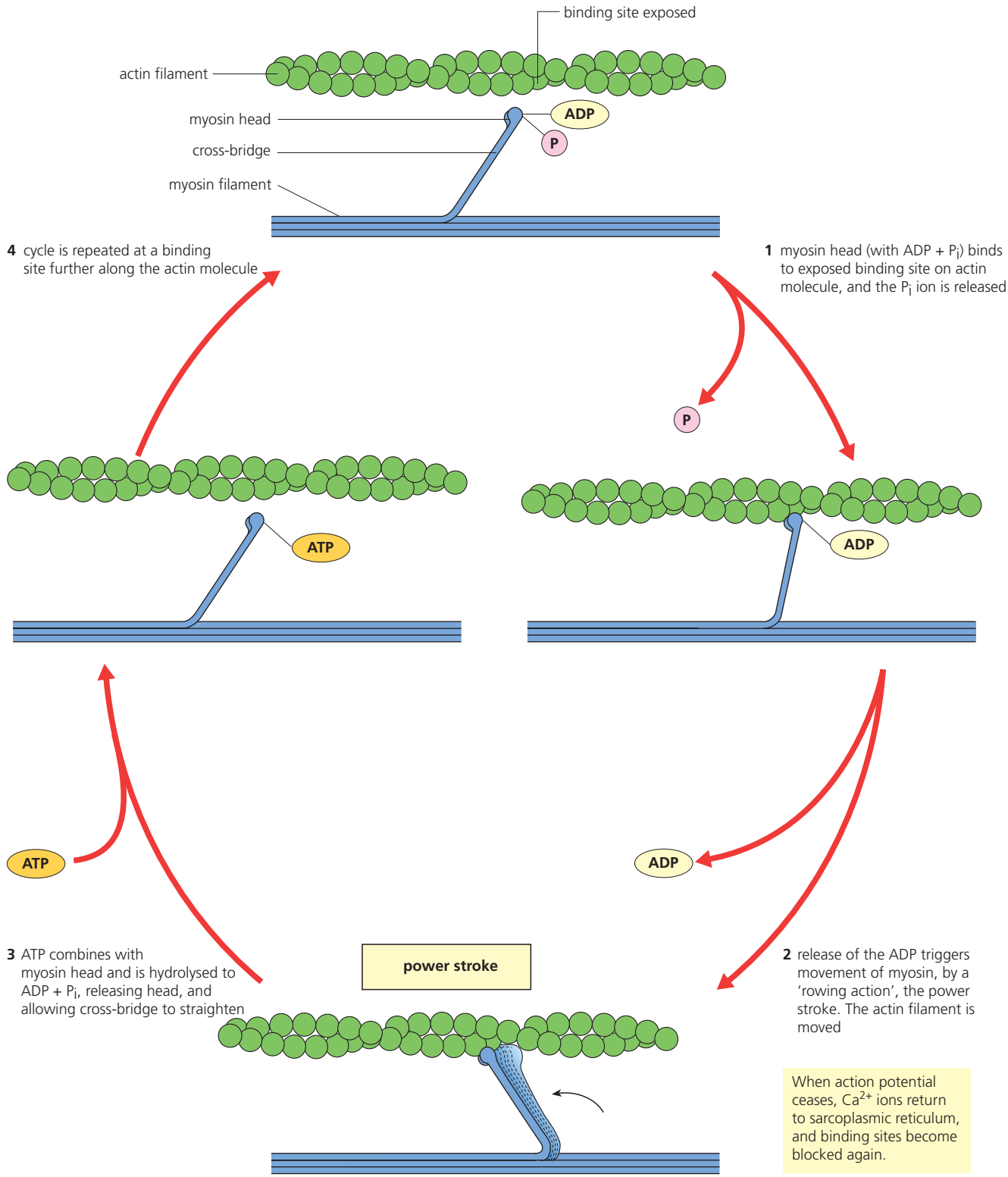


Figure 12.17 The sliding-filament hypothesis of muscle contraction

Arrival of action potential at myofibril releases Ca^{2+} ions from sarcoplasmic reticulum.

Ca^{2+} ions react with a protein (troponin), activating it. Activated troponin reacts with tropomyosin at the binding sites on the actin molecules, thereby exposing the binding sites.

Each myosin molecule has a 'head' that reacts with ATP → ADP + P_i which remain bound.



When action potential ceases, Ca^{2+} ions return to sarcoplasmic reticulum, and binding sites become blocked again.

2 Each bulbous head to which ADP and P_i are attached (called a **charged bulbous head**) reacts with a binding site on the actin molecule beside it. The phosphate group (P_i) is shed at this moment.

3 The ADP molecule is then released from the bulbous head, and this is the trigger for the **rowing movement** of the head, which tilts by an angle of about 45°, pushing the actin filament along. At this step, the **power stroke**, the myofibril has been shortened (**contraction**).

4 Finally, a fresh molecule of **ATP binds** to the bulbous head. The protein of the bulbous heads includes the enzyme ATPase, which catalyses the hydrolysis of ATP. When this reaction occurs, the ADP and inorganic phosphate (P_i) formed remain attached, and the **bulbous head is now 'charged'** again. The charged head detaches from the binding site and straightens.

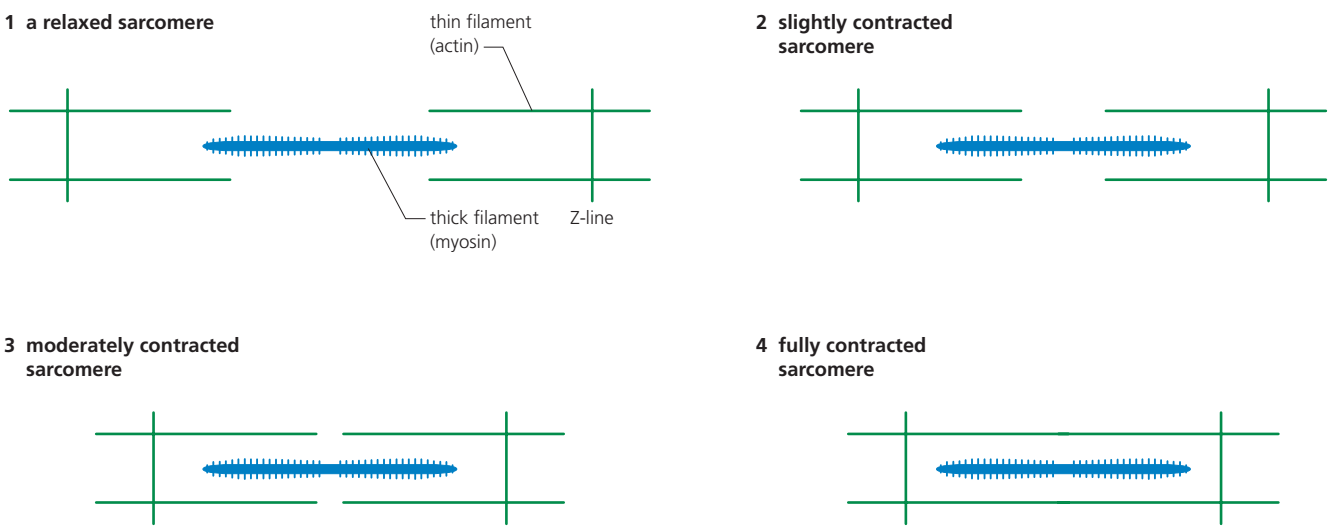
This cycle of movements is shown is Figure 12.17. The cycle is repeated many times per second, with thousands of bulbous heads working along each myofibril. ATP is rapidly used up, and the muscle may shorten by about 50% of its relaxed length.

Muscles, controlled movements, and posture

Muscles are involved in maintaining body posture and in subtle, delicate movements, as well as in vigorous or even violent actions. Consequently, nervous control of muscle contraction may cause relaxed muscle to contract slightly, moderately or fully, depending on the occasion. In these differing states of contraction, the overall lengths of the sarcomeres are changed accordingly. These relative changes are illustrated diagrammatically in a single sarcomere in Figure 12.18. Below them is a representation of part of a myofibril, seen at a particular stage of contraction. This representation is not diagrammatic, but is based on an interpretation of a TEM. (See the TEMs in Figure 12.16, for example.) SAQ 10 is concerned with analysing the state of contraction of myofibrils.

10 a Identify the approximate state of contraction illustrated in the sketch of a TEM of a myofibril shown in Figure 12.18.
b Sketch a similar myofibril, fully contracted. Label your drawing (sarcomere, Z lines, light band and dark band).

Figure 12.18 Analysing states of contraction in striated muscle fibres



sketch representing a myofibril at a particular stage in contraction, based on interpretation of a TEM of a sample of striated muscle

The kidney – excretion and osmoregulation

11.3.1–11.3.9

The chemical reactions of metabolism produce by-products, some of which would be toxic if allowed to accumulate in the organism.

Excretion is the removal from the body of the waste products of metabolism. It is a characteristic activity of all living things. Metabolites present in excessive concentrations are also excreted.

In mammals, excretion plays an important part in the process by which the internal environment is regulated to maintain more or less constant conditions (homeostasis). Associated with excretion, and very much part of homeostasis, is the process of osmoregulation.

Osmoregulation is the maintenance of a proper balance of water and dissolved substances in the organism.

Why is the issue of nitrogenous excretion so important in animals?

The reason is that animals break down excess proteins and amino acids by a process called **deamination** – the first step is the removal of the amino groups. Once removed, a likely initial product is **ammonia**, which is extremely toxic to cells if it is allowed to accumulate. So, the amino group is converted into some other nitrogenous excretory product for safe disposal in most animals. **Urea** is a chemical compound synthesised from carbon dioxide and ammonia. In dilute solution, urea is safely excreted from the body in mammals.

The human kidney – an organ of excretion and osmoregulation

The kidneys regulate the internal environment by constantly adjusting the blood composition. The waste products of metabolism are transported from the metabolising cells by the blood circulation, removed from the blood in the kidneys, and excreted in a solution called urine. The concentrations of inorganic ions, such as Na^+ and Cl^-, and of water in the body are also regulated in the kidneys, which are, therefore, also organs of osmoregulation.

This function of the kidneys is another example of **homeostasis by negative feedback** (page 218), a characteristic of the body of mammals. For a summary of osmoregulation by the kidney see Table 12.4, page 376.

The position of the kidneys in humans is shown in Figure 12.19. Each kidney is served by a renal artery and drained by a renal vein. Urine from the kidney is carried to the bladder by the **ureter**, and (occasionally) from the bladder to the exterior by the **urethra**, when the bladder

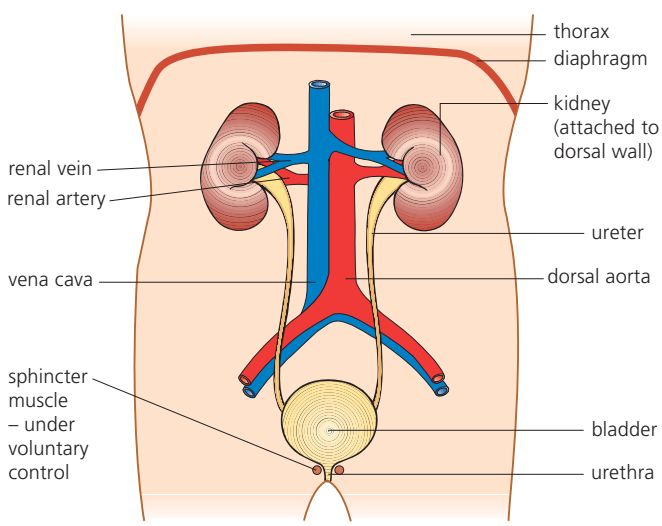
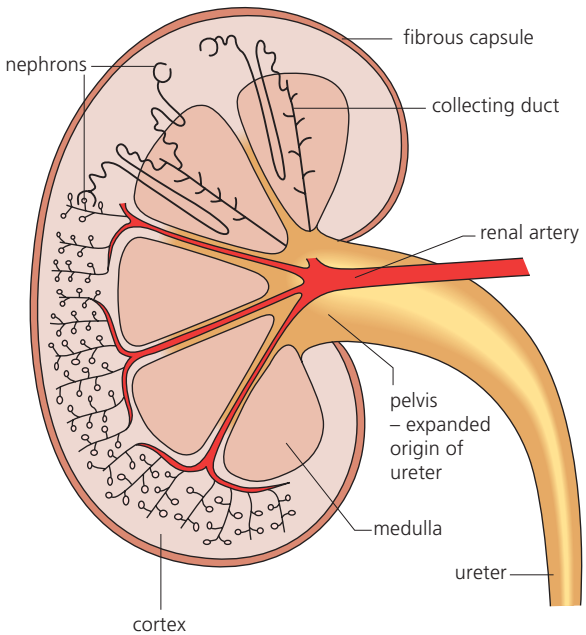


Figure 12.19 The human urinary system

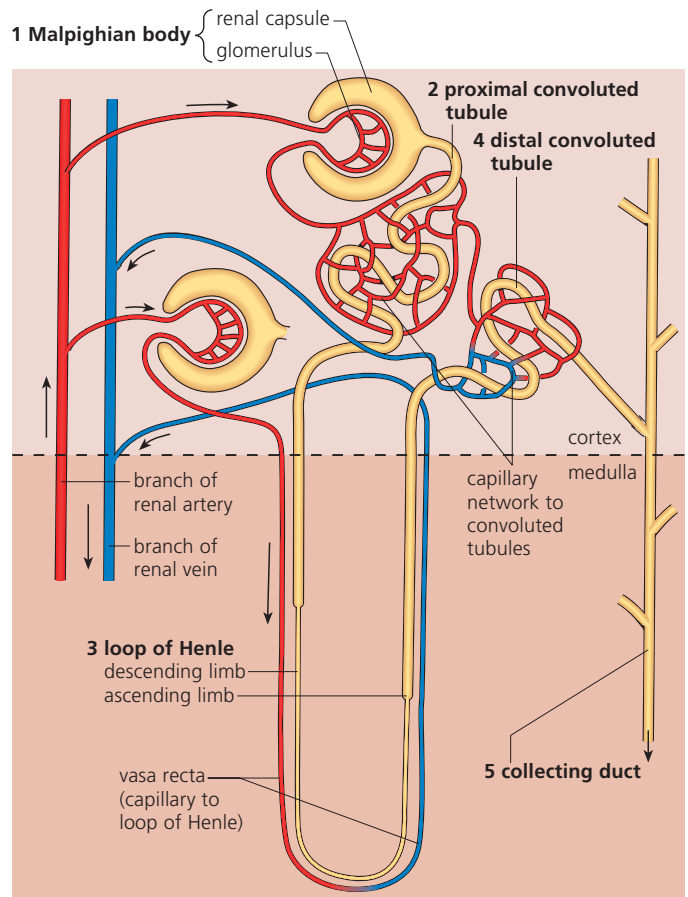
sphincter muscle is relaxed. Together these structures are known as the urinary system. In section, a kidney can be seen to consist of an outer **cortex** and inner **medulla**. These are made up of a million or more tiny tubules, called **nephrons**, together with their blood supply. Part of a nephron is in the cortex and part in the medulla. A nephron is a thin-walled tubule about 3 cm long. **Capillary networks** are crucial to the functioning nephron (Figure 12.20).

Figure 12.20 The kidney and its nephrons – structure and roles

LS through kidney showing positions of nephrons in cortex and medulla



nephron with blood capillaries



Roles of the parts of the nephron

1 Malpighian body = ultrafiltration
2 proximal convoluted tubule = selective reabsorption from filtrate
3 loop of Henle = water conservation
4 distal convoluted tubule = pH adjustment and ion reabsorption
5 collecting duct = water reabsorption

11 Distinguish between excretion, egestion, osmoregulation and secretion by means of both definitions and examples.

The formation of urine

In humans, about 1.0–1.5 litres of urine are formed each day, typically containing about 40–50 g of solutes of which **urea** (about 30 g) and **sodium chloride** (up to 15 g) make up the bulk. The nephron produces urine in a continuous process which we can conveniently divide into five steps, to show how the blood composition is so precisely regulated. These steps are discussed in turn below.

Step 1: Ultrafiltration in the renal capsule

In the glomerulus, water and relatively small molecules of the blood plasma, including useful ions, glucose and amino acids, are forced out of the capillaries, along with urea, into the lumen of the capsule. This is described as **ultrafiltration** because it is powered by the pressure of the blood, which drives substances through an extremely fine sieve-like structure.

The **blood pressure** here is high enough for ultrafiltration because the input capillary (afferent arteriole) is wider than the output capillary (efferent arteriole). The 'sieve' is made of two layers of cells (the endothelium of the capillaries of the glomerulus and the epithelium of the capsule), between which is a basement membrane.

You can see this arrangement in Figure 12.21.

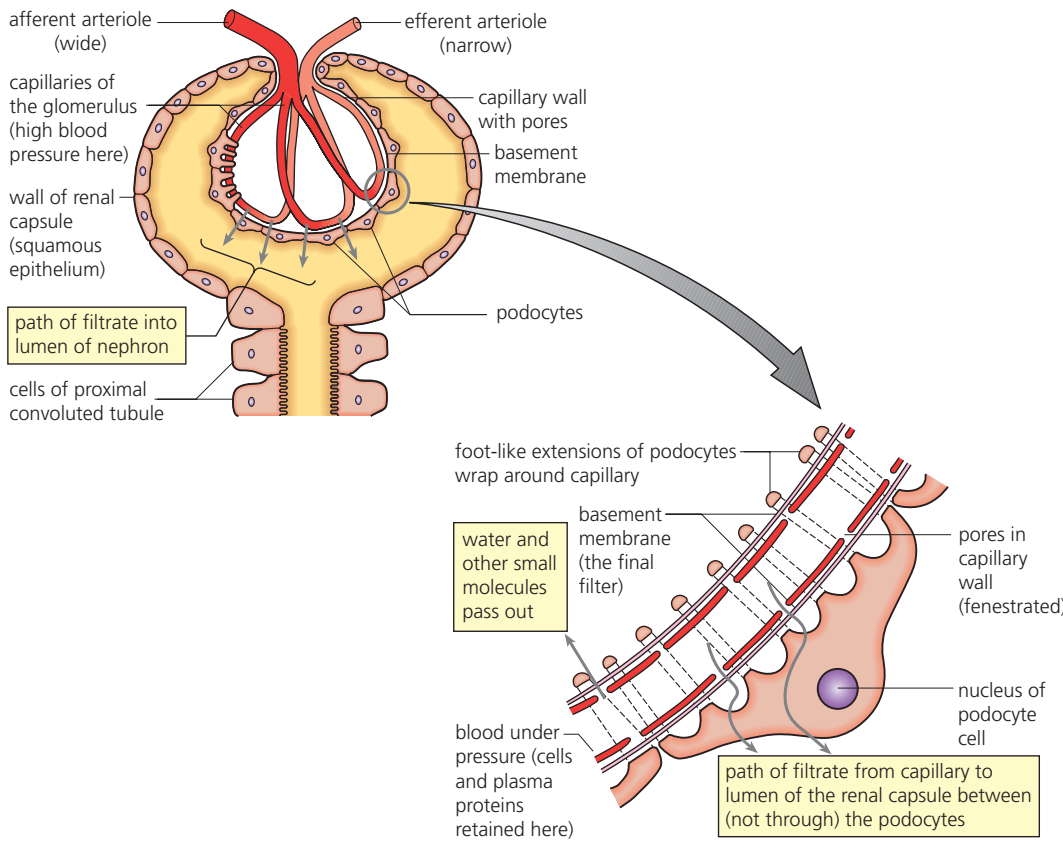


Figure 12.21 The site of ultrafiltration

Notice that the cells of the capsule wall are called **podocytes** because they have foot-like extensions that form a network with tiny slits between them (a situation we call **fenestrated**). Similarly, the endothelium of the capillaries has **pores**, too. This detail has only become apparent from studies using the electron microscope – these filtration gaps are very small indeed.

12 State the source of energy for ultrafiltration in the glomerulus.

The entire contents of blood are *not* forced out. Not only are blood cells retained, but the majority of blood proteins and polypeptides dissolved in the plasma are also retained in the circulating blood. This is because of the presence of the **basement membrane**.

Step 2: Selective reabsorption in the proximal convoluted tubule

The proximal convoluted tubule is the longest section of the nephron. The walls are one cell thick and are packed with mitochondria (ATP is required for the active transport). The cell membrane in contact with the filtrate has a brush border of microvilli which enormously increase the surface area for reabsorption. A large part of the filtrate is reabsorbed into the capillary network here (Figure 12.22).

The individual mechanisms of transport are:

- movement of water by **osmosis**;
- **active transport** of glucose and amino acids across membranes;
- movement of mineral ions by a combination of **active transport**, **facilitated diffusion**, and some **exchange of ions**;
- **diffusion** of urea;
- movement of proteins by **pinocytosis**.

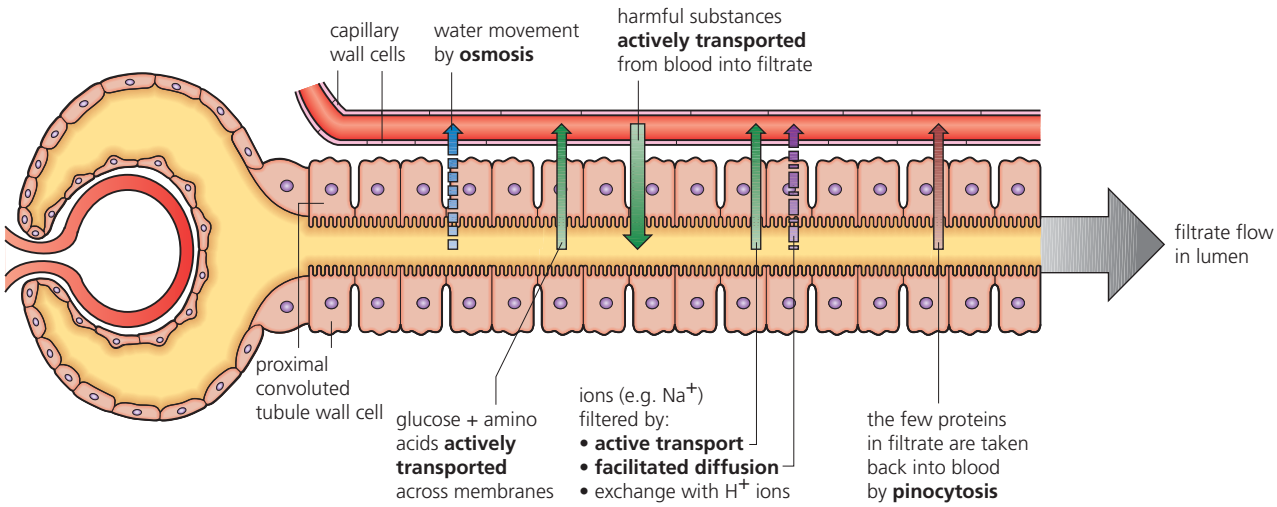


Figure 12.22 Reabsorption in the proximal convoluted tubule

13 Cells of the walls of the proximal convoluted tubule have a brush border. **Describe** what this means, and **explain** how it helps in tubule function.

Step 3: Water conservation in the loop of Henle

Urea is expelled from the body in solution, so water loss in excretion is inevitable. However, mammals are able to form urine that is more concentrated than the blood (when necessary), thereby reducing the water loss to a minimum. The role of the loop of Henle with its **descending** and **ascending limbs**, together with a parallel blood supply, the **vasa recta**, is to create and maintain a high concentration of salts in the tissue fluid in the **medulla of the kidney**. This is brought about by a **countercurrent multiplier mechanism**. It is the building up of a high concentration of salts in the tissue of the medulla that causes water to be reabsorbed from the filtrate in the collecting ducts. The collecting ducts run through the medulla.

The roles of the vasa recta are to:

- absorb water that has been absorbed into the medulla at the collecting ducts;
- remove carbon dioxide and deliver oxygen to the metabolically active cells of the loop of Henle without removing the accumulated salts from the medulla.

Figure 12.23 explains how the countercurrent mechanism works. Notice that the descending and ascending limbs lie close together.

Look first at the second half of the loop, the ascending limb.

Here, sodium and chloride ions are pumped out into the medulla but water is retained inside the ascending limb. Opposite, the descending limb is permeable here, so sodium and chloride ions diffuse in. Water passes out into the medulla tissue, due to the salt concentration in the medulla. As the filtrate flows down the descending limb, this water loss increases the salt concentration in the loop, making the filtrate more concentrated.

Consequently, sodium ions and chloride ions diffuse out down their concentration gradient, around the 'hairpin' zone at the base of the descending limb, adding to the concentration of ions in the medulla. How this concentration helps in the formation of concentrated urine is explained in step 5 below.

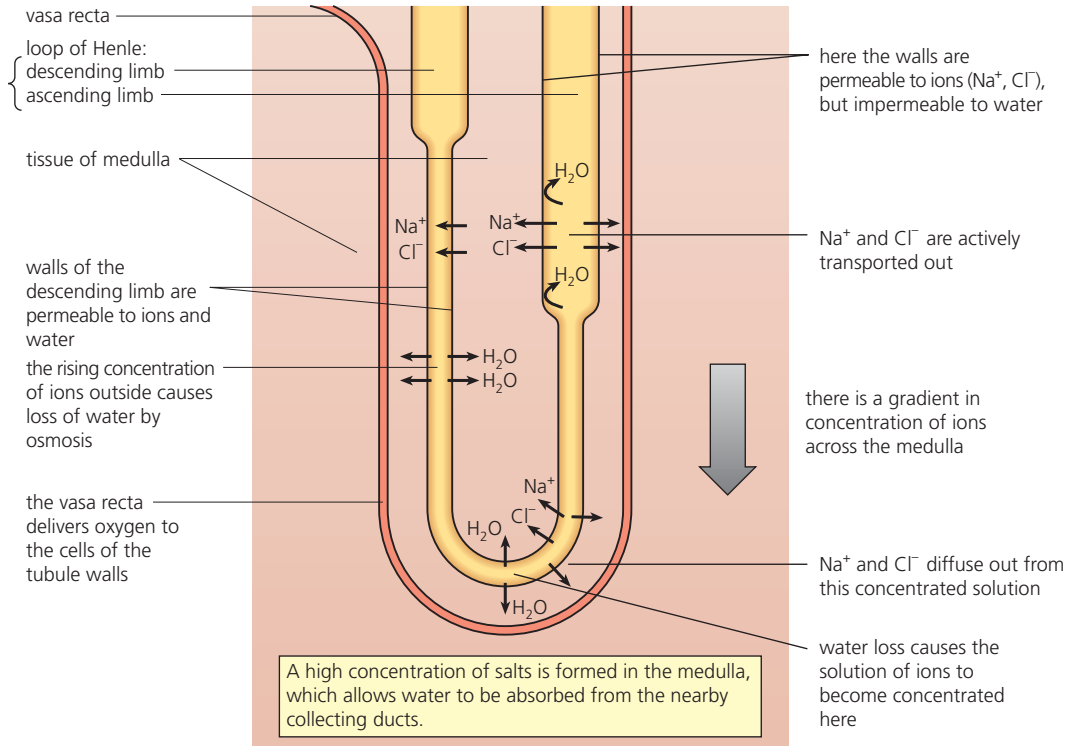


Figure 12.23 The functioning loop of Henle

Step 4: Blood pH and ion concentration regulation in the distal convoluted tubule

Here the cells are of the same structure as those of the proximal convoluted tubule, but their role is to adjust the composition of the blood, and in particular the **pH**. An initial tendency for the pH of the blood to change is buffered by the blood proteins, but if the blood does begin to deviate from pH 7.4, then the concentration of hydrogen ions and hydroxyl ions in the blood is adjusted, along with the concentration of hydrogencarbonate ions. Consequently, blood pH does not vary outside the range pH 7.35–7.45, but the pH of urine varies from pH 4.5 to pH 8.2.

Also in the distal convoluted tubule, the selective reabsorption of ions useful in metabolism occurs from the filtrate.

Step 5: Water reabsorption in the collecting ducts

The collecting ducts are where the **water content of the blood** (and therefore of the whole body) is regulated (Figures 12.24 and 12.25). When the water content of the blood is low, **antidiuretic hormone** (**ADH**) is secreted from the posterior pituitary gland. When the water content of the blood is high, little or no ADH is secreted.

The permeability of the walls of the collecting ducts to water is variable (a case of facilitated diffusion) – the presence of ADH causes the walls of the collecting ducts to be fully permeable. This allows water to be withdrawn from the filtrate of the tubule into the medulla, due to the high concentration of sodium and chloride ions there (see step 3 above). This water is taken up and redistributed in the body by the blood circulation, and only small amounts of concentrated urine are formed. Meanwhile, the ADH circulating in the blood is slowly removed at the kidneys.

When no ADH is secreted, the walls of the collecting ducts become less permeable. The result is that large quantities of very dilute urine are formed.

14 Predict in what circumstances in the body ADH is released.

Figure 12.24 Water reabsorption in the collecting ducts

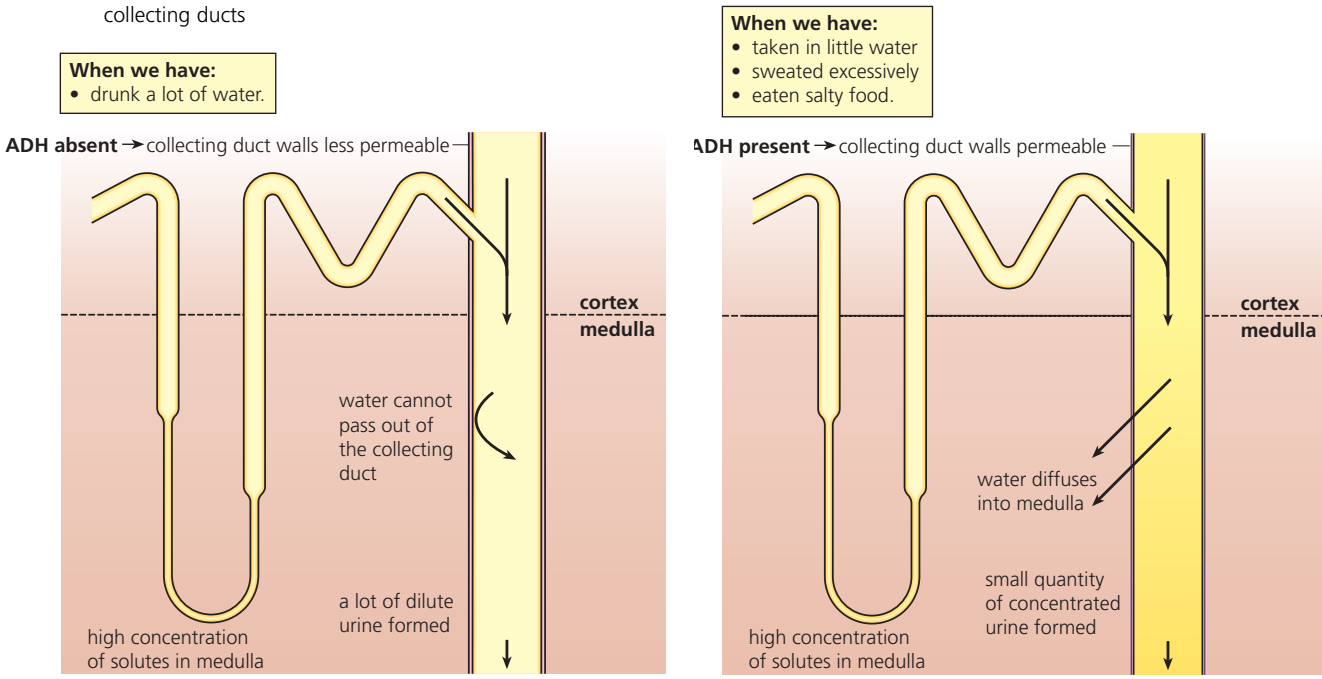
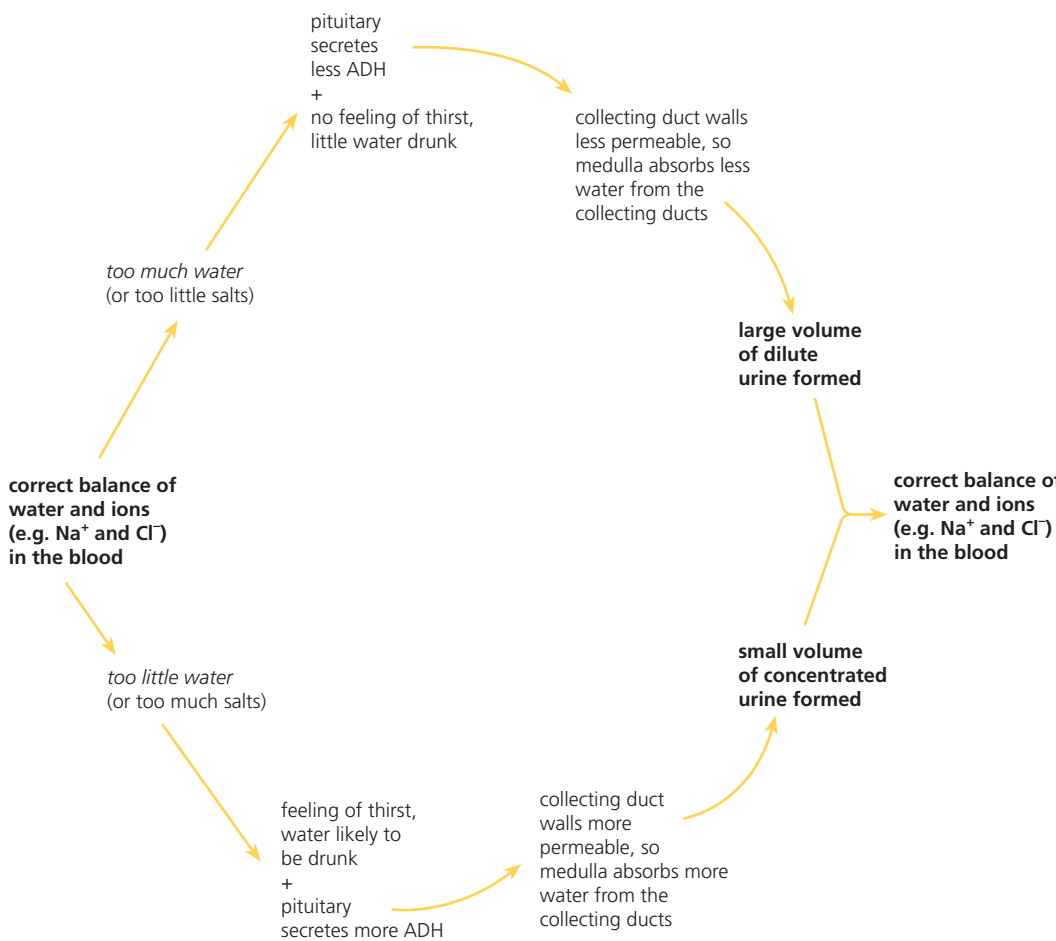


Figure 12.25 Homeostasis by osmoregulation in the kidneys – a summary



Differences in composition of blood plasma, glomerular filtrate and urine

The composition of urine excreted from the body is inevitably very variable. It is greatly influenced by our diet (including salt intake and amount of protein consumed), water intake, degree of physical activity, environmental conditions, water loss by other routes (particularly in sweat), and of course by our state of health, as the urine of diabetics illustrates (see below).

On the other hand, the composition of the blood is held more or less constant. This is due to the efficiency of the homeostatic mechanisms of the body, particular those operating at the kidney tubules.

Similarly, the composition of the filtrate passing into the renal capsule is remarkably constant. This is a result of ultrafiltration in the glomerulus, the pressure of the blood, and the sizes of blood proteins and polypeptides dissolved in the plasma – they are mostly too large to be filtered.

In Table 12.4, we have evidence of the power of the ultrafiltration mechanism and the scale of selective reabsorption. And we see the largest divergence in concentration when the composition of the filtrate is compared with composition of urine, formed in one day.

Substance	Total amount in the plasma	Filtered (per day)	Reabsorbed (per day)	Urine (excreted per day)
water	3 litres	180 litres	178–179 litres	1–2 litres
glucose	2.7 g	162.0 g	162.0 g	0.0 g
urea	0.9 g	54.0 g	27.0 g	27.0 g**
proteins + polypeptides	200.0 g	2.0 g	1.9 g	0.1 g
Na^+ ions	9.7 g	579.0 g	575.0 g	4.0 g
Cl^- ions	10.7 g	640.0 g	633.7 g	6.3 g

Table 12.4 Composition of blood, glomerular filtrate and urine

**some is also secreted by the cells of the tubule into the urine

The composition of urine of diabetic patients

The diseases known as diabetes were introduced in Chapter 7, page 224. In either form – early-onset diabetes (type I diabetes) and diabetes mellitus (type II diabetes) – blood glucose levels are erratic and generally frequently well above the normal blood glucose level of about 90 mg glucose per 100 cm^3 of blood. This is especially the case after a meal has been digested and the products of digestion absorbed in the small intestine.

An outcome of this condition is the failure of the kidney tubules to reabsorb all the glucose that is forced out of the blood plasma during ultrafiltration. Consequently, the urine of diabetic patients typically has raised and erratic concentrations of glucose. This is one symptom that indicates the likelihood of a patient being a diabetic.

Reproduction

11.4.1–11.4.15

In sexual reproduction, sex cells (**gametes**) fuse to form a zygote which then grows into a new individual. The gametes are produced in paired glands called gonads – male gametes or **sperms**, are formed in **testes** (Figure 7.49, page 225); female gametes, ova or **oocytes** (singular, ovum or **oocyte**) are formed in **ovaries** (Figure 7.50, page 226).

15 **Suggest** what evidence there is that, in nature, only some individuals of any species succeed in reproducing themselves.

The process of gamete formation, known as **gametogenesis**, involves not only mitosis but also meiotic division, thereby halving the normal chromosome number. That is, gametes are haploid. **Fertilisation** restores the diploid number of chromosomes.

The structure of the gonads and the steps of gamete formation are what we consider first.

Gametogenesis

In gametogenesis, many gametes are produced, although relatively few of them are ever used in reproduction. The processes of gamete formation in testes and ovaries have a common sequence of phases.

First, there is a **multiplication phase** in which the gamete mother cells divide by mitotic cell division (Figure 1.33, page 32). This division is then repeated to produce many cells with the potential to become gametes.

Secondly, each developing sex cell undergoes a **growth phase**.

Third and finally, comes the **maturation phase**. This involves meiosis and results in the formation of the haploid gametes. The products of meiosis I are secondary **spermatocytes** and secondary **oocytes**, and the products of meiosis II are **spermatids** and **ova**. The steps of meiosis are described on page 328, Figure 11.1.

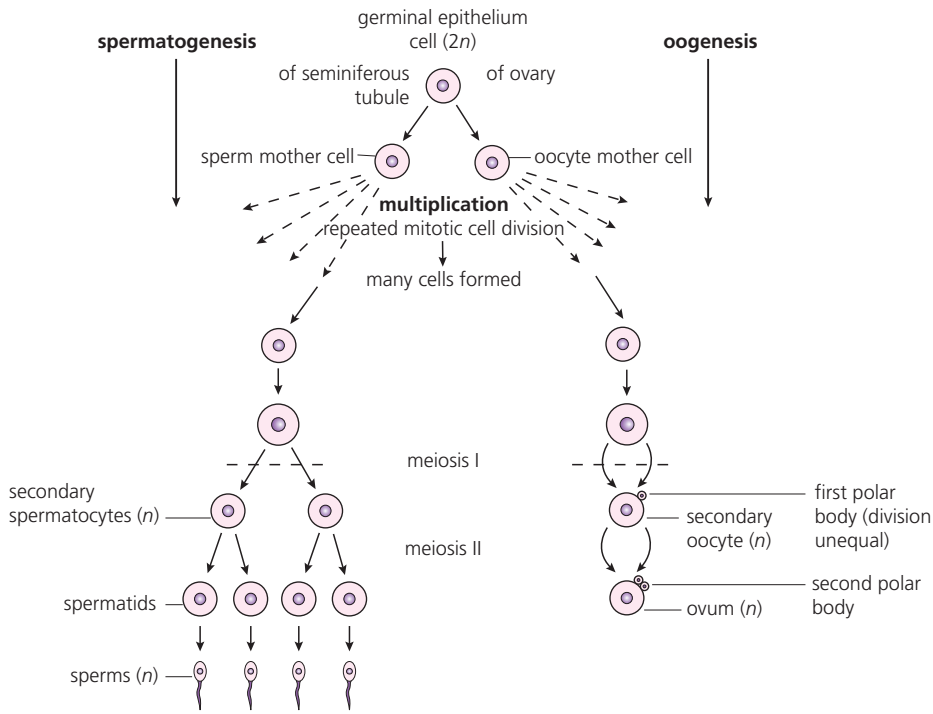
These phases in sperm and ova production are summarised in Figure 12.26; the differences between gametogenesis in testis and ovary are listed in Table 12.5 (page 382).

Figure 12.26 The phases and changes during gametogenesis

multiplication phase

growth phase

maturation phase



The structure and functioning of the testis

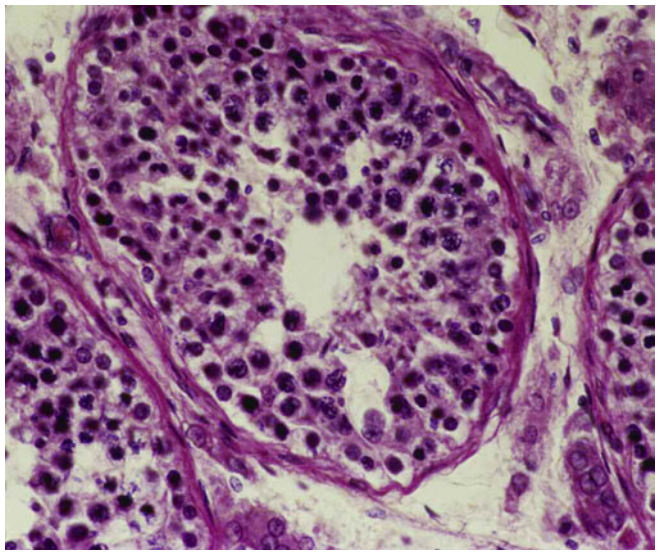
In the human fetus, testes develop high on the posterior abdominal wall and migrate to the scrotum in about the seventh month of pregnancy. In the scrotum, the paired testes are held at a temperature 2–3 °C below body temperature after birth. This lower temperature is eventually necessary for sperm production – testes that fail to migrate do not later produce sperms.

Spermatogenesis begins in the testes at puberty and continues throughout life. Each testis consists of many **seminiferous tubules**. These are lined by germinal epithelial cells which divide repeatedly. Tubules drain into a system of channels leading to the epididymis, a much coiled tube which leads to the **sperm duct**. Between the individual seminiferous tubules is connective tissue containing blood capillaries, together with groups of **interstitial cells**. These latter cells are hormone-secreting (the testis is also an endocrine gland). Testes are suspended by a spermatic cord containing the sperm duct and blood vessels (Figure 12.27).

In the seminiferous tubules, the **germinal epithelium cells** are attached to the basement membrane, along with the **nutritive cells**. Cells from the subsequent steps of sperm production (spermatogonia, primary spermatocytes, secondary spermatocytes and spermatids) occur lodged in the surface of Sertoli cells (nutritive cells) on which they are dependent until they mature into spermatozoa (sperms) (Figures 12.28 and 12.29).

Figure 12.27 Photomicrograph of testis tissue in section, and interpretive drawing

photomicrograph of TS of seminiferous tubule, LP (×500)



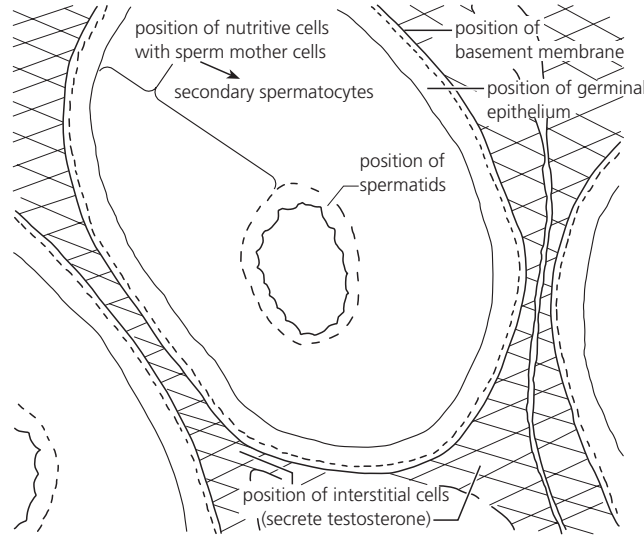
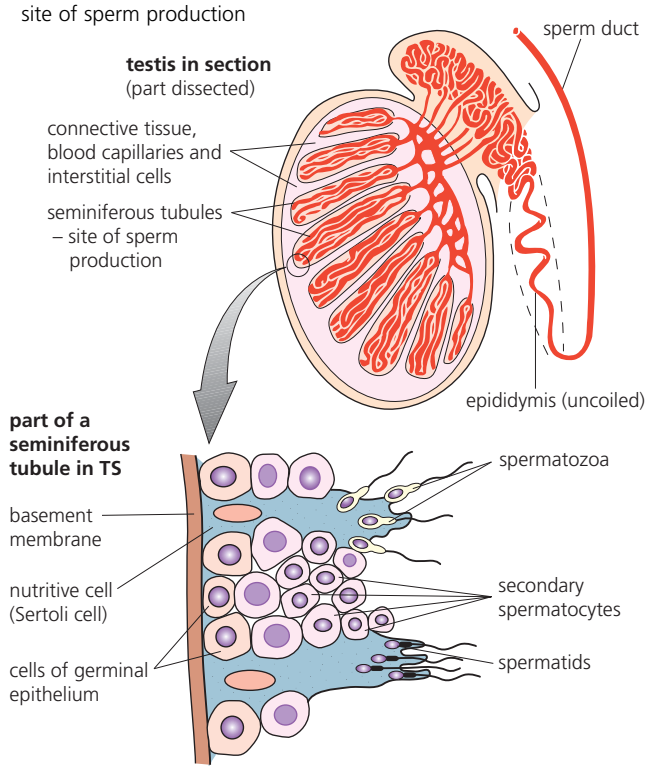
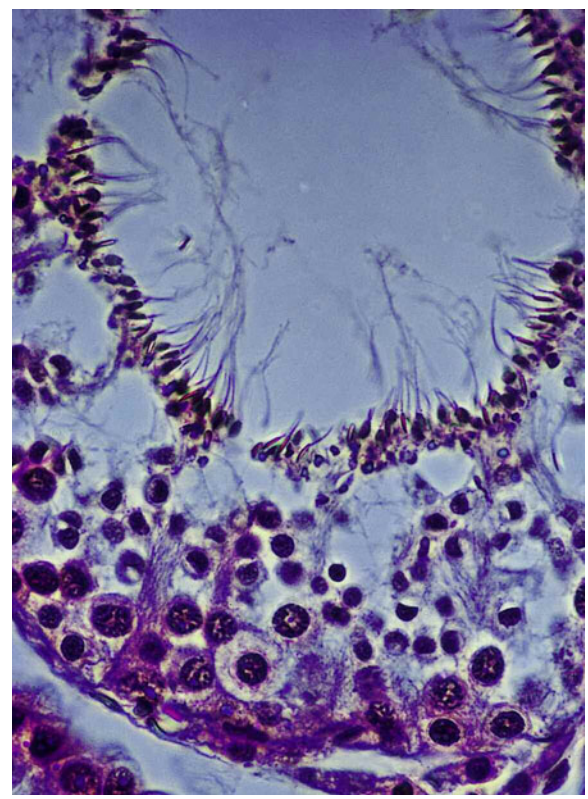


Figure 12.28 Structure of a seminiferous tubule – site of sperm production



photomicrograph of TS of seminiferous tubule, HP (×1000)



The mature sperms, and the production of semen

The sperms are immobile when first formed. From the seminiferous tubules they pass into the much-coiled epididymis where maturation is completed and storage occurs. During an ejaculation, the sperms are moved by waves of contraction in the muscular walls of the sperm ducts. Sperms are transported in a nutritive fluid secreted by glands, mainly the seminal vesicles and prostate gland. These glands add their secretions just at the point where the sperm ducts join

with the urethra, below the base of the penis (Figure 7.49, page 225). As well as providing nutrients for the sperms, semen is a slightly alkaline fluid, the significance of which we will return to later. During an ejaculation, the sphincter muscle at the base of the bladder is closed.

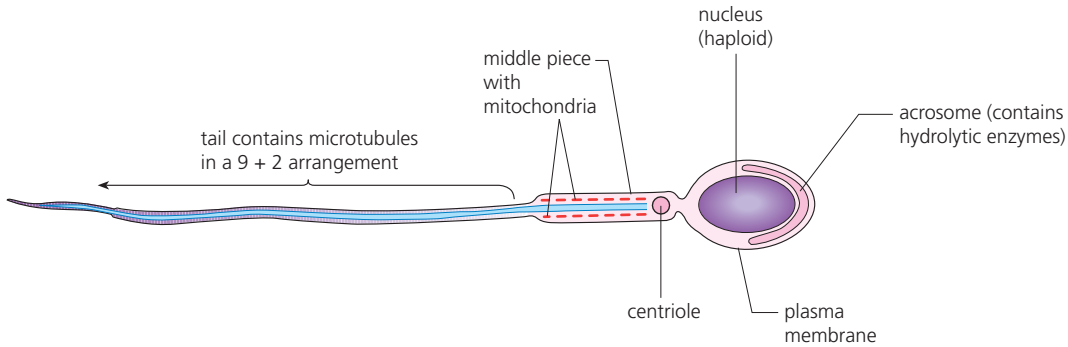


Figure 12.29 The structure of a mature spermatozoon

The roles of hormones in spermatogenesis

The onset of puberty is triggered by a part of the brain called the hypothalamus. Here, production and secretion of a releasing hormone causes the nearby **pituitary gland** (the master endocrine gland) to produce and release into the blood circulation two hormones, known as **follicle-stimulating hormone (FSH)** and **luteinising hormone (LH)**. These hormones are so named because their roles in sexual development in humans were first discovered in the female reproductive system. However, FSH and LH operate in males also (Figure 12.30).

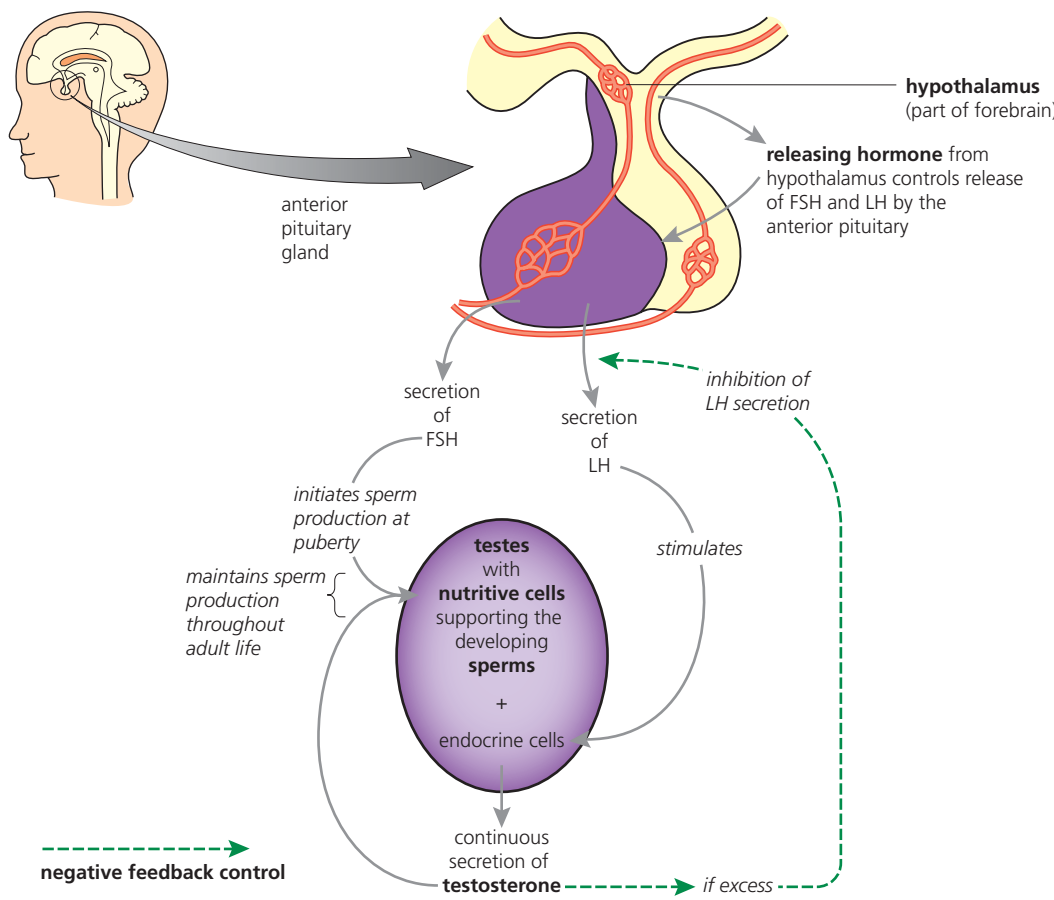


Figure 12.30 Hormone regulation of sperm production

So, at puberty, hormone from the hypothalamus triggers the secretion of FSH and LH by the anterior lobe of the pituitary. In the male, the first effect of FSH is to initiate sperm production in the testes. LH stimulates the endocrine cells of the testes to secrete testosterone. Subsequently, testosterone and FSH together maintain continued sperm production and the growth of the essential Sertoli cells that support sperms with nutrients as they grow and develop in the testes.

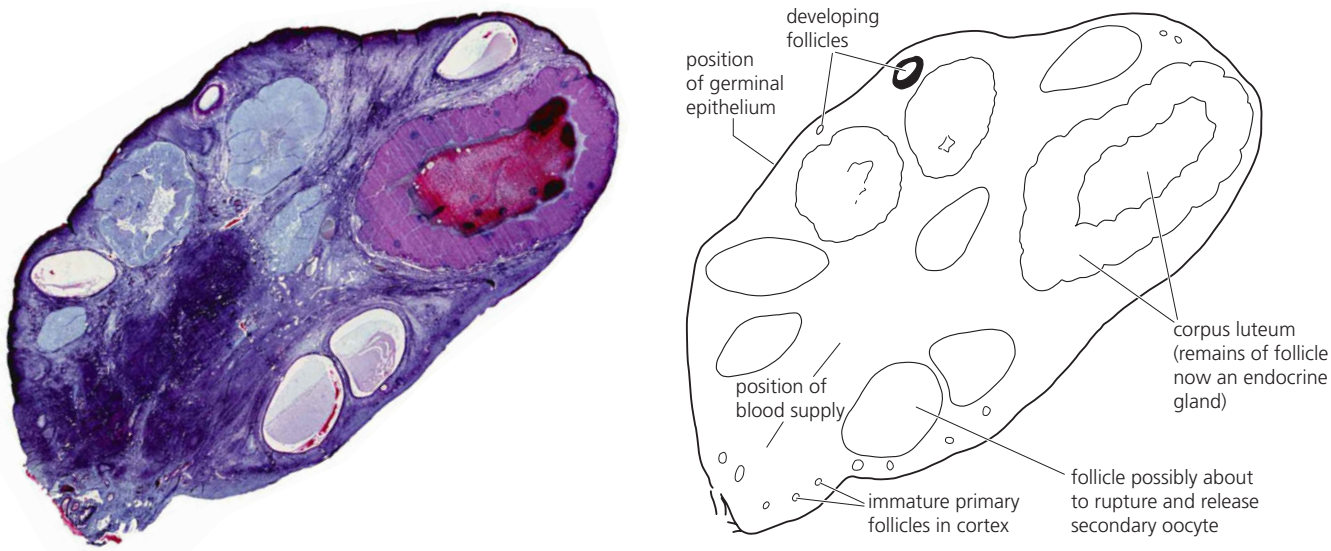
Subsequently, secretion of testosterone continues throughout life. Over-activity of testosterone is regulated by **negative feedback control**, as an excessively high level of testosterone in the blood inhibits secretion of LH. Only when the concentration of LH in the blood has fallen significantly will testosterone production recommence. (Similarly, over-activity of the nutritive cells inhibits secretion of FSH for a while.)

The structure and functioning of the ovaries

In the female, the ovaries are about 3 cm long and 1.5 cm thick. These paired structures are suspended by ligaments near the base of the abdominal cavity. As well as producing egg cells, the ovaries are also endocrine glands. They secrete the female sex hormones **oestrogen** and **progesterone**. A pair of oviducts extend from the uterus and open as funnels close to the ovaries. The oviducts transport oocytes, and are the site of fertilisation. In the event of fertilisation, development of the fetus will occur in the uterus.

The steps of oogenesis occur in the ovary. Ovulation, the process by which an egg is released to the oviduct, occurs at the secondary oocyte stage. Development of a secondary oocyte into an ovum is triggered in the oviduct if fertilisation occurs. Consequently, a thin section through a mature ovary examined by light microscopy shows developing oocytes at differing stages (Figure 12.31).

Figure 12.31 Photomicrograph of an ovary in section, and interpretive drawing



The structure of the ovary and the steps of oogenesis

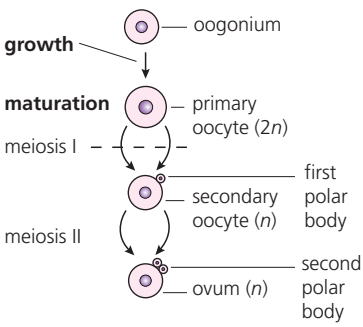
Oogenesis begins in the ovaries of the fetus before birth, but the final development of oocytes is only completed in adult life (Figure 12.32). The germinal epithelium, which lines the outer surface of the ovary, divides by mitotic cell division (page 32) to form numerous oogonia. These cells migrate into the connective tissue of the ovary, where they grow and enlarge to form oocytes. Each oocyte becomes surrounded by layers of follicle cells, and the whole structure is called a **primary follicle**.

By mid-pregnancy, production of oogonia in the fetus ceases – by this stage there are several million in each ovary. Very many degenerate, a process that continues throughout life. At the onset of puberty, the number of primary oocytes remaining is about 250 000. Less than 1% of these follicles will complete their development; the remainder never become secondary oocytes or ova.

Between puberty at about 11 years and the cessation of ovulation at menopause, typically about 55 years of age, primary follicles begin to develop further. Several start growth each month, but usually only one matures. Development involves progressive enlargement, and at the same time, the follicles move to the outer part of the ovary. The primary follicle then undergoes **meiosis I** (page 328), but the cytoplasmic division that follows is unequal, forming a tiny polar body and a **secondary oocyte**. The second meiotic division, **meiosis II**, then begins, but it does not go to completion. In this condition the **egg cell** (it is still a secondary oocyte) is released from the ovary (ovulation), by rupture of the follicle wall (Figure 12.32).

Figure 12.32 The ovary, and stages in oogenesis

summary of changes from oogonium to ovum – steps in the growth and maturation phases of gametogenesis in the ovary



Secondary oocyte **begins** meiosis II but this does not complete until sperm nucleus penetrates cytoplasm of oocyte.

diagrammatic representation of the sequence of events in the formation of a secondary oocyte for release and the subsequent changes in the ovary

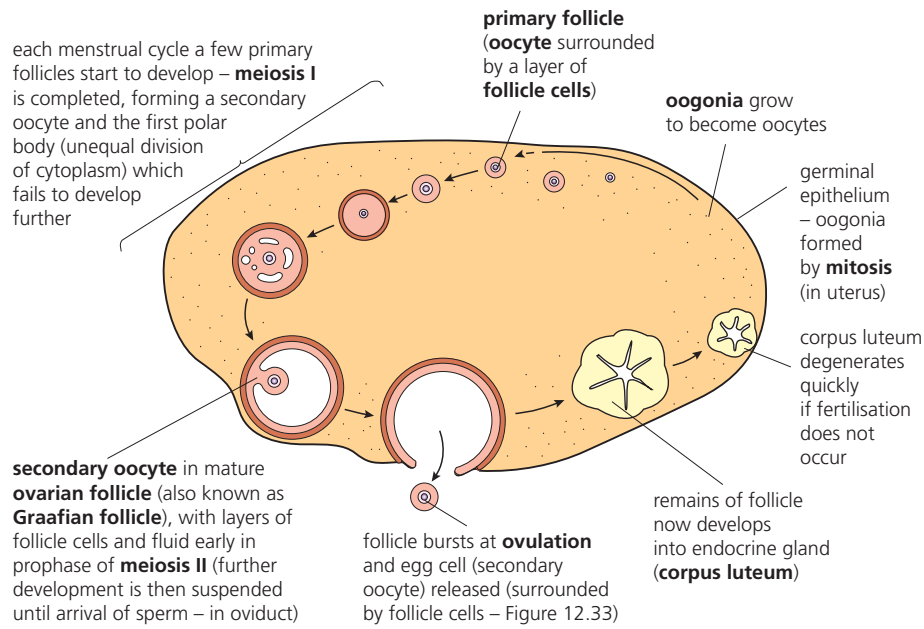
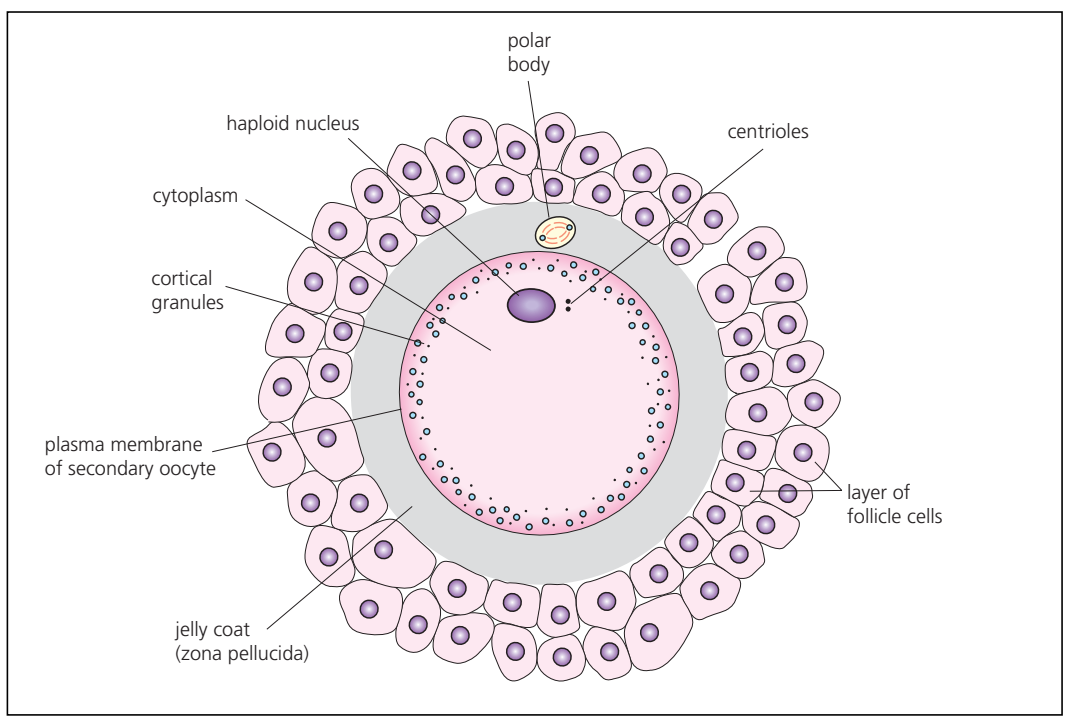


Figure 12.33 The structure of a mature secondary oocyte



16 Distinguish between the roles of FSH in sperm and ovum production.

Ovulation occurs from one of the two ovaries about once every 28 days. Meanwhile, the remains of the primary follicle immediately develop into the yellow body, the **corpus luteum**. This is an additional but temporary endocrine gland with a role to play if fertilisation occurs (see below).

Gametogenesis in testis and ovary are compared in Table 12.5.

Spermatogenesis	Oogenesis
Spermatogonia formed from the time of puberty, throughout adult life.	Oogonia formed in the embryonic ovaries, long before birth.
All spermatogonia develop into sperms, nurtured by the nutritive cells of the seminiferous tubules of the testes.	Oogonia become surrounded by follicle cells, forming tiny primary follicles, and remain dormant within ovary cortex. Most fail to develop further – they degenerate.
Millions of sperms are formed *daily*.	From puberty, a few primary oocytes undergo meiosis I to become secondary oocytes *each month*. Only one of these secondary oocytes, surrounded by a much enlarged follicle, forms a Graffian follicle – the others degenerate.
Four sperms are formed from each spermatogonium.	One ovum is formed from each oogonium (the polar bodies degenerate, too).
Sperms released from the body by ejaculation.	Graffian follicle releases secondary oocyte into oviduct at ovulation.
Meiosis I and II go to completion during sperm production.	Meiosis II reaches prophase and then stops until a male nucleus enters the secondary oocyte, triggering completion of meiosis II.
Sperms are small, mobile gametes.	Fertilised ovum is non-motile and becomes lodged in the endometrium of the uterus where cell divisions lead to embryo formation.

Table 12.5 Spermatogenesis and oogenesis compared

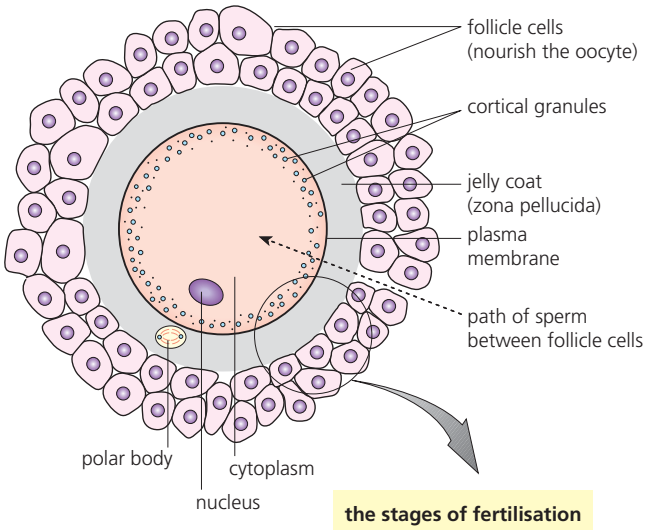
Fertilisation

In mammals, fertilisation is internal and occurs in the upper part of the oviduct. The sperms are introduced into the female during sexual intercourse. The erect penis is placed in the vagina, and semen may be ejaculated (3–5 cm^3, in humans) close to the cervix. Typically, more than one hundred million sperms are deposited. The pH of the vagina is quite acid, but the alkaline secretion of the prostate gland, a component of the semen, helps to neutralise the acidity and provides an environment in which sperms can survive.

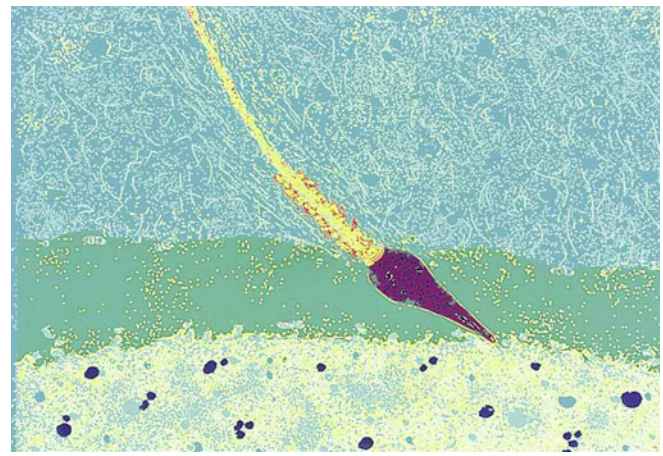
Waves of contractions in the muscular walls of the uterus and the oviducts assist in drawing semen from the cervix to the site of fertilisation. In this way, a few thousand of the sperms reach the upper uterus and swim up the oviducts. One or more of the few sperms that reach a secondary oocyte pass between the follicle cells surrounding the oocyte. Next, the coat that surrounds the oocyte, which is made of glycoprotein and called the zona pellucida, has to be crossed (Figure 12.34). This is made possible by hydrolytic enzymes which are packaged in the tip of the head of the sperm, called the **acrosome**. In contact with the zona pellucida, these enzymes are released and digest a pathway for the sperm to the oocyte membrane. This process is part of the activation processes, called **capacitation**, in which sperms are prepared for fertilisation.

The head of the sperm, containing the male nucleus, is then able to fuse with the oocyte membrane. The male nucleus enters the oocyte. As this happens, granules in the outer cytoplasm of the oocyte release their contents outside the oocyte by exocytosis. The result is that the oocyte plasma membrane cannot be crossed by another sperm.

As the sperm nucleus enters the oocyte, completion of meiosis II is triggered, and the second polar body is released. The male and female haploid nuclei come together to form the diploid nucleus of the zygote. Fertilisation is completed.



this human spermatozoon has just penetrated the zona pellucida (×3200)



the stages of fertilisation

1 sperm passes between the follicle cells and arrives at the jelly coat surrounding the secondary oocyte

2 'capacitation': changes to the surface of the head of the sperm, releasing hydrolytic enzymes

3 enzyme digestion creates a tiny path for sperm to reach the plasma membrane of the secondary oocyte

4 head of sperm is engulfed by the secondary oocyte's plasma membrane

jelly coat

cytoplasm with cortical granules

plasma membrane of oocyte

5 cortical granules in outer layers of cytoplasm of secondary oocyte pass across plasma membrane by exocytosis, preventing entry of a second sperm

6 nucleus of secondary oocyte is stimulated to complete meiosis II

7 the nuclei undergo mitosis

haploid male nucleus

haploid female nucleus

polar bodies

8 division of the cytoplasm follows, forming the first two diploid cells of the embryo

Figure 12.34 Fertilisation of a human secondary oocyte

Early development and implantation

Fertilisation occurs in the upper oviduct. As the zygote is transported down the oviduct by ciliary action, mitosis and cell division commence. The process of the division of the zygote into a mass of daughter cells is known as **cleavage**. This is the first stage in the growth and development of a new individual. The embryo does not increase in mass at this stage. By the time the embryo has reached the uterus it is a solid ball of tiny cells called **blastomeres**, no larger than the fertilised egg cell from which it has been formed. Division continues and the blastomeres organise themselves into a fluid-filled ball, the blastocyst (Figure 12.35).

In humans, by day 7 the blastocyst consists of about 100 cells. It now starts to become embedded in the endometrium, a process known as **implantation**. Implantation takes from day 7 to day 14 approximately. At this stage, some of the blastomeres appear grouped as the **inner cell mass**, and these cells will eventually become the fetus. Once implanted, the embryo starts to receive nutrients directly from the endometrium of the uterus wall (Figure 12.36).

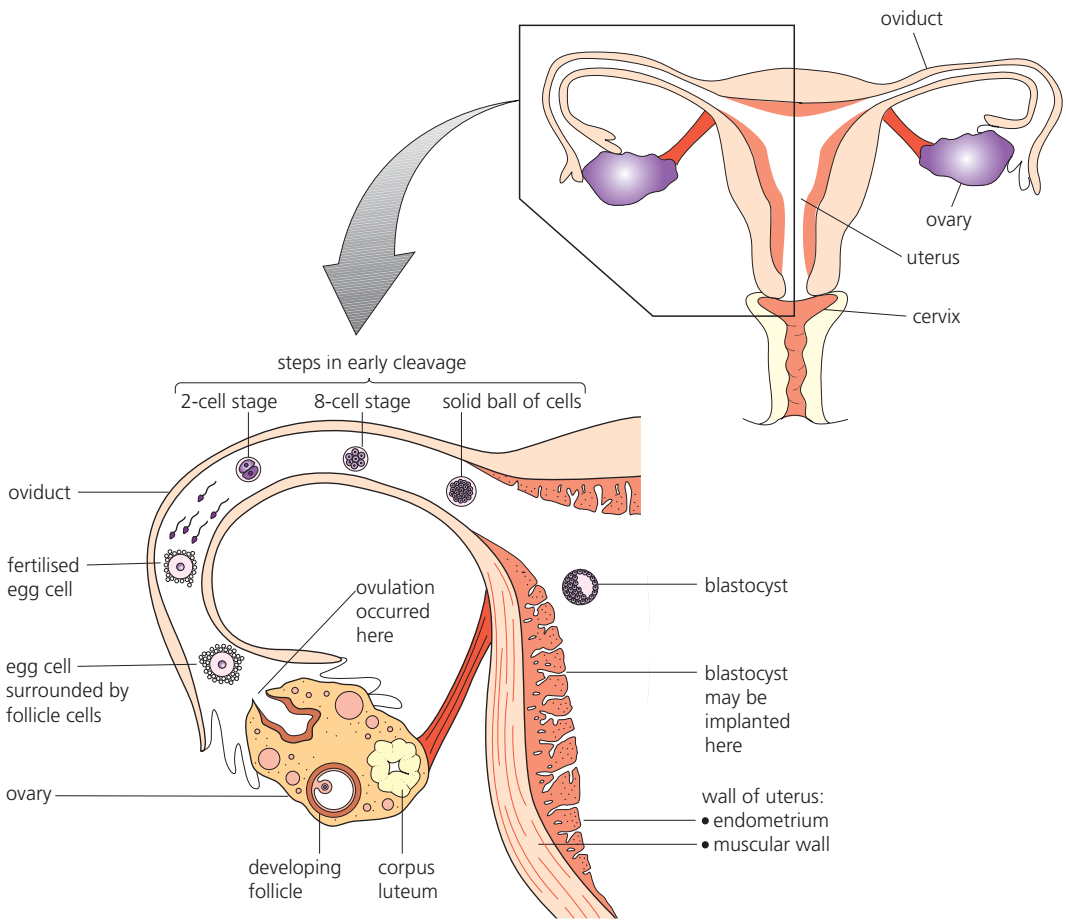


Figure 12.35 The site of fertilisation and early stages of development

Gestation – zygote to embryo to fetus in humans

The period of development in the mother's body, lasting from conception to birth, is known as **gestation** (40 weeks in humans). The rate of growth and development during gestation is much greater than in any other stage of life. In the first two months of gestation, the developing offspring is described as an **embryo**.

From early in the development of the embryo, this tiny, delicate structure is contained, supported and protected by the **amniotic sac** and **amniotic fluid** (Figure 12.37). It is the outer layers of the tissues of the embryo that grow and give rise to the **membranes**, and also the **placenta** (see below). By the end of two months' development, the beginning of the principal adult organs can be detected, and the placenta is operational. During the rest of gestation, the developing offspring is called a **fetus**.

Figure 12.36 The embryo at implantation

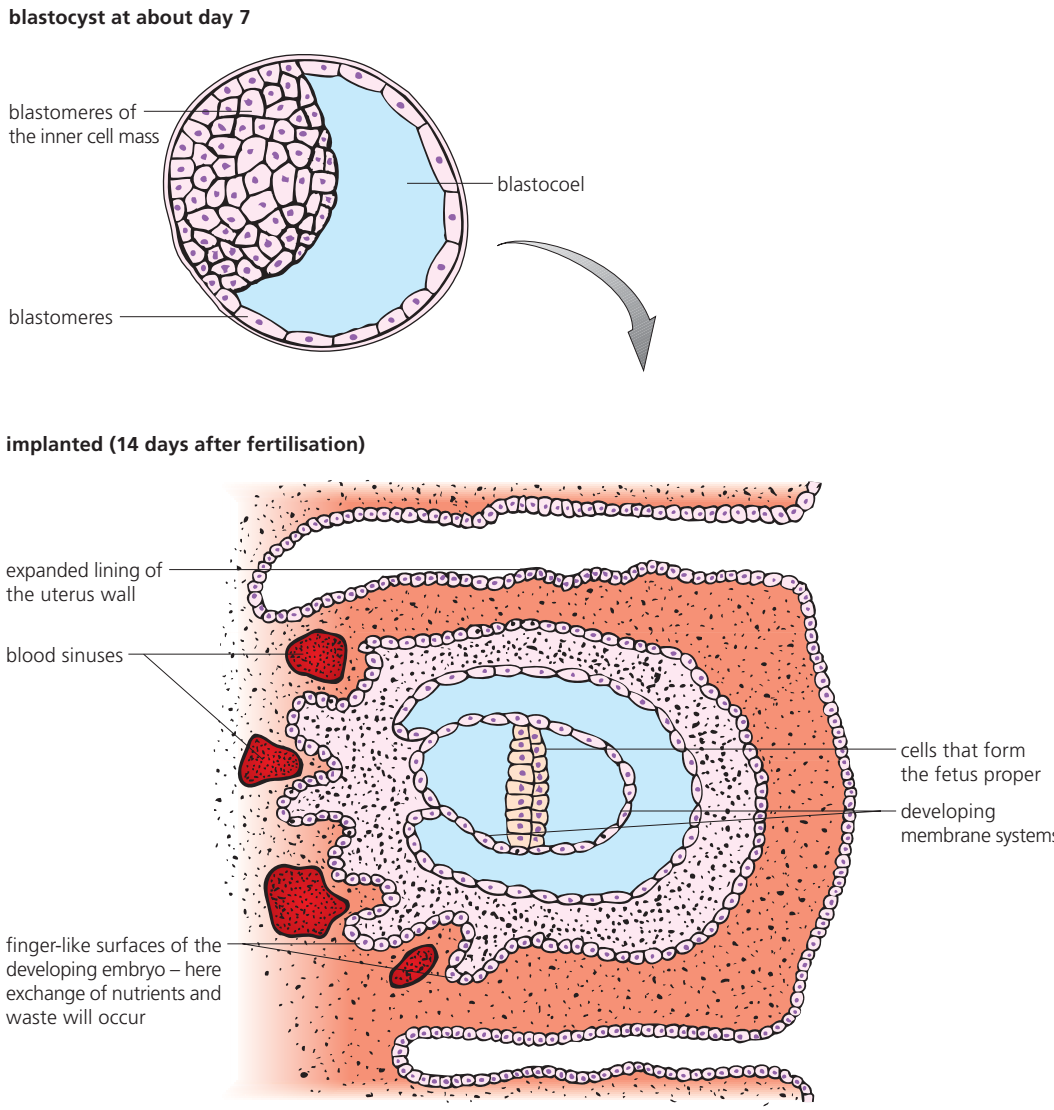
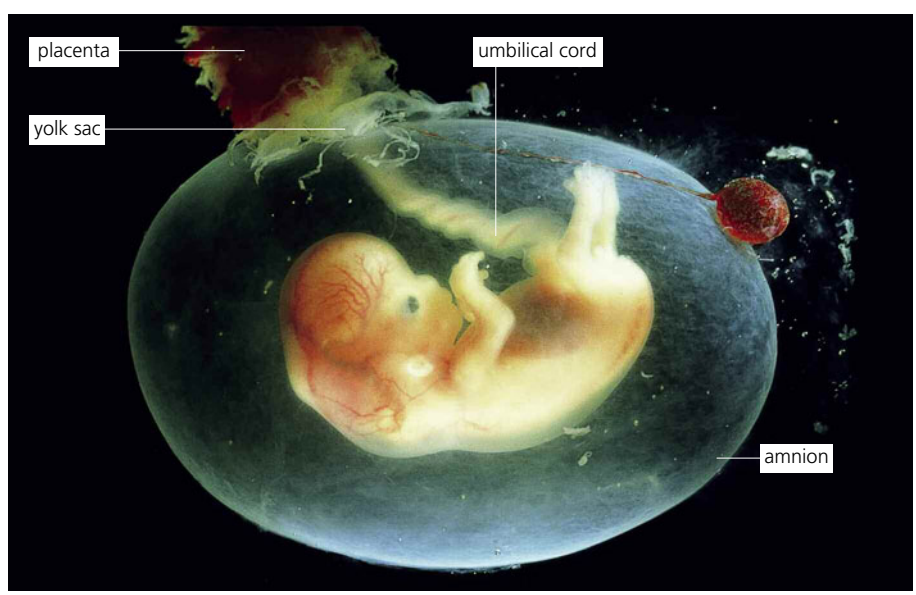


Figure 12.37 Human embryo at the six-week stage



The placenta – structure and function

The **placenta** is a disc-shaped structure composed of maternal (endometrial) and fetal membrane tissues. Here the maternal and fetal blood circulations are brought very close together over a huge surface area, but they do not mix. Placenta and fetus are connected by arteries and a vein in the umbilical cord (Figure 12.38).

Exchange in the placenta is by diffusion and active transport. Movements across the placenta involve:

- **respiratory gases**, which are exchanged; oxygen diffuses across the placenta from the maternal haemoglobin to the fetal haemoglobin (page 669), and carbon dioxide diffuses in the opposite direction;
- **water**, which crosses the placenta by osmosis; **glucose**, which crosses by facilitated diffusion; and **ions** and **amino acids**, which are transported actively;
- **excretory products**, including urea, leaving the fetus;
- **antibodies** present in the mother's blood, which freely cross the placenta so the fetus is initially protected from the same diseases that the mother has protection from (**passive immunity**, page 351).

The placenta is a barrier to bacteria, although some viruses can cross it.

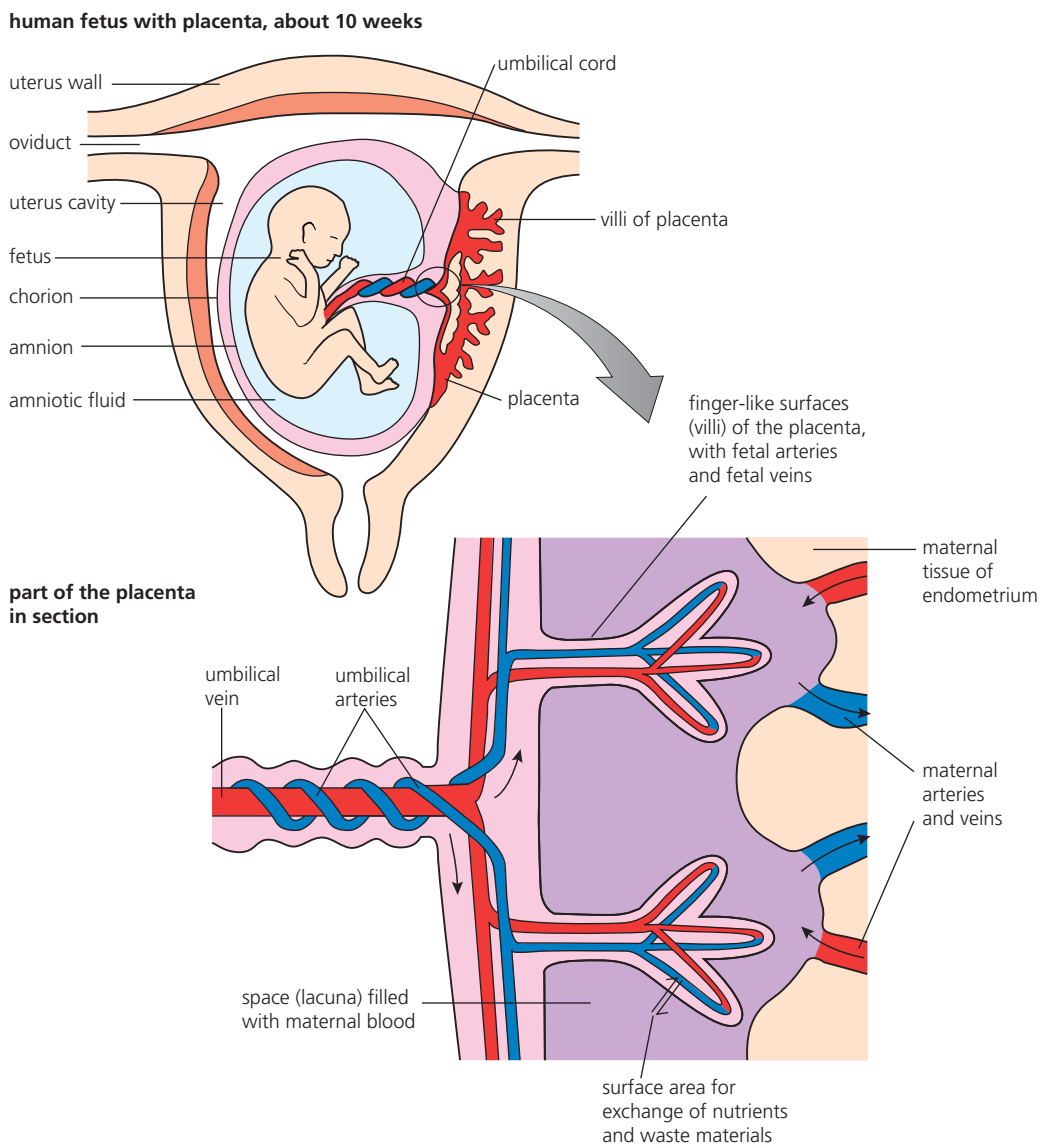


Figure 12.38 The placenta – site of exchange between maternal and fetal circulations

17 **Explain** why it is so important that the blood of mother and offspring do not mix together in the placenta.

18 **List** the features of structure of the placenta which contribute to efficient exchange and **explain** why each is important.

The placenta as endocrine gland

The placenta is also an endocrine gland, initially producing an additional sex hormone known as **human chorionic gonadotrophin (HCG)**. HCG appears in the urine from about seven days after conception. We have already noted that it is the presence of HCG in a sample of urine that is detected using monoclonal antibodies in a pregnancy-testing kit (Figure 12.8, page 358).

HCG is initially secreted by the cells of the blastocyst, but later it comes entirely from the placenta. The role of HCG is to maintain the corpus luteum as an endocrine gland (secreting oestrogen and progesterone) for the first 16 weeks of pregnancy. When the corpus luteum eventually does break down, the placenta itself secretes oestrogen and progesterone (Figure 12.39). Without maintenance of these hormone levels, conditions favourable to a fetus are not maintained in the uterus, and a spontaneous abortion results.

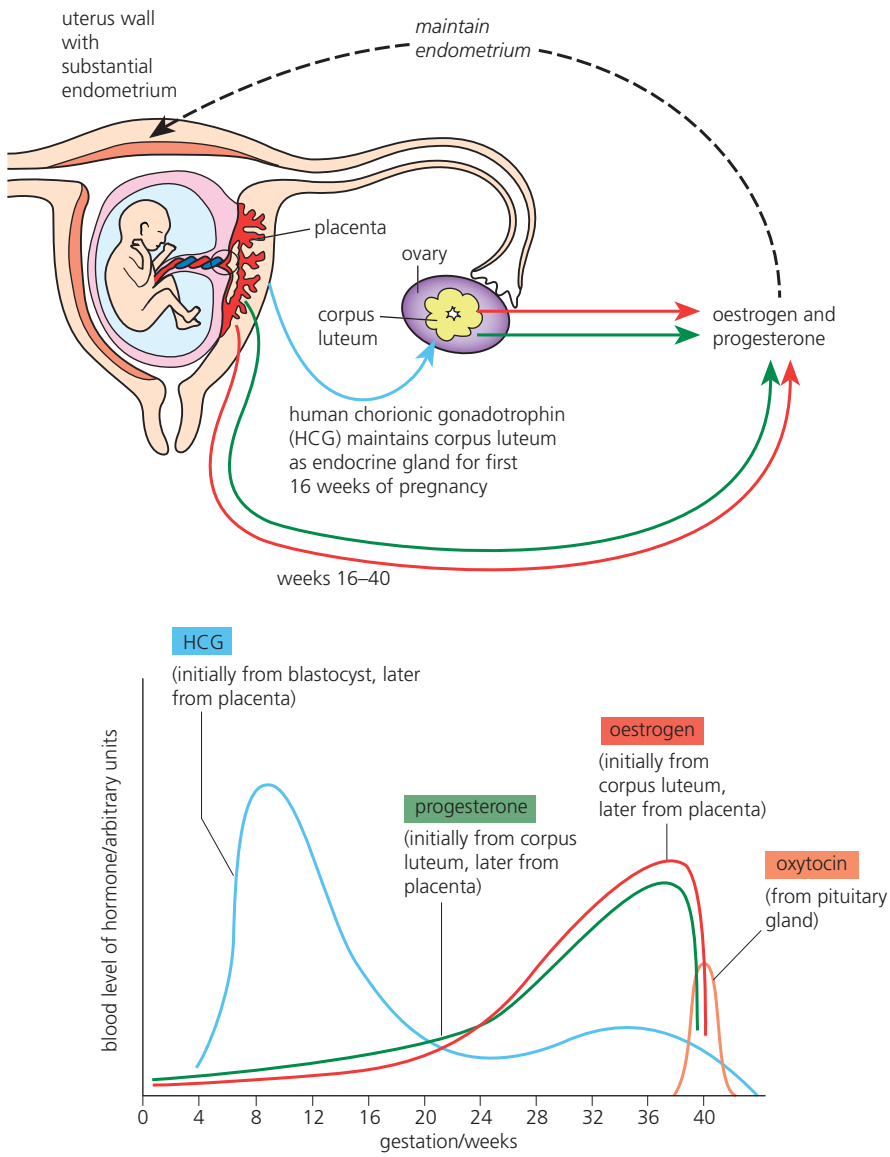


Figure 12.39 Blood levels of sex hormones during gestation

19 Suggest why the immediate production of HCG by the embryo while it still consists of relatively few cells is significant in a successful outcome to gestation.

The process of birth and its hormonal control

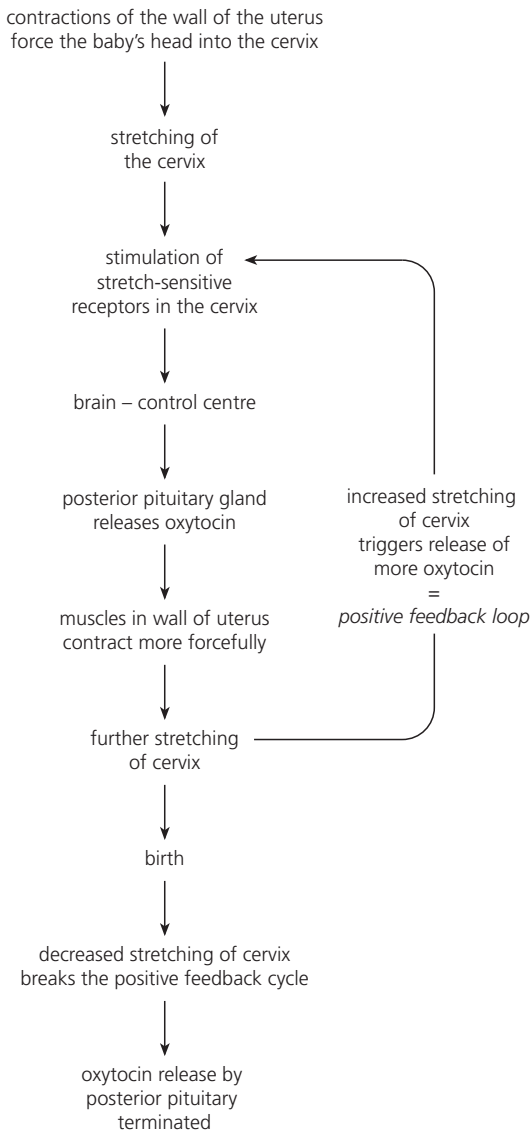


Figure 12.40 The positive feedback loop in the control of labour

Immediately before birth, the level of **progesterone** declines sharply. As a result, progesterone-driven inhibition of contraction of the muscle of the uterus wall is removed.

At the same time, the posterior pituitary begins to release a hormone, **oxytocin** (Figure 12.39). This relaxes the elastic fibres that join the bones of the pelvic girdle, especially at the front, and thus aids dilation of the cervix for the head (the widest part of the offspring) to pass through. Oxytocin also stimulates rhythmic contractions of the muscles of the uterus wall. Subsequently, control of contractions during birth occurs via a **positive feedback loop** (Figure 12.40). The resulting powerful, intermittent waves of contraction of the muscles of the uterus wall start at the top of the uterus and move towards the cervix. Progressively during this process (known as labour), the rate and strength of the contractions increase, until they expel the offspring.

Finally, less powerful uterine contractions separate the placenta from the endometrium, and cause the discharge of the placenta and remains of the umbilicus as the afterbirth.

20 Distinguish between negative and positive feedback processes.

Extension: Lactation

During pregnancy, the mammary glands are prepared for milk production by the action of hormones. Just before birth, lactation commences. Lactation is the production, secretion and ejection of milk.

In the first 2–3 days the milk, at this stage called **colostrum**, provides sugar and protein, but no fat. Also present are **antibodies** that aid survival during first exposures to potentially dangerous microorganisms.

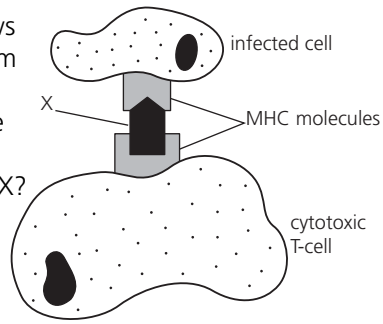
Milk formed later is an almost complete diet, providing 1.5–2.0% protein, 3.5% fat, 6.5% milk sugar (lactose), and 0.3% minerals such as Ca^{2+} ions, vitamins A, B, C and D, and water. Milk is deficient in iron, and offspring have to rely on iron stored in their liver until their diet changes and develops.

Examination questions – a selection

Questions 1–6 are taken from past IB Diploma biology papers.

Q1 The diagram shows the immune system identifying an infected cell in the body. What is the structure labelled X?

A antigen
B antibody
C IgA
D IgM



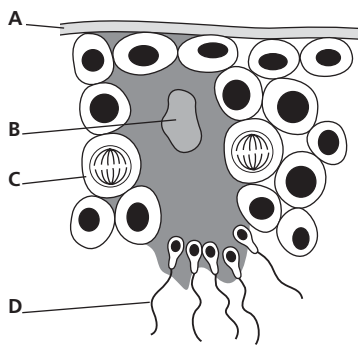
Higher Level Paper 1, May 02, Q31

Q2 Which structure of the kidney responds to ADH by reabsorbing water?

A proximal convoluted tubule
B loop of Henle
C glomerulus
D collecting duct

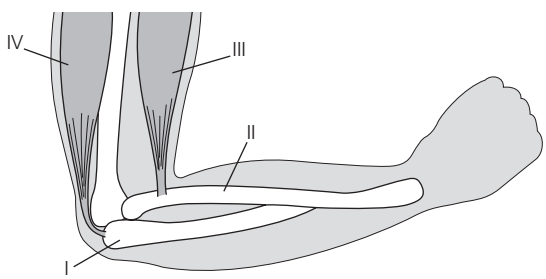
Higher Level Paper 1, May 02, Q37

Q3 The diagram shows the structure of testis tissue as seen using a light microscope. Which is the primary spermatocyte?



Higher Level Paper 2, May 03, Q35

Q4 The diagram below shows a human elbow joint.



Which response correctly identifies the ulna and the extensor muscle?

	Ulna	Extensor muscle
A	I	III
B	I	IV
C	II	IV
D	II	III

Higher Level Paper 1, May 03, Q36

Q5 Which of the following parts of the male reproductive system contribute to the production of semen?

I epididymis
II seminal vesicle
III bladder
IV prostate

A II only
B II and IV only
C I, II and IV only
D I, II, III and IV

Higher Level Paper 1, May 04, Q32

Q6 Which of the following occur(s) at birth in the mother's body?

I increase in oxytocin
II increase in uterine contractions
III increase in levels of progesterone

A I only
B I and II only
C II and III only
D I, II and III

Higher Level Paper 1, May 06, Q24

Questions 7–10 cover other syllabus issues in this chapter.

Q7 **a** Distinguish between passive and natural immunity. (4)
b Outline the sequence of steps by which the human body may acquire naturally active immunity to a viral infection. (6)

Q8 Muscles contract when the fibres shorten. This shortening is brought about when myosin and actin filaments of the sarcomeres slide past each other. Explain by means of annotated drawings how ATP and calcium ions (Ca^{2+}) enable the myosin and actin filaments to interact to shorten a muscle. (10)

Q9 **a** Outline the principle of a countercurrent mechanism. (4)
b Describe the roles of:
i the vasa recta
ii the descending and ascending limbs of the loop in water conservation by the loop of Henle. (8)

Q10 **a** Identify the stage in gestation at which the placenta forms, and the tissues involved. (4)
b Explain why is it essential that maternal and fetal blood circulations do not mix in the placenta. (4)
c Describe what substances are required by the fetus and how each is transferred to the fetal blood at the placenta. (6)
d Outline the additional role of the placenta as an endocrine gland. (3)

Answers to self-assessment questions (SAQs) in Chapters 1–12

1 Cells – the building blocks

page 1

1 The processes characteristic of living things are:

- transfer of energy (respiration)
- feeding or nutrition
- metabolism
- excretion
- movement and locomotion
- responsiveness or sensitivity
- reproduction
- growth and development

page 2

2 **a** 1 mm □ 1000 □m.
Since $^{1000}/_{100}$ □ 10, 10 cells of 100 □m will fit along a 1 mm line.

b The drawing of *E. coli* is 64 mm in length (actual length 2.0 □m). Therefore, magnification = 64 □ 1000/2 = □32 000.

page 5

3 **a**

Dimensions/ mm	Surface area/ mm^2	Volume/ mm^3	SA:V ratio
1 □ 1 □ 1	6	1	6:1 □ 6.0
2 □ 2 □ 2	24	8	24:8 □ 3.0
4 □ 4 □ 4	96	64	96:64 □ 1.5
6 □ 6 □ 6	216	216	216:216 □ 1.0

b Small cells and organisms have a large surface area to volume ratio (that is, the surface area available for diffusion). As the cell increases in size, the surface area to volume ratio decreases very rapidly (i.e. less of the cytoplasm has access to the cell exterior).

c As cell size increases, the efficiency of diffusion for the removal of waste products decreases.

page 8

4 **Emergent properties** of humans may include behavioural traits such as playfulness (a quality typical of many mammals) and anticipation of danger, and affective (emotional) traits such as loyalty and sadness, perhaps.
What have you listed?

page 12

5 The maximum length of the image of the *Amoeba* in the photomicrograph is 129 mm. From the scale bar shown here, 28 mm = 0.1 mm = 100 □m. So the actual length of the *Amoeba* is 129/28 □ 100 □m, which is approximately 460 □m.

page 13

6 The magnification is □60 (6 □ 10).

page 14

7 See **Magnification and resolution of an image**, page 13.

page 15

8 The electron microscope has powers of magnification and resolution that are greater than those of an optical microscope. The wavelength of visible light is about 500 nm, whereas that of the beam of electrons used is 0.005 nm. At best the light microscope can distinguish two points which are 200 nm (0.2 □m) apart, whereas the transmission electron microscope can resolve points 1 nm apart when used on biological specimens. Given the sizes of cells and of the organelles they contain, it requires the magnification and resolution achieved in transmission electron microscopy to observe cell ultrastructure – in suitably prepared specimens.

page 17

9 The length of the image is approximately 60 mm. From the scale bar, 1 □m = 23 mm. Therefore, the actual length of *E. coli* = 60/23 □m = 2.6 □m. The magnification of the image = 60 × 1000/26 = approximately ×23 000.

page 19

10 The characteristics *not* shown by viruses are:

- transfer of energy (respiration)
- feeding or nutrition
- movement and locomotion
- excretion
- responsiveness or sensitivity
- metabolism

11 **a** A **cell wall** is characteristic of plant cells. It is entirely external to the cell, surrounding the plasma membrane. The wall is not an organelle. Plant cell walls are primarily constructed from cellulose, an extremely strong material. When a growing plant cell divides, a cell wall is laid down across the old cell, dividing the contents. This primary cell wall has more layers of cellulose added, forming the secondary cell wall. In some plant cells the secondary layers of cellulose become very thick indeed. Cell walls are absent from animal cells. Prokaryotes have cell walls, too, but chemically different from those of plants.

The **plasma membrane** is the cell membrane that surrounds and contains the cytoplasm of all cells. It is constructed almost entirely of protein and lipid. The lipid of membranes is phospholipid, arranged as a bilayer. The proteins of cell membranes are globular proteins which are buried in and across the lipid bilayer, with most protruding above the surfaces. The whole structure is described as a fluid mosaic. See **Figure 1.19**, page 21.

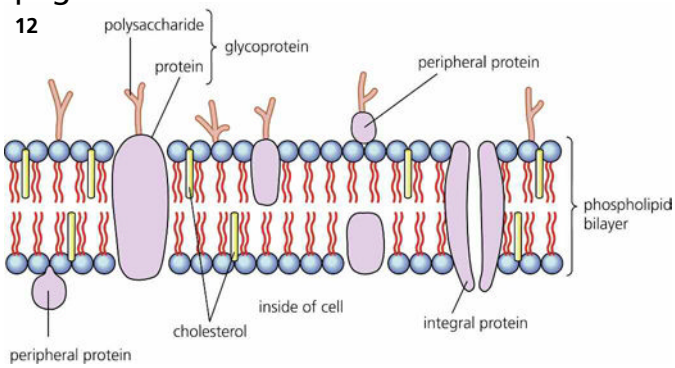
b The **nucleus** is the largest organelle in the eukaryotic cell, typically 10–20 □m in diameter, and contains the chromosomes. These thread-like structures are visible at the time the nucleus divides; at other times they are dispersed as a diffuse network, called chromatin. The nucleus is surrounded by a double membrane, which contains many tiny pores, each only 100 nm in diameter. Pores are so numerous that they make up about one-third of the nuclear membrane's surface area. This suggests that communication between nucleus and cytoplasm is important. The appearance of the nucleus in electron micrographs is shown in Figure 1.13, page 15.

The **nucleoid** is the genetic material of a prokaryote, and consists of a single, circular chromosome of DNA (helix), located in the cytoplasm, but attached at one point to the inner surface of the plasma membrane. Here the DNA is not supported by histone protein, as is the case with eukaryotic chromosomes. See Figure 1.15, page 16.

c **Flagella** and **pili**, see **Figure 1.15**, page 16.

page 21

12



Diagrammatic cross-section of the fluid mosaic membrane

page 22

13 A **lipid bilayer** is demonstrated in **Figure 1.20**, page 22, in which, in the presence of sufficient water, molecules of lipid arrange themselves as a bilayer, with the hydrocarbon tails facing together. This latter is the situation in the plasma membrane.

Several organelles of eukaryotic cells have a **double membrane**, including chloroplasts and mitochondria, in which there are present an outer and an inner membrane, both of which consist of lipid bilayers.

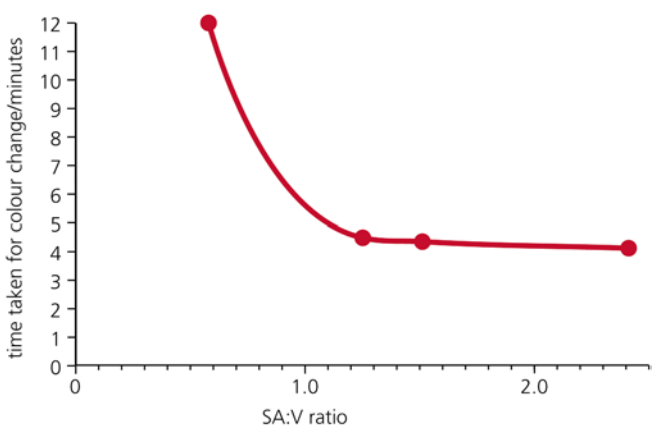
So the difference between a lipid bilayer and a double membrane lies in the number of lipid bilayers present.

page 26

14 a

Dimensions/mm	SA:V ratio
10 × 10 × 10	600/1000 = 0.6
5 × 5 × 5	150/125 = 1.2
4 × 4 × 4	96/64 = 1.5
2.5 × 2.5 × 2.5	37.5/15.6 = 2.4

b



Graph of time taken for colour change against SA:V ratio

c Hydrogen ions enter the gelatine blocks by diffusion and there cause the observed colour change. Relatively large blocks (10 mm cubes) have a low SA:V ratio – meaning that little of the interior matter is close to the external environment, and the diffusion path is a long one. Here, colour change is slowest. The reverse is true of the smallest gelatine blocks (2.5 mm cubes).

page 26

15 Compare **Figure 1.26**, page 25, and **Figure 1.27**, page 25. The difference is the permeability of the membrane traversed; in facilitated diffusion this is due to the properties of the substance that passes – it triggers the opening of pores through which diffusion occurs.

page 27

16 The concentrated solution of glucose (where the concentration of free water molecules is low) will show a net gain of water molecules at the expense of the dilute glucose solution.

page 29

17 Uptake of ions is by active transport involving metabolic energy (ATP) and protein pump molecules located in the plasma membranes of cells. As with all aspects of metabolism, this is a temperature-sensitive process and occurs more speedily at 25 °C than at 5 °C.

Individual ions are pumped across by specific, dedicated protein molecules. Because there are many more sodium ion pumps than chloride ions pumps, more of the former ion is absorbed.

page 30

18 a See pages 28–30 for the differences between proteins and lipids.

b See pages 47 and 51 for the differences between active transport and bulk transport.

c See page 30 for the difference between endocytosis and exocytosis.

page 34

19

Stage of mitosis	a % of dividing cells at each stage of mitosis*	b Time taken by each step of mitosis (given an overall length of 60 minutes)
prophase	70	70/100 × 60 = 42 min
metaphase	10	10/100 × 60 = 6 min
anaphase	5	5/100 × 60 = 3 min
telophase	15	15/100 × 60 = 9 min

*Your pie chart should comprise sectors with the following angles at the centre: prophase 252°, metaphase 36°, anaphase 18° and telophase 54°.

page 35

20 Cancer cells divide uncontrollably, forming a tumour. The cells of a malignant tumour invade and damage surrounding organs. Cancer cells may detach and be transported by the blood circulation (metastasise) to distant sites in the body. Normal cells do not behave in this way.

The switch to uncontrolled growth of cancer cells is caused by the accumulation of harmful genetic changes (mutations) in a range of genetically controlled mechanisms that regulate the cell cycle.

The cell cycle is a regulatory loop. It ensures that DNA is faithfully copied and that the replicated chromosomes move to the new (daughter) cells. Further, cell division is restricted – only a selected range of cells can divide again. Normally, if any cell is severely damaged, or grows abnormally, or becomes redundant in the further development of an organ, it undergoes programmed cell death. Thus, many controls prevent cancerous cell behaviour in normal cells.

So cancer cells arise as a result of the coincidence of mutations in several normal growth and behaviour genes. Once some cancers are established, malignant tumour cells typically release a specific protein that dictates the formation of a network of blood vessels supplying the tumour. This occurs at the expense of the supply of nutrients to surrounding healthy cells, which enhances abnormal cell division and growth, and leads to further metastasis.

2 Chemistry of life

page 38

1 An **atom** is the smallest part of an element that can take part in a chemical change. The atoms of different elements are of different sizes, but all atoms are incredibly **small**. Because of the small size of atoms we cannot refer to their mass by a standard unit, like the gram, for example. Instead, we compare the mass of an atom relative to an agreed standard. For this purpose the reference atom is that of carbon (see diagram below). The carbon atom is given a **relative atomic mass** of 12 (A_r □ 12). By comparison, atoms of hydrogen (A_r □ 1) are much lighter than carbon, but atoms of nitrogen (A_r □ 14) are slightly heavier, and atoms of potassium (A_r □ 39) substantially heavier.

Atoms themselves are made of three kinds of smaller particle. They consist of a nucleus of **protons** (positively charged particles) and, usually, **neutrons**, which are uncharged particles. (The exception is the nucleus of the hydrogen atom which contains no neutron, only a single proton.) Around the nucleus occur incredibly tiny particles called **electrons** (negatively charged particles), moving in shells.

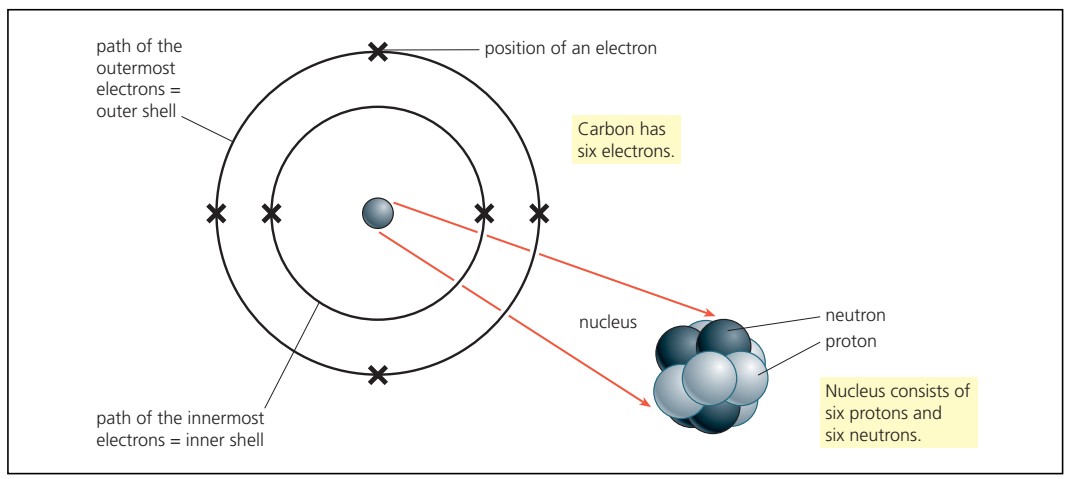
Protons and neutrons have virtually the same mass, and together make up the mass of the nucleus. Electrons have almost no mass at all. Actually, the neutron is made of two particles, a proton (□) and an electron (□), which explains why the charge on a neutron is neutral. **The atom is electrically neutral** because the number of protons in the nucleus is equal to the number of electrons in the orbits around. (Note that nucleus of an atom must not be compared or confused with the cell nucleus.)

Ions, on the other hand, are the **charged particles** formed when atoms combine together in **ionic bonding**.

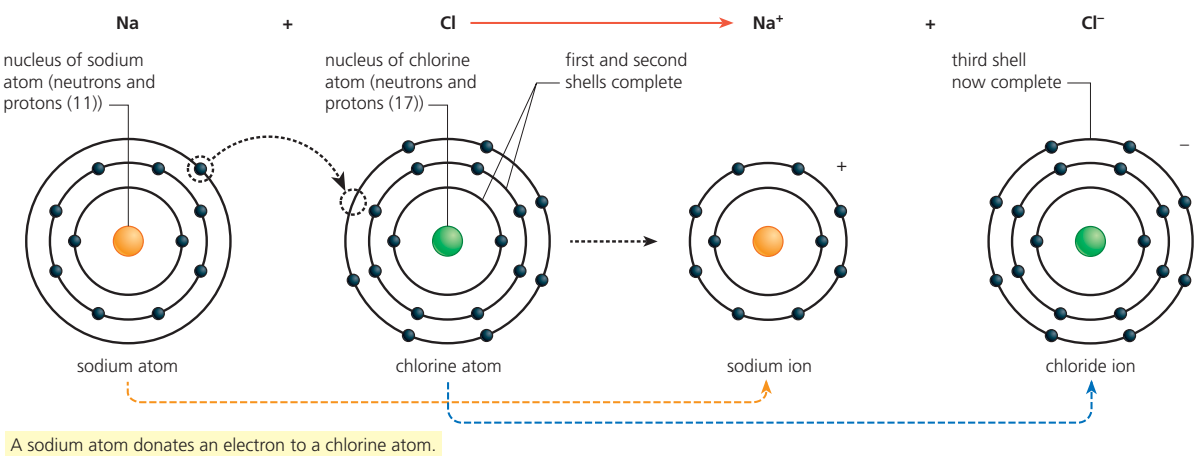
Atoms combine in ways that produce a **stable arrangement of electrons in the outer shell** of each atom – atoms are most stable when their outer shell of electrons is complete. The first electron shell can hold up to 2 electrons and then it is full. The second shell can hold up to a maximum of 8 electrons, the third shell can hold up to 18 electrons, and the fourth shell can hold up to 32 electrons. (There are further shells in the largest atoms, but they do not concern us here because the elements that make up living things have atoms that are among the smallest and lightest in the Earth's crust.)

So, ions are very **stable** because their outermost electron shell is complete. Note that when sodium and chlorine ionically bond, the sodium ion is formed by losing an electron (and so is no longer neutral – it is positively charged), and the chlorine atom has gained an electron to become the chloride ion (and so is negatively charged), as shown in the diagram below.

The structure of a carbon atom



The formation of ions by ionic bonding



Positively charged ions are called **cations**, and negatively charged ions are called **anions**. The table below shows some important ions that occur widely in living cells, and which have some specialised roles in functioning cells.

Positively charged ions (cations)	
Na^+ – sodium ion	Involved in the setting up of the action potential of a nerve fibre, and the flow of the action potential (impulse) (page 212).
K^+ – potassium ion	Involved in the setting up of the action potential of a nerve fibre, and the flow of the action potential.
Ca^{2+} – calcium ion	Involved in the contraction of the muscle myofibrils by combining with blocking molecules, so that the myosin head of the cross bridge can attach to the actin (page 368).
Negatively charged ions (anions)	
NO^{3-} – nitrate ion	Plants reduce nitrate to ammonia and combine it with an organic acid, forming an amino acid from which all 20 amino acids used to manufacture proteins are derived.
PO_4^{3-} – phosphate ion	The phosphate ion is combined with ADP to form ATP. ATP is the energy currency of cells, involved in energy-requiring reactions and processes, like protein synthesis, and muscle contraction.
HCO^{3-} – hydrogencarbonate ion	The form in which carbon dioxide is transported in the blood (plasma and red cells). It is formed when CO_2 reacts with water, catalysed by carbonic anhydrase enzyme.

2 Atoms bond in ways that produce a stable arrangement of electrons in the outer shell of each atom. The outer electron shells are completed either by **sharing electrons** to form **covalent bonds**, or by **giving and taking electrons** to form **ionic bonds**.

Ionic bonding to form sodium chloride (sodium ions and chloride ions) is illustrated above (page B392). Ionic bonding is the electrostatic attraction between oppositely charged particles. In solid sodium chloride (crystals of the salt), ionic bonds hold the ions together in a regular arrangement (known as a crystal lattice). In solution, however, water molecules surround the sodium and chloride ions, causing them to be separated and dispersed.

In **covalent bonding**, electrons are shared between atoms. Covalent bonds are the strongest bonds occurring in biological molecules. This means they need the greatest input of energy to break them. So, covalent bonds provide great stability to biological molecules, many of which are very large and elongated. Bonding of this kind is common in non-metal elements such as hydrogen, nitrogen, carbon and oxygen.

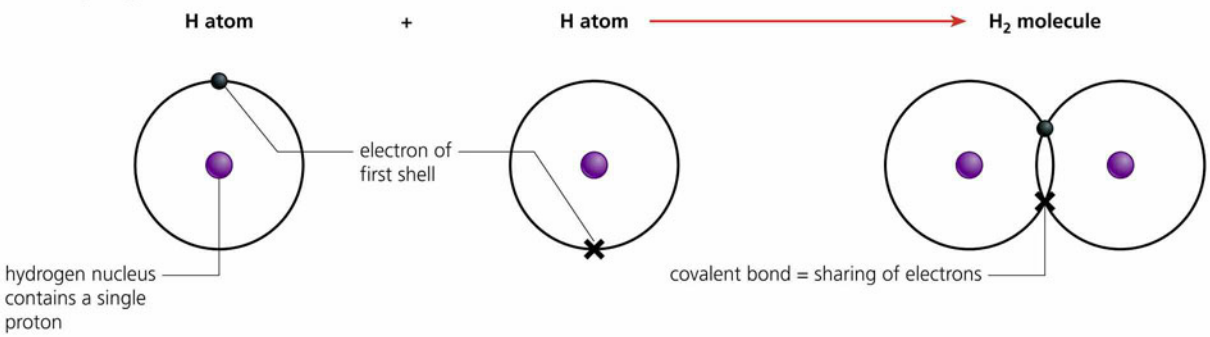
page 42

3 In a solution of glucose, water is the solvent and sucrose is the solute.

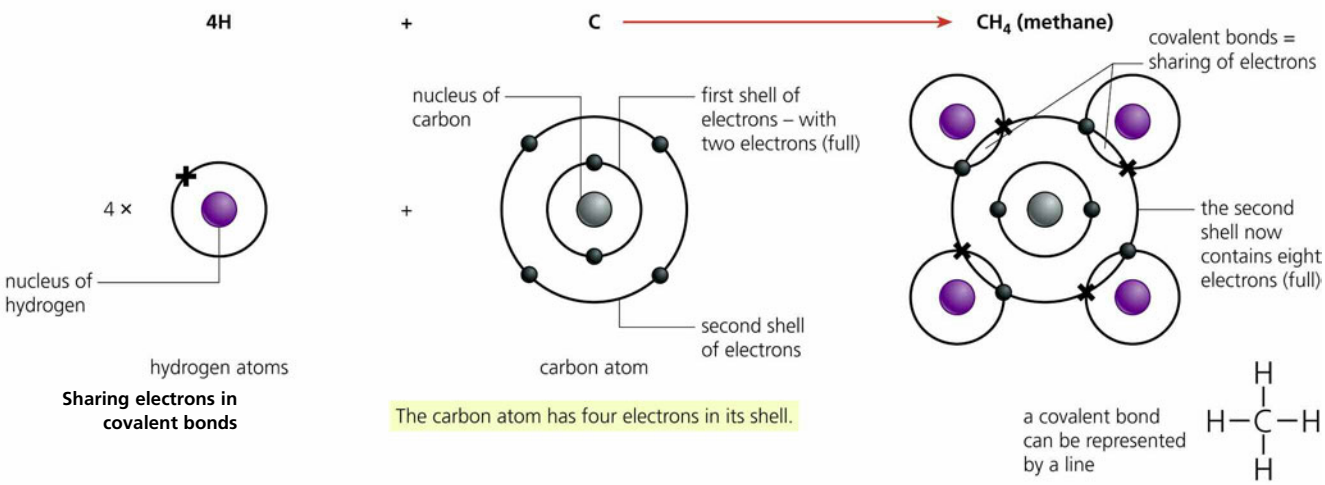
page 43

4 The causes and reasons for the **properties of water** – due to **H-bonds** – are explained in **Water** and **Figure 2.1** (pages 38–39). The mechanisms of the **cooling and solvent properties of water** are explained subsequently (page 39 and page 42).

1 In the **hydrogen molecule** a covalent bond is formed.



2 In **methane**, four single bonds are formed with hydrogen atoms to make the outer shell of carbon up to eight electrons.



Sharing electrons in covalent bonds

The carbon atom has four electrons in its shell.

page 44

5

Non-organic C	Found in the biosphere
carbon dioxide	in the atmosphere and dissolved in fresh water and in sea water as hydrogencarbonate ions
carbonates (e.g. calcium and magnesium carbonate)	in the shells of species of non-vertebrate animals, mostly marine, but also of some fresh water and terrestrial habitats, and chalk and limestone rocks, principally formed from fossilised shells

6 The presence of sugar in the cytoplasm or vacuoles of cells has a powerful osmotic effect (see **Osmosis – a special case of diffusion**, page 26), whereas if glucose molecules are condensed to starch (or glycogen, in animal cells), these insoluble carbohydrate reserves have no effect on cell water relations. Meanwhile, starch and glycogen may be speedily hydrolysed to sugar if and when the need arises.

page 46

7 A **polymer** is a large organic molecule made of repeating subunits known as **monomers**, chemically combined together. Among the carbohydrates, starch, glycogen and cellulose are polymers built from a huge number of molecules of glucose (the monomer). The glucose monomers are combined together in different ways in the three, distinctive polymers.

page 48

8 See **Disaccharides** and **Condensation and hydrolysis reactions**, **Figure 2.8**, pages 45–46, and **Figure 2.11**, page 49.

page 50

9 The **hydrophobic properties** of a lipid are combined with the **hydrophilic properties** of an ionised phosphate group, resulting in a 'head and tail' molecule.

page 52

10 In a polypeptide of only five amino acids there can be 20^5 different types, or 3 200 000.

So in a polypeptide of 25 amino acids, 20^{25} different types, and in one of 50, 20^{50} different types of polypeptide can be formed. A polypeptide of more than 50 amino acid residues is by convention called a protein. So, there is virtually unlimited structural variation possible in proteins of cells. Remember, as the sequence of amino acids changes, so do their structure and properties.

page 54

11 Enzymes work by binding to their substrate molecule at a specially formed pocket on the enzyme – the active site. Most enzymes are large molecules, and the active site takes up a relatively small part of the total volume of the enzyme molecule. Nevertheless, the active site is a function of the overall shape of the globular protein, and if the shape of an enzyme changes for whatever reason, the catalytic properties may be lost.

page 56

12 Pre-incubation is required to ensure that when the reactants are mixed, the reaction occurs at the known, pre-selected temperature.

page 57

13 **pH** is a measure of acidity or alkalinity of a solution. Strictly, pH is a measure of the hydrogen ion concentration. The pH scale runs from 0 to 14, where pH 7 is neutral. This is the pH of pure water, where the concentrations of hydrogen ions (H^+) and hydroxide ions (OH^-) are low and equal in number. A solution of pH of less than 7 is an acidic solution, and strong acids have a pH of 0–2. Solutions of pH of more than 7 are alkaline.

A **buffer solution** acts to resist pH change when diluted or a little acid or alkali is added. Many buffers used in laboratory experiments contain a weak acid (such as ethanoic acid – vinegar) and its soluble salt (e.g. sodium ethanoate). In this case, if acid is added, the excess hydrogen ions are immediately removed by combination with ethanoate ions to form undissociated ethanoic acid. Alternatively, if alkali is added, the excess hydroxide ions immediately combine with hydrogen ions to form water. At the same time, more of the ethanoic acid dissociates, adding more hydrogen ions to the solution. The pH does not change in either case.

In the body of the mammal, the blood is very powerfully buffered by the presence of a mixture of phosphate ions, hydrogencarbonate ions and blood proteins. The blood is held between pH 7.35 and 7.45. pH is very important in living organism, largely because pH affects the shape of enzymes, almost all of which are proteins.

page 59

14 Approximately 0.23 $cm^3 s^{-1}$

page 60

15 **a** see **Figure 2.22**, page 60.

b Working with an excess of substrate molecules (a relatively high concentration of substrate), the effect of increase in the concentration of enzyme is to increase the rate of reaction. This is because, at any moment, proportionally more substrate molecules are in contact with an enzyme molecule.

page 61

16 *in vitro* → biological processes occurring in cell extracts (literally 'in glass')

in vivo → biological processes occurring in living organisms (literally 'in life')

page 62

17 A catalyst is a substance that alters the rate of a chemical reaction, but remains unchanged at the end.

Inorganic catalysts	Enzymes
typically a metal like platinum in a finely-divided state (e.g. as platinised mineral wool)	typically made of protein (but occasionally RNA acts as an enzyme), so easily denatured (e.g. by high temperature)
able to withstand high temperatures, high pressures and extremes of pH, if necessary	extremely specific – most are specific to one type of substrate molecule, for example

page 64

18 **Nitrogenous base**

Nitrogenous bases are organic compounds. The carbon backbones of nitrogenous bases are ring molecules containing two or more nitrogen atoms, covalently bonded together. They are derived from one of two parent compounds: purine (a double-ring compound) or pyrimidine (a single-ring compound).

Inorganic base
An inorganic base is a substance that can accept a hydrogen ion and so neutralise an acid. Strong bases are substances like sodium hydroxide. They are ionised compounds.

page 71

19

RNA involved in transcription	
messenger RNA (mRNA)	See **Transcription** – **the first step in protein synthesis**, page 69.
RNA involved in translation	
messenger RNA (mRNA)	mRNA is a linear molecule in the cytoplasm along which ribosomes move, 'reading' the genetic code and transcribing the information into a linear sequence of amino acid residues (i.e. primary structure of the protein)
transfer RNA (tRNA)	See **Amino acid activation** and **Figure 2.32**, pages 70–71.
ribosomal RNA	RNA is a major component of the ribosomes – organelles found in the cytoplasm and attached in rough endoplasmic reticulum (rER); here, the mRNA is 'read' and the protein molecules are formed (**Figure 2.33**, page 72). The structure of a ribosome is shown in **Figure 8.11**, page 248.

page 74

20

Proteins other than enzymes	**Roles**
'active' proteins in cell membranes	■ active uptake pumps (e.g. for ions) ■ hormone receptor sites ■ antigens of immune system
'active' proteins in body fluids (e.g. blood)	■ blood proteins (pH buffers) ■ blood clotting proteins ■ antibodies
structural proteins	■ hair of mammals ■ collagen fibres of tendons and ligaments
storage proteins (plants only)	■ grains of proteins deposited in the cells of some seeds and plant storage organs – available for re-use on germination and growth

3 Energy transfer in cells

page 78

1 See **ATP, the universal energy currency**, page 77.

page 80

2 Glycolysis (the conversion of glucose to pyruvate) which occurs in the cytosol (the aqueous part of the cytoplasm that surrounds organelles such as the mitochondria) does not require oxygen for completion.

page 81

3 Lactate and a limited amount of ATP are the final products of anaerobic respiration in muscle fibres.

page 82

4 In the respirometer, the far side of the U-tube manometer is the control tube (A). Here, conditions are identical to those in the respirometer tube, but in the former, no living material is present. However, any change in external temperature or pressure is equally experienced by both tubes, and their effects on the level of manometric fluid are equal and opposite, and they cancel out.

5

Mammals		**Flowering plants**
product of digestion of food items, particularly from carbohydrates hydrolysis of glycogen reserves in muscle cells	**source of glucose**	product of photosynthesis in the chloroplasts, mainly situated in the green leaves hydrolysis of stored starch

page 84

6 If a naked flame (Bunsen burner) is used to heat the water (Step 1), then the flame must be switched off before the propanone (acetone) is poured out – to avoid fire.

Care is needed when fresh leaves are immersed in boiling water using blunt forceps – to avoid a scalding accident.

The contents of centrifuge tubes to be placed in the centrifuge (Step 4) must be balanced in contents, so that they are of equal mass – to prevent damage to the centrifuge or to the tubes during centrifugation.

Pigment solution should be labelled with the contents and particularly the solvent, and stored in cool dark conditions -– so that they are not confused with any other materials, and so that solvent vapour is not accidentally exposed to a naked flame.

page 87

7 **a** Carbon: from carbon dioxide by reducing it to carbohydrate
b Hydrogen: from water by photolysis to form hydrogen and oxygen gas (waste product)

page 88

8 Starch in the green leaves held in the dark is converted to glucose and then to sucrose and the latter is carried (translocated) in the phloem sieve tubes to storage sites, typically in the stem and root cells – where it is converted back to starch and stored.

page 89

9 A thermometer is required in the glass tube with the *Elodea* (in a position that does not interfere with the supply of light to the plant), so that the temperature at which photosynthesis is measured is known. Also, it is advisable to check that the *Elodea* sample is not subjected to an unplanned rise in temperature as the intensity of light is increased, for example.

page 90

10 The biochemical steps of photosynthesis are dependent on the action and activity of numerous (protein) enzymes. These are progressively denatured and rendered inactive as the temperature is raised.

4 Genetics

page 92

1 See **Chromosomes occur in pairs**, page 91.

page 94

2 See **The significance of mitosis**, page 35, and **Abrupt change in hereditary information – mutations**, particularly **Chromosome mutations**, page 98.

page 100

3 **Figure 4.10** shows the karyotype of a male patient showing non-disjunction at chromosome 21.

page 102

4 See **Gene mutations**, page 98.

page 105

5 The presentation of a breeding experiment, such as that shown in **Figure 4.15**, is a prediction of the likely outcome. It represents the probable results, provided that:
- fertilisation is random
- there are equal opportunities for survival among the offspring
- large numbers of offspring are produced.

What is actually observed in a breeding experiment may not necessarily agree with the prediction. For example, there is a chance in this particular cross that:
- more pollen grains of one genetic constitution may fuse with egg cells than another
- more developing seeds of one type are predated and destroyed by insect larvae of species attacking the plant (so fewer zygotes of one type complete development)
- the cross produced too few progeny in total.

page 106

6 Below is a table showing how the Law of segregation relates to meiosis.

Mendel's monohybrid cross	**Feature of meiosis**
Within an organism there are breeding factors controlling characteristics like 'tall' and 'dwarf'. These factors remain intact from generation to generation.	Each chromosome holds a linear sequence of genes. A particular gene always occurs on the same chromosome in the same position after each nuclear division.
There are two factors for each characteristic in each cell.	The chromosomes of a cell occur in pairs, called homologous pairs.
One factor comes from each parent. (A recessive factor is not expressed in the presence of a dominant factor.)	One of each pair came originally from one parent and the other from the other parent.
Factors separate in reproduction; either can be passed to an offspring. Only one of the factors can be in any gamete.	At the end of meiosis each cell contains only one of the homologous pair of chromosomes present in the parent cell.
Mendel recognised the 3:1 ratio is the product of randomly combining two pairs of unlike factors (e.g. **T** and **t**) – the product of the binomial expression (where a recessive factor is not expressed in the heterozygote): (**T** + **t**)(**T** + **t**) = 1**TT** + 2**Tt** + 1**tt**	The arrangement of bivalents at the equatorial plate of the spindle is random; maternal and paternal homologous chromosomes are independently assorted. In a large number of matings all possible combinations of chromosomes will occur in equal numbers.

page 107

7 The layout of your monohybrid cross will be as in **Figure 4.17** (page 107), but the **parental generation** (**P**) will have the genotypes (if you have chosen C = allele for coat colour):
$\mathbf{C^RC^R}$ × $\mathbf{C^WC^W}$
where $\mathbf{C^R}$ represents the allele for red coat, and $\mathbf{C^W}$ represents the allele for white coat. The gametes the parental generation produces will be:
$\mathbf{C^R}$ and $\mathbf{C^W}$
The **offspring (F_1)** will have genotype $\mathbf{C^RC^W}$
and the phenotype will be roan.
In a **sibling cross of the F_1 generation** the gametes of both siblings will be:
$\frac{1}{2}\mathbf{C^R}$ + $\frac{1}{2}\mathbf{C^W}$
From an appropriate Punnett grid, the offspring (F_2) to be expected and the proportions are as follows.

genotypes:	$\mathbf{C^RC^R}$	$\mathbf{C^RC^W}$	$\mathbf{C^WC^W}$
ratio:	1	2	1
phenotypes:	red	roan	white

page 109

8 The **Jones** were A × B which might produce any one of:
AA × BB → AB group
AO × BO → AB or O groups
AA × BO → AB or A groups
AO × BB → AB or B groups
So the Jones could be the parents of *any* of the four children.
The **Lees** were B × O which might produce either of:
BB × OO → B group
BO × OO → B or O groups
So the Lees might be the parents of *either* the B group or the O group child.
The **Gerbers** were O × O which can produce only:
OO × OO → O group
So the **Gerbers were parents of the O group child** (and therefore the **Lees were the parents of the B group child**).
The **Santiagos** were AB × O which might produce:
AB × OO → A or B groups.
So the **Santiagos were the parents of the A group child** (since the Lees were the parents of the B group child).
So by elimination, the **Jones were the parents of the AB group child**.

9 **a** Liz and Diana
b (**i**) David and Anne (**ii**) James
c Eight: Richard and Judith, Anne, Charles, Sophie, Chris, Sarah, and Gail
d James and William, Arthur and Diana, etc.

page 111

10 Given a homozygous normal-handed parent (**nn** – producing **n** gametes only) crossed with a heterozygous brachydactylous parent (**Nn** – producing 50% **N** and 50% **n** gametes), the probability of an offspring with brachydactylous hands is 50% or 0.5, as your genetic diagram and Punnett grid will show.

page 114

11 See **Sex linkage** and **Figure 4.24**, pages 113–114.

page 115

12 See **Sex linkage**, **Haemophilia** and **Figure 4.25**, pages 113–115.

5 Genetic engineering and biotechnology

page 117

1 The **genome** is the complement (genes) of an organism or of an individual cell. (See also **Table 4.2**, page 109, for other genetics terms.)

page 119

2 Protease enzymes hydrolyse proteins to shorter chain polypeptides, and finally to the individual amino acids which made up the primary structure of the protein.

3 See **The DNA double helix**, pages 64–66.

page 123

4 Complementary base pairing is explained on page 64 and illustrated in **Figure 2.27**, page 65.

5 The minimum length (in terms of nucleotides) of the artificial gene is:
14 □ 3 □ 42 to code for the constituent amino acids
2 □ 3 □ 6 to code for the 'stop' and 'start' codons
plus 6 for recognition by the sticky end restriction enzyme.
Total □ 54

page 124

6 See **Figure 5.8**, page 125.

7 When the bacterium invades a host cell, the Ti plasmid becomes inserted into a chromosome, introducing genes for the synthesis of a plant growth substance (causing proliferation of host cells □ tumour) and other genes present (or previously inserted) on the plasmid. In effect, the bacterium carries out genetic engineering on its own behalf.

page 128

8 The fruits of the cultivated grasses (which include wheat and rice) are the staples of human diets all over the world – providing the bulk of essential energy-rich foods. The ability of leguminous plants to fix atmospheric nitrogen for amino acid and protein production enables these plants to grow well without the addition of (expensive) nitrogen-based fertilisers. Food products from leguminous plants are also relatively rich in proteins. GM cultivated grasses that could also fix nitrogen (should they come about) would grow well without nitrogen-based fertilisers – and might add even more to human nutrition, contributing to the overcoming of protein deficiency as well as meeting energy needs.

page 131

9 **a** Genotype and genome: see answer to SAQ1, Chapter 5, above.
b Restriction endonuclease and ligase: see **Figure 5.2** (page 119) and **Figure 5.6** (page 122).
c Blunt ends and sticky ends: see **Figure 5.2** (page 119).
d Bacterial chromosome and a plasmid: see **Vectors for cloning**, page 000. (Note that the bacterial chromosome is known as a nucleoid – see **Figure 1.15**, page 16.)

page 132

10 Alleles coding for proteins that cause genetic disease typically arise by mutation. Mutations occur at a steady rate in human populations. Consequently, genetic diseases are maintained at low levels in populations.

6 Ecology, evolution and biodiversity

page 138

1 The dominant plants set the way of life for many other inhabitants, largely by:
- providing the principal source of nutrients
- determining the living spaces (habitats/microhabitats)
- influencing the environmental conditions.

page 139

2 **b** all the frogs of the lake – population
a the whole lake – ecosystem
f the mud of the lake – habitat
c the flow of water through the lake – abiotic factor
g the temperature variations in the lake – abiotic factor
d all the plants and animals present – community
e the total mass of vegetation growing in the lake – biomass

page 140

3

	Producer	Primary Consumer	Secondary consumer	Tertiary consumer
Food chain 1	phyto-plankton	zooplankton	common mussel	herring gull
Food chain 2	sea lettuce	grey mullet	pollack	seal

page 142

4 See **Feeding relationships – producers, consumers and decomposers**, page 138.

5 Primary consumers – humans who eat food of plant origin, only. All consumer trophic levels – humans who eat a diverse range of foods (e.g. hunter-gatherers).

page 144

6 See **Energy flow**, pages 142–144.

page 147

7 See **Figure 6.6 B**, and follow the 'cycling of materials' arrow.

page 148

8 **a** In the atmosphere: carbon dioxide gas
b In the hydrosphere: hydrogencarbonate ions
c In the lithosphere: carbonates (e.g. Ca_2CO_3)

page 150

9 **a** The % change in mean atmospheric carbon dioxide 1960–2000: in 1960, 315 ppm of CO_2; in 2000, 370 ppm of CO_2.
Increase in these 40 years was 55 ppm.
% gain was $\frac{55}{315}$ □ 100 □ 17.5%
b The atmospheric carbon dioxide varies between high and low values within each 12-month period of the graph due to photosynthesis on lands in the northern hemisphere – the peaks are winter levels, and troughs are summer levels of atmospheric CO_2.

page 155

10 a Favourable abiotic factors such as a continuing good food supply. Favourable biotic factors such as the absence of predators, and freedom from parasite attack.
b Intraspecific competition for resources and space as the population grows.

page 156

11 a Two soluble nutrients essential to phytoplankton growth are nitrate and phosphate ions.
b The numbers of phytoplankton decrease significantly by May and June due to a shortage of essential nutrients.
c The increasing quantity of nutrients in October came from the decay of the dead phytoplankton and zooplankton.

page 157

12 Sedimentary rocks are formed by deposition of the products of erosion of other rocks, most frequently when this material is swept down rivers to the sea, or enters large fresh water lakes, and settles in layers on the bottom.
See **Figure 6.19**, page 157.

page 159

13 The breeding of domesticated plants and animals has created varieties with little external resemblance to their wild ancestors. Darwin bred pigeons, and noted there were more than a dozen distinctive varieties of pigeon, all of which were descended from the rock dove (**Figure 6.20**). Darwin argued that if so much change can be induced in so few generations, then species must be able to evolve into other species by the gradual accumulation of minute changes, as environmental conditions alter and natural selection operates.

14 All these differing varieties of domesticated dog are capable of interbreeding to produce fertile offspring.

page 161

15 See **Neo-Darwinism**, page 160.

page 162

16 a Exposure of pathogenic bacteria to sub-lethal doses of antibiotic may increase the chances of resistance developing in that population of pathogens.
b By varying the antibiotics used, there is increased likelihood of killing all the pathogens in a population, including any now resistant to the previous antibiotics used. This approach works until multiple-resistance strains have evolved, such as in strains of *Clostridium difficile* and *Staphylococcus aureus*.

page 164

17 By 'evolution' we mean 'the gradual development of life in geological time'. For example, we know that life appeared on Earth about 3500 million years ago, and that most of the great diversity of living forms have appeared subsequently. Before early geologists realised that the Earth is extremely old, biblical calculations suggested that life had been created in 4004 BC, or 6000 years ago. If this were still believed, then there would be insufficient time for evolution by natural selection.

18 The impact of a meteorite or asteroid from space or the violent eruption of a volcano are examples of events that might have caused violent and speedy habitat change over a substantial part of the surface of the Earth.

page 165

19 Such names are precisely defined and internationally agreed; they facilitate cooperation between observers by ensuring the exact species that is being investigated and reported on.

page 166

20 Your answer depends largely on the types of plant growing in gardens where you are, and perhaps on the main uses they are put to; for example:
- vegetables (of root, leaf, fruit or seed origin)
- flowers (of particular seasons or perhaps soil types).

7 Human physiology, health and reproduction

page 178

1 Saprotrophic nutrition: in the decomposer organisms, mainly bacteria and fungi.
Parasitic nutrition: by organisms living on or in a host organism for much or all of their life cycle, and feeding on the host's tissues.

page 181

2 a See **Digestion, where and why**, page 178.
b Water, vitamins, minerals.
c Cellulose.

page 182

3 Protease enzymes of digestion are secreted in inactive forms to prevent autolysis of the cells of the gastric glands and pancreas in which they are formed.

page 184

4 Absorption: uptake into the body (blood circulation or lacteals) of the useful products of digestion, from the gut lumen.
Assimilation: uptake of nutrients into cells and tissues.

page 185

5 For example, it means that blood cannot be directed to a respiratory surface immediately before or after servicing tissues that are metabolically active (and so have a high rate of respiration). Rather, the blood circulates randomly around the blood spaces and blood vessels.

page 187

6 Phagocytic leucocytes function anywhere in the body that an infection occurs, but in addition to their presence in the blood plasma, lymph and lymph glands, they are always present in the airways and alveoli of the lungs, and in the liver, lining the rows of liver cells (hepatocytes) past which the blood flows.

page 192

7 These non-elastic strands keep the heart valve flaps pointing in the direction of the blood flow. They stop the valves turning inside out when the pressure rises abruptly within the ventricles.

page 193

8 Adrenaline is secreted in 'flight or fight' situations, when the individual is suddenly startled or attacked, or otherwise believes itself to be in danger. It prepares the body for exertion and high physical and mental performance.

page 194

9

Bacterial cell (see Figure 1.15, page 16)	**Virus particle (see Figure 1.17, page 19)**
small cells typically 5–10 μm	crystal structure when outside host cells (i.e. not cellular) all extremely small, 20–400 nm.
cell wall of complex biochemical composition (not cellulose)	no cell wall
plasma membrane surrounding cytoplasm	no cytoplasm or plasma membrane
nucleoid (circular length of DNA) in the cytoplasm, attached to the plasma membrane	no nucleoid; has a core of nucleic acid (DNA or RNA) surrounded by a protein coat called the capsid
no envelope present	some viruses have an additional envelope of protein and lipid – acquired as they leave the previous host cell

page 195

10 Antibiotics work by interfering with specific metabolic processes, typically the synthesis and laying down of new wall materials (see **Figure 7.15**, page 194). Bacteria have cell walls that are serviced metabolically (and a general metabolism within their cytoplasm), but viruses have neither.

page 196

11 This parasite has evolved a sucker (**Figure 7.16**, page 195) that is capable of burrowing through human skin submerged in infected water, if contact is maintained for a sufficiently long period.

12 A ciliated epithelium with numerous goblet cells lines the trachea and bronchi. The mucus secreted by the goblet cells traps any dust particles in the incoming air, thus protecting the alveoli in the lungs.

page 197

13 Antibodies are proteins secreted by the lymphocytes. Consequently, we know they are produced by ribosomes attached to endoplasmic reticulum (**rER**, page 250).

page 203

14 See **The onset of AIDS**, page 200.

15

Antigen	a substance capable of binding specifically to an antibody	**Antibody**	a protein produced by blood plasma cells derived from the B-lymphocytes when in the presence of a specific antigen; it binds, aiding destruction of the antigen
Antibiotic	an organic compound produced by microorganisms which selectively inhibits or kills other microorganisms	**Vaccine**	a preparation of microorganisms (or/ their antigens) which can induce protective immunity against the microorganism (pathogenic bacterium or virus), but which does not itself cause disease
Vector	an organism that transmits a disease-causing organism *or* a device for transferring genes during genetic engineering	**Host**	organism in or on which a parasite spends all or part of its life cycle

page 205

16

Characteristic	**How it influences diffusion**
a large, thin surface area	the greater the surface area and the shorter the distance that gases (O_2 and CO_2) have to diffuse, the quicker gas exchange occurs
a ventilation mechanism that moves air (or water) over the respiratory surface	the higher the concentration of oxygen on the 'supply side' of the respiratory membrane, the quicker gas can diffuse across the surface
a blood circulation that speeds up the removal of dissolved oxygen, with a respiratory pigment that increases the gas-carrying capacity of the blood	the quicker that oxygen is picked up and transported away from the gas exchange surface, the greater the rate of diffusion

The relationship of these factors is summarised by **Fick's Law of Diffusion**:

$$\text{rate of diffusion} \propto \frac{\text{surface area} \times \text{difference in concentration}}{\text{length of diffusion pathway}}$$

page 206

17

Inspiration		**Expiration**
relax	**internal intercostal muscles**	contract
contract	**external intercostal muscles**	relax
move upwards and outwards	**effects on rib cage**	move downwards and inwards

page 209

18 Carbon dioxide is an acidic gas which, if it were to accumulate in the blood, would alter the pH of the plasma solution. The normal pH of the blood is 7.4, and for life to be maintained it cannot be allowed to vary more than within the range pH 7.0–7.8. This is largely because blood pH affects the balance of essential ions which are transported in the plasma solution. Efficient removal of respiratory CO_2 at the lungs is as important to life as efficient uptake of O_2.

19 See **Gaseous exchange**, page 204.

20 The surfaces of the alveoli are moist, and oxygen gas dissolves in the surface film of water. It is oxygen in solution that diffuses into the red cells and combines with the haemoglobin there. Consequently the outgoing air contains water vapour that has evaporated during gaseous exchange.

page 213

21 **a** See **The resting potential**, page 211.
b See **The action potential**, page 212.

page 216

22 **a** See **Junctions between neurones**, page 214.
b See **Figures 7.35** and **7.36**, pages 215 and 216.

page 220

23 Body temperature falls significantly during sleep, but rises above normal during periods of physical activity.

24 **Endergonic reactions** require energy input, because the products have more potential energy than the reactants – see also **Figure 8.28**, page 261. **Note:** the alternative type of reaction is one in which the products have less potential energy than the reactants, and these reactions transfer energy as heat and work. They are called exergonic reactions.

page 223

25 The liver cells receive oxygen from the blood delivered by the hepatic artery.

26 Mitochondria are the type of organelle would you most expect to see when a liver cell is examined by EM.

page 230

27 See **Roles of hormones in the control of reproduction**, page 228.

page 233

28 *This is a personal issue which needs handling with great sensitivity. When you discuss a response to this question with your peers it is best if each individual explains their own approach and is listened to carefully. It may be that a consensus concerning a suitable response can be arrived at – but there are circumstances over issues of this sort where individuals cannot agree. In this situation, it is highly desirable to understand the approach of those you do not agree with rather than to have conflict.*

8 Nucleic acids and proteins

page 235

1 In summary:
organic base □ pentose sugar □□□□□□□□□ **nucleoside**
nucleoside □ phosphoric acid □□□□□□□□□ **nucleotide**
many nucleotides condensed together □□□□□ **polynucleotide (nucleic acid)**

See **Nucleotides** and **Figure 2.25**, page 63; **Nucleotides become nucleic acid** and **Figure 2.26**, page 64; **Table 8.1**, page 243.

page 237

2 Given the length of the DNA in each chromosome (of which there are 46 in human nuclei), together with the movements required of the chromosomes for a successful outcome of nuclear division (particularly mitosis, but also meiosis), compact packaging is essential. See **Chromosome structure and the packaging of DNA**, page 235; **Mitosis** and **Figure 1.33**, page 32.

page 238

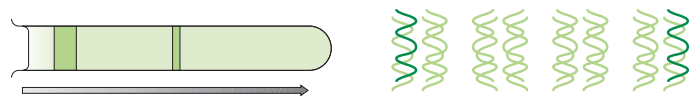
3 Hershey and Chase would have expected that only the bacteria infected with virus labelled with ^{35}S (an element in proteins but not in nucleic acids) would have produced radioactive virus. Remember, they were not aware that the protein coat of the virus remained outside of the host cell at the time of infection.

page 240

4 A hydrogen bond is an electrostatic attraction between the positively charged region of one molecule and the negatively charged region of a neighbouring one; the attraction gives rise to a relatively weak bond (compared to a covalent bond). Collectively, the H-bonds of DNA hold the polynucleotide strands together. See **Hydrogen bonds** and **Figure 2.1**, page 38.

page 241

5 After three generations, 75% of the DNA would be 'light'.



The experimental results that would be expected if the Meselson–Stahl experiment were carried on for three generations

page 245

6 See **Base pairing and 'direction' in the DNA molecule** and **Figure 8.5**, page 238–239.

page 250

7 See **Figure 2.15**, page 52.

8 These are illustrated in sequence in **Figure 8.8**, page 246.

page 251

9 **a** The sequence of amino acids coded for by the sequence of bases is found using the genetic code (**Figure 8.14**):

GGU	AAU	CCU	UUU	GUU	ACU	CAU	UGU
Gly	Asn	Pro	Phe	Val	Thr	His	Cys

b The sequence of bases in the antisense strand of DNA from which this mRNA was transcribed would have been:

CCA	TTA	GGA	AAA	CAA	TGA	GTA	ACA

page 252

10

Structural proteins supporting DNA	Enzymes
■ histones of nucleosomes ■ non-histone protein of 'scaffold'	■ helicase – unwinds DNA helix; holds strands apart for replication ■ DNA polymerase III, RNA primase, and polymerase I; involved in DNA replication ■ RNA polymerase; involved in transcription

See **Chromosome structure and the packaging of DNA**, page 235	See **Replication – DNA copying itself**, page 240, and **The sequence of events at transcription**, page 245

page 253

11 See **The central dogma** and **Table 8.3**, pages 252–253.

page 257

12 See **Figure 8.19**, page 255.

page 260

13 Membrane proteins are important structurally (see **Figure 1.19**, page 21) and as receptors, enzymes and molecular pumps (see **Figure 1.22**, page 23).

page 264

14

Lock and key hypothesis	Induced fit hypothesis
Refers to the pocket or crevice on the enzyme surface (at the active site) where the substrate molecule binds to the enzyme.	Binding occurs at the active site where the arrangement of certain amino acid residues matches certain groupings on the substrate molecule.
Binding is a prelude to catalysis, but this hypothesis does not explain the simultaneous chemical change as binding occurs.	As the complex forms, an essential, critical change of shape is induced in the enzyme molecule which raises the substrate molecule to the transitional state from which the catalysed reaction follows.

9 Energy transfer in cells II

page 272

1 The following are produced during glycolysis:
NADH
ATP
pyruvate.

2 Length = 6.25 □m, width = 2.1 □m.

page 274

3 a **Dehydrogenases**
In respiration, all the hydrogen atoms are gradually removed from glucose, catalysed by dehydrogenase enzymes. They are added to hydrogen acceptors, usually NAD (nicotinamide adenine dinucleotide, page 271), which itself is reduced. We can write this addition of hydrogen to its carrier as:
NAD □ 2H □ □ □ □ □ □ NADH$_2$
but what actually happens is
NAD□ □ 2H □ □ □ □ □ □ NADH □ H□
So NAD is a coenzyme that works with specific dehydrogenase enzymes in the oxidation of substrate molecules by the removal of hydrogen.

b **Decarboxylases**
Decarboxylation is the removal of carbon from organic compounds by the formation of carbon dioxide. For example, glucose consists of six carbon atoms. All six carbon atoms are removed at different stages of respiration, one at a time, and given off as carbon dioxide. A specific decarboxylase enzyme is involved in each case. The first decarboxylation in aerobic respiration occurs in the reaction linking glycolysis with the Krebs cycle, when pyruvate is converted to a 2-carbon molecule. The other decarboxylation reactions of aerobic respiration occur in steps in the Krebs cycle.

page 275

4 In the absence of oxygen, reduced NAD (NADH$_2$) accumulates and oxidised NAD reserves are used up. In the absence of NAD□, pyruvate production by glycolysis slows and stops, so subsequent steps in respiration stop too.

5 See **Figure 9.7**, page 275. Protons move from the inter-membrane space to the matrix (the interior) of the mitochondrion.

page 276

6 a Substrate and intermediate:
Substrate □ molecule that is the starting point for a biochemical reaction; it forms a complex with a specific enzyme.
Intermediate □ metabolite formed as a component of a metabolic pathway.

b Glycolysis and the Krebs cycle:
See **Cellular respiration: Glycolysis** page 271, and **The mitochondria and the steps to the Krebs cycle**, page 272.

c Oxidation and reduction:
See **Respiration as a series of redox reactions**, page 270.

page 279

7

Photosystem I	Role
many accessory pigment molecules	harvest light energy and funnel the energy to a single chlorophyll molecule of the reaction centre
reaction centre of chlorophyll a molecule absorbing light of 700 nm wavelength	has ground-state electrons that are raised to an excited state when light energy is received from accessory pigments – excited electrons passed to oxidised NADP
electron-carrier molecules	receive excited electrons and pass these on
Photosystem II	**Role**
many accessory pigment molecules	harvest light energy and funnel the energy to a single chlorophyll molecule of reaction centre
reaction centre of chlorophyll a molecule absorbing light of 680 nm wavelength	has ground-state electrons that are raised to an excited state when light energy is received from accessory pigments – excited electrons passed to reaction centre of photosystem I
water-splitting enzyme	catalyses splitting of water into hydrogen ions, electrons and oxygen atoms
electron-carrier molecules	receive excited electrons and pass these on; simultaneous pumping of protons from stroma to thylakoid compartment

See **Figure 9.12**, page 280.

page 280

8 Electrons displaced from the reaction centre of photosystem II in an excited state are first passed to the reaction centre of photosystem I. Here they are again raised to an excited state and this time they are passed to oxidised NADP ($NADP^+$) to form reduced NADP ($NADPH + H^+$).

page 281

9 An **isotonic** solution is of the same osmotic concentration as the chloroplasts which therefore are not disrupted by excess movement of water into or out of their delicate structure. As a **buffer** solution, this medium maintains the pH at that of the cell contents, thereby ensuring enzyme action is not interfered with.

The presence of **ice** keeps the solution and organelles at or just above 0 °C, thereby preventing unwanted biochemical reactions while the extracted organelles are held before investigation.

page 284

10 *Chlorella* cells in suspension function biochemically like mesophyll cells but they can be cultured in suspension, sampled without interference of the biochemistry of the other cells, and supplied directly and quickly with light, intermediates and inhibitors (if required) in such a way that all the cells are treated identically. Gaseous exchange and diffusion occurs without the complexity (and delays) of the intact leaf with its air spaces and stomata.

page 287

11 **a** Light-dependent reactions and light-independent reactions:
See **The reactions of photosynthesis**, and **Figure 9.10**, page 278.

b Photolysis and photophosphorylation:
See **The reactions of photosynthesis**, page 278.

12 The starting materials (carbon dioxide and water) are in an oxidised (low energy) state. The end product of photosynthesis – glucose – is in a reduced (higher energy) state.

13 The experimental material is a (more or less) intact plant, rather than isolated chloroplasts where the light is absorbed and exploited in the light reaction.

Difficult to precisely control the wavelength of light received by the chloroplasts, in the mesophyll cells, deep inside the leaves.

Difficult to ensure that the plant behaves identically at each wavelength over the extended period of time needed to collect a measurable sample of oxygen gas (evidence of the rate of respiration).

page 288

14 **a** Chlorophyll and chloroplast:
See **Investigating chlorophyll**, page 83, and **Chloroplast, grana and stroma – the venue for photosynthesis**, page 287.

b Stroma and grana:
See **The steps of photosynthesis** and Figure 9.9, page 277.

c Absorption spectrum and action spectrum:
See **Light and photosynthesis**, page 287.

page 291

15 In cyclic photophosphorylation, photoactivation of photosystem I and production of ATP occur. Reduction of $NADP^+$ does not occur (cyclic photophosphorylation occurs when NADPH has accumulated in the illuminated green leaf because the fixation of carbon dioxide is prevented), see **Figure 9.24**, page 290.

10 Plant science

page 297

1 The features you are likely to emphasise include:
Leaves – whether they are simple or the leaf blade is divided into leaflets; the arrangement of the 'veins' of the blade, whether they are parallel or in a network.
Stem – its length, and whether it branches or not, and how the leaves are attached.
Root – whether a tap root or a fibrous system.
Flowers – whether solitary or grouped together (an inflorescence), numbers of parts (typically in 3s, 4s, or 5s or multiples), and whether they are hermaphrodite flowers (they typically are).

page 298

2 The shape of the cellulose polymer allows close packing into long chains (**Figure 1.14**, page 16) held together by hydrogen bonds. These are laid down in porous sheets, and have great tensile strength. The monomer from which cellulose is assembled by condensation reaction is glucose, of which green plants typically have an excellent supply.

page 303

3 Light has diverse effects on plants, including:

- being required for the manufacture of sugar in photosynthesis
- inhibiting extension growth in length of stems (so plants growing in the dark 'bolt', whereas plants in full light are short and sturdy)
- causing positive phototropic growth and sun-tracking
- promoting expansion of leaf blades
- being required for the formation of chlorophyll (so plants in the dark are yellow)
- triggering the switch from vegetative growth to flowering in many species, in response to some particular regime of light and dark (day and night) cycles
- triggering germination in a few species.

page 305

4 The features of root hairs that facilitate absorption from the soil are:

- they greatly increase the surface area of the root
- they grow in close contact with the film of soil water that occurs around mineral particles where the essential resources (water and soluble ions) occur
- their walls are of cellulose and are in intimate contact with cells of the cortex in a region that is permeable to water.

page 307

5 See **Figure 10.13**, page 306.

6 Plant growth is dependent on a supply of chemically combined nitrogen (for amino acid and protein synthesis) of which nitrates are typically the most readily available. However, nitrates are also taken up by microorganisms, and may be released in the soil at times other than when plant demand is at its peak. They are also very soluble and are easily leached away into ground water in heavy rain. By taking up nitrate whenever it becomes available (and hording it in cells), plants can maintain growth at peak times.

7 Plant roots are metabolically very active and require oxygen for aerobic cell respiration and ATP formation (for ion uptake, for example). Waterlogged soil lacks soil air and the essential oxygen gas.

page 311

8 See **Transport of water through the plant**, and **Figure 10.18**, pages 310–311.

page 312

9 See **Stomata and transpiration**, and **Figures 10.19** and **10.20**, pages 312–313.

page 313

10 See **Stomata and transpiration**, and **Figures 10.19** and **10.20**, pages 312–313.

page 317

11 Transpiration is a direct consequence of plant structure, plant nutrition, and the mechanism of gaseous exchange in leaves. In effect, the living green plant is like a wick that steadily dries the soil around it. Put like this, transpiration is an unfortunate consequence of plant structure and metabolism, rather than a valuable process. It means that a constant, adequate supply of water is critical to plant growth; if the water supply fails, plants cannot move away.

12 In the veins and vascular bundles of the leaf, the water-conducting tissue (xylem) is on the upper side (with phloem below – because of their relative positions in stem vascular bundles with which they connect directly). So, looking at this feature of Marram grass leaf, we see the outer epidermis is strictly the lower epidermis.

page 319

13 See **Transport of water through the plant**, page 310 and **The stem and organic solute transport**, page 000.

page 321

14 See **Pollination and fertilisation**, page 320.

page 322

15

Ovary structure	Fruit or vegetable
a an ovary containing one seed	plum, peach, avocado
b an ovary containing many seeds	pea pod, runner bean
c several ovaries fused together, containing many seeds	gooseberry, cucumber, tomato

11 Genetics II

page 333

1

The process	Mitosis	Meiosis First division	Second division
Prophase	■ chromosomes become visible, finally as 2 chromatids joined at the centromere ■ nuclear membrane breaks down ■ spindle forms	■ chromosomes become visible ■ homologous chromosomes pair to form bivalents ■ nuclear membrane breaks down ■ spindle forms	■ chromosomes reappear as 2 chromatids joined at the centromere ■ spindle forms
Metaphase	■ chromosomes at equator of spindle, attached by centromeres to fibre	■ bivalents at equator of spindle, centromeres attached to a fibre	■ chromosomes at equator of spindle, attached by centromeres to fibre
Anaphase	■ centromeres divide ■ chromatids move to opposite poles	■ chromosomes (each two sister chromatids joined at the centromeres) move to opposite poles	■ centromeres divide ■ chromatids move to opposite poles
Telophase	■ spindle breaks down ■ nuclear membrane reforms ■ cytoplasm divides	■ spindle breaks down ■ nuclear membrane may reform ■ cytoplasm may divide or may be delayed	■ spindle breaks down ■ nuclear membrane reforms around both nuclei ■ cytoplasm divides
The products			
	■ 2 identical cells, each with the diploid chromosome number	■ 2 non-identical nuclei each with the haploid chromosome number (1 of each homologous pair per nucleus)	■ 4 non-identical cells, each with the haploid chromosome number
The consequences and significance			
	■ produces identical diploid cells ■ permits growth within multicellular organisms, and asexual reproduction	■ produces haploid cells ■ contributes to genetic variability by: – reducing chromosome number by half, permitting fertilisation, and the combination of genes from two parents – permitting random assortment of paternal and maternal homologous chromosomes – recombination of segments of individual maternal and paternal homologous chromosomes during crossing over	

table continues

page 334

2 From the information available in **Figure 11.6**:
a the heterozygotes were the F_1 progeny – round and yellow;
b the recombinants (due to reassortment) among the F_2 progeny were 'round and green' and 'wrinkled and yellow'.

3 These genes are on separate chromosomes (so there is no chance of recombinants due to crossing over, incidentally). Because the

alleles of these genes are on separate chromosomes, the dihybrid ratio was obtained by Mendel. He would have surely been confused if (by chance) he had chosen two pairs of contrasting characters controlled by genes on the same chromosome.

page 337

4 See **Abrupt change in hereditary information – mutations**, page 98.

page 338

5 Using the lay-out given in **Figure 11.7**:

P normal wing ♀ vestigial wing
normal body ebony body
homozygous homozygous
WWGG **wwgg**

gametes **WG** ♂ **wg**

offspring (F_1) **WwGg**
genotype heterozygous
phenotype normal wing
normal body

sibling cross **WwGg** ♀ **WwGg**

gametes **WG Wg wG wg** ♂ **WG Wg wG wg**

(**Punnett grid**)

offspring (F_2)

genotypes	$\frac{1}{16}$ **WWGG**	$\frac{1}{16}$ **WWgg**	$\frac{1}{16}$ **wwGG**	$\frac{1}{16}$ **wwgg**
	$\frac{1}{8}$ **WWGg**	$\frac{1}{8}$ **Wwgg**	$\frac{1}{8}$ **wwGg**	
	$\frac{1}{8}$ **WwGG**			
	$\frac{1}{4}$ **WwGg**			
phenotypes	normal wing normal body	normal wing ebony body	vestigial wing normal body	vestigial wing ebony body
phenotype ratio	9 ::	3 :	3 :	1

page 340

6 In order to create gene maps of linkage groups, it is necessary to carry out crosses between known genotypes that yield large samples of offspring. From the outcomes it is possible to calculate statistically significant recombination frequencies. Obviously, it is quite impossible to conduct such breeding experiments with humans. Rather, it is the exercise to map the human genome (**Human Genome Project**, page 131) that has produced maps of the positions of the genes on each of our chromosomes.

page 343

7 The recombinants would be:

Tb **tB**
tB **tB**

page 344

8

Phenotypes	white	light	medium	dark	black
Range of genotypes	$\frac{1}{16}$ **aabb**	$\frac{2}{16}$ **Aabb** $\frac{2}{16}$ **aaBb**	$\frac{1}{16}$ **AAbb** $\frac{4}{16}$ **AaBb** $\frac{1}{16}$ **aaBB**	$\frac{2}{16}$ **AaBB** $\frac{2}{16}$ **AABb**	$\frac{1}{16}$ **AABB**
Phenotype ratio	1	4	6	4	1

12 Human physiology, health and reproduction II

page 350

1 See **Figure 12.1**, page 349.

page 351

2

Inflammation response	rapid, localised response to cuts, blows or bites – involves vasodilation of capillaries and increased permeability of blood vessels, bringing phagocytic white cells to site, for example
Clotting of blood	sealing of gaps due to local haemorrhage, cuts or breaks in vessels, see **The blood clotting mechanism**, page 349
Immune response	production of specific antibodies, see **The immune system and the response to invasion**, page 350

page 352

3 See **What recognition of 'self' entails**, page 350, and **Steps of antibody production**, page 351.

page 355

4 Antigens may be present more or less anywhere in the body that becomes contaminated from outside. Antibodies exist in the blood and lymph, and may be carried in the plasma solution anywhere that blood 'leaks out' to – including sites of invasions.

page 356

5 See pages 353–355, points 6 and 7.

page 360

6 The existence of memory cells avoids the steps in the production of activated T-cells.

The particular memory cell, once re-activated by the re-invading antigen, switches into production of an excess of appropriate plasma cells and helper T-cells (**Figure 12.4**, page 354).

page 361

7 Immunity is resistance to the onset of disease after infection by harmful microorganisms or internal parasites. Long-lived specific immunity is a result of the action of the immune system; it may be acquired naturally by previous infection, or can be induced by vaccination. See **Types of immunity**, page 351.

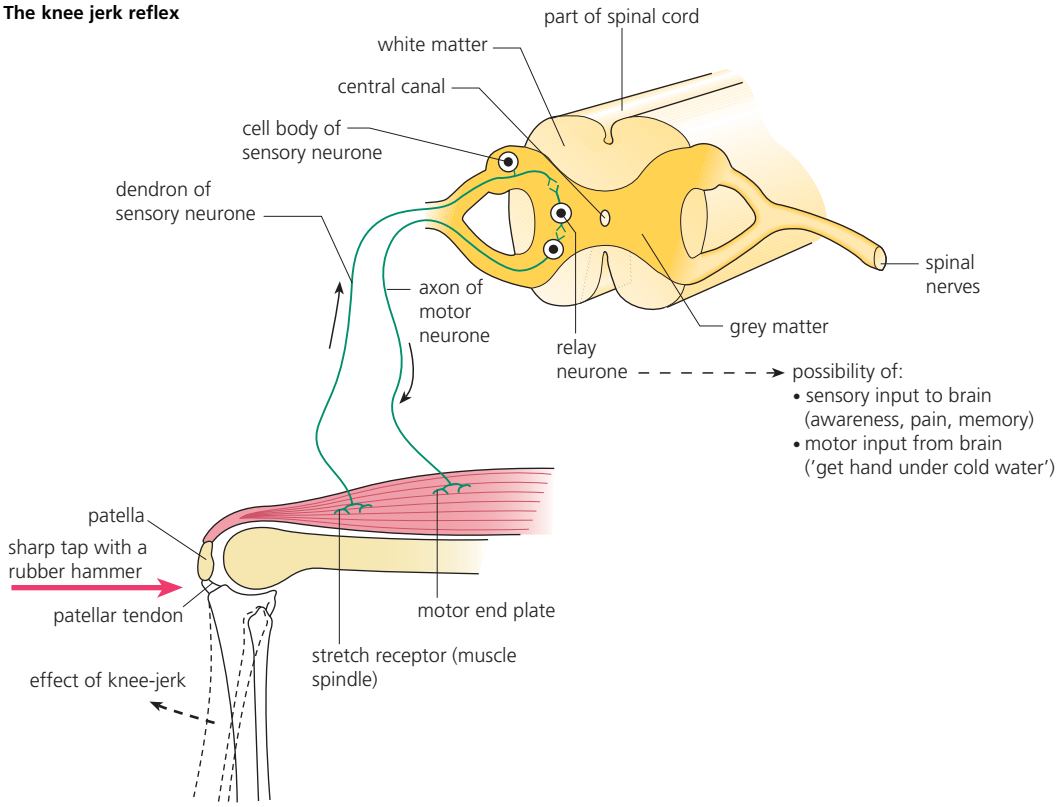
page 365

8 See diagram at the top of page B405.

page 367

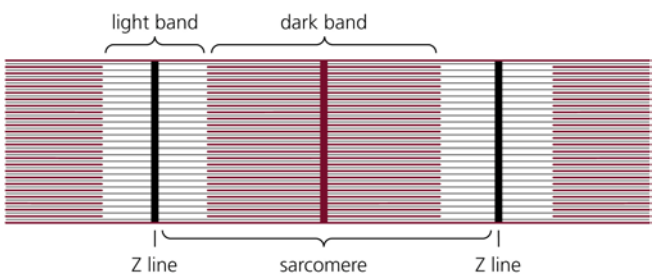
9 See **The ultrastructure of skeletal muscle**, and **Figures 12.13** and **12.15**, page 000.

The knee jerk reflex



10 a The myofibril shown in the sketch has the same proportions of 'light' and 'dark' bands as the slightly contracted sarcomere shown above.

b Sketch of a fully contracted myofibril



page 371

11

Excretion	elimination of the waste products of metabolism from the body (e.g. removal of urea from the blood and the formation of urine in the kidney tubules)
Egestion	disposal of waste from the body (e.g. defecation)
Osmoregulation	regulation of water potential – osmotic concentration – of body fluids by the control of water and salt content (e.g. actions of distal convoluted tubule and collecting duct of kidney)
Secretion	material produced and released from glandular cells (e.g. saliva by salivary glands)

page 372

12 The source of energy for ultrafiltration in the glomerulus is respiration of heart muscle that 'fuels' contractions of the ventricles and thus creates blood pressure in the glomerulus. This pressure is heightened by the efferent arteriole of the glomerulus being of smaller diameter than the afferent arteriole (see **Figure 12.21**).

page 373

13 See **Selective reabsorption in the proximal convoluted tubule**, and **Figure 12.22**, pages 372–373.

page 374

14 See **Figure 12.25**, page 375.

page 376

15 An excess of offspring are typically produced by breeding pairs, yet the numbers in most species remain fairly constant over time. When an excess of offspring do survive, there is generally strong competition for resources resulting in a speedy return to normal population numbers.

page 382

16 The role of FSH in sperm production, see **Figure 12.30**, page 379. The role of FSH in ovum production, see **Role of hormones in the control of reproduction**, and **Figure 7.52**, page 228.

page 386

17 The fetus is 'foreign' tissue to the mother (for example, they may not be of the same blood group) and carries antigens foreign to her (see **What recognition of 'self' entails**, page 350).

18

Structural feature	Importance
large, disc-shaped structure, composed of tissue of two two different organisms (fetal membranes and part of uterus wall)	large area for exchange of essential metabolites between different organisms
maternal and fetal blood circulations brought close together	conditions favour diffusion without meeting of the two blood circulations
finger-like projections of fetal membranes grow into the endometrium and become bathed by maternal blood	surface area enormously increased

page 387

19 The chief effect of HCG is to maintain the corpus luteum as an endocrine gland, at least for the first 16 weeks of the pregnancy. The maintenance of the corpus luteum means that the endometrium persists, to the benefit of the implanted embryo, and menstruation is prevented. (The premature demise of the corpus luteum, leading to degeneration of the endometrium during early pregnancy, is a possible cause of a miscarriage.)

page 388

20 In **negative feedback** the effect of a deviation from the normal or set condition is to create a tendency to eliminate the deviation. Negative feedback is a part of almost all control systems in living things. The effect of negative feedback is to reduce further corrective action of the control system once the set-point value is reached.

In **positive feedback** the effect of a deviation from the normal or set condition is to create a tendency to reinforce the deviation. Positive feedback intensifies the corrective action taken by a control system, so leading to a vicious-circle situation. Imagine a car in which the driver's seat was set on rollers (not secured to the floor), being driven at speed. The slightest application of the foot brake causes the driver to slide and to press harder on the brake as the car starts to slow, with an extreme outcome.

Biological examples of positive feedback are rare, but one can be identified at the synapse. When a wave of depolarisation (a nerve impulse) takes effect in the post-synaptic membrane, the entry of sodium ions triggers the entry of further sodium ions at a greater rate. This is a case of positive feedback. The depolarised state is established, and the impulse moves along the post-synaptic membrane.

Glossary

Entries that are IB action verbs and entries that are IB syllabus-required definitions are coloured red and blue respectively. Entries are *aides-mémoire*, rather than formal definitions.

A

abiotic factor a non-biological factor (e.g. temperature) that is part of the environment of an organism

abscisic acid a plant growth substance tending to inhibit growth

absorption spectrum range of a pigment's ability to absorb various wavelengths of light

acetylcholine a neurotransmitter, liberated at synapses in the CNS

acid rain the cocktail of chemical pollutants that may occur in the atmosphere

action potential rapid change (depolarisation) in membrane potential of an excitable cell (e.g. a neurone)

action spectrum range of wavelengths of light within which a process like photosynthesis takes place

activation energy energy a substrate molecule must have before it can undergo a chemical change

active site region of enzyme molecule where substrate molecule binds

active transport movement of substances across a membrane involving a carrier protein and energy from respiration

adenine a purine organic base, found in the coenzymes ATP and NADP, and in nucleic acids (DNA and RNA) in which it pairs with thymine

adenosine diphosphate (ADP) a nucleotide, present in every living cell, made of adenosine and two phosphate groups linked in series, and important in energy transfer reactions of metabolism

adenosine triphosphate (ATP) a nucleotide, present in every living cell, formed in photosynthesis and respiration from ADP and P_i, and functioning in metabolism as a common intermediate between energy-requiring and energy-yielding reactions

adrenaline a hormone secreted by the adrenal medulla (and a neurotransmitter secreted by nerve endings of the sympathetic nervous system), having many effects, including speeding of heart beat, and the breakdown of glycogen to glucose in muscle and liver

aerobic respiration respiration requiring oxygen, involving oxidation of glucose to carbon dioxide and water

alimentary canal the gut; a tube running from mouth to anus in vertebrates, where complex food substances are digested and the products of digestion selectively absorbed into the body

allele an alternative form of a gene, occupying a specific locus on a chromosome

allele frequency the commonness of the occurrence of any particular allele in a population

alpha cell (pancreas) glucagon-secreting cell of the islets of Langerhans in the pancreas

alveolus air sac in the lung

amino acid building block of proteins, of general formula $R.CH(NH_2).COOH$

anabolism the building up of complex molecules from smaller ones

anaerobic respiration respiration in the absence of oxygen, involving breakdown of glucose to lactic acid or ethanol

analogous structure similar in structure but of different evolutionary origin

analyse interpret data to reach a conclusion

anion negatively charged ion

annotate add brief notes to a diagram, drawing or graph

anther part of the stamen in flowers, consisting of pollen sacs enclosed in walls that eventually split open, releasing pollen

antibody a protein produced by blood plasma cells derived from B lymphocytes when in the presence of a specific antigen, which then binds with the antigen, aiding its destruction

antibiotics organic compounds produced by some microorganisms which selectively inhibit or kill other microorganisms

anticodon three consecutive bases in tRNA, complementary to a codon on RNA

antidiuretic hormone (ADH) hormone secreted by the pituitary gland that controls the permeability of the walls of the collecting ducts of the kidney

antigen a substance capable of binding specifically to an antibody

apoplast collective name for the cell walls of a tissue or plant

apply use an idea, equation, principle, theory, or law in a new situation

aqueous humour fluid between lens and cornea of the eye

arteriole a very small artery

artificial classification classifying organisms on the basis of few, self-evident features

artificial selection selection in breeding exercises, carried out deliberately, by humans

asexual reproduction reproduction not involving gametes and fertilisation

assimilation uptake of nutrients into cells and tissues

atherosclerosis deposition of plaque (cholesterol derivative) on inner wall of blood vessels

atrio-ventricular node mass of tissue in the wall of the right atrium, functionally part of the pacemaker mechanism

atrio-ventricular valve tricuspid or bicuspid valve

atrium (plural, **atria**) one of the two upper chambers of the mammalian four-chambered heart

autolysis self-digestion
autotrophic (organism) self-feeding – able to make its own elaborated foods from simpler substances
autonomic the involuntary nervous system
auxin plant growth substance, indoleacetic acid
axon fibre carrying impulses away from the cell body of a neurone

B

bacillus a rod-shaped bacterium
bacteriophage a virus that parasitises bacteria (also known as a phage)
baroreceptor a sensory receptor responding to stretch, in the walls of blood vessels
basement membrane the thin fibrous layer separating an epithelium from underlying tissues
beta cell (pancreas) insulin-secreting cells of the islets of Langerhans in the pancreas
bicuspid valve valve between atrium and ventricle on the left side of the mammalian heart
bile an alkaline secretion of liver cells which collects in the gall bladder in humans, and which is discharged into the duodenum periodically
binary fission when a cell divides into two daughter cells, typically in reproduction of prokaryotes
binomial system double names for organisms, in Latin, the generic preceding the specific name
biological pest control control of pests and weeds by other organisms
biomagnification the process by which chemical substances become more concentrated at each trophic level
biomass total mass of living organisms in a given area (e.g. a quadrat)
biome a major life-zone over an area of the Earth, characterised by the dominant plant life present
biosphere the inhabited part of the Earth
biotechnology the industrial and commercial applications of biology, particularly of microorganisms, enzymology and genetic engineering
biotic factor the influence of living things on the environment of other living things
bivalent a pair of duplicated chromosomes, held together by chiasmata during meiosis
blastocyst embryo as hollow ball of cells, at the stage of implantation
blind spot region of the retina where the optic nerve leaves
body mass index (BMI) body mass in kg/(height in m)^2
bone marrow tissue special connective tissue filling the cavity of certain bones
boreal forest northern coniferous forests (example of a biome)
bovine somatotrophine (BST) hormone produced by the pituitary, controlling milk production
brain the coordinating centre of the nervous system
breed (animal) the animal equivalent of a plant variety
bronchiole small terminal branch of a bronchus
bronchus a tube connecting the trachea with the lungs
brush border tiny, finger-like projections (microvilli) on the surface of epithelial cells of the small intestine
buffer a solution which minimises change in pH when acid or alkali are added
bundle of His bundles of long muscle fibres that transmit myogenic excitation throughout the ventricle walls

C

C_3 pathway the light-independent reaction in photosynthesis, producing as its first product, a 3-carbon compound, glycerate 3-phosphate
C_4 plants plants with an additional carbon dioxide-fixation pathway that augments the supply of this raw material of photosynthesis at the chloroplast
calculate find an answer using mathematical methods
Calvin cycle a cycle of reactions in the stroma of the chloroplast by which some of the product of the dark reaction is re-formed as the acceptor molecule for carbon dioxide (ribulose biphosphate)
carrier an individual that has one copy of a recessive allele that causes a genetic disease in individuals that are homozygous for this allele
carrier protein one of the types of protein in plasma membranes, responsible for active transport across the membranes
cardiac cycle the stages of the heart beat, by which the atrial and then the ventricle walls alternately contract (systole) and relax (diastole)
carnivore flesh-eating animal
cartilage firm but plastic skeletal material (e.g. cartilage over bones at joints)
Casparian strip band of cells with impervious walls, found in plant roots
catabolism the breaking down of complex molecules in the biochemistry of cells
catalyst a substance that alters the rate of a chemical reaction, but remains unchanged at the end
cellular respiration controlled release (transfer) of energy from organic compounds in cells to form ATP
cellulase enzyme capable of hydrolysing cellulose
cellulose an unbranched polymer of 2000–3000 glucose residues, the major ingredient of most plant walls
central dogma the idea that transfer of genetic information from DNA of the chromosome to mRNA to protein (amino acid sequence) is irreversible
centromere constriction of the chromosome, the region that becomes attached to the spindle fibres in division
centrosome organelle situated near the nucleus in animal cells, involved in the formation of the spindle prior to nuclear division
cephalisation development of a head at the anterior of an animal
cerebellum part of hindbrain, concerned with muscle tone, posture and movement
cerebral cortex superficial layer of grey matter on extension of forebrain, much enlarged in humans and apes

cerebral hemispheres (cerebrum) the bulk of the human brain, formed during development by the outgrowth of part of the forebrain, consisting of densely packed neurones and myelinated nerve fibres

chemoautotroph an organism that uses energy from chemical reactions to generate ATP and produce organic compounds from inorganic substances

chemoheterotroph an organism that uses energy from chemical reactions to generate ATP and obtains organic compounds from other organisms

chemoreceptor a sense organ receiving chemical stimuli

chemosynthesis use of chemical energy from oxidation of inorganic compounds to synthesise organic compounds, typically from carbon dioxide and water

chiasma (plural, **chiasmata**) site of crossing over (exchange) of segments of DNA between homologous chromosomes

chloroplast organelle that is site of photosynthesis and contains chlorophyll

chlorophyll the main photosynthetic pigment of green plants, occurs in the grana membranes (thylakoid membranes) of the chloroplasts

cholesterol a lipid of animal plasma membranes; a precursor of the steroid hormones, in humans, formed in the liver and transported in the blood as lipoprotein

chromatid one of two copies of a chromosome after it has replicated

chromatin a nuclear protein material in the nucleus of eukaryotic cells at interphase; forms into chromosomes during mitosis and meiosis

choroid layer of blood vessels lying below the retina

chromosome visible in appropriately stained cells at nuclear division, each chromosome consists of a long thread of DNA packaged with protein; chromosomes replicate prior to division, into chromatids. Contents of nucleus appears as granular chromatin between divisions

chyme partly digested food as it leaves the stomach

cilium (plural, **cilia**) motile, hair-like outgrowth from surface of certain eukaryotic cells

citric acid cycle see *Krebs cycle*

clade the branch of a phylogenetic tree containing the set of all organisms descended from a particular common ancestor which is not an ancestor of any non-member of the group

cladistics method of classifying living organisms that makes use of lines of descent only (rather than phenotypic similarities)

climax community the mature (stable) stage of a succession of communities

clone a group of genetically identical individuals (or cells)

coccus spherical bacterial cell

CNS see nervous system

codominant alleles pairs of alleles that both affect the phenotype when present in a heterozygous state

codon three consecutive bases in DNA (or RNA) which specify an amino acid

coleoptile protective sheath around emerging leaves of germinating grass seeds

colon part of the gut, preceding the rectum

colostrum first milk secreted by the mother, after birth of young

commensalism a mutually beneficial association between two organism of different species

comment give a judgement based on a given statement or result of a calculation

community a group of populations of organisms living and interacting with each other in a habitat

compare give an account of similarities and differences between two or more items (e.g. by using a table)

compensation point the point where respiration and photosynthesis are balanced

condensation reaction formation of larger molecules involving the removal of water from smaller component molecules

cone (retinal cell) a light-sensitive cell in the retina, responsible for colour vision

conjugate protein protein combined with a non-protein part

connective tissue tissues that support and bind tissues together

conservation applying the principles of ecology to manage the environment

contractile vacuole a small vesicle in the cytoplasm of many fresh water protozoa that expels excess water

construct represent or develop in graphic form

cornea transparent covering at the front of the eye

corpus luteum glandular mass that develops from an ovarian follicle in mammals, after the ovum is discharged

cotyledon the first leaf (leaves) of a seed plant, found in the embryo

covalent bond bond between atoms in which electrons are shared

cristae folds in the inner membrane of mitochondria

crossing over exchange of genetic material between homologous chromosomes during meiosis

crypt of Lieberkuhn endocrine cells within the pancreas

cuticle layer of waxy material on outer wall of epidermis

cyanobacteria photosynthetic prokaryotes

cytokinesis division of cytoplasm after nucleus has divided into two

cytology study of cell structure

cytoplasm living part of the cell bound by the plasma membrane, excluding the nucleus

cytosol what remains of cytoplasm when the organelles have been removed

D

data recorded products of observations and measurements
qualitative data observations not involving measurements
quantitative data precise observations involving measurements

deamination the removal of NH_2 from an amino acid

deciduous loss at the end of the growing season (e.g. of leaves from broadleaved trees)

decomposer organisms (typically microorganisms) that feed on dead plant and animal material, causing matter to be recycled by other living things

degenerate code the triplet code contains more codons than there are amino acids to be coded, so most amino acids are coded for by more than one codon

deduce reach a conclusion from the information given

define give the precise meaning of a word or phrase as concisely as possible

denaturation a structural change in a protein that results in a loss (usually permanent) of its biological properties

dendrite a fine fibrous process on a neurone that receives impulses from other neurones

depolarisation (of axon) a temporary and local reversal of the resting potential difference of the membrane that occurs when an impulse is transmitted along the axon

derive manipulate a mathematical equation to give a new equation or result

describe give a detailed account including all relevant information

desertification the conversion of marginal cultivated land into desert, caused by climate change or by over-grazing or inferior cultivation

design produce a plan, object, simulation or model

determine find the only possible answer

detrital chain a food chain based on dead plant matter

detritivore an organism that feeds on detritus (dead organic matter)

dialysis separation of large and small molecules in solution by the inability of the former to pass through a selectively permeable membrane

diaphragm a sheet of tissues, largely muscle, separating thorax from abdomen in mammals

diastole relaxation phase in the cardiac cycle

dichotomous key one in which a group of organisms is progressively divided into two groups of smaller size

dicotyledon class of Angiospermophyta having an embryo with two seed leaves (cotyledons)

diffusion passive movement of particles from a region of high concentration to a region of low concentration

dihybrid cross one in which the inheritance of two pairs of contrasting characters (controlled by genes on separate chromosomes) is observed

diploid condition organisms whose cells have nuclei containing two sets of chromosomes

disaccharide a sugar that is a condensation product of two monosaccharides (e.g. maltose)

discuss give an account including, where possible, a range of arguments, assessments of the relative importance of various factors or comparisons of alternative hypotheses

distinguish give a difference between two or more different items

disulphide bond S—S bond between two S-containing amino acid residues in a polypeptide or protein chain

diuresis increased secretion of urine

division of labour the carrying out of specialised functions by different types of cell in a multicellular organism

DNA a form of nucleic acid found in the nucleus, consisting of two complementary chains of deoxyribonucleotide subunits, and containing the bases adenine, thymine, guanine and cytosine

dominant allele an allele that has the same effect on the phenotype whether it is present in the homozygous or heterozygous state

double bond a covalent bond involving the sharing of two pairs of electrons (rather than one)

double circulation in which the blood passes twice through the heart (pulmonary circulation, then systemic circulation) in any one complete circuit of the body

double fertilisation a feature of flowering plants in which two male nuclei enter the embryo sac, and one fuses with the egg cell and one with the endosperm nucleus

draw represent by means of pencil lines (with labels added)

duodenum the first part of the intestine after the stomach

E

ecology the study of relationships between living organisms and between organisms and their environment – a community and its abiotic environment

ecosystem a natural unit of living (biotic) components and non-living (abiotic) components (e.g. temperate deciduous forest)

edaphic factor factor influenced by the soil

effector an organ or cell that responds to a stimulus by doing something (e.g. a muscle contracting, a gland secreting)

egestion disposal of waste from the body (e.g. defecation)

egg cell an alternative name for an ovum

electron microscope (EM) microscope in which a beam of electrons replaces light, and the powers of magnification and resolution are correspondingly much greater

electron-transport system carriers that transfer electrons along a redox chain, permitting ATP to be synthesised in the process

embolism a blood clot blocking a blood vessel

embryo the earliest stages in development of a new animal or plant, from a fertilised ovum, entirely dependent on nutrients supplied by the parent

embryo sac occurs in the ovule of flowering plants, and contains the egg cell and endosperm nucleus

emulsify to break fats and oils into very tiny droplets
endemic species restricted to a particular region
endergonic reaction metabolic reaction requiring energy input
endocrine glands the hormone-producing glands that release secretions directly into the body fluids
endocytosis uptake of fluid or tiny particles into vacuoles in the cytoplasm, carried out at the plasma membrane
endoplasmic reticulum system of branching membranes in the cytoplasm of eukaryotic cells, existing as rough ER (with ribosomes) or as smooth ER (without ribosomes)
endosperm the stored food reserves within the seeds of flowering plants
endoskeleton an internal skeleton system
endothermic generation of body heat metabolically
endothelium a single layer of cells lining blood vessels and other fluid-filled cavities
enzyme mainly proteins (a very few are RNA) that function as biological catalysts
epidemiology the study of the occurrence, distribution and control of disease
epidermis outer layer(s) of cells
epiglottis flap of cartilage that closes of the trachea when food is swallowed
epiphyte plant living on the surface of other plants
epithelium sheet of cells bound strongly together, covering internal or external surfaces of multicellular organisms
erythrocyte red blood cell
estimate find an approximate value for an unknown quantity, based on the information provided and scientific knowledge
etiolation the condition of plants when grown in the dark
eukaryotic (cells) cells with a 'good' nucleus (e.g. animal, plant, fungi and protoctista cells)
evaluate assess the implications and limitations
evolution cumulative change in the heritable characteristics of a population
ex situ not in its original or natural position or habitat
excretion removal from the body of the waste products of metabolic pathways
exergonic reaction metabolic reaction releasing energy
exocytosis secretion of liquids and suspensions of very fine particles across the membrane of eukaryotic cells
exocrine gland gland whose secretion is released via a duct
exoskeleton skeleton secreted external to the epidermis of the body
exothermic chemical reaction that releases energy as heat (an endothermic reaction requires heat energy)
explain give a clear account including causes, reasons or mechanisms
expiratory emitting air during breathing
extensor muscle a muscle that extends or straightens a limb

F

F_1 generation first filial generation – arise by crossing parents (P), and when selfed or crossed via sibling crosses, produce the F_2 generation
facilitated diffusion diffusion across a membrane facilitated by molecules in the membrane (without the expenditure of metabolic energy)
fermentation anaerobic breakdown of glucose, with end-products ethanol and carbon dioxide or lactic acid
fetus a mammalian embryo when it becomes recognisable (e.g. the human embryo from 7 weeks after fertilisation)
fertilisation the fusion of male and female gametes to form a zygote
field layer the layer of herbaceous plants in a forest or wood
filter-feeding feeding on tiny organisms which are strained from the surrounding medium
fimbria (singular, **fimbrium**) thin, short filaments protruding from some bacteria, involved in attachment
flaccid state of a tissue with insufficient water, as in wilting leaves
flagellum (plural, **flagella**) a long thin structure, occurring singly or in groups on some cells and tissues, and used to propel unicellular organisms, and to move liquids past anchored cells (flagella of prokaryotes and eukaryotes are of different internal structure)
flexor muscle a muscle that on contraction bends a limb (or part of a limb)
flower develops from the tip of a shoot, with outer parts (e.g. sepals, petals) surrounding the male and female reproductive organs
fluid mosaic model the accepted view of the structure of the plasma membrane, comprising a phospholipid bilayer with proteins embedded but free to move about
food chain a sequence of organisms within a habitat in which each is the food of the next, starting with a producer, which is photosynthetic
food web interconnected food chains
founder effect genetic differences that develop between an original breeding population and a small isolated interbreeding group of these organisms
fovea point on a retina of greatest acuity of vision
free energy part of the potential chemical energy in molecules that is available to do useful work when the molecules are broken
frequency commonness of an occurrence
fruit forms from the ovary after fertilisation, as the ovules develop into seeds
functional group the chemically active part of a member of a series of organic molecules
fungus heterotrophic, non-motile, multicellular (usually) eukaryotic organism with 'plant' body – a mycelium of hyphae with cell walls of chitin; the fungi constitute a separate kingdom

G

gall bladder sac beside the liver that stores bile, present in some mammals (e.g. humans)
gamete sex cell (e.g. ovum, sperm)
ganglion part of a nervous system, consisting of nerve cell bodies
gaseous exchange exchange of respiratory gases (oxygen, carbon dioxide) between cells/organism and the environment
gastric relating to the stomach
gene a heritable factor that controls a specific characteristic
gene mutation change in the chemical structure (base sequence) of a gene resulting in change in the characteristics of an organism or individual cell
gene pool all the genes (and their alleles) present in a breeding population
gene probe an artificially prepared sequence of DNA made radioactive with ^{14}C, coding for a particular amino acid residue sequence
gene therapy various mechanisms by which corrected copies of genes are introduced into a patient with a genetic disease
generator potential localised depolarisation of a membrane of a sensory cell
genetic code the order of bases in DNA (of a chromosome) that determines the sequence of amino acids in a protein
genetic counselling genetic advice to potential parents on the risks of having children with an inherited disease
genetic engineering change to the genetic constitution of individuals or populations by artificial selection
genome the genetic complement (genes) of an organism or of an individual cell – the whole of the genetic information of an organism
genotype the genetic constitution of an organism – the alleles of an organism
genus a group of similar and closely related species
germination the resumption of growth by an embryonic plant in seed or fruit, at the expense of stored food
gland cells or tissues adapted for secretion
global warming the hypothesis that the world climate is warming due to rising levels of atmospheric carbon dioxide, a greenhouse gas
glomerulus network of capillaries which are surrounded by the renal capsule
glycocalyx long carbohydrate molecules attached to membrane proteins and membrane lipids
glycogen a much-branched polymer of glucose, the storage carbohydrate of many animals
glycogenesis the synthesis of glycogen from glucose (the reverse is glycogenolysis)
glycolysis the first stage of tissue respiration in which glucose is broken down to pyruvic acid, without use of oxygen
glycoprotein membrane protein with a glycocalyx attached
glycosidic bond a type of chemical linkage between monosaccharide residues in polysaccharides
goblet cell mucus-secreting cell of an epithelium
Golgi apparatus a stack of flattened membranes in the cytoplasm, the site of synthesis of biochemicals
gonad an organ in which gametes are formed
gonadotrophic hormone follicle-stimulating hormone (FSH) and luteinising hormone (LH), secreted by the anterior pituitary, which stimulate gonad function
granum (plural, **grana**) stacked disks of membranes found within the chloroplast, containing the photosynthetic pigments, and the site of the light-dependent reaction of photosynthesis
grey matter regions of the brain and spinal cord consisting largely of nerve cell bodies
growth more or less irreversible increase in size and amount of dry matter
gut the alimentary canal

H

habitat the locality or surroundings in which an organism normally lives or the location of a living organism
haemoglobin a conjugated protein, found in red cells, effective at carrying oxygen from regions of high partial pressure (e.g. lungs) to regions of low partial pressure of oxygen (e.g. respiring tissues)
half-life the time taken for the ionising radiation emitted by a radioactive isotope to fall to half maximum
hallucinogen a drug capable of causing hallucinations
halophyte a plant adapted to survive at abnormally high salt levels (e.g. seashore or salt marsh plant)
haploid (cells) cells having one set of chromosomes, the basic set
heart rate number of contractions of the heart per minute
hepatic associated with the liver
herb layer layer of herbaceous plants (mainly perennials) growing in woodland
herbaceous non-woody
herbicide pesticide toxic to plants
herbivore an animal that feeds (holozoically) exclusively on plants
hermaphrodite organism with both male and female reproductive systems
heterotroph an organism incapable of synthesising its own elaborated nutrients
heterozygous having two different alleles of a gene
hexose a monosaccharide containing six carbon atoms (e.g. glucose, fructose)
hibernation passing the unfavourable season in a resting state of sleep
histology the study of the structure of tissues
histone basic proteins (rich in the amino acids arginine and lysine) that form the scaffolding of chromosomes
holozoic ingesting complex food material and digesting it
homeostasis maintenance of a constant internal environment
homeotherm organism that maintains a constant body temperature

homologous chromosomes chromosomes in a diploid cell which contain the same sequence of genes, but are derived from different parents

homologous structures similar due to common ancestry

homozygous having two identical alleles of a gene

hormone a substance, formed by an endocrine gland and transported in the blood all over the body, but triggering a specific physiological response in one type of organ or tissue

host an organism in or on which a parasite spends all or part of its life cycle

humus complex organic matter, the end-product of the breakdown of the remains of plants and animals, which covers the mineral particles of soil

hybrid an individual produced from a cross between two genetically unlike parents

hybridoma an artificially produced hybrid cell culture, used to produce monoclonal antibodies

hydrocarbon chain a linear arrangement of carbon atoms combined together and with hydrogen atoms, forming a hydrophobic tail to many large organic molecules

hydrogen bond a weak bond caused by electrostatic attraction between a positively charged part of one molecule and a negatively charged part of another

hydrolysis a reaction in which hydrogen and hydroxide ions from water are added to a large molecule causing it to split into smaller molecules

hydrophilic water loving

hydrophobic water hating

hydrophyte an aquatic plant

hydrosere a plant succession that originated from open water

hydrostatic pressure mechanical pressure exerted on or by liquid (e.g. water) also known as pressure potential

hyperglycaemia excess glucose in the blood

hypertonic solution a more concentrated solution (one with a less negative water potential) than the cell solution

hypha the tubular filament 'plant' body of a fungus, which in certain species is divided by cross walls into either multicellular or unicellular compartments

hypoglycaemia very low levels of blood glucose

hypothalamus part of floor of the rear of the forebrain, a control centre for the autonomic nervous system, and source of releasing factors for pituitary hormones

hypothesis a tentative (and testable) explanation of an observed phenomenon or event

hypotonic solution a less concentrated solution (one with a more negative water potential) than the cell solution

I

identify find an answer from a number of possibilities

immunisation (e.g. inoculation/vaccination) the injection of a specific antigen, derived from a pathogen, to confer immunity against a disease

immunity resistance to the onset of a disease after infection by the causative agent

- **active immunity** immunity due to the production of antibodies by the organism itself after the body's defence mechanisms have been stimulated by antigens
- **passive immunity** immunity due to the acquisition of antibodies from another organism in which active immunity has been stimulated, including via the placenta, colostrum, or by injection of antibodies

immunoglobin proteins synthesised by the B lymphocytes of the immune system

immunology study of the immune system

immunosuppressant a substance causing temporary suppression of the immune response

implantation embedding of the blastocyst (developed from the fertilised ovum) in the uterus wall

impulse see *action potential*

imprinting process occurring soon after birth, causing young birds follow their mother

in situ in the original place (in the body or organism)

in vitro biological processes occurring in cell extracts (literally 'in glass')

in vivo biological process occurring in a living organism (literally 'in life')

inbreeding when gametes of closely related individuals fuse leading to progeny that is homozygous for some or many alleles

incubation period period between infection by a causative agent and the appearance of the symptoms of a disease

incus tiny, anvil-shaped bone, the middle ossicle of the middle ear in mammals

industrial melanism increasing proportion of a darkened (melanic) form of an organism, in place of the light-coloured form, associated with industrial pollution by soot

infectious disease disease capable of being transmitted from one organism to another

inhibitor (enzyme) a substance which slows or blocks enzyme action (a competitive inhibitor binds to the active site; a non-competitive inhibitor binds to another part of the enzyme)

inhibitory synapse synapse at which arrival of an impulse blocks forward transmissions of impulses in the post-synaptic membrane

innate behaviour behaviour that does not need to be learned

innervation nerve supply

inspiratory capacity amount of air that can be drawn into the lungs

intelligence the ability to learn by reasoning and to solve problems not yet experienced

interferon proteins formed by vertebrate cells in response to virus infections

intermediates metabolites formed as components of a metabolic pathway

interphase the period between nuclear divisions when the nucleus controls and directs the activity of the cell
interspecific competition competition between organisms of different species
intestine the gut
intracellular enzymes enzymes operating inside the cell
intraspecific competition competition between organisms of the same species
intron a non-coding nucleotide sequence of the DNA of chromosomes, present in eukaryotic chromosomes
invagination the intucking of a surface or wall
ion charged particle formed by the transfer of electron(s) from one atom to another
ionic bonding strong electrostatic attraction between oppositely charged ions
iris circular disc of tissue, in front of the lens of the eye, containing circular and radial muscles
irreversible inhibition inhibition by inhibitors that bind tightly and permanently to an enzyme, destroying its catalytic properties
islets of Langerhans groups of endocrine cells scattered through the pancreas
isomers chemical compounds of the same chemical formula but different structural formulae
isotonic being of the same osmotic concentration and therefore of the same water potential
isotopes different forms of an element, chemically identical but with slightly different physical properties, based on differences in atomic mass (due to different numbers of neutrons in the nucleus)

J

joule the SI unit of energy

K

keratin a fibrous protein found in horn, hair, nails, and in the upper layer of skin
kinesis random movements maintained by motile organisms until more favourable conditions are reached
kinetic energy energy in movement
kingdom the largest and most inclusive group in taxonomy
Krebs cycle part of tissue respiration

L

label add labels to a diagram
lactation secretion of milk in mammary glands
leaching washing out of soluble ions and nutrients by water drainage through soil
learned behaviour in animals, behaviour that is consistently modified as a result of experiences
leucocyte white blood cell
lichens permanent, mutualistic associations between certain fungi and algae, forming organisms found encrusting walls, tree trunks and rocks
ligament strong fibrous cord or capsule of slightly elastic fibres, connecting movable bones
light-independent step part of photosynthesis occurring in the stroma of the chloroplasts and using the products of the light-dependent step to reduce carbon dioxide to carbohydrate
light-dependent step part of photosynthesis occurring in grana of the chloroplasts, in which water is split and ATP and $NADPH_2$ are regenerated
lignin complex chemical impregnating the cellulose of the walls of xylem vessels, fibres and tracheids, imparting great strength and rigidity
lipid diverse group of organic chemicals essential to living things, insoluble in water but soluble in organic solvents such as ether and alcohol (e.g. lipid of the plasma membrane)
linkage group the genes carried on any one chromosome
lipoprotein a complex of lipid and protein of various types which are classified according to density (e.g. LDL, HDL)
list give a sequence of names or other brief answers with no elaboration
liver lobule polygonal block of liver cells, a functional unit within the liver structure
locus the particular position on homologous chromosomes of a gene
loop of Henle loop of mammalian kidney tubule, passing from cortex to medulla and back, important in the process of concentration of urine
lumen internal space of a tube (e.g. gut, artery, etc.) or sac-shaped structure
lymph fluid derived from plasma of blood, bathing all tissue spaces and draining back into the lymphatic system
lymph node tiny glands in the lymphatic system, part of the body's defences against disease
lymphatic system network of fine capillaries throughout the body of vertebrates, which drain lymph and return it to the blood circulation
lymphocyte type of white blood cell
lysis breakdown, typically of cells
lysosome membrane-bound vesicles, common in the cytoplasm, containing digestive enzymes

M

macromolecule very large organic molecule – rmm 10 000+ (e.g. protein, nucleic acid or polysaccharide)
macronutrients ions required in relatively large amounts by organisms
Malpighian body glomerulus and renal capsule of mammalian nephron
mandibles the lower jaw of vertebrates; in arthropods paired, biting mouthparts
matrix ground substance of connective tissue, and the innermost part of a mitochondrion
measure find a value for a quantity
mechanoreceptors a sensory receptor sensitive to mechanical stimulus
meiosis nuclear division with daughter cells containing half the number of chromosomes of the parent cell
melanic pigmented
menstrual cycle monthly cycle of ovulation and menstruation in human females
meristem plant tissue capable of giving rise to new cells and tissues

mesentery connective tissue holding body organs (e.g. gut) in position
mesophyll parenchyma cells containing chloroplasts
mesosome an invagination of the plasma membrane of a bacterium
metabolic pathway sequence of enzyme-catalysed biochemical reactions in cells and tissues
metabolic water water released within the body by oxidation, typically of dietary lipids
metabolism integrated network of all the biochemical reactions of life
metabolite a chemical substance involved in metabolism
metaphase stage in nuclear division (mitosis and meiosis) in which chromosomes become arranged at the equator of the spindle
microhabitat the environment immediately surrounding an organism, particularly applied to tiny organisms
micronutrient ions required in relatively small (trace) amounts by organisms
microtubule tiny, hollow protein tube in cytoplasm (e.g. a component of the spindle)
microvillus one of many tiny infoldings of the plasma membrane, making up a brush border
middle lamella a layer of pectins between the walls of adjacent cells
mitochondrion (plural, **mitochondria**) organelle in eukaryotic cells, site of Krebs cycle and the electron-transport pathway
mitosis nuclear division in which the daughter nuclei have the same number of chromosomes as the parent cell
mitral valve left atrio-ventricular valve
mode the most frequently occurring value in a distribution
monoclonal antibody antibody produced by a single clone of B lymphocytes; it consists of a population of identical antibody molecules
monocotyledon class of angiosperms having an embryo with a single cotyledon
monocyte large phagocytic white blood cell
monohybrid cross a cross (breeding experiment) involving one pair of contrasting characters exhibited by homozygous parents
monosaccharide simple carbohydrate (all are reducing sugars)
morphology form and structure of an organism
motile capable of moving about
motor area area of the brain where muscular activity is coordinated
motor end plate the point of termination of an axon in a voluntary muscle fibre
motor neurone nerve cell that carries impulses away from the central nervous system to an effector (e.g. muscle, gland)
mRNA single-stranded ribonucleic acid formed by the process of transcription of the genetic code in the nucleus, that then moves to ribosomes in the cytoplasm
mucilage mixture of various polysaccharides that become slippery when wet
mucosa the inner lining of the gut
mucus a watery solution of glycoprotein with protective and lubrication functions
muscle spindle sensory receptor in muscle, responding to stretch stimuli
mutagen an agent that causes mutation
mutant organism with altered genetic material (abruptly altered by a mutation)
mutation a change in the amount or the chemical structure (i.e. base sequence) of DNA of a chromosome
mutualism a case of symbiosis in which both organisms benefit from the association
mycelium a mass or network of hyphae
mycology the study of fungi
mycorrhiza a mutualistic association between plant roots and fungi, with the mycelium restricted to the exterior of the root and its cells (ectotrophic), or involving a closer association between hyphae and root cell contents (endotrophic)
myelin sheath an insulating sheath of axons of nerve fibres, formed by the wrapping around of Schwann cells
myelinated nerve fibre nerve fibre insulated by a lipid sheath formed from membranes of Schwann cells
myofibril contractile protein filament from which muscle is composed
myogenic originating in heart muscle cells themselves, as in generation of the basic heart beat

N

natural classification organisms grouped by as many common features as possible, and therefore likely to reflect evolutionary relationships
nectary group of cells secreting nectar (dilute sugar solution) in a flower
nematocyst stinging cell of cnidarians (coelenterates) (e.g. *Hydra*)
Neolithic revolution the period of human development involving the first establishment of settled agriculture practices, and including the breeding and cultivation of crop plants and herd animals
nephron the functional unit of a vertebrate kidney
nerve bundle of many nerve fibres (axons), connecting the central nervous system with parts of the body
nerve cord in non-vertebrates, a bundle of nerve fibres and/or nerve ganglia running along the length of the body
nervous system organised system of neurones which generate and conduct impulses
autonomic nervous system (ANS) the involuntary nervous system
central nervous system (CNS) in vertebrates, the brain and spinal cord
parasympathetic nervous system part of the involuntary nervous system, antagonistic in effect to the sympathetic nervous system
peripheral nervous system (PNS) in vertebrates, neurones that convey sensory information to the CNS, and neurones that convey impulses to muscles and glands (effector organs)

sympathetic nervous system part of the involuntary nervous system, antagonistic in effect to the parasympathetic nervous system
neurone nerve cell
neurotransmitter substance chemical released at the pre-synaptic membrane of an axon, on arrival of an action potential, which transmits the action potential across the synapse
neutrophil a type of white blood cell
niche both the habitat an organism occupies and the mode of nutrition employed
node of Ranvier junction in the myelin sheaths around a myelinated nerve fibre
noradrenaline neurotransmitter substance in the sympathetic nervous system
nuclear division first step in the division of a cell, when the contents of the nucleus are subdivided by mitosis or meiosis
nuclear membrane double membrane surrounding the eukaryotic nucleus
nuclear pores organised gaps in the nuclear membrane, exit points for mRNA
nucleic acid polynucleotide chain of one of two types, deoxyribonucleic acid (DNA) or ribonucleic acid (RNA)
nucleus largest organelle of eukaryotic cells; controls and directs the activity of the cell
nucleolus compact region of nucleus where RNA is synthesised
nucleoside organic base (adenine, guanine, cytosine, thymine) combined with a pentose sugar (ribose or deoxyribose)
nucleotide phosphate ester of a nucleoside – an organic base combined with pentose sugar and phosphate (P_i)
nutrient a chemical substance found in foods that is used in the human body – any substance used or required by an organism as food
nutrition the process by which an organism acquires the matter and energy it requires from its environment

O

obesity condition of being seriously over-weight (BMI of 30+)
oestrous cycle reproductive cycle in female mammal in the absence of pregnancy
oestrous period of fertility (immediately after ovulation) during the oestrous cycle
olfactory relating to the sense of smell
omnivore an animal that eats both plant and animal food
oncogene a cancer-initiating gene
oocyte a female sex cell in the process of a meiotic division to become an ovum
oogamy union of unlike gametes (e.g. large ovum and tiny sperm)
opsonin type of antibody that attacks bacteria and viruses, facilitating their ingestion by phagocytic cells
order a group of related families
organ a part of an organism, consisting of a collection of tissues, having a definite form and structure, and performing one or more specialised functions
organelle a unit of cell substructure
organic compounds of carbon (except carbon dioxide and carbonates)
organism a living thing
osmoreceptor sense cells or organ stimulated by changes in water potential
osmoregulation control of the water balance of the blood, tissue or cytoplasm of a living organism
osmosis diffusion of free water molecules from a region where they are more concentrated (low solute concentration) to a region where they are less concentrated (high solute concentration) across a partially permeable membrane
outline give a brief account or summary
ovarian cycle the monthly changes that occur to ovarian follicles leading to ovulation and the formation of a corpus luteum
ovary female reproductive organ in which the female gametes are formed
ovarian follicle spherical structures found in the mammalian ovary, containing a developing ovum with liquid surrounded by numerous follicle cells, and from which a secondary oocyte is released at ovulation
ovum (plural, **ova**) a female gamete
ovulation shedding of ova from the ovary
ovule in the flowering plant flower, the structure in an ovary which, after fertilisation, grows into the seed
oxygen dissociation curve a graph of % saturation (with oxygen) of haemoglobin against concentration of available oxygen
oxyntic cells cells in the gastric glands secreting hydrochloric acid

P

pacemaker structure that is the origin of the myogenic heart beat, known as the sino-atrial node
Pacinian corpuscles sensory receptors in joints
pancreas an exocrine gland discharging pancreatic juice into the duodenum, combined with endocrine glands (islets of Langerhans)
parasite an organism that lives on or in another organism (its host) for most of its life cycle, deriving nutrients from its host
parenchyma living cells, forming the greater part of cortex and pith in primary plant growth
pathogen an organism or virus that causes a disease
partial pressure the pressure exerted by each component of a gas mixture, proportional to how much of the gas is present in the mixture; the partial pressure of oxygen in air is represented by the symbol pO_2 and is expressed in kilopascals (kPa)
pentadactyl having all four limbs (typically) terminating in five digits
pentose a 5-carbon monosaccharide sugar
peptide a chain of up to 20 amino acid residues, linked by peptide linkages
peptide linkage a covalent bonding of the α amino group of one amino acid to the carboxyl group of another (with the loss of a molecule of water)

perception the mental interpretation of sense data (i.e. occurring in the brain)
pericardium a tough membrane surrounding and containing the heart
peristalsis wave of muscular contractions passing down the gut wall
pesticide a chemical that is used to kill pests
petal modified leaf, often brightly coloured, found in flowers
phagocytic cells cells that ingest bacteria etc. (e.g. certain leucocytes, *Amoeba*)
phenotype the characteristics or appearance (structural, biochemical, etc.) of an organism
pheromone volatile chemical signal released into the air
phloem tissue that conducts elaborated food in plant stems
phosphate (P_i) phosphate ions, as involved in metabolism
phospholipid formed from a triacylglycerol in which one of the fatty acid groups is replaced by an ionised phosphate group
photoautotroph an organism that uses light energy to generate ATP and to produce organic compounds from inorganic substances
photoheterotroph an organism that uses light energy to generate ATP and obtains organic compounds from other organisms
photomorphogenesis effects on plant growth of light
photoperiodism day-length control of flowering in plants
photosynthesis the production of sugar from carbon dioxide and water, occurring in chloroplasts and using light energy, and producing oxygen as a waste product
photophosphorylation the formation of ATP, using light energy (in the light-dependent step of photosynthesis in the grana)
phototropism a tropic response of plants to light
phylogenetic classification a classification based on evolutionary relationships (rather than on appearances)
phylum a group of organisms constructed on a similar general plan, usually thought to be evolutionarily related
physiology the study of the functioning of organisms
phytoplankton photosynthetic plankton, including unicellular algae and cyanobacteria
pinocytosis uptake of a droplet of liquid into a cell involving invagination of the plasma membrane
pituitary gland the master endocrine gland, attached to the underside of the brain
placenta maternal and fetal tissue in the wall of the uterus, site of all exchanges of metabolites and waste products between fetal and maternal blood systems
plant growth substance substances produced by plants in relatively small amounts, that interact to control growth and development
plasma the liquid part of blood
plasma membrane the membrane of lipid and protein that forms the surface of cells (constructed as a fluid mosaic membrane)
plasmid small circular DNA that is independent of the chromosome in bacteria (R plasmids contain genes for resistance to antibiotics)
plasmolysis withdrawal of water from a plant cell by osmosis (incipient plasmolysis is established when about 50% of cells show some shrinkage of cytoplasm away from the walls)
plankton very small, aquatic (marine or fresh water) plants and animals, many of them unicellular, that live at or near the water's surface
plastid an organelle containing pigments (e.g. chloroplast)
platelets tiny cell fragments that lack a nucleus, found in the blood and involved in the blood clotting mechanism
pleural membrane lines lungs and thorax cavity and contains the pleural fluid
polarise the setting up of an electrical potential difference across a membrane
polarised light light in which rays vibrate in one plane only
pollen microspore produced in anthers (and male cones), containing male gamete(s)
pollen tube grows out of a pollen grain attached to a stigma, and down through the style tissue to the embryo sac
polygenic inheritance inheritance of phenotypic characters (such as height, eye colour in humans) that are determined by the collective effects of several different genes
polynucleotide a long, unbranched chain of nucleotides, as found in DNA and RNA
polymer large organic molecules made up of repeating subunits (monomers)
polypeptide a chain of amino acid residues linked by peptide linkages
polyploidy having more than two sets of chromosomes per cell
polysaccharides very high molecular mass carbohydrates, formed by condensation of vast numbers of monosaccharide units, with the removal of water
polysome an aggregation of ribosomes along a molecule of mRNA strand
population a group of organisms of the same species which live in the same area (habitat) at the same time
portal vein vein beginning and ending in a capillary network (rather than at the heart)
post-synaptic neurone neurone 'downstream' of a synapse
potential difference separation of electrical charge within or across a structure (e.g. a membrane)
potential energy stored energy
predator an organism that catches and kills other animals to eat
predict give an expected result
pre-synaptic membrane membrane of the tip of an axon at the point of the synapse
pre-synaptic neurone neurone 'upstream' of a synapse
prey–predator relationship the inter-relationship of population sizes due to predation of one species (the predator) on another (the prey)

proboscis a projection from the head, used for feeding
producer an autotrophic organism
productivity the amount of biomass fixed by producers (photosynthetically)
gross productivity total amount of organic matter produced
net productivity the organic matter of organisms less the amount needed to fuel respiration
prokaryote tiny unicellular organism without a true nucleus; they have a ring of RNA or DNA as a chromosome (e.g. bacteria and cyanobacteria)
prophase first stage in nuclear division, mitotic or meiotic
proprioceptor an internal sensory receptor
prosthetic group a non-protein substance, bound to a protein as part of an enzyme, often forming part of the active site, and able to bind to other proteins
protein a long sequence of amino acid residues combined together (primary structure), and taking up a particular shape (secondary and tertiary structure)
Protoctista kingdom of the eukaryotes consisting of single-celled organisms and multicellular organisms related to them (e.g. protozoa and algae)
protoplast the living contents of a plant cell, contained by the cell wall
protozoan a single-celled animal-like organism, belonging to a sub-kingdom, the Protozoa, of the kingdom Protoctista
pseudopodium a temporary extension of the body of an amoeboid cell, by which movement or feeding may occur
pulmonary circulation the circulation to the lungs in vertebrates having a double circulation
pulmonary ventilation rate breathing rate
pulse a wave of increased pressure in the arterial circulation, generated by the heart beat
pumps proteins in plasma membranes that use energy directly to carry substances across (primary pump) or work indirectly from metabolic energy (secondary pump)
pupil central aperture in the eye through which light enters
pure breeding homozygous, at least for the gene(s) specified
Purkinje fibres fibres of the bundle of His that conduct impulses between the atria and ventricles of the heart
pyloric sphincter circular muscle at the opening of the stomach to the duodenum
pyruvic acid a 3-carbon organic acid, $CH_3.CO.COOH$; product of glycolysis

Q

quadrat a sampling area enclosed within a frame

R

radical a short-lived, intermediate product of a reaction, formed when a covalent bond breaks, with one of the two bonding electrons going to each atom
radioactive dating using the proportions of different isotopes in fossilised biological material to estimate when the original organism was alive
reaction centres protein–pigment complexes in the grana of chloroplasts, sites of the photochemical reactions of photosynthesis
receptor a sense organ
recessive allele an allele that has an effect on the phenotype only when present in the homozygous state
reciprocal cross a cross between the same pair of genotypes in which the sources of the gametes (male and female) are reversed
recombinant a chromosome (or cell or organism) in which the genetic information has been rearranged
recombinant DNA DNA which has been artificial changed, involving joining together genes from different sources, typically from different species
recycling of nutrients the process by which materials from dead organisms are broken down and made available for re-use in the biosphere
Red Data Book an internationally produced record of actions for endangered species
redox reaction reaction in which reduction and oxidation happen simultaneously
reductive division meiosis, in which the chromosome number of a diploid cell is halved
reflex a rapid unconscious response
reflex action a response automatically elicited by a stimulus
reflex arc a functional unit in the nervous system, consisting of sensory receptor, sensory neurone, (possibly relay neurones), motor neurone and effector (e.g. muscle or gland)
refractory period the period after excitation of a neurone, when a repetition of the stimulus fails to induce the same response, divided into periods known as absolute and relative
relative atomic mass the ratio of the mass of an atom of an element to the mass of a carbon atom
renal capsule the cup-shaped closed end of a nephron which, with the glomerulus, constitutes a Malpighian body
renewable energy energy that comes from exploiting wave power, wind power, tidal power, solar energy, hydroelectric power or biological sources such as biomass
replication duplication of DNA by making a copy of an existing molecule
semi-conservative replication each strand of an existing DNA double helix acts as the template for the synthesis of a new strand
reproduction formation of new individual by sexual or asexual means
residual volume volume of air remaining in the lungs after maximum expiration
respiration the cellular process by which sugars and other substances are broken down to release chemical energy for other cellular processes
respiratory centre region of the medulla of the brain concerned with the involuntary control of breathing

respiratory pigment substance such as haemoglobin, which associates with oxygen

respiratory quotient ratio of the volume of carbon dioxide produced to the oxygen used in respiration

respiratory surface a surface adapted for gaseous exchange

respirometer apparatus for the measurement of respiratory gaseous exchange

response the outcome when a stimulus is detected by a receptor

resting potential the potential difference across the membrane of a neurone when it is not being stimulated (repolarised)

restriction enzymes enzymes, also known as endonucleases, that cut lengths of nucleic acid at specific sequences of bases

retina the light-sensitive layer at the back of the eye

retroviruses viruses which, on arrival in a host cell, have their own RNA copied into DNA which then attaches to the host DNA for a period

ribosome non-membranous organelle, site of protein synthesis

ribonucleic acid (RNA) a form of nucleic acid containing the pentose sugar ribose, found in nucleus and cytoplasm of eukaryotic cells (and commonly the only nucleic acid of prokaryotes), and containing the organic bases adenine, guanine, uracil and cytosine

rod cell one of two types of light-sensitive cell in the retina, responsible for non-colour vision

roughage indigestible matter (such as cellulose fibres) in our diet

ribulose bisphosphate the 5-carbon acceptor molecule for carbon dioxide, in the light-independent step of photosynthesis

S

saliva secretion produced by salivary glands

saltatory conduction impulse conduction 'in jumps', between nodes of Ranvier

saprotroph organism that feeds on dead organic matter (saprotrophic nutrition)

sarcolemma membranous sheath around a muscle fibre

sarcomere a unit of a skeletal (voluntary) muscle fibre, between two Z-discs

sarcoplasm cytoplasm around the myofibril of a muscle fibre

sarcoplasmic reticulum network of membranes around the myofibrils of a muscle fibre

saturated fat fat with a fully hydrogenated carbon backbone (i.e. no double bonds present)

Schwann cell cell which forms the sheath around nerve fibres

sclera the opaque, fibrous coat of the eyeball

secondary sexual characteristic sexual characteristic that develops under the influence of sex hormones (androgens and oestrogens)

secondary succession a plant succession on soil already formed, from which the community had been abruptly removed

secretion material produced and released from glandular cells

sedentary organism living attached to the substratum (e.g. rock or other surface)

seed formed from a fertilised ovule, containing an embryonic plant and food store

segmentation body plan built on a repeating series of similar segments (e.g. as in annelids)

selection differential survivability or reproductive potential of different organisms of a breeding population

self-pollination transfer of pollen from the anther to the stigma of the same plant (normally the same flower)

selfing self-pollination or self-fertilisation

semilunar valve half-moon shaped valves, preventing backflow in a tube (e.g. a vein)

seminiferous tubule elongated tubes in the testes, the site of sperm production

sense organ an organ of cells sensitive to external stimuli

sensory area an area of the cerebral cortex of the brain receiving impulses from the sense organs of the body

sensory neurone nerve cell carrying impulses from a sense organ or receptor to the central nervous system

sensory receptor a cell specialised to respond to stimulation by the production of an action potential (impulse)

sepal the protective outermost parts of a flower, usually green

seral stage/sere stages in a seral succession, the whole succession being known as a sere

sex chromosome a chromosome which determines sex rather than other body (soma) characteristics

sex linkage genes carried on only one of the sex chromosomes and which therefore show a different pattern of inheritance in crosses where the male carries the gene from those where the female carries the gene

sexual reproduction involves the production and fusion of gametes

show give the steps in a calculation or derivation

shrub layer the low-level (below trees) woody perennials growing in a forest or wood, normally most numerous in clearings (e.g. where a full-grown tree has died)

sibling offspring of the same parent

sieve tube a phloem element, accompanied by a companion cell, and having perforated end walls known as sieve plates

simple sugar monosaccharide sugar such as a triose sugar (3C), pentose sugar (5C), or hexose sugar (6C)

single access key contrasting or mutually exclusive characteristics are used to divide the group of organisms into progressively smaller groupings until individual organisms (species) can be identified

sino-atrial node cells in the wall of the right atrium in which the heart beat is initiated, also known as the pacemaker

sinus a cavity or space

sketch represent by means of a graph showing a line plus labelled but unscaled axes and with important features (e.g. intercepts) clearly indicated

solar energy electromagnetic radiation derived from the fusion of hydrogen atoms of the Sun, reaching Earth from space

solve obtain an answer using algebraic and/or numerical methods

somatic cell (soma) body cell – not a cell producing gametes (sex cell)

specialisation adaptation for a particular mode of life or function

speciation the evolution of new species

species a group of individuals of common ancestry that closely resemble each other and that are normally capable of interbreeding to produce fertile offspring

sperms motile male gametes of animals

spermatogonia male germ cells (stem cells) which make up the inner layer of the lining of the seminiferous tubules, and give rise to spermatocytes

spermocyte cell formed in seminiferous tubules of testes; develops into sperm

spindle structure formed from microtubules, associated with the movements of chromosomes in mitosis and meiosis

spiracle hole in the side of an insect (thorax and abdomen) by which the tracheal respiratory system connects with the atmosphere

spiral vessel protoxylem vessel with spirally arranged lignin thickening in lateral walls

spirometer apparatus for measurements of lung capacity and breathing rates

spore a small, usually unicellular reproductive structure from which a new organism arises

standing crop the biomass of a particular area under study

stamen male reproductive organ of the flower, consisting of filament and anther, containing pollen sacs where pollen is formed and released

state give a specific name, value or other brief answer (no supporting argument or calculation is necessary)

steroid organic molecule formed from a complex ring of carbon atoms, of which cholesterol is a typical example

stigma part of the carpel receptive to pollen

stimulus a change in the environment (internal or external) that is detected by a receptor and leads to a response

stoma (plural, **stomata**) pore in the epidermis of a leaf, surrounded by two guard cells

stretch receptor sensory receptor in muscles

stroke volume volume of blood pumped out by the heart per minute

stroma the membranous matrix of the chloroplast, site of the light-independent reaction in photosynthesis

style found in the female part of the flower (carpel), linking stigma to ovary

subthreshold stimulus a stimulus not strong enough to trigger an action potential

substrate a molecule that is the starting point for a biochemical reaction and that forms a complex with a specific enzyme

succession the sequences of different communities developing in a given habitat over a period of time

sugars compounds of a general formula $C_x(H_2O)y$, where x is approximately equal to y, and containing an aldehyde or a ketone group

suggest propose a hypothesis or other possible answer

summation combined effect of many nerve impulses

- **spatial** many impulses arriving from different axons
- **temporal** many impulses arriving via a single axon

suspensory ligament attaches lens to ciliary body in the vertebrate eye

symbiosis literally 'living together'; covering parasitism, commensalism and mutualism

symplast the pathway (e.g. of water) through the living contents of cells

synapse the connection between two nerve cells; functionally a tiny gap, the synaptic cleft, traversed by transmitter substances

synaptic knob the terminal swelling of a pre-synaptic neurone

synergism acting together and producing a larger effect than when acting separately

synovial fluid secreted by the synovial membrane at joints, having lubricating role

systematics the study of the diversity of living things

systemic circulation the blood circulation to the body (not the pulmonary circulation)

systemic pesticide pesticide that is absorbed and carried throughout the body

systole contraction phases in the cardiac cycle

T

target organ organ on which a hormone acts (although broadcast to all organs)

taste bud sense organ found chiefly on the upper surface of the tongue

taxis response by a motile organism (or gamete) where the direction of the response is determined by the direction of the stimulus

taxon a classificatory grouping

taxonomy the science of classification

telophase a phase in nuclear division, when the daughter nuclei form

template (DNA) the DNA of the chromosome, copied to make mRNA

tendon fibrous connective tissue connecting a muscle to bone

terminal bud bud at the apex of the stem

test cross testing a suspected heterozygote by crossing it with a known homozygous recessive

testa seed coat

testis male reproductive gland, producing sperms

thermogenesis generation of heat by metabolism

testosterone a steroid hormone, the main sex hormone of male mammals

thorax in mammals, the upper part of the body separated from the abdomen; in insects, the region between head and abdomen

threshold of simulation the level of stimulation required to trigger an action potential (impulse)
thrombosis blood clot formation, leading to blockage of a blood vessel
thylakoid membrane system of chloroplast
thyroid gland an endocrine gland found in the neck of vertebrates, site of production of thyroxine and other hormones influencing the rate of metabolism
tidal volume volume of air normally exchanged in breathing
tight junction point where plasma membranes of adjacent cells are sealed together
tissue collection of cells of similar structure and function
tissue fluid the liquid bathing cells, formed from blood minus cells and plasma proteins
tissue respiration biochemical steps by which energy is released from sugars
tonoplast membrane around the plant cell vacuole
total lung capacity volume of air in the lungs after maximum inhalation
toxic poisonous
toxin poison
toxoid inactivated poison
trachea windpipe
tracheal system system of tubes by which air is passed to tissues in insects
tracheole branch of the trachea
trait a tendency or characteristic
transcription when the DNA sequence of bases is converted into mRNA
transect arbitrary line through a habitat, selected to sample the community
transfer RNA (tRNA) short lengths of specific RNA that combine with specific amino acids prior to protein synthesis
translation the information of mRNA is decoded into protein (amino acid sequence)
translocation transport of elaborated food via the phloem
transmitter substances substances released into the synaptic cleft on arrival of an impulse at the pre-synaptic membrane to conduct the signal across the synapse
transpiration loss of water vapour from the aerial parts of plants (leaves and stem)
tricarboxylic acid (TCA) cycle the stage in tissue respiration in which pyruvate is broken down to carbon dioxide, and hydrogen is removed for subsequent oxidation
tricuspid valve right atrio-ventricular valve
triglyceride fatty acid ester of the 3-carbon alcohol, glycerol – forms into globules because of its hydrophobic properties
triose a 3-carbon monosaccharide
tripeptide a peptide of three amino acid residues
trophic level a level in a food chain defined by the method of obtaining food and in which all organisms are the same number of energy transfers away from the original source of the energy (photosynthesis)
tropism a growth response of plants in which the direction of growth is determined by the direction of the stimulus
tumour abnormal proliferation of cells, either benign (if self-limiting) or malignant (if invasive)
turgid having high internal pressure

U

ultrafiltration occurs through the tiny pores in the capillaries of the glomerulus
ultrastructure fine structure of cells, determined by electron microscopy
unisexual of one or other sex
unsaturated fat lipid with double bond(s) in the hydrocarbon chain
urea NH_2CONH_2, formed from amino groups deaminated from excess amino acid
ureter tube from kidney to bladder
urethra tube from bladder to exterior
uterine cycle cycle of changes to the wall of the uterus (approximately 28 days)
uric acid an insoluble purine, formed from the breakdown of nucleic acids and proteins
urine an excretory fluid produced by the kidneys, consisting largely of a dilute solution of urea
uterus the organ in which the embryo develops in female mammals

V

VO_2 the amount of oxygen being used in the body ($cm^3 kg^{-1} min^{-1}$); with increasingly vigorous exercise, VO_2 will increase, initially
VO_2max the maximal oxygen uptake by the body ($cm^3 kg^{-1} min^{-1}$) – even if the maximum physical effort is maintained, a situation is reached where further increase is impossible
vaccination conferring immunity from a disease by injecting an antigen (of attenuated microorganisms or inactivated component) so that the body acquires antibodies prior to potential infection
vascular bundle strands of xylem and phloem (often with fibres) separated by cambium; the site of water and elaborated food movements up and down the stem
vacuole fluid-filled space in the cytoplasm, especially large and permanent in plant cells
vagus nerve 10th cranial nerve; supplies many internal organs, including the heart
variety a taxonomic group below the species level
vasa recta capillary loop supplying the loop of Henle
vascular tissue xylem and phloem of plants
vasoconstriction constriction of blood supply to capillaries (of skin)
vasodilation dilation of blood supply to capillaries (of skin)
vector an organism that transmits a disease-causing organism, or a device for transferring genes during genetic engineering
venous return volume of blood returning to the heart via the veins per minute
vein vessel that returns blood to the heart
ventilation rate number of inhalations or exhalations per minute

ventral the underside
ventricle chamber, either of the centre of the brain, or of the heart
venule branch of a vein
vertebrate animal with a vertebral column
vesicle membrane-bound sac
vestibular apparatus the semicircular canals of the inner ear, concerned with balance
vestibular canal upper compartment of the cochlea
vestigial small, imperfectly developed structure
virus minute, intracellular parasite, formed of protein and nucleic acid
vital capacity the total possible change in lung volume – the maximum volume of air that can be exhaled after a maximum inhalation
vitalism theory early idea that organic compounds could only be produced in living cells
vitreous humour clear jelly of inner eye

W

water potential the tendency of water molecules to move
water table level of ground water in the Earth
wax complex form of lipid
weathering breakdown of rock
white matter nerve fibres wrapped in their myelin sheaths

X

xeromorphic modified to withstand drought
xerophyte plant showing modifications to withstand drought
xerosere succession of plants starting from dry terrain
xylem water-conducting vessels of plants

Y

yolk food stores of egg cells, rich in proteins and lipids
yolk sac membranous sac with numerous blood vessels, developed by vertebrate embryos around the yolk (e.g. in birds and reptiles) or as a component of the placenta (in mammals)

Z

zonation naturally occurring distribution of organisms in zones
zygote product of the fusion of gametes
zymogenic cells cells of gastric glands, secreting pepsinogen

Acknowledgements

The publishers have made every effort to contact copyright holders. If any have been inadvertently overlooked, they will be pleased to make the necessary arrangements at the earliest opportunity.

Photo credits

Thanks are due to the following for permission to reproduce copyright photographs.

Cover © David Aubrey/Science Photo Library; **p.6** Fig. 1.4 *l* Science Photo Library/J.C.Revy; **p.6** Fig. 1.4 *r* Gene Cox; **p.11** Fig. 1.8 Gene Cox; **p.12** Fig. 1.10 Science Photo Library/Andrew Syred; **p.13** Fig. 1.11 *l* Biophoto Associates; **p.13** Fig. 1.11 *r* from S.W. Hurry *The Microstructure of Cells*, John Murray, 1965, courtesy of Mr A.D. Greenwood; **p.14** Fig. 1.12 Science Photo Library/J.C.Revy; **p.15** Fig. 1.13 Dr Kevin S. Mackenzie, School of Medical Science, Aberdeen University; **p.16** Fig. 1.14 *t* Science Photo Library/Biophoto Associates; **p.16** Fig. 1.15 *b* Science Photo Library/ Kwangshin Kim; **p.19** Fig. 1.17 Science Photo Library/Barry Dowsett; **p.22** Fig. 1.21 *l* Science Photo Library/NIBSC; **p.22** Fig. 1.21 *r* Science Photo Library/Dr D. Fawcett; **p.32** Fig. 1.33 all Biophoto Associates; **p.34** Fig. 1.35 Gene Cox; **p.41** Fig. 2.3 Biophoto Associates; **p.41** Fig. 2.4 Science Photo Library/C.Nuridsany & M.Perennou; **p.43** Fig. 2.6 NHPA/B Jones & M Shimlock; **p.47** Fig. 2.9 Biophoto Associates; **p.50** Fig. 2.12 Gene Cox; **p.66** Fig. 2.28 *t* Science Photo Library/A. Barrington Brown; *bl* Science Photo Library; *br* Science Photo Library; **p.77** Fig. 3.2 *l* Andrew Lambert; *r* Getty Images/Mike Hewitt; **p.92** Fig. 4.1 *l* Biophoto Associates; *r* Biophoto Associates; **p.99** Fig. 4.8 *l* Biophoto Associates; *r* Science Photo Library/Hattie Young; **p.100** Fig. 4.10 Photolibrary/Scott Camazine; **p.103** Fig. 4.13 Science Photo Library; **p.110** Fig. 4.20 Science Photo Library/Dr P. Marazzi; **p.111** Fig. 4.21 Science Photo Library/Chris Bjornberg; **p.114** Fig. 4.24 Science Photo Library/Adam Hart-Davis; **p.123** Fig. 5.7 Science Photo Library/Will & Dent McIntyre; **p.125** Fig. 5.8 Science Photo Library/C.C.Studio; **p.126** Fig. 5.9 Frank Lane Picture Agency/N.Cattlin; **p.130** Fig. 5.11 Science Photo Library/David Parker; **p.134** Fig. 5.13 Roslin Institute, Edinburgh; **p.152** Fig. 6.15 *l* Still Pictures/N.Dickinson; **p.153** Fig. 6.16 *t* NHPA/Kevin Schafer; *b* NHPA/B&C Alexander; **p.157** Fig. 6.19 Visuals Unlimited/Ken Lucas; **p.158** Fig. 6.21 *tl* NHPA/Andy Rouse; *tr* NHPA/Henry Ausloos; *bl* Getty Images/Martin Ruegner; *br* Alamy/Kathy Wright; **p.183** Fig. 7.5 Gene Cox; **p.186** Fig. 7.8 Gene Cox; **p.188** Fig. 7.9 Gene Cox; **p.199** Fig. 7.19 Science Photo Library/Eye of Science; **p.206** Fig. 7.27 Alamy/David R. Frazier Photolibrary, Inc; **p.208** Fig. 7.29 Gene Cox; **p.215** Fig. 7.35 Science Photo Library/Prof S. Cinti; **p.218** Fig. 7.38 *tl* Natural Visions/Heather Angel; *tc* NHPA/Laurie Campbell; *tr* Science Photo Library/Tomas Friedman; *bl* Natural Visions/Heather Angel; *br* NHPA/Stephen Dalton; **p.219** Fig. 7.39 Nickel-Electro Ltd.; **p.222** Fig. 7.44 Gene Cox; **p.224** Fig. 7.47 Science Photo Library/Saturn Stills; **p.236** Fig. 8.2 Biophoto Associates; **p.237** Fig. 8.3 Science Photo Library/Biozentrum, University of Basel; **p.257** Fig. 8.22 Wellcome Photo Library; **p.258** Fig. 8.23 Science Photo Library/Alfred Pasieka; **p.259** Fig. 8.25 Science Photo Library/Phantatomix; **p.272** Fig. 9.4 Science Photo Library/CNRI; **p.277** Fig. 9.9 Science Photo Library/Dr Kari Lounatmaa; **p.296** Fig. 10.2 Gene Cox; **p.297** Fig. 10.3 Gene Cox; **p.309** Fig. 10.16 Gene Cox; **p.310** Fig. 10.17 Dr R.P.C. Johnson, Aberdeen University (retired); **p.312** Fig. 10.19 *r* Gene Cox; *l* Gene Cox; **p.313** Fig. 10.20 both from C.P. Whittingham (1964) *The Chemistry of Plant Processes*, Methuen, London; **p.316** Fig. 10.24 *l* Dr C.J. Clegg; *r* Gene Cox; **p.317** Fig. 10.25 Alamy/Keith Pritchard/fstop2; **p.319** Fig. 10.27 NHPA/Gerry Cambridge; **p.320** Fig. 10.28 *t* Corbis/Bill Ross; *b* Science Photo Library/Jane Sugarman; **p.330** Fig. 11.3 Biophoto Associates; **p.338** Fig. 11.9 NHPA/Stephen Dalton; **p.341** Fig. 11.12 Photolibrary/Rowan Isaac; **p.349** Fig. 12.1 Science Photo Library/CNRI; **p.359** Fig. 12.9 *l* Corbis/Bettmann; *r* Science Photo Library/Jean-Loup Charmet; **p.365** Fig. 12.13 Gene Cox; **p.366** Fig. 12.14 *t* Mediscan/University of Aberdeen; **p.366** Fig. 12.15 *b* Science Photo Library/Biology Media; **p.367** Fig. 12.16 *l* Biophoto Associates; *r* Biophoto Associates; **p.378** Fig. 12.27 *l* Science Photo Library/Astrid & Hans-Frieder Michler; **p.378** Fig. 12.28 *r* Gene Cox; **p.380** Fig. 12.31 Science Photo Library/Jean-Claude Revy-A. Goujeon, ISM; **p.383** Fig. 12.34 Science Photo Library/Petit Format; **p.385** Fig. 12.37 Science Photo Library/Moscoso; **p.400** Fig. 13.8 FLPA/Angela Hampton; **p.401** Fig. 13.9 *r* Science Photo Library; *l* Science Photo Library/Dr P. Marazzi; **p.403** Fig. 13.10 reproduced by permission of Sainsbury's Supermarkets Ltd; **p.407** Fig. 13.13 *bl* Getty Images/Raymond K. Gehman/National Geographic; *tr* FLPA/Nigel Cattlin; *tl* Corbis/Dung Vo Trung/Sygma; *br* Corbis/Owen Franken; **p.408** Fig. 13.14 *b* Corbis/Wolfgang Kaehler; *t* Alamy/Robert Fried; **p.410** Fig. 13.16 Science Photo Library/Prof S. Cinti; **p.414** Fig. 13.19 *l* Corbis; *r* Rex Features/Sabah Arar; **p.418** Fig. 13.21 *tr* Dr C.J. Clegg; *br* NHPA/E.A.Janes; *l* PA Photos/DPA; **p.420** Fig. 13.22 Science Photo Library/BSIP VEM; **p.427** Fig. 14.4 Getty Images/Toru Yamanaka/AFP; **p.434** Fig. 14.7 PA Photos/Steve Mitchell; **p.437** Fig. 14.12 Getty Images/Dr Gladden Willis; **p.439** Fig. 14.14 *tl* Corbis/Gary Hershorn/Reuters; *tr* Getty Images/Farjana K. Godhuly/AFP; **p.439** Fig. 14.15 PA Photos/DPA; **p.458** Fig. 16.3 *b* Corbis/Theo Allofs/zefa; **p.463** Fig. 16.6 Dr C.J. Clegg; **p.464** Fig. 16.7 *r* Corbis/Owen Franken; *l* NHPA/Nick Garbutt; **p.465** Fig. 16.8 *l* Science Photo Library/Tony Wood; *r* GardenWorld Images/D.Gould; **p.466** Fig. 16.9 Getty Images/Nina Leen/Time Life Pictures; **p.467** Fig. 16.10 Prof A. Gray; **p.481** Fig. 16.22 *t* Getty Images/Chris McGrath; *b* Getty Images/Arthur Tilley; **p.485** Fig. 16.27 Science Photo Library/John Reader; **p.492** Fig. 16.33 la Grotte de Niaux, Niaux, Tarascon-sur-Ariège, Pyrenees, France; **p.507** Fig. 17.3 Gene Cox; **p.509** Fig. 17.4 NHPA/Nick Garbutt; **p.512** Fig. 17.6 Gene Cox; **p.513** Fig. 17.7 Gene Cox; **p.521** Fig. 17.12 Dr C.J. Clegg; **p.523** Fig. 17.13 Gene Cox; **p.529** Fig. 17.20 Science Photo Library/Omikron; **p.537** Fig. 17.24 all Prof Robert Turner, Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London; **p.542** Fig. 17.27 *l* FLPA/Fritz Polking; *r* FLPA/Silvestris Fotoservice; **p.543** Fig. 17.28 *l* FLPA/Jurgen & Christine Sohns; *r* Getty Images/Adrian Bailey/Aurora; **p.545** Fig. 17.29 Alamy/Penny Boyd; **p.547** Fig. 17.31 *t* FLPA/Nigel Cattlin; *b* FLPA/Foto Natura Stock; **p.548** Fig. 17.33 *l* FLPA/Frans Lanting/Minden Pictures; *tr* Nature Picture Library/Peter Oxford; *br* FLPA/Peggy Heard; **p.557** Fig. 18.3 *l* Science Photo Library/Dr Kari Lounatmaa; *r* Corbis/Ralph White; **p.559** Fig. 18.6 Gene Cox; **p.565** Fig. 18.15 Gene Cox; **p.578** Fig. 18.24 Corbis/Macduff Everton; **p.579** Fig. 18.25 *t* Anthony Blake Photo Library; *b* Anthony Blake Photo Library/ Gerrit Buntrock; **p.584** Fig. 18.32 Gene Cox; **p.592** Fig. 18.36 Science Photo Library/Matt Meadows/Peter Arnold Inc; **p.595** Fig. 18.39 Rex Features; **p.603** Fig. 19.3 *t* Educationphotos.co.uk/Walmsley; **p.603** Fig. 19.4 *b* Nature Picture Library/Chris Gomersall; **p.608** Fig. 19.9 *b* Alamy/Daniel L. Geiger/SNAP; *t* Alamy/WildPictures; **p.610** Fig. 19.10 *tl* NHPA/Rod Planck; *tr* NHPA/Martin Harvey; *bl* Premaphotos; *br* Dr C.J. Clegg;

p.616 Fig. 19.16 *l* Corbis/Steven G. Smith; *r* Corbis/Eric and David Hosking; **p.621** Fig. 19.20 Dr C.J. Clegg; **p.625** Fig. 19.25 *l* Science Photo Library/John Devries; *r* NHPA/Stephen Dalton; **p.628** Fig. 19.28 Alamy/CuboImages; **p.630** Fig. 19.30 Science Photo Library/Dr G. Feldman/NASA/GSFC; **p.631** Fig. 19.31 Alamy/Natural History Museum, London; **p.647** Fig. 20.7 *t* Gene Cox; *b* Science Photo Library/Innerspace Imaging; **p.648** Fig. 20.8 Dr Kevin S. Mackenzie, School of Medical Science, Aberdeen University; **p.652** Fig. 20.10 Gene Cox; **p.652** Fig. 20.11 Dr Kevin S. Mackenzie, School of Medical Science, Aberdeen University; **p.662** Fig. 20.19 Corbis/Michael Keller; **p.672** Fig. 20.27 Science Photo Library/Dr Jeremy Burgess; **p.673** Fig. 20.28 Getty Images/Tim Graham; **p.686** Fig. 22.1 Gary Seston; **p.687** Fig. 22.2 Gary Seston; **p.689** Fig. 22.3 Gary Seston.

Artwork credits

The publishers would like to thank the following for permission to reproduce copyright material:

p.118 Fig. 5.1 Modified, with permission, from Dr P. Moore (1994) *Recombinant DNA Technology* © The Biological Society; **p.149** Fig. 6.13 *Climate Change 2001: The Scientific Basis* © The Intergovernmental Panel on Climate Change, quoted in F.Press, R.Siever, J.Grotzinger, T.H.Jordan (2004) *Understanding Earth* (4th edition); **p.152** Fig. 6.15 cartoon © Scott Willis, San Jose Mercury News; **p.163** Fig. 6.24 Adapted from A.D.Bradshaw and T.McNeilly (1981) *Evolution and Pollution*, Institute of Biology, *Studies in Biology* No. 130, by permission of Cambridge University Press; **p.220** Fig. 7.41 Graph of human body temperature from J.H.Green (1963) *An Introduction to Human Physiology*, by permission of Oxford University Press; Isotherm diagrams from K.Schmidt-Nielson (1983) *Animal Physiology: Adaptation and Environment* (3rd edition) by permission of Cambridge University Press; **p.251** Fig. 8.5 Adapted from E.J.Wood, C.A.Smith and W.R.Pickering (1997) *Life Chemistry and Molecular Biology* © Portland Press Ltd; **p.393** Fig. 13.2 Images of children adapted from A.F.Walker (1990) *Cambridge Social Biology Topics: Human Nutrition*, by permission of Cambridge University Press; **p.413** Fig. 13.18 Adapted from an article by Helen Carter (1998) copyright Guardian News and Media Limited; **p.426** Fig. 14.3 David Graham, adapted from K.Weston, N.Wiggins-James, G.Thompson and S.Hartigan (2005) *Sport and PE* (3rd edition) Hodder Arnold; **p.430** Table 14.2 Adapted from Colin Clegg (revised Stephen Ingham) (1995) *Exercise Physiology* © Feltham Press; **p.434** Fig. 14.8 David Graham, adapted from K.Weston, N.Wiggins-James, G.Thompson and S.Hartigan (2005) *Sport and PE* (3rd edition) Hodder Arnold; **p.449** Fig. 15.4 Adapted from Figure 12.13 in C.J.Clegg, with D.G.Mackean (2000) *Advanced Biology: Principles and Applications* (2nd edition) John Murray; **p.451** Fig. 15.6 Adapted with permission from a method published on the Science and Plants for Schools (SAPS) website (see www.saps.org.uk); **p.457** Fig. 16.2 Adapted from John Maynard Smith and E.Szathmáry (1999) *The Origins of Life*, by permission of Oxford University Press; **p.488** Fig. 16.30 Map adapted from R.Lewin (1989) *Human Evolution: An Illustrated Introduction* © Blackwell Publishing Ltd.; Skull images reproduced from R.Leakey (1994) *The Origin of Humankind* © Weidenfeld & Nicolson, a division of The Orion Publishing Group, and John Brockman Associates; **p.489** Fig. 16.31 Stuart Bailey, adapted from Paul Arnold 'Human Evolution', *Biological Sciences Review*, Volume 8, Number 4 (1996) © S.E.R.Bailey; **p.497** Table 16.7 M.Ridley (1993) *Evolution* © Blackwell Publishing Ltd; **p.523** Fig. 17.13 and **p.524** Fig. 17.14 and Fig. 17.15 Adapted from J.Cheverton (1994) 'Animals in Action', *Biological Sciences Review*, Volume 7, Number 2, Philip Allan Updates, Philip Allan Publishers; **p.526** Fig.17.17 Adapted from G.G.Simpson, C.S.Pittendrigh, L.H.Tiffany (1957) *Life, An Introduction to Biology*, Routledge; **p.545** Fig.17.29 Adapted from *The Animal Mind* by James L. Gould and Carol Grant Gould. Copyright 1994 by Scientific American Library. Reprinted by permission of Henry Holt and Company, LLC.; **p.546** Fig. 17.30 Adapted from G.G.Simpson, C.S.Pittendrigh, L.H.Tiffany (1957) *Life, An Introduction to Biology*, Routledge & Kegan Paul Ltd, © William S. Beck, Elizabeth Simpson Wurr, Helen S. Vishniac, Joan S. Burns; **p.549** Fig. 17.34 Adapted from J.D.McCarthy (1966), *The Study of Behaviour*, Institute of Biology, *Studies in Biology* No. 3 (Edward Arnold), by permission of Cambridge University Press; **p.599** Fig. 19.1 T.J.King (1989) *Selected Topics in Biology – Ecology* (2nd edition) Thomas Nelson & Sons Ltd; **p.619** Fig. 19.19 Alastair Gray (Ed.) (1985), *World Health and Disease*, reproduced with kind permission of the Open University Press; **p.621** Fig. 19.20 Data supplied by Dr John Cheverton, Head of Biology Department, Oakham School, Rutland, UK and the Upper Sixth Form Biologists; **p.624** Fig. 19.23 Michael Madigan, John Martinko (2006) *Brock Biology of Micro-organisms* (11th edition) © Reprinted by permission of Pearson Education Inc., Upper Saddle River, NJ; **p.625** Fig. 19.24 Adapted from M.Wade, L.Child and N.Adachi (1996) 'Japanese Knotweed – a cultivated coloniser', *Biological Sciences Review*, Volume 8, Number 3, © Biological Records Centre, ITE, Monks Woods; **p.626** Fig. 19.27 Graph from the paper by Dr I. Newton of the Institute of Terrestrial Ecology, *Britain Since Silent Spring*, Institute of Biology Symposium; **p.629** Fig.19.29 David L. Hawksworth and Francis Rose (1976) 'Lichens as Pollution Monitors', *Studies in Biology* No. 66, by permission of Cambridge University Press; **p.676** Fig. 21.1 From C.J.Clegg (1967), 'Studies in tissue metabolism in stored potato tissue', MSc Thesis, University of London.

Thanks are due to the following sources of inspiration and data for illustrative figures:

p.141 Fig. 6.3b Adapted from *Introducing Ecology: Nature at Work*, British Museum (Natural History) London, and Cambridge University Press, 1978; **p.195** Fig. 7.16 Alastair Gray (Ed.) (1985), *World Health and Disease*, Open University Press; **p.239** Fig. 8.5 Adapted from E.J.Wood, C.A.Smith and W.R.Pickering (1997) *Life Chemistry and Molecular Biology*, Portland Press, London; **p.286** Fig. 9.18 Adapted from D.O.Hall and K.K.Rao (1992) *Photosynthesis* (4th edition) Cambridge University Press; **p.313** Fig. 10.20 C.P.Whittingham (1964) *The Chemistry of Plant Processes*, Methuen; **p.483** Fig. 16.25 Matrix adapted from R.Lewin (1984) *Human Evolution*, Blackwell Scientific Publications, Boston; map adapted from R.Leakey (1994) *The Origin of Humankind*, Weidenfeld & Nicolson, a division of The Orion Publishing Group, and John Brockman Associates; **p.501** Fig. 16.39 and **p.502** Fig. 16.40 Adapted from *Man's Place in Evolution*, British Museum (Natural History), 1980, Cambridge University Press; **p.527** Fig. 17.19 Adapted from D.McFarland (1981) *The Oxford Companion to Animal Behaviour*, Oxford University Press; **p.612** Fig. 19.11 Julian Cremona (1988) *A Field Atlas of the Sea Shore*, Cambridge University Press; **p.618** Fig. 19.18 T.King (1980) *Ecology*, Nelson; **p.622** Fig. 19.21 Adapted from Stuart L. Pimm and Clinton Jenkins (2005) 'Sustaining the Variety of Life' © *Scientific American* Inc. 293:3. All rights reserved.

Text credits

p.623 Fig. 19.22 Adapted from an article in *The Times*, 2 February 2006, © Nick Meo.

Examination questions credits

The publishers would like to thank the International Baccalaureate Organization for permission to reproduce its intellectual property.

Index

Note: blue page numbers refer to material on the accompanying CD-ROM.

Option A

13 Human nutrition and health

STARTING POINTS

- Of the vast number of **organic compounds** that make up living things, many fall into one of three groups – the **carbohydrates**, **lipids** and **proteins**.
- By **nutrition**, organisms obtain energy to maintain the functions of life, and matter to **build and maintain their structures**, using carbohydrates, lipids and proteins as bulk **raw materials**.
- Animals (and fungi) take **complex food molecules** and break them down by **digestion**, to form molecules small enough to be **absorbed** into body cells.
- Digestion involves **hydrolytic enzymes**. These are mostly enzymes packaged for discharge from the cells in which they are produced.
- **Health** is a state of physical, mental and social well-being, not merely the absence of disease. Non-communicable diseases are those illnesses which are not caught from other organisms. **Malnutrition** is an example of a non-communicable disease.
- This chapter extends study of aspects of **biological molecules** begun in Chapter 2 (pages 37–75), and of **human nutrition and health** in Chapter 7 (pages 178–234).

There is a division in the living world over how organisms obtain the materials they require. Green plants manufacture all their own nutrients, but most other organisms cannot do this. Instead, they rely on ready-made food in the form of other organisms, dead or alive, and the waste products of organisms. Their nutrition is described as **heterotrophic nutrition**.

One form of heterotrophic nutrition is the **holozoic nutrition of mammals**, in which food is taken into a specialised structure, the **alimentary canal**. The alimentary canal is also referred to as the gut. It is a long muscular tube beginning at the mouth and ending at the anus. Along the gut are several glands, and the whole structure is specialised for the movement and **digestion** of food and for the **absorption** of the useful products of digestion.

Mechanical digestion occurs by the action of the jaws and teeth in the mouth, and through the churning action of the muscular walls as the food is moved along the gut. Chemical digestion is by **enzymes**. The useful products of digestion are then absorbed into the body and **assimilated** into cells and tissues. Undigested matter is eliminated from the body as **faeces**. The regions of the gut in the human are shown in Figure 7.1, and the stages to holozoic nutrition are defined in Table 7.1 (page 180).

In this chapter, we first focus on the components of the **human diet**, their roles, and the effects of their absence from the food eaten. Subsequently, issues arising from the sources and supply of **energy in foods**, of excessive and insufficient energy intake, of the control of our **appetite** for food, and of ethnic differences in diet are all discussed. Finally, **special issues in human nutrition**, including ones of health and disease, together with certain ethical issues of diet, are considered.

Nutrients of the human diet

A1.1–1.7

The diet is the total of food a holozoic organism such as the human must take in to provide an appropriate amount of required nutrients on a regular basis.

A nutrient is a chemical substance found in foods that is used in the human body.

The **nutrients** that food must provide on a regular basis are as follows.

- **Metabolic fuel** is often supplied from carbohydrates and fats (lipids), but this is not necessarily always the case. Metabolic fuel may also be supplied from proteins. Chemical **energy** in all of these fuels is transferred into a required form (such as movement in a contracting muscle) by respiration.

- **Combined nitrogen** (nitrogen combined in compounds, rather than N_2 gas) is required for the building of **proteins**. It is available to the body as proteins which have to be hydrolysed to their constituent amino acids. Dietary protein is the chief source of the 20 amino acids from which the mammal's own proteins are built up.
- **Vitamins** are organic compounds that are required in only tiny amounts. Most function as coenzymes in the body. Typically, vitamins cannot be manufactured in the body, so their absence from the diet tends to have marked effects, known as **deficiency diseases**.
- **Minerals**, including major minerals in the form of ions, like calcium, iron and phosphate, are needed for the construction of body tissues or are combined in metabolites essential for many metabolic processes. Also required are micronutrients such as manganese, which are often co-factors in the functioning of particular enzymes.
- **Water** is essential because 70–90% of the body is water.

Dietary fibre is also a necessary component of food, although it is not a nutrient. Dietary fibre will be discussed later.

A diet that provides the necessary nutrients in correct amounts is often described as a **balanced diet**. A diet deficient in one or more essentials, or which provides the wrong balance of nutrients, will lead to the condition of **malnutrition**. We will return to the issue of malnutrition in humans later in this chapter.

The idea of essential nutrients

Above are listed the nutrients that food must provide on a regular basis. Some of these components are also described as **essential nutrients** – but not all of them.

Essential nutrients are the components of a diet that must be present because they cannot be entirely or adequately produced by the body from other substances taken in as food, in sufficient quantities for survival and good health. The nutrients essential in our diet are:

- **amino acids**, obtained from proteins;
- **fatty acids**, obtained from lipids;
- **minerals**;
- **vitamins**;
- **water**.

Notice that carbohydrates are omitted. This is because energy can be obtained from fatty acids and from proteins. Although carbohydrate is not an essential element of the diet, most diets not only include it, but many people gain the bulk of their essential metabolic energy from it. Carbohydrates are an extremely common source of metabolic fuel.

Issues concerning the supply of **metabolic fuel** are discussed in a later section of this chapter. The first issue to consider is exactly what essential nutrients we must have.

Amino acids obtained from proteins

We have seen that the proteins in our diet are first digested, a process which begins in the stomach and is completed in the small intestine. They are broken down by proteases to their constituent amino acids (page 51), and these are then absorbed into the body. There, they contribute to the pool of amino acids from which new proteins are built.

Organisms need a supply of 20 different amino acids to synthesise all the proteins they require. Plants are able to manufacture all their amino acids, but mammals such as humans are able to make only about half of them, using other amino acids. They can do this *provided their diet includes sufficient protein, in total*. The remaining amino acids must be obtained already assembled. Eight amino acids are **essential amino acids** in human diets. They must be available to the body from the proteins digested, if malnourishment is to be avoided (Figure 13.1).

Although the primary purpose of food protein is to provide the amino acids for the production and maintenance of the body's proteins, we must remember that food protein may also be broken down to form urea and then lead to the provision of energy. In other words, a secondary role of proteins, when present in excess, is as a supplementary energy source.

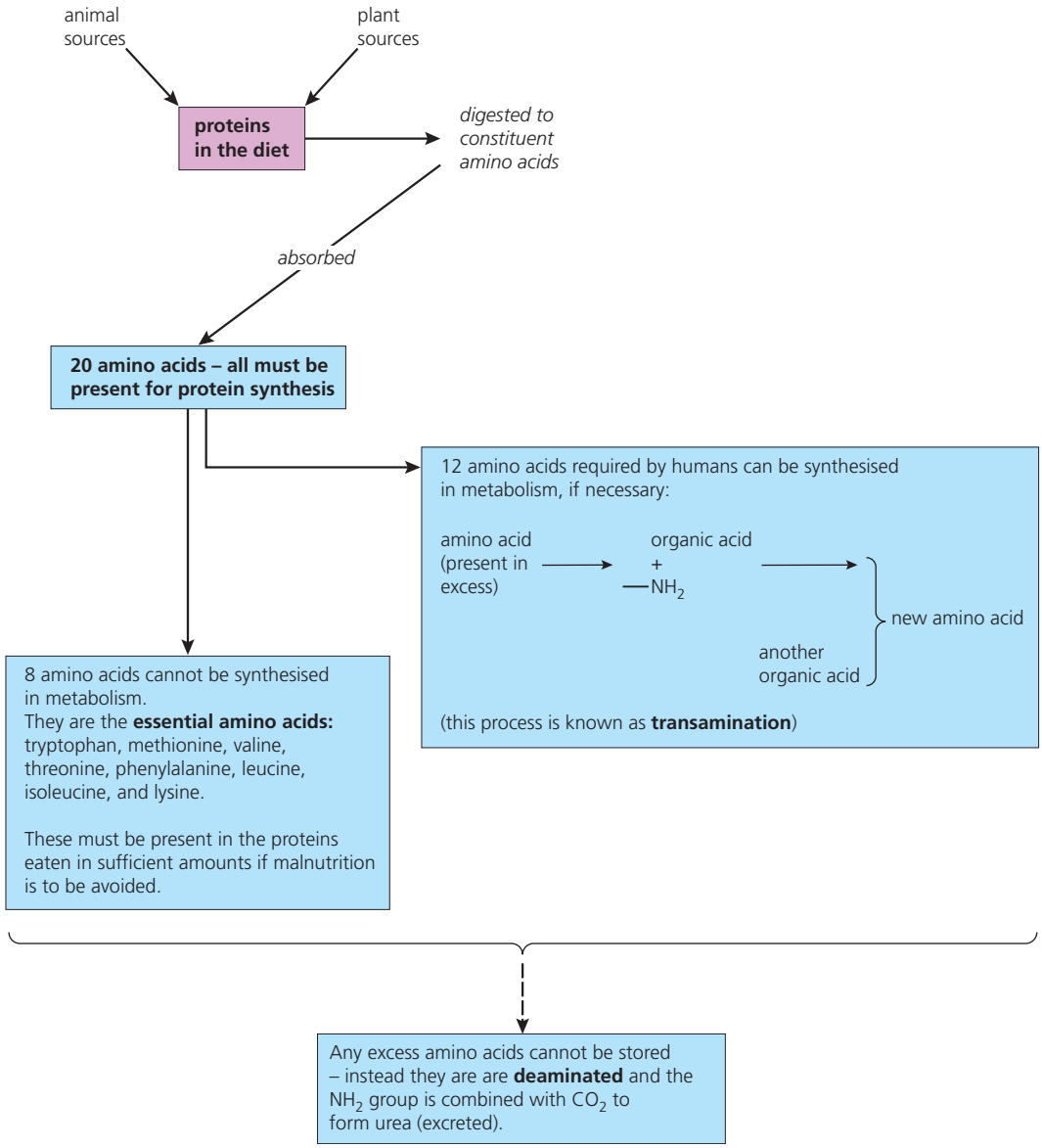


Figure 13.1 The supply of amino acids in human metabolism

1 Deduce the biochemical step that must first occur before proteins and amino acids can be respired.

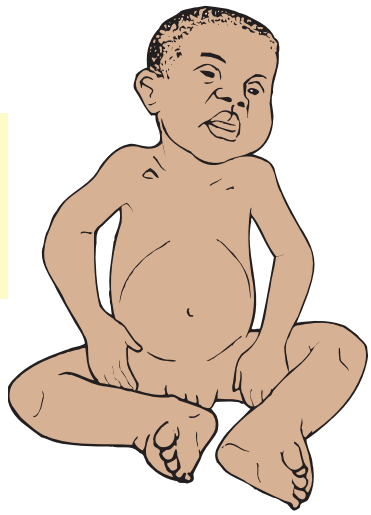
Consequences of protein deficiency malnutrition

Protein deficiency in the diet leads to a shortage of one or more essential amino acids. Surprisingly perhaps, the resulting symptoms are complex and variable. Extreme cases of protein deficiency show symptoms referred to as **kwashiorkor** and **marasmus**, the clinical signs of which are quite different. A child with marasmus has low weight for his age, thin arms and legs, and a wizened appearance. A child with kwashiorkor has swelling of the legs (oedema), sparse hair and a 'flaky paint' appearance (Figure 13.2).

Figure 13.2 Kwashiorkor and marasmus – the clinical signs

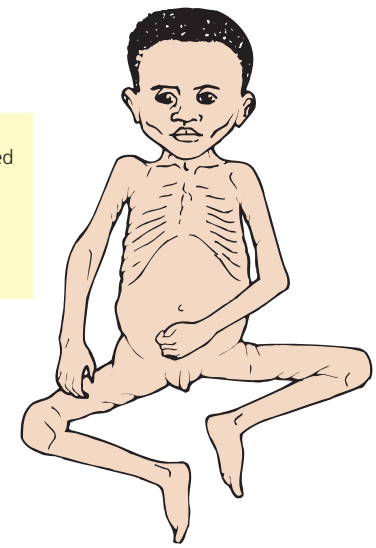
kwashiorkor

- swelling of legs (oedema)
- sparse hair
- 'moon face', with little interest in surroundings
- flaky appearance of skin
- swollen abdomen
- thin muscles, but fat present



marasmus

- normal hair
- 'old man' or wizened appearance
- thin limbs with little muscle or fat
- very underweight body



At one time, kwashiorkor was described as due to **protein deficiency**, while marasmus was thought to be caused by a **lack of energy** in the diet. Remember, when a human body is starved of food generally, reserves of glycogen (a storage carbohydrate) are used up first, and then fat stores around the body are mobilised. With continuing starvation and the absence of any remaining reserves, muscle protein starts to be used as a source of energy, causing, among other things, extremely thin arms and legs.

However, it was then observed that, with two children similar in condition and age and of the same sex, both eating the same poor diet, one might succumb to marasmus and the other to kwashiorkor. Yet both diets contained the same (minimal) amount of protein.

What is clear is that in all cases of protein deficiency symptoms, energy intakes are well below the **recommended daily allowances** (RDAs). As a result, the use of protein for growth and maintenance is subordinated to its use as an energy source – the amino acids obtained by digestion are respired before muscle proteins are mobilised.

Oedema, the presence of excess body fluid in the tissues, is characteristic of kwashiorkor. This is caused by a marked lowering in the level of the blood protein albumin. Albumin has a role in our blood plasma of drawing water from the tissues osmotically, back into the blood.

When the energy intake of the diet is sufficient, most diets supply sufficient protein, whether it comes from (whole grain) rice, whole wheat, or fish and meat sources (more expensive than plant sources, and more concentrated). The exceptions are diets based on the tropical crops cassava (a root crop), plantain (a type of banana) and sweet potato – unfortunately, all these sources are very low in protein.

Phenylketonuria

Phenylketonuria (PKU) is a genetic error of protein metabolism. It is caused by a mutation in a gene coding for a protein that forms the enzyme that converts the amino acid phenylalanine into tyrosine. Phenylalanine is commonly present in the diet in excess of requirements for protein synthesis. In people born with this mutation, phenylalanine obtained in the diet, and not immediately used by the body in synthesis of new proteins, starts to build up in the blood. An excess of phenylalanine in the blood causes many unpleasant and dangerous side-effects. These include vomiting, seizure, growth deficiency, and (eventually) severe mental disability.

Babies need to be screened at birth for this disorder. Where PKU is detected, the symptoms can largely be avoided by restricting the diet to one supplying only the amount of phenylalanine that the body requires in protein synthesis, but unfortunately, the patient may still develop learning difficulties.

2 **Explain** the term 'mutation'.

Fatty acids from lipids

Lipids include fats (in animals) and oils (in plants), and both consist of compounds called **triglycerides** (strictly, triacylglycerols, page 47). These are formed by reactions in which water is removed (**condensation reactions**) between fatty acids (monocarboxylic acids) and an alcohol called **glycerol** (Figure 2.11, page 49). When three fatty acids combine with one glycerol molecule, a triglyceride is formed. Note that the differences between one fat or oil and another are largely due to the different fatty acids combined in each.

In **digestion**, triglycerides are hydrolysed to their constituent fatty acids and glycerol by enzymes called lipases in the small intestine. These components are then absorbed in the villi of the small intestine (page 183), and transported to respiring cells or to the liver, where they are further metabolised or respired. It is the fatty acids which we are particularly concerned with here.

So, a range of different fatty acids are found, combined in triglycerides. Each fatty acid consists of a chain of carbon atoms (each with hydrogen atoms attached) with a carboxyl group at the end. The length of the carbon chain varies between 4 and 24 carbon atoms, but the commonest fatty acids the body uses have 16 or 18 carbon atoms. In fact, the fatty acids present in the diet can be divided into three distinct groups.

1 Saturated fatty acids

Here the carbon atoms are combined together by single bonds, so the carbon chain consists of —CH_2—CH_2—, repeated again and again. Examples include palmitic acid and stearic acid, both of which are major constituents of butter, lard, suet and cocoa butter. Butyric acid is present in small amounts in milk fat and butter, but plays a major part in creating their characteristic flavours.

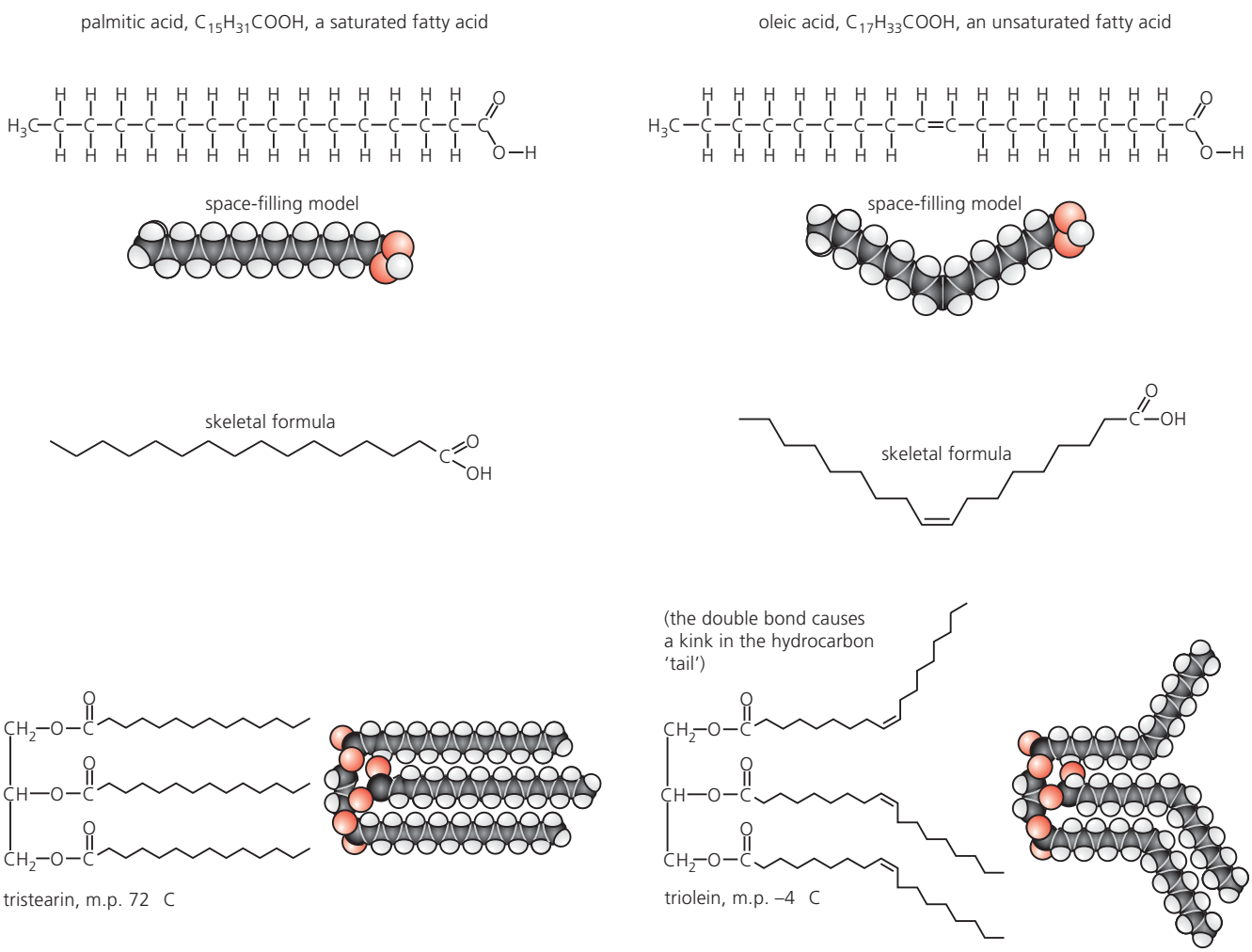
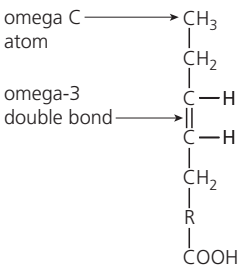


Figure 13.3 Saturated and unsaturated fatty acids

2 **Monounsaturated fatty acids**
Here there is one double bond in the carbon chain (so they have two missing hydrogen atoms when compared with the corresponding saturated fatty acid). Note that carbon atoms not only react with each other to form extended chains, but pairs of carbon atoms may share two electrons to form a double bond. The carbon atoms concerned are described as unsaturated. Oleic acid is an example of an unsaturated fatty acid, occurring in significant quantities in most fats, but in olive oil and rapeseed oil it makes up about 70% of the fatty acids present.

3 **Polyunsaturated fatty acids**
Here there are two or more double bonds in the carbon chain. Linoleic acid, which has two double bonds, occurs in large amounts in vegetable seed oils, such as maize, soya and sunflower seed oils. Linolenic acid, which has three double bonds, occurs in small amounts in vegetable oils. Fats with unsaturated fatty acids melt at a lower temperature than those with saturated fatty acids, because their unsaturated hydrocarbon tails do not pack so closely together (Figure 13.3) in the way that saturated fats do. This difference between saturated and polyunsaturated fats is important in the manufacture of margarine and butter spreads, since these spread easily even when taken straight from the refrigerator. It is also an important difference in terms of human health, as we shall discuss later.

Figure 13.4 An omega-3 fatty acid



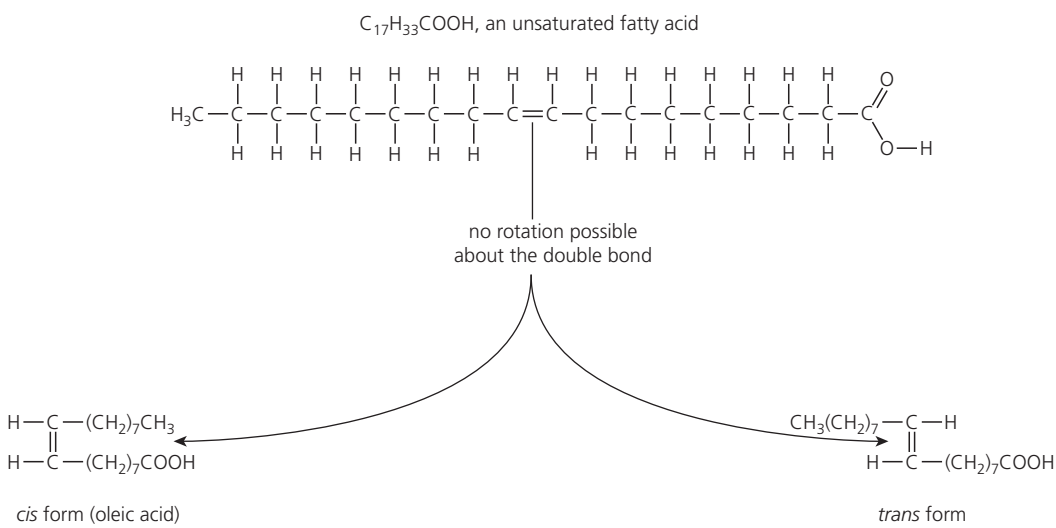
Omega-3 fatty acids

Omega-3 fatty acids are a particular group of naturally occurring polyunsaturated fatty acids. They are chemically special in that they have between three and six double bonds in the hydrocarbon tail, and the first double bond is always positioned between the third and the fourth carbon atom from the opposite (omega) end of the hydrocarbon chain to the carboxyl group, as shown in Figure 13.4. A number of omega-3 fatty acids occur in plant and fish oils, and are thought to be particularly beneficial to health, of which more later.

Cis and *trans* fatty acids

In many organic molecules, rotation of one part of the molecule with respect to another part is possible about a single covalent —C—C— bond. However, when two carbon atoms are joined by a double bond there is no freedom of rotation at this point in the molecule. We can demonstrate the significance of this in a monounsaturated fatty acid of chemical formula $C_{17}H_{33}COOH$, where the double bond occurs in the mid-point of the hydrocarbon chain, between carbon atoms 9 and 10 (Figure 13.5). In one possible form of this molecule, the two parts of the hydrocarbon chain are on the same side of the double bond. This molecule is *cis* oleic acid. The alternative form has the two parts of the hydrocarbon chain on opposite sides. This is the *trans* form of the molecule, *trans* oleic acid.

Figure 13.5 *Cis* and *trans* fatty acids



3 Explain what part of an enzyme can recognise differences in the chemical structure of the substrate molecule.

The significances of this complication to our understanding of fatty acids are two-fold. First, enzymes can recognise the difference between *cis* and *trans* forms of molecules at their active sites (page 52). Enzymes of lipid metabolism generally recognise and can trigger metabolism of the *cis* forms, but not the *trans* forms. Secondly, there are important health implications for human diets, and we will return to this point shortly.

Essential fatty acids

A wide range of different fatty acids are used in the cells, tissues and organs of the mammal's body. They are required in the production of very many different types of molecule. For example, they are used in biosynthesis of many metabolic intermediates, of membrane lipids including very many receptors, of blood lipids, and many hormones. The sex hormones are of lipid origin, as are others that are critically involved in the regulation of metabolism and the operation of the body's physiology (Figure 13.6).

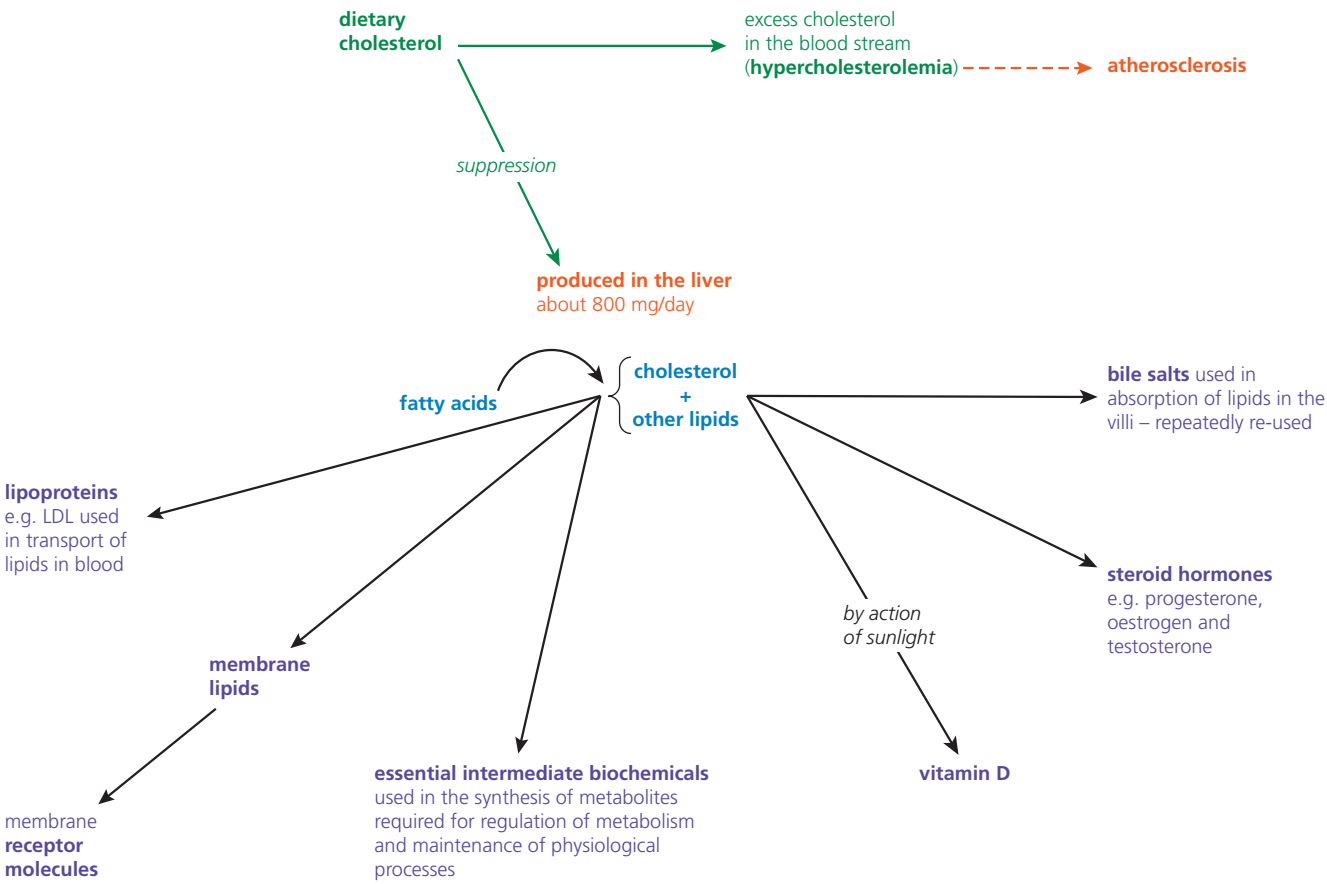


Figure 13.6 Roles of fatty acids in biosynthesis

Some of the polyunsaturated fatty acids involved in all this biosynthesis are **essential fatty acids (EFAs)** because they cannot be produced by the body. They must be taken in with the diet. EFAs include linoleic acid, arachidonic acid, and possibly linolenic acid although very little of the latter is used. About 10 g of EFAs are needed daily, but human diets are rarely deficient in them.

Health consequences of diets rich in different fatty acids

Diets of affluence

Affluent diets tend to contain an excess of lipids and fatty acids and so provide more energy-rich items than the body requires. People in such situations are in danger of becoming over-weight and then obese. This is because the excess fat is stored in fat cells (Figure 13.16) that make up the adipose tissue found around body organs and under the skin.

There are disputes among dieticians and medical researchers as to just what adverse medical conditions may follow for the individual from being chronically over-weight, but there may be an enhanced likelihood of acquiring **type II diabetes** (page 224) and possibly high blood pressure (**hypertension**, page 665). Also, in an affluent, active society there are social and psychological pressures that may cause an obese person to become depressed. The popular image of the jolly fat person is often misplaced.

Clearly, there are some health consequences of dietary fat consumption; we need to consider three issues in particular – cholesterol, omega-3 fatty acids, and ketosis.

Cholesterol

Cholesterol, a steroid, is a lipid with a different chemical structure from that of the fatty acids. The skeleton of a steroid is a set of complex rings of carbon atoms; the bulk of the molecule is hydrophobic, but the polar —OH group is hydrophilic (Figure 13.7).

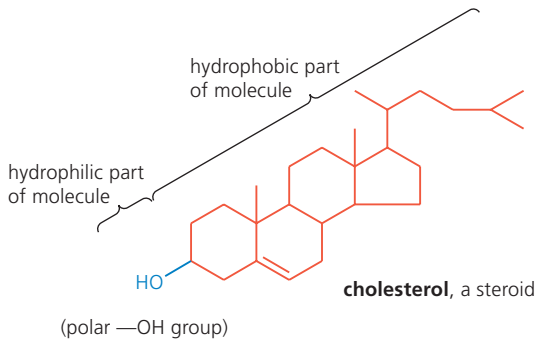


Figure 13.7 The steroid cholesterol

Cholesterol is an essential component in the plasma membranes of all cells. In mammals, cholesterol has additional roles (Figure 13.6). For example, the sex hormones progesterone, oestrogen and testosterone and also certain growth hormones are produced from it. Bile salts, compounds involved in lipid transport in the blood plasma, are also synthesised from cholesterol. Consequently, cholesterol is essential for normal, healthy metabolism.

Cholesterol is made in the liver of mammals, formed from ethanoyl coenzyme A (previously known as acetyl CoA, see Figure 9.5, page 273) and certain fatty acids – many steps and enzymes are involved in the synthesis. An adult human on a low-cholesterol diet forms about 800 mg each day. However, the production pathway is highly sensitive to an excess of dietary cholesterol; if we eat a high-cholesterol diet, the liver's production is suppressed. Diets high in fats, and especially those containing saturated fatty acids (animal fats), often cause this, and lead to abnormally high blood-cholesterol levels. On the other hand, diets where the fatty acids released are mostly polyunsaturated and in the *cis* forms lead to very low levels of blood cholesterol. The benefits of reducing dietary cholesterol are discussed later in this chapter.

Omega-3 fatty acids

The omega-3 polyunsaturated fatty acids do not influence the levels of blood cholesterol (apart from if they displace saturated fats in our diet), but they appear to help prevent heart disease by reducing the tendency of the blood to form clots, and by giving us healthy, well-functioning plasma membranes around cardiac muscle fibres.

This claim is based on studies of people who have a high daily intake of these fatty acids, such as Greenland Inuit (such studies are known as **cohort studies**). Here, diets are exceptionally rich in oily fish meat. Associated with this diet (and life style) is a very much reduced likelihood of an acute myocardial infarction (heart attack) compared to other human groups.

It is estimated by dieticians who favour cohort studies that we need about 0.2 g per day (1.5 g per week) of omega-3 fatty acids, which will be taken in if we eat sufficient oily fish, regularly. Walnuts are a plant source rich in omega-3 fatty acids.

However, other studies and trials (referred to as **randomised control trials**) suggest it may be difficult to show the clear benefits of omega-3 fatty acids. This issue may remain a controversial one for some time.

Ketosis

Fat deposits in the body are not inert stores; rather they are constantly being broken down and re-formed. This allows for interchange of the fatty acids in the triglycerides, by which method the body finds particular fatty acids required in syntheses.

However, during fat metabolism small amounts of acetoacetic acid, butyric acid and acetone are produced. These three are known as **ketone bodies** and if they accumulate in the blood, ketosis is said to occur.

Provided ketone bodies are quickly respired their presence is beneficial, but if they accumulate they may reach toxic levels and their acidity may dangerously lower the blood pH (from 7.4 to 7.2 or lower). People with an excessively lipid-rich diet, and those with severe diabetes, may be in danger from ketosis. Blood at a much lowered pH causes a condition known as acidosis which may trigger a coma, and eventually leads to death.

Extension: Diets consistently very low in energy-rich foods, including lipids, cause major health risks

Such diets miss out on sufficient fatty acids, so people in these situations respire the amino acids derived from protein digestion, rather than using them to build and maintain tissues. On a continuing low-energy diet, muscle proteins are broken down, and the person's body wastes away. For the nutritionally deprived members of communities in less-developed countries this is a constant danger. The recommendation is that they consume more foods rich in fatty acid, not less (should they be available).

4 **Suggest** likely reasons why small changes in blood pH may lead to a coma.

Minerals, vitamins and dietary supplements

A1.8–1.14

Minerals are essential elements, present in ionic form. About 15 of them are known to be essential for a healthy body. They are obtained from food sources where they are present in low concentrations. The functions performed by these minerals are threefold:

- some are **major constituents of structures** such as bones and teeth, including calcium, phosphorus and magnesium;
- some are used in the **control of the composition of body fluids**, including sodium and chloride ions;
- some are minor but **essential constituents** of specific enzymes or components of proteins or hormones concerned in metabolism – for example, iodine (page 402).

Vitamins, on the other hand, are organic compounds, but they too are required in only tiny amounts. Nevertheless, they are essential constituents of the diet as most cannot be synthesised by the body. Diseases called **deficiency diseases** are associated with a shortage of specific vitamins. Vitamins have complex chemical structures, are mostly quite different from each other, and do not belong to a single group of chemical compounds. Consequently, they are put in groups: vitamin A, vitamins of the B group, and so on, but some are also known by their specific names. The functions of vitamins are as diverse as their structures, although several function as coenzymes in the body. Some of the vitamins are fat-soluble; others are water-soluble, of which vitamin C is a good example.

Vitamin C and health

Vitamin C is a relatively small organic molecule of formula $C_6H_8O_6$. The molecule is known as ascorbic acid, but it does not contain a free carboxyl group (—COOH) as true organic acids do. However, it is 'sharp' to the taste and does form salts. It is a good reducing agent. Ascorbic acid is water soluble, and it is destroyed by cooking, particularly in alkaline conditions. Humans are among the few animals unable to form their own vitamin C (as are the monkeys, primates and guinea pigs). We require it in foods, but it is not widely distributed in food items.

Ascorbic acid occurs mainly in foods of plant origin (fruits such as blackcurrants, rosehips, strawberries, oranges and lemons), and in vegetables (green vegetable such as sprouts, cabbage, cauliflower and water cress, and in potatoes). As much as 75% of the ascorbic acid in green vegetable is **lost in cooking**. There is a very small amount in cow's milk (which is reduced by pasteurisation) but significantly more in human breast milk (page 414).

The importance of vitamin C

The importance of a vitamin is first recognised when it is absent from a diet. Vitamin C is low in many diets worldwide, and the early deficiency signs are an increased susceptibility of the mouth and gums to infection and a slowing down in the rate at which wounds and tissue damage heal. Prolonged deficiency in vitamin C leads to a condition known as **scurvy**. This takes the form of haemorrhages under the skin and other tissues, the gums become swollen and spongy, and teeth loosen and fall out. Deficiencies occur in infants when the mother's milk is substituted by cow's milk which has not had addition of supplementary vitamin C (or has been boiled before drinking). Scurvy in infants causes pain in the lower limbs and changes in bone structure.

In the body, ascorbic acid is involved in the synthesis of collagen, the key structural protein in connective tissues, which makes up about 30% of the total of protein in the body. It is in the synthesis of the amino acid hydroxyproline from proline that vitamin C is involved; hydroxyproline makes up about 15% of collagen. It is also involved in the synthesis of lipoproteins, by which lipids are made soluble for transport in the blood plasma. Lack of this vitamin is linked to iron deficiency and to anaemia (due to low blood haemoglobin levels).

5 **Explain** why failures in the maintenance of body collagen can lead to haemorrhages.

The cells involved in the formation of bone, enamel and dentine fail to function properly without adequate vitamin C, but exactly how this vitamin participates in all these processes is not known.

The discovery of vitamin C

An early report of a mysterious illness (with symptoms we now recognise as **scurvy**) was made by members of the crew of a ship that docked in Quebec in 1545. By chance, those sailors benefited from the use of a local 'herbal cure' – a juice made from the leaves of a local tree, obtained when they moved ashore. All were cured.

In 1753, another sea captain reported this disease could be avoided or the symptoms cured by supplementing the diet with fruits such as oranges and lemons. After that, lime juice came to be issued to the British Navy when away at sea.

Until 1907, scurvy was thought to be specific to humans. Attempts to induce the symptoms in laboratory rats and mice were entirely unsuccessful. Then it was discovered that a small mammal, the guinea pig (*Cavia porcellus*), was also unable to manufacture ascorbic acid. Guinea pigs, fed on diets free of this vitamin, developed symptoms of the disease. Scientists, using this animal model, established that scurvy was due to a deficiency of vitamin C (Figure 13.8). However, it was not until 1930 that the chemical, ascorbic acid, was isolated and named.

With the role of ascorbic acid in mammalian health established, the question arose as to the minimum dose required to avoid the symptoms of scurvy in humans. Between 1942 and 1944, experiments were conducted so that the role of vitamin C could be better understood. Initially, all the volunteers received a substantial daily dose of vitamin C within their balanced diets. Then, for a prolonged experimental period, one group continued on the same dietary regime, the second group received a diet with reduced vitamin C and the third group received a diet without any vitamin C. The outcomes are summarised in Figure 13.8.

Figure 13.8 Experiments that established the role of Vitamin C

The guinea pig (*Cavia porcellus*) was an inspired choice for experiments on the importance of vitamin C – it is one of only very few animals unable to synthesise the vitamin.

working with the guinea pig

Alex Holst (Professor of Hygiene and Bacteriology, University of Oslo)
and
Theodor Frolich (paediatrician)

↓ fed guinea pigs a diet of bread baked from different grains

↓ scurvy-like symptoms developed

↓ diet supplemented with fresh cabbage or lemon juice

↓ full health returned

↓ published their results in 1907

↓ scientific uproar resulted – the concept of nutritional deficiencies was unknown

Vitamin C was identified by 1930.

working with human volunteers (1942–1944)

Pattern of experimental treatments		
6 week preliminary treatment	**32 week – experimental treatments**	
all volunteers given diet including 70 mg vitamin C per day	Group A: 3 volunteers kept on the diet with 70 mg per day	none of these volunteers developed scurvy
	Group B: 7 volunteers had their diet adjusted to supply only 10 mg per day	
	Group C: 7 volunteers had their diet adjusted to provide no vitamin C*	all developed scurvy
* When the symptoms of scurvy were evident, these volunteers were subjected to superficial cuts of length 3cm in the skin of their thighs. These were then surgically stitched up and kept under observation. The wounds failed to heal		

The recommended daily dose of vitamin C

The basis for current estimates of an appropriate daily dose of vitamin C for humans has come from earlier experiments and from long-term observations of the health of communities on particular diets. Today, the term 'recommended daily allowance' has been replaced by the concept of dietary reference value (DRV). The figures for vitamin C intake per day vary between 25–30 mg for infants and young children to 40 mg for adults. Pregnant and lactating mothers require additional amounts.

TOK Link

The investigations once conducted on people are viewed as unethical and many people have reservations about animal experimentation.

1 If patients taking part in carefully planned experiments are volunteers, does this influence your judgement as to whether their treatments are ethical or not?
2 Are alternative investigations using animals acceptable in some situations?

The controversy over high doses of vitamin C

Some people have proposed much higher daily intakes – of between 10 times and 100 times the DRV. Extremely large doses – in excess of 500 mg per day, are claimed to protect against upper respiratory tract infections such as the common cold, against cancer (remember, this is not a single disease), and against arterial and heart disease.

However, the results of these experimental doses are doubtful, and it appears the megadose hypothesis may prove to have little scientific basis. At very high doses, some people have diarrhoea. Above a saturation level, excess vitamin C is excreted immediately. If a person has been taking an excessive dose for some while, the body may develop temporary dietary deficiency symptoms after returning to a normal daily dose – a case of so-called rebound malnutrition.

Vitamin D and health

Vitamin D is found almost exclusively in foods of animal origin, but there are few really rich food sources of this vitamin (Table 13.1).

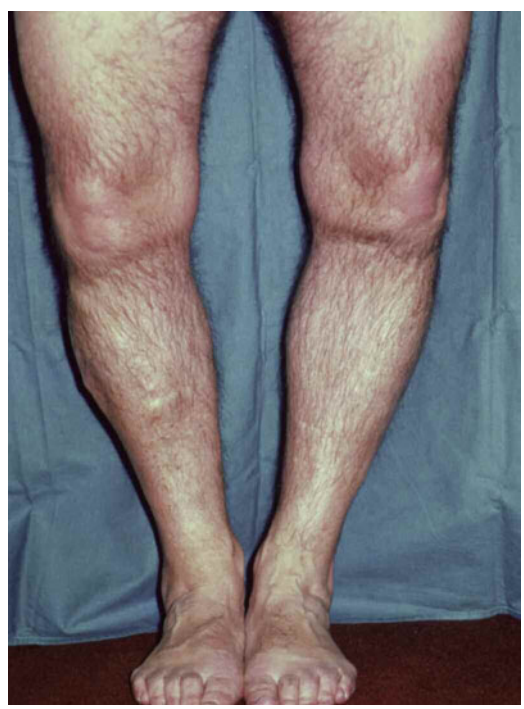
Actually, for very many people, the majority of vitamin D they require comes from the action of sunlight on a naturally occurring precursor chemical, related to cholesterol, which occurs in the skin (Figure 13.6). Problems of supply normally only arise for people exposed to relatively little sunlight – for example, those who are elderly and housebound, and members of ethnic communities who by tradition wear totally enveloping clothes. Children and pregnant or lactating mothers have especially high requirements, and these groups also need food containing sufficient vitamin D.

Source	Vitamin D/μg g^{-1}	Source	Vitamin D/μg g^{-1}
halibut liver oil	up to 100	cornflakes, fortified	0.028
cod liver oil	2.0–7.5	milk	0.0003–0.0010
herrings, sardines	0.05–0.45	butter	up to 0.02
salmon	0.04–0.30	cheese	0.003
low-fat spread	0.075	eggs	0.01–0.015

Table 13.1 Foods in which vitamin D occurs

Figure 13.9 Patient who had childhood vitamin D deficiency

This patient experienced prolonged vitamin D deficiency as a child and has developed rickets – the weak bones have bent permanently under normal body weight



X-ray of lower leg bones of a healthy patient with bones strong enough to sustain normal body weight and activities (false colour radiograph)

Vitamin D is concerned with the **absorption of calcium and phosphorus** by the body. In the absence of sufficient of this vitamin, these elements, naturally present in the food and available for absorption in the lower gut, pass out with the faeces. So, indirectly, vitamin D influences bone mineralisation. Infants deprived of vitamin D develop rickets, and have deformed bones, often too weak to support them (Figure 13.9). In the elderly, deficiency in vitamin D leads to osteomalacia, a disease in which the bones soften.

Vitamin D from sunlight versus the danger of malignant melanoma

Exposure of skin to sunlight brings with it potential dangers, as well as the advantage of generation of vitamin D. Excessive exposure to sun is the major cause of skin cancer. **Malignant melanoma** is just one form of this type of cancer; it accounts for only 2% of skin cancer cases, but it is the most life-threatening, if it is contracted. This is because, once the disease is established, it develops quickly, the malignant cells moving about the body, colonising other tissues and growing rapidly (metastasising).

The outer layer of the skin consists mainly of keratinised cells, but about 8% of the cells here are pigmented cells called melanocytes. These contain the brown–black pigment melanin that contributes to our skin colour. These cells absorb damaging ultra-violet light, and are the cells which may become carcinogenic. The current rise in skin cancer is attributed partly to people spending more time in the sun, and partly to depletion of the ozone layer (page 627). People with light-coloured skin who burn in sunshine rather than developing a sun-tan are vulnerable. So too are people who live in the sunny climates and work outdoors, together with those at high altitude, where UV light is more intense.

The benefits of artificial dietary supplementation

We have seen that certain common food items are supplemented with additional vitamins. From Table 13.1 it is clear that in cornflakes, a breakfast cereal manufactured from maize 'seeds' (they are actually fruits), the vitamin content has been supplemented. Supplementation of breakfast cereals is mostly by the use of malt extract as a necessary addition in the manufacturing process (Figure 13.10).

Similarly, popular low-fat spreads – 'healthier' substitutes for butter – have several added vitamins. For many people, this will be an important source of some of the additional vitamin D they require.

The addition of minerals

The thyroid is an endocrine gland at the base of the neck. Thyroid hormones, including **thyroxine**, are manufactured from the amino acid tyrosine with the addition of the mineral **iodine**. Thyroid hormones affect all body cells. They cause an increase in the rate of energy production, known as the **basal rate of metabolism**.

Iodine deficiency is common in many countries; most foods are naturally short of this mineral. In fact, sea foods are about the only good source. Otherwise it is drinking water that provides a natural dietary source of iodine, when present as the iodide and iodate ions. In iodine-deficient areas, the soil and rocks contain little of this element.

If iodine is not added in the diet, people may develop a distinctly swollen thyroid gland, which causes goitre (Figure 13.11). The common method of combating this deficiency is the addition of iodine as the iodate ion to salt used for cooking and on the table.

Vitamin additions to diet

There has been a progressive increase in the use, by people in the developed world, of dietary supplements – typically vitamin pills. One example already discussed – those who take a megadose of vitamin C – illustrates how the use of dietary supplements may be popular, despite little evidence of usefulness. Groups with whom this practice has gained popularity include some competitive athletes and exercise enthusiasts, together with elderly people concerned about decreasing health and loss of faculties.

Figure 13.10 Additives in breakfast cereal

Nutrition

Typical values	Per 30g with 125ml semi-skimmed milk	Per 100g
Energy	730 kJ	1596 kJ
	172 kcal	**376 kcal**
Protein	**6.3g**	**7.3g**
Carbohydrate	**31.5g**	**84.0g**
of which sugars	9.0g	8.9g
of which starch	22.5g	75.1g
Fat	**2.3g**	**1.2g**
of which saturates	1.4g	0.4g
of which mono-unsaturates	0.7g	0.2g
of which polyunsaturates	0.3g	0.6g
Fibre	**0.9g**	**3.0g**
Salt	**0.6g**	**1.6g**
of which sodium	0.3g	0.6g
Vitamin D	**1.5μg (30% RDA)**	**5.0μg (100% RDA)**
Thiamin (B1)	**0.5mg (34% RDA)**	**1.4mg (100% RDA)**
Riboflavin (B2)	**0.7mg (44% RDA)**	**1.6mg (100% RDA)**
Niacin	**5.5mg (31% RDA)**	**18.0mg (100% RDA)**
Vitamin B6	**0.7mg (34% RDA)**	**2.0mg (100% RDA)**
Folic Acid	**128.0μg (64% RDA)**	**400μg (200% RDA)**
Vitamin B12	**0.8μg (80% RDA)**	**1.0μg (100% RDA)**
Pantothenic Acid	**2.2mg (37% RDA)**	**6.0mg (100% RDA)**
Iron	**4.3mg (30% RDA)**	**14.0mg (100% RDA)**

RDA means recommended daily allowance

Ingredients

Maize, Sugar, Salt, Dextrose, Barley Malt Extract, Emulsifier: Mono- and Diglycerides of Fatty Acids; Iron, Niacin, Pantothenic Acid, Vitamin B6, Riboflavin (B2), Thiamin (B1), Folic Acid, Vitamin D, Vitamin B12.

Figure 13.11 The condition of goitre (prevented by iodine supplement)

adult with an untreated goitre

Countries with the highest incidence of goitre are typically inland areas where iodine levels in soils are low, and sea food is uncommon or unavailable.

Countries where high levels of goitre have been reported in recent years include:

- Syria, Central African Republic, Guinea, Zambia and Bangladesh (over 50% in junior age groups)
- Paraguay, Nepal, Lesotho, Rwanda and Albania (over 40% in junior age groups).

Of course, if individuals have a deficiency in particular minerals or vitamins, supplementary pills are a sound way of remedying it. Multivitamin capsules contain the recommended quantity of each vitamin (Figure 13.12). Care in their use is required; once the body is saturated with what is required, excess vitamin may accumulate. Many vitamins are water-soluble, and any excess intake is rapidly excreted, but with vitamins that accumulate in the body, then potentially toxic doses may build up.

Figure 13.12
Multivitamin preparation

One-A-Day complete multivitamin food supplement

Storage

KEEP OUT OF THE REACH OF YOUNG CHILDREN. REPLACE LID SECURELY.
STORE IN A COOL, DRY PLACE, BELOW 25°C.

Usage

RECOMMENDED DAILY INTAKE.

Swallow one tablet a day with a cold drink. Do not chew or take with hot drinks.

Do not exceed the recommended intake. Women who are pregnant or planning a pregnancy should consult their doctor before taking vitamin or mineral supplements. Food supplements are intended to supplement the diet and should not be regarded as a substitute for a varied diet.

Nutrition

Typical Values	Per Tablet	% RDA
Niacin	18 mg	100
Vitamin A	800 µg	100
Vitamin D	5 µg	100
Vitamin E	10 mg	100
Vitamin C	60 mg	100
Thiamin (Vitamin B1)	1.4 mg	100
Riboflavin (Vitamin B2)	1.6 mg	100
Vitamin B6	2 mg	100
Folic Acid	200 µg	100
Vitamin B12	1 µg	100
Biotin	0.15 mg	100
Pantothenic acid	6 mg	100
Calcium	90 mg	11
Phosphorus	38 mg	5
Iron	5 mg	36
Magnesium	45 mg	15
Zinc	5 mg	33
Iodine	100 µg	67
Manganese	1.2 mg	*
Chromium	25 µg	*
Molybdenum	25 µg	*
Selenium	30 µg	*
Chloride	4.5 mg	*
Potassium	5 mg	*

Dietary fibre

Dietary fibre, or **roughage**, is mostly of plant origin, typically consisting of cellulose from cell walls. Foods high in dietary fibre include vegetables, fruits and nuts. The important quality of dietary fibre is that it is not digested in the gut of a mammal. It is therefore not present for direct nutritional benefits, unlike all the other dietary constituents.

Initially, the value of dietary fibre within the gut was seen simply as a component of the otherwise diminishing food mass that persists, retains water, and provides bulk to the gut contents in the later stages of digestion and during absorption. Its presence thereby stimulates movement by peristaltic activity (see page 181) through the entire alimentary canal.

The present high degree of interest in dietary fibre follows comparative studies of diets that are low and high in fibre, and the different disease patterns that appear to be associated with them. The study of community disease patterns of this sort is called **epidemiology**. Such investigations cannot prove relationships because there are so many uncontrolled variables which cannot be resolved. For example, as more fibre is added to a diet, the proportions of other components also change.

However, it does appear that the presence of substantial dietary fibre is consistently associated with the absence of the diseases of affluence, namely **constipation**, **diverticular disease**, **cancer of the colon**, **appendicitis**, **varicose veins**, **haemorrhoids**, **venous thrombosis**, **atherosclerosis** and **coronary heart disease**, **diabetes**, and **peptic ulcers**, to varying degrees.

How may such dramatic advantages be brought about?

The answer is that it is not entirely clear how, but the impact of diets high in fibre on gut function and our nutrition may be due in part to the following advantageous features of such diets.

- They tend to be low in fat, refined sugar and salt, and generally contain more protein of plant origin than animal protein.
- They may have a satiety effect because their bulk may be giving us the impression of being well fed while consuming less of digestible foods.
- They decrease the time taken for the meal to pass through the entire gut.
- They cause the production of faeces in the form of large, soft stools that pass readily and quickly through the large intestine without strain or damage to mucosa (linings) of the thick muscular walls there.

Energy issues in human diets

A2.1–2.3

Energy content

An essential component of a balanced diet is food that supplies **metabolic energy**, and it is this aspect of diet and health that we now consider. The energy we need to sustain the processes of metabolism, growth, repair, and all other activities of cells, tissues and organisms, is normally supplied from carbohydrates and fats (lipids), but remember, metabolic fuel can also be obtained from proteins. Chemical energy is released by respiration in living things, but proteins must first be deaminated. This means that the combined nitrogen as the amino group (—NH_2) is removed first (page 411), before respiration occurs.

The energy value of foods is expressed in **joules** and **kilojoules** (**J** and **kJ**, 1 kJ = 1000 J). The joule is defined in terms of heat energy: 4.18 J of heat energy are required to raise 1 g of water through 1 °C. The energy typically provided by individual nutrients is estimated experimentally, using the bomb calorimeter (Figure 3.1, page 76). The amounts of energy that major food substances might yield when they are completely respired, are as follows (in kJ g^{-1}):

- **carbohydrates** 16.0–17.6 ;
- **lipids** (fats and oils) 37.0–40.0 ;
- **protein** 17.0–17.2 .

An earlier term for the amount of energy in food substances was the calorie. This term is often still found on commercial packs of food (1 calorie = 4.18 J).

Why do lipids transfer more energy than carbohydrates?

Triglycerides are the form in which lipids are respired. When triglycerides are oxidised in respiration, a lot of energy is transferred and used to make ATP. Mass for mass, fats and oils release more than twice as much energy as carbohydrates do when they are respired. This is because fats are more reduced than carbohydrates – more of the oxygen in the respiration of fats comes from the atmosphere, whereas in the oxidation of carbohydrate, more oxygen is present in the carbohydrate molecule, itself. Fat, therefore, forms a concentrated energy store.

6 **Distinguish** between
a oxidation and reduction
b condensation and hydrolysis reactions.

Ethnicity and energy source

Diets around the world are largely influenced by local **readily available food sources**, regardless of whether food is obtained by some form of hunting and gathering or supplied by farming and horticulture. In all cases, local seasons and climate, soil and natural dominant vegetation clearly influence local food supplies.

Additionally, **cultural and religious traditions** have become influential as to which of the available foods are selected and used. The impact of this factor is illustrated by reference to traditional ethnic and religious rules (Table 13.2).

Group	Animal protein	Other factors
Hindus	■ no beef ■ rarely eat fish	■ mostly vegetarian ■ no alcohol ■ fasting common
Jews	■ no pork ■ animals slaughtered by licensed person, then salted ■ only fish with scales or fins (vertebrates)	■ meat and dairy products not eaten together
Muslims	■ no pork ■ meat dedicated to God by Muslim present at slaughter	■ regular fasting as during Ramadan (fast during daylight hours for 1 month)
Rastafarians	■ vegetarians (organic preference) ■ no animal products apart from milk	■ food must be living (not processed) ■ no salt ■ no alcohol
Sikhs	■ no beef ■ animals slaughtered by 'one blow to head'	■ no alcohol ■ otherwise less rigid rules than Hindus and Muslims

Table 13.2 Dietary restrictions practised by ethnic and religious groups (adapted from MAFF *Manual of Nutrition* 10th edn HMSO, London. p. 121)

7 Predict the nutritional advantages of diets low in or free from alcohol.

How the staple energy source of different ethnic groups varies around the world is illustrated in Table 13.3 and in Figures 13.13 and 13.14.

Source	Ethnic group example	Environmental features / food value
rice	*Oryza sativa* grown extensively throughout Asia (e.g. China; 35% of world production)	cultivated in China for 5000+ years grown in paddy fields of standing water – grains provide 30%+ starch and about 2.5% protein
wheat	*Triticum durum* (durum wheat) grown in central Europe (e.g. Italy)	requires moderate moisture and cool weather for early growth, then sunny, dry months – provides flour for pasta of 70% starch and 12% protein
cassava	*Manihot esculenta* grown under slash-and-burn cultivation in tropical African countries (e.g. Nigeria, Congo and Rwanda)	survives in poor soil with little attention, providing foliage for livestock (30% protein), and swollen roots of 30% starch (no protein or lipid) which are boiled and then pounded into a thick paste
maize	*Zea mays* originated in South America (modern Mexico) now cultivated widely to be a mainstay food in many communities	modern varieties are hybrid seeds, requiring advanced agricultural practices – modern cultivation, fertilisers and pesticides, and careful harvesting; most varieties yield starch and oil, but some, protein also
fish	sea fishing for marine fish by isolated communities with basic hand-made equipment (e.g. Tupi Indians of Brazil's north-east coast)	fish flesh typically provides approximately 20% protein and 10+% oil over-fishing has led to serious depletion in fish stocks where modern, industrial boats with huge nets have been devised
meat	herds of sheep and goats, taken out and supervised in daylight, then enclosed at night – by Moroccan (Arab and African) families	animals feed by foraging on the few plants, sparse shrubs and bushes growing on slopes (e.g. of High Atlas mountain ranges) bordering the nearby deserts yield milk for yoghurts and cheese, and meat – typically about 20% protein and 8% fats

Table 13.3 Main dietary sources of different ethnic groups

Figure 13.13 Cultivation of four staple crops in ethnic context

Top left: cassava on land made productive by 'slash-and-burn' cultivation – Senegal, Africa

Top right: durum wheat being harvested on steep slopes – Tuscany, Italy

Bottom left: rice grows in paddy fields cut into limestone hillsides – southern China

Bottom right: harvested maize crop being 'husked' – Ecuador, South America

Figure 13.14 Herding and fishing for staple energy sources

Morocco – Berber shepherd with a flock of sheep.

Brazil – Caboclo people fishing on the Amazon River.

Health consequences of energy-rich diets

Excess carbohydrates

We have seen that carbohydrates are organic compounds formed from the three elements, carbon, hydrogen and oxygen, with hydrogen and oxygen present in the ratio 2:1, as they are in water. Carbohydrates occur in three groups, the **monosaccharides**, **disaccharides** and **polysaccharides** (Table 2.4, page 44). The monosaccharides or simple sugars are the basic unit of structure of carbohydrates; disaccharides and polysaccharides are formed by condensation reaction of simple sugars. Common sources of carbohydrates in the diet are shown in Table 13.4.

Table 13.4 Common dietary carbohydrates in human nutrition, and the percentages of various carbohydrates contained by some common food sources

Carbohydrate groups	Example	Common food sources
monosaccharides	glucose	tomatoes (1%), onions (2%), grapes (7%), honey (30%)
	fructose	honey (35%), sweet fruit juices
disaccharides	sucrose	sugar cane, sugar beet, honey
	lactose	milk – cows (5%), human (7%)
	maltose	malt extract, by hydrolysis of starch
polysaccharides	starch	bread and pasta, potatoes, cassava, rice, maize (corn)
	glycogen	liver
	cellulose is not digested – it is the chief source of dietary fibre	

Sugars and starch – the major products of green plant metabolism – are the chief source of food energy for humans all over the world. Dietary glucose is absorbed directly, but starch is first digested to glucose.

How are sugar and starch used in the body?

Once absorbed into the blood, glucose may be used in cell respiration in all tissues. Additionally, glucose is commonly used in the body to metabolise other substances. For example, it may be converted to the carbohydrate in the glycoprotein found on membranes (fluid mosaic membrane, page 21). It is also metabolised into the two 5-carbon sugars, ribose and deoxyribose, and built into nucleotides and nucleic acid, RNA and DNA (page 63). Sugars are also commonly metabolised into certain organic acids. These then provide the carbon skeletons to which an amino group is added in the production of specific amino acids required for protein synthesis (page 392).

However, in diets **rich in excess carbohydrates**, excess glucose does not remain in the blood. Blood glucose levels are tightly controlled (page 222) except in patients with diabetes. Excess glucose is rapidly stored, rather than excreted.

First, glucose is converted to glycogen, the polysaccharide found in mammals. Glycogen itself is easily broken down to glucose, as and when required. A certain amount of glycogen is stored around the body, mainly in the liver and muscles (the brain cannot store glycogen).

Then, given continuing excesses of blood glucose, the long-term store is as fat, laid down under the skin (subcutaneous fat), and also in connective tissue around body organs. Thus, the outcome of maintenance of a high carbohydrate diet is the condition of being first over-weight, then obese. These terms are defined in terms of the body mass index (BMI – see below). Figure 13.15 shows a way of measuring body fat.

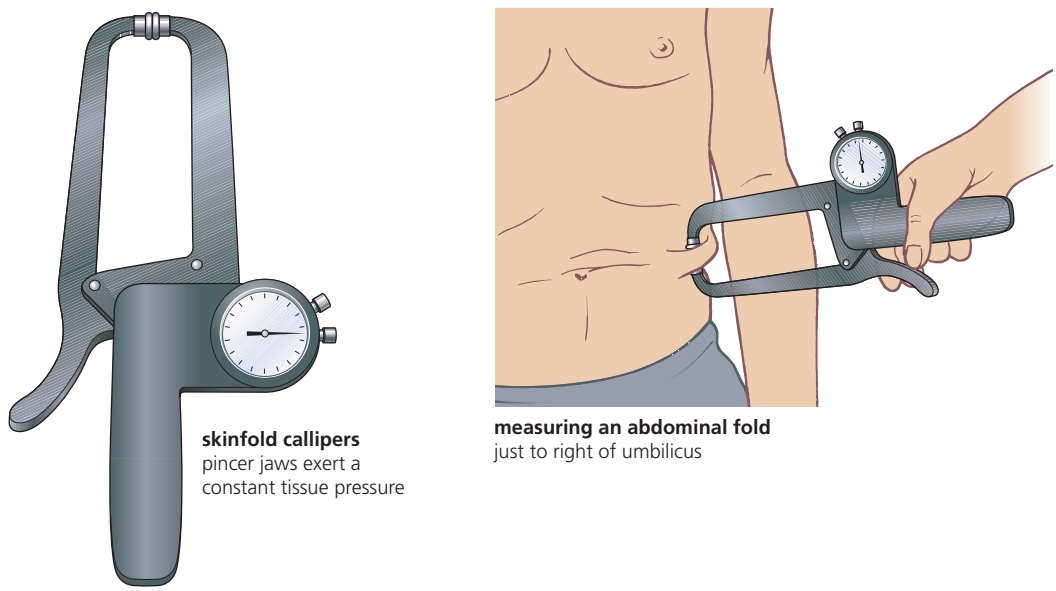


Figure 13.15 Measuring body fat in humans

8 Outline the dangers to the body of glucose occurring in the blood circulation at:

a too high a concentration
b too low a concentration.

Excess lipids

We have already seen that lipids, too, are organic compounds of carbon, hydrogen and oxygen, but here the proportion of oxygen is less than in carbohydrates (page 48). In effect, the carbon of lipids is more reduced, and therefore, gram for gram, lipids are more concentrated sources of energy (see above). We may obtain dietary lipids in the form of oils from plant sources and fats from animal sources; examples of foods rich in lipids are listed in Table 13.5.

Source	Lipid	%	Source	Lipid	%
cream (from cow)	fat	19–48	low-fat spread	oil	40
milk	fat	4	cheese	fat	34
herrings	oil	13	avocado	oil	19
beef	fat	11	peanut butter	oil	54
chicken	fat	18	almond nuts	oil	56

Table 13.5 Examples of food sources of lipids

How are lipids used in the body?

In the steps of cell respiration, fats and oils are broken down to fatty acids and glycerol, and these intermediates feed directly into the pathways of cell respiration. Another form of lipid in cells is as phospholipids. These are key components of all cell membranes. Nerve fibres in mammals are insulated from each other by a lipid sheath, made of the cell membranes of Schwann cells (Figure 7.31, page 211). The general role of fatty acids is summarised in Figure 13.6 (page 396).

Figure 13.16 SEM of mammalian fat cells – here the cytoplasm is packed with lipid

Because fats and oils are insoluble in water (we say they are hydrophobic molecules), they can be safely stored in cells and tissues without osmotic consequences. Fats are stored in the bodies of animals in fat cells (Figure 13.16). We have already noted that excess glucose may be converted into fats and stored. The huge stores of fat that build up in the bodies of marine mammals may be seen as an insulation against loss of body heat into the surrounding, intensely cold waters, since the adipose tissues (fat stores) have a limited blood supply. However, all who continue with a high-lipid diet, in excess of what is required, will become over-weight and then obese.

Excess proteins

We have seen that proteins are relatively large molecules, built from amino acids by condensation reactions. Amino acids contain the element nitrogen, in addition to carbon, hydrogen and oxygen. Some contain sulphur, too. We refer to proteins as our source of combined nitrogen; this essential element is approximately 80% of the air we breathe, but molecular nitrogen is not available to animal cells. Only when it is chemically combined in molecules like amino acids can our cells use nitrogen. Examples of food sources for proteins (and the percentage of protein they contain) are:

- lean beef 20%;
- herring 20%;
- Quorn (mycoprotein) 12%;
- almond nuts 21%.

How are proteins used in the body?

Dietary proteins, whether from animal or plant sources, are first digested to their constituent amino acids, then these are absorbed.

The absorbed amino acids are used to build the body's own proteins which are required for the growth and functioning of cells and tissues, and in their repair, when necessary. There are about 20 amino acids from which a mammal's own proteins are built up. The important roles of proteins include their functions as:

- enzymes (biological catalysts);
- antibodies in our defences against disease;
- contractile fibres in muscles;
- a component of membranes.

A protein-rich diet may provide excess protein and amino acids. These cannot be stored in the animal's body. Instead, the amino group (—NH_2) is first removed in a step known as **deamination**. The amino group is then converted to urea by combination with carbon dioxide. This occurs in the liver. Urea is carried in dilute solution in the blood plasma, and removed from the blood in the kidneys. It is eventually excreted from the body in the urine. The remainder of the amino acid molecule, after deamination, functions as a source of metabolic fuel. However, if it is not immediately respired, an excess builds up which we convert to fat (via the action of certain enzymes of glycolysis and the Krebs cycle). So the effect of excess protein in the diet will be the excretion of urine rich in urea and the steady deposition of excess fatty acids as body fat, leading to the conditions of being over-weight and obese.

Appetite control

A2.4–2.8

Our brain houses a control centre for the appetite, located in the hypothalamus, which is a part of the floor of the fore brain (Figure 13.17). Here our appetite is regulated. Actually, this part of the brain also regulates our thirst and body temperature, operating by means of impulses despatched to specific body organs and organ systems, via nerves and the spinal cord. The actual mechanism of appetite control is understood in part.

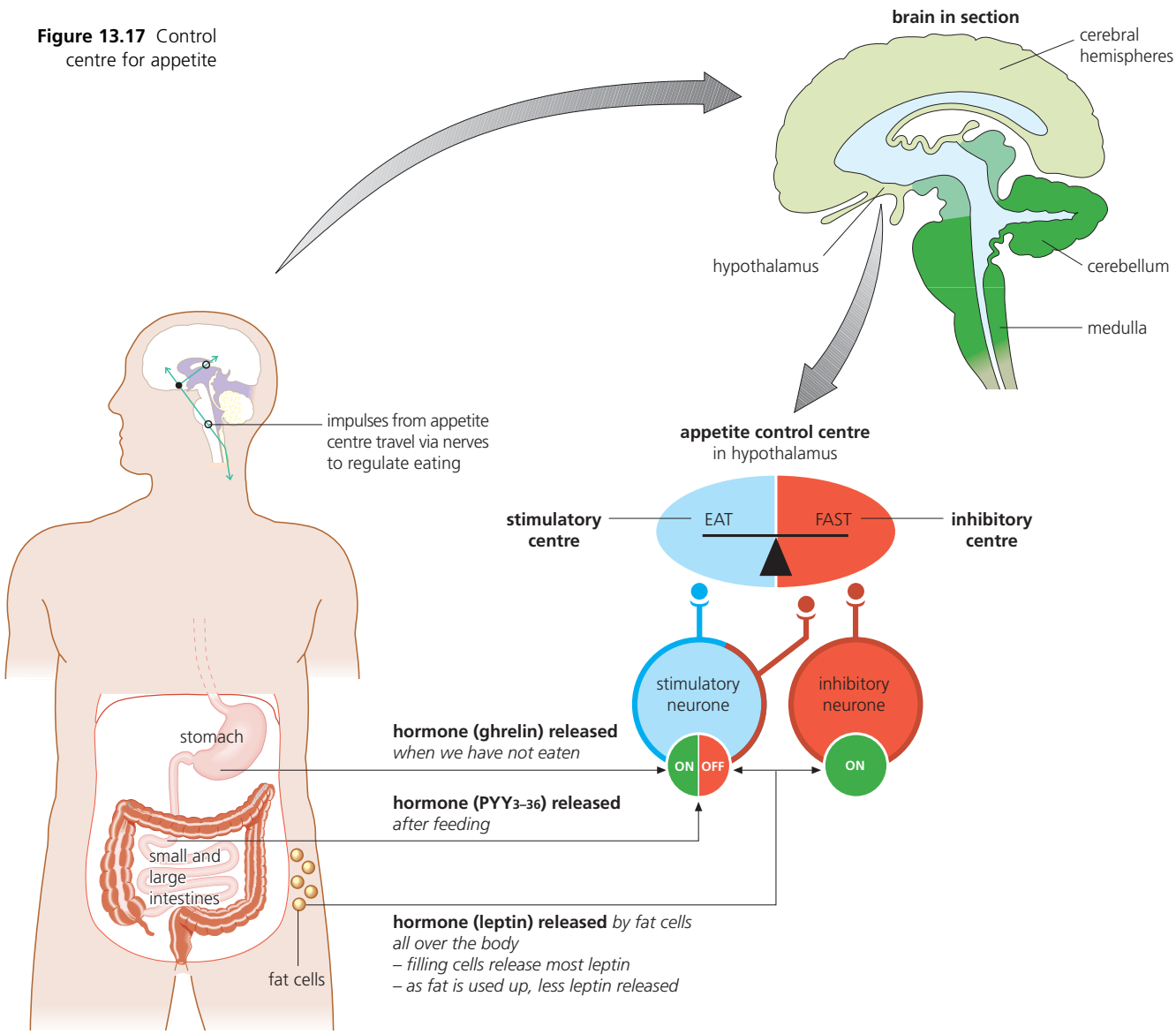


Figure 13.17 Control centre for appetite

How is this centre kept informed of the hunger state of the body?

The answer is that the appetite centre is chiefly stimulated by specific **hormones** from tissues and organs in the body. (It is possible that nerve impulses from parts of the gut may play a part, too.)

There are three hormones involved in the control process. One, known as **leptin**, plays a continuous role. In adult life, the number of fat cells does not change significantly. If we over-eat they fill up with lipids; when we are short of food, reserves are drawn on and the fat cells empty.

Now, as the fat cells fill up, they secrete more leptin. Like all hormones, leptin circulates in the blood. On reaching the appetite centre, leptin suppresses the sensation of hunger. On the other hand, when fat cells empty and shrink, they secrete less leptin, and the sensation of hunger is experienced in the brain. Clearly, leptin is associated with long-term regulation of eating.

More immediately, control is influenced by additional hormones. As the stomach empties, a hormone (known as **ghrelin**) is secreted, and this hormone stimulates the appetite control centre to create a wish to feed – our appetite, in effect. Increasingly the hunger sensation is experienced.

When we have eaten, another hormone (known as $\mathbf{PYY_{3–36}}$) is released from the upper intestines and pancreas. This gut hormone is present in the blood from early in the digestion processes, and on reaching the appetite centre, suppresses the hunger sensation. The drive of hunger and the urge to feed is overcome, for the time being.

Body mass issues

Body mass index

We need to be able to quantify the conditions of the body we believe are under-weight, normal, over-weight and obese for different people if individuals are to be able to satisfactorily regulate their body weight.

How can this quantification be done?

To accurately and consistently quantify body weight in relation to health, the **body mass index (BMI)** has been devised. We calculate our BMI according to the following formula:

$$\text{BMI} = \frac{\text{body mass in kg}}{(\text{height in m})^2}$$

You can also visit nhlbisupport.com/bmi to calculate your BMI at the click of a mouse. Table 13.6 shows the relationships between various results for BMI and weight status.

Table 13.6 The boundaries between under-weight, normal and over-weight

BMI	Status
below 18.5	under-weight
18.5–24.9	normal
25.0–29.8	over-weight
30.0 and over	obese

Extension: Health and shape

A basic calculation of the ratio between waist and hip size has recently been suggested as a more accurate indicator of the risk of having a heart attack in adult men and women than BMI. People with waist to hip ratios of 0.9–1.0 or less in males and of 0.85–0.9 or less in females have a significantly lower risk of a cardiac infarction (heart attack). This correlation is attributed to abdominal fat cells being the main source of metabolites that damage the insulin-production system in the body.

Clinical obesity

The condition of clinical obesity is defined as having a BMI of 30 and over. The incidence of obesity has substantially increased over the past 20 years. Studies by the World Health Organization, published in 2003, estimate that, worldwide, 300 million people are clinically obese, and in a developed country like the UK:

- clinically obese men 13–17%
- clinically obese women 16–19%.

Why has this trend developed, and why so quickly, too?

Obesity arises for two reasons; because of the types and amount of food we eat and because of the amount of physical exercise we undertake. We might define these as issues of life style.

Diet issues behind obesity are summarised in the phrase 'an unhealthy diet'. People in the developed world have access to food that is not highly priced – often it is extremely cheap and certainly very plentiful. Some of this food is described as junk food because, while it is extremely attractive, it is high in fats and carbohydrates – all in an easily digested condition. For example, in the UK about 70% of school children eat biscuits, sweets and chocolate daily, and only 12% consume the recommended daily intake of fruit and vegetables. Many adults do little better. Portions of food served in fast food outlets and many other convenient eating places are larger than is necessary.

Physical activity issues are equally problematic. In many Northern European communities, for example, there are few children who walk to school or college; even fewer cycle. Physical education activities in the curriculum of some countries have apparently decreased. Physically inactive children become physically inactive adults. While metal-handling industries and generally physically demanding work have decreased, desk-bound jobs have expanded. Car ownership continues to increase, and so walking and cycling become even less common practices. Computer games and television are far more popular pastimes.

Anorexia nervosa

Not all weight problems are due to people being over-weight – the conditions known as **anorexia nervosa** and **bulimia nervosa** are thought to be on the increase, particularly among young Caucasian females from middle or upper social classes. In anorexia, deliberate dieting, and sometimes deliberate vomiting, lead to serious weight loss and even the loss of consecutive menstrual cycles. However, these conditions are not exclusive to females. Patients have an obsessive fear of gaining weight or becoming fat; they see themselves as much fatter than they actually are. In bulimia, periods of excessive eating (binge eating) are followed by self-induced vomiting and use of laxatives to achieve weight control. Here, patients do not necessarily lose excessive weight and, if female, their menstrual cycles remain normal.

These two conditions are believed to have more to do with anxiety about maturation and sexuality than with diet. People in their teens are possibly more concerned about their appearance and self-image than at any other times in their lives. If this becomes linked to anxieties about their own physical development and how they will relate to their peers, particularly those of the opposite sex, then anxiety may cause some people to lose a realistic evaluation of themselves. A study of this issue in UK young people is shown in Figure 13.18 – it emphasises the importance of good role models.

Figure 13.18 'Girls gripped by worry over looks' *The Guardian* 16 November 1998

Many teenage girls worry about their appearance more than anything else in their lives. Weight and appearance are by far the main concerns for girls, eclipsing family problems, difficulties with friends, health, careers and school, according to an **Exeter University survey**.

'The overwhelming majority of those who say they wish to lose weight have no medical weight problem at all, and some are underweight' said researcher John Balding.

Researchers from the Schools Health Education unit questioned 37,500 young people aged 12–15 years, and concluded that girls were much more worried about their appearance than boys.

The study concluded that, to help counter the problem, a greater diversity of body shapes among actresses and models needed to be seen on television and in newspapers.

by journalist Helen Carter
First published in *The Guardian*, 16/11/98

Special issues in human nutrition

A3.1–3.7

Milk for infants

Mammals characteristically feed their young on milk from mammary glands. Human milk is produced in small, sac-like glands within the breasts. These glands develop when stimulated by specific hormones (oestrogen, progesterone, and others) during pregnancy. The human breasts do not store a large volume of milk as does the cow's udder. The suckling stimulus by the infant on the nipple stimulates both milk production and further hormone releases that bring about the movement and release of the milk.

Cow's milk by itself is inappropriate for human babies less than one year old. While cow's milk contains most of same components as breast milk, these components are not present in the same proportions, and there are other subtle differences, too. Cow's milk totally lacks the mother's antibodies (Table 13.7).

Breast milk		Cow's milk
lactose, about 6.6%	**sugar**	glucose, about 4.4%
mainly lactalbumin, about 1.9%	**protein**	casein, about 6.5% (much harder for a baby to digest than lactalbumin)
triglycerides of particular essential fatty acids	**fats**	same quantity of lipid, but triglycerides of different composition
necessary vitamins in correct amounts	**vitamins**	similar vitamins, in different proportions (less of vitamins A and C, more of D)
necessary minerals including calcium at 0.33 mg cm^{-3}	**minerals**	similar minerals, but calcium at much higher concentration
same as mother's immune system has produced	**antibodies**	no antibodies

Table 13.7 Breast milk and cow's milk compared

Figure 13.19 Preparing a baby's bottle in contrasting conditions – the elaborate facilities for sterilisation of bottle-feeding equipment available in developed countries contrasts with the resources to hand for a young Iraqi mother who sits in the street with her baby, in the shadow of an abandoned government building

Since cow's milk is inappropriate for young babies, in the absence of breast milk a **commercial substitute formulation** is recommended, made up with boiled and cooled tap water. This is similar in composition to breast milk but contains no antibodies. However, most babies survive adequately on it, provided the parents are in a position to make up the babies' feeds at the right concentration, and can do so with sterile water in sterilised feeding bottles (Figure 13.19).

Companies that manufacture these preparations have, on occasions, been accused of persuading mothers who can breast feed to use formula milk instead. Often this has been by attractive advertisements. Where economic, social or environmental circumstances prevent parents from using formula milk at the full concentration and under hygienic circumstances, their babies' lives have been put in jeopardy.

The benefits of breast feeding

- **Colostrum** is the milk formed for the first five days. It is of different composition. For example, it contains about half the immune factors (**antibodies**, page 351) that pass naturally from mother to infant (the remainder have already arrived via the placenta). The baby is able to resist infection.
- **All the necessary nutritional ingredients in the right amounts** for healthy growth are present in breast milk. It contains appropriate amounts of carbohydrates, protein, and fat, and it also provides the digestive enzymes, vitamins and hormones the infant requires. It provides minerals (apart from iron – infants have adequate iron stored in their liver to last until about 8 months of age). The milk is easy for the baby to digest and absorb.
- A **lower mortality rate** occurs in infants who are breast fed. In the developed world, such children are less likely to suffer infant cot death (sudden infant death). Further, breast milk is low in bacteria. Human milk contains anti-microbial factors, so breast-fed infants have **fewer infections** such as gastroenteritis.
- It is **always safe to take** because it cannot be made up in the wrong concentration (or using contaminated water), it does not cause allergies, and is not in danger of introducing harmful bacteria into the infant's gut. In the developing world, mortality rates for bottle-fed infants are very high indeed.
- The **close bond between mother and child** begun in the uterus is reinforced during breast feeding. This is especially beneficial for emotional and social development of the new individual, possibly throughout life.

Diabetes

The production of insulin occurs in the β-cells of the islets of Langerhans in the pancreas (Figure 7.44, page 222). Insulin is the hormone that brings about the maintenance of the blood glucose concentration at or about 90 mg per 100 cm^3. Failure to correctly regulate blood glucose leads to a disease known as **diabetes**. In diabetics, the level of blood glucose tends to be erratic and, generally, permanently raised. Glucose is regularly excreted in the urine. The condition requires careful monitoring and treatment.

In fact there are two forms of diabetes.

- **Type I diabetes**, also known as early onset diabetes, is due to destruction of the β-cells of the islets of Langerhans by the body's immune system (a so-called **autoimmune disease**) or perhaps by an infection. This is a disease contracted in childhood, or at least before the age of 20 years.
- A much more common form of diabetes is late-onset or **type II diabetes**, which is the form we are concerned with here. This arises most commonly in people aged over 35 years, often in those who have become over-weight or obese. In type II diabetics, the β-cells usually still produce insulin, and commonly in quantities that ought to be sufficient. However, the insulin receptors on target cells have become insensitive. It is also possible that in these patients, the fat cells may release substances that interfere with the insulin–receptor reaction.

Patients have a generally raised blood glucose level, but other symptoms of diabetes (such as thirst, needing to urinate more frequently, undiminished hunger) are only mildly experienced or are absent (such as weight loss, and limb-threatening ulcer). However, as with other forms of diabetes, there is a raised risk of circulatory disorders, renal failure, blindness, strokes or heart attacks, if the condition is not diagnosed and treated.

Susceptibility to diabetes between ethnic groups

The incidence of type II diabetes is not evenly distributed throughout the human population; certain ethnic groups are more prone to this complaint. For example, the Pima Indians of central Arizona have the highest incidence – typically about 50% of people aged 30 years and above in the communities are affected. This level has developed since they abandoned their traditional, frugal existence along the river banks, and became city-fringe dwellers outside Phoenix, replacing traditional diets with biscuits, cheeseburgers and Coca Cola.

In fact, most ethnic groups that have been discovered to have a high incidence of type II diabetics have made a recent switch from their traditional or generally minimal diet to a more lavish one. Associated with this, they may have also abandoned regular physical activity for a sedentary existence (often described as a Western life style). Other affected groups include native Americans of Pacific Island communities, Ethiopians living in Israel, Indians living in Britain, and almost every other non-European group that has switched to a Western life style.

How is type II diabetes treated, and why?

The raised blood glucose may be successfully treated by permanent adjustments to diet. The patient needs to **completely avoid** eating items that liberate a high concentration of sugar into the body quickly. Sweets and sweet biscuits are obvious examples. Carbohydrates are not excluded from the diet completely as such a move often leads to consumption of more fats – harmful in their own right. The patient needs to consume forms of carbohydrate that are slow to release sugar, as digestion proceeds. Then the levels of blood sugar generated may be successfully processed.

The patient needs to permanently control diet very carefully so that if they are already overweight they quickly return to a normal body weight and then hold body weight at that level. So, all energy-rich items are kept to the essential minimum, and alcohol is avoided. A good level of healthy physical activity is also needed.

If the patient does have a lowered production of insulin as well as decreased insulin sensitivity, then a drug that stimulates the ability to secrete insulin can be prescribed.

9 In type II diabetes, the body's insulin receptors are defective. **Outline**:

a what you understand by 'a receptor'
b where receptors may be found
c what normally follows contact between insulin and receptor.

Foods of animal origin – ethical issues

Issues of meat and meat-products in our diet centre around concerns about our relationships with animals and with animal welfare generally, and about the causes and resolution of world hunger for many, set against the direction in which human diet has naturally evolved.

Mammals that are carnivores (meat eating) or herbivores (plant eating) evolved distinctive, contrasting dentitions. Humans, on the other hand, have evolved a dentition part way between these extremes, with features in common with other **omnivores** – eaters of meat and vegetable matter (Figure 13.20). Our diet naturally includes foods of animal origin.

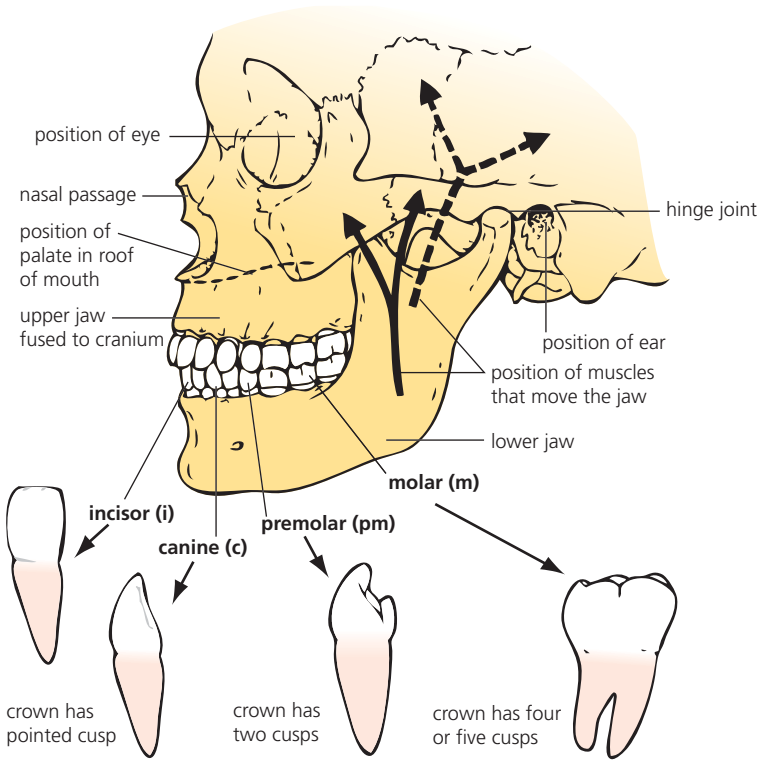
Whether as hunter-gathers, scavengers, or farmers, meat has always been a component of human diet. The earliest humans (few in number) survived in a hostile world, under constant attack from larger and stronger carnivorous hunter mammals. The likely initial human diet might have been a mix of the results of scavenging among the remains of carnivore kills (the extremely nutritious marrow of long bones could be accessed with a heavy stone, for example) and from the proceeds of plant hunting.

Figure 13.20 Human dentition contrasted with carnivore and herbivore dentition

Human dentition
Most animals have **two sets of teeth** in their life-time, a deciduous set followed by a permanent set. The number of teeth of each type in a dentition in one side of the jaw are represented in a dental formula. The dental formulae of humans are:

Deciduous set: ('milk teeth') $i\ \frac{2}{2}\ c\ \frac{1}{1}\ m\ \frac{2}{2}$

Permanent set: $i\ \frac{2}{2}\ c\ \frac{1}{1}\ pm\ \frac{2}{2}\ m\ \frac{3}{3}$



Sheep dentition
If you are able to, examine a prepared sheep's skull (or that of a similar herbivore) in your biology laboratory.

Dentition of sheep:

$i\ \frac{0}{3}\ c\ \frac{0}{1}\ pm\ \frac{3}{3}\ m\ \frac{3}{3}$

- incisors and canine of lower jaw bite against horny pad in the upper jaw
- premolars and molars form a grinding mill
- the surfaces of the teeth wear to form sharp ridges
- during chewing the jaw moves sideways

Dog dentition
If you are able to, examine a prepared dog's skull (or that of a similar carnivore) in your biology laboratory.

Dentition of dog:

$i\ \frac{3}{3}\ c\ \frac{1}{1}\ pm\ \frac{4}{4}\ m\ \frac{2}{3}$

- incisors are sharp, and used to nip, grip and tear
- canine teeth are curved and sharply pointed; used to seize and kill prey
- premolars and molars are used to cut and crush
- some premolars and molars form a shearing unit (like scissor blades)

Then the practical skills of animal and plant breeding made possible the first great revolution in human history, the Neolithic revolution. It started after the last Ice Age, about 10 000 years ago, and quickly, profoundly changed human nutrition. Groups of *Homo sapiens* stopped their nomadic, hunter-gatherer existence and become farmers in settled communities. At this stage, the human population started to grow – a rise made possible by the products of the breeding of domesticated animals and the cultivation of many plant crops, all from the wild stock around. This technology, now called artificial selection, was developed without knowledge of heredity (the transmission of characteristics or traits from one generation to another). Many plant and animal species of modern farming were first produced at this time.

The ethical issue of malnourishment

Today, human population numbers are huge. This places enormous demands on food supplies, and on the resources that generate them. So much so that, around the world, there are local populations of people with too little to eat. World hunger is a major problem of which we are all aware, living as we do in the 'global village'.

In response to the ethical challenges for well-nourished people living alongside other humans who may starve and die prematurely due to malnourishment, some humans have opted to be vegetarian in diet. Vegetarians do not eat meat and most do not eat fish, but the majority consume animal products such as milk, cheese and eggs.

The rationale for this type of vegetarianism is that when we eat meat rather than food of plant origin, we **extend food chains by a least one trophic level**.

At this point you may need to remind yourself of the nature of a food chain, and of the pyramid of biomass (Figure 6.6, page 144). Only about 10% of what is eaten by a consumer is built into the organism's body, and so is potentially available to be transferred on in predation or browsing (or feeding of captive livestock). In effect, vegetarians observe that we waste energy by choosing to eat animal products rather than matter of plant origin, thereby leaving less for others. This is one important issue in considering the eating of animal products.

The ethical issue of our stewardship of animals

Other humans argue that we must not hurt or harm animal life around us at any time. They particularly object to the ways in which animals are reared, treated and slaughtered for our food. They might point out that, apart from taking animal lives to provide meat or fish, we get milk by interrupting the natural life of cows and their calves (at birth, calves are removed and fed with milk substitutes, so that the mother cow can be milked, providing milk for human consumption two or three times a day). In some parts of the developed world (but not in the EU) certain natural hormones in lactating cows are augmented to increase milk yields. The cow may be treated as little more than a milk-production factory in these circumstances. Meanwhile, calves may be fast-reared for veal, sometimes under conditions more akin to industry than traditional farming. In another scenario, hens may be intensively reared, sometimes in battery cages, for egg laying (Figure 13.21). Other examples are flagged by the Compassion in World Farming Trust, in their literature and videos (www.ciwf.org).

Figure 13.21 Examples of intensive farming practices

top right: single crops are grown in large fields, in 'monoculture' conditions

bottom right: damage to the crops by pests is prevented or minimised by the application of pesticides

egg production by hens in battery cages under automated environmental conditions

The range of products we may take from animals in our search for food is quite extensive (Table 13.8). One concerned response to this situation has been that of **vegans – people who choose to eat no foods of animal origin** at all.

The components of a balanced diet, together with all essential nutrients, can be obtained from a diet composed entirely of plant foods, with the exception of vitamin B_{12} (which can be obtained from several manufactured dietary supplement preparations, including yeast extract). Clearly, humans can survive and maintain a good quality of life without taking or interrupting animal lives, should they choose to. Of course, such a view places a different importance on the lives of plants – presumably because of the absence of nerves and sense organs that plants show.

Table 13.8 Foods of animal origin

Animals	Products
cows, sheep, pigs, domestic hens (and other birds), goats, horses (and other mammals)	meats – various forms of muscle tissue and offal (organs such as liver, kidney)
cows, sheep, goats	milk, liquid or manufactured into cream, cheeses, yoghurts etc.
domestic hens (and other birds)	eggs
marine and fresh-water fish (vertebrates)	'meat' – muscle tissue or other organs (e.g. 'roes')
shellfish (marine non-vertebrates)	'meat' – muscle tissue or whole organism
snails (non-vertebrates)	'meat' – muscle tissue or whole organism
bees (insects)	honey (pollen and nectar from flowers, collected and mixed with bee saliva)

Reducing dietary cholesterol

We have seen that cholesterol, a type of lipid which is taken in from foods rich in saturated fatty acids, is essential for normal, healthy metabolism. Cholesterol is also made in the liver; on a low-cholesterol diet the liver of adults forms about 800 mg of cholesterol each day.

Lipids are insoluble in water and have to be transported about the body in the plasma, associated with proteins. Cholesterol is transported in this way, in a form known as low-density lipoprotein (LDL). Excess blood cholesterol, present as LDL, may cause atherosclerosis – the progressive degeneration of the artery walls. Healthy arteries have pale, smooth linings; in unhealthy arteries, the walls have strands of yellow fat deposited under the endothelium (Figure 13.22). With the fatty streaks are laid down fibrous tissue as well, and the whole structure, known as plaque, starts to impede blood flow. Blood pressure is raised. Progressive reduction of the blood flow to the heart muscle impairs oxygenation of this organ, leading to chest pains, known as angina, usually brought on by physical exertions.

Where the smooth lining actually breaks down, the blood is exposed to the fatty, fibrous deposits. Blood platelets collect at the exposed roughened surface, releasing factors that trigger blood clotting (page 349). A blood clot within the vessel may form. It is known as a **thrombus**, until it (eventually) breaks free and circulates in the blood stream, whereupon it is called an **embolus**.

An embolus may be swept into a small artery or arteriole which is narrower than the diameter of the clot, and so cause a blockage. Immediately, the blood supply to the tissue downstream of the blockage is deprived of oxygen. Without oxygen, the tissue dies. The arteries supplying the heart, the coronary arteries, are especially vulnerable. When heart muscle dies, the heart ceases to be a fully effective pump – a heart attack has occurred (known as a myocardial infarction).

When an embolus blocks an artery in the brain a **stroke** occurs. Neurones of the brain depend on a continuous supply of blood for oxygen and glucose. Within a few minutes of the blood supply being lost, the neurones die. Neurones cannot be replaced, so the result of the blockage in the brain is a loss of some body functions controlled by that region of the brain.

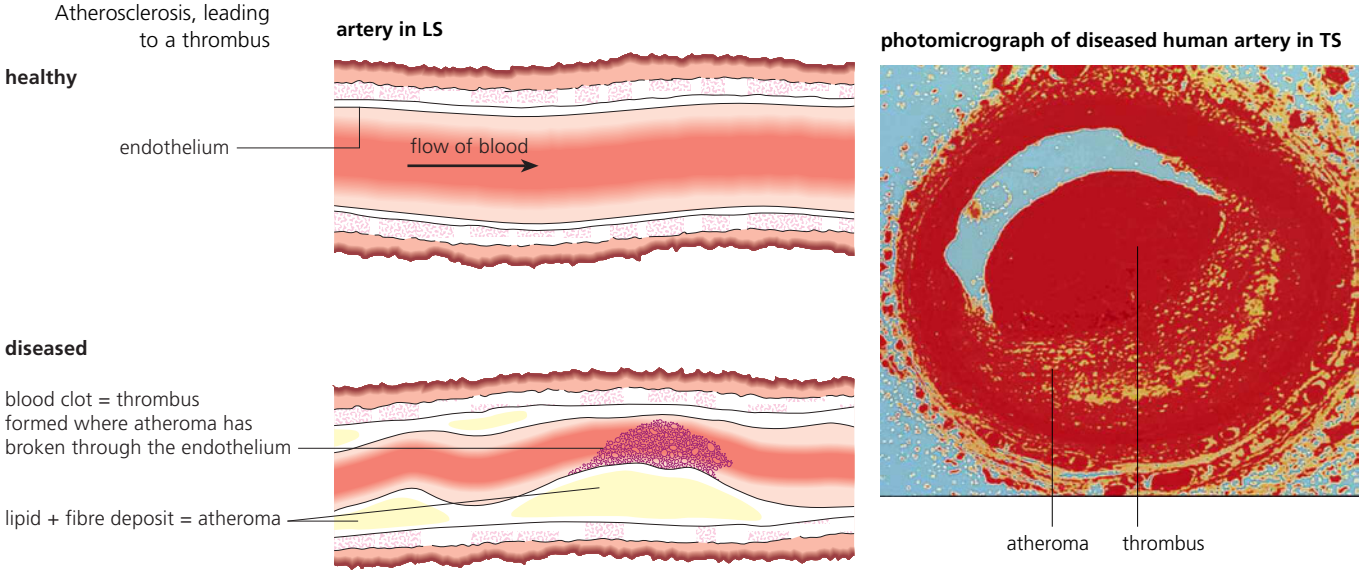


Figure 13.22 Atherosclerosis, leading to a thrombus

Finally, there is another danger from a persistent excess of blood cholesterol. In arteries where the wall has been weakened by atherosclerosis, the remaining layers may be stretched and bulge under the pressure of the blood pulses. Ballooning of the wall like this is called an aneurysm. An **aneurysm** may burst at any time. Clearly the benefits of maintaining a low-cholesterol diet are the lowered risks of dangerous vascular accidents.

10 Deduce the advantages to the body of the tendency of platelets to release clotting factors at the site of damaged artery walls.

Extension: Statins and a revolution in the treatment of heart disease

In 1970, a scientist working for a Tokyo pharmaceutical firm discovered a substance that lowered the level of blood cholesterol. He had been searching systematically for natural fungal products that are capable of changing the rate of particular chemical reactions within our bodies. He had chosen fungi as his source because of the earlier discovery of penicillin – an apparently irrelevant molecule to the fungus *Pencillium notatum*, but lethal to certain invading bacteria when we ingest it as a medicine.

The new fungal product, discovered in Tokyo, turned out to be a specific inhibitor of an enzyme central to the pathway of cholesterol synthesis in the liver. This molecule was a forerunner of a series of similar compounds, now called statins. Statins have revolutionised the prevention of heart disease (heart attacks and strokes).

Food miles – the issues

'Food miles' is the term coined to represent the average distances travelled by food items between the sites of production and the point of consumption – that is from farm to plate. It is a highly significant and growing issue today in the developed world because of changes in food retailing and in our diet, changes in work and leisure practices, and because of developing environmental problems. These are:

- current approaches to the home and garden and leisure activities, including the abandonment of home vegetable production in developed countries;
- a preference for food items to be available all the year round rather than merely when in season locally;

- frequent selection of pre-packaged and manufactured foods, including frozen foods, rather than traditional alternatives;
- centralisation of retailing in large supermarkets that operate in a highly competitive market, and which themselves centralise their processing and preparation of foods;
- positioning of supermarkets at some distance from town centres, often difficult to access by public transport;
- the practice of restricting food shopping to once a week, involving quantities that require the private car to carry it, and the use of modern home storage facilities such as freezer and refrigerator;
- people preferentially living in cities and towns and their home waste materials (including food waste and packaging) having to be transported some distance for appropriate disposal;
- the growing concern about climate change and the impacts of road and air transport on atmospheric carbon dioxide levels.

As a consequence, although retailing of food in the developed world is often conducted in an assertively advertised atmosphere of competitive price-cutting and an emphasis on cheapness, there are **major hidden costs** for society.

In the UK, food miles are mounting. A new report by the government's Department for the Environment, Food and Rural Affairs (Defra) has calculated that food miles rose by 15% between 1992 and 2002. Agriculture and food account for nearly 30% of the goods transported on the roads.

Why does food travel so far?

One reason is that, because of competitive labour costs, fish caught in UK waters is sometimes despatched to China to be processed and packaged (because labour costs are much lower in China) before being transported back to the UK. This is an extreme example, but local products such as milk and potatoes are now usually carried to centralised packaging depots before being distributed, including back to the local communities where they were originally produced. Food processing, too, may require items to be moved several times between industrial centres, before being distributed from central facilities. Today, consumer choices and a preference for a more varied diet in many north European countries has contributed to increases in food miles; in the UK, typically 95% of fruit is imported and 50% of vegetables, too.

Why do food miles matter?

In the UK, for example, food transport accounted for an estimated 30 billion vehicle kilometres in 2002. Food transport produced 19 million tonnes of carbon dioxide in 2002, but transport of food by air has the highest carbon dioxide emissions per tonne and is the fastest growing mode. The direct environmental, social and economic costs of food transport are over £9 billion each year and are dominated by congestion.

Additionally, the transport of animals is an important animal welfare issue. Today, live animals are transported often for great distances from farms to large, centralised abattoirs and meat-processing plants. They are also exported and imported to and from other countries.

Finally, freshly picked crops and locally slaughtered animals are most likely to have higher food values; for the consumer there is a question of quality.

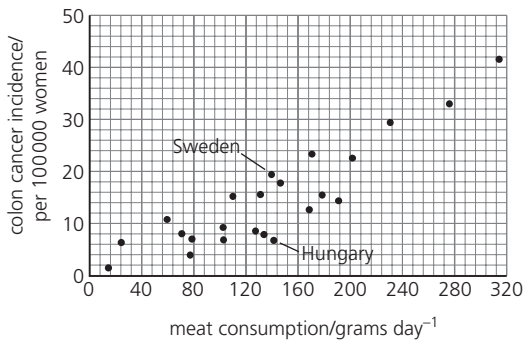
What can be done?

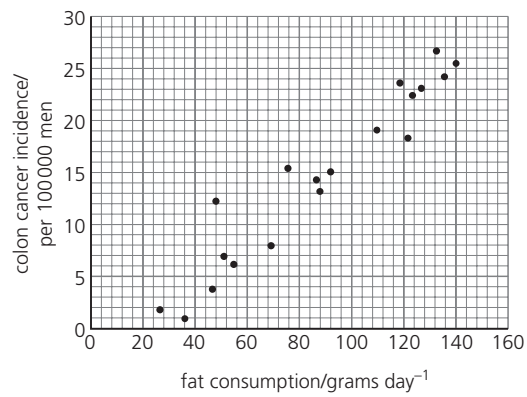
There are campaigns to get local food used where ever possible, including in schools and hospitals. Also people are encouraged to buy locally, by the development of local farmers' markets. Associated with campaigns to tackle obesity is encouragement to cook items at home, with fresh ingredients, thus avoiding the additives (such as very high salt levels) often employed in manufactured and processed foods to prevent bacterial contamination. People are encouraged to buy food with as little packaging as possible, and to compost waste in their own garden where possible. Another way individuals can have influence is to buy fruit and vegetable items only when they are in season locally.

Examination questions – a selection

Questions 1–3 are taken from past IB Diploma biology papers.

Q1 Dietary factors are known to influence the incidence of colon cancer. The graphs below show the correlation between meat consumption (eating meat) and colon cancer in sample countries and the correlation between fat consumption and colon cancer.





a i State the relationship between daily meat consumption and the incidence of colon cancer in women. (1)

ii Using the data in the two graphs, suggest reasons for the relationship between daily meat consumption and the incidence of colon cancer in women. (2)

b i Calculate the difference in colon cancer incidence between Hungarian women and Swedish women. (1)

ii Discuss whether meat consumption causes colon cancer in Sweden and Hungary. (3)

Standard Level Paper 3, May 06, QA1

Q2 **a** State **two** main functions of carbohydrates absorbed from food. (2)

b Suggest how environmental conditions cause malnutrition. (2)

Standard Level Paper 3, November 04, QA2

Q3 **a** Outline the fate of the products of ingested proteins. (2)

b Discuss whether meat is essential in the human diet. (4)

Standard Level Paper 3, November 02, QA2

Questions 4–10 cover other syllabus issues in this chapter.

Q4 Diabetes is a complex disease condition. Describe the various causes of diabetes, types I and II, and outline how they may be treated to maintain health. (6)

Q5 Explain the differences between the idea of a 'balanced diet' and the term 'essential nutrients' by reference to the range of different food materials required for good health in humans. (6)

Q6 Describe the consequences for body structure and function of protein deficiency malnutrition:

a at an initially mild level

b as a result of a severe and persistent level of deficiency. (8)

Q7 From your understanding of body structure and functions (support and movement, nervous communication, blood and internal transport, for example), outline the roles played by **four** minerals essential to the maintenance of the working body. (4)

Q8 Discuss how medical observations and experiments have contributed to our knowledge of the roles of vitamins such as vitamins C and D in health and growth. (6)

Q9 It is apparent that indigestible matter in food is a significant component of diet. Suggest possible roles that 'roughage' may have in the maintenance of health. (4)

Q10 Explain how negative feedback operates in the control of human appetite for food. (6)

Option B

14 Physiology of exercise

STARTING POINTS

- In multicellular organisms most **cells** are adapted to perform **specialised functions**, and are organised into **tissues and organs**.
- Organisms need usable **energy** to maintain cells and carry out their **activities and functions**. **Respiration** is the process that transfers energy in living cells.
- When cells are respiring aerobically, **oxygen** is taken in and **carbon dioxide** is given out. Mammals have specialised respiratory surfaces, the **lungs**, where **gaseous exchange** occurs.
- In a large, multicellular organism an **internal transport system** is required. In mammals, internal transport is by a system of vessels through which **blood** is pumped by the **heart**, and circulated to all parts of the body.
- In the contexts of **exercise**, **training** and **fitness**, this chapter extends study of the specialisation of cells, tissues and organs, of the process of respiration in muscles to sustain movement, and the functioning of lungs and the circulatory system, begun in chapter 1 (pages 1–36), chapter 3 (pages 76–90) and chapter 7 (pages 178–234).

The support system of an animal's body is its **skeleton**. This helps maintain the distinctive form and shape of the animal by which it is recognised. Most animal skeletons are quite rigid structures which also provide some protection to certain internal tissues and organs. However, the chief role of the animal skeleton is to allow **locomotion**.

Mammals and other vertebrates have an internal skeleton (an **endoskeleton**) of many component parts, mostly made of **bone**. Attached to bones are **muscles**, and between many of the bones are **joints**. Muscles, bones and movable joints combine as **systems of levers**, allowing parts of the body to be pushed down on or against the environment, to bring about **movement**.

In this chapter, our body's **apparatus for locomotion**, skeletal and muscular, is examined first, and the mechanism of **muscle contraction** explained. Then the effect of **sports training** on both **pulmonary** and **cardiovascular** systems is discussed, and the **physiology and biochemistry of energy transfer** for movement are examined. Finally, issues of **fitness**, **training** and **sports injury** are considered.

Muscles and movement

B1.1–1.8

The content of this section is that of IB Topic 11 (11.2.1–11.2.8) and is addressed in Chapter 12, 'Muscles and movement', pages 361–9.

Training and the pulmonary system

B2.1–2.3

Training for successful performance can improve and enhance fitness levels. This section looks into how we study and analyse the working of the human pulmonary system and how exercise may improve performance of the thorax and lungs.

Our pulmonary system consists of the **lungs** housed in the chest or **thorax**, a chamber formed by the rib-cage and its muscles, with a domed floor, the diaphragm.

The lungs are organs that connect with the pharynx at the rear of the mouth by means of a tube called the **trachea**. The trachea divides into two **bronchi**, and within the lungs the bronchi divide into smaller **bronchioles**.

The bulk of the lung tissue consists of tiny air sacs or **alveoli**, arranged in clusters, each served by a tiny bronchiole. Here, gaseous exchange occurs.

A vast **capillary system** wraps around the clusters of alveoli. Each capillary is connected to a branch of the **pulmonary artery** and drained by a branch of the **pulmonary vein**. Blood in the capillaries becomes oxygenated by the lungs, and it is then returned, via the pulmonary veins, to the heart, to be pumped to the rest of the body.

The structure of the thorax is illustrated in Figure 7.26 (page 205) and the site of gaseous exchange in Figure 7.29 (page 208).

Ventilation of the lungs – breathing

Air is drawn into the alveoli when the air pressure in the lungs is lower than atmospheric pressure, and it is forced out when pressure in the lungs is higher than atmospheric pressure. Since the thorax is an airtight chamber, pressure changes in the lungs occur when the volume of the thorax changes, due to the alternate relaxation and contraction of the diaphragm and intercostal muscles.

In the pulmonary system, air flow is tidal – air enters and leaves by the same route. Consequently there is a residual volume of air that cannot be expelled because the system is closed at one end. Incoming air mixes with and dilutes the residual air, rather than replacing it.

The ventilation mechanism of the lungs is detailed in Figure 7.28 (page 207).

Investigating human breathing

The changes in lung volume during breathing are investigated using an apparatus called a **spirometer** (Figure 14.1). This consists of a Perspex lid enclosing the spirometer chamber, hinged over a tank of water. This chamber is connected to the person taking part in the experiment via an interchangeable mouthpiece and flexible tubing. As breathing proceeds the lid rises and falls as the chamber volume changes. With the spirometer chamber filled with air, the capacity of the lungs when breathing at different rates can be investigated. Incidentally, if the spirometer chamber is filled with oxygen, and a carbon-dioxide-absorbing chemical such as soda lime is added to a compartment on the air return circuit, this apparatus can be used to measure oxygen consumption by the body, too.

From 'traces' printed out from investigations of human breathing under different conditions we have a picture of how extensively breathing may vary (Figure 14.2).

Examine the lung trace now.

Notice that the volume of air breathed in and out during normal, relaxed, rhythmical breathing, the **tidal volume**, is typically 400–500 cm^3.

The tidal volume is the volume of air taken in and out with each inhalation or exhalation.

However, we have the potential for an extra large intake (maximum inspiratory capacity) and an extra large expiration of air (expiratory reserve volume), when required. These, together make up the vital capacity of our lungs – about 4.5 dm^3.

The vital capacity is the volume of air that can be exhaled after a maximum inhalation.

When the residual volume is added to the vital capacity, we arrive at the total capacity of our lungs, typically about 6 dm^3.

The total lung capacity is the volume of air in the lungs after a maximum inhalation.

The spirometer can also be used to investigate steady breathing over a short period of time, and in these cases the *y*-axis of the spirometer trace represents time. In this way, the rate at which the lungs are ventilated can be investigated.

The ventilation rate is the number of inhalations or exhalations per minute.

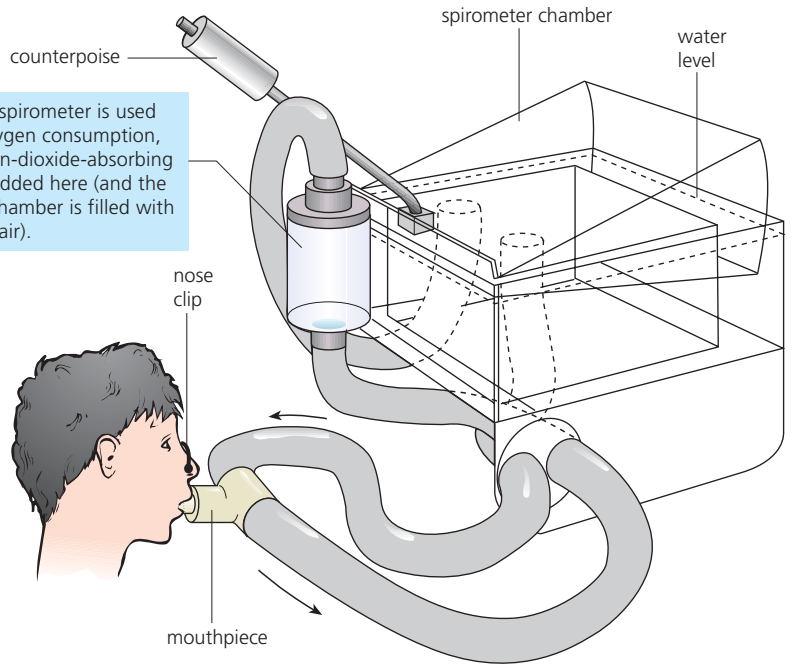
The effects of exercise on breathing

Physical activity affects oxygen consumption and carbon dioxide production more than any other physiological stress the body experiences.

During exercise, the demand for oxygen by the working muscles increases (and carbon dioxide output rises), and as a result, both the **tidal volume** and the **ventilation rate** increase and so maintain an adequate gas concentration in the lungs for rapid gaseous exchange. However, the pattern of change is quite complex, and it varies with the level of exercise (Figure 14.3).

Figure 14.1 Investigating breathing with a spirometer

A recording spirometer is used to analyse the pattern of change in lung volume during breathing.



The movements of the lid of the airtight spirometer chamber are recorded by a position transducer, connecting box and computer. The results (e.g. inspiratory capacity and tidal volume) are printed out using appropriate control software.

Figure 14.2 Lung volumes shown by a spirometer trace

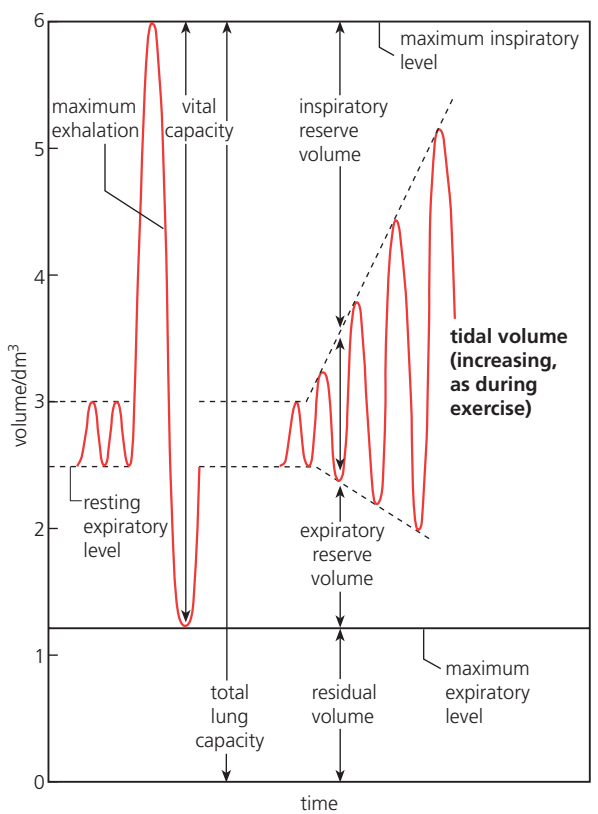
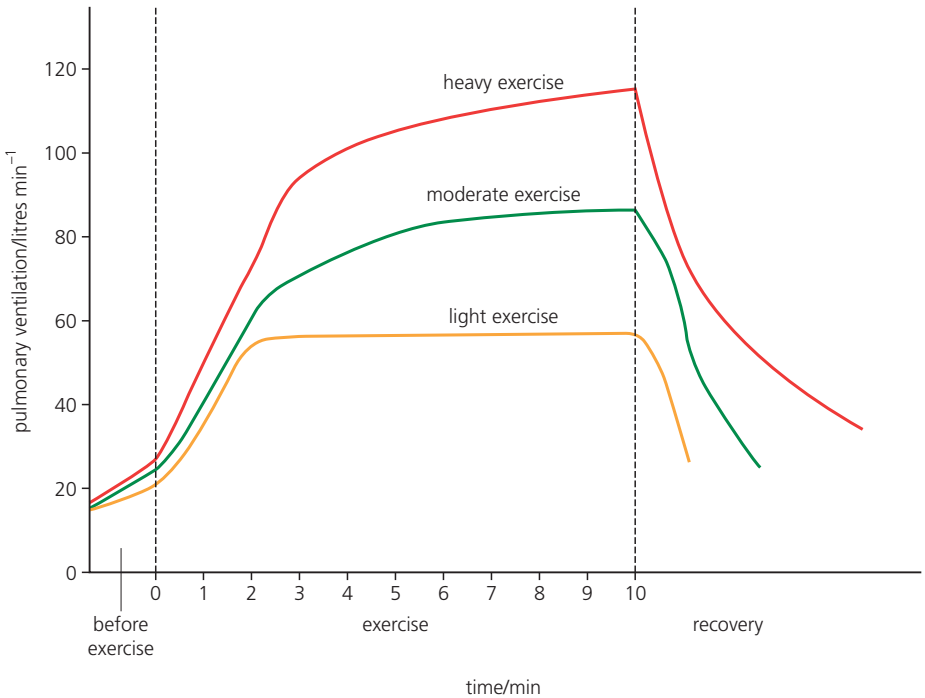


Figure 14.3 Respiratory response to exercise

Note: increases in pulmonary ventilation are brought about by a marked increase in ventilation rates and an increase in tidal volume.



To start with, **planned physical exercise** is preceded by a slight increase in ventilation. This is an anticipatory rise, no doubt due to the effect of the hormone adrenaline on the respiratory centre that controls breathing. It happens because we are planning the activity, and are preparing for it.

When exercise commences, there is a rapid rise in ventilation under nervous control. However, in physical activity that is less than the maximum the body can achieve (**sub-maximal exercise**), the sudden increase in ventilation begins to slow and normally reaches a plateau or steady state. The oxygen reaching the muscles is now sufficient to support a level of respiration that transfers sufficient energy to the contracting muscle fibres.

During **maximal exercise**, on the other hand, the steady state does not occur, and ventilation continues to increase until exercise is complete.

The respiratory centre in the brain is sensitive to the concentration of carbon dioxide in the blood and responds to rising levels by increasing ventilation. Also, when the necessary energy cannot be obtained by aerobic respiration alone during continuing maximal exercise, anaerobic respiration in muscles produces lactic acid. The respiratory centre is also stimulated by the level of blood lactate.

The effects of training on the pulmonary system

A successful training programme should be **relevant and appropriate** to the sport for which the individual is training, so that the same muscle groups are exercised in training as are used for the sport activity, and under similar internal physiological conditions to those the sport activity generates.

The training aim for the pulmonary system is to achieve long-term, favourable physiological adaptation. In this way, the body is better able to cope with the stresses the chosen sport activity will impose during performances. For example, several weeks of appropriate training during sub-maximal exercise has been found to reduce oxygen consumption by the muscles of the pulmonary system (intercostal muscles and diaphragm muscles). This enhances endurance because it reduces the fatigue effect of exercise on the ventilatory muscles, and it allows oxygen freed from use by the ventilatory muscles to be available to the locomotory muscles.

What are the mechanical effects of training on the pulmonary system, and what are the consequences?

Studies with both adolescents and adults involved with sub-maximal exercise training for sport activity have established that their tidal volume increases and ventilation rate decreases during their training exercise regime. Consequently, air remains in their lungs longer between breaths.

The result is that the exhaled air from trained people during exercise contains only 14–15% oxygen, whereas that from untrained people under identical conditions contains an average of 18% oxygen. Untrained people ventilate more frequently to achieve the same oxygen uptake, in effect.

1 **Construct** a table to summarise the effects of training on the ventilation rate at rest, the maximum ventilation rate, and the vital capacity.

At the same time, the effect of training on lung volume is slight or non-existent; vital capacity remains largely unchanged. On the other hand, under moderate or heavy exercise the maximum ventilation rate is increased – possibly from about 40 to 45 breaths per minute.

Clearly, a trained pulmonary system contributes to physical performance, and improves conditions for the cardiovascular system. We examine the effects of training on the cardiovascular system next.

Training and the cardiovascular system

B3.1–3.5

Blood is the transport system of our body. The composition of the blood and the system of tubes (blood vessels) by which it is circulated around the body are introduced in Chapter 7 (page 185). The specific roles of the individual components of the blood circulation system are summarised in Tables 7.5 and 7.6.

Mammals have a **double circulation** in which the right side of the heart pumps deoxygenated blood to the lungs. The arteries, veins and capillaries serving the lungs are known as the **pulmonary circulation**. At the same time, the left side of the heart pumps oxygenated blood to the rest of the body. The arteries, veins and capillaries serving the body are known as the **systemic circulation**. The structure of the heart, a four-chambered organ with left and right sides separate but operating in tandem, is shown in Figure 7.12. The general layout of the blood circulation is summarised in Figure 7.11 (page 189).

We shall now examine the pumping action of the heart in more detail, and compare the effects of physical activity and periods of rest on the circulation of the blood.

Figure 14.4 A natural marathon runner has a low resting heart rate – training for endurance events further improves heart performance

The pumping heart at work

We have noted that the basic rate of the heart beat is controlled by the pacemaker (page 192) which generates an intrinsic rhythm of about 50 beats per minute (bpm). However, the working of the heart is governed by events happening elsewhere in the body. Nervous, hormonal and other factors override the basic rhythm because the body's demands on the circulatory system change. As a result, the **heart rate** is constantly adjusted – it may change to values between about 50 and 200 bpm.

The heart rate is the number of contractions of the heart per minute.

The average resting heart rate is 72 bpm. Incidentally, a low resting heart rate may indicate a high level of endurance fitness. Some highly trained endurance athletes, such as those who make a success of the marathon, have even been reported to have a heart rate as low as about 30 bpm, an exceptional figure (Figure 14.4).

2 **Suggest** why it is advantageous for an athlete to have a low heart rate when at rest.

Each time the heart beats, a volume of blood is ejected from the ventricles – the left side drives blood out through the aorta, and the right side through the pulmonary arteries. This volume is known as the **stroke volume**, and it is the difference between the volume of blood present before and after ventricle contraction.

The stroke volume is the volume of blood pumped out with each contraction of the heart.

The amount of blood flowing from the heart each minute is known as the **cardiac output**.

The cardiac output is the volume of blood pumped out by the heart per minute.

The cardiac output is generally measured from the left ventricle, and is equal to the product of the stroke volume and heart rate. So, for example, the average sedentary man at rest may have a stroke volume of about 71.4 cm^3 and a heart rate of 70 bpm. His cardiac output is calculated as follows.

stroke volume × heart rate = cardiac output
71.4 cm^3 × 70 bpm = 4.998 cm^3

In other words, the cardiac output in this case is close to 5 litres. This is approximately the total blood volume of a typical adult male. We see immediately that the **entire volume of our blood** flows through the pulmonary and systemic circulations each minute, *while physically inactive*.

Shortly, we shall look into the effect of exercise on circulation. Obviously the cardiac output is changed by physical activity – we can feel out hearts beating faster, for one thing!

Stroke volume and cardiac output for women average about 25% below the values for men. Here, the average stroke volume is 50–60 cm^3. This gender difference is related to the body size of women, which is generally smaller than that of men.

Finally, we must note that when the body is at rest, about 70% of our total volume of blood is in the veins. This represents a large reservoir of blood which is returning to the heart. Since the heart can only pump out the quantity of blood it receives, the cardiac output is dependent on **venous return**.

The venous return is the volume of blood returning to the heart via the veins per minute.

Any rapid increase in venous return will enable a significant increase in cardiac output.

Next, we must examine the effects of exercise.

Immediate effects of exercise on cardiac output and venous return

During periods of persistent exercise, particularly if the exercise is strenuous, the amount of oxygen in the blood (the oxygen tension) falls markedly. This is because the demand for oxygen in active muscles, with their raised level of respiration, causes removal of oxygen from the blood faster than it is being loaded in the alveoli of the lungs. Similarly, the amount of dissolved carbon dioxide (the carbon dioxide tension) is raised. Since carbon dioxide is a slightly acidic gas, the pH of the blood falls slightly at these times.

We saw earlier that the control centre for the heart is located in the hindbrain, in a structure called the **medulla** (Figure 7.14, page 192). As the blood circulates through capillaries in this cardiovascular control centre, the changing composition of the blood due to strenuous exercise – particularly the slightly lowered pH – is detected. As a consequence, impulses that stimulate heart rate are transmitted to the pacemaker. Cardiac output is immediately raised, as a result.

During strenuous exercise, the skeletal muscles are extremely active. For example, antagonistic pairs of muscles alternately contract and relax, and, among other things, these movements keep up a constant alternating squeezing and relaxation on the veins of the limbs. We have seen that pressure from surrounding tissues is an important factor in moving blood in veins (Figure 7.10, page 188). Since the many valves found in veins prevent backflow, strenuous activity significantly increases the return of the blood to the heart. So, the volume of blood returning, the venous return, is increased by the body's activity.

3 **Suggest** why the changes in pH of the blood are only slight changes.

The effects of training on the heart at rest and during activity

During periods of physical activity both cardiac output and the volume of venous return are increased.

With regular periods of training, what are the long-term effects?

Two important changes are brought about. First, because cardiac muscles are slightly elastic, the walls of the chambers of the heart (particularly the ventricle walls) are stretched during each cardiac cycle (Figure 7.13, page 191), when the ventricles are filling. The cardiac muscle walls are also progressively thickened with training. Thus, athletes develop larger ventricle cavities and thicker ventricle walls. The heart of an athlete is larger than that of a non-athlete. This phenomenon is called **cardiac hypertrophy**.

Secondly, cardiac hypertrophy is accompanied by a decreased resting heart rate. Note that there are two independent nerves running from the **cardiac control centre** to the **pacemaker** (Figure 7.14). During periods of training, the pacemaker in the athlete's heart is under greater influence (more impulses) from the nerve from the control centre that slows heart rate (known as the parasympathetic nerve). At the same time, the nerve that causes the heart to beat faster (known as the sympathetic nerve) has less influence (fewer impulses). As a result, with persistent and regular strenuous physical training activity, the heart rate is steadily lowered in the resting heart. This is called **bradycardia**.

What are the consequences of these changes?

When the trained athlete is at rest, the increased ventricular cavities allow the heart to fill with more blood than it does in an untrained person so when contraction occurs, the strengthened cardiac muscle (compared to the non-athlete's heart), causes a substantially increased stroke volume. The heart of a trained athlete delivers the same cardiac output as that of an untrained person, but with fewer beats per minute.

When the trained athlete undertakes strenuous activity, the heart rate rises rapidly in response to the changing needs of the body, particularly of the skeletal muscles. Under these conditions, the raised heart rate will be similar, whether or not a person is trained. However, the trained athlete has an increased stroke volume and so the cardiac output of the athlete during continued training or performance is substantially higher than that of an untrained person.

These changes are summarised in Table 14.1.

	Heart rate	Stroke volume	Cardiac output
Athlete at rest	slower than that of an untrained person – typically 50 bpm (rather than 70 bpm)	greater than that of an untrained person – typically 100 cm^3 (rather than 71 cm^3)	about 5 litres per minute whether or not a person is trained
Athlete during exercise	rises to 195 bpm (whether or not a person is trained)	rises, typically to 179 cm^3 (rather than 113 cm^3 in an untrained person)	about 35 litres per minute (compared to 22 litres per minute in an untrained person)

Table 14.1 Effects of training on the heart at rest and during exercise

The redistribution of blood during exercise

It has been calculated that if all the blood vessels of the body were fully dilated at any time, they would hold about 20 litres of blood. Actually, our blood circulation holds a mere 5–6 litres in males (4–5 litres in females) but this is not a problem for the circulation system.

Why not?

Circulation to various organs and organ systems is finely regulated by the mechanism of localised **vasodilation** and **vasoconstriction**. The capacity of particular organs such as the skin, gut and skeletal muscles is varied according to particular circumstances (Figure 14.5). For example, we have already seen that when the body experiences intense cold and there is a danger the core temperature of the body may fall, the arteries and arterioles that supply the skin are severely constricted, thus cutting heat loss by radiation and conduction from the body surface (we look extremely pale when we are cold, Figure 7.43, page 221).

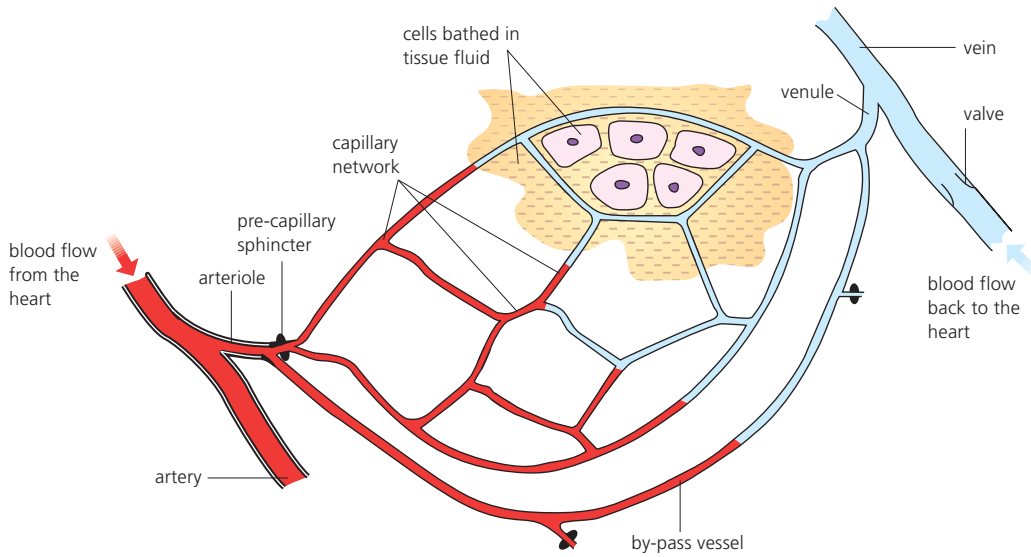


Figure 14.5 The 'by-pass' principle in blood redistribution

Tissues are 'by-passed' by vasoconstriction of the capillaries and by contraction of pre-capillary sphincter muscles under involuntary (unconscious) control of nerves conducting impulses from centres in the brain.

However, not all parts of the circulation system are equally variable. For example, the supply to the brain is unvarying, with the most obvious advantages for our survival in all situations. Similarly, the supply to the lungs is never permitted to fall, and the supply to the kidneys remains about the same, too, in a quite wide range of circumstances, although under certain conditions it is reduced, as we shall shortly see.

When vigorous exercise commences, the blood supply to the skeletal muscles is greatly increased immediately, mostly at the expense of that to the gut and, to a lesser extent, to the kidneys. Table 14.2 details the changes in blood volumes to organs under conditions of maximum physical effort (expressed as a percentage of the total volume of blood circulating per minute).

Organ	Blood flow at rest		Blood flow under vigorous exercise	
	cm^3 per minute	%	cm^3 per minute	%
brain	750	15	750	2.5
skeletal muscles	1 000	20	26 000	88
liver and gut	1 250	25	375	1.25
skin	500	10	750	2.5
coronary vessels	250	5	1 200	4
kidneys	1 000	20	300	1
whole body (total)	5 000	100	30 000	100

Table 14.2 The pattern of redistribution of the blood circulation during activity

4 **Suggest** and comment on the best advice to give to people planning to swim shortly after eating lightly, and after dining at a banquet!

Special measures to improve sport performance

The oath sworn on behalf of Olympic athletes commits all competitors to participation 'without doping and without drugs in the true spirit of sportsmanship'.

Unfortunately, not all sports persons keep this promise; there are drug scandals in competitive sports from time to time, as we are all aware.

As a result, there are accredited laboratories that test body fluid samples, randomly collected. They test for the presence of chemicals known to enhance physical performance in competitive sports, directly or indirectly, and which have been banned. **Banned substances** include androgenic anabolic steroids, amphetamines and related stimulants, and caffeine, among a host of substances that are considered performance enhancing. We will return to this issue later.

Blood doping

Another development in the quest for enhanced performance is known as blood doping, or **red blood cell infusion**.

This technique involves withdrawing samples of the sports person's blood (450–1800 cm^3 of blood) and placing it in frozen storage for future return to the athlete's body, normally within a week or less of the competitive event. In the meantime, the lost blood volume is automatically made good – the body withdraws extra red cells from store within the bone marrow to return to normal red blood cell count. When the source of red cells is added back, prior to a competitive event, the added blood volume contributes to a raised cardiac output while the additional red blood cells increase the oxygen-carrying capacity of the blood.

There is a significant performance benefit detected in some cases, but there are risks as well. For example, any increase in blood viscosity may compromise the flow of blood through the smaller capillaries including in the heart muscles, and so there is an enhanced risk of heart attack or stroke.

Blood doping is illegal under committee rules.

Hormonal blood boosting

To achieve a similar effect to blood doping without the associated lengthy and complex procedures, some athletes are using a synthetic form of the natural hormone **erythropoietin** (**EPO**).

EPO is normally produced by the kidneys and regulates red blood cell production within the bone marrow. A 12% increase in haemoglobin typically follows a six-week period of pre-performance treatment with EPO.

However, the corresponding increase in blood viscosity that occurs is potentially life-threatening. There is increased likelihood of stroke, heart attack, or pulmonary oedema (swelling due to accumulation of tissue fluid). For example, it has been alleged that 18 deaths (attributed to heart attacks) among competitors in gruelling cycle races have been associated with EPO-pretreatments.

EPO cannot be detected in the urine so suspected abuse is tested for by looking for abnormally high red blood cells counts.

No doubt all these measures and moves to enhance physical achievement will be the subject of continuing experiment, publicity and comment by sports regulating bodies in the future.

Exercise and respiration

B4.1–4.7

We have seen that for muscle contraction to occur, a supply of ATP is required by the sarcomeres of the myofibrils. To sustain continuing muscle contraction, the supply of ATP needs to be continuous, because the reserves of ATP in the tissue are finite and a great deal is used up (and ADP and P_i formed) in contraction movements of the thin filaments.

Between the myofibrils are packed many mitochondria, the organelles in which ATP production occurs during aerobic respiration. The mitochondria continually need an adequate supply of oxygen and the respiratory substrates, glucose and fatty acids. Although the blood supply to muscle is the main source, muscle tissue also contains reserves of these chemicals, as we shall see shortly.

The uptake and utilisation of oxygen

Molecular oxygen combines with haemoglobin of red cells in the alveoli of the lungs and oxygen is transported to the respiring tissues as oxyhaemoglobin. Arterial blood is relatively rich in oxygen. In the respiring tissues, oxyhaemoglobin dissociates into oxygen and haemoglobin because here the oxygen tension is much lower than in the lungs where the loading of haemoglobin occurred. So blood returning to the heart and lungs contains much less oxygen.

In fact, the difference in oxygen content between arterial and venous blood at any time is a measure of the demand and uptake of oxygen in the respiring tissues. This difference is known as VO_2.

VO_2 = the amount of oxygen being used in the body ($cm^3\ kg^{-1}\ min^{-1}$).

When we start physical exercise, conditions in the respiring muscles begin to change. For one thing, their oxygen consumption will increase. But we have already seen that the breathing rates will change, and the amount of blood circulating to the muscles changes too. Obviously the amount of oxygen in transport and being taken in by muscles increases (VO_2 increases).

With increasingly vigorous exercise, VO_2 will continue to increase. However, the situation will be reached where further increase is impossible, even if the maximum physical effort is maintained. This point is **VO_2max**.

VO_2max is the maximal oxygen uptake by the body ($cm^3\ kg^{-1}\ min^{-1}$).

The muscles hold reserves to sustain respiration

1 **Glycogen** is held in reserve in muscle fibres, available for hydrolysis to glucose when required. Remember, glycogen is an insoluble polysaccharide, a branched polymer of glucose, with a structure rather similar to the form of starch known as amylopectin. In fact, glycogen has been called 'animal starch'. It occurs, stored in most tissues in our bodies (but not in the brain). Fortunately, glycogen is hydrolysed to glucose relatively quickly, by the appropriate enzyme.

2 **Oxygen** reserves are held in muscles, too, associated with a molecule called **myoglobin**, a substance only found in muscle cells in the body. Myoglobin is a respiratory pigment molecule, similar to haemoglobin found in red cells, but it is only one quarter the size. Although smaller, myoglobin has a much higher affinity for oxygen than haemoglobin, so in normal conditions in muscle the myoglobin is saturated with oxygen, and exists as **oxymyoglobin** – effectively a store of oxygen held just alongside where it may be needed in future.

When muscle is very active for a prolonged period, the oxygen concentration eventually falls to the point when oxymyoglobin will dissociate, releasing molecular oxygen for aerobic respiration to continue. In the muscle at rest and relaxed, oxymyoglobin reserves are re-formed.

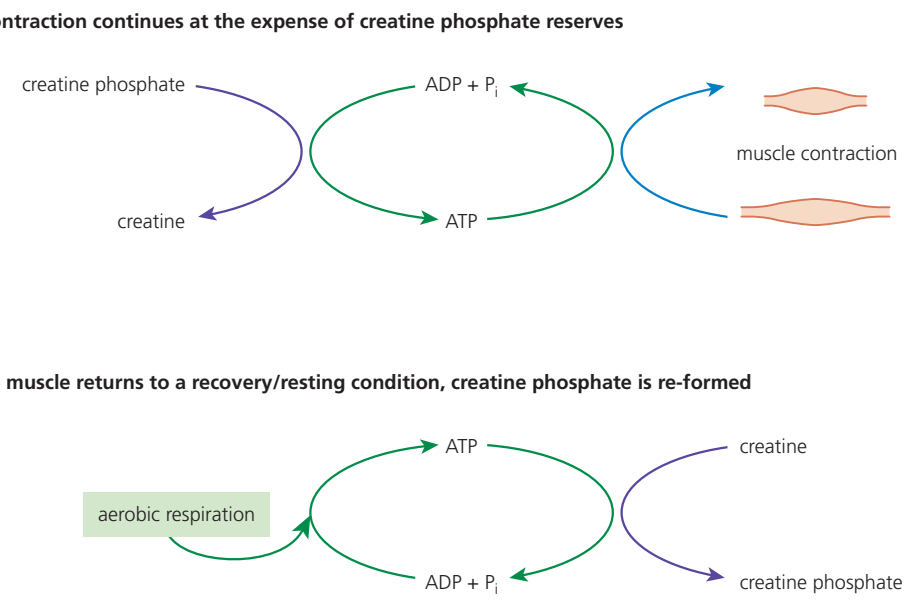


Figure 14.6 Creatine phosphate as an emergency energy store

3 **Creatine phosphate** is effectively a reserve form of ATP, present in muscle tissue. Creatine phosphate is a molecule that can be broken down by the enzyme creatine kinase into creatine and free phosphate. The energy transferred by this reaction is then used to resynthesise ATP from ADP and P_i. This occurs when:

a the reserves of ATP in the muscles are mostly hydrolysed to ADP and P_i;
b ATP cannot be resynthesised in the surrounding mitochondria fast enough to sustain muscle contraction.

This ATP is then used in muscle contraction (Figure 14.6). When the muscle returns to a resting state, creatine phosphate is re-formed from creatine and fresh ATP. Prior to this, reserves of ATP are re-established by aerobic respiration in the mitochondria.

The pattern of ATP production in muscles under varying exercise regimes

In the muscle fibres, mitochondria occur between the microfibrils. When skeletal muscle is at rest, aerobic respiration in the mitochondria generates the ATP that is needed to maintain **muscle tone**. Muscle tone is the degree of muscle contraction required to maintain posture and body position, and for any minor movements of the body that occur. Under these conditions, ATP reserves are maintained at their maximum.

In the event of intense physical exercise commencing abruptly, there is no possibility of a steady build-up in oxygen supply to the muscles. Additional oxygen is needed immediately to facilitate increased aerobic respiration, so that ATP reserves may be speedily replenished.

In this situation, the first reserves to be drawn down are the creatine phosphate reserves. We have seen that this molecule reacts with ADP and forms ATP, which is then used for continuing muscle contraction.

Of course, creatine phosphate reserves are significant, but they are finite too, and they are rapidly exhausted when intense physical exercise is maintained. This occurs in 8–10 seconds, as shown in the graph in Figure 14.9 (page 435).

In this situation, skeletal muscle can continue to obtain ATP **only** by lactic acid fermentation (anaerobic respiration, page 80). This is wasteful of glucose resources (aerobic respiration produces far more ATP from each molecule of glucose respired than is produced by lactic acid fermentation).

Also, the effect of the accumulating lactic acid may ultimately be harmful to the tissues. The body tolerates only limited accumulation of intramuscular H^+ ions before the lowered pH starts to interfere with cell biochemistry. Muscle fatigue follows.

If physical activity continues, but **less intensely** for example, then the percentage of glucose that is respired anaerobically steadily decreases. The rate of oxygen supply to the muscles begins to catch up with demand.

5 **Explain** the roles of:

a myoglobin
b creatine phosphate

in muscle tissue.

Eventually, the oxygen supply is sufficient for aerobic respiration to replace anaerobic respiration completely. Once again, all the ATP needed for muscle contraction is provided by aerobic respiration. Next, the reserves of creatine phosphate are re-formed. The lactate that accumulated is disposed of too, of which more later.

The use of dietary supplements containing creatine phosphate

Skeletal muscle contains the bulk of our body's creatine phosphate. The body manufactures a small amount daily, but the bulk of our intake comes in our diet (in meat including poultry, and fish). Unfortunately, vegetarian diets are low in creatine phosphate. Creatine supplements to diets are available and are not banned by sports governing bodies.

Is taking creatine in its various available forms advantageous to the athlete?

Taking creatine supplements at the recommended levels has been shown to improve muscular strength and augment endurance in short bursts of muscular overload (Figure 14.7). Taking a high dose of creatine may help replenish muscle creatine reserves following intense exercise, enabling athletes to maintain repeated bursts of high-intensity exercise.

Figure 14.7 Commercial sources of creatine popular with athletes

Doses of creatine taken orally by athletes may significantly increase the creatine content of skeletal muscles, and accordingly improve performance in periods of intense physical activity. The only adverse effects reported have been anecdotal evidence of the inducement of cramp during lengthy competition periods. However, only very limited information exists on the long-term effects of creatine supplementation for athletes. Currently we can say that there are clear advantages, but 'the jury is still out'! It does appear this supplement is widely experimented with.

Intense exercise, VO_2 and respiratory substrates

The effects of intense activity on a subject's VO_2 can be illustrated by a hypothetical example of a cyclist with the task, starting from rest, to cycle at a steady speed up a series of progressively steeper hills (Figure 14.8). As the first slope is tackled, the volume of oxygen being used increases sharply but quickly levels to a steady state. The additional aerobic respiration required to meet the demands on the cyclist's skeletal muscles now has the necessary additional oxygen supply required.

A similar situation arises as the next, steeper hill is tackled. This time the new steady state of VO_2 is at an inevitably higher level, but oxygen supply is sufficient to sustain the additional aerobic respiration required. This is repeated in the same way, as each new, steeper hill is tackled.

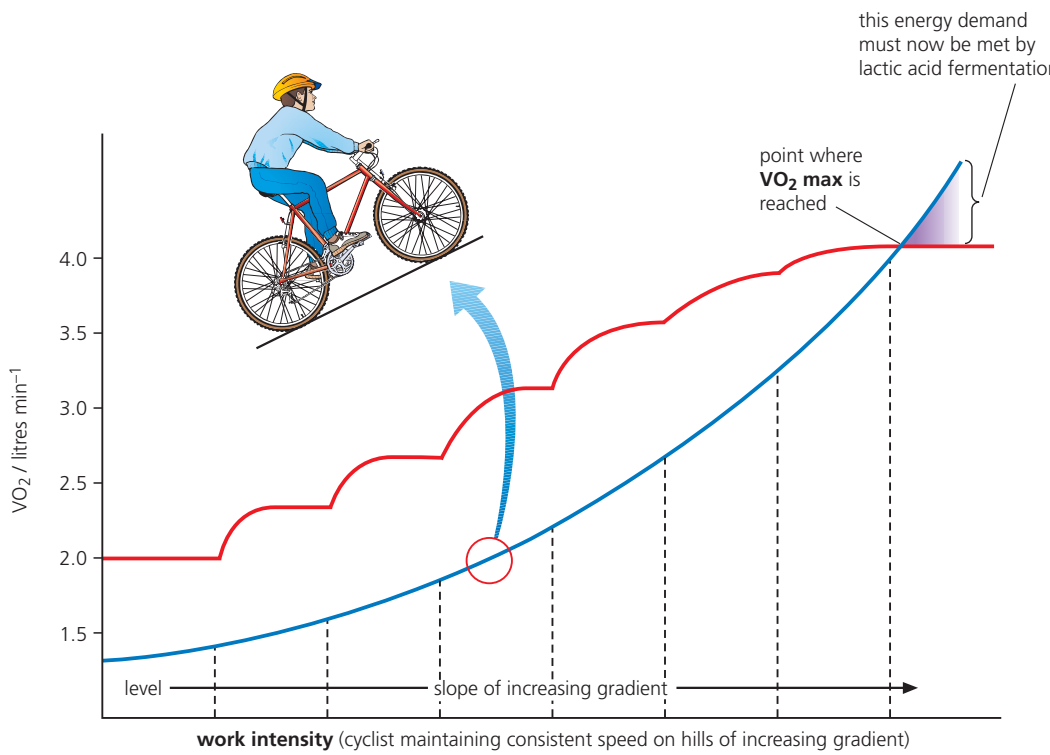


Figure 14.8 Oxygen consumption during exercise of progressive intensity

Remember, we require our subject cyclist to maintain the same speed throughout this demonstration.

The body's mechanism for delivery of adequate oxygen to skeletal muscles cannot continue to match continually rising demand. The graph shows the point where the maximum volume of oxygen that can be taken in, transported and used by working muscles per minute is reached. **This is VO_2max for this subject.**

The subject can continue with the set task, but now the additional demand for transfer of energy must be supplied by anaerobic respiration. Lactate and an excess of H^+ ions will accumulate in the muscles, and some will reach the blood circulation immediately. Our subject starts to experience muscle fatigue. How long the task is continued will depend to some extent on the degree of training previously undertaken (see below), and of course, on determination and will-power. However, the supply of respiratory substrates will become a factor, too.

How does the selection of respiratory substrate change during an extended period of intense activity?

Studies of how the body's energy transfer demands are met show a characteristic pattern (Figure 14.9).

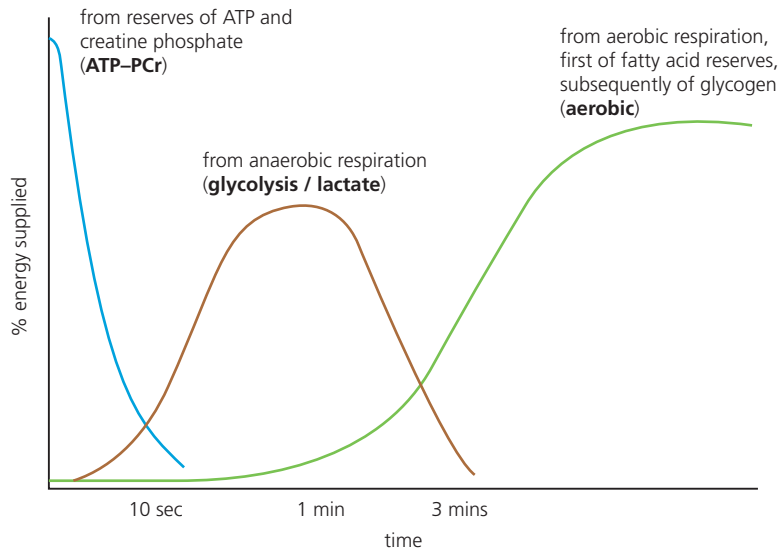


Figure 14.9 Contributions of energy transfer systems with duration of exercise

The most immediate demands of muscular contraction are met by using the small pool of ATP available in the resting muscles and the reserves of creatine phosphate (**ATP–PCr**). They may be exhausted in a matter of seconds, certainly in less than half a minute.

Meanwhile, the body's response to the sudden need for additional oxygen cannot immediately deliver adequate oxygen to the muscle mitochondria. In effect there is a lag period before aerobic respiration can operate at the necessary level. In this phase, the short-term energy system involves anaerobic breakdown of glucose reserves existing in the muscles (**glycolysis/lactate**).

The third phase, the aerobic energy system (**aerobic**) progressively operates effectively. Typically, within two minutes of continuing intense activity, the bulk of energy transfer is by this route. The resources available in and to the tissues are twofold: a supply of fatty acids and of glycogen (carbohydrate). It is the fatty acids that are preferentially oxidised by respiring muscles. Only when these are significantly depleted or exhausted does the breakdown product of glycogen become the respiratory substrate. At this point, glycogen reserves start to be depleted too.

The fate of lactate

Lactate that accumulates in muscles under anaerobic conditions is eventually transferred to the liver by the blood stream. In the liver, lactate is converted back to glucose, a process that requires energy from respiration, and thereby requires extra oxygen. Consequently, after lactic acid fermentation by muscles, an oxygen debt builds up in the body. The body also requires additional oxygen to recharge the oxymyoglobin reserves in muscles. Consequently, after a period of heavy activity, oxygen uptake only very slowly returns to resting levels. First, the oxygen debt is repaid (Figure 14.10).

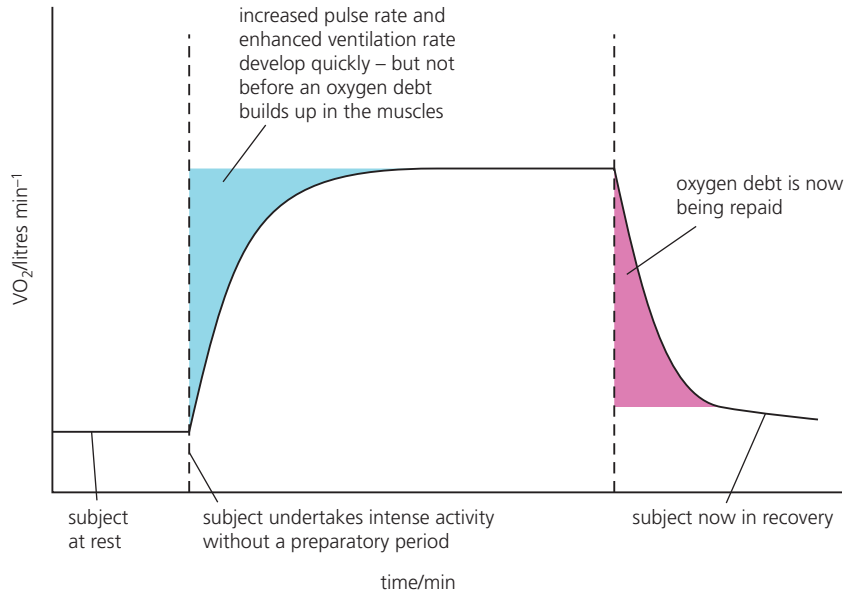
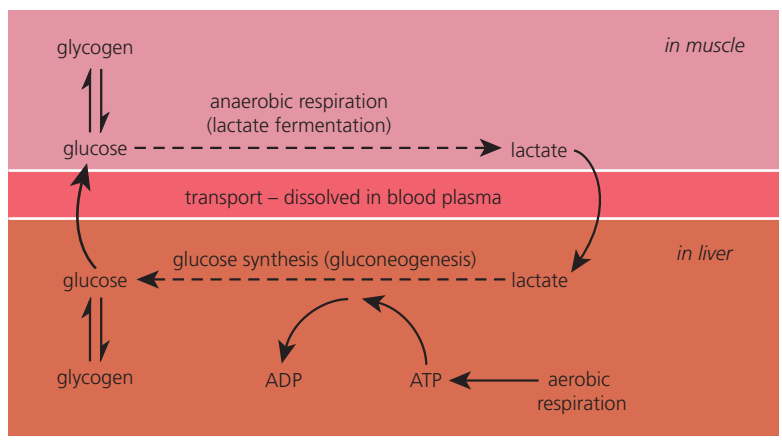


Figure 14.10 Lactate metabolism in the liver and the oxygen debt

biochemical changes in muscles and liver as (1) lactate accumulates and (later) (2) lactate is metabolised to glucose / glycogen:



Fitness and training

B5.1–5.5

We might say that a fit person has adapted successfully to cope with the stresses of their life style. Defined like this, fitness is rather a broad concept.

The reason is that 'life style' is so general. Such fitness is difficult or impossible to measure. In relation to training, for example, we must define fitness in terms of some specific activities.

Fitness is the ability to undertake specific activities without undue fatigue.

Of course, this definition can be applied to athletes and non-athletes, and within the field of athletics itself it is necessary to specify the type of activity that is the focus of training.

This is because fitness for a marathon runner will be a product of different qualities and abilities than many of those of a sprinter. Nevertheless, we can identify components of fitness, even though some will relate more to one activity than another. Furthermore, the factors fall into two broad categories: health-related and skill-related (Figure 14.11).

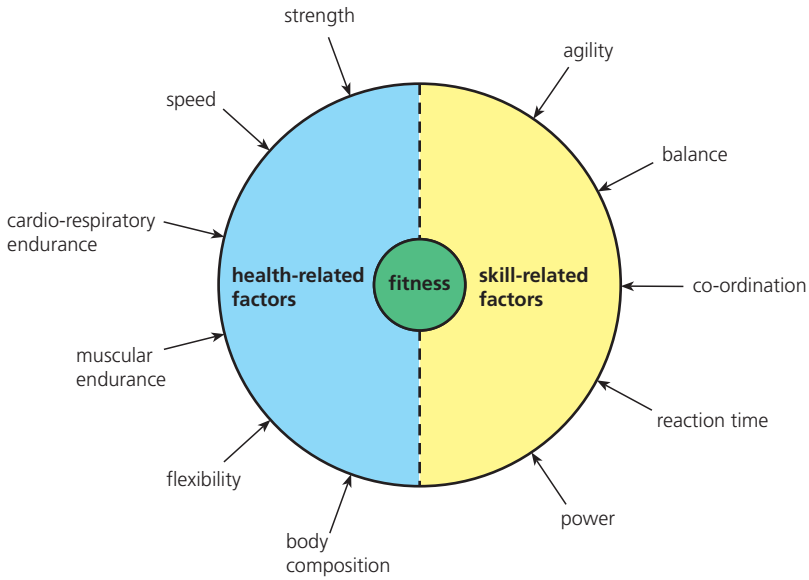


Figure 14.11 The components of fitness

Speed, stamina and muscle fibre composition

The basic structure of skeletal muscle and of the fibres that compose the muscle blocks that occur, surrounded by connective tissue, are illustrated in Figures 12.13 and 12.14 (page 366). In fact, two distinct types of skeletal muscle fibre exist in humans. These go by the rather intriguing names of **fast-twitch fibres** and **slow-twitch fibres**, and they possess differing physiological properties which we will discuss first (Table 14.3).

What are the properties of these two types of fibre?

Fast-twitch muscle fibres generate high force for a short time. The fibres contract rapidly, a property made possible by the high concentration of ATPase enzyme (observed by staining – Figure 14.12) among the cross-bridges of the myosin filaments.

Fast-twitch fibres rely on the glycolysis/lactate energy system, which we describe as the anaerobic energy source. This is a short-term supply route for ATP, but it is easily exhausted. However, the fibres have sarcoplasmic reticulum that is adapted for both rapid release and uptake of calcium ions, and this, too, contributes to their being able to contract rapidly (Figure 14.13).

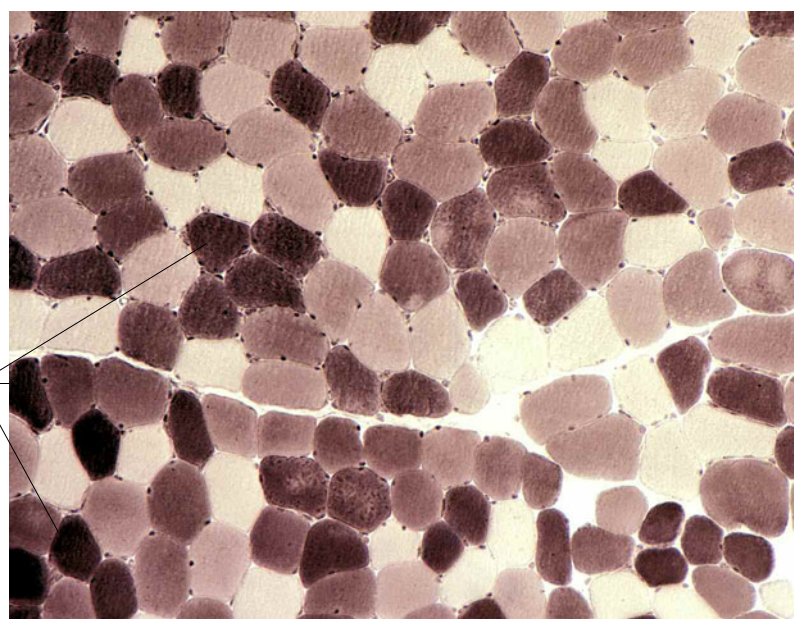


Figure 14.12 Photomicrograph of muscle fibres in cross-section, stained for the presence of ATPase activity. Note that the muscle fibres rich in ATPase activity have been stained darkly

This photomicrograph shows **muscle tissue from a marathon runner.** Here there is a predominance of slow-twitch fibres (stained lightly). On the other hand **muscle tissue from a sprint runner**, for example, would have a marked predominance of fast-twitch fibres (stained darkly).

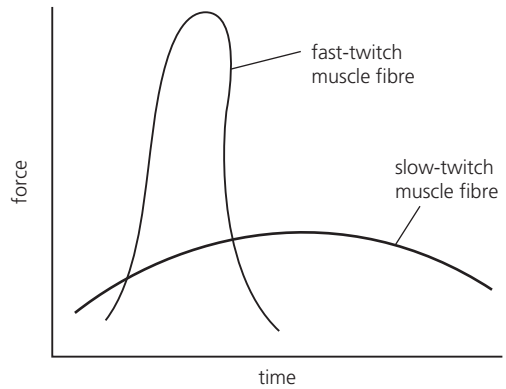


Figure 14.13 Muscle fibres' response profiles

On the other hand, slow-twitch muscle fibres achieve less force but sustain contraction for a longer time (Figure 14.13). The fibres have a relatively low concentration of ATPase enzyme among their myosin filaments.

Slow-twitch fibres have a high concentration of myoglobin (the oxygen-reserve pigment found in muscles), plus many mitochondria, and a high concentration of the mitochondrial enzymes needed to sustain aerobic respiration. They rely little on the glycolysis/lactate pathway. Their sarcoplasmic reticulum has slow calcium ion shunting qualities when compared to that in fast-twitch fibres, and this, too, contributes to their slow shortening characteristics.

Finally, we should note that skeletal muscles of human limbs typically consist of about 45–55% of slow-twitch muscle fibres. However, the precise proportions of fast-twitch and slow-twitch muscle fibres any individual has in the muscles of their arms and legs is apparently **genetically determined**. People have different proportions of the two. We might say that some of us are born for the marathon, some for a sprint, and some for activities requiring a balance of both. Both fibre types occur in equal quantities in able middle-distance runners, swimmers and footballers, for example. These activities combine demands for both muscle responses and so depend on both aerobic and anaerobic energy transfer during performance.

Fast-twitch fibres	Quality	Slow-twitch fibres
contract quickly, with fast release and removal of Ca^{2+} ions by sarcoplasmic reticulum	**speed of contraction**	contract slowly, with slow release and removal of Ca^{2+} ions by sarcoplasmic reticulum
glycolysis/lactate pathway predominates	**energy source**	aerobic respiration pathway sited in the many mitochondria
little dependence on oxygen for fast contraction – little myoglobin present (light in colour)	**dependence on oxygen supply**	highly dependent – well endowed with myoglobin oxygen store (so red in colour)
can contract with great force, but power quickly exhausted	**power and duration of contraction**	can contract repeatedly but with limited power
sprint runners typically have 80% of these fibres in some skeletal muscles	**occurrence**	marathon runners typically have 80% of these fibres in some skeletal muscles

Table 14.3 Characteristics of the contrasting muscle-fibre types – a summary

Speed and stamina as measures of fitness

Speed is just one aspect of fitness for an athlete, but obviously an important one. We can define speed as the rate that someone moves over a given distance, such as the movement required for sprinting.

In some sporting activities, this movement is combined with the speed of movement of a particular limb, within the whole performance, as in the launching of a javelin, or the bowling arm movements of a fast bowler (Figure 14.14).

The ability to achieve speed is genetically determined in part, since development of speed in practice is partly dependent on the proportion of fast-twitch muscle fibres the individual possesses.

Stamina is the **ability to endure prolonged physical or mental strain**. In this context, we might say that a person has developed staying power in relation to particular tasks or situations.

In terms of development of fitness, we see that cardio-respiratory endurance and muscular endurance (shown in Figure 14.11) are components of stamina – as must be a positive mental attitude.

Figure 14.14 Speed in performance – sprinter and fast bowler in action

Thus, achieving fitness involves developing the ability of both muscle-fibre types, including the ability to sustain repeated contractions for an extended period, and the related ability to provide and sustain energy transfer aerobically.

How can these attributes be effectively trained?

Training programmes devised to achieve fitness must relate to the particular sporting activity the person has chosen, and focus on both the muscle-fibre type and the dominant energy transfer system. In doing this, an appropriate programme focuses on:

- **frequency** of training – the issue of balance between exercise periods and rest;
- **intensity** of training – activity to develop heart rate and VO_2max, for example; in fact, moderately intense activity stimulates slow-twitch muscle fibres, whereas high-intensity activity stimulates the development of fast-twitch muscle fibres;
- **duration** of training – probably a corollary to intensity issues;
- **type** of training – issues resolved by the particular activity chosen.

6 **Design** a training programme relevant to the needs of a middle-distance runner, swimmer or footballer (see also 'The role of warm-up routines', page 441).

Ethics of performance-enhancing substances

Figure 14.15 Is it the gruelling physical demands, intense competition and the international accolades that success may bring that drive some athletes to resort to performance-enhancing substances?

Performance-enhancement refers to the practice of doping in sport. By 'doping' we mean:

> *The use (or the distribution for use) of certain substances which may have the effect of improving artificially an athlete's performance: these might act directly on muscles, nerves, blood circulation or breathing, or indirectly, via their effects on mental state.*

The International Association of Athletics Federations has ruled that doping is forbidden, and the organisations responsible for individual sports have been placed in control of testing for abuses (Figure 14.15). Chemists working in dedicated laboratories analyse samples of body fluids, chiefly blood and urine samples, typically collected at random, and check for evidence of use of particular groups of drugs (Figure 14.16).

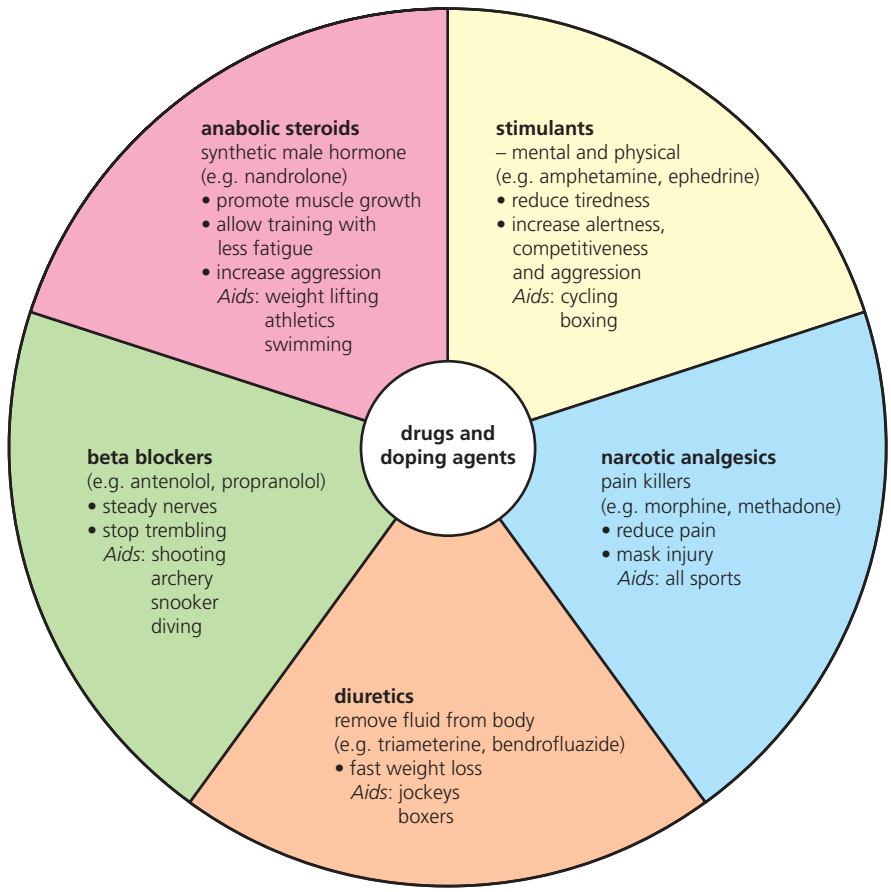


Figure 14.16 Illegal substances used for performance enhancement

However, it seems that in recent years, drug taking has become a common part of many individuals' body training programmes as well as those of athletes competing at national and international levels. Actually, the practice probably originated with athletes in ancient Greece who used stimulants to enhance their performances. As fast as the testers find means of detecting the use of performance-enhancing substances, today's suppliers seek for newer substances which leave little or no detectable trace in the samples taken for analysis.

The role and use of anabolic steroids

The naturally occurring sex hormones testosterone and oestrogen are chemicals known as steroids, and are manufactured in our bodies from cholesterol (Figure 13.7, page 397). Testosterone is the more powerful of the steroid hormones in producing pronounced anabolic effects, sought by many who wish to enhance body size and strength.

Anabolism is the building of large molecules, such as proteins, including muscle protein.

The use of anabolic steroids leads to enlarged muscles, and incidentally, an increased speed of recovery between training sessions. Another anabolic effect of testosterone and the other sex hormones is to slow down the natural breakdown and reabsorption of old bone, and promote the deposition of new bone. The skeleton is strengthened as a result.

Testosterone also has **androgenic effects**, summarised as masculising of the body. Facial hair growth increases and the voice deepens – unfortunate consequences for female athletes who choose to use it.

Testosterone has to be administered by injection and even then is short-lived in effects, for it is rapidly broken down in the body and excreted.

Laboratory-synthesised forms of testosterone with minor changes to the molecular structure have yielded a group of drugs called **anabolic steroids**. The effects of these on the body are almost identical to testosterone. Their use is banned by the International Olympic Committee

and their sale in some countries is a criminal offence. However, it is alleged that sales of anabolic steroids are second only to sales of cannabis among the list of illegal substances. Those who sell anabolic steroid to users (including to some gym users keen to increase muscle size) are able to import from countries where their sale is legal, including from Mexico.

With such widespread abuse, reasons for the popularity of anabolic steroids, and the dangers of their use, are worth considering: the ethics of anabolic steroid use are summarised in Table 14.4.

Why anabolic steroids attract some competitors	Why taking anabolic steroids should be banned
■ allow exceptional development of skeletal muscles	■ not fair competition; the user is cheating/being immoral
■ enable competitors to train harder	■ health risks and a danger of death
■ increase aggression and competitiveness	■ dangerous role model for young fans
■ readily obtained, and if use is monitored, health risks can be avoided	■ drug development should focus on disease prevention and cure
■ suppress fears of not 'making it', given the high expectations of peers, press, etc.	■ gives technologically developed countries an unfair advantage

Table 14.4 Issues with the use of drugs, such as anabolic steroids

Injuries

B6.1–6.2

Inevitably, strong or very sudden physical activity can occasionally result in unexpected damage to parts of the body. Even more likely are injuries incurred in contact sports, especially when these are played vigorously and in a highly competitive atmosphere, typical in contests where local or national pride is at stake. These causes apart, whenever humans are physically active, all manner of accidents are possible, and occasionally happen.

The role of warm-up routines

Warm-up routines are an integral part of appropriately planned training routines. They are, similarly, essential precursors to competitive participation – in both cases because they prepare the body for exercise. Warm-ups can help avoid injuries, too.

How do they achieve this?

- Warm-ups stimulate the release of the hormone adrenaline. This increases the heart rate and dilates the capillaries in the skeletal muscles, thereby allowing the speedier delivery of oxygenated blood to the muscles.
- Warm-ups decrease the viscosity of the blood as a result of the increases in muscle temperature that occurs. This speeds delivery of nutrients and oxygen to the muscles.
- Warm-ups increase muscle metabolism – another consequence of the temperature rise there. This improves the speed and strength of muscle contractions.
- Warm-ups increase production of synovial fluid in the joints, improving efficient movement.

A typical pattern of warm-up activity is:

1 general physical activity designed to raise heart rate;
2 mobility and stretching exercises;
3 sport-specific and skill-related components.

Common injuries to muscles and joints

Even with a carefully implemented warm-up routine, sports injuries arise from time to time. The pattern of common injuries that may be incurred during sporting activities include the following conditions.

Sprains

A sprain is brought about by a forcible wrenching or twisting movement across a joint. The result is a stretched (or even torn) ligament, but the bones of the joint are not dislocated. For a sprain to occur, the ligaments are stressed beyond their normal capacity. Serious sprains may also cause damage to surrounding blood vessels, muscles, tendons, or nerves. The whole area may become swollen and be very painful for a while.

Torn muscles

There are more than 300 muscles at work in our body all the time, enabling the maintenance of posture and body position, and bringing about locomotion when we choose. When these muscles are correctly exercised, giving them regular stretching and strengthening through use, they rarely give us problems. However, a direct blow on a muscle or sudden, violent exertion may bring about a muscle tear.

A tear may be partial or complete. If less than 5% of the muscle fibres are torn then only mild pain is experienced. The injury is often called a **pulled muscle**.

When the quantity of ruptured muscle fibre is much greater, then a bump appears below the skin at the site of damage, which is where pain is felt. The muscle can only be partially contracted and movement involving that muscle is painful. Most partial tears heal within four weeks with careful management such as a graduated physical exercise programme.

A **complete rupture** of the muscle immediately incapacitates, and severe pain is felt. A large lump appears at the site of the tear and much internal bleeding occurs there too. The injury may require surgery to heal properly, but the immediate response includes the application of an icepack. As recovery eventually begins, uninjured parts of the body have to be kept active, but in ways that exclude any use of the torn muscle, as long as it is painful.

Torn ligaments

A torn ligament is a serious injury to, or an actual break in, a ligament. This might be either in a ligament that holds particular bones together, or in one that forms part of a specific joint capsule (Figure 12.11, page 363). Torn ligaments may result from sudden force being applied or from a direct blow to a joint, or by a deep cut. Sometimes a 'snap' is felt or heard, but the site typically becomes painful, and movement unstable at the joint. Immediate generalised swelling occurs in the area and bruising develops soon after the injury.

A partial tear usually heals in several weeks of treatment – movements at the joint may require restriction, for example with splints, to aid the natural repair processes. A complete tear may require surgery.

Dislocation of joints

A dislocation occasionally results in adults and teenagers, when such extreme force is locally applied to the body that it causes particular bones that were being held together, to separate. For example, at a ball and socket joint, the ball may come completely out of its socket – as is the case with the most commonly dislocated joint in our body, that at the shoulder. In children, dislocations are rarer because, under similar pressures, their bones (which have less strength than those of adults) are more likely to fracture first.

Dislocations can be extremely painful events, and once they have occurred, are more likely to happen again. Most likely this is a result of the excessive stretching that is experienced by the surrounding ligaments, which may even become torn in the process. It is possible that there is also damage inflicted on surrounding soft tissues, including nerves and blood vessels. Immediate treatment with an icepack is recommended to reduce bleeding and decrease the pain experienced.

A dislocated joint requires speedy manipulation to reposition the bones, if the joint does not quickly self-correct. Sometimes effective manipulation requires a general anaesthetic and muscle relaxants. In more serious cases, surgery is required. Subsequently, the damaged joint may require complete immobilisation for a period of two weeks or more as part of the restorative programme of rest and selective exercises.

Intervertebral disc damage

Between each vertebra of the backbone (vertebral column) is a functional shock-absorber known as the intervertebral disc (Figure 14.17). These discs are constructed of an outer, thick collar of fibrous cartilage surrounding a soft, pulpy centre. During physical activity, the discs permit vigorous movements of the backbone and they absorb vertical shock. Under compression they flatten, broaden and even bulge out slightly, in extreme conditions.

Damage to an intervertebral disc may occur – it is common between certain of the vertebrae of the lower back (known as the lumbar region of the vertebral column) when the body is under the greatest mechanical stress. Such damage, often referred to as a 'slipped disc', involves the fibrous coat becoming torn so that the soft interior protrudes. The protrusion puts pressure on the spinal cord or spinal nerves, which may cause acute pain (Figure 14.17). In some cases, surgery is needed to repair the damaged disc, but usually a programme of rest and of appropriate exercise is sufficient to restore the area to normal function.

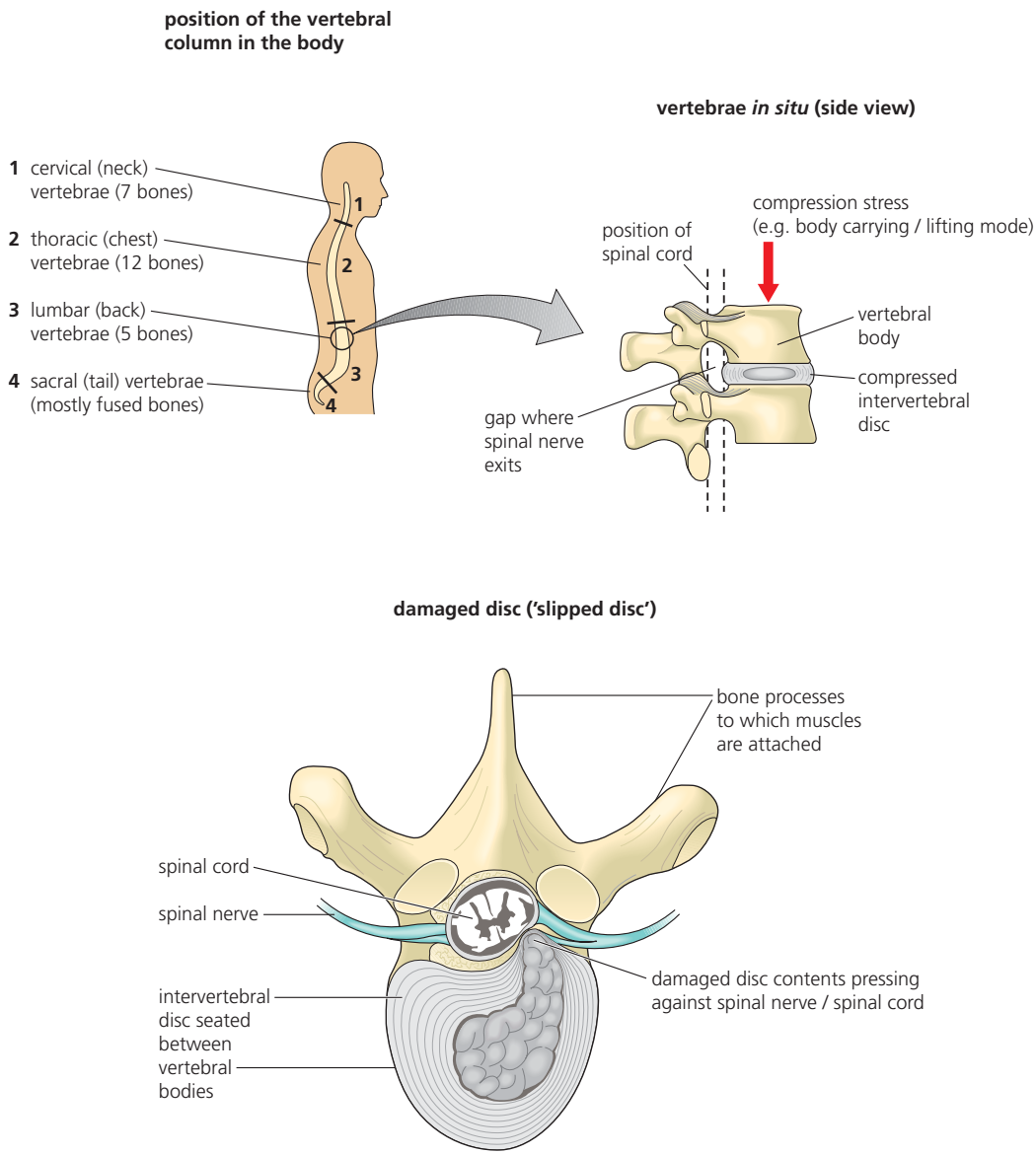


Figure 14.17 The vertebral column, vertebrae, and intervertebral discs

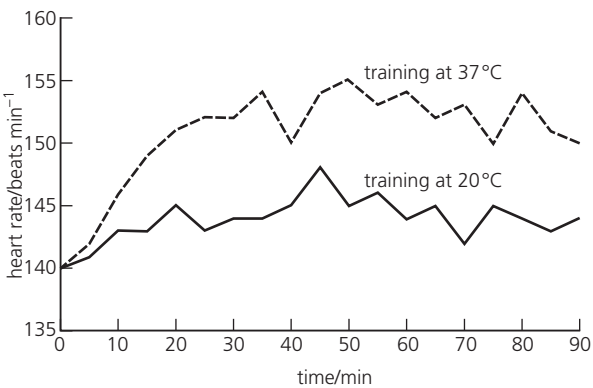
Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 Tests were carried out on 10 male cyclists to see if training at higher than normal temperatures improved their fitness. The cyclists were divided into two groups, A and B. Their training schedules were as follows:

Days 1–12	Groups A and B trained at 20 °C
Days 13–19	Group A trained at 20 °C and Group B trained at 37 °C
Day 20	Groups A and B had their heart rates measured during a time trial
Days 21–32	Groups A and B trained at 20 °C
Days 33–39	Group B trained at 20 °C and Group A trained at 37 °C
Day 40	Groups A and B had their heart rates measured during a time trial

The graph below shows the average results of the time trial on days 20 and 40 for the cyclists. In both trials the cyclists had to cycle 40 km in exactly 90 minutes at 20°C, with their heart rates being measured every 5 minutes.



a State the time of the highest heart rate reached by the cyclists training at 37 °C. (1)

b State the heart rate of the cyclists training at 20 °C after 50 minutes. (1)

c Compare the results for the heart rate at each training temperature. (3)

d Discuss whether the results support the hypothesis that training at high temperatures improves cardiovascular fitness. (2)

e Suggest why on days 33 to 39 both Group A and Group B trained at different temperatures from the temperature they trained at on days 13 to 19? (1)

Standard Level Paper 3, November 05, QB1

Q2 Explain the change in ventilation with exercise. (3)

Standard Level Paper 3, November 03, QB3

Q3 **a** Define the term *fitness*. (1)

b Describe how training relates to repaying the oxygen debt after vigorous exercise. (3)

Standard Level Paper 3, November 02, QB3

Q4 **a** State the type of muscle fibres that would be found in higher percentages in well trained marathon runners. (1)

b Explain why this type of muscle fibre is suitable in marathon runners. (3)

Standard Level Paper 3, November 05, QB2

Questions 5–10 cover other syllabus issues in this chapter.

Q5 Blood circulation to body organs is regulated, but not all parts of our circulation system are equally variable. Discuss this fact in relation to the pattern of blood supply to our:

a brain (2)

b skeletal muscles (2)

c skin. (2)

Q6 Outline the chief reserve substances held in skeletal muscle to sustain respiration and describe the sequence in which each is used by muscle fibres during sudden, vigorous activity. (6)

Q7 Despite the banning of the taking of performance-enhancing substances in sports, there is evidence suggesting that the anabolic steroid nandrolone has been used in this way, for example. Explain the reasons why the use of such a substance is potentially dangerous to athletes, as well as being illegal. (6)

Q8 Outline the components of an appropriate warm-up routine for a sporting activity of your choice, and explain the contribution of each stage to the avoidance of common sports injuries. (6)

Q9 **a** Outline **four** ways in which the structure of the thorax facilitates gaseous exchange in humans. (4)

b Describe the main effect of training on:

i tidal volume (2)

ii ventilation rate (2)

iii vital capacity. (2)

Q10 **a** Explain what you understand by

i**i** a 'slipped disc' (2)

ii a dislocation at the shoulder. (2)

b Predict the type of tissue that may arise as a direct result of the injuries listed in **a**. (4)

Option C

15 Cells and energy

STARTING POINTS

- **Proteins** are large molecules, built by **condensation reactions** between **amino acids**. Many hundreds or thousands of amino acid residues form a typical protein. Amino acids have a basic **amino group (—NH_2)** and an acidic **carboxyl group (—COOH)** attached to a carbon atom to which the rest of the molecule (—R) is also attached.
- Of the many amino acids known, **only 20** are built up to make the proteins of living things. Amino acids combine by **peptide linkages** between the carboxyl group of one molecule and the amino group of the other to form a **polypeptide chain**.
- In cells and organisms, the properties and roles of proteins are determined by the choice of **amino acids they are built from** and the **sequence in which the amino acids occur**. The sequence of amino acids is the primary structure of the protein, and is controlled by the DNA of the nucleus.
- **Respiration** occurs in every living cell, transferring energy to maintain cells and to carry out activities and functions. Cellular respiration requiring oxygen, **aerobic respiration**, involves the oxidation of sugar to carbon dioxide and water.
- Green plants manufacture glucose by **photosynthesis**, using the raw materials carbon dioxide and water and energy from sunlight. Oxygen is produced as a waste product.
- **Chloroplasts** are the photosynthetic organelles. Chloroplasts contain chlorophyll needed for the transfer of light energy into chemicals like glucose during photosynthesis.
- This chapter extends study of proteins and enzymes begun in Chapter 2 (pages 37–75) and Chapter 3 (pages 76–90), and of cellular respiration and photosynthesis, begun in Chapter 3.

Proteins, which make up about two-thirds of the total dry mass of a cell, are built from **amino acids**. Peptides and proteins typically consist of several hundred or even thousands of amino acid molecules combined together. The terms 'polypeptide' and 'protein' can be used interchangeably, but when a polypeptide is about 50 amino acid residues long it is generally agreed to have become a protein. Once the chain is constructed, a protein takes up a specific shape. Shape matters with proteins – their shape is closely related to their function. This is especially the case in proteins that are **enzymes**, as we shall shortly see.

Living things require **energy** to build and repair body structures, and to maintain all the activities of life, such as movement, reproduction, nutrition, excretion and sensitivity. Energy is transferred in cells by the breakdown of nutrients, principally of carbohydrates like glucose. The process by which energy is made available from nutrients in cells is called **cellular respiration**. Organisms require a supply of nutrients to sustain respiration, but they obtain their nutrients in very different ways. Green plants synthesise sugars from simpler substances by **photosynthesis**.

In this chapter, the **structure and roles of cell proteins**, particularly those that function as **enzymes**, are examined. Then the **biochemical events of respiration** and **photosynthesis** are considered in detail.

Proteins

C1.1–1.4

The content of this section is that of IB Topic 7 (7.5.1–7.5.4). It is addressed in Chapter 8, on pages 254–60.

Enzymes

C2.1–2.5

The content of this section is that of IB Topic 7 (7.6.1–7.6.5). It is addressed in Chapter 8, on pages 261–7.

Cellular respiration

C3.1–3.6

The content of this section is that of IB Topic 8 (8.1.1–8.1.6). It is addressed in Chapter 9, on pages 269–76.

Data to explain relating to cell respiration

C3.7

Oxidation/reduction

Reactions in which oxygen is added to a molecule, or hydrogen is removed, or electrons are transferred from one reactant to another, are called oxidation–reduction reactions (redox reactions).

1 In the following redox reactions, **deduce** which reactant is oxidised and which is reduced, and **list** your decisions in a copy of Table 15.1.

When an iron (Fe) nail is placed in copper (Cu) sulphate solution, the nail becomes coated in copper. The reaction can be presented in two parts:

$$Fe \rightarrow Fe^{2+} + 2^{e-}$$
$$Cu^{2+} + 2^{e-} \rightarrow Cu$$

$$Cu^{2+} + Fe \rightarrow Cu + Fe^{2+}$$

When ethanol (alcohol) is converted to ethanal (acetaldehyde) in the presence of the hydrogen accepter NAD^+, the reaction can be presented in two parts:

$$CH_3CH_2OH \rightarrow CH_3CHO + 2H$$
$$NAD^+ + 2H \rightarrow NADH + H^+$$

$$CH_3CH_2OH + NAD^+ \rightarrow CH_3CHO + NADH + H^+$$

When glucose is converted to carbon dioxide and water in the presence of oxygen in aerobic respiration, the reactions can be represented in a single (summary) equation:

$$6O_2 + C_6H_{12}O_6 \rightarrow 6CO_2 + 6H_2O$$

Reaction	Oxidised reactant	Reduced reactant
Iron reacting with copper sulphate		
Ethanol conversion involving NAD^+		
Aerobic respiration – summary		

Table 15.1 Which was oxidised and which reduced?

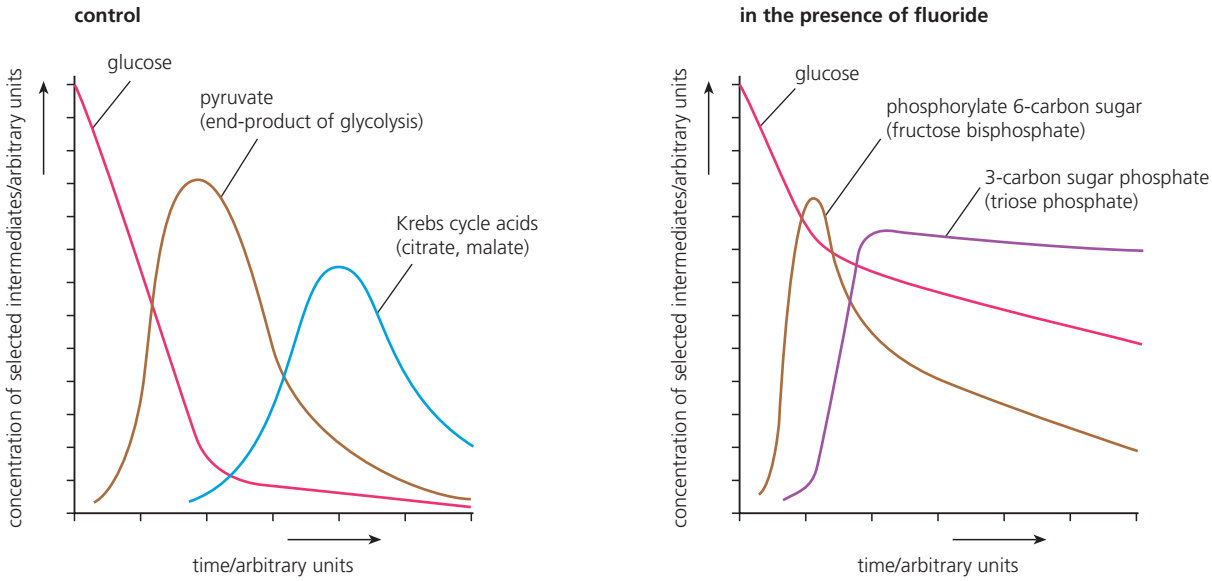
Glycolysis

To work out the steps within biochemical pathways such as glycolysis (Figure 9.3, page 271) (and the Krebs cycle, Figure 9.5, page 273), biochemists have experimented with actively respiring tissues of plant and animal origins. Investigations have been designed:

- to discover whether the predicted **intermediates are present**;
- to confirm the **presence of the necessary enzymes**;
- to check that **addition of intermediates** to respiring tissue starved of glucose leads to an immediate rise in their rate of respiration, as indicated by ATP accumulation or oxygen consumption (known as **feeding experiments**);
- to investigate the effect of adding a known **inhibitor** of a particular respiratory enzyme; does it causes the intermediates formed earlier in the pathway to accumulate and those formed later in the pathway not to appear?

In Figure 15.1, the result of an investigation of the effects of a respiratory inhibitor are summarised. The known inhibitor used in dilute solution was sodium fluoride. This inhibitor specifically inactivates one particular enzyme of glycolysis. Examine the graphs of the sequence of intermediates that accumulated in treated and control tissue. Then answer the questions.

Figure 15.1 Inhibition of respiring tissue using sodium fluoride



2 **Explain** how an enzyme brings about a particular chemical change in a substrate molecule.

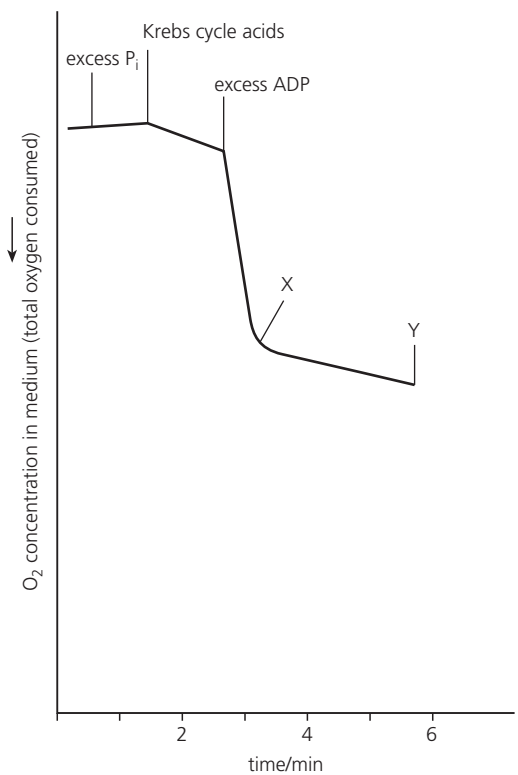
3 Fluoride is a non-competitive inhibitor in this experiment. **Describe** how it makes an enzyme inactive.

4 **Deduce** the specific step in glycolysis that is inhibited by fluoride.

5 **Predict** how the balance in the concentrations of ATP and ADP would change in the presence of the inhibitor, by contrast to the situation in the control tissue. **Outline** your reasons.

The Krebs cycle

Figure 15.2 Oxygen consumption by isolated mitochondria



Mitochondria can be extracted intact from respiring cells, washed free of any soluble enzymes from the cytoplasm (such as the glycolysis enzymes), and incubated in a buffer solution in the presence of selected intermediates. These kinds of experiments allow the metabolism of mitochondria to be investigated.

For example, using a research instrument known as an oxygen electrode, oxygen uptake by the mitochondria can be accurately measured. Additional oxygen is consumed by mitochondria when excess phosphate (P_i) and ADP are present, and when Krebs cycle acids are added in modest amounts.

Examine the graph of part of the results of this experiment (Figure 15.2), and then answer the questions over the page.

6 **State** which steps of aerobic respiration occur within mitochondria.

7 **Suggest** why oxygen consumption is delayed until the addition of Krebs cycle intermediates (in the presence of P_i and ADP).

8 **Predict** what condition has caused the change in total oxygen consumed at X. **Explain** the reason for your prediction.

9 **Explain** why mitochondria require both P_i and ADP before they may consume oxygen.

10 **Deduce** what would be the outcome at Y if there was now added to the mitochondria some:

- **a** more Krebs cycle acids
- **b** glucose
- **c** pyruvate
- **d** triose phosphate.

Terminal oxidation and the formation of ATP

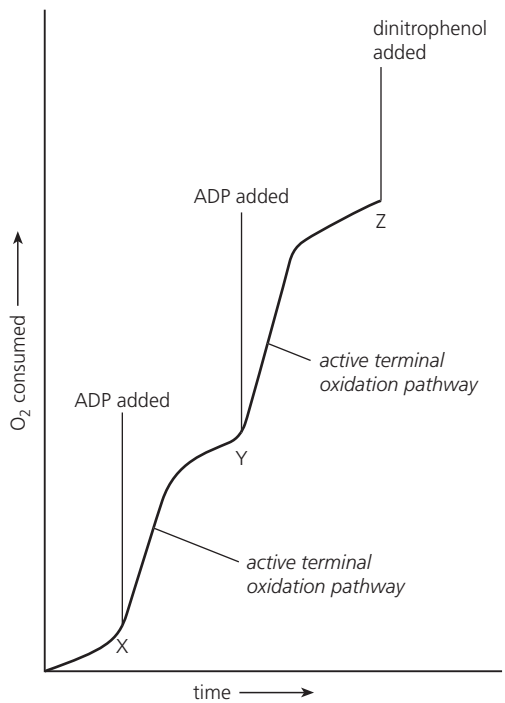


Figure 15.3 Terminal oxidation and ATP synthesis in isolated mitochondria

In Figure 9.6, page 274, we see that terminal oxidation (the transfer of electrons to oxygen to form water) is linked to the formation of ATP from ADP and P_i. These reactions have been investigated in isolated, washed mitochondria, when placed in contact with an oxygen electrode in specialised research apparatus. Examine the graph in Figure 15.3 and answer the following questions.

11 **State** where reduced NAD^+ ($NADH + H^+$) is produced in the respiratory pathway.

12 **Suggest** the fate of ATP formed in terminal oxidation in mitochondria in intact, growing cells.

13 **Describe** how ATP and ADP pass into and out of the mitochondria with their double membrane.

14 **Suggest** why the consumption of oxygen is dependent upon the regular addition of ADP (at X and Y).

15 Dinitrophenol is an inhibitor used by biochemists to isolate (detach) the process of ATP formation from the steps of terminal oxidation (the formation of water). These normally interconnected processes are uncoupled by this inhibitor. **Predict** the effect of addition of dinitrophenol at point Z on oxygen consumption.

16 **Suggest** the mechanism by which the rate of respiration is naturally regulated in cells:

- **a** to be rapid in cells undergoing growth and development
- **b** to be minimal in cells in an inactive state.

Photosynthesis

C4.1–4.8

The content of this section is that of IB Topic 8, (8.2.1–8.2.8). It is addressed in Chapter 9, on pages 277–92.

Data to explain relating to photosynthesis

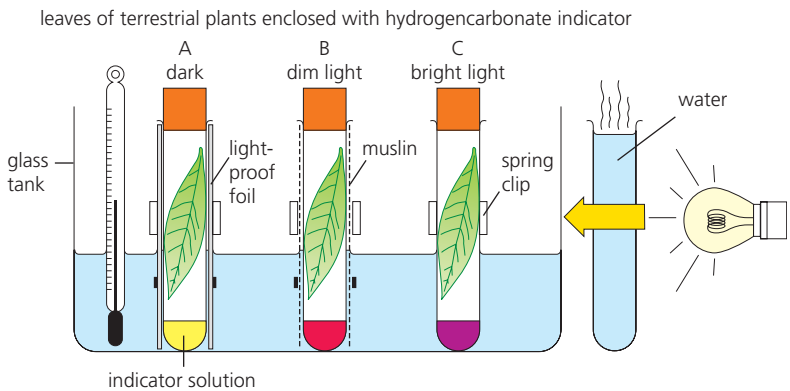
C4.9

Gaseous exchange in green leaves

Leaves of a terrestrial plant were placed in test tubes (Figure 15.4) to which about 15 cm^3 of hydrogencarbonate indicator solution had been added previously. The tubes were sealed and arranged in the water bath as shown, so that one tube was fully illuminated, one was in dim light, and the contents of the other were in the dark.

The hydrogencarbonate indicator solution had first been equilibrated with atmospheric air. This turns the solution a cherry red colour – the colour of the indicator in each tube at the start. If the concentration of carbon dioxide subsequently changes, the indicator will change colour. In fact, the hydrogencarbonate indicator solution turns yellow if it becomes more acidic. It turns purple if it becomes less acidic.

The outcomes of this experiment are quite rapid colour changes in two tubes, as stated in Figure 15.4.



Result:
- the indicator solution in A (dark) goes yellow
- the indicator solution in B (dim light) remains cherry-red
- the indicator solution in C (bright light) goes purple

Figure 15.4 The effect of light on the atmosphere surrounding green leaves

17 **State** the component in air which may be detected by a pH indicator. **Explain** why it has this effect.

18 **Suggest** why, in the apparatus shown, the light was arranged to pass through a tube containing water, prior to exposure of the experimental tubes.

19 **Explain** fully the causes of the colour changes in tubes A and C.

20 The evidence suggests that no change has occurred in tube B. **Comment** on this observation as fully as you are able.

Temperature and the rate of photosynthesis

The effect of temperature on the amount of gas given off by an illuminated aquatic plant shoot (stem and leaves) was investigated, using the apparatus shown in Figure 15.5. A freshly cut shoot of pondweed was used (*Elodea* is an example) which, when inverted, produced a vigorous stream of gas bubbles from the base. The bubbles confirmed the pondweed was actively photosynthesising.

The apparatus contained a dilute solution of sodium hydrogencarbonate, which supplied the carbon dioxide (as —HCO_3 ions) required by the plant for photosynthesis. The quantity of oxygen evolved in a given time (30 minutes) is a measure of how much photosynthesis had occurred in that period. This was estimated by drawing the gas bubble that collects, into the capillary tube, and measuring the length of the bubble. This length is then converted to a volume.

The effect of temperature on the rate of photosynthesis at both high and low light intensities was measured. Light intensity was varied by having the light source at different distances from the pondweed.

The result of one experiment is given in Table 15.2.

Temperature / °C	Oxygen produced in 30 minutes/cm³	
	at high light intensity	at low light intensity
5.0	26	20
10.0	39	25
15.0	66	28
17.5	95	30
20.0	128	33
22.5	110	30
25.0	76	28

Table 15.2 The effect of temperature on photosynthesis at high and low light intensity

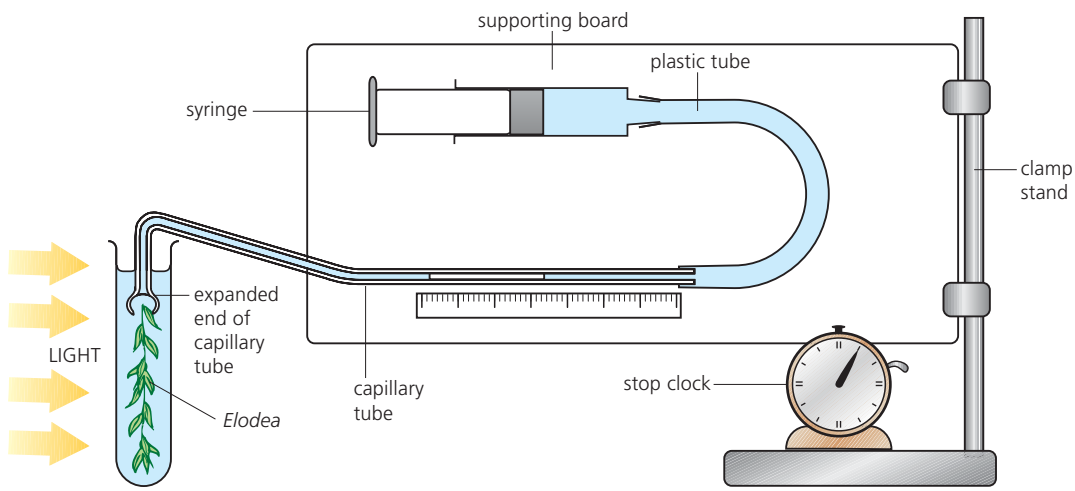


Figure 15.5 Apparatus for the measurement of the gas produced by illuminated pondweed

21 Using the figures in Table 15.2, **construct** a graph to show the effects of temperature on gas production at high and low light intensity.

22 **Distinguish** precisely, in your own words, between the effect of a rise in temperature from 10 °C to 20 °C on gas production by *Elodea* in high light intensity, and its effect in low light intensity.

23 **Compare** the effects of temperature on a photochemical reaction (such as when a silver halide photographic film is exposed), and on chemical reactions, with the observed results of this experiment. **Predict** the types of reaction that make up photosynthesis.

24 **Suggest** the likely cause of the change in rates of photosynthesis at the higher temperatures.

25 The result of this experiment was interpreted as showing photosynthesis is made up of two sequential reactions. **Explain** what these reactions are.

Investigating the behaviour of leaf discs

In Figure 15.6, steps of an experiment carried out on leaf discs cut from a terrestrial plant are illustrated. Notice that discs were first placed in a dilute solution of hydrogencarbonate in the barrel of a 10 cm^3 plastic pipette.

Examine Figure 15.6 and read the steps carefully.
Think about this experiment and the outcome before answering the questions below.

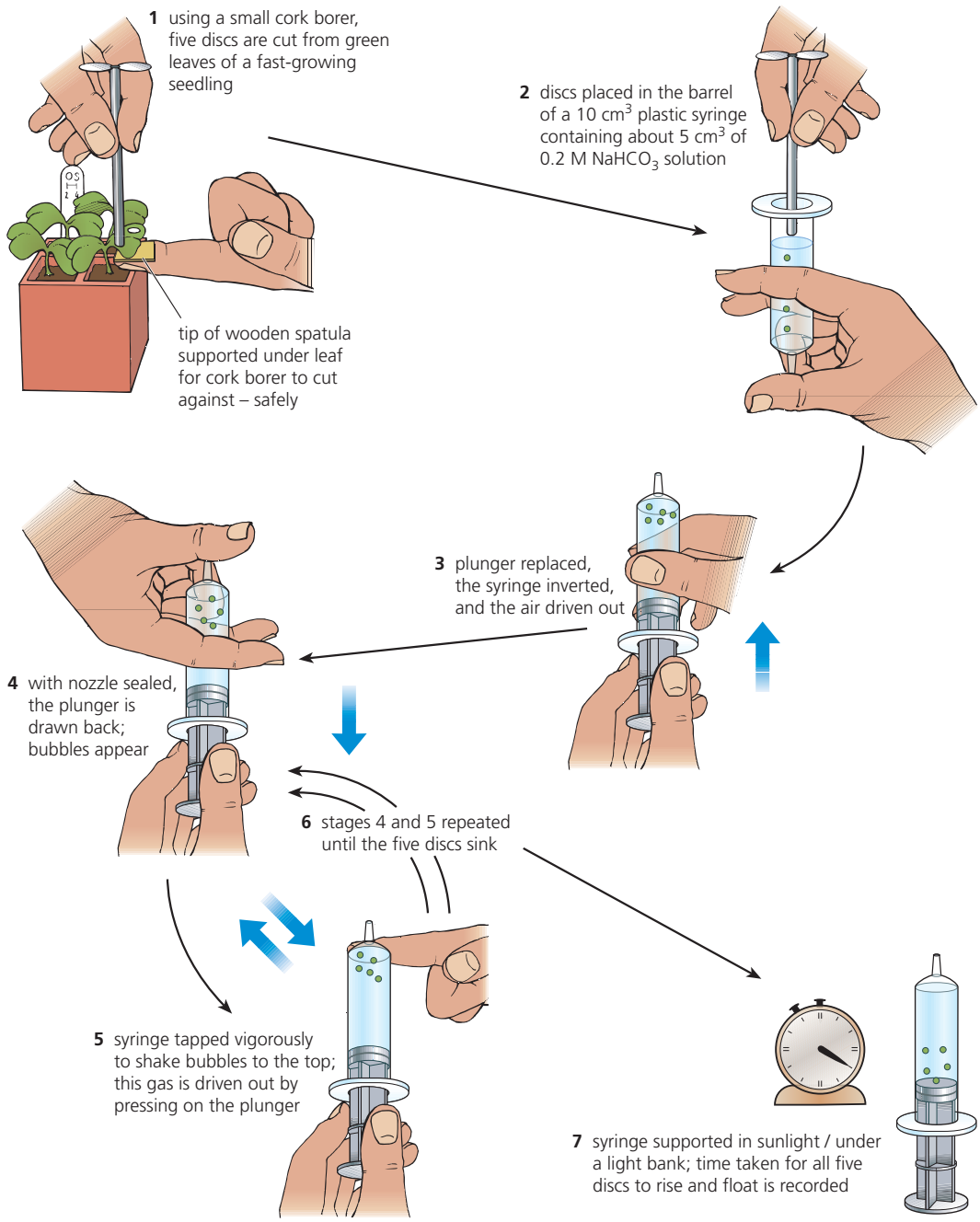


Figure 15.6 An investigation of factors limiting the rate of photosynthesis

26 **Explain** why the leaf discs initially float in the hydrogencarbonate solution.

27 **Identify** where the gas bubbles come from when the plunger is drawn back and the barrel contents are placed under tension (negative pressure).

28 **Describe** why the illuminated discs rise in the barrel after illumination.

29 **Design** an experiment to investigate the effect of different carbon dioxide concentrations on the rate of photosynthesis, using this simple procedure.

30 **Annotate** point A to explain the significance of this point on the graph.

31 **Identify** the stages of the curves at which light intensity is limiting the rate of photosynthesis.

32 **Identify** the stages of the curves at which light intensity is not limiting the rate of photosynthesis.

33 **Explain** the conclusion you can draw from the results of repeating the investigation at a higher concentration of carbon dioxide.

34 **List** the factors that may be limiting the rate of photosynthesis at E.

Limiting factors

A limited supply of either carbon dioxide or light may limit the overall rate of photosynthesis, since both are essential for photosynthesis. Temperature is also a limiting factor in photosynthesis.

Examine the graph in Figure 15.7 which records the results of an investigation of the effect of varying light intensity on the rate of photosynthesis at different concentrations of carbon dioxide. Make a copy of this graph so that your answers to the questions (left) may be added as annotations.

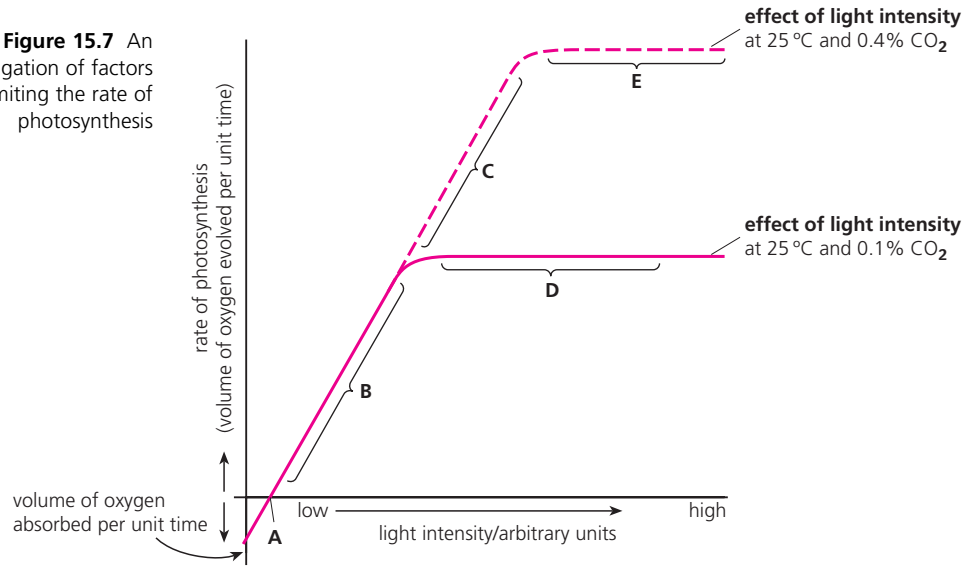
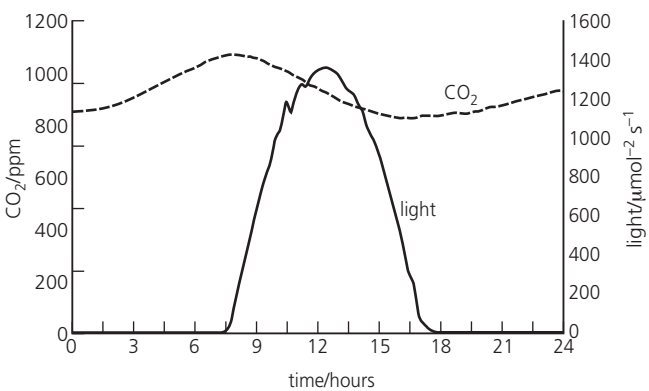


Figure 15.7 An investigation of factors limiting the rate of photosynthesis

Examination questions – a selection

Questions 1–5 are taken from past IB Diploma biology papers.

Q1 Biosphere 2, an enormous greenhouse built in the Arizona desert in the USA, has been used to study five different ecosystems. The effects of different factors, including changes in carbon dioxide concentration in the greenhouse, were studied. The data shown below were collected over the course of one day in January 1996.



a **i** Identify the time of day the sun rose. (1)

ii Identify the time of minimal CO_2 concentration (1)

b Determine the maximum difference in the concentration of CO_2 over the 24-hour period. (1)

c Suggest reasons for changes in CO_2 concentration during the 24-hour period. (2)

Standard Level Paper 3, November 04, C1

Q2 **a** Outline the induced fit model of enzyme activity. (3)

b **i** State **two** products of the light-dependent reactions of photosynthesis. (3)

ii Explain the **light-independent** reactions in photosynthesis. (4)

Standard Level Paper 3, November 04, C3

Q3 **a** Distinguish between oxidation and reduction. (2)

b State where in the mitochondrion the enzymes for the Krebs cycle are found. (1)

c Outline the processes involved in the Krebs cycle. (4)

Standard Level Paper 3, November 05, C2

Questions in Chapters 8 and 9 (pages 268 and 293) cover other syllabus issues in this chapter.

Option D

16 Evolution

STARTING POINTS

- By '**evolution**' we mean 'the development of life in geological time'.
- The **mechanism of evolution** is by **natural selection** of chance variations. Some phenotypes are better able to survive and reproduce in a particular environment than others. Selection determines the survivors and the genes that are perpetuated.
- **Variations** arise by mutations of genes and chromosomes, by the reshuffling of genes (independent assortment) that occurs in meiosis, and as a result of the random nature of fertilisation.
- By '**species**' we refer to 'a group of individuals of common ancestry that closely resemble each other and are normally capable of interbreeding to form fertile offspring'.
- The process of **classification** involves naming organisms by the **binomial system** so that each has a generic name and a specific name. Genera are grouped into a **hierarchical classification** of families, orders, classes, phyla and kingdoms.
- This chapter extends study of aspects of genetics begun in Chapter 4 (pages 91–116) and of evolution begun in Chapter 6 (pages 137–77).

The word '**evolution**' is used widely, but in biology the term specifically means 'the process that has transformed life on Earth from its earliest beginnings to the diversity of forms we know about today, both living and extinct'. Biologists now believe that **organic evolution by natural selection** accounts for the major steps in evolution. These are **macroevolution** – major developments such as the origin of the eukaryotic cell, the origin of multicellular organisms, and the origin of vertebrates from non-vertebrates; and **microevolution** – the relatively minor changes that arise and lead to the appearance of new, but closely related species.

In some cases, we still know little about how major change came about; certain current ideas remain speculative. However, the timing of many of them has been dated relatively precisely. It is possible that the details of some of these issues will remain partly a mystery, although modern evidence for evolution from comparative biochemical studies of DNA and of cell proteins, for example, has added new and fertile sources of evidence – both on the timing of changes and the existence of common ancestors.

In this chapter, the possible spontaneous **origin of life**, the likely **mechanisms for the evolution of new species** from existing forms, and the path of **human evolution** are all discussed in some detail.

In the Additional Higher Level extension, the formula for **detection of changes in the composition of the gene pool** in a local population (the basis of evolutionary change) is explained. Finally, the processes of **classifying organisms** and the **evidence on which our classifications are based** are shown to contribute to our knowledge of the **common ancestries of organisms**.

Origin of life on Earth

D1.1–1.8

The **Earth** is one of the smallest planets grouped in the Solar System around a central star, the **Sun**. The fact that the planets all revolve in the same plane supports the theory that the Sun and planets were all formed from the condensation of a single **revolving disc of matter**. It is likely the Earth originated from masses of molten rock that collided and coalesced. With cooling, a **crust** formed but the restless surface was initially continuously disturbed as other matter collided. Heat from impacts and from the decay of radioactive elements such as uranium was probably sufficient to melt matter and keep it molten.

In the liquid state, the bulk of heavy elements, particularly iron, formed the Earth's liquid **core** of dense matter. Radioactive elements, though present in small amounts, have had enormous effects on the Earth's geological evolution and they continue to keep the interior hot.

The surface of the Earth eventually cooled to 100 °C and below, and an **atmosphere** developed. The gravitational field on Earth was strong enough to retain this atmosphere, unlike that of the Moon. The major constituents of the atmosphere would have been:

- **nitrogen**, **water vapour** and **carbon dioxide**;
- smaller amounts of **methane**, **ammonia**, **carbon monoxide**, **sulphur dioxide**, **hydrogen sulphide** and **hydrogen cyanide**.

These are all products of the effects of heat on the lighter chemical elements of the crust, and of **lightning** and **ultra-violet radiation**. (The arrival of **comets** was a possible alternative source for some of the gases, particularly **water vapour**.)

The atmosphere was virtually **without oxygen** – in fact, any trace of free oxygen would have immediately reacted with the large quantity of iron present.

1 **Explain** why we can expect that, of all the fossils found in sedimentary rock, those of the lowest strata may bear the least resemblance to present-day forms.

The rock of the Earth's **crust** is a relatively thin layer. It is divided into huge **plates** that move about on the surface, and where they meet, one or both turn under and become part of the **mantle** layer below.

As the Earth continued to cool, the water vapour in the atmosphere condensed and returned to the surface as rain, forming rivers and lakes. **Seas** formed. Now the process of erosion began to mould the landscape, and the eroded debris became the first **sedimentary rocks**.

The spontaneous origin of life on Earth

Life in the form of **living cells** may have developed spontaneously in evolving conditions similar to those described above. If so, the following steps would have been involved:

- the non-living synthesis of **simple organic molecules**, such as sugars and amino acids;
- the assembly of these molecules into **polymers**;
- the development of **self-replicating molecules**, such as nucleic acids;
- the packaging of these molecules within **membranous sacs**, so that an internal chemistry can develop, different from the surrounding environment.

Are such developments possible to imagine?

Charles Darwin speculated about the origin of 'the living' from inorganic sources after this issue was raised as a criticism of his theory – *On the Origin of Species by means of Natural Selection*, first published in 1858. He originated the concept of a **primordial soup** – a 'warm little pond' as the venue.

However, the environmental conditions he imagined at that time are now recognised as totally unrealistic. For example, before oxygen gas was present in the Earth's atmosphere there could be no ozone layer in the upper atmosphere. Today we have in excess of 20% oxygen in our atmosphere, and consequently an ozone layer shields against **ultra-violet radiation (UV)** – relatively little UV light reaches life's terrestrial environments. In the absence of an ozone layer, life could not survive because it would be exposed to continuous bombardment by UV radiation.

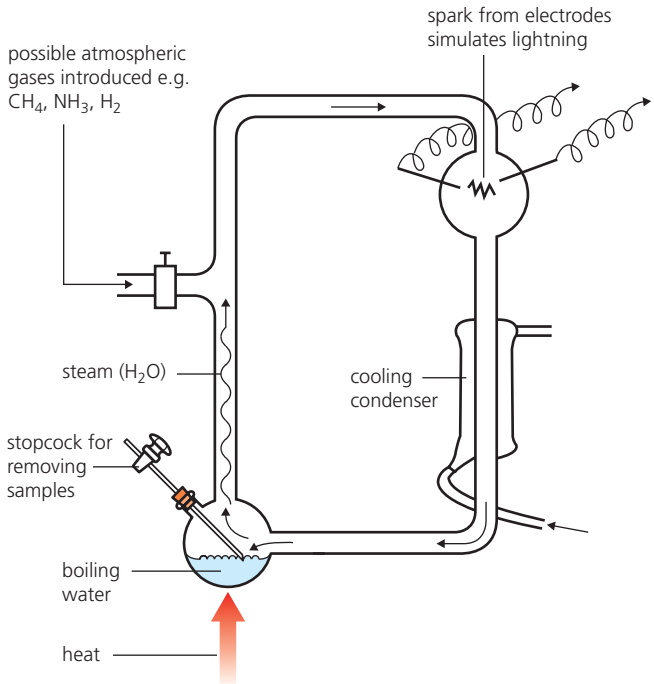
We have seen that the molecules that make up living things are built from carbon, hydrogen, oxygen, with some nitrogen, phosphorus, sulphur, and a range of atoms of relatively few other elements also present. Today, living things synthesise and metabolise these molecules by the action of **enzymes** in their cells, but for life to originate from the non-living, the first step was the **non-living synthesis of simple organic molecules**.

Experimental evidence for the origin of organic molecules

Experimental evidence of how simple organic molecules might have arisen from the ingredients thought to be present at the time before there was life on Earth was produced by **S. L. Miller** and **H. C. Urey in 1953** (Figure 16.1). They set up a reaction vessel in which particular

Figure 16.1 Apparatus for simulating early chemical evolution

Apparatus like this has been used with various gases to investigate the organic molecules that may be synthesised.



environmental conditions could be reproduced. Here, strong electric sparks (simulating lightning) were passed through mixtures of methane, ammonia, hydrogen and water vapour for a period of time. They discovered that amino acids were naturally formed (some of them known to be components of cell proteins) as well as other compounds.

This approach confirmed that **organic molecules can be synthesised outside cells**, in the absence of oxygen. The experiment has subsequently been repeated, sometimes using different gaseous mixtures and other sources of energy (UV light, in particular), in similar apparatus. The products have included amino acids, fatty acids, and sugars such as glucose. In addition, **nucleotide bases** have been formed, and in some cases, **simple polymers** of all these molecules have been found.

To summarise, we can see how it is possible that a wide range of organic compounds could have formed on the pre-biotic Earth, including some of the building blocks of the cells of modern organisms.

TOK Link

To what extent can you argue that Miller and Urey's experimental response to a seemingly insoluble issue was uniquely a scientific response?

An alternative source of organic molecules

Francis Crick, the co-discoverer of the structure of DNA, was a modern supporter of a suggestion that organic molecules, the essential precursors of living cells, may have emerged on another planet or moon and 'hitched a ride' to Earth on a comet. The idea that life did not originate on Earth but arrived in some form from an extraterrestrial source is known as **panspermia** (Greek for 'all seeds' – it was a Greek philosopher who 2500 years ago proposed that all life originated from combinations of tiny seeds pervading the cosmos).

Currently, this idea is being researched by astrobiologists and planetary geologists in America. NASA scientists have confirmed that early in the history of our Solar System, conditions essential for life were present elsewhere. For example, on Mars, water flowed intermittently, and life may have existed there. Also, Europa, the fourth-largest moon of Jupiter, appears to possess liquid water under an icy surface. Titan, the largest satellite of Saturn, is rich in organic compounds.

The expanse of interplanetary space has been crossed in ways that may have transported organic matter. For example, about 30 meteorites found on Earth originated from Mars. Biological matter is more likely to survive travel in the interior of meteorites, either in the form of RNA alone or assembled with ribosomes in 'protein factories'. As yet, though, there is no evidence it happened.

Assembly of the polymers of living things

In order for polymers to be assembled *in the absence of cells and enzymes*, the steps required include the concentration of the biologically important molecules such as monosaccharides (the monomer building blocks for polysaccharides), amino acids (building blocks for proteins), and fatty acids (for lipid synthesis). They would need to come together in 'pockets' where further chemical reactions between them were possible.

How could this have come about?

Speculative answers, in the absence of direct evidence, include the following suggestions:

- in situations where clay particles have accumulated – here the tiny spaces between the huge surface areas of clay particles and their sheets of atoms may have favoured the organisation and reactions of existing simple molecules;
- in seas close to deserts or the lava flows of volcanoes, where pockets of water occurred, frequently vulnerable to the drying and concentration of the contents;
- in areas of the ocean surfaces where foam and bubbles persist and provide enclosed spaces;
- in the solutions under deep ice caps that do not freeze – here the conditions could have included high pressure, at times, which favours reactions;
- in the vents of sub-marine volcanoes where the environment is hot, the pressure is high, and the gases being vented are often rich in sulphur and other compounds.

Origin of self-replicating molecules

For the evolution of life from a mixture of polymers and their monomers, two special situations need to emerge. These are:

- a **self-replication system**;
- an **ability to catalyse chemical change**.

These are essential ingredients of living, functioning cells. Today, in living cells, these essentials are achieved by our **DNA**, the home of the genetic code, and our **enzymes**, which are typically large, globular proteins.

However, no-one has yet been able to synthesise DNA and globular proteins in any of the reported experiments repeating Miller and Urey's demonstration of how biological important molecules could be synthesised in the pre-biotic world.

So what may have filled the roles of DNA and enzymes in the origin of life?

A possible answer was found in the unexpected by-product of genetic engineering experiments involving *in vitro* investigation of the enzymes required to patch and join short lengths of RNA (a process that genetic engineers call splicing). These experiments showed, to everyone's surprise, that when the naturally occurring protein enzymes that catalyse RNA patching (obtained from cells) were omitted from the reaction mixtures, the RNA fragments still spliced *on their own*.

It had been assumed that the RNA-patching enzyme (a protein) was the essential catalyst. This was the first demonstration that short lengths of **RNA**, as well as being 'information molecules', also **function as enzymes**. These catalytic RNA molecules have been named **ribozymes**.

Perhaps short lengths of RNA filled the dual roles of information molecules and enzymes in the evolution of life.

Now we have experimental evidence that short lengths of RNA can also function as enzymes, although they may rarely do so in modern cells.

In present day eukaryotic cells, **messenger RNA (mRNA)** carries the **genetic code** between nucleus and the site of protein synthesis, the **ribosomes** (themselves another form of RNA). Other RNA, known as **transfer RNA (tRNA)** brings the amino acids to the ribosome for the building of the protein. However, the enzymes that catalyse the chemical reactions involved throughout are proteins.

2 Suggest likely chemical changes that would have occurred in the first cells, and that would have required a catalyst.

Further investigations show ribozymes to be fairly inefficient enzymes – slow and unpredictable at times, but that they work satisfactorily with polynucleotide substrates. They can catalyse simple replications, although they do this in an error-prone way, on occasions. Thus, ribozymes may catalyse the formation of DNA, for example.

The discovery of ribozymes completes the story of a possible and credible route from the pre-biotic soup to living things, simply because this form of RNA is an information molecule that both replicates and may function as enzymes.

Protobionts to prokaryotes – a possible transition?

The first cells were prokaryotes. This we know from the fossil record. Were they preceded by a 'lower' or lesser level of organisation – some form of **protobiont**?

A limited number of lipid molecules, once formed, arrange into a monolayer on the surface of water (page 22). When more lipids become available, the whole re-forms into lipid bilayers – the basis of plasma membranes today. If such bilayers formed and linked up into **microspheres** that surrounded a small amount of the pre-biotic soup of polymers and monomers, perhaps these were the fore-runners of cells (Figure 16.2)?

Microspheres might be dubbed 'membrane systems with a distinctive internal chemistry', for the contents have the potential to develop a chemical environment different from the surroundings.

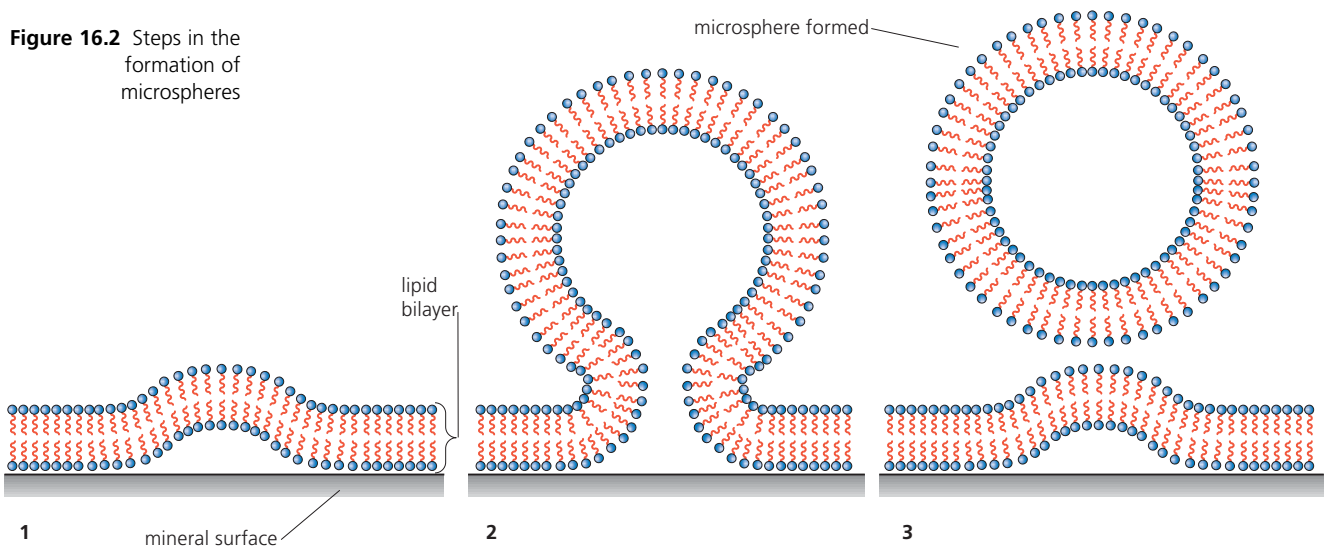


Figure 16.2 Steps in the formation of microspheres

Also observed are structures called **coacervates**. These are formed from dilute solutions of two substances each having large polymer molecules carrying opposite charges. The two most commonly studied are gelatine and gum arabic. At certain concentration, these separate into sol (liquid) and gel (solid) phases. Each phase contains both polymers but at different concentrations. However, both contain large amounts of other molecules in solution.

Complex coacervates have been observed, one gel droplet within another. If such droplets came into existence and contained enzymes, they would form a model for the biochemistry of the cell. The Russian biochemist **Oparin** (1894–1980), who pioneered the chemical approach to the origin of life, attached great importance to coacervates in the evolution of life.

A **prokaryote cell** differs from these models in a number of ways. For example, attached to the plasma membrane in the prokaryote cell is a single **circular chromosome** of DNA, known as a nucleoid (page 17). Also, a cell wall of complex chemistry is secreted outside the membrane barrier to the cell contents. However, both protobionts and the first prokaryotes could have survived nutritionally on the organic molecules of the pre-biotic soup. In this early life environment, with a wealth of simple organic molecules surrounding simple cells, digestion and respiration would have demanded only limited enzymic machinery. Biochemical sophistications would have to evolve with time – if life originated in this manner.

The contribution of prokaryotes to an oxygen-rich atmosphere

Some of the earliest prokaryote fossils contain cells very similar to modern cyanobacteria (modern prokaryotes that are photosynthetic). They are present in large mounds known as **stromatolites**, fossilised examples of which are common, and of which there are also still living examples (Figure 16.3). Stromatolites are formed in shallow waters, the mounds built of layer upon layer of bacterial mats. The earliest fossil stromatolites date from some 3500 million years ago.

Figure 16.3 Stromatolites today – these living structures of ancient origin were recently photographed at Shark Bay, Western Australia

In stromatolite mounds, the outer layer is of filamentous cyanobacteria – **photosynthetic bacteria** that absorb light, produce carbohydrates, and release **oxygen**. Below is a layer of purple bacteria that also absorb light and also manufacture carbohydrate, but do so without releasing oxygen. Further below is a layer of other bacteria that are saprotrophic. Some are able to fix atmospheric nitrogen into combined nitrogen of amino acids, for example. The combined components of stromatolites are a biochemically able assortment.

Photosynthetic prokaryotes began the process by which free oxygen accumulated in the Earth's atmosphere. With free oxygen in the atmosphere, the **formation of an ozone layer** in the upper atmosphere commenced. Once formed, the ozone layer began to reduce the incidence of UV light reaching the Earth's surface. Terrestrial existence (rather than life restricted to below the water surface) became a possibility.

Meanwhile other prokaryotes, more akin to modern aerobic bacteria, simply 'fed' on the organic molecules available in their environment. However, these bacteria had evolved aerobic respiration (only possible as a result of the free oxygen from photosynthetic cyanobacteria, now present in the atmosphere) and so had the enzymes not only of glycolysis, but also of the Krebs cycle and terminal oxidation.

3 **Outline** the significance for the development of life on Earth of the accumulation of oxygen in the atmosphere as a result of photosynthesis.

Prokaryote to eukaryote

We know that the first cells were prokaryotes. It is likely that some larger prokaryote cells came to contain their chromosome (whether of RNA or DNA) in a sac of infolded plasma membrane (Figure 16.4). If so, a distinct nucleus was now present. But how might the other organelles have originated? Remember, membranous organelles are a feature of eukaryotes, in addition to their discrete nucleus.

The cell as a habitat – a possible origin for mitochondria and chloroplasts

Both mitochondria and chloroplasts contain a ring of DNA double helix, just like that contained by a prokaryote. They also contain the small ribosomes, like those of prokaryotes. These features have caused some evolutionary biologists to suggest that some organelles are descendants of free-living prokaryotic organisms that came to inhabit larger cells. It seems a fanciful idea, but not an impossible one.

Present day prokaryotes are similar to fossil prokaryotes, some of which are 3500 million years old. By comparison, the earliest eukaryote cells date back only 1000 million years. Thus eukaryotes must have evolved, surrounded by prokaryotes that were long-established organisms. It is possible that, in the evolution of the eukaryotic cell, prokaryotic cells (which at one stage were taken up into food vacuoles for digestion) came to survive as organelles instead. If so, with time they would have become integrated into the biochemistry of their host cell. This concept is known as the **endosymbiotic origin of eukaryotes**. It is examined further in Table 16.1.

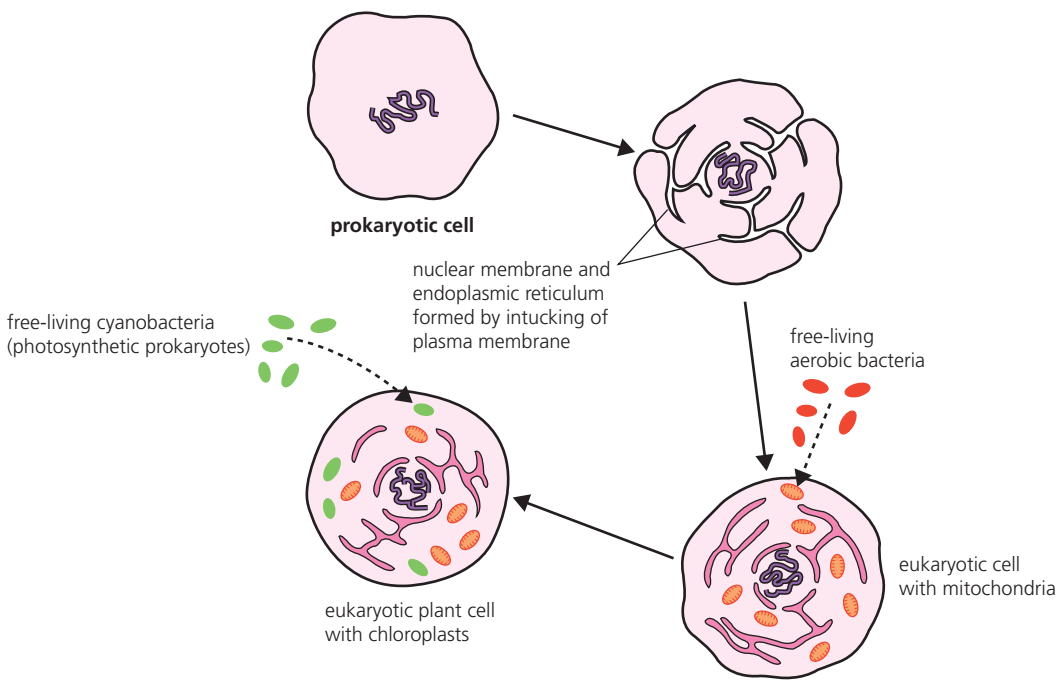


Figure 16.4 Origin of the eukaryotic cell

Arguments in favour …	However …
Prokaryotes are known to inhabit eukaryotic cells.	Chloroplasts and mitochondria could have arisen by invagination and specialisation of the plasma membrane (see Figure 16.4).
Chloroplasts and mitochondria have similarities to true bacteria, including size, method of reproduction (binary fission), presence of smaller ribosomes, circular DNA (not associated with histone proteins), similar enzymes and membrane proteins.	Chloroplasts and mitochondria are not genetically autonomous – most of their proteins originate in the cytoplasm outside them, by translation of mRNA that has been transcribed from genes in the nucleus.

Table 16.1 The endosymbiotic hypothesis examined

Extension: Special creation – an alternative explanation of the origin of life?

The Christian idea that each species was created by God, based on the biblical account in Genesis, was widely accepted at one time. Associated with this idea was the belief that species were unchanging (**immutable**).

Also, from the chronology of the biblical account, life on **Earth was thought to be merely a few thousand years old**. In 1654, a biblical scholar calculated that the world was created in the year 4004 BC. This figure was accepted in Europe as a fact, at least until well into the nineteenth century.

Some accepted these ideas on authority alone; others reasoned an 'argument for design'. In the eighteenth century, a theologian, William Paley, based a case for 'creation' on the complexity of living things. He argued that if one found a watch but knew nothing of such objects, one would conclude the existence of a watch-maker on discovering the intricate watch mechanism.

Today, many Christians may view the bible in a different light. Nevertheless, they typically hold to involvement of a supernatural force in the origin of living things, even if they accept that living things may change with time. They may have some affinity with William Paley, perhaps?

But Christianity is only one of several world religions. Table 16.2 briefly summarises the central ideas of other religions and belief systems in relation to the origin (and perhaps the changeability) of living things. Inevitably, this summary is only introductory, but it does acknowledge a range of views about the nature of life.

Ethical system	Ideas on the origin of life and organic evolution
Judaism	A single god who not only created the Universe, but who continues to work in the world.
Islam	There is only one god (Arabic name: Allah) who created everything and rules everything. The concept of evolution is not accepted.
Hinduism	Spirituality is a principle rather than a personality. The Universe is one divine entity, one god embodying the principles of Brahman the Creator, who is continuing to create, and Vishnu the preserver.
Sikhism	Perhaps a blend of Hinduism and Islamic ideas in origin. Holds belief in an eternal creator god who governs the Universe absolutely.
Buddhism	A tradition focused on personal spiritual development. Buddhists strive for deep insight into the true nature of life and do not worship deities.
Rastafarianism	No formal creed. A religion of oppressed Black people living in exile.
Humanism	Accept current science view and experimental approach. View the concept of god as a human invention, and may be critical of its value.

Table 16.2 Views on the origin of life

We can conclude that the idea of creation – albeit in different forms – has been, and still is, held as a matter of faith in many different cultures. Perhaps the origin of the idea was from speculations by early humans about the world of nature which surrounded them, and became a religious idea, later? Be that as it may, **special creation is not a scientific theory**. The consequence of this belief may be significant to personal religious views, but it does not lend itself to experimental investigation, enquiry and falsification. This is a field in which science is not applicable.

4 **Outline** the significance to evolutionary theory of the realisations by geologists that the Earth was more than a few thousand years old.

Speciation – the basis of microevolution D2.1–2.11

Present-day flora and fauna have arisen by change from pre-existing forms of life. Most biologists believe this. This process has been variously called 'descent with modification', 'organic evolution', and 'microevolution', but perhaps **speciation** is appropriate here because it emphasises that species change.

So, what is a species?

The Swedish botanist Karl Linnaeus (1707–78) devised the binomial system of nomenclature (page 164) in which every organism has a double name consisting of a Latinised generic name (**genus**) and a specific adjective (**species**). There was no problem in Linnaeus' day in defining species because it was believed that each species was derived from an original pair of animals created by God. Since species had been created in this way, they were fixed and unchanging.

In fact, the fossil record provides evidence that changes do occur in living things. Humanoid and human fossils alone illustrate this point, as we shall see. Taxonomists now use as many different characteristics as possible in order to define and identify a species. The three main characteristics used are:

- external and internal structure (**morphology** and **anatomy**);
- **cell structure** (whether cells are eukaryotic or prokaryotic);
- **chemical composition** (comparisons of nucleic acids and proteins and the immunological reactions of organisms, page 497).

A species is a group of organisms of common ancestry that closely resemble each other structurally and biochemically, and which are members of natural populations that are actually or potentially capable of breeding with each other to produce fertile offspring, and which do not interbreed with members of other species.

The last part of this definition cannot be applied to self-fertilising populations or to organisms that reproduce only asexually. Such groups are species because they look very similar (morphologically similar), and because they behave and respond in similar ways, with bodies that function similarly (they are physiologically similar).

But however we define the term, since species may change with time (mostly a slow process), there is a time when the differences between members of a species become significant enough to identify separate **varieties** or **subspecies**. Eventually these may become new species. All these points are a matter of judgement.

5 **Comment** on the differences between a variety and a species. **State** an example of each.

Related organisms form populations

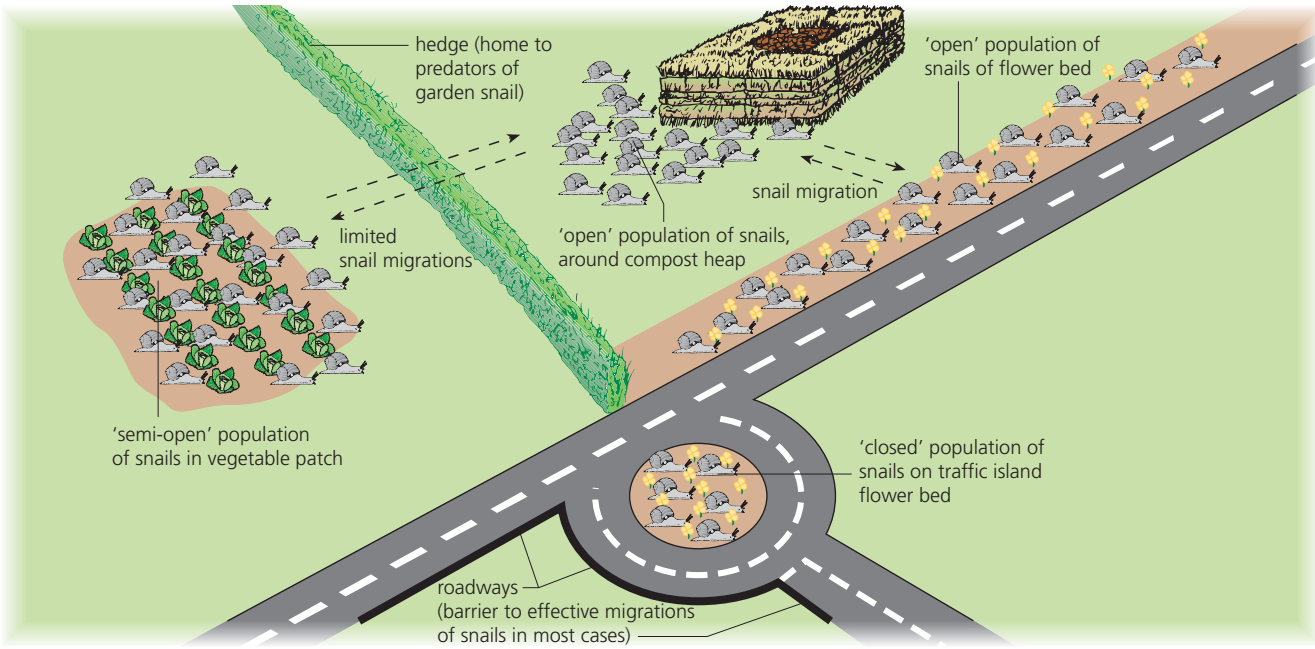
In nature, organisms occur in local populations. Therefore, we can look to local populations as the venue for evolution.

A population is a group of individuals of a species, living close together, and able to interbreed.

So a population of garden snails might occupy a small part of a garden, say around a compost heap (Figure 16.5). A population of thrushes (snail-eating birds) might occupy several gardens and surrounding fields. In other words, the area occupied by a population depends on the size of the organism and on how mobile it is, for example, as well as on environmental factors (e.g. food supply, predation, etc.).

The boundaries of a population may be hard to define. Some populations are fully open, with individuals moving in or out, from nearby populations. Alternatively, some populations are more or less closed – that is, isolated communities almost completely cut off from neighbours of the same species. Obviously, the fish found in small lakes are a good example of the latter.

Figure 16.5 The concept of population



Population genetics

Population genetics is the study of genes in populations. In any population, the total of the alleles of the genes located in the reproductive cells of the individuals make up a **gene pool**.

A gene pool consists of all the genes and their different alleles, present in an interbreeding population.

When breeding between members of a population occurs, a sample of the alleles of the gene pool will contribute to the genomes (gene sets of individuals) of the next generation, and so on, from generation to generation. Remember, an allele is one of a number of alternative forms of a gene that can occupy a given locus on a chromosome. The frequency with which any particular allele occurs in a given population will vary.

Allele frequency is the commonness of the occurrence of any particular allele in a population.

When allele frequencies of a particular population are investigated they may turn out to be static and unchanging. Alternatively, we may find allele frequencies changing. They might do so quite rapidly with succeeding generations, for example.

When the allele frequencies of a gene pool remain more or less unchanged, then we know that population is static as regards its inherited characteristics. We can say that the population is not evolving.

However, if the allele frequencies of the gene pool of a population are changing (the proportions of particular alleles are altered – we say they are disturbed), then we may assume that evolution is going on. For example, some alleles may be increasing in frequency because of an advantage they confer to the individuals carrying them. With possession of those alleles, the organism is more successful. It may produce more offspring, for example. If we can detect change in a gene pool we may detect evolution happening, possibly even well before a new species is observed.

Note for 'Higher' students: The **Hardy–Weinberg formula** may be used to detect change in allele frequency, in practical situations (page 493).

Speciation

Speciation, the evolution of new species, requires that allele frequencies change with time in populations. Next we look into some of the processes known to bring about significant change, leading to the eventual appearance of a local population of organisms that are a new species, unable to breed successfully with members of the population from which they originated.

6 **Explain** what is meant by genetic difference.

Speciation by isolation

A step towards speciation may be when a local population becomes **isolated** from the main bulk of the population, so the local gene pool is completely cut off and permanently isolated. The result is reproductive isolation within the original population. Even when reproductive isolation has occurred, many generations may elapse before the composition of the gene pool has changed sufficiently to allow us to call the new individuals a different species. However it does happen, and isolation that is effective in leading to genetic change can occur in **space** (geographical isolation), **time** (temporal isolation) and as a product of **behaviour** (behavioural isolation).

Geographical isolation

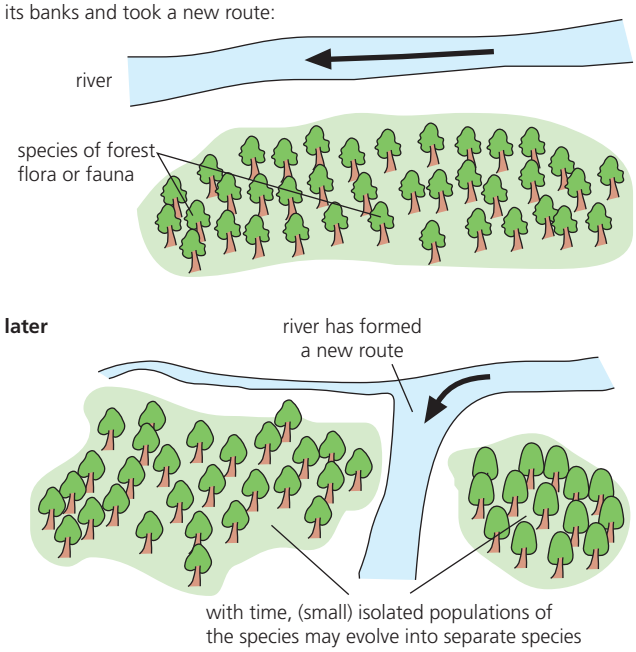
This is the consequence of the development of a barrier within a local population. Today, both natural and human-imposed barriers can occur abruptly, sharply restricting movement of individuals (or their spores and gametes, in the case of plants) between divided populations (Figure 16.6). Before separation, individuals shared a common gene pool, but after isolation, 'disturbing processes' like natural selection, mutation and random genetic drift may trigger change.

Genetic drift is random change in gene frequency in small isolated populations.

For example, a new population may form from a tiny sample that became isolated and separated from a much larger population. While numbers in the new population may rapidly increase, the gene pool from which they formed might have been totally unrepresentative of the original, with many alleles lost altogether. The outcome of these processes may be marked divergence between populations, leading to their having distinctly different characteristics.

Figure 16.6
Geographical barriers

1 isolation by a new, natural physical barrier
A natural habitat became divided when a river broke its banks and took a new route:



2 isolation by a human-imposed barrier
The by-pass at Newbury cuts through established habitats, separating local populations

Geographic isolation also arises when motile or mobile species are dispersed to isolated habitats – as, for example, when organisms are accidentally rafted from mainland territories to distant islands. The 2004 tsunami generated examples of this in south-east Asia. Violent events of this type have surprisingly frequently punctuated world geological history.

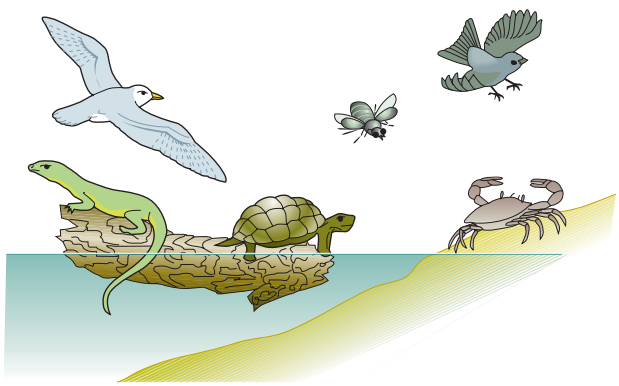
Charles Darwin visited the isolated islands of the Galapagos, off the coast of South America, during his voyage with *The Beagle* in 1831–36 (Figure 16.7). The islands are 1000 kilometres from the South American mainland. The origin of these islands is volcanic, and they appeared out of the sea about 16 million years ago, so they were uninhabited, initially. Today they have a flora and fauna which relates to mainland species.

Darwin encountered examples of population divergence on the Galapagos. For example, the tortoises found on these islands had distinctive shells. With experience, an observer could tell which individual island an animal came from by its appearance, so markedly had the local, isolated populations diverged since their arrival from the mainland.

Figure 16.7 The Galapagos Islands and species divergence there

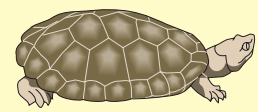
Many organisms (e.g. insects and birds) may have flown or been carried on wind currents to the Galapagos from the mainland. Mammals are most unlikely to have survived drifting there on a natural raft over this distance, but many large reptiles can survive long periods without food or water.

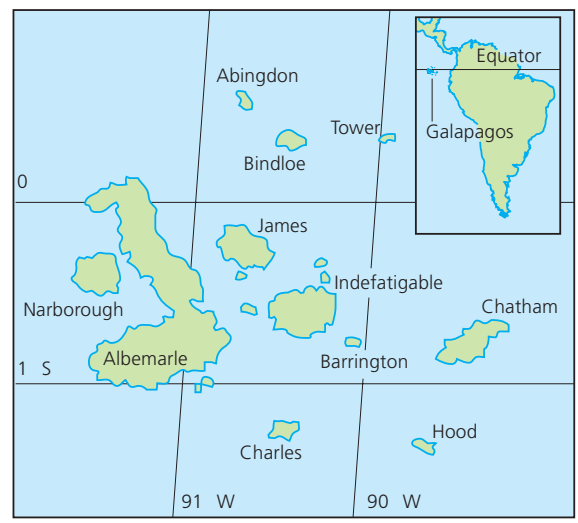
immigrant travel to the Galapagos



The Galapagos Islands

Today the tortoise population of each island is distinctive and identifiable.





The **giant iguana lizards** on the Galapagos Islands became dominant vertebrates, and today are two distinct species, one still terrestrial, the other marine, with webbed feet and a laterally flattened tail (like the caudal fin of a fish).

terrestrial iguana

marine iguana

The iguana lizard here had no mammal competition when it arrived on the Galapagos. It became the dominant form of vertebrate life, and was extremely abundant when Darwin visited. By then two species were present, one terrestrial and the other fully adapted to marine life. The latter is assumed to have evolved locally as a result of pressure from overcrowding and competition for food on the islands (both species are vegetarian) driving some members of the population out of the terrestrial habitat.

Temporal isolation

This is illustrated when two very closely related species occupy the same habitat and differ only in the time of year that they complete their life cycles. Reproductive isolation may develop in this situation within a local population so that some members produce gametes at distinctly different times of the year from others; thus, two distinctive gene pools start to evolve.

Examples of the outcome of temporal isolation include two members of the genus *Pinus* found in Californian forests (Figure 16.8).

Figure 16.8 Related species, the products of temporal isolation

Right: ***Pinus radiata*** (Monterey pine) produces fertile cones in February

Far right: ***Pinus attenuata*** (knobcone pine) produces fertile cones in April

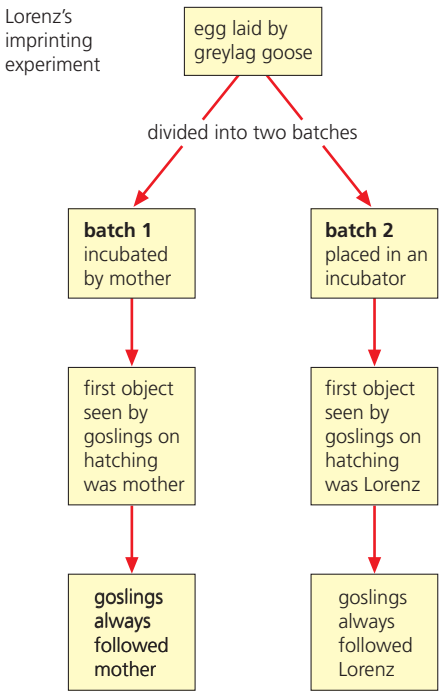
Behavioural isolation

This type of isolation results when members of a population acquire distinctive behaviour routines in their growth and development, courtship or mating process that are not matched by all individuals of the same species. An example occurs in the imprinting behaviour of the young of geese, swans and other birds. When chicks of these species hatch out of the egg, the adult birds are in the vicinity, caring for them. The young imprint the image of their parents as they relate to and learn from them. They associate socially only with their own species (or variety), and as adults, they will eventually only bond with and breed with their own species.

Imprinting became apparent when a goose chick, on hatching, was placed with swan adults as parents. That goose, when an adult, bred with a swan, and the offspring was an infertile 'Gwan' (Figure 16.9). Clearly, the swan and goose are related species that have evolved apart for long enough for their progeny to be infertile, but not long enough to exclude the formation of a hybrid.

Other examples of behavioural isolation are demonstrated by closely related species of fish, including in guppies (*Poecilia* spp.) with different, distinctive body markings by which pairs select their mates, and in four species of gull of the Canadian arctic (*Larus* spp.) with distinctive plumage by which they are identified during breeding periods.

Konrad Lorenz (1903–1989) discovered imprinting – the establishment of a more or less instantaneous bond (working relationship).



Imprinting underpins reproductive isolation in geese and swans. If a fertile goose's egg is added to a swan's nest and incubated with the swan's eggs, then on hatching, the goose chick imprints on the swan 'parent' birds and is brought up by them. Later, the goose may mate with a swan. If so the progeny will be a 'gwan'.

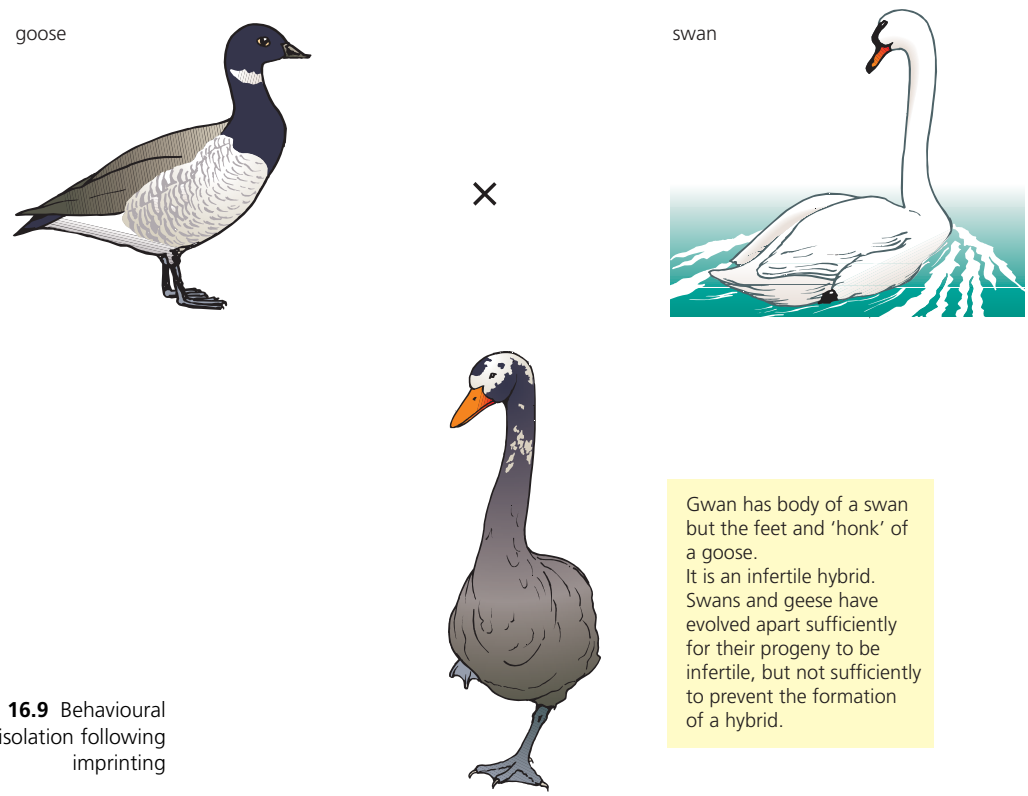


Figure 16.9 Behavioural isolation following imprinting

Speciation by polyploidy – a form of mutation

A sudden change in the genetic information of an organism, known as a **mutation**, may be heritable. **Chromosome mutations** involve a change in the structure or number of chromosomes. In fact, the types of chromosome mutation that may lead to a new species generally involve an alteration in the number of whole sets of chromosomes. An organism with more than two sets of chromosomes is called a **polyploid**.

Polyploids are largely restricted to plants and (some) animals that reproduce asexually (the sex determination mechanism of vertebrates prevents polyploidy).

In plants, polyploidy is a relatively common occurrence, and one which more or less **instantly creates a new species** (provided the polyploid survives the early period when its numbers are incredibly low).

The additional set(s) of chromosomes of a polyploid may come from a member of the same species – a situation known as **autopolyploidy**. Typically, this occurs if the spindle fails in meiosis, causing diploid gametes to be formed. A well known and economically very important example is the origin of the cultivated potato *Solanum tuberosum* ($2n = 48$), an autopolyploid of the smaller, wild *Solanum brevidens* ($2n = 24$).

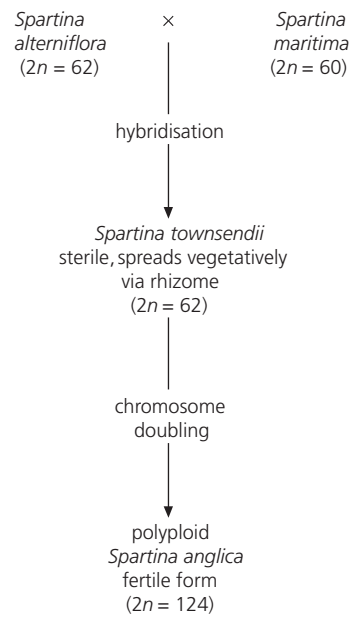
Alternatively, the sets of chromosomes of a polyploid may come from two species by hybridisation – a situation known as **allopolyploidy**. The best known example of a natural allopolyploid is the cord-grass (*Spartina townsendii*). It originated in Southampton Water, UK, around 1870, from a natural cross between a local species, *S. maritima*, and an introduced American species, *S. alterniflora*. (Marine and estuarine species are often inadvertently transported between the coasts of continents by shipping. For example, the practice of filling tankers with ballast water for return journeys inevitably carries into the ship specimens of local aquatic organisms, some of which may survive the journey.)

While *S. townsendii* is an infertile hybrid, it spreads by vigorous growth of its underground stem (rhizome). Further, an autopolyploid evolved, *S. anglica*, and this new species is both fertile and of vigorous vegetative growth. These plants stabilise mud flats of shore lines and estuaries, and may cause the blocking of water channels, too (Figure 16.10).

7 **Explain** what is meant by the statement 'all polyploids are instantly isolated from the parent plant by a barrier'.

Figure 16.10 Polyploid cord-grass, *Spartina anglica*

Spartina anglica in its natural habitat



Speciation – a summary

Apart from cases of instant speciation by polyploidy discussed above, species do not evolve in a simple or rapid way. The process is usually gradual, taking place over a long period of time. In fact, in many cases speciation may occur over several thousand years. Complex though it is, we can recognise that all cases of speciation require 'isolation'.

A deme is the name we give to a small, isolated population.

The individuals of a deme are not exactly alike, but they resemble one another more closely than they resemble members of other demes. This similarity is to be expected, partly because the members are closely related genetically (similar genotypes), and partly because they experience the same environmental conditions (which affect their phenotype).

The ways demes become isolated have been discussed already. Reviewing these, we see they fall into two groups, depending on the way isolation is brought about.

- Isolating mechanisms that involve special separation are known as **allopatric speciation** (literally 'different country').
- Isolating mechanisms involving demes in the same location are known as **sympatric speciation** (literally 'same country').

So, isolation may result from a deme becoming spatially separated from the rest of the local population, or it may occur within a local population. Either way, natural selection may come to act differently on the demes and, if this continues over a large number of generations, complete divergence may be the final outcome. In Table 16.3, allopatric speciation and sympatric speciation are compared.

Table 16.3 Comparison of allopatric and sympatric speciation

Allopatric speciation	Sympatric speciation
■ due to physical separation of the gene pool preventing organisms of related demes or their gametes from meeting	■ due to an isolating mechanism within a gene pool preventing production of viable offspring between members of related demes in the same locality
■ by geographic isolation – when motile or mobile species are dispersed to isolated habitats	■ by reproductive isolation due to: – temporal mechanisms – behavioural mechanisms – polyploidy
■ important in plant speciation	■ important in animal and plant speciation

Adaptive radiation

Evidence for the occurrence of evolution comes from diverse sources, of which adaptive radiation is one (Figure 16.11).

What is meant by adaptive radiation?

Frequently, structures present in a range of different organisms appear fundamentally similar. For example, the limbs of vertebrates appear to relate to a common plan we call the pentadactyl limb (meaning 'five-fingered'). These limbs are all described as **homologous structures** (page 159), because they occupy similar positions in the organism, have an underlying basic structure in common, but may have evolved different functions (Figure 16.12). They relate to the plan, but may also show modification from it. The underlying similarity is attributed to adaptive radiation, and these organisms share a common ancestry, however long ago they diverged.

Figure 16.11 The range of evidence for evolution

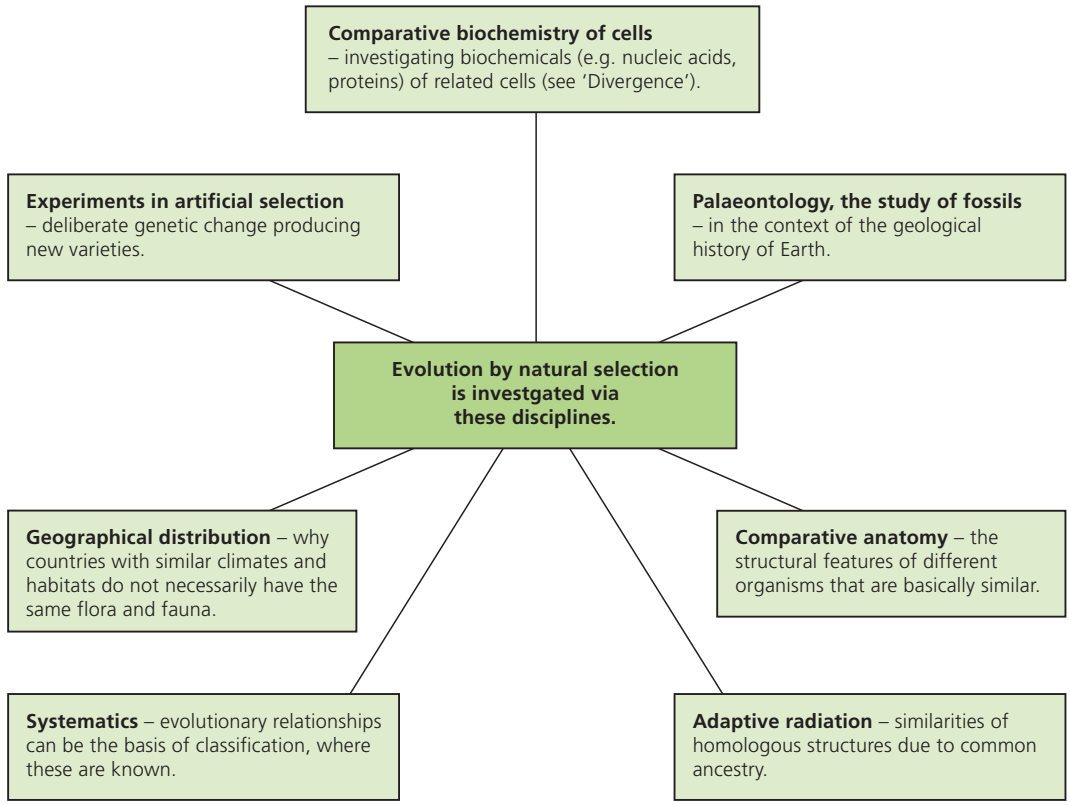


Figure 16.12 Homologous structures demonstrate adaptive radiation

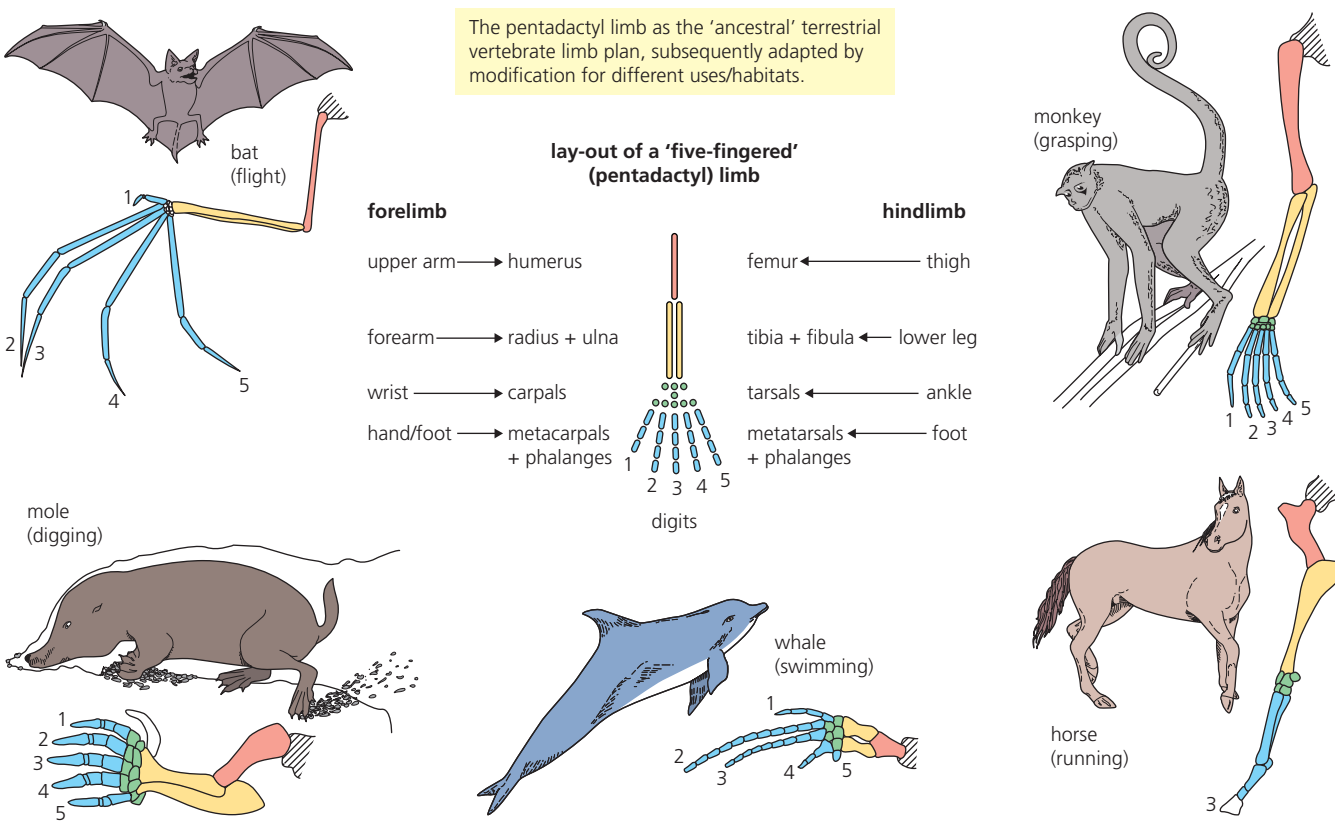
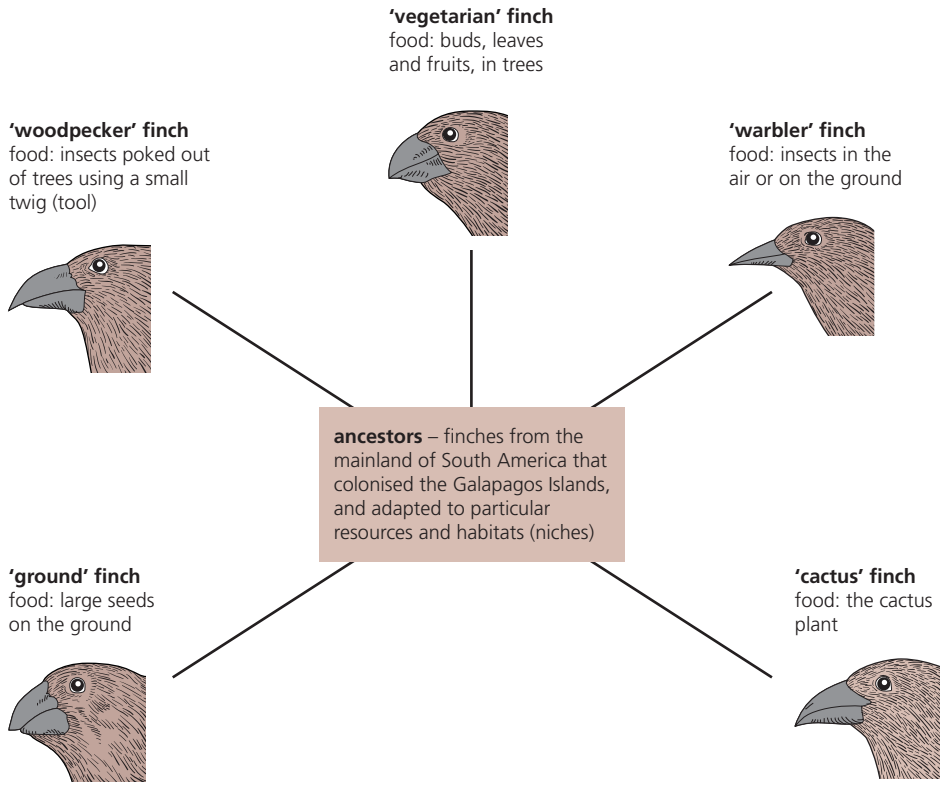


Figure 16.13 Adaptive radiation in Galapagos finches

The adaptations by the finches that reached these islands off South America seem to have minimised competition between them.



Adaptive radiation of the beaks of the finches of the Galapagos Islands caught Charles Darwin's attention, and the study of these birds was followed up by the ornithologist David Lack (Figure 16.13).

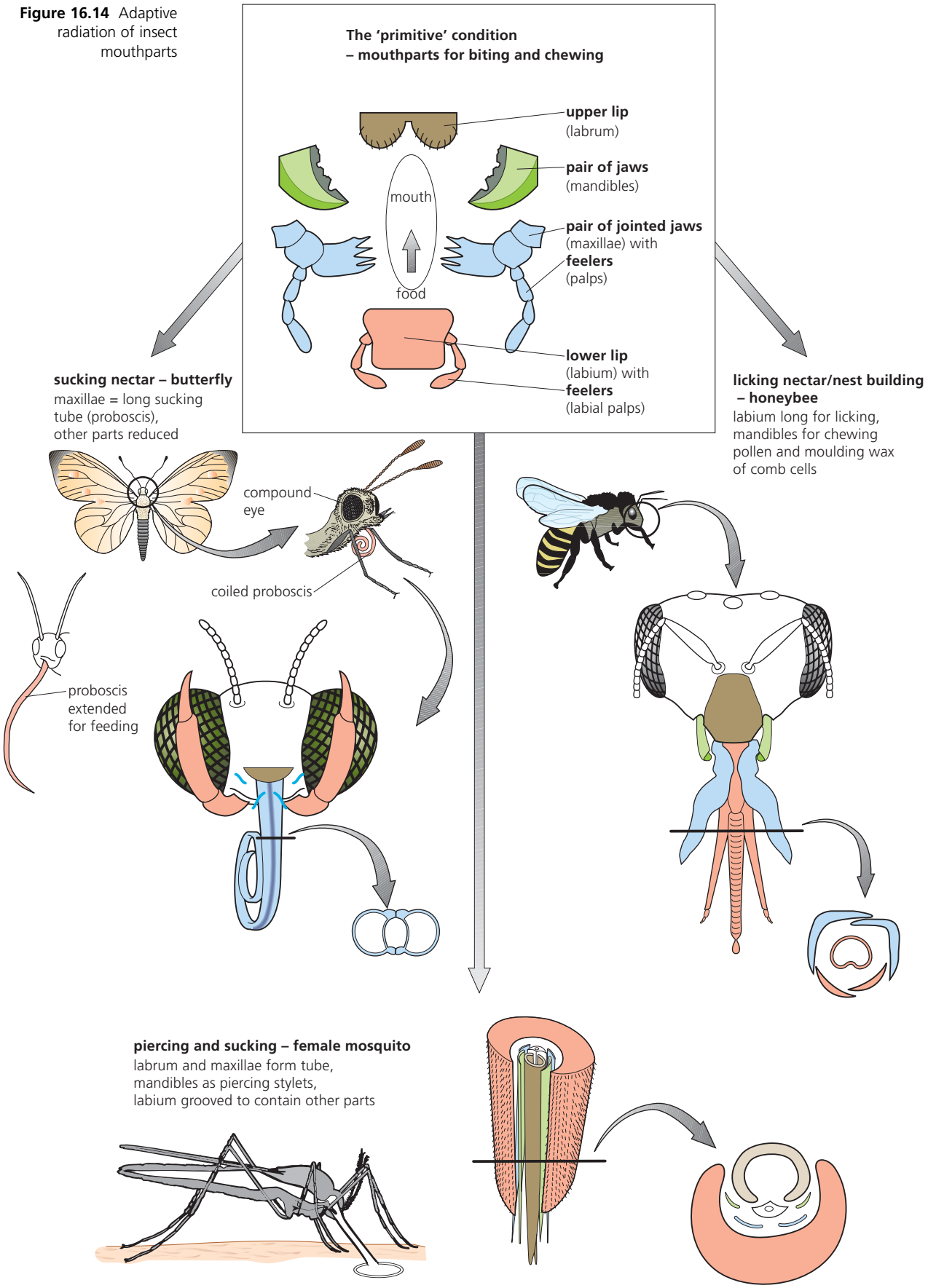
The **mouthparts of insects** are thought to have originated from paired limbs, one pair on each of the segments that developed into the head region of the insect's body (insects are members of the phylum of animals known as the arthropods, having segmented bodies, typically with one pair of jointed limbs per segment – page 172). The basic plan of the insect's mouthparts is of **lips** (labrum), **a pair of jaws** (mandibles), **a pair of jointed jaws** (maxillae), and **lower lips fused** (labium). With adaptation to alternative diets (adaptive radiation) these structures have become modified, illustrated in Figure 16.14 by the mouthparts of the butterfly, honeybee, and the female mosquito.

Convergent evolution

Some organisms resemble each other in appearance or in the way they function, or both, and yet are not closely related. These unrelated organisms show only superficial similarities, which are known as **analogous structures** (page 159).

A most dramatic example of this comes from a comparison of the mammals (fossil and living members) of Australia with those of North America and Europe. Australia became isolated (an island continent) about 120 million years ago (in the Cretaceous period), early in mammal evolution. This isolation was a product of the movement of the plates of the Earth's crust (plate tectonics). At this time, **eutherian mammals** (placental mammals) were diverging from **marsupial mammals** (without a placenta; the young are born at a more immature stage, migrate to a pouch on the mother's body, and complete development there). In Australia, only marsupial mammals developed, and for many millions of years, adapted to an environment without competition from eutherian mammals. Nevertheless, there were many similarities (in effect,

Figure 16.14 Adaptive radiation of insect mouthparts



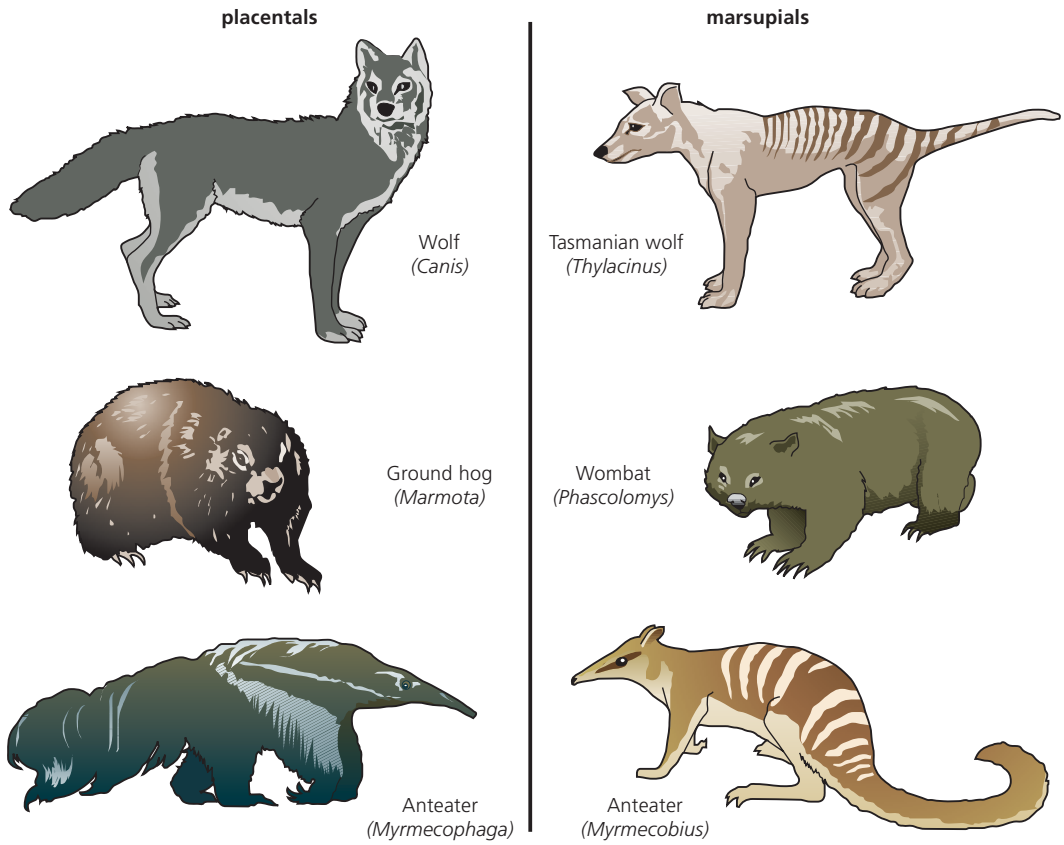


Figure 16.15 Convergent evolution of placental and marsupial mammals

parallels) between the environments in Europe and North America and in Australia, and there are corresponding similarities in the types of mammal found in comparable niches (Figure 16.15). A **niche** is the habitat an organism occupies and the mode of nutrition it employs (Chapter 19, page 605).

Convergent evolution does not arise by chance, and yet natural selection is a process without a purpose or a plan. The environment, made up of both living (biotic) and non-living (abiotic) forces, is the driving force for natural selection. We know that variation among progeny arises randomly, **but natural selection is not a random process**.

By natural selection, organisms with characteristics of structure or physiology that favour their survival are most likely to reproduce and generate offspring themselves. Many of the offspring will have those characteristics, too. In this way, we can expect many similarities to be shown by the successful organisms that come to occupy comparable niches for long periods of geological time. The result may be parallel adaptive radiation, as shown by the selection of marsupial and eutherian mammals, and may lead to convergent evolution.

Divergent evolution

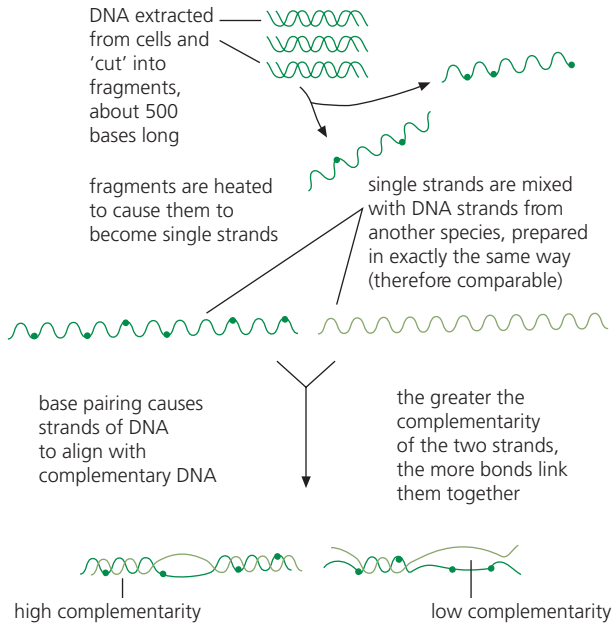
We have seen that vertebrates have limbs relating to a common plan, the pentadactyl limb (Figure 16.12). Where a structure of a distinctive type, like this, is shared by a wide range of organisms we can assume that it first appeared in a common (early) ancestor, in the history of the group. Variants of the pentadactyl limb now depart from the original by varying degrees in different vertebrates, depending how far they have diverged and diversified to survive.

The degree of divergence in the evolution of related organisms is investigated experimentally today by studies in comparative biochemistry. It is often an important source of evidence for the pathways evolution has taken. For example, investigations of the degree of compatibility in immunological proteins of mammals (page 497) and of the genetic differences between the DNA of various organisms (Figure 16.16) give us data on degrees of divergence.

8 By means of a table, **compare** convergent and divergent evolution.

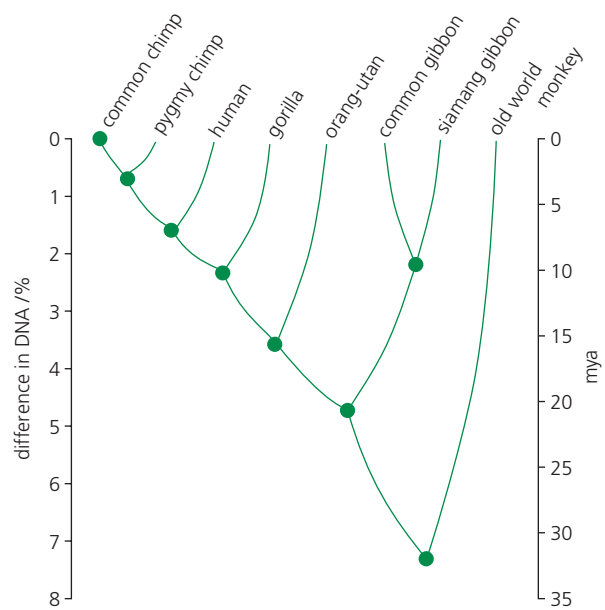
Figure 16.16 Genetic difference between DNA samples

DNA hybridisation is a technique that involves matching the DNA of different species, to discover how closely they are related.



The closeness of the two DNAs is measured by finding the temperature at which they separate – the fewer bonds formed, the lower the temperature required.

The degree of relatedness of the DNA of **primate species** can be correlated with the estimated number of years since they shared a common ancestor.



Pace of evolution: gradualism *versus* punctuated equilibria

Since geologists estimate the age of the Earth as being 4500 million years, and that life originated about 3500 million years ago (mya), the timescale over which evolution has occurred has seemed almost unimaginably long. Furthermore, the fossil record provides evidence of the long evolutionary history of most major groups. This observation of evolution by natural selection as being an exceedingly gradual process is known as **gradualism**. Indeed, from the theory of evolution by natural selection we might expect species to only gradually disappear, and be replaced by new species at a similar slow rate.

Instead, this may not have always been the case. Some new species have appeared in the fossil record relatively quickly (in terms of geological time), and then have tended apparently to remain unchanged or little changed, for millions of years. Sometimes, periods of stability were followed by periodic mass extinctions, all evidenced by the fossil record (Figure 16.17).

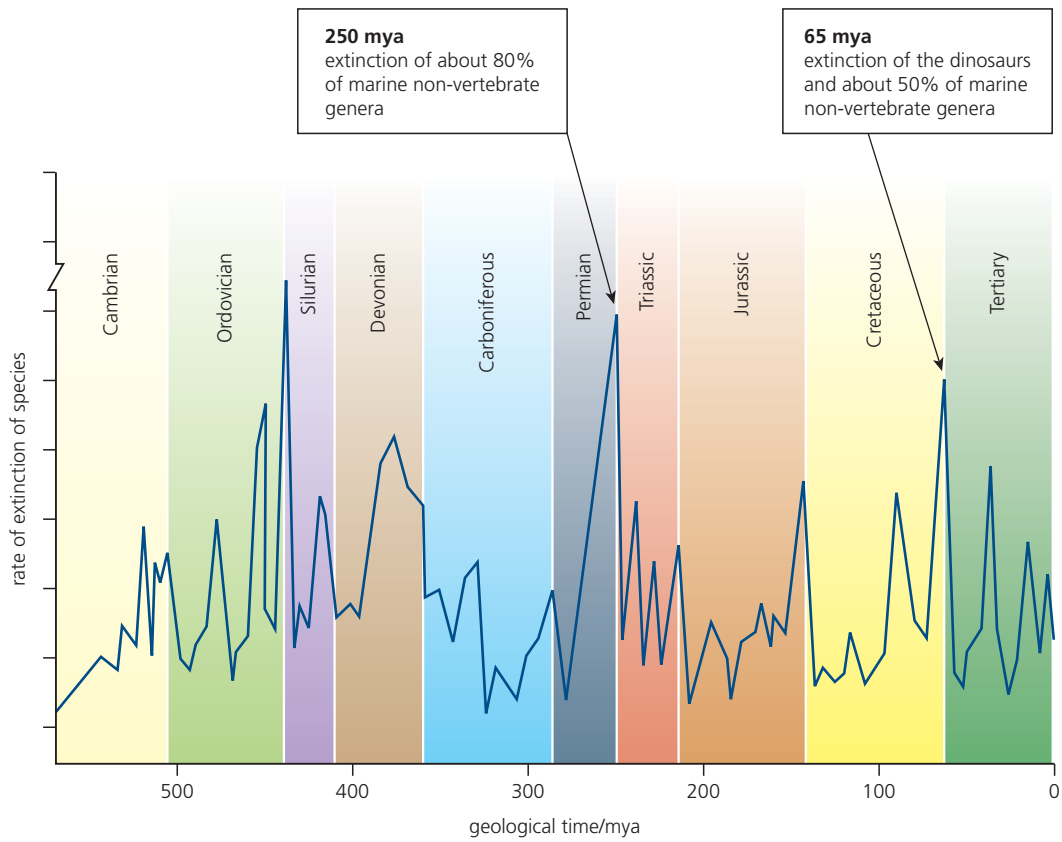


Figure 16.17 The extinction sequence in geological time

Some say the fossil record looks like this because we have a partial (distorted) fossil record, when compared to the numbers of organisms that have lived. This is quite possible; we have no way of being certain the fossil record is fully representative of life in earlier times. This is a possible explanation.

However, two evolutionary biologists, Niles Eldredge and Stephen Gould, proposed an alternative explanation. They argue that the fossil record for some groups is not significantly incomplete, but rather, accords with their hypothesis of the origins of new species, which they called **punctuated equilibria**. This hypothesis holds that:

- When environments become unfavourable, populations attempt to migrate to more favourable situations.
- If the switch to adverse conditions is very sudden or very violent, then a mass extinction occurs. Major volcanic eruptions or major meteor impacts can throw so much detritus into the atmosphere that the Earth's surface is darkened for many months, cooling the Earth and killing off much plant life.

- Populations at the fringe of a massive disturbance may be sheltered or protected from the worst effects of extreme conditions, and survive.
- Members of these populations may become small, isolated reproductive communities, from which repopulation eventually occurs.
- The surviving group(s) may have an unrepresentative selection of alleles of the original gene pool. If one becomes the basis of a repopulation event and adapts to the new conditions quickly, then abrupt genetic changes may occur. This phenomenon is known as the **founder effect**.

So there are alternative proposals for the ways natural selection has operated in practice in the establishment of life in geological time. In fact, gradualism and punctuated equilibria may not be alternatives; both may have contributed to the pattern of life on Earth in geological time.

Natural selection and polymorphism

Organisms that exist in two forms are examples of polymorphism. This phenomenon is well illustrated by **industrial melanism**.

As a result of the industrial revolution in Britain, in areas of heavy industry and the surrounding countryside, the pollutant chimney gases (e.g. SO_2) and soot (particles of unburnt carbon) killed off the tiny plants that grow on the surfaces of trees, walls and roofs, including lichens, algae and mosses. Exposed surfaces were blackened.

Dark-coloured forms of the peppered moth (*Biston betularia*), known as the **melanic forms**, tended to increase in these areas, but their numbers were low in unpolluted countryside, where **pale speckled forms** of the moth were far more common.

This effect was explained as due to natural selection – one or other form of the moth was effectively camouflaged from predation by insectivorous birds, and became the dominant species. By a twist of fate this was confirmed when, after Clean Air Acts were introduced in the UK, pollution of industrial cities was reduced, and surfaces were eventually re-colonised by epiphytic plants. Pale forms of the moth have returned as the dominant form in these areas.

The initial effect of natural selection was to favour the evolution of melanic forms. This is described as an example of disruptive selection – two extremes of a characteristic were produced, without intermediate forms. As the environmental crisis caused by pollution passed, the population of *B. betularia* returned to the original form, and so this is a case of **transient polymorphism** (Table 16.4 and Figure 16.18).

Melanic form	Environmental change	Pale speckled form
Naturally camouflaged on polluted surfaces, conferring a selective advantage. The proportion of alleles for the melanic form in the gene pool increases.	Industrialisation led to extensive deposition of soot and dust on surrounding environment and countryside, including on twigs and tree branches where moths rest.	Selective predation (when resting on heavily polluted surfaces) by insectivorous birds causes the proportion of alleles for 'pale' form to decrease
Now selectively vulnerable to predation by insectivorous birds when resting on natural surfaces.	Clean Air Acts led to gradual loss of black deposits on surfaces around areas of former heavy industry. Gradual return of epiphytic algae, lichens and moss on surfaces, and loss of soot.	Now naturally camouflaged on non-polluted surfaces, conferring a selective advantage. The proportion of alleles for the 'pale' form in the gene pool increases.

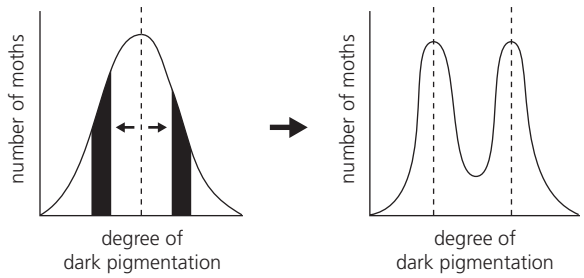
Table 16.4 *Biston betularia*; environmental change and selection pressure

A contrasting case of polymorphism is the **sickle cell trait**, the product of a gene mutation (page 100). Here, the gene that codes for the amino acid sequence of one of the components of the haemoglobin molecule is prone to a base substitution which triggers the substitution of one amino acid in the protein chain. The effect on the haemoglobin molecule is to cause clumping of the molecules in the red cell, producing sickle-shaped red cells. In this condition, the cells transport little oxygen and may even block smaller vessels. People who are heterozygous for the condition have less than 50% sickle haemoglobin. The person is said to have sickle cell trait, and they are only mildly anaemic.

Figure 16.18 The peppered moth and industrial melanism – a case of transient polymorphism

In the peppered moth (*Biston betularia*), environmental conditions have, at different times, favoured either the pale or the melanic forms.

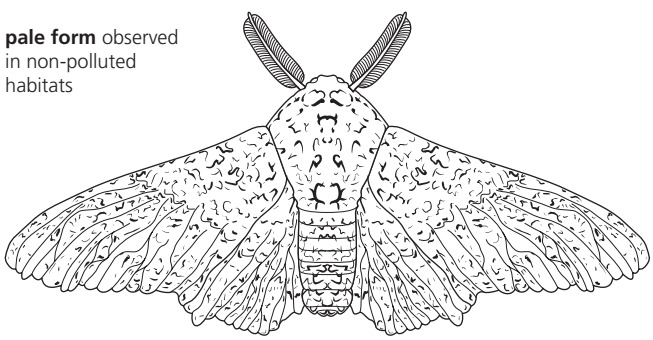
In these circumstances, the effect of natural selection on the gene pool (in the form of selective predation of moths resting on exposed surfaces by insectivorous birds) has been '**disruptive**'.



Disruptive selection favours two extremes of the 'chosen' character at the expense of intermediate forms.

Biston betularia

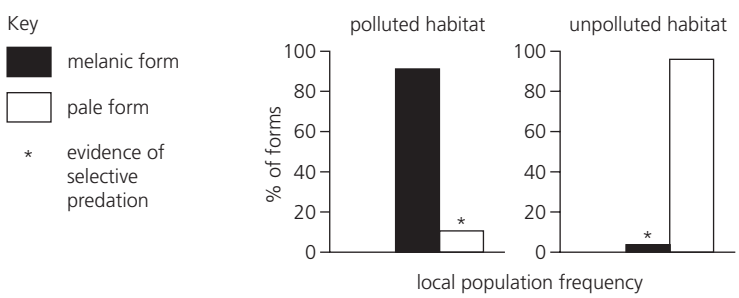
pale form observed in non-polluted habitats



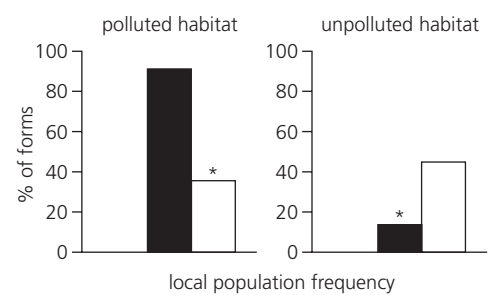
melanic form observed in industrially polluted habitats

experimental evidence that establishes transient polymorphism

results of frequency studies in polluted and unpolluted habitats



mark–release–capture experiments using laboratory-reared moths of both forms, in polluted and unpolluted habitats



There is an advantage in having sickle cell trait where malaria is prevalent. The malarial parasite completes its life cycle in red cells, but it cannot do so in sickle cells. People with sickle cell trait are protected to a significant extent. Where malaria is endemic in Africa, possession of one mutant gene (the person is heterozygous and has sickle cell trait, not full anaemia) is advantageous. Under conditions of 'survival of the fittest', this allele is consequently selected for. Fewer of the alleles for normal haemoglobin are carried into the next generation (Figure 4.11, page 101).

9 **Outline** what is meant by saying that the environment can direct natural selection.

Because of this selective advantage, the sickle cell condition is an example of **balanced polymorphism** – the stable co-existence of two (or more) distinct types of individual in a species (or population). The proportion of both alleles is maintained by natural selection.

Human evolution

D3.1–3.10

Humans belong to the mammalian order Primates. This order contains three distinctive groups of animals, namely the apes (which includes the genus *Homo*), the monkeys, and the prosimians (a name meaning 'before the monkeys'). These are mostly tree-dwelling species with grasping hands and feet. The range of animals that constitute the Primates and how they are related are summarised in Figure 16.19.

Apart from humans, who have achieved worldwide distribution, most primates live in tropical and sub-tropical regions. An interesting feature of primates is their relatively unspecialised body structure, combined with some highly sophisticated behaviour patterns, as we shall see.

How do we know about the (recent) development of the primates?

The relatedness of organisms is investigated by **comparative biochemical studies** (page 499), particularly of mitochondrial DNA which in each generation is passed from mother to offspring unchanged. This type of DNA undergoes a steady rate of mutation – it changes as a function of time alone. The degree of difference between mitochondrial DNA samples discloses how recently groups of organisms shared a common ancestor.

Also, we know something of the history of the primates from the fossil record. Much of primate evolution took place in what is today the continent of Africa. The rocks in which many fossils have been found can be dated by analysis of **naturally occurring radioisotopes**.

Dating rocks by radioisotope analysis

Atoms of certain elements exist in more than one form, and these are known as **isotopes** of that element. Isotopes are classified as stable or unstable. Stable isotopes persist in nature because they do not undergo radioactive decay. Oxygen-16 and oxygen-18 are stable isotopes of the element oxygen.

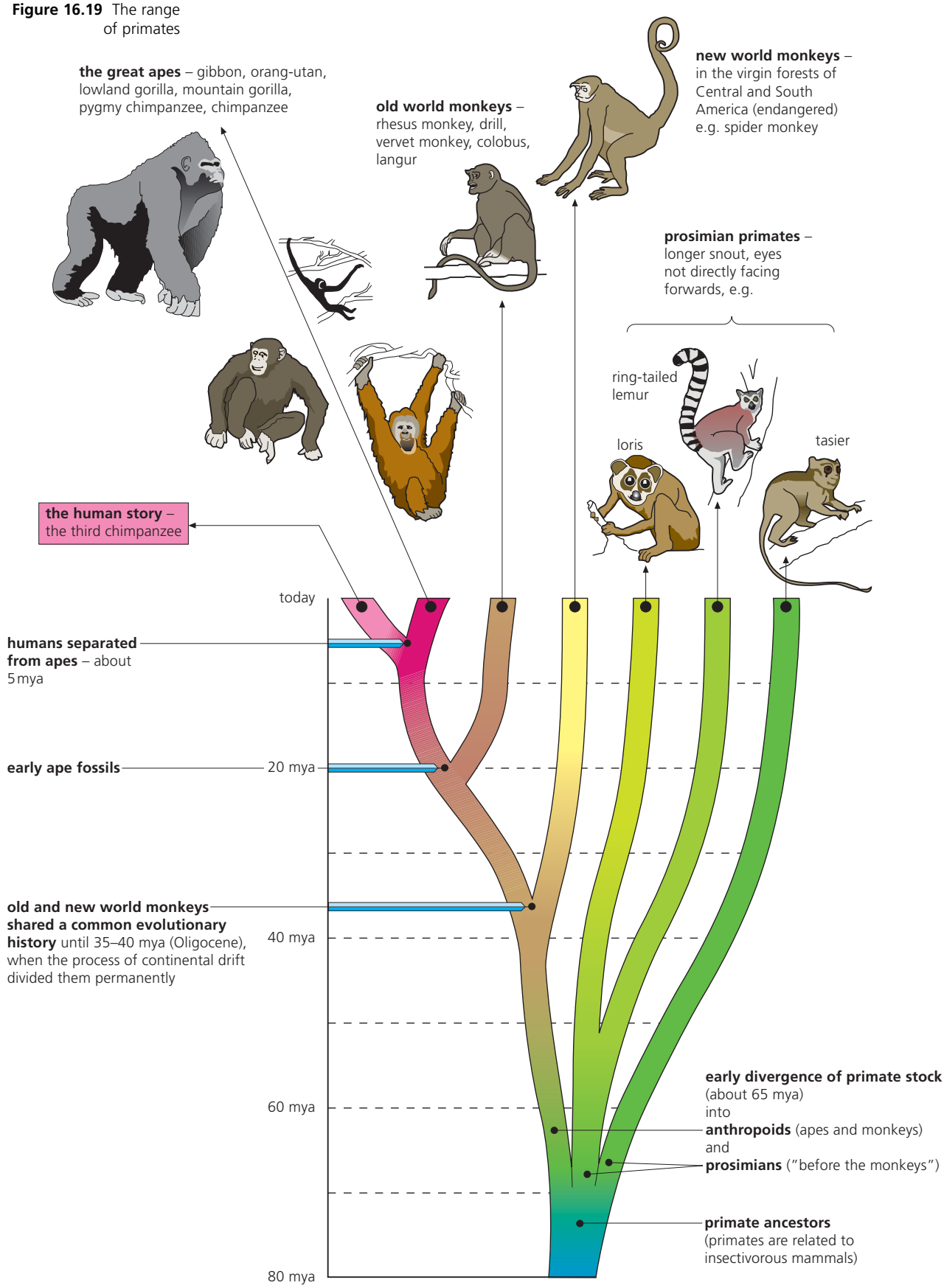
Unstable isotopes spontaneously disintegrate, often emitting α or β particles. The product of radioactive decay is a stable isotope. The decay process is slow, but it is independent of temperature and pressure that the isotope is exposed to.

Radioactive isotopes exist in the matter from which the Earth formed. As the original liquid rock cooled and crystallised, radioactive elements became trapped in the crystals formed. An example is **radioactive uranium**, which slowly decays to **lead**. In fact, radioactive uranium decays exceedingly slowly – it takes 4500 million years for half of the uranium to be converted to lead. Consequently, certain ancient rocks with varying amounts of uranium present in them can be dated by determining the ratio of uranium to lead present. However, rocks that were exposed for the first thousand million years are now submerged, so the uranium:lead ratio cannot be used to date the age of the Earth itself.

This way of estimating the age of rock is called **radiometric dating**. By comparing the amount of original isotope in a rock sample to the amount of decay product, it is possible to estimate its age. The rate of decay of a radioactive isotope is expressed as its **half-life** (Figure 16.20).

The half-life of a radioactive isotope is the time taken for the amount of radioactive isotope to fall by half.

Figure 16.19 The range of primates



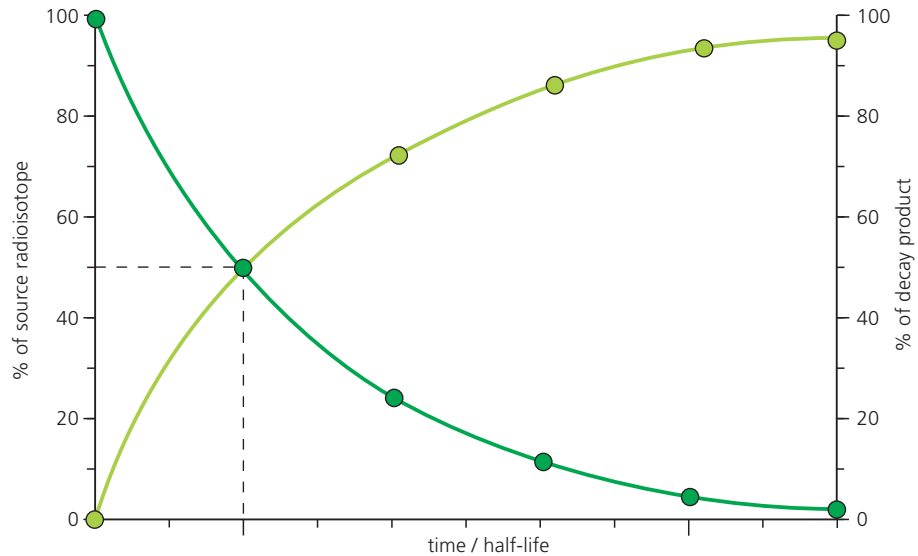


Figure 16.20 The decay curve of a radioisotope

With radiometric methods, the absolute age of rock samples (and fossils trapped in rocks) can often be estimated. The technique gives precise answers. Radioactive isotopes that take millions of years to decay may be especially useful in dating fossils or rocks.

We can examine two types of radiometric dating.

Using ^{14}C

Most carbon is ^{12}C, but due to cosmic radiation ^{14}C is formed at a low, steady rate.

While alive, organisms absorb carbon in the ratio of ^{12}C:^{14}C present in the environment around them. After death, accumulation of radioactive (and other) atoms stops. Meanwhile, ^{14}C steadily breaks down:

$$^{14}C \xrightarrow{\text{half-life of } 5.6 \times 10^3 \text{ years}} {}^{14}N$$

So the ratio of ^{14}C:^{12}C in a fossil decreases with age; the less ^{14}C, the older the fossil. **This technique gives good dates for fossils of the last 60 000 years.**

Using the ratio of ^{40}K:^{40}Ar

The pyroclastic rocks flowing out of volcanoes may contain radioactive isotopes such as potassium-40 which decays to argon-40, as shown:

$$^{40}K \xrightarrow{\text{half-life of } 1.3 \times 10^9 \text{ yrs}} {}^{40}Ar \text{ (gas)}$$

In hot lava, argon gas boils away into the atmosphere. Once lava has solidified by cooling, which occurs quickly after volcanic eruptions, the argon gas that is then formed by radioactive decay is trapped in the rock. By measuring the ratio of ^{40}K:^{40}Ar in lava deposits, the exact ages of the lava and the approximate age of the sedimentary rocks (and their fossils) below and above lava layers are estimated.

This technique spans the whole of geological time back to the Cambrian period (580 million years ago), but it is too slow to give reliable results over the most recent half million years.

10 **Distinguish** between radioactive potassium dating and radioactive carbon dating.

Humans as mammals and primates

To the biologist, humans are primate mammals. By this we mean that humans show many of the characteristics of other mammals, the general characteristics common to other primates, and many of the features shown by the great apes to which we are most closely related (Table 16.5, and Figures 16.21, 16.22, 16.23 and 16.24).

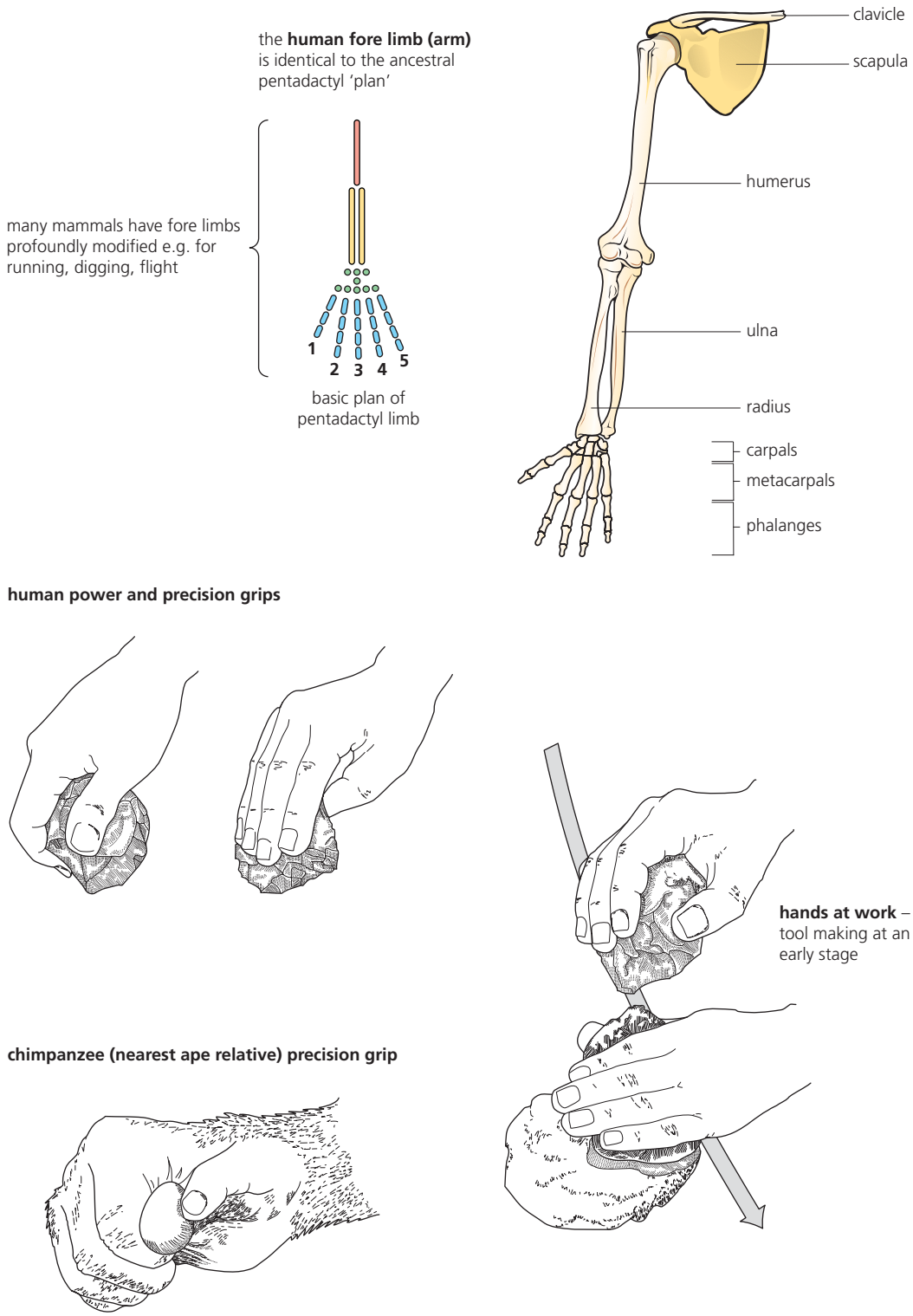
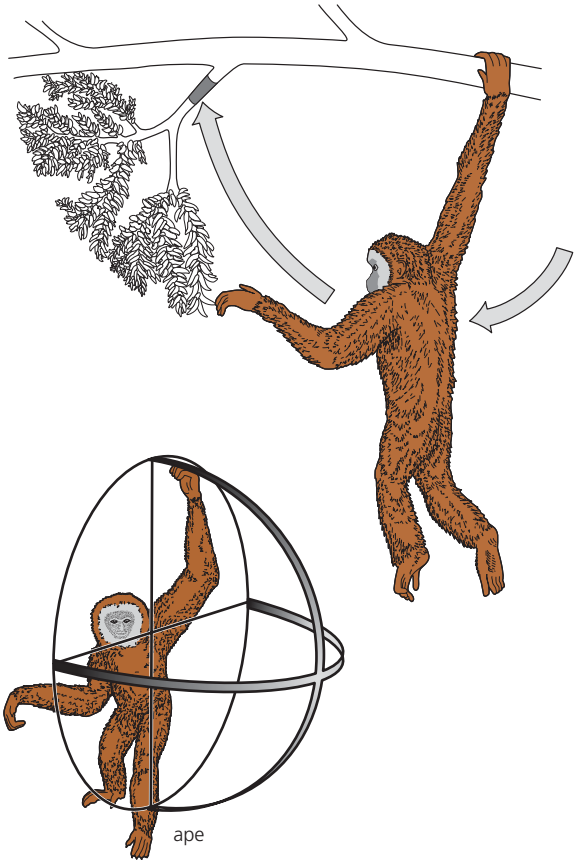
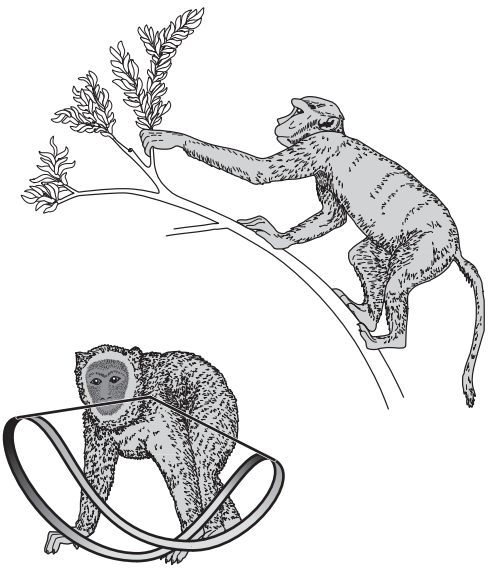


Figure 16.21 Humans as primates – powerful, precision hands

Figure 16.22 Humans as primates – mobile limbs



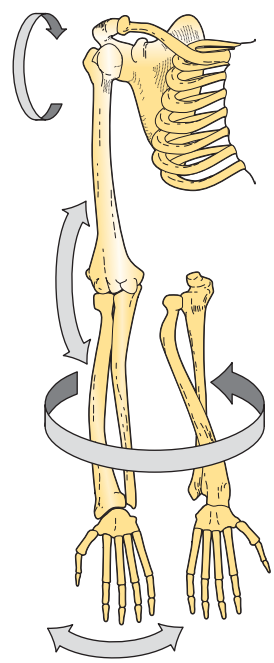
Range of arm movement
Compared with an ape, the monkey has limited movements of the arm. The monkey walks among the canopy, reaching down from the branches strong enough to take its weight, and sometimes jumping, but not brachiating.



Human arm has similar flexibility to that of the ape fore limb

Swinging from arm to arm (brachiation) is possible because:

- pectoral girdle (shoulder blade) and clavicle (collar bone) protrude beyond rib cage, so ball and socket joint of arm allows 360° movement
- elbow joint allows arm to straighten completely
- lower arm is able to twist by maximum rotation of radius and ulna
- wrist allows extensive movement of hand.



humans exploit their ability to brachiate

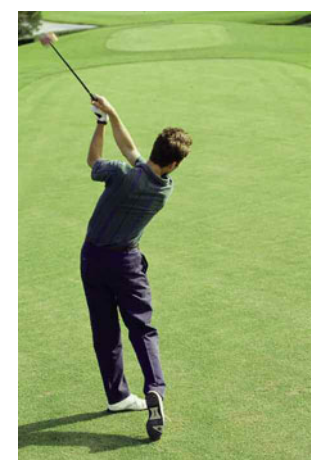
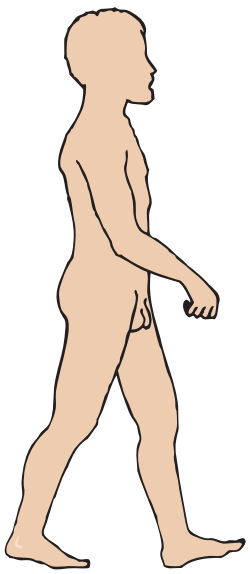
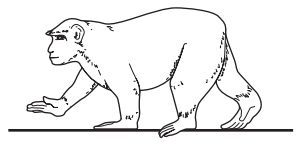


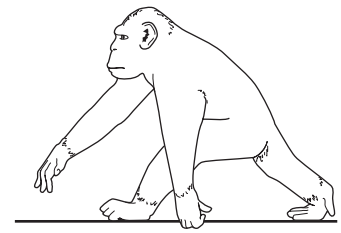
Figure 16.23 Humans as primates – upright gait and straight walking

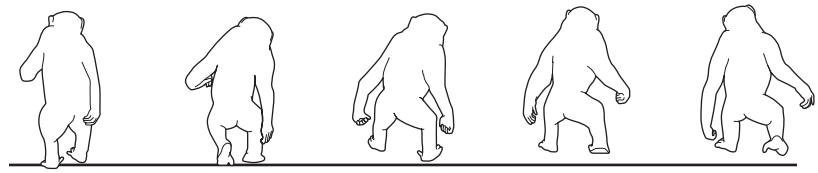


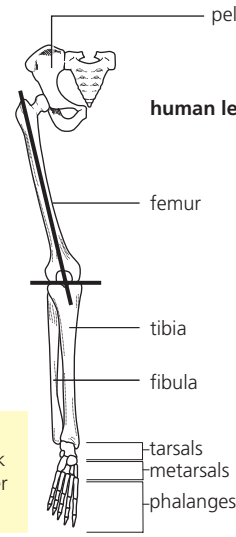
Humans walk upright, straight, and smoothly, with legs swinging below the body, supporting body weight and the propelling forward movement, alternately.



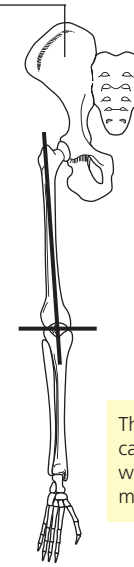
Apes run on four legs, or 'knuckle walk', but when they stand on hindlegs they can only 'waddle'.







The angle of the femur allows a human to walk upright, with feet under the centre of gravity.

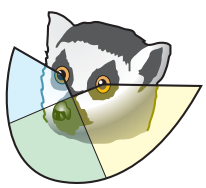


The angle of the femur causes an ape to 'waddle' when attempting bipedal motion.

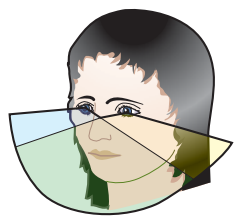
Figure 16.24 Humans as primates – developed eyes and nose

In the apes and human there is an **emphasis on vision**:

- data from the eyes is processed in the large visual cortex area of the brain, situated in the cerebral hemispheres
- impulses from the visual cortex to other parts of the brain integrate incoming information with other sensory data, with memory and with activity
- both eyes are used to view (largely) the same visual field (depth perception – stereoscopic vision)
- retina is developed to increase acuity (detail discrimination) and colour discrimination



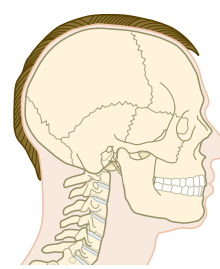
prosimian field of view



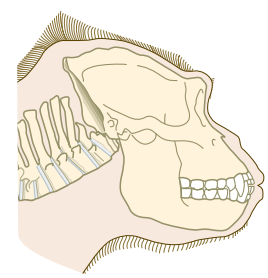
human field of view

Development of senses as a whole supports acrobatic skills in trees, searches for food and the demands of social communication (facial expressions, gestures, and – in humans – speech).
In humans, hearing (the ears) and sight (the eyes) are more advanced than the sense of smell. In human skulls, the nasal area is reduced, and a 'snout' feature is completely lost.

human head and gorilla head in side views



'compressed jaw' supports fewer teeth



Mammalian characteristics	Humans as primates
The young are nourished by milk which is secreted by modified skin glands, known as mammary glands.	Humans retain ancestral pentadactyl limbs with five digits and marked mobility of the digits, opposable finger and thumb, and power and precision grips (Figure 16.21).
Skin is covered by hair which functions as an extension of the sensory system (hair mainly functions as a heat-insulating layer in most mammals).	Free mobility of limbs; the arrangement of bones at shoulder permits brachiation movement; the radius and ulna are unfused (Figure 16.22).
Skin has sweat glands (secreting sweat) and sebaceous glands (secreting a greasy liquid onto hairs).	Development of upright gait with extensive head rotation (Figure 16.23).
The cranial cavity (brain box) is large. Below it, the secondary palate separates the nasal passage from the mouth, allowing food to be retained and chewed by teeth, before swallowing.	Development of nervous system permits precise rapid control of musculature. Brain is large and complex, especially parts for vision, tactile inputs, muscle co-ordination, memory and learning.
Brain has large cerebral hemispheres and a large cerebellum. The senses of sight, hearing and smell are often acute.	Enlargement of eyes allows an increased amount of light to be received. There is development of retina for sensitivity to low levels of illumination, and for colour vision.
There are three bones of the middle ear, and there is an external ear.	The eyes look forwards with overlapping visual fields giving stereoscopic vision.
Thoracic and abdominal cavities are separated by a muscular diaphragm. Ribs and diaphragm assist in breathing.	Reduction in the apparatus and function of smell, with flattened snout, and a reduction in the number of teeth (Figure 16.24).
Mammals are highly active, with a high metabolic rate, a tendency to play and learn, and a marked commitment to parental care.	Lengthening of the pre-natal and post-natal dependency period. Humans exist in persisting social groupings – family life.

Table 16.5 Humans – their features as mammals and as primates

From apes to humans in 35 million years

The earliest fossils which we confidently identify as anthropoids (apes) have been found at many sites in Africa (Figure 16.25). They date from about 35 mya. Humans clearly demonstrate one form of anthropoid body organisation, so we can say **the human story has taken about 35 million years to unfold**.

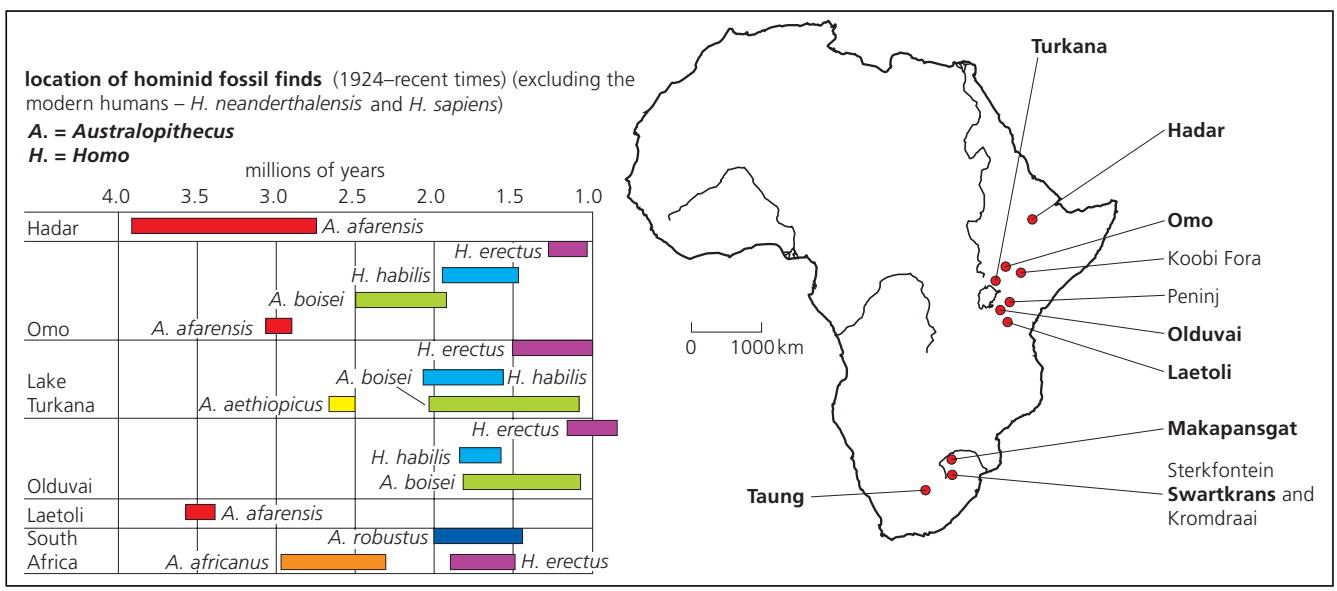


Figure 16.25 Major fossil sites

Apes are distinguished from monkeys by several important features which are reflected in their fossils. These early anthropoids had become agile, rainforest dwellers, mostly daytime feeders on fruit and vegetation. Today, many of these features have persisted in the surviving apes (chimpanzees, gibbons, gorillas, humans and orang-utans). One of the most important is brachiation (progression between trees by arm-swinging – Figure 16.22).

At the time that apes and humans were evolving, the climate was changing dramatically (with long periods of unchanged conditions, too). One reason for this was the drifting movements of the Earth's tectonic plates on which continents are located. For example, it was only 16–17 mya that Africa collided with Eurasia, and took up a position close to its current one. A direct consequence of this changing climate was that, over this period, the extent of the African forests, including rainforest, changed. Sometimes they covered the continent, but at other times they were reduced to a few isolated pockets about the size and in the position of modern Zaire. Environmental pressures of these proportions would have driven those apes that could adapt to move to the savannah, a factor that must have contributed to the development of bipedalism.

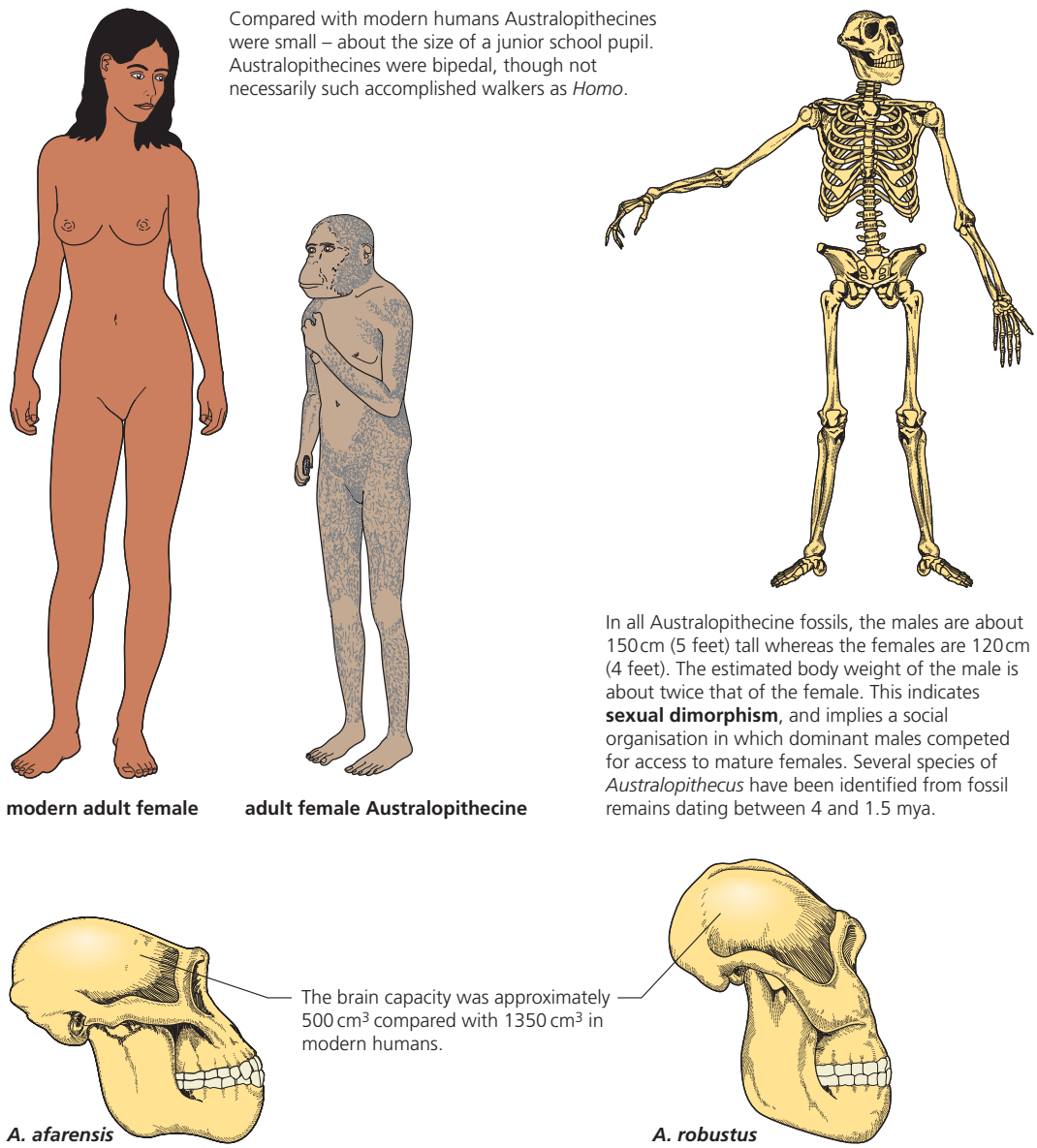


Figure 16.26 Reconstructions of Australopithecine species

'Southern apes' – the first hominids

Australopithecines (the name means 'southern apes') lived in Africa from 5 to 1.5 mya (Figure 16.26). The oldest known southern ape is *Australopithecus ramidus*. Remains of this oldest known hominid form were discovered at a site in Ethiopia in 1994. The rocks at the site are dated at 4.4 mya. Parts of skulls of this genus have been uncovered in various locations (including at Taung, in 1924), prior to the discovery of a new hominid fossil, first known as Lucy, at Hadar in Ethiopia in 1974.

Lucy is identified as *Australopithecus afarensis*. She was ape-like in that she had the same limited brain capacity as ape species of the period, but hominid-like in that she was a powerful, upright walker (the pelvis was of characteristically human form) and had no long muzzle. We now recognise that upright walking (known as **bipedalism**) was an early stage in the evolution of the hominids. The Lucy fossil was laid down 3 mya.

11 **Outline** how the ages of fossils sandwiched between layers of solid volcanic lava can be accurately dated (page 479).

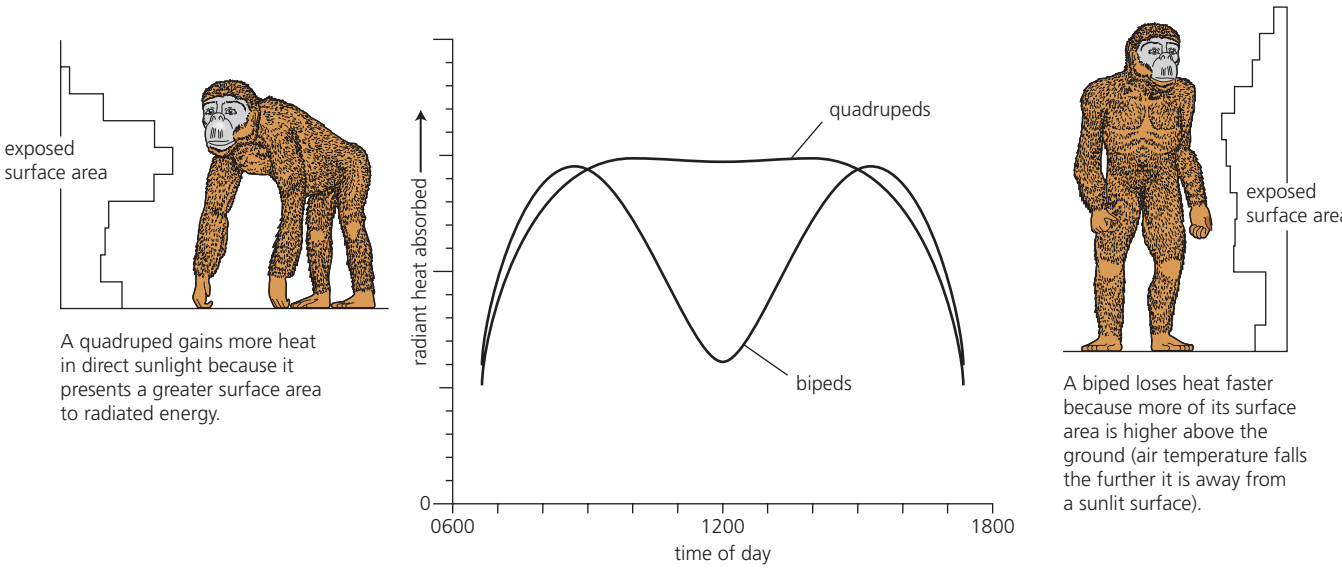
We are confident about bipedalism at this time, because of the discovery of the footsteps at Laetoli, imprinted in volcanic ash, 3.6 mya (Figure 16.27). The soft ash was presumably moistened by rain (no additional prints added), immediately baked into hard rock, and then buried by soil blown in. The footsteps were discovered in 1976. Two adults had walked in line, in a northerly direction, with a youngster who later ran off to one side. Being volcanic ash, this trace fossil can be dated precisely by the potassium:argon ratio method.

Figure 16.27 The Laetoli footprints

12 **Suggest** an advantage of keeping the head region cooler, even if the body temperature has to rise slightly.

One advantage of bipedalism (perhaps the chief advantage, initially) is as a mechanism to prevent the head region of the body overheating at the high midday temperatures of equatorial latitudes (Figure 16.28). Australopithecines lived in mosaic environments: part tropical rainforest, part woodland and tree-savannah, part scrub. Wherever they lived, no doubt they preferred to shelter at times of greatest temperature. But they may have often needed to travel to new venues, visit water holes, or scavenge and collect food at times when faster and stronger predatory animals were most likely to be resting. If so, being bipeds gave them an advantage.

Figure 16.28 Bipedalism – how to stay cool in the equatorial sun



Another critical advantage of bipedalism is that hands are freed for obtaining and carrying food. Apes breed slowly, producing few offspring at a time. A male ape that had mastered bipedalism could improve his mate's reproductive capacity by feeding her, thus freeing her to concentrate on the production and rearing of young. The genes of apes with a tendency for bipedalism will have had a better chance of replication in future generations. This would have been particularly effective in male–female pairs, rather than in troops of primates where males invested time and energy maintaining dominance over the females. On this account, hominids would have tended to be monogamous apes with lessened sexual dimorphism (males the same size as females).

Early humans – enlarged brains and busy hands

Homo habilis

The first fossil sufficiently human-like to be placed in the genus *Homo* that we know about was *H. habilis*. This hominid occurred in Africa from 2 to about 1.5 mya. Habilines, as they are known, are clearly distinguished from Australopithecines by lighter cranial bones and by an enhanced brain capacity (typically about 750–800 cm^3).

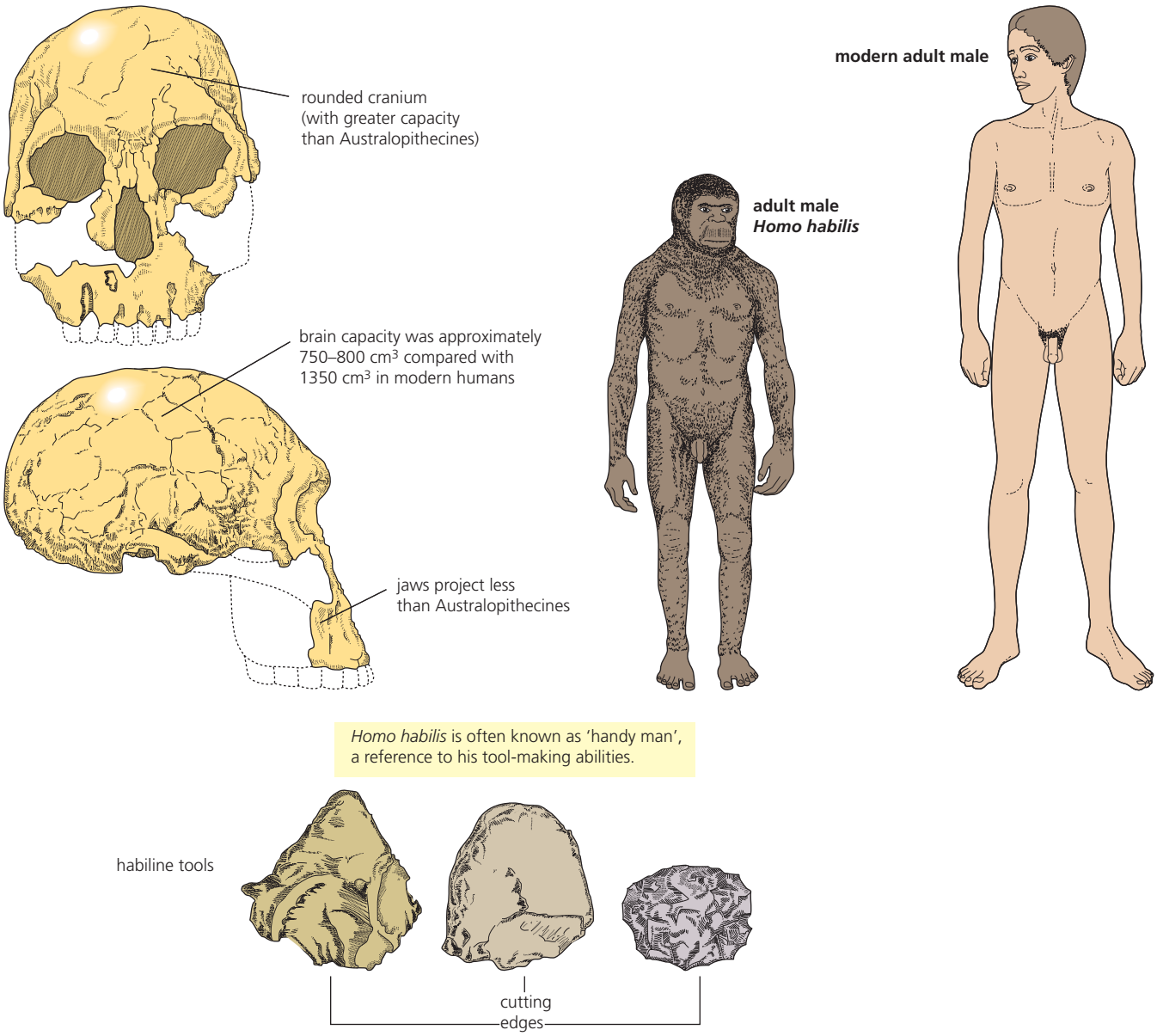
Habilines were the first hominids to be associated with tools – they used large pebbles, chipped in at least two directions, as sharpened implements to crush, break and cut. Their additional brain capacity had resulted in advanced manual dexterity (Figure 16.29). It was applied to the making and using of simple tools (selected strong stones) to chip pebbles, for a purpose. Using tools to make tools (i.e. the development of a tool industry) is what distinguishes hominid tool-makers from all other tool-users in the living world.

13 **Suggest** examples of the use of simple tools by animals other than human.

Homo erectus

Fossil remains of *Homo erectus* have been discovered in strata dating from 1.8 mya up to as recently as 200 000 years ago (ya). This was the first hominid to have extended its range beyond Africa, and there are records of its habitation in east and north Asia, and over much of Europe (Figure 16.30). *Homo erectus* is also distinguished by its body size (adults were typically 150–180 cm high) and by the size of its brain (900–1100 cm^3).

Figure 16.29 *Homo habilis* fossils, and reconstruction of early Habilines



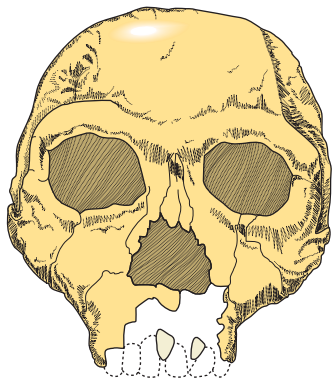
Skull endocasts (casts of the inside of the brain case of the skull) show that the areas of the brain associated with speech and language are significantly developed, so we can assume that cultural evolution (page 491) was also under way. This was also the first hominid to use fire consistently, which will have aided the colonisation of areas so far north of equatorial Africa, and also with its habit of eating meat.

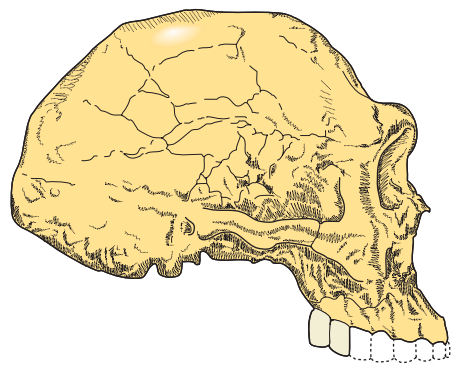
By modern human standards, *H. erectus* had a marked brow-ridge and protruding jaws, but the pronounced sexual dimorphism of earlier hominids was reduced – adult males were now only about 20–30% larger than females.

Figure 16.30 *Homo erectus* hominids, their diaspora and artefacts

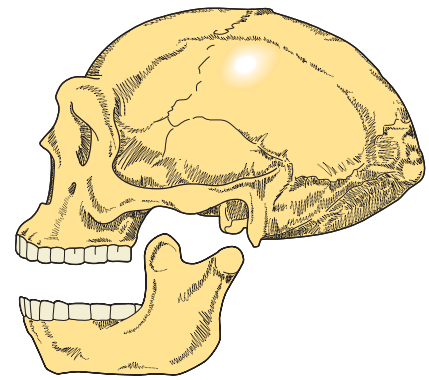
Fossil records show *H. erectus* lived from 1.8 mya (Africa) to as recently as 250 000–200 000 ya (in Asia). Skulls show enlarged cranium, but retain the pronounced brow-ridge. They lack the 'chin' of *H. sapiens*.

Skull from Turkana, Kenya site
lived about 1.8 mya
brain size = 850 cm^3

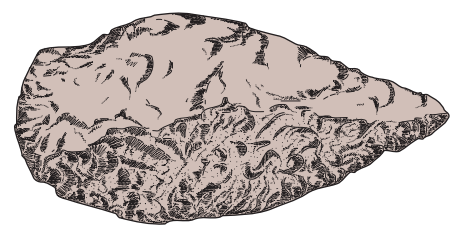




Skull from China (Peking Man)
lived 0.8 mya
brain size = 1000 cm^3

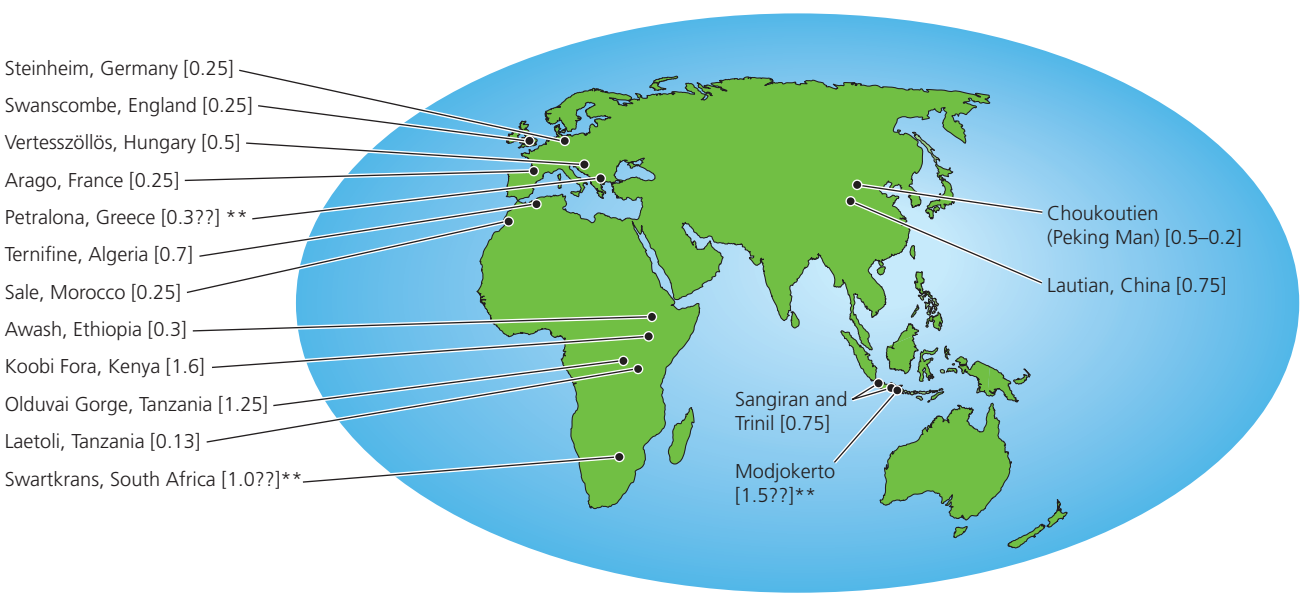


Tear-drop shaped hand axes, beautifully fashioned – the hallmark of *H. erectus.*



Diet likely to have been more meat based – a hunting culture, hence the need to travel more widely? *H. erectus* was the first hominid to have become dispersed all over Europe and Asia (the first 'out of Africa' event in human evolution).

Major sites where *H. erectus* fossil remains have been found
and dated, except ** (figures = mya)



The origin of modern humans

Recent evidence has confirmed that modern humans arose in Africa, probably about 200 000 ya. A gap in the fossil record of the earliest modern humans has recently been filled by a fossil discovery at Herto, a site north-east of Addis Ababa, in Ethiopia. 'Herto man' has been dated to between 160 000 and 154 000 ya. Before this new discovery, biochemical evidence obtained from mitochondrial DNA (page 499) indicated that all of today's human populations are related to an African human ancestor. That ancestor lived between 290 000 and 140 000 ya, and these modern human populations were first leaving Africa from 180 000 to 90 000 ya. Additional evidence about our origins comes from analysis of the Y chromosome, exclusive to male humans. This evidence also suggests a clear line of descent for modern humans from Africa-inhabiting ancestors at least 150 000 ya (Figure 16.31).

Neanderthals before *Homo sapiens*

Homo sapiens populations were preceded by at least one earlier form of human, known as *Homo neanderthalensis*. The Neanderthals originated about 250 000 ya, and became fully established 125 000 ya, but disappeared quite abruptly, between 45 000 and 32 000 ya. Their fossils are found in western and central Europe and part of the middle east, and they were probably confined to this area.

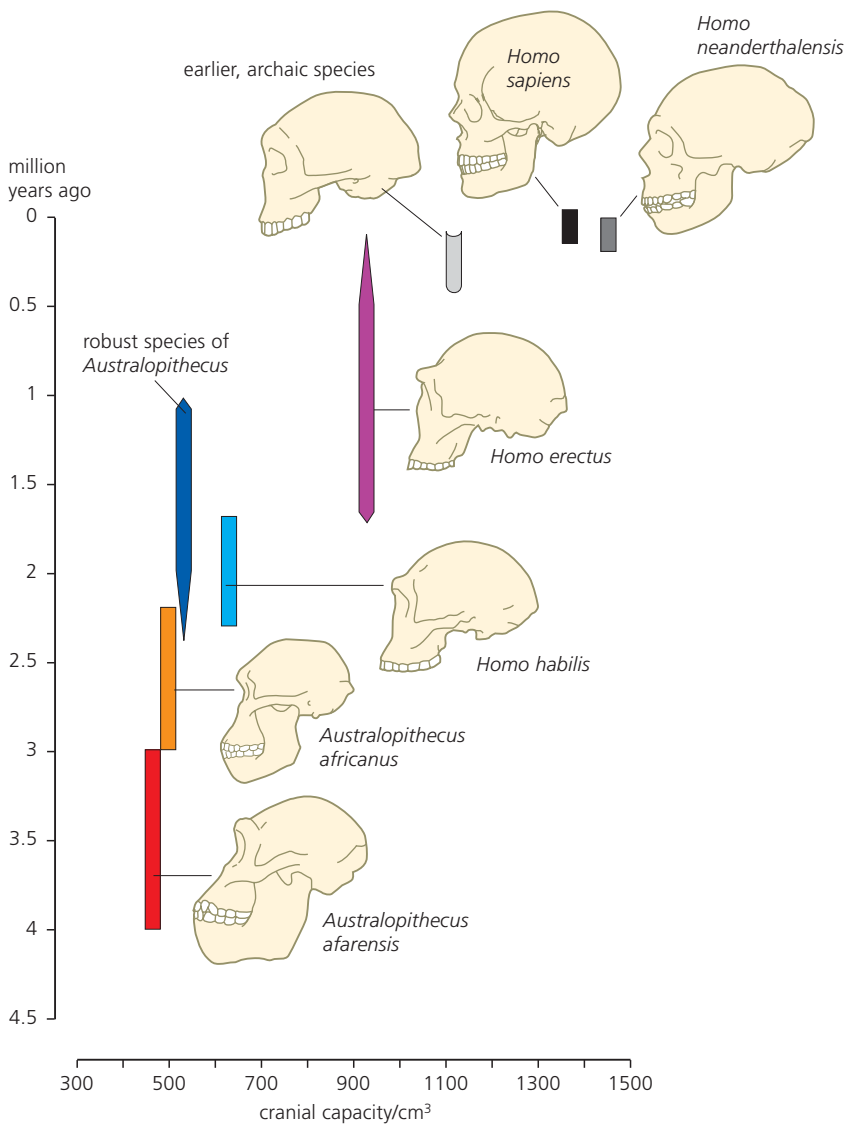


Figure 16.31 The major evolutionary stages in hominid evolution

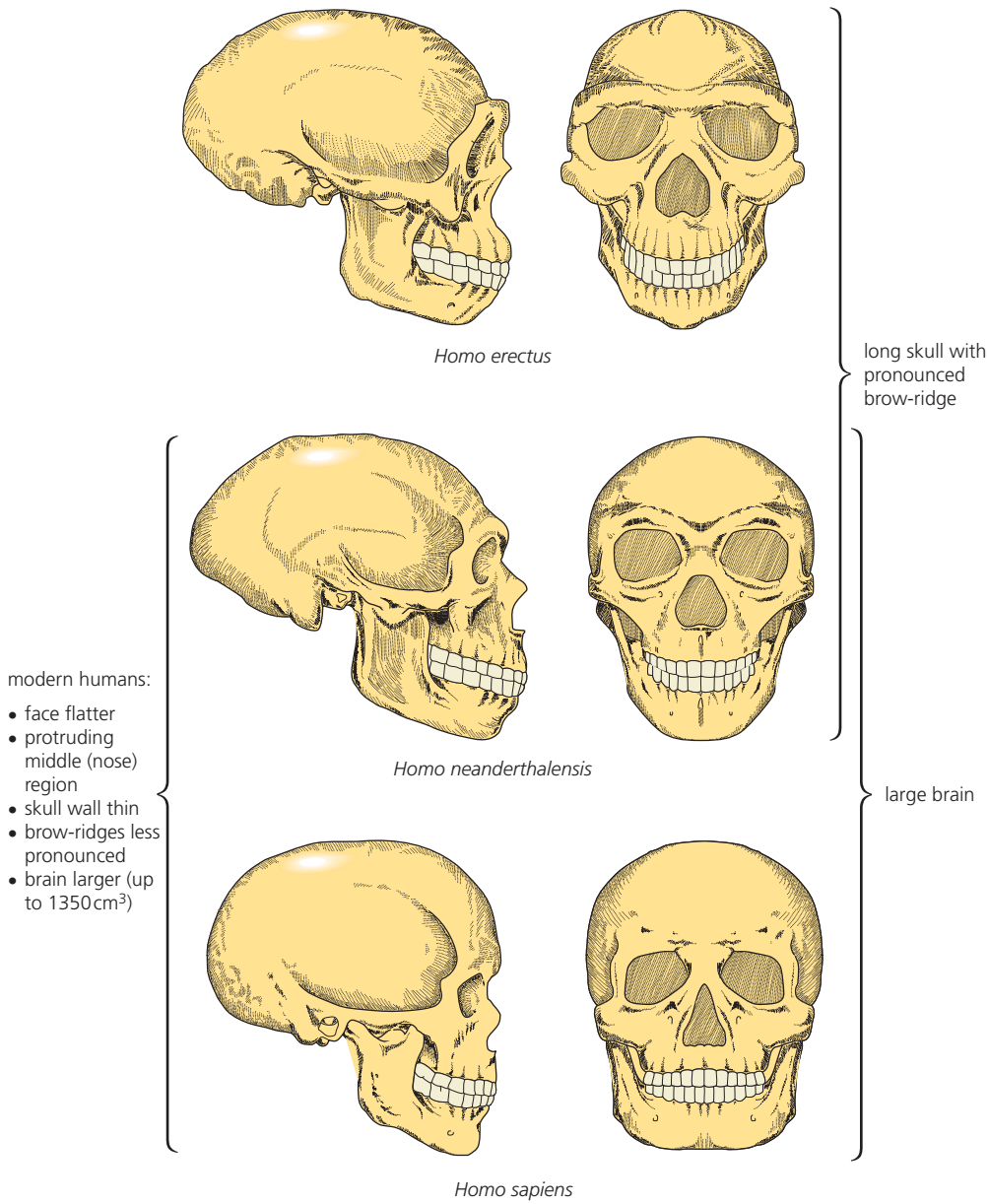


Figure 16.32 The major evolutionary stages in hominid evolution

The brain size of neanderthals was larger than that of modern humans (Figure 16.32). This may reflect the requirements of controlling the large musculature, because they were heavier and more muscular than *H. sapiens*. The latter are more slightly built, but taller and longer-limbed. As members of this modern population migrated out of Africa they lived alongside other *Homo* populations, but ultimately completely replaced them. It is clear that at various stages in hominid evolution, **different species of *Homo* have coexisted together at times**.

Human population growth at the rate the world has experienced in recent times was not an issue, initially. Indeed, until the Neolithic revolution (page 492), human communities, although widely dispersed, probably survived only with the greatest difficulty. Humans were then hunter–gatherers but, compared with many of the competing wild animals, not especially strong or fast. Scavenging would have been a major source of nutrients, at least initially. Only with the development of agriculture and other advances in technology (e.g. brewing, cheese making) did humans move into circumstances in which population sizes grew significantly, and they could start to dominate their environment and become secure.

Diet and brain size

The issue of the actual diet at each stage may have been a critical factor. This is because brains are metabolically expensive. Our brains make up 2% of our body mass but respire about 20% of our energy budget. The human brain is about three times the size of that of an equivalently sized ape.

We may conclude that the expansion in the brains that we have noted in the succeeding species of *Homo* will have demanded enhanced energy supplies. This means that human evolution must have been increasingly dependent on a reliable supply of protein and fat. However, it is not dependent on advanced hunting skills, at least not from the outset. Hominids will have discovered that the long bones of herbivorous mammals (discarded by the large carnivorous mammal hunters around them) were a rich source of bone marrow. Bone marrow is rich in protein and fat and could be accessed by nothing more sophisticated than a heavy blow from a rock onto the bone shaft, held against a hard place.

The incompleteness of the fossil record

Anthropologists disagree about the origin of modern humans from time to time. They use evidence from fossil remains, from artefacts like stone tools that can be associated with particular hominids, and the record in animal bones that surrounded their habitations and which indicate diet. Fresh evidence of these types is frequently discovered, and existing data are sometimes re-interpreted. Re-interpretation occurs in the light of new biochemical evidence or the development of new analytical techniques. For example, until quite recently, another theory about the origin of modern humans vied with the current 'out of Africa' theory.

The alternative was a **multiregional model**, in which *H. sapiens* emerged wherever populations of *H. erectus* had become established, in Africa, Europe and Asia. This made *H. neanderthalensis* only one example of an archaic hominid form, intermediate between *H. erectus* and modern humans. According to this model, there was ongoing genetic exchange between populations of various archaic forms until *H. sapiens* emerged and replaced all others. Currently, the body of evidence is increasingly against this theory.

Controversy will continue because of the inevitable incompleteness of the fossil record. **Fossilisation is an extremely rare, chance event.** This is because predators, scavengers and bacterial action normally break down dead plant and animal structures long before they can be fossilised. Of the relatively few fossils formed, most remain buried, or if they do become exposed, are often overlooked or may be accidentally destroyed before discovery. Nevertheless, numerous fossils have been found, and as more hominid fossils are discovered, so our knowledge may change and our understanding of our past be advanced. This is yet one more branch of science where the frontier of knowledge is entirely open. You can follow the debate from now on. Perhaps you may contribute to it, too.

Genetic and cultural evolution

Genetic evolution refers to the changes in allele frequencies that result in changes in individuals and therefore in populations, brought about by natural selection. In outline, these are due to:

- Genetic variations, which arise via mutations, random assortment of paternal and maternal chromosomes in meiosis, recombination of segments of maternal and paternal homologous chromosomes during crossing over that occurs in meiosis in gamete formation, and the random fusion of male and female gametes in sexual reproduction.
- When genetic variation has arisen in organisms, it is expressed in their phenotypes. Some phenotypes are better able to survive and reproduce in a particular environment, and natural selection operates to determine the survivors and the genes that are perpetuated in a population. In time, this process may lead to new varieties and new species.

By **cultural evolution** we refer to the development of the customs, civilisation and achievements of people. The development and transmission of human culture has a biological basis.

Key to this was the extension of the **period of parental care**, delayed onset of puberty, and the resulting long period of childhood when the next generation of a population are trained and schooled as they develop essential survival skills – all features of the evolution of the genus *Homo*.

The **development of language** is the most important human characteristic central to the evolution of culture. Endocasts give a slight impression of the areas of the brain that developed and were enlarged (the chief neural machinery for speech in most modern humans is found in the left hemisphere). Also critical is the position of the vocal folds in the neck. On both counts it seems likely that only Neanderthals and *H. sapiens* achieved the structures necessary for elaborate vocal communication. In particular, the high palate and high larynx found in *H. sapiens* allowed a greater range of resonance for complex word sounds.

Once established, verbal communication allowed advantageous developments (for example, in the form of new ideas) to be passed on rapidly. The potential speed of development of this form of cultural evolution contrasts markedly with change brought about by slow inherited accumulation of advantages by genetic evolution. Today's latest cultural-sharing breakthroughs – the internet and the human genome project – are cases in point.

The **developments in tool technology** were also dependent on the development of a large brain. Compared to the achievements of the Habilines in this, from about 35 000 years ago, modern humans made spectacular advances. Bone and antler were added to the list of raw materials, and advances in the skills of fashioning stone flakes and blades into finely worked scrapers, chisels, drills, arrowheads and barbs were spectacular. Tool-kits comprised items for engraving and sculpture. Functional implements like spears became decorated with life-like animal carvings.

Figure 16.33 Head of bison, Niaux cave, France

The latter point relates to human use of the brain, powers of detailed observation, and manual dexterity, all of which underpin cultural development. *Homo sapiens* as observers and artists achieved incredible feats at the earliest phase of their development.

We have a remarkable record of the artistic skills of our first human ancestors in the cave paintings from this period that have been discovered. The drawings, produced by human communities from 25 000 to 10 000 years ago, show contemporary animals in scientific detail (Figure 16.33). The pictures demonstrate perspective representation. At one time, art historians regarded perspective as a technique invented in Renaissance Florence.

Relative importance of genetic and cultural evolution

Genetic evolution has given rise to the diversity of living things, including human beings. However, this is a process that has taken thousands of millions of years. The special features that humans have developed, mostly unique to them, have been the basis of **cultural evolution**. For example, with the development of agriculture and other technologies, humans have changed their immediate environment with the creation of settlements and then gone on to evolve communal living. Enlarged populations have been both necessary to the new way of life, and sustained by it. Rules and laws have succeeded basic customs, and individuals have acquired rights and responsibilities. Consequently, the conditions for genetic evolution have been progressively sidelined as the processes of cultural evolution have taken over.

A feature of our cultural evolution has been its speed. For example, can you find out the approximate time that elapsed between the appearance of the first hominids and the Neolithic revolution that occurred in the fertile crescent; between the Neolithic revolution and the industrial revolution; and finally, between the industrial revolution and the silicon revolution of today's world? Is it not a shrinking time span?

Today, humans in their more altruistic and caring modes, seek to remove the illnesses and inadequacies that humankind experience (and that individuals often perished from before being able to reproduce and pass on their genes) by the development and deployment of technological solutions – perhaps the highest achievement of their cultural evolution?

What do you think?

The Hardy–Weinberg principle

D4.1–4.3

We have noted that in any population, the total of the alleles of the genes located in the reproductive cells of the individuals make up a gene pool. A sample of the alleles of the gene pool will contribute to form the **genomes** (gene sets of individuals) of the next generation, and so on, from generation to generation.

When the gene pool of a population remains more or less unchanged, then we know that population is not evolving. However, if the gene pool of a population is changing (i.e. the proportions of particular alleles are altered – we say 'disturbed' in some way), then evolution may be going on.

How can we detect change or constancy in gene pools?

The answer is, by a mathematical formula called the Hardy–Weinberg formula (Figure 16.34). Independently, this principle was discovered by two people in the process of explaining why dominant characteristics don't take over in populations, driving out the recessive form of that characteristic. For example, at the time, people thought (wrongly) that human eye colour was controlled by a single gene, and that an allele for blue eyes was dominant to the allele for brown eyes. They wanted to answer the question, Why doesn't the population become blue-eyed?

The general formula to represent the frequency of dominant and recessive alleles is:

$p + q = 1$

where p = frequency of the dominant alleles, and q = frequency of the recessive alleles.

This equation was developed by Hardy and Weinberg to describe stable gene pools:

p^2	+	$2pq$	+	q^2	= 1
frequency of dominant homozygous individuals		frequency of heterozygous individuals		frequency of homozygous recessive individuals	TOTAL

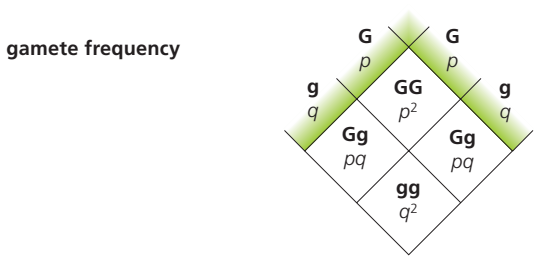
Figure 16.34 Deriving the Hardy–Weinberg formula

Let the frequency of the dominant allele (**G**) be p, and the frequency of the recessive allele (**g**) be q.

The frequency of alleles must add up to I, so $p + q$ = I.

This means in a cross, a proportion (p) of the gametes carry the **G** allele, and a proportion (q) of the gametes carry the **g** allele.

The offspring of each generation are given by the Punnett grid.



Hardy–Weinberg formula

If the frequency of one allele (**G**) is p, and the frequency of the other allele (**g**) is q then the frequencies of the three possible genotypes:

GG, **Gg** and **gg**

are respectively: p^2, $2pq$ and q^2.

So the progeny are respectively:
p^2 = frequency of **GG** homozygote
$2pq$ = frequency of **Gg** heterozygote
q^2 = frequency of **gg** homozygote

The main problem in estimating gene frequencies is that it is not possible to distinguish between homozygous dominants and heterozygotes, based on their appearance or phenotype, as has been noted previously (pages 106 and 366).

However, using the above equation, it is possible to calculate allele frequency **from the number of homozygous recessive individuals in the population**. This is q^2.

By taking the square root of this we can find q. The result tells us the frequency of the recessive allele, and this can then be substituted into the initial equation $p + q = 1$ to find the frequency of the dominant allele. Values for p and q can then be used with the Hardy–Weinberg formula to calculate the proportions of homozygous dominant and heterozygous individuals in the population.

Using the Hardy–Weinberg formula

The absence of the skin pigment, melanin, is a condition called albinism (Figure 4.20, page 110), a genetically controlled characteristic. An albino has the genotype **pp** (homozygous recessive), whereas people with normal pigmentation are homozygous (**PP**) or heterozygous (**Pp**).

In a large population, only one person in 10 000 is albino. From the equation above, the frequency of homozygous recessives (**pp**) = q^2.

Thus:

$q^2 = 0.0001$, so $q = \sqrt{0.0001} = 0.01$.

So, the frequency of the non-melanin-secreting allele (**p**) in the population = 0.01 (or 1%). Substituting into the initial equation $p + q = 1$

$p + 0.01 = 1$, therefore $p = 0.99$

So, the frequency of the melanin-secreting allele (**P**) in the population = 0.99 (or 99%).

The Hardy–Weinberg formula has allowed us to find the frequencies of alleles **P** and **p** in a population.

Incidentally, it has also shown that the frequency of carriers of an allele for albinism in the population (**Pp**) is quite high (about 1 in 50 of the population) despite the fact that albinos make up only 1 in 10 000. In other words, very many more people carry around an allele for albinism than those who know they may do so.

14 In a small isolated population, the following incidence of genotypes was found:

AA 109, **Aa** 252, **aa** 39.

Calculate the frequency of the alleles **A** and **a** in this local population.

The Hardy–Weinberg principle and disturbing factors

The Hardy–Weinberg principle predicts that the gene pool in a population does not change in succeeding generations. That is, genes and genotype frequencies normally remain constant in a breeding population. This condition, known as a **genetic equilibrium**, will occur, provided that:

- the breeding population under investigation is a large one;
- there is random mating, with individuals of any genotype all equally likely to mate with individuals of any other genotype (e.g. no one genotype is being selectively predated);
- no mutations are occurring;
- there is no immigration or emigration occurring.

Of course, gene pools do sometimes change. Changes in the gene pool may be the forerunner of evolution.

Phylogeny and systematics

D5.1–5.10

Phylogeny is the study of evolutionary relationships, and **systematics** is the study of the identification and classification of organisms.

The process of classification, known as **taxonomy**, involves:

- giving every organism an agreed name – the **binomial system** of naming (page 164);
- imposing a scheme on the diversity of living things by using the **hierarchical scheme of classification** (page 165).

Classification is essential to biology because there are too many different living things to sort out and compare unless they are organised into manageable categories. The scheme of classification has to be entirely flexible, allowing newly discovered living organisms to be added into the scheme where they fit best. It should also be able to accommodate fossil organisms as they are discovered, since we believe living and extinct species are related.

The **quickest** way to classify living things is on their **immediate and obvious similarities and differences**. For example, we might classify together animals that fly, simply because the essential organs – wings – are so easily seen. Such a group would include almost all birds and many insects

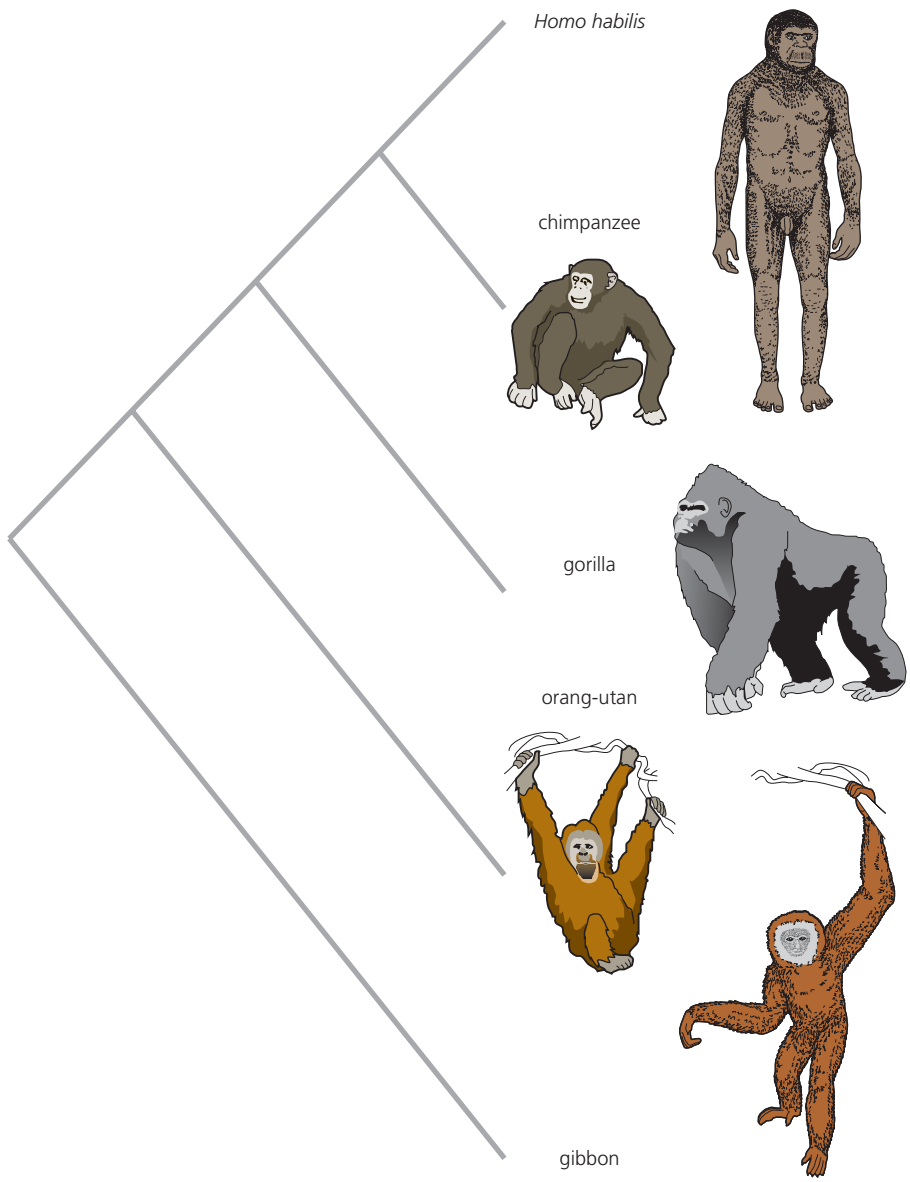


Figure 16.35 Phylogenetic tree of modern humans and their closest relatives

(as well as the bats, and certain fossil dinosaurs). However, resemblances between the wings of a bird and the wings of an insect are superficial. While both are aerofoils (structures that generate lift when moved though the air), they are built from different tissues and have different origins in development of the body. We say that the wings of birds and insects are **analogous structures**. Analogous structures resemble each other in function but differ in their fundamental structure (Table 16.6). The similarities in structure are due not to being derived from a common ancestor, but to being derived from different ancestors by convergent evolution towards a similar purpose. A classification based on analogous structures is an **artificial classification**.

The alternative is a classification that may be based on **similarities and differences due to close relationships between organisms because they share a common ancestor**. So, for example, the limbs of all vertebrates suggest they are modifications of a common plan, the pentadactyl limb (Figure 16.12, page 469), and indicate that vertebrates can be classified together. Structures built to a common plan, but adapted for different purposes, are **homologous structures** (Table 16.6).

Table 16.6 Analogous and homologous structures

Analogous structures	Homologous structures
■ resemble each other in function	■ are similar in position and development, but not necessarily in function
■ differ in their fundamental structure	■ are similar in fundamental structure
■ illustrate only superficial resemblances	■ are similar because of common ancestry

A classification based on homologous structures is believed to reflect evolutionary relationships, and is called a natural or **phylogenetic classification**. A diagram illustrating evolutionary relationships between species is called a **phylogenetic tree**. The word 'tree' is employed because the branching pattern shows when different species 'split off' from others (Figure 16.35).

The **value of classifying organisms** is that it aids identification to organise profoundly similar organisms together. A scheme of classification based on evolutionary relationships may have a predictive value too. If most members of a group share a particular characteristic, then other members are likely to show that characteristic as well. With an effective classification system in use, it is easier to organise our ideas about organisms and to make generalisations.

Biochemical evidence and phylogeny

All living things have DNA as their genetic material, with a genetic code that is virtually universal. The processes of 'reading' the code and of protein synthesis, using RNA and ribosomes, are very similar in prokaryotes and eukaryotes. Processes such as respiration involve the same types of steps and similar or identical intermediates and biochemical reactions, similarly catalysed. ATP is the universal energy currency. Among the autotrophic organisms, the biochemistry of photosynthesis is virtually identical, as well.

This biochemical commonality suggests a common origin for life, as the biochemical differences between the living things of today are limited. Some of the earliest events in the evolution of life must have been biochemical, and the results have been inherited widely. However, large molecules like nucleic acids and the proteins they may code for are subject to change with time, and this change may be an aid to the study of evolution and relatedness. It is possible to measure the relatedness of different groups of organisms by the amount of difference between molecules such as DNA and proteins, and enzyme systems – the amount of difference is a function of the time since the particular organisms under investigation shared a common ancestor.

Variations in protein molecules indicating phylogeny

Haemoglobin, the β chain of which is built of 146 amino acid residues, shows variations in the sequence of those amino acids between the different species in which it occurs. Haemoglobin structure is determined by inherited genes, so the more closely related species are, the more likely their amino acid sequence is to match (Table 16.7). Variations are thought to arise by mutations of an ancestral gene for haemoglobin. If so, the longer it is since species diverged from a common ancestor, the more likely it is that differences will have arisen.

Table 16.7 Number of amino acid differences in β chain of haemoglobin compared to human haemoglobin

Species	Number of differences	Species	Number of differences
human	0	kangaroo	38
gorilla	1	chicken	45
gibbon	2	frog	67
rhesus monkey	8	lamprey	125
mouse	27	sea slug (mollusc)	127

Similar studies have been made of the differences in the polypeptide chains of other protein molecules, including ones common to all eukaryotes and prokaryotes. One such is the universally occurring electron-transport carrier, cytochrome *c*.

Biochemical variation used as an evolutionary clock

Biochemical changes like those disclosed above may occur at a constant rate, and if so, could be used as a molecular clock. If the rate of change can be reliably estimated, then they do record the time that has passed between the separations of evolutionary lines.

But do these changes occur in response to natural selection or as a result of random change – the accumulation of mutational change?

It is most likely that evolutionary changes in organisms do *not* occur at a constant rate. Think, for example, of changes in a particular group of animals in response to environmental change. Surely among these changes, some will have arisen quickly, some more slowly, while other features may have hardly changed at all.

In the case of the haemoglobin of vertebrate animals (which evolved from a form present in an early common ancestor), studies indicate a different situation. Analyses of the haemoglobins of current vertebrates suggest that a natural drift has occurred between them, whether or not the animal species concerned have evolved little or significantly from a common ancestor. The haemoglobin 'clock' does appear to 'tick' regularly.

Immunological studies are another means of detecting differences in specific proteins of species, and therefore (indirectly) their relatedness. **Serum** (plural, **sera**) is the liquid produced from blood samples when blood cells and fibrinogen have been removed. Protein molecules present in the serum act as antigens if the serum is injected into animals with an immune system that lacks these proteins.

Typically, a rabbit is used when investigating relatedness to humans. The injected serum causes the production of antibodies against the injected proteins. Then, serum produced from the treated rabbit's blood (now containing antibodies against human proteins) can be tested against sera from a range of animals. The more closely related the animal is to humans, the greater the precipitation observed (Figure 16.36).

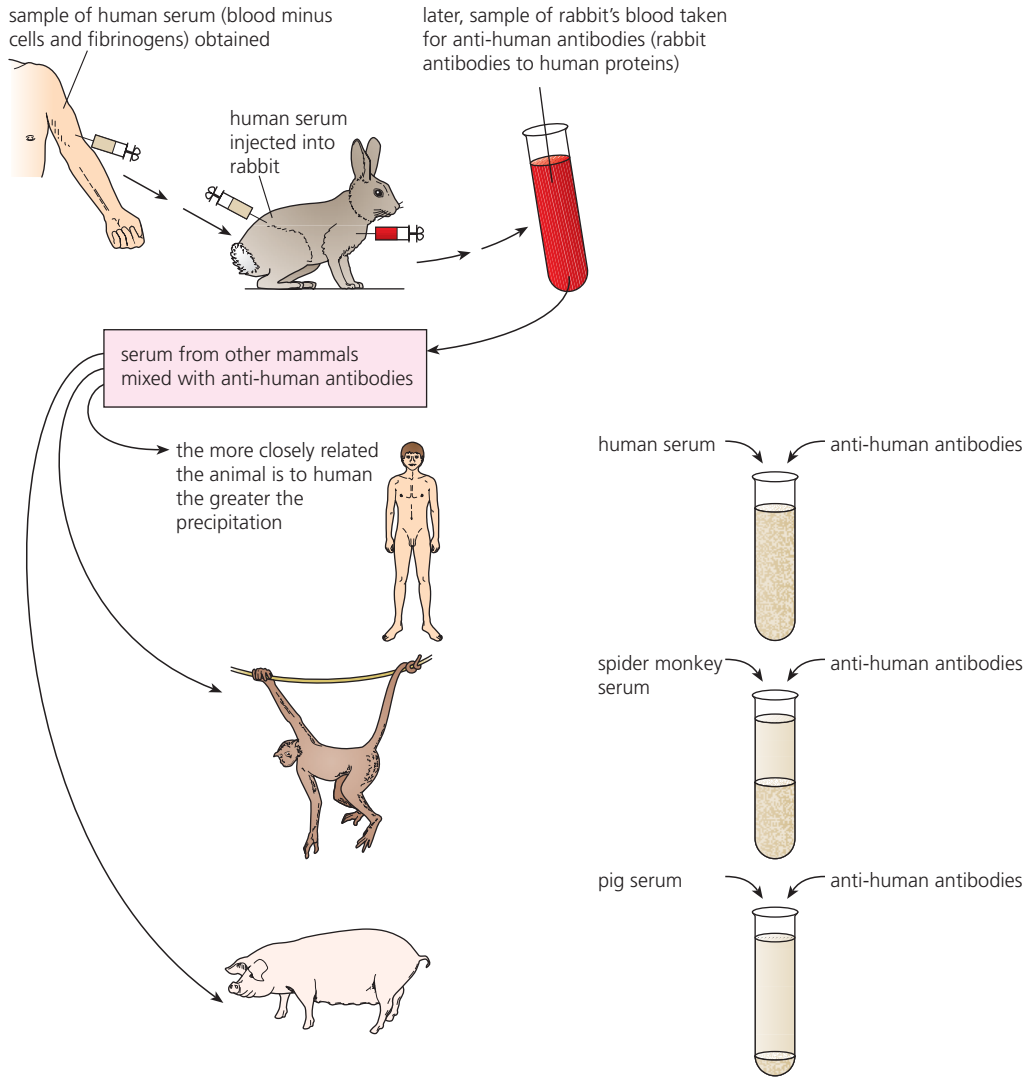
The precipitation produced by reaction with human serum is taken as 100%. For each species in Table 16.8, the greater the precipitation, the more recently the species shared a common ancestor with humans. This technique, called comparative serology, has been used by taxonomists to establish phylogenetic links in a number of cases, in both mammals and non-vertebrates.

Of course, we do not know of the common ancestor of these animals, and the blood of that ancestor is not available to test anyway. But if the 584 amino acids that make up blood albumin change at a constant rate, then the percentage immunological 'distance' between humans and any of these species will be the distance 'back' from humans to the common ancestor plus the distance 'forward' again to that species. Hence the differences between a listed species and humans can be halved to gauge the differences between a modern form and the common ancestor.

Since the radiation of the Primates is known from geological/fossil evidence, the forward rate of change since the lemur gives the rate of the molecular clock – namely 35% in 60 million years, or 0.6% every million years. This calculation can now be applied to all the data (Table 16.8, right-hand column).

Figure 16.36 The immune reaction and evolutionary relationships

Immunological studies are a means of detecting differences in specific proteins of species, and therefore (indirectly) their **relatedness**.



Species	Precipitation / %	Difference from human	Difference from common ancestor (half difference from human) / %	Postulated time since common ancestor/millions of years
human	100	—	—	—
chimpanzee	95	5	2.5	4
gorilla	95	5	2.5	4
orang-utan	85	15	7.5	13
gibbon	82	18	9	15
baboon	73	27	13.5	23
spider monkey	60	40	20	34
lemur	35	65	32.5	55
dog	25	75	37.5	64
kangaroo	8	92	46	79

Table 16.8 Relatedness investigated via the immune reaction

DNA as a molecular clock

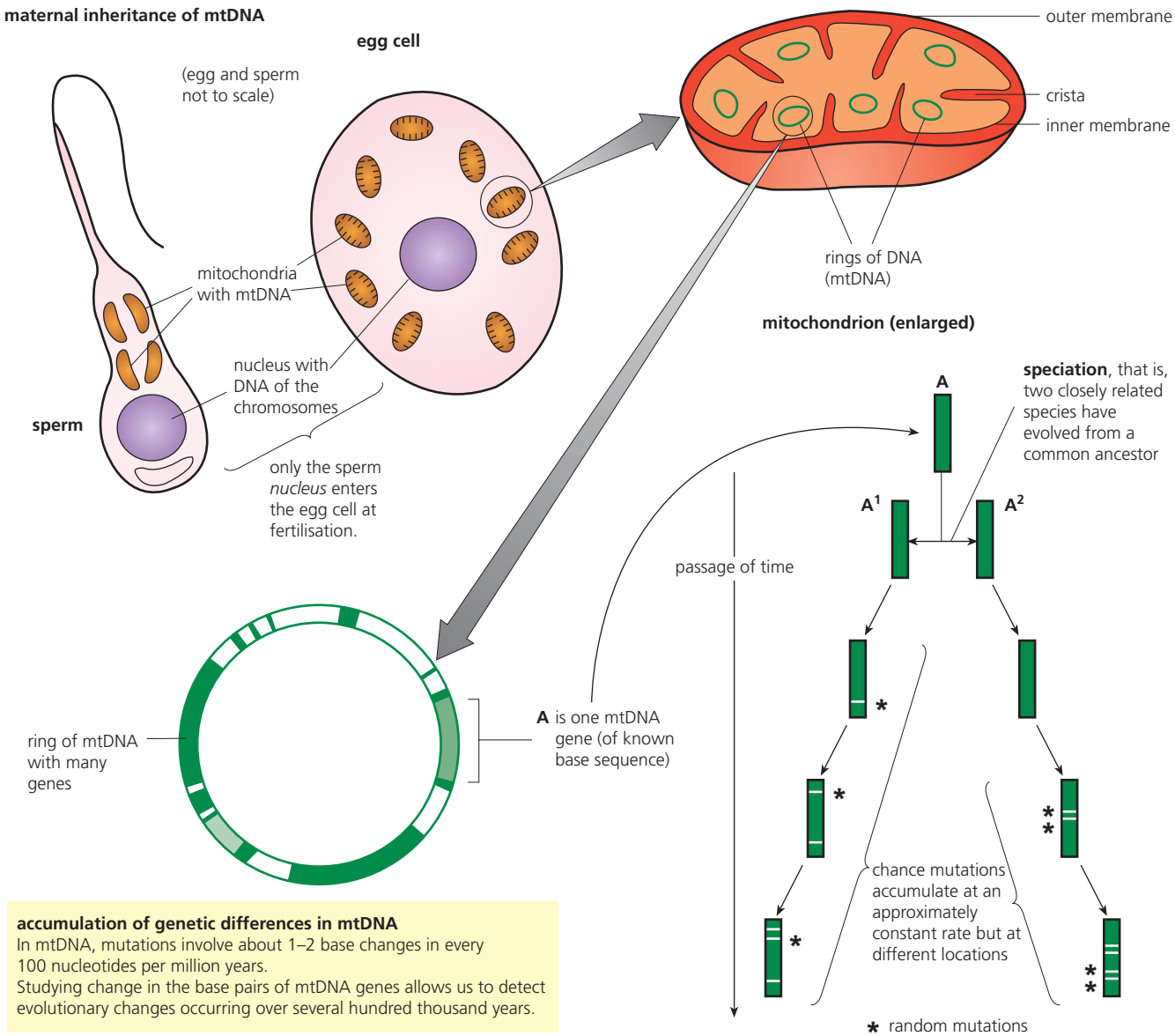
DNA also has potential as a molecular clock. DNA in eukaryotic cells occurs in chromosomes in the nucleus (99%) and in the mitochondria. **Mitochondrial DNA (mtDNA)** is a circular molecule, very short in comparison with nuclear DNA. Cells contain any number of mitochondria, typically between 100 and 1000.

Mitochondrial DNA has approximately 16 500 base pairs. Mutations occur at a very slow, steady rate in all DNA, but chromosomal DNA has with it enzymes that may repair the changes in some cases. **These enzymes are absent from mtDNA.**

Thus, mtDNA changes 5–10 times faster than chromosomal DNA, involving 1–2 base changes in every 100 nucleotides per million years. Consequently, the length of time since organisms belonging to different but related species have diverged can be estimated by extracting and comparing samples of their mtDNA (Figure 16.37).

Furthermore, at fertilisation, the sperm contributes a nucleus only (i.e. no cytoplasm). All the mitochondria of the zygote come from the egg cell. There is no mixing of mtDNA genes at fertilisation, and so the evidence about relationships from studying differences between samples of mtDNA is easier to interpret in the search for early evidence of evolution.

Figure 16.37 The use of mitochondrial DNA in measuring evolutionary divergence



accumulation of genetic differences in mtDNA
In mtDNA, mutations involve about 1–2 base changes in every 100 nucleotides per million years.
Studying change in the base pairs of mtDNA genes allows us to detect evolutionary changes occurring over several hundred thousand years.

Cladistics

Taxonomists seek to use evolutionary relationships in the schemes of classification of living things they propose. That is, phylogenetic trees are sought, rather than any form of artificial (superficial) classification. Phylogenetic trees (Figure 16.35, page 495) have two important features:

- **branch points** in the tree – representing the **time** that a divide between two taxa occurred;
- the **degree of divergence** between branches – representing the **differences** that have developed between the two taxa since they diverged.

Taxonomists ponder the question, *Which of these two features is the more significant in a phylogenetic taxonomy, and so receives the greater emphasis in devising a scheme?*

In **phenetics**, the answer to this question is that the measurable similarities and differences of anatomy should be used to arrange species into dichotomously branching trees. The product is a **dendrogram**.

In **cladistics**, classification is based on when the branches arise in the taxonomic tree. Cladistics is a system of analysis of relatedness. The product is a **cladogram**.

A clade is a branch of a phylogenetic tree containing all the organisms descended from a particular common ancestor.

The differences between phenetics and cladistics are illustrated by lizards, crocodiles and birds, for example. Lizards and crocodiles resemble each other more than either resemble the birds, but crocodiles and birds share a common ancestor (Figure 16.38).

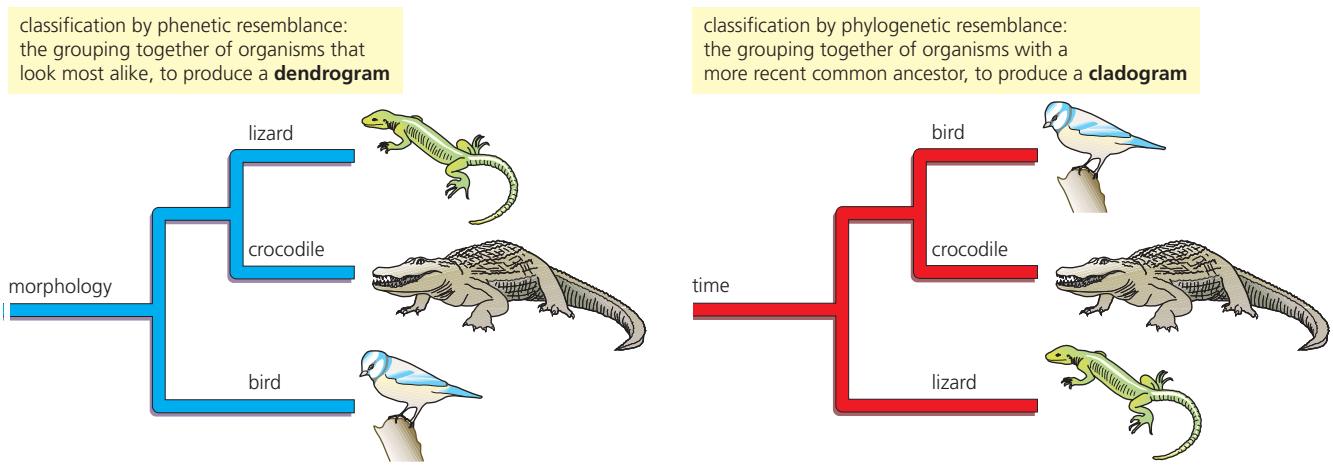


Figure 16.38 Phenetics and cladistics – two ways of classifying

The construction of cladograms

The cladistic school of taxonomists aims to express the relationships among species regardless of how similar or different they are, by seeking to trace their descent from a common ancestor. Each species to be classified has a mixture of characteristics. Some of these characteristics will be called primitive, because they existed in a common ancestor. The sharing of primitive characteristics tells us nothing about subsequent evolutionary divergence. Other characteristics will have appeared subsequently, and some may define a branching point between the individual species to be placed in a phylogenetic tree. Characteristics may be morphological or biochemical, according to the species concerned or the available evidence.

For example, we can construct a cladogram based on morphological features of many of the primates represented in Figure 16.19 (page 478) – this illustration is very similar to a cladogram. However, some of the morphological differences are minor and relatively trivial. For example, it makes more sense to differentiate the old world and new world monkeys simply on where they occur.

Alternatively, the use of biochemical evidence to investigate relatedness is illustrated by the use of the immune reaction. The data in Table 16.8 may be used to create a cladogram.

15 **Construct** a cladogram based on the biochemical data in Table 16.8.

Analysing cladograms in terms of phylogenetic relationships

The cladogram required in answer to SAQ 15 is based on a single characteristic – the immune response proteins of the organisms studied. It illustrates a point about cladograms, namely that cladistic analysis can be based on as much or as little information as the researcher selects.

Currently, cladograms are typically based on a wide variety of information. Differences in a sufficiently large number of characteristics are assumed to lessen the impact of similarities due to convergent evolution. Remember, in convergent evolution unrelated organisms show superficial similarities known as analogous structures.

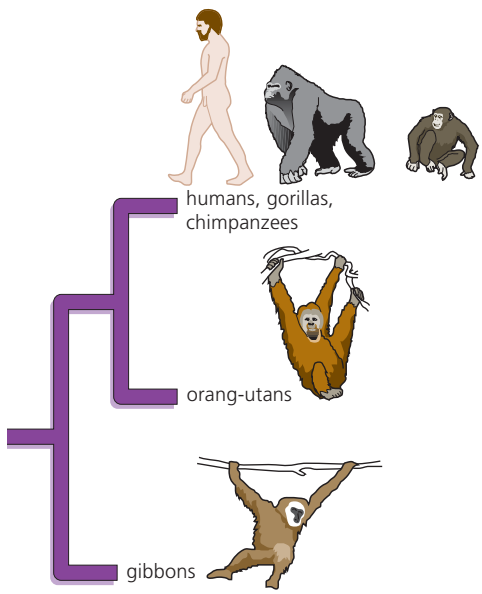


Figure 16.39
Relatedness of gibbons, orang-utans, gorillas, chimpanzees and humans

We can now attempt to analyse a cladogram regarding the issue of phylogenetic relationships among primates. Which of the primates are our closest relatives, the gibbons, orang-utans, gorillas or chimpanzees? The cladogram in Figure 16.39 indicates that the gibbons and orang-utans are not as closely related to *Homo* as are the chimpanzees and gorillas. The characteristics used here were:

- possession of a frontal sinus (a cavity in the skull just above the eyes) – a homology that gibbons and orang-utans lack;
- degrees of relatedness of certain blood-clotting proteins found in the blood plasma.

So, if chimpanzees and gorillas are our closest relatives, which is closest? Figure 16.40 shows three cladograms representing possibilities. Here a list of seven characteristics is analysed to resolve this issue.

Study Figure 16.40.

Does the evidence presented there resolve the question of our true phylogenetic relationship with the gorilla and the chimpanzee? Which cladogram is supported by most evidence?

Today, differences based on molecular data are likely to be the most reliable.

DNA sequencing has become relatively easy and is widely applied, and the wealth of data it generates can be handled by computer software. The outcomes increasingly shed new light on phylogenetic relationships. For example, if data on differences in DNA were added to the analysis of the cladograms in Figure 16.40, it is likely that chimpanzees would prove to be our closest relatives. Today, humans have been described as the 'third chimpanzee'.

The relationship between cladograms and classification of living things

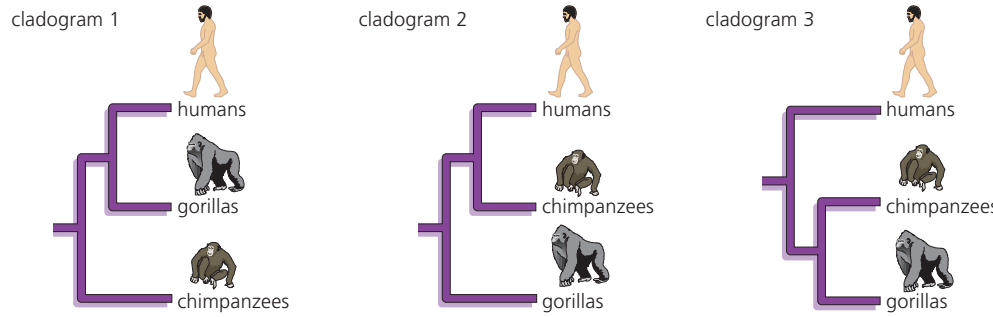
The classifications used in practice in biology may be described as a natural classification or, more grandly, a classical evolutionary taxonomy. This system seeks to use the products of *both* phenetic data (where many anatomical similarities and differences are used without differentiation between homology and analogy), together with cladistic data (based on branching events).

Where a conflict arises in these alternative sources of evidence, then a subjective judgement has to be made about which evidence should be given priority. Thus, in the case of the crocodile, lizard and bird (Figure 16.38), we classify birds in their own class because although birds and crocodiles shared a common ancestor more recently, the classification scheme we use recognises the ability to fly is an evolutionary breakthrough (one that has evolved independently several times).

While biochemical evidence of relatedness may increasingly yield more precise information on when related organisms shared a common ancestor, we are likely to rely on morphological evidence as well to help us maintain a practical 'filing system' for the range of living things.

Figure 16.40 Analysing cladograms in terms of phylogenetic relationships

gorilla or chimpanzee – which is our closest relative?



shared characteristics by which to analyse the above cladograms

	humans	gorilla	chimpanzee	other primates	*which cladogram does this characteristic support?*
bones and teeth					
limb length	arms shorter than legs	legs shorter than arms	legs shorter than arms	arms and legs of equal length	**3**
canine teeth	small	large	large	large	**1, 2 or 3**
thumbs	long	short	short	long	**3**
soft parts of the body					
head hair	long	short	short	short	**1, 2 or 3**
chromosomes					
total number	46	48	48	42 or more	**3**
structure of chromosomes 5 and 12	like other primates	different from other primates	different from other primates		**3**
biochemical evidence					
haemoglobin	'reference molecule'	one amino acid different	identical to humans	several differences	**2**
sequence of amino acids in myoglobin	generally like other primates	like chimpanzees	like gorillas		**3**

Examination questions – a selection

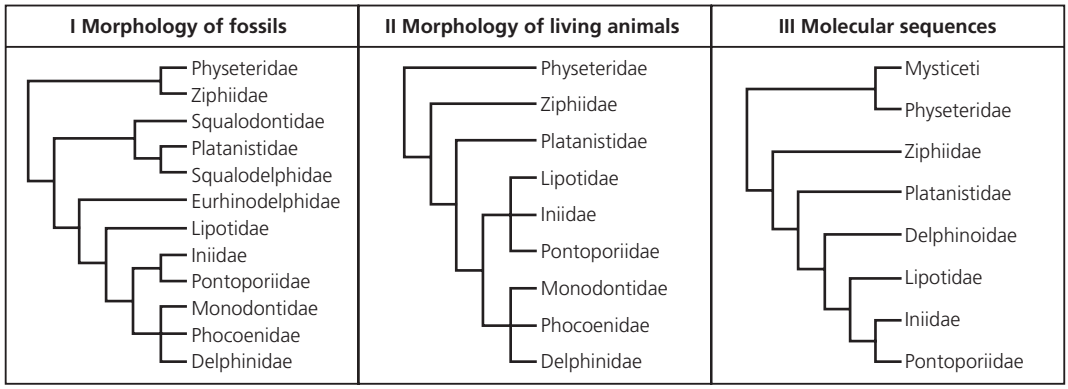
Questions 1–5 are taken from past IB Diploma biology papers.

Q1 River dolphins live in fresh-water habitats or estuaries. They have a number of features in common which distinguish them from other dolphins: long beaks, flexible necks, very good echo-location and very poor eyesight. Only four families of river dolphins have been found in rivers around the world.

River	River dolphin family
Amazon, Brazil	Iniidae
La Plata, Argentina	Pontoporiidae
Yangtze, China	Lipotidae
Indus and Ganges, India	Platanistidae

Evolutionary biologists have tried to determine how closely related these river dolphins are to one another. River dolphins are members of the group, the toothed whales. Three lines of evidence were analysed, producing three cladograms (family trees) for all the toothed whales. The evidence to construct these cladograms came from the morphology (form and structure) of fossil toothed whales (I), the morphology of living toothed whales (II) and the molecular sequences from living toothed whales (III).

a Suggest reasons why there are more families present in cladogram I, produced from the morphology of fossils, than for the other cladograms. (1)



b Using only the data from cladogram III, identify which other family of river dolphins is most closely related to Platanistidae. (1)

c State what material would be used to produce cladogram III, based on the molecular sequences of living toothed whales. (1)

The tree using the data from the morphology of living animals (II) indicates that the families are more closely related than the tree using molecular sequences (III) from the same animals.

d Explain how these dolphins can look so similar when in fact they may not be so closely related.

These cladograms show the species that share common ancestors, but do not show how long ago they diverged from one another.

e Outline further evidence that would be needed to determine when these families of toothed whales diverged.

Standard Level Paper 3, November 03, QD1

Q2 State **three** conditions thought to have been present on the pre-biotic Earth. (3)

Standard Level Paper 3, November 03, QD3

Q3 **a** State **one** radioisotope used for the dating of rocks and fossils. (1)

b Define the term *half-life*. (1)

c Explain how the approximate age of a fossil could be determined using a radioisotope. (2)

Standard Level Paper 3, November 02, QD2

Q4 **a** State the class in which human beings are placed. (1)

b Describe some of the physical features that define human beings as primates. (3)

c State **two** differences and **one** similarity between genetic and cultural evolution. (3)

Standard Level Paper 3, November 05, QD2

Q5 **a** **i** State the name given to Darwin's theory of evolution. (1)

ii Describe Darwin's theory of evolution. (4)

b Discuss the possible origins of prokaryotic cells. (3)

Standard Level Paper 3, November 04, QD3

Questions 6–10 cover other syllabus issues in this chapter.

Q6 Outline how experimental evidence has been obtained for the possible origin of a wide range of organic compounds on the pre-biotic Earth. (6)

Q7 Suggest the possible roles of RNA in the origins of self-replicating molecules before the first cells formed. (4)

Q8 **a** Define *gene pool* and *gene frequency*. (2)

b Gene frequencies in a population may be found to change over many generations. Explain what this suggests to an evolutionary biologist, and why. (4)

Q9 Outline mechanisms by which:

a allopatric speciation

b sympatric speciation

may be brought about. (6)

Q10 The apes, monkeys and humans are groups of primates. Describe **three** features that humans exhibit that suggest they are more closely related to apes than to monkeys. (6)

Option E

17 Neurobiology and behaviour

STARTING POINTS

- Specialised cells called **neurones**, such as the motor neurone and others, are assembled to form the **nervous system**.
- The mammalian nervous system consists of a brain and spinal cord (**central nervous system**, **CNS**) and nerves serving the tissues and organs of the body (**peripheral nervous system**, **PNS**).
- The nervous system links **receptors** (e.g. sense organs) to **effectors** (muscles or glands).
- An impulse or **action potential** is a temporary reversal of the electrical potential difference that is maintained across the **membrane of the nerve fibres**. Conduction of an action potential is extremely fast.
- Action potentials are transmitted between neurones across tiny gaps at **synapses**. Transmission here is **chemical**, involving diffusion of a specific **transmitter substance**.
- This chapter extends study of aspects the nervous system begun in Chapter 7 (pages 178–234).

Living things are able to detect changes and respond appropriately. This ability, known as **sensitivity**, is a characteristic of all living things. It is essential for their survival. Sensitivity is just as much a feature of single-celled organisms as it is of mammals, as we shall shortly see. Animal responses are typically quick movements, and their responses to change may involve adjustment of their behaviour. In multicellular animals such as the mammals, the detection of and response to external and internal changes are brought about by the **nervous system** (the reactions of which are rapid), and also the endocrine (hormone) system (the responses of which are generally very much slower).

In this chapter, the working of the **nervous system** and the **perception of stimuli** are examined, followed by discussion of **animal behaviour**. The functioning of **neurotransmitters at the synapse** is then reviewed; this provides the basis for understanding of the **effects of psychoactive drugs**.

In the Additional Higher Level extension, the structure and functioning of the **human brain** is explored, followed by discussion of **further issues in animal behaviour**.

Stimulus, response and reflex

E1.1–1.4

Sensitivity is the ability to detect change and to respond to it. Changes that are detected and lead to a response are called stimuli (singular, stimulus).

A stimulus is a change in the environment (internal or external) that is detected by a receptor, and elicits a response.

The **response** to a stimulus might be the movement of the whole animal (perhaps in pursuit of food or away from danger), or movement of part of the animal (such as a limb jerking away from a hot object). A response to an internal stimulus might be the secretion of enzymes in response to the presence of food in the gut.

A response is the activity of a cell or organism in terms of movement, hormone secretion or enzyme production, for example, as a result of a stimulus.

Some responses are immediate, quick, and short-lived. Such responses are automatic and involuntary (unconscious). They are referred to as reflex actions or **reflexes.**

A reflex is a rapid, unconscious response.

What structures bring about reflex actions and what are their roles?

The response of the body to a stimulus involves a **receptor organ**, **neurones** of the nervous system, and an **effector organ** arranged in a functional unit called a **reflex arc**. In the reflex arc, an impulse generated in a receptor is transmitted via neurones to the effector which brings about a response, in the sequence:

stimulus
↓
receptor → sensory neurone → CNS → motor neurone → effector
↓
response

More than one type of neurone is involved in a reflex arc. We will examine the structure and function of neurones first.

Neurones of the reflex arc

The nervous system is built up from specialised cells called **neurones**. Three types of neurone are involved in a reflex arc (Figure 17.1). Each neurone has a **cell body** and a number of extensions, the **nerve fibres**. The three types are:

- **motor neurones**
  These have many fine **dendrites** (meaning 'little trees') which bring impulses towards the cell body, and a single long nerve fibre called an **axon** which carries impulses away from the cell body. The structure of a motor neurone was introduced on page 211.
- **sensory neurones**
  These have a single long nerve fibre, a **dendron**, which brings impulses towards the cell body, and a single long **axon** which carries impulses away.
- **relay neurones** (also known as **interneurones**)
  These have numerous, short nerve fibres.

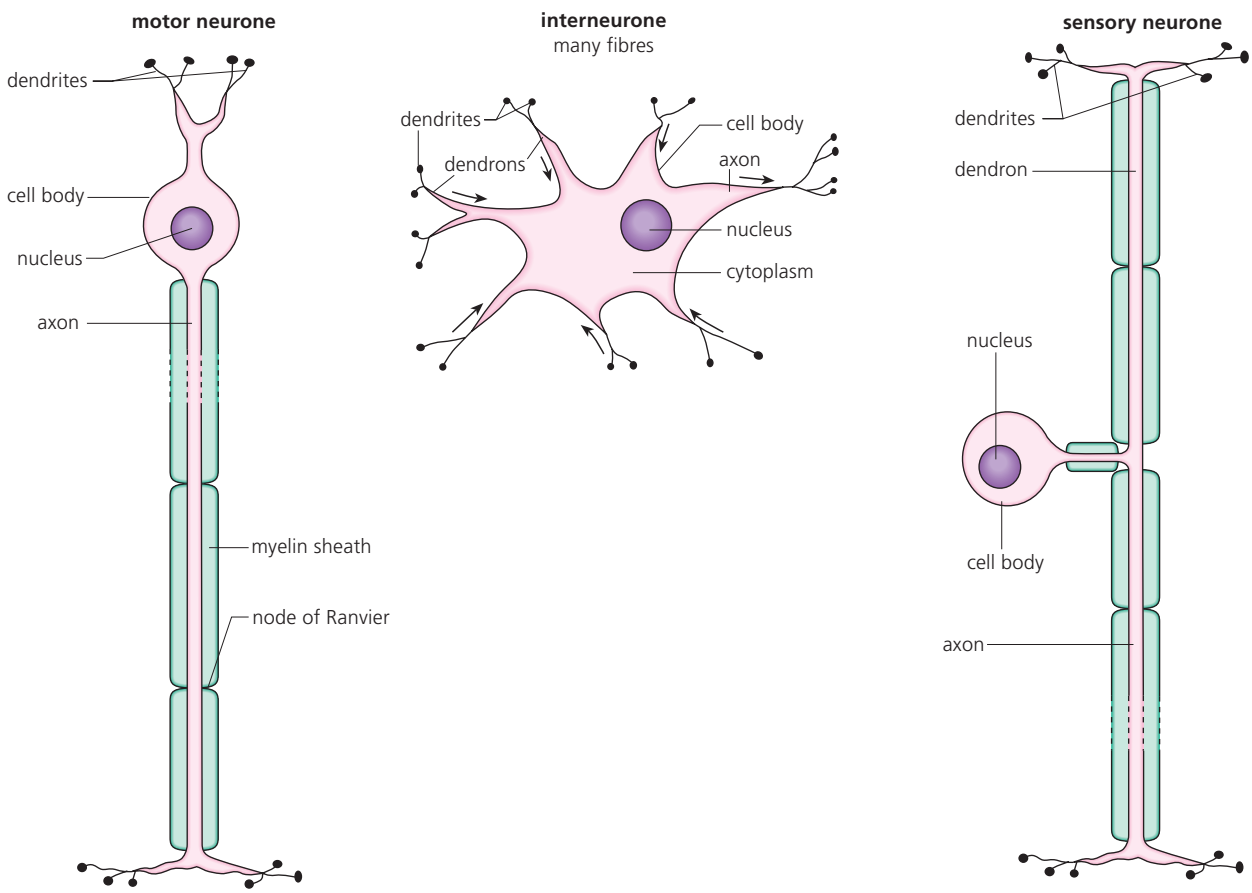


Figure 17.1 Neurones of the nervous system

The axons of motor neurones and dendrons of sensory neurones are surrounded by supporting cells called **Schwann cells**. Schwann cells wrap themselves around these fibres, forming a protective structure called a **myelin sheath** (Figure 7.31, page 211). Frequent gaps occur along a myelin sheath, between the individual Schwann cells. The gaps are called **nodes of Ranvier**. The myelin sheath with its gaps actually helps increase the speed at which impulses are conducted.

Incidentally, many of these fibres are very long – some of the longest in your body run from the base of the spinal cord to the tip of your feet, for example.

1 Compare motor, sensory and relay neurones by means of a concise table.

Nerve fibres are specialised for the transmission of electrochemical impulses, known as **action potentials** (page 212). An action potential is transmitted in a few milliseconds, so nervous co-ordination is extremely fast (and responses are virtually immediate). Because of the lengths of nerve fibres, action potentials can be transmitted over considerable distances within the body, travelling to specific points in the body they serve. So the effects of impulses are targeted and localised. Nervous control is precise. Once generated, the action potential is transmitted along the fibres of a sequence of neurones of the reflex arc to a muscle which is caused to contract, or to a gland which is then activated.

Receptor organs and effector organs

Associated with the neurones of the nervous system are the receptor organs and effector organs.

Receptor organs are the sense organs that detect change. In a sense organ, a stimulus in the form of energy (such as sound, light, or mechanical pressure) is transferred into an action potential in the nerve fibre of the neurone that serves the sense cell. Receptors are typically sensitive to one type of stimulation only (such as differences in temperature, light, touch or chemicals). Some sense organs consist of a sensitive nerve ending. Others consist of an individual cell or small group of cells, while some are complex organs like the eye (page 511) or ear (page 516), containing elaborate receptor cells within a complex supporting structure.

Effector organs are muscles and glands. On receipt of an action potential a muscle may contract and a gland may secrete. The outcome is the body's responses to the stimulus.

Reflex arcs and the response of animals to stimuli

Many of the body's neurones are assembled into pathways of impulse transmission called reflex arcs. The layout of a reflex arc is shown in Figure 17.2.

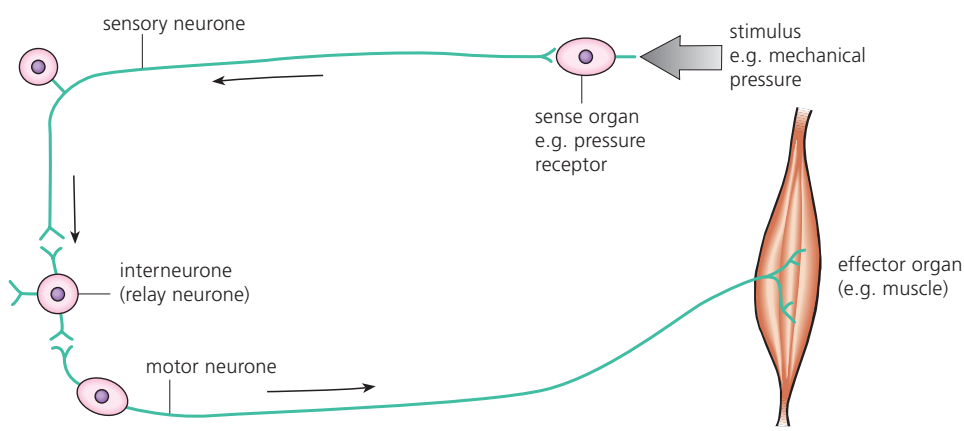


Figure 17.2 The layout of a reflex arc

In animals, by means of a specific reflex arc, a particular stimulus produces the same immediate, quick, involuntary responses – a reflex action – every time. In humans, an example of a reflex action is the jerking away of our hands from scalding hot water, or the withdrawal of a limb from pain (Figure 17.3).

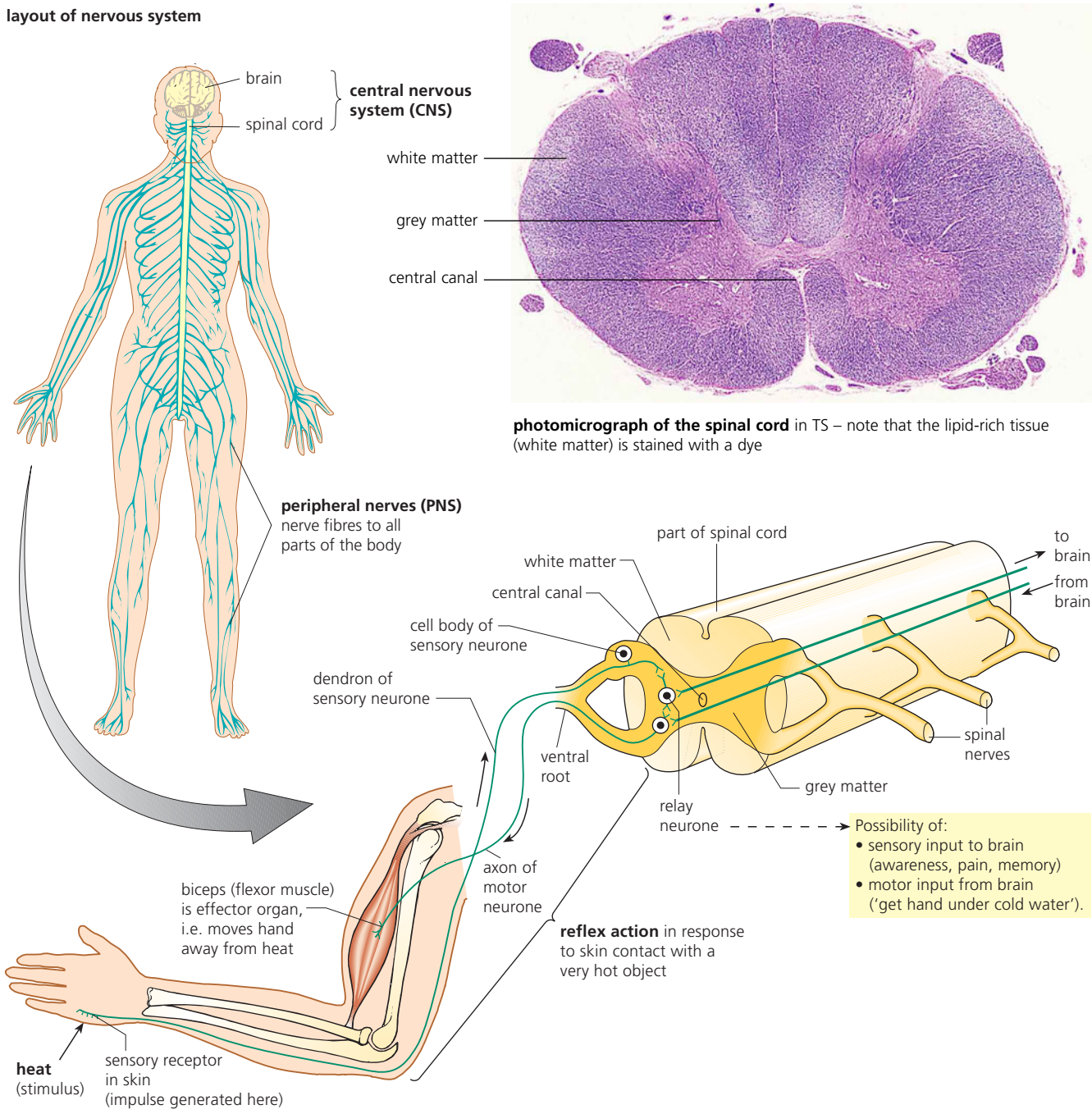


Figure 17.3 Reflex arc of the withdrawal of a limb from pain

2 Outline the source of energy used to:

a establish the resting potential
b power an action potential (page 212).

Extension: Reflex arcs and the nervous system

In vertebrates, and particularly in mammals, there is a complex nervous system. Within the nervous system are very many reflex arcs.

In addition, many neurones connect reflex arcs with a control centre, the brain. The brain contains a highly organised mass of relay neurones, connected with the rest of the nervous system by motor and sensory neurones.

With a nervous system of this type, complex patterns of behaviour are common, in addition to many reflex actions. This is because:

- impulses that originate in a reflex arc also travel to the brain;
- impulses may originate in the brain and be conducted to effector organs.

Consequently, much activity is **initiated** by the brain, rather than being merely responses to external stimuli. Also, reflex actions may be **over-ruled** by the brain, and the response modified (for instance, we may decide not to drop an extremely hot object that is very valuable).

So, we can see that the nervous system of an animal such as a mammal has roles in:

- quick and precise communication between the sense organs that detect stimuli and the muscles or gland that cause changes;
- complex behaviour patterns that animals display.

Animal responses and natural selection

Inherited traits or characteristics may be passed from one generation to the next. Does a particular trait help an organism to survive, feed, breed and produce offspring? If it does, then many of the offspring will also benefit from that trait. Those offspring that do are more likely to become parents themselves. This is the process of natural selection of favourable characteristics.

Remember, **natural selection** involves:

1 **Genetic variation**, which may arise via
   a mutations, including chromosome mutations and gene mutations;
   b random assortment of paternal and maternal chromosomes in meiosis;
   c recombination of segments of maternal and paternal homologous chromosomes during crossing over;
   d the random fusion of male and female gametes in sexual reproduction.
2 **Expression of genetic variation in the phenotype**. Some phenotypes are better able to survive and reproduce in a particular environment; **natural selection operates**, determining the survivors and the genes that are perpetuated.

Natural selection is just as applicable to an **inherited behavioural trait** as it is to structural characteristics. Such behavioural traits might be responses that favour survival in the face of food shortage, predator attack or other external danger.

The **European hedgehog** (*Erinaceus europaeus*) typically defends itself from attack by rolling up into a ball, presenting an almost impenetrable exterior of sharp spines. In the presence of a large predator, muscles underlying the skin act like the strings of a draw-string bag to bring about this response. The body and limbs are contained within the spiny shield (Figure 17.4). A hedgehog may remain in this state for hours.

In fact, a population of these nocturnal animals shows variation in their responses. Some just raise their spines, peer out from under, and check how the danger develops. If the danger persists, these hedgehogs may run away. They only roll up if they finally find themselves unable to escape.

With hedgehogs whose habitat is crossed by busy roads, the response to danger by running away may be advantageous. If 'rollers' are more likely to be crushed under the wheels of cars, vans and lorries, then road traffic is a natural selection force in behavioural responses of the hedgehog. Among local populations of hedgehogs frequenting busy roads, 'rollers' may be selectively culled during their first summer. If so, the outcome will be an increased frequency of 'runners' in the population.

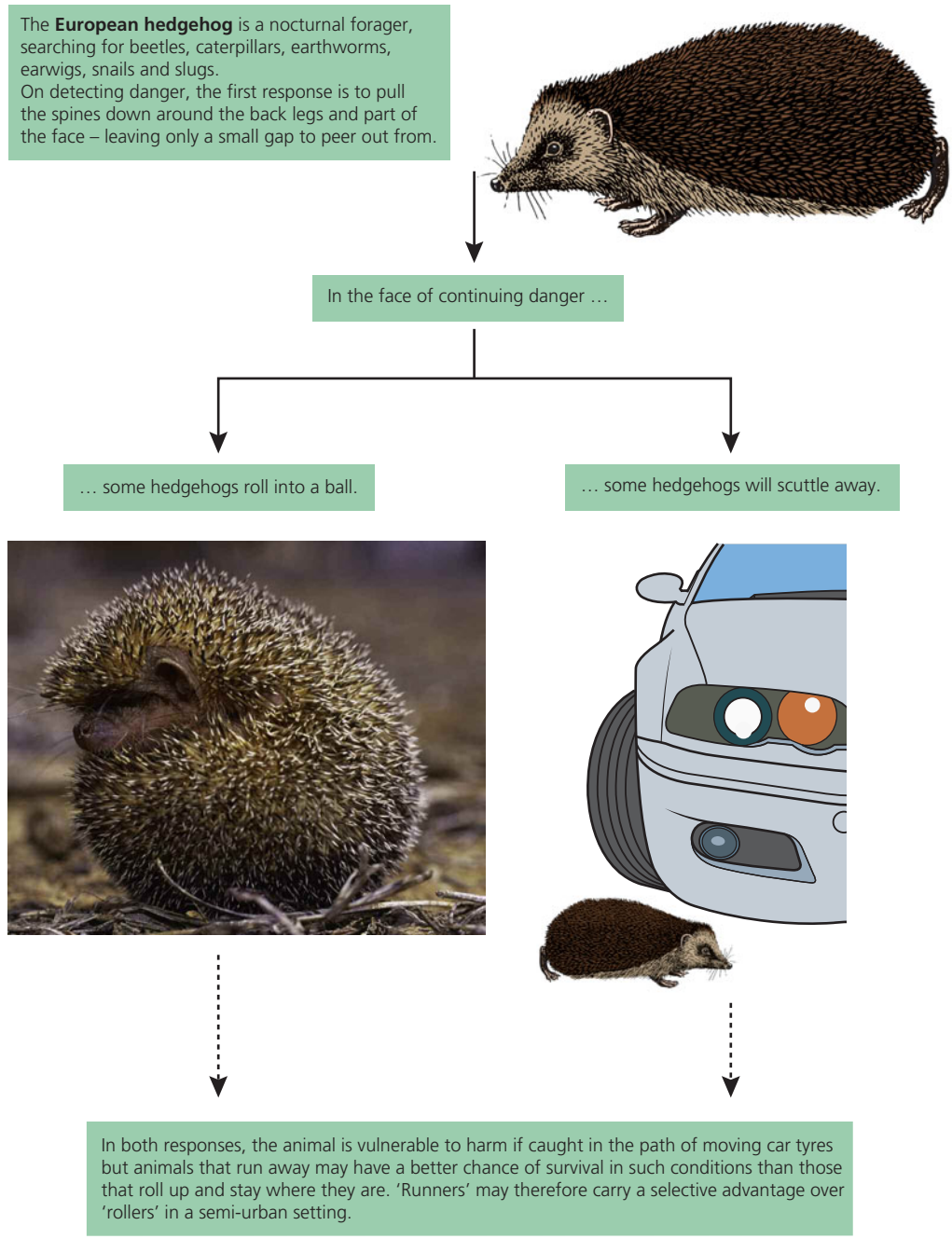


Figure 17.4 The responses to danger by hedgehogs

The fighting behaviour of male **marine iguana** (*Amblyrhynchus cristatus*) of the Galapagos Islands is another behavioural trait favoured by natural selection.

Fighting between members of the same species is almost universal among vertebrate species. This trait serves the important function of spacing out individuals into non-overlapping territories, large enough to support an animal, its mate, and their offspring. Overcrowding is prevented.

Fighting also results from competition for a mate, resulting in breeding by the stronger, more determined males. Consequently, fighting is as common among herbivorous animals as among carnivores.

However, fighting between individuals of the same species rarely ends in death – or even injury. **Fights are highly ritualised.** Evolution exerts a selective force for aggressive competition but against harmful injury to males of the same species. Harm is prevented when the loser (the weaker or younger or less determined combatant) adopts a submission posture, signalling that defeat is accepted (Figure 17.5). At the earliest opportunity, a retreat is effected.

3 Suggest one example of an animal response apparently affected by natural selection, using an animal (such as a bird species, perhaps) local to your home, school or college.

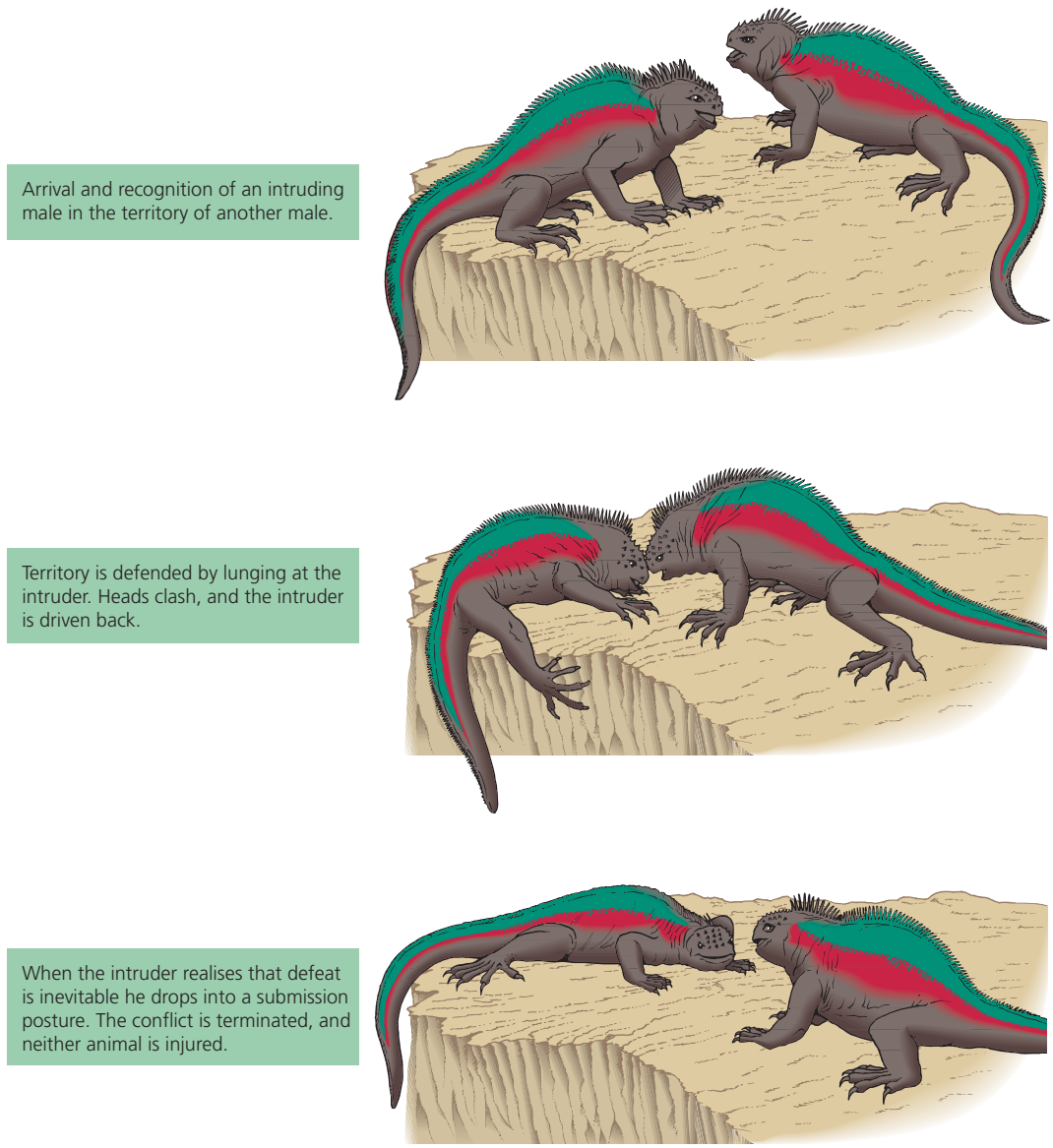


Figure 17.5 Ritualised combat to maintain territory among marine iguana

Perception of stimuli

E2.1–2.7

All cells are sensitive to changes in their environment, but **sense cells** are specialised to detect stimuli and to respond by producing an action potential. Specialised sense cells are called **receptors**.

A diversity of sensory receptors exists in the body. Some are merely sensitive nerve endings, such as the pressure receptors found below the skin and at joints in the body. Other receptors consist of an individual cell or small groups of cells, while yet others are complex organs like the eye, containing elaborate receptor cells within a complex supporting structure.

The property of a sense cell is to transfer the energy of a particular type of stimulus into electrochemical energy of an action potential (impulse), which is then conducted to other parts of the nervous system. The stimulus that the sense cell responds to is some form of energy, **mechanical**, **chemical**, **thermal** or **light** (**photic**).

In Table 17.1 are listed the sense organs of mammals under headings of the stimuli they respond to.

Sense data (form of energy)	Type of receptor	Location in the body
Mechanoreceptors, responding to mechanical stimulation		
light touch	touch receptors	mostly in dermis of skin
touch and pressure	touch and pressure receptors	dermis of skin
movement and position	stretch receptors (e.g. muscle spindles, proprioceptors)	skeletal muscle
sound waves and gravity	sensory hair cells	cochlea, etc. of inner ear
blood pressure	baroreceptors	aorta and carotid artery
Thermoreceptors, responding to thermal stimulation		
temperature change in the skin	nerve endings	dermis of skin
internal temperature change	cells of hypothalamus	brain
Chemoreceptors, responding to chemical stimulation		
chemicals in the air	sense cells of olfactory epithelium	nose
taste	taste buds	tongue
blood O_2, CO_2, H^+	carotid body	carotid artery
osmotic concentration of the blood	osmoregulatory centre in hypothalamus	brain
Photoreceptors, responding to electromagnetic stimulation		
light	rod and cone cells of retina	eye

Table 17.1 The sense organs of mammals

4 **Identify** and **list** the various types of stimuli (sense data) originating from conditions within the human body that are detected by particular receptors, using Table 17.1.

The human eye and the sense of sight

The eyes of mammals are protected in deep, bony sockets called orbits. The eyes supply information from which the brain perceives the size, shape, movement, and (sometimes) the colour of objects in the environment, and also information about the direction and intensity of light. In those mammals which have their eyes directed forwards so that the visual fields of the eyes overlap (as in the primates), the brain also resolves the slightly different information from the two retinas into a single, three-dimensional image. This is known as stereoscopic vision. The structure of the eye is shown in Figure 17.6.

The working retina – rods and cones

The retina of the eye is sensitive to light in the wavelength range 380–760 nm, known as the visible range of the electromagnetic spectrum. The structure of the retina with its two types of light-sensitive cell, the **rods** and **cones**, is shown in Figure 17.7.

Rods and cones are very elongated cells, with an outer part called the **outer segment** consisting of flattened membranous vesicles in which light-sensitive pigment is housed. The **inner segment** contains many mitochondria.

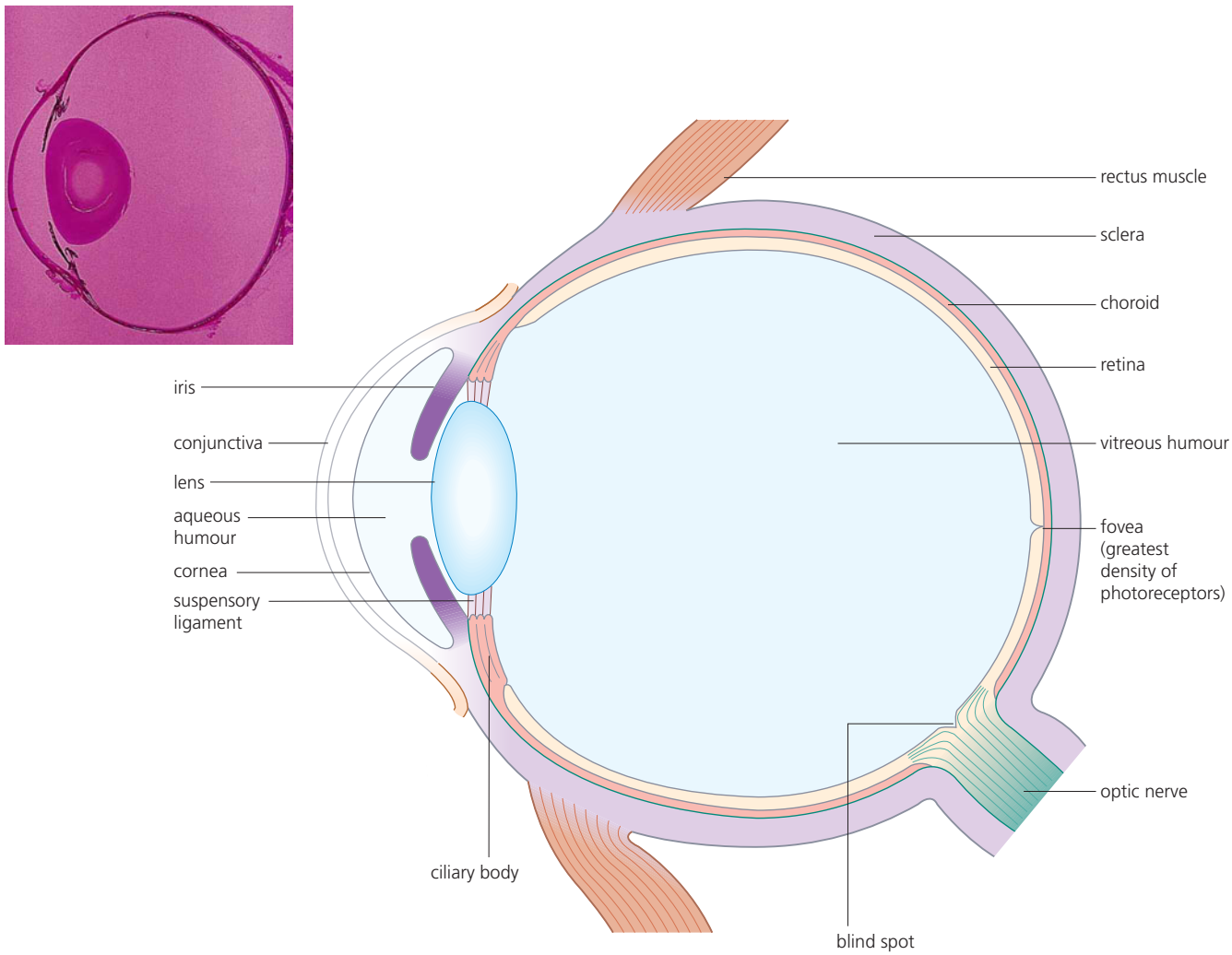
Rods are far more numerous than cones; the human retina contains about 120 million rods compared with 6 million cones. Rods are distributed evenly throughout the retina, apart from at the fovea, where they are absent.

Cones are found all over the retina but are particularly concentrated at and around the fovea, the area of most accurate vision. It is the fovea that we use when we look directly at an object.

Note that the retina is 'inverted' in that the light passes through the neurones synapsing with the rod and cone cells **before** reaching the outer segments of rods or cones.

5 **Distinguish** the structural differences between the fovea and the blind spot.

Figure 17.6 The eye



The role of rod cells

Rod cells are all the same; a rod cell is a light-sensitive cell that responds to light of all visible wavelengths, but at low light intensities. Rod cells are principally used for dim light and night vision. The visual pigment housed in the rods is called visual purple or **rhodopsin**. The rhodopsin molecule is a combination of a protein (opsin), and a light-absorbing compound derived from vitamin A, called retinal. (A diet deficient in vitamin A causes night blindness.)

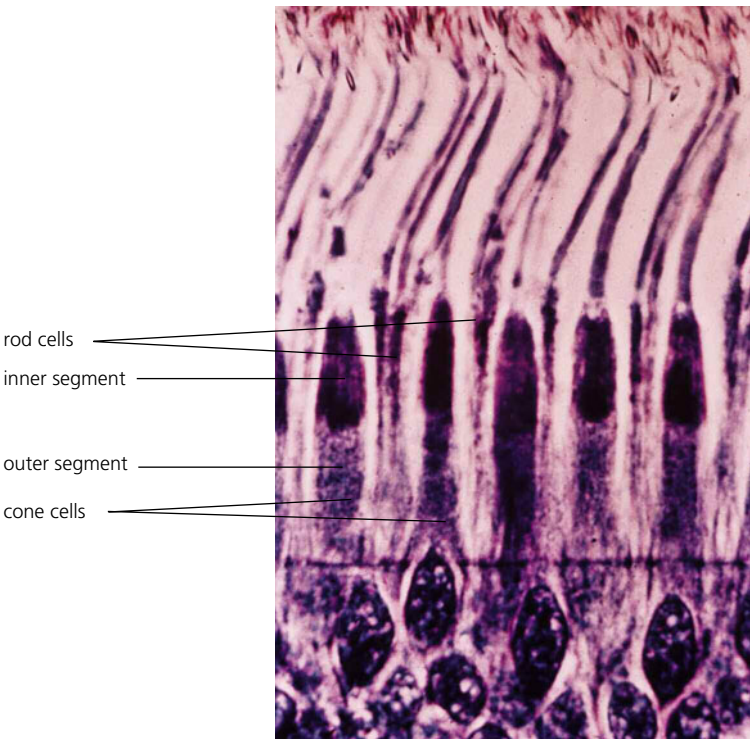
Light falling on the retina causes a reversible structural change in rhodopsin (called bleaching). This immediately affects the permeability of the cell surface of the outer segment where rhodopsin is embedded, and hence the pattern of ion movements in the rod cell is changed. As a result, the secretion of a special transmitter substance by the rod cell is altered and an **action potential** is generated in a neurone of the optic nerves serving the rod cell. This action potential is transmitted to the visual cortex of the brain.

Note that action potentials from several rod cells are fed to a single neurone of the optic nerve (via bipolar neurones – Figure 17.7), as described later.

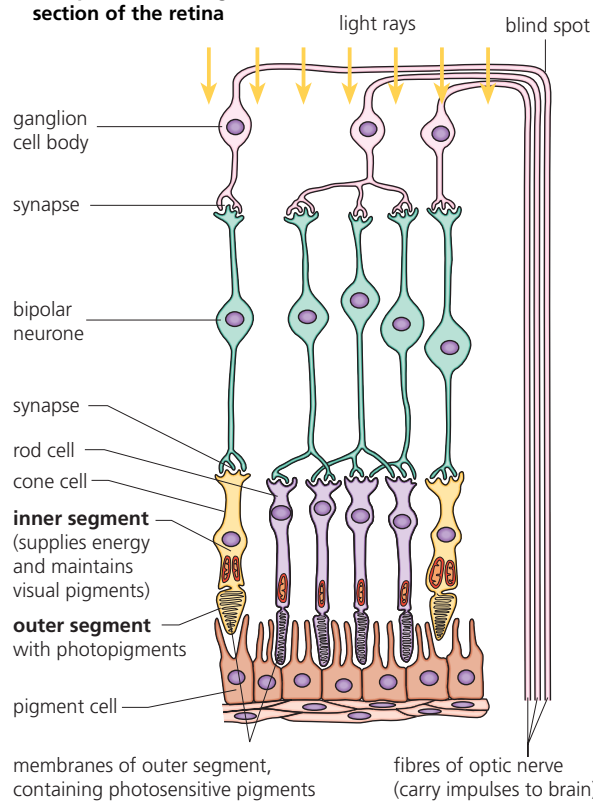
Meanwhile, following bleaching, the rhodopsin molecule is rebuilt, using energy from ATP. Of course, in very bright light, all the rhodopsin is bleached. In these conditions we are using our cone cells, so the state of the visual pigment in rod cells is not of immediate consequence. We are not aware it is temporarily bleached. But if we then move from bright to very dim light (where cone cells are useless) it takes time for sufficient reversing of bleaching to occur, and we are temporarily blinded. We say our eyes are 'adapting to the dark'.

Figure 17.7 The structure of the retina

photomicrograph of a thin section of retina, stained to show cellular structure



interpretive drawing of section of the retina



The role of cone cells

Animals that have cone cells in their retinas are able to **distinguish colours**. It is not the case for all mammals, but the human eye does contain cones, concentrated in the fovea where light is most sharply focused. Cone cells operate on the same principle as the rod cells, but with a different pigment, called **iodopsin**. This is less readily broken down; it needs more light energy. Cones work only in high light intensities; we cannot see colours in dim light.

According to the **trichromatic theory** of colour vision, there are three types of cone cell present in the retina, each with a different form of iodopsin. These absorb different wavelengths of light – in the blue, green and red regions of the spectrum. White light stimulates all three types equally, but different colours are produced by the relative degree of stimulation of the three types of cone.

Processing visual stimuli in the retina

There is a further, very important difference between the arrangements of cones and rods in the retina. The action potentials from each individual cone cell are fed to a single neurone of the optic nerve (via a singular bipolar neurone), whereas many rod cells synapse with a single bipolar neurone. This latter condition is known as convergence, for obvious reasons. This is shown in Figure 17.7.

Check this detail again now.

This difference has implications for **resolution** (visual acuity) – how much detail we can see. Since cone cells, when they occur in the retina, synapse with a single bipolar neurone which synapses with a single ganglion cell, each part of an image is detected by a separate cell, and there is no blurring of boundaries. With cones, there is a high degree of resolution.

On the other hand, since very many rod cells synapse with a single bipolar cell, parts of an image falling on rod cells will be poorly resolved. Convergence gives poor resolution. However, there is also a distinct advantage in this arrangement. Visual sensitivity at low light intensity increases with convergence since information from many rods is pooled to generate an action potential (summation). We can still see at very low light intensities.

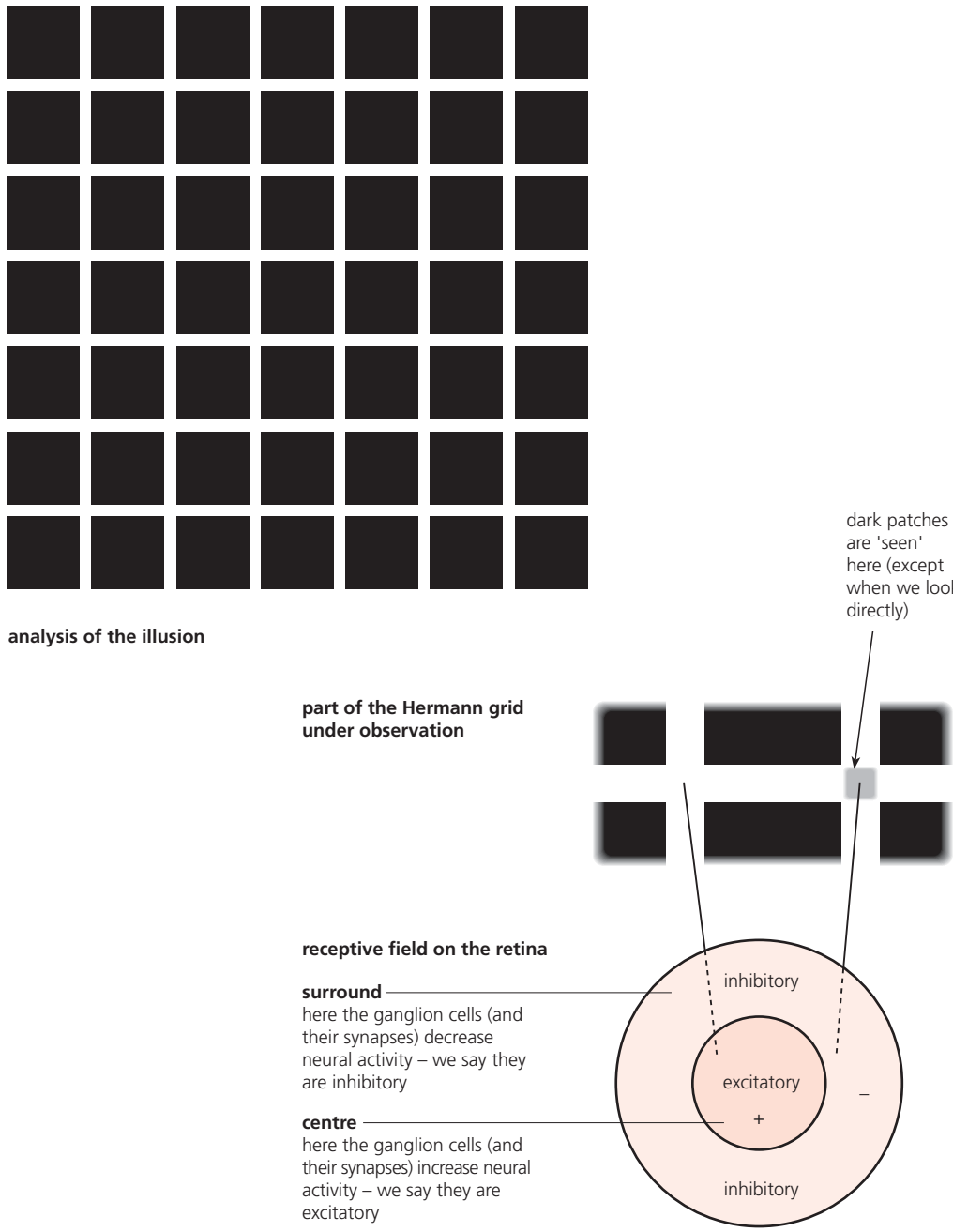


Figure 17.8 Demonstration of edge enhancement with the Hermann grid illusion

Another phenomenon, known as edge enhancement, occurs within the retina. Edge enhancement is best demonstrated by the Hermann grid illusion (Figure 17.8).

When we examine a particular grid of black squares on a light background we immediately have the impression that there are ghost-like grey blobs at the intersections of the surrounding white lines. However, the grey blobs disappear as soon as you look directly at any one of them.

Why does this happen?

The answer is that the analysis of the massive amount of information from our visual images begins in the retinas of our eyes. The edge enhancement phenomenon is a product of this process.

Each eye's receptive field is a circular patch of the retina there. The ganglion cells of the optic nerve neurones (and particularly their synapses) respond differently, depending on whether the impulses are from light falling on the centre of the circle (the fovea) or on the surrounding area. The on-centre ganglia increase their activity, but those of the outer ring decrease their activity. The decreased activity of the surrounding area is known as lateral inhibition.

As a consequence, intersections in the grid viewed here appear grey (until examined directly). Thus, nerve cells of the retina associate and interact with each other to send a map of our visual field to the brain, highlighted where there are changes in the levels of illumination – that is, at the edge of objects.

Processing action potentials from the retina, received in the brain

Observations of the three-dimensional world about us are reduced to two-dimensional images on the surface of the retina. As a consequence, action potentials generated in the rods and cones are carried by neurones of the optic nerves to the visual cortex of the brain. While each eye views left and right sides of the visual field, the brain receives and interprets action potentials from the right and left visual fields on the opposite side of the visual cortex. This is known as **contralateral processing** (Figure 17.9).

The messages from interpretation of these action potentials are combined by the brain to produce a single impression, our sight. Seeing, therefore, occurs in the brain, and the seeing process, known as **perception**, is complex.

Perception involves the interpretation of sense data from the retina in terms of our existing and past experiences, and our expectations. In large part, it is a subjective process. Human vision is, therefore, a complex process, and the phenomenon of perception has implications for the nature and reliability of visual sense data.

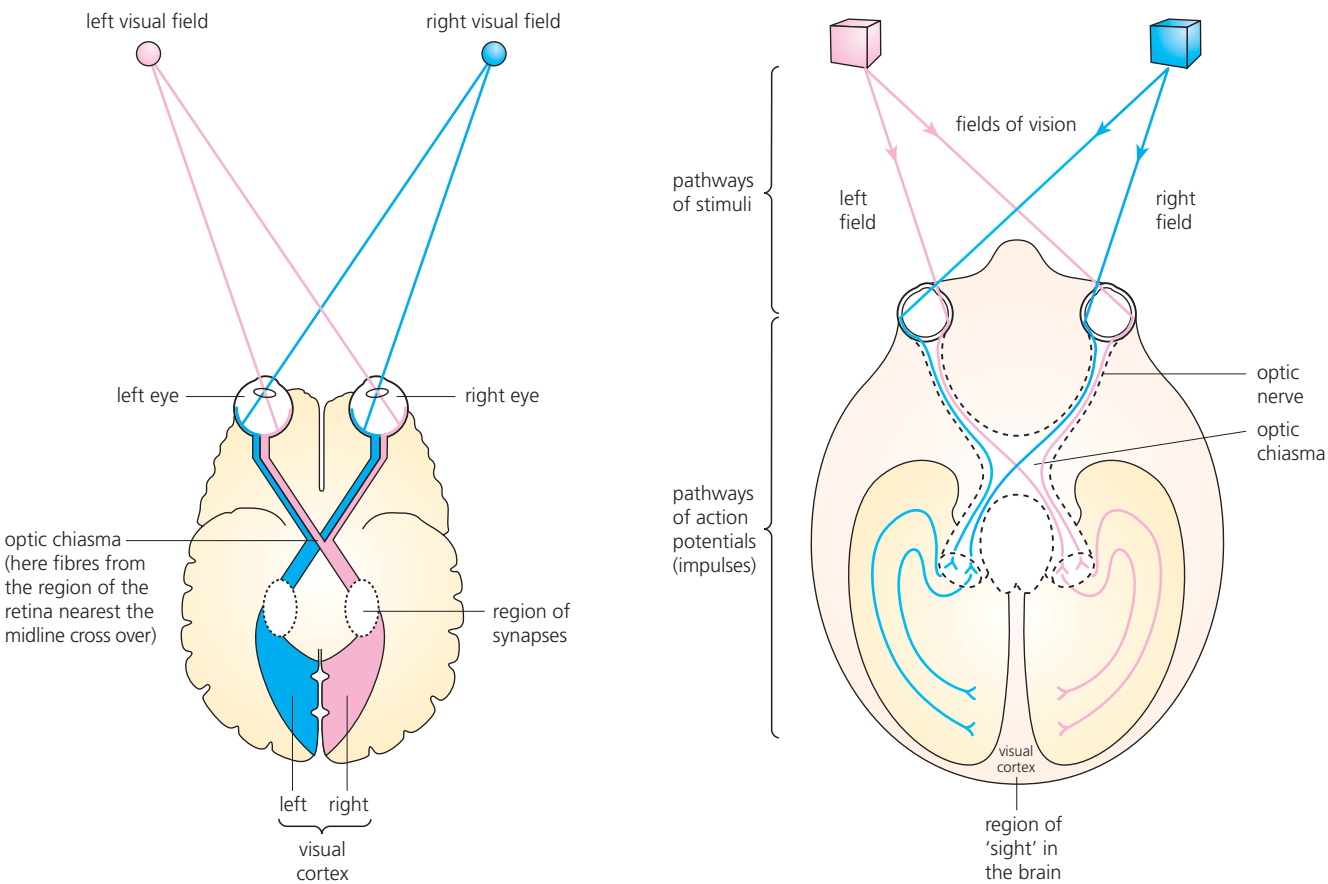


Figure 17.9 The eyes and the visual cortex – the pathways of impulses

TOK Link

What are the implications for the reliability of science data of the realisation that we see largely in our brains, rather than our eyes?

Extension: Strokes (brain lesions) and abnormal vision

Brain cells are critically dependent on blood supply – interruption of blood flow for a few minutes causes damage to neurones that is almost always irreversible. We say a stroke has occurred.

The effects of strokes are very variable, depending on which region of the brain they occur in. Damage to motor areas may trigger paralysis; damage to the visual cortex may interfere with sight. Other mental faculties normally remain intact.

What we know about the complexity of the interpretations of impulses from the retinas has been learned, *in part*, from the experiences of such patients. For example, a stroke affecting the visual cortex of the left cerebral hemisphere may deprive the patient of colour vision in the right half of the field of view (everything in that half is in shades of grey). We know, as a result of these cases, that colour is analysed separately from other aspects of vision. Other patients have lost the facility to detect motion – so motion is analysed in its own separate area. There are many other aspects of vision that have been detected as a result of similar vascular accidents in the brain.

6 Compare by means of a table, the similarities and differences between rod and cone cells of the retina.

The human ear and the sense of hearing

The ear performs two distinct and major sensory functions, those of hearing and balance. There are three regions of the ear: the outer, middle and inner ear. Each region has a distinct role. The structure of the ear is shown in Figure 17.10.

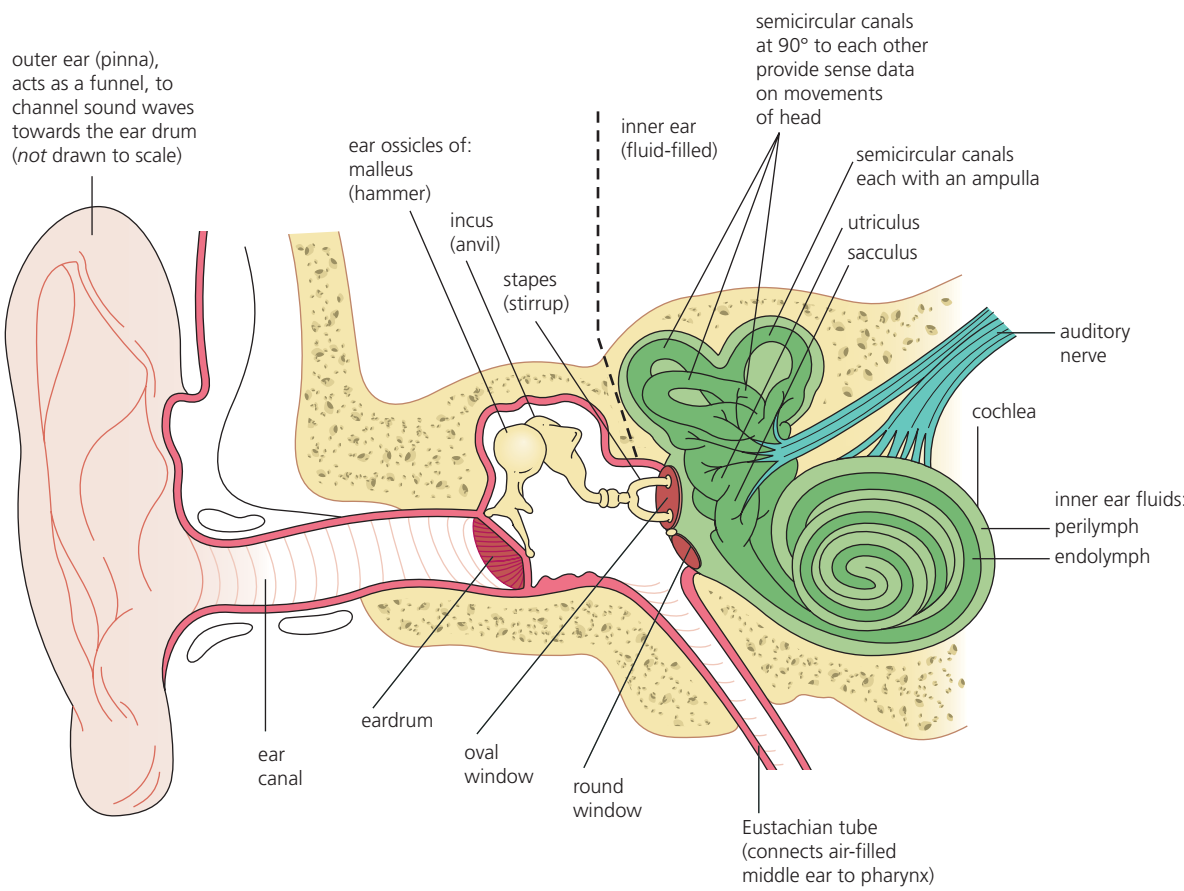


Figure 17.10 Structure of the ear

How sound is perceived in the ear

Look at Figure 17.10. Locate the three regions of the ear.

The **outer ear** (pinna) is a flap of elastic cartilage (you can feel how flexible yours is) covered by skin, and a tube (ear canal) leading to the eardrum. The pinna acts as a funnel, channelling sound waves towards the eardrum. Dogs, and many other mammals (but not humans), have mobile pinnae, and use them to locate the source of a sound.

The **middle ear** is an air-filled chamber cut off from the outer ear by the eardrum, and from the fluid-filled inner ear by the oval and round windows. The Eustachian tube connects the middle ear with the pharynx (back of the throat), but is normally closed. It opens briefly when swallowing. When it is open, it ensures that the air pressure is equal on both sides of the eardrum. The **eardrum** is a thin, strong sheet of elastic connective tissue that vibrates in response to sound wave pressure.

The smallest bones in the body occur in the middle ear. These bones, known as the hammer (malleus), anvil (incus) and stirrup (stapes), together traverse the middle ear, from eardrum to oval window. The **bones of the middle ear** form a lever system, increasing the effective pressure on the oval window by 20 times.

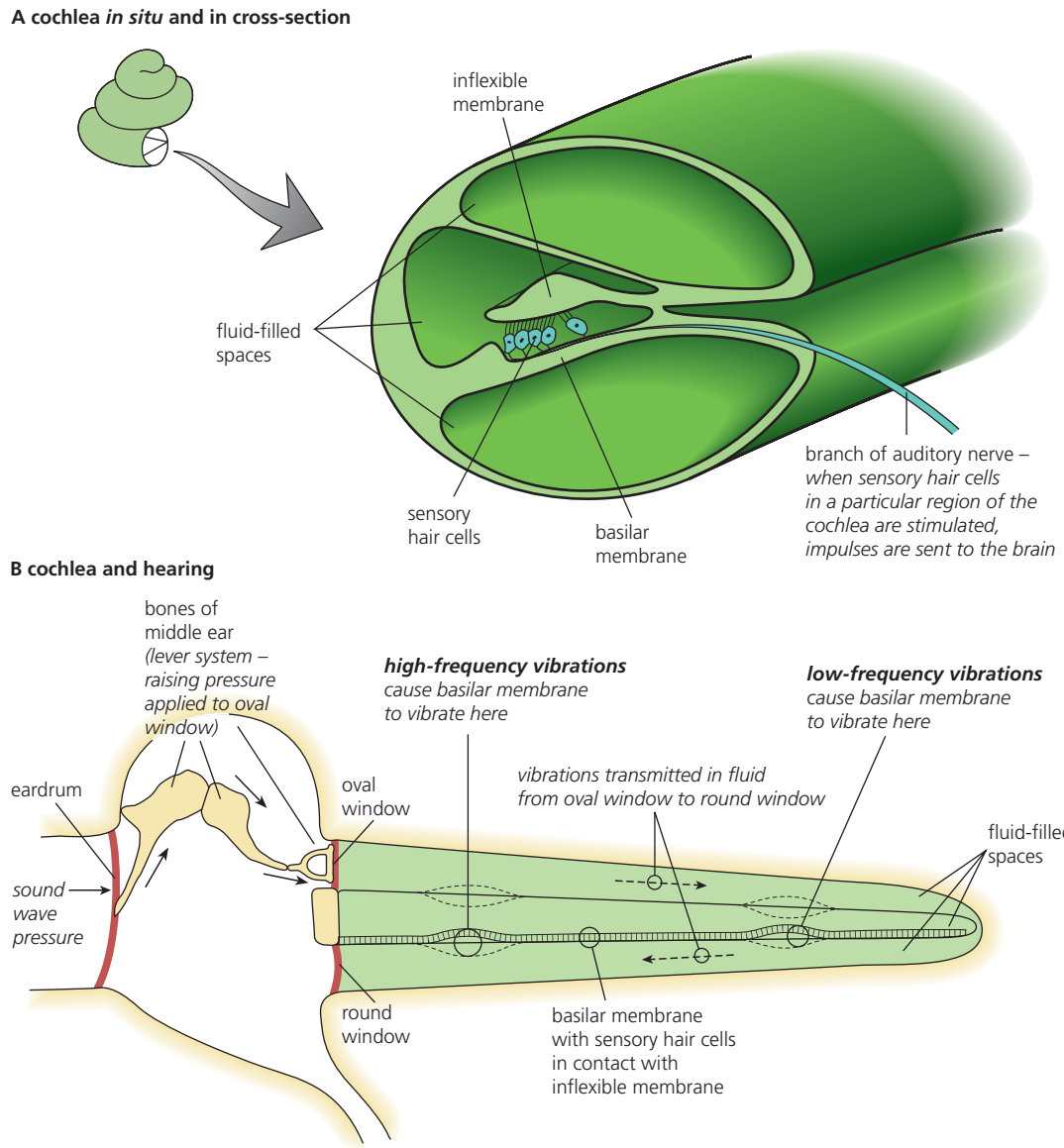


Figure 17.11 The cochlea and hearing

The **inner ear** is a fluid-filled cavity of membranous canals, surrounded by extremely hard bone. The inner ear consists of the cochlea (concerned with transfer of sound waves to nerve impulses) and the semicircular canals (concerned with balance).

The **cochlea** (Figure 17.11 A) consists of a spirally coiled, fluid-filled tube, which is divided longitudinally into three compartments, separated by membranes. The upper and lower compartments communicate at the tip of the cochlea. The upper compartment is in contact with the oval window. The lower compartment is in contact with the round window. Between the two canals is the middle compartment. This consists of the **basilar membrane**, and projecting into the canal is an **inflexible membrane** parallel with the basilar membrane, and running the full length of the cochlea. Immediately beneath the inflexible membrane, the basilar membrane supports the sensitive **hair cells of the cochlea**. These sensory cells connect with the **auditory nerve**, and they have sensory hairs projecting upwards, making contact with the inflexible membrane. There are about 25 000 hair cells in total in the cochlea.

Vibrations from the eardrum enter the cochlea via the bones of the middle ear. Since liquid cannot be compressed, the vibrations cause movements in the fluid of the inner ear, and are eventually transmitted to the round window. These pressure waves cause the hair cells attached to the basilar membrane to rub or pull against the inflexible membrane. The resulting movements of the sensory hairs cause the production of an action potential that is transmitted to the brain via a branch of the auditory nerve. In different regions along the cochlea, different wavelengths cause the basilar membrane to vibrate (Figure 17.11 B). In this way, different wavelengths are detected.

Innate and learned behaviour

E3.1–3.6

Behaviour is defined as the way organisms respond to the environment and to other members of the same species. Here we are concerned with aspects of animal behaviour. The activities of animals enable them to survive, to seek out favourable environments, and to reproduce. Behaviour is based on feedback, using the control and co-ordination machinery of the body (this includes the sense organs, nervous system, and the effector organs).

Behaviour is sometimes said to be either **innate behaviour** (instinctive – that is, automatically triggered in certain circumstances) or **learned behaviour**. Innate behaviour includes behaviour that is due to a reflex action. Learning occurs when experiences are retained and used to modify behaviour on future occasions.

Innate behaviour develops independently of the environmental context, whereas learned behaviour develops as a result of experience.

However, in the natural world, the differences between innate and learned behaviour are not always clear cut. Rather, many animals display a range of behaviour, some with innate features and some which is clearly learned in part, at least. In vertebrates and especially in mammals, very complex behaviour patterns are common. Study of behaviour at this level is largely outside the scope of this book.

Innate behaviour in non-vertebrates – an experimental approach

Responses of motile organisms (or motile gametes) to external stimuli may enable the individual organism to position itself favourably in the environment. For example, many non-vertebrate animals have very simple patterns of behaviour in which the direction or rate of movement is a response to a stimulus. Further, their actions appear to be largely determined by reflexes.

Two types of such behaviour can be demonstrated under conditions where a single stimulus can be applied under controlled conditions (such as a laboratory experiment), as follows:

1 **Kinetic movements** or kineses are movements in which the **rate of movement is related to the intensity of the stimulus**, but its direction is random. Examples include woodlice, which are seen to move randomly and quickly in dry conditions, but slow down and stop when their movements bring them into a humid area. Another example of a kinesis is the slow, random movement of the tentacles of *Hydra* that may bring the tentacles in contact with a food source, and which speeds up when a food source is nearby.
2 **Tactic movements** or taxes occur when the **direction of the stimulus determines the direction of the response**. Examples include the flatworm *Planaria* moving towards food (chemotaxis), and the photosynthetic unicellular protoctistan *Euglena* moving towards light (phototaxis).

Investigating the response of woodlice to humidity – a kinesis

The response of organisms to environmental conditions such as humidity may be investigated in a simple **choice chamber apparatus** made from a plastic Petri dish. The atmosphere on one half of the dish would need to be very humid, and that of the other half dry, so that across the dish there is a **gradient in humidity**. In such an experimental set-up, rapid movements of the woodlice in dry areas would increase the probability that they discover moist regions – and become less active there.

Guidance in designing an experiment

Designing an experiment, the **aim** of which is to investigate the effect of a humidity gradient on the distribution of woodlice, would include the following issues.

Risk assessment
- In your experiment, would good laboratory practice be sufficient to avoid a hazard?

Method
- What design of Petri dish choice chamber would be most appropriate?
- How could a moist atmosphere be maintained in one half, and low humidity in the other half, of a lower compartment to the chamber?
- On what type of surface would the woodlice be free to move about while experiencing the gradient in humidity within the chamber?
- How many woodlice would be appropriate for each experiment?
- Under what conditions should the woodlice be maintained prior to use?
- How and where should the woodlice be introduced into the chamber?
- How frequently would you need to examine and record the distribution of the woodlice?
- What other factors would you maintain constant (control)?
- How long should the investigation be carried on?
- Is any form of automated data-logging possible and appropriate in this enquiry?
- How should the experimental animals be treated after the experiment?

Data recording, presentation and analysis
- Should a pre-designed table be used to record observations?
- To make any trends or correlations in the data self-evident, should data presentation involve graphs, bar charts, histograms, kite diagrams, scatter graphs or pie charts?
- Are there statistical tests that should be applied to the numerical relationships the data indicate?

Discussion of the results
- Was the response of the woodlice best described as a kinesis or as a taxis, and why?
- How does the observed behaviour pattern relate to the way of life of the woodlice?
- How would you modify your method if you were to repeat the experiment in future?
- What further investigations might be appropriate in the light of the outcomes?

Investigating the response of blowfly larvae to light – a taxis

Blowflies (also known as bluebottles, *Calliphora* sp.) are insects that lay eggs, some species in the carcasses of dead animals, others in animal dung, and at least one species in any damaged skin of living animals, such as sheep. When the larvae hatch out (maggots, known as gentles to fishermen who may use them as bait), they eat their way through their chosen substrate. When fully grown, the maggots pupate and, in time, hatch as blowflies.

Blowfly maggots can be negatively phototactic (this might help in the maintenance of a feeding position) or positively phototactic (which might help at pupation).

Guidance in designing an experiment

Designing an experiment, the **aim** of which is to investigate the response of blowfly larvae to light varying in direction, would include the following issues.

Risk assessment

- If experiments are to be carried out in darkness or very low light, are there safety implications?
- Does handling the organism raise any risk of infection that needs guarding against?
- Otherwise, would good laboratory practice be sufficient to avoid a hazard?

Method

- Under what conditions should blowfly larvae be maintained, prior to use?
- How many larvae would make a significant sample?
- How may light be significantly reduced or excluded from the experimental area?
- On what surface could the investigation be best carried out, such that the path of a moving larva can be tracked and recorded?
- How may a unilateral beam of light be introduced for a controlled period?
- How might the intensity and direction of light be varied, if at all?
- How are the movement responses of larvae when exposed to a light stimulus to be recorded?
- Should time (i.e. speed) be recorded as a factor, in addition to distance and direction of movement?
- What other factors would you maintain constant (control)?
- How long should each investigation be carried on for?
- Is any form of automated data-logging possible and appropriate in this enquiry?
- How should the experimental animals be treated after the experiment?

Data recording, presentation and analysis

- How can the results of individual experiments be pooled to summarise the response of a significant sample size of larvae?
- To make any trends or correlations in the data self-evident, should data presentation involve graphs, bar charts, histograms, kite diagrams, scatter graphs or pie charts?
- Are there statistical tests that should be applied to the numerical relationships the data indicate?

Discussion of the results

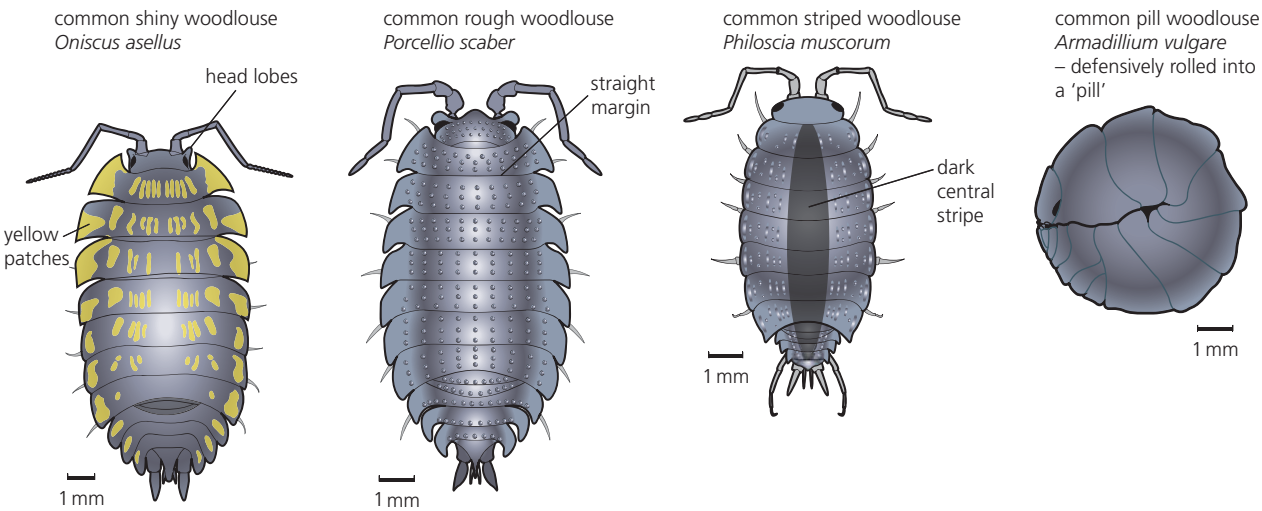
- Can the response of blowfly larvae be described as positively or negatively phototactic, and why?
- How does the behaviour of the larvae relate to their way of life?
- How would you modify your method if you were to repeat the experiment in future?
- What further investigations might be appropriate in the light of the outcomes?

7 Design a simple experimental procedure to investigate the response of a culture of *Euglena* (a unicellular green protoctistan, common in fresh water) to a unidirectional light stimulus.

Analysis of data from behaviour experiments with non-vertebrate animals

The environment provides a wealth of opportunities and conditions favourable to non-vertebrates, but is equally the source of threats and dangers. Such threats may be both abiotic (such as excess heat, drought and danger of desiccation), and biotic (such as attack by predators). Orientation behaviour in non-vertebrate animals may improve the individual's chances of locating favourable conditions and remaining there, and avoiding dangers. Thus they improve the chance of survival and reproduction. We will examine two examples of this.

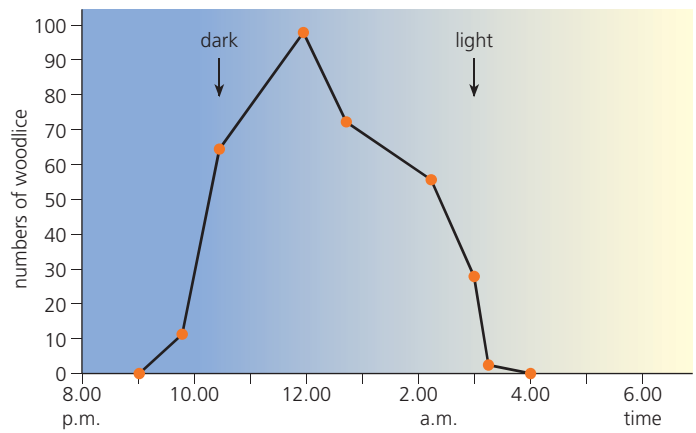
Figure 17.12 Behaviour patterns in woodlice



marked length of garden wall where woodlice are observed to browse and scavenge at particular times

In the hours of daylight the woodlice were found along the base of the wall, under stones, logs and rotting leaves.

numbers of woodlice moving on the surface of the wall on nine occasions spaced out between dusk and dawn (study undertaken on a warm, humid night)



Nocturnal activity of woodlice

Woodlice are small oval animals that occur widely, often found inside our buildings, but typically discovered under logs and stones, in bark crevices, and among dead leaves and rotting plant material wherever this occurs. The bulk of the diet of most species consists of dead and decaying vegetable material. In their turn, they frequently fall prey to shrews, toads, ground beetles, centipedes and some spiders. However, many species can move quickly, and woodlice may stray far and wide in their foraging, in the hours of darkness.

Woodlice belong to a class of the **arthropod** phylum (animals with a hard external skeleton covering the segmented body, typically with a pair of jointed limbs per segment, page 172) known as **crustaceans**. Crustaceans are predominantly aquatic arthropods and include the crabs and lobsters.

Woodlice now live on land but have not fully adapted to terrestrial conditions. For example, their skins are not completely waterproof and almost all species of woodlice are confined to damp or humid places. They are nocturnal.

The first five pairs of limbs on the central part of the woodlouse body are unlike the other legs – they consist of two, leaf-like flaps, very thin and well supplied with blood. These act as gills when surrounded by a film of water. In very dry weather, the gills fail – so woodlice suffocate as well as desiccate in extremely dry conditions. Submerged in water, such as short-lived pools or gullies of rain water, they easily drown – woodlice are just as easily killed by conditions that are too moist as too dry.

All these factors affect their chances of survival and reproduction, and so can be seen to influence their choice of habitat and their behaviour.

Analysis of the graph in Figure 17.12 demonstrates the role of orientation behaviour in the survival of woodlice.

- In the day, woodlice move away from light (negatively phototactic) and maintain themselves in a humid environment (kinesis).
- After dark, they engage in feeding movements (chemotactic movements) which bring them onto walls (in this case). It may be that fewer woodlice are predated on these types of surface. They are less likely to drown in sudden, heavy rain in this type of foraging territory too.
- As soon as light starts to return, they once again respond negatively phototactically.

With this distinctive behaviour pattern, the woodlice are more likely to survive and reproduce.

Foraging by bumble bees

Bumble bees are one of several groups of insect that regularly visit flowers to collect nectar and/or pollen as food, and which incidentally bring about pollination and, frequently, cross-pollination.

The evolution of the whole range of flowering plants (angiosperms), many with flowers whose petals are fused into long tubes, has been a co-operative venture between certain insect groups and flowering plants, to mutual advantage. One of the insect groups involved is the bees.

So, bees and flowers have evolved together, each gaining advantage for themselves with the minimum of investment and expense. We can illustrate this relationship by observation of bumble bee visits to the inflorescences of foxgloves (Figure 17.13).

In the bumble bee's life cycle, the availability of nutrients (nectar and pollen) in flowers is critical. The large bumble bees seen in springtime are mated queen bees that have hibernated over winter (the only members of a colony to survive the unfavourable season).

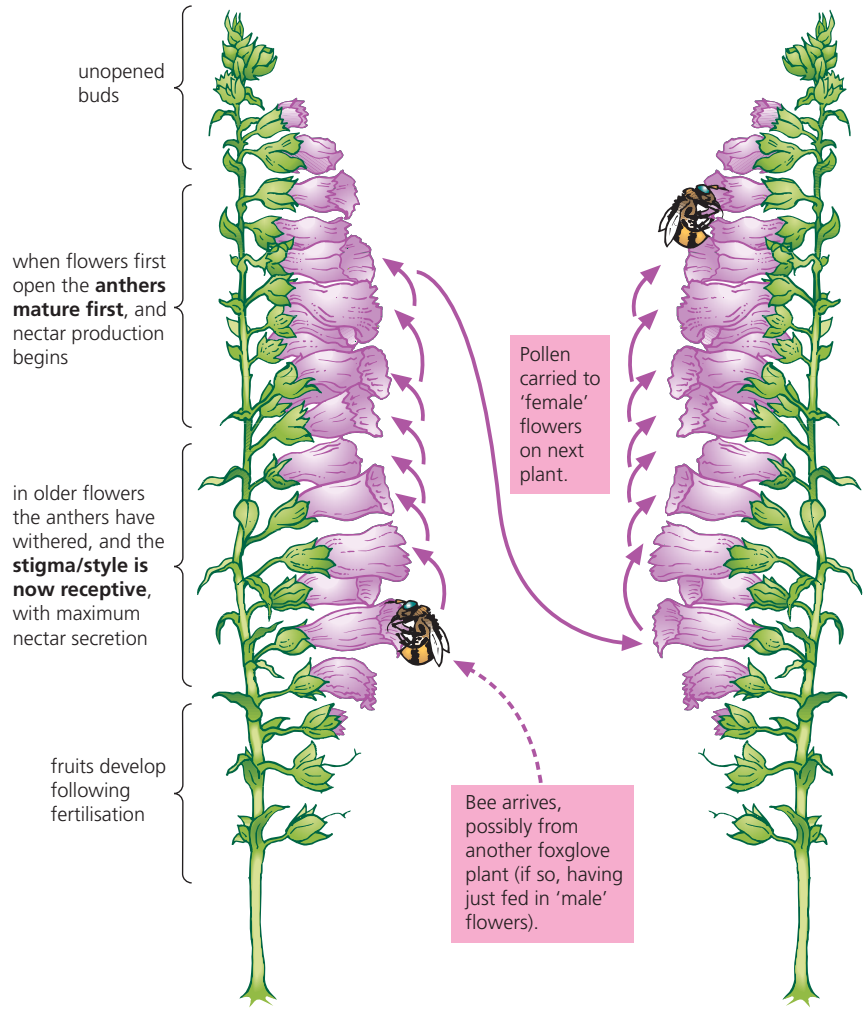
Once flowers become abundant in late spring, each queen builds an underground nest where she lays up a food store and adds eggs. They are then surrounded by a wall of wax that she produces.

When the new bees emerge (as grubs, which grow into bees) they eventually become new worker bees. They forage for nectar and pollen, and take over the job of feeding succeeding generations of young. The queen continues to lay eggs, and later in the year these develop into new queens (fertile females) and male bumble bees. The new queens then mate, and eventually hibernate to continue the cycle the following year, if they survive. This whole social colony is maintained on resources foraged from flowers in spring and summer.

Figure 17.13 Typical foraging visits of bumble bees to foxglove flowers

Foxglove (*Digitalis purpurea*) is a species found in woodlands on acid soils, throughout the UK and on mainland Europe from Norway to Spain, and in Morocco (Africa). (The drug digitalin, obtained from this plant is used extensively for heart complaints.)

typical foraging path of bumble bee presented with a group of foxglove inflorescences



The foxglove inflorescence maximises the chances of cross pollination. The bumble bee's behaviour minimises the investment of energy in flight and maximises the gain of nectar.

Optimal foraging is essential if a queen (or a worker bee) is to maximise its acquisition of essentials when visiting inflorescences. What path will it take, when presented with flowers grouped together in an inflorescence, each flower being a potential source of nectar and pollen? Figure 17.14 shows four patterns of artificial flowers, put together as they frequently appear in nature, by an experimental behavioural scientist.

Bumble bees normally start at the lowest point in an inflorescence. Plot the route you think most likely to deliver the maximum yield for a bumble bee. Then compare your solution with the observed results obtained by the experimental scientist who undertook this original investigation (Figure 17.15). Is the bumble bee's response to the foraging possibilities likely to maximise chances of survival and reproduction?

Figures 17.14 and 17.15 are shown over the page (on page 524). When you turn over, first cover the lower illustration (Figure 17.15) before you look at Figure 17.14 and work out your answer.

8 Construct a list of the characteristics of 'instinctive behaviour' as you now understand it.

Figure 17.14 Four patterns of artificial flowers, set up to study foraging by bumble bees

Artificial 'flowers' were created and arranged in different patterns on a perspex board (simulating an inflorescence in each case), to investigate a bumble bee's chosen foraging flight path.
The flowers were filled with 5 µl of a dilute sucrose solution – simulating nectar.

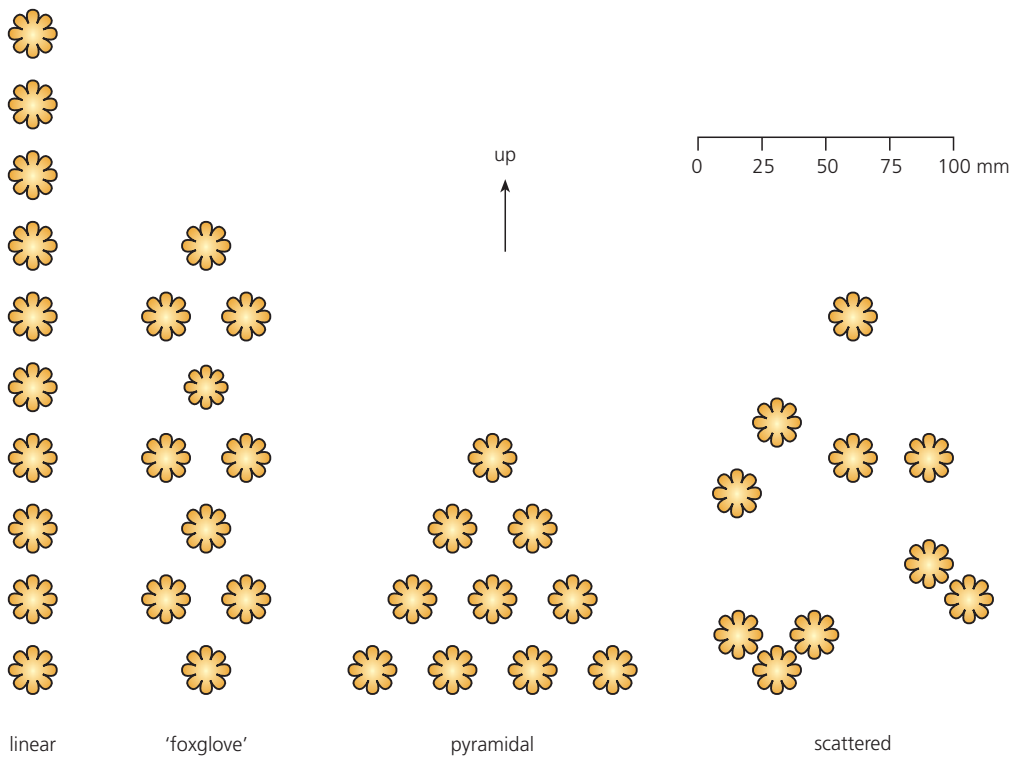
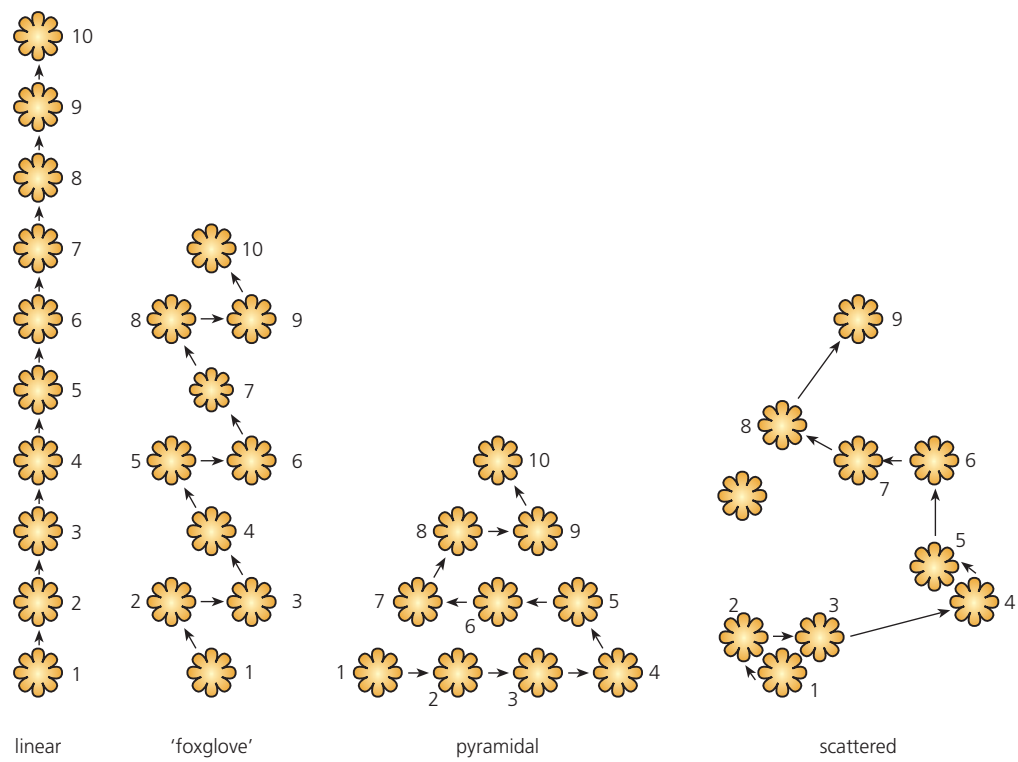


Figure 17.15 A bumble bee's chosen path over the patterns of artificial flowers

summary of experimental data obtained with the apparatus shown in Figure 17.14



Learned behaviour and conditioning

When an animal changes its behaviour in response to some development in its environment, the change may be due to **learning**.

Learning permits quick adaptations in changing circumstances; it is acquired by experience and modified in the light of further experience. Since learning is based on experience, learned behaviour is typically individual-specific, rather than being a characteristic of the whole species. It depends on what happens to an organism as to what is learned. Various forms of learning behaviour are shown by animals. We look at these, next.

In **habituation**, repeated application of a stimulus results in decreased responsiveness. This is the simplest form of learned behaviour. Habituation is illustrated when a crawling snail is touched with a leaf and retreats into its shell. Soon it re-emerges. Every time the snail is touched in this way it withdraws, but for a shorter period. Eventually it does not respond protectively at all. Thus, a repeated stimulus that brings no danger can safely be ignored, so saving loss of feeding time and thereby increasing the chances of survival and reproduction for the snail.

Similarly, a flock of birds, such as pigeons, are driven away from agricultural crops when a bird-scaring device that sounds like a gunshot is first installed, but later the birds feed undisturbed, despite the continuing noise.

Imprinting is a form of learning that occurs in an early, very receptive stage in the life of birds and mammals. This was first demonstrated by Konrad Lorenz, working with greylag geese (Figure 16.9, page 466). The advantage of imprinting is the establishment of a more-or-less instantaneous bond (working relationship) with parents who go on to impart the essential skills for survival, including, for example, feeding and communication.

Conditioning is a form of learning associated with a reward (or punishment). The work of the Russian experimental physiologist, **Ivan Pavlov** (1849–1936) in St Petersburg first demonstrated this phenomenon, as part of an investigation of digestion.

It is a common observation that a hungry dog may salivate (saliva drips from the mouth) on sight or smell of food. Pavlov studied this response under laboratory conditions that ensured the dog received no unintended additional stimulus (Figure 17.16).

In his experiment, a second stimulus, not directly related to food or feeding, was then introduced. In Pavlov's first series of investigations this stimulus was the ringing of a bell (later he used the ticking of a metronome). The sound of the bell produced no response initially – the dog continued to salivate only on sight or smell of food.

However, after several experiences of consistently hearing the ringing of a bell **at the same time as food was seen or smelt**, the dog became conditioned to salivate whenever the bell was rung, even without the food stimulus (Table 17.2). This response was described as a **conditioned response**. Pavlov called the salivation at the sound of the bell a **conditioned reflex**.

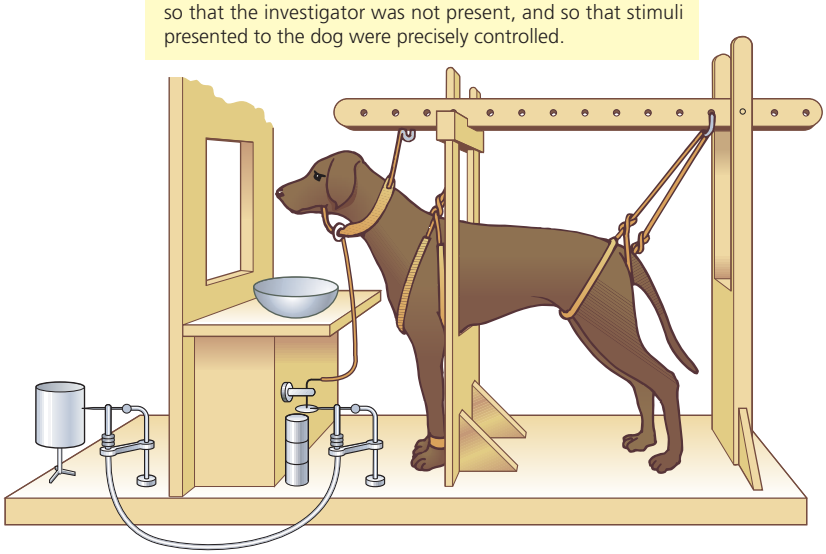


Figure 17.16 Pavlov's 1902 experiment on conditioned reflexes

In these experiments Pavlov devised this complex arrangement so that the investigator was not present, and so that stimuli presented to the dog were precisely controlled.

Experiments on animals

The idea of carrying out an experiment that is expected to cause pain to an animal is unacceptable to most people. This revulsion has led to the introduction of laws that attempt to eliminate any unnecessary suffering. Even so, some people believe that no experimentation on animals can ever be justified.

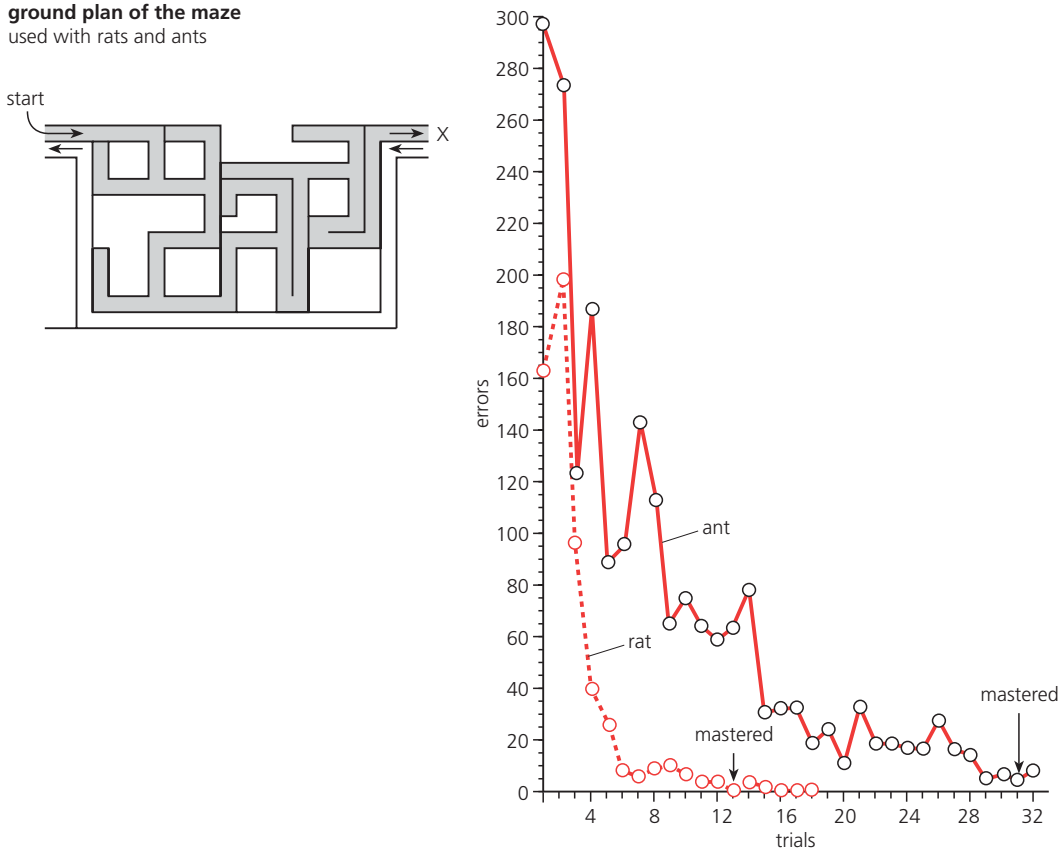
Pavlov's experimental procedure would not be acceptable to most biologists today, but it is described here in acknowledgement that much of our understanding of human physiology has been derived from experiments on animals.

Are there any circumstances in which you would accept a need to experiment with animals in a way that would cause them pain?

Table 17.2 Stimuli and responses in Pavlov's experiment

	Stimuli	Responses
Unconditioned	sight or smell of food	salivation
Conditioned	bell sound	salivation without food stimulus

Figure 17.17 Trial-and-error learning investigated in a maze



When **trial-and-error learning** is investigated, the observations are typically made in the laboratory, using a maze. A maze is a series of pathways with one or more points where the animal has to choose which way to go. A wrong choice leads to a blind end and no reward. The simplest type of maze has a Y or T pattern. A maze is mastered when an animal can consistently pass through without making a wrong turn. Animals with a well-developed nervous system can quickly learn a quite complex maze. For example, ants are found to learn a complex maze very quickly, although not as quickly as rats (Figure 17.17).

Maze-learning is based on the exploratory behaviour typical of many animals. This is usually exhibited at a high level in a new situation, but if this initial level of interest is to be maintained, then success must be rewarded (for example, with food for a hungry animal).

In **insight learning**, also known as **reasoning**, problems are solved without use of trial and error (Figure 17.18). We describe such activities as intelligent behaviour – the most sophisticated form of learning. In this, humans exploit currently received sense data together with experiences held in memory. We abstract general principles (which we call concepts) from concrete experiences, and from other situations we have learned about without necessarily having had experience of them. We may also use previous trial-and-error learning (some of which we may call play). We can analyse a new situation, and our perception of that situation is the basis of our response to it. All these mental facilities are based on the efficient working of the brain. Clearly, intelligent behaviour is likely to markedly improve chances of survival.

9 List the differences between trial-and-error and insight learning.

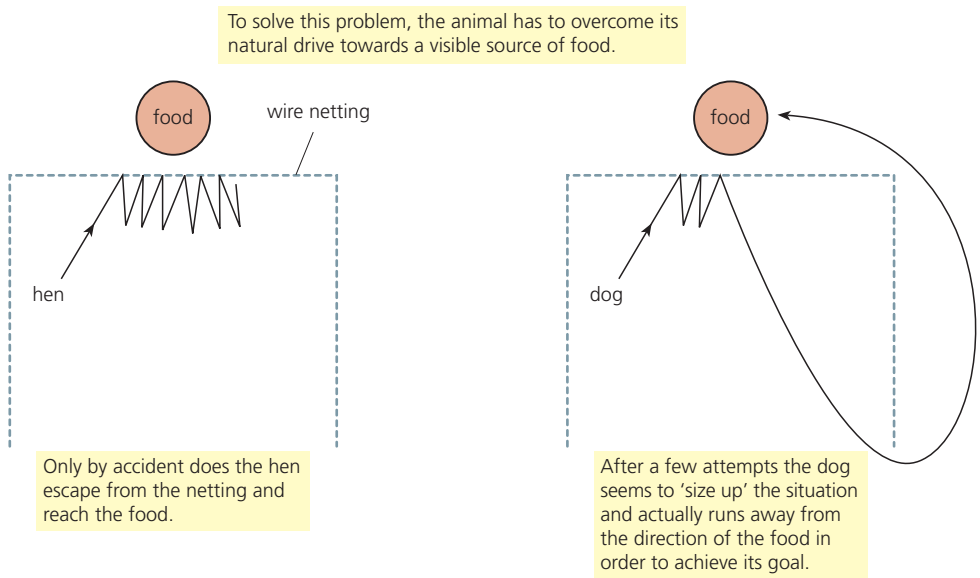


Figure 17.18 Insight learning

The roles of inheritance and learning in the development of birdsong

The environment is rarely silent, and one of the many sources of sound are those that are deliberately made by animals, largely or solely as a form of communication.

While these sounds may give valuable information to members of the same species, they may also unhelpfully alert nearby predators. Consequently, sounds produced tend to be highly specific, and produced in controlled circumstances.

Birds are one of many groups of animals that communicate with sound. In their case their voice box (site of the vocal cords) is where the trachea splits into two bronchi, just outside the lungs. Birds sing with their mouths open during expiration of air, and are able to vary their song by altering the way air is expired, altering the position of head and neck (which alters the shape of the trachea), and by opening and closing the mouth. The result is the possibility of a rich range of notes.

How do birds acquire their specific song?

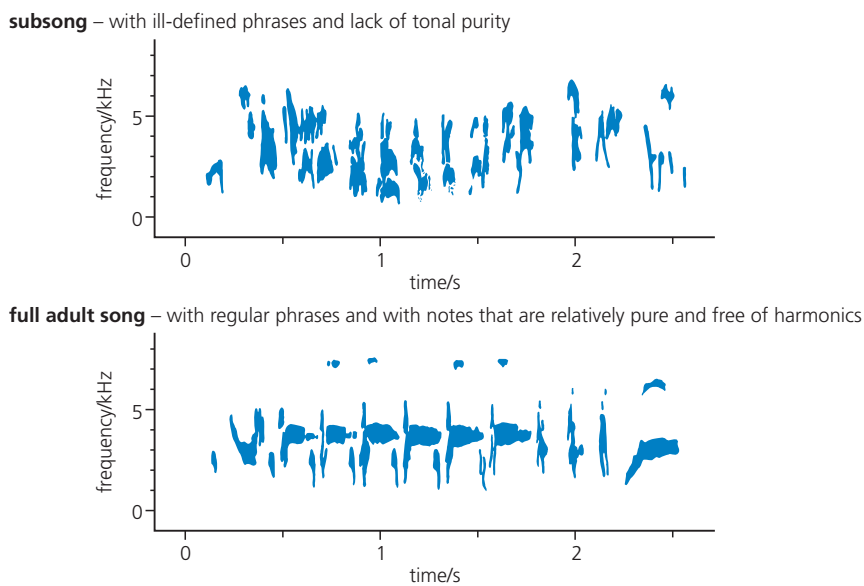


Figure 17.19 Sonogram of the song of the canary (*Serinus canaria*)

In many birds, learning and experience are important. In only a very few bird species do young birds *never* hear the song of their parents as they are reared and become fledglings. One such is the cuckoo, which is reared by foster parents after the female cuckoo has planted her egg in place of an egg of a host bird (a bird of an entirely different species). The host incubates this, and later feeds and rears the resulting cuckoo chick, in place of its own young. We must assume the cuckoo's song has a totally instinctive origin – it is inherited.

There are some other bird species in which the song is also instinctively acquired. We know this because when these birds have been (abnormally) reared in total isolation, their song as eventual adults is exactly that of others reared normally.

But there is good evidence that in many bird species only a basic song 'template' is inherited. When these birds are very young, their song (described as subsong) is quieter than that of adults of the same species, and is described as unrefined. It has long phrases, for example.

As these young birds are reared by their vocal parents, their song is slowly changed with practice. These birds learn to sing the song their parents perform (Figure 17.19).

Neurotransmitters and synapses

E4.1–4.6

We have seen that action potentials are transmitted from one neurone to another in the nervous system at a structure called a **synapse** (Figure 7.35, page 215). At the majority of synapses, **transmission is chemical**, and it is with this type of synapse that we are concerned here.

In the introduction to the working chemical synapse, an **excitatory synapse** was described. That is, the incoming action potential excited the post-synaptic membrane and generated an action potential that was then transmitted along the post-synaptic neurone (Figure 7.36, page 216).

Now, some synapses have the opposite effect, and these are known as **inhibitory synapses**. Here, release of the transmitter into the synaptic cleft triggers the opening of ion channels in the post-synaptic membrane through which chloride ions enter, or channels through which potassium ions leave. In either case, the interior of the post-synaptic neurone becomes more negative (we say it is **hyperpolarised**). This makes it more difficult for the post-synaptic cell to generate a nerve impulse in response to excitation by other synapses (excitatory synapses) present on the same post-synaptic cell.

In general, at a synapse, an action potential will only be generated in the post-synaptic neurone if the combined effects of the excitatory action potentials and inhibitory action potentials exceed the threshold level. This additive effect of post-synaptic potentials is known as **summation**. Several impulses may arrive at the synapse in quick succession from a single axon and cause an action potential in the post-synaptic neurone (**temporal** summation). Alternatively, impulses from several different axons may contribute to the total (**spatial** summation). Summation contributes to decision-making processes of the brain (Figure 17.20).

Decision making in the central nervous system

We have seen that our nervous system consists of a **central nervous system** (CNS – the brain and spinal cord) and the **peripheral nerves** that carry action potentials from sense organs to the CNS, and action potentials from the CNS to effector organs (muscles and glands) (Figure 17.25, page 538).

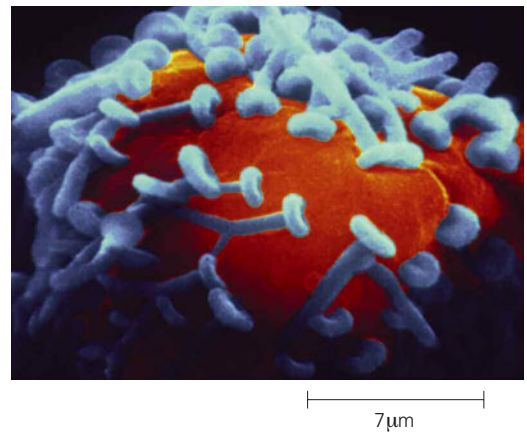
The CNS consists of vast numbers of interconnected neurones; in fact, our brain typically contains at least 10^{11} neurones (that is 100 000 000 000) linked by 10^{14} synapses.

The fundamental activity of the brain is to **co-ordinate and control** all body functions (except those under control of simple reflexes). The brain also stores information and builds up the **memory bank** of past experiences. The brain may **initiate activity**; it is the brain that enables us to imagine, to create, to plan, to calculate and to predict. Various forms of **abstract reasoning** are conducted here, too.

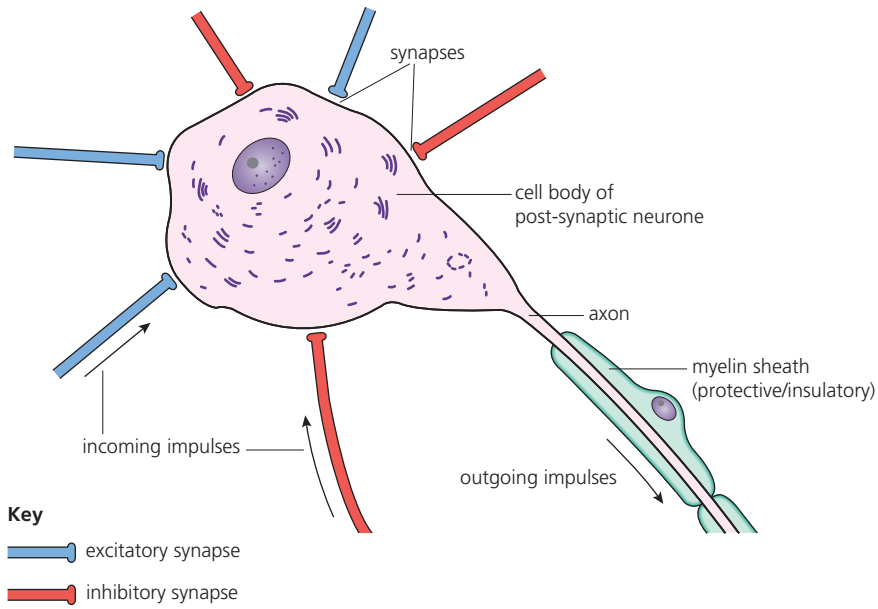
How does the brain achieve all these outcomes?

Figure 17.20 Integration of multiple synaptic inputs

SEM of synaptic junctions with the cell body of a post-synaptic neurone



excitory and inhibitory synapses



Brain function remains an area of study with a large amount of uncertainty, but one at which much study is directed. Progress is slow, although many brains work on the problem.

However, we do know that different areas of the brain carry out particular functions (page 535) – we can think of there being centres where aspects of body control and behaviour are processed. Such special centres are connected by nerve fibres via the spinal cord and peripheral nerves to relevant areas of the body.

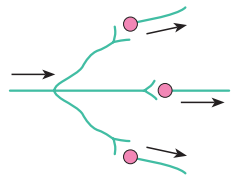
In addition to impulses being received and generated here, there is an on-going process which we can call **integration**. In effect, decisions are being taken in these centres, based on integration of incoming data with memory data, at least in many cases.

Further, two other aspects – one functional and one structural – are central to the decision-making processes of the brain, namely:

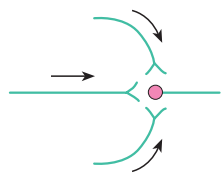
- interaction between the activities of excitatory and inhibitory pre-synaptic neurones at the synapses, operating within unimaginably numerous connections that occur between the vast numbers of neurones present;
- different types of connection pathways are found between neurones: some of these relay impulses to other pathways, some trigger a single response from many inputs, some ensure persistence of an original input, and some produce a strong, precise result (Figure 17.21).

Figure 17.21 Neural connections observed between brain cells

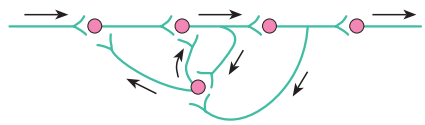
Each neurone has many terminal synaptic knobs. The different types of connections these make provide the structural basis of the decision-making facility of the brain.



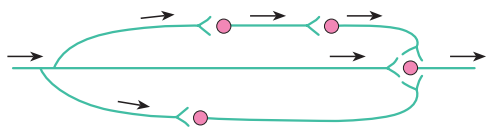
Divergent connections – in which information from one pathway is relayed to several others. This forms the basis of a variety of responses.



Convergent connections – in which information from several pathways comes together and is focused on a single (or fewer) pathway(s). This results in strong excitation or inhibition, or triggers a single response as a consequence of different stimuli.



Circular/reverbatory connections – in which the signal returns to its source. This results in reinforcement or ensures the signal persists for some time.



Parallel, after-discharge connections – in which the post-synaptic neurone sends out a stream of impulses without feedback. The result may be a precise, strong response.

Incidentally, monitoring of brain activity using functional magnetic resonance imaging (fMRI, page 536) on working, healthy brains has enabled researchers to locate decision-making activities in particular areas of the brain associated with particular skills and body functions.

10 **Suggest** what advantages may result from the convergence of axon endings from several neurones on one post-synaptic neurone.

The brain and psychoactive drugs

A psychoactive drug is one that affects the mind. These drugs have their effects on synapses in the brain, the performance of which may be profoundly altered.

How do these drugs interfere with the activity of synapses?

Some drugs **amplify** the processes of synaptic transmission; in effect, they increase post-synaptic transmission. Nicotine and atropine have these effects.

Other drugs **inhibit** the processes of synaptic transmission, in effect decreasing synaptic transmission. Amphetamines and β blocker drugs have these effects (Figure 17.22).

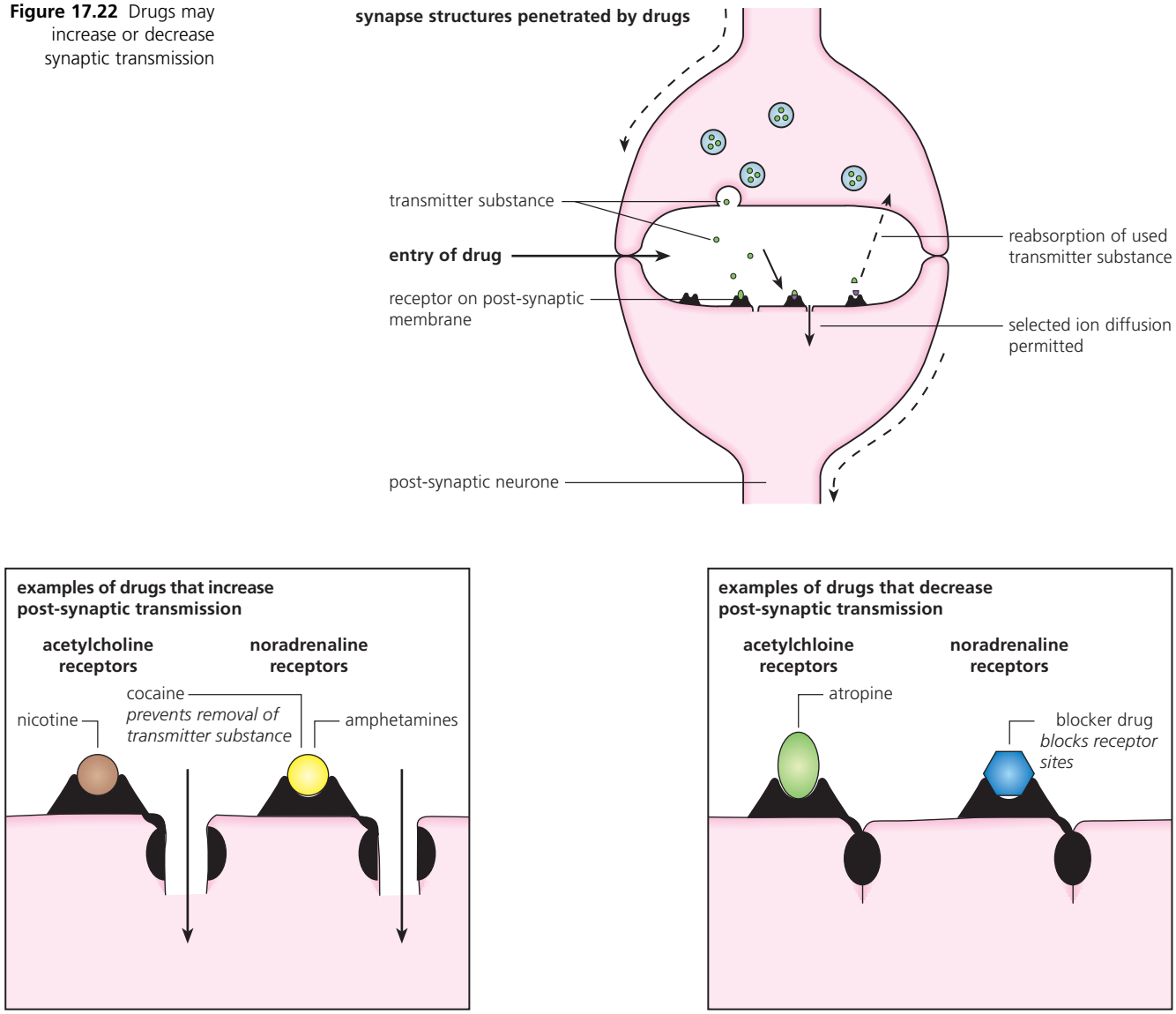


Figure 17.22 Drugs may increase or decrease synaptic transmission

Why may psychoactive drugs affect personality?

We have noted the significance of synaptic transmission in brain function. Also, our brain has a central role in the co-ordination of body functions (except those under control of simple reflexes), in the storage of information, in maintenance of our memory bank, in enabling us to imagine, create, plan, calculate and predict, and is the seat of our abstract reasoning powers. For all of these reasons, the use of psychoactive drugs may have a profound effect on behaviour, and possibly on personality.

The **wide use of drugs** of this type has involved their applications to medicine, to horticulture (insecticides) and as nerve gases (weapons). For example, substances known as **painkillers** (analgesics) and **tranquillisers** (such as librium, valium and mogadon) activate inhibitor synapses in the brain, making the brain neurones resistant to excitation. The effects of the insecticide **malathion** and the nerve gas **Sarin** are to inhibit enzymic breakdown of transmitter substances after they have attached to their receptors on the post-synaptic membrane. This effect may disrupt the entire nervous system and lead to death at relatively low doses of the chemical.

The **use of psychoactive drugs in a social, recreational context** is popular with many people in diverse societies all over the world. However, these uses often have dangerous or tragic consequences. Examples of such drugs are listed in Table 17.3.

Excitatory drugs	Inhibitory drugs
Nicotine – similar in structure to synapse transmitter substance acetylcholine and fits acetylcholine receptor on post-synaptic membranes, but is not broken down by the enzyme that inactivates acetylcholine. It remains on the receptors, prolonging the effects.	**Benzodiazepines** – enhance effects of the natural brain synapse transmitter known as **GABA** by combining with and slowing down GABA receptors in the post-synaptic membrane. GABA slows brain action, and benzodiazepines exert an extra (often excessive) inhibitory effect.
Amphetamines – cause increased release of synapse transmitter substance noradrenaline leading to enhanced activation of post-synaptic neurones in the brain. A high state of mental arousal may result.	**Ethanol** (alcohol) – while initially a stimulant, as the level of blood alcohol rises it has a sedative effect by slowing activity of neurones. In effect, it comes to act on the brain as an anaesthetic. (It also is harmful to liver, pancreas and heart muscle function.)
Cocaine – prevents removal of noradrenaline from receptors on the post-synaptic membrane, resulting in the continuous transmission of action potentials in the post-synaptic neurone.	**THC in cannabis** – this drug crosses the blood–brain barrier and binds to particular receptors (cannabinoid receptors) that occur in selected structures, including in the cerebellum (responsible for initiating movements) and particular parts of the cerebral cortex (reasoning and perception). Cannabis slows or prevents movement, and also impairs reasoning powers.

Table 17.3 Examples of excitatory and inhibitory psychoactive drugs

11 **Identify** the dangers for general health of the effects of alcohol abuse, including effects on the liver (use as your sources recent articles in, for example, *Biological Science Review*, *New Scientist*, *Scientific American*, or an appropriate website). Acknowledge your sources.

Examining the effects of cannabis and cocaine

Marijuana or cannabis

This is reported to be the most commonly used illicit drug. Also referred to as pot, grass or weed, it is obtained from the hemp plant (*Cannabis sativa*).

In cannabis, the chief mind-altering chemical is delta-9-tetrahydrocannabinol (**THC**). Cannabis products are usually smoked. The THC content is variable, typically 1–4%, but rising to 7.5% in some preparations (and as high as 24% in hashish, a resin form).

The **effects of cannabis on mood** are as a mild hallucinogen. It has some of alcohol's disinhibiting properties. As well as inducing a sense of well-being and a dreamy state of relaxation, it encourages fantasies. Users typically become highly suggestible and therefore vulnerable, especially if in a position to drive a car or ride a motorbike or bicycle. These effects are generally felt within a few minutes and peak in 10–30 minutes. Effects wear off in 2–3 hours. Stronger doses may induce disturbing, more persistent mood changes, including a state of paranoia.

The **effects of cannabis on behaviour** include the limiting of the capacity to absorb and retain information. The danger from this is greatest for young teenagers, in whom failures in learning – of social skills as well as school work – through persistent use, make it harder for them to reach a secure, healthy maturity. Inappropriate life-style choices are a common outcome for a proportion of users. For example, cannabis is a 'gateway' drug; use of cannabis makes subsequent resort to more disabling drugs (including cocaine) more likely. Before this stage is reached, however, adolescent users of cannabis are at greater risk of being a victim of car accidents, or of unwanted pregnancies, or of acquiring sexually transmitted disease.

Cocaine

This is a strongly addictive, stimulant drug. It is an alkaloid found in the leaves of the coca shrub (*Erythroxylon coca*), a native of South America. A coca leaf typically contains 0.1–0.9% cocaine. If chewed in this form it rarely generates problems for the user, social or medical. When the leaves are soaked and mashed, however, a coca paste can be extracted, and this is purified to 60–80% cocaine hydrochloride, by extraction with an organic solvent.

This was the predominantly used recreational form, taken by ingestion or snorting or intravenous injection (for extra 'kicks'), until crack cocaine was devised. For crack, cocaine paste is heated in a solution of baking powder until the water has evaporated. The resulting drug

vaporises at low temperature, and so is inhaled from a heated pipe (making a cracking sound on heating); this leads to much faster absorption.

Cocaine has its effect on brain neurones containing the neurotransmitter dopamine. Dopamine is associated with the body's pleasure responses, particularly an inclination to feel euphoric. Cocaine is a strong nervous system stimulant that works by interfering with the reabsorption of dopamine. It triggers prolonged stimulation of the neurones by dopamine, which is the cause of the extended pleasure response perceived by the user.

The **effects of cocaine on mood** are dramatic. Cocaine induces euphoric effects, which include hyperstimulation, reduced fatigue, and mental clarity. Cocaine users experience an intensity of pleasure outside the normal range of human experience, talking of a most wonderful experience of consciousness, particularly if using crack.

In fact, the faster the drug is absorbed into the body the more intense are the sensations (and the quicker addiction follows).

However, the initial short-lived euphoria is followed by a crash experience, and some users experience restlessness, irritability and anxiety, particularly with prolonged use. Some respond by increasing the dose to prolong the effects. Use of cocaine in a 'binge' during which the drug is taken repeatedly may result in paranoid psychosis in which contact with reality is lost. Auditory hallucinations may be experienced.

The **effects of cocaine on behaviour** arise from the addiction that quickly develops from regular use and indulgence in increasing dose size. Once an intense craving for more cocaine develops, stereotyped, compulsive and repetitive patterns of behaviour develop. For example, sensations of insects crawling under the skin lead to severe depression, agitated delirium, and paranoid psychosis. (A **psychosis** is a severe mental derangement, such as a loss of contact with external reality.)

The social consequence of cocaine addiction is destructive. Family and friends are alienated as the hooked user becomes isolated and suspicious. Resources of money and time are invested in obtaining yet more cocaine. To obtain crack, addicts will often lie, cheat, steal, and commit crimes of violence. Once-loved families (partners and children) are put aside.

The causes of addiction

Addiction is a state of taking a mood-altering drug habitually and of being unable to give it up without experiencing very unpleasant side-effects. An addict is someone who (apparently) cannot control or abandon their drug use.

We have seen that a major cause of addiction is the mere experimentation with a (possibly) less harmful gateway drug like THC in cannabis, in an environment that encourages further experimentation – moving on to involvement with some form of cocaine, for example. The interference with **dopamine metabolism** that this particular drug triggers produces a state of dependence. More and more drug is needed to produce the same effect, and the user comes to feel that the effect cannot be lived without. Thus, a major cause of addiction is the habit-forming nature of the drugs themselves. However, there are other factors.

Genetic predisposition may be a factor in some addicts, but we need to differentiate this from the environmental aspects of being exposed (particularly as a young person) to a culture of drug use in the immediate family environment. Undoubtedly, some people have a metabolic state in which drugs (like alcohol) are very effective within their body (they lack the level of enzymes to more promptly dispose of the drug, for example), and this condition may be inherited.

Another possible genetically controlled factor may be a personality type which is either especially hedonistic in inclinations, or one that is inclined towards unnecessary risk-taking, in general. However, these tendencies too may be due, to a greater or lesser extent, to acquired family values, perhaps?

Social factors such as poor diet, high unemployment, and limited access to education and training that will lead to rewarding employment, together with little opportunity for personal fulfilment of all types, can all generate a sense of hopelessness in which drug use seems a possible, practical escape.

12 **Discuss** social factors that relate to the uptake and use of recreational drugs by young people in your community, keeping in mind international issues – for example, concerning origin and supply, if relevant.

The human brain

E5.1–5.7

The vertebrate brain develops in the embryo from the anterior end of a simple tube, the **neural tube**. This tube enlarges during embryological development to form three primary structures, known as the **forebrain**, **midbrain** and **hindbrain**. The various parts of the mature brain develop from these by selective thickening and folding processes of their walls and the roof (Figure 17.23).

These enlargement processes are most pronounced in mammals, and a striking feature is the enormous development of the **cerebral hemispheres**, an outgrowth of the forebrain. The human brain contains about 10^{11}–10^{12} relay neurones and the same number again of supporting cells (known as neuroglia cells). The majority of these neurones occur in the cerebral hemispheres, where each relay neurone is in synaptic contact with about a thousand other neurones. Mammals are the most intelligent of all animals, and their memory, complex behaviour and subtle body controls are linked to this development.

White and grey matter in the brain

When tissue of the CNS is examined, the parts where cell bodies are grouped together appear grey, and are known as **grey matter**. Other areas of the CNS, where nerve fibres occur, each wrapped around by a lipid sheath (known as myelinated nerve fibres), appear as **white matter**.

White and grey matter is present in the brain, as in the spinal cord (Figure 17.3, page 507). Grey matter makes up the interior of the brain, white the exterior. However, in the cerebral hemisphere and cerebellum (see below) there are additional layers of grey matter (that is, additional neurones).

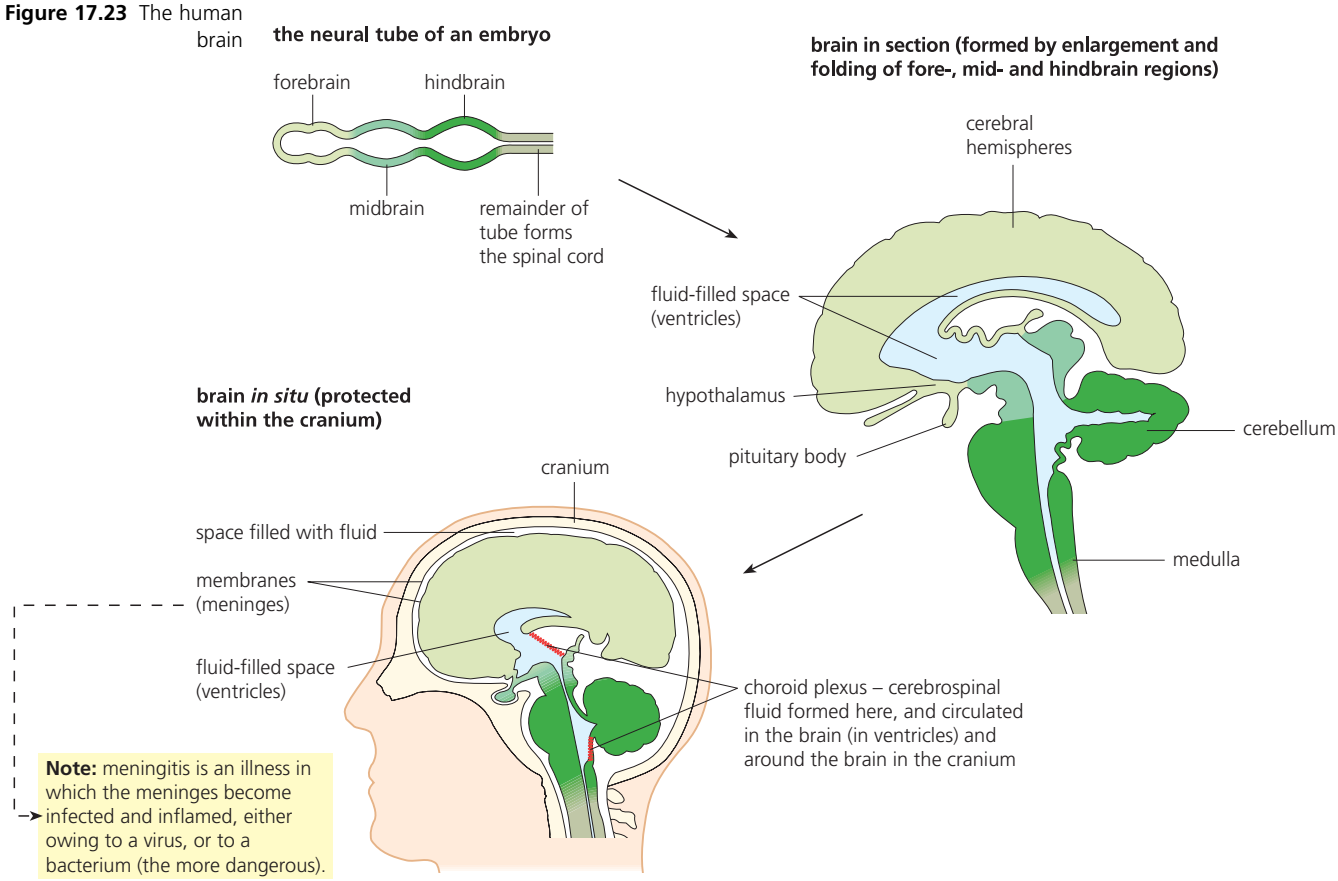


Figure 17.23 The human brain

Extension: Blood–brain barrier

Blood capillaries are present throughout the nervous tissue. However, in the brain the capillary walls form a barrier to many of the dissolved substances in the blood. Normally, only the essential substances, oxygen and glucose, cross the blood–brain barrier; other substances dissolved in the plasma that might affect activity of the synapses are largely excluded. Some amino acids present in the blood are neurotransmitters in brain synapses. This barrier is important for maintenance of normal brain function.

Brain function

The human brain controls body functions (apart from those functions under the control of simple spinal reflexes) by:

- receiving impulses from sensory receptors;
- integrating and correlating incoming information in association centres;
- sending impulses to effector organs (muscles and glands) causing bodily responses;
- storing information and building up an accessible memory bank;
- initiating impulses from its own self-contained activities – the brain is also the seat of personality and emotions, and enables us to imagine, create, plan, calculate, predict, and abstractly reason.

We have previously speculated on decision-making processes in the brain (page 528).

Within the brain, tasks and roles are localised.

The **hypothalamus** is part of the floor of the forebrain, and is the main controlling link between nervous and endocrine (hormone) systems. For example, it is the control centre for the **autonomic nervous system**, which consists of neurones that convey impulses to smooth muscle, cardiac muscle and glands – all systems not under our conscious control.

The hypothalamus, which is exceptionally well supplied with blood vessels, is also concerned with the maintenance of a constant internal environment (**homeostasis**). For example, it is here that body temperature, and the levels of sugars, amino acids, and ions in the blood, are monitored and controlled. Feeding and drinking reflexes, and aggressive and reproductive behaviour, are also controlled here.

Also in the hypothalamus, hormone levels in the blood are monitored. The hypothalamus is attached to the **pituitary gland** (the master hormone-producing gland) and the two work together in controlling the release of most hormones. From the anterior lobe of the pituitary gland are released many hormones that regulate different body functions, and from the posterior lobe of the pituitary are released two hormones that are actually produced by the hypothalamus.

The **cerebral hemispheres**, an extension of the forebrain, form the bulk of the human brain. They consist of densely packed nerve cells and myelinated nerve fibres. The ratio of the size of the cerebral hemispheres to the size of the whole nervous system is larger in humans than in any other mammal.

In the cerebral hemispheres are co-ordinated very many of the body's **voluntary activities** (conscious activities), together with many involuntary (unconscious) ones. It is an **integrating centre** of memory, learning, emotions and other complex functions.

The hemispheres have a **vastly extended surface**, achieved by folding with deep grooves. The surface, the **cerebral cortex**, is covered by grey matter (nerve cell bodies) to a depth of 3 mm. This cerebral cortex is **densely packed with non-myelinated neurones**. These have a mass of dendrites and an almost unimaginable number of synaptic connections.

The areas of the cortex with special sensory and motor functions have been mapped out. The **sensory areas** receive impulses from receptors all over the body. The **motor areas** are where motor impulses, serving the whole body via the spinal cord, originate.

The **cerebellum**, part of the hindbrain, also has an external surface layer of grey matter. It is here that involuntary muscle movements of posture and balance are controlled. It is also where all precise, voluntary manipulations involved in hand movements, speech and writing are co-ordinated (rather than initiated).

13 **Explain** why death is immediate if brain and spinal cord are severed below the medulla.

The **medulla oblongata**, the base of the hindbrain, houses the regulatory centres concerned with heart rate, ventilation of the lungs, blood pressure, swallowing, digesting and vomiting. In the medulla, the ascending and descending pathways of nerve fibres connecting the spinal column and brain cross over. As a consequence, the left side of our body is controlled by the right side of the brain, and vice versa.

How brain function is investigated

The earliest known reference to the human brain is on an Egyptian papyrus, dated to approximately 3800 years ago. It records the treatment of a sword wound to the head. In the long story of the discovery of the structure and functions of the human brain, detective work on people who have survived after a severe but local head wound has played a part.

By **investigating brain-damaged patients** it has sometimes been possible to correlate precisely, specific areas of the brain with the performance of particular functions. The site of control of various facilities has been located. For example, the initial discovery of the role in sight of the visual cortex (on the rear of the cerebral hemispheres – Figure 17.9, page 515) came from the experiences of soldiers surviving bullet wounds in the rear of the skull.

However, external wounds that penetrate to the brain are often quickly fatal. On the other hand, **cerebrovascular accidents** (strokes – page 419) are relatively common events, particularly in older people. Many strokes are not immediately fatal. The outcome may be paralysis in a part of the body, or loss of a specific sensation or facility. In subsequent post-mortem investigation it has been possible to discover the particular area of the brain involved.

Another source of information on brain function has been **animal experiments**. These have involved mammals and other vertebrates, and the removal of parts of a healthy brain or the severing of connections within the brain. The resulting altered behaviour has provided insights into the roles of parts of the brain.

For example, the severing of the fibres that cross over in the centre of the brain below the two halves of the cerebral hemispheres gave clues to the interaction of left and right halves of the brain. Cats were used for that investigation.

Many of these approaches, using animals, would be unacceptable today (and unnecessary – see below); as a society we are reluctant to sanction non-essential surgery that may cause pain.

By **magnetic resonance imaging (MRI)** the precise parts of a living, healthy, functioning brain that are activated when a particular body activity occurs can be accurately mapped. In this technique, nothing has to be injected into the body – it is entirely non-invasive. Furthermore, it can detect activity anywhere in the brain, and with high resolution. Results of a scan are quickly to hand.

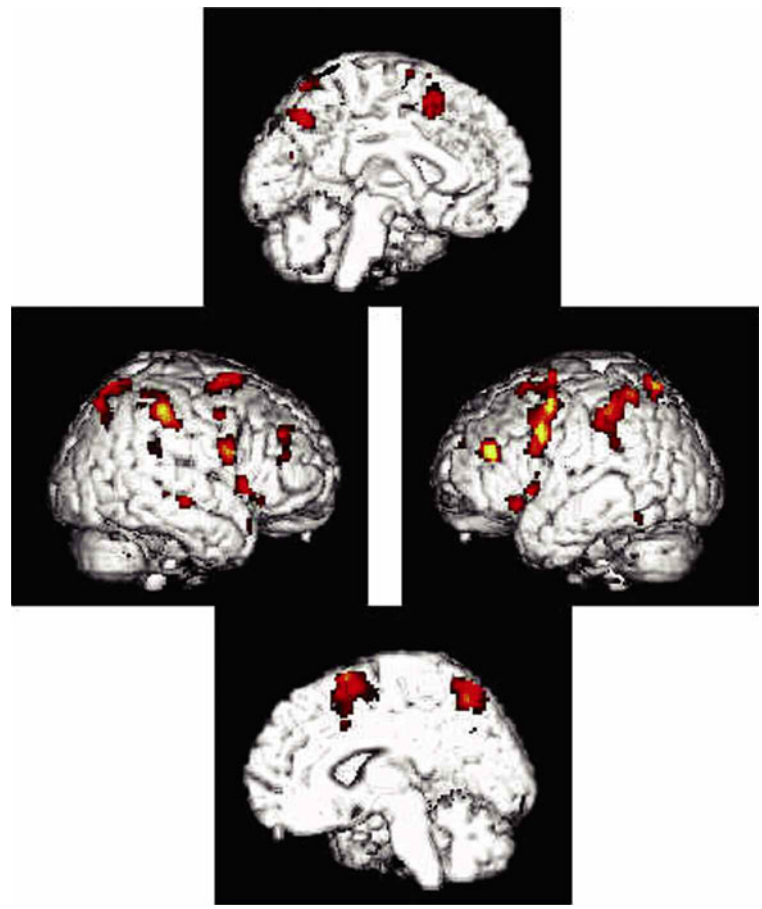
MRI uses a strong magnet to produce detailed images. It works by measuring the way the vast number of hydrogen atoms present in the body absorb and then emit electromagnetic energy. The nucleus of each hydrogen atom is, in effect, a tiny magnet; in a strong magnetic field, these line up – as compass needles do in a magnetic field. In the process of a scan, a pulse of radio waves is then applied, sufficient to cause hydrogen nuclei to change orientation. When the pulse is switched off, the nuclei revert to their original orientation and each nucleus gives off energy (at radio frequencies). From this signal, the scanner can work out the location of each nucleus.

Functional magnetic resonance imaging (fMRI) is an advanced form of MRI that detects the parts of the brain that are active when the body performs particular tasks. Brain cells require energy and a good supply of oxygen at all times but during periods of intense activity, the demand for these resources increases locally. The scanner can detect an increase in red blood cell oxygenation at the site of special neural activity; the technique for this is known as blood oxygen level dependent (BOLD) contrast. The increase in blood flow to the most active areas is disclosed because what is actually detected is the difference between signals arising from hydrogen nuclei in water molecules in the neighbourhoods of (a) oxyhaemoglobin and (b) deoxyhaemoglobin. When the concentration of oxyhaemoglobin increases, the fMRI signal rises.

Figure 17.24 An advanced fMRI scanner in use

subject undergoing a series of brain scans while occupied on particular tasks – the regions of the cerebral hemispheres that are momentarily the sites of special neural activity are observed and recorded via the computer screen

fMRI images of volunteer undertaking a specific thinking task – scans of the brain show the right and left hemispheres (side views) and the brain from the midline, looking at left hemisphere (upper) and right hemisphere (lower image)



In Figure 17.24 we see the scanner in use with a volunteer who is undertaking particular tasks. The data obtained from scans are transformed into three-dimensional images that record the regions of the brain which are most active. Here, the bright parts of the image are those that are particularly activated.

In this case, a volunteer is involved in a thinking task (generating random numbers between 1 and 9, at a given pace). It can be seen that particular parts of the cerebral hemispheres show increased activity of red blood cells, confirming the roles of these areas.

This new brain mapping technique permits the elucidation of the spatial organisation of human brain function down to a submillimetre level.

The autonomic nervous system

We have seen that the nervous system consists of the CNS (brain and spinal cord) and the peripheral nerves. The peripheral nervous system (PNS) is organised in a distinctive way, and this is what we examine next. The rather complex layout of the PNS is summarised in Figure 17.25.

There is a diversity of components of the PNS and we have met with parts of this system already (the neurones that carry impulses from sense organs to the CNS, and the somatic nervous system of neurones are shown at work in Figure 17.3).

Check this diagram again to note their particular pathways and roles.

The PNS also contains the **autonomic nervous system (ANS)** – autonomic means 'self-governing'. The ANS is special because it controls activities and structures inside the body that are mostly under unconscious (involuntary) control. In effect, the ANS regulates the working of the interior of the body more or less without our knowledge.

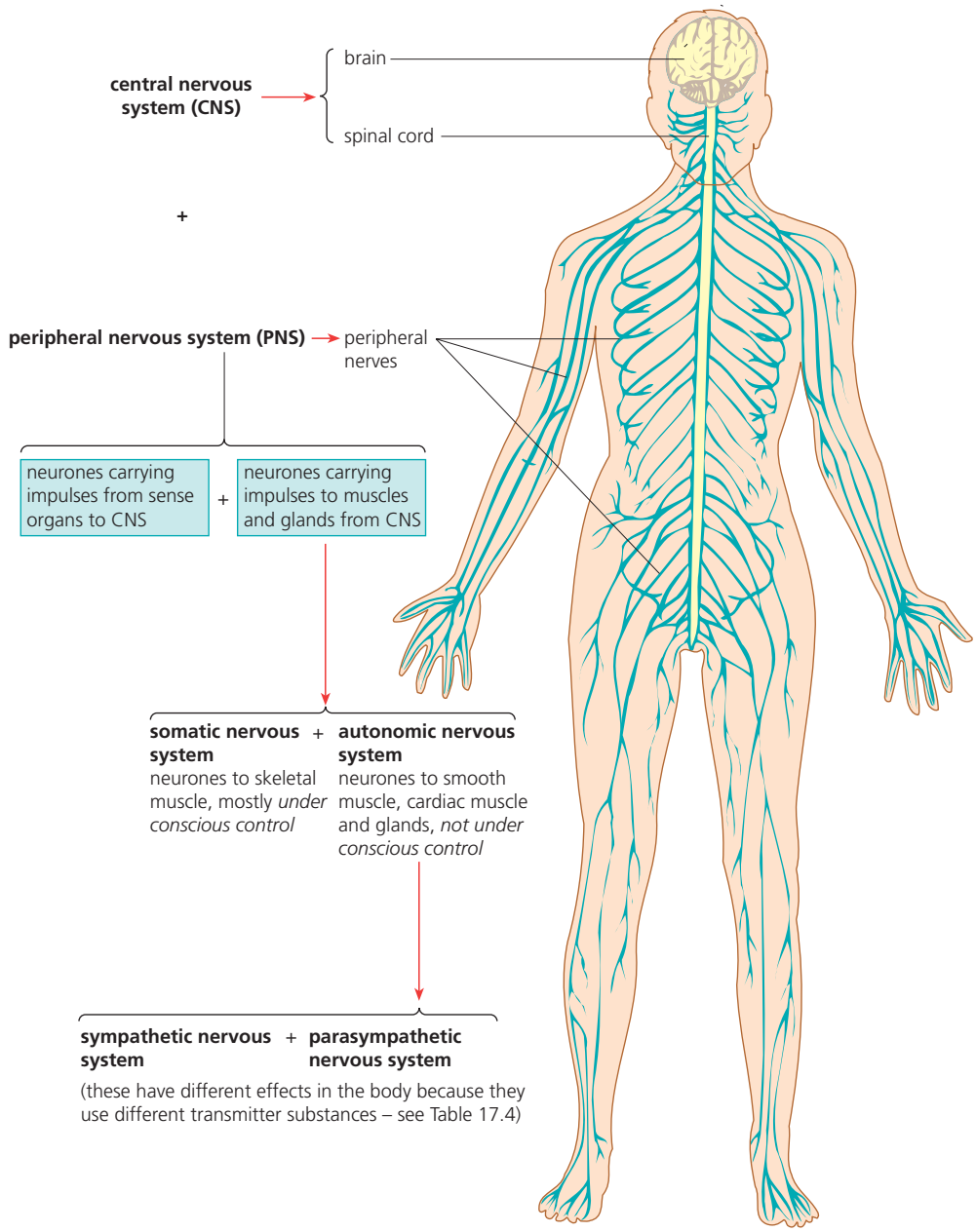


Figure 17.25 The organisation of the peripheral nervous system

In the ANS, nerve fibres emerge from the brain and spinal cord and pass to specific tissues, organs and glands. In fact, the fibres are those of motor neurones, and they either pass directly from the CNS or from bead-like structures called ganglia that are linked to the CNS. Ganglia contain synapses. ANS fibres serve the smooth muscle (involuntary muscle, not the striped muscles of the skeleton) of the internal organs, and the various glands.

The ANS is divided into two parts:

- the **sympathetic nervous system**;
- the **parasympathetic nervous system**.

Table 17.4 illustrates the key differences between the sympathetic nervous system and the parasympathetic nervous system. Note that:

- the ANS is coordinated by **unconscious regions of the brain**, namely the medulla and the hypothalamus (but some of its activities are regulated by conscious areas of the brain; for example, control over the sphincter muscles of the bladder and anus);
- many of the functions of the two systems are generally **antagonistic in their effects** (but in a few cases they may have the same effect).

Sympathetic system		Parasympathetic system
more active in times of stress to produce the 'flight or fight' responses	**when active**	in conservation of energy and the replacement of body reserves
at synapses with effectors, neurones release noradrenaline	**neurotransmitter**	at synapses with effectors, neurones release acetylcholine
Some of the responses of the two systems		
accelerates heart rate	**effect on heart**	slows down heart rate
causes widening (dilation) of the pupils	**effect on iris**	causes narrowing (constriction) of the pupils
constricts blood flow to arterioles	**effect on gut**	maintains normal blood flow to arterioles

Table 17.4 The working of the autonomic nervous system

14 **Identify** the control centre for the autonomic nervous system, and **suggest** three different sense organs from which sensory fibres will run to that centre.

The iris reflex

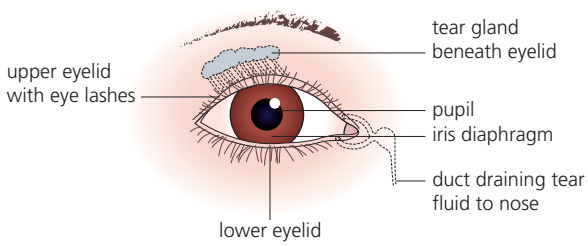
The working of the ANS may be illustrated by examining control of the iris diaphragm reflex.

Controlling the amount of light entering the eye is critical. If too little light reaches the retina, the cones may not be stimulated sufficiently for effective vision – so, a prey animal might not see an approaching predator. But with too much, the cells of the retina may be over-stimulated or even damaged – again, sight would be temporarily lost.

Fortunately, the muscles of the iris regulate pupil size precisely; circular muscle fibres are innervated by the parasympathetic nervous system, and when contracted, reduce the size of the pupil and, therefore, the amount of light reaching the retina. The radial muscle fibres are innervated by the sympathetic nervous system, and when contracted, increase pupil size (Figure 17.26).

Figure 17.26 The role of the iris diaphragm in controlling light entry

face view



iris diaphragm muscles

In bright light:
- circular muscles contracted
- radial muscles relaxed
- pupil diameter decreased.

In dim light:
- radial muscles contracted
- circular muscles relaxed
- pupil diameter increased.

Extension: Atropine and pupil size

The drug atropine (an alkaloid, obtained from the plant known as deadly nightshade) prevents the pupils from contracting – it blocks the acetylcholine receptors on the circular muscles of the iris at the neurone/muscle fibre junctions. This drug is sometimes applied to the eye in tiny quantities prior to surgery on the retina.

Regulation of the iris clearly demonstrates the autonomic system at work. The mechanism is a reflex action involving neurones of the optic nerve, relay neurones in the brain and spinal cord, and neurones of the parasympathetic nervous system and sympathetic nervous system. Although this response involves neurones in the brain, it is not subject to conscious control – it is a true reflex action (automatic response – Table 17.5).

In bright light	In dim light
■ many photoreceptor cells of the retina are stimulated so many impulses pass to brain via sensory neurones	■ few photoreceptor cells of the retina are stimulated so few impulses pass to brain via sensory neurones
■ via relay neurones, impulses pass along motor neurones of parasympathetic system to circular muscle fibres of the iris	■ via relay neurones, impulses pass along motor neurones of sympathetic system to radial muscle fibres of the iris
■ circular muscles of iris contract and radial muscles are passively stretched – pupils are constricted, and less light enters eye	■ radial muscles of iris contract and circular muscles are passively stretched – pupils are dilated, and more light enters eye
incidentally: depth of focus is increased	incidentally: depth of focus is decreased

Table 17.5 The iris diaphragm mechanism

The concept of brain death and the role of the pupil reflex

Death appears to be self-evident when, in a large organism such as a mammal, the organism's heart and breathing stop. In the case of a human, the medical term for this state is **cardiac death**.

However, modern resuscitation devices can maintain the functions of the heart and lungs for a considerable period (possibly as long as days) after the centres of the brain responsible for life-processes like these have stopped functioning, and to all intents and purposes, the patient appears dead. As a consequence, a further state of death is recognised, known as **brain death**. Brain death is defined as the irreversible cessation of all brain functions.

The importance of these definitions is self-evident since transplant surgery has become a possibility, requiring as it does, a supply of viable organs. All concerned require agreed medical criteria for brain death that has legal status.

The **agreed criteria for brain death** are that there is **absence of all brain functions**, resulting in **no purposeful movement or response to any stimulus**, including:

- pupils remain in the mid position, and do not react to light (pupil reflex is absent);
- the eyes do not blink when touched;
- the eyes do not rotate in their sockets when the head is moved;
- the eyes do not move when iced water is placed in the outer ear canal;
- there is no cough (or gagging) when a suction tube is placed well into the trachea;
- breathing does not commence when the patient is taken off the ventilator.

A point for further discussion

Could such criteria be misleading in the cases of patients in either a seriously hypothermic state, or unconscious due to the effects of certain drugs?

How pain is perceived and the role of endorphins in the brain

We have noted that sense organs and sensitive nerve endings are the **receptors of the nervous system**. In response to appropriate stimuli, receptors generate action potentials in a sensory nerve fibre, so the presence of the stimulus is registered by the nervous system as a whole, and may be responded to.

Receptors are typically sensitive to one type of stimulation only, such as differences in temperature, light, touch or chemicals. Some sense organs are complex organs (such as the eye, page 511), but others consist of an individual cell or small groups of cells, while some are **sensitive nerve endings**, such as pressure receptors and **pain receptors**.

Pain is indispensable for survival. Pain receptors occur in the skin and all other parts of the body tissues except the brain. When these are stimulated, action potentials pass to sensory areas in the cortex of the cerebral hemispheres. It is in this area of the brain where the feeling of pain arises. Our bodies experience two types of pain:

- **Fast pain** occurs rapidly after the stimulus is received, generating a feeling of acute pain, such as is experienced when the skin is cut. This experience is localised. This type of pain is not felt deep within the body.
- **Slow pain** is generated after a slight delay. It gradually increases in intensity but typically builds to be a chronic, burning or throbbing pain. It can be felt superficially or deep within the body, and it is often diffuse, appearing to come from a large area.

How does the brain respond to pain sensations?

It has been discovered that our bodies are able to produce opiate compounds known as **endorphins**. They bring about pain relief and a sense of well-being. Their discovery was made in 1975, and they are so named as a contraction of 'endogenous morphine', meaning 'morphine produced in the body'. **Endorphins block the transmission of impulses at synapses involved in pain perception.**

Endorphins may be released into the blood from the pituitary gland and into the spinal cord and into the brain by the hypothalamus (remember, endorphins in the blood cannot cross the blood–brain barrier). So, for example, it is found that in the bodies of patients enduring chronic pain, the level of endorphins may be high.

Prolonged, continuous exercise has also been found to induce a raised level of endorphins, giving us a welcome sense of euphoria (and good health) following vigorous exercise. Similarly, the good feeling one gets from an orgasm is partially attributed to the release of endorphins.

When pain sensations are out of all proportion to the level of tissue damage or persist for no reason, creating unreasonable discomfort, then the administration of pain relief, known as analgesia, is appropriate. Available drugs for pain relief range from aspirin to morphine and other opiate drugs. While the patient may still be aware of the pain, its acute intensity is removed by an effective analgesic.

Further studies of behaviour

E6.1–6.7

Social organisation

Very few animals are totally solitary; most meet together with others of the same species, if only during courtship and mating. However, some animals live in quite complex **social groups** on a permanent and structured basis. These groupings show varying degrees of differentiation in the roles of their members. Two examples are social insects such as the honey bee, and primates such as the baboon.

Honey bee (*Apis mellifera*)

A honey bee community consists of a reproductive female (the **queen**), many thousands of sterile females (**workers**), and a few hundred fertile males (**drones**), together with the brood (**eggs** and **larvae** the colony is rearing), living together in a hive. Altogether, the community consists of 20 000–80 000 individuals, of which the vast majority are the workers. These are the ones we see visiting flowers (Figure 17.27).

The queen lays about 1500 eggs per day, each in a wax cell prepared for her by the workers. Most eggs are fertilised. Most of the larvae, when they hatch out, are fed on pollen and nectar; they develop into **workers**. Workers survive for about six weeks, and they undertake a sequence of duties in the colony. First they are nurse bees tending to the growing larvae and building fresh

Figure 17.27 Honey bees

Left: Honey bees form cells in sheets, hanging from the roof of some cavity. Here, young are reared and honey is stored.

Right: Worker bee on flower (source of nutrients), collecting nectar via its tubular mouthparts, and storing the pollen that collects on its body in 'pollen baskets' on the hindlegs.

comb cells. Later they become outside workers, surveying for feeding sites, communicating about new food sites to other workers, guarding the hive entrance, and foraging for water and food.

A very few larvae are fed on royal jelly. This is a protein-rich food, a secretion of nurse bees. These larvae develop into **new queens**. A queen typically survives for 2–5 years, but she may mate only once, since the sperms stored in her abdomen following mating are sufficient for her life-time. After this, she establishes a new colony.

Occasionally eggs that are unfertilised are laid by the queen. These still develop into larvae (with a single set of chromosomes – that is, they are haploid). These larvae are fed on royal jelly, pollen and nectar and they develop into drones. Drones survive for about five weeks. They do not work in the hive, but when a new queen leaves the hive they accompany her, compete to mate with her, and then they die.

Common baboon (*Papio cynocephalus*)

The baboon is a ground-living primate found in Africa south of the Sahara. The principal habitats of baboons are woodland savannah, grassland, acacia scrub and sub-desert. Each night the troop sleeps in tall trees (or on a rock face), away from danger. The groups (of 10–200 animals) are permanent and cohesive. They are held together by a diversity of individual relationships, and intimate group knowledge of the territory they occupy. Groups are self-sufficient, their territories being large, overlapping home ranges. Different groups meet infrequently, tending to avoid each other. Territories are not defended. Social control, as far as it exists, is based on physical dominance. Communications are largely gestural, and concerned with immediate situations.

Most of a baboon's life is spent within a few feet of other baboons. Individuals are strongly emotional and highly motivated. Grooming (a significant health event in itself) develops and reinforces social bonds based on mutual dependence between members. Individuals afford themselves protection from predators via mutual alarm systems. These operate as they fan out across the savannah and search for food. Once an individual becomes isolated from the group,

Figure 17.28 Baboons

Left: Grooming assists both in health and in bonding of the group.

Right: The pack escapes to safety from an attack by ground-living predators, by taking to high trees.

the chances of death are high. They have an almost entirely vegetarian diet; there is no food sharing, and no division of labour. Males are more peripheral to the group, and may confront minor predators, but usually the whole group flees to safety (Figure 17.28).

Females are sexually mature at five years, and they become receptive for a few days of each cycle. The males (mature at ten years of age) compete for access to these females, who are promiscuous. Females spend most of their mature lives pregnant or lactating. Since their reproductive cycles are not seasonal, some females are available for mating all of the time. Mothers suckle their young for 6–8 months, and carry them until they are old enough to move independently.

How may natural selection operate at the community level?

Natural selection operates on individuals, or rather on their phenotypes. Phenotypes are the product of a particular combination of alleles, interacting with the effects of the environment of the organism. Consequently, natural selection causes changes to gene pools of populations. For example, individuals possessing a particular allele or combination of alleles, may be more likely to survive, breed, and pass on their alleles than other, less well-adapted individuals are. This process is also referred to as differential mortality. However, natural selection does not require the death of the less fit, but rather that their genes are not perpetuated. In this way, natural selection operates to change the composition of gene pools.

So the effect of natural selection on a community may be expected to be similar to its effect on a particular population. It is a process that brings about adaptation and evolution – but not quickly.

In the **baboon colony**, since there is no home base, all individuals must keep up with the troop. Individuals not inclined to remain within the troop (and any individual that becomes unwell or incapacitated and therefore falls behind), is rapidly lost to predators operating at the fringes. Social cohesion is reinforced by natural selection, despite the very limited social organisation and structure that the baboon troop exhibits.

Looking at a **honey bee colony**, the bulk of bees present are the worker bees, and they are sterile. Workers apparently act 'unselfishly', for the benefit of other bees only (altruistic tendency). Nevertheless, survival of the whole colony and of the reproducing members is dependent on the workers.

How can natural selection operate at the level of the colony in this unusual situation?

In the bee colony there are three phenotypes present (queen, drones and workers), but only two genotypes:

- drones are the community's males – they develop from unfertilised eggs, and so they are haploid;
- the queen and the workers develop from fertilised eggs, are diploid, but possess the same genome (the difference arises from the diet at the larval stage; diet determines whether it is the worker-making genes or the queen-making genes of the genome that are activated).

Normally, genetic affinity between parent and offspring is $\frac{1}{2}$, and between siblings it is also $\frac{1}{2}$. However, in the bee colony the picture is different. Here each worker receives exactly the same genetic material from the father because only one drone fertilises a queen (who then stores those sperms before use). The drone father was haploid and produced only one type of sperm (meiosis was not involved in the male bee's gametogenesis). The queen on the other hand, being diploid, produces gametes that are not identical (there is random assortment of homologous chromosomes during meiosis in egg formation – page 97).

Considering together all the genes received from queen and drone, the workers share on average half the genes from their mother but all of the genes they received from their father ($\frac{1}{2}$ from the drone and $\frac{1}{2} \times \frac{1}{2}$ from the queen). So the **workers have $\frac{3}{4}$ (0.75) of their genes in common**.

Workers are more closely related to each other than their queen is to *either* the drones *or* the new queens she begets. Sterile though they are, the workers have more genes in common with the next generation (0.75) than if they had been able to breed with a drone and reproduce themselves (0.5).

Arrangements in the bee colony distinctly favour survival of the genes of the workers.

The evolution of altruistic behaviour

An altruistic act decreases the altruist's chance of surviving, while increasing the survival chances of some other organism. The most familiar examples of this are acts by mothers who show near-suicidal courage in protecting a defenceless offspring against large, aggressive predators, if this can be described as truly altruistic.

How can altruistic behaviour have evolved if it reduces the reproductive success of the self-sacrificing individual?

We need to look at examples of altruism to answer that question.

In the honey bee colony, one stage in a worker bee's 'work experience profile' is the guarding of the hive against intruders. Any larger, aggressive organism attempting to steal the honey (a not uncommon event in the wild) is set upon by the guard bees, many of whom may sink their sting into the predator. The bee sting is a barbed structure, a little like a fish hook. It goes in easily, but it cannot be withdrawn, so the worker bee is forced to tear away part of her abdomen and leave the tissues behind, often including abdomen musculature. This continues to pump the cocktail of unpleasant chemicals of the sting into the aggressor, but the damaged worker crawls away to die.

Cases like that of the altruistic worker bee are described as **kin-directed altruism**. This is because the giver shares many genes with the beneficiaries, probably including the altruistic genes. This is certainly the case with worker honey bees, as we have already noted.

Most animal altruism is 'selfish' in benefiting the genes of the giver directly or indirectly. The great geneticist J. B. S. Haldane (1892–1964) responded to the question of whether he would lay down his life for his brother by saying he would do that for two brothers or four cousins! His apparently flippant reply pre-dated the idea of kin-directed altruism – a prescient observation.

In other cases, the giver is helping a wider community, including many less closely related members of the species. This may be seen as **reciprocal altruism** – all individuals in the group aid and support each other. They all benefit sooner or later from their neighbours' acts (almost a case of 'You scratch my back and I'll scratch yours' although not consciously thought out, of course). This is the case in vervet and baboon troops, in many instances.

Vervet monkeys of East Africa may feed on the ground, rather like baboons do. Vervets have a range of alarm calls – distinct vocalisations that warn of sudden impending dangers to the whole troop in earshot (Figure 17.29). Hearing the 'eagle alarm' results in all vervets looking up (if on the ground) or descending from the heights of trees, against a danger from the sky. On hearing

the 'leopard alarm', vervets leap up into trees, clear of a ground-approaching danger; on hearing the 'snakes alarm' they search around themselves in the grass (apparently, few of the snakes present a danger, but some are serious threats). Had the first observer that gave the warning cry taken cover immediately rather than first warning the troop, that monkey would have had a better chance of personal survival.

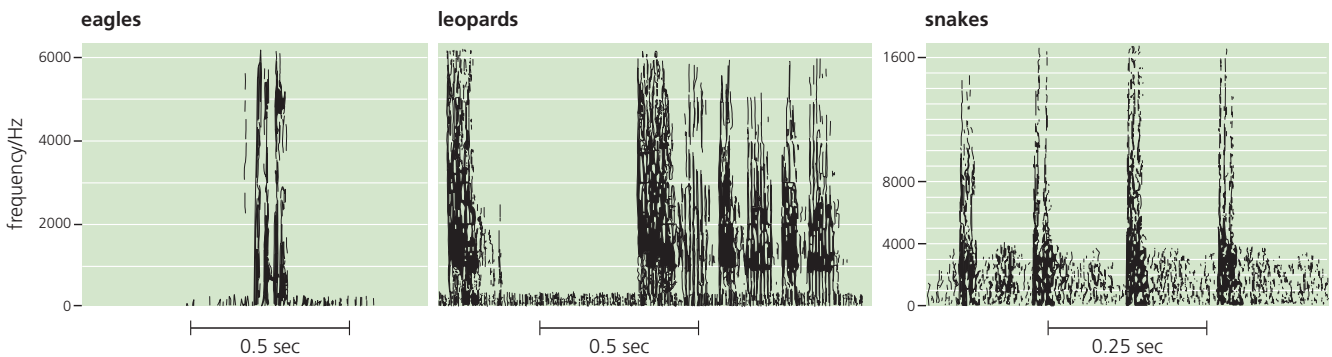


Figure 17.29 The alarm calls of vervets are specific to the type of predator – the sonograms show the calls elicited by eagles, leopards and snakes

Foraging behaviour and food intake

Animals need to find food and feed in order to survive. 'Foraging' refers to the processes of searching for, obtaining, and then consuming food. The need for food often results in curious patterns of behaviour.

It is useful to remind ourselves that foraging animals divide into herbivores (which consume plants), carnivores (which consume other animals), and omnivores (with a diet that includes both plants and animals). Within these major divisions we find specialist feeders which rely on one type of food or feed on a particular species, and generalist feeders with a more varied diet.

Whatever the food source, food is generally rarely distributed uniformly, and when located, different sources may be of different qualities. Consequently, foraging animals need to optimise the return on their investment of time and energy in obtaining food. Two examples of a response to this challenge are presented here.

Foraging by the honey bee

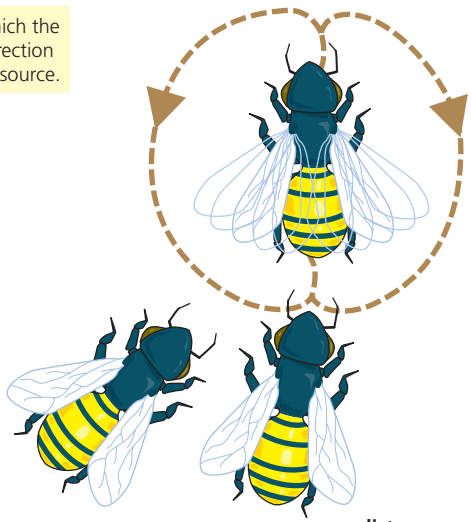
In a honey bee, the mouthparts are modified to form a long, sheathed sucking tube (Figure 16.14, page 471). The bee draws nectar from flowers back into the honey sac. From here it can be regurgitated for the benefit of the bee community, back in the hive. The other item of a bee's diet is pollen. This adheres to the hairy body, and can be collected by combing actions of the legs. Pollen is carried back to the hive as pellets attached to the pollen baskets of the hindlegs (Figure 17.27, page 542).

Foraging for nectar and pollen is the chief duty of worker bees at a later stage in their working life. At an earlier stage in duties outside the hive, an individual worker bee is responsible for surveying for feeding sites (populations of recently opened flowers, rich in nectar and pollen), and reporting back to the main body of workers in the hive.

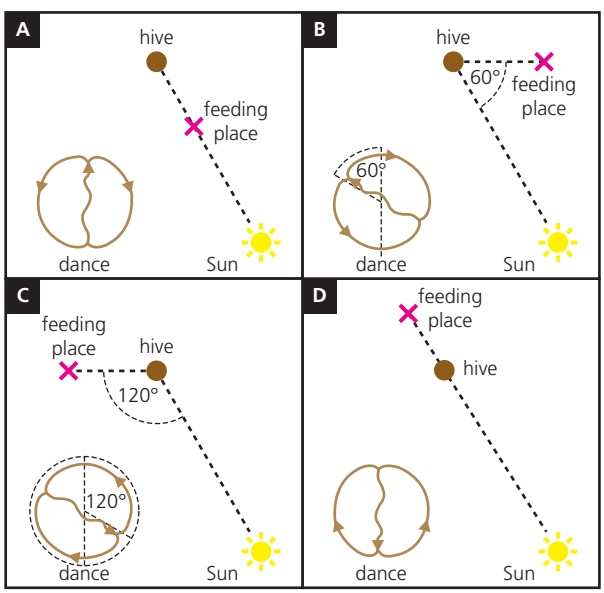
The **waggle dance of a worker honey bee** is the way it communicates the location of new food sources. The dance is performed on the vertical comb surface. It is a figure-of-eight dance, performed in darkness, surrounded by sister workers. The dancing bee emits buzzing sounds, and vibrates its wings and body (Figure 17.30). The speed of the dance (defined as the number of runs down the diameter of her circle per 15 seconds) is the way **distance** is communicated. The angle between the straight and the vertical (gravity) represents the angle between the Sun, the hive and the food source. Clearly, the waggle dance optimises food intake by the hive community.

Figure 17.30 The language of the honey bee

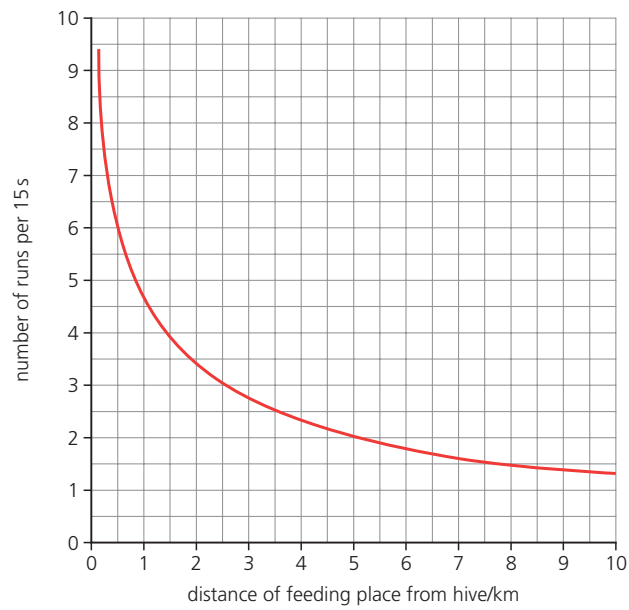
The **waggle dance** is the means by which the exploring worker bees communicate direction and distance of a favourable new food source.



direction communication
– the angle between the Sun, the hive and the food source



distance communication
– number of runs along the waggle dance diameter per 15 s



15 **Distinguish** the extent to which the bee language of the waggle dance best fits kin-directed altruism or reciprocal altruism.

Bluegill sunfish foraging for *Daphnia*

The bluegill sunfish (*Lepomis macrochirus*) is a species of fish from the still, fresh-water ponds and lakes in the western Northern Hemisphere, particularly in regions of the USA (Figure 17.31). Its diet is chiefly the tiny crustacean *Daphnia*, common in these habitats, but vulnerable to predation by a range of predators. *Daphnia* itself feeds on the saprotrophic bacteria populations that break down the faeces of many aquatic non-vertebrates (themselves variable in numbers, too). Consequently, populations of *Daphnia* are very variable in size, even in apparently favourable conditions.

The feeding strategy of the bluegill sunfish has been observed to vary according to the abundance and size of *Daphnia*. The observed patterns of feeding behaviour are known to behavioural scientists as **optimal foraging theory**.

When *Daphnia* are present **in excess and in a range of sizes**, feeding is less likely to be a random process. In these conditions, bluegill sunfish are often seen to pursue the largest *Daphnia*, even when this is at a relative distance and will require significant energy expenditure in the pursuit and achievement of the goal. Observers assume that the extra energy the larger prey provides exceeds the additional energy investment in the capture. Meanwhile, small and medium-sized *Daphnia*, also present in abundance and within more immediate reach, are more likely to be ignored.

When its **prey is in short supply**, the predator is likely to eat any size of *Daphnia* that it is fortunate enough to encounter. In fact, this is the case, in contrast to their behaviour at high prey densities when they are more likely to be selective.

Figure 17.31 The bluegill sunfish (top) and its prey – bluegill sunfish grow to about 20 cm in length, whereas their prey, *Daphnia*, range in length from 0.2 to 5 mm

Mate selection and behaviour traits

In animal species that reproduce sexually, the quality of the mate may be critical to reproductive success. So it is not surprising to find that animals seldom mate indiscriminately – various mechanisms ensure some selectivity in the sexual process.

Sexual selection is the name Charles Darwin gave to the struggle between individuals of one sex (normally the males) for the possession of access to individuals of the opposite sex. The outcome for a loser of this struggle is few or no offspring.

Victory in the struggle may depend on the use made of special features of structure or behaviour. The long-term outcome has been the evolution of **exaggerated traits** that draw attention to a potential mate and markedly increase the possibility of reproductive success.

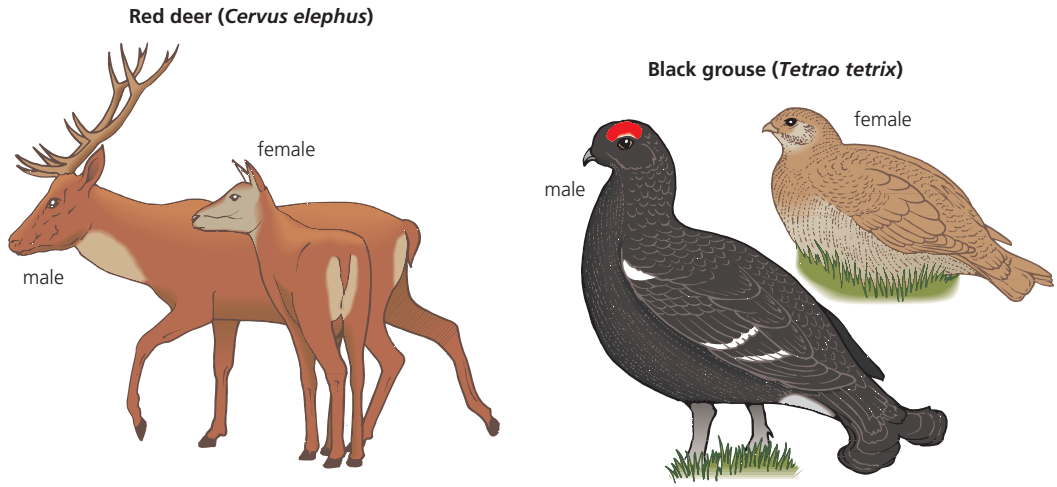


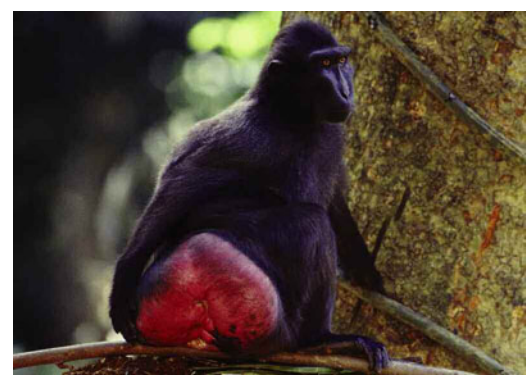
Figure 17.32 Sexual dimorphisms due to sexual selection

Fortunately, mate selection is usually the initial step in the reproductive cycle, and mating typically follows promptly. This is advantageous where the winning trait also makes that competitor more vulnerable to predation. Exaggerated traits advantageous to mate selection may not necessarily favour long-term survival. However, they enhance the chance of reproductive success and so they have been selected for.

Today we assume that differences in size, shape or colour between males and females of the same species, occupying the same niche (known as **sexual dimorphism**), are typically due to sexual or mate selection in situations where one male seeks the opportunities to mate with individuals or even take possession of a group of females, to the exclusion of other males (Figure 17.32).

Mate selection may also be the outcome of exploiting special features such as antlers aggressively (e.g. in male deer) or distinctive plumage assertively (e.g. in the peacock – Figure 17.33 A).

A male peacock



B female Macaque monkey

C European robins

Figure 17.33 Exaggerated traits associated with mate selection

Successfully mating is also dependent on the appropriate physiological state of the partners. Once this state is attained, animals may advertise their state of preparedness, sometimes in highly distinctive ways. As the female enters oestrous, this may be advertised by marked bodily changes (e.g. the Macaque monkey – Figure 17.33 B). Additionally, the secretion of volatile hormones (**pheromones**) generates detectable odours that attract potential partners.

Mate selection may be dependent on the identification of a substantial territory in which the offspring may be fed and reared. Once this territory has been established it may be advertised to aid mate selection (and defended against encroachment) by otherwise uncharacteristic noisy vocalisations and assertive or aggressive behaviour (e.g. bird display and territory-proclaiming song – Figure 17.33 C).

Rhythmical variations in activity in animals

Rhythmical behaviour patterns are common in animals, including daily (circadian) and annual rhythms. These patterns have adaptive value – aiding survival of the organisms concerned.

Circadian rhythms

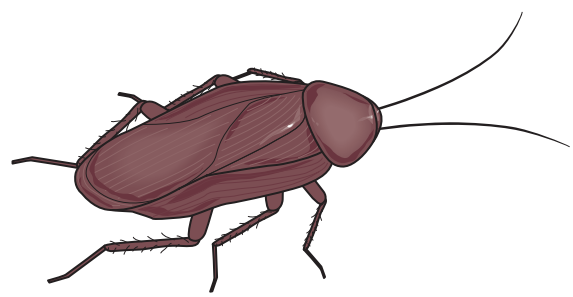
Animals are active for only a part of the 24-hour cycle. Some function at dusk or dawn (crepuscular), some in the night (nocturnal), and some in the day (diurnal). The time period favoured for activity can be shown to provide favourable conditions for feeding and survival.

For example, the cockroach, *Periplaneta*, experiences favourable humidity and decreased danger from human presence in times of darkness. The cockroach is nocturnal. It can be shown to adjust its behaviour to light/dark stimulation so as to be active only in darkness.

So, when the cockroach with its marked daily rhythm is deprived of light/dark clues of time of day, the animal is found to operate on a clock that is close to the 24-hour clock, but slightly longer, by perhaps one hour. Clearly, there is an internal clock mechanism – we say the rhythm is endogenous in that it is apparently independent of the outside stimulus to which it can be related.

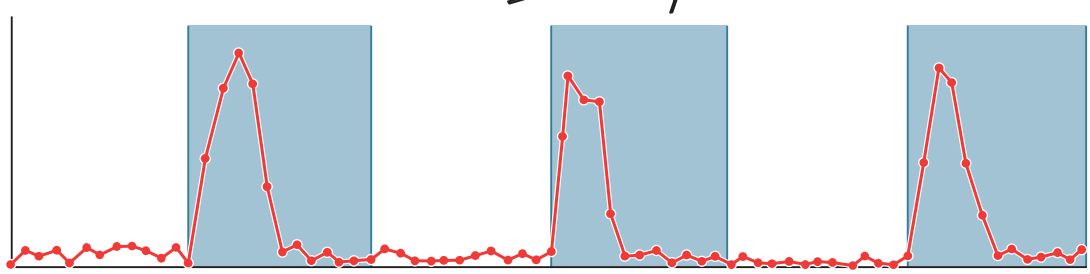
Figure 17.34 A circadian rhythm in the cockroach, *Periplaneta*

The cockroach (*Periplaneta*) is a nocturnal scavenger of human habitation, common in kitchens, bake houses and other warm places where food matter and waste left by humans may collect. They have a similar body construction to the locust, but are dorso-ventrally flattened. However, they do not have well developed back legs for jumping, but may escape from danger by being able to run fast.



study of the activity / inactivity of the cockroach under a regime of 12 hr light / 12 hr darkness

Scavenging commences as the light goes, and is completed during the dark period (in approximately 6 hours).



the dark period was advanced by 7 hours

After a brief delay scavenging activity is advanced to the new start of darkness, but is maintained for approximately the same time.

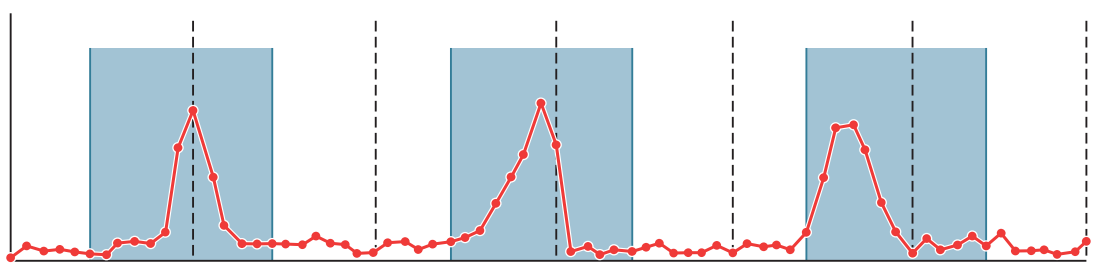
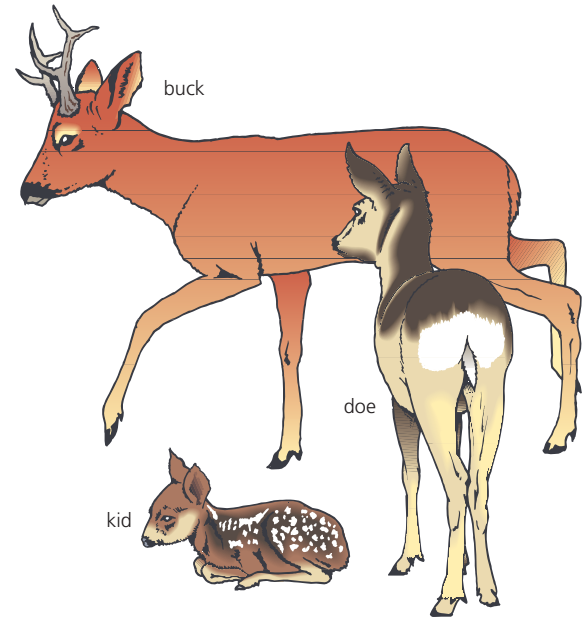
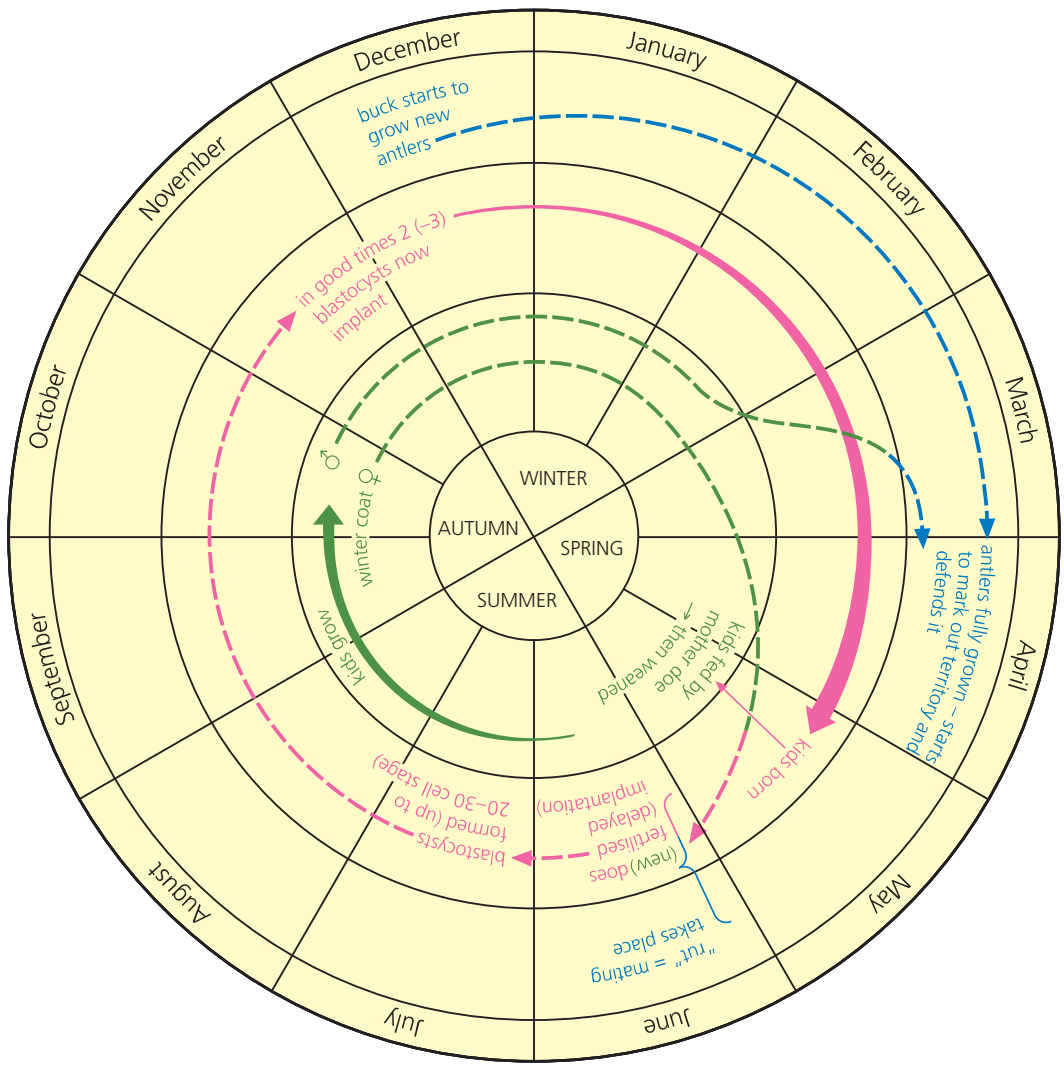


Figure 17.35 Roe deer – their life cycle in an environmental context

In the life cycle, growing conditions in summer and autumn and the severity of the winter determine the general condition of bucks and does – as they recover from the rut, or pregnancy and lactation, and prepare for winter. Preparations for rearing of kids take place in late winter/spring, but the kids are born and reared in late spring/summer when conditions are most favourable – illustrating the adaptive value of the seasonal activity pattern in roe deer.





However, endogenous rhythms are responsive to outside factors. We see this in the experimental situation of cockroaches subjected to an extension of the dark period (Figure 17.34). There is no immediate effect, but quite quickly the animal advances its light/dark activity cycle in response to earlier darkness, so that the peak in activity is maintained at the same relationship to the onset of darkness whenever that regularly occurs.

What is the adaptive value of this behaviour pattern?

The dark period and the activity response are associated with exploratory behaviour, mainly feeding. The dark, when it comes, ensures a more humid environment than daylight, and may afford the insect protection from predation. But if the period of darkness is variable in some way, it is clearly advantageous for the endogenous behaviour pattern to be adjustable to changing circumstances.

Annual rhythms

Most animals, unlike humans, do not have the potential to breed all the year round. Instead, they produce young in a season favourable for rearing and feeding. The hormonal control of breeding cycles is discussed on page 228.

The roe deer (*Capreolar capreolus*) is a well known inhabitant of Eurasian forests, occurring from Britain in the west to Manchuria and China in the east. They are ruminant herbivores, highly adaptable, and found in a wide variety of habitats – anywhere that affords the privacy and shelter they need and a variety of plant food. They live in loose family groups or are solitary – they never form into herds. Bucks are frequently solitary, but does are accompanied by kids for much of the year. There are three subspecies, showing some variation in the timing of their annual cycle that attunes with climate variations in different parts of their extent. Here, the pattern seen in north/western populations is outlined (Figure 17.35).

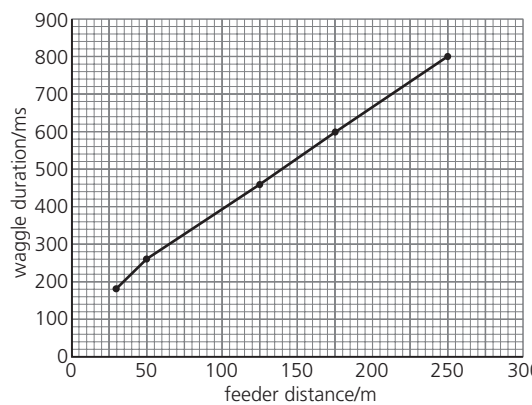
From the chart in Figure 17.35 we can see that the kids are born at a time when the flush in vegetation produced by improved summer weather is at its height and the mother can easily maintain her diet and the quality of her lactation, on which the kid is dependent.

Also, while rutting takes place in autumn when the does have recovered their maximum weight from spring and summer feeding, their fertilised eggs (now blastocysts) do not implant in the uterus wall until December or January. In good years, two or even three implant, and twins are most frequently born in the following May or June. In harsh winter conditions, when survival of the mother and any newborn may be at greater risk, fewer blastocysts implant. Clearly, rhythmical variations in activity in the roe deer life cycle have adaptive value.

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 Entomologists investigating communication between worker bees (*Apis mellifera*) observed workers collecting sugar solutions from feeder stations placed at different distances from the hive. The workers' waggle dances were filmed as they returned to the hive and the duration of the waggle was measured. The results are shown in the graph.



a State the relationship between the duration of the waggle dance and the distance of the feeder from the hive. (1)

b State another piece of information that worker bees communicate to one another. (1)

Worker bees were then trained to collect sugar solution by flying down an 8 m long tunnel running south from the hive. The tunnel was placed so that the entrance was 3 m from the hive. The floor and walls of the tunnel were covered with a black and white chequered pattern but the roof was covered in netting so the bees could see the sky.

c Calculate the actual distance of the sugar solution at the end of the tunnel from the hive. (1)

When the bees returned to the hive, the duration of the waggle dance was recorded and found to have a mean of 350 ms.

d Using the graph, determine the distance to the sugar solution indicated by the duration of the waggle dance by worker bees returning from the tunnel. (1)

e Deduce why there is a difference between the real distance flown by the bees and the information that they passed on to their fellow workers in the hive. (2)

f Suggest a modification to the experiment that would help to test your answer to part **e**. (1)

Standard Level Paper 3, November 03, QE1

Q2 **a** Compare rod cells and cone cells in the retina. (3)

b **i** State one example of altruistic behaviour. (1)

ii Explain the role of the example given in **i** in the social organisation of the population. (2)

Standard Level Paper 3, November 04, QE2

Q3 **a** Outline the value of quantitative data in studies of behaviour. (4)

b Explain, using examples, how inhibitory psychoactive drugs affect brain physiology. (6)

Higher Level Paper 3, May 05, QE3

Q4 Outline Pavlov's experiments on the conditioning of dogs. (3)

Higher Level Paper 3, May 04, QE2

Questions 5–10 cover other syllabus issues in this chapter.

Q5 **a** Describe the different types of neurones of the mammalian nervous system. (6)

b Explain how neurones are arranged to form a reflex arc. (4)

Q6 Sketch the structure of the human ear and annotate your drawing to show the role of the component parts in the receipt of sound waves and their translation into impulses in the auditory nerve. (8)

Q7 **a** Identify the key steps to synaptic transmission of an action potential. (5)

b Explain how a pre-synaptic neurone may:

i enhance

ii inhibit

post-synaptic transmission. (4)

Q8 **a** Annotate a drawing of the human brain in section to outline the main roles of the different parts. (6)

b Explain how pain is perceived in the body and how endorphins are involved in natural pain relief. (6)

Q9 **a** Explain the difference between *kin-directed altruism* and *reciprocal altruism*. (4)

b Suggest how the practice of reciprocal altruism may be favoured by natural selection. (4)

Q10 **a** By means of examples, show the difference between circadian and annual rhythms observed in animal behaviour. (6)

b Identify the chief environmental factors by which an 'internal clock' may be regulated in the examples of rhythmic behaviour quoted in **a**. (2)

Option F

18 Microorganisms and biotechnology

STARTING POINTS

- **Biodiversity** is a term for the vast numbers of living things that exist. **Classification** involves naming organisms (genus and species), and then grouping them into a hierarchical classification.
- **Prokaryotes** include the bacteria and cyanobacteria – all organisms with very small cells and **without a true nucleus**. Prokaryotes are largely unicellular.
- **Eukaryotes** – plants, animals and fungi – have cells containing many organelles, and a nucleus **contained by a nuclear membrane**. Some eukaryotes are unicellular, but many are multicellular organisms.
- There are two forms of **nucleic acid**, **DNA** and **RNA**. They differ in the 5-carbon sugar and the bases they contain, and in their roles in cells. Also, RNA is single-stranded, DNA is double-stranded.
- **Biotechnology** is the industrial application of biological processes. It has ancient origins, but today biotechnology is often linked with **genetic engineering**, particularly of microorganisms.
- This chapter extends study of aspects of cell structure begun in Chapter 1 (pages 1–36), of environmental biology begun in Chapter 6 (pages 137–77), and of genetic engineering and biotechnology begun in Chapter 5 (pages 117–36).

At one time, the living world seemed to divide naturally into two kingdoms, the **plants** and **animals**. Then, with the use of the electron microscope in biology came the discovery of the two fundamentally different types of cellular organisation, namely **prokaryotic** and **eukaryotic**. As a result, the divisions of living things were reorganised, eventually into **five kingdoms**, namely the prokaryotes (bacteria and cyanobacteria), and the four eukaryotic kingdoms of the protoctistans, fungi, plants, and animals (Figure 6.43, page 176).

More recently, studies of life under extreme environmental conditions, including extremes of temperature or salinity, and at great ocean depth, have led to discovery of a huge and growing range of previously unknown microorganisms. A systematic hunt for **extremophiles**, life forms that live in extreme places, is under way.

The biochemistry of the cells of extremophiles has proved distinctly different from existing known life forms, too. As a consequence, the study of microorganisms has fresh consequences and **applications in many fields of biology**, including ecological and environmental biology, biotechnology, and health.

In this chapter, the **diversity of microorganisms** is explored, and their **roles in the environment** discussed. Then the applications of microbiology in **biotechnology and food production** are examined.

In the Additional Higher Level extension, the **metabolism of microorganisms** is reviewed further, and finally, issues of **microbes and disease** are explored.

The diversity of microorganisms

F1.1–1.9

Vast numbers of microorganisms occur in the biosphere immediately around us, and many of them exist in environmental conditions similar to those of plants and animals. However, many more microorganisms flourish in the most unlikely places, where they survive and prosper despite the extremely challenging environmental conditions. All of these latter microorganisms are collectively referred to as the extremophiles. The numbers and different types of known extremophiles are increasing as the importance of these strange habitats becomes better understood. Extremophiles include microorganisms that:

- are salt-loving (**halophiles**), found in salt lakes and where sea water has become concentrated and salt has crystallised;

- require extremely alkaline conditions above pH 10 (**alkalinophiles**), found in soda lakes;
- survive abnormally high temperatures (**thermophiles**);
- withstand 250 atmospheres pressure (**barophiles**), as well as extremely high temperatures;
- survive extremely cold habitats (**psychrophiles**).

The microorganisms of extreme habitats have cells that we can identify as prokaryotic (page 16). That is, there is no true nucleus present, nor is the cytoplasm packed with the range of organelles so typical of eukaryotic cells.

However, the **biochemistry of the cells of extremophiles** has proved distinctly different from existing known life forms. In particular, it is the larger **RNA molecules present in the ribosomes** that have proved tell-tale molecules.

How are the RNA molecules of extremophiles analysed?

First, ribosomes are isolated from the cytoplasm of cells, and then the RNA that makes up these tiny organelles (remember, ribosomes are sites of protein synthesis) is extracted. On analysis, the sequence of nucleotides present (i.e. the sequence of the bases cytosine, guanine, adenine and uracil in this RNA) is determined.

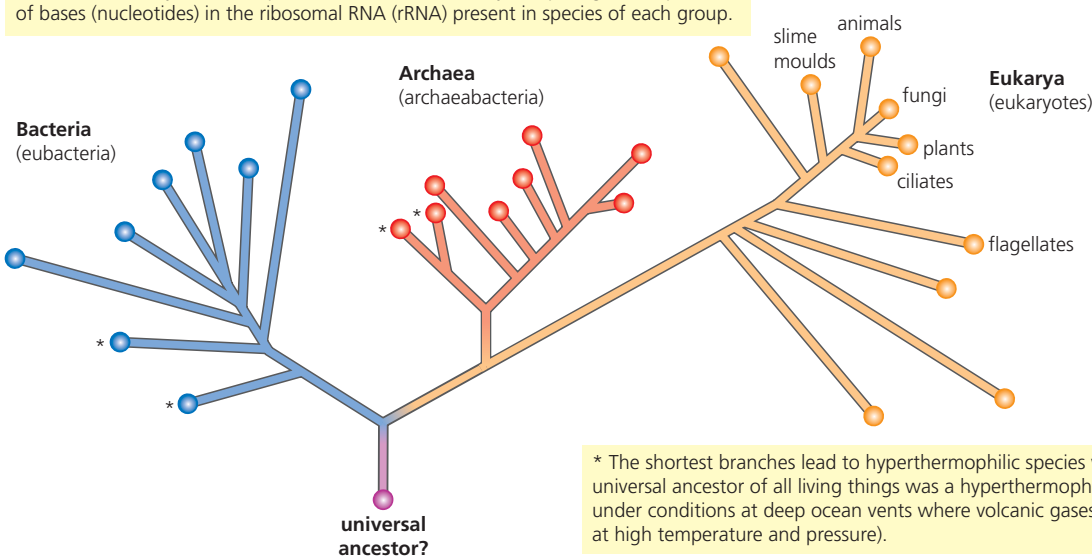
Comparisons of the base sequences in RNA molecules of the ribosomes in these cells with those of previously known bacteria and eukaryotes have led to the discovery of **new evolutionary relationships**.

The outcome has been the development of a new scheme of classification of living things. We now recognise **three major forms of life**, and these are called **domains** (Figure 18.1). The organisms of each domain share a distinctive, unique pattern of ribosomal RNA which establishes their close evolutionary relationship. These domains are:

- the **Archaea** or **Archaeabacteria** (the extremophile prokaryotes);
- the **Bacteria** or **Eubacteria** (the true bacteria – that is, the other prokaryotes);
- the **Eukarya** (all eukaryotic cells – the protoctista, fungi, plants and animals).

Figure 18.1 The classification of living organisms into domains

These evolutionary relationships have been established by comparing the sequence of bases (nucleotides) in the ribosomal RNA (rRNA) present in species of each group.



* The shortest branches lead to hyperthermophilic species which suggests that the universal ancestor of all living things was a hyperthermophile (possibly 'assembled' under conditions at deep ocean vents where volcanic gases are discharged into water at high temperature and pressure).

1 **Distinguish** the different forms of RNA present in cells by means of their specific roles in protein synthesis.

The characteristics of the three domains

The Archaeabacteria are organisms with which we are least familiar, so we concentrate on their unique features first. Then we can compare their structure with that of cells of organisms of the other two domains (Table 18.1).

The **Archaeabacteria** are a diverse group in both structure and physiology. Some are single cells, others are filamentous or form other types of cellular aggregate. We have noted that many are found in extreme conditions (the extremophiles), some in terrestrial habitats, others in water. For example, Archaea represent about a third of all prokaryotic biomass in polar coastal waters. However, these microorganisms occur very widely indeed in the natural world. A few live symbiotically in the gut of animal species – as do vast numbers of true bacteria, too. They probably occur everywhere.

As prokaryotes, the Archaeabacteria have **cell walls**, but the chemistry of their walls is different from the chemistry of the walls of the true bacteria. Often a variety of complex polysaccharides are present, but none contains the peptidoglycan or particular amino acids characteristic of the Eubacteria. Some have a substantial layer of protein or glycoprotein external to the polysaccharides, and some have walls of protein only.

Another distinguishing feature is the **lipids** of the cell membranes found in the Archaeabacteria. Glycolipids are present, but these have branched chain **hydrocarbons** attached to glycerol (in place of the **fatty acids** of the lipids of the Eubacteria and eukaryotes). Polar lipids are also present, and they all combine in different ways to form membranes of variable thickness and rigidity.

The Archaeabacteria, like the Eubacteria, have a **single circular chromosome** that is without **introns** (non-coding sequences of nucleotides, page 245). However, they generally possess far fewer genes than the true bacteria. Their nucleic acid is extremely variable in the proportions of the two bases guanine (G) and cytosine (C) that it contains. In species where the nucleic acid has been sequenced, the genes are unlike those found in true bacteria and eukaryotes. Also, there are relatively few **plasmids** present.

The **ribosomes** of Archaea are small (70S) like those of the Eubacteria, but they are more variable in shape and chemistry. Understanding of the evolutionary relationships of members of the three domains is based on the biochemical analysis of the ribosomes. Finally, the **physiology**, **metabolism** and **life styles** of Archaea are very variable, as described shortly. For the purposes of comparison, the structure of a typical bacterium (*Escherichia coli*) is shown in Figure 1.15 (page 16), and of a human liver cell – typical of a eukaryotic cell – in Figure 1.13 (page 15).

Look those up again now, to help you make sense of Table 18.1.

	Archaeabacterial cell (Figure 18.2)	Eubacterial cell such as *E. coli* (Figure 1.15)	Eukaryotic cell such as liver cell (Figure 1.13)
Ribosomes: size and chemical analysis	■ 70S size ■ unique biochemistry	■ 70S size ■ unique biochemistry	■ 70S size ■ unique biochemistry
Histones with nucleic acid of chromosome(s)	■ absent	■ absent	■ present, forming nucleosomes
Introns within chromosome(s)	■ absent	■ absent	■ present between exons
Plasmids in cytoplasm	■ few present	■ many present	■ typically absent
Cell walls: present or absent, and chemistry	■ present ■ chemically different from those of Eubacteria	■ present ■ chemically different from those of Archaea	■ present in green plant cells (cellulose) and in fungi (chitin) ■ absent in animal cells
Cell membranes: composition of lipids	■ glycolipids with hydrocarbon chains, in plasma membrane	■ lipids with fatty acids, in plasma membrane	■ lipids with fatty acids, in all membranes present

Table 18.1 The cellular characteristics of the three domains

Diversity within the Archaeabacteria

The wide diversity in the habitats of the Archaeabacteria may be illustrated by contrasting those of the methanogens, thermophiles and halophiles.

Methanogens occur in microhabitats where oxygen is absent – we say they are obligate anaerobes (Figure 18.2). In fact they require anaerobic environments rich in organic matter.

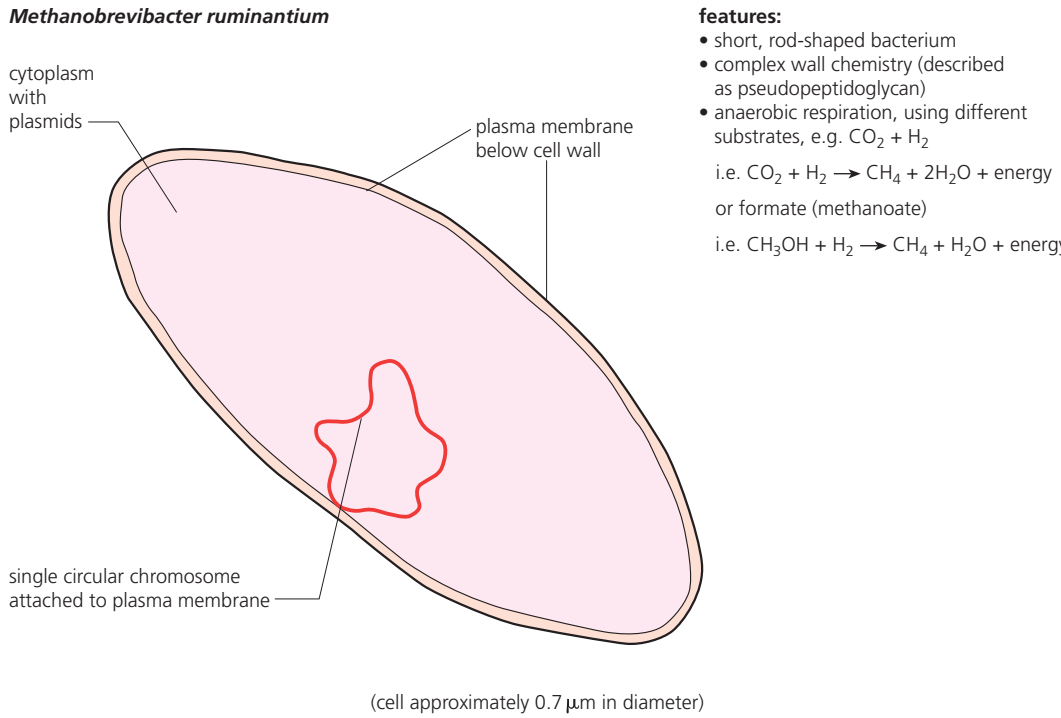


Figure 18.2 Typical methanogenic bacterium of the Archaea (drawing based on an interpretation of a TEM)

Here they produce methane as a waste product of the metabolism of a variety of respiratory substrates, including combinations of methanoate (formate), methanol, ethanoate (acetate), carbon dioxide and hydrogen.

These Archaeabacteria are found in the rumen and intestinal system of animals (typically, in ruminants such as cows, sheep and deer), in marine and fresh-water sediments rich in organic matter, and in marshes. They also occur in hot springs and in the anaerobic sludge digesters of sewage works (Figure 18.16, page 567).

A kilogram of organic matter may yield up to 600 litres of methane, so methanogens have potential in bio-fuel production in the future, when fossil fuels are scarce and more expensive. However, methane is also a powerful greenhouse gas – methanogens indirectly contribute to global warming (page 149).

Thermophiles grow at extremely high temperatures. Prokaryotes with this property were first discovered in hot volcanic springs such as occur in Yellowstone Park in the USA. The temperature they flourished at, 70 °C, seemed exceptional at the time.

Now we know that the deep ocean-floor volcanic vents that continuously discharge hot, sulphurous volcanic gases are home to **hyperthermophiles** that require a minimum of 70 °C, and typically thrive at over 100 °C, albeit at very high pressures, so the water is not boiling (Figure 18.3).

These habitats have been suggested as sites where life itself may have originally evolved, as conditions there may have favoured the spontaneous formation of complex organic molecules, and then assisted these molecules (including polymers) to become assembled into the simplest living cells (page 457).

Halophiles grow in salt lakes and anywhere that sea water has evaporated leaving a highly concentrated solution of salts or even salt crystals. One such organism, *Halobacterium halobium*, is found among the salt crystals in salt mines.

Closely related to halophiles are members of the Archaeabacteria that require not only a high concentration of salt, but also extremely alkaline conditions (pH 10 or higher). For example, in soda lakes, *Natronococcus* occurs naturally and thrives. A high concentration of salt is maintained in the cytoplasm of this halophile, preventing dehydration of the cell.

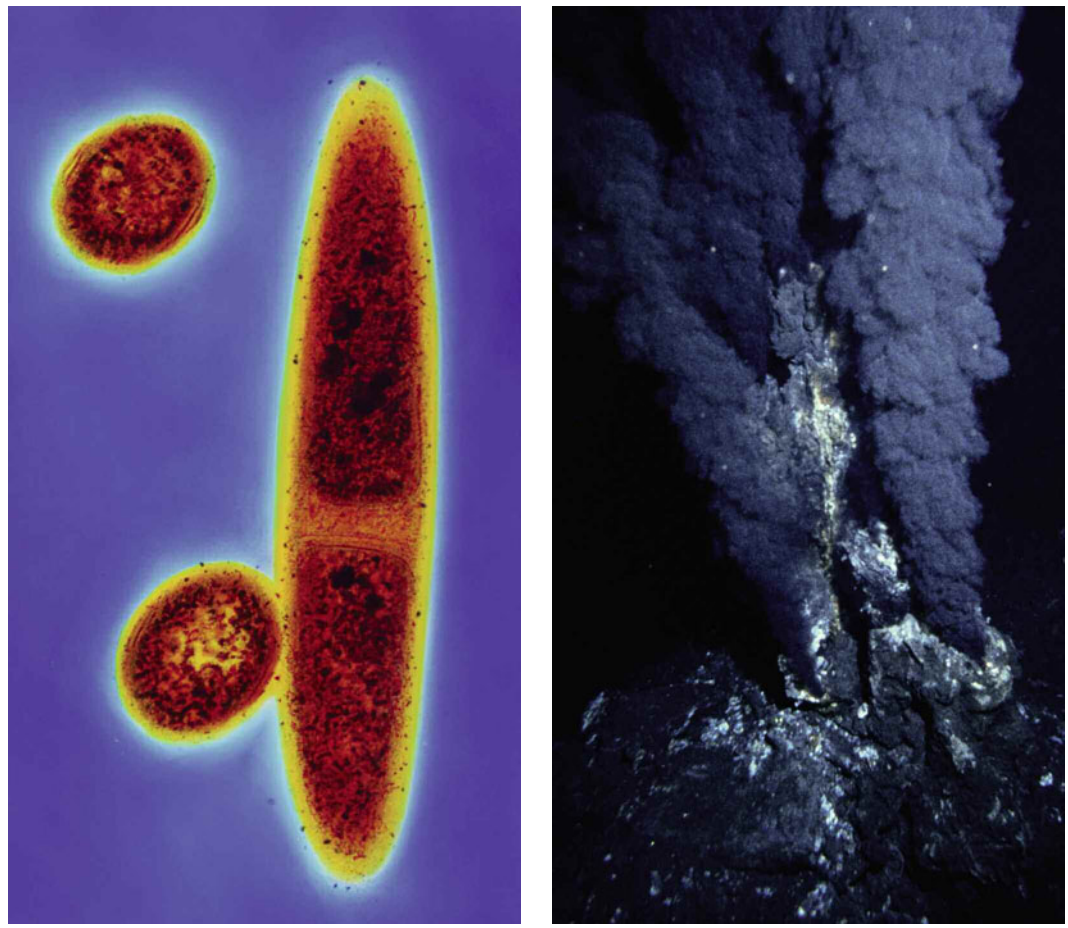


Figure 18.3 A hyperthermophile Archaeabacterium and its deep sea habitat

Left: TEM of *Methanopyrus* (false colour), size 0.5 × 2–14 μm.

Right: A deep ocean volcanic vent (upper temperature about 110 °C) from which volcanic gases bubble out, rich in sulphur and heated by geothermal energy. Extremophile Archaea found here use the sulphur compounds as a source of energy.

Diversity within the Eubacteria (true bacteria)

The structure of a bacterium, *Escherichia coli*, a very common gut commensal, is illustrated in Figure 1.15. Note that the outline shape of *E. coli* is a rod.

The very many thousands of different bacteria that have been isolated, identified and named by bacteriologists (genus and species), have been organised into the traditional hierarchical classification system used in biology:

Phylum ← Class ← Order ← Family

Alternative, commonly used ways of dividing bacteria include simple groupings based on the shape of their cells and the chemical structure of their cell walls.

Shapes of bacterial cells

Bacteria are classified into five groups on the basis of their shape; basically bacteria are rods, spheres, comma-shaped, spirals, or corkscrew-shaped. In fact, the majority are rods or spheres. The cells maintain these shapes because bacterial cells have rigid walls.

Although basically unicellular, bacteria may often appear clumped together and species showing this form of growth are named accordingly. For example, bacteria growing in pairs have the prefix diplo- attached to the generic name. Where the cells occur as a chain, the prefix strepto- is used, and where they occur in grape-like clusters the generic prefix is staphylo- (Figure 18.4).

Wall structure in bacteria

We have seen that the wall of a bacterium gives a permanent shape to the cell because of its rigidity. It also protects the cell contents against rupture due to osmosis, for example, and it helps to protect against harm by other organisms.

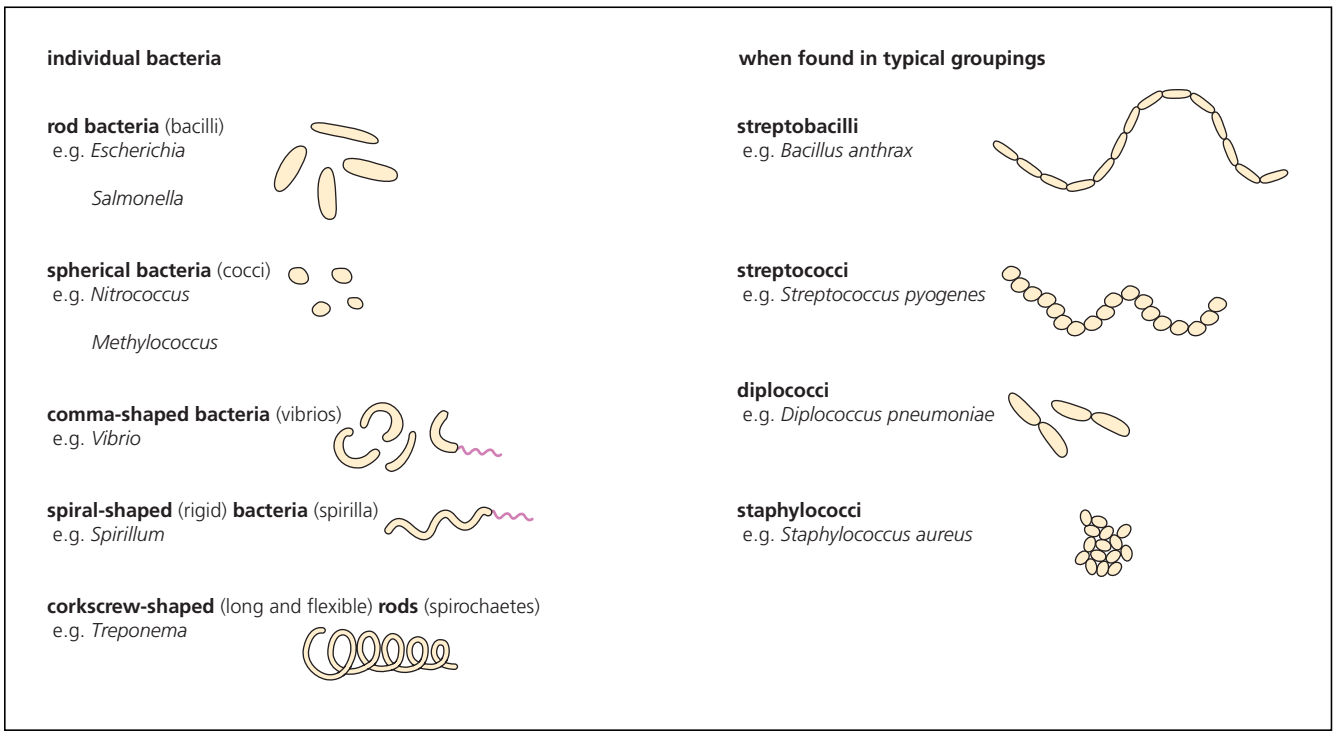


Figure 18.4 Classification of bacteria by shape

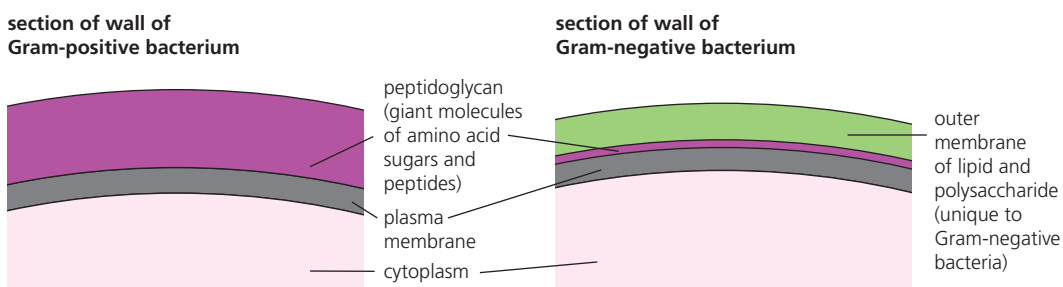
Bacterial cell walls contain giant molecules (polymers) built of amino-sugars and peptide monomers. The resulting polymer is known as **peptidoglycan**, and is a totally different substance from cellulose of plant cell walls. Bacteria have walls of this substance, but some have additional layers on the outer surface of their wall, and these additional layers change the staining property of the wall.

It was a Dane, Hans Gram (1884), who first showed that some bacteria can be stained purple with a particular dye called crystal violet. Such bacteria are now known as **Gram-positive species** (Figure 18.5). They have a wall made largely of peptidoglycan.

Other bacteria cannot retain crystal violet in their walls, and they are known as the **Gram-negative species** (Figure 18.5). They have a wall that contains peptidoglycan, but external to that layer is a shield layer largely made of lipid.

Thus, the Gram-staining effect is due to an important difference in cell-wall chemistry. It has proved to be a valuable tool for recognising bacteria and for separating them into two groups. Incidentally, the Gram-positive bacteria tend to be destroyed by penicillin and certain other drugs, whereas Gram-negative species have a wall that resists these drugs.

Figure 18.5 Gram-positive and Gram-negative bacteria walls



Some bacteria form aggregates with distinctive characteristics

Bioluminescence

The bacterium *Vibrio fischeria*, a Gram-negative, rod-shaped bacterium, is able to emit light – a phenomenon known as **bioluminescence**. The bacterium occurs in marine habitats, and large numbers are found in a mucus matrix in the light organ of a squid species and also of the

Atlantic flashlight fish. Light is generated by the action of an enzyme, luciferase, acting upon a reduced respiratory coenzyme. In effect, light is generated by deflecting electrons from the electron-transport chain and ATP synthesis (page 274). When the bacterium is in high-density populations, the organisms interact to trigger light emission.

Biofilms

When certain bacteria that are pathogens reach populations of sufficient density, they interact to produce toxins. One such is the rod-shaped bacterium *Pseudomonas aeruginosa*, which when present in large numbers in the respiratory tract of patients with cystic fibrosis (page 576), produces toxins that may overwhelm and kill the host.

Diversity within the microscopic eukaryotes

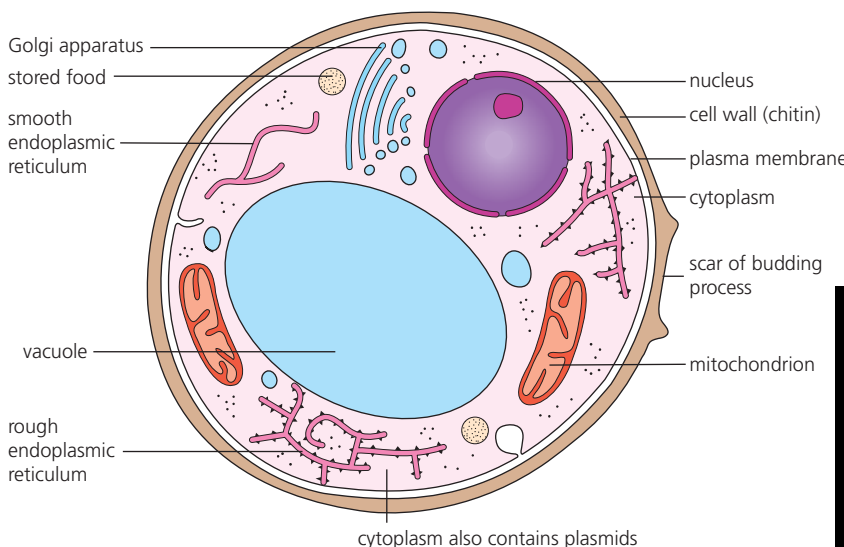
We can now consider the diversity found among microscopic eukaryotes, concentrating particularly on their nutrition, locomotion and wall structure. The eukaryotic cell is structurally distinct from that of prokaryotes; for one, eukaryotic cells are mostly very much larger, and contain an array of organelles not found in prokaryote cells.

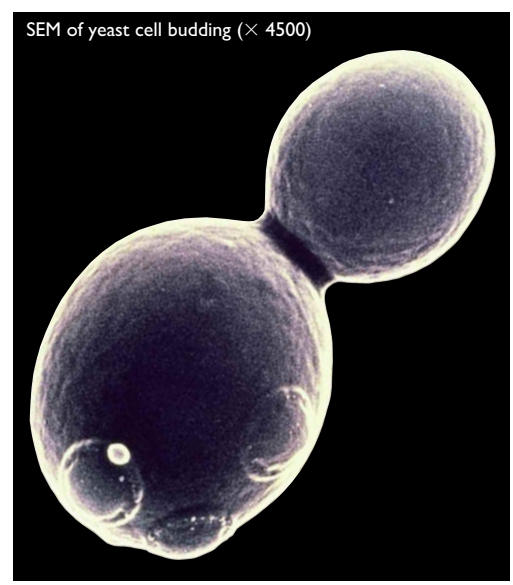
While eukaryotic microorganisms share this characteristic, they also show great diversity. This is illustrated by the six eukaryotic unicells described below. The differences are then summarised in Table 18.2.

Saccharomyces (yeast)

The name *Saccharomyces* means 'sugar fungi'. Yeasts are saprotrophic, unicellular fungi that occur everywhere sugar solution is available – in flowers, on the surface of fruits, on leaves and stems where sap is exuded. They are also found in the soil and on animal mucous membranes. Yeasts are unicellular (Figure 18.6), but under favourable growing conditions, the cells reproduce asexually and divide (called budding) so quickly they may temporarily form long branching chains of connected cells. The cell wall is composed of fungal cellulose (chitin).

Figure 18.6 The structure of *Saccharomyces*





Yeast cells are non-motile, but are readily dispersed by larger organisms visiting sites where yeasts grow. Yeast spores are carried around in air currents, too.

The yeasts respire by alcoholic fermentation and produce carbon dioxide and ethanol as waste products. The ethanol cannot be metabolised further by yeast, and it may accumulate in the growth medium. Yeast is of great economic importance because:

- waste products of fermentation are exploited in brewing, wine making, and in the baking of bread dough;
- they are easily cultured in liquid media in fermenters, and so can be useful in modern biotechnology industries, too;
- like many bacteria, they contain plasmids in the cytoplasm (a very unusual feature in eukaryote cells), and so yeast cells may be genetically modified to carry and express additional genes.

Amoeba (a protozoan)

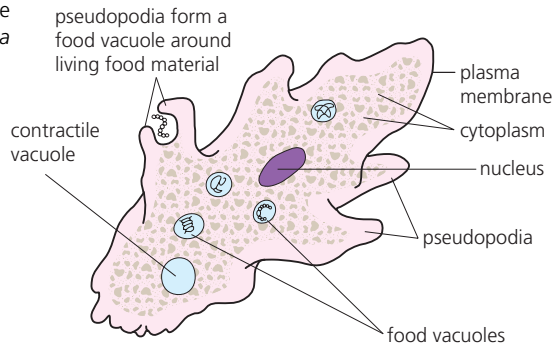


Figure 18.7 The structure of *Amoeba*

Amoeba is a protozoan, just visible to the naked eye, which occurs on the bottom of shallow lakes and ponds. Observed by light microscopy, the cell is colourless and transparent, and has an irregular and constantly changing outline (Figure 18.7). Movement occurs by flowing movements of the cytoplasm, leading to the building of almost tubular pseudopodia and the dissembling of others; this is appropriately known as amoeboid movement. No cell wall is present outside the flexible plasma membrane.

Nutrition is holozoic; *Amoeba* feeds on smaller protozoa and algae. Prey is surrounded by pseudopodia in the shape of a cup, enclosed in a food vacuole, and then digested. You are already familiar with this process from study of the activities of phagocytic white cells – part of the human body's defence against disease (page 196).

Plasmodium (malaria parasite)

Plasmodium is a parasitic protozoan that occurs in human hosts, but is unwittingly transported to new hosts by the female mosquito (*Anopheles*) when a blood meal is taken from humans. *Plasmodium* lives in liver cells at first, and then moves into red blood cells (Figure 18.8).

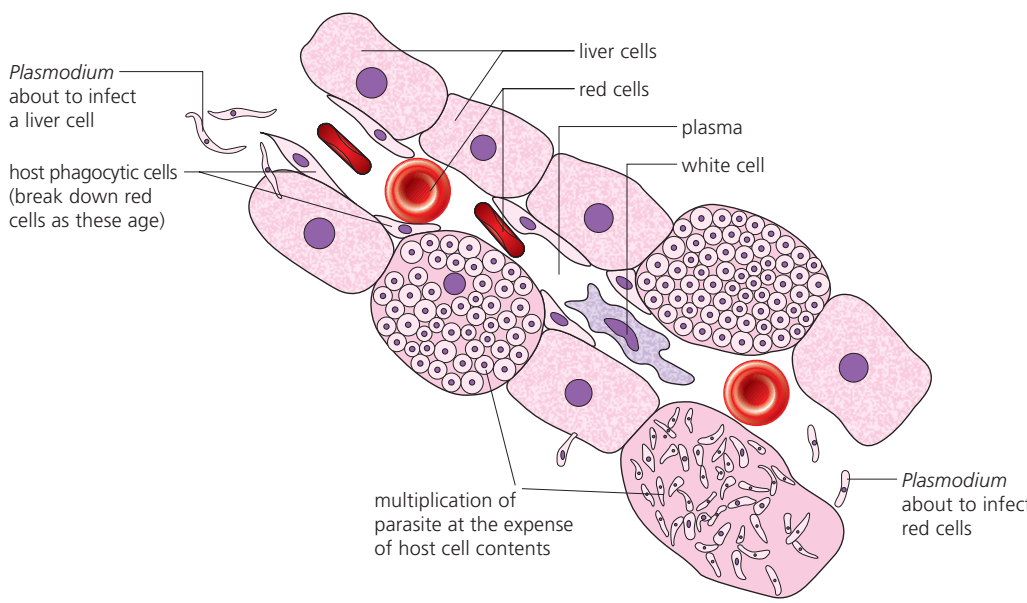


Figure 18.8 *Plasmodium*, in and out of host cells

In both tissues it feeds at the expense of the cell cytoplasm, and, at the same time, hides from the host's immune system. As the parasite escapes from host cells, toxins are released into the blood stream, and these trigger the symptoms of malarial fever, so harmful to the host.

Paramecium (slipper animalcule, a ciliate)

Paramecium is a protozoan found in ponds rich in decaying organic matter and numerous bacteria. Nutrition is largely holozoic, with bacteria the main component of the diet. This protozoan has a definite, unchanging shape because the plasma membrane is a slightly rigid pellicle through which rows of cilia project (Figure 18.9). *Paramecium* swims through the water, rounded end first, by the beating action of the cilia. Food vacuoles form at the top of a permanent gullet (cytopharynx) around acceptable microbes that have been swept up by ciliary action.

Figure 18.9 *Paramecium*

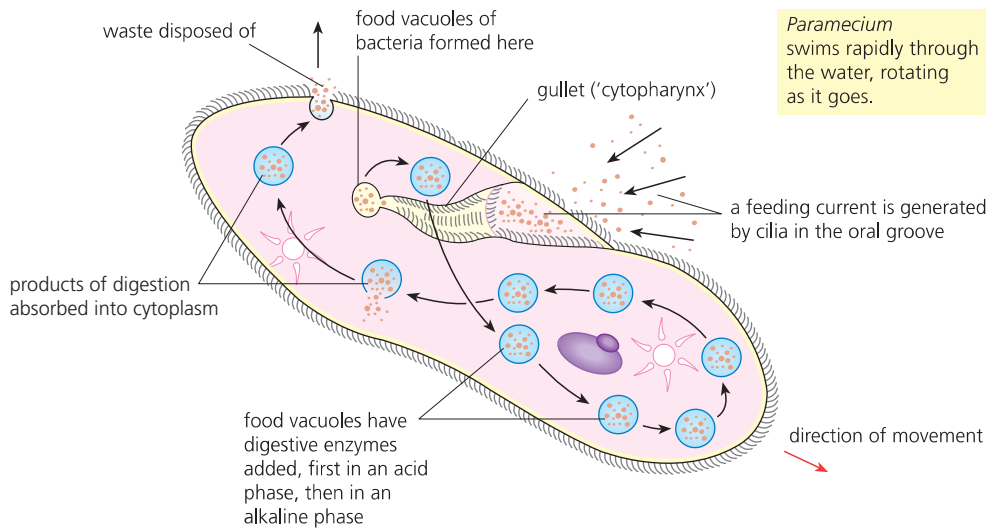
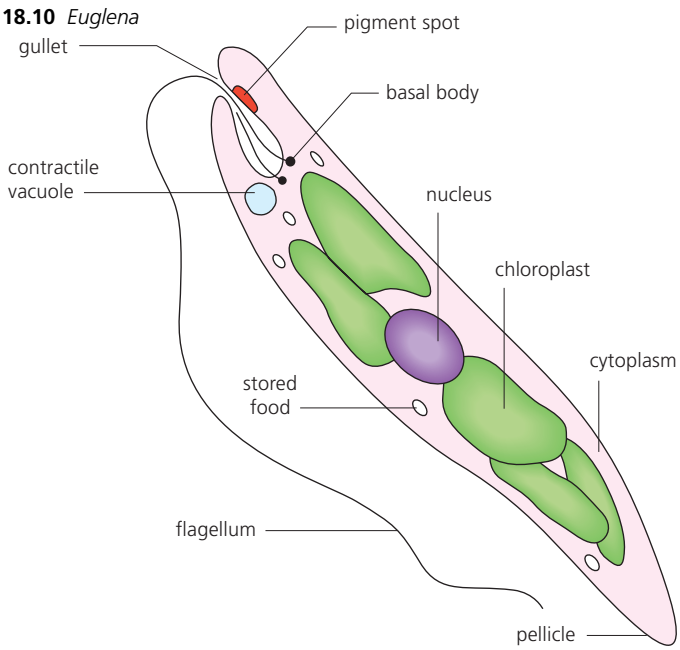


Figure 18.10 *Euglena*



Euglena (a flagellate)

Euglena is described as a plant-like protozoan. It is common in stagnant water of ponds and ditches, typically at densities that colour the water green. This tiny, spindle-shaped cell is not surrounded by a cellulose cell wall; rather the cytoplasm is bounded by a firm but flexible plasma membrane (pellicle) (Figure 18.10). In swimming movements, the long flagellum that emerges from the terminal gullet and trails beside the cell, propels the organism forwards. Movement may also occur by flowing movements of the cytoplasm. Nutrition is largely by photosynthesis and by absorption of essential ions from water for biosynthesis of essential metabolites using sugar. However, in some circumstances, *Euglena* can exist saprotrophically.

Chlorella (a green alga)

Chlorella is a unicellular alga of fresh-water ponds, present at such a density as to colour the water green. *Chlorella* is a popular organism for research because it is easily cultured in laboratories for the study of green plant metabolism. It is an immobile unicell contained within a cellulose cell

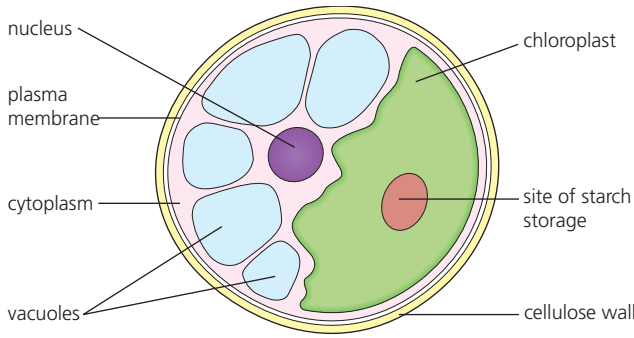


Figure 18.11 *Chlorella*

wall, and is passively transported by water currents (Figure 18.11). It has photosynthetic metabolism, and manufactures all metabolites required from the sugar formed, and from ions absorbed from the medium it lives in. Metabolically, this unicellular alga is indistinguishable from a green plant cell – hence its value as a research organism, since the supply of its nutrients and the environmental conditions it experiences are easily controlled.

Table 18.2 Diversity in microscopic eukaryotes

	Cell structure	Nutrition	Motility
Saccharomyces	■ cell wall – chitin (polymer of glucose derivative containing N)	■ saprotrophic – feeds on nectar, via fermentation	■ non-motile – chance transport by pollinators, or as spores via air currents
Amoeba	■ cell wall – absent	■ holozoic – feeds after trapping food in food vacuoles	■ motile – moves by cytoplasmic streaming
Plasmodium	■ cell wall – absent	■ parasitic – feeds on contents of liver cells and red cells	■ motile – burrows its way into host cells
Paramecium	■ cell wall – absent ■ plasma membrane – strengthened by additional protein (a pellicle)	■ holozoic – feeds after trapping food in food vacuoles	■ motile – driven by the many cilia that cover the pellicle
Euglena	■ cell wall – absent ■ plasma membrane – strengthened by additional protein (a pellicle)	■ largely photosynthetic with chloroplasts	■ motile – driven by long flagellum
Chlorella	■ cell wall – present, largely cellulose	■ photosynthetic with chloroplast	■ non-motile

2 **Compare** cell structure in eukaryotes and prokaryotes.

Introducing the viruses

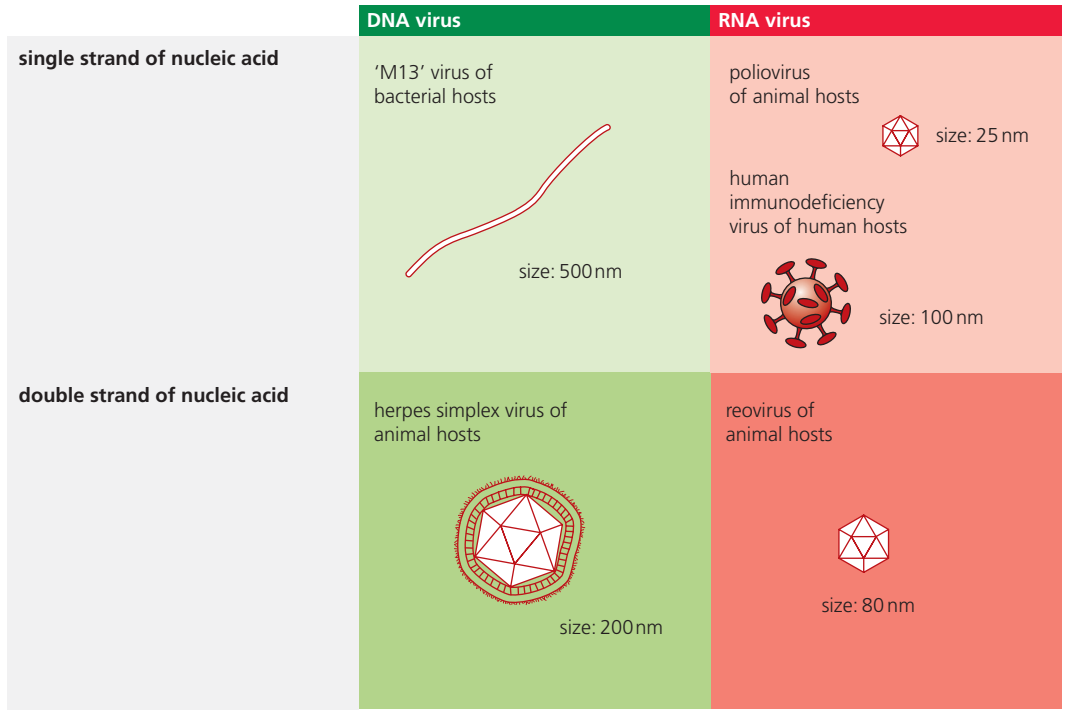
Viruses are disease-causing agents, rather than microorganisms. The distinctive features of viruses are:

- they are not cellular structures, but consist of a core of nucleic acid surrounded by a protein coat called a **capsid** (Figure 18.12);
- in some viruses there is an additional **external envelope** of membrane made of lipids and proteins (e.g. in the influenza and herpes viruses, Figure 18.13);
- they are **extremely small** when compared with bacteria (most viruses are in a size range of 20–400 nm (0.02–0.4 μm); they become visible only by means of the electron microscope;
- they can reproduce only inside specific living cells, so they function as **endoparasites** in their **host** organism;
- they have to be **transported** in some way between hosts;
- viruses are **highly specific to particular host species** – some to plant species, some to animal species and some to bacteria;
- viruses are classified by the type of nucleic acid they contain, either **DNA** or **RNA**, and whether they have a **single or double strand of nucleic acid**.

Virus structures compared

A **single-strand RNA virus** is illustrated by the influenza (flu) virus (Figure 18.13). Here, eight short single strands of RNA (a segmented genome), ranging from about 800 to 2400 nucleotides long, are seen. The protein coat (the capsid) around the nucleic acid is itself surrounded by an external layer of lipid. This combined coat is flexible, so these viruses are of no fixed shape. However, a constant feature of this viral capsid are the protein 'knobs' (known as H and N), projecting through the lipid layer. These function as antigens in the human host.

Figure 18.12 A classification of viruses

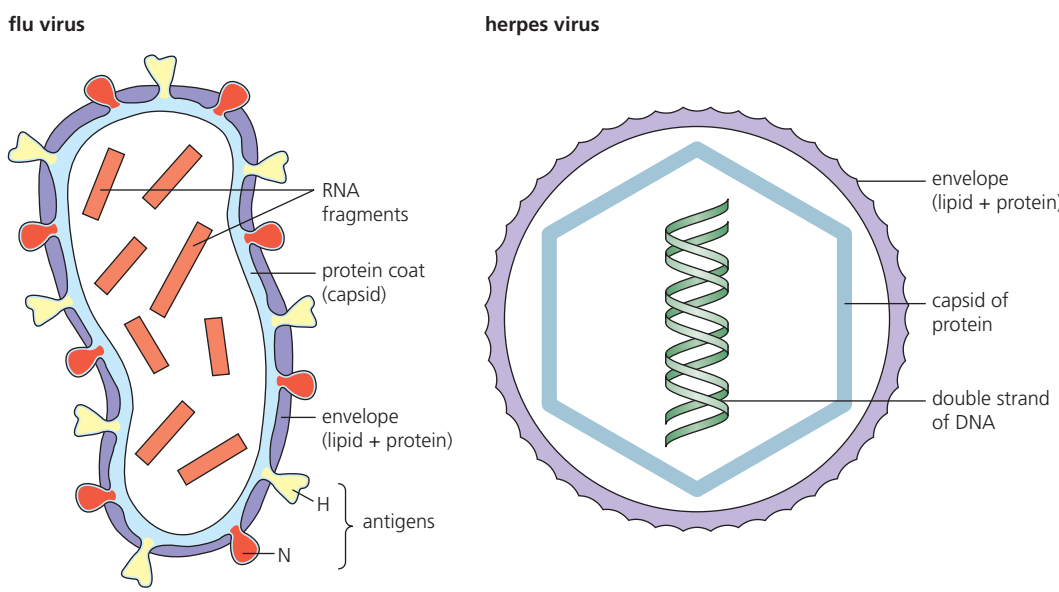


Extension: So are viruses living at any stage?

Viruses are an assembly of complex molecules, rather than a form of life. Isolated from their host cell they are inactive, and best described as crystalline. Within susceptible host cells they are highly active genetic programmes that typically take over the biochemical machinery of host cells for replication. Their component chemicals are synthesised, and then assembled to form new viruses. On breakdown (lysis) of the host cell, viruses are released, and may cause fresh infections. So, viruses are **not living organisms**, but may become **active components of host cells**.

A **double-strand DNA virus** is illustrated by the herpes virus (Figure 18.13). The protein coat (capsid) is a regular shape with twenty faces (icosahedral), surrounded by an envelope of protein and lipid, and derived (as in the case of the flu virus envelope) from the previous host cell's membrane. The envelope is acquired as the virus particle departs the host cell.

Figure 18.13 Virus structure compared



3 **Explain** why, for the replication of a virus, it is not necessarily essential for the protein capsid of the virus to enter the host cell.

Microorganisms and the environment

F2.1–2.8

Microorganisms typically have significant impacts on their environment, partly because of their numbers but also due to their forms of nutrition and metabolism. As a consequence, an alternative means of classifying them can be built on the basis of their roles within ecosystems – in effect, on their environmental impacts.

On this basis we can recognise three categories; **producers**, **nitrogen fixers** and **decomposers**. This division is based on their metabolism and nutrition, as defined in Table 18.3.

Table 18.3 Roles of microorganisms in the environment

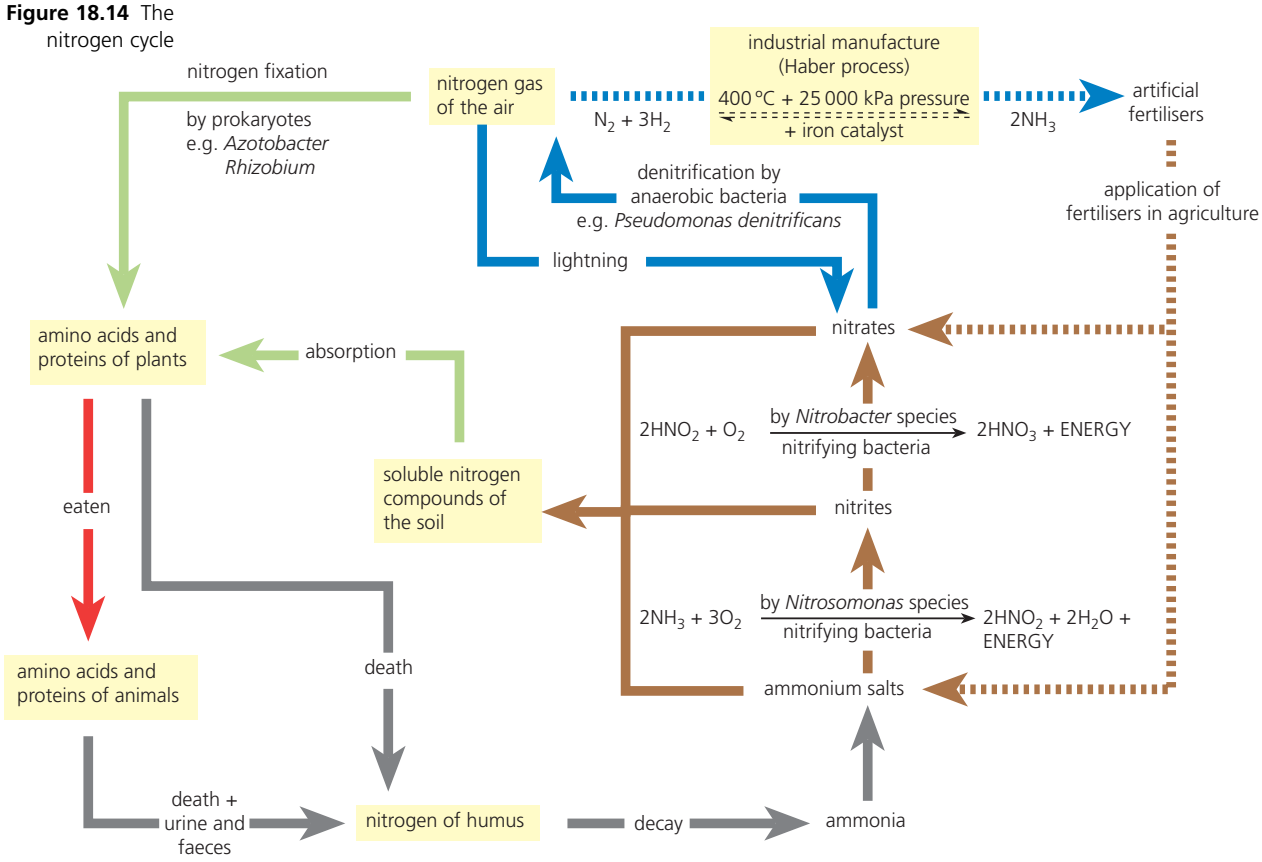
	Examples / environmental effect
Producers	■ photosynthetic bacteria such as cyanobacteria ■ produce sugar and oxygen from CO_2, using light energy
Nitrogen fixers	■ free-living and symbiotic N-fixing bacteria ■ convert nitrogen gas from the atmosphere to ammonia using energy from sunlight (free-living N-fixers) or energy from sugar obtained from host (symbiotic N-fixers)
Decomposers	■ saprotrophic bacteria (and fungi) ■ break down dead organic matter and waste matter from organisms to CO_2, H_2O and ions

So, when we examine a significant environmental process such as the nitrogen cycle, for example, we can detect microorganisms of these three categories present (Figure 18.14).

The nitrogen cycle

Although the element nitrogen in the form of di-nitrogen molecules (N_2) makes up about 80% of the Earth's atmosphere, nitrogen compounds available for use by living things are relatively scarce within the biosphere. The nitrogen cycle summarises the all-important steps by which the element nitrogen cycles between the soil, the atmosphere and living things.

Figure 18.14 The nitrogen cycle



Microorganisms of the nitrogen cycle

Nitrogen-fixing microorganisms

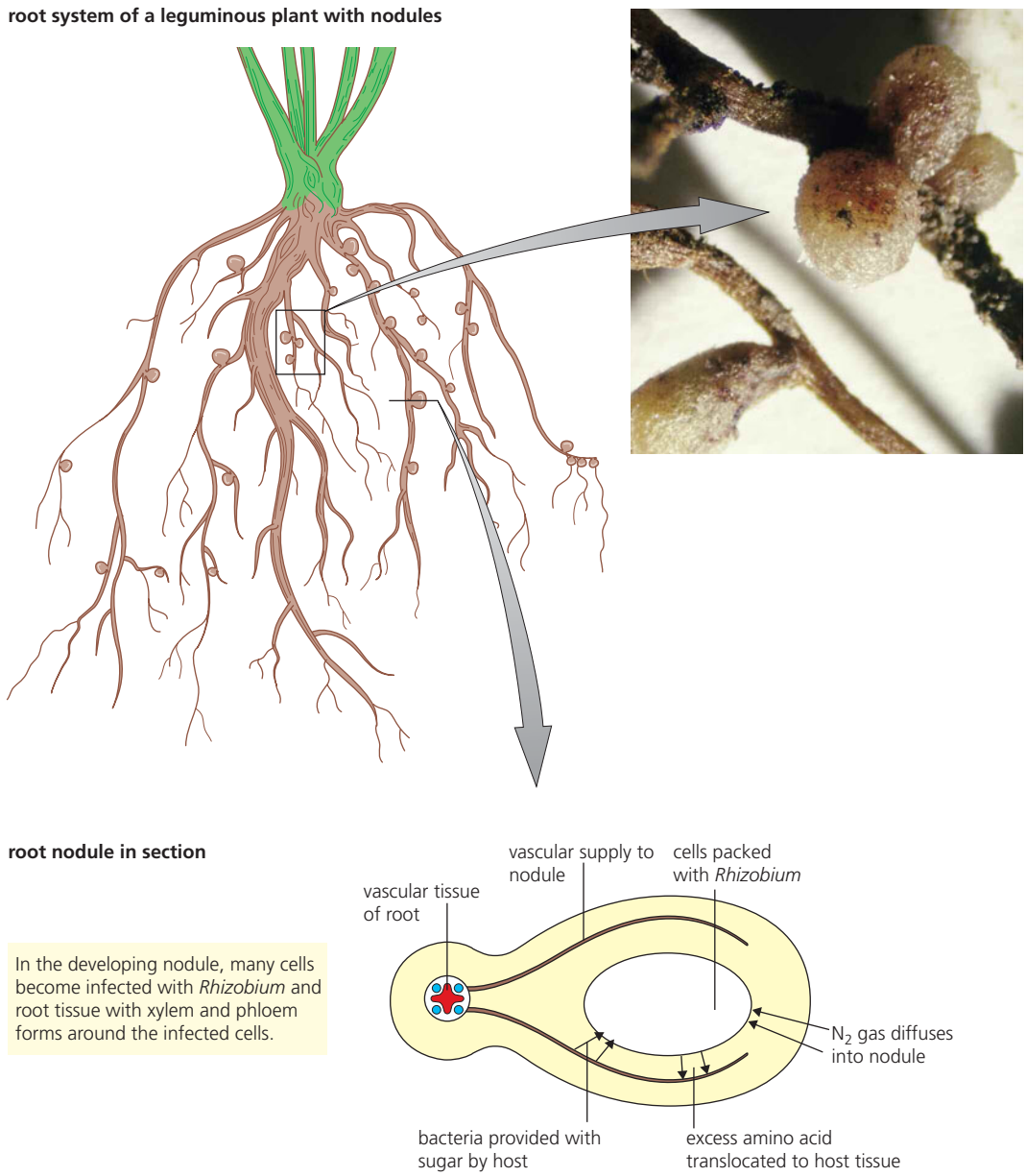
Some microorganisms can take in nitrogen from the air and fix this atmospheric nitrogen (they chemically reduce it) to form ammonia. The ammonia is then combined with organic acids, forming amino acids. This is known as **nitrogen fixation.**

Nitrogen fixation requires energy (as ATP) and hydrogen (reducing power from $NADH_2$) from respiration; they need a great deal of both for this reduction. The enzyme involved is nitrogenase. The steps of N-fixation are summarised below.

1 nitrogen fixation **2** amino acid synthesis

$$N_2 \xrightarrow[\text{ATP + NADH}_2\text{ from respiration}]{\text{nitrogenase}} NH_3 \xrightarrow{\text{organic acid}} \text{amino acid}$$

Figure 18.15 Root system with root nodules, and the functioning nodule



Some nitrogen-fixing organisms are **free-living**, meaning they exist independently in the environment. Examples include certain species of bacteria, such as *Azotobacter* spp. In *Azotobacter*, the energy (ATP) and reducing power ($NADH_2$) are supplied by aerobic respiration in the bacterium.

Some nitrogen-fixing organisms live **mutualistically in host plants**. Mutualism is the name for a close association between two organisms, in which both organisms benefit. The bacterium *Rhizobium* commonly occurs in the soil around plant roots where it feeds saprotrophically on dead organic matter. But it is also able to enter the roots of leguminous plants (clover, peas, beans, soya and many others), if growing nearby. Here it causes the host tissues to form into a nodule around the cells containing the bacterium (Figure 18.15). In this environment, the *Rhizobium* is able to reduce nitrogen gas to ammonia. From this, amino acids are immediately formed in the nodules and some, possibly most, pass out to the surrounding cells and are used by the host plant.

4 Identify the source of reducing power and energy used by *Rhizobium* in nitrogen fixation.

Because of their association with *Rhizobium*, leguminous plants can grow and flourish in soils low in nitrates. As crop plants they are found to be rich in proteins. Consequently, leguminous crops are vitally important to communities otherwise suffering from shortage of protein in their diets, such as those living by subsistence farming on very poor soils. In fact, rich and poor communities alike, all around the world, make good use of leguminous plants.

Formation of soil nitrates by nitrifying bacteria

A range of saprotrophic bacteria and fungi of the soil community break down organic matter in the soil (dead remains of organisms, waste matter excreted by animals, humus around soil particles, etc.). The final products of breakdown include carbon dioxide, water and a range of metal (e.g. Ca^{2+}, K^+) and non-metal ions (e.g. PO_4^{3-}). Vast numbers of microorganisms exhibit this type of nutrition.

Meanwhile, the combined nitrogen present is reduced to ammonia (NH_3). This is known as **ammonification**.

Ammonia occurs as the ammonium ion (NH_4^+), dissolved in the soil solution, particularly in acidic and neutral soils. Because ammonia is volatile, some loss can occur from soil, especially in basic soils (e.g. soil formed from chalk rock).

However, losses of ammonia and ammonium ions from soils is prevented when rapid **nitrification** (oxidation of NH_4^+ to NO_3^-), brought about by nitrifying bacteria, follows ammonifications.

The **first step to nitrification** is the oxidation of ammonium ions to nitrite ions. This is an aerobic process and an exothermic reaction. It occurs in the soil, carried out by enzymes of bacteria of the *Nitrosomonas* genus.

$$2NH_3 + 3O_2 \xrightarrow{\textit{Nitrosomonas}\text{ spp.}} 2HNO_2 + 2H_2O + \text{ENERGY}$$

The **second step to nitrification** is the oxidation of nitrite ions to nitrate ions. This also is an aerobic process and an exothermic reaction. It occurs in the soil, carried out by enzymes of bacteria of the *Nitrobacter* genus.

5 Identify how the combined nitrogen formed by *Azotobacter* and *Rhizobium* is made available to organisms growing in the same habitat.

$$2HNO_2 + O_2 \xrightarrow{\textit{Nitrobacter}\text{ spp.}} 2HNO_3 + \text{ENERGY}$$

Note that *Nitrobacter* and *Nitrosomonas* use the energy released from the chemical reactions they catalyse in their nutrition. The energy is coupled to the synthesis of sugar from carbon dioxide and water in a process known as chemosynthesis (chemoautotrophic nutrition, page 584).

Loss of nitrates from soil by bacteriological action – denitrification

Breakdown of soil nitrates to nitrogen gas may occur, for example, in waterlogged soils where the oxygen normally present has been driven out. The breakdown is due to the metabolism of certain anaerobic bacteria, particularly of the *Bacillus* and *Pseudomonas* genera (e.g. *Pseudomonas denitrificans*). These organisms flourish in anaerobic conditions. The effect of the anaerobic metabolism of these bacteria in soils is harmful to the growth of green plants because of this loss of nitrate from the soil.

What soil conditions favour nitrification and denitrification?

The bacteria of nitrification are strongly aerobic, and function best in soils in which air is present – the normal condition for most soils. However, if bad drainage conditions develop and water accumulates, then air is driven out of soil. We say the soil is waterlogged. In the absence of air, denitrification takes over.

Sewage treatment and the roles of microorganisms

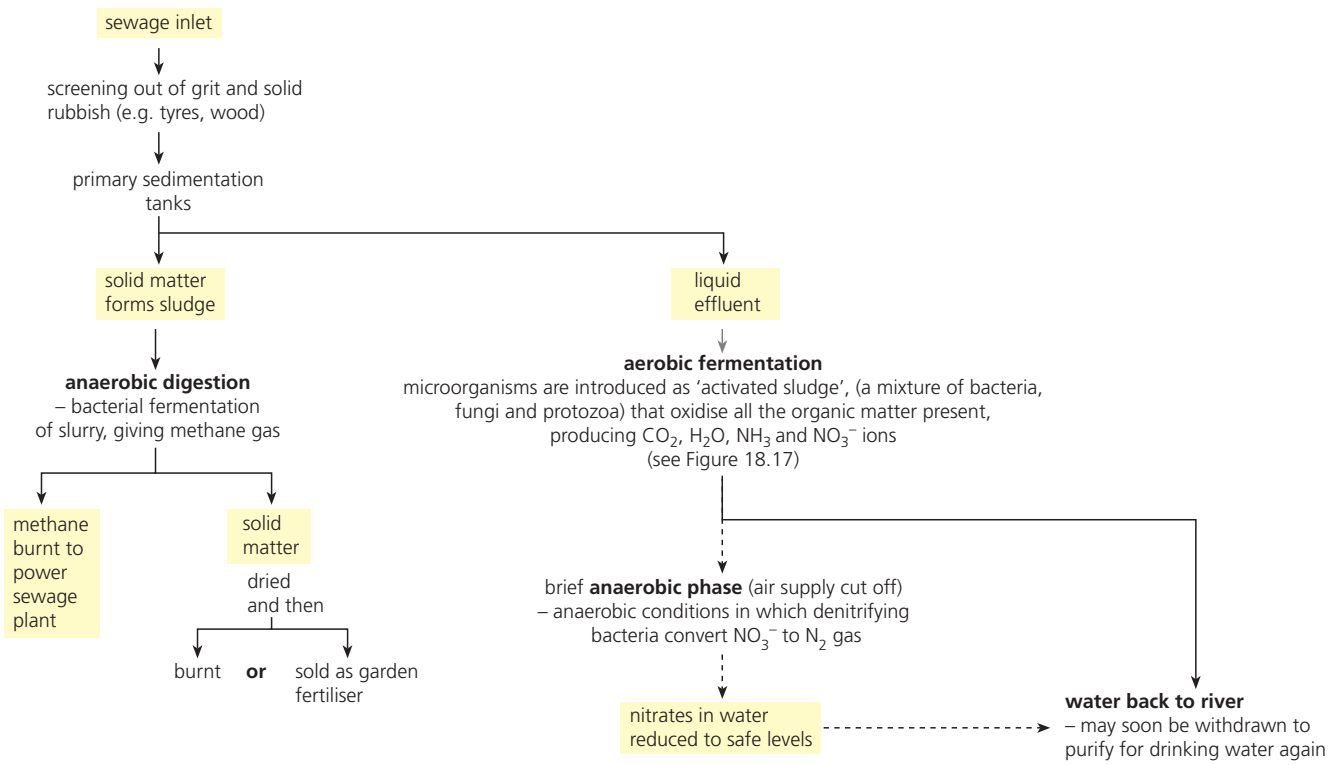
Sewage is the fluid waste of human communities, from houses and commercial properties. It is preferably piped to sewage works for processing. Sewage consists largely of used water, faeces and urine, and is rich in organic matter with ammonium, nitrate and other ions present. It contains vast numbers of microorganisms, many harmless saprotrophic ones, but also pathogenic bacteria.

Treatment of sewage to produce water safe to return to the environment involves the metabolism of vast numbers of microorganisms. During treatment (Figure 18.16), sewage is separated into **liquid effluent** and **solid matter**.

Liquids are cleaned by the aerobic metabolism of a mixture of bacteria, fungi and protozoa that is known in the industry as activated sludge. It requires vigorous efforts by water engineers to maintain the concentration of oxygen in the liquor throughout the process. There are various ways of doing this.

Solids are anaerobically digested to break down the organic matter. Liquid effluent is converted to clean water, and solid matter is converted to methane and carbon dioxide, and to solids usable as soil fertiliser (or that are disposed of by burning).

Figure 18.16 Sewage treatment



Roles of saprotrophic microorganisms in sewage

Provided aerobic conditions are maintained, saprotrophic microorganisms of the activated sludge quickly digest the entire dissolved organic matter present, producing carbon dioxide, ammonia, nitrate ions and other ions, including phosphates – and water. The water can be returned to the river largely without harm to the flora and fauna present, and may subsequently be withdrawn for re-use.

However, there is a persisting problem in the presence (and steady build up) of dissolved ions, such as nitrates and phosphates, as water resources are used again and again. We return to this issue later.

How is sufficient dissolved oxygen provided, given that oxygen gas is only slightly soluble in water, and there is a huge demand for oxygen by the saprotrophs that flourish in sewage effluent?

Dissolved oxygen is maintained in the liquid effluent by:

- **pumping air** through the liquid effluent as it passes slowly through shallow open tanks (Figure 18.17);
- **spraying the effluent** over beds of solid, porous clinker, and continually re-circulating the liquor (Figure 18.17);
- **passing the effluent through beds of reeds** selected to release air into the water around their roots (Figure 18.17).

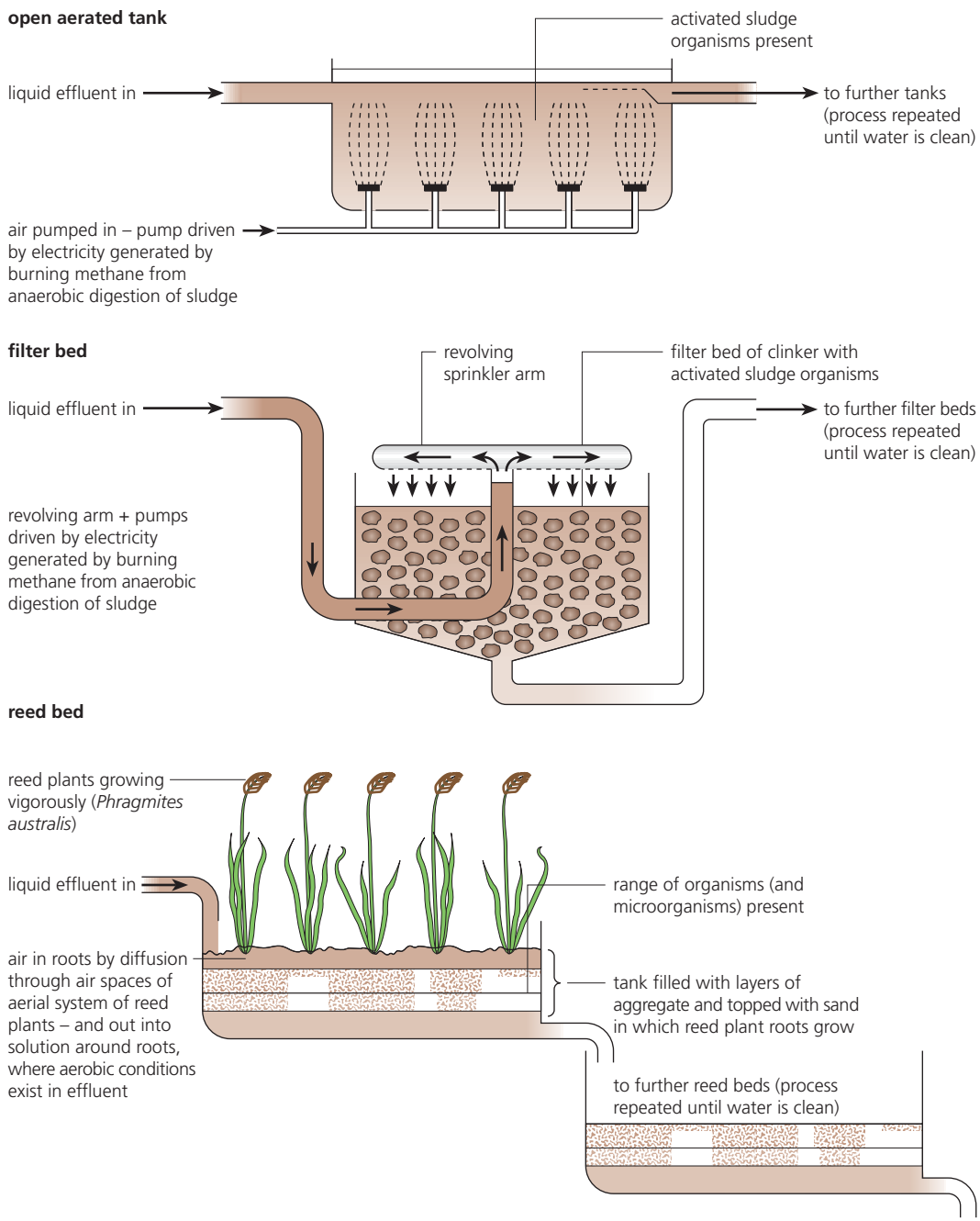


Figure 18.17 Alternative (aerobic) treatments of liquid effluent

The reed plant used in the latter case is a species of *Phragmites*, which maintains strongly aerobic conditions around the roots because air diffuses through the air spaces of leaves, stem and roots. Thus, aerobic conditions are maintained in the sand and effluent where aerobic, saprotrophic microorganisms are actively breaking down the organic matter.

Because the reed beds involve higher plants, they have the advantage of taking up ions favourable for reed growth, such as nitrates and phosphates. This reduction of ions in the water is not achieved by other methods. However, reed beds take up a huge space compared with other methods, and space is often at a premium in large cities where many people live, and where huge quantities of sewage need processing.

Sewage contamination of river water

A river that becomes contaminated with raw sewage is an environmental disaster (Figure 18.18). This is because the water is made excessively rich in organic matter at the point of contamination. Initially, saprotrophic bacteria flourish and sewage starts to break down – much as it does in a sewage works. However, in natural river conditions only limited oxygen is dissolved in the water, and there is no mechanism to pump oxygen in and maintain aerobic conditions. The explosion in the populations of aerobic bacteria that the organic matter of the sewage triggers, leads quickly to anaerobic (anoxic) conditions.

The tendency of water contaminated with organic matter to take up oxygen is called its **biochemical oxygen demand (BOD)**. BOD is defined as the amount of oxygen taken up by water over five days when incubated in the dark at 20 °C. Clean water takes up less than 5 ppm; heavily polluted water, 10 ppm or more.

The complete absence of any dissolved oxygen in river water quickly kills obligatorily aerobic organisms present in the river immediately downstream of the pollution, including the fish. As a consequence, **anaerobic bacteria flourish**. A waste product of the anaerobic decay of all the organic matter in these conditions is the unpleasant smelling gas **hydrogen sulphide**. This highly soluble gas is extremely poisonous. The remaining flora and fauna are also killed.

Slowly, the gases escape into the atmosphere, other pollutants are diluted by the river, and are metabolised by a range of organisms able to live at low oxygen concentrations. Eventually, **the river community will recover**.

Sewage is likely to have contained **pathogenic bacteria** and their spores, which may be present in the faeces. There is, therefore, an additional health hazard generated if the river is later used for bathing, or as a source of drinking water for communities downstream. Treatment with chlorine is essential in the treatment of river water for drinking purposes if public health is to be maintained.

Agricultural effluent and water-way pollution

Another common source of river water pollution is agricultural run-off from arable land that is treated with fertilisers or manure. The steps taken to maintain soil fertility under heavy cropping regimes inevitably result in a steady build up of ions such as nitrates (and phosphates) in water of the water table. Slowly, these leach into ditches, streams and rivers as rain water drains from the fields.

In water enriched with these inorganic ions, plant growth is normally luxuriant. In the growing season, water temperatures rise and then the algae of lakes and rivers undergo a population explosion, known as an **algal bloom**. The enrichment of waters with inorganic nutrients has caused an excess of aquatic plant life. This process is known as **eutrophication**. It is not restricted to rivers; the seas around estuaries are often similarly affected.

Later, after the algal bloom has died back, the organic remains of these plants are decayed by saprotrophic aerobic bacteria. As a result, the water becomes deoxygenated, exactly as when raw sewage is discharged into a river (Figure 18.18). The resulting anaerobic conditions, and the hydrogen sulphide produced, kill many aquatic organisms.

Note that in the absence of pollution, eutrophication still occurs as a natural process, but only very slowly. In rivers and lowland ponds and lakes, the amount of dissolved nutrients steadily increases. By contrast, the water of mountain lakes or tarns, and the streams that feed them, are typically low in dissolved nutrients. Here the surrounding land is often stony, with poor soil. The result is few dissolved ions are present and aquatic plants show little growth.

Figure 18.18 The effect of sewage pollution of river water

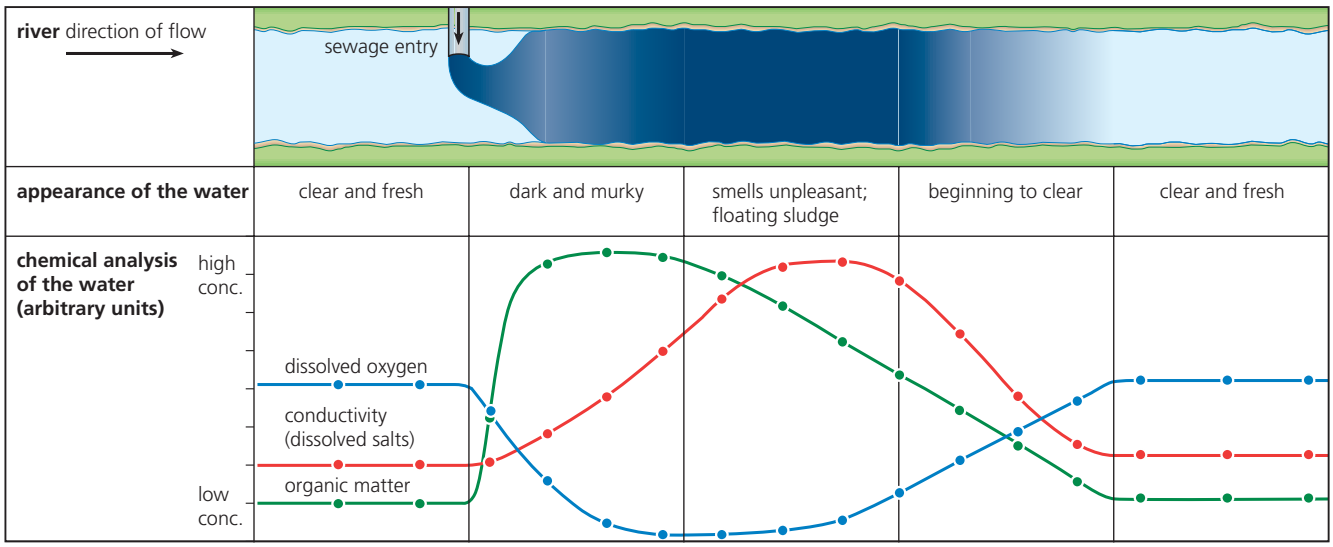
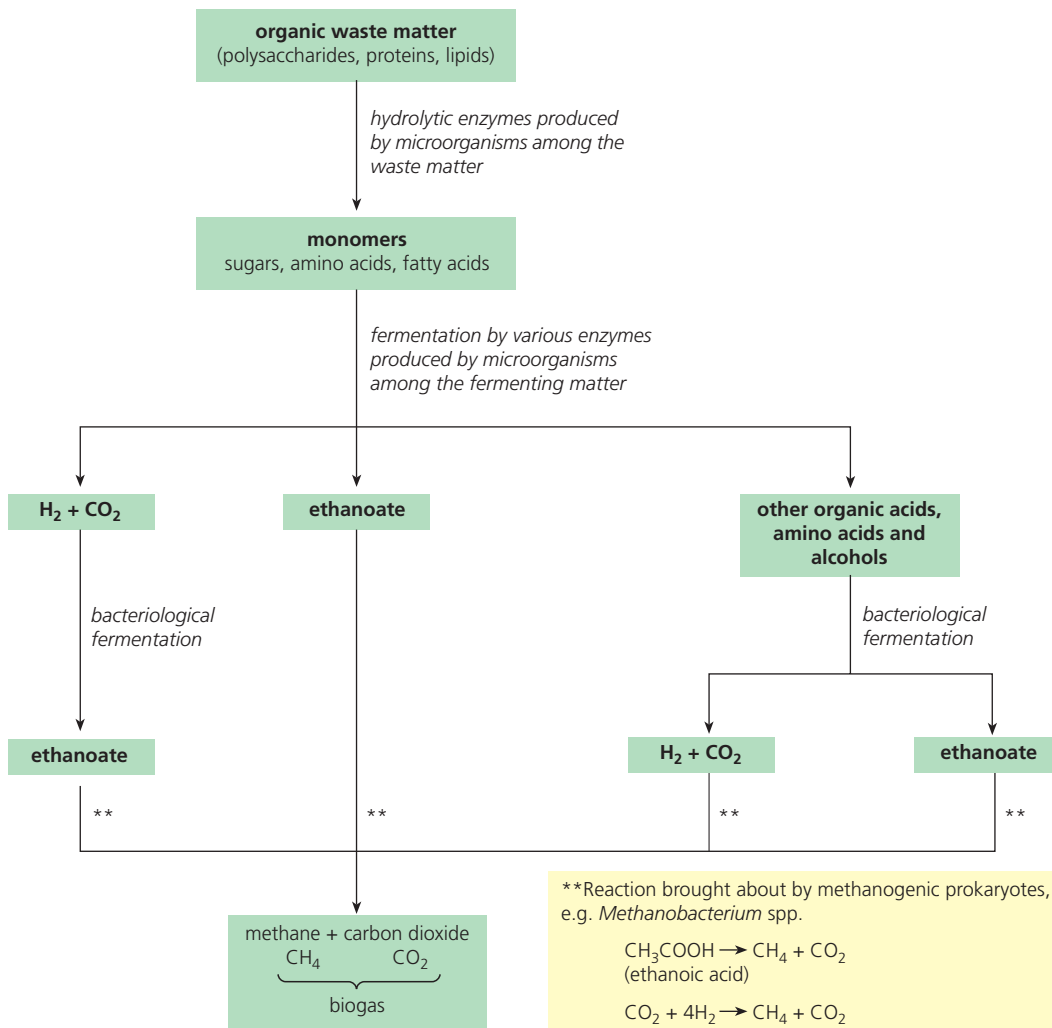


Figure 18.19 Pathways of methane synthesis from organic waste matter



Fuels from biomass

Oil, gas and coal are known as fossil fuels because these resources were laid down long ago, during the Carboniferous period of Earth's history. Dependence on fossil fuels generates problems because:

- they are non-renewable sources of energy – stocks will eventually run out;
- burning of fossil fuels releases into the atmosphere carbon dioxide gas that has been locked away for over 300 million years – consequently, the concentration of carbon dioxide (one of the greenhouse gases) in the Earth's atmosphere is rising (page 149).

As a result, there are potentially good economic and environmental reasons for developing alternative fuels from organic matter; such sources are referred to as biomass.

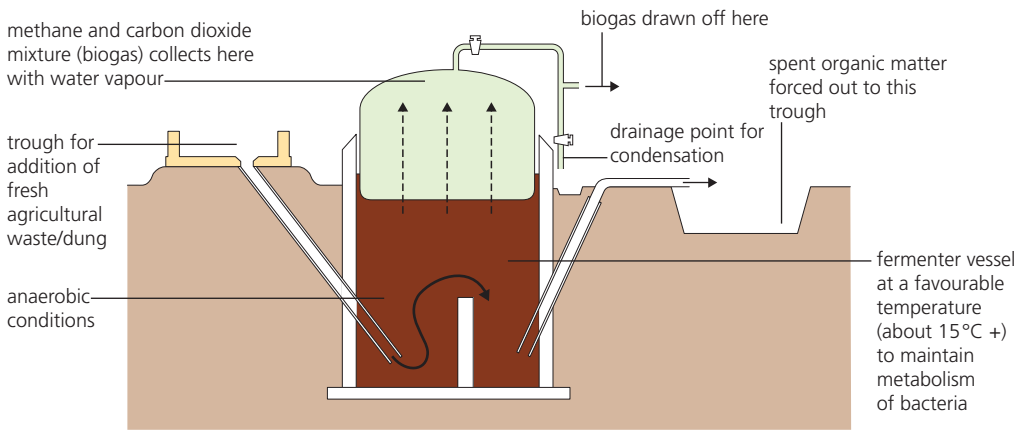
One approach involves **energy farming**, the name given to the growing of crops like sugar cane specifically for fuel production. The extracted sugar is fermented using yeast, and the resulting aqueous **ethanol** solution has to be distilled to a solution pure enough (only 4.4% water) to power cars adapted to burn ethanol. In Brazil, this fuel is called Alcool.

Alternatively, organic waste matter may be **anaerobically digested** by a combination of microorganisms to produce a **methane-rich gas**, often referred to as biogas (Figure 18.19). We have already seen that such biogas is a natural product of anaerobic digestion of sewage sludge in fermenter tanks at sewage works, where it has been used to generate electricity to power the machinery.

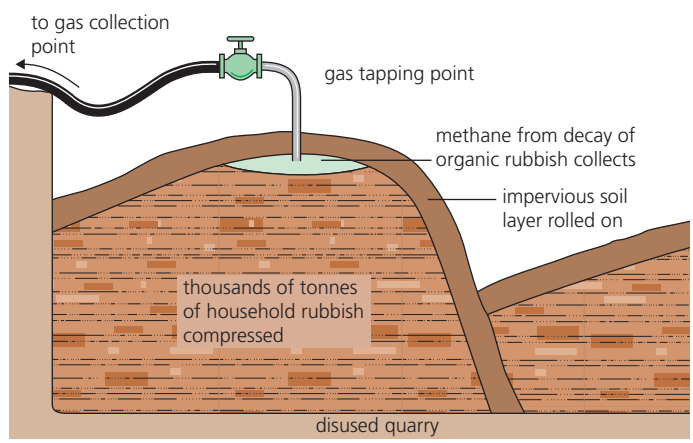
The pathway of methane synthesis from organic waste matter involves hydrolysis of the polysaccharides, proteins and lipids present to their monomers (sugars, amino acids, fatty acids and alcohols), and the fermentation of these to form organic acids such as ethanoate (acetate).

Figure 18.20 Biogas production in contrasting circumstances

a developing world biogas plant



biogas from an urban land-fill site



Then, methanogenic prokaryotes such as *Methanobacterium* spp. convert ethanoic acid and ethanoate (acetate) to methane and carbon dioxide:

$$CH_3COOH \longrightarrow CH_4 + CO_2$$

Methane gas from waste matter in the less-developed world

In some rural communities that are under-resourced with fuel of any sort, biogas digesters are in regular use. Into the digester goes animal dung and/or agricultural waste for anaerobic fermentation. Simple digesters operate more reliably in climates where the contents of the digester will be at or above a temperature of 15 °C. The biogas formed is about 50–80% methane, 15–45% carbon dioxide and about 5% water vapour (Figure 18.20). (This compares reasonably well with natural gas from oil and gas fields, which is about 80% methane.)

6 **Outline** the range of microorganisms involved in the conversion of complex organic waste into molecules which methanogenic prokaryotes may exploit as substrates.

Methane gas from waste matter in the developed world

Urban communities in the developed world typically dump huge quantities of domestic waste, much of which is organic matter, into disused pits or quarries called landfill sites. The rubbish becomes compressed, anaerobic conditions quickly develop, and a cocktail of microorganisms break down the organic matter, giving off biogas. Completed landfill sites have to be vented so that these gases can safely escape. Schemes now operate to tap the biogas and pipe it to where it can be burnt to generate heat and/or power, wherever this is practical (Figure 18.20).

Microbes and biotechnology

F3.1–3.5

Biotechnology is the industrial and commercial application of biological science, particularly of microbiology, enzymology and genetics. Modern biotechnologies are enormously important to industry, medicine and research, and these technologies are often presented as a recent development. In fact, their origins go back several millennia; the earliest biotechnologies were cheese making and alcoholic fermentation such as brewing, for example. However, it is certain modern applications of biotechnology that are illustrated here.

Genetic engineering involves change in the genetic constitution of a living organism brought about by artificial means (meaning other than by conventional breeding). In genetic engineering procedures, individual genes in the form of fragments of DNA from cells of an organism may be isolated and then spliced into the genome of other organisms (page 121). Very often, but not exclusively, the gene is added to a bacterium. The result is the **genetic modification of organisms**.

This technology and the diverse techniques it has already spawned have many applications, including:

- the manufacture of drugs such as human insulin from bacteria;
- the production of genetically modified, fast-growing tree species yielding soft fibres for industrial paper production with lowered environmental harm;
- the production of genetically modified sheep that secrete human proteins required in the treatment of disease;
- the possible repair of faulty genes by the techniques of gene therapy;
- the identification of a person (or other organism) from a sample of their DNA by a process known as genetic fingerprinting.

In this revisiting of the subject of genetic engineering, there are two particular aspects to consider: the genetic engineer's use of reverse transcriptase enzyme, and the issue of gene therapy. However, we should first acknowledge that sometimes there is only a very small sample of DNA available to the genetic engineer, perhaps too little to progress with. The discovery of the **polymerase chain reaction (PCR)** has solved this. By means of the PCR, fragments of DNA may be copied repeatedly, faithfully and speedily, in a process that has been fully automated. Details of the steps of the PCR are not needed here.

Reverse transcriptase

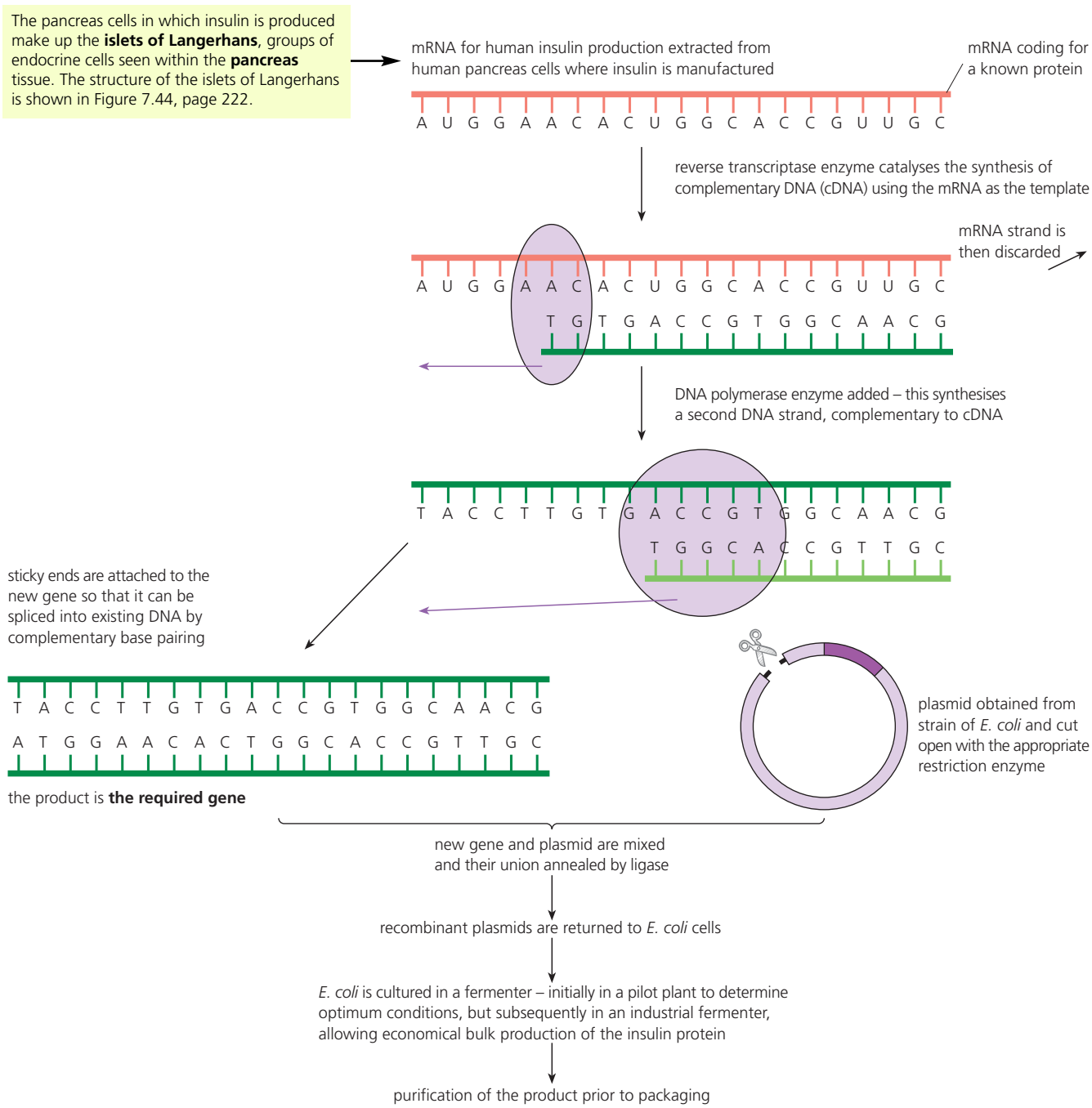
We have seen that the normal flow of genetic information is a one-way flow, from the DNA of the genes to messenger RNA (mRNA) in the cytoplasm. That is, the genetic code in DNA is transcribed into messenger RNA which passes out into the cytoplasm and is translated into the sequence of amino acids by which a protein is assembled in the ribosomes. This concept is called the **central dogma of molecular biology** (page 252).

Since Watson and Crick made this original observation, retroviruses have been discovered.

Retroviruses have RNA as their genetic material rather than DNA, yet they are able to translate the information in their RNA into double-stranded DNA. In this latter form, their genes may be added to the (eukaryotic) host's DNA, in the nucleus. The **AIDS virus is an example of a retrovirus** (page 253).

Figure 18.21 Insulin production by genetically modified (GM) bacteria

The pancreas cells in which insulin is produced make up the **islets of Langerhans**, groups of endocrine cells seen within the **pancreas** tissue. The structure of the islets of Langerhans is shown in Figure 7.44, page 222.



This conversion is possible because the virus contains an enzyme, known as **reverse transcriptase**, which it introduces into the host's cell, at the time of the initial infection. Reverse transcriptase catalyses the production of **DNA from RNA**. Today, reverse transcriptase enzyme is an essential part of the genetic engineer's tool-kit.

Why is reverse transcriptase such a useful enzyme in genetic engineering?

A eukaryotic gene, in place in a chromosome, typically consists of meaningful DNA (called **exons**) interrupted periodically along its length by apparently meaningless lengths of DNA (called **introns**). The phenomenon of exons and introns was introduced on page 247.

Apparently, most if not all eukaryotic genes have these units of non-gene DNA within their boundaries. Remember, when such a gene is transcribed into mRNA, the mRNA itself initially contains the sequence of introns, exactly as they occur in the DNA.

In the next stage of **post-transcriptional modification**, the introns are removed and disposed of. This occurs before the mRNA passes out into the cytoplasm. The resulting (shortened) length of mRNA is described as **mature**.

The presence of exons is a challenge to genetic engineers who want to isolate a eukaryotic gene because they have no easy way of cutting out the introns. Now there is an alternative approach, avoiding the problem of introns altogether. This involves **making a copy of the gene from its mRNA**, using reverse transcriptase.

For example, mRNA coding for the human hormone **insulin** can be extracted from the cells of the pancreas (page 222). Then, using reverse transcriptase, a DNA copy of the gene for insulin can be produced – a copy of the gene without introns. When this gene is then spliced into DNA that can be introduced into a bacterium, such as an appropriate strain of *E. coli*, then the bacterium can be induced to synthesise human insulin (Figure 18.21).

This is the actual mechanism by which human insulin is synthesised by the pharmaceutical industry, today. When adding genes to a bacterium, the addition is made to plasmids (page 122) which the bacterium can be induced to take up.

7 Comment on the challenge to Mendelian genetics of the existence of reverse transcriptase.

Gene therapy

Genetic diseases are heritable conditions that are caused by a specific defect in a gene. This type of disease affects about 1–2% of the human population. Common genetic diseases include cystic fibrosis, sickle cell disease (page 101), Duchenne muscular dystrophy, thalassaemia, severe combined immunodeficiency (SCID), familial hypercholesterolaemia, and haemophilia (page 114).

These conditions mostly arise from a mutation involving a single gene. The mutant allele that causes the disease is commonly recessive, so a person must be homozygous for the mutant gene for the condition to be expressed (but people with a single mutant allele are carriers). For example, cystic fibrosis is due to a recessive gene on chromosome 7. Duchenne muscular dystrophy and haemophilia are due to recessive alleles on the X chromosome, so these two conditions are sex-linked (page 113).

Somatic and germ-line therapy

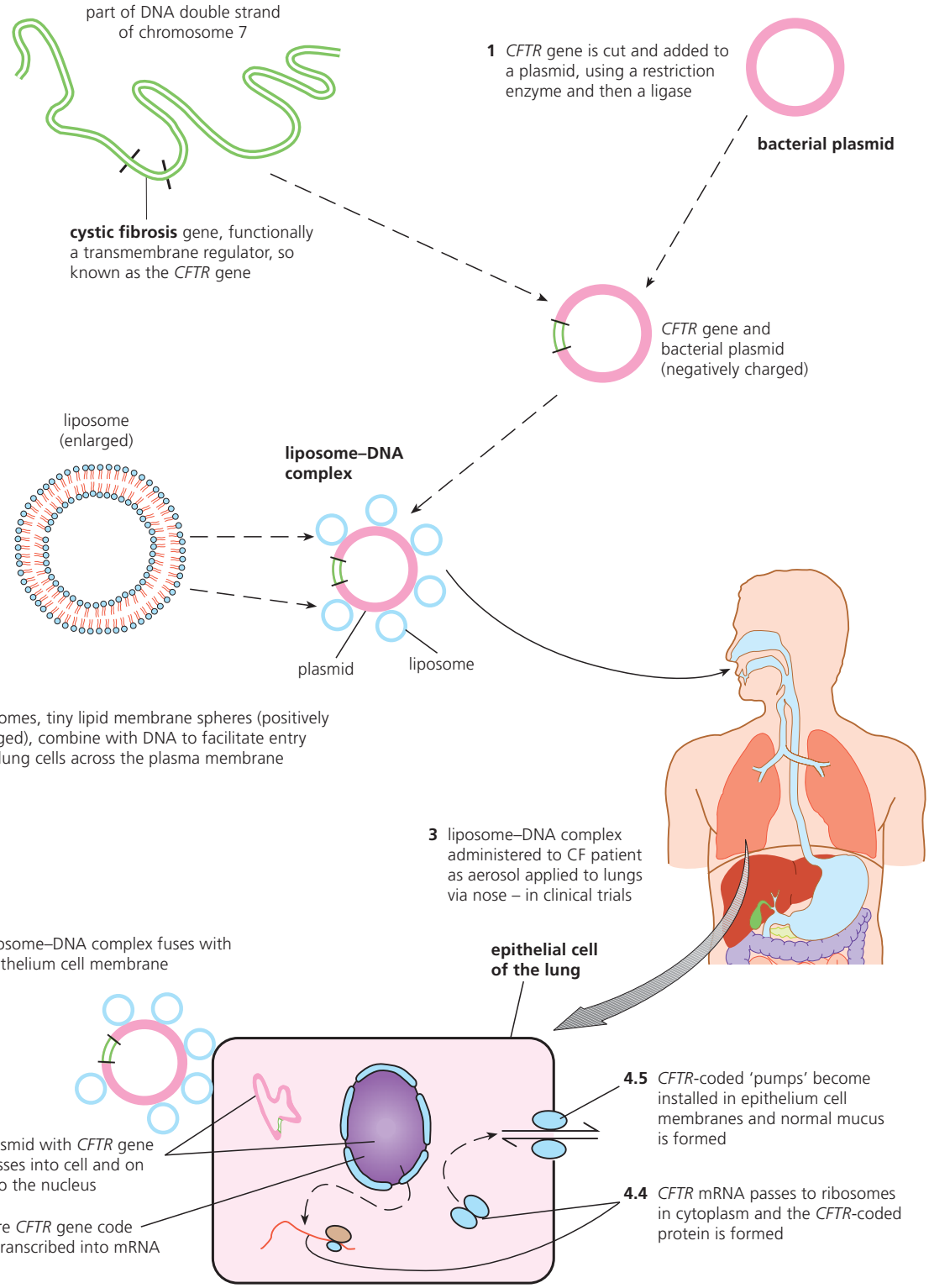
Gene therapy aspires to use recombinant DNA technology to overcome genetic disease, where this is thought safe and ethically sound. A permanent and acceptable solution is to supply the missing gene to body cells in such a way that it remains permanently functional. This solution is known as **somatic therapy** (Figure 18.22 and 18.23).

On the other hand, it is not considered safe or ethical to attempt to tamper with germ cells (cells which give rise to gametes in the testes and ovaries). This banned approach is called germ-line therapy.

8 Comment on the observation that medical treatments with expensive drugs, skilled surgery, or by gene therapy receive much attention in our media, but virtually no consideration is given to the possibility of directing funds to everyday measures to improve health and prevent disease in the first place.

The cystic fibrosis gene codes for a membrane protein that occurs widely in body cells, and pumps ions (e.g. Cl^-) across cell membranes.

Figure 18.22 Cystic fibrosis (CF); getting the healthy gene into functioning cells in the lungs



In recent clinical trials some 20% of epithelium cells of CF patients were *temporarily* modified (i.e. accepted the *CFTR* gene), but the effects were relatively short-lived. This is because our epithelium cells are continually replaced at a steady rate, and in CF patients the genetically engineered cells are replaced with cells without *CFTR*-coded pumps. Patients would require periodic treatment with the liposome–DNA complex aerosol to maintain the effect permanently.

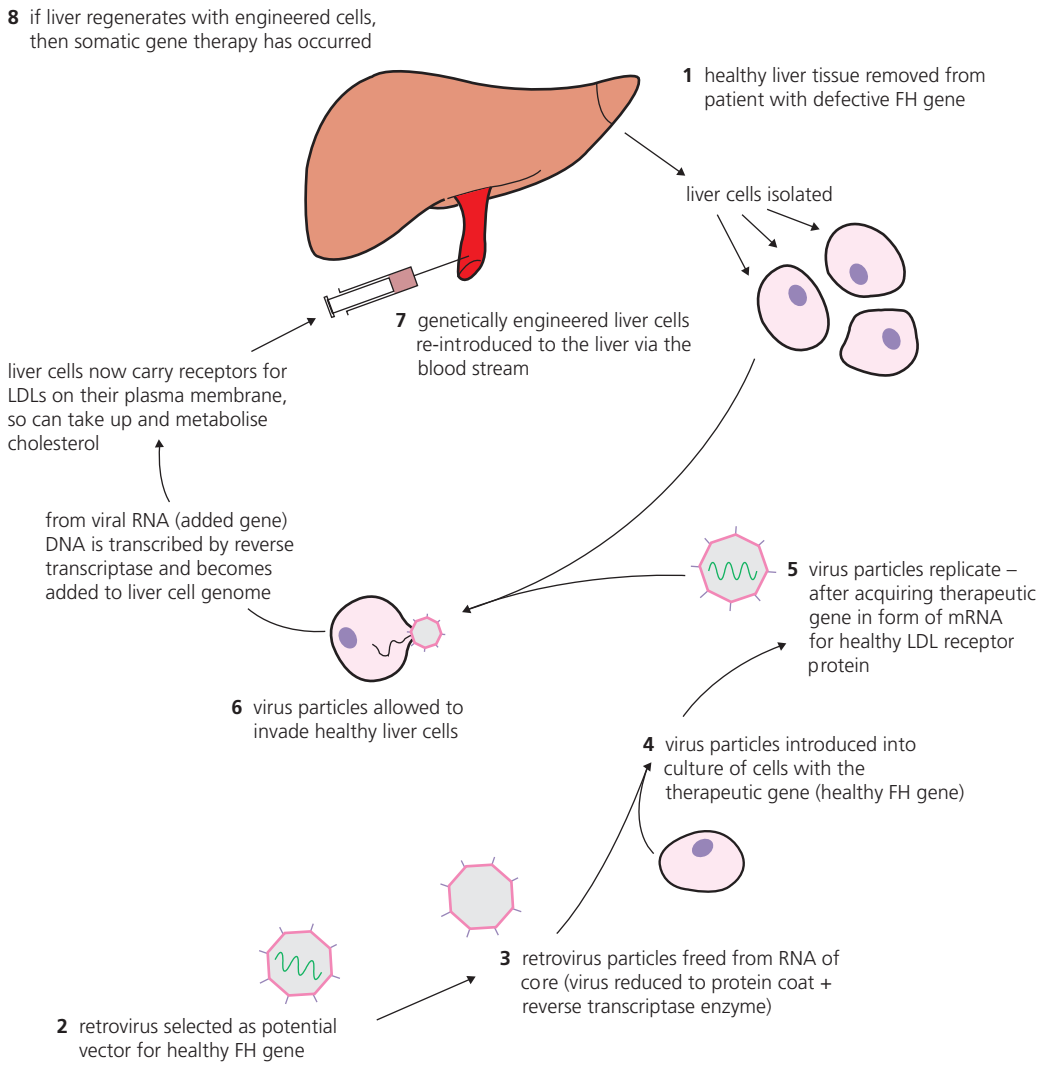
Two approaches to somatic gene therapy

Cystic fibrosis affects the epithelial cells of the body and is the most common genetic disorder in Northern Europe. The **CF** gene codes for a protein (called CFTR) that functions as an ion pump. The pump transports chloride ions across membranes and water follows. Consequently, epithelia are kept smooth and moist.

In cystic fibrosis patients, the mutant gene (**cfcf**) coding for this protein has a changed base sequence and either codes for no protein or for a faulty protein. As a result, the epithelia remain dry, and thick sticky mucus builds up. The effects are felt in the pancreas (secretion of digestive juices) and the sweat glands (salty sweat formed), for example. But for people with cystic fibrosis, a life-threatening consequence arises in the lungs which may become blocked by mucus and are prone to infection.

Figure 18.23 Familial hypercholesterolaemia (FH); correcting liver cell function by virus vector

Principle of this approach
Engineering a healthy gene into a sample of liver cells so that genetically corrected cells may be re-established in the liver and restore correct functioning.



Another possible therapy involves adding the missing genes by **injection of (harmless) retroviruses with engineered genes** directly into the patient's blood stream.

Gene therapy here involves getting copies of the healthy gene to the cells of the lung epithelia, delivered in an aerosol spray (Figure 18.22). The spray contains tiny lipid bilayer droplets called liposomes to which copies of the healthy gene are attached. Liposomes fuse with cell membrane lipid and deliver the gene to the epithelial cell. In trials, the treatment is effective, but only until the epithelial cells are routinely replaced. The treatment has to be regularly repeated. The cure can only be more permanent when it is targeted on the cells that make epithelial cells.

Familial hypercholesterolaemia (FH) is a failure in cholesterol metabolism in the body. Cholesterol travels in the blood in particles called low-density lipoproteins (LDLs), each containing about 1500 cholesterol molecules. A person with FH has defective receptors for LDLs on the membranes of cells that normally take in cholesterol and metabolise it (e.g. use it in membrane synthesis). As a result, people with FH have abnormally high concentrations of cholesterol in their blood – they may have yellow deposits of cholesterol under their skin as a result. FH causes early cardiovascular disease by causing the depositing of cholesterol in the walls of blood vessels, leading to atherosclerosis (page 420).

Gene therapy for FH involves engineering a healthy gene into a sample of liver cells so that the genetically corrected cells may be re-established in the liver and restore correct functioning. Here gene therapy has been attempted using a disabled **virus as the vector** (transporter) of the healthy gene (Figure 18.23). This is chosen because a virus may attach its nucleic acid to the host cell chromosome. If a healthy FH gene has been added to the viral genome, then the healthy FH gene may also become part of the cell's genome.

The risks of gene therapy

In 1990, two children who had SCID and were living in a sterile bubble in order to survive, were given a working copy of the gene they lacked which had been inserted into a sample of their own white cells. These people now lead normal healthy lives.

This success established that somatic gene therapy can be made to work. However, there are many failures, too, and the process carries risks for the patient. Risks include:

- the use of liposomes as the vector is inefficient because the gene cannot (as yet) be made to insert itself into chromosomes in the nucleus in cells; therefore, the treatment must be repeated at regular intervals and all medical treatments carry risk;
- disabled viruses are more efficient vectors of a healthy, working gene, but the virus coat, on entry, often triggers an immune response, causing tissue damage;
- retroviruses are the most efficient vectors so far attempted, but when they integrate a new gene into the patient's genome it may be at a position that inactivates other genes – for example, the effect of insertion may be to trigger the activation of proto-oncogenes with the result that a fatal cancer infection is caused at the same time that the inherited genetic disease is cured.

There will probably be other problems and disappointments on the way to the development of safe, reliable and effective treatments for inherited diseases.

9 The human disease SCID is also treated by genetic modification of extracted body cells, using a virus vector. When the corrected cells are returned to the body the patient may be cured for life (body cells are now able to produce the missing enzyme). **Explain** why this genetic modification (GM) treatment of SCID in a child patient, for example, is described as a case of somatic therapy only.

Microorganisms and food production

F4.1–4.4

The exploitation of microorganisms first occurred several millennia ago, in the processes of bread production, wine production and beer brewing, for example. These processes exploit the unicellular fungus *Saccharomyces* (yeast) (Figure 18.6, page 559).

Yeasts respire by alcoholic fermentation (page 80) and produce carbon dioxide and ethanol as waste products (waste in the sense that the *Saccharomyces* cells cannot use them further). The ethanol is exploited in brewing and wine making; the carbon dioxide is required to make leavened (raised) bread.

Brewing

Beer is made from barley grains (or rice or maize) which are first allowed to germinate. In this first step, the enzymes of the grains are used to convert the stored starch and proteins to sugars and amino acids. To the resulting sugar solution, yeast is added and produces alcohol (Figure 18.24).

Figure 18.24 Brewing of beer and lager

harvesting of barley

↓

malting of barley by inducing germination of grains (moisture + warmth); starch within the grains is converted to sugar, proteins to amino acids

↓

cracking – germination stopped by heating of grains to a light roasting; passed between rollers to open the germinated grains

↓

mashing with hot water to remove sugar and amino acids; remainder forms 'brewers grains', sold for animal feed

↓

boiling of the wort: additional sugar may be added – hops are added to boiling wort, adding flavour (bitterness) and anti-microbial substances

↓

cooling follows

↓

fermentation – yeast is added to convert the sugar to ethanol and CO_2

- *Saccharomyces cerevisiae* – a top-fermenting yeast (distributed in wort as fermentation proceeds, but carried to surface of the wort by CO_2 given off); used to produce **beers and ales** at temperatures of 14–23 C over a period of 5–7 days
- *S. carlsbergensis* – a bottom-fermenting yeast (settles to bottom of tank); used to produce **lager** at a temperature of 6–12 C over a period of 8–14 days

↓

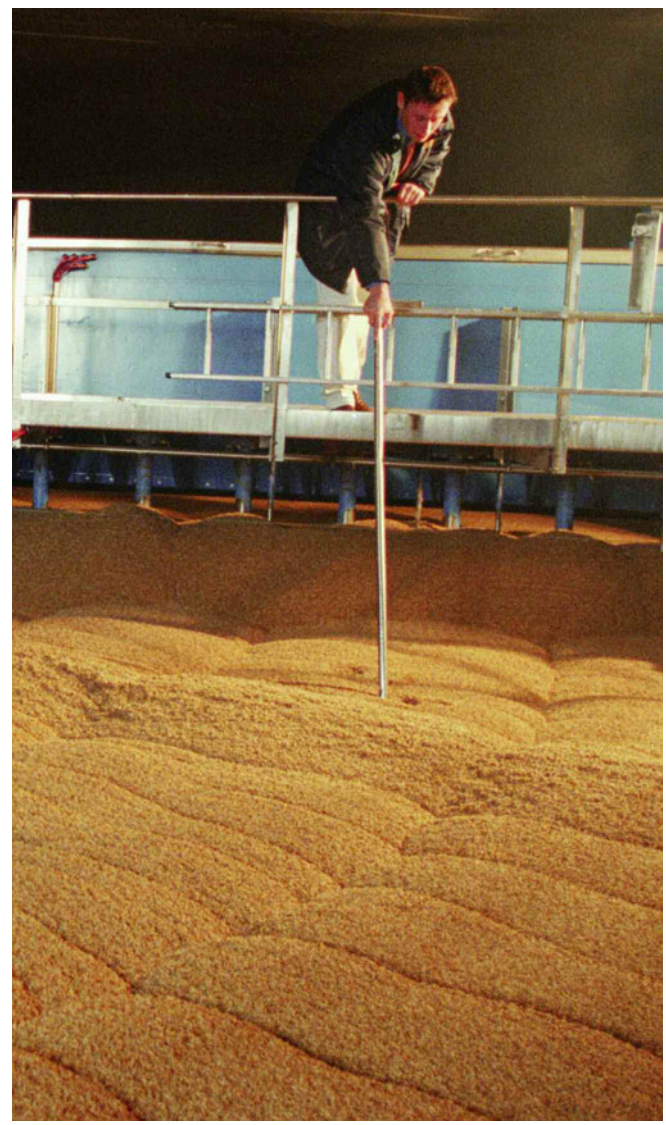
(by-products are CO_2, sold as dry ice, and yeast for yeast extract manufacture)

↓

conditioning of the beer, stored in casks; yeast settles out, beer cleared by filtration; standardisation of colour and flavour

↓

bottled or sold in barrels – a solution of ethanol at 3.5–8.0% ethanol (10% in barley wine); some beers are pasteurised



the fermentation stage in the brewing of beer, using a top-fermenting yeast

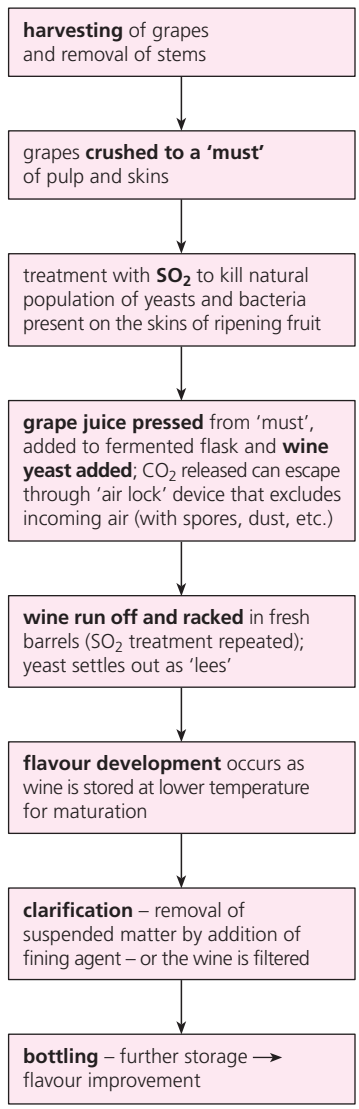
10 **Explain** why grain reserves of starch and protein must be hydrolysed by the grain's enzymes, prior to the addition of yeast.

Flavours of particular brews are due to the ingredients selected (e.g. the malt, water, hops and yeast), the proportions used, and the final cask conditioning of the live beer allowed by different brewers. The difference between lagers and beers is due to the yeasts used. Market variety is also due to the industrial development of keg beers – processed, pasteurised beers, served using nitrogen and carbon dioxide gas mixtures. These keg beers are viewed unfavourably by real ale enthusiasts.

Wine making

Wine production begins with the harvesting of ripe grapes, followed by crushing of the fruit to produce a must (a slurry of pulp, skins and pips). Today, the must is treated with sulphur dioxide to kill the flora of wild yeasts and bacteria that lived on the skins of the ripening fruit. Then a chosen strain of yeast is added to ferment the sugar. There are many variations to the basic process of wine production (Figure 18.25) by which the distinctive flavours and aromas (known as bouquet) of different wines are produced.

Figure 18.25 Red and white wine production



White wine is made from white grapes (or juice only from red grapes).
Red wine is made from red grapes – the 'must' (with skins) is fermented.

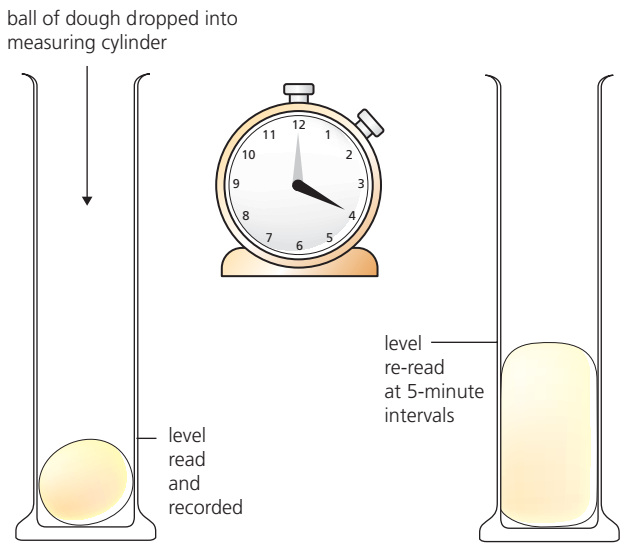
Bread making

Figure 18.26 Yeast and the rising of dough

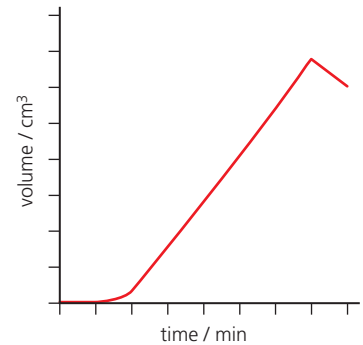
Dough is a mixture of flour, sugar, water and bakers' yeast. Worked into a ball and then kept in a warm place, the dough expands. This is due to the CO_2 released from the yeast respiration and trapped in the dough.

Dough mixture:
- 1 part bakers' yeast
- 1 part sucrose
- 6 parts flour
- 4 parts water.

Work into a dough.

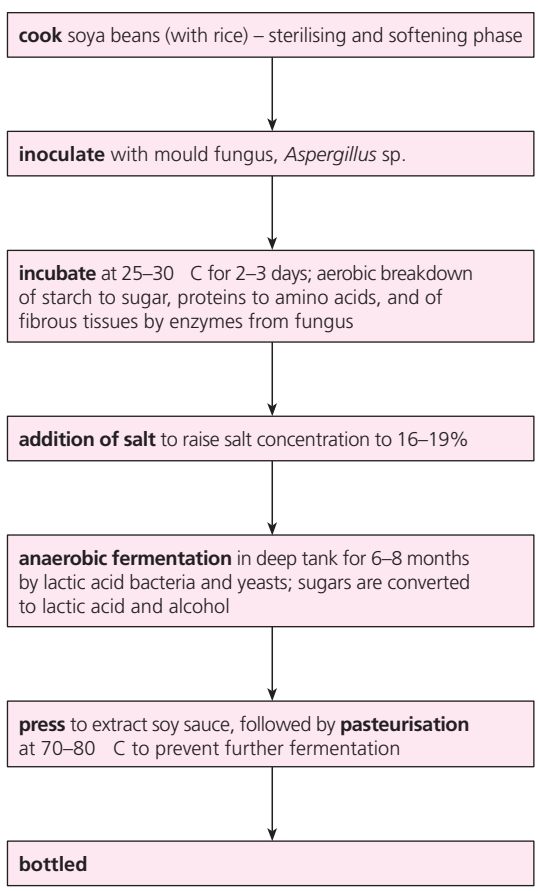


change in volume of dough against time



What would be the effect of:
- using different sugars?
- varying the temperature (measuring cylinder clamped in water bath)?
- using yeast killed by heating?

Figure 18.27 Production of soy sauce



Bread is made from flour and water, and smaller proportions of sugar, salt, yeast, vegetable fat, emulsifier and vitamin C. By baking, bread becomes a solid foam, having many tiny pockets of carbon dioxide distributed throughout its structure. First, ingredients are mixed or kneaded. Then they are left to ferment, during which process, starch in the flour is largely turned to sugars, and fermentation of sugars is begun. Also, the proteins of flour, known as gluten, become hydrated and form silky, elastic fibrils. Bubbles of carbon dioxide are retained in the dough, due to properties imparted by the hydrated gluten.

The dough is cut into loaf-sized pieces and allowed to ferment further. Subsequently the loaves are baked, during which all ethanol in the dough is driven out into the atmosphere. The effect of fermentation on the dough can be demonstrated and investigated experimentally (Figure 18.26).

Soy sauce production

Soya bean (*Glycine max*) is one of the oldest cultivated crops, originating in South East Asia. Soya is an example of a leguminous plant, housing *Rhizobium*

bacteria in root nodules. This has important consequences for the plant (and for those who feed on its seeds). Today it is grown very widely, and the beans are used mostly for oil, animal feed, and soya flour preparations for humans.

Various fermented foods have been derived from soya beans. Soy sauce is a food flavouring and colouring agent derived from soya beans by fermentation brought about by the common mould fungus *Aspergillus* sp. (Figure 18.27).

Food preservation

Microorganisms are ubiquitous, so their presence on food is inevitable. Consequently, if food items are not consumed as soon as they have been prepared, decay by microorganisms must be prevented in some way if the food is not to deteriorate. Figure 18.28 summarises possible approaches to food processing for preservation. For any particular type of food, the method of preservation chosen must maintain the original character (flavour and appearance), as well as nutritive value, as far as possible.

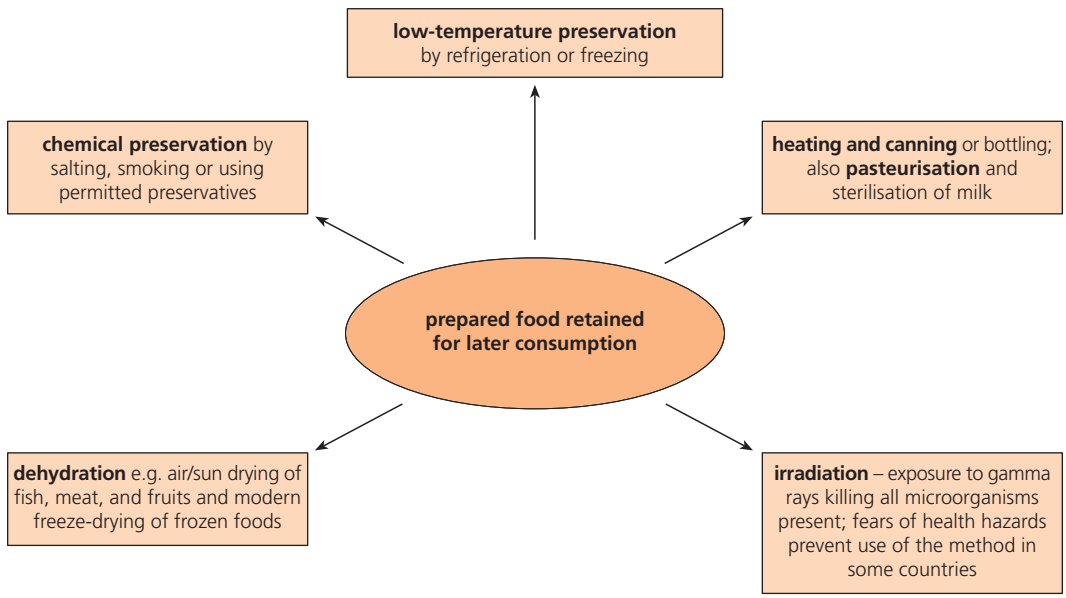


Figure 18.28 Techniques of food preservation

We can illustrate the process of food preservation by examining the uses of **salt**, **sugars** and **acids**.

Common salt, **sodium chloride**, is probably the oldest preservative, used especially with meat and fish. Traditionally, foods are soaked in a concentrated solution of rock salt. Sodium chloride (the major ingredient of the rock salt) dissociates to form sodium ions and chloride ions, and these ions strongly attract water molecules (page 27). Much more water diffuses from the cells and tissues than diffuses in and the food material is dehydrated. Any bacterium present (and any entering subsequently) is destroyed by dehydration, because the salt penetrates deep into the food.

Rock salt contains small amounts of nitrates. Bacteria in meat turn nitrates to nitrites, and nitrite ions combine with haemoglobin, forming nitrohaemoglobin. This substance gives the meat a strong pink colour, even when cooked. The permanent pink colour of meat tissue preserved in this way shows that the salts have reached all tissues, and the meat is safe to eat. Nitrates and nitrites are frequently used in the preservation of meat products.

Sugars are used as a preservative of fruits, especially in jams. A high concentration of sugar causes dehydration by osmosis and thus destroys any microorganisms that land on the jam and attempt to grow.

Pickling in **vinegar** (**ethanoic acid solution**) is another traditional method of preserving food. Here it is the very low pH that inhibits the growth of microorganisms, particularly bacteria. Most food-spoilage bacteria do not grow at pH values below 5. Foods commonly pickled include vegetables such as cabbage (sauerkraut), and some meats and fruits.

An alternative to adding acid is to allow acidity to develop in the food by the action of bacteria, such as the lactic acid forming bacteria, ethanoic acid bacteria, and proprionic acid bacteria. However, these organisms cannot lower the pH below pH 4, so the shelf-life of food preserved in this way is limited. Yoghurts, for example, are kept under refrigeration.

11 **Identify** the possible human health risks of specifically using sugar, salt and an organic acid as food preservatives.

Food poisoning (enteritis) caused by a bacterium

Enteritis is an **acute disease** (i.e. arising suddenly, but of short duration) that results from the ingestion of food containing certain microorganisms (usually bacteria) or microbial toxins. The organisms involved may reproduce themselves in food (e.g. when prepared food is not refrigerated) or in the host (the human patient). The linings of the small and large intestine are where the bacteria or its toxins have their effects. Typically, the patient develops abdominal pain and diarrhoea, with or without vomiting and fever.

The rise and rise of food poisoning

The number of reported cases of food poisoning has been rising at an accelerating rate in many developed countries (Figure 18.29). Possible reasons include:

- fewer meals are prepared and cooked at home, and more ready-meals are bought in;
- more fast-foods are eaten while on the move, with little opportunity to wash hands;
- demand for cheaper food has lead to intensive farming of livestock, followed by slaughter at centralised abattoirs with larger throughputs of carcasses;
- changes (mutations) in the microorganisms themselves;
- greater public awareness of food poisoning symptoms, improved diagnosis by doctors, and better detection at hospital pathology labs, may mean more cases are now reported and recorded.

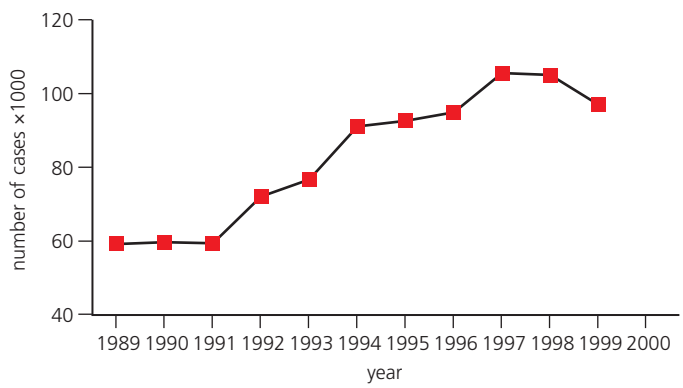


Figure 18.29 Food poisoning cases in a Western European country in recent years

Staphylococcus aureus – a food poisoning organism

Staphylococcus aureus is a Gram-positive coccus, capable of producing toxins in contaminated foods. People who handle cooked foods (e.g. cream-filled pastries, chicken and other meat products, puddings and creamy salad dressings) with contaminated hands may introduce the bacteria. However, if this food is kept refrigerated after preparation, the bacteria will fail to multiply. But if the food is left in a warm place (such as a kitchen or out of doors at a picnic), then *Staphylococcus* will grow at an alarming rate, and will secrete toxins into the food. If the contaminated food is eaten, nausea, vomiting and diarrhoea occur in 1–6 hours (Figures 18.30 and 18.31). Note that re-cooking the contaminated food does not remove these toxins – they are heat stable.

Figure 18.30 How *Staphylococcus aureus* causes food poisoning

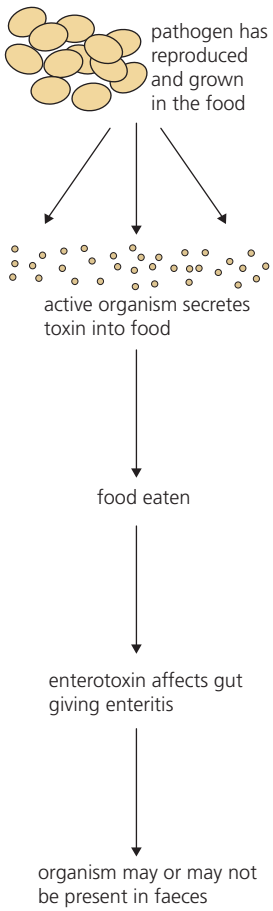
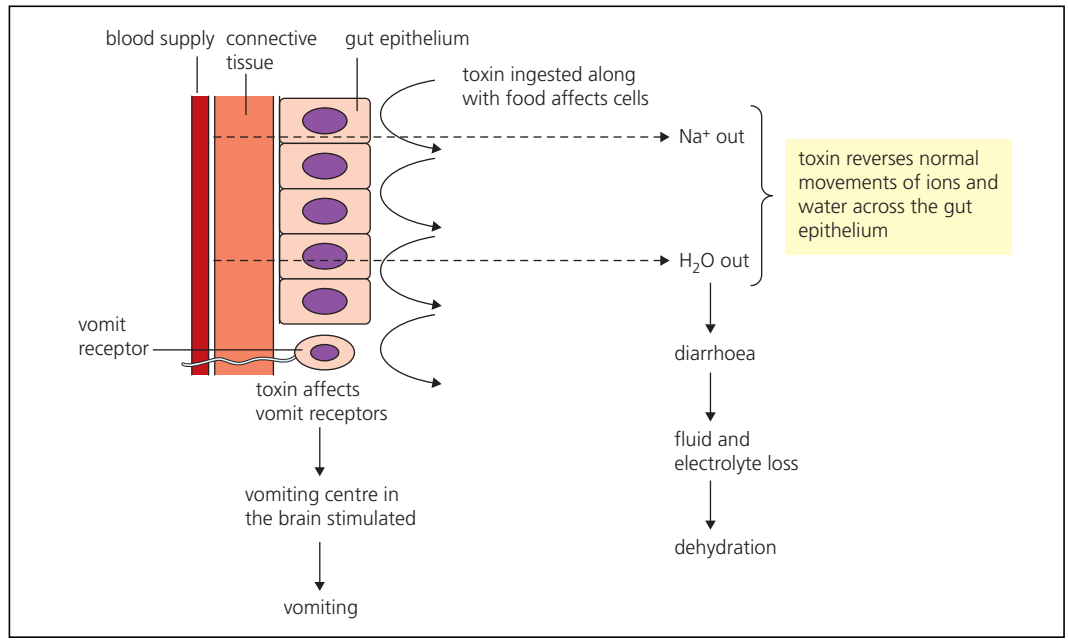


Figure 18.31 What happens when *Staphylococcus aureus* toxin is ingested



Treatment of food poisoning

Patients with food poisoning suffer from mild (or possibly severe) dehydration if vomiting and diarrhoea continue for some time. The need is to minimise discomfort from the fever experienced, and to maintain body fluids as and when possible.

In extreme cases, providing clean water may not be enough and oral administration of a dilute solution of electrolytes may be needed to make good the fluid and ions lost from the body. Any additional treatment with antibiotics is normally inappropriate – the bacterium has had its effects through the action of its toxins, which will progressively subside. Fluid (and electrolyte) replacement alone should restore normal body functions, with time.

12 Explain why significant losses in ions and salts in cases of enteritis pose a serious short-term threat to health.

TOK Link

A huge number of microorganisms inhabit the human body, together with occasional visitors, including unpleasant pathogens. Given the plethora of microorganisms about our bodies, how may a particular disease be correlated with the presence of a causative pathogenic bacterium?

Extension: Prevention is better than a cure!

In the issue of food poisoning, prevention requires high standards of hygiene during storage of food and during the preparation of meals to ensure that when consumed, food is free from pathogenic bacteria. The importance of using hygienic methods in food handling arises from the ubiquitous presence of dangerous microorganisms and in the speed with which infected food can become a source of toxins. The practical steps concern the avoidance of initial contamination, the prevention of multiplication of microorganisms, and the thorough cooking of food to be eaten hot (or later, when cool).

Metabolism of microorganisms

F5.1–5.6

Nutrition is the means by which organisms obtain the **energy** they require, and a source of **carbon** for the organic compounds they need. In biology of the eukaryotes, our focus has largely been restricted to the contrasting modes of nutrition of green plants (photosynthesis), and of animals and fungi (various forms of heterotrophic nutrition). Diversity in metabolism is a major feature of prokaryotes (both the Eubacteria and Archaeabacteria); indeed, they show greater variety in nutrition than the eukaryotes do.

Depending on how they obtain nutrients, prokaryotes can be divided into four major categories: photoautotrophs, photoheterotrophs, chemoautotrophs and chemoheterotrophs.

Photoautotrophs are organisms that use light energy to generate ATP and to produce organic compounds from inorganic substances.

An example is the cyanobacterium *Anabaena* (Figure 18.32). This cyanobacterium is found in fresh-water habitats. (*Anabaena* is also an example of a free-living nitrogen-fixing organism.)

Photoheterotrophs are organisms that use light energy to generate ATP, but obtain organic compounds from other organisms.

An example is the purple non-sulphur bacterium *Rhodospirillum*, an organism found in mud, lake water and sewage.

Chemoautotrophs are organisms that use energy from chemical reactions to generate ATP and produce organic compounds from inorganic substances.

An example is *Nitrobacter*, a soil bacterium that releases energy from the oxidation of nitrites to nitrates, and so plays a role in the nitrogen cycle (page 564).

Chemoheterotrophs are organisms that use energy from chemical reactions to generate ATP and obtain organic compounds from other organisms.

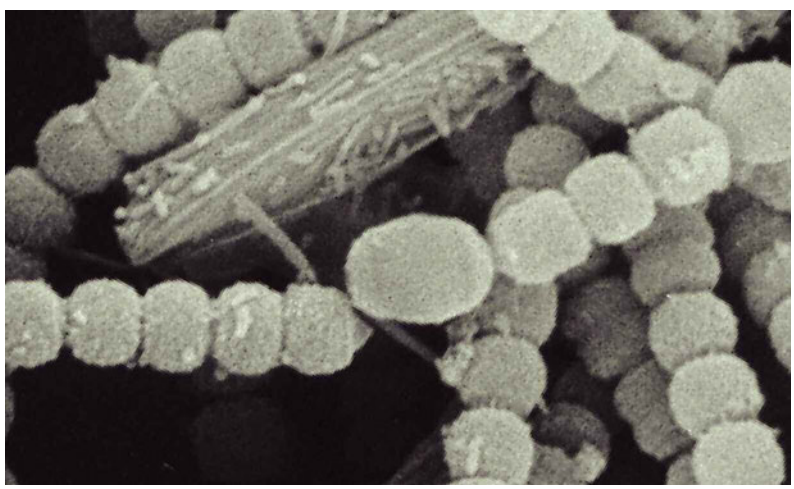


Figure 18.32 SEM of an *Anabaena* filament showing photosynthetic cells and occasional, larger heterocysts (site of nitrogen-fixation)

Very many of the bacteria belong in this huge group of microorganisms that decompose organic matter as and when it becomes available to them, ultimately enabling the recycling of nutrients for the benefit of the remainder of the food chain. *Escherichia coli* is an example – just one of the very many bacteria found in the human gut. Here, the metabolism of *E. coli* involves the fermentation of sugars.

13 **Draw** and **label** a diagram of the filamentous cyanobacterium, *Anabaena*, as shown in the SEM in Figure 18.32. (The guidelines for drawing scientific illustrations on page 10 may be helpful.)

Comparing photoautotrophy and photoheterotrophy

Photoautotrophs, such as the cyanobacterium *Anaebena*, and photoheterotrophs, such as the purple non-sulphur bacterium *Rhodospirillum*, are both prokaryotes and so do not contain chloroplasts. Their photosynthetic pigments (mainly chlorophyll *a* in the cyanobacteria, and bacteriochlorophyll in the non-sulphur bacteria) are located in simple membrane systems called mesosomes that are intuckings from the plasma membrane.

The steps of photosynthesis in these prokaryotes differ, too (Table 18.4).

In the **cyanobacteria**, carbon dioxide is the source of carbon, and this molecule is reduced to carbohydrate, using electrons obtained by the splitting of water. Oxygen is evolved as a waste product. The cyanobacteria are aerobic organisms.

Photosynthesis in the cyanobacteria (in summary):

$H_2O + CO_2 \rightarrow [CH_2O] + O_2$ where $[CH_2O]$ represents carbohydrate

In the **purple non-sulphur bacteria**, light energy is used to synthesise ATP (cyclic photophosphorylation, page 290), but **organic carbon is the carbon source**. The purple non-sulphur bacteria are anaerobic organisms, typically found in habitats where oxygen is absent or at very low concentrations. They also cannot survive in high concentrations of hydrogen sulphide, hence their name.

Table 18.4 A comparison of photoautotrophic and photoheterotrophic metabolism

	Photoautotroph (e.g. the cyanobacterium *Anabaena*)	Photoheterotroph (e.g. purple non-sulphur bacteria *Rhodospirillum*)
Energy source	■ light energy is used to obtain electrons from H_2O ■ light energy is also used to generate ATP from ADP + P_i ■ respiration is aerobic	■ light energy to remove electrons from organic molecules ■ light energy is also used to generate ATP from ADP + P_i, by cyclic photophosphorylation ■ respiration is anaerobic
Carbon source	■ CO_2 is reduced to carbohydrate	■ a variety of sources (e.g. organic or amino acids) are reduced to carbohydrate

Comparing chemoautotrophy and chemoheterotrophy

The **chemoautotrophs**, like all autotrophs, use an external energy source to synthesise carbohydrates from carbon dioxide (Table 18.5). They are sometimes referred to as the chemosynthetic bacteria. They differ from photosynthetic bacteria in that they take energy from inorganic chemical reactions (*), not light energy, in order to manufacture carbohydrate from carbon dioxide and water:

$CO_2 + H_2O + ENERGY^* \rightarrow [CH_2O] + O_2$ where $[CH_2O]$ represents carbohydrate

As a consequence, they are commonly found growing in the absence of light.

It is some particular exergonic chemical reaction that each species of chemoautotroph catalyses. Examples of economically important chemoautotrophs include the iron bacteria that live deep in iron ore deposits, and the nitrifying bacteria found in soils, which are illustrated here:

$$2NH_3 + 3O_2 \xrightarrow{\textit{Nitrosomonas}\text{ spp.}} 2HNO_2 + 2H_2O + ENERGY^*$$

$$2HNO_2 + O_2 \xrightarrow{\textit{Nitrobacter}\text{ spp.}} 2HNO_3 + ENERGY^*$$

Chemoheterotrophs, like all heterotrophs, obtain complex nutrients from food molecules taken in from their environment (Table 18.5). These complex organic molecules are their source of carbon (and energy) – they metabolise these molecules into all the other metabolites the cell requires. (Chemoheterotrophs may be further classified on the basis of the particular range of organic molecules on which they are dependent.)

Energy for metabolism is obtained by respiration of glucose and other metabolites absorbed. Very many chemoheterotrophs are aerobic, obtaining the energy required (as ATP) by aerobic respiration. Others are anaerobes, and obtain ATP by fermentation of various substrates.

Table 18.5 A comparison of chemoautotrophic and chemoheterotrophic metabolism

	Chemoautotrophs (e.g. *Nitrosomonas*)	Chemoheterotrophs (e.g. *Acetobacter* – aerobic; *Clostridium* – anaerobic)
Energy source	■ inorganic chemical reactions catalysed as the source of energy for synthesis of carbohydrate ■ respiration is the source of ATP for metabolism	■ complex organic molecules taken into the cell ■ respiration is the source of ATP for metabolism
Carbon source	■ CO_2 is reduced to carbohydrate	■ complex organic molecules taken into the cell

Bacteria and the bioremediation of the environment

Bioremediation is the process of exploiting microorganisms in the removal of pollutants from the environment. For example, we see this activity in the spontaneous removal of engine oils that drip from machines and engines, cars and other vehicles. These commonly seen deposits accumulate on nearby surfaces, but are normally removed (almost unnoticed by us), due to their degradation by naturally occurring (chemoheterotrophic) bacteria, provided the deposits are kept moist by occasional rain. A wide range of naturally occurring, ubiquitous prokaryotes and unicellular eukaryotes are able to hydrolyse and oxidise mineral oil to carbon dioxide (Figure 18.33).

When a major spill from an oil tanker occurs, leading to massive oil slicks that eventually reach seashore habitats, it has been found that about 80% of the non-volatile components of the oil may be oxidised and removed within a year of the pollution event. While vast numbers of marine and littoral species of non-vertebrates are destroyed by the hydrocarbon pollutants in the initial disaster, removal of oil slicks may be speeded by spraying the oil with essential inorganic nutrients, chiefly phosphates and nitrates, which aid the saprotrophic bacteria, and thus speed up bacteriological oxidation.

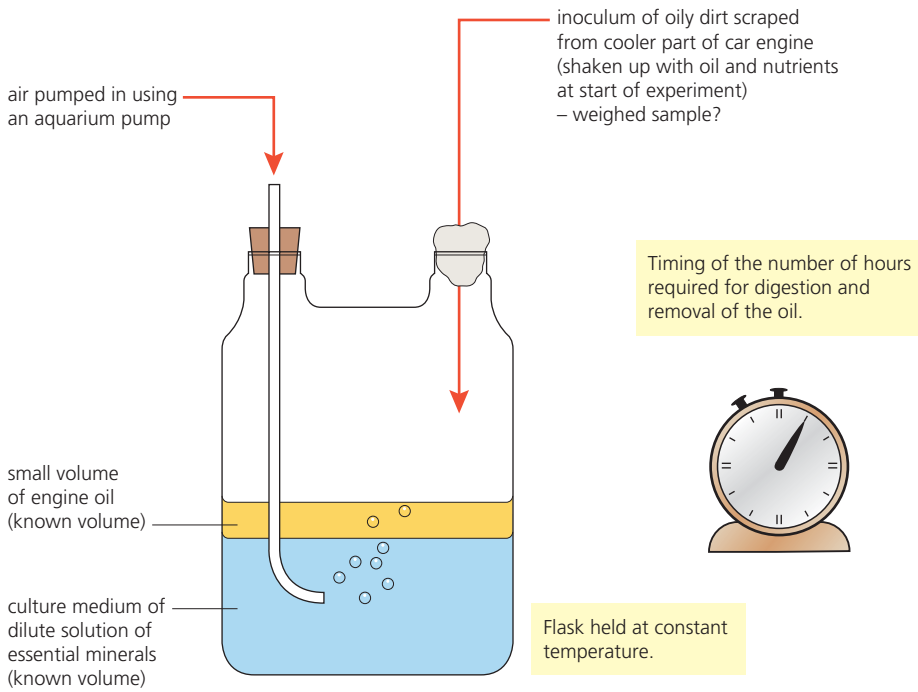


Figure 18.33 Investigating the degradation of mineral oil spills

Bioremediation of soils

Pesticides are products of the agrochemical industry that are very widely used to control harmful organisms that are a danger to crops or herds. Their use extends to the horticultural industries that are not organically orientated, of course, and the use of pesticides has been enthusiastically adopted by many people who maintain gardens in developed countries.

Modern pesticides are substances designed to kill specific types of pest, including plant weeds (herbicides), insects (insecticides), fungi (fungicides), and slugs and snails (molluscicides). Pesticides have enormously improved productivity in agriculture, but their use has generated problems in the environment. Herbicides are the most widely used group of pesticides, but insecticides are also widely used and they may cause the greater problems for humans.

Pesticides are a wide range of different compounds. Many of them are organic molecules suitable as carbon sources and electron donors for various soil microorganisms, but others are not. Those that can be biodegraded are eventually removed from soils (Table 18.6), but their breakdown may not be entirely due to the activity of microorganisms. Some substances leach away into underground water, and others may spontaneously degrade with time.

Substance	Time for disappearance from soil
Herbicide: 2,4-D	4 weeks
Herbicide: 2,4-T	20 weeks
Insecticide (chlorinated compound): DDT	4 years
Insecticide (chlorinated compound): Aldrin	3 years
Insecticide (organophosphate compound): Malathion	1 week
Insecticide (organophosphate compound): Parathion	1 week

Table 18.6 Persistence of a selection of pesticides

Industrial solvents, now widely used in developed countries at least, are organic compounds such as dichloroethylene, trichloroethylene, chloroform, and some brominated and fluorinated related compounds, to name but a few. These toxic substances (some are also suspected of being carcinogens) are frequently detected as persistent contaminants of ground waters. This is a well documented problem in parts of the USA, for example.

A variety of different bacteria are able to break down these molecules, including the removal of the chlorine component. For example, the bacterium *Dehalobacterium* uses the compound dichloromethane in its metabolism, producing essential respiratory substrates, formate and acetate, as follows (in summary):

$$3CH_2Cl_2 + CO_2 + 4H_2O \rightarrow 2CHOOH + CH_3COOH + 6HCl$$

Substrate-level phosphorylation ($ADP + P_i \rightarrow ATP$, page 272) occurs in the steps of both formate and acetate formation.

In areas where the underground waters are seriously polluted, the activities of natural populations of bioremedial microorganisms may be enhanced at the site by drilling wells and vents so that micronutrients and sources of oxygen (as compressed air or peroxide solutions) can be added for the benefit of the microorganisms. The natural clean-up reactions are accelerated in this way, although the process may take months and years to complete. No doubt the dispersal of the pollutants and the depth of the geological strata they occupy is also a factor.

Selenium and bacteriological action

Reactions between microorganisms and certain inorganic elements and compounds, are not always advantageous to humans. For example, some microorganisms may alter the physical characteristics of metals and metalloid elements like arsenic and selenium, in the course of their metabolism.

The metalloid element selenium is not itself toxic to microorganisms. However, some may inadvertently convert inorganic selenium to a methylated form in the processes of their metabolism. In this form, selenium is easily taken up by organisms of local food chains. When methylated selenium is absorbed by mammals and other vertebrates, it is transported in the blood stream and is able to cross the blood–brain barrier. In the brain, this compound causes neurological damage that can be fatal.

Microbes and disease

F6.1–6.10

Disease is generally defined as an unhealthy condition of the body. Many diseases are due to parasitic organisms that invade the body, and are referred to as **pathogens**. The diseases caused by them are called **communicable diseases** because the pathogen may be passed from host to host.

Microorganisms cause the majority of communicable diseases, but of course, relatively few of the vast numbers of microorganisms are pathogens (most are free-living saprotrophs). Despite this, pathogenic microorganisms generally appear to have earned microorganisms a bad name; people often equate bacteria with disease, despite the many ways our quality of life and environment are maintained by activities of microorganisms.

The infectious disease cycle and the transmission of pathogens

The steps of an infectious disease event are:

source of pathogen → transmission → establishment in host → exit from host/fresh infection.

A major feature of this cycle of events is the transmission between hosts and the process of infection. Very few pathogens are able to move between host organisms by their own means; most rely on some method of transport to a new host. The methods of disease transmission are listed in Table 18.7.

Method of transmission	Examples
contact	■ clothing–skin contacts or direct skin contact (e.g. **ringworm**, caused by a fungus, *Microsporum*) ■ sexual contact (e.g. **chlamydia**, caused by a bacterium, *Chlamydia trachomatis*) ■ contact with bathing water with aquatic larvae of parasite (e.g. **schistosomiasis**, caused by blood fluke, *Schistosoma*)
droplet – in air currents	■ sneezes from infected people (e.g. **influenza**, caused by influenza virus)
dust – in air currents	■ sneezes by infected people creating contaminated dust that persists in ill-ventilated enclosures (e.g. **tuberculosis** (**TB**), caused by a bacterium, *Mycobacterium tuberculosis*)
food and drink	■ prepared, contaminated food left in warm conditions (e.g. **food poisoning**, caused by a bacterium *Staphylococcus aureus*) ■ drinking water contaminated by sewage (e.g. **cholera**, caused by a bacterium *Vibrio cholerae*)
via other vehicles	■ contaminated surgical instruments such as hypodermic needles (e.g. **hepatitis B**, caused by a virus of same name) ■ contaminated blood transfusion (e.g. **AIDS**, caused by human immunodeficiency virus)
vector	■ external transmission (e.g. **salmonella food poisoning**, by a housefly transferring a bacterium such as *Salmonella enteridis*) ■ internal transmission (e.g. **malaria**, by a female mosquito taking blood meal and transferring protozoan parasite, *Plasmodium vivax*)

Table 18.7 How pathogens are transmitted between hosts

Profiles of infections

Some pathogens cause disease only after successful penetration of host cells (intracellular pathogen). Some diseases are caused by toxins from pathogens, so the pathogen merely has to be present in an area of the body, such as the gut (extracellular pathogen).

Chlamydia, an intracellular pathogen

Chlamydia trachomatis is an intracellular bacterial parasite of cells, initially cells of the urinogenital tract, and is transmitted during sexual intercourse. It causes a condition known as chlamydia. This has become the commonest sexually transmitted disease of the developed world. Symptoms vary, depending on the precise sites of infection, but typically infection with this organism makes urination painful, and if the conjunctiva of the eyes are infected, results in painful red eyes. Infection of the ovarian tubes causes abdominal pain; infection of the testes causes swelling and pain.

Streptococcus, an extracellular pathogen

Streptococcus pyrogens is an extracellular bacterial parasite widely distributed in humans. Most infected individuals are merely carriers who show no symptoms of disease. When patients succumb to a 'strep' infection it may be an infection of the skin (such as impetigo or other skin infection), or scarlet fever, or a sore throat, or a form of pneumonia, for example.

The types of toxin that may cause disease

Many disease-causing microorganisms have their adverse effects via the toxins they produce. Disease-causing toxins may originate within the pathogen and be secreted, or exist on its outer surface. Either way, toxins are molecules that interfere with normal host body cell functioning, sometimes doing this at the site of infection, and sometimes more widely in the body.

Toxins are classified as **exotoxins** and **endotoxins** (Figure 18.34).

Exotoxins are soluble proteins (often enzymes) secreted by the pathogen and causing specific symptoms. Exotoxins may travel from the site of infection to other tissues where they have their effects. Some of the most lethal substances known are exotoxins. For example, **diseases caused by exotoxins** include **botulism** and **tetanus**.

- The neurotoxin **botulin** is secreted by *Clostridium botulinum*, an anaerobic bacterium of soil and pond mud. When consumed in contaminated food that has been inadequately cooked, botulin binds to the synapses of motor neurones and prevents release of neurotransmitter substances. As a result, muscles cease to contract and paralysis results.
- Tetanus is caused by *Clostridium tetani*, an anaerobic bacterium of soil and animal faeces. When it contaminates deep anaerobic wounds, it releases a neurotoxin that interferes with synapses in the spinal cord and motor nerves. The result is uncontrolled muscle contraction leading to muscular spasms that affect the whole body, but which in particular cause lock jaw and breathing difficulties.

Endotoxins are lipopolysaccharides on the **outer surface of the bacterium wall** of Gram-negative bacteria (page 558). Endotoxins are heat stable, but generally toxic only at high doses. They have their effects in several ways. Typically they may cause fever, shock, blood coagulation, diarrhoea, or inflammation. **Diseases caused by endotoxins** include **Salmonella food poisoning**:

- *Salmonella enteritidis*, a pathogen that invades the gut and introduces a toxin there, causes a food-based infection. The symptoms are a sudden onset of headaches, chills, vomiting and diarrhoea. At this stage, the toxin is mostly still attached to the bacterial walls. This is followed by a fever that lasts a few days, and is caused when the dislodged toxin reaches the blood circulation. Even at this stage, the bacterium itself remains in the lumen of the gut, without invading cells.

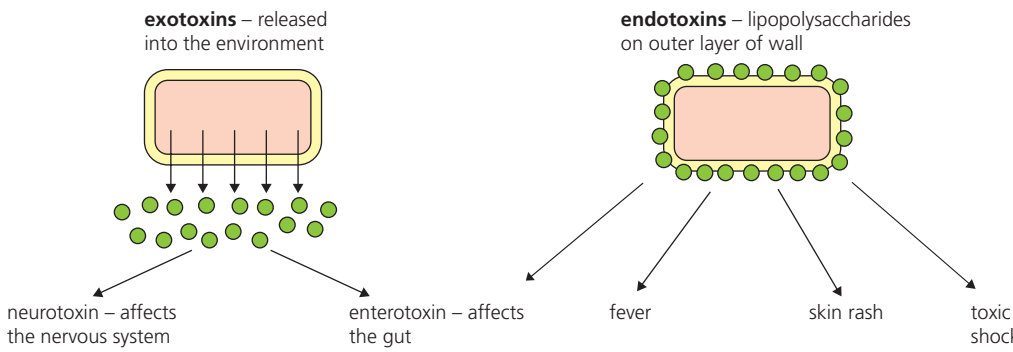


Figure 18.34 Exotoxins *versus* endotoxins

Disease prevention by avoidance of contamination

The prevention of infection by pathogenic microorganisms requires their inhibition and removal from the environments of people (or animals) vulnerable to the diseases they cause. Similarly, in laboratories where microorganisms are cultured, and places where food is manufactured and packaged, bacterial contamination has to be prevented. In premises where drugs, surgical equipment and related items are prepared and packaged, the same issue is uppermost. There are a number of ways this is tackled.

Disinfectants

Disinfectants are chemical agents used to check and prevent growth of microorganisms, typically on the surface of objects and equipment likely to be contaminated. Ideally, a disinfectant is effective against a wide variety of microorganisms, Gram-positive and Gram-negative bacteria and others, for example, together with dormant spores, fungi and viruses.

At the same time, disinfectants should not be toxic to humans and should not react with or corrode surfaces or equipment.

A wide spectrum of substances have been used. **Phenol** was the first disinfectant (and antiseptic), used by Joseph Lister in 1867, when he introduced the idea of preventing contamination of wounds during surgery (antiseptic surgery). Today, Lysol, a mixture of phenolics, is available as a commercial disinfectant. Phenolics have the advantage of remaining active on surfaces for a long time after application, but they leave a strong unpleasant odour and cause skin irritation.

Alcohols are some of the most widely used disinfectants today, often ethanol or isopropanol at 70–80% concentration. The five **halogen elements** (particularly, chlorine, bromine and iodine) are important antimicrobials, too. **Chlorine** is used in public water supplies, and has applications in food industries, for example. **Iodine**, complexed with an organic carrier to form iodophor, is widely used in hospitals and laboratories.

Antiseptics

Antiseptics are chemical substances applied to living tissues to prevent infection; they kill microorganisms present and inhibit growth of any pathogen. An antiseptic should not irritate the patient's skin, nor kill tissues, so necessarily, antiseptics are generally less toxic substances than disinfectants (or used at lower concentration). Most antiseptics are used for hand washing, and for treating surface wounds. Common antiseptics are alcohols at 60% or higher concentration, hydrogen peroxide at 3% solution, some phenol-containing compounds and iodine preparations.

Pasteurisation

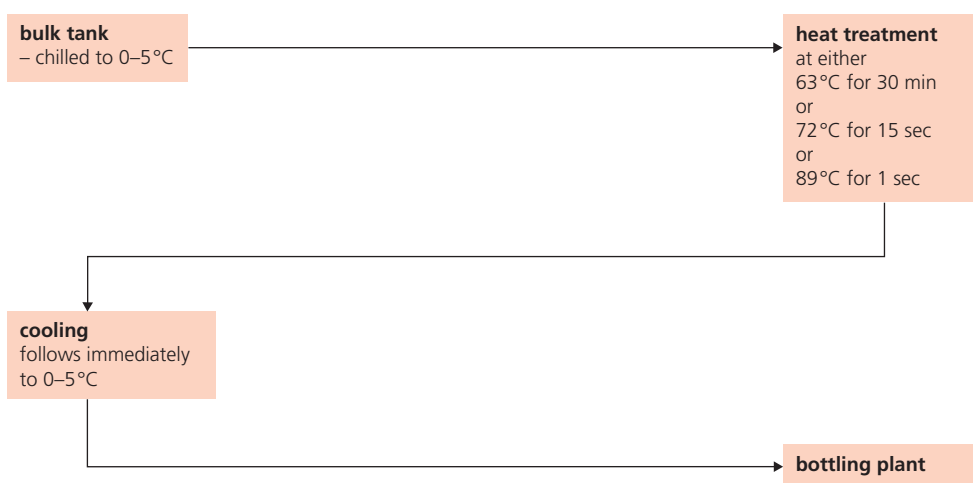
Pasteurisation is the process of reducing the microbial population of fresh milk and other heat-sensitive liquid foods. The name of this procedure acknowledges the genesis of this technique – first devised by Louis Pasteur to kill microorganisms in wine.

Heat treatment of milk is necessary since it is an ideal medium for the growth of microorganisms, and it is a difficult product to keep free from contamination from the air. Occasionally, an unhealthy cow may be the source of harmful bacteria, too, including *Mycobacterium tuberculosis* (causing TB), *Brucella abortus* (causes brucellosis), or *Streptococcus pyogenes* (causing sore throats and scarlet fever).

Pasteurisation of milk (treatment at 72 °C for 15 seconds, then rapidly cooled to below 10 °C) kills over 99% of the bacteria present. Pasteurised milk keeps longer, too, because bacteria that turn lactose (milk sugar) into lactic acid (which causes milk to sour) are also killed off (Figure 18.35).

Alternatively, milk can be rendered bacteriologically sterile by ultra-high temperature (UHT) treatment (homogenised, then heated to 132 °C for 3 seconds, then cooled). However, this treatment alters the protein content of the milk.

Figure 18.35 Pasteurisation of milk



Irradiation

Some forms of electromagnetic radiation are harmful to microorganisms, including two forms of ionising radiation:

- **X-rays**, produced by an X-ray source;
- **gamma rays**, emitted during certain forms of radioactive decay.

High levels of ionising radiation kill microorganisms outright. This form of energy brings about many disabling chemical changes in cells, but the destruction of DNA is the most damaging. Its usefulness as a sterilisation agent is its penetration deep into objects. Gamma radiation emitted from a cobalt-60 source is used to sterilise drug solutions and equipment, including disposable supplies such as syringes, following their manufacture. It can be used to sterilise food, too, but this is not widely applied to date.

Ultra-violet radiation, particularly at 260 nm wavelength, also kills all forms of microorganisms, also by inactivating their nucleic acids. Actually, little UV light below 300 nm wavelength reaches the Earth's surface because of the action of the upper ozone layer. However, damage to this ozone shield has followed on certain forms of atmospheric pollution in recent times (page 627). UV lamps are used in bacteriological laboratories, particularly in inoculation cabinets, but care is taken to ensure humans are not exposed, particularly since the radiation will damage skin and eyes. UV light does not penetrate through glass or far into water, but it is effective in the destruction of microorganisms in water when a thin film is passed under an appropriately powered source.

14 Design an experiment to demonstrate that microbes, freely circulating in the air, are able to contaminate exposed matter.

Antibiotics – their mechanisms of action

Many bacterial infections can be treated with antibiotics because antibiotics in low concentration inhibit the growth of microorganisms. Most antibiotics are naturally occurring chemical substances obtained mainly from certain fungi and bacteria commonly found in the soil.

The first antibiotic to be discovered, isolated and developed (not an easy task – it took from 1929 to 1944) was penicillin. Over 400 different antibiotics have since been isolated, but only about 50 have proved not toxic to patients (they show **selective toxicity**), and so have achieved wide usage.

Antibiotics effective against a wide range of pathogenic bacteria are called **broad-spectrum antibiotics**, and these include chloramphenicol and tetracyclines. Others, including penicillin and streptomycin, are **effective against a limited range of bacteria**.

How do useful antibiotics terminate a bacterial infection without harming the mammalian host?

Antibiotics specifically damage bacterial pathogens by disrupting one of three major aspects of the growth or metabolism of bacteria that occur in ways that are more or less specific to prokaryotes. Since the particular components, metabolites or enzymes concerned are not found in eukaryotic cells in the same form, the antibiotic is not toxic to the mammalian host tissues.

Cell wall synthesis inhibition

The most effective antibiotics work by interfering with the synthesis of bacterial cell walls (Figure 7.15, page 194). Once the cell wall is destroyed, the delicate plasma membrane of the bacterium is exposed to the destructive force generated by excess uptake of water by osmosis, and possibly also to attack by antibodies and phagocytic macrophages.

Several antibiotics, including penicillin, ampicillin, and bacitracin, bind to and inactivate specific wall-building enzymes. These are the enzymes required to make essential cross-links between the linear polymers of the walls, in particular species. In the presence of the antibiotic, wall polymers continue to be synthesised by the pathogens, but the individual strands are not linked and bound together. The walls fall apart.

Protein synthesis inhibition

Other antibiotics inhibit protein synthesis by binding with ribosomal RNA. The ribosomes of prokaryotes (known as 70S) are made of particular RNA subunits, together with many polypeptides (which mainly function as enzymes). The ribosomes of eukaryotic cells are larger (80S), and are built with different RNA molecules and polypeptides. Antibiotics like streptomycin, chloramphenicol, tetracyclines, and erythromycin all bind to prokaryotic ribosomal RNA subunits that are unique to bacteria. Here their presence causes protein synthesis to be terminated.

Nucleic acid synthesis inhibition

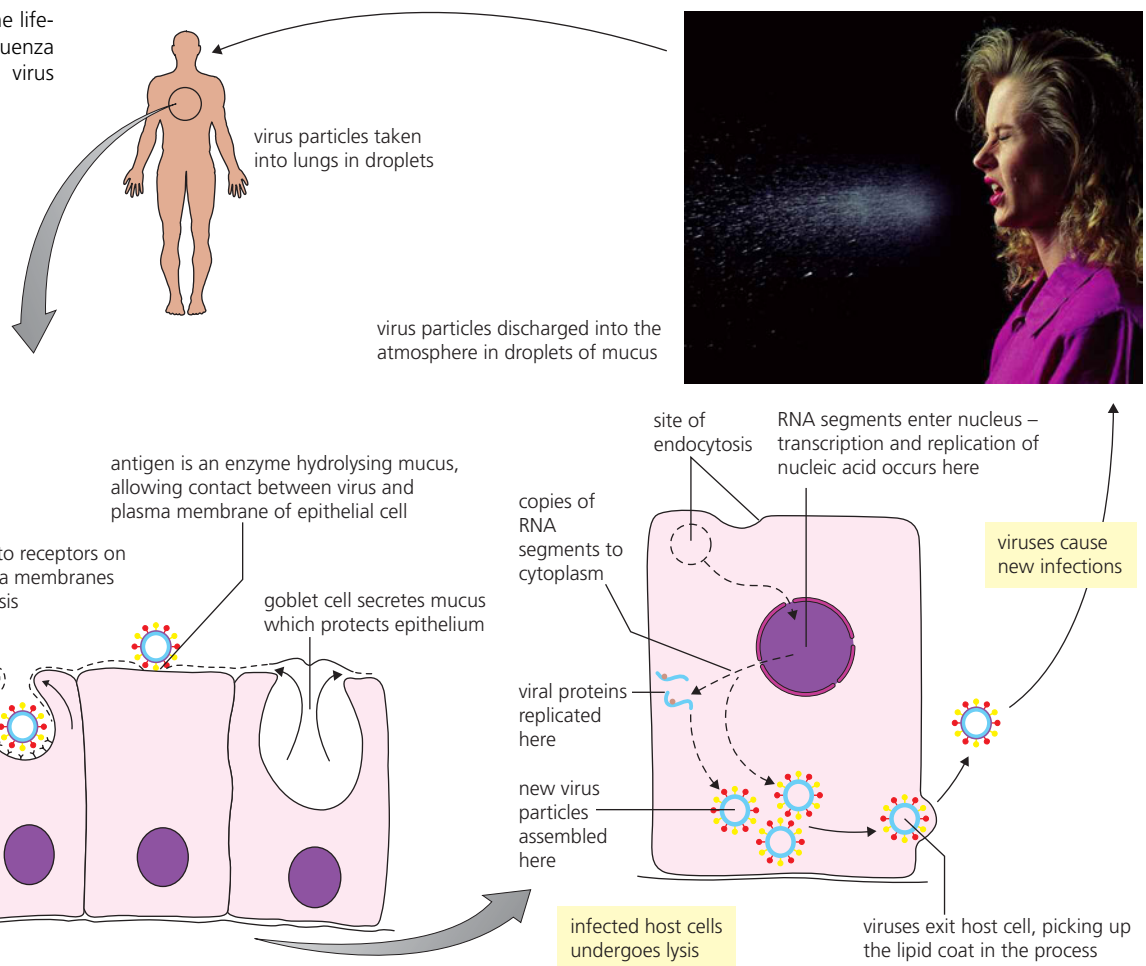
A few antibiotics interfere with DNA replication or transcription, or they block mRNA synthesis. These antibiotics – for example, the quinolones – are not as selectively toxic as other antibodies. This is because the processes of replication and transcription do not differ so greatly between prokaryotes and eukaryotes as wall synthesis and protein synthesis do.

The life cycle of the influenza virus

We have seen that the influenza (flu) virus contains eight short strands of RNA (a segmented genome), surrounded by a protein capsid, and external to this is a layer of lipid (Figure 18.13, page 563). Protein 'knobs' (antigens with specific names, traditionally shortened to H and N) project through the lipid layer.

The virus enters the human host's lungs via droplets inhaled through the mouth and nose. The epithelial cells lining the bronchus and bronchioles are the cells that the virus parasitises. Entry into epithelial cells occurs by endocytosis (Figure 18.36).

Figure 18.36 The life-cycle of the influenza virus



Then, replication of the virus occurs – new RNA segments are produced in the nucleus and the capsid proteins are produced in the cytoplasm of the host cell. New viral particles are assembled, and the lipid capsule is added as the virus exits the dying host cells.

The host cells now break down (lysis) and toxins are released at this stage. Destruction of the epithelial cells paves the way for serious secondary bacterial infections of the host lung tissues, typically bacterial pneumonia.

The flu life cycle (*if we can talk about a life cycle in a disease-causing agent that is not strictly described as living*) is described as lytic.

Influenza viruses are classified into antigenic groups according to the H and N proteins of their capsids. The chemistry of these antigenic proteins is controlled by the genome of the virus, which is potentially very variable. Alterations of the virus genome frequently arise, and they lead to new antigenic types and therefore to new strains of flu in humans. Sometimes, humans (or other hosts) have little or no resistance to the new strain. The immune system will not have previously met the new antigens. So, no memory cells (page 354) exist in the lymph nodes.

Because of this, flu is frequently a major epidemic disease; epidemics occur almost every year.

An epidemic disease is one of widespread occurrence in a community at a particular time.

However, strains arise from time to time with exceptional genes for virulence and transmissibility. A strain with just the right mix of virulence and transmissibility leads to a pandemic in which millions of humans die.

A pandemic disease is one that is prevalent over a whole country, continent, or the world.

During the last century there were three pandemics (Table 18.8), and it is inevitable that flu pandemics will occur again, from time to time.

Pandemic	Profile
Spanish flu 1918–9	■ Killed about 40 million people (compare the 10 million victims of the First World War – at the time called The Great War). ■ Nations struggled to cope; the end of the war was a time when resources were exceptionally stretched, and viruses were not understood. ■ Vaccines were targeted at bacteria. ■ Civilians were put in quarantine (but troop movements continued between continents). ■ High levels of personal hygiene were advocated.
Asian flu 1957–8	■ Killed about 1 million people. ■ Medical knowledge was more advanced than in 1918, but also, the strain was substantially less virulent. ■ Strain was identified and vaccines were produced, but it was not possible to produce sufficient. ■ Quarantine measures were used, again to little effect.
Hong Kong flu 1968–9	■ Killed about 1 million people. ■ By this date the WHO existed, and its global flu surveillance network gave early warning of an imminent pandemic as the disease spread from its origins in South East Asia. ■ Vaccines were developed quickly, but not enough could be produced in time to meet the full demand.

Table 18.8 The most recent past flu pandemics

The epidemiology of a flu pandemic

Epidemiology is the study of the occurrence, distribution and control of diseases.

There are three pre-requisites for a flu pandemic:

- **a novel virus subtype must reach human hosts** from its point of origin, and human immune systems must be unfamiliar with it;
- the virus must **replicate in humans** and cause disease there;
- the virus must be **efficiently transmitted between humans**.

You may remember the public discussions and anxiety about the bird flu epidemic that threatened us in 2005–06. That strain of flu was identified as H5N1. The interlocking flight paths of wild birds, migrating between seasonal feeding grounds in different parts of the globe, transmitted the infection among wild populations of birds, and sometimes passed it to farmed bird stocks. (Clearly there are many adverse features that underpin the onset of a pandemic which cannot be controlled.)

Some unfortunate people in countries spread as widely apart as the Far East and Eastern Europe, who had direct close contact with infected birds (and sometimes indirect contact), contracted the disease. For some, this exposure proved fatal.

The H5N1 strain of 2005, although extremely virulent, failed the third pre-requisite listed above – it was not a strain that was quickly and easily transmitted between individual humans, and so no pandemic ensued (at that time). Had the virus adapted in that way too, then it is highly likely that the next flu pandemic would have occurred at that point.

What measures are required in the face of a possible flu pandemic?

The international success experienced in controlling the SARS (severe acute respiratory syndrome) epidemic of 2003 confirmed the value of a **supranational health initiative**. Here, the role is to maintain oversight of the whole world scene, exchange updated information, warn countries most likely to next be in the path of the infection wave, and extol best practice. Currently, the World Heath Organization (WHO) and the technical agencies it liaises with fill this role. We have the essential early warning and response system that allows timely exchange of essential information.

The technical difficulties in the **production of influenza vaccine** to the new antigens, as soon as these can be identified, at the required speed and in volumes that would be required, are the subject of continuing research. So, too, is the development of newer and more effective prophylactic drugs to maintain the resistance to infection of key medical and pharmaceutical staff, at the very least.

Equally important is the planning by governments concerning the **maintenance of vital national services** such as food and fuel supplies, and public safety, given that 25% or more of the adult population may be incapacitated at the height of a pandemic.

15 Distinguish between:

- **a** antibiotics and vaccines
- **b** inflammation and immunity
- **c** vector and host.

Malaria – caused by a protozoan, transmitted by mosquitoes

Malaria is the most important of all insect-borne diseases. About 80% of the world's malaria cases are found in Africa south of the Sahara, and here some 90% of the fatalities due to the disease occur (Figure 18.37). It is estimated that about 400 million people are infected, of whom 1.5 million (mostly children under 5 years old) die each year.

Malaria is caused by *Plasmodium*, a protozoan, which is transmitted from an infected person to another person by the blood-sucking mosquito *Anopheles*.

Transmission of *Plasmodium* by the mosquito and its effects on the host

Only the female mosquito is the vector (the male mosquito feeds on plant juices). *Anopheles* is a fluid-feeding heterotroph (not an ectoparasite, as sometimes stated). It detects its human host, lands, and inserts mouthparts (which are formed into a long thin proboscis) into a blood vessel below the skin surface. A meal of blood is taken quickly – there is a danger that an active, alert human will swat the insect. Mosquitoes tend to feed at night, on sleeping victims, but those that alight on patients already ill with malaria are likely to be able to feed unhindered (Figure 18.38).

For any blood-sucking insect, the mammalian blood-clotting mechanism presents a problem, and has to be overcome. As the mosquito's proboscis penetrates the vein, a secretion from its salivary glands passes down the hypopharynx and inhibits clotting. At this point *Plasmodium* may enter its human host (if the *Anopheles* carries an infective stage). Meanwhile, the *Anopheles* loads up with a blood meal. This is essentially 80% water, and excess water has to be lost from the body as a priority.

The effects of *Plasmodium* on the human host depend partly on the species of *Plasmodium* involved, and partly on how many previous bouts of malaria the patient has had and hence the degree of immunity that has developed.

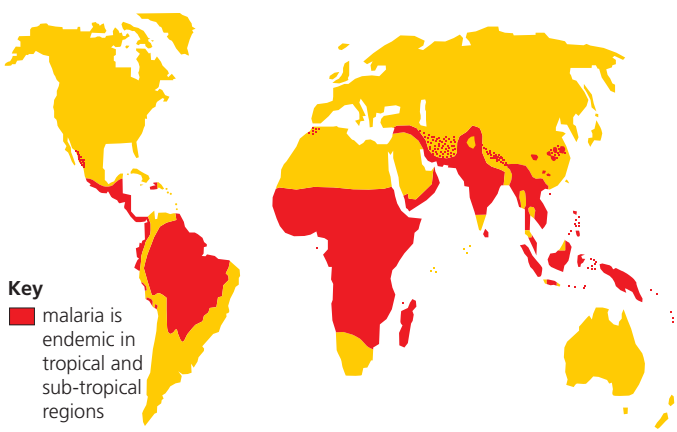


Figure 18.37 World distribution of malaria

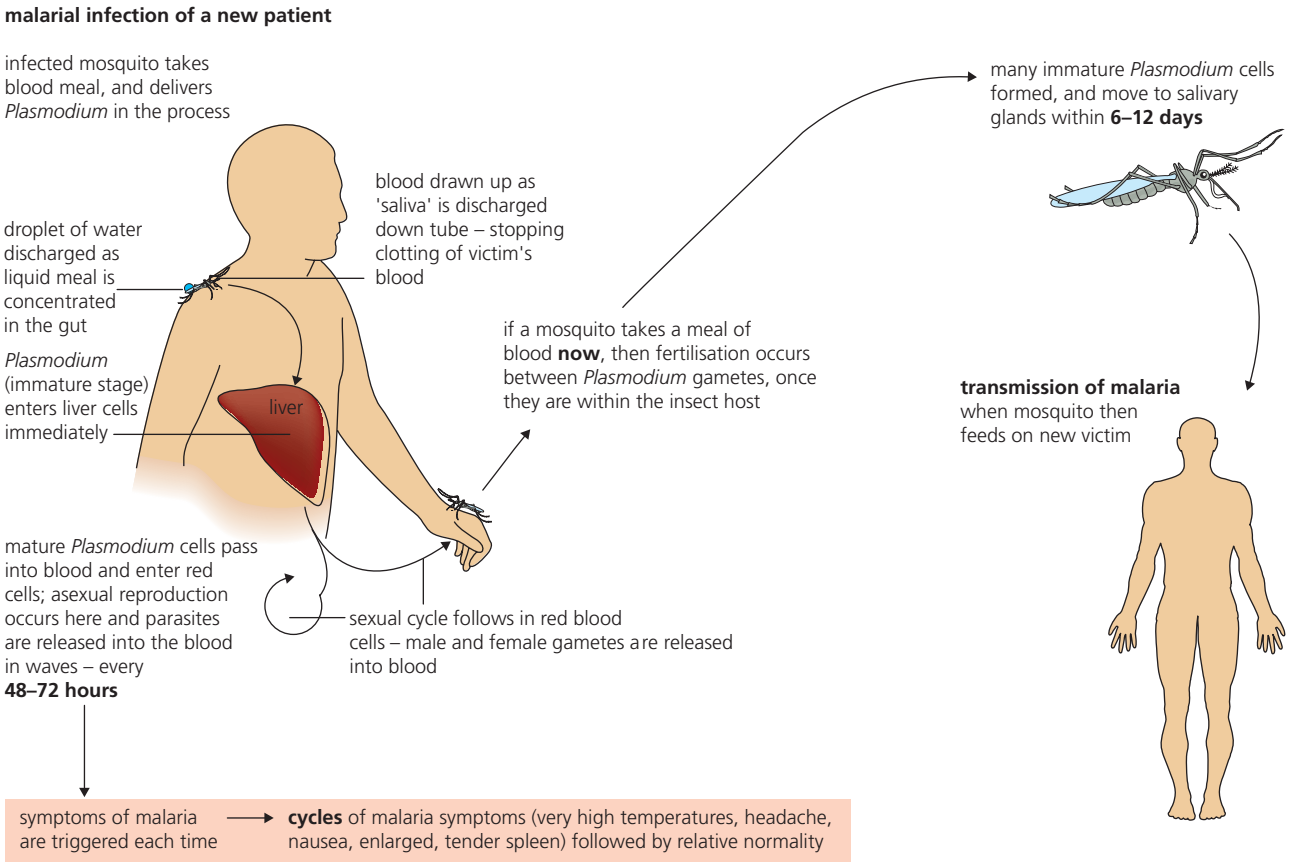


Figure 18.38 Anopheles feeding and the transmission of malaria

During the peak of infection, parasites are released from red cells into the plasma of the blood stream every two to three days, together with toxins. The body temperature rises to 40–40.5 °C, with intense fever symptoms. The spleen becomes enlarged and painful. Afterwards, the body temperature falls below normal, and there is profuse sweating.

Prions and the cause of encephalopathies

Proteins called **prions** are believed to be the agents that cause diseases known as encephalopathies in which the brain becomes spongy and forms holes where once there were neurones. The term 'prion' is a contraction, derived from **pr**ote**i**nace**o**us i**n**fectious particle.

The affected organisms, which may be human (with Creutzfeldt–Jacob disease – CJD), sheep (with scrapie), or cattle (with bovine spongiform encephalopathy – BSE), lose physical condition and eventually become totally unco-ordinated. The cow in Figure 18.39 had difficulty in standing. In humans, the memory is lost, as well as body control, prior to death.

Figure 18.39 A cow with BSE

What are prion proteins and how may they cause disease?

Prion proteins are natural components of cells of mammals and also of yeasts, but they probably occur more widely. Prions are, therefore, assumed to have some important role, but this is not yet understood. What has been established is that the normal tertiary structure of prion protein ($\mathbf{PrP^c}$) consists of multiple α helices, but that this large molecule

can unfold and form into a different molecule (**PrP^{sc}**) where two of the helices become β pleated sheets (Figure 8.20, page 256). The prion is said to have flipped (Figure 18.40).

Once this change in shape has occurred, **PrP^{sc}** can trigger the same change in shape in normal **PrP^{c}** with which it is in contact. Then the mass of **PrP^{sc}** molecules apparently coalesce into insoluble fibrils. When this happens in neurones of the brain, neurone destruction occurs and the encephalopathy condition follows.

In 1997, Stanley Prusiner of the University of California at San Francisco was awarded a Nobel Prize for his work on his prion theory of encephalopathies. Other workers held that an as-yet undiscovered virus was involved. However, the infectious agent for encephalopathies remains infectious even after bombardment with radiation sufficient to destroy all DNA or RNA present (essential to a virus). Since proteins are not destroyed by radiation, prions are most likely to be responsible.

Encephalopathies are cannibalistic in origin

The first known encephalopathy disease was *kuru*, observed in people of Papua, New Guinea, whose custom it was to honour their dead by eating them. Men ate muscle tissue, but women and children received brain tissue. Only the latter eventually died of *kuru* – when the ancestor had also died of that disease.

Another prion disease with a long history is scrapie. Until recently, encephalopathies had not been known to jump the species barrier, say from sheep to cattle. Recent farming practices of using offal from sheep or cows in manufactured animal feed are likely to have spread encephalopathies among cattle. Also, pituitary extract prepared by vets for injection into other cattle may have been a source.

What allows the prions of one species to infect some but not all other susceptible species?

It may depend on how similar the primary structures of different prion proteins are. If the size and amino acid sequence of prions of two species are sufficiently similar, then they may infect both.

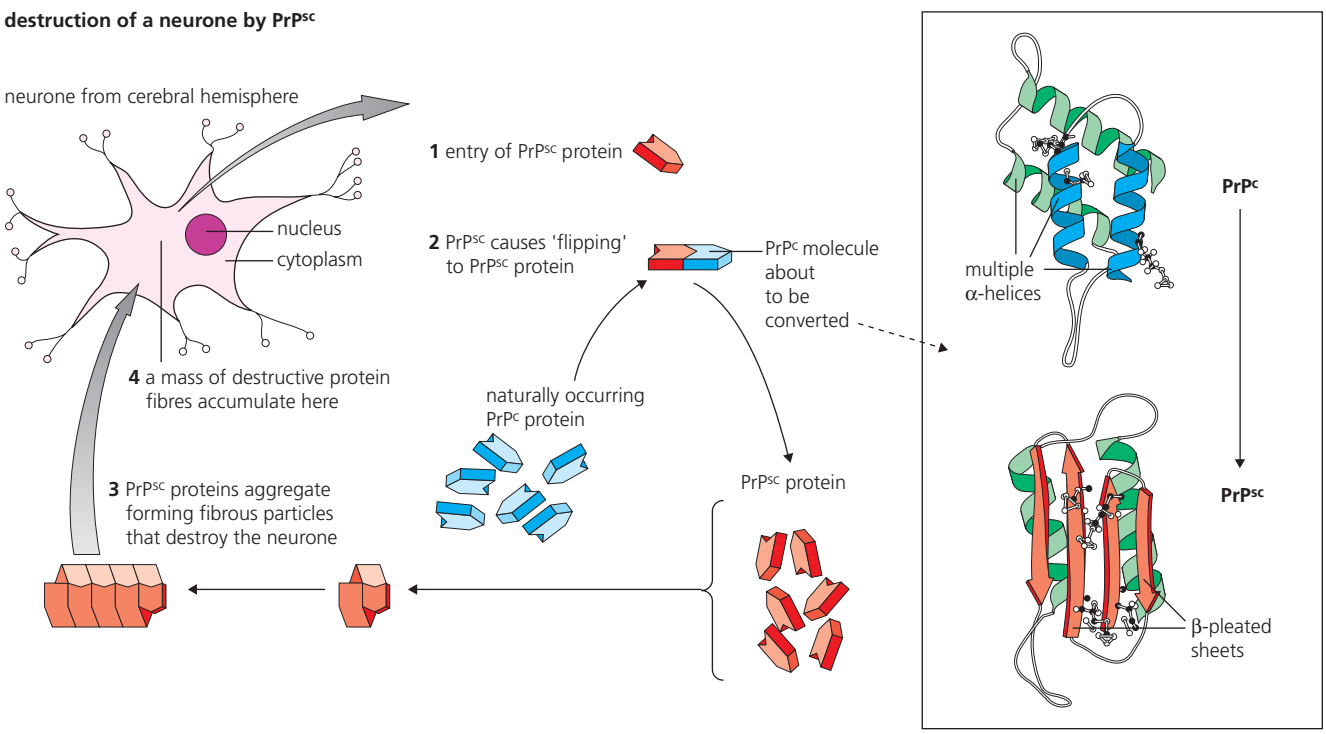
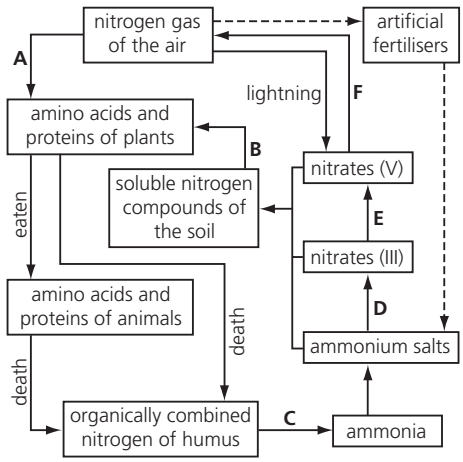


Figure 18.40 Prion protein flipping and the triggering of an encephalopathy

Examination questions – a selection

Questions 1–6 cover syllabus issues in this chapter (a new Option)**.**

Q1 The flow chart below represents the main events of the nitrogen cycle, in several stages of which microorganisms play a central role.



a State the different types of organism that may be involved in **step A** and where they occur within the environment. (6)

b Outline how the events of **step B** are brought about. (6)

c Identify the types of organism involved in **step C** and state where they may be found within the environment. (4)

d Describe how ammonium salts may be converted to nitrites (**step D**) and nitrites to nitrates (**step E**) in the soil. (6)

e Suggest typical soil conditions that bring about **step F**. (1)

f Part of the nitrogen cycle flow diagram is represented by dotted lines. Explain how these steps differ fundamentally from the remainder of the cycle. (2)

Q2 **a** Distinguish between *division* and *domain* in the classification system of living things. (4)

b Compare the characteristic features of cell structure and biochemistry by which the organisms of the three domains are differentiated. (6)

c Outline the distinguishing features of methanogens as archaebacteria. (4)

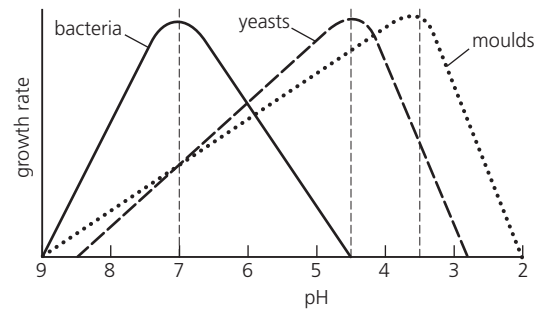
d Suggest why viruses are **not** classified within any domain. (2)

Q3 **a** Describe **two** ways in which the metabolism of saprotrophic microorganisms is exploited in the conversion of human sewage to safely disposable products. (8)

b Outline **one** method by which biomass may provide a source of fuel. (4)

Q4 The steps to genetic engineering are brought about with the aid of several naturally occurring enzymes – sometimes known as the genetic engineer's 'tool-kit'. Identify **four** of these enzymes, explaining their roles *in vivo* and in genetic modification processes. Present your answer in the form of a table. (12)

Q5 The graph below shows the effect of pH on the growth rate of three types of microorganism.



a Discuss what is meant by *growth rate* of a microorganism. (2)

b State the pH range in which yeasts grow and what their optimum pH is. (2)

c Calculate the range of pH tolerance of the group of microorganisms whose growth is least inhibited by differing conditions of acidity and alkalinity. (1)

d In acid soils, state which microorganisms are more likely to be responsible for the decay of dead organic matter, and which microorganisms are likely to be least active in decay processes. (2)

e In the preservation of food using vinegar, explain which of these groups of microorganisms will be least inhibited. (1)

f State one alternative technique of food preservation to vinegar and explain how this substance inactivates the growth of microorganisms that may contaminate. (3)

Q6 **a** Explain why carbon and nitrogen are defined as macronutrients but calcium and magnesium are micronutrients. (3)

b Define *photoautotroph* and *photoheterotroph* by reference to the way that energy is transferred and the source of carbon used. (4)

c Construct a labelled diagram of the carbon cycle and annotate it to emphasise the ecological significance of the metabolism of chemotrophic prokaryotes. (6)

Option G

19 Ecology and conservation

STARTING POINTS

- **Ecology** is the study of organisms in relation to their environment.
- An **ecosystem**, such as a lake, woodland, or salt-marsh, is a stable, settled unit of nature. A **habitat** is the surroundings in which an organism lives.
- A **population** consists of all the living things of the same species in a habitat. A population may initially grow exponentially, but it is prevented from doing so indefinitely by environmental constraints.
- A **community** consists of all the populations of organisms interacting with each other and with their surrounding physical and chemical environment within a habitat.
- A **food chain** is a sequence of organisms within a habitat in which each is the food for the next, starting with **producers** (photosynthetic green plants – **autotrophs**), to primary **consumers** (herbivores), to secondary and tertiary consumers (carnivores). Each level of the food chain is called a **trophic level**.
- While energy flows through food chains, from sunlight to becoming heat energy lost into space, **nutrients** (particularly minerals, but also carbon dioxide) are endlessly recycled. These are released from dead organisms and their waste matter by the actions of **saprotrophic** bacteria and fungi (**decomposers**).
- This chapter extends study of aspects of ecology and of the **impact of humans on the environment** begun in Chapter 6 (pages 137–77).

About 150 years ago, the common approach to biology was to study a range of different species, each in isolation, as examples of different levels of biological organisation. This work established a great deal of information about the structure (morphology and anatomy) and function (physiology) of individual organisms. It largely ignored the ways in which **organisms interact with each other**.

Environmental biology or **ecology** is the study of relationships between living things and their environment, and of the impact of human actions on them. Humans have brought about change everywhere. One outcome may be the enhanced rate of **extinctions of species**. However, attempts at conservation often clash with the needs of productive industries. Conservation initiatives have international dimensions, and are the subject of national government policies and international action.

In this chapter, the **distribution of organisms in a community**, and the ways this may be **studied experimentally** are discussed. **Productivity of ecosystems** and the **successions** of organisms in the world's biomes (major life zones) are then investigated. Then, in these contexts, **the impacts of humans** on ecosystems are examined.

In the Additional Higher Level extension, issues in the **conservation** of biodiversity are explored further, and an experimental approach to **population ecology** is established.

Community ecology

G1.1–1.8

Ecology is the study of the relationships of living things with their non-living surroundings – their environment. Factors which have an effect on the life of organisms are called **environmental factors**.

Environmental factors, and there are many of them, divide naturally into those due to the living (**biotic**) components, and those due to the non-living (**abiotic**) components. At any one moment, one particular environmental factor may appear more influential than others. In practice, however, environmental factors are interrelated, and it is usually difficult to isolate the influence of any one (Figure 19.1). We can illustrate this interaction of factors first.

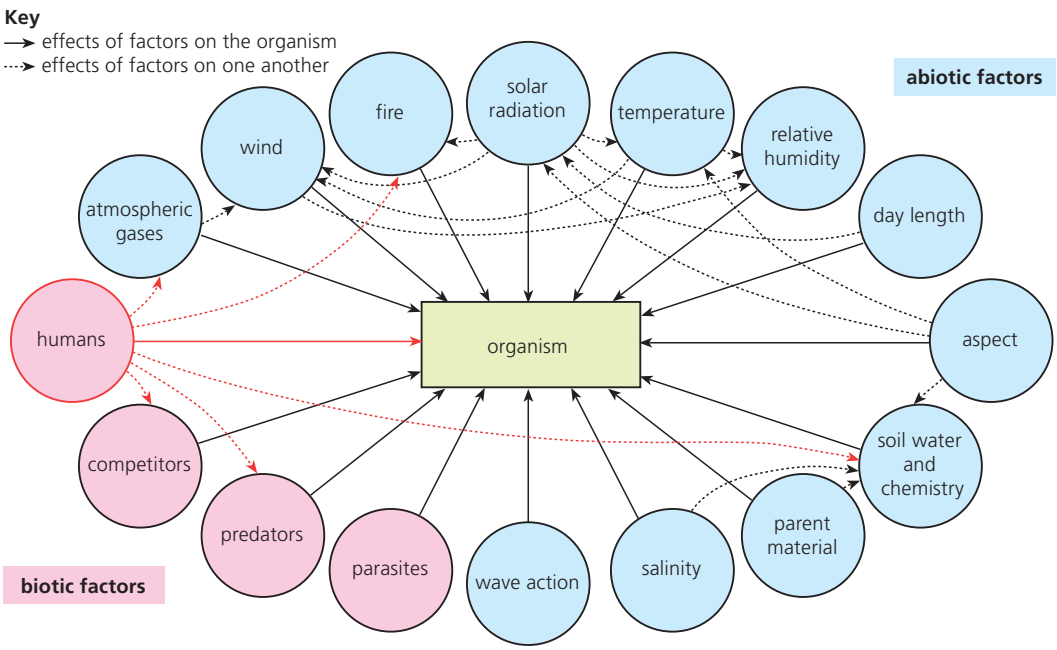


Figure 19.1 Abiotic and biotic factors and their interactions – a summary

Introducing factors that affect the distribution of plants

The individuals of a species are not distributed at random in the environment; each species is restricted to certain geographical areas, and there, in particular habitats. Some of the more important environmental factors that determine the distribution of green plants are the light regime (length of day, seasonality), temperature range, water availability, soil pH, salinity and mineral nutrition.

Light is the ultimate source of energy in ecosystems; plants grow where there is sufficient light for their **autotrophic nutrition**. This need for light has an effect on the structure of plant communities. For example, woodlands and forests are stratified into layers, from the canopy above, to the shrub layer below, the field layer (herbaceous plants) and ground layer (mosses). Each layer has a different degree of exposure to light, and it has plant life adapted to that light regime.

Also, the length of day is the dependable environmental clue for many plants (and animals) for the setting of their daily and seasonal rhythms; the duration of illumination is the environmental trigger for inducing flowering of many plants, for example. Sunlight is the major source of heat, too. Through its effects on temperature, light intensity also influences humidity.

Temperature determines the rate of biochemical reactions in organisms. Most eukaryotic green plants function within a temperature range of 0–50 °C. Within this range, individual plant species may have an optimum temperature, together with minimum and maximum temperatures they normally tolerate. Remember, higher plants are not mobile, and the plant body has a temperature approximately the same as that of the environment. In fact, temperature acts in conjunction with light and water in most habitats as determining factors. For example,

temperature determines the rate of evaporation from plants (transpiration), given they have an adequate water supply, and their stomata are open (page 312) in the light.

Water is vital to life; the unique properties of water (page 43) make life possible. Water availability in association with light and temperature largely determines the distribution of plants in terrestrial habitats. Where water is plentiful, vegetation tends to grow densely, and where water is in short supply, vegetation is sparse. Most of the water that enters plants is lost by evaporation, and death from dehydration is a major threat to the rooted plant. Land plants have extensive external surfaces (due to their mode of nutrition) over which evaporation may occur, so the plant body shows adaptations to reduce water loss, including a waxy cuticle covering the surface of leaves and stems. The plant attempts to replace all water lost via an extensive root system that is in intimate contact with the soil solution that occurs as a film surrounding individual soil particles.

Soil pH also exerts a strong influence on plant distribution, chiefly through its effect on the availability of essential ions, taken up by roots. The pH of soil also influences the population and activity of soil bacteria that decompose organic matter. So, the pH of the soil affects the release of ions from dead organic remains and waste matter in and on the soil. Generally, soils at or near pH 7 support a greater diversity of plants (and animals) than soils at or near pH 4 and pH 9. However, some plant species have adapted to particular soil environments, and many species are restricted to acid or alkaline soils.

Salinity is a factor in the distribution of water plants where salts and ions have accumulated to high levels. In natural fresh water, derived mainly from rainfall, the salt concentration is low (typically 0.02% dissolved salts), whereas in sea waters the salinity is relatively high, at about 35 parts per thousand (3.5%) dissolved salts.

In some low-lying coastal regions, salt marshes occur. In a salt marsh the salinity is variable. At high tides, salinity is at or close to that of the sea water that floods in. Once the tide turns, the salt concentration may change quickly. For example, salinity may exceed that of sea water when high temperatures and winds cause evaporation, and the remaining sea water is concentrated. Alternatively, when there is heavy rainfall, or when the flow of river water dilutes the sea water, the concentration of salts will fall well below that of sea water.

Plants that can survive in conditions of high salinity are known as **halophytes**. These generally survive by retaining enhanced levels of ions in their cells, thereby resisting loss of water to their environment, by osmosis. Such plants are typical of the salt marsh flora.

Mineral nutrients occur dissolved in water of the soil solution – generally at low concentrations. Essential elements for plants (and animals) are released by decay of dead organisms and their waste matter. For plant growth, the mineral salts required in substantial quantities – the macronutrients – include metal ions such as potassium and calcium, and the non-metal ions nitrate and phosphate. Other inorganic nutrients, required in much smaller amounts, are the micronutrients or trace elements (page 38).

As ions, micronutrients and macronutrients are absorbed and re-used by the plant. Nutrients are endlessly cycled, but some are lost from the system from time to time, for example, by leaching from the soils. Stocks of ions may be added to by weathering of parent rocks, but nitrates are generated from atmospheric nitrogen gas by the actions of nitrogen-fixing microorganisms, and as an outcome of lightning.

Introducing factors that affect the distribution of animals

The nutrition of animals is heterozoic, and many show **holozoic nutrition**, as humans do. Actually, the **study of food chains** (page 139) established the dependence of animals on plant life, either directly or indirectly, for their sources of energy and carbon. So very often it is the dominant plants in ecosystems that determine – or at the very least, strongly influence – animal distribution. In addition to being the principal source of nutrients, the dominant plant community typically determines the sorts of habitat that exist and the critical abiotic conditions.

As a consequence, the factors that affect the distribution of plant species are indirectly highly influential on animal distribution, and need to be kept in mind when investigating animal distribution.

Some of the more important environmental factors that determine the distribution of animals are temperature, water, food supply, territory and breeding sites. However, these factors seldom exert their effects alone.

Temperature variations both govern the rate of animal metabolism and growth, and also influence animal behaviour. Temperature changes may also be correlated with other physical factors, particularly light.

Animals may be divided into the **homeotherms** that regulate their body temperature, and the **poikilotherms** which maintain less control over their internal temperature. The latter have activities more determined by external temperatures, whereas the former remain active despite quite wide variation in environmental temperature.

Water is required by animals for use in ways similar to those of plants – simply because for them, too, water makes up the bulk of their bodies. In the processes of gaseous exchange and in regulation of body temperature, a great deal of water is lost by evaporation. Water is also lost in the disposal of waste matter. Terrestrial animals that possess adaptations to reduce water loss, such as an impervious body covering (like that of insects, reptiles, birds and mammals) occupy a wider range of habitats than animals with a body surface that permits a great deal of water loss by evaporation (like that of earthworms, woodlice and amphibians). However, so great is the need for water that no animals stray far from a source of fresh water for long. Reliable water sources (water holes, springs, and stream banks) are consequently places where carnivorous, hunting animals often concentrate their feeding activities.

Food supply is a factor influencing the existence and size of a local population. Some animals are dependent on plant matter (herbivores), some on animals (carnivores), and others eat both plant and animal matter (omnivores). Whatever the diet, when food supplies are short, individuals compete.

Competition is the term for interactions between individuals arising from their shared need for any limited resource. Either individuals of the same species compete (**intraspecific competition**), or individuals of different species do (**interspecific competition**). Competition for food is a major factor regulating population size. No population can be sustained for any length of time if it predates its prey population so severely that the organisms are threatened. (The same applies to herbivores that browse on particular plants.) Generally populations are regulated well below that level, because any food source that is over-grazed or over-predated, would itself crash. The result would be that the predator or browser population would then also crash – and so on, up the food chain.

Along with competition for food goes competition for space. **Territory** is a defended area of a habitat. Territoriality is a feature of the life style of many species of birds, mammals and a few non-vertebrates. A territory may be established by an individual, by breeding pairs, or by family groups. Many territories are established for a breeding period, but the same individuals then form into co-operative flocks at other times. Typically, we think of territory as a permanently held space, but some nomadic herds effectively mark out their territory in time, using scent markings as they travel, letting competitors know how recently they passed through the area.

Defence of a territory, particularly by higher animals, is through active behaviour, but fighting may be kept to a minimum. Territory owners assert their ownership by conspicuous display or song. Aggression, when it arises, is most often ritualised, with body language positively displayed to declare displeasure, threat, and (eventually) submission on the part of one party to the dispute.

Breeding sites include the areas – nests or homes – in which the young are fed and reared. Typically they are placed at the heart of a territory. The amount of protection a site affords, and the accessibility of food resources, affect the survival chances of the young.

Estimating population size and distribution in a habitat

An important step in the study of a habitat is to collect accurate information on the size of populations present. For practical reasons, a random sample is selected for this, involving either a part of the area of the habitat, or a restricted number of the whole population (Figure 19.2). By means of random sampling, every individual of the population has an equal chance of being selected, and so a representative sample is assured.

Figure 19.2 Random locating of quadrats

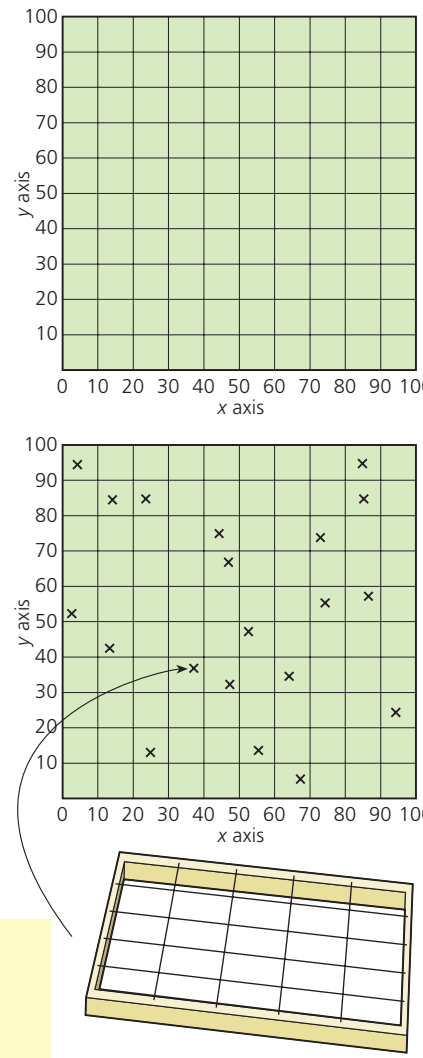
1 A map of the habitat (e.g. meadowland) is marked out with gridlines along two edges of the area to be analysed.

2 Co-ordinates for placing quadrats are obtained as sequences of random numbers, using computer software, or a calculator, or published tables.

3 Within each quadrat, the individual species are identified, and then the density, frequency, cover or abundance of each species is estimated (see Figure 19.3).

4 Density, frequency, cover, or abundance estimates are then quantified by measuring the total area of the habitat (the area occupied by the population) in square metres. The mean density, frequency, cover or abundance can be calculated, using the equation:

$$\text{population size} = \frac{\text{mean density (etc.) per quadrat} \quad \text{total area}}{\text{area of each quadrat}}$$



Random sampling of plant populations, using quadrats

Quadrats are commonly used to estimate plant populations (Figure 19.3). A quadrat is a square frame which outlines a known area for the purpose of sampling. The choice of size of quadrat varies depending on the size of the individuals of the population being analysed. For example, a 10 cm^2 quadrat is ideal for assessing epiphytic *Pleurococcus*, a single-celled alga, commonly found growing on damp walls and tree trunks. Alternatively, a 1 m^2 quadrat is far more useful for analysing the size of two herbaceous plant populations observed in grassland, or of the earthworms and the slugs that can be extracted from between the plants or from the soil below.

Quadrats are **placed according to random numbers** after the area has been divided into a grid of numbered sampling squares (Figure 19.2). The different plant species present in the quadrat may be identified. Then, using the quadrat, an observer may estimate the **density, frequency, abundance** or **cover** of plant species in a habitat. The steps are outlined in Figure 19.3, including the issue of how many quadrats may need to be placed for an accurate assessment of population size. An area of chalk grassland is illustrated in which the use of quadrats would give estimates of the population sizes of two species of flowering plants (Figure 19.4).

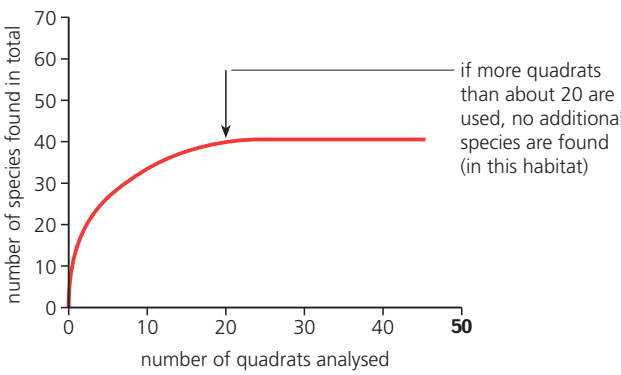
Figure 19.3 Estimating plant population size using a quadrat

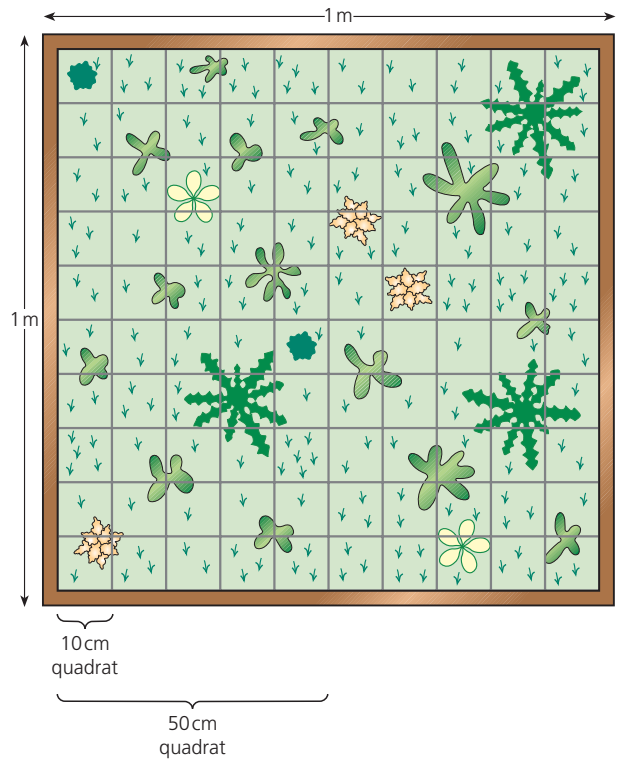
Use of the quadrat:
- positioned at random within habitat being investigated
- different species present are then identified
- without destroying the plants present and the microhabitats beneath them, plant species' density, frequency, abundance or cover can be estimated.

density = mean numbers of individuals of each species per unit area (time-consuming and may be hard to assess separate individuals)
frequency = number of quadrats in which a species occurs, expressed as % (rapid and useful for comparing two habitats)
cover = the % of ground covered by a species (useful where it is not possible to identify separate individuals)
abundance = subjective assessment of species present, using the DAFOR scale: D = dominant, A = abundant, F = frequent, O = occasional, R = rare (same observer must make 'abundance' judgements, which may be useful as comparisons of two or more habitats, rather than objective scores)

What is the optimum size of quadrat? This varies with the habitat, and the size of plants found. Look at the example here. In the 1 m quadrat there are six species present. How many different species are counted in the quadrat of sides 10, 20, 30, 40, 50, 60, 70, 80 and 90 cm? The optimum quadrat size is reached when a further increase in size adds no or very few further species as present.

How many quadrats? When there is no further increase in the number of species found, sufficient quadrats have been analysed in that habitat.





Estimating species distribution by means of a transect

Whereas some communities are relatively uniform over a given area and are suitable for random analysis (such as the grassland habitat shown in Figure 19.4), others show a trend in **variation in a particular direction**. Examples include seashore, pond or lake margin, salt marsh or even an area where there is a change from dry soil to wet land. The appropriate technique to study such a trend of variation is the **transect**.

Figure 19.4 An area of chalk grassland with a rich flora

Transects are a means of sampling biotic (and abiotic) data at right angles to the impact of unidirectional physical forces. Although there may only be time to study one transect in detail (and this may be sufficient as a demonstration of the zonation of communities), a single transect may not provide an adequate sample or give an indication of differences from place to place. Transects should, therefore, preferably be replicated several times.

Where transects are carried out across a habitat where the land changes in height and where level is an important factor (such as a seashore, salt marsh or pond margin), then the changes in level along the transect line can be measured and recorded as a **profile transect**. The surveying for this requires the use of survey poles and a field level device (Figure 19.5).

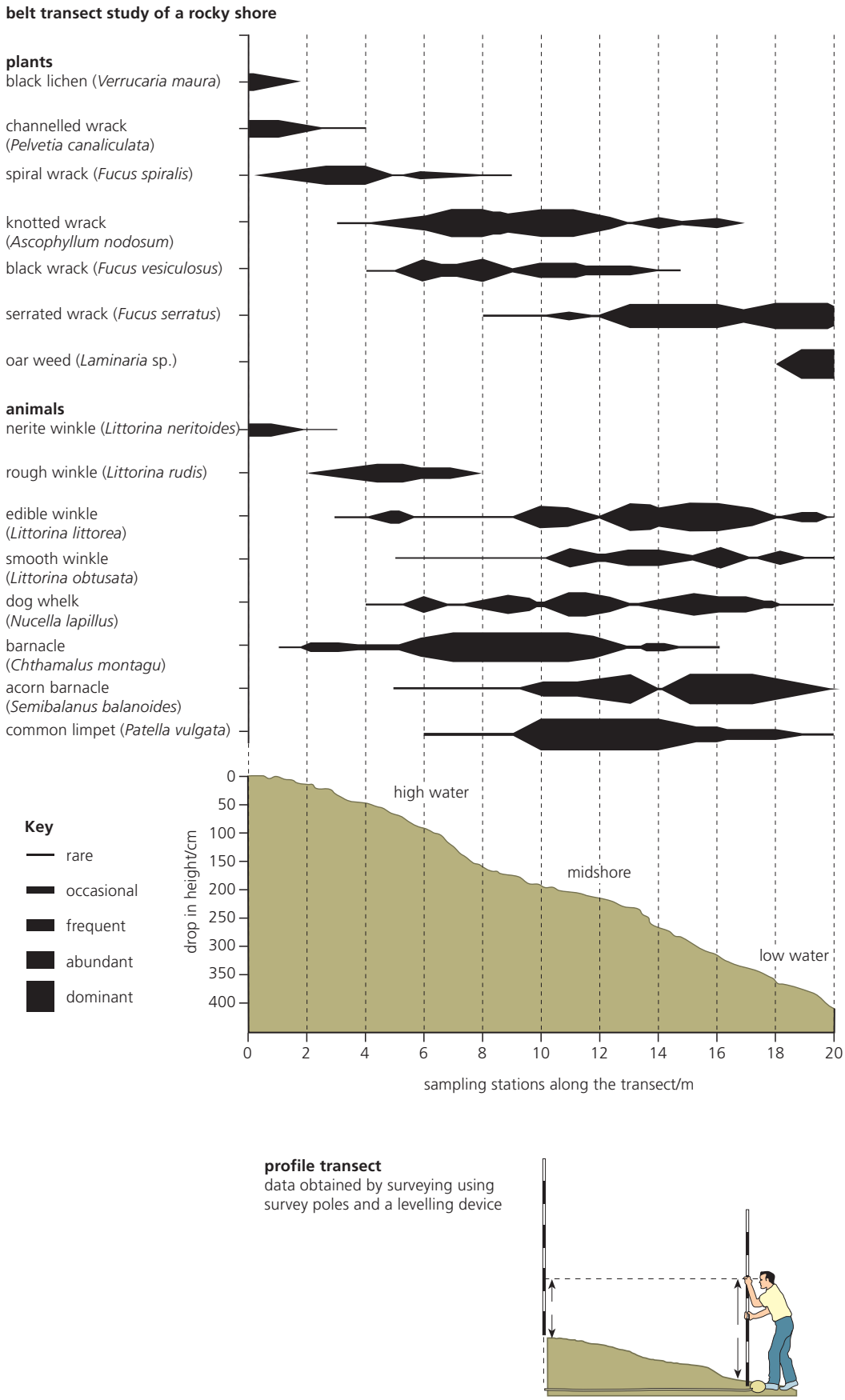


Figure 19.5 Profile and belt transect analysis of a rocky shore community

The community present along a transect can be analysed from a straight line such as a measuring tape, laid down across an apparently representative part of the habitat. The positions of every organism present that touches the line are recorded either all the way along the line or else at regular intervals. The result is a **line transect**.

A **belt transect** is a broad transect, usually half a metre wide. To produce it, a tape measure or rope is laid as for a line transect, but this time the organisms in a series of quadrats of half-metre width are sampled at (say) metre intervals. If the community changes little along the transect, then quadrats can be placed less frequently. Along with data on the biota, data on abiotic variables can be measured and recorded along the transect. For example, along a terrestrial transect the pH of the soil might be measured.

The results of a belt transect study of a seashore community are shown in Figure 19.5. The seashore (known as the littoral zone) is a part of the extreme margins of continents and marine islands periodically submerged below sea water, and so affected by tides. Tides are the periodic rise and fall of the sea level due to the attractions (gravitational pull) of the Moon and Sun. The shore is an area very rich in living things, and almost all are of marine origin.

The higher an organism occurs on the shore, the longer the daily exposure to the air. Exposure brings the threat of desiccation and wider extremes of temperature than those experienced during submersion. Exposure is an abiotic factor that influences distribution of organisms on the seashore.

1 From the data in Figure 19.5, **deduce** a plant and an animal species that appear to be well adapted to the degree of exposure experienced at:

a a high water location of the shore

b a low water location of the shore.

The niche and competitive exclusion

The **niche** is an ecological term that defines just how an organism **feeds**, where it lives, and how it **behaves** in relation to other organisms in its habitat. The niche concept is useful because it identifies precise conditions which a species needs.

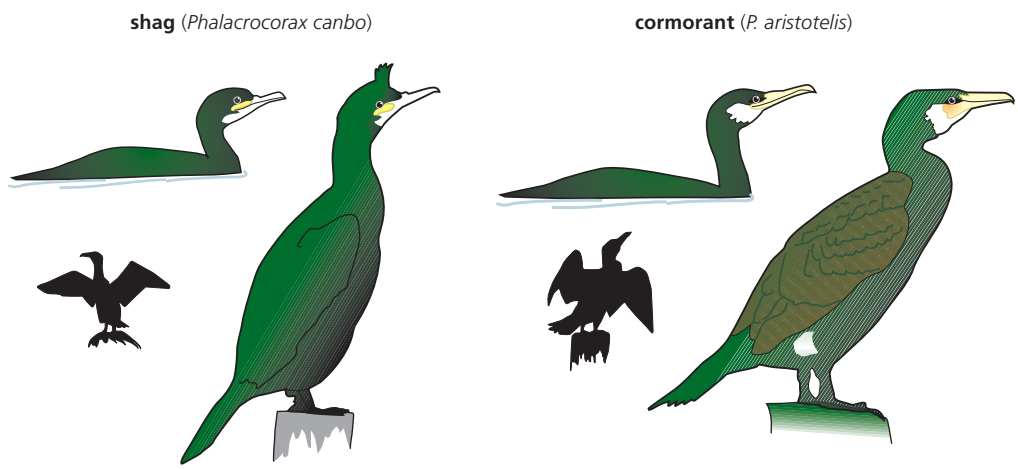


Figure 19.6 Different niches of the cormorant and the shag

diet is a key difference in the niches of these otherwise similar birds

prey		% of prey taken by shag	% of prey taken by cormorant
surface-swimming prey	sand eels	33	0
	herring	49	1
bottom-feeding prey	flatfish	1	26
	shrimps, prawns	2	33

The principle of distinct niches can be illustrated by two common and rather similar seabirds, the cormorant and the shag (Figure 19.6). These birds live and feed along the coast line, and they rear their young on similar cliffs and rock systems. We can say that they apparently share the same habitat. However, their diet and behaviour differ. The cormorant feeds close to the shore on seabed fish, such as flatfish. The shag builds its nest on much narrower cliff ledges. It also feeds further out to sea, and captures fish and eels from the upper layers of the waters. Since these birds feed differently and have different behaviour patterns, although they occur in close proximity, they avoid competition. They have different niches.

Competition refers to the interaction between two organisms striving for the same resource. For competing organisms, the more their niches overlap, the more both will strive to secure a finite resource. Hence, potential competitors have evolved different niches.

However, another possible outcome may be an equilibrium situation in which neither competitor succeeds as it might, yet both survive. This was found to be the case with two species of clover (*Trifolium* spp.) in an experiment in which they were grown together and separately, and the resulting growth rates compared (Figure 19.7).

2 **Suggest** the most likely finite resources that the two species of *Trifolium* (Figure 19.7) were competing for.

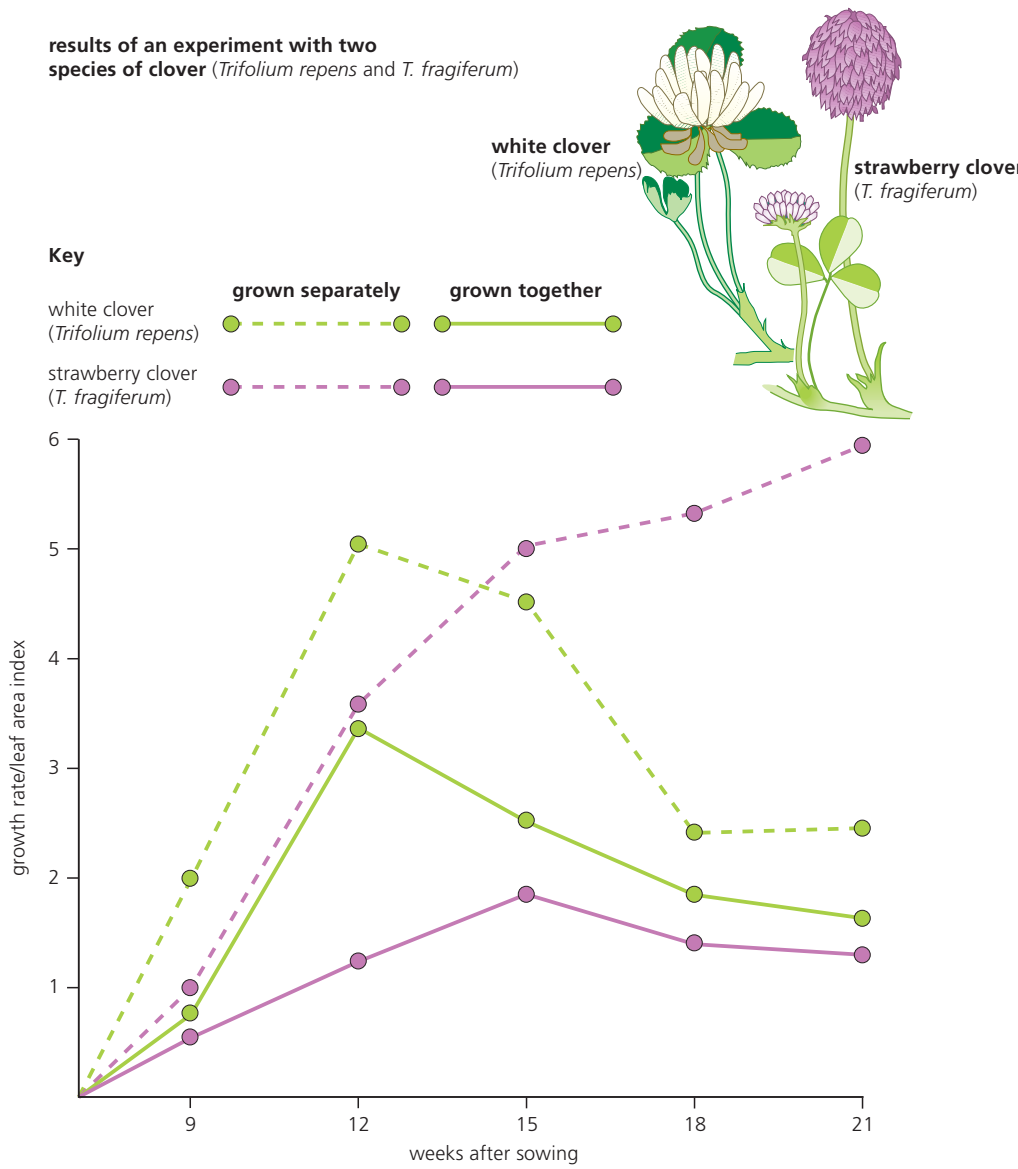


Figure 19.7 Competing organisms in equilibrium

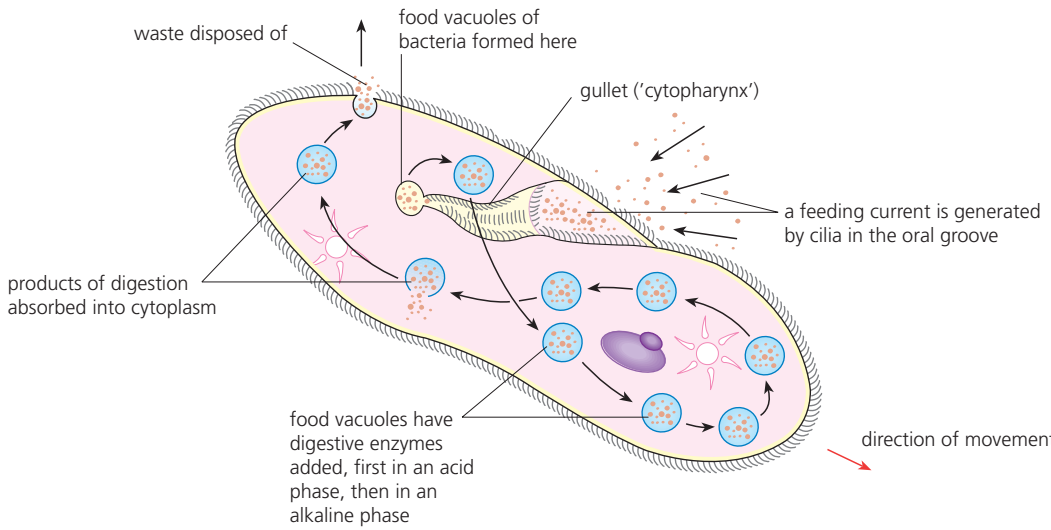
Fundamental and realised niches

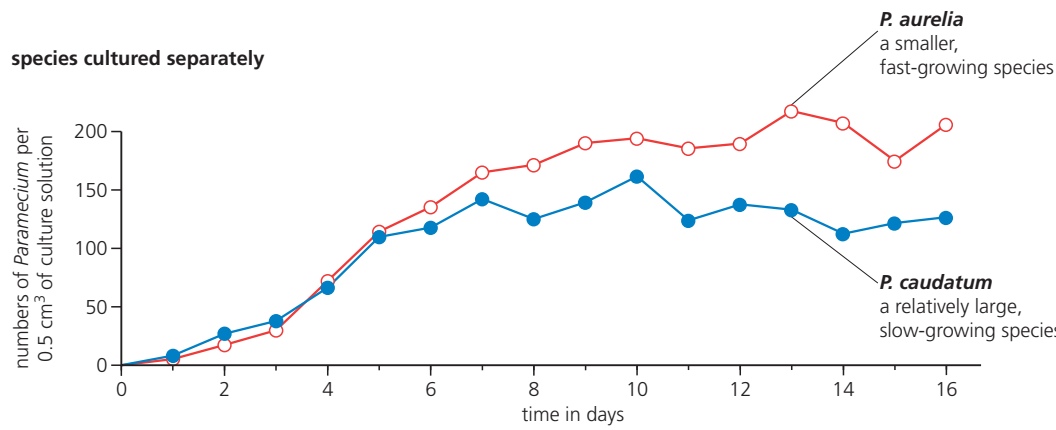
Most species of plants and animals can survive in a range of conditions – in effect, their niches are fairly unrestricted ones. For example, animals can mostly eat a wide range and variety of food sources. This range of conditions an organism can tolerate is defined as its **fundamental niche**.

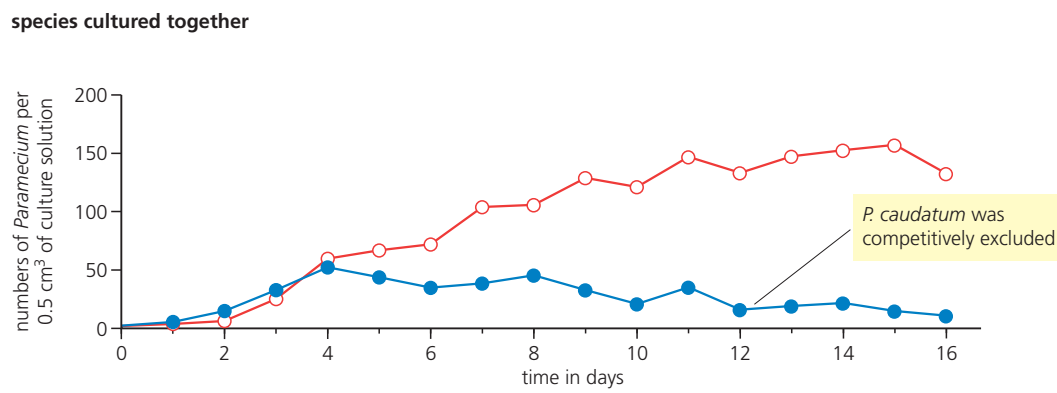
The fundamental niche of a species is the potential mode of existence, given the adaptations of the species.

Figure 19.8 Population growth curves for two species of *Paramecium*

An experiment carried out by G. C. Gause in 1934 using species of *Paramecium*, a large protozoan common in fresh water. It feeds on plankton, the food source used in these experiments.







Only when organisms experience resources and conditions completely outside this range are they fatally threatened by their environment, and so die out. In practice, it is the experience of most organisms that only part of their fundamental niche is available to them because of a degree of competition with other organisms. The portion of the niche that is occupied in these cases is the **realised niche**.

The realised niche of a species is the actual mode of existence, which results from its adaptations and competition with other species.

Competitive exclusion principle

Competition between two organisms of different species (interspecific competition) which results in one competitor taking the resource exclusively, and the other being driven out and away, or even dying out, is another possible outcome wherever two species compete intensively. When two species cannot co-exist, this is known as **competitive exclusion**.

The **competitive exclusion principle** is the idea that ecological separation of closely related or otherwise similar species is the inevitable outcome. So, if two species share the same resource at the same place and the same time, then the dominant species will out-compete the other, which will either die out or move away to avoid the competition. This principle was first suggested by a Russian biologist, G. F. Gause in 1934, based on his experiments culturing different species of *Paramecium* in the laboratory (Figure 19.8).

We can see that competitive exclusion could be the pressure that causes closely related species living in close proximity to have evolved clearly defined but separate niches in which competition between them is at an end. This would occur over an extended period of time. For example, the competitive exclusion principle would account for the difference in niches between the cormorant and the shag that we noted above.

Since it was first proposed, the competitive exclusion principle has been experimentally tested in many laboratory studies in which pairs of species have been reared together under constant conditions. Much convincing evidence has been obtained.

***Chthamalus* sp.** is a barnacle able to withstand prolonged exposure – it survives higher on rocky shore lines, where other species of barnacle cannot compete.

***Semibalanus* sp.** is a barnacle easily crowded out from more exposed positions on the shore. It is mostly found lower down, in the mid to lower shore region.

Figure 19.9 The growth of two species of barnacle on the seashore. Barnacles have a free-swimming larval stage and tend to settle and colonise all surfaces of intertidal zones of rocky shores. Where they are subsequently found in large numbers depends on their ability to survive periods of exposure (between high tides)

However, the demonstration of its operation in the environment is more difficult. Two species of barnacle that occur on the seashore do appear to be a case in point (Figure 19.9). One species (*Chthamalus*) is able to withstand prolonged exposure when the tides recede, conditions that kill off the other species (*Semibalanus*). But before the sedentary adult stage, the free-swimming larval stages of these crustaceans colonise any surface in the intertidal zone on which they happen to settle. Only later, as the degree of exposure is experienced by the growing barnacles, are *Chthamalus* barnacles crowded out from lower zones, and *Semibalanus* barnacles crowded out from exposed zones. We see here that it is a difference in resistance to periodic desiccation that has determined where each species survives best.

3 **Illustrate** and **explain** the competitive exclusion principle by reference to a species in a habitat you have studied in your locality.

The range of interactions between species within a community

Resources of many sorts are often in limited supply in the environment, and so organisms must compete for them. For example, plants may compete for space, light and mineral ions. Animals may compete for food, shelter and a mate. Whenever resources are in short supply, they become a limiting factor.

Competition between individuals of the same species, **intraspecific competition**, occurs when individuals compete for a mate, for example.

Competition between individuals of different species, **interspecific competition**, may be so intense as to cause the exclusion of one species by another. Examples of interspecific competition are species of *Paramecium* competing for plankton, and species of clover (*Trifolium*) competing for mineral ions, as already noted.

What other forms of interactions between species occur?

Herbivory is the feeding on plant matter (browsing, in effect) by animal species (Figure 19.10). Herbivorous mammals include the ruminant species, some of which like cattle and sheep have been domesticated, while others are wild animals of grasslands and savannah, such as the wildebeest. Non-vertebrate herbivores include the leaf-eating caterpillars (the larval stage of a butterfly) such as that of the monarch butterfly, feeding on the leaves of milkweed plants. The monarch is a migratory insect of the American continent.

Predation is preying of one animal on another for food (Figure 19.10). Predation frequently requires patient stalking of prey and a determined final attack because most prey species invest time and care in their own or the whole herd's defence. We see this, for example, in the hunting of a juvenile or an isolated or ailing wildebeest, part of a large herd, by a pride of lions living beside their prey among the long grass of the African savannah. Not all cases of predation are so dramatic. A ladybird insect alights on a plant stem where aphids are feeding, their probosces inserted deep into the plant stem tissues, and systematically eats its fill of the defenceless aphids.

Parasitism involves two organisms living together for all or most of the parasite's life cycle, and feeding at the expense of the host's living tissues (Figure 19.10). The relationship is, at the very least, unhelpful or possibly very **harmful to the host**. The parasite has a supply of nutrients on hand, but may experience resistance by the host which has to be overcome. Movement between hosts is another period of danger to parasites. Both plant and animal parasites exist.

One example of an animal parasite is the protozoan *Plasmodium*, which lives in blood and liver cells of humans and causes malaria (see Figure 18.38, page 595). Chance transfer of *Plasmodium* between hosts may be achieved when a female mosquito takes a blood meal from an infected human, and later feeds on another human, not yet infected with malaria. *Plasmodium* is an **endoparasite** as it resides within its host.

Ticks are small arachnids (related to spiders) that attach themselves to the skin of endothermic (warm-blooded), furry mammals. They are **ectoparasites**. They bite into the skin sufficiently to draw out blood meals. Ticks on dogs, rabbits, hedgehogs, and many similar hosts tend to drop off when satiated with blood, often where the mammal rests. They re-attach later, and they continue to feed and breed in the vicinity of their hosts.

Mutualism is a form of symbiosis in which the two organisms of different species, living in intimate association, do so to **mutual advantage** (Figure 19.10). Mutualism is an example of favourable interaction between organisms. Simple examples are the egrets that pick ticks off the backs of water buffalo, or cleaner fish that remove sea lice from big fish.

One good example of mutualism is the huge community of microorganisms, mainly bacteria and fungi, that live in the rumen of ruminant mammals. The rumen is a large, stomach-like fermentation vessel. Here the vegetation the herbivore has eaten and begun to chew up is received. The microorganisms break down the cellulose and other polymers in the plant matter, producing organic acids (and a great deal of methane). The ruminant is dependent on this enzymic digestion in order to gain nutritional benefit from its diet. For the microorganisms, the rumen is their habitat (until digested by ruminant enzymes as they pass further down the gut of the herbivore host).

Another example of mutualism is the **mycorrhizal association** between the roots of plants, particularly trees, and the hyphae of particular species of soil fungi. The fungus benefits from sugar supplied by the plant, and in turn, feeds back to the tree's roots the excess mineral ions, like phosphate, potassium and nitrate, that the fungus absorbs from the soil immediately they are released there (often at times of the year when the tree's growth is minimal). Given the importance of trees as a resource for human activities, the mycorrhizal relationship is of great ecological and economic significance.

Figure 19.10 Interactions between species

Top left: Herbivory – caterpillars of the monarch butterfly feeding on milkweed leaves

Top right: Predation – African lion at the moment of capture of prey (kudo – a savannah herbivore)

Bottom left: Parasitism – sheep tick (an ectoparasite) attached to the skin of a cat where it has fed on a blood meal

Bottom right: Mutualism – mushroom of fly agaric fungus takes sugars and amino acids from a tree's roots in return for essential ions, via its hyphae attached below ground

Ecosystems and biomes

G1.9–1.10 and G2.1–2.11

Ecosystems are stable, settled units of nature consisting of a community of organisms, interacting with each other and with their surrounding physical and chemical environment. Examples of ecosystems are ponds or lakes, woods or forests, seashores or salt marshes, grassland, savannah or tundra. From this list you can see that any particular ecosystem is variable in size, but each represents a space favourable for the growth of a community of organisms, all of which constitute the biomass of the ecosystem.

Biomass is the total weight (or volume, or energy equivalent) of living organisms in a given area (e.g. a quadrat).

The rate at which biomass is produced by an ecosystem is an expression of the productivity of the ecosystem. Biomass is commonly expressed as mass or energy per unit area; for example, kilojoules per square metre per year (kJ m^{-2} yr^{-1}).

Productivity within an ecosystem

Productivity has two components: **primary productivity**, the production of new organic matter by green plants (autotrophs), and **secondary productivity**, the production of new organic matter by consumers (heterotrophs). Both of these can be divided into gross productivity and net productivity.

Gross productivity is the total amount of organic matter produced.
Net productivity is the organic matter of organisms less the amount needed to fuel respiration.
Gross production – respiration = net production.

Measuring the biomass of different trophic levels in a community

The level at which an organism feeds in a food chain is called its **trophic level** (page 142). Producers (green plants) are designated as trophic level 1, because their energy has been transferred once, from Sun to plant. Herbivores constitute level 2, because here energy has been transferred twice, and so on. The different trophic levels in a community are:

- producers — trophic level 1
- primary consumers — trophic level 2
- secondary consumers — trophic level 3
- tertiary consumers — trophic level 4.

How can we measure the biomass at different trophic levels?

To do this in the field first involves estimating the numbers of organisms of each type at each trophic level in a representative part of the community. Then, if the mass at each trophic level is to be expressed (as it commonly is) as grams dry mass, then it is necessary to find the **dry mass** of a representative sample (at least) of each type of organism.

The dry mass is found by heating a weighed sample to a temperature of about 80 °C. This drives off the water, but is not hot enough to burn away any organic matter. Cooled samples are weighed and heated, cooled and re-weighed. This is repeated until two consecutive readings give the same mass, showing that all the water has been driven off.

Of course, to obtain dry mass measurements the representative samples of organisms have first to be killed. Consequently, communities of organisms are destroyed by this technique, and the area of the habitat they occupied is reduced to being a 'desert' – at least until it is repopulated from surrounding communities.

This destructive technique raises challenging ethical issues, and these are of increased pertinence at a time that biodiversity is under threat. A less destructive approach would involve finding the **fresh mass** of a small representative sample of organisms all or most of which can then be returned to the environment. However, it is less accurate.

4 **Identify** the reasons why fresh mass is a much less reliable measure of biomass than dry mass.

Calculating productivity

A study of the ecological productivity of a fresh-water site in Florida was published in 1957 by Professor Eugene Odum of the University of Georgia, USA (Table 19.1). In this painstaking work, species present were divided into their four trophic levels. Data were obtained on the biomass organisms at each level acquired by feeding (herbivory or predation). Losses by primary, secondary and tertiary consumers due to respiration were calculated. Results were expressed in energy values (kJ m^{-2} yr^{-1}).

	Biomass expressed in terms of energy content (kJ m^{-2} yr^{-1})		
Trophic level	**Matter retained**	**Lost in respiration**	**Exported due to herbivory or predation by organisms of next trophic level**
producers (energy input in the form of sunlight)	1 695	50 112	35 263
primary consumers	6 506	27 154	1 602
secondary consumers	192	1 322	88
tertiary consumers	33	54	

Table 19.1 Ecological productivity by the fresh-water community at Silver Springs, Florida

We can use these data to calculate values for gross productivity and net productivity by the consumers present.

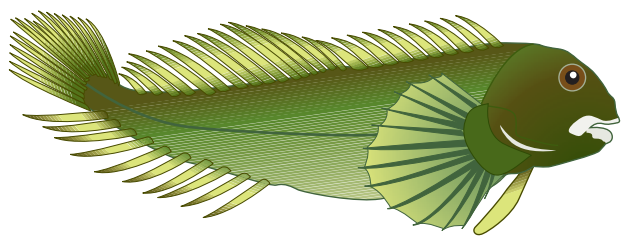
The gross production by the primary consumers (herbivores that browsed on the producers – the green plants) was the biomass that is recorded as exported by the primary producers, that is 35 263 kJ m^{-2} yr^{-1}.

The respiratory loss by the primary consumers was 27 154 kJ m^{-2} yr^{-1}.

So the net production by the primary consumers was:

35 263 – 27 154 = 8109 kJ m^{-2} yr^{-1}.

5 Calculate net production by the secondary and the tertiary consumers of the fresh-water community at Silver Springs (Table 19.1).



common blenny (*Blennius pholis*) ×2

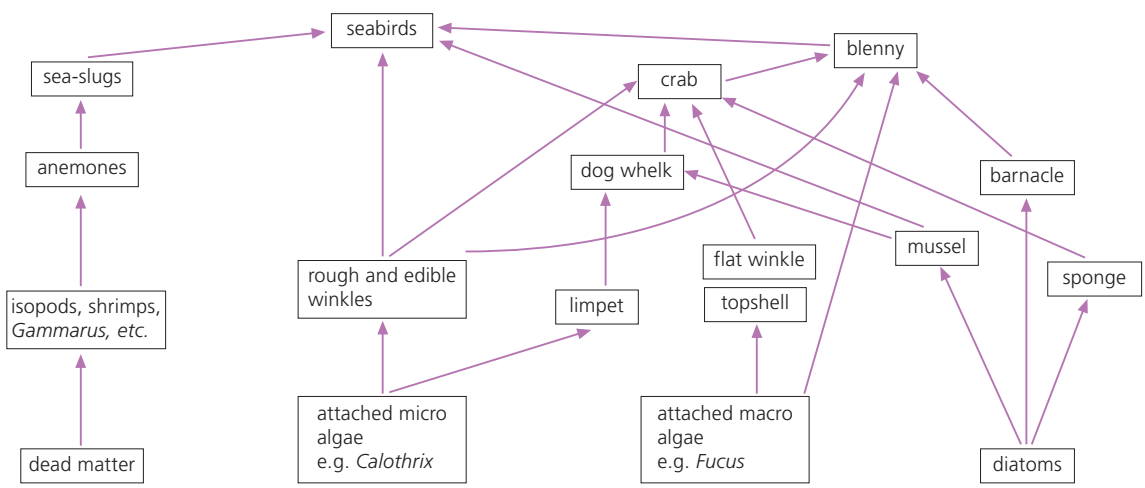


Figure 19.11 Simplified food web for a community of a rocky shore

Issues of trophic levels

The **trophic level** is a level in a food chain defined by the method of obtaining food, and in which all organisms are the same number of energy transfers away from the original source of energy (the Sun and photosynthesis).

For studies of the productivity of whole communities, it is first necessary to designate to each organism identified, a trophic level. This is clearly a straightforward task in the case of the producers (they are level 1), but after that, problems may arise.

For example, for many organisms, **the trophic level may vary with each item of the diet**. In the simplified food web for a rocky shore (Figure 19.11) the blenny (a small marine fish of inshore waters) feeds as a primary consumer (of *Fucus*, brown seaweed), as a secondary consumer (of winkles), and as a tertiary consumer (of crabs). In terrestrial food webs, too, it is common to find the top carnivores feed at trophic levels between 2 and 5, at any one time.

6 **Analyse** the number of trophic levels occupied by seabirds in the food web in Figure 19.11.

A similar difficulty arises with **omnivores** in all food webs. Omnivores necessarily feed at two trophic levels, at the very least. By definition, they are primary consumers and secondary consumers.

Another pertinent factor in the collection of data over days, weeks or months is that the **diets of many animals change** dramatically at different stages in their growth and development.

Again, the consumers of dead organic matter, the **detritivores and decomposers** that break up dead organisms and decay complex organic matter to simple, inorganic nutrients, are not featured in food webs. Yet up to 80% of plant matter produced may be consumed in this way.

So, for experimental ecologists, the quantification of food webs has many unresolved difficulties.

TOK Link

It is said that 'observations provide a secure basis from which knowledge may be derived'. Does this apply to the apparently simple task of classifying organisms into trophic levels?

Constructing a pyramid of energy

Pyramids of numbers of organisms present, or of biomass, or of energy, give a proportional representation of the organisms present at successive trophic levels. Quantitative measurements are required, and are then represented on graph paper by areas proportional to the data being used – basically, numbers, biomass or energy of the organisms sampled in a defined area.

We can use the data assembled in Table 19.1 to draw up a pyramid of energy for the Silver Springs fresh-water community (Table 19.2 and Figure 19.12). Note that the energy acquired at each of the four trophic levels is the total of energy recorded in the table as matter retained, lost in respiration, and exported.

Table 19.2 Trophic level and total energy (from Table 19.1)

Trophic level	Energy ($kJ\ m^{-2}\ yr^{-1}$)
1	87 070
2	35 262
3	1 602
4	87

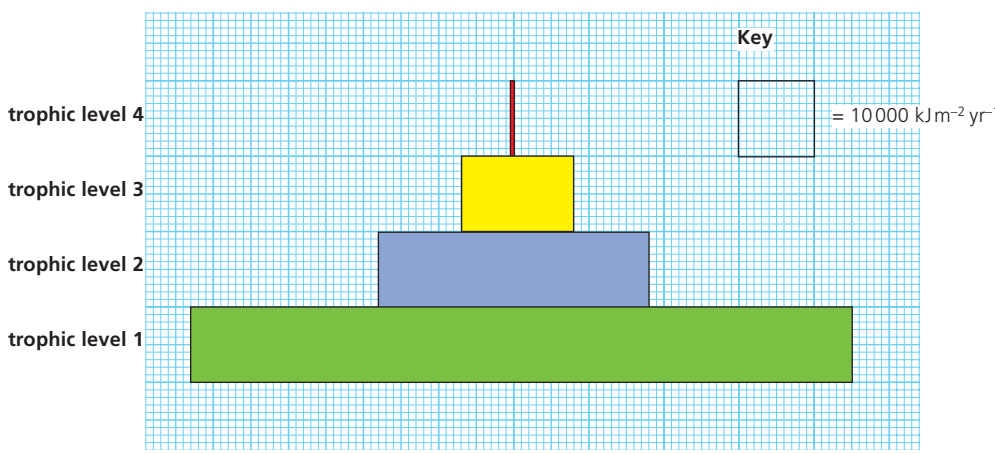


Figure 19.12 Pyramid of energy for the Silver Springs fresh-water community

Ecological pyramids, such as the pyramid of energy for the Silver Springs community (Figure 19.12), convey a visual impression of the relative importance of the trophic levels in the ecosystem. Particularly evident are the low energy levels at the higher trophic levels – reflecting the relatively small biomass and low numbers of organisms at these levels.

Why are the numbers of organisms low at higher trophic levels?

Largely, this occurs because there are significant losses of energy in each trophic level. This occurs as a result of respiration – animals invest efforts into finding food, digesting food, and in processes like reproduction and the caring of their young. All these activities require a supply of energy. Also, a proportion of the biomass at each level is not eaten by the next trophic level for various reasons (e.g. deep roots of producers are rarely available to primary consumers). In addition, losses in the form of the faeces (including undigested food matter) and urine have to be taken into account. Deep roots and faeces mostly pass directly to decomposers. Consequently, typically only 10% of the matter and energy at any one level becomes matter in the organisms of the next level (Figure 19.13).

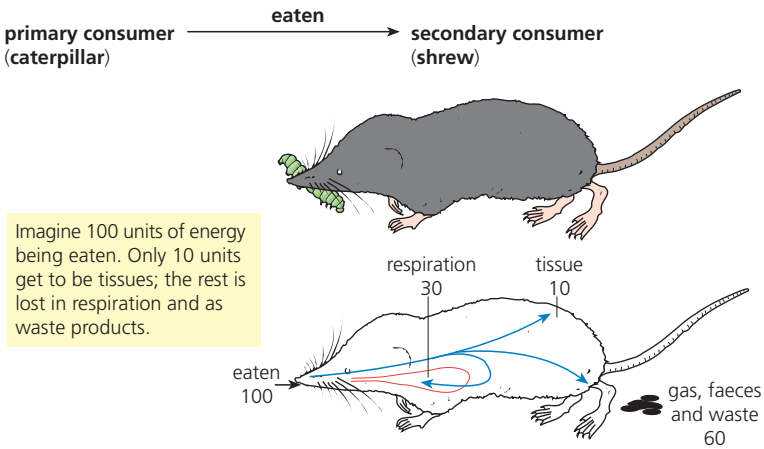


Figure 19.13 Loss of energy between trophic levels

Additionally, the top carnivores may hunt for their food, which often has to be killed. In the chase, more energy is transferred, prey are sometimes injured but escape to die elsewhere, and perhaps become food for other organisms, at a lower trophic level. So, at each boundary in a food chain, the principal feature is that 90% of what is eaten at one trophic level is not available to the next level in the chain.

Ecological successions

When new land is exposed it is quickly invaded and colonised by organisms. In fact, a sequence of communities develops with time by a process known as **ecological succession**. The stages in a succession are a series of plant and animal communities of increasing complexity, that develop with time. These stages are called **seral stages**, and the whole process is termed a **sere**. A **climax community** finally results (if the site remains undisturbed) which is characteristic of the area. We illustrate this in typical primary and secondary successions.

Primary and secondary successions

When the succession sequence starts on entirely new land **without established soil**, then the process is known as a **primary succession**.

New land is formed on the Earth's surface at river deltas, at sand dunes, and from cooled volcanic lava, for example. Primary successions also develop in aquatic habitats, such as in a pond formed and fed by a spring. In all cases, the first significant development in a primary succession is the formation of soil.

Alternatively, sometimes established communities are suddenly disrupted and totally destroyed. This occurs, for example, when fire destroys a large area of vegetation; occasionally it occurs as a result of human activities. In these situations, soil is already formed and present – it is the existing biota that has been abruptly removed. A succession that starts from **existing soil** is known as a **secondary succession**.

Successions and seral stages

At the initial site of a primary succession, all that may be present is **parent rock**, from which the bulk of the **soil** is formed by erosion. Erosion is the breaking of solid rock into smaller particles by the effects of extremes of temperature, by the action of wind and water, and by chemical reactions that occur, such as when slightly acid rain falls. Mineral particles may also be blown or washed in from elsewhere. The resulting **mineral skeleton** is of particles of a wide range of sizes from small stones and coarse sand to the finest clay particles.

Soil, when fully formed, has organic matter called **humus** wrapped around the particles of the mineral skeleton. Humus is a substance derived from dead plant and animal remains, together with animal faeces, that have been decomposed by the actions of microorganisms. Humus is a dark-coloured, sticky substance, that continues to be decayed, releasing mineral nutrients. Humus also helps the soil to hold **water**. Between mineral particles and humus linings are innumerable pockets of **air**. Also present in a developed soil is a huge community of **microorganisms** and small animals, including **earthworms**, all adapted to life in this habitat.

Humus is first added by plant invaders of the primary succession, known as **pioneer plants**. These are typically lichens and cushions of moss, and after them, tiny herbaceous plants, many with features that help them survive where water is scarce.

Until the soil is fully formed it retains little water, even when water is freely available. Plants able to survive drought are called xerophytes, and their special features are called xeromorphic features (page 316). When a sere starts from dry conditions, then the sere is called a **xerosere** (Figure 19.14).

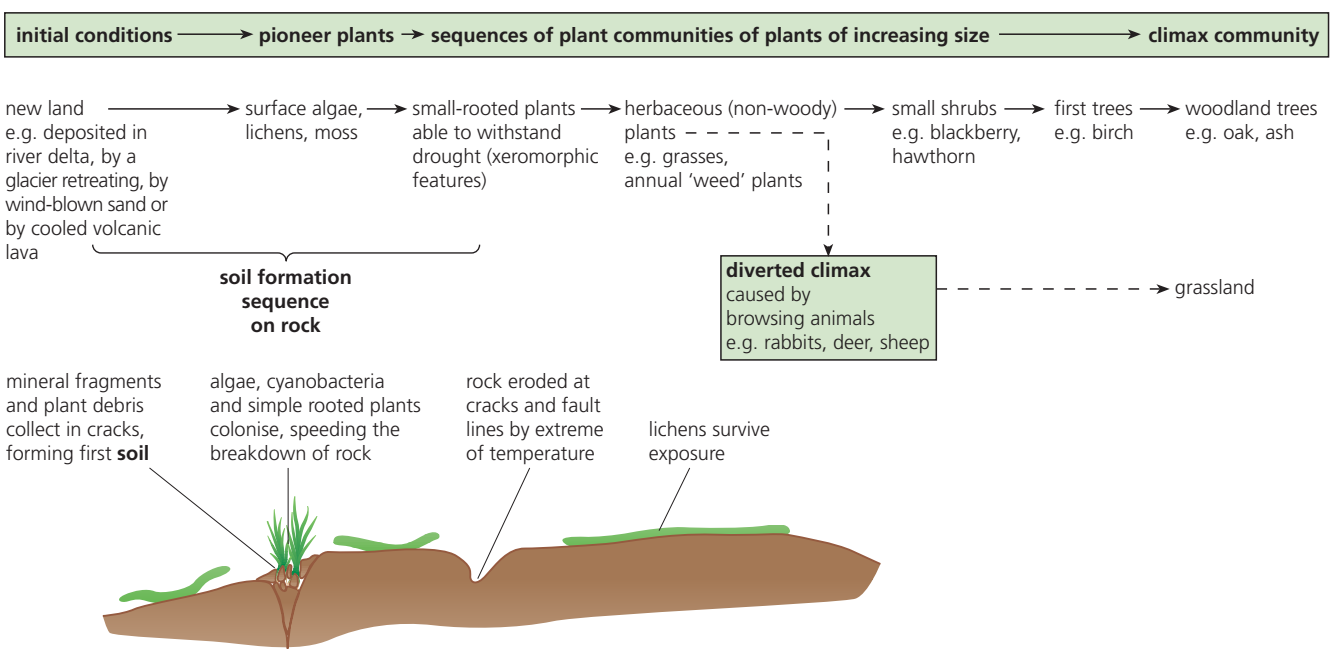
The growth and death of the early plant communities continue to add humus, and so more soil water is retained. Nutrients are added to the soil when organisms die, and the range of nutrients available to plants increases steadily. Nutrients are the **ions essential for plant growth**, such as nitrates, phosphates, and a range of micronutrients. Different plants now grow – various herbaceous weeds, for example, may start to shade out the pioneers. The conditions in the soil are increasingly favourable to microorganisms and soil animals, which continue to invade the habitat from the surroundings. Herbaceous plants are followed by shrubs and small trees, all growing from seeds that are carried in by wind, water or the activities of animals.

The first-formed soil accumulates relatively slowly, because much is vulnerable to erosion and dissipation by wind and heavy rains. Increasingly, as plant life becomes established, the growth of plant roots and the cover provided by plant vegetation prevent or reduce **loss of soil by erosion**.

So, succession can be seen as a directional change in a community with time. Initially, **abiotic factors** have the greater influence on the survival and growth of organisms. Later, as the numbers of living organisms build up, **biotic factors** increasingly affect survival too.

Figure 19.14 A primary succession on dry land – a xerosere

A **xerosere** = succession under dry, exposed conditions where water supply is an abiotic factor limiting growth of plants, at least initially.



This succession sequence is not a rigid ecological process, but an example of what may happen. It is influenced by factors like:
a how quickly humus builds up and soil forms
b rainfall or drought, and the natural drainage that occurs
c invasions of the habitat by animals and seeds of plants.

The eventual outcome of the succession is perhaps a stable woodland community (an example of a **climax community**) where once bare rock or dry sand stood. Whether or not this specific community is the outcome, an important feature of a succession is the **progressive increase in the number of species present**. As more and more species occur in a habitat, the food webs are likely to be more diverse and complex. If so, in the event that one population crashed, such as when a disease sweeps through the members (or predators of a population have a temporary population explosion), then alternative food chains may be sufficient to supply the higher trophic levels.

If the primary successions develop in aquatic habitats, then the sequence of pioneer plants differs from a dry land primary succession, but the result may well be a woodland climax community, too (Figure 19.15). The succession is called a **hydrosere**. Plants adapted to aquatic or permanent swamp conditions are known as hydrophytes.

Figure 19.15 A primary succession from open water – a hydrosere

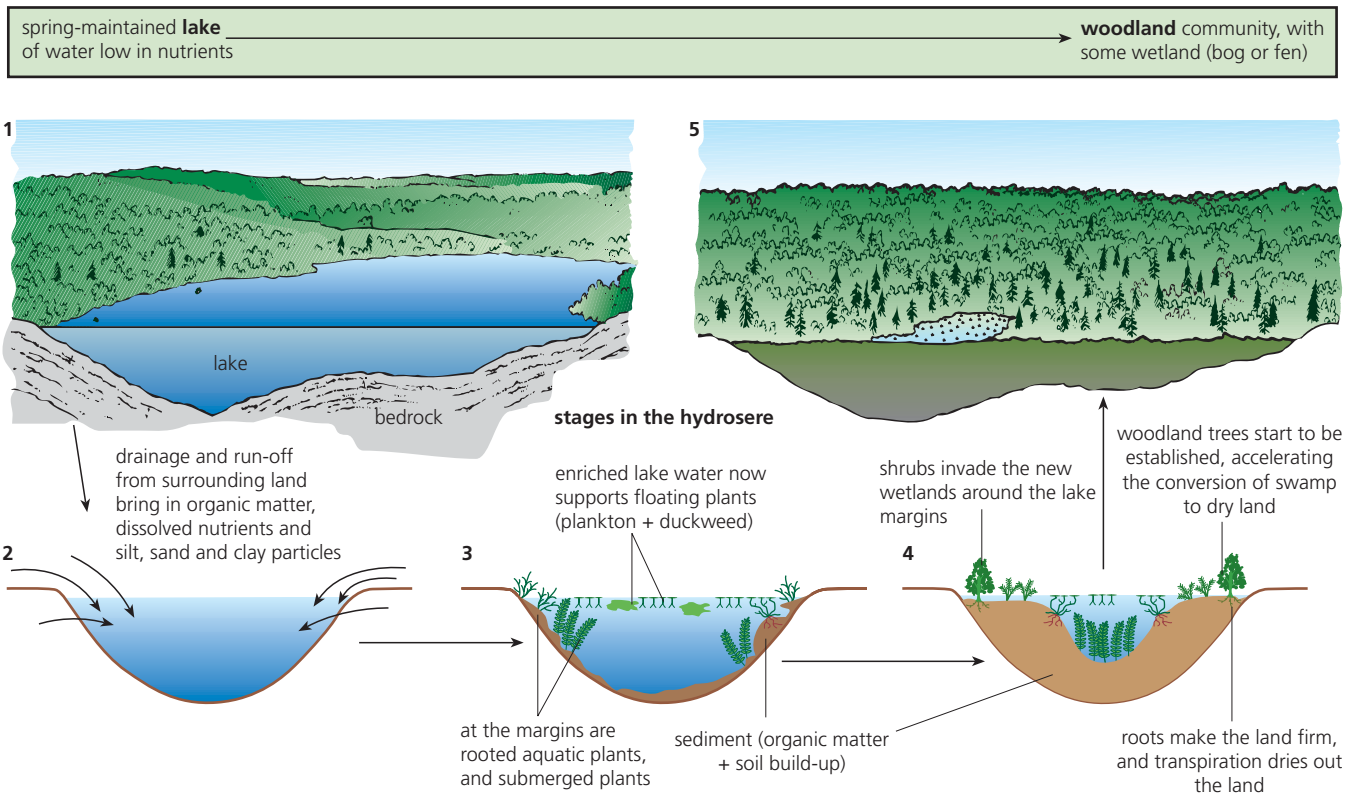


Figure 19.16 An example of a secondary succession – recolonisation of woodland after fire

7 **Describe** the ways secondary succession on burnt forest land differs from primary succession established from a barren, rocky shore.

Secondary successions normally happen quite quickly, since the necessary soil for plant life is already present. Plant communities are established in succession, as spores and seeds are blown in, or carried in by visiting animal life, or as they grow in from the surrounding, unharmed climax communities. After forest fires, for example, the soil is quickly covered by moss species that favour scorched soil habitats. The carpet of moss reduces soil erosion, starts to contribute to the supply of humus, and provides conditions favourable to the lodging and germination of seeds of higher plants (Figure 19.16).

Introducing biomes

Only a part of planet Earth, its land, oceans and atmosphere, is inhabited. In fact, the majority of organisms occur in a narrow belt from upper soil to the lower atmosphere; if marine, many – perhaps most – occur near the ocean surface.

This restricted zone which living things inhabit is called the **biosphere**. Ecology sets out to explain, among other things, the distribution patterns of living things within the biosphere. Important patterns, often visible on a global scale, are the large, stable vegetation zones on the Earth (Figure 19.17). The zones are called biomes. A **biome** is a major life zone characterised by the dominant plant life present. Examples include tropical rainforest and grassland. Here the climax community typically extends unbroken over thousands of square kilometres.

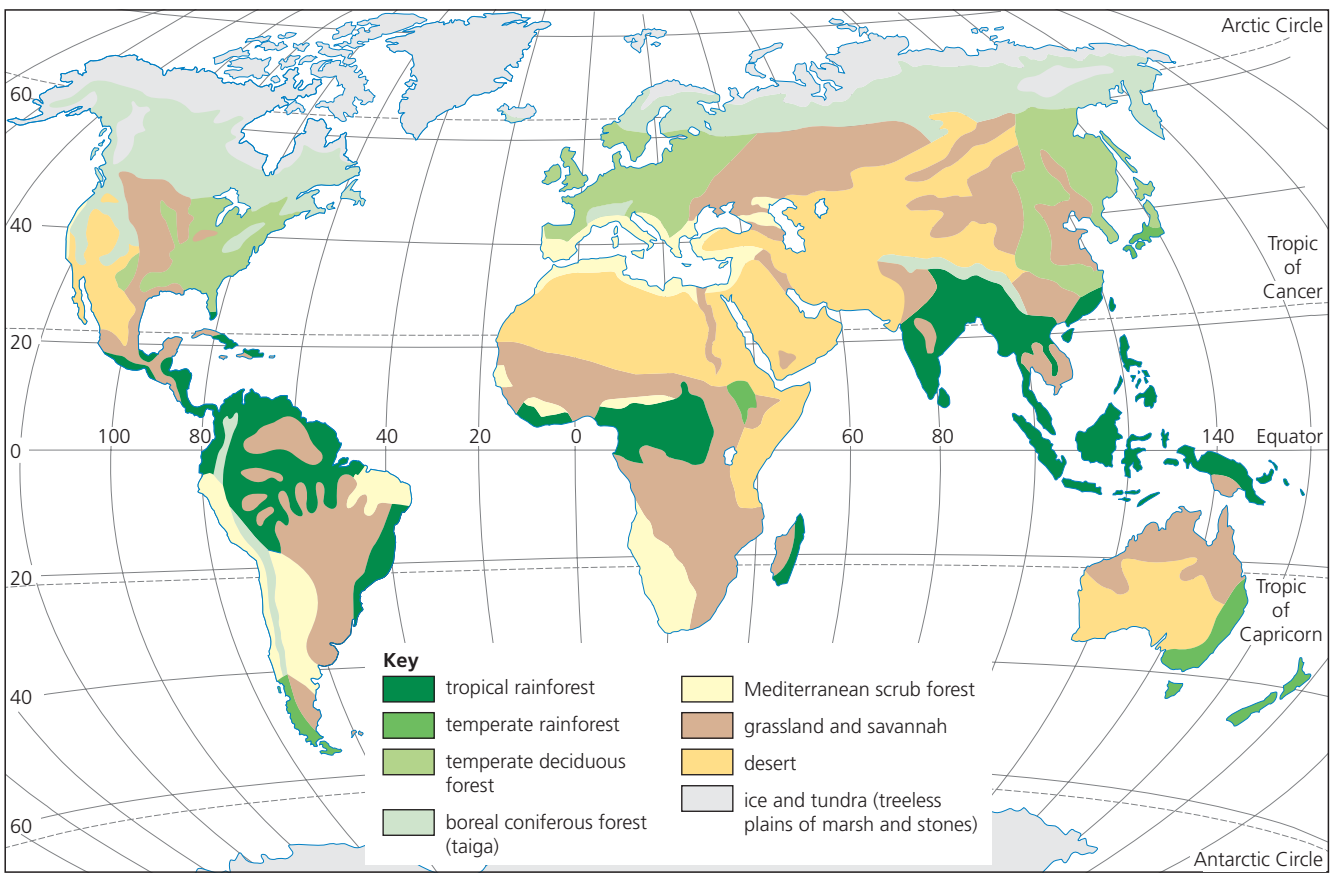


Figure 19.17 The biomes of the world

Why are particular biomes found where they are?

The abiotic factors which appear to determine the occurrence and distribution of the biomes are **rainfall** and **temperature** (Figure 19.18). The interaction of these two factors, which vary with latitude, longitude, position within land masses and proximity to the sea, are critical.

The reason why temperature is influential on distribution of organisms is mainly because of its effects on their metabolism. Many plants (and animals too) have phases in their life cycles that require particular temperature conditions for success. Seed germination is a case in point.

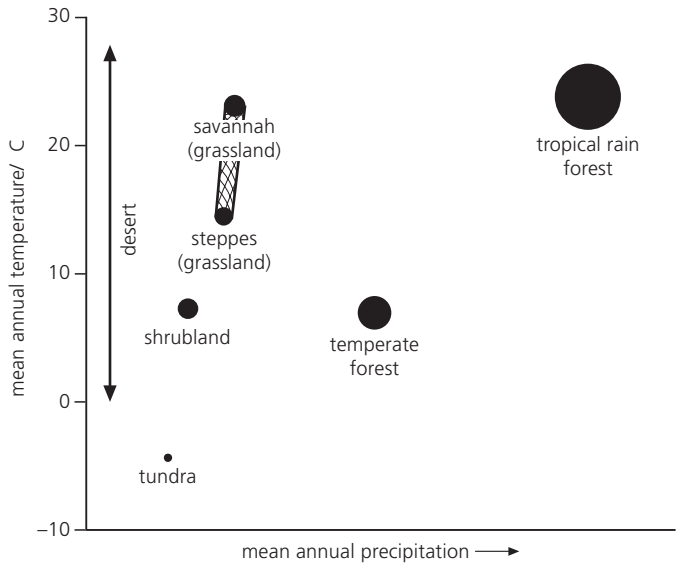


Figure 19.18 The relationship of biome type to climate (mean annual temperatures and precipitation levels)

Similarly, the availability of water in the soil is critical for both the growth of the plant and its nutrition. The availability of water varies greatly, and the adaptations of plants to growth under differing water-level regimes (in relation to various ambient temperatures) influence where different species grow successfully.

The biomes

Examine the map in Figure 19.17 to note the various locations where each of the following biomes is found.

1 **Tundra**, particularly **Arctic tundra**, is a biome found where temperatures are low and the ground is typically continuously frozen, a condition known as permafrost. Plant roots cannot penetrate the soils, which are described as poor. The soils are saturated with water that is unable to evaporate away, and so soil air is limited. The growing season is short – winters are long, dark (and dry). In the brief summer period, conditions improve sufficiently for plant growth and flowering to be spectacular. Growth of the flora is immediately accompanied by the equally brief appearance of vast numbers of insects.

   Alpine tundra occurs on the highest mountains, well above the tree-line. Night-time temperatures here are below freezing, but daylight typically varies little around the year. Day-time temperatures typically rise above freezing, sufficient for the plant life here to show steady, very slow growth, more or less throughout the year.

2 **Grassland** (also called temperate grassland) occurs as **steppe, prairie** and **pampas** around the world where climates are moderately dry, summers are hot and winters cold. The principal plants are perennial grasses, together with some broad-leaved flowering plants that flower either at the start or the end of the growing season when the grasses are less dominant and over-shadowing. Grassland provides natural pasture for grazing animals, and these areas have often been converted into grain-crop grassland, where rainfall is sufficient, as human populations have expanded and their agriculture has become mechanised.

   Savannah is tropical grassland with scattered, individual trees. Typically, here there are three distinct seasons: cool and dry, hot and dry, and warm and wet. This environment typically supports an abundance of wildlife, including large herds of herbivores and their predators, many from the cat family. Frequent fires triggered by electric storms, and the presence of large, destructive herbivores like the elephant, are important factors limiting the growth and spread of trees.

3 **Desert** occurs in the sub-tropical zones where the land receives **very limited** and **unpredictable rain**. The limited plant life is either of ephemerals (dormant until rain, then completing their life cycle in days or weeks) or succulents that survive above ground all the year round. The limited animal life responds to the danger of desiccation in various ways. Camels are large mammals that have adapted to this terrain. Their many adaptations include the ability to allow body temperature to vary by up to 7 °C, allowing conservation of water in daylight hours.

4 **Tropical rainforest** grows in areas of heavy rainfall near the equator. Despite the more or less continuous growth of the dominant plants (trees 30–50 metres high) which the continuous moisture and warmth make possible, the resulting forest is not impenetrable. The height of the trees, their substantial canopies, and the dense growth of epiphytes along the branches, all reduce the light that can reach the forest floor. The flora and fauna is extremely rich – many species present are unknown to science, because it is here that the greatest diversity of any biome is to be found. Much of the animal life is confined to the canopy, specialising on a diet of the abundant fruits present, or else on the leaves. The soil below and the canopy above support a huge fauna of non-vertebrates.

5 **Temperate deciduous forests** are dominated by broad-leaved deciduous trees growing in conditions of adequate moisture. Typically these forests have several layers, with sufficient light penetrating to the lowest layer, the herbaceous plants of the ground level. Between the tree canopy and the ground layer are found a shrub layer, and sometimes an understorey of lower trees. All this vegetation supports a diversity of animal life at all levels. These forests have an annual rhythm to their growth, in which buds break in spring leading to growth of new foliage. The leaves fall at the end of summer, and the trees are, in effect, dormant during the winter. Leaf-loss like this is only possible in temperate zones where the soil holds the essential nutrients that will be required for fresh growth, when it comes in spring.

6 **Scrubland** includes chaparral and heathland. They occur in mild temperate regions with abundant winter rainfall but dry summers. Typical climax vegetation here consists of evergreen shrubs. The origins of this biome are disputed, some ecologists holding that it forms from temperate forest ravaged by fire or the effects of over-grazing, others that it is a diverted climax community.

Biodiversity, and the impact of humans on ecosystems

G3.1–3.11

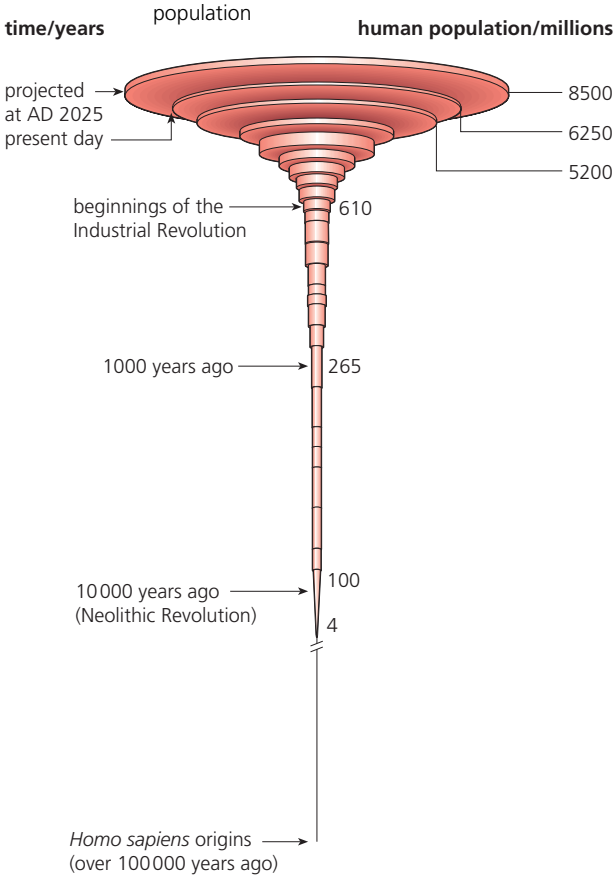


Figure 19.19 The changing pattern of the estimated world human population

There are vast numbers of living things in the world. The word **biodiversity** is a contraction of 'biological diversity', and is the term we use for this abundance of different types or species. Today, the issue of **human influence on biodiversity** is of major concern.

Why?

In the long history of life on Earth, humans are very recent arrivals. Life originated about 3500 million years ago, but our own species arose only a little over 100 000 years ago. Initially, human activities had little impact on the environment. For one thing, during early human prehistory, population numbers were low. *Homo sapiens* was a rather struggling species, living among many very successful ones. At this stage, survival must have been a very chancy affair.

The first significant increase in the human population occurred with the development of settled agriculture – a change began in the fertile crescent in the Middle East about 10 000 years ago. It is known as the **Neolithic (new stone age) Revolution**. After that time, the human population started to increase in numbers, but only at a very slow rate. The current human population explosion began around the beginning of the **Industrial Revolution**, some 200 years ago. The human population explosion continues still, and no-one is certain when the rate of population growth will slow down, as it surely must. When we plot world human population against time we see a steeply rising J-shaped curve (Figure 19.19). This curve can be compared to the first half of the sigmoid curve of growth of microorganisms, shown in Figure 6.17 (page 155).

Human population growth appears to be in an **exponential** (log) **phase of growth**, due to high birth rates and lower death rates, leading to rising life expectancy (people are living very much longer).

Today, the impact of humans on the environment is very great indeed – there is virtually no part of the biosphere which has not come under human influence and been changed to some extent. It continues to be changed now, both the physical and chemical (abiotic) environment, and the living organisms (the biota). Many of these changes threaten biodiversity.

Next we look at how biodiversity can be studied in a local habitat, and then consider some of the ways biodiversity is threatened by us.

Simpson diversity index

Early on in the development of a succession – for example, at the pioneer plants stage on new land (Figure 19.14, page 615) – the number and diversity of species are low. At this stage, the populations of organisms present are usually dominated by abiotic factors. For example, if extreme, unfavourable abiotic conditions occur (prolonged, very low temperatures, perhaps), the number of organisms may be severely reduced.

On the other hand, in a stable climax community many different species are present, many in quite large numbers. In this situation, adverse abiotic conditions are less likely to have a dramatic effect on the numbers of organisms present. In fact, in a well-established community the dominant plants set the way of life for many other inhabitants. They provide the nutrients, determine the habitats that exist, and influence the environmental conditions. These plants are likely to modify and reduce the effects of extreme abiotic conditions, too.

Thus the **diversity of species** present in a habitat is also an indicator of the **stability** of the community. Species diversity of a community may be measured by applying the formula known as the **Simpson diversity index**:

$$\text{diversity} = \frac{N(N-1)}{\sum n(n-1)}$$

where N = the total number of organisms of all species found
n = the number of individuals of each species

8 In a vegetable plot left fallow (left unsown for a year) three weed species appeared. Individual plants were counted:

Species	*n* – number of individuals
Groundsel (*Senecio vulgaris*)	45
Shepherd's purse (*Capsella bursa-pastoris*)	40
Dandelion (*Taraxacum officinale*)	10
Total (*N*)	95

Calculate the Simpson diversity index for this habitat.

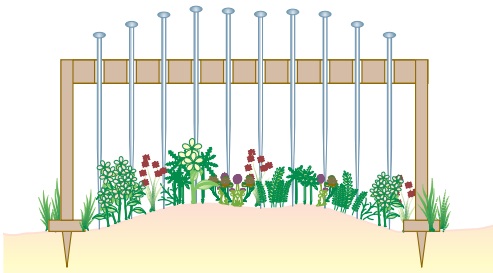
9 Figure 19.20 is a study of the vegetation of a system of developing sand dunes. Note how both the species diversity and physical parameters change as the dunes age. **Calculate** the diversity indices of the fore dune and semi-fixed dune (using a spreadsheet). **Comment** on the results.

The case for conservation of biodiversity – rainforests as a special case

Rainforests cover almost 2% of the Earth's land surface, but they provide the habitats for almost 50% of all living species. It has been predicted that if all non-vertebrates occurring in a single cubic metre of tropical rainforest soil were collected for identification, there would be present at least one completely previously unknown species. It is the case that tropical rainforests contain the greatest diversity of life of any of the world's biomes.

Sand deposited by the sea and blown by the wind, builds into small heaps around pioneer xerophytic plants, such as marram and couch grass, at coasts where the prevailing wind is on-shore. The tufts of leaves growing through the sand accelerate deposition, and gradually drifting sands gather a dense cover of vegetation. Fixed sand dunes are formed.

point frame quadrat in use



The frame is randomly placed a large number of times, the 10 pins lowered in turn onto vegetation and the species (or bare ground) recorded.

Behind the fore dune the fixed dunes can be seen.

Figure 19.20 Sand dune community development – a student field study exercise

Study of a sand dune succession from embryo state to fixed dune

	Site 1	**Site 2**	**Site 3**	**Site 4**
Stage in succession	Embryo dune	Fore dune	Semi-fixed dune	Fixed dune
Number of pins dropped	220	220	220	220
sea couch grass	169	9	0	0
marram grass	1	123	19	0
fescue grass	0	0	126	182
spear-leaved orache	4	0	0	0
prickly saltwort	2	0	0	0
bindweed	1	44	0	0
bird's foot trefoil	0	0	0	6
biting stonecrop	0	0	0	1
buttercup	0	0	0	1
cat's ear	0	0	5	2
clover	0	0	0	68
common stork's bill	0	0	0	11
daisy	0	0	2	8
dandelion (common)	0	0	5	1
eyebright	0	0	0	4
hawkbit	0	0	20	1
ladies bedstraw	0	0	86	35
medick	0	0	0	62
mouse-ear chickweed	0	0	0	2
ragwort	0	0	8	1
restharrow	0	0	25	15
ribwort plantain	0	0	0	37
sand sedge	0	0	32	17
stagshorn plantain	0	0	0	36
tufted moss	0	0	0	119
yellow clover	0	0	0	5
Yorkshire fog	0	0	2	0
wild thyme	0	0	0	82
Total live hits	**177**	**176**	**330**	**706**
Bare ground hits	**48**	**68**	**6**	**0**
% bare ground	27	38	3	0
soil moisture (%)	5.0	8.5	16.0	14.3
soil density (g cm^{-3})	1.6	0.9	0.6	0.5
soil pH	8.0	7.5	7.0	7.0
wind speed (m s^{-1})	10.0	9.3	2.4	1.3

Sadly, tropical rainforests are being rapidly destroyed. Satellite imaging of the Earth's surface provides the evidence that this is so – even where no other reliable sources of information are available. The world's three remaining tropical rainforests of real size are in **South America** (around the Amazon Basin), in **West Africa** (around the Congo Basin) and in the **Far East** (particularly but not exclusively on the islands of Indonesia).

The current rate of destruction is estimated to be about one hectare (100 m × 100 m – a little larger than a football pitch) every second. This means that each year an area larger than the British Isles (31 million hectares) is cleared. While extinction is a natural process, the current rate is on a scale equivalent to that at the time of the extinction of the dinosaurs (an event 65 million years ago, at the boundary of the Cretaceous and Tertiary periods – Figure 16.17, page 474).

Extinctions have been a feature of the history of life. Why should we strive to conserve biodiversity?

Conservation involves applying the principles of ecology to manage the environment. The reasons for conservation of rainforest are outlined below.

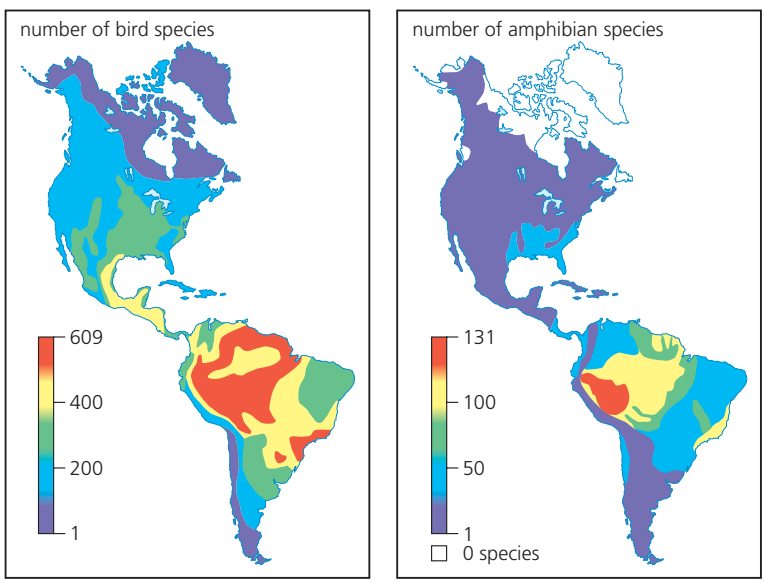


Figure 19.21 Most species have small ranges and these are concentrated unevenly

1 Ecological

a Most species of living things are not distributed widely, but instead, are restricted to a narrow range of the Earth's surface. In fact, the majority of species living today occur in the tropics (Figure 19.21). So when tropical rainforests are destroyed, the only habitat of a huge range of plants is lost, and with them very many of the vertebrates and non-vertebrates dependent on them.

b In effect, the rainforests are critically important 'outdoor laboratories' where we learn about the range of life that has evolved, the majority of which consists of organisms as-yet unknown.

c The soils under rainforest are mostly poor soils that cannot long support an alternative biota. To destroy rainforest is to remove the most productive biome on the Earth's surface – in terms of converting the Sun's energy to biomass.

d If destroyed but later left to regenerate, only species that have not become extinct may return. Re-grown rainforest will be deprived of its variety of life.

2 Economic

a The whole range of living things is functionally a gene pool resource, and when a species becomes extinct its genes are permanently lost. The destruction of rainforests decreases our genetic heritage more dramatically that the destruction of any other biome. As a consequence, future genetic engineers and plant and animal breeders are deprived of a potential source of genes. Many new drugs and other natural products, in some form or another, are manufactured by plants. The discovery of new, useful substances often starts with rare, exotic or recently discovered species.

b About half of the original rainforest present before clearance programmes started has been cleared – 7 million km^2 of humid tropical forest. Much is cleared to make way for the production of crops for food (Figure 19.22). Yet only 2 million km^2 have remained in agricultural production. Mostly, the soil does not sustain continued cropping.

c Rainforest is a continuing resource of hardwood timber, which if selectively logged can be productive, but when cleared as forest, is lost as a source of timber for the future.

d Trees help stabilise land and prevent disastrous flooding downriver. Huge areas of productive land are washed away once mountain rainforest has been removed.

e Trees in general are carbon dioxide sinks that help reduce global warming. Without these trees, other ways of reducing atmospheric carbon dioxide must be found.

Figure 19.22 Rainforest land for agriculture

Adapted from: *THE TIMES* Thursday February 2 2006

Rainforest Eden faces death by chainsaw – to make margarine

A £4.4bn plantation could destroy an ecological treasure, reports **Nick Meo** from Borneo.

IN THE cool of dawn Betung Kerihun could almost be an English wood until the honks of a rhinoceros hornbill echo around the great creeper-festooned trees.

The forest is an almost untouched Eden. Orang-utans and gibbons live high in the canopy. On the forest floor clouded leopards and eight-metre pythons hunt wild boar and deer and are themselves hunted by one of the last truly nomadic forest peoples, the Penan. But this rainforest, which has survived for millions of years, may now be doomed.

A £4.2 billion plan proposed by a Chinese bank and backed by Jakarta politicians would clear 1.8 million hectares of this wilderness over the next six years to grow oil palms to feed the world's growing appetite for margarine, ice-cream, biscuits and biodiesel fuel.

Until now Betung Kerihun, technically a protected national park, has been saved by its remoteness despite the network of roads that illegal loggers have begun to push inside its boundaries.

What survives is a biological treasure that staggers scientists newly arrived from Europe. It is home to thousands of tree frogs, bats and orchids. More than 1,000 insects have been identified in a single tree. In one ten-year period 361 new plant, animal and insect species were discovered in Borneo.

Charles Darwin, who explored the giant island before writing *The Origin of Species*, called it "one great untidy luxuriant hothouse made by nature for herself".

Reaching the virgin jungle from the coast takes days by river or logging road across a landscape of tree stumps, and the most common sound is that of the chainsaw. Only half of the island is now covered with forest, compared with three quarters in the 1980s.

Conservationists believe that the real motive for the oil palm scheme is to gain access to the valuable timber along a great swath of the Indonesian-Malaysian border.

"It's a scam," said Stuart Chapman, of the WWF conservation group. "Palm oil is a lowland equatorial crop and not suited to steep upland soils. There are two million hectares of idle land, already cleared of forest, in lowland Borneo which is suitable for planting. but putting this plantation in the island's centre would give logging interests the excuse they need to cut down trees."

10 There are other reasons for conservation of diversity, besides those listed above. **Suggest** at least one more reason under one or more of the categories listed here.

3 Aesthetic

a These habitats are beautiful and exhilarating venues to visit and be inspired by.

b They are part of the inheritance of future generations which should be secured for future people's enjoyment, too.

4 Ethical

a This biome is often home to forest people who have a right to their traditional ways of life.

b Similarly, many higher mammals, including relatively close relatives of *Homo*, live exclusively in these habitats. The needs of all primates must be respected.

The impacts of alien species and of invasive species on ecosystems

Alien species are introduced plants and animals that have been accidentally or deliberately transferred from habitats where they live, to new environments where the abiotic conditions are also suitable for them. If the alien takes over in an apparently aggressive way, detrimental to the food chains of their new habitat, then they are described an **invasive species**.

Examples of alien species that have become invasive are:

- the **rabbit** (*Oryctolagus* sp.) deliberately introduced from Europe, into Australia, and the **myxoma virus** disease of rabbits, from South America, introduced as a biological control;
- **Japanese knotweed** (*Fallopia japonica*) deliberately introduced into northern Europe in the early nineteenth century as an ornamental plant for garden ponds and lakes;
- **American grey squirrel** (*Sciurus carolinensis*) accidentally introduced into Britain in the nineteenth century.

The profile of the impact of an alien species can be illustrated by the introduction of the European wild rabbit to Australia in 1859. This voracious herbivore spread rapidly. Clearly, no

natural predators in Australia were threats to numbers (**interspecific competition** failed) because rabbits quickly over-ran large parts of that continent. Grassland, available to herds of herbivores, principally cows and sheep, was seriously damaged. In desperation, and as a **biological control measure**, myxoma virus that had been discovered in South American rabbits (a different species from the European rabbit), was introduced in 1950.

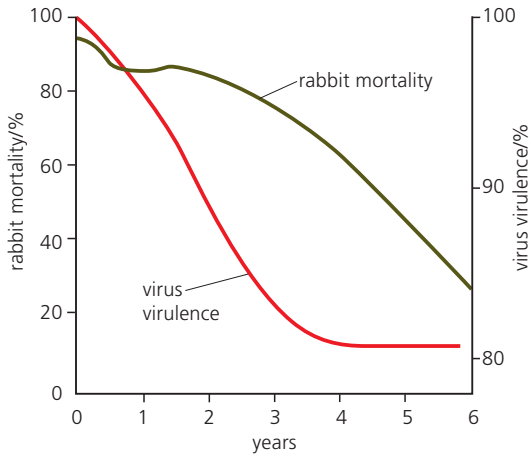


Figure 19.23 Changing rabbit mortality and virus virulence, following introduction of the myxoma virus as a biological control measure

Myxoma virus causes a parasitic disease of rabbits, **myxomatosis**, but on its original continent, the effects were mild. The changing effects of this virus on rabbit mortality (1950–56) are shown in Figure 19.23, as is the resulting change in virulence of the virus. We may assume that initially, a small number of the huge population of rabbits had immunity. As their vulnerable relatives were killed off the immune rabbits prospered from the diminished competition for grass. Rapidly, the bulk of the population were immune, and any that failed to develop immunity were quickly taken out.

11 Describe how the rabbits' immune system confers immunity to the effects of a virus.

In its original habitat, an alien species will have evolved in the company of **natural parasites** and **predators** – and come to exist in balance with other organisms of the community. In a new habitat, the aliens' natural enemies may be absent, so they may grow at the expense of native species, crowding them out.

This is the case with **knotweed**, which in Britain is not browsed by natural herbivores as it escapes into natural waterways. Similarly, plant parasites here do not attack the plant.

Knotweed plants grow into dense, submerged thickets that over-shadow and crowd out native water plants, block waterways, and block public access to stream banks. Yet back in its native Japan, the plant grows in a balanced relationship with other water plants and causes no significant problems. The reason for the difference is that knotweed was introduced in Britain without its natural populations of predators. The chemicals in its leaves and roots discourage predation by leaf browsers or root parasites. In Japan, there are natural populations that browse on knotweed and keep its growth in check. The **invasive pattern** of this alien's growth is shown in Figure 19.24.

The invasive pattern of the spread of the grey squirrel in Britain is proving a serious threat to the indigenous **red squirrel**, which is **now threatened with extinction** if its current rate of decline is maintained (Figure 19.25). Since it arrived, the grey squirrel has spread rapidly and it appears to be driving out the red squirrel, chiefly by being able to consume a wide range of locally growing nuts (including the very common woodland oak tree's acorn – this contains polyphenolic substances that the new arrival can digest, but which the red squirrel cannot). With its wider diet, the grey has out-bred the red, and multiplied much faster. This enlarged population has successfully competed for hazel nuts and pine cones too (the limited diet of the red squirrel).

Pesticide pollution and food chains

Wild plants and animals have evolved alongside their predators and parasites and most live in balance with them. Meanwhile our cultivated crops and herds have been selected to be especially productive and high-yielding, but typically have limited resistance to local parasites and predators, especially given the way they are grown – in intensive monocultures.

Pesticides are substances used to control harmful organisms that are a danger to crops or herds. Pesticides have enormously improved productivity in agriculture, but their use has generated problems in the environment.

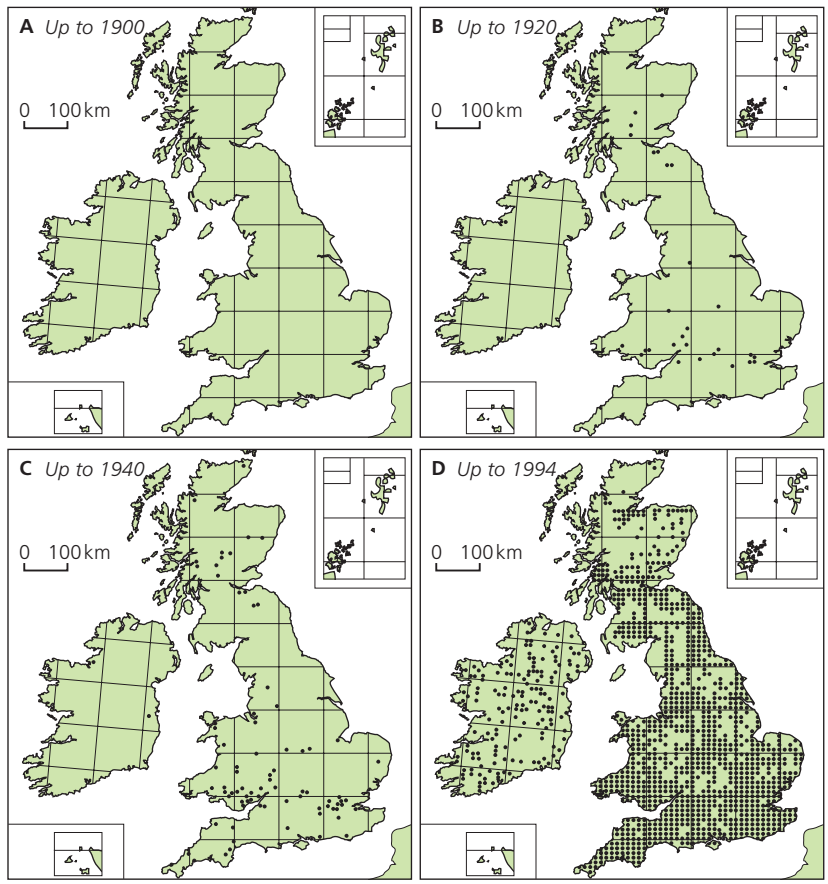


Figure 19.24 Distribution of knotweed in the UK, 1900–94

Figure 19.25 Red and grey squirrels

The revolution in the chemical control of insect pests came when a particular substance known as dichlorodiphenyltrichloroethane (**DDT**) was found to be a very effective insecticide.

The DDT molecule has chlorine atoms attached to hydrocarbon rings, and was the first of a family of **organochlorine insecticides**. DDT is a nerve poison to insects, causing rapid death even when applied in low concentrations. DDT was economical to use because the molecule was **stable** when dispersed in the biosphere, remained **lethal for a very long time**, and was effective against a very wide range of insect pests (broad-spectrum insecticide). It did not do any harm to vertebrates in concentrations needed to kill insects, and so it was liberally used.

During the 1939–45 war and subsequently, DDT was especially effective against the mosquito, the vector for malaria (page 595). Of course, very many insect predators of pest species were also killed by DDT.

Figure 19.26 Biomagnification of DDT

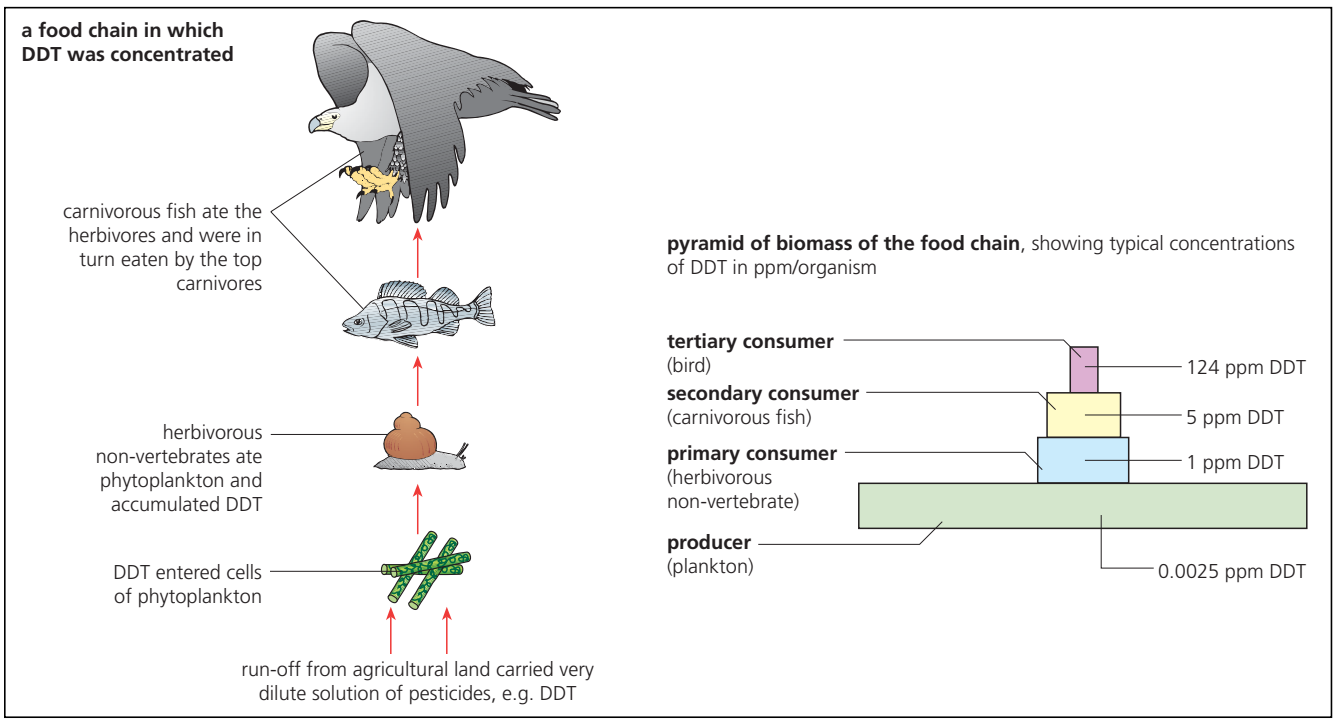
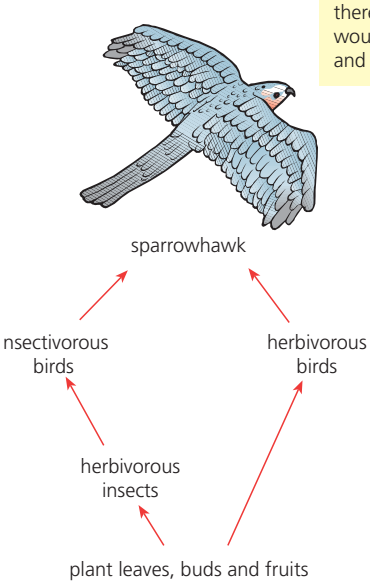
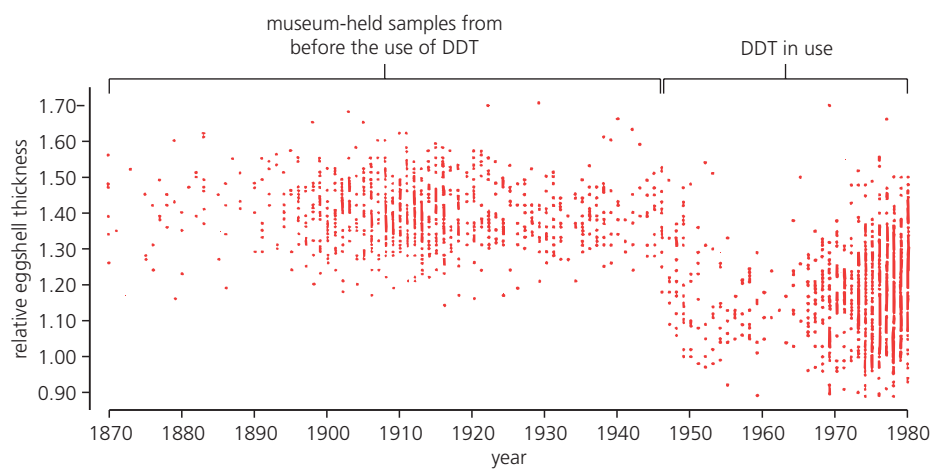


Figure 19.27 Eggshell thickness studies in sparrowhawks

Food web of the sparrowhawk
When DDT was used as an insecticide, there was a likelihood this pesticide would bioaccumulate at each step, and biomagnify in the top carnivore.



In the eggshell thickness study of British sparrowhawks 1870–1970, more than 2000 clutches of eggs were measured; each dot represents the mean shell thickness of a clutch (typically five eggs).



DDT is fat soluble and is **selectively retained in fatty tissues** of animals, rather than circulating in their blood to be excreted by the kidneys. As a result, at each stage of the food chain, DDT becomes concentrated, a process known as biomagnification.

Biomagnification is a process in which chemical substances become more concentrated at each trophic level.

In fact, in non-vertebrates, fish and birds, for example, DDT concentrations sometimes reached toxic levels (Figure 19.26). It began to be concentrated in top carnivores with devastating consequences (Figure 19.27). Although DDT is not a nerve poison in birds and mammals, in breeding birds it does inhibit the deposition of calcium in the egg shell. Affected birds lay thin-shelled eggs that easily crack. There was a rapid decline in numbers of birds of prey, in areas where DDT had become widely used in agriculture.

With this harmful effect of DDT discovered, the quality of 'stability' was renamed 'persistence'. Once the wider effects of organochlorine pesticides on wildlife were recognised, a ban was imposed, at least in the developed countries.

Instead, insecticides that are **biodegradable** in the biosphere (not persistent) and which are more specific in their actions were sought. Initially, organophosphate insecticides were favoured because they break down in the soil. They work by blocking the synapses of the insect nervous system, but are now suspected of being harmful to humans. If tiny quantities of pesticide persist in the food chain, these residues may accumulate in humans.

Subsequently, other substances, synthetic derivatives of pyrethrum, have been developed. These are much less toxic to mammals and also biodegradable. However, they are lethal to fish and must not be used near streams, rivers or lakes.

12 **Analyse** the steps by which pesticides, applied at apparently safe concentrations, may cause death of organisms seemingly unrelated to the crops or herds for which the pesticides are designed.

Atmospheric pollution and the ozone of the stratosphere

Ozone is a form of oxygen that contains three atoms of oxygen combined together. Today it occurs naturally in the Earth's atmosphere as an **ozone layer** found in the region of the atmosphere called the **stratosphere**. This is at a height of 15–40 km above the Earth's surface, so it is described as **high-level ozone**.

Ozone (O_3) is formed in the upper atmosphere by the action of ultra-violet (UV) radiation on molecular oxygen (O_2), which is split into two highly reactive atoms of oxygen. Each highly reactive oxygen atom reacts with a molecule of oxygen to form ozone:

$$O_2 \xrightarrow{+\ UV\ light} O\cdot + O\cdot \qquad O\cdot + O_2 \xrightarrow{+\ UV\ light} O_3$$

At the top of the stratosphere there is much incoming UV radiation but little oxygen. Closer to the Earth's surface there is much oxygen but little UV radiation. Consequently, the highest concentration of ozone is at a midpoint of the stratosphere.

Ozone of the stratosphere is maintained by the action of UV light, and this ceaseless cycle of changes leaves the composition of the atmosphere unchanged, *and most of the incoming UV light absorbed*. The stratosphere is slightly warmed by the reactions, but the heat is lost to space:

$$O_3 \underset{+\ UV\ light}{\overset{+\ UV\ light}{\rightleftharpoons}} O\cdot + O_2$$

UV radiation that reaches the Earth's surface is harmful to living things because it is absorbed by the organic bases (adenine, guanine, thymine, cytosine and uracil) of nucleic acids (DNA and RNA) and causes them to be modified (mutation, page 100).

Consequently, the maintenance of the high-level ozone layer is important to the survival of life.

Atmospheric pollution destroys high-level ozone

Gaseous pollutants are threatening the ozone layer. These pollutants are chemicals made by industry – substances such as chlorofluorocarbons (**CFCs**). CFCs are very unreactive, stable molecules, deliberately manufactured to use as propellants in aerosol cans and as the coolant in refrigerators. However, gases escape into the atmosphere from these sources from time to time, and are slowly carried up to the stratosphere. It takes as long as five years for this to happen.

In the stratosphere, CFCs are exposed to high levels of UV light, and are broken down. Highly reactive chlorine atoms are then released, and these break down ozone molecules in a cyclic reaction:

$$O_3 + Cl\cdot \longrightarrow ClO + O_2$$

$$ClO + O\cdot \longrightarrow Cl\cdot + O_2$$

The outcome is that in the presence of CFCs, ozone molecules are broken down **faster than they can be reformed** by the natural reaction between molecular oxygen and UV light that we noted above.

Large quantities of CFCs were released into the atmosphere before this danger was realised. Because of the time taken for CFCs to reach the upper atmosphere, ozone depletion continues, despite the current steps to replace CFCs by safer chemicals.

The thinning of the ozone layer is greater over some countries than others, but for all organisms exposed to sunlight on land, the thinning ozone layer (called an **ozone hole**) is a potential problem. Protective clothing and UV blocking creams are necessary to avoid the danger of skin cancer, for people exposed to sunlight. This is potentially a major danger for people in Chile, South Africa, New Zealand and Australia, for example.

Conservation of biodiversity

G4.1–4.6

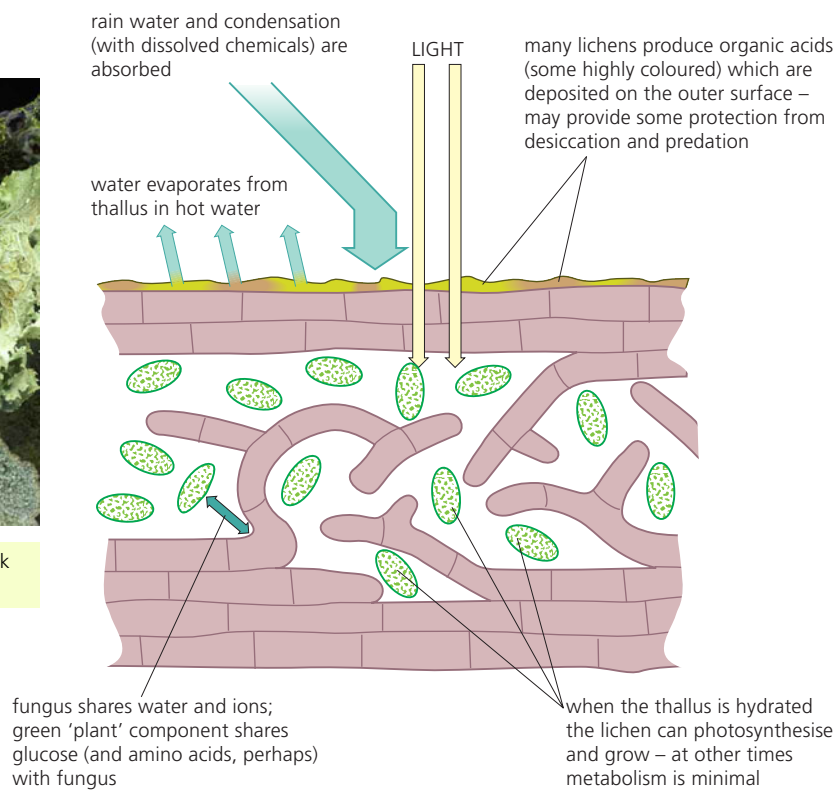
The destruction of tropical rainforest has already been noted as an existing major threat to biodiversity (page 619). In fact, human activities are changing many biomes worldwide, and in far-reaching ways, although it is not always as self-evident as the process and outcomes of deforestation are. The first step to effective conservation of natural and semi-natural environments is the early detection of environmental change.

Figure 19.28 The physiology of lichen mutualism

lichen (*Evernia prunastri*) common on trees, fences, rocks and walls

The presence and metabolism of lichen may speed rock erosion.

lichen thallus in section



Fungal hyphae retain water and ions as and when they are available.

When the green 'plant' component is a cyanobacterium, atmospheric nitrogen may be fixed to form amino acids.

How can environmental change be detected and monitored?

Some organisms are particularly sensitive or vulnerable to change in their environment. If the numbers and condition of such species in threatened habitats can be monitored, their predicament (their health and well-being) can function as biological indicators (**biotic indices**) of impending environmental damage. For example, lichens (and mosses) are ideal organisms for the early detection of atmospheric pollution – in effect, they are potential **indicator species** of environmental change.

Lichens as pollution monitors

Lichens are dual organisms; their body (called a thallus) is made of a fungal component and an algal (or photosynthetic bacterium, cyanobacterium) component living together for mutual benefit (Figure 19.28). Some lichens are leaf-like, some crust-like, and others are described as shrubby or beard-like. There are several thousand different species of lichen, often occurring in quite hostile habitats. Compact fungal hyphae make up the bulk of the lichen, and this component absorbs and retains water and ions. The algal component carries out photosynthesis.

Lichens (and mosses) are especially susceptible to air-borne pollutants dissolved in rain water because their surfaces are not protected by a waxy cuticle. The lichen thallus absorbs and accumulates various pollutants, and samples of lichens may be analysed for contamination by heavy metal ions. The effects of pollution on the extent of their growth may also be a valuable indicator (Figure 19.29).

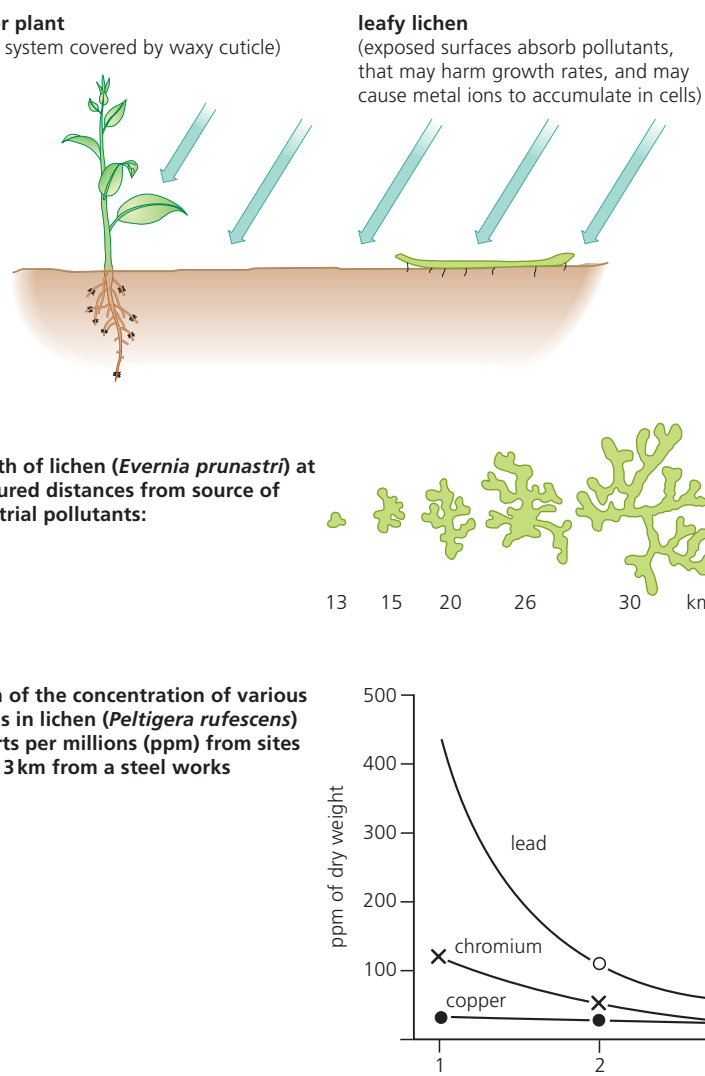


Figure 19.29 Lichens as pollution indicators

Monitoring fresh water pollution

In water enriched with inorganic ions, aquatic plant growth becomes abnormally luxuriant. Ion enrichment may be due to accidental pollution by raw sewage, or from stockyard effluent (manure or silage liquors). Alternatively, an excessive or incorrect use of fertilisers on farm crops results in excess soluble fertiliser being leached from the soil by rain. Of the many ions beneficial to plants, an increased concentration of ammonium, nitrate and phosphate ions particularly increases plant growth.

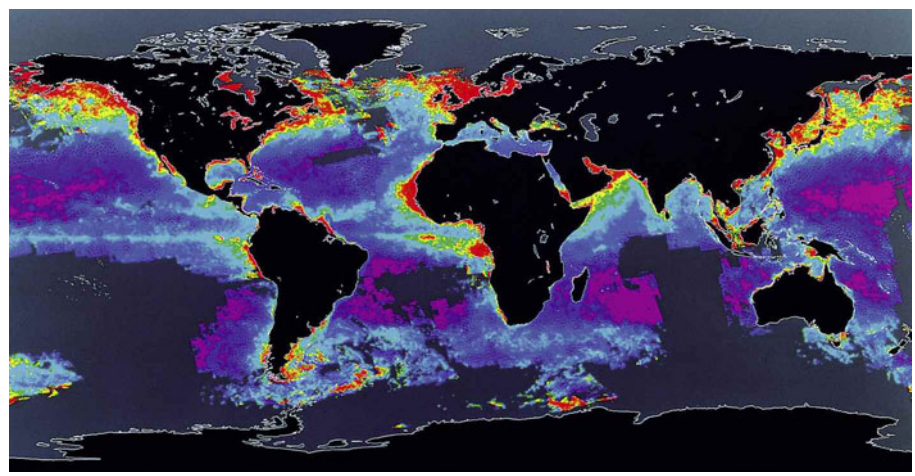
In summer, in waters that are ion-enriched, with raised water temperatures, there occurs a phenomenal plant population explosion, often referred to as an **algal bloom** (because algae are involved, too, and as many are unicellular, their growth rates are spectacular). An excess of aquatic plant life in polluted waters is an example of **eutrophication**.

Figure 19.30 Detection of aquatic pollution following eutrophication

In marine habitats

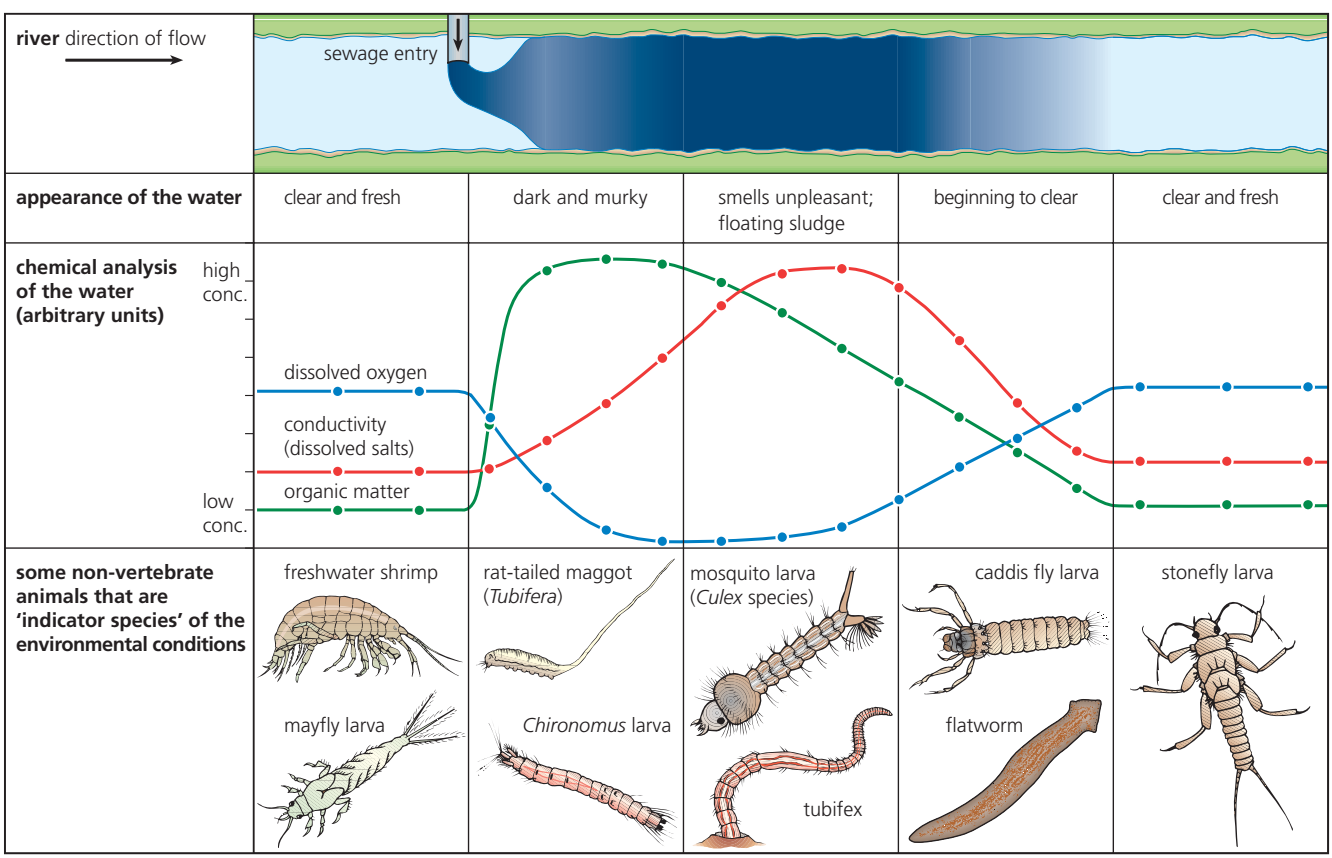
Extensive seasonal algal blooms of the oceans can be observed in satellite images.

This 'false colour' photograph of the Earth's surface, taken from a satellite, shows where **algal blooms** (orange) are most common, and where algal growth is least (purple areas).



In fresh-water habitats

The effect of pollution of a river with untreated sewage can be observed by eye and measured by chemical analysis.



Later, after the algal bloom has died back, the organic remains of the plants are rapidly decayed by saprotrophic, aerobic bacteria. As a result, the water becomes deoxygenated, causing anaerobic decay and then, hydrogen sulphide formation. The few organisms that can survive in these conditions prosper, but the death of many aquatic organisms occurs, including fish, due to the total absence of dissolved oxygen and the presence of hydrogen sulphide. The sequence of events when a river is polluted in this way is summarised in Figure 19.30. Note that the changing populations of non-vertebrate aquatic animals are indicators of the degree of pollution and recovery.

Much of the waters from the land eventually drain into the seas, where the phenomenon of algal blooms is repeated. **Satellite photography** is a way of monitoring pollution in marine habitats as well as rainforest destruction.

Red Data Books as listings of indicator species

Currently, the rate of extinctions is exceptionally high. Environmentalists seek the survival of endangered species by initiating and maintaining local, national and international action. The International Union for the Conservation of Nature and Nature Reserves co-ordinates the updating of **Red Data Books**. These list endangered species, identifying those for which special conservation efforts are needed. The health and general well-being of populations of these organisms are indicators of environmental changes.

Practical conservation steps include the designation and maintenance of nature reserves, botanical and zoological gardens (with their captive breeding programmes), and the establishment of viable seed banks; all are measures to combat loss of endangered species and of the habitats that support them. We look into this issue next.

Extinctions and conservation of species

New species evolve, but other species, less suited to their environment become extinct, as much of the fossil record throughout geological time records. One example of a well documented extinction is the dodo.

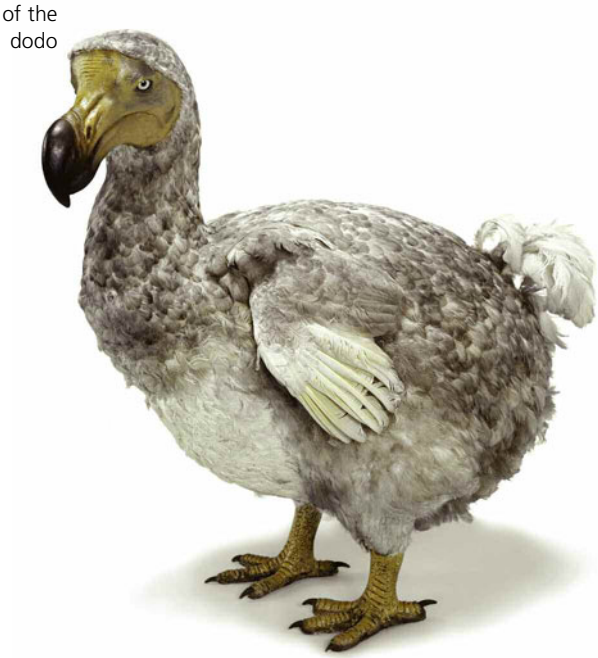


Figure 19.31 A reconstruction of the dodo

The dodo

The dodo (*Raphus cucullatus*) was an inhabitant of the island of Mauritius in the Indian Ocean (Figure 19.31). It was a bird related to modern pigeons. Over geological time, it had evolved to master a terrestrial habit. In the process it became large bird (about a metre long, with a mass of approximately 20 kg).

The dodo nested on the ground, and reared its young there. The diet was one of seeds and fruits that had fallen from the forest trees. It was one of many forest-dwelling birds on the island – one of the 45 species for which there are early records, of which only 21 species have survived to this day.

What factors contributed to the extinction of the dodos?

Mauritius is a medium-sized island, extremely far from any mainland. The dodo's misfortune was that Mauritius got involved in developments in what we know as the **spice trade**.

Spices have been brought from the East, where they are grown and produced, to European centres since ancient times. The rarity of spices in the West and their importance at times when food was difficult to preserve fresh gave them great value. In order to break the monopoly in spice trading held by the people of Venice from 1200 to 1500, other European explorers (first from Portugal, and then from other west European sea-going countries) sailed the sea-routes from early in the sixteenth century. They made stop-over visits to ports and islands to restock (fresh water and fresh meat were required). Visits led to population migrations, and people brought their farm stock with them. Mauritius was a popular location for this, and later was also used secretly to grow imported spice plants. Seeds and plants were smuggled in from previous supply centres, in order to further break monopolies. The aim was to reduce prices set by traditional suppliers.

Records show that the dodo became extinct on Mauritius by 1681. The specific factors that contributed to this were:

1 Visiting **sailors killed the dodo** as a source of fresh meat to restock their ships. The dodo proved an easy victim, unfamiliar with humans (or other mammals). In fact, it had previously lacked significant natural enemies.
2 Later, **settlers brought cats, dogs and pigs**, and inadvertently, rats – all alien species which fed on the occupant of the dodo's nest: the eggs and growing chicks. The dodo population numbers were no longer maintained.
3 **Natural habitats on the island were deliberately destroyed** as land was cleared for agriculture and the growth of introduced spice plant species. The forests with their fruit and seed-yielding trees were removed, so the remaining dodos lost their source of food. Indeed, many other species are likely to have become extinct with the forest clearance.

Conservation by promotion of nature reserves

Biological conservation can be attempted by setting aside land for restricted access and controlled use, to allow the maintenance of biodiversity, locally. This is not a new idea; the New Forest in southern Britain was set aside for hunting by royalty over 900 years ago. An incidental effect was a sanctuary for many forms of wildlife.

Today, this solution to extinction pressures on wildlife includes the setting up of nature reserves (such as areas of chalk downland in northern European countries) and National Parks (such as Yellowstone, one of the original National Parks, set up in North America in 1872; and African game parks, more recently established). In total, these sites represent habitats of many different descriptions, in many countries around the world. Some of the conservation work they achieve may be carried out by volunteers.

What biogeographical features of a reserve best promote conservation?

In a nature reserve, the area enclosed is important (Figure 19.32). In fact, there is an area of reserve too small to be effective, but as size increases there comes a point where further increase secures no greater diversity. The actual dimensions of an effective reserve vary with species, size, and life style of the majority of the threatened species it is designed to protect.

Also, for a given reserve there is an **edge effect** (Figure 19.32). A compact reserve with minimal perimeter is more effective than one with an extensive interface between its perimeter and the surroundings. But whatever the size, the uses of the surrounding area are important, too. If it is managed sympathetically it may indirectly support the reserve's wildlife.

Another feature is **geographical isolation** (Figure 19.32). Reserves positioned at great distances from other protected areas are less effective than reserves in closer proximity. Where reserves are small, and a larger provision would help, it has been found that connecting corridors of land may overcome the disadvantage. In agricultural areas, these may simply take the form of hedgerows protected from contact with pesticide treatments that nearby crops receive.

Active management of nature reserves

Conservation involves using knowledge of the ecology of habitats, including of both the biota and the abiotic factors operating, in order to manage the environment of the nature reserve, and so maintain biodiversity. Conservation is an active process, not simply a case of preservation.

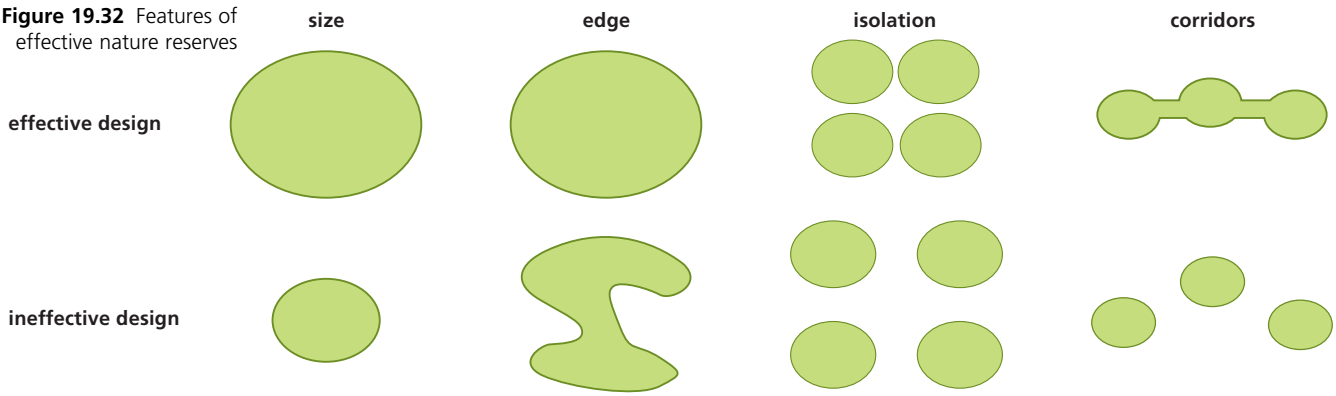


Figure 19.32 Features of effective nature reserves

There are five key steps to active management of a nature reserve:

1 **Continuous monitoring** of the reserve so that causes of change are understood, change may be anticipated, and measures taken early enough to adjust conditions without disruption, should this be necessary. Changes particularly arise from natural succession tendencies in habitats that are not climax communities. Alternatively, changes in human industries and land use in surrounding areas, or devastating natural disasters due to unseasonal weather conditions might all cause havoc. Information is essential to keep the reserve from changing.
2 **Habitat conservation** to maintain stable habitats essential for threatened species. For example, diverted climax communities such as grassland may be quickly displaced by the growth of scrub followed by wood, if the natural invading plants are not checked. A diverted climax has often arisen by particular human farming practices (such as land grazing by ruminant herds), and this may need to be re-introduced and periodically maintained.
3 **Maintenance of effective boundaries**, and the limiting of unhelpful human interference. The enthusiastic involvement of the local human community should be facilitated, at the same time. This often involves the provision of vantage points for over-viewing the reserve, hides for observation of shy animals, and seasonal information on the life of the community – all encourage understanding of the ecology that underpins the purpose of the reserve. Opportunities for local volunteers communicate the message that everyone has a part to play in conservation.
4 **Measures to facilitate the successful completion of life cycles** of any endangered species for which the reserve is home, together with supportive conditions for vulnerable and rare species, too. These may involve the supply and maintenance of nesting boxes, access to ponds or flowing water, and the planting of essential food plants to sustain critical food chains.
5 **Restockings and re-introductions of once-common species** from stocks produced by captive breeding programmes of zoological and botanical gardens. Introduced organisms from these sources do not necessarily make a successful transition, so their initial progress requires monitoring carefully.

13 **Evaluate** the particular challenges in management of a nature reserve that is maintained near your home, school or college.

Active management measures vary greatly with individual reserves. Local reserves need to be consulted to appreciate the significance of the challenges of conservation in your locality.

Endangered species – an appraisal of *in-situ* conservation *versus ex-situ* conservation

Endangered species typically have very low population numbers and are in serious danger of becoming extinct. Our own self-interest requires the preservation of endangered species. The reasons are numerous.

When a particular species becomes extinct, its genes are permanently lost, and the total pool of genes on which life operates is diminished. Wild organisms contribute to a pool of genetic diversity useful to genetic engineers.

Many wild plants are sources of compounds with medicinal value. Some wild organisms are more efficient energy converters than existing crops and herbs in particular ecosystems.

It is also essential to maintain a diverse flora and fauna to ensure the continuation of the processes of evolution of new species in response to the changing environments of the Earth.

International action

As noted above, the International Union for the Conservation of Nature and Nature Reserves coordinates the updating of the Red Data Books which list endangered species, identifying those for which special conservation efforts are needed. Problems are well documented.

How may endangered species best be preserved?

In effect, the alternative practical approaches (as immediate responses) to the problem are:

- **habitat preservation** (*in-situ* conservation) – this involves the setting up and maintenance of nature reserves, national parks of representative habitats, and conservation areas all over the world;
- **captive breeding programmes** (*ex-situ* conservation) in zoological and botanical gardens, and the building up of seed banks.

Terrestrial and aquatic nature reserves (*in-situ* conservation)	**Captive breeding programmes of zoological and botanical gardens and seed banks (*ex-situ* conservation)**
Habitats that are already rare are especially vulnerable to natural disaster – rare habitats themselves are easily lost, if a range of examples are not preserved as nature reserves.	Originally, zoos were collections of unfamiliar animals kept for curiosity, with little concern for any stress caused to the animals. Today, captive breeding programmes make good use of these resources.
When a habitat disappears the whole community is lost, threatening to increase total numbers of endangered species.	Captive breeding maintains the genetic stock of rare and endangered species.
A refuge for endangered wildlife allows these species to lead natural lives in a familiar environment for which they are adapted. They are also able to fit into their normal food chains.	The genetic problems arising from individual zoos having very limited numbers to act as parents is overcome by cooperation between zoos (and artificial insemination in some cases).
The biota of a reserve may be monitored for early warning of any further deterioration in numbers of a threatened species, so that remedial steps can be taken.	Animals in zoos tend to have significantly longer life-expectancies, and are available to participate in breeding programmes for much longer than wild animals.
The offspring of endangered species are nurtured in their natural environment and gain all the experiences this normally brings, including the acquisition of skills from parents and peers around them.	Captive breeding programmes, for most species, have been highly successful. However, the young do not grow up in the wild, so there is less opportunity to observe and learn from parents and peers.
There is an established tradition of maintaining reserves and protected areas in various parts of the world, so there is much experience to share on how to manage them successfully.	Captive breeding programmes generate healthy individuals in good numbers for attempts at re-introduction to natural habitats, a particularly challenging process, given that natural predators abound in these locations.
Nature reserves are popular sites for the public to visit (in approved ways), thereby maintaining public awareness of the environmental crisis due to extinctions, and individual responsibilities that arise from it.	Zoos and botanical gardens are accessible sites for the public to visit (often sited in urban settings where many may have access), contributing effectively to public education on the environmental crisis.
Reserves are ideal venues to return endangered individuals to (the products of intensive breeding programmes) as they provide realistic conditions for re-adaptation to habitat, and progress can be monitored.	Seed banks are a convenient and efficient way of maintaining genetic material of endangered plants, with similarities to the ways seeds may survive long periods in nature.

Table 19.3 *In-situ* and *ex-situ* conservation of endangered species

Population ecology

G5.1–5.6

Population ecology is concerned with the study of factors influencing the numbers and structure of a population. Three aspects are considered here: reproduction strategies in relation to habitat and environment; a method used by ecologists to estimate population size in mobile animal species in the wild; and finally an applied aspect of population in the issue of fish stock sizes in the face of commercial fishing practices.

Reproduction in an ecological context – r-strategies and K-strategies

Some species reproduce themselves **rapidly**, and they produce **large numbers of offspring**. We say they demonstrate high fertility, and they have a short generation time. Ecologists define these organisms as having an opportunistic way of life. Typically, such species are **pioneer organisms** – they move into new habitats, breed early, and often **achieve rapid colonisation**. Migration and dispersal are important to their survival, their mortality rates are high but their **lifespans are short**. Species adopting this distinctive reproductive strategy and life style are called **r-species**.

Can you identify an animal and a plant found in habitats near you, your home or school or college, that show many of the features of r-species?

On the other hand, there are many species that reproduce **slowly** and have a **long lifespan**. These species are seen to move into habitats that are settled and stable. Once established, they **compete successfully for the resources**. Population numbers of these organisms are more or less **constant** – ecologists say they occur at the **carrying capacity** for the habitat (where carrying capacity is defined as the maximum number of individuals of a particular species that can be supported indefinitely by a given part of the environment). Species with this type of life cycle are called **K-species**.

Perhaps these strategies (r-strategy and K-strategy) are two possible extremes of evolutionary strategy?

How representative are they of most organisms' life styles?

Before answering this question, look at Table 19.4, where the profiles of species employing r-strategy and K-strategy are compared in the context of the characteristics of the likely environments that favour each.

r-strategy species	Characteristic	K-strategy species
variable; frequently new or changed by environmental crisis (e.g. fire) in which they are 'pioneer' organisms	**favoured environment**	stable and settled habitats, typically with a fairly wide and varied range of established species
reproduce early and often, producing many offspring	**rate of reproduction**	reproduce slowly, producing few offspring
very variable, but often well below the maximum that a habitat will support	**size of population**	typically small; significant time and effort is invested in rearing young
emigration and recolonisation are common events	**tendencies for emigration and immigration**	typically settled, stable communities – individuals unlikely to move on
short	**lifespan**	long
typically poor competitors	**response to competition**	typically strong competitors
often high and variable	**mortality rate and survival qualities**	often low and typically regular

Table 19.4 r-strategies and K-strategies compared in context

r-strategy and K-strategy as extremes?

When biologists attempt to compare possible examples of organisms which display r- and K-strategies, agreements are not always easily achieved.

If we take the case of human populations, for example, it seems to depend on circumstances. In countries in developed parts of the world, many, if not most, humans conform as K-species. However, in less developed parts, many families exhibit several r-species characteristics. It seems that ecological disruption favours r-strategies.

Can you think of examples? Does this seem true to you?

There are some organisms that do conform to one or other of these strategies – after all, parasites and pests (and many weed species) certainly exhibit r-strategies.

However, it is probable that most organisms have life histories that are to varying degrees intermediate between these two extreme situations. And some organisms switch strategies depending on environmental conditions – for example, the vinegar fly, *Drosophila*, and perhaps also humans.

Techniques for estimating population size

We have already noted that an important step in the study of populations is the collection of accurate information on the size of populations present in a habitat.

A total count of all the members of a population is called a **census**, and would give us the most accurate data. However, the taking of a census is normally impractical. This may be because of the size of the habitat, or because of the quick movements of many animals, or because of species that are only active at dusk or after dark.

The capture, **mark, release and recapture (MRR)** technique, also known as the Lincoln index, is a practical method of estimating population size (N) of mobile animals, such as small mammals, woodlice, or insects that can be captured and marked with a ring, tag, or dab of coloured paint or nail varnish.

For the first sample caught for marking (n_1), it is essential to use a method of marking that is totally resistant to removal by moisture or the deliberate actions of the animals. It must also be a method that does no harm to the captured animals. For example, they must not be more visible and therefore more vulnerable to predation than normal. It is essential to trap relatively large samples if the results are to be significant.

The method then requires that marked individuals, on their release, are free to distribute themselves randomly in the whole population. After randomisation has occurred, a second sample is caught (n_2), some of which will be from the first sample (n_3), recognised by their marking. The size of the population is estimated by the formula:

$$N = n_1 \times n_2 / n_3$$

MRR may be demonstrated on populations of woodlice discovered sheltering under stones and flower pots, etc. in an area of a garden or woodland. Alternatively, it can be demonstrated on populations of night-flying moths caught with light traps, or on populations of small mammals trapped in Longworth small mammal traps (provided sufficient traps are available).

Prior to experimenting on natural populations, the technique can be learned and understood by application to non-living models. Possible models include beans dispersed in sawdust or a population of dried peas in a hessian bag. Here, samples are taken by hand, the sampler being effectively blindfolded as they cannot see the individual beans or peas. These trials enable the accuracy of the MRR method to be tested because the actual population size of beans or dried peas is easily known. It will quickly be evident that the sizes of samples captured determine the accuracy of the technique; small samples may give wildly inaccurate estimates. The steps of the MRR technique are outlined in Figure 19.33.

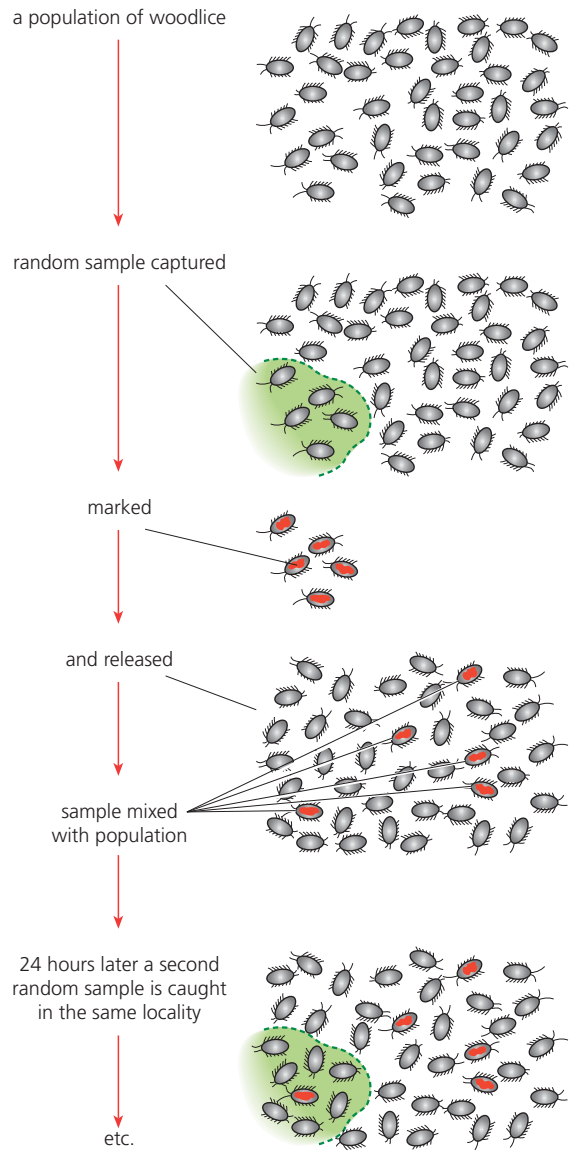


Figure 19.33 Estimating animal populations using mark, release and recapture

In 'mark, release, recapture' the size of the animal population is estimated by the formula:

$$N = n_1 \times n_2 / n_3$$

where

N = the population being estimated

n_1 = number captured, marked and released

n_2 = total number captured on the second occasion

n_3 = number of marked individuals recaptured

Here: $N = 5 \times 8 / 1 = 40$

14 Using MRR, relatively large samples must be caught for significant results. **Calculate** the estimated size of the population in Figure 19.33 if the second sample had included two marked woodlice, not one.

15 As a new habitat was colonised, the size of a population of the common rough woodlouse *Porcellio scaber* was investigated by mark, release and recapture, with the following result.

Day	**1/2**	**15/16**	**30/31**	**45/46**	**60/61**	**75/76**
Captured, marked and released (n_1)	7	14	25	19	15	19
Captured in the second sampling (n_2)	6	12	24	16	9	14
Marked individuals recaptured (n_3)	2	2	3	2	1	2

a **Calculate** the size of the population at days 2, 16, 31, 46, 61 and 76.

b **Plot** a graph of estimated population size against time.

c **Annotate** your graph as fully as you are able.

Fishing and commercial fish stocks

Fish are an important source of food for many human populations. The majority of fish populations that are exploited commercially as food sources are marine species, and it is commonly the case that the populations are shared among a number of harvesting countries. Fishery vessels from many nations take their harvest, and over-fishing is increasingly a problem (Figure 19.34). Conflict between economic and conservation interests arises – demand for the harvest of the seas and the need for economic returns for fishermen have to be balanced with our obvious need to sustain the resources themselves. In fact, the scale of commercial fishing in the seas around Western Europe in recent years has seriously depleted many stocks.

If populations of fish are persistently over-fished (as they regularly are), stocks will be rapidly depleted to the point where they collapse and can no longer support a commercial fishery. The extent to which species are fished needs to be based on population dynamics that take into account their reproductive and growth rates.

The **maximum sustainable yield (MSY)** of a stock represents the maximum average catch that a stock can sustain over a long period of time. This catch corresponds to optimum balance between the reproductive rate and growth rate of the stock and the deaths due to harvesting and natural mortality. Typically, harvesting at MSY requires much lower fishing rates than occur in many fisheries.

Calculation of MSY requires good knowledge of the relationship between the size of the stock and the number of juveniles produced each year. The natural variation in the number of juveniles produced annually makes establishing this relationship difficult. It usually requires a very long series of data – at least 20 years. Nevertheless, there are reliable methods that provide an adequate approximation to MSY.

Estimating the size of commercial fish stocks

Information on the state of fish stocks is based on four parameters which are measured or assessed by the fishery research services of national governments. These are:

- **fishing mortality** – the proportion of fish stocks taken each year during commercial fishing;
- **spawning stock biomass** (SSB) – the total mass of mature fish in the population;
- **recruitment** – the number of young fish produced each year which survive to enter the spawning stock;
- **landings** – the total annual tonnage of fish landed by the fishing fleet.

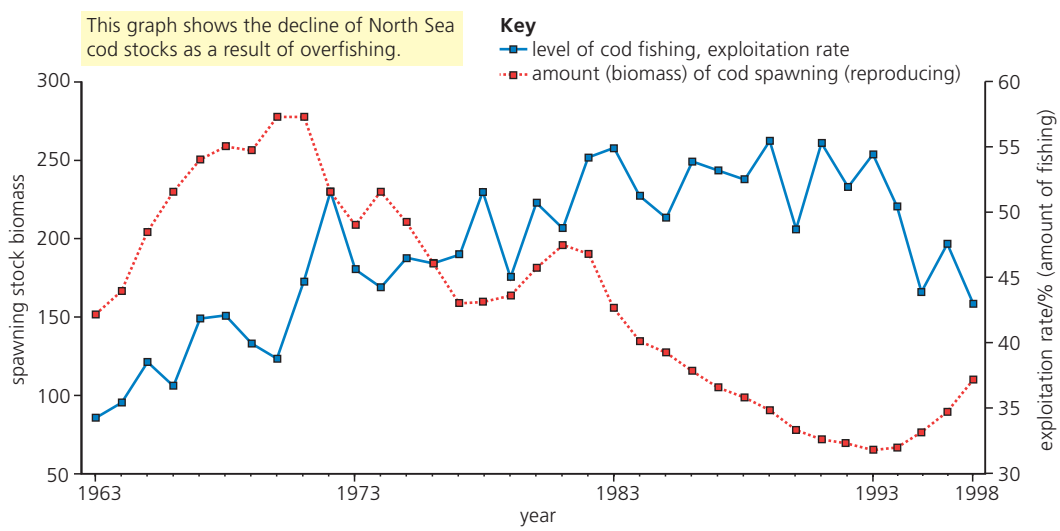


Figure 19.34 Over-fishing in European waters

How do marine fishery research scientists obtain this information?
Every year, teams of marine scientists collect data via three routes:

1 **Market sampling**, in which the landings of fish are sampled to compile information on fish length and age. Fish are aged by examination of the ear otolith, taken from the head of the animal. These consist of layers of bone which are added annually, in light and dark bands. The age is obtained by counting the number of bands (like counting tree rings in the trunks of trees).

   Data from many samples allow the age structure of the whole population to be estimated. Year by year, a picture of how the population is changing in the face of the fishing demands emerges. For example, if there are very few old fish in the stock, then over-fishing is likely as the figures indicate that survival rate is very low and few fish mature to breed.
2 **Discard sampling**, in which the mass of fish caught and then discarded is measured. Many fish caught at sea are discarded, for example, because they have no commercial value, or because they are below the legal minimum landing size, or because quota limits have been exceeded.

   These data are combined with landings data to give a more complete picture of the total catch from the stocks.
3 **Research vessel surveys**, in which distribution and abundance of adult stocks and of recruits are investigated. Surface species (such as herrings) are assessed by acoustic surveys which use sonar equipment to detect fish and convert the signal into a biomass estimate.

   Bottom-dwelling species (such as cod and haddock) are monitored by trawling surveys where the number of fish caught per hour gives an index of abundance.

With these data to hand, ecological computer models are used to create a picture of the total biomass of living fish across the whole area studied, and of the trend in any changes detected, year by year.

International measures to promote the conservation of fish

Because the scale of fishing in the seas around Western Europe has been so intensive as to over-exploit the resource, fishing nations have implemented a wide variety of control measures, including bans on fishing of certain species. Where these measures are respected, breeding stocks have recovered.

Currently, the European Union is regulating fishing around the shores of member states and in adjacent seas by a number of enforced measures.

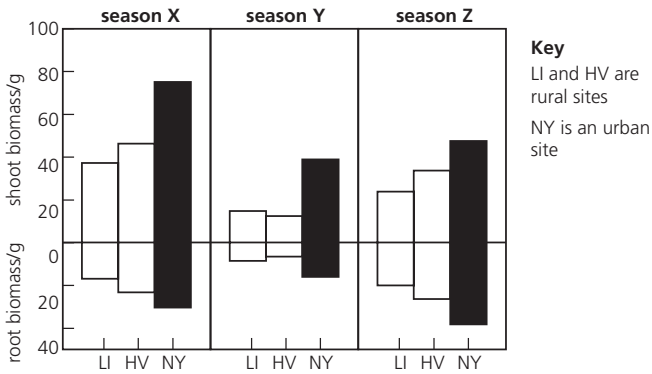
1 **Total allowable catches (TACs) and quotas**
   TACs – figures for the overall weight of fish which fishing boats can land – are agreed by member states annually. This is based on advice from scientists in the International Council for the Exploitation of the Sea (ICES) on levels of catch if breeding stocks are to be sustained. The TACs for each type of fish harvested are then divided between member states. **Each state receives a quota for individual fish stocks**, reflecting levels of dependence on fishing.
2 **Technical conservation (TC)**
   TC measures provide a further line of protection for breeding stocks. These are rules that aim to make fishing itself more selective and so reduce the discarding of young fish. These measures:
   a set **minimum landing sizes** for different species of fish to discourage the catch of small fish;
   b require the use of **specific mesh sizes** in some circumstances so that small fish can escape;
   c require the use of **separator devices** as an integral part of the gear to prevent the catch of more vulnerable stocks;
   d restrict the type of fishing gear that may be used in certain localities;
   e close down some areas to certain types of fishing, either permanently, or for a part of the year.

Meanwhile, funding is maintained for national marine biology stations to carry out the necessary research to gain a clear picture of the ecology of the fishing grounds and of the key stages in life cycles of organisms that make up the relevant food chains.

Examination questions – a selection

Questions 1–4 are taken from past IB Diploma biology papers.

Q1 A study was conducted to investigate the growth factors affecting plants in urban areas compared to rural areas. Fast-growing clones of Eastern cottonwood trees (*Populus deltoides*) were grown in both urban and rural sites. The results of three successive growing seasons (X, Y and Z) are shown below.



a Identify the site which most favours the growth of cottonwood trees. (1)

b Calculate the ratio of shoot biomass to root biomass in site L1 during season X. (1)

c Analyse the data for growth patterns over three years of the study. (3)

A further study showed that differences in light, temperature, water, CO_2 concentration, and the soil could not account for the differences in growth of the cottonwoods in the urban and rural areas.

d Suggest a reason which could account for the growth differences. (1)

Higher Level Paper 3, May 05, QG1

Q2 **a** Define the term *biomass*. (1)

b State **two** organisms that interact together by mutualism. (1)

c Explain why the biomass at higher trophic levels tends to be small. (3)

Standard Level Paper 3, November 05, QG2

Q3 **a** Describe the use of *ex situ* conservation measures. (3)

b **i** Define the term *niche*. (1)

ii Explain the niche concept using named organisms. (4)

Standard Level Paper 3, November 04, QG3

Q4 **a** Simpson's Index is given by the following equation:

$$D = \frac{N(N-1)}{\Sigma n(n-1)}$$

where D = the diversity index, N = the total number of all species found and n = the number of individuals of a particular species.

i State what happens to this index if the number of all species found increases but the number of individuals in a species stays the same. (1)

ii State what a high value of D suggests about an ecosystem. (1)

b Explain the use of biotic indices in monitoring environmental changes. (3)

Standard Level Paper 3, November 03, QG3

Questions 5–10 cover other syllabus issues in this chapter.

Q5 **a** Outline the effects of **three** environmental factors that affect the distribution of plants. (6)

b Explain how:
i temperature **ii** water supply
may effect animal distribution. (6)

c Discuss the special value of transects in analysing the distribution within a **named** community of organisms that continually experiences the impact of unidirectional abiotic factors. (8)

Q6 **a** Outline the difference between a *primary succession* and a *secondary succession*, using **named** examples. (6)

b Distinguish between the ecological concepts of *biosphere* and *biome*. (4)

c Compare the main features of a *tropical rainforest* and a *temperate deciduous forest*, by means of a table. (6)

Q7 Evaluate **three** reasons in support of our striving to maintain species diversity in the face of increasing extinctions, today. (9)

Q8 **a** Describe how the *ozone layer* in the stratosphere is created and maintained. (6)

b Explain the importance of the ozone layer for terrestrial organisms. (3)

Q9 **a** Discuss the ecological significance of r- and K-strategies in an ecological context. (6)

b Suggest how you might classify human communities in terms of r- and K-strategies. (4)

Q10 **a** Describe **one** technique used to estimate animal population size. (4)

b Outline the methods used to estimate the size of commercial fish stocks in marine waters. (4)

c Explain what international measures might best promote the conservation of marine fish stocks. (4)

Option H

20 Further human physiology

STARTING POINTS

- In mammals, **internal communications** by which responses to appropriate stimuli are brought about, involve both the nervous system and **hormones**.
- **Digestion** occurs in the alimentary canal or gut, a long muscular tube. Chemical digestion is by **enzymes** which may be secreted onto the broken-up food in particular regions of the gut, under controlled conditions of pH.
- The products of digestion are absorbed in the **small intestine**, where the surface area is vastly increased by **villi**. Some of the products of digestion enter the blood stream immediately and are transported to the **liver**.
- The four-chambered **heart** of mammals is divided into right and left halves. The right side pumps **blood to the lungs** and the left side pumps **blood to the rest of the body**. From the heart, blood is delivered to **capillaries** in the tissues by **arteries** and returns to the heart in **veins**.
- In mammals, **gaseous exchange** occurs in the lungs in the thorax cavity, which moves rhythmically to ventilate the lungs. The **circulation system transports oxygen all over the body**, combined with **haemoglobin** of the red cells.
- This chapter extends study of aspects of human physiology begun in Chapter 7 (pages 178–234), and developed in Chapter 12 (pages 348–89).

The essence of physiology is questioning **how and why the parts of the body function as they do**. The outcome, our understanding of the working body, underpins exercise physiology and sports science, and all activities where the body may have to function under extreme conditions and up to the limits of its potential. Equally, the physiology of the normal, healthy body is the standard by which the impact of disease and conditions of malfunction are assessed by the medical profession and pharmaceutical industry. It was the Frenchman, Claude Bernard (1813–78) who established physiology as an experimental science, distinct from anatomy, in the mid-nineteenth century. His most important contribution was the discovery of a regulated and constant internal environment – maintained by **homeostasis**.

In this chapter, study of control and communications within the body, which was previously focused on the nervous system, is now extended to the roles of **hormones**. Nutrition – the process by which the organism obtains energy to maintain the functions of life and matter to build and maintain the structures – is investigated further, particularly **chemical digestion** and the **fates of absorbed nutrients**. Then the functioning of the **heart as a pump**, the **control of the heart beat**, and heart disease are explored. Finally, the **transport of respiratory gases** and the **regulation of breathing rate** are discussed in normal function, at high altitudes, and in a disease condition.

Hormonal control

H1.1–1.5

Hormones are **chemical messengers** produced and secreted from the cells of the ductless or **endocrine glands**. They are transported indiscriminately in the blood stream, but they act only at specific sites, appropriately called **target organs**. Although present in small quantities, hormones are extremely effective in the control and co-ordination of several body activities.

Hormone control of body function is quite different from nervous control; the latter is concerned with quick, precise communication, whereas hormones mostly work by causing specific changes in metabolism and development, often over an extended period of time. However, these contrasting systems are co-ordinated by the brain – nervous system and hormones work together.

The **structure of endocrine glands** can be contrasted with that of the ducted glands (exocrine glands, page 647), such as the salivary glands and sweat glands. Ducted glands deliver their secretions down tubular ducts, whereas endocrine glands are ductless glands that secrete hormones directly into the blood stream.

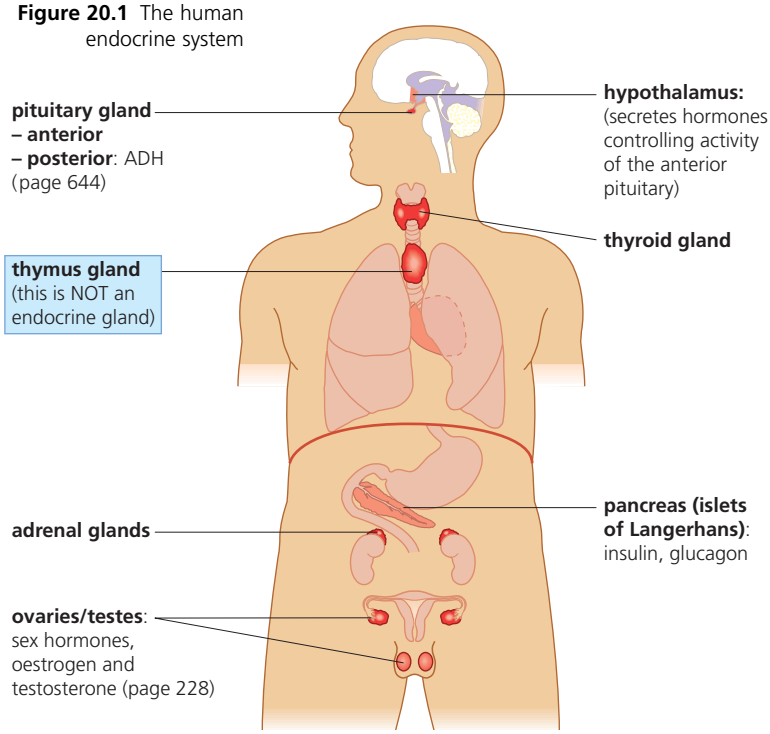


Figure 20.1 The human endocrine system

Once released, hormones typically work by triggering changes to specific metabolic reactions in their target organs. Examples of hormones that have already been introduced are **insulin** and **glucagon**, released from the islets of Langerhans in the pancreas (page 222) and **antidiuretic hormone** (**ADH**, page 374), which has its effects in the kidneys.

A hormone circulates in the blood stream only briefly because they are continually being broken down in the liver and any breakdown products that are no longer of use to the body are excreted in the kidneys. So, long-acting hormones must be secreted into the blood stream continuously to be effective.

The position of the chief endocrine glands and the hormones produced are summarised in Figure 20.1.

Chemical analysis of animal hormones shows that they fall into three groups:

- some are **steroids** (e.g. the sex hormones, oestrogens and testosterone);
- some are **peptides** or **proteins** (e.g. ADH);
- some are derived from the amino acid **tyrosine** (e.g. thyroxin from the thyroid gland).

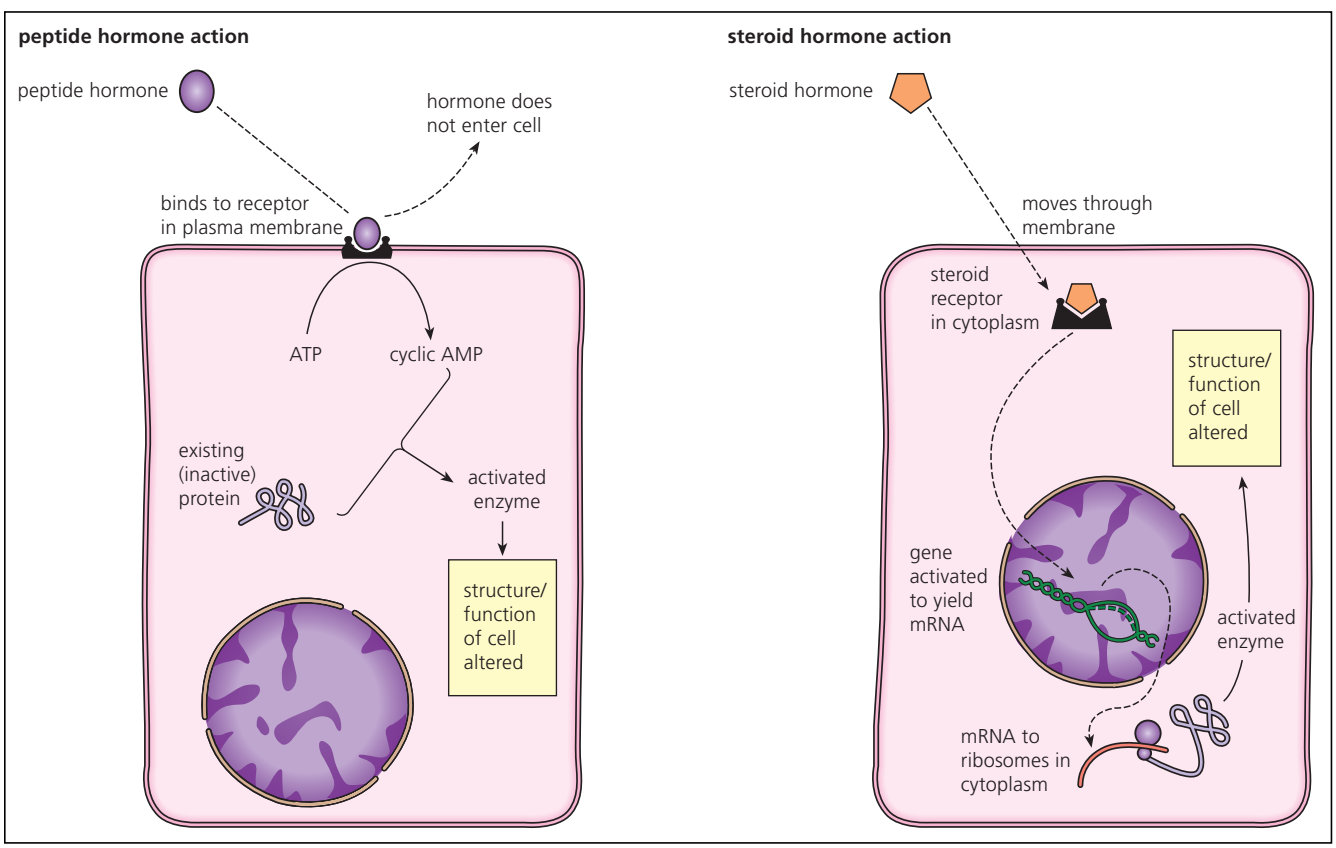


Figure 20.2 How hormones influence target cells

Two modes of hormone action

Since hormones are carried in the blood stream, they reach all cells of the body. The effects of a hormone are to alter the metabolic reactions only in certain cells. These cells possess specific receptor molecules on the external surface of their plasma membranes; if a cell lacks an appropriate receptor site, it will not respond to the hormone.

However, it turns out that there are **alternative ways** in which hormones have an impact on the metabolism of target cells, depending on whether the hormone is a steroid or is made of protein or peptide (Figure 20.2).

Steroid hormones cross the plasma membrane of their target cells, and cause the **activation of a specific gene** on a chromosome in the nucleus. The effect on the gene is to trigger the synthesis of a particular protein (most likely an enzyme) that will then bring about change.

Protein or peptide hormones, on the other hand, first **bind to a receptor** in the plasma membrane of the target cell. This type of hormone remains outside the cell, but the activated receptor causes release of a secondary messenger on the inside of the plasma membrane. The secondary messenger then causes the activation of an existing protein in the cytoplasm, thereby bringing about a specific change in metabolism.

Roles and relationships of the hypothalamus and pituitary gland

The organisation and layout of the nervous system is shown in Figure 7.30 (page 210), including the central nervous system of brain and spinal cord. The brain itself consists of forebrain, midbrain and hindbrain regions within its compact structure, as shown in Figure 17.23, page 534. Here, we are going to consider a tiny, but vital part of the brain that has endocrine functions: the hypothalamus.

The **hypothalamus** is part of the floor of the forebrain. Most importantly, it is exceptionally well supplied with blood vessels and is the site of special neurones. The hypothalamus has a key role in the monitoring and control of many aspects of body function. Partly, this is achieved by the constant monitoring of blood composition as it circulates through the capillary networks of the hypothalamus. This data, and that which comes from sensory receptors located in key organs in the body and sensory neurones via the spinal cord, enables the hypothalamus to regulate many body activities concerned with maintenance of a constant internal environment – homeostasis (page 217).

The **pituitary gland** is situated below the hypothalamus, but is connected to it. This gland consists of **two parts**, the anterior pituitary and the posterior pituitary lobes. The pituitary gland

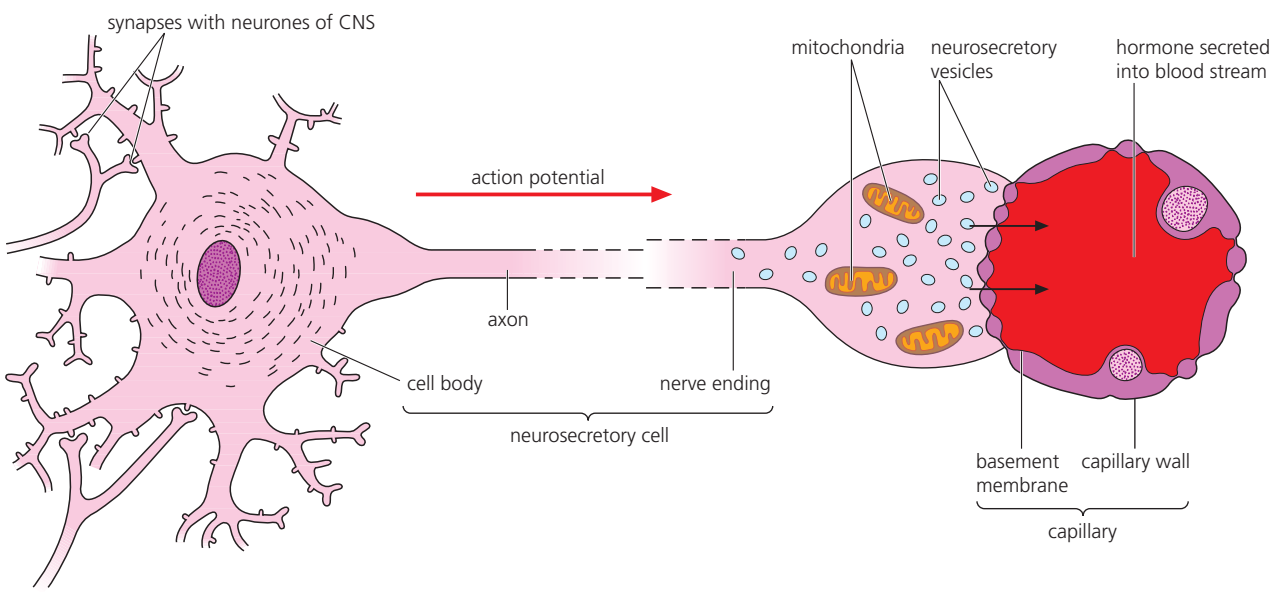


Figure 20.3 A working neurosecretory cell of the hypothalamus

is part of the endocrine system, and it has been called the master hormone gland. However, it is the hypothalamus that largely controls the endocrine activity of the pituitary gland. The hypothalamus does this by releasing a number of hormones from its special **neurosecretory cells**, into the **portal vein** running between hypothalamus and anterior lobe (Figure 20.3), as well as nerve impulses via other neurones that connect with the pituitary.

A neurosecretory cell is a special type of neurone that secretes chemical messengers that travel round the body via the blood circulation.

The **anterior pituitary** produces and secretes hormones that regulate a wide range of body activities and functions, from growth to reproduction. The release of these hormones is triggered by **releasing hormones** from the hypothalamus (via neurosecretory cells) and inhibited by **inhibiting hormones**, also from the hypothalamus.

The **posterior pituitary** does not synthesise hormones but it does store and release two hormones, one of which is the **antidiuretic hormone** (**ADH**). These hormones are produced in the hypothalamus and stored in vesicles at the ends of neurosecretory cells in the posterior pituitary. Specific nerve impulses from the hypothalamus trigger release of specific hormones which pass into the capillary networks in the posterior pituitary and are then circulated in the blood stream (Figure 20.4).

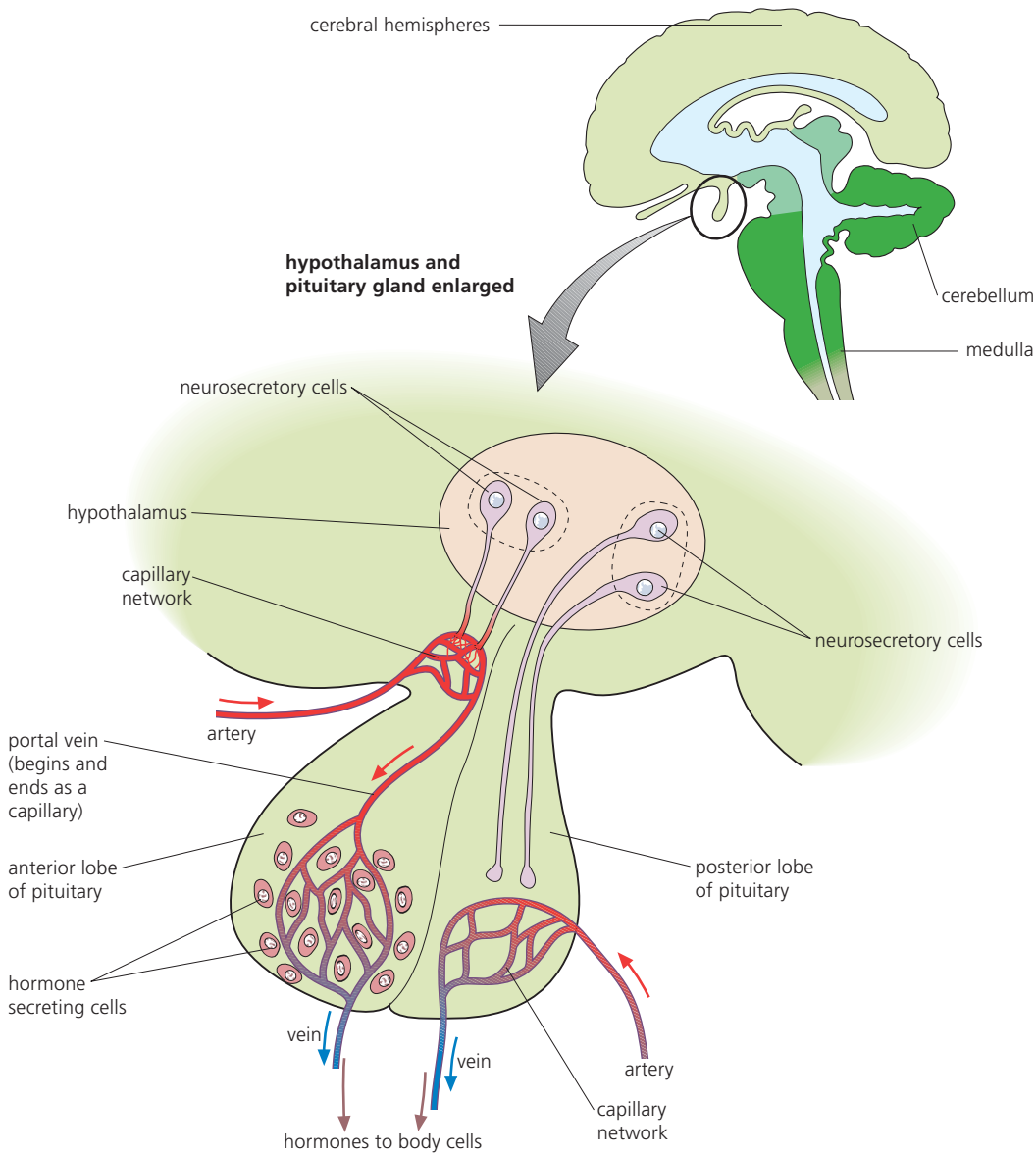


Figure 20.4 The hypothalamus and the pituitary gland

1 In the nervous system, impulses reach a specific muscle or gland. **Explain** how the effects of hormones are restricted to particular cells or tissues.

So, the endocrine system and the nervous system work in distinctive and different ways in the control and co-ordination of body activities. However, their **activities are co-ordinated** by the pituitary gland working in tandem with the hypothalamus of the brain. We can illustrate this relationship by examining how the water balance of the body is regulated.

Hypothalamus, pituitary and kidneys in the control of water balance

The structure of the **kidneys** and their kidney tubules, known as **nephrons**, was introduced in Figure 12.20, page 371.

Remind yourself of nephron structure, now.

The site of the regulation of the water content of the body is the kidneys, particularly the **collecting ducts** into which the nephrons drain. However, the **balance of water** in the body (whether there is too little or too much – known as its osmolarity) is detected in the hypothalamus.

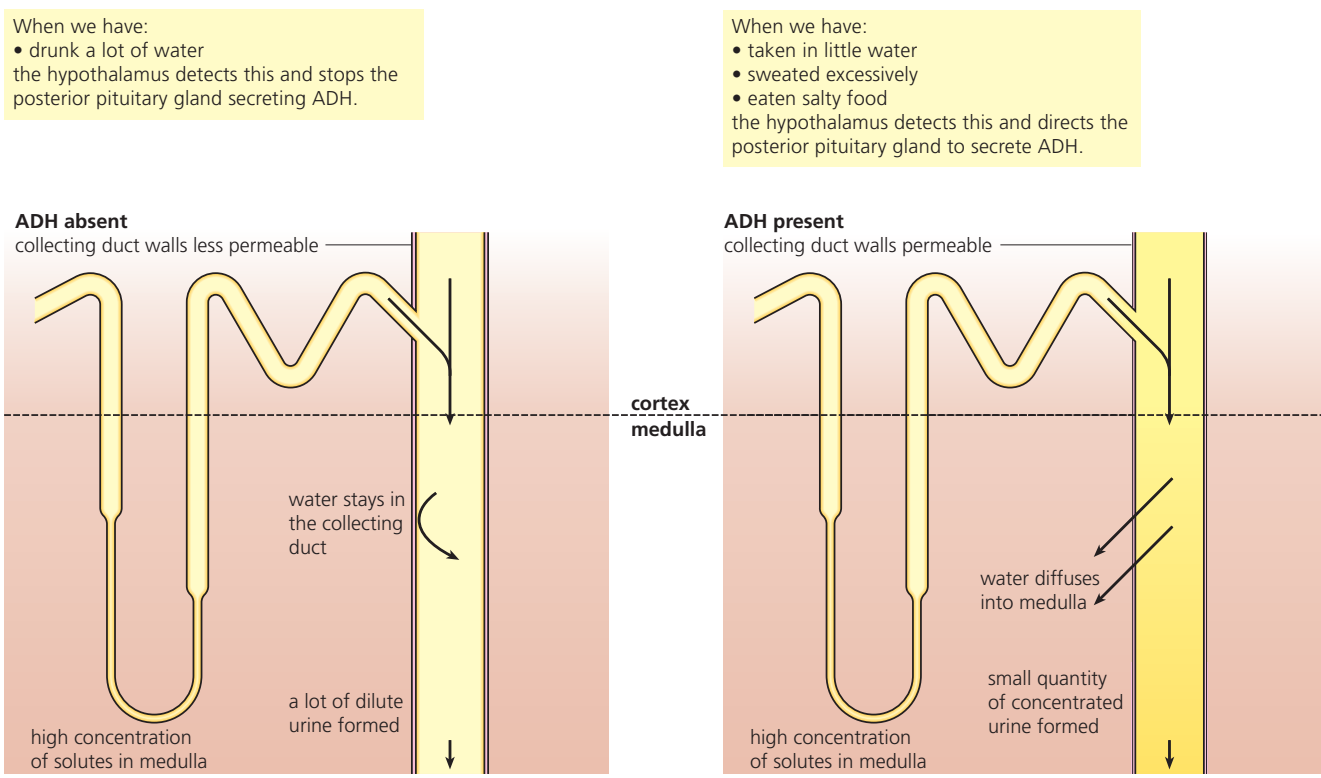
In the hypothalamus, changes in the solute potential of the blood plasma are detected by a group of cells, known as **osmoreceptors**. The composition of the blood in the capillaries, particularly the osmolarity, is constantly monitored as the blood passes through the hypothalamus.

If the body has too much water, then more water must be lost from the body in the urine (a process we call **diuresis**). If the body has too little water, then more water must be retained, and highly concentrated urine formed. The sequence of events by which the hypothalamus, posterior pituitary gland and collecting ducts work together in this task is as follows:

- When the water content of the blood is low, more ADH is secreted by the pituitary gland. Release of ADH is triggered by nerve impulses from the hypothalamus, down the axons of neurosecretory cells, the bulbous ends of which contain stored ADH. As a result, ADH is released and transported in the blood circulation, eventually reaching the collecting ducts.

  In the **presence of ADH** in the blood stream, the walls of the collecting ducts become fully permeable. This allows water to be withdrawn from the filtrate passing down the tubular collecting ducts, back into the medulla, due to the high concentration of sodium and chloride ions maintained in the medulla by the actions of the loops of Henle (page 374). This water is taken up and redistributed in the body by the blood circulation. The outcome at the kidneys is that only **a small amount of concentrated urine is formed** (Figure 20.5).

Figure 20.5 Water reabsorption in the collecting ducts



- When the water content of the blood is high, little or no ADH is secreted from the posterior pituitary gland.

  In the **absence of ADH**, the walls of the collecting duct become less permeable. The result is that **a large quantity of very dilute urine is formed** (Figure 20.5).

How does ADH change the permeability of the walls of the collecting ducts?

The plasma membranes of the cells that form the walls of the collecting ducts contain proteins as an integral part of their fluid mosaic structure, as do other cells in the body. However, sited in the plasma membranes are a high proportion of channel proteins capable of having an open pore running down their centre. You can see the structure of a fluid mosaic membrane and its channel proteins in Figure 1.19, page 21.

When there is an excess of ADH in the blood circulating past the kidney tubules, this hormone causes the protein channels present in the collecting duct membranes to be open. As a result, much water diffuses out into the medulla and very little diffuses from the medulla into the collecting ducts.

When ADH is absent from the blood circulating past the kidney tubules, the protein channels in the collecting duct plasma membranes are closed. Water retention by the medulla tissue is now minimal.

Meanwhile, as ADH circulates in the blood to all organs of the body, the actions of the liver continually remove this hormone and inactivate it. This means that the presence or absence of fresh ADH release has an immediate regulatory effect.

The mechanism of osmotic concentration control of the blood is an example of **negative feedback**. The role and activity of the osmoreceptors in the hypothalamus in maintaining constant monitoring activity there, together with the responsiveness of the protein channels of the collecting duct walls to the presence or absence of ADH, and the prompt inactivation of blood ADH in the liver, all ensure sensitive negative feedback control of blood water level (Figure 20.6).

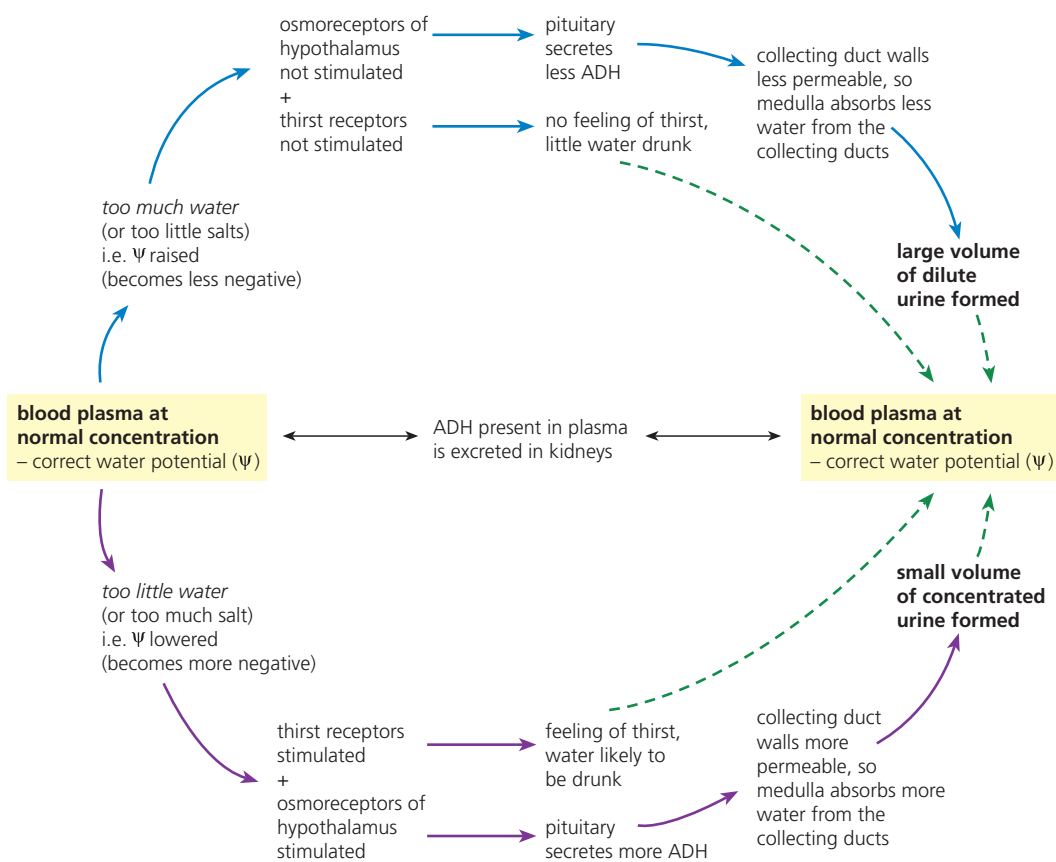


Figure 20.6 Osmoregulation by negative feedback

2 **Describe** what is meant by negative feedback control, giving a second example, and **explain** how it differs from positive feedback.

Digestion and the absorption of digested food

H2.1–2.9, H3.1–3.4

In the nutrition of mammals, food is taken into a specialised structure called the **alimentary canal** (**gut**) – a long muscular tube beginning at the mouth and ending at the anus. Along the gut are several glands, and the whole structure is specialised for the movement and digestion of food and for the absorption of the useful products of digestion. The digestion of food by enzymes and the way the useful products are taken up into the body are discussed next.

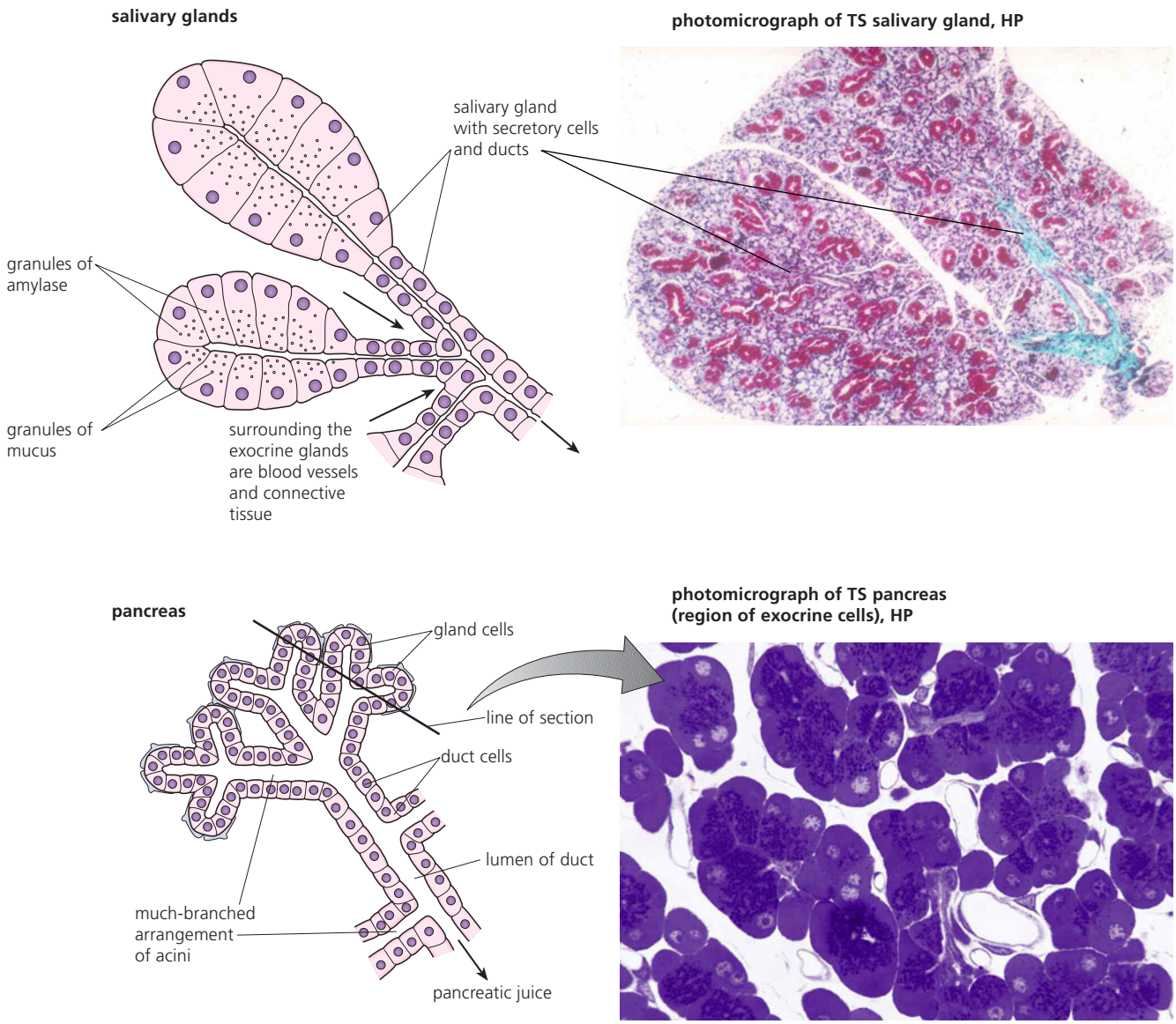
Digestive enzymes are proteins. They are secreted onto the food in the gut, and these enzymes (together with those in cells in the walls of the small intestine), complete digestion of the food. The major **glands secreting digestive enzymes** into the alimentary canal are:

- the **salivary glands** in the mouth;
- the **gastric glands** in the stomach;
- the **exocrine glands** of the pancreas;

together with

- the **cells of the walls of the small intestine**.

Figure 20.7 Structure of exocrine digestive glands



The structure and function of exocrine glands in digestion

Glands that secrete digestive enzymes are **exocrine glands** because they secrete into a space or lumen. The clustered **secretary cells** of an exocrine gland, arranged around the space into which secretion takes place, are called **acini** (Figure 20.7).

The secretory cells themselves have a distinctive structure. The cytosol is packed with **rough endoplasmic reticulum** (**rER**). This is evident from electron microscope images (Figure 20.8). In rER, the attached **ribosomes** are the sites of the synthesis of proteins to be packaged for export from the cells (the **secretory vesicles**), so their presence in excess there is no surprise. Also present are many **mitochondria** – the source of the ATP necessary for the protein synthesis.

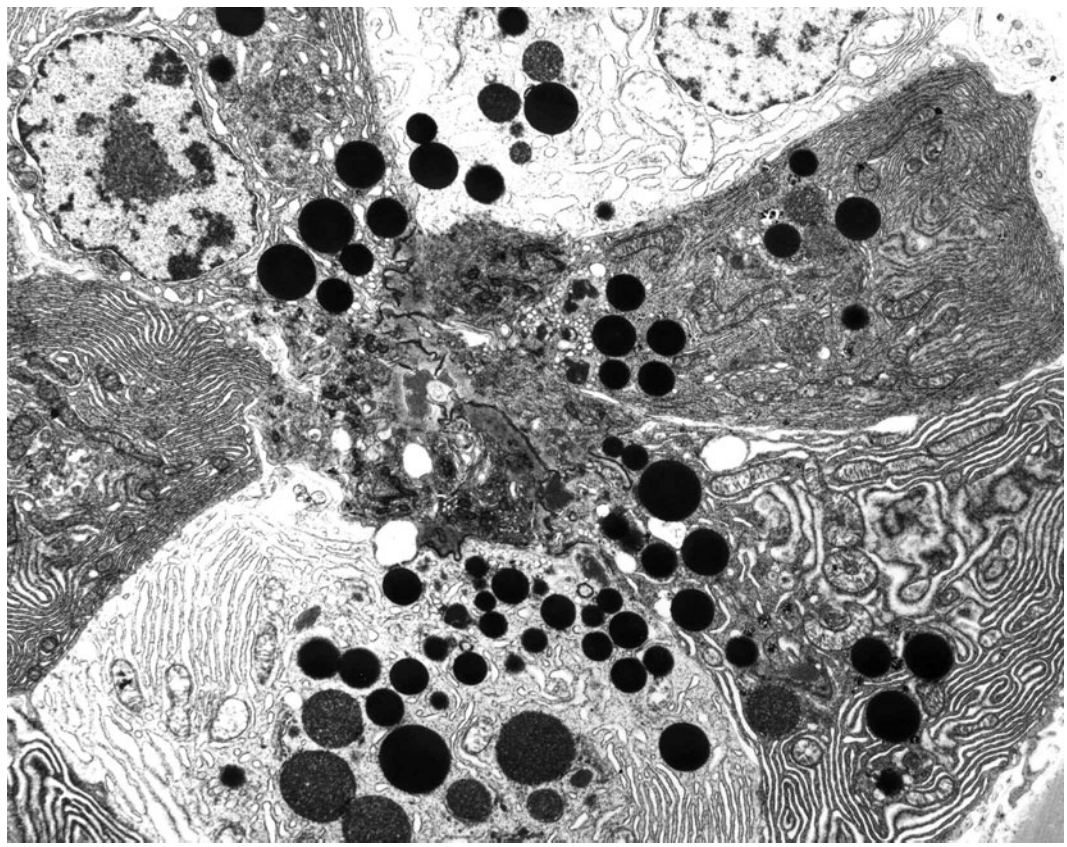


Figure 20.8 TEM of secretory cells of the exocrine glands of the pancreas

Here we see parts of six gland cells arranged around a central space that leads into a duct, down which pancreatic juice may flow (the whole known as acini of the pancreas – Figure 20.7, page 647)

Interpretation of electron micrographs – a skill to develop
Make a representative drawing of an exocrine cell showing key organelles recognised, including:
- mitochondria
- nucleus
- rough endoplasmic reticulum (rER)
- lysosomes
- plasma membrane

The **contents** of the secretions flowing from the salivary glands, gastric glands, and the pancreas are compared in Table 20.1.

Note that an important, additional role of saliva (and possibly its chief one in the primate mammals with their communication by vocalisation) is the lubrication of the mouth and throat. Think how difficult it is to talk with a dry throat!

Another point to note is the pronounced **pH changes** that occur in the environment of digestion along the gut.

After the mouth, the **stomach** contents are turned strongly acidic, but on entry to the first stage of the **small intestine** (it is known as the duodenum), the pH returns to mildly alkaline conditions. Acidity in the stomach is brought about by the hydrochloric acid component of gastric juice itself.

However, in the duodenum although pancreatic juice itself is alkaline, the strong acid present in the digesting meal on arrival (called chyme) is neutralised by the addition of **bile from the gall bladder**. Since bile contains no enzymes and the gall bladder is not a digestive gland, it is not listed in the table.

Finally it is also useful to observe that in the duodenum and subsequently, all essentials of the diet that require digestion before absorption (that is the carbohydrates, fats and proteins) are digested simultaneously. Digestion began in earnest in the stomach, but it was only the digestion of protein that commenced there.

		Active ingredients, conditions required, and outcome			
Secretion and gland	**Site of action**	**pH**	**enzymes and non-enzyme components**	**substrate and effect**	**product**
saliva	mouth	6.5–7.5	amylase	starch (polysaccharide)	maltose (disaccharide)
salivary glands			mucus	lubricates	
gastric juice	stomach	2.0	pepsin	proteins	polypeptides
gastric glands in wall of stomach			rennin (young mammals only)	coagulates milk protein	
			hydrochloric acid	creates acidic environment that kills bacteria	
pancreatic juice	small intestine	7.0	amylase	starch	maltose
pancreas			proteases (trypsin and chymotrypsin)	proteins	polypeptides
			peptidases	polypeptides	peptides and amino acids
			lipases	triglycerides	fatty acids and glycerol
			nucleases	nucleotides	pentose sugars, P_i and bases

Table 20.1 Digestive secretions, sites and actions

Proteases secreted as inactive precursors

Proteases are powerful enzymes which, if present in an active state in the cells in which they are produced, would hydrolyse the cell protein, and disrupt the cell (a phenomenon known as self-digestion or **autolysis**). In fact, cells synthesise these enzymes in **inactive forms**, and they are secreted also in an inactive form. Only in the gut, in the presence of food, are they activated.

There are two **peptidase** enzymes that may be present in the gastric juice secreted by the gastric glands of the stomach, both in an inactive form.

Pepsin is secreted as **pepsinogen**, and is activated by the hydrochloric acid as they mix in the gastric juice. Pepsin breaks protein chains into shorter length polypeptides by hydrolysing peptide linkages where tyrosine and phenylalanine residues occur. Pepsin is known as an endopeptidase (endo- means 'within') because it catalyses the breakdown of proteins at numerous points within the long protein chain, breaking large proteins into small polypeptides. (Exopeptidases hydrolyse peptide linkages at the ends of peptides, releasing individual amino acids – see below.)

Rennin is produced only in the gastric glands of young mammals. Its inactive form is called **prorennin**. The conversion of prorennin to rennin is also triggered by the hydrochloric acid of gastric juice. Rennin commences the digestion of the milk protein caseinogen by coagulating it. In the coagulated form, milk protein is digested by pepsin. It is milk that forms the complete diet of very young mammals.

3 **Predict** the effect on the pancreatic enzymes if the pH of the chyme in the duodenum remained the same as that of the stomach contents.

In the duodenum, two more proteases are secreted, also in their inactive forms. One is **trypsinogen**, which is activated to **trypsin**. This activation is brought about by the enzyme enteropeptidase, secreted by the epithelial cells in the villi. The other protease present in the pancreatic juice that is secreted into the duodenum is **chymotrypsinogen**. This is activated to **chymotrypsin** by trypsin. Both these enzymes hydrolyse proteins to polypeptides. Also present in pancreatic juice are a number of other peptidases that eventually hydrolyse the polypeptides to tripeptides, dipeptides and free amino acids.

The control of secretion of digestive juices

Digestive juices are not secreted continuously, but are co-ordinated with the presence of food in a particular part of the gut. This co-ordination is regulated by nervous or hormonal mechanisms, or both. For example, saliva, which mainly aids swallowing, is produced by a reflex action triggered by the sight, smell, taste or thought of food.

Control of the secretion of gastric juice is more complex. First, as with saliva, the sight or smell of food may trigger secretion by the gastric glands before food is taken into the mouth. Then, when the arrival of food in the stomach stretches the stomach wall, this mechanical stimulation triggers further secretion of gastric juice by reflex action. At the same time, cells in the stomach lining are stimulated to secrete a hormone, **gastrin**, into the blood stream. On reaching the gastric glands via the blood circulation, this stimulates further secretion of gastric juice. The contributions of these three components to the regulation of gastric juice secretion are shown in the graph in Figure 20.9.

Hormones continue to regulate secretion in the subsequent stage of digestion, too. In the duodenum, the pancreatic juice (and bile) are regulated by hormones (secretin and cholecystokinin) produced by cells in the wall of this part of the gut.

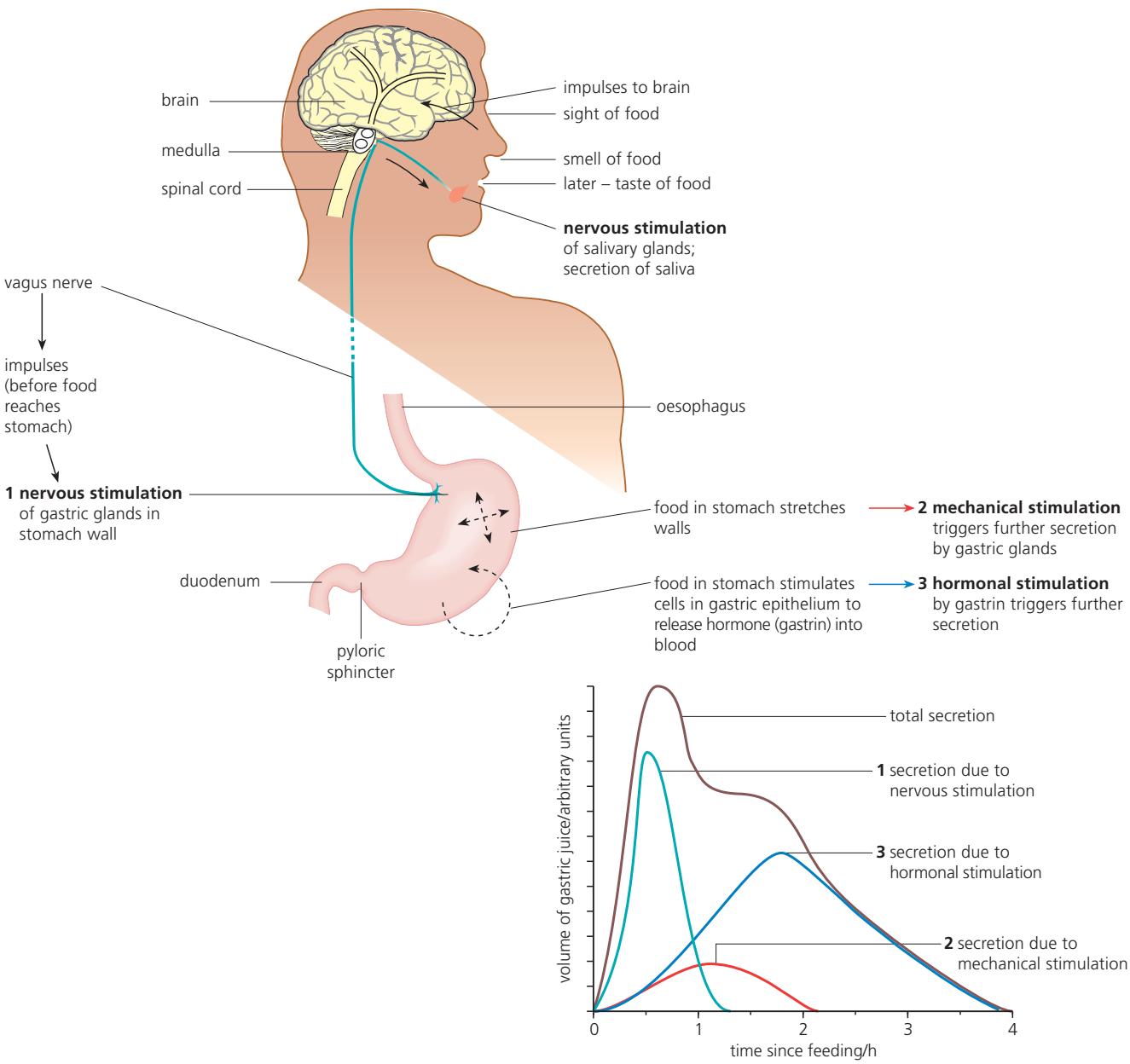


Figure 20.9 The control of gastric secretion

Digestion in the small intestine

The small intestine is long, about 5–6 metres in total, but the first region, the **duodenum**, is only about 30 cm long. From the stomach, the chyme passes into the duodenum where the bulk of digestion is completed. Into the duodenum runs bile from the **gall bladder** and **pancreatic juice** from the pancreas. From Table 20.1 we can see that enzymes present in the pancreatic juice include those that digest carbohydrates, lipids and proteins.

Bile contains no enzymes. What is it for?

Lipid in a hydrophilic medium – the roles of bile

In the aqueous environment of the gut, the lipids of our diet, being strongly hydrophobic substances, resist break-up into small droplets. Lipid molecules tend to coalesce.

Yet a first step in digestion is the break-up of food into smaller units with a huge surface area for the action of digestive enzymes to work on. It is only at the lipid–water interface that lipid is accessible to lipase. Lipase, like all digestive enzymes is water soluble, but has an active site to which its hydrophobic substrate binds.

The teeth start the break-up processes in the mouth. In the stomach the muscular contractions of the wall continue this, but as the lipids are freed from other components of the diet, they naturally coalesce into quite large globules. In an aqueous medium, they continue to resist break-up.

This problem is overcome in the duodenum, due to the action of the special fluid, bile. Bile is produced in the liver cells (page 657). In humans (and some other mammals) bile collects in the gall bladder, and intermittently passes into the duodenum via the bile duct. Bile is a yellow-green, mucous fluid containing the bile salts, bile pigments, cholesterol and salts. This strongly alkaline fluid plays a major part in neutralising the acidity of the chyme.

The other main role of bile is to lower the surface tension of large fat globules, causing them to break into tiny droplets which enormously increases their surface area. It is the bile salts in the bile that bring these changes about, for bile salts are molecules with both hydrophilic and hydrophobic properties. Tiny spheres of lipid are formed, with the hydrophobic part of bile salts embedded in their surfaces, and the hydrophilic parts exposed. The droplets in this condition, known as **micelles**, remain suspended in an aqueous medium. This process is called **emulsification**. It greatly speeds digestion by lipase.

The structure of the small intestine and its villi

The wall of the small intestine, as elsewhere in the gut, contains involuntary muscle tissue, and its inner surface is lined with an epithelium layer liberally enriched with goblet cells, called a **mucosa**. The goblet cells of the mucosa secrete **mucus**, a lubricating secretion that adheres to the cells lining the gut (Figure 1.8, page 11). Mucus protects the gut from mechanical damage (and helps resist self-digestion, too).

In the small intestine, the inner layers of the gut wall are shaped into finger-like projections called villi. Needless-to-say, the epithelium of villi contains many goblet cells, too. The structure of the wall of the small intestine, the distinctive villi it supports, and their roles in aiding absorption were introduced in Figure 7.5 and Table 7.4 (page 183–4).

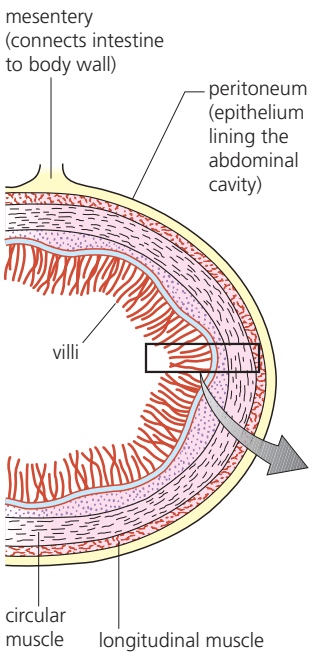
It would help to refresh your memory of that table, now.

Here, a TS of the small intestine is shown (Figure 20.10), together with a labelled drawing of a tissue map. The structure of an individual epithelial cell from a villus is shown in Figure 20.11.

With the aid of the transmission electron microscope, we are able to see that the epithelium cells lining the villi have important structural features. Several of these features have parts to play in the absorption of the useful products of digestion that occurs here including:

- **microvilli** – these tiny, finger-like infoldings of the cell surface facing the lumen of the gut greatly increase the surface area in contact with material to be absorbed;
- **mitochondria** – these organelles are present in large numbers, suggesting a significant demand for ATP in these cells;
- **pinocytotic vesicles** – these are the site of pinocytosis by which fluid is taken up or released in tiny vesicles, across the plasma membrane of a cell;
- **tight junctions** – these bind together the individual epithelial cells, so that the only way into the tissues of the body is through the epithelium.

Figure 20.10 The small intestine in TS



photomicrograph of the small intestine wall in TS, LP

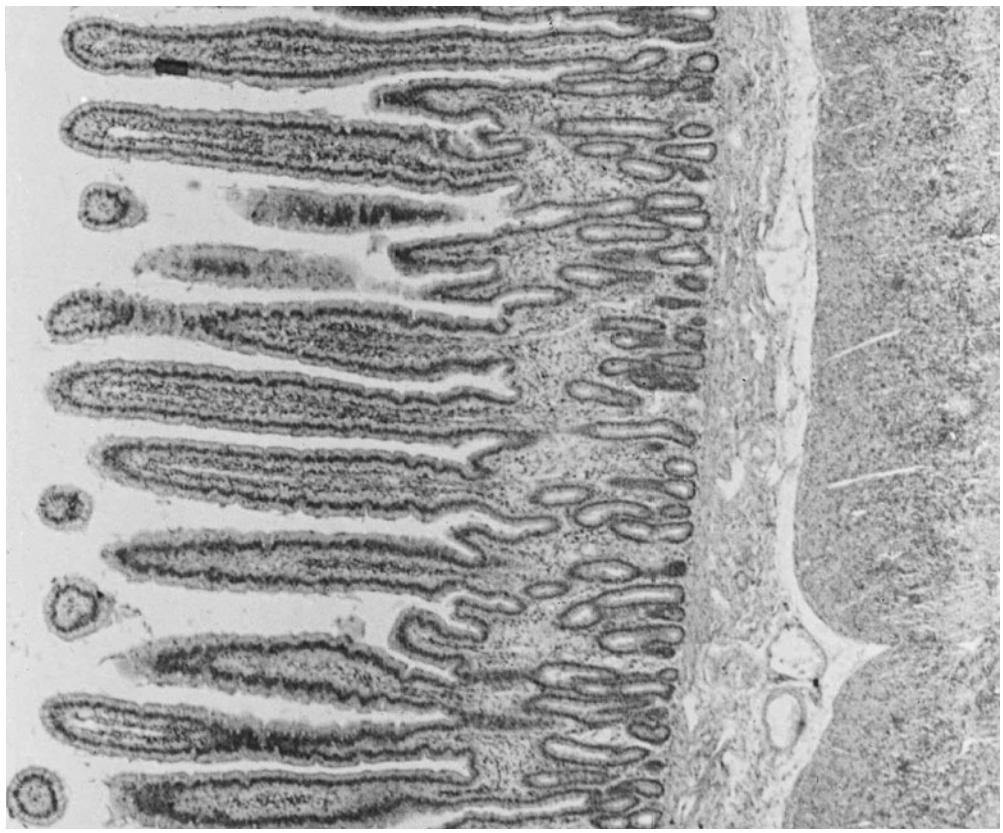
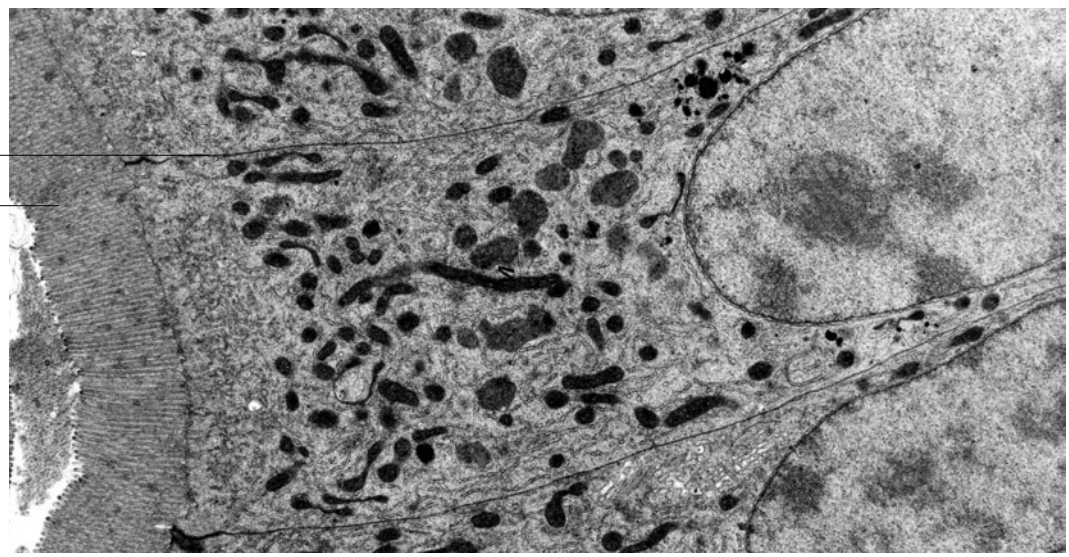


Figure 20.11 TEM of an epithelium cell of a villus (×8000)



What other organelles do you recognise?

Completion of digestion in the small intestine

The specific digestive processes, begun in the duodenum by the enzymes of the pancreatic juice, continue in the remainder of the small intestine because these enzymes are still active in the chyme as it is slowly moved along.

However, here in the small intestine there is an additional source of enzymes. Proteases and carbohydrases occur on the **outer surface of the epithelial cells**, bound to the plasma membrane of the microvilli. These enzymes catalyse the final stages of hydrolysis of protein and carbohydrate digestion. Thus, a final act of digestion may occur during transport of the products of digestion across the plasma membrane, into the epithelial cells (Figure 20.12 and 20.13).

Incidentally, the epithelial cells of the villi are short-lived. They are constantly replaced as they become dislodged from the basement membrane and mix in with the remains of the chyme. Fresh epithelial cells migrate out of the intestinal glands between the villi. However, the disintegrating cells that have been discarded, continue to contribute to digestion, for the enzymes that were bound to their plasma membranes remain active among the contents of the small intestine.

The combined effects of all these enzymes complete digestion.

Absorption in the small intestine

Very little uptake occurs in the gut prior to the small intestine. However, drugs such as aspirin and alcohol are absorbed as they enter the gut, and absorption of them is completed in the stomach. Water is taken up by osmosis, and while this may occur anywhere along the gut, much water uptake occurs in the large intestine.

Efficient uptake occurs in the small intestine because it has a huge surface area, due to the vast number of villi, together with their microvilli. The epithelial cells contain mitochondria in abundance, the source of ATP. Transport of many of the nutrients is an active process, for which ATP is required. These features aid the absorption processes as the products of digestion make contact with the epithelial cells of the villi.

How is each component of digestion absorbed?

Of the products of **carbohydrate digestion**, the monosaccharides **glucose** and **galactose** are actively transported into the epithelial cell, combined with sodium ions, by a protein known as a secondary pump.

Fructose enters epithelial cells by facilitated diffusion (page 26).

The remaining disaccharides are hydrolysed as they are absorbed. For example, **sucrose** and **maltose** attach to specific enzymes of the plasma membrane that complete their digestion to monosaccharides and then pump the monosaccharide products into the epithelial cells. The same may be true of **lactose** (milk sugar).

Extension: Lactose

Lactase is present in the small intestine of young mammals during the period their diet consists exclusively of milk. After that stage in development, differences arise. Some people fail to produce much or any lactase; this results in a condition known as **lactose intolerance**. For these people, lactose taken in the diet remains in the faeces in the rectum, causing water to be retained there that would otherwise be reabsorbed into the blood. Also, some of this lactose is subjected to bacterial fermentation resulting in additional gas production. The result is that diarrhoea and flatulence give the lactose-intolerant person discomfort, after consumption of dairy products. If milk remains an item of their diet, they need dietary supplements to aid lactose digestion.

All the products of digestion of sugars pass from epithelial cells to the blood stream by facilitated diffusion. These processes are summarised in Figure 20.12.

Of the products of **protein digestion**, the amino acids are actively transported into the epithelial cells by the action of membrane protein pumps.

However, the contents of the small intestine also include short-chain peptides, such as **dipeptides** and **tripeptides**. These peptides become attached to specific proteases of the plasma membrane that complete their digestion to amino acids and then pump the amino acids into the epithelial cells (Figure 20.13).

Of the products of **lipid digestion**, the **short-chain fatty acids** (and glycerol) are absorbed by simple diffusion into the epithelial cells. Then, from the interior of the epithelial cells, the short-chain fatty acids diffuse out into the capillaries, and are transported in the blood stream, just as monosaccharides and amino acids are.

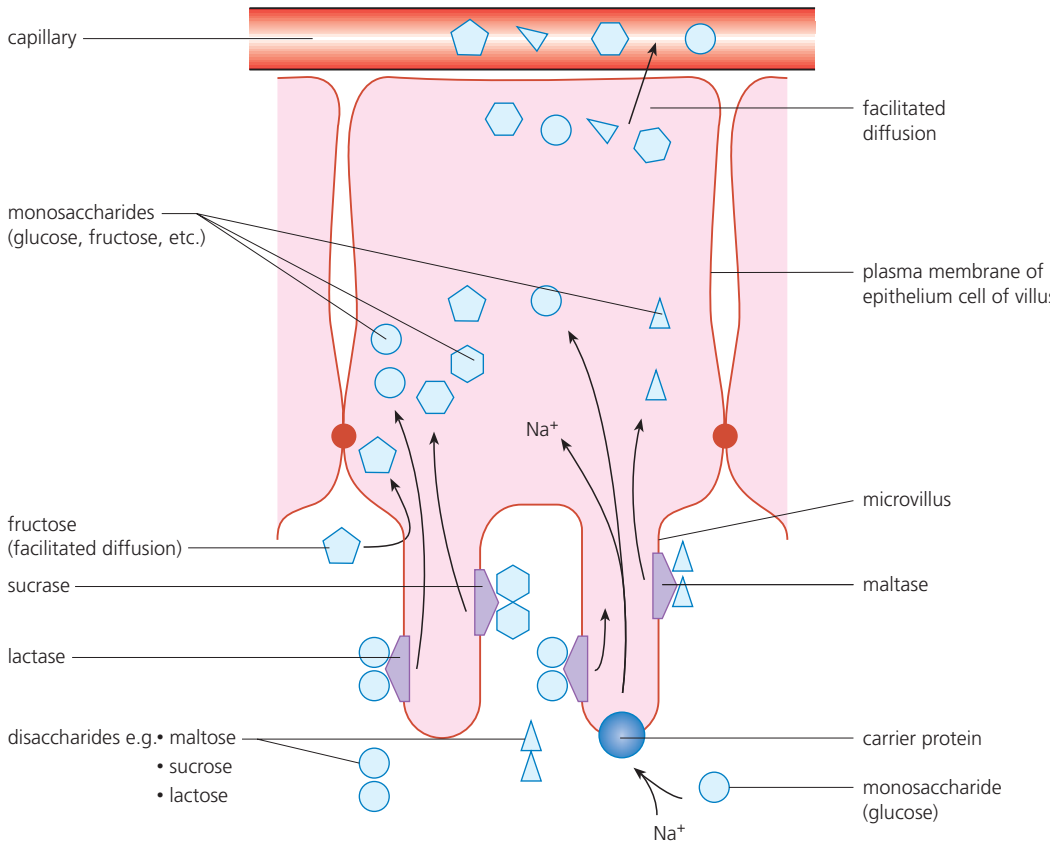


Figure 20.12 The uptake of sugars in the villi

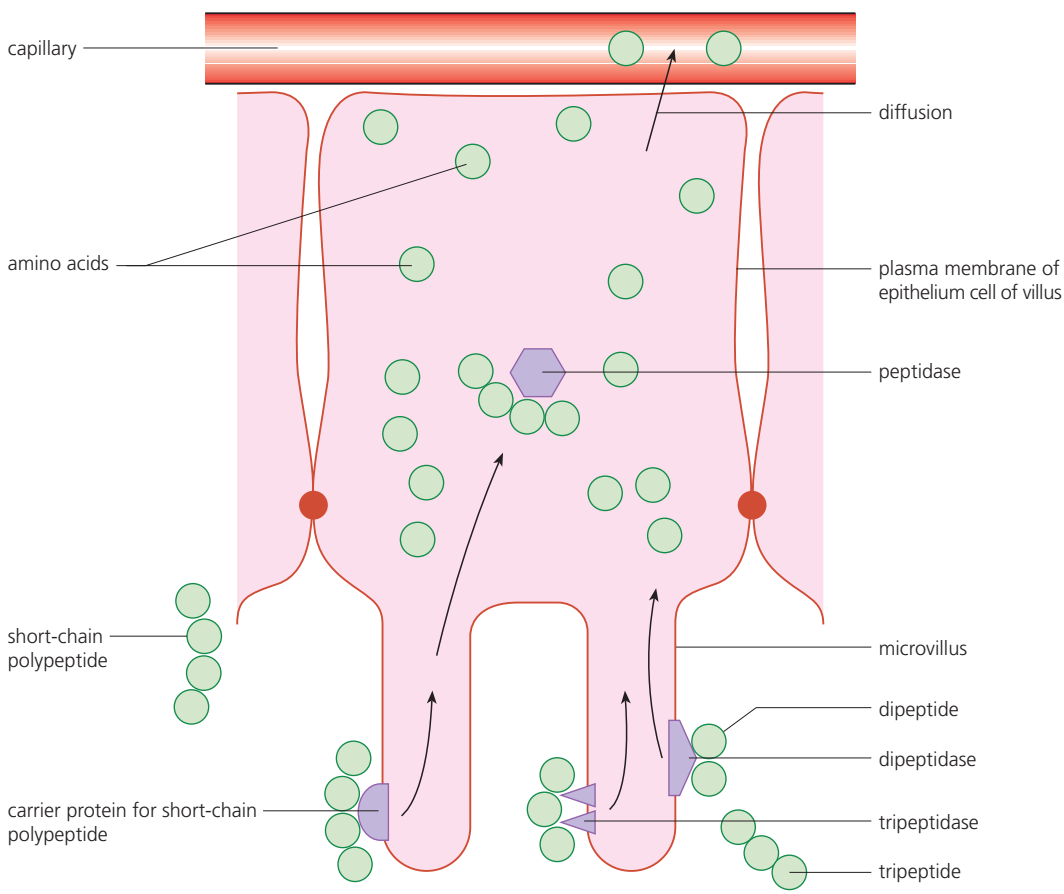


Figure 20.13 The uptake of amino acids and short-chain polypeptides

However, many of the dietary lipids yield **long-chain fatty acids**, when hydrolysed by lipases. The uptake of the long-chain fatty acids across the plasma membrane, also by diffusion, is aided by the bile salts. Subsequently, the bile salts are transported back to the liver in the blood circulation and are re-used in the production of fresh bile. At the same time, the long-chain fatty acids are combined with glycerol to re-form triglycerides.

These triglycerides are then coated with protein, forming spherical masses called **chylomicrons**. The site of these events is the smooth endoplasmic reticulum (sER) of the epithelial cells. Then, chylomicrons exit the epithelial cells by reverse pinocytosis (exocytosis) and enter the lymph in the lacteal vessels in the villi (Figure 20.14). Lacteals are part of the lymphatic system. Lymph drains into the blood circulation via veins at a point close to the heart.

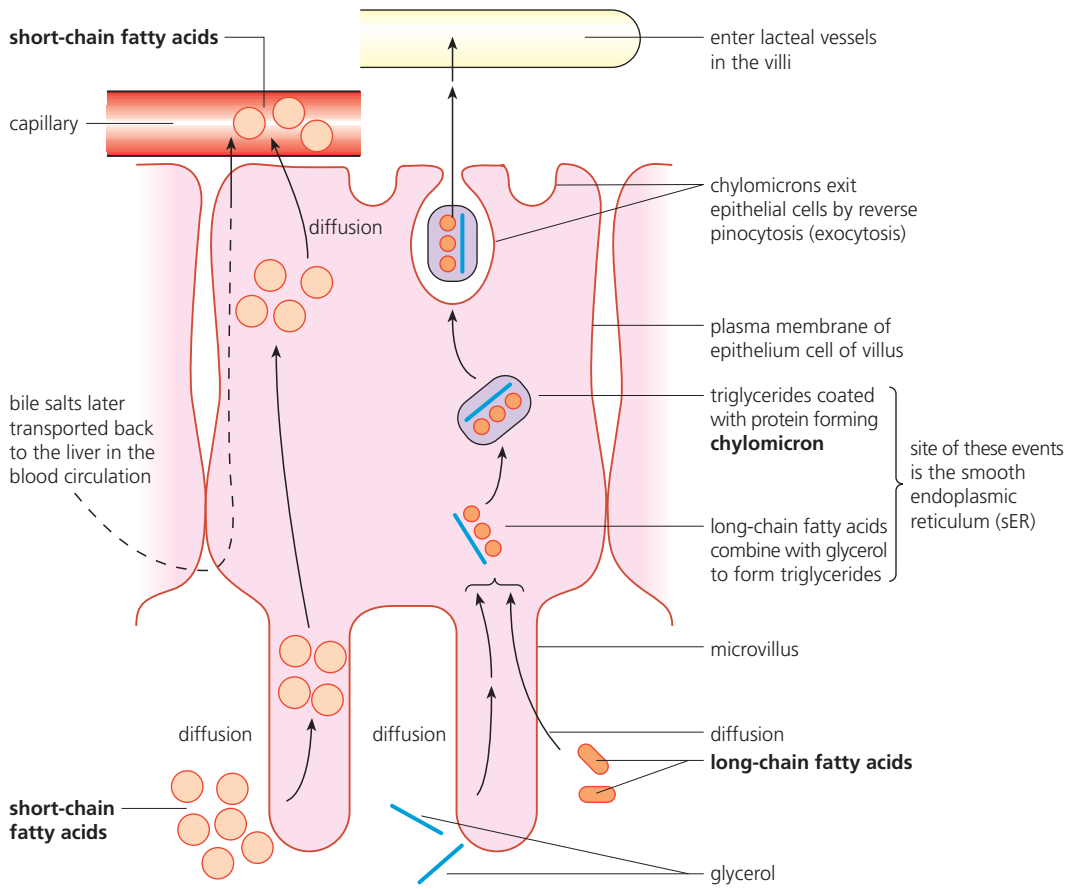


Figure 20.14 The uptake of the products of lipid digestion

4 **Explain** the practical effects of the presence of villi in the intestine.

The nature and fate of undigested matter

The remaining matter that was eaten but has not been digested, passes on into the large intestine by waves of peristalsis (page 181). By this stage of the digestion process, most of the useful products of digestion have been absorbed. Of what remains, the mineral ions and much of the water are yet to be absorbed in the large intestine. Remember, in addition to water present as a component of our diet, many litres of water have been secreted onto the chyme in the form of digestive juices. Much of this water is conserved by the body before the faeces are formed.

The materials that are not absorbed and are egested are cellulose and lignin from plant matter, the remains of intestinal epithelial cells, bile pigments, and bacteria. Bacteria flourish in the conditions in the large intestine. Mostly these microorganisms neither help nor harm us, and this sort of relationship is described as **commensalism**.

5 **Identify** the components of faeces that are waste products of the metabolism of cells of the gut.

What remains is now referred to as faeces. Bacteria compose about 50% of the faeces. Bile pigments (excretory products formed from the routine breakdown of red cells) which were added in the duodenum, colour the faeces uniformly. The rectum, the final region of the gut, is a short muscular tube which terminates at the anus. Discharge of faeces from the body at the anus is controlled by sphincter muscles.

Digestion of cellulose

Cellulose is the most abundant organic compound in the biosphere; it makes up more than 50% of all the organic carbon. It exists in green plant cell walls. Humans, along with all other mammals do not have cellulase enzymes and so are not able to digest cellulose. Instead, it makes up the roughage or dietary fibre that is an essential of our diet. The value of dietary fibre is as bulk that stimulates movement of the chyme through the gut. Ultimately, it becomes merely a component of the undigested matter our bodies dispose of.

6 **Suggest** why the relationship between gut microorganisms found in the rumen of certain herbivorous mammals and their host is described as mutualistic.

Herbivorous mammals are also unable to digest plant cell wall materials. However, many bacteria (and many fungi and protozoa) do produce cellulose enzymes and are able to hydrolyse cellulose to glucose or to organic acids. The herbivorous mammals exploit this facility of bacteria to digest cellulose. Ruminant mammals (e.g. cows and sheep) have a large fermentation chamber called the **rumen** as part of their gut. Here, microorganisms are housed and cellulose is hydrolysed. In non-ruminant mammals (e.g. rabbits and horses) the cellulose-digesting bacteria are housed in the caecum and appendix, at the end of the large intestine.

An unwelcome stomach visitor – *Helicobacter pylori*

Because of the strong acidity of the stomach contents during digestion, it was assumed that the interior of this part of the gut is largely sterile. However, a Gram-negative, spiral-shaped bacterium, *Helicobacter pylori*, is able to survive there. It was discovered first in certain patients, associated with their stomach ulcers.

Helicobacter pylori does not invade cells, but survives by attaching to receptors on the plasma membrane of the cells of the stomach mucosa, below the mucus lining. Here it is largely protected by the mucus layer that lines the inner surface throughout the entire gut. (In the stomach the mucus plays perhaps its most important role by shielding cells against the strong acid in gastric juice.) Actually, any hydrogen ions that penetrate to the bacterium are neutralised by hydrogencarbonate ions and ammonium ions the bacterium produces by the action of the enzyme urease on urea – apparently an adaptation to the bacterium's abnormal habitat.
In the persistent presence of this bacterium on the exterior of cells of the stomach, the body's immune system becomes sensitised. Antibodies are produced in the vicinity of the infection, and killer cells of the immune system accumulate there, too. Since these agents of the body's defences cannot reach the invading cells on the exterior of the plasma membrane they are ineffective.

Nevertheless, an inflammation reaction occurs at the site of infection – a condition known as **gastritis**. A further outcome may be the progressive failure of goblet cells in the infected area. If this occurs, cells of the stomach lining become exposed to the protease and hydrochloric acid of the gastric juice. The result may be an unpleasant stomach ulcer (**gastric ulcer**).

Stomach cancer may also be associated with *Helicobacter pylori* infection. The inflamed cells of the mucosa are more likely to undergo abnormal growth leading to malignancy. More than 65% of Japanese people above the age of 50 are infected with this stomach bacterium, and Japan has the highest rate of **stomach cancer**. Of course, the presence of the bacterium does not imply that cancer will result. The risk of cancer depends on when the infection was (unknowingly) contracted – if it was in the patient's childhood, then stomach cancer later in life is more likely. Clearly, gastritis and stomach ulcers are infections that require prompt treatment.

Functions of the liver in health and disease

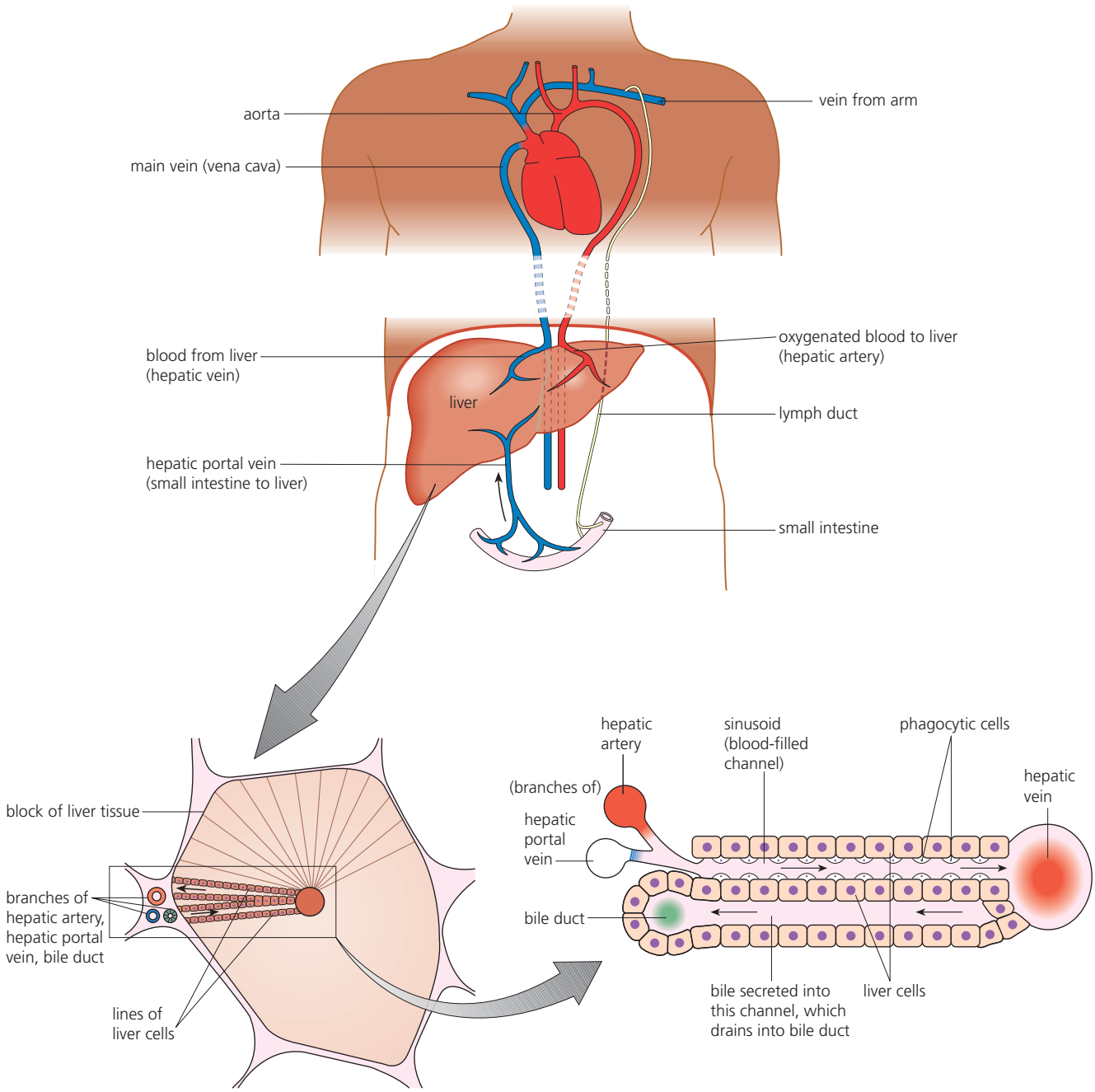
H4.1–4.7

The liver is a large, lobed organ located immediately below the diaphragm, and above and around the stomach. It is served by the **hepatic artery** which delivers oxygenated blood, and it is drained by the **hepatic vein**. In addition, there is a portal vein, the **hepatic portal vein** that brings blood to the liver directly from the small intestine (Figure 20.15).

Within the liver, arterial blood mixes with that from the hepatic portal vein as it flows through blood-filled channels known as **sinusoids**, past lines of liver cells, and then on to a branch of the hepatic vein. The sinusoids differ from the capillaries that we see in most other organs, in that the former are without walls separating blood from the liver cells. In the sinusoids, blood and liver cells are in direct contact. However, lining the sinusoids are many phagocytic cells – an issue we will return to later.

Between two rows of liver cells is another channel, a bile channel, which is entirely isolated from the blood supply. Into these channels, bile from the liver cells flows. The bile channels merge into **bile ducts**, carrying the bile to the gall bladder.

Figure 20.15 Liver cells and their blood supply



7 **Identify** what liver cells require from the blood arriving in the hepatic artery that they are unlikely to receive from the blood in the hepatic portal vein, and explain why.

The roles of the liver

In the regulation of nutrient levels in the blood

Glucose, amino acids and short-chain fatty acids first reach the liver via the hepatic portal vein, after absorption in the villi. Here their concentrations in the plasma are regulated. If this were not so, then the composition of the blood might vary wildly (and dangerously, perhaps) at different times of the day.

The normal level of **blood glucose** in humans is about 90 mg per 100 cm^3 (90 mg 100 cm^{-3}), but the concentration actually varies between 70 mg (when the body has been without food for a prolonged period) and up to about 150 mg (in the hepatic portal vein after a carbohydrate-rich meal has been digested and is being absorbed). On arrival in the liver sinusoids, excess glucose is withdrawn from the plasma solution and used in metabolism or stored as glycogen (animal starch, page 47). Glycogen reserves are also stored elsewhere in the body, particularly in the skeletal muscles.

Respiring tissues of the body receive glucose supplies from the blood circulation. For most tissues, it is a principal substrate for respiration. Glucose is one of the substrates of skeletal muscles (along with fatty acids), but in the brain, glucose is the *only* fuel molecule absorbed (neurones are unable to store glycogen and they cannot respire lipids). As the level of blood glucose falls due to respiration in tissues, glycogen reserves in the liver are converted back to glucose to maintain the normal plasma concentration (Figure 20.16). Remember, the hormones insulin and glucagon that control blood glucose levels (page 223).

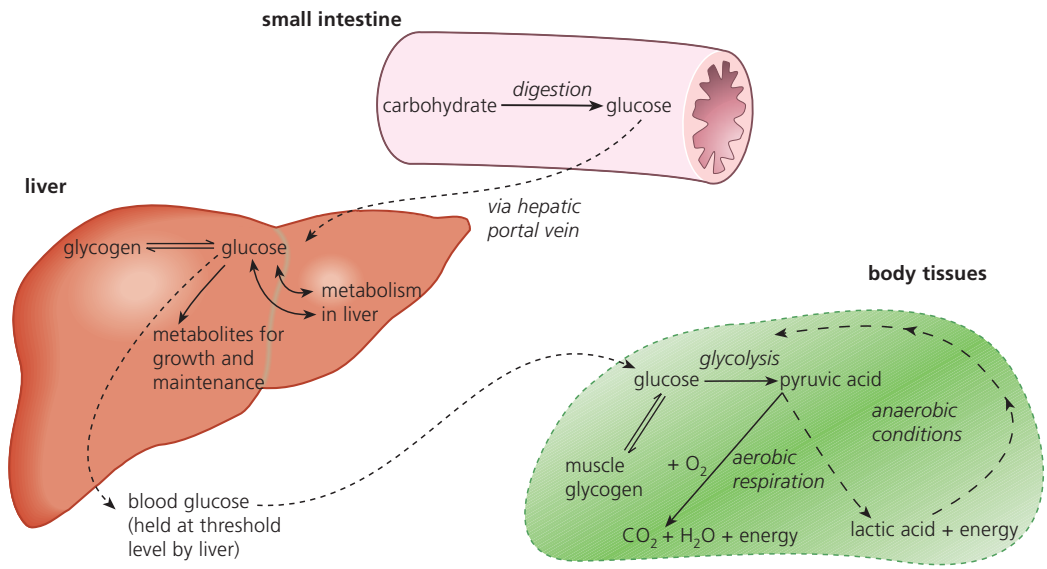


Figure 20.16 The role of the liver in the supply of glucose for respiration

The level of **amino acids** is also adjusted by the liver cells as the blood passes along the liver sinusoids. A pool of amino acids is maintained in the plasma, in the liver and in other tissues undergoing rapid protein synthesis. Amino acids are constantly being built up into proteins, which then function as enzymes, components of membranes, and structural components (e.g. collagen fibres, keratin). Remember, at least 2 million red cells are formed every day, and there are several other body tissues where many new cells are formed daily. The demand for new proteins on a daily basis is very high indeed.

Most proteins are short-lived, and are broken down and contribute to the amino acid pool again, from which the new proteins are made. However, the body cannot store amino acids, should there be an excess. Instead, excess amino acids are deaminated in the liver. The organic acid part of each amino acid is removed and respired, or converted to fat or carbohydrate. The —NH_2 (amine) group(s) of each amino acid are converted to ammonia and combined with carbon dioxide to form urea:

$$2NH_3 + CO_2 \rightarrow (NH_2)_2C{=}O + H_2O$$

By this deamination process, the liver ensures that soluble ammonia is not formed and released in the tissues. Urea is removed from the blood in the kidneys.

The **fatty acids** (and glycerol) that reach the liver are combined to form triglycerides. These are combined with proteins in the liver, and may be stored there. Alternatively they are transported in the blood plasma, mostly as low-density lipoproteins (LDLs), to the tissues. Here lipids may be stored as food reserves (fat), or immediately broken down and respired as a source of energy. Heart muscle preferentially respires fatty acids, and skeletal muscle tissue readily respires fatty acids.

8 Explain why the body makes heavy and constant demands on its pool of amino acids.

As a site of syntheses

The liver is the site of synthesis of all the blood proteins, including globulins, albumin, prothrombin and fibrinogen. Also, most of the cholesterol required by the body on a daily basis is manufactured in the liver (but the remainder is taken in as part of the diet).

In detoxification

The liver detoxifies harmful substances such as alcohol (see below), or renders drugs and toxins that have entered the blood stream into harmless forms for excretion from the blood circulation in the kidneys. Drugs such as the antibiotics penicillin and erythromycin are handled in this way, as are sulphonamides. Hormones such as thyroid hormone, and steroid hormones such as oestrogen, testosterone, and aldosterone are similarly inactivated, ready for removal from the blood.

In the breakdown of red cells

Another important task of the liver cells is the breakdown of the redundant red cells. Red cells loose their nucleus on formation, and remain functional in the blood circulation for only a limited time before they are broken down and replaced.

First, the worn-out cells are taken up by **phagocytosis** by macrophage cells, such as those along the walls of the sinusoids. Then the globin (protein) and haem (non-protein part containing iron ions) of haemoglobin are split apart, and the **globin is hydrolysed** by proteases to amino acids. From the **breakdown of the haem**, iron in the form of iron III ion (Fe^{3+}) is removed and attached to a carrier protein. In this form, iron is stored in the liver cells and, sooner or later, exported via the blood plasma to the bone marrow where the new red cells are formed. The remainder of the haem part is converted into **bile pigments** in the liver cells, and these are then secreted into the bile channels, between the sinusoids, as a component of the bile formed by the liver.

9 Draw and label a phagocytic cell in operation.

Thus, most of the breakdown products are held in the liver for re-use, or are further metabolised there. The role of the liver in haemoglobin breakdown and formation of bile is summarised in Figure 20.17.

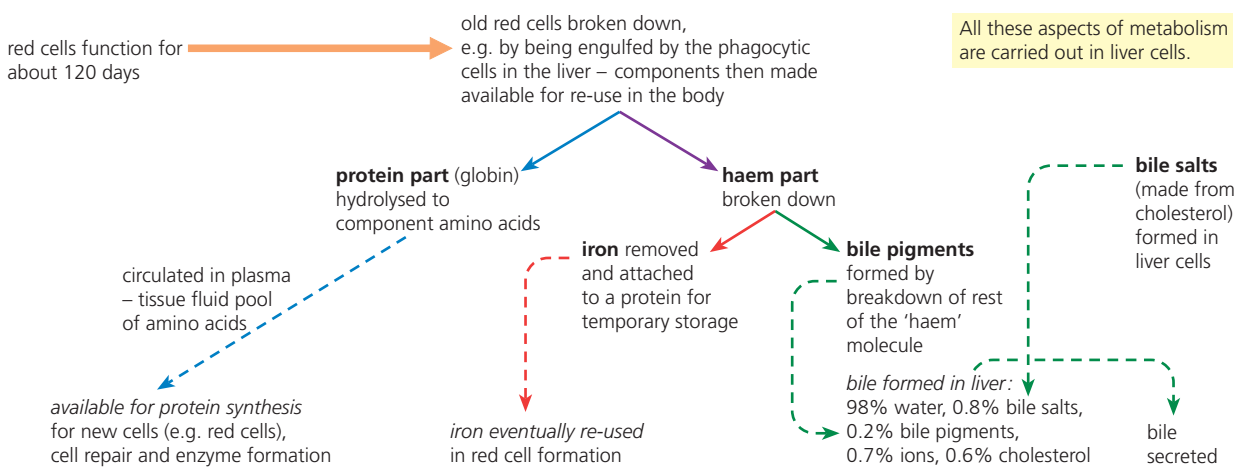


Figure 20.17 Red cell breakdown and the formation of bile

10 **Define** the meaning of the statement that most reactions of liver cells are endergonic.

As a storage organ

We have noted already that the liver cells are sites where iron and carbohydrate (as glycogen) are stored. Fats, too, are stored here, along with many other sites in the body. In addition, certain vitamins (A – as retinal, B_{12}, D – as calciferol, E, and K) are stored in the liver. These vitamins, with the exception of B_{12}, are all fat-soluble. They are released from the liver when needed elsewhere in the body.

The liver and alcohol

Alcohol has been a feature in the life of human communities from a very early stage in their development. For example, there are records of beer drinking in ancient Egyptian artefacts, dating from 4000–5000 years ago. In the absence of a clean water supply, a dilute solution of ethanol might be a safer drink than water contaminated with bacteria or gut parasites.

When alcohol is drunk, it is quickly absorbed into the blood, mostly from the stomach. Since alcohol is a more concentrated source of energy than carbohydrate, it must be considered as a food stuff. Indeed, one potentially harmful effect of an excessive intake of alcohol is obesity.

However, alcohol is basically considered to be a drug because of its effects on the central nervous system. When a small amount is consumed, especially when drunk slowly and taken with food, alcohol is a mild stimulant in the brain. Indeed, alcoholic drink, when restricted to one or two units a day, appears to confer a small health advantage regarding coronary heart disease (compared with an alcohol-free diet) for many people. A **unit of alcohol** is 10 cm^3 of alcohol, and is found in half a pint of beer, one glass of wine or fortified wine (port or sherry), or a single measure of spirits (whisky, brandy, rum, vodka, gin).

Binge drinking, which is defined by a high-speed intake of a considerable amount of alcohol, is particularly unhealthy. As more alcohol is consumed, especially if this occurs quickly, the alcohol becomes a sedative, slowing down the activity of the brain. Restraints on behaviour are progressively removed (a feature which initially helps relaxation and eases social relationships), resulting in excesses in conversation, loud laughter, and sometimes aggression. Judgement is affected, making it hard to see straight or judge distance. Muscle co-ordination is reduced, leading to clumsiness, staggering and slurred speech. If alcohol consumption continues, the effect is as an anaesthetic – the drinker-to-excess becomes unconscious. A blood alcohol concentration of about 500 mg per 100 cm^3 is considered a lethal dose.

After intake, alcohol is distributed all over the body (it is able to cross the blood–brain barrier). However, alcohol is progressively metabolised in the liver.

First, alcohol (ethanol) is oxidised to acetaldehyde (ethanal) by the enzyme **alcohol dehydrogenase**. Acetaldehyde is further oxidised to acetic acid (ethanoic acid) by acetaldehyde dehydrogenase. The acetic acid then becomes part of the pool of metabolites that are oxidised to carbon dioxide and water by various biochemical pathways. The long-term and adverse effects of these intermediates, and of the direct effects of alcohol on gut, brain, heart and liver include:

- **cirrhosis** of the liver – a chronic inflammation of the liver in which liver cells are destroyed and replaced by fibrous or adipose (lipid-containing) connective tissue;
- **obesity**;
- **gastritis** – alcohol irritates the stomach mucosa;
- **malnutrition** – excess alcohol affects the appetite, so a balanced diet is not taken;
- **dementia** – for alcohol destroys neurones;
- **weakened heart muscles**.

The huge **social effects** of excessive alcohol consumption can be as much if not more of a problem, and affect the lives of a wider group.

The transport system

H5.1–5.5

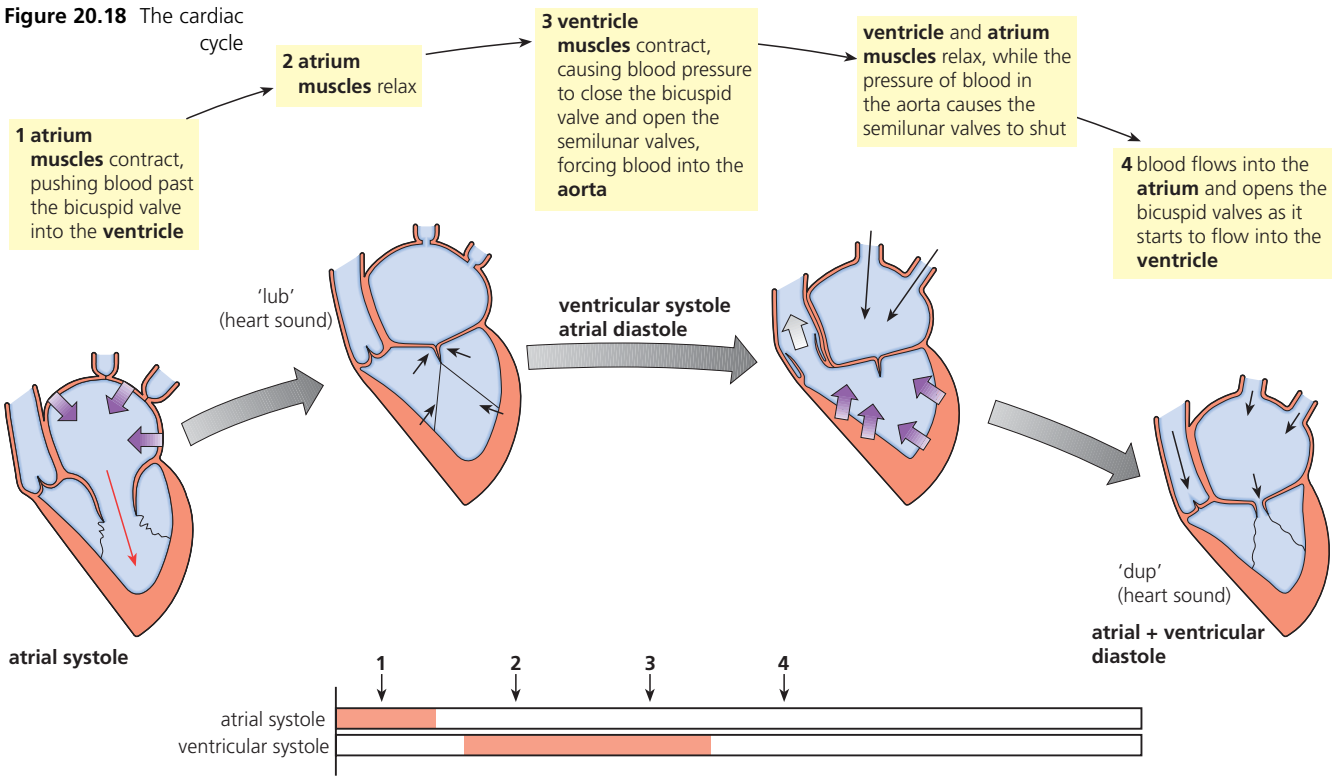
A complex internal transport system has evolved in many animals, especially in those that are large and of compact build. Throughout such organism's lives, materials are constantly required to be moved to and from all parts of their body.

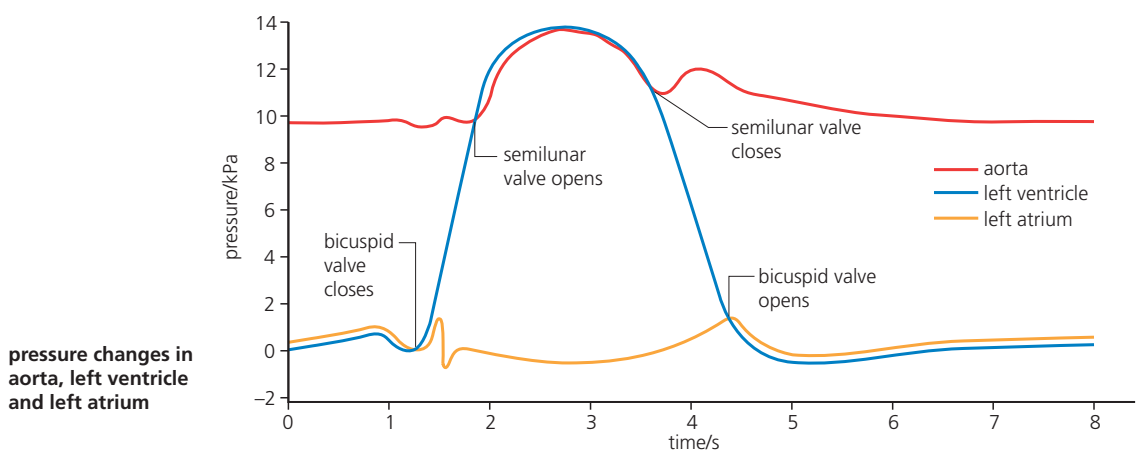
Internal transport systems in multicellular organisms are examples of mass flow. The **mammalian closed system of blood vessels**, by which the blood is circulated to all the tissues of the body, pumped by a powerful heart (page 185), is an example. The blood circulation system delivers oxygen, essential nutrients, water and hormones; it also distributes heat, and removes waste products for disposal. We now examine the way the heart works (the cardiac cycle), and the way heart activity is controlled.

The cardiac cycle

We have seen that the heart is divided into four chambers (Figure 7.12, page 190). When the muscular walls of a chamber contract, the volume of the chamber decreases and the pressure on the blood in the chamber increases. The outcome is that blood is forced to a region where the pressure is lower. Valves of the heart prevent blood flowing backwards to a region of low pressure, so blood always flows on through the heart.

Figure 20.18 The cardiac cycle





pressure changes in aorta, left ventricle and left atrium

The **cardiac cycle** is the name given to the sequence of events of a heart beat. The heart beats at a rate of about 75 times per minute, so each cardiac cycle lasts for about 0.8 seconds. The events of a heart beat are divided into two stages: systole and diastole. In **systole**, heart muscle contracts, and during **diastole**, heart muscle relaxes.

Look at the steps in Figure 20.18. Here the cycle is illustrated on the left side of the heart only, but **both sides function together**, in exactly the same way.

Start with contraction of the atrium (**atrial systole** – lasting about 0.1 sec). As the walls of the atrium contract, any backflow of blood is prevented by the contraction of the atrial wall which also has the effect of sealing off the pulmonary veins. Blood pressure in the atrium rises, and blood is pushed past the atrio-ventricular valve, into the ventricle where the contents are under lower pressure.

The atrium now relaxes (**atrial diastole** – lasting about 0.7 sec).

Next the ventricle contracts (**ventricular systole** – lasting about 0.3 sec) and contraction of the ventricles is very forceful indeed. The high pressure this generates slams shut the atrio-ventricular (bicuspid) valve and opens the semilunar valves. Blood is forced into the aorta. Another pulse has been generated. The pulse is the wave of high pressure passing along the arteries that we can detect at points in the body where an artery passes over a bone, just beneath the skin surface (Figure 20.19).

This contraction is followed by relaxation of the ventricle (**ventricular diastole** – lasting about 0.5 sec). In ventricular diastole, the high pressure developed in the aorta and the falling pressure in the ventricle causes a slight backflow of blood, closing the semilunar valves, preventing any further backflow.

Each contraction of cardiac muscle is followed by relaxation and elastic recoil, due to the presence of connective tissue fibres in the muscle, which are temporarily distorted in the contraction.

The changing pressure of blood in the atria, ventricles, pulmonary artery and aorta (shown in the graph in Figure 20.18) automatically opens and closes the valves, and the closing of valves generates the heart sounds.

Heart sounds

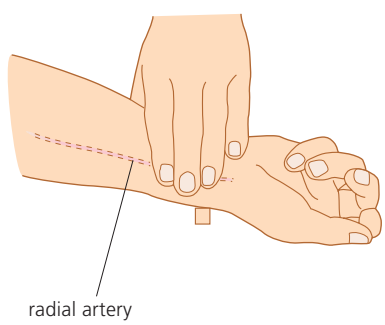
The valves of the heart close passively whenever there is a tendency for blood to flow in the reverse direction, and the sound of their closing can be heard by using a stethoscope:

- the first sound, **lub** (*spoken softly*), is due to the simultaneous closure of the atrio-ventricular valves;
- the second sound, **dup**, is due to closure of the semilunar valves;
- as the heart pumps, the sequence of sounds from a healthy heart is **lub dup** *pause* **lub dup** *pause* and so on.

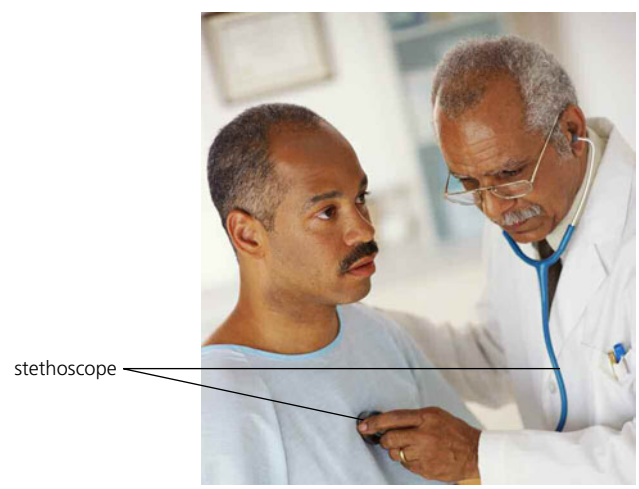
The **value of listening to heart sounds** is that it allows early detection of valve malfunction (Figures 20.18 and 20.19).

Figure 20.19 Measuring pulse (heartbeat rate) and listening to the heart sounds

taking the pulse at the wrist
(note that a finger tip is used, not the thumb)



the heart sounds can be heard at various points on the chest



11 Examine the data on pressure change during the cardiac cycle in Figure 20.19. **Determine** or **suggest** why:

- **a** pressure in the aorta is always significantly higher than that in the atria
- **b** pressure falls most abruptly in the atrium once ventricular systole is under way
- **c** the semilunar valve in the aorta does not open immediately that ventricular systole commences
- **d** when ventricular diastole commences, there is a significant delay before the bicuspid valve opens, despite rising pressure in the atrium
- **e** it is significant that about 50% of the cardiac cycle is given over to diastole.

The control of the heart beat

We have already noted that the heart beat is of **myogenic origin** (page 192).

The heart beat originates in a structure in the muscle of the wall of the right atrium, called the **sino-atrial node (SAN)**, also known as the natural **pacemaker**. Muscle fibres radiating out from the SAN conduct the impulses to the muscles of both atria, triggering atrial systole.

Then a second node, the **atrio-ventricular node**, situated at the base of the right atrium, picks up the excitation and passes it to the ventricles though modified muscle fibres, the **Purkinje fibres** (Figure 20.20). Ventricular systole is then triggered.

After every contraction, cardiac muscle has a period of insensitivity to stimulation, known as a **refractory period** (in effect, a period of enforced non-contraction which we may call a rest), when the heart refills with blood. This period is a relatively long one in heart muscle, and doubtless an important feature, enabling the heart to beat throughout life.

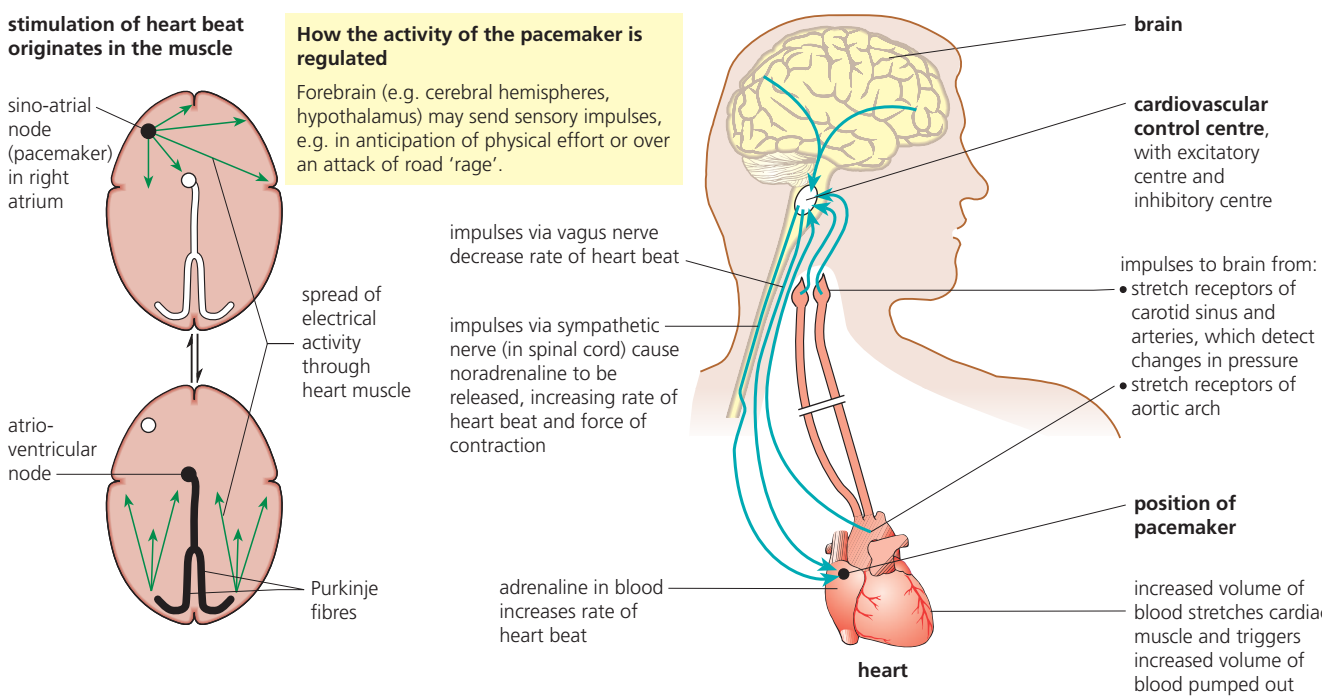
The heart's own rhythm, set by the SAN, is about 50 beats per minute, but conditions in the body can override this basic rate and increase heart performance. The action of the pacemaker is modified according to the needs of the body. For example, it is increased during physical activity, from the 75 beat per minute of the at-rest heart, up to 200 beats a minute in strenuous exercise.

How is the pacemaker regulated?

Look at Figure 20.20 now. Nervous control of the heart is by reflex action. The heart receives impulses from the cardiovascular centre in the medulla of the hindbrain, via two nerves:

- a **sympathetic nerve**, part of the sympathetic nervous system, which speeds up the heart;
- a **branch of the vagus nerve**, part of the parasympathetic nervous system, which slows down the heart.

Figure 20.20 Control of heart rate



The sympathetic and parasympathetic nervous systems are part of the **autonomic nervous system**, not under conscious control (Figure 17.25, page 538).

Since the sympathetic nerve and the vagus nerve have opposite effects in this matter of regulation of heart beat, we say they are **antagonistic**.

The cardiovascular centre also has nerves supplying it. For example, it receives impulses from **stretch receptors** located in the walls of the aorta, in the carotid arteries, and in the wall of the right atrium, when changes in blood pressure occur at these positions.

When blood pressure is high in the arteries, the rate of heart beat is lowered by impulses from the cardiovascular centre via the vagus nerve.

When blood pressure is low, the rate of heart beat is increased.

The rate of heart beat is also influenced by impulses from the higher centres of the brain. For example, emotion, stress and anticipation of events can all cause impulses from the sympathetic nerve to speed up heart rate.

In addition, the hormone **adrenaline**, which is secreted by the adrenal glands and carried in the blood, causes the pacemaker to increase the heart rate.

12 During vigorous activity, the heart beats more quickly. **Outline** the sequence of events that causes this raised heart rate.

Coronary heart disease (CHD)

Diseases of the blood vessels are the cause of more premature death in the developed world than any other single cause. Most of these are due to **atherosclerosis** – the progressive degeneration of the artery walls. The structure of an artery wall was introduced in Figure 7.9 (page 188).

Healthy arteries have pale, smooth linings, but in arteries that have become unhealthy, the walls have strands of yellow fat deposited under the endothelium. This fat builds up from certain of the lipoproteins and from cholesterol that may be circulating in the blood. Fibrous tissue is laid down with the fatty streaks. These deposits start to impede blood flow, and contribute to raised blood pressure. Progressive reduction of the blood flow to the heart muscle impairs oxygenation to this organ leading to chest pains, known as **angina**, usually brought on by physical exertions.

How can degeneration of artery walls lead to fatal heart disease?

Where the smooth lining actually breaks down, the blood is exposed to the fatty, fibrous deposits. Blood platelets collect at the exposed roughened surface, releasing factors that trigger blood clotting (page 349). A blood clot within the vessel may form (Figure 13.22, page 420). It is known as a **thrombus**, at least until it breaks free and is circulated in the blood stream, where upon it is called an **embolus**.

An embolus may be swept into a small artery or arteriole which is narrower than the diameter of the clot, causing a blockage. Immediately, the blood supply to the tissue downstream of the block is cut off and the tissue is deprived of oxygen. Without oxygen, the tissue dies.

The arteries supplying the heart are the coronary arteries. These arteries are especially vulnerable, particularly those to the left ventricle. When heart muscle dies in this way, the heart may cease to be an effective pump. We say a heart attack has occurred (known as a **myocardial infarction**). Coronary arteries that have been damaged can be surgically by-passed (Figure 20.21).

When an embolus blocks an artery in the brain, a **stroke** occurs. Neurones of the brain depend on a continuous supply of blood for oxygen and glucose. Within a few minutes of the blood supply being lost, the neurones affected will die. Neurones cannot be replaced, so the result of the blockage is a loss of some body functions controlled by that region of the brain.

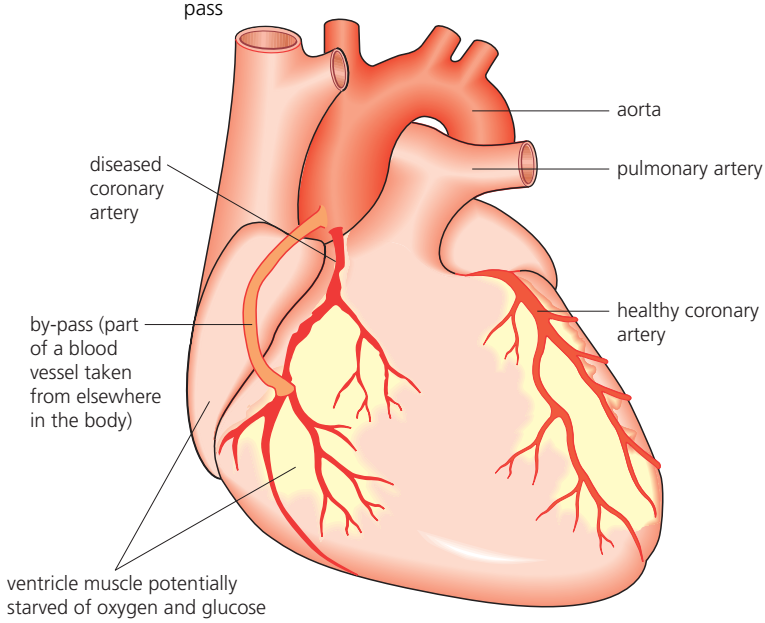


Figure 20.21 A heart by-pass

In arteries where the wall has been weakened by atherosclerosis, the remaining layers may be stretched and bulge under the pressure of the blood pulses. Ballooning of the wall like this is called an aneurysm. An **aneurysm** may burst at any time.

Factors affecting the incidence of CHD

In the developed world, cardiovascular disease tops the list of most serious health problems. In many countries, CHD is the primary cause of death, and many of these deaths are classified as premature, given that life expectancy is otherwise high and rising. Another feature of this type of disease is that the condition may go largely unnoticed for many years – often until it is too late.

What factors contribute to the high incidence of CHD?

1 Age

The risk of coronary atherosclerosis increases with age, but the evidence is that the genesis of the condition develops early in life. Post-mortem studies of soldiers killed in action have disclosed early but well-developed fatty plaques in the arteries of young males of average age 22 years, suggesting that these commenced development in their bodies during their adolescence. Consequently, for individuals brought up in the developed world, after the ages of 35 in men and 45 in women, the chances of dying from CHD increase dramatically.

2 Genetic factors

Heart attacks that occur at a *relatively early age* frequently run in families, suggesting that there are genes that may confer vulnerability to this disease. This aspect will become clear as knowledge of the human genome is developed to identify the role of individual genes and their effects on metabolism.

One genetic factor is beyond dispute. **Gender**, in the form of possession of a Y chromosome (being a **male**), predisposes to greater risk of CHD than females carry.

3 Hypertension

Persistently high blood pressure is defined as systolic pressure greater than 140 mmHg and diastolic pressure greater than 90 mmHg (remember, a blood pressure of 120/80 mmHg is normal in an adult).

Hypertension is known as a silent killer because of the damage it does to the heart, blood vessels, brain, and kidneys without causing noticeable discomfort. It accelerates onset of atherosclerosis, increases the workload of the heart, and makes a brain haemorrhage more likely. Once detected, hypertension can be successfully treated using drugs, including the following types:

a diuretics (reduction of blood pressure by decreasing blood volume through increased elimination of water);

b ACE inhibitors (reduction of blood pressure by vasodilation);

c β blockers (reduction of blood pressure by reducing heart rate).

4 Smoking

After hypertension, the habit of cigarette smoking generates the greatest risk of fatal ill-health, and especially from cardiovascular diseases. There have been huge investments in the tobacco industry via the growth of the crop and the manufacturing of tobacco products, for a long time. The enthusiastic endorsement of cigarette smoking in the developed world (manifested first by the availability of cigarettes to the troops in two world wars) has now spread to developing countries as fresh markets for tobacco products have opened up. Consequently, the dangers of smoking have not had the attention they deserve, and some people have even suggested the evidence against tobacco is equivocal.

Yet from the earliest statistical studies there has been no room for doubts.

Dr Richard Doll and colleagues, working at St Thomas's Hospital, London from 1947, investigated the cause of death of a sample of 3500 people admitted to hospital for treatment. Whatever the symptoms these people had, the vast majority of those who were **smokers later died** from a cardiovascular disease.

This study was followed up by one of a group of 40 000 healthy, working doctors (among whom the habit of smoking was widespread at that time) which was more conclusive still. Some of Richard Doll's data from his studies with doctors is shown in Table 20.2. Few doctors smoke today.

Cause of death	Non-smokers	Continuing smokers
heart disease	606	2067
stroke	245	802
aneurysm	14	136
atherosclerosis	23	111
Total	888	3116

Table 20.2 Mortality from cardiovascular disease caused by cigarette smoking

5 High fat or cholesterol diets, leading to blood lipid abnormalities

An abnormal blood lipid level is a critical factor in the cause of atherosclerosis. Today, statins are very effective drugs used to regulate blood cholesterol. These inhibit the natural synthesis of cholesterol by the liver. Combined with a diet that controls and limits the intake of saturated fats, inappropriate lipid levels can be avoided.

Lack of exercise and obesity – being over-weight to the point of being obese (having a BMI in excess of 30, page 412) – greatly increase the stress on heart and arteries. Regular, appropriately strenuous exercise not only helps to combat obesity, it also improves heart performance.

TOK Link

Media stories on obesity are extremely common in developed countries. Here the prevalence of obesity has roughly doubled since 1980 among adults and tripled among children. But although numbers of deaths caused by diabetes have risen, the predicted increases in mortality from heart disease and strokes have not materialised. Is this evidence of oversimplification of research results in order to reinforce (popular) public prejudices and superstitions?

Gaseous exchange

H6.1–6.7

Living things require energy for the activities of life. Energy is transferred in cells by the breakdown and oxidation of sugars and other substances, in the process of cellular respiration. As a consequence of respiration, gases are exchanged between respiring cells and their environment, in a process known as **gaseous exchange**.

We have seen that the blood circulation is the transport system of the body, most importantly circulating nutrients to all the cells. It also serves the needs of cell respiration directly by delivering oxygen and removing carbon dioxide (as well as other waste products of metabolism).

It is the issues of **transport of respiratory gases** and **gaseous exchange** that we focus on here.

Transport of oxygen

Earth's atmosphere is a mixture of gases, as we already appreciate. The composition of the air on a dry day is:

- nitrogen 78.6%
- oxygen 20.9%
- carbon dioxide 0.04%
- other gases 0.06%
- water vapour 0.4%.

Nitrogen and oxygen make up by far the greatest part; so much so, we often round up or approximate the concentration of these gases to: nitrogen = 79%; oxygen = 21%.

On these figures, the amount of carbon dioxide is exceedingly small. Nevertheless, carbon dioxide is a highly significant component of our atmosphere for several reasons. We return to the issues of carbon dioxide later.

As terrestrial animals living at or near sea level, we live at the bottom of a sea of air. The atmosphere that surrounds us exerts a significant pressure. The air we breathe is a mixture of gases, and in a gas mixture, each gas exerts a pressure.

In fact, the pressure of a mixture of gases is the sum of the pressure of the component gases. Consequently, the pressure of a specific gas in a mixture of gases is called its **partial pressure**. The symbol for partial pressure is *P*, and the partial pressure for a gas x is *p*X. So, for example, pO_2 denotes the partial pressure of oxygen.

Partial pressure is the fraction of the total gas pressure that is exerted by a particular constituent gas.

The unit of pressure is the **pascal (Pa)** and its multiple the **kilopascal (kPa)**. At one time pressure was quoted in 'atmospheres' and the alternative units were mmHg (millimetres of mercury). In fact, medical textbooks still use mmHg (and a doctor or nurse will quote your blood pressure as 120/80 mmHg. In order to convert from these (earlier) units we need to know that:

1 atmosphere = 101.3 kPa
1 mmHg = 133.3 Pa.

What is the partial pressure of the oxygen in the air around us?

At sea level, the atmospheric pressure is typically about 101.3 kPa. So, the partial pressure of oxygen is given by:

$$\frac{101.3 \times 20.9}{100} = 21.2 \text{ kPa}$$

The significance of this figure lies in the properties of the respiratory pigment haemoglobin that transports oxygen in the blood. We now need to look into this.

Role of haemoglobin

Haemoglobin occurs in the red cells. Each red cell contains about 280 million molecules of haemoglobin.

The haemoglobin molecule is built of four interlocking subunits (Figure 20.22). These subunits are composed of a large **globular protein** with a **non-protein haem group** attached, containing iron.

One molecule of oxygen will combine with each haem group, at the concentration of oxygen that occurs in our lungs. This means each haemoglobin molecule is able to transport four molecules of oxygen:

haemoglobin + oxygen → oxyhaemoglobin

$$Hb + 4O_2 \rightarrow HbO_8$$

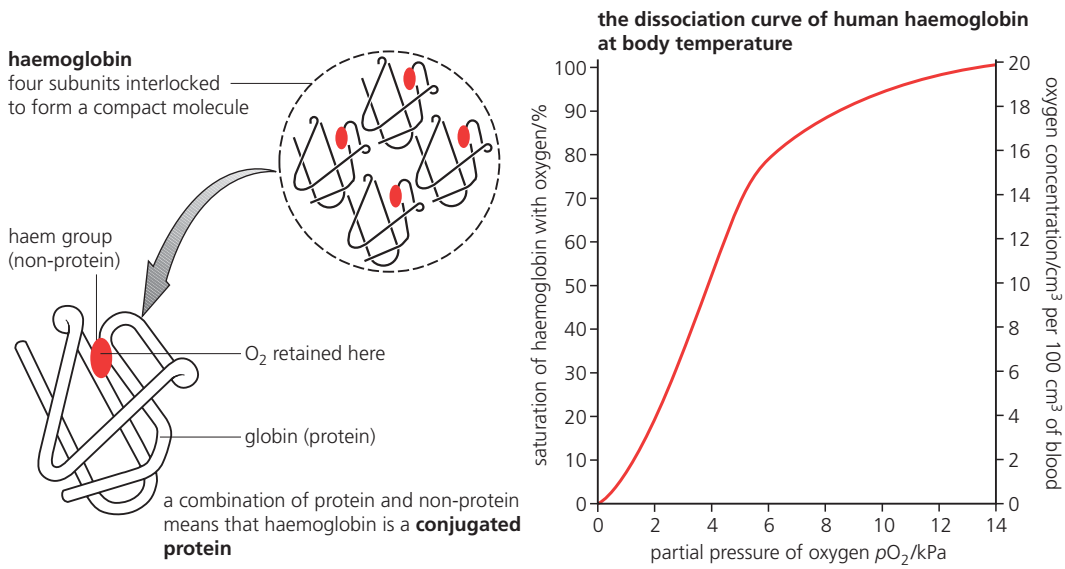


Figure 20.22 The structure of haemoglobin and its affinity for oxygen

The **affinity of haemoglobin for oxygen** is measured experimentally by finding the percentage saturation with oxygen of blood exposed to air mixtures containing different partial pressures of oxygen. The result is called an oxygen dissociation curve (Figure 20.22).

Notice that the oxygen dissociation curve is S-shaped. This tells us that in the complex haemoglobin molecule, the first oxygen molecule attaches with difficulty, but once it has, the second combines more easily, and so on until all four are attached and the molecule is saturated. In other words, the amount of oxygen held by haemoglobin depends on the partial pressure of oxygen.

What is the significance of the oxygen dissociation curve in the working body?

In the body, too, the amount of oxygen held by haemoglobin depends on the partial pressure. In the lungs, air is saturated with water vapour, and so the partial pressure of the component gases is different from that outside, in dry air (Table 20.3).

Table 20.3 Partial pressures of the components of air in the alveoli

Component gases	% composition	partial pressure/kPa
nitrogen	75.5	76.4
oxygen	13.1	13.3
carbon dioxide	5.2	5.3
water vapour	6.2	6.3

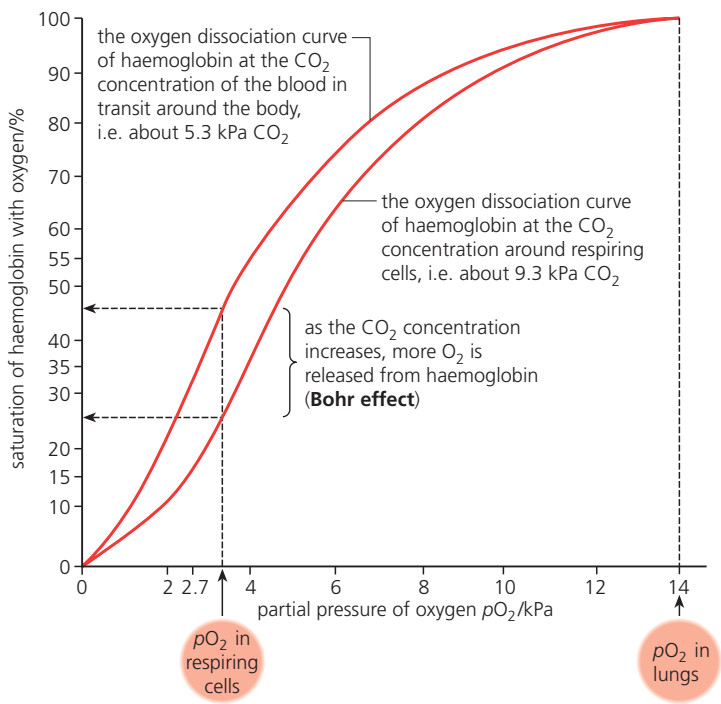
From the oxygen dissociation curve, we can see that the haemoglobin in red cells in the capillaries around the alveoli in the lungs will be about 95% saturated.

However, in respiring tissues the **oxygen partial pressure is much lower**, due to aerobic respiration there. In fact, the oxygen partial pressure in actively respiring tissues may be 0.0–4.0 kPa. At these partial pressures, oxyhaemoglobin breaks down, releasing oxygen in solution and this rapidly diffuses into the surrounding tissues. Clearly the chemistry of haemoglobin makes it an efficient vehicle for oxygen transport.

The effect of carbon dioxide on oxygen transport – the Bohr shift

The blood circulation also transports carbon dioxide from respiring tissues (where it is at relatively high partial pressures) to the lungs. The mechanism of carbon dioxide transport will be discussed shortly. For the moment, we need to see what effect the presence of this carbon dioxide has upon oxygen transport itself.

Figure 20.23 How carbon dioxide favours release of oxygen in respiring tissues



In respiring cells, the concentration of carbon dioxide is approximately 9.3 kPa, whereas in the lungs, as we see in Table 20.3, it is 5.3 kPa.

The effects of these partial pressures of carbon dioxide on the oxygen dissociation curve of haemoglobin is marked (Figure 20.23).

Look at the curves in Figure 20.23, now.

An increased carbon dioxide concentration shifts the oxygen dissociation curve to the right. That is, where the carbon dioxide concentration is high (obviously in the actively respiring cells), oxygen is released from oxyhaemoglobin even more readily. This very useful outcome for living tissues is known as the Bohr effect.

13 **Deduce** the change in percentage saturation of haemoglobin if the oxygen partial pressure drops from 4 kPa to 2.7 kPa when the partial pressure of CO_2 is 5.3 kPa (Figure 20.23).

Myoglobin

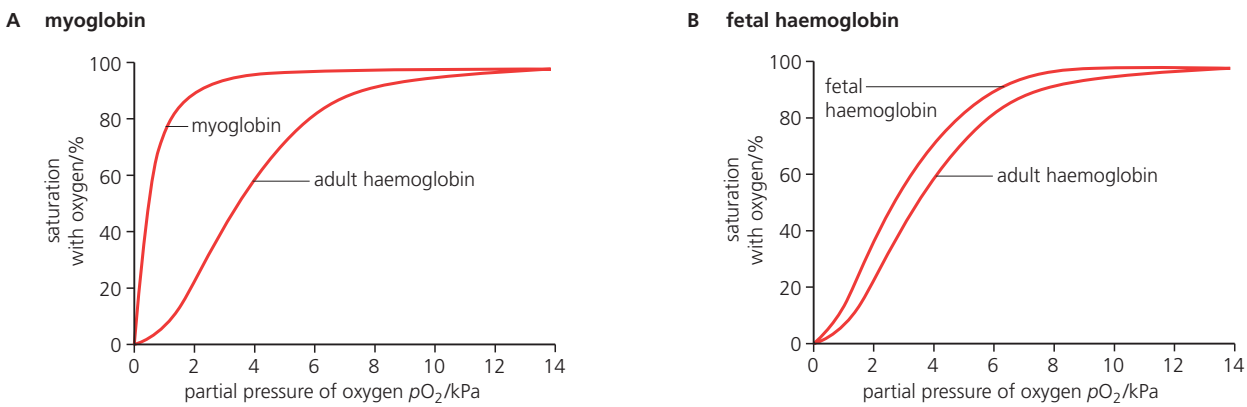
Myoglobin is a respiratory pigment built of a single haem–globin unit, similar to the four units in haemoglobin. It is only found in skeletal muscle cells, where it acts as a **reserve of oxygen**. This reserve is drawn on during intense muscle contraction when the oxygen supply would otherwise be insufficient.

Myoglobin has a much higher affinity for oxygen than haemoglobin (Figure 20.24 A), so in normal conditions in muscle the myoglobin is saturated with oxygen. It functions as an oxygen store. When muscle is very active for a prolonged period, the oxygen concentration in the muscle tissue may fall below 0.5 kPa. When this happens, oxymyoglobin will dissociate and supply oxygen, allowing aerobic respiration to continue. Finally, if muscle contraction continues and all the myoglobin has yielded its oxygen, then muscle tissue will switch to anaerobic respiration by lactic fermentation (page 80) and muscle contraction can continue for longer.

Fetal haemoglobin and oxyhaemoglobin

The fetus obtains oxygen from its mother's blood through the placenta, where the maternal and fetal circulations come very close together, but do not mix (page 386). Now the haemoglobin of the adult mammal and the haemoglobin in the fetal circulation differ slightly in their chemistry. Fetal haemoglobin has the higher affinity for oxygen (Figure 20.24 B). This means that the haemoglobin present in the circulation of the fetus combines with oxygen more readily than the maternal haemoglobin at the same partial pressure. You can guess why it is advantageous that fetal haemoglobin has this property, given that the only access to an oxygen supply for the fetus is via the placenta. If fetal haemoglobin had a lower affinity than its mother's haemoglobin has, oxygen would pass from the fetus to the mother.

Figure 20.24 The oxygen dissociation curves of fetal haemoglobin and myoglobin



The transport of carbon dioxide in the blood

Carbon dioxide is transported from the respiring tissues to the lungs as hydrogencarbonate ions in the plasma and in red cells (Figure 20.25). The steps of this process are as follows, beginning in the respiring tissues:

1 A little carbon dioxide reacts with water of the plasma to form carbonic acid. A slightly larger quantity of carbon dioxide reacts with plasma proteins.
2 Red cells contain the enzyme **carbonic anhydrase** that greatly accelerates the formation of hydrogencarbonate ions from carbon dioxide and water. Most carbon dioxide is transported by this mechanism:

$$CO_2 + H_2O \xrightarrow{\textit{carbonic anhydrase}} HCO_3^- + H^+$$

3 The hydrogen ions are **buffered** by the plasma proteins and haemoglobin, preventing the blood from becoming acidic. This also assists in the release of oxygen from oxyhaemoglobin, in the respiring tissues.

4 Much of the hydrogencarbonate diffuses out into the plasma and chloride ions diffuse into the red cell (**chloride shift**), maintaining charge neutrality.
5 The **pumping action of the heart** causes the deoxygenated blood in the respiring tissues to return to the lungs (via the right side of the heart).
6 In the capillaries of the lungs, these **processes are reversed**.

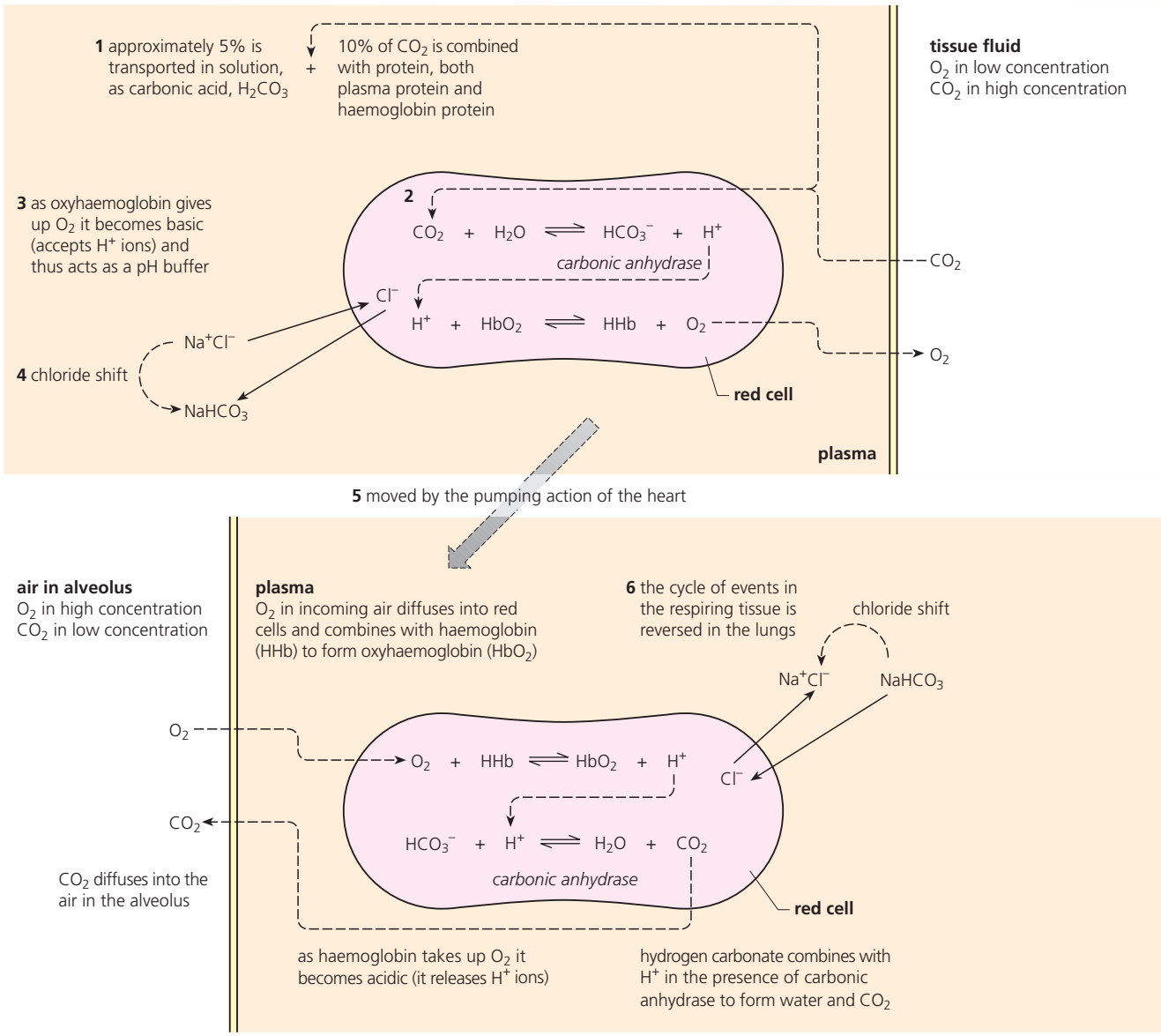


Figure 20.25 The transport of carbon dioxide between respiring tissues and lungs

14 **Explain** the events and the significance of the chloride shift in the transport of carbon dioxide in the red cells.

The control of breathing – the ventilation of the lungs

The **respiratory centre**, situated in the medulla of the hindbrain, controls the rate at which we breathe. Here, two adjacent and interacting groups of nerve cells (neurones), known as the **inspiratory centre** and the **expiratory centre** respectively, bring about ventilation movements by reflex action. Breathing occurs automatically (involuntarily).

- The inspiratory centre sends impulses to increase rate and depth of breathing.
- The expiratory centre sends impulses to inhibit the inspiratory centre and stimulate expiration.
- Alternating impulses from these two centres cause rhythmic breathing.

The breathing rate is also continually adjusted. On average, our normal rate of breathing is about 15 breaths per minute. Since the tidal volume is typically 400 cm^3, the volume of air taken into the lungs in one minute (ventilation rate) is about 6 litres. We can consciously override this breathing rate with messages sent from the cerebral hemispheres, as when we prepare to shout, or sing or play woodwind or brass instruments.

Breathing rates may also be adjusted without conscious thought. This occurs during increased physical activity, when voluntary muscles use much more oxygen and more carbon dioxide is produced and transported in the blood. The main stimulus that controls breathing is the concentration of carbon dioxide in the blood. Blood carbon dioxide level is detected by the chemoreceptors present in the carotid arteries and aorta (Figure 20.26).

When carbon dioxide level increases, as during strenuous physical activity, the chemoreceptors that are **hydrogen ion detectors** (CO_2 is an acid gas, in solution) send impulses to the inspiratory centre. In response, this centre sends additional impulses to the intercostal muscles and diaphragm, causing an increase in their contraction rates. (To a lesser extent, lowered oxygen concentration is also detected.)

After strenuous exercise stops, the concentration of carbon dioxide in the blood falls (and the concentration of oxygen rises). These changes are detected and the ventilation rate is regulated accordingly.

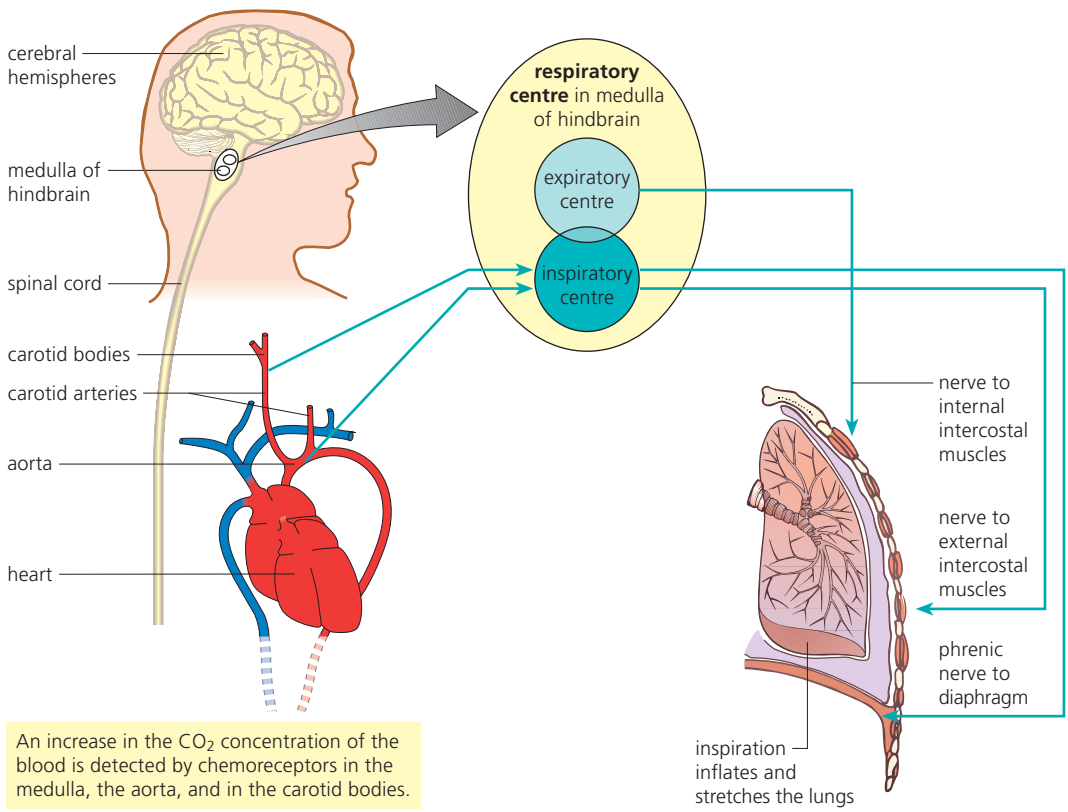


Figure 20.26 The control of ventilation rate

Gaseous exchange under stress

Asthma

Asthma is a disease of the airways of the lungs. In an asthma attack, these become narrow due to contraction of the smooth muscle in the walls of the trachea, bronchi, and bronchioles (Figure 7.29, page 208). Breathing becomes difficult. Symptoms also include coughing, wheezing and a feeling of tightness in the chest. An abnormal rapid heart rate may also occur. Extra mucus of a particular viscid nature, produced as part of the asthma condition, exaggerates the symptoms.

But asthma is a disease involving the immune system (as are the conditions of hay fever and eczema). It is the immune response – an acute inflammatory response – which causes the symptoms. Release of histamine triggers an asthma attack, and also attracts an excess of phagocytic cells to the site.

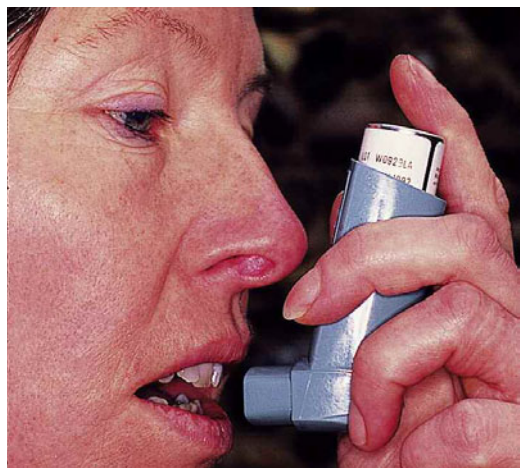


Figure 20.27 Asthma patient using an inhaler

The whole experience of asthma leaves the patient fatigued and anxious. A chronic long-term effect is the build-up of fibrous tissue in the lungs – replacing the normal epithelium of alveoli. This reduces the gas exchange surfaces.

Asthma attacks are triggered by irritants like pollen, dust from pets, droppings from house-dust mites, by certain viruses, or oxides of nitrogen present in vehicle exhaust fumes. People diagnosed as susceptible carry an inhaler to treat themselves (Figure 20.27). Long-term therapy strives to suppress the underlying inflammation by the administration of anti-inflammatory drugs.

Asthma is increasing in incidence in the developed world. Some argue this is because we are underexposed to disease-causing organisms as affluence influences life style. Perhaps our immune system, under-challenged, has become hypersensitive and over-reacts to certain conditions we may experience in our lungs? This is just one hypothesis.

High altitude

It is estimated that more than 40 million people live and work at altitudes of 3000–5500 m, mainly in the Andes and Himalayas. The problems of gaseous exchange at high altitude do not arise because the percentage of oxygen is lower up there – it isn't; the percentage of oxygen does not vary significantly between sea level and high altitudes. With increasing altitude, atmospheric pressure falls, and so the partial pressure of oxygen also falls (Table 20.4).

Altitude/m above sea level	Atmospheric pressure/kPa	Oxygen content/%	Partial pressure of oxygen/kPa
0	101.3	20.9	21.2
2 500	74.7	20.9	15.7
5 000	54.0	20.9	11.3
7 000	38.5	20.9	8.1
10 000	26.4	20.9	5.5

Table 20.4 Change in partial pressure of oxygen at altitude

The result of these changes is that as altitude increases, it becomes increasingly hard for haemoglobin in red cells in the lungs to load oxygen. Once the percentage saturation with oxygen is lowered, this is detected by chemoreceptors. The response of the respiratory centre is to cause the taking of extra-deep breaths. As a result of these, more carbon dioxide is lost from the body, which causes a small but significant rise in the pH of the blood. Now the chemoreceptors become ineffective. Ventilation regulation is hampered.

The body cannot adapt to high altitude immediately; sudden, prolonged exposure at these altitudes by unacclimatised people can be fatal. However, progressively the following changes take place.

- A more alkaline urine is secreted by the kidney tubules via the collecting ducts, and the pH of the blood returns to normal; as a result, the carbon dioxide chemoreceptors become sensitive again, and normal ventilation is maintained.
- Bone marrow tissue, the site of red cell formation, is produced and releases more red cells, thereby enhancing the oxygen-carrying capacity of the blood (Table 20.5).

	Altitude/m above sea level	Red cell count/× 10^{12} dm^{-3}
Human	0 (sea level)	5.00
	5000+ as a temporary visitor	5.95
	5000+ as a resident	7.37
Rabbit	0	4.55
	5000+	7.00
Sheep	0	10.5
	5000+	12.05

Table 20.5 Acclimatisation and adaptation of mammals to breathing at high altitude

Animals that have evolved at high altitude have a form of haemoglobin in their red cells that loads more easily at lower partial pressures, as shown by an oxygen saturation curve obtained with samples of their blood (Figure 20.28). So, for example, a llama from habitats high in the Andes mountains of South America, has a much lower loading partial pressure for its haemoglobin than a mammal from the lowlands of South America.

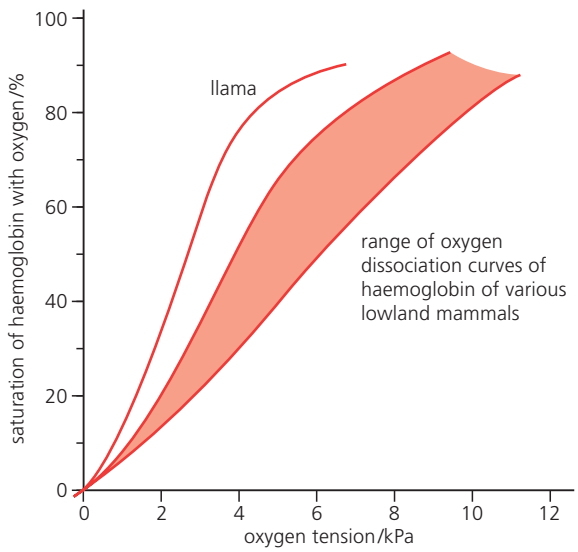


Figure 20.28 Dissociation curve of haemoglobin of llama from mountains in South America

15 **Analyse** and list the changes that occur to the breathing rate and its control as the body makes long-term adaptations to high altitude.

Examination questions – a selection

Questions 1–5 are taken from past IB Diploma biology papers.

Q1

	CO_2 Transport in blood plasma at rest and during exercise		
		Rest	Exercise
Form of transport	**Arterial/mmol l^{-1}**	**Venous/mmol l^{-1}**	**Venous/mmol l^{-1}**
dissolved CO_2	0.68	0.78	1.32
hydrogencarbonate ion	13.52	14.51	14.66
CO_2 bound to proteins	0.3	0.3	0.24
Total CO_2 in plasma	14.50	15.59	16.22
pH of blood	7.4	7.37	7.14

A major requirement of the body is to eliminate carbon dioxide (CO_2). In the body, carbon dioxide exists in three forms: dissolved CO_2, bound as the hydrogencarbonate ion, and bound to proteins (e.g. haemoglobin in red blood cells or plasma proteins). The relative contribution of each of these forms to overall CO_2 transport varies considerably depending on activity, as shown in the table above.

a Calculate the percentage of CO_2 found as hydrogencarbonate ions in the plasma of venous blood at rest. (1)

b **i** Compare the changes in total CO_2 content in the venous plasma due to exercise. (1)

ii Identify which form of CO_2 transport shows the greatest increase due to exercise. (1)

c Explain the pH differences shown in the data. (3)

Higher Level Paper 3, November 02, QH1

Q2 **a** Outline the contents of saliva and gastric juice. (4)

b Explain how the ileum absorbs and transports sugar and lipids. (6)

Higher Level Paper 3, November 05, QH3

Q3 **a** Describe the mechanisms that control the heart beat. (4)

b Explain why ventilation rate varies with exercise. (6)

Higher Level Paper 3, November 03, QH3

Q4 **a** Describe the mode of action of steroid hormones and peptide hormones. (4)

b Explain the control of ADH secretion. (6)

Higher Level Paper 3, May 03, QH3

Q5 **a** State where bile is synthesised. (1)

b Explain the role of bile in digestion. (2)

Higher Level Paper 3, May 02, QH2

Questions 6–10 cover other syllabus issues in this chapter.

Q6 **a** Describe the roles of the capillary networks in the hypothalamus. (3)

b Draw a diagram of the hypothalamus and pituitary glands to show the neurosecretory cell connections between them. Annotate your drawing to outline the functions of the structures shown. (10)

Q7 **a** Describe the role of the hepatic portal vein. (2)

b Explain what is meant by *threshold level* in relation to the regulation of blood sugar by the liver. (2)

c Outline the steps involved in the metabolism of amino acids by the liver, including any that are in excess to body requirements. (4)

Q8 **a** State which heart valves are:

i opened

ii closed

by blood pressure changes during *ventricular systole*. (2)

b Explain how degeneration of artery walls may lead to a fatal heart disease. (6)

c Define hypertension, and identify three factors that contribute to this condition. (4)

Q9 **a** Describe the structure of haemoglobin and the way that oxygen is retained by this molecule in transit in the blood circulation. (6)

b Define *partial pressure*. (2)

c Explain how the body adapts to overcome the special problems of gaseous exchange at high altitude. (4)

Q10 Discuss the ways in which carbon dioxide produced in cell respiration affects the transport of oxygen by the blood. (8)

21 Statistics

STARTING POINTS

- In scientific investigations, observations are recorded as **data** (singular, **datum**). There are different forms of data, but they are all potentially useful. No one type of data that is relevant to an enquiry is necessarily superior to any other type, provided the data have been accurately made and recorded.
- We can define the types of data we collect as:
  qualitative (or descriptive) observations such as, in behaviour studies, the feeding mechanism of honey bees visiting flowers (Figure 17.13, page 523) or nesting behaviour of a species of bird. Qualitative data may be recorded in written observations or notes, or by photography or drawings
  quantitative (or numerical) observations such as the size (numbers, length, breadth or area) of an organism, or of organs such as the leaves of a plant in shaded and exposed positions, or the pH values of soil samples in different positions (Figure 19.20, page 621).
- Quantitative data may be discrete or continuous:
  discrete data are whole numbers, such as the number of eggs laid in a nest – no nest ever contains 2.5 eggs!
  continuous data can be any values within some broad limit, such as the heights of the individuals of a population.
- In this chapter, a means of representing the **variability** of graphical data is illustrated, and statistical tests are examined. These concern the calculation of **means** and of **standard deviation**, and discussion of their usefulness, followed by application of the ***t*-test**.

Recording variability of data – error bars

1.1.1

In experimental science, the outcomes of investigations are checked to confirm they are reproducible. So, for example, when a leading laboratory makes an important discovery and publishes results in a paper, details of the experimental methods are given so that others may repeat the work. Incidentally, if other laboratories fail to confirm the results, then a controversy breaks out. The results are not accepted. Subsequent investigations on both sides of the ensuing argument eventually lead to a resolution of the difference.

Most often, initial results are confirmed because investigations are repeated several times, at the outset, before results are published.

In particular experiments that are part of your course, time may sometimes be too limited for you to repeat readings as you might wish. However, groups of fellow students carrying out the same experiment may be able to pool results. If so, then you may be able to see how variable or consistent a particular result is. So when you display data as part of your record of an investigation, using a graph for example, you can record the degree of variability in readings that the student group obtained in total. To do this you use **error bars**.

An example of error bars in use

A preliminary investigation of the effect of aerobic and anaerobic pre-treatment of tissue discs on their subsequent gain in mass is shown in Figure 21.1. Thin discs of plant tissue are often used because this technique allows all the cells in a sample to receive more or less identical conditions. (You can see the use of leaf tissue discs in an experiment in Figure 15.5, page 451.)

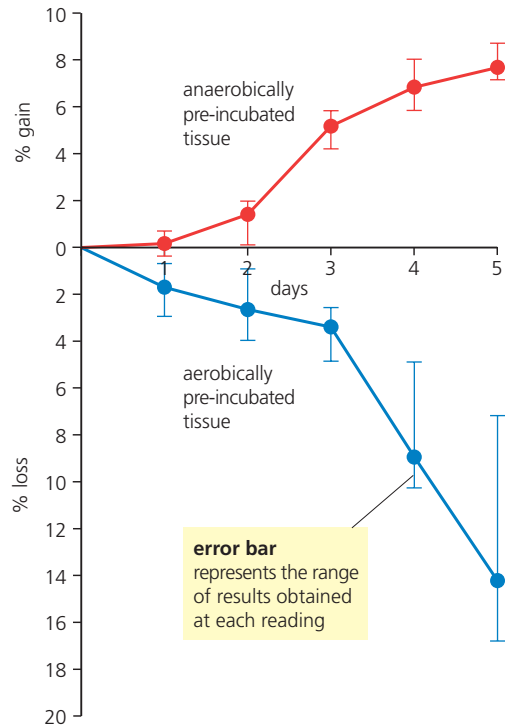


Figure 21.1 Change in fresh mass of tissue discs after aerobic and anaerobic pre-treatment

The results of this enquiry indicate that anaerobic pre-treatment leads to subsequent gain in mass, whereas aerobic pre-treatment leads to loss in mass. This experiment was based on five batches of ten discs for both treatments, and the variability of the results is recorded in error bars. Each error bar indicates the range of values (readings) from the highest to the lowest. In Figure 21.1, for example, the error bars draw attention to the much greater variation in results from aerobically pre-treated tissue.

Sometimes, the error bars shown in a graphical representation of data record variability as the standard deviation of a result. The calculation of standard deviations is discussed shortly.

Summarising data – the mean

Points on the curves in the graph in Figure 21.1 are based on five tissue batches (each of ten tissue discs) for each treatment. The value of each is the average or **arithmetic mean value** of individual batches. The value of means is that they convey the 'middleness' of the data.

How are the averages or means of the readings calculated?

To calculate a mean value, all the values from a particular treatment or observation are summed, and the total divided by the number of these values.

The formula for the arithmetic mean is:

$$\bar{x} = \frac{\Sigma x}{n}$$

where
$\bar{x}$ = arithmetic mean
Σx = sum of all the measurements
n = the total number of measurements

The value of means in other experimental situations

For example, let us imagine you have completed an investigation on the effects of the application of pesticide on the numbers of a common species of soil organism.

The outcome is that your field notebook now contains a large number of counts from randomly placed quadrats (page 601), some from treated soils, and some from untreated soil (see table in Figure 21.2).

In Figure 21.2, the data are also presented graphically, with the number of worms per quadrat on the *x*-axis, and the frequency of quadrats with each number of worms on the *y*-axis.

Look at the spread of data from both treatments.

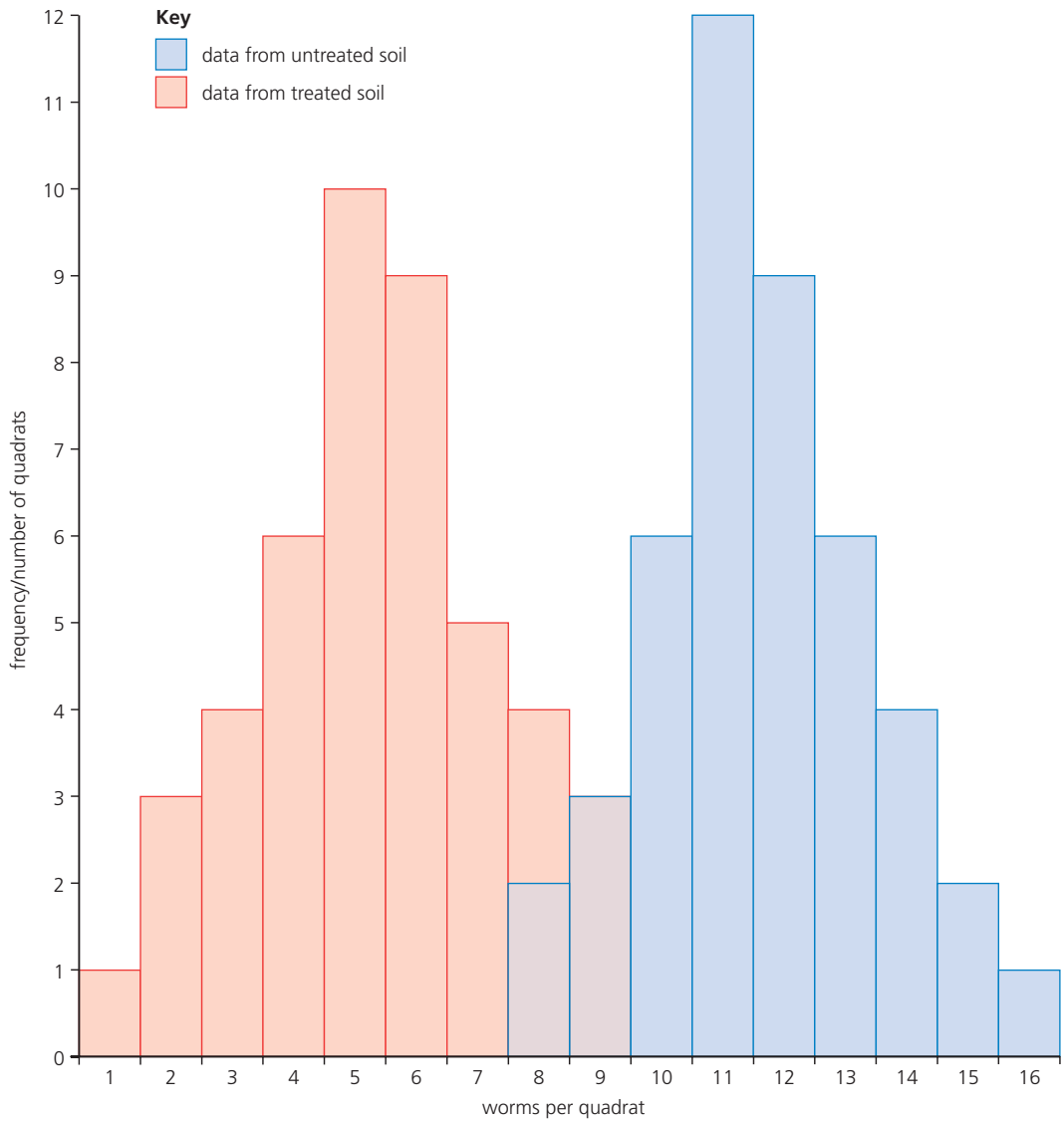
With these data presented as a graph, two characteristic bell-shaped curves result (which only slightly overlap). These are referred to as **normal distributions**. These arise, given a large enough number of observations or measurements, if the data may have an exactly symmetrical spread. However, data rarely exactly conform completely to a bell-shaped curve – an approximation is usual, given a finite number of observations. This is what we see in Figure 21.2.

Figure 21.2 An investigation of the effects on soil worm populations of pesticide treatment

experimental results

Quadrats on soil treated with pesticide			Quadrats on untreated soils		
Worms per quadrat	Frequency	Total	Worms per quadrat	Frequency	Total
0		**0**	7		**0**
1	𝍸	**1**	8	𝍸	**2**
2	𝍸	**3**	9	𝍸	**3**
3	𝍸	**4**	10	𝍸 𝍸	**6**
4	𝍸 𝍸	**6**	11	𝍸 𝍸 𝍸	**12**
5	𝍸 𝍸	**10**	12	𝍸 𝍸	**9**
6	𝍸 𝍸	**9**	13	𝍸 𝍸	**6**
7	𝍸	**5**	14	𝍸	**4**
8	𝍸	**4**	15	𝍸	**2**
9	𝍸	**3**	16	𝍸	**1**
10		**0**	17		**0**

graph of frequency against numbers of earthworms per quadrat



How can the data best be summarised so that a comparison of the effect of the alternative treatments can be made?

The answer is in different ways, one of which is to find the **mean value** of soil worms for each treatment. To do this, all the values from quadrats on treated soil are summed, and the total divided by the number of values. In this way, we have an average value for the effect of this soil treatment. The mean will convey the 'middleness' of the data. The counts from quadrats on untreated soil are treated in the same way too, of course. *You could calculate the means for the data in Figure 21.2, for both treated soil and untreated soil.*

Extension: Mean, median and mode

In the handling of experimental data, you may come across two other terms, namely **mode** and **median**. These are not alternatives to the mean value; rather they have different meanings:

- **mode** is the most frequent value in a set of values;
- **median** is the middle value in a set of values arranged in ascending order.

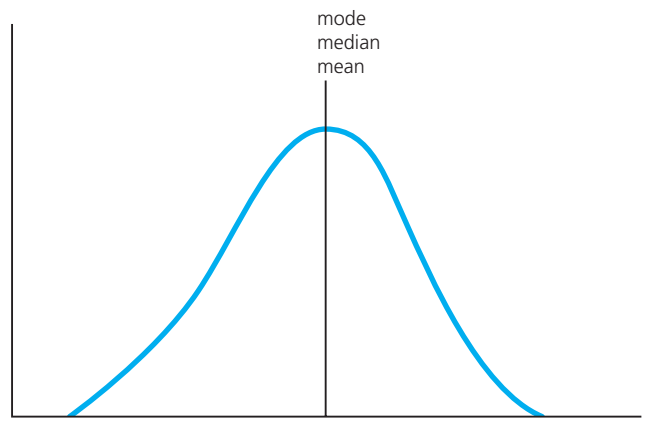
The graphs in Figure 21.3 illustrate how mean, mode and median relate in a **normal distribution** and in **skewed data**. You can see that it is in skewed data that they have particular significance.

Figure 21.3 Frequency distributions of symmetrical and skewed data

Normal distribution curve

Most biological data shows variability, but with values grouped symmetrically around a central value.

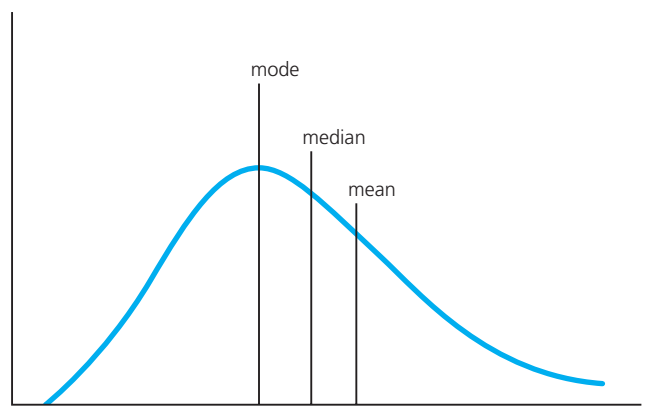
Here the mode, median and mean coincide.



Skewed distribution

Values reduce in frequency more rapidly on one side of the most frequently obtained value than the other.

Here the difference between the mean and mode is a measurement of 'skewness' of the data.



Calculating standard deviations

1.1.2–1.1.4

The **standard deviation** (**SD**, **s** or σ) of the mean tells us how spread out are the readings (the 'spreadoutness' of the data).

A small standard deviation indicates that the data is clustered closely around the mean value. A large standard deviation indicates a wider spread around the mean.

So, standard deviations are a measure of the variation in the data from the mean value of a set of values.

The five steps to calculating the standard deviation of a data set are:

1 Calculate the mean ($\bar{x}$)
2 Measure the deviations ($x - \bar{x}$)
3 Square the deviations $(x - \bar{x})^2$
4 Add the squared deviations $\sum(x - \bar{x})^2$
5 Divide by the number of samples (n).

Calculating standard deviations – an example

An ecologist investigated the reproductive capacity of two species of buttercup, *Ranunculus acris* (meadow buttercup) and *R. repens* (creeping buttercup).

Figure 21.4 Data on achene (fruit) production in two species of *Ranunculus*

Number of achenes	Frequency	
	R. repens	*R. acris*
15	1	0
16	1	0
17	1	0
18	2	1
19	4	1
20	4	1
21	8	1
22	7	1
23	9	3
24	10	4
25	16	4
26	9	5
27	10	5
28	4	6
29	5	8
30	3	14
31	1	12
32	1	10
33	2	7
34	1	3
35	1	2
36	0	3
37	0	2
38	0	3
39	0	2
40	0	2
41	0	0

The latter species spreads vegetatively via strong and persistent underground stems. Would this investment be reflected in a lowered production of fruit (the product of sexual reproduction) compared with fruit production by the meadow buttercup, which reproduces more or less exclusively by sexual reproduction?

Using comparable sized plants growing under similar conditions in the same soil, the numbers of achenes (fruits) formed in 100 flowers of each species were counted and recorded. The results are given in Figure 21.4, and calculations of the SDs are shown in Figure 21.5. Note that you are not expected to know the formula for calculating SD. The purpose of presenting the steps to the calculation is to take away the mystery of a calculation normally carried out by a scientific calculator or programmed spreadsheet.

Alternative methods of calculating means and SDs

Rather than carry out all these steps manually – for example, using a dedicated worksheet as in Figure 21.5 – the value of the standard deviation may be obtained using a scientific or **statistics calculator**, or by means of a **spreadsheet** incorporating formulae, or by using Merlin (page 683).

Using a scientific calculator or spreadsheet, you can calculate the SDs of both the data of frequencies of worms on quadrats on soil treated with pesticide and on untreated soils, at this point, if you wish.

Once obtained, the value may be applied to the normal distribution curve, as shown in Figure 21.6. Note that 68% of the data occurs within ± 1 SD, and more than 95% of the data occurs within ± 2 SDs.

Figure 21.5 Calculating the means and SDs of the data in Figure 21.4

achene production in *Ranunculus acris*

Values obtained in ascending order x	Frequency f	fx	Deviation of x from the mean $(x - \bar{x})$ [= d]	d^2	fd^2
17	0				
18	1	18	–12	144	144
19	1	19	–11	121	121
20	1	20	–10	100	100
21	1	21	–9	81	81
22	1	22	–8	64	64
23	3	69	–7	49	147
24	4	96	–6	36	144
25	4	100	–5	25	100
26	5	130	–4	16	80
27	5	135	–3	9	45
28	6	168	–2	4	24
29	8	232	–1	1	8
30	14	420	0	0	0
31	12	372	1	1	12
32	10	320	2	4	40
33	7	231	3	9	63
34	3	102	4	16	48
35	2	70	5	25	50
36	3	108	6	36	108
37	2	74	7	49	98
38	3	114	8	64	192
39	2	78	9	81	162
40	2	80	10	100	200
41	0				
	$\Sigma f = 100$	$\Sigma fx = 2999$			$\Sigma fd^2 = 2031$

Mean of data = $\frac{\Sigma fx}{\Sigma f} = \frac{2999}{100} = 29.99$

SD = $\sqrt{\left(\frac{\Sigma fd^2}{\Sigma f - 1}\right)} = \sqrt{\frac{2031}{99}} = \sqrt{20.51} = 4.53$

Thus the mean of the sample *Ranunculus acris* = 29.99, and the SD = 4.53.

achene production in *Ranunculus repens*

Values obtained in ascending order x	Frequency f	fx	Deviation of x from the mean $(x - \bar{x})$ [= d]	d^2	fd^2
14	0				
15	1	16	–9	81	81
16	1	17	–8	64	64
17	1	36	–7	49	98
18	2	76	–6	36	144
19	4	80	–5	25	100
20	4	168	–4	16	128
21	8	154	–3	9	63
22	7	207	–2	4	36
23	9	240	–1	1	10
24	10	400	0	0	0
25	16	234	1	1	9
26	9	270	2	4	40
27	10	112	3	9	36
28	4	145	4	16	80
29	5	90	5	25	75
30	3	31	6	36	36
31	1	32	7	49	49
32	1	66	8	64	128
33	2	34	9	81	81
34	1	35	10	100	100
35	1	36	11	121	121
36	0				
	$\Sigma f = 100$	$\Sigma fx = 2479$			$\Sigma fd^2 = 1479$

Mean of data = $\frac{\Sigma fx}{\Sigma f} = \frac{2479}{100} = 24.79$

SD = $\sqrt{\left(\frac{\Sigma fd^2}{\Sigma f - 1}\right)} = \sqrt{\frac{1479}{99}} = \sqrt{14.93} = 3.86$

Thus the mean of the sample *Ranunculus repens* = 24.79, and the SD = 3.86.

The value of calculating standard deviation

We have noted that a standard deviation of low value indicates that the observation differs very little from the mean, and that high values of SD indicate a wider spread around the mean.

Thus the SDs can be used to help to decide whether the differences between the two related means are significant or not, such as those shown in Figure 21.2 (page 677).

If the SDs are much larger than the difference between the means, then the differences in the means are highly unlikely to be significant.

On the other hand, when SDs are much smaller than the differences between the means, then the differences between the means is almost certainly significant.

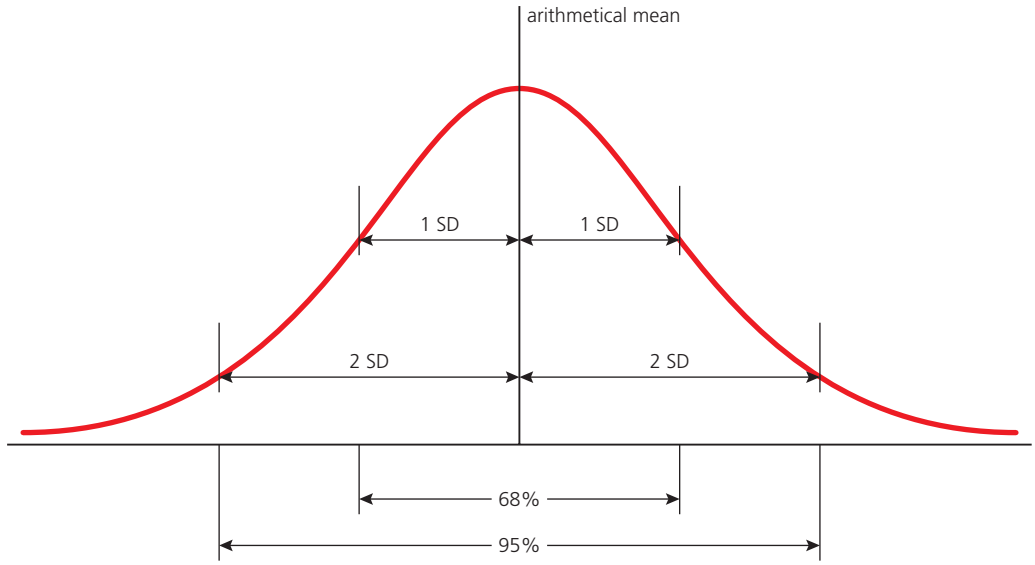


Figure 21.6 The normal distribution and its SD

Another statistical test – the *t*-test

1.1.5

Statistical tests typically compare large, randomly selected representative samples of normally distributed data. In practice, it is often the case that data can only be obtained from quite small samples. The ***t*-test** may be applied to sample sizes of more than 5 and less than 30 of normally distributed data. It provides a way of measuring the overlap between two sets of data – a large value of *t* indicates little overlap and makes it highly likely there is a significant difference between the two data sets. An example will illustrate the method. However, you should note that you are not expected to calculate values of *t*.

Applying the *t*-test

An ecologist was investigating woodland microhabitats, contrasting the communities in a shaded position with those in full sunlight. One of the plants was ivy (*Hedera helix*), but relatively few occurred at the locations under investigation. The issue arose: were the leaves in the shade actually larger than those in the sunlight?

Leaf widths were measured, but because the size of the leaves varied with the position on the plant, only the fourth leaf from each stem tip was measured. The results from the plants available are shown in Table 21.1.

Size-class/mm	Leaves in sunlight (a)	Leaves in shade (b)
20–24	24	
25–29	26, 26	26
30–34	30, 31, 31, 32, 32, 33	33, 34
35–39	37, 38	35, 35, 36, 36, 36, 37
40–44	43	41, 42
45–49		45

Table 21.1 Sizes of sun and shade leaves of *Hedera helix*

Steps to the *t*-test

1 The null hypothesis (negative hypothesis) assumes the difference under investigation has arisen by chance. In this example, the **null hypothesis** is:
'There is no difference in size between leaves in sunlight and leaves in shade.'
The role of the *t*-test is to determine whether to accept or reject the null hypothesis. If it is rejected here, we can have confidence that the difference in the leaf sizes of the two samples is statistically significant.

2 Next, check that the data are normally distributed. This is done by arranging the data for leaves in sunlight and leaves in shade as in Table 21.1 (and plot a histogram, if necessary).

3 *You are not expected to calculate values of t.* This statistic can be found by using a scientific or statistics calculator, or by means of a spreadsheet incorporating formulae.

Actually, a formula for the *t*-test for unmatched samples (data sets a and b) is:

$$t = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$$

where
$\bar{x}_a$ = mean of data set a
$\bar{x}_b$ = mean of data set b
s_a^2 = standard deviation for data set a, squared
s_b^2 = standard deviation for data set b, squared
n_a = number of data in set a
n_b = number of data in set b
$\sqrt{}$ = square root of

4 Once a value of *t* has been calculated (here $t = 2.10$), we determine the **degrees of freedom (df)** for the two samples, using the formula:
df = (total number of values in both samples) – 2
$= (n_a + n_b) - 2$
In this case:
df = (12 + 12) – 2 = 22.
Now we consult a table of critical values for the *t*-test.

5 A table of critical values for the *t*-test is given in Figure 21.7. Look down the column of significance levels (p) at the 0.05 level until you reach the line corresponding to df = 22. You will see that here, $p = 2.08$.

6 Since the calculated value of *t* (2.10) exceeds this critical value (2.08) at the 0.05 level of significance, it indicates that there is a lower than 0.05 probability (5%) that the difference between the two means is solely due to chance. Therefore, we can reject the null hypothesis, and conclude the difference between the two samples is significant.

For the experimenter, the significance of all this, is that there is a reason for the difference in the means, which can now be further investigated and fresh hypotheses proposed.

Extension: Merlin – statistical software available to biology students

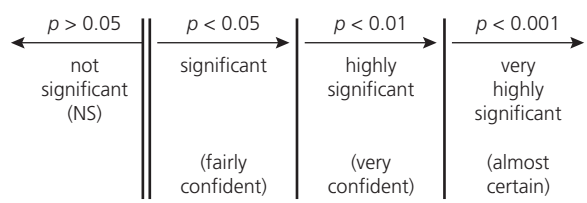
Merlin is a statistical package produced by Dr Neil Millar of Heckmondwike Grammar School, UK. This software is an add-in for Microsoft Excel and is easy to use. Merlin is available free of charge for educational and non-profit use. The package may be copied to laboratory and student computers from the school's website, currently at
www.heckgrammar.kirklees.sch.uk/index.php?p=310

URLs do change periodically. Merlin can also be located by means of a Google search under 'merlin+statistics'.

Once you have data from laboratory or field investigations, Merlin can be used to carry out a statistical test. There are no calculations required, and no look-up tables – no maths, no mistakes, in effect! Merlin also includes a basic introduction to statistics for biology students and a 'test chooser'. Here, in answer to a series of questions, Merlin selects the right test. Also with Merlin, Excel can display data in a range of graphs and charts, as appropriate.

Figure 21.7 Critical values for the *t*-test

	decreasing value of *p* →			
Degrees of freedom	***p* values**			
(df)	**0.10**	**0.05**	**0.01**	**0.001**
1	6.31	12.71	63.66	636.60
2	2.92	4.30	9.92	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.89	2.36	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
12	1.78	2.18	3.05	4.32
14	1.76	2.15	2.98	4.14
16	1.75	2.12	2.92	4.02
18	1.73	2.10	2.88	3.92
20	1.72	2.09	2.85	3.85
22	1.72	2.08	2.82	3.79
24	1.71	2.06	2.80	3.74
26	1.71	2.06	2.78	3.71
28	1.70	2.05	2.76	3.67
30	1.70	2.04	2.75	3.65
40	1.68	2.02	2.70	3.55
60	1.67	2.00	2.66	3.46
120	1.66	1.98	2.62	3.37
∞	1.64	1.96	2.58	3.29



Correlations do not establish causal relationships

1.1.6

By 'correlation' we mean 'a mutual relation between two (or more) things', or 'an interdependence of variable quantities'. The belief that because things have occurred together, one must be connected or related to the other in the sense that one is the cause of the other is an easily and commonly made mistake. Because two events (A and B) regularly occur together, it may appear to us that A causes B. This is not necessarily the case. In fact, there may be a common event that causes both, for example, or it may be an entirely spurious correlation. For example, some infants (very few) develop the symptoms of autism shortly after the normal time in childhood when the MMR inoculation is administered. Some parents of autistic children who had arranged for their child to be inoculated came to blame the vaccination for the child's condition. This confusion caught on for several years, many parents became anxious, and the practice of having the triple injection became unpopular. The numbers of vaccinated children fell to dangerously low levels. It was some time before detailed studies could convince the

majority of parents that the two events, MMR inoculation and the onset of autism, were not causally linked (page 360).

The fact that correlation does not prove cause was one of the reasons why Richard Doll's amassing of statistical evidence of a link between smoking and ill health was successfully resisted by the tobacco industry for an exceptionally long time (page 666). Now we know the various reasons why cigarette smoke triggers malfunctioning of body systems, ill heath and diseases of various sorts.

So, having applied statistical tests that indicate the possibility of a correlation, we cannot then assert that one event is the cause of the other. What we can do is have confidence that the events may well be linked, and so go on to investigate the mechanisms of the linkage – if there is one.

For example, a persistent condition of hypertension is directly linked to the raised incidence of coronary heart disease and vascular accidents of other sorts (page 665). In this case, the statistical relationship has been followed up, enabling us to understand why hypertension has these effects. Once the connections between events or conditions are understood, the relationship has been established. In other words, just because a correlation does *not* prove the cause does *not* mean there cannot be a causal relationship.

So, statistical confidence in the possibility of a causal link is a springboard to further investigation, not proof of a relationship.

TOK Link

Is the idea of 'cause and effect' generally uncritically accepted?

Does it permeate our culture?

Examine some current newspapers or journals that are frequently read in your country. Can you find examples of assumptions about cause and effect that are stated, but which are unproved and possibly dubious?

22 Teaching and learning IB Diploma biology

guest author Gary Seston

Whether you are a teacher or a student, you can feel both privileged and excited to be part of the International Baccalaureate Organisation (IBO) and its Diploma programme. This model of education is rapidly gaining popularity throughout the world and is already the preferred choice for many universities. It is easy to see why the programme has gained such respect when one appreciates its philosophical roots. The IB furthers the education of the whole person, not just by requiring knowledge and skills but also by engendering personal attributes and states of being to create **lifelong learners**.

This approach is encapsulated in a document available from IBO called the **IB Learner Profile**. While this document initially seems to be aimed at students, it is clear that the best IB teachers are those who consider themselves also to be lifelong learners, so the IB Learner Profile is also personally useful to teachers. If you are a teacher who is new to IB, then prepare to be a learner, too.

To all IB learners – both teachers and students

This chapter is designed to help students and teachers new to IB Diploma biology to develop an overview of the course and appreciate the necessary approaches for success. It enables you to visualise what the IB Diploma experience will entail and helps you to make the most of it by:

- briefly summarising the philosophy and approach of the IB Diploma programme and the place of IB Diploma biology within it;
- describing the components of the IB Diploma biology course as experienced by students and teachers;
- providing straightforward practical advice to ensure your success.

The sections below are intentionally brief and to the point. Further details of procedures can be obtained directly from the IBO via the school's IB co-ordinator.

What distinguishes the IB Diploma biology approach?

Perhaps more than with other pre-university biology courses, you will get the distinct impression this course is one important component within a much greater scheme, the IB Diploma. This philosophy of **integration** binds all aspects of the programme. Biology will not stand alone but will be an important vehicle for exploring and developing the various aspects of the IB learner profile. There will be a clearly **international** aspect to your study in that it will consider global issues as well as local case studies. Central to all subjects will be the work done on the **IB Theory of Knowledge** (TOK) course, which creates a philosophical backdrop against which to discuss issues raised within the biology course. Specific reference to TOK issues have been included within the chapters of this book to ensure that you make these links. For example, within Topic 6 (Chapter 7, page 203), students are asked to consider the response of individuals and of governments to the rapid increase in sexually transmitted diseases such as AIDS. This **broadening** and often **trans-disciplinary** influence, together with IBO's periodic syllabus review cycle, ensures that the syllabus is both **contemporary and relevant** by including important issues of the time.

This course is not just a matter of knowing some biology facts. You will find that developing informed viewpoints, engaging in discussion and evaluating evidence is inherent in the approach needed for success at IB Diploma biology (Figure 22.1). You will also need to appreciate that definite, clear answers are not always possible. That's what makes it so much fun for both students and teachers.

Figure 22.1 IB students at United World College of South East Asia engaged in discussion

What will I experience as a student or teacher of IB Diploma biology?

Whether you are a student or teacher, you will discover that the IB way is one of partnership. Teachers and students share the experience in an open and co-constructive approach in which all are aware of the learning outcomes expected and modes of assessment employed. The principles embodied in the Assessment for Learning initiative can be put to good use in IB Diploma biology. The initiative maintains that students should work together with teachers and each other to focus on a continuous progress cycle of setting and reviewing specific learning targets on the path towards their final assessment. As an indication of this, the internal assessment of practical skills allows students to be made fully aware of the assessment criteria applied and then assesses students based on their best scores for each skill, as opposed to an average of all scores over the course.

Your more tangible experience of IB Diploma biology will include:

- **theory** – developing knowledge and understanding of biological theory as dictated by the syllabus;
- **practical** – developing the methodological (practical) skills needed to be successful in the internal assessment of practical work;
- **Group 4 project** – engaging in collaborative, constructivist research with other science students;
- **extended essay** (perhaps) – engaging in an independent research project on a biological topic.

The most successful IB biology courses are those where these elements are integrated as far as is practically possible (teacher advice, page 690).

What will each of these components involve?

Studying biological theory towards sitting the external examination

This exciting syllabus leads you to an understanding of life at such diverse levels as:

- the behaviour of electrons passing across proteins within membranes to produce energy potentials equivalent to those within a lightning bolt;
- the impact of humans on the world's climate which might alter the course of evolution.

The subject matter is divided into:

- the **Core syllabus** – studied by all students to provide both breadth and depth of understanding;
- the **Additional Higher Level (AHL) syllabus** – studied only by higher-level (H-level) students to give further depth;
- **two options topics** – studied in great depth; three sorts of option exist – those available to standard-level (S-level) students, those available to H-level students, and those available to both S-level and H-level students with additional higher-level material.

The syllabus is arranged into conceptual areas focusing on the understanding of biological processes (such as homeostasis, photosynthesis) rather than being a survey of the living world from the simple to the complex.

These areas of knowledge are bound by four overarching themes:

- structure and function;
- universality and diversity;
- equilibrium and systems;
- evolution.

The IB biology subject guide available from IBO provides excellent guidance and includes an extremely detailed syllabus with assessment statements, time allocations and teachers' notes, which are useful to both teacher and students throughout the course. The assessment statements in particular give a clear reference for the material which will be assessed in the final examinations. Once understood, these statements act as a valuable checklist during examination preparations. Students who cover the syllabus closely and understand it well will be successful. Obvious, but all too often forgotten. Students who are determined not to let weekly uncertainties pass them by, but instead develop a routine of reviewing recent material and asking for help so that they understand as they proceed, are guaranteed success. The chapters of this book are designed to follow the syllabus very closely and have self-assessment and examination-style questions throughout. Students can expect to be periodically tested by their teacher, using examples of IB questions to check understanding and to enable them to gain practice at the various styles of question. Past IB papers can be obtained from the IBO.

Developing methodological (practical) skills towards internal assessment (IA)

Alongside the syllabus content, students are required to develop methodological (practical) skills which will be assessed formatively and then summatively for the final assessment (Figure 22.2) This work leads to the internal assessment (IA) component of the course which contributes 24% towards the final grade.

The work will involve a mixture of laboratory experiments and fieldwork and may also include surveys, models, evidence of presentations, posters; in fact, any activity can be used to assess students on the IB Group 4 (science) internal assessment criteria, given for assessment of the skills of:

- designing (planning) an investigation;
- collecting, processing and presenting data;
- concluding and evaluating;
- manipulation (carrying out an investigation competently and safely);
- personal skills (perseverance, ethical work, team work and reflection).

Figure 22.2 IB biology students on task in practical work

Each student is expected to spend a certain amount of time on practical work (regardless of number of practical activities performed). Currently this is 60 hours for AHL students and 40 hours for standard students. This time includes all active practical work but not the time to write the reports. Standard-level and higher-level students are assessed in exactly the same way and to the same level, the only difference being the amount of time spent. The school's practical scheme of work should be designed to spread the practical activities over the main part of the course and involve the majority of the syllabus topics.

It is a good idea to start by developing each skill separately using dedicated practical activities (they can be regarded as practices but still contribute to the time spent on practical work). In this respect, 'designing an investigation' may not be the best skill to begin with as it is the one that most students find the hardest. The more familiar skills of 'manipulation' and 'data collection and processing' are a good introduction to using the criteria. Once skills have been developed and students are confident, more demanding practical activities will allow students to score well.

Not all skills will be assessed for each practical. One practical activity might be used to assess manipulation, data collection, data processing and presentation (e.g. determining the concentration of potato cell sap by finding the external solution which causes no mass change – an activity which can be given as a prescribed procedure). A different practical activity might be used to assess planning, conclusion and evaluation (e.g. an investigation into the factors affecting water loss (transpiration) rates from leaves – for which students would be asked to generate their own research question, etc.).

Eventually, by the end of the course, a small number of scores (currently, the best two for each skill) will be used in calculating the summative (final) grade. This means that, as a student, you will probably be assessed many more times than minimally necessary. In essence, you can 'mess up' on some practical activities and still do well if you improve during the course.

To drive this improvement, as a student, you should expect to:

- be aware of which skills are being assessed during each practical activity;
- be given the assessment criteria to work towards;
- be given specific feedback on ways of getting improved scores in future.

This allows a formative approach in which students can learn from their mistakes by receiving specific feedback from teachers. To this end, teachers are encouraged to include criteria-specific comments and grade(s) on students' work, both for the student's benefit and to aid the IB moderator.

Examples of such comments might be:

- 'Design skills, aspect C is partially completed – you need to plan to collect sufficient data by carrying out repeats.'
- 'Data collection and processing skills, aspect C is partially complete – your averages were based on a number of readings which varied somewhat; you need to include error bars for each averaged data point on your graph before you can confidently interpret it.'

Students should try to keep a record of their own grades, set themselves specific targets and review and track their own progress. However, all evidence of the completion of the required number of hours of student practical work (marked practical reports with student instruction sheets attached, digital photographs, etc.) must be kept securely in the student's practical portfolio, which may later be required for moderation purposes by IBO. Teachers may prefer to keep these safely locked away in school.

Similarly, a record of all internal assessment scores must be kept by the teacher from the very beginning of the course. The IB will require the final internal assessment grades and moderation samples 1–2 months before the IB examinations begin; therefore, schools will normally complete and internally moderate practical work around 2–3 months before. Further details of the current internal assessment procedures, skills and assessment criteria can be found in the current biology subject guide available from the IBO.

The Group 4 (science) project

This is an opportunity for some creative and exciting collaboration as part of the practical scheme of work. The emphasis is on process more than product. Biology students will work together with students from other Group 4 subjects such as chemistry, physics and design technology on a co-constructivist research project which is conceived and planned entirely by the students themselves. The time taken to do this will count toward the time allocation for internal assessment and is normally around ten hours. Because of the collaborative nature of the project, the skills of design and data collection cannot be assessed and, in fact, most schools assess only personal skills.

Figure 22.3 Students collecting data as part of their Group 4 project, later to be presented for peer review

A typical organisational plan is:

1. Group 4 students from all subjects brainstorm together to select a theme for study (e.g. Science in sport).
2. The group generates individual research questions to be studied by the individual subject teams comprising biologists, chemists, physicists or technologists. These research questions should be interlinked in some way under the overall theme (e.g. 'physiological response to exercise at different temperatures' for biologists, 'thermal properties of different sports clothing material' for physicists, and so on).
3. Subject teams work on their individual investigations over a number of days.
4. Subject teams report findings to the group emphasising links to the other teams' studies, often brought together in a group presentation or display (Figure 22.3).

Group 4 projects are often carried out midway through the course, after students have developed their methodological skills and become familiar with a range of practical techniques but before the pressures of finalising coursework and preparing for examinations.

The extended essay

Each IB Diploma student is required to carry out an individual piece of research in a chosen subject area, such as biology, and working under the guidance of a teacher supervisor. This culminates in the production of a 4000-word essay. It is an opportunity to follow an area of personal interest, which might well be related to the student's plans for further education.

Students who choose to write the extended essay in biology must first ensure that the topic is clearly biological (rather than psychological or geographical). For example, although a title such as 'A survey of people's attitudes to body shape related to their exposure to fashion magazines', is clearly related to diet and nutrition issues, the overall essay is more likely to be sociological rather than biological. Secondly, the essay should be based on a practical investigation in order to gain a good grade. A description of the mode of transmission of the human immunodeficiency virus (HIV) and finding novel ideas for blocking it, while eminently worthwhile in itself, would not allow for scoring on the biological components of the extended essay assessment criteria.

The study must start with a well-focused research question and a realistic plan generated by the student. It must use either novel techniques of the student's own design or modified standard practical procedures that can be applied to the particular study. Undertaking a simple, well known textbook investigation will not gain a good grade. For example, 'Determining the vitamin C (ascorbic acid) content of lemon juice by titration with DCPIP' is unlikely to yield the opportunities needed.

Furthermore, as the analysis of data contributes a large part to the final grade, it is essential to plan a study that can be guaranteed to yield sufficient data over a reasonable time frame. Avoid investigations that rely on long growth or development periods under novel conditions. For example, 'Investigating birth rates in small fish (e.g. guppies) at different pH values' needs too long a study period; a better study might be 'An investigation into the stimuli provoking the aggressive response of gill flaring in Siamese fighting fish' or 'An investigation into the effects of spices on bacterial growth' or 'An investigation into the effects of ultra-violet radiation on algal growth'.

Another important (though not biological) element in the final grade is the student's ability to present the essay clearly and in accordance with the conventions for research papers (for example, referencing). The IBO produces guidance documents listing general and subject-specific assessment criteria which all students should have access to before planning their extended essay.

Many students find extended essays in biology very attractive and rewarding although some also find that the nature of practical work, using living systems, can be excessively time consuming unless the research question is chosen very carefully indeed.

How will the course be assessed?

The final grade awarded for IB Diploma biology (7 highest, 1 lowest) will be based on the internal assessment of practical work (24%) and on the final external examinations (76%). The final examinations currently include multiple-choice questions, data-analysis questions, short-answer questions and essays. There will be one paper (paper 3) dedicated to the option topics, and which includes data analysis and shorter answer questions (often not very short in practice and requiring in-depth understanding).

Students who practise past examination questions and are then given model answers or mark-schemes become familiar with the format and this becomes a major contributor to their success. In addition, an emphasis on understanding the option topics thoroughly and providing detailed answers often leads to students gaining the best grades.

To new IB teachers

This section is designed to provide further guidance to teachers who are new to IB Diploma biology. It assumes that the previous section has already been assimilated.

You are probably very keen to obtain all the important details of syllabus, internal assessment procedures and examinations, and start planning your course. In this section, we will first look at ways of obtaining the detailed information needed and then at ideas for developing an IB Diploma biology course appropriate for your particular circumstances.

How do I find out more about IB Diploma biology now and in the future?

You will quickly come to understand that the IBO is a dynamic organisation which continually evolves according to a regular development review cycle. As such, it is important that you are familiar with ways of finding out more now and of keeping informed in the future:

- The first point of contact for subject teachers within a school is the school's IB co-ordinator, who will probably act as the sole route of communication with the IBO. He or she will obtain all necessary documents and will also receive periodic coordinator notes, which will include news about developments in Group 4 subjects, including biology.
- The main document to first refer to is the IB biology subject guide. This specifies all current details about the subject.
- To underline their commitment to lifelong learning and continuing professional development, the IBO organises regular regional IB conferences which include IB subject workshops. These are extremely useful for teachers new to the IB and are advertised in the IB co-ordinator notes.

- The IBO provides excellent online help through its online curriculum centre (OCC) which includes all IB documentation, subject forums and resource-sharing facilities. Access details are obtainable from your IB co-ordinator.
- Schools new to IB often make links with an established IB school nearby. Contacting experienced IB biology teachers can often prove to be very useful indeed.

How should we organise and deliver the biology course in our school?

Schools all over the world operate different course-delivery models as dictated by local circumstances such as size of cohort or responsibility to another course (e.g. their national curriculum). The IB requires that IB Diploma biology students complete 240 hours of study for AHL and 150 hours of study for standard level. This includes all theory and practical work. How these hours are organised is for each school to decide.

The delivery model

Different schools organise these hours in different ways. The main difference is with regard to the degree of integration (higher students with standard; theory work with practical) deemed desirable or possible. By far, the preferred model is to have higher and standard students timetabled separately over a two-year course within laboratories so that practical work can be integrated seamlessly within the relevant theory. This is the model favoured by the IBO. If this describes your circumstance, then you are on a good footing. Some schools are unable to arrange for this; other delivery models include:

- The theory delivered in classrooms with a separate, weekly laboratory-based practical session.
- Both H-level and S-level students timetabled together to cover the core material with H-level students attending an extra lesson covering the AHL material. Some options can also be covered this way since they involve a shared standard/higher section followed by an additional higher section. One standard option (Option C: Cells and energy) can be almost fully integrated since it is covered almost totally by the AHL material within IB Topics 7 and 8 as covered in Chapters 8 and 9 of this book.
- The core syllabus, standard options material and 40 hours of internal assessment covered in year 1 so that S-level students sit the examination at the end of the first year while H-level students continue into year 2 to cover the AHL material, complete the option topics and finish their internal assessment work.

The options

Once the delivery model has been decided on, the next important decision is to select the option topics from those on offer by the IBO (if your model allows for this). Options can be chosen:

- by the teacher;
- by consensus among students following which all students study the chosen options;
- by students individually who then work independently on these options.

In most schools, the teacher decides on the two options to be studied and delivers the material in a similar way to all other sections of the syllabus. The more aware students will be quick to point out that, since they are to be examined on both options in paper 3, the options do, in fact, then become entirely non-optional for the student.

The scheme of work

Once the options have been chosen, the scheme of work and teaching order can be developed. The IB biology syllabus has not been designed as a scheme of work. Each biology team will wish to develop a scheme of work that suits them. However, below are some suggestions and things to think about.

The options are best placed or even integrated within the schemes of work where they exploit links to core issues (Table 22.1).

Option	Relates to topic
A Human nutrition and health	6 Human health and physiology
C Cells and energy	3 The chemistry of life
D Evolution	5 Ecology and evolution
E Neurobiology and behaviour	6 Human health and physiology
H Further human physiology	11 Human health and physiology II

Table 22.1 Options and schemes of work

There are clear links, and a certain linearity in understanding, among the syllabus areas which would probably lead us to start with biochemistry and cellular biology and work upwards towards broader systems and concepts such as ecology or evolution. It would, however, be a mistake to expose students who are new to IB biology to an initial topic which feels unfamiliar. Consequently, many schools agree that cell biology is a comfortable starter. It also makes sense to place the more experimental topics, such as biochemistry, ecology and physiology, across the course but not near to the end since internal assessment will already have been finalised.

Table 22.2 shows two examples a sequence in which the chapters of this book might be studied, one for S-level students and one for H-level students. **These are in no way prescriptive** – each biology team will wish to pioneer an individual approach, and to modify this in the light of experience, no doubt.

A scheme for teaching S-level students IB Diploma biology	A scheme for teaching H-level students IB Diploma biology
1 Cells – the building blocks	1 Cells – the building blocks
2 Chemistry of life	2 Chemistry of life
3 Energy transfer in cells	3 Energy transfer in cells
7 Human physiology, health and reproduction Option A: Human nutrition and health	9 Energy transfer in cells II
6 Ecology, evolution and biodiversity Option E: Neurobiology and behaviour	8 Nucleic acids and proteins
4 Genetics	4 Genetics
5 Genetic engineering and biotechnology	11 Genetics II
	7 Human physiology, health and reproduction
	12 Human physiology, health and reproduction II Option H: Further human physiology
	10 Plant science
	6 Ecology, evolution and biodiversity Option D: Evolution

Table 22.2 Two teaching schemes for IB Diploma biology

Now it is just a matter of finding those eager students to engage in learning.