CAMBRIDGE
UNIVERSITY PRESS

Biology

for the IB Diploma

COURSEBOOK

Brenda Walpole

2023 CAMBRIDGE DEDICATED TEACHER AWARDS

Teachers play an important part in shaping futures.
Our Dedicated Teacher Awards recognise the hard work that teachers put in every day.

Thank you to everyone who nominated this year; we have been inspired and moved by all of your stories. Well done to all of our nominees for your dedication to learning and for inspiring the next generation of thinkers, leaders and innovators.

CONGRATULATIONS TO OUR INCREDIBLE WINNERS!

WINNER Central & Southern Africa
Akeem Badru
St Michael R.C.M Primary School, Ogunpa Lunloye, Nigeria

Regional Winner: East & South Asia
Gaurav Sharma
FirstSteps School, India

Regional Winner: North & South America
Nathalie Roy
Glasgow Middle School, United States

Regional Winner: Australia, New Zealand & South-East Asia
Goh Kok Ming
SJKC Hua Lian 1, Malaysia

Regional Winner: Middle East & North Africa
Uzma Siraj
Future World School, Pakistan

Regional Winner: Europe
Selçuk Yusuf Arslan
Atatürk MTAL, Turkey

For more information about our dedicated teachers and their stories, go to **dedicatedteacher.cambridge.org**

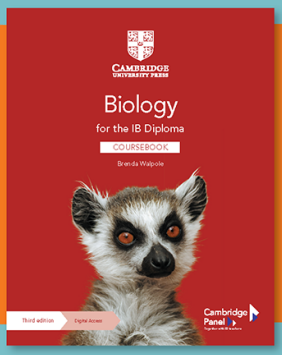
CAMBRIDGE UNIVERSITY PRESS

Brighter Thinking
Better Learning

Contents

How to use this series

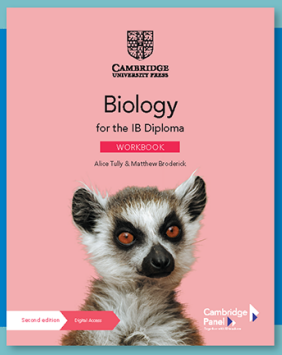
This suite of resources supports students and teachers of the IB Biology Diploma course. All of the books in the series work together to help students develop the necessary knowledge and scientific skills required for this subject.

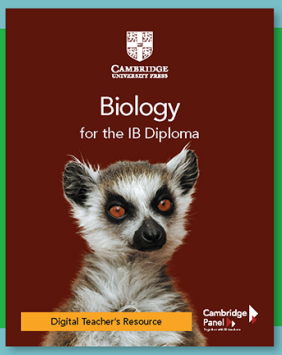


The coursebook with digital access provides full coverage of the latest IB Biology Diploma course.

It clearly explains facts, concepts and practical techniques, and uses real world examples of scientific principles. A wealth of formative questions within each chapter help students develop their understanding, and own their learning. A dedicated chapter in the digital coursebook helps teachers and students unpack the new assessment, while exam-style questions provide essential practice and self-assessment. Answers are provided on Cambridge GO, supporting self-study and home-schooling.

The workbook with digital access builds upon the coursebook with digital access with further exercises and exam-style questions, carefully constructed to help students develop the skills that they need as they progress through their IB Biology Diploma course. The exercises also help students develop understanding of the meaning of various command words used in questions, and provide practice in responding appropriately to these.





The Teacher's resource supports and enhances the coursebook with digital access and the workbook with digital access. This resource includes teaching plans, overviews of required background knowledge, learning objectives and success criteria, common misconceptions, and a wealth of ideas to support lesson planning and delivery, assessment and differentiation. It also includes editable worksheets for vocabulary support and exam practice (with answers) and exemplar PowerPoint presentations, to help plan and deliver the best teaching.

How to use this book

Throughout this book, you will find lots of different features that will help your learning. These are explained below.

UNIT INTRODUCTION

A unit is made up of a number of chapters. The key concepts for each unit are covered throughout the chapters.

LEARNING OBJECTIVES

Each chapter in the book begins with a list of learning objectives. These set the scene for each chapter, help with navigation through the coursebook and indicate the important concepts in each topic.

- A bulleted list at the beginning of each section clearly shows the learning objectives for the section.

GUIDING QUESTIONS

This feature contains questions and activities on subject knowledge you will need before starting this chapter.

The content in this book is divided into Standard and Higher Level material. A vertical line runs down the margin of all Higher Level material, allowing you to easily identify Higher Level from Standard material.

Link

These are a mix of questions and explanation that refer to other chapters or sections of the book.

Key terms are highlighted in **orange bold** font at their first appearance in the book so you can immediately recognise them. At the end of the book, there is a glossary that defines all the key terms.

KEY POINTS

This feature contains important key learning points (facts) to reinforce your understanding and engagement.

EXAM TIP

These short hints contain useful information that will help you tackle the tasks in the exam.

SCIENCE IN CONTEXT

This feature presents real-world examples and applications of the content in a chapter, encouraging you to look further into topics. You will note that some of these features end with questions intended to stimulate further thinking prompting you to consider some of the benefits and problems of these applications.

NATURE OF SCIENCE

Nature of Science is an overarching theme of the IB Biology Diploma course. The theme examines the processes and concepts that are central to scientific endeavour, and how science serves and connects with the wider community.

Throughout the book, there are ‘Nature of Science’ features that discuss particular concepts or discoveries from the point of view of one or more aspects of Nature of Science.

THEORY OF KNOWLEDGE

This section stimulates thought about critical thinking and how we can say we know what we claim to know. You will note that some of these feature end with questions intended to get you thinking and discussing these important Theory of Knowledge issues.

INTERNATIONAL MINDEDNESS

Throughout this Biology for the IB Diploma course, the international mindedness feature highlights international concerns. Science is a truly international endeavour, being practised across all continents, frequently in international or even global partnerships. Many problems that science aims to solve are international and will require globally implemented solutions.

EXTENSION

The feature highlights information in the book that is extension content and is not part of the syllabus.

TEST YOUR UNDERSTANDING

These questions appear within each chapter and help you develop your understanding. The questions can be used as the basis for class discussions or homework assignments. If you

can answer these questions, it means you have understood the important points of a section.

WORKED EXAMPLE

Many worked examples appear throughout the text to help you understand how to tackle different types of questions.

REFLECTION

These questions appear at the end of each chapter. The purpose is for you as a learner to reflect on the development of your skills proficiency and your progress against the objectives. The reflection questions are intended to encourage your critical thinking and inquiry-based learning.

EXAM-STYLE QUESTIONS

Exam-style questions at the end of each chapter provide essential practice and self-assessment. These are signposted in the print coursebook and can be found in the digital version of the coursebook.

SELF-EVALUATION CHECKLIST

These appear at the end of each chapter as a series of statements. You might find it helpful to rate how confident you are for each of these statements when you are revising. You should revisit any topics that you rated 'Needs more work' or 'Almost there'.

I can	Subsection	Needs more	Almost there	Confident to move

		work		on

Free online material

Additional material to support Biology for the IB Diploma course is available online.

This includes Assessment guidance – a dedicated chapter in the digital coursebook helps teachers and students unpack the new assessment and model exam specimen papers. Additionally, answers to the Test your understanding and Exam-style questions are also available.

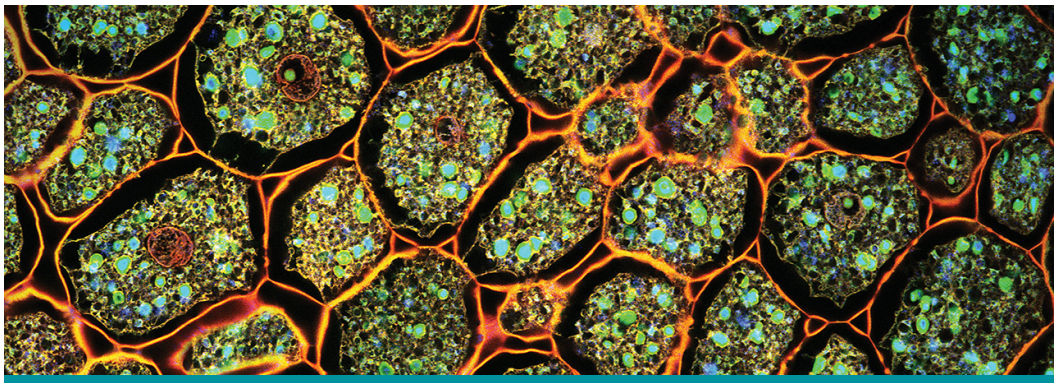
Visit Cambridge GO and register to access these resources.

Unit 1
Molecular organisation

INTRODUCTION

Living things are made up of many different elements and molecules. They are bonded together to build the millions of complex organisms we can study. All life is based on carbon compounds and, as molecules interact in different ways, they build the variety of carbohydrates, proteins, lipids and nucleic acids that we see. These molecules control the composition of the bodies of both single-celled and multicellular organisms.

In this unit we will describe how elements and molecules are arranged and how interactions between molecules form webs of chemical reactions that build up substances or break them down. These reactions are the basis of all the processes of life. Molecules of DNA and protein enable organisms to grow, reproduce and change so that living things can respond to their environments. They also give every living thing its unique characteristics, which can be passed on to the next generation.



Chapter 1

Elements, molecules and water

A1.1, A1.2, B1.1, B1.2

INTRODUCTION

Molecular biology examines the structures and reactions of the chemical substances that are important to life. Living things are composed of many chemical elements. Most elements are bonded together in organic, carbon-containing molecules and compounds. Chemical compounds are divided into two groups: organic and inorganic. Organic compounds include all the complex compounds of carbon that are found in living organisms. The four groups of macromolecules that build

living things are carbohydrates, proteins, lipids and nucleic acids. Water molecules are not organic but they are vital for life and water makes up about 70% of most organisms.

1.1 Elements in living organisms

LEARNING OBJECTIVES

In this section you will:

- learn that organic molecules must contain carbon
- learn that most organic molecules also contain hydrogen and oxygen and small amounts of nitrogen, phosphorus and sulfur

GUIDING QUESTIONS

- Which are the most abundant elements and molecules in living organisms?

1.1.1 Organic molecules

Molecular biology explains the life processes that we observe and all the chemical substances that are involved and the reactions that occur between them. There are almost 100 naturally occurring elements and 25 of these are present in living organisms. The six most common elements in living things are carbon, hydrogen, oxygen, nitrogen, phosphorus and sulfur. Carbon, hydrogen and oxygen are found in all the vital organic compounds: proteins, carbohydrates, nucleic acids and lipids that build up living organisms. Nitrogen is always found in proteins and nucleic acids.

EXAM TIP

A mnemonic is a set of letters that helps us to remember something. The mnemonic CHON is a good way to remember the four key elements found in all organisms: carbon, hydrogen, oxygen and nitrogen.

The **organic compounds** form a vast group that includes gases, liquids and solid substance. Every organic compound in living organisms contains two or more atoms of carbon. As carbon atoms can easily bond with each other, organic compounds can be formed from carbon chains that differ in shape and length. Carbohydrates, proteins, lipids and nucleic acids are the main types of carbon-containing molecules on which life is based (Table 1.1.1).

Molecule	Units	Elements present
Carbohydrate	Sugar monomers	C, H, O

Proteins	Amino acids	C, H, O, N, S
Lipids	Glycerol and fatty acids	C, H, O
Nucleic acid	Nucleotide	C, H, O, N, P

Table 1.1.1: Elements present in different biological molecules.

Any compound that does not contain carbon is an **inorganic compound**. A wide variety of inorganic substances is found in living things and are important to the structure and function of different organisms.

KEY POINTS

inorganic compounds are compounds that do not contain the element carbon.

organic compounds are compounds that do contain carbon atoms. Some compounds contain carbon, but they are not organic compounds. These are carbon dioxide, carbon monoxide and carbonates.

1.1.2 Elements needed in small quantities and larger amounts

The four key elements carbon, oxygen, nitrogen and hydrogen combine and form many thousands of large molecules which make up cell structures and functional molecules (Table 1.1.1). In all organisms, the proportion of these elements is far greater than all others. They are the most abundant (that is, present in the largest quantities) by mass and the number of atoms. Table 1.1.2 shows the percentages of different elements in a human body. About 99% of the human body is made up of the four key elements plus calcium and phosphorus. Five other elements – potassium, sulfur, sodium, chlorine and magnesium – make up just under 1%. These 11 elements are necessary for life.

Element	Symbol	Percentage mass	Percentage atoms
Oxygen	O	65.0	24.0
Carbon	C	18.5	12.0
Hydrogen	H	9.5	62.0
Nitrogen	N	3.2	1.1
Calcium	Ca	1.5	0.22
Phosphorus	P	1.0	0.22
Potassium	K	0.4	0.03
Sulfur	S	0.3	0.38
Sodium	Na	0.2	0.37
Chlorine	Cl	0.2	0.24
Magnesium	Mg	0.2	0.07

Table 1.1.2: The mass and the number of atoms of key elements in the human body, as a percentage of the total mass and the total number of all atoms.

Similar proportions of the same elements occur in all species, from large ocean mammals to tiny single-celled organisms with very different forms. The bodies of all organisms are built from the same essential elements.

1.1.3 Trace elements

The elements iron (Fe), copper (Cu), cobalt (Co), manganese (Mn) and zinc (Zn) are **trace elements** that are needed by animals.

KEY POINT

trace elements are chemical elements that are required only in tiny amounts by living organisms for normal life.

Iron

Iron is a key trace element that almost all living organisms need for their metabolism. It is essential for cell respiration, energy production, DNA synthesis and cell division. Iron also forms part of the hemoglobin molecule (Figure 1.1.1) contained in red blood cells that transport oxygen in the blood of animals (Section 8.3). In mitochondria, which are structures in the cell where energy is released, iron is part of the electron transfer chain that allows eukaryotes to respire (Section 3.2). It is also an important component of both respiratory proteins and enzymes.

Despite the importance of its role, iron forms a tiny proportion of the body's mass. In a human, it accounts for only about 0.006% of body mass.

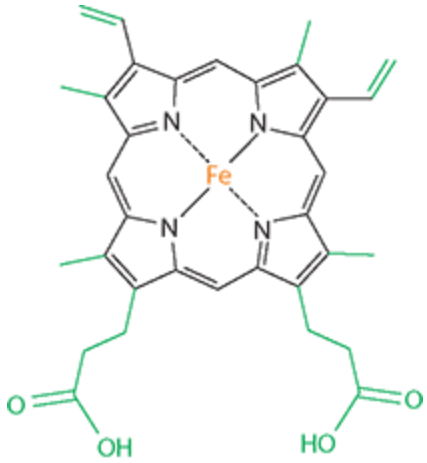


Figure 1.1.1: Hemoglobin contains iron at the centre of its molecule.

Plants also need iron for chlorophyll to be made (or synthesised) and to maintain the structure of their chloroplasts. Almost all the iron in a plant is found in chloroplasts, with a small amount present in the cytoplasm and organelles. Without iron, photosynthesis cannot take place.

Elements' functions in different organisms

All organisms need certain essential elements but the elements may have very different functions in different organisms. For example, plants need iron for photosynthesis, but animals need iron for oxygen transport. Sodium is important in all cell membranes, but it has a special role in sending (or transmitting) the nerve impulses in animals.

Some important roles of inorganic elements are shown in Table 1.1.3.

Element	Example of role in prokaryotes	Example of role in plants	Example of role in animals

sulfur (S)	a component of two amino acids	a component of two amino acids	a component of two amino acids, needed to make some antibodies
calcium (Ca)	co-factor in some enzyme reactions	co-factor in some enzyme reactions	important constituent of bones, needed for muscle contraction
phosphorus (P)	a component of ATP, DNA and phospholipids	a component of ATP, DNA and phospholipids	a component of ATP, DNA and phospholipids
iron (Fe)	a component of cytochrome pigments used in aerobic respiration	a component of cytochrome pigments used in aerobic respiration	a component of hemoglobin and cytochrome pigments used in aerobic respiration
sodium (Na)	important in membranes, changes solute concentration and affects osmosis	important in membranes, changes solute concentration and affects osmosis	important in membranes, changes solute concentration and affects osmosis; also important in transmission of nerve impulses
magnesium (Mg)	Important in ATP and nucleic acid formation	Important in ATP and nucleic acids. In plants,	Important in ATP and nucleic acid formation

		central part of chlorophyll molecules	

Table 1.1.3: Roles of inorganic elements in living organisms.

NATURE OF SCIENCE

The impact of science

The methods used in science can have important ethical implications. This means they can have effects that are considered to be morally wrong. What is considered acceptable by one generation may not accepted by another generation. Experiments carried out in 1928 made the important discovery that copper was an essential trace element needed by living things. Scientists fed various diets to rats. The diets contained, or did not contain, different elements. Without iron the rats were unable to produce red blood cells. But with a diet that contained iron but which did not include copper, the rats became anemic, meaning that they did not have enough red blood cells to carry oxygen in their blood. When copper was added to the rats' diet, the animals recovered.

To consider:

1 Do you think experiments like this would be considered ethical today?

2 Discuss the arguments for and against such experiments.

1.1.4 Toxicity of some elements

Some essential trace elements become toxic if organisms ingest them in high doses. Heavy metals are among the most toxic elements. The most dangerous (or hazardous) include chromium (Cr), nickel (Ni), copper (Cu) and zinc (Zn). Heavy metals are released from natural sources during the weathering of rocks or volcanic eruptions, but industrial processes and agriculture have increased the amounts in the environment. **Toxic elements** remain in the environment and can accumulate, or build up, in the bodies of organisms if they are transferred through food chains and webs (Sections 12.1.1 and 12.2.2).

TEST YOUR UNDERSTANDING

1 What are the four key elements that are found in living organisms?

2 List two functions for elements that are required in very small quantities by organisms.

3 What are the main sources of toxic elements that pollute the environment?

KEY POINT

toxic elements are substances that contaminate drinking water, food and the air, making them poisonous or harmful.

Links

- How are molecules changed through metabolic processes? (Chapter 3)
- Which elements and molecules do all living things need? (Chapter 4)

1.2 Water

LEARNING OBJECTIVES

In this section you will:

- learn that water is a polar, covalent molecule that can form hydrogen bonds between its molecules
- understand that water is an important solvent and transport medium
- learn that blood and plant sap are mainly water and that they are used to transport substances in animals and plants
- describe how water has emergent properties
- learn that water has cohesive properties, which create surface tension
- learn that water has adhesive properties that are important at exchange surfaces and in plant transport
- understand that the thermal properties of water are important in controlling the temperatures of living organisms
- understand why ice has a lower density than water and has insulating properties
- recall that water is transparent and provides a habitat for living things
- learn that water is a metabolite in biochemical reactions.

GUIDING QUESTIONS

- How do the structure and properties of water make it essential for life?

1.2.1 The structure of water

Hydrogen bonds

Covalent bonds form between the hydrogen atoms and oxygen atoms in water molecules. Each of these bonds contains a shared pair of electrons. But the atoms do not share these electrons equally. Oxygen has a greater pull on the shared electrons because it has more protons and therefore a greater positive charge in its nucleus. The oxygen area of the molecule is slightly more negative than the hydrogen end.

In diagrams that represent water molecules, the slight difference in charge is shown by the Greek letter delta (δ).

Figure 1.2.1: Diagram of a single water molecule.

Water molecules are unusual because they have a small positive charge on the two hydrogen atoms and a small negative charge on the oxygen atom. Because of this arrangement, water is said to be a polar molecule.

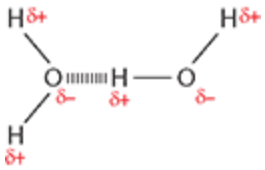


Figure 1.2.2: Hydrogen bond joins two water molecules.

KEY POINTS

covalent bonds are chemical bonds that involve the sharing of electron pairs between atoms.

hydrogen bonds are bonds that form between water molecules because they have polarity.

polar molecules are molecules that have an unevenly distributed electrical charge so that there is a positive region and a negative region. Water, sugars and amino acids are polar molecules.

A weak bond can form between the negative charge of one water molecule and the positive charge of another, as shown in Figure 1.2.2. This type of bond, known as a **hydrogen bond**, is responsible for many of the properties of water. The bond is known as an intermolecular hydrogen bond because it forms between the same types of molecules in the same substance. Hydrogen bonds are constantly forming and reforming between water molecules, which gives water its fluid property.

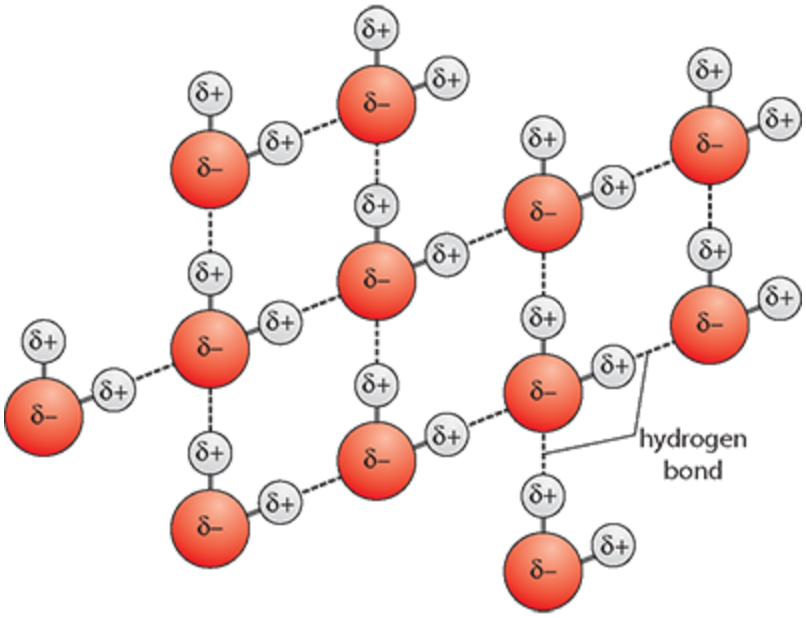


Figure 1.2.3: Hydrogen bonding in water.

1.2.2 Solvent properties of water

Water is sometimes known as a universal **solvent**. Its polarity makes it an excellent solvent for other **polar molecules**. Most inorganic ions, such as sodium, potassium and chloride ions, dissolve well in water. This is because the positive or negative charges of the ions are attracted to the charges of water molecules (Figure 1.2.4). In liquid water many water molecules are bonded together, but approximately 20% are free and able to bond with other chemical substances.

KEY POINT

solvent a substance that will dissolve a solute; for example, water is a solvent in which salt (a solute) will dissolve.

The positive and negative charges of water molecules attract ions with negative or positive charges. This means that the ions position, or orientate, themselves between water molecules and dissolve.

Polar organic molecules, such as amino acids and sugars, are also soluble in water. Water is the medium in which most biochemical reactions take place. This is because almost all substances involved dissolve well in water. Protein synthesis and most of the reactions of photosynthesis and respiration take place in an aqueous (water) solution.

Substances are classified into two groups according to their solubility in water. **Hydrophilic** substances such as sugars and salts dissolve easily.

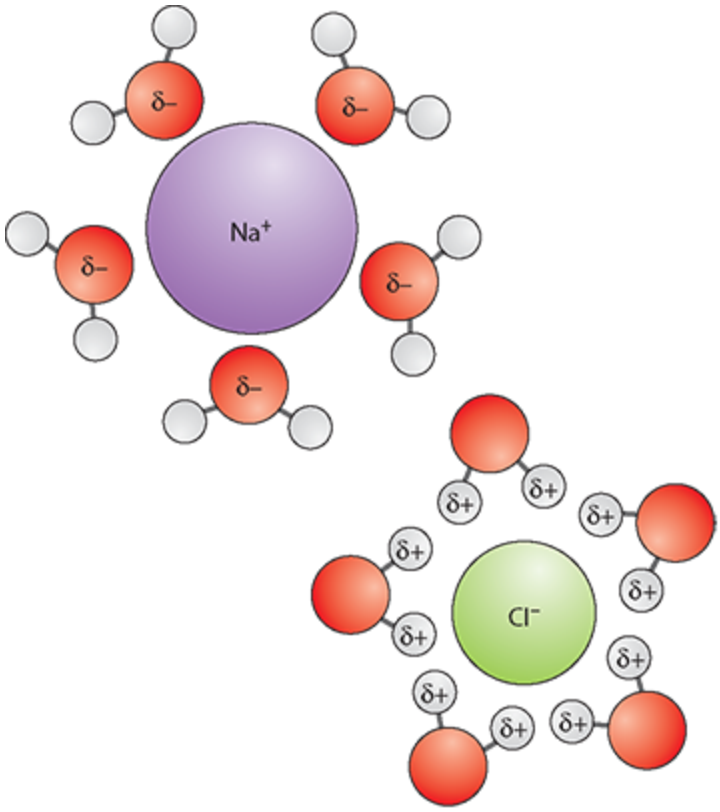


Figure 1.2.4: Diagram showing the orientation of water molecules around a positively charged sodium ion and a negatively charged chloride ion.

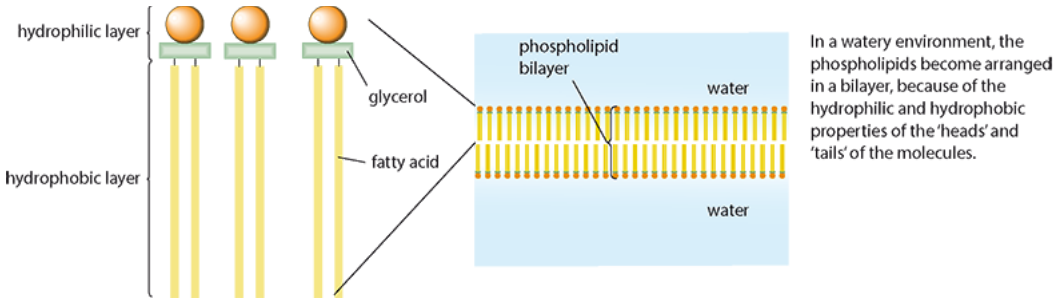


Figure 1.2.5: A phospholipid molecule includes a phosphate, glycerol and two fatty acids. In diagrams the phospholipid molecule is often simplified and shown as a circle with two tails.

THEORY OF KNOWLEDGE

The ‘memory’ of water

Homeopathy is a form of complementary medicine, which means it is not a treatment offered by conventional modern medicine. Homeopathy began in the 1790s, when a German doctor, Samuel Hahnemann, proposed the idea of ‘like cures like’ and suggested that a substance that causes symptoms can be used to treat those symptoms. A central principle of homeopathy is a process of dilution.

The ‘memory of water’ is a phrase that is usually associated with homeopathy. It was coined by Jacques Benveniste (1935–2004) who claimed that water retained a memory of substances that had been dissolved in it. Homeopathic remedies are prepared by diluting ingredients to such a low dilution that, in some cases, no molecules of the original substance are found in the solution. There is no scientific evidence to support the claim that water has a memory. The subject of homeopathy is controversial and many scientists reject it completely.

To consider:

What criteria can be used to distinguish scientific claims from false, or pseudoscientific, claims?

EXAM TIP

Make sure you can draw a single water molecule and its charges, and a group of molecules hydrogen-bonded together.

Amino acids with polar side groups also dissolve well. **Hydrophobic** substances do not dissolve in water. Hydrophobic

substances are usually uncharged, and examples include fats and oils, cholesterol and some large proteins.

Non-polar substances are not very soluble in water because water molecules would rather remain hydrogen-bonded to each other, than to allow non-polar molecules to come between them. Non-polar substances such as cholesterol are packaged in spherical particles called lipoprotein (see the section on Water as a transport medium). This allows them to be carried in the blood.

KEY POINTS

hydrophobic refers to water hating substances.

hydrophilic are water loving substances.

Water as a metabolite

Most metabolic reactions take place in water because more kinds of substance dissolve in water than any other liquid, and it is the most effective solvent we know. **Metabolites** are substances that are formed in or are needed for metabolism. Metabolites are found in water, but water is also a reactant in photosynthesis and a product of respiration. Water is needed for hydrolysis reactions such as digestion to take place and also is produced during condensation reactions (Section 1.3).

Water as a transport medium

As water is such an excellent solvent, substances can be transported in solution around the bodies of animals and around plants. Water also has a low viscosity, which means that it flows easily through narrow tubes such as blood capillaries or the xylem of a plant.

Vertebrates all have blood that is more than 90% water and which is enclosed in a system of blood vessels. Invertebrates, such as insects, crustaceans and many molluscs, have a type of blood known as hemolymph. This fluid, which is also predominately water, is not kept in vessels but flows around the animals' bodies and into the blood spaces of their transport systems.

Blood is ideal to carry many dissolved soluble solutes, such as glucose, sodium and chloride ions, amino acids and vitamins, from the **digestive system** to other organs of the body. Blood can also transport dissolved nitrogenous waste and hormones.

Gases, such as oxygen and carbon dioxide, are not very soluble in water because they are essentially non-polar. Oxygen can dissolve in water, which allows aquatic life to exist, but its solubility is very low. Oxygen and carbon dioxide are carried in the blood, but must be held by (or bound to) hemoglobin or other pigments, or converted to soluble bicarbonate ions, to be carried in blood plasma (Section 8.3).

Many large molecules, such as lipids and proteins (Sections 1.3 and 1.4), are also mostly non-polar but can be carried through the aqueous environment of the blood. For these molecules to be sufficiently soluble, they must have some polar groups exposed on the outside of the molecule. Polar groups on the outside of soluble proteins interact with polar water molecules and make the entire protein soluble.

Cholesterol is only slightly soluble in water and dissolves in the blood in very small amounts. For this reason, cholesterol is transported by the circulatory system in lipoproteins, which are small spheres that have an outer surface made up of **amphipathic** proteins and lipids (Figure 1.2.6). The outward-

facing surfaces of these molecules are water-soluble and their inward-facing surfaces are lipid-soluble. Triglycerides (fats) are carried inside them while **phospholipids** and cholesterol, being amphipathic, are transported in the surface layer of the lipoprotein particle.

KEY POINT

amphipathic a molecule that has both polar and non-polar regions which determine how it interacts with other molecules. Phospholipids that build membranes have non-polar tails and polar heads.

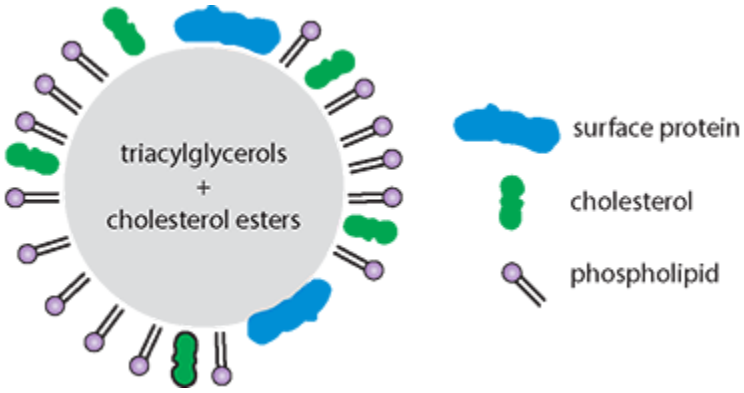


Figure 1.2.6: Simple diagram of a lipoprotein.

Substance	Solubility	Properties
sodium chloride	very soluble	ions are attracted to water molecules
glucose	very soluble	polar molecule
amino acids	very soluble	have both positive and negative charges
oxygen	low	if bound to hemoglobin it can travel in

	solubility	blood
fats and cholesterol	insoluble	fats are non-polar and cholesterol is only slightly charged. To be carried in blood fats must travel in lipoprotein complexes

Table 1.2.1: Table summarising the solubility of important molecules.

The transport system of a plant includes two types of thin, tube-like vessels called the xylem and the phloem. These tubes contain a fluid called sap, which is mostly water. The xylem carries dissolved minerals from the roots to the leaves, and the phloem transports soluble sugars up and down the plant. Liquid in the xylem contains less than 1% solutes, mostly organic acids and mineral salts. Phloem sap may contain up to 25% solutes, mainly sucrose and amino acids.

1.2.3 The physical properties of water

Emergent properties are defined as properties of a complex system that arise from simple interactions of individual component parts. In the case of water, these properties are due to interactions between individual molecules.

KEY POINT

emergent properties of water these properties are due to interactions between individual molecules.

The polar properties of water molecules, which are joined by hydrogen bonds, give water important emergent properties including:

- Cohesion of its molecules to one another
- Adhesion of its molecules to other molecules
- High specific heat capacity
- Surface tension

These properties have important consequences for many organisms, especially those which live in aquatic habitats.

1.2.4 Cohesion and its impact on organisms

Cohesion

Hydrogen bonding between water molecules holds them together in a network, resulting in a phenomenon known as cohesion (Figure 1.2.7), which gives water many of its biologically important properties. **Cohesive forces** allow water to form droplets and are also responsible for surface tension.

In the xylem of a plant, water can be drawn up inside the stem because cohesive forces keep the water together as a continuous column (Figure 1.2.7). Strong pulling forces are produced as water evaporates from the leaves at the top of tall trees. This draws water and dissolved minerals great distances up to the tips of branches high above the ground.

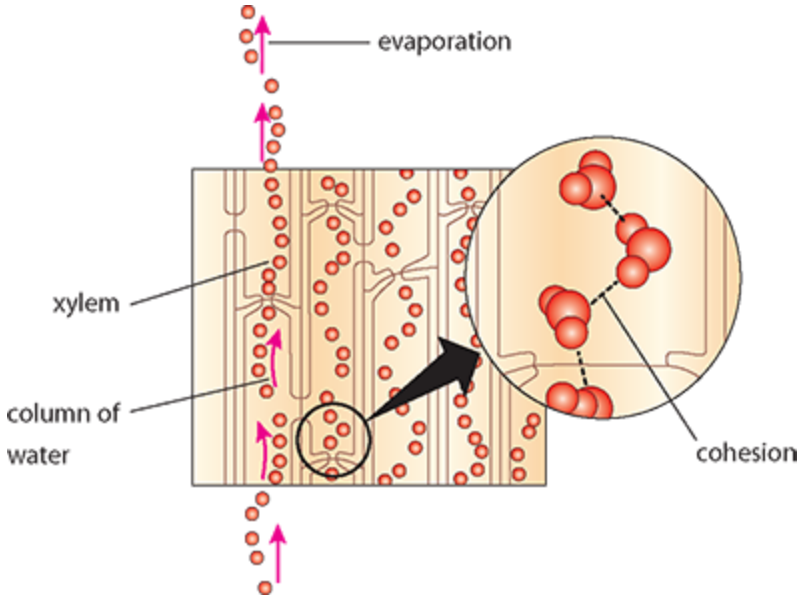


Figure 1.2.7: Cohesive forces hold water molecules in a column.

Cohesion is also responsible for surface tension. At the surface of water where it meets the air, there is a greater attraction of water molecules to one another than to the air. This forms a strong surface which enables some small organisms to walk on water. This cohesion contributes to the thermal properties of water too (Figure 1.2.8).

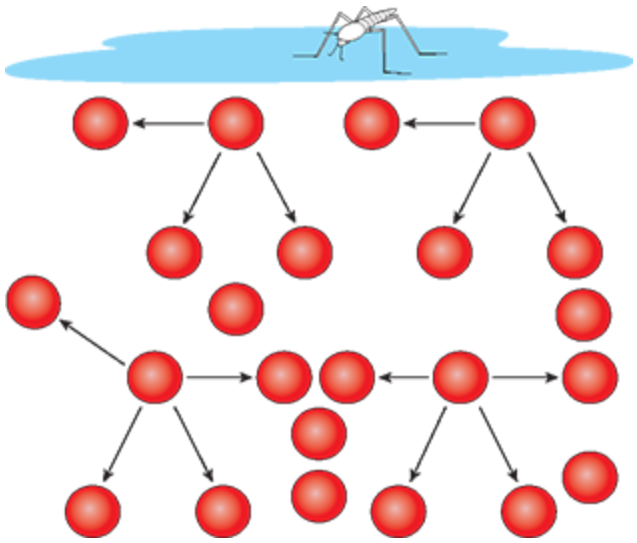


Figure 1.2.8: Surface tension attracts water molecules to one another. Pond skaters use long slender legs to distribute their body mass evenly and have tiny hydrophobic hairs on their legs.

Adhesion

Water is attracted to other polar or charged molecules. There are forces of attraction, known as **adhesive forces**, which occur between water molecules and different molecules in vessels that contain the water. This means that water tends to be attracted to and stick, or adhere, to the walls of its container. The surface of glass is polar, so it forms adhesive bonds with water.

Adhesion of water to the sides of a tube or container creates an upward force on the edges of the liquid. For this reason the surface of the water isn't flat. Instead, the surface curves

downwards towards the centre of the tube, which is called a meniscus. This is because surface tension keeps the surface intact. If the adhesive force to the sides of a narrow tube is stronger than the cohesive force between molecules, water will be drawn upward against the pull of gravity. This is known as **capillary action**.

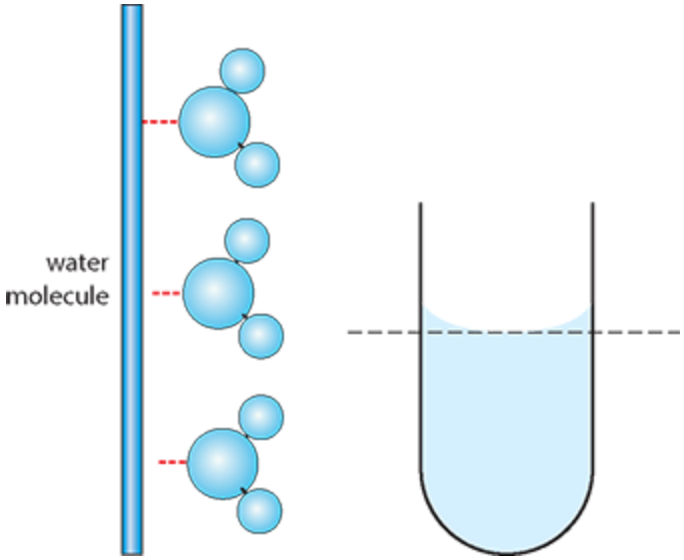


Figure 1.2.9: Adhesive forces attract water molecules to the sides of a glass container.

Water is drawn through plant cell walls and into the roots of a plant by capillary action. Adhesive forces are also important as they attract water molecules to the sides of the xylem. These forces allow water to be draw up the stem of a plant. Adhesive forces are greater in a narrow tube because relatively more water molecules are in contact with the sides. This means that adhesive forces are able to 'hold up' and support a substantial mass of water in the fine xylem vessels. Cohesive forces hold the water column together.

1.2.5 Thermal properties of water

Water is liquid at most temperatures at which life exists so it forms a useful habitat for living things to live in and on. Water and air are both heated through convection currents and radiation from the sun, but the ability of water to absorb heat is far greater than that of air. It is said to have a high specific heat capacity. This makes an aquatic habitat a far more stable environment for living things because the temperature of water varies less than the temperature of air.

KEY POINT

specific heat capacity is the amount of heat energy needed to raise the temperature of 1 cm^3 of water by 1 °C.

Thermal properties explained

Thermal properties are properties that materials or liquids have that are related to their ability to conduct heat. Water has special thermal properties. Water is unusual among small molecules because it is a liquid at most temperatures found on Earth. A large amount of energy is needed to break the many weak hydrogen bonds between the water molecules. This gives water a high **specific heat capacity** and means that water can absorb or give out a great deal of heat energy without its temperature changing very much. A stable temperature is important to living things because the range of temperatures in which biological reactions can occur is quite narrow. The thermal properties of water allow it to keep an organism's temperature fairly constant. In the body, water can act as a temperature regulator. Water is a major component of blood, which carries heat from warmer parts

of the body, such as the liver, to cooler parts such as the feet or to the skin where is lost as radiant heat.

For liquid water to evaporate and become vapour, many hydrogen bonds between the molecules must be broken, so evaporation requires a lot of energy. As a result, water is a liquid at most temperatures found on Earth, and it has a high boiling point. When water evaporates, it carries a great deal of heat with it. For example, when sweat evaporates from the skin surface of a mammal, the process of evaporation acts cools the mammal's body.

A lot of heat must be removed from water before it freezes. This means that cell contents of organisms and the water in aquatic environments will not freeze easily in cold conditions.

Special properties of ice

At temperatures greater than 0 °C hydrogen bonds between water molecules are easily made and broken because water is a liquid and molecules move freely within in. When the temperature falls, ice crystals form because the hydrogen bonds become permanent. Ice has a hexagonal framework of molecules with an open structure. Ice takes up more space than the volume of water that formed it, so ice is less dense than water and floats on the surface of ponds, lakes and the ocean. A layer of ice on the surface insulates the water beneath it, keeping it warmer than the cold environmental conditions above the ice. The water beneath the ice is protected from freezing, so plants and animals can survive in the slightly warmer water beneath the ice.

The properties of water are summarised in Table 1.2.2.

Property	**Reason**	**Consequence/benefits to living organisms**

cohesion	Hydrogen bonds hold water molecules together.	Water can travel in continuous columns – for example, in the stems of plants – and act as a transport medium.
adhesion	Water molecules are attracted to other different molecules.	A column of water can be held up in the narrow xylem of a plant.
solvent	The polar molecules of water can interact with other polar molecules and ions.	Ions dissolve easily. Large molecules with polar side groups, such as carbohydrates and proteins, can also dissolve. So water acts as an excellent transport medium and as a medium for metabolic reactions.
thermal	Water has a high heat capacity. Large amounts of energy are needed to break hydrogen bonds and change its temperature.	The temperature of organisms tends to change slowly. Fluids such as blood can transport heat round their bodies.
	Water has a high boiling point compared with other solvents because hydrogen bonds need large amounts of	Water is liquid at most temperatures at which life exists, so is a useful medium for metabolic reactions.

	energy to break them.	
	Water evaporates as hydrogen bonds are broken and heat is taken from the water.	Sweating and transpiration enable animals and plants to lose heat. Water acts as a coolant.
viscosity	Water has a low viscosity, molecules slide easily past one another.	Water flows easily through tiny capillaries and other very small spaces such as spaces in cell walls or in the soil.
transparency	Water allows light to pass through it.	Light can reach plants and animals below the surface of water. Plants can photosynthesise and animals such as fish, birds and seals can hunt using sight.
buoyancy	Water exerts an upthrust (force) which opposes the weight of a partially submerged body.	Water can support the weight of floating and submerged organisms, providing a habitat and also reducing the need for large supporting structures such as skeletons.

Table 1.2.2: The properties of water

1.2.6 Life on water, land and in the air

All species have body shapes and lifestyles that adapt them to the habitats in which they live. A few species live on land, in the air and in water and must adapt to life in all three areas. One bird that does this is the Black-throated loon (*Gavia arctica*) Figure 1.2.10. It is a large, streamlined diving bird that can float on water and easily dive to a depth of 5 m to hunt for fish. The bird's body shape is well-suited to diving and they have waterproof feathers for protection. They also have webbed feet with toes joined by a membrane which are suited to swimming. They nest on land around sheltered coasts in cold and temperate Arctic areas but out of water. They walk awkwardly because their legs are so far back on their bodies. Black throated loons are large birds and must run over water to take off and fly, even though they have a large wingspan, the wing surface area is small compared to its body size.

Figure 1.2.10: The black-throated loon develops black feathers on its throat in the breeding season.

The ringed seal (Figure 1.2.11) is a small earless seal that lives in the Arctic and sub-Arctic regions. Ringed seals live in frozen waters and are well-adapted to ice-covered areas. Water in various forms is essential to the seals which use liquid water to hunt in and solid water (ice) to live in and on. Seals are mammals and need to breathe air above the ice. They have strong claws on their front flippers that they use to cut breathing holes through the ice. Ringed seals stay on the ice most of the year; their streamlined bodies allow them to slide across the solid ice and swim underneath it to hunt. Thick layers of fat protect them from the cold. They dig snow caves for breeding in late winter and spring. The ice and snow caves provide protection for their young from extreme cold and polar bears which are their main predators. Loss of sea ice and snow cover on the ice poses the main threat to this species.

Figure 1.2.11: Ringed seal (*Pusa hispida*) streamlining, thick blubber and strong claws adapt this species to life on the ice

NATURE OF SCIENCE

Using theories and models to explain the properties of water

Scientific models are developed to explain patterns and observations. Models cannot be proven but are useful to explain processes that are not directly observable.

The properties water has are due to its molecular structure and energy. The detail about how molecules in water interact is a question that has been studied by scientists for many years. Techniques including infrared absorption, neutron scattering and nuclear magnetic resonance imaging (NMRI) have all been used to study the structure of water. The results along with data from theoretical calculations have led to models, theories and computer simulations that try to describe the structure of water and explain its properties.

Observations have shown that water is a small, simple molecule (H_2O) in which each hydrogen atom is covalently bonded to the central oxygen atom by a pair of electrons that are shared between them (Figure 1.2.12). Only two of the six outer-shell electrons of each oxygen atom are used to form these covalent bonds, leaving four electrons in two non-bonding pairs. These non-bonding pairs remain closer to the oxygen atom and exert a strong repulsion against the two covalently bonded pairs. The two hydrogen atoms are pushed closer together. Overall, water molecules are electrically neutral, but this model of the water molecule results in small positive and negative charges unevenly distributed over the molecule. When H_2O molecules are crowded together in liquid water, the forces between the atoms produce the properties of water that we see. Properties that are unique to

water arise from the cage-like, tetrahedral hydrogen bonding among molecules that are next to each other.

New models of water are now being designed to investigate the structure of water when it contains chemicals and biomolecules. Molecular models may help us create new technologies for producing clean water, reclaiming polluted water, predicting climate and designing new drugs to cure diseases.

To consider:

1 How are models used to help scientists explain evidence that is observed or recorded?

2 How do models change when new techniques or discoveries are made?

3 Why do you think that the detail of the structure of water has been so difficult to understand?

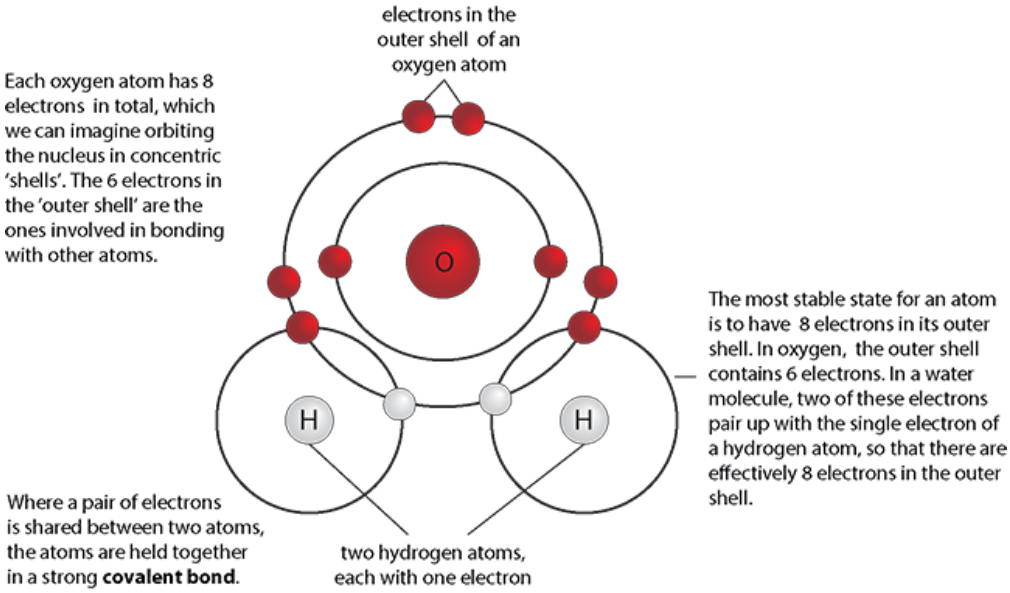


Figure 1.2.12: The two hydrogen atoms in a water molecule are pushed together on one side because of the repulsive effect of the two pairs of non-bonding electrons in the outer shell of the oxygen atom.

1.2.7 Origins of water on Earth

Most scientists believe that the water we have on Earth today originated from outside our planet. Evidence for the origin of water has come mainly from the study of material found in asteroids. Recent studies of an asteroid called 25143 Itokawa that was brought to Earth by a Japanese robot probe Hayabusa have used atom-probe tomography to examine atoms in grains of dust from the asteroid and the results suggest that at least half the water on Earth has come from interplanetary dust. The grains from the asteroid contained water that was probably created by solar wind, streams of particles that flow from the Sun. Over billions of years, as the solar system developed, the particles in the solar winds interacted with oxygen atoms in dust clouds of the solar system. As Earth orbited the Sun and passed through the clouds, water molecules slowly accumulated around and on its surface, held down by Earth's gravity. It is unlikely that all the water on Earth arrived in this way; an equally important source would probably have been the ice in comets and asteroids that crashed on Earth's surface. Evidence for this theory comes the fact that ice from comets and asteroids contain higher amounts of the hydrogen isotope deuterium compared with water on Earth, but solar dust contains relatively low levels of the isotope. If we consider the two sources together we find that the isotope balance matches that of Earth's water.

Water has remained on Earth for two important reasons; the first is the pull of Earth's gravity which retains not only water, but also water vapour drawing it back to the surface and down into reservoirs and oceans. The second reason is temperature. Temperatures are cold enough for water to condense and certain regions of the atmosphere are very cold, at high altitude the

temperature can be -60o C, so that water forms crystals of ice that fall back to Earth's surface.

Over billions of years of our planet's history water has allowed life to evolve on Earth. But our understanding of water's origins also suggests that other parts of the solar system could have water, possibly as ice and this knowledge is important in the search for life in other parts of the galaxy. An ice deposit is believed to have formed near the south pole of our Moon and NASA is aiming to explore this in the Artemis programme.

TEST YOUR UNDERSTANDING

4 Describe the significance of water to living organisms.

5 Define the term 'hydrophilic'.

6 Explain how hydrogen bonding affects the force of cohesion.

7 Discuss three key properties of water that make it an ideal habitat for living organisms.

8 Why is sweat a good coolant for the body?

- **a** The small blood vessels that transfer water to sweat move closer to the skin surface when it is hot
- **b** Breaking H bonds between water molecules in sweat requires energy from body heat
- **c** Sweat contains minerals such as sodium chloride
- **d** Sweat is non-polar.

9 Explain what is meant by The Goldilocks Zone.

Extra terrestrial life and water

So far Earth is the only planet with life as we know it but there are potentially habitable worlds elsewhere. The ‘Goldilocks Zone’ or habitable zone is a term that astronomers use to narrow the search for such worlds. The zone was given its name from the children’s fairy tale in which Goldilocks looked for porridge that was ‘just right’ not too hot and not too cold. Scientists define The Goldilocks Zone as the range of distance from the Sun, or other star, that an object can be before water on its surface boils away or freezes. Planets in this zone have the right temperature for water to remain liquid. There are 1,780 planets beyond our solar system that we know of. Of these about 16 are located in their star’s habitable zone where conditions are not too hot or too cold to support life. A planet must also be just the right size. Too large and the atmosphere is too compacting, whilst a planet that is too small cannot maintain an atmosphere. A recently discovered planet Kepler-186f, is close to the size of Earth and orbits in its solar system’s habitable zone. It is 493 light years from Earth.

Links

- How does the solubility of oxygen affect the distribution of aquatic animals? (Chapter 12)
- What are the roles of water in photosynthesis and respiration? (Chapter 3)
- How do the properties of water influence temperature control in animals? (Chapter 8)

1.3 Organic molecules in living organisms

LEARNING OBJECTIVES

In this section you will:

- learn that carbon atoms can form four covalent bonds and they can produce a wide range of different stable compounds
- understand that life is based on carbon compounds including carbohydrates, proteins, lipids and nucleic acids
- define monomers as small units that are built up into polymers
- define polymers as molecules that are built up during condensation reactions and broken down by hydrolysis
- describe glycerol and fatty acids as components of triglycerides built up by condensation reactions and triglycerides as compounds that are broken down during hydrolysis reactions
- learn that all living organisms use nucleotides to synthesise nucleic acids

> learn that functional groups such as phosphates, amines and carboxyl groups are found in many molecules and give them specific properties

- understand that the arrangement of bonds around a carbon atom allows for structural isomers with different three-dimensional forms to be made.

GUIDING QUESTIONS

- Which types of molecules are found in all living organisms?
- How are small molecules made into large ones in living organisms?
- How is carbon essential to all living organisms?

1.3.1 The importance of carbon atoms

Carbon is found in all organic molecules and forms a wide range of different compounds. Figure 1.3.1 shows how other elements can be added to a single carbon atom in one of four directions. In this way, complex three-dimensional molecules can be built up.

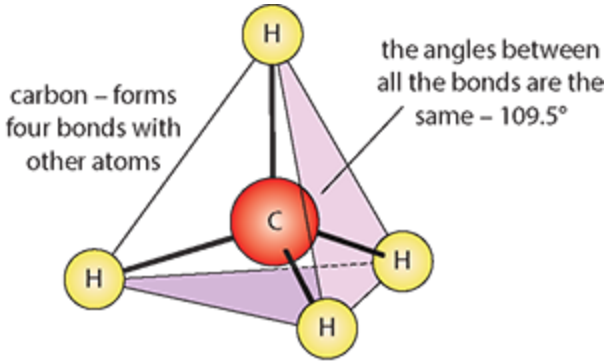


Figure 1.3.1: Hydrogen atoms (H) can bond to a single carbon atom (C) in four different directions.

Every organic compound contains two or more atoms of carbon. Carbon atoms are joined or attached (bonded) together by covalent bonds. A carbon atom can form four covalent bonds. Carbon compounds that differ in shape and length can be made, and in this way different organic molecules are created. Carbon atoms can also form double and triple bonds with other atoms. This increases the variety seen in the molecular structure of organic compounds.

KEY POINT

covalent bond is the type of bond that joins together non-metal atoms. The outer electrons of each atom are shared to fill the outer shell and make the molecule stable.

1.3.2 Carbon compounds: the building blocks of life

Carbon compounds – that is, carbohydrates, lipids, proteins and nucleic acids – form the basic molecules for all life. Carbohydrates are compounds that usually contain only the elements carbon, hydrogen and oxygen and are the most abundant group of biological molecules. Lipids contain the same three elements but with much less oxygen than a carbohydrate of the same size. Lipids may also contain small amounts of other elements such as phosphorus. Proteins, unlike carbohydrates and lipids, always contain nitrogen. Sulfur, phosphorus and other elements are also often present. Nucleotides always contain nitrogen and phosphorus (Table 1.1.1).

Many organic molecules are very large and complex but they are built up of smaller parts (subunits), which can be relatively simple. Figure 1.3.2 shows some of these building blocks. Small subunits called monomers are built into larger complex molecules called polymers in a process known as polymerisation.

Carbohydrates

Carbohydrates are the most abundant type of molecule in living things. In both plants and animals, carbohydrates have an important role as a source of energy, and in plants they also have a structural function. Carbohydrates occur in different forms. **Monosaccharides**, with the general formula $(CH_2O)_n$, where n = the number of carbon atoms in the molecule, are monomers; that is, single sugars made up of just one subunit. **Disaccharides** are sugars that have two subunits joined together by a condensation reaction (Section 1.3.3) and **polysaccharides** are long molecules consisting of a chain of monosaccharides linked together.

Proteins

Proteins are built up of units called amino acids. The atoms occurring at the fourth bond (known as the R group; see Figure 1.3.2) differ in different amino acids and give each one its own properties. The simplest amino acid is glycine, in which R is a hydrogen atom, whereas the R group in the amino acid alanine is CH_3. There are more than 100 naturally occurring amino acids but only 20 are used in the bodies of living things. Amino acids bond together in condensation reactions (Section 1.3.3) to form polypeptide chains that, in turn, are built up into proteins.

Lipids

Lipids are fats, oils, waxes and steroids. All are organic compounds that are insoluble in water but they do dissolve in organic solvents such as ethanol. An important role of lipids in living organisms is as energy storage molecules. One group known as **triglyceride** lipids includes fats and oils. Solid lipids are generally referred to as fats, whereas lipids that are liquid are known as oils. Animals store energy as fat, whereas plants store oils. Examples of plant oils include sunflower oil and olive oil. Lipid contains about twice as much energy per gram as carbohydrates such as starch, but each type of storage molecule has its own advantages. The second group of lipids includes steroids (Figure 1.3.3), which consist of four linked rings of carbon atoms. Vitamin D and cholesterol are two well known examples of steroids.

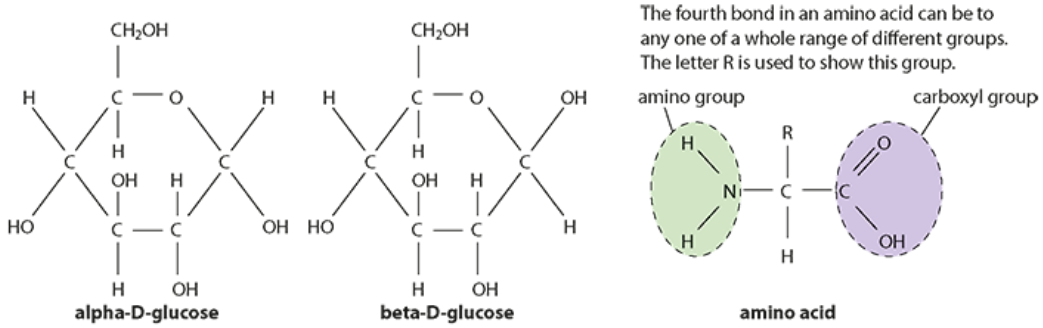
KEY POINTS

monomer refers to a small molecule that can bond to other similar small molecules, to make up repeating chains that form larger polymers.

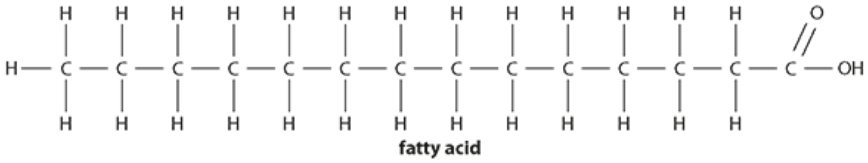
polymer is a large, complex molecule built up of a series of monomers. Formed by condensation reactions in a process called polymerisation.

Nucleic acids

Nucleic acids are found in all living cells and viruses. Two types of nucleic acid found in cells are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) (Figure 1.3.4). DNA is found in the nucleus, mitochondria and chloroplasts of eukaryotes, whereas RNA is usually found in the cytoplasm. Nucleic acids are long molecules consisting of chains of units called nucleotides. Each nucleotide consists of a pentose sugar (which is ribose in RNA and deoxyribose in DNA) and the pentose is linked to phosphoric acid and an organic base. Nucleic acid chains are longer than those found in proteins. They are vital to inheritance and development; these aspects are discussed in Chapter 5.



The alpha and beta forms of glucose are readily interconvertible, but molecules made from the different forms have different properties. Cellulose is made of beta glucose molecules while amylose is a polymer of alpha glucose molecules.



This fatty acid is saturated because there are no double bonds between the 14 carbon atoms in the chain.

Most naturally occurring saturated fatty acids have between 4 and 28 carbon atoms.

Figure 1.3.2: The basic structures of glucose, amino acids and fatty acids.

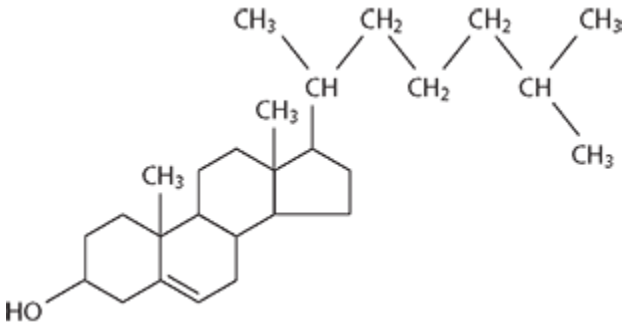


Figure 1.3.3: Like all steroids, cholesterol has four rings of carbon atoms. Other steroids differ in the side groups attached to them.

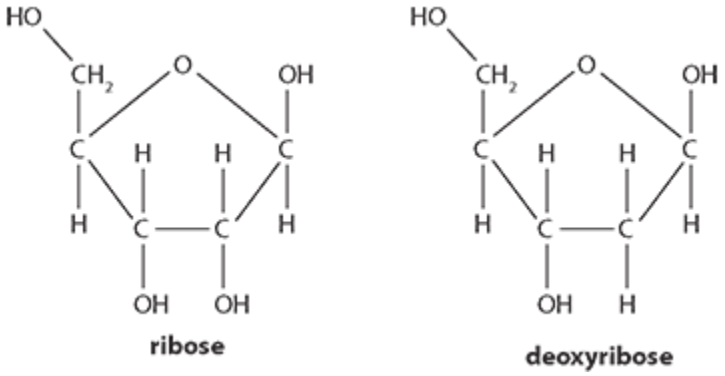


Figure 1.3.4: Ribose and deoxyribose are the building blocks of nucleic acids.

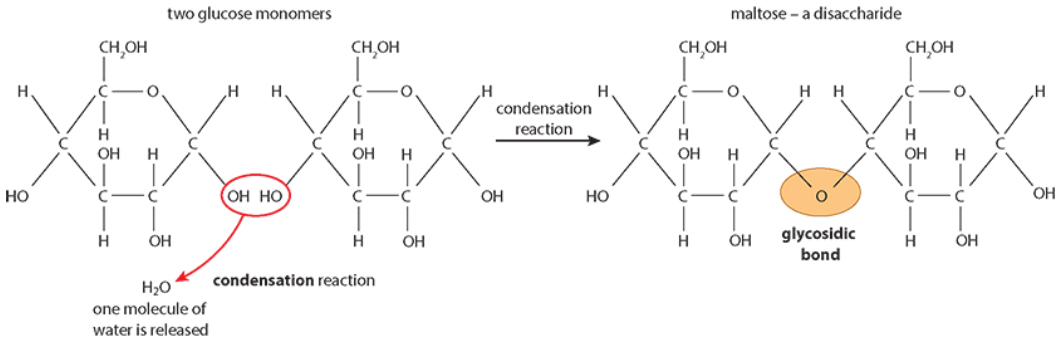


Figure 1.3.5: Monosaccharide subunits (glucose in this case) are joined in a condensation reaction, forming a disaccharide (maltose) and water.

1.3.3 Monomers and polymers

Condensation: building larger molecules

In a **condensation reaction**, two molecules are joined together by strong covalent bonds to form a larger molecule. Condensation is an example of an **anabolic** reaction, which builds up monomers to form macromolecules. Each condensation reaction requires an enzyme to catalyse the process and it produces one molecule of water. The condensation of two monosaccharide monomers produces a disaccharide. For example:

glucose	+	galactose	→	lactose + water
(monosaccharide)		(monosaccharide)		(disaccharide)

KEY POINTS

condensation reaction refers to an addition reaction in which two molecules combine to form a single molecule, and water is removed and released.

anabolic is a type of reaction in which large molecules are built up from small ones.

If further monosaccharides are added to a disaccharide, a polysaccharide is formed.

In a similar way, two amino acids can be linked to form a **dipeptide** (Figure 1.3.6):

amino acid + amino acid → dipeptide + water

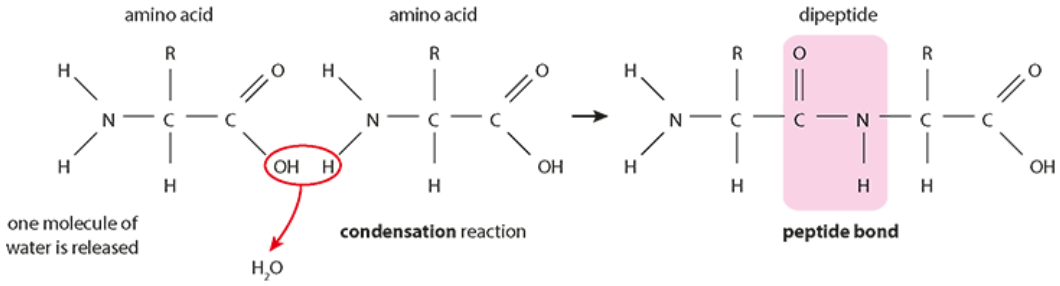


Figure 1.3.6: Two amino acids combine to form a dipeptide.

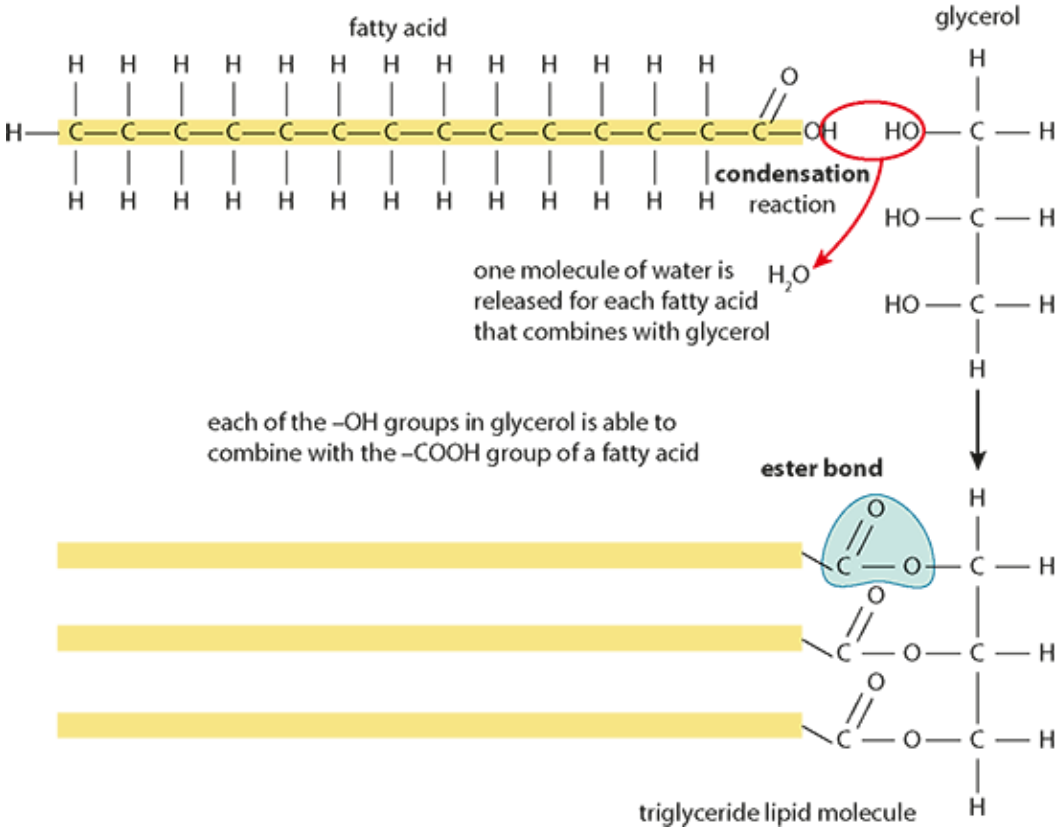


Figure 1.3.7: How a triglyceride lipid is formed from glycerol and three fatty acids.

When more than two amino acids are joined in this way, a **polypeptide** is formed. Polypeptide chains form protein molecules.

In another condensation reaction, glycerol links to fatty acids to produce triglyceride lipid molecules (Figure 1.3.7):

glycerol + 3 fatty acids → triglyceride lipid + water

EXAM TIP

Notice how the carbon atoms in the pentose sugar are labelled 1 to 5. These numbers will help you remember how other molecules bond in the nucleotide.

Nucleotides build nucleic acids

Nucleic acids are large polymers that are essential to all forms of living organism. The term nucleic acid includes DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). Each monomer subunit of a DNA or RNA molecule is called a nucleotide (Figure 1.3.8). All organisms use nucleotides in the synthesis of their nucleic acids (Section 2.7).

Nucleotides are the building blocks of DNA. Each pentose sugar contains a ring of five carbon atoms. Two carbon atoms are linked to either a phosphate group or a pentose sugar, and one carbon atom is bonded to one of four nitrogenous bases. These bases are adenine, guanine, cytosine and thymine.

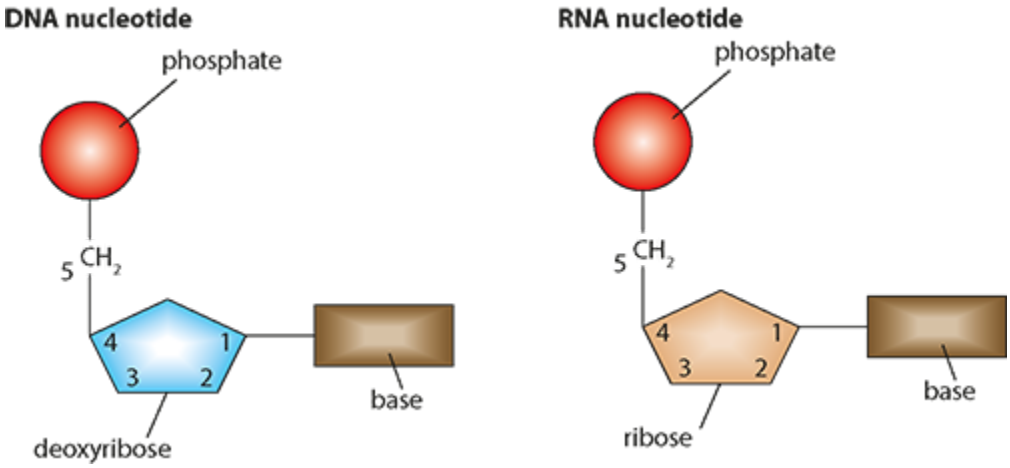


Figure 1.3.8: The general structure of DNA and RNA nucleotides.

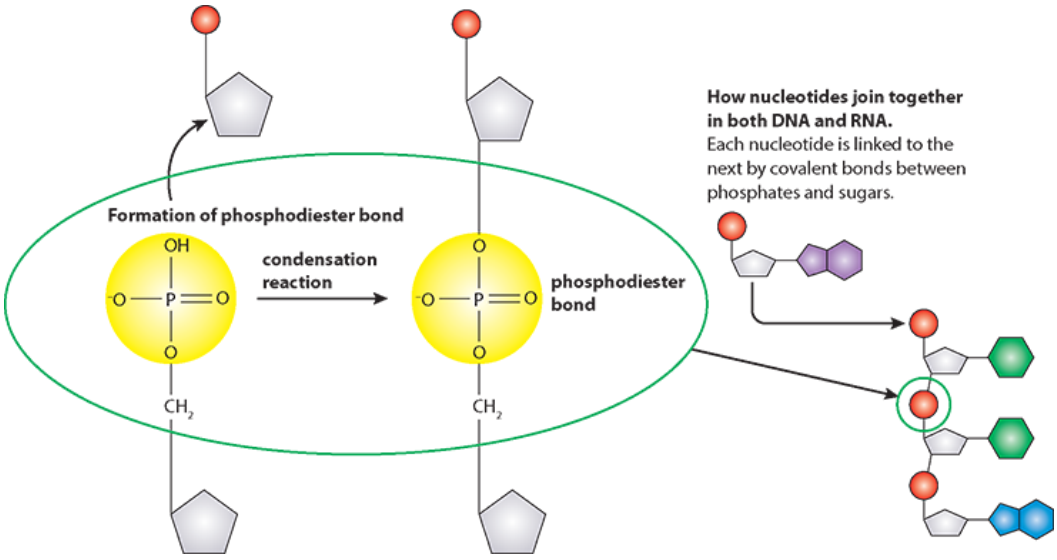


Figure 1.3.9: As nucleotides are joined together, a phosphate group is bonded to a sugar molecule, and water is released. This makes a sugar phosphate that forms the backbone of RNA and DNA. This is another example of a condensation reaction.

RNA nucleotides have a similar structure but the pentose sugar found in RNA is ribose, a five-carbon sugar that has an hydroxyl (–OH) group attached to the second carbon atom. DNA has a

single –H in this position. Another difference is that RNA nucleotides have uracil instead of thymine as one of their nitrogenous bases.

Condensation reactions bond nucleotides together to form a DNA molecule. The phosphate group of one nucleotide links to the deoxyribose ring of the adjacent molecule to form a chain of nucleotides, as shown in Figure 1.3.9.

The sugar and phosphate groups are identical all along the nucleotide chain and form the 'backbone' of DNA and RNA molecules, as shown on the right of Figure 1.3.9.

Hydrolysis: breaking down macromolecules

Hydrolysis reactions break down polymers into smaller units. They are the reverse of condensation reactions. Hydrolysis reactions separate polysaccharides, polypeptides and triglycerides into their monomers.

Reactions like this take place every time food is digested. Hydrolysis is an example of a catabolic reaction in which macromolecules are broken down into monomers. Water molecules are used in **hydrolysis reactions**. Enzymes are required to catalyse the reactions:

- Hydrolysis of starch (a polysaccharide) uses water and produces many molecules of glucose.
- Hydrolysis of protein (made of polypeptide chains) uses water and produces many amino acids.
- Hydrolysis of a triglyceride (a lipid) uses water and produces fatty acids and glycerol molecules.

TEST YOUR UNDERSTANDING

10 State the number of covalent bonds that can be formed by a carbon atom.

11 State what is meant by the term monomer.

12 Name the type of reaction which builds monomers into polymers.

13 What are the three components that organisms use in the synthesis of nucleic acids?

14 Fill in the gaps in this sentence:

Triglycerides are broken down into ..... and ..... during hydrolysis reactions.

15 Which molecule do all organisms use as a source of energy?

KEY POINTS

hydrolysis reactions are those reactions that break down polymers into smaller units. Water is used in the reaction. Hydrolysis reactions are made faster (or catalysed) in living organisms by proteins called enzymes.

catabolism is a type of reaction in which complex molecules are broken down into simpler ones.

1.3.4 Functional groups

Three examples of important **functional groups** in living organisms are **carboxyl groups**, amines and phosphates.

Carboxyl groups

A carboxyl group is a functional group that binds to larger molecules and gives them specific properties. It occurs in carboxylic acids, which have important roles in biological systems. The carboxyl group consists of a carbon bonded to oxygen and an hydroxyl (OH) group and is usually written as –COOH or $–CO_2H$ (Figure 1.3.10).

Carboxyl groups are **polar** and strongly hydrophilic so they can participate in creation of **hydrogen bonds** and many other important reactions. The R in Figure 1.3.10 can be a single H atom or one of many carbon-containing molecules. Carboxyl groups are important in protein synthesis (Section 4.2).

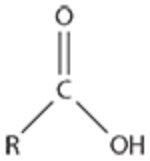


Figure 1.3.10: The structure of a carboxyl group.

KEY POINTS

functional group is a group of atoms in a molecule that has similar chemical properties in every compound in which it appears.

carboxyl group is a functional group consisting of a carbon atom linked by a double bond to an oxygen atom and by a single bond to a hydroxyl group.

polar means relating to a molecule in which the distribution of electrons is not even, so that one part of the molecule is more positively charged than another part.

hydrogen bond is a bond found in macromolecules formed by the attraction of a small positive change on a hydrogen atom and a small negative charge on an oxygen or nitrogen atom.

Every amino acid contains both a carboxyl group and an amino group. Bonds form between the amino group of one amino acid and the carboxyl group of the next. The bond is called a peptide bond and links amino acids together to form polypeptides. Carboxyl groups are also found in a large variety of other molecules including fatty acids.

Amines

Amines are functional groups that contain a nitrogen atom with a lone pair of electrons in their outer shell (Figure 1.3.11, shown in red). Amines are derived from (meaning they come from) ammonia (NH_3), with one or more hydrogen atoms replaced by a different group. The most important group that contains amines are the amino acids. The $–NH_2$ group is called an amino group.

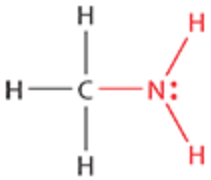


Figure 1.3.11: The structure of an amine group.

Phosphate groups

A phosphate is an ion that contains one phosphorus and four oxygen atoms (usually written PO_4^{3-}). When it is attached to a carbon-containing molecule it is called a phosphate group (Figure 1.3.12).

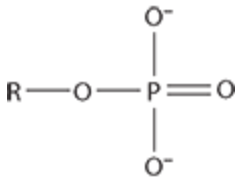


Figure 1.3.12: The arrangement of atoms in a phosphate group.

Phosphate groups are found in DNA and RNA and also in adenosine triphosphate (ATP), which provides energy to cells. Phosphate groups also form part of phospholipid molecules in cell membranes (Section 6.1). Each phospholipid is made up of a lipid molecule and a phosphate group. Phosphate is also an important nutrient resource in ecosystems, especially in freshwater environments (Section 12.2).

KEY POINT

isomers are molecules that have the same molecular formula, but different molecular shapes. There are two major classes of isomers: structural isomers and stereoisomers.

Phosphate groups activate proteins through phosphorylation (the addition of a phosphate group) so that the proteins can carry out their functions in cells. De-phosphorylation is the removal of a phosphate group and this deactivates proteins. ATP is made up of adenosine and three phosphate groups. One phosphate group

is removed and then replaced during the process of energy release.

EXTENSION

Phosphate acts as important buffer in cells. A buffer keeps the pH neutral (not acidic or alkaline), a state that living cells require because most biological processes only occur at a neutral pH. Phosphate-buffered saline, a solution containing water, salt and phosphate, is often used in research into cell activities.

TEST YOUR UNDERSTANDING

16 Define a functional group.

17 Draw a carboxyl group and state one type of molecule that contains it.

NATURE OF SCIENCE

As you read this book and other scientific texts you will come across many examples of scientific conventions that are agreed internationally. One of these is the international system (SI) of metric units and prefixes which are used by all scientists worldwide.

SI – International System of length and volume

1 metre (m) = 1m

1 millimetre (mm) = 10^{-3}m

1 micrometre (mm) = 10^{-6}m

1 nanometre = 10^{-9}m

1 centimetre cubed = 1 cm^3

1 decimetre cubed = 1 dm^3

The convention is also applied to time which is measured in seconds (s), minutes (mins) and hours (h) and to concentration which is measured in mol dm^{-3}

Links

Relatively few elements are found in living organisms and yet a wide range of different molecules can be made from them. How can this be? (Chapter 1)

- How does the solubility of oxygen affect the distribution of aquatic animals? (Chapter 12)
- What are the roles of water in photosynthesis and respiration? (Chapter 2)
- How do the properties of water influence temperature control in animals? (Chapter 8)

1.4 Carbohydrates

LEARNING OBJECTIVES

In this section you will:

- learn that carbohydrates exist in different forms, which give them a variety of properties
- understand that monomers are linked together by condensation reactions to form disaccharides and polysaccharides
- understand that larger carbohydrates are less soluble in water
- compare the storage polysaccharides glycogen and starch which have compact molecular structures
- learn that cellulose is a structural polysaccharide with a molecular structure that is related to its function
- understand that carbohydrates are an efficient short-term energy store
- recall that monosaccharides are metabolised to release energy

> learn that different monosaccharides, including glucose, ribose and deoxyribose, have different properties determined by the way in which they bond together

- understand how the hydroxyl groups of monosaccharides allow the condensation reactions that form polymers
- understand how the arrangement of branching in polysaccharide molecules is related to their structural and energy storage roles.

GUIDING QUESTIONS

- How do living organisms use carbohydrates as storage and structural molecules?
- Why are carbohydrates good sources of energy?

1.4.1 Carbohydrates

Carbohydrates are the most abundant category of molecule in living things. In both plants and animals carbohydrates have an important role as a source of energy, and in plants they also have a structural function. Carbohydrates occur in various forms and different carbohydrates have different properties and roles in living organisms. Monosaccharides have the general formula $(CH_2O)_n$, where n = the number of carbon atoms in the molecule. They are monomers, which means that they are single sugars made up of just one subunit. Glucose, fructose and galactose are three examples of common monosaccharides. Glucose is the most common monosaccharide and it has the chemical formula $C_6H_{12}O_6$.

Linking together monosaccharide monomers to build up polymers produces different types of carbohydrate. The condensation reactions involved in this process result in the production of either disaccharides or polysaccharides. Disaccharides consist of two monomers and polysaccharides are formed from long chains of monosaccharide monomers (Figure 1.4.1). The strong covalent bond between two monomers in a carbohydrate is known as a glycosidic bond and a water molecule is released in the condensation reaction.

In a condensation reaction, two molecules are joined together by strong covalent bonds to form a larger molecule. Condensation is an example of an anabolic reaction, which builds up monomers to form macromolecules. Each condensation reaction requires an enzyme to catalyse the process and it produces one molecule of water.

The condensation of two monosaccharide monomers produces a disaccharide.

For example:

glucose + galactose → (monosaccharide) (monosaccharide)
lactose + water (disaccharide)

Glycosidic bonds can also form between a sugar molecule and another molecule such as a lipid or amino acid to produce glycolipids and glycoproteins. Glycolipids and glycoproteins are found on the outer surface of cell membranes and are important in enabling cells to recognise one another. You can read more about cell to cell recognition in Chapter 6.

KEY POINT

glycosidic bond a type of covalent bond that links a carbohydrate molecule to another carbohydrate, or another group.

In any condensation reaction, two molecules can be joined to form a larger molecule, held together by a covalent bond. Condensation is an example of an anabolic reaction, in which monomers are built up to form macromolecules. Each condensation reaction requires an enzyme to catalyse the process and every reaction produces one molecule of water.

Different combinations of monosaccharide monomers produce a range of disaccharides. When a bond is formed between two glucose monomers, a disaccharide called maltose is produced. Maltose is found in seeds such as barley. Other monosaccharides include fructose, found in fruits, and galactose, which is present in milk. The equations here show some disaccharides that can be formed.

NATURE OF SCIENCE

Healthy eating and scientific literacy

Scientists have a responsibility to inform people about their work. Responsible citizens need to be able to evaluate information critically. What do you know about healthy eating?

It is difficult to say which parts of our diet are most likely to cause obesity. Studies show that the answer is not simple. Too much carbohydrate in a person's diet can cause them to become overweight, but this explanation may be too simple. The type of carbohydrate a food contains is important as well.

Simple carbohydrates (sugars, such as glucose and sucrose) are typically found in processed foods with low nutritional value, such as sugary drinks or food with added sugar. Adding sugar to food increases its calorie content, but does not provide any additional nutrition. But simple carbohydrates are not only found in low-nutrient foods. Milk and milk products contain lactose, which is a type of simple carbohydrate, but milk is also rich in protein, calcium and vitamin D. Other simple carbohydrates are present in fruits and vegetables that contain a variety of other vitamins and minerals, and nutrients such as fibre.

Complex carbohydrates, such as starch, are found in food such as bread and pasta. They contain longer chains of sugar molecules than simple carbohydrates. The body converts these molecules into glucose, which it uses for energy. As complex carbohydrates have longer chains, they take longer to digest and provide a more lasting source of energy than simple carbohydrates. Complex carbohydrates found in wholefoods tend to be highly nutritious. For example, wholegrain foods

such as brown rice, barley and oats also provide fibre, vitamin B and E. Wholegrain foods may also reduce the risk of serious health conditions, such as type 2 diabetes and cardiovascular disease.

But not all complex carbohydrates are healthier choices: they are also found in processed foods such as white flour and white rice.

This means that some forms of simple carbohydrates are healthier than some complex carbohydrates. As you choose what to eat, consider the overall nutrition in each food, not just the type of carbohydrate it contains.

To consider:

1 How do you make informed choices about your own diet?

2 Discuss the importance of labelling on processed foods. Do labels help people learn about healthy options?

glucose	+	glucose	→	maltose + water
(monosaccharide)		(monosaccharide)		(disaccharide)
glucose	+	fructose	→	sucrose + water
glucose	+	galactose	→	lactose + water

If further monosaccharides are added to a disaccharide, a polysaccharide is formed, as you can see in Figure 1.4.1. Types of carbohydrate are summarised in Table 1.4.1.

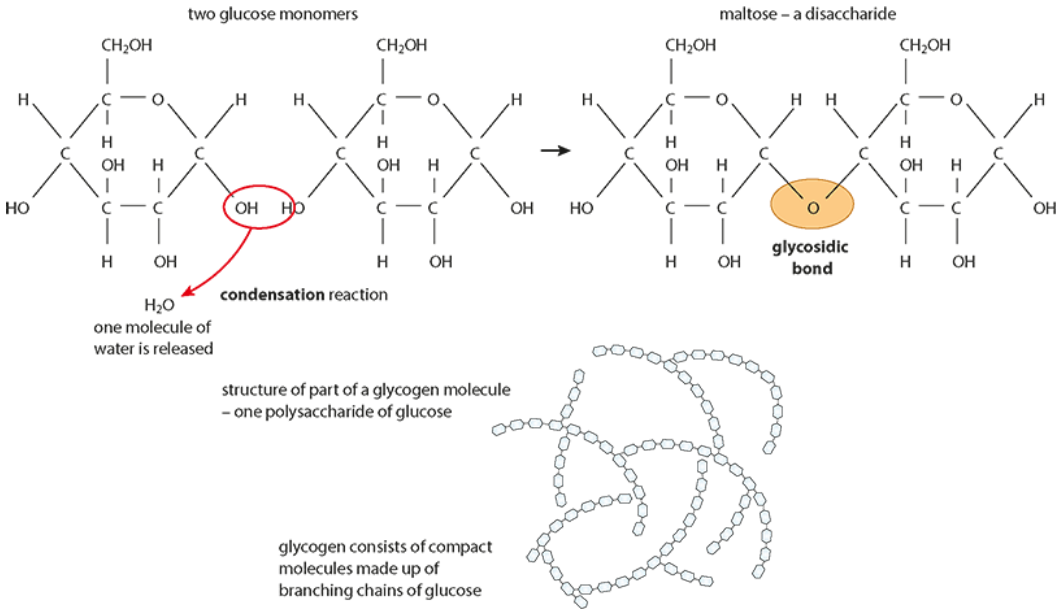


Figure 1.4.1: Monosaccharide subunits (glucose in this case) are joined in a condensation reaction, forming a disaccharide (maltose) and water. Glycogen is a polysaccharide, formed from long chains of glucose subunits.

Form of carbohydrate	Examples	Example of use in plants	Example of use in animals
monosaccharide	glucose, galactose, fructose	fructose is found in many fruits, making them taste sweet and attracting animals to eat them; this helps seeds in the fruit to be dispersed, or spread	glucose is the source of energy for cell respiration it is obtained from the digestion of carbohydrate foods
disaccharide	maltose,	sucrose is	lactose is

	lactose, sucrose	transported from leaves to storage tissues and other parts of the plant to provide an energy source	found in milk and provides energy for young mammals
polysaccharide	starch, glycogen, cellulose	cellulose is a structural component of plant cell walls starch is used as a food store	glycogen is the storage carbohydrate of animals, found in the liver and muscles

Table 1.4.1: Examples and roles of carbohydrates.

Form and function of monosaccharides

Monosaccharides can be either pentoses or hexoses. Pentose have a ring of 5 carbon atoms (Fig 1.4.9), while hexoses such as glucose are formed of a ring of 6 carbon atoms.(Figure 1.4.1)

1.4.2 Size, solubility and energy storage

Small carbohydrates such as glucose are very soluble in water, but larger molecules are less soluble. Glucose and other monosaccharides are an important source of energy for metabolism, their solubility makes them easy to transport. As part of larger molecules, they can form storage molecules. Large polysaccharides are those which contain from 40 to 1000 monomers.

Starch, glycogen and cellulose are all polymers of glucose monomers. Glycogen and starch are **storage carbohydrates** that act as an energy reserve.

Glycogen is stored in animals, fungi and prokaryotes, and starch is stored in plants. Both molecules have a compact shape and are insoluble because of their large size (Figure 1.4.1). Glycogen is made of branching chains of glucose monomers (Figure 1.4.1) and starch has long chains of glucose that coil into a helical shape (Figure 1.4.2).

Glycogen and starch are excellent for energy storage because they are easily put together, or assembled, and easily broken down by enzymes. In animals glycogen is stored in the liver and muscles. It can be used quickly and efficiently to release energy if it is needed. In plants starch is stored in all cells and is a useful reserve for the plant. Starch also provides food for animals in the form of potatoes and seeds, such as wheat and maize.

Both glycogen and starch take up less space than the same number of free glucose molecules. Small differences in the bonding between monomers produce the differences in shape between glycogen and starch.

Starch is a mixture of **amylose** and **amylopectin**. Amylose is a linear molecule with between 500 and 20 000 of glucose molecules. Amylopectin is branched and contains over 1 million glucose molecules (Figure 1.4.2).

When glycogen and starch are hydrolysed (broken down; see Section 1.3) glucose monomers are released. Glucose can be metabolised in cells in the reactions of respiration that release the energy living organisms need.

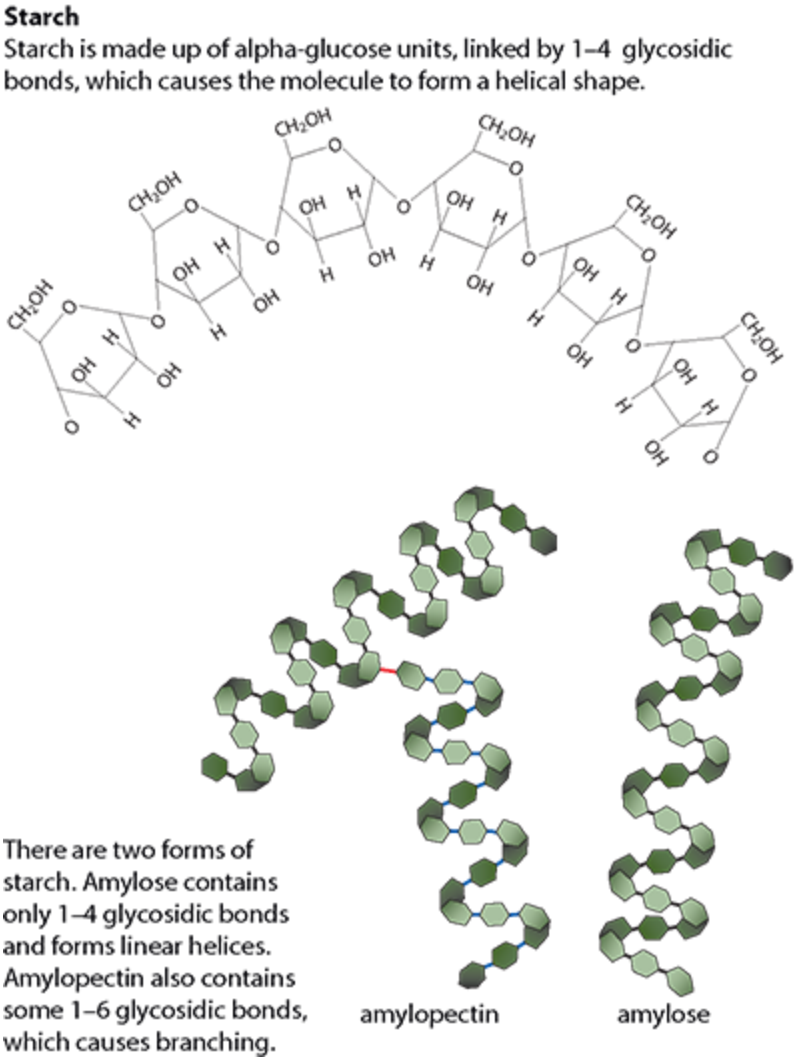


Figure 1.4.2: Starch is a polysaccharide found in plants.

Cellulose

Cellulose is polysaccharide with a different arrangement of bonding between glucose monomers. As a result, cellulose has very different properties from starch and glycogen. Cellulose is a **structural polysaccharide** that is used to build the cell walls of plants. It is made up of long, straight chains of glucose molecules but the glycosidic linkages between glucose monomers produce flat, linear strands. Hydrogen bonds can form cross-links between adjacent, parallel chains, which gives the polysaccharide its strength and structural properties (Figure 1.4.3). Cellulose strands form tightly packed, rigid bundles that support stems and hold plants upright.

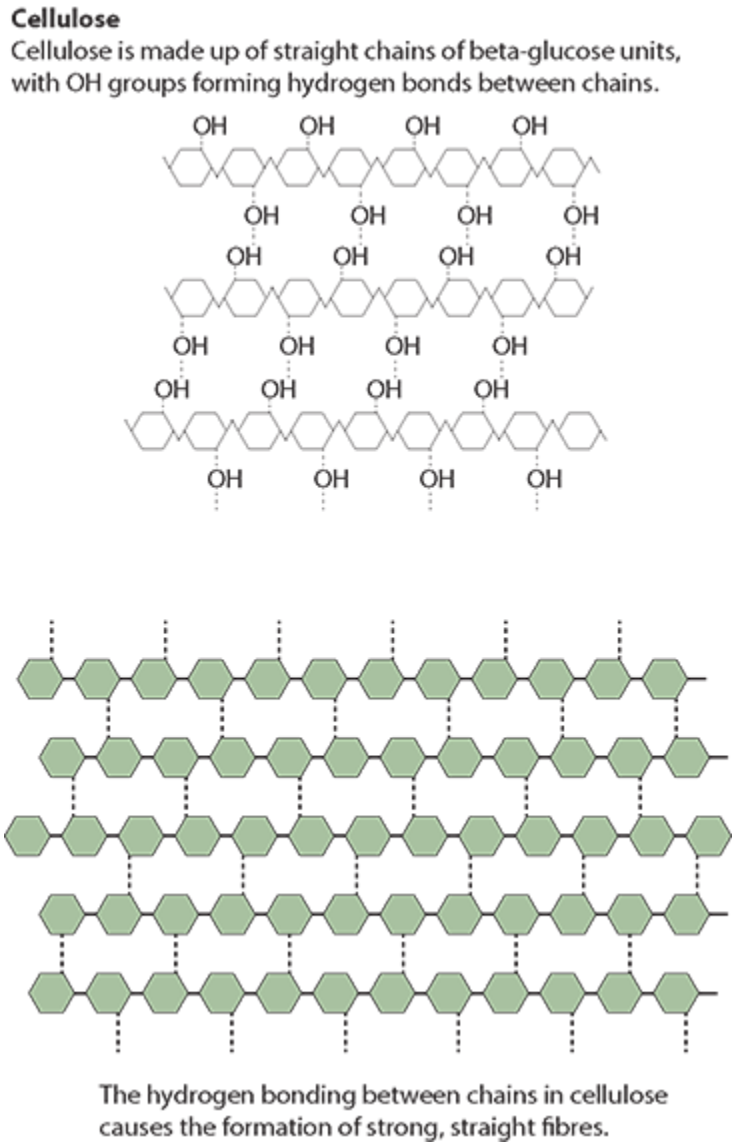


Figure 1.4.3: Cellulose is a structural polysaccharide found in plants.

SCIENCE IN CONTEXT

The abundance of cellulose

Cellulose, found in the cell walls of all plants, is probably the most abundant organic molecule on Earth. Paper and cotton are almost pure cellulose and wood contains cellulose and lignin.

Cotton is a fluffy fibre which surrounds the seeds of cotton plants to help them disperse. Cotton is found naturally in tropical and subtropical regions of Africa, Egypt, India and the Americas.

Cotton has been spun and woven into fabric for more than 2500 years: fragments of cotton from 500 BCE have been found in Peru and dyed cotton clothing was used in ancient India, Egypt and China.

Today cotton is the most commonly used natural fibre and around 25 million tonnes are produced annually. India is the largest producer and the USA is the largest exporter of cotton. In recent years genetically modified cotton plants have been developed in an effort to reduce the use of pesticides. It was recently estimated that 25 million hectares of land, almost 70% of the total, are planted with GM cotton plants.

Long ago, paper was also made from cotton rags and linen from flax plants. Cotton paper from the year 751 has been found in Samarkand in Uzbekistan, but by the turn of the 20th century wood pulp became the primary material used. Cotton is still used to make some speciality paper for important documents because it will last for many years and not deteriorate.

To consider:

1. Thousands of years ago, cotton plants were cultivated independently in several part of the world. Investigate the reasons why cotton was such a useful plant to domesticate.

2. Why do you think that waste fabric is not used to make paper today?

3 Thinking about your own clothing, are you wearing any items that are made from cotton fibres?

TEST YOUR UNDERSTANDING

18 State two examples of disaccharides.

19 Which of the following statements is true?

All carbohydrates:

A are polymers

B are simple sugars

C consist of one or more simple sugar

D are soluble in water

20 Fill in the gaps in this description of the formation of polysaccharides.

Two monosaccharide monomers are linked together by .....reactions to form a ……..... . If further monomers are added a ……. is formed.

21 Give two examples of polysaccharides and outline their functions.

22 Explain why plants store starch and animals store glycogen rather than glucose molecules.

EXTENSION

Bonding arrangements of monosaccharides

Isomers of glucose

Glucose has two structural isomers, alpha (α) and beta (β) glucose (Figure 1.4.4).

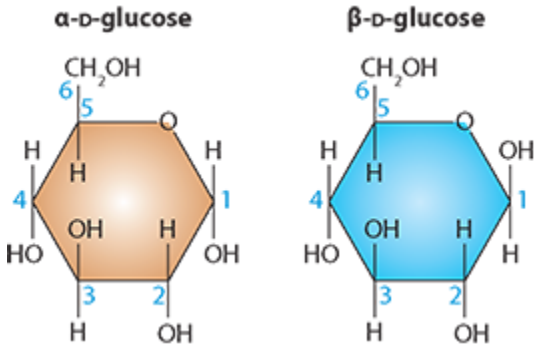


Figure 1.4.4: The isomers of glucose.

Notice how the carbon atoms in the molecules are numbered. These numbers are used to describe the linkages that the molecules can form. The positions of the side groups in the two molecules give the molecules abilities to bond in different ways. Bonding patterns produce the different polysaccharides that form amylose and amylopectin in starch molecules, and cellulose. Starch is built up of alpha-D-glucose monomers and cellulose of beta-D-glucose monomers.

Amylose (Figure 1.4.2) consists of a linear, helical chains of between 500 and 20 000 alpha-D-glucose monomers linked together through glycosidic bonds between the 1 carbon on one molecule and the 4 carbon on the next molecule. The bonds are called 1–4 glycosidic bonds. The hydroxyl (–OH) groups allow **condensation reactions** to link adjacent monomers and release water as they form.

Amylopectin molecules are much larger polymers each containing over a million glucose monomers. Amylopectin is

branched because it forms 1–6 glycosidic bonds as well as 1–4 bonds (Figure 1.4.6).

Condensation reactions occur between sugar monomers. Here a 1–4 linkage is made between two alpha glucose monomers to form maltose. A glycosidic bond is made and water is released from the hydroxyl groups.

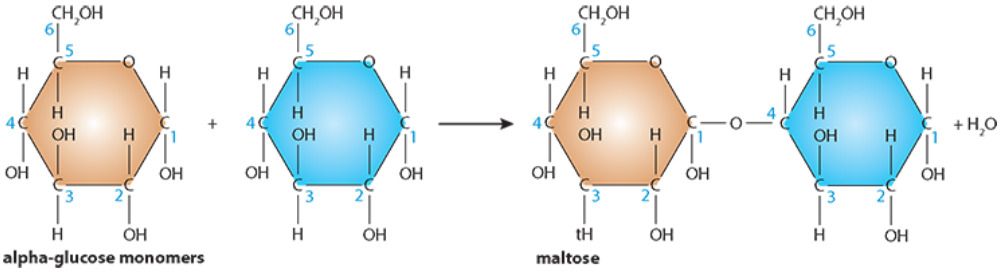


Figure 1.4.5: Two alpha glucose monomers are joined in a condensation reaction to produce maltose.

Cellulose molecules are linear chains of hundreds or thousands of beta-D-glucose monomers joined through 1–4 linkages (Figure 1.4.7). Notice that in cellulose alternate glucose monomers are 'upside down'. Compare this structure with the structure of glucose which has alpha glucose monomers joined through 1–4 linkages.

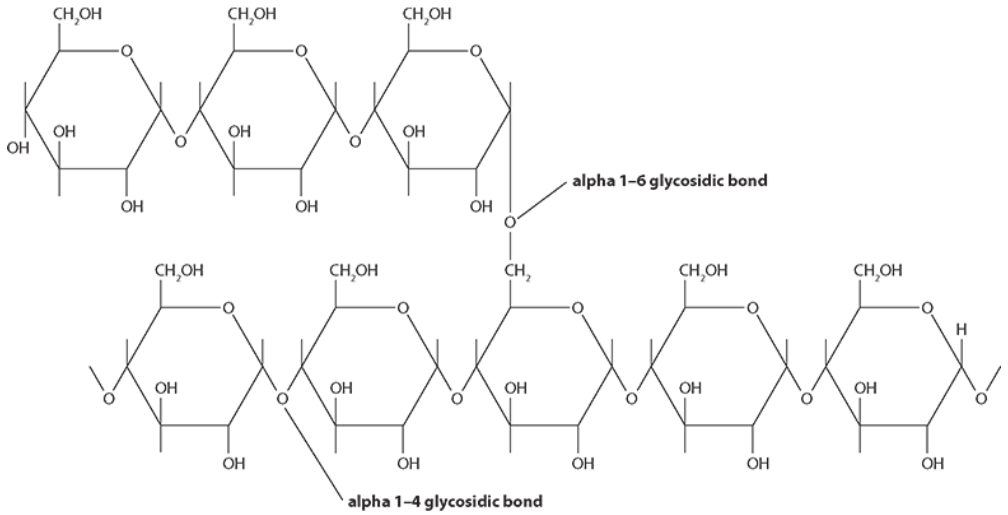


Figure 1.4.6: Amylopectin.

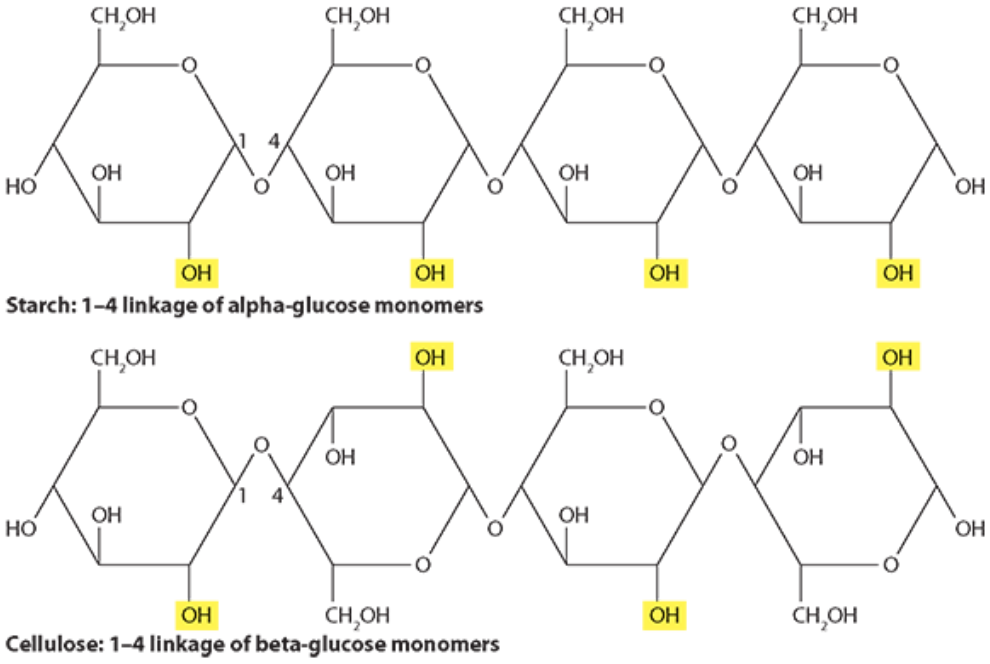


Figure 1.4.7: Comparing the linkages in starch and cellulose.

Table 1.4.2 summarises the different structures and functions of the important polysaccharides amylose, amylopectin, cellulose and glycogen.

Polysaccharide	Structure	Function
amylose	Linear, helical chains 500–20 000 alpha-D-glucose monomers linked together through the bonds are called 1–4 glycosidic bonds.	Energy storage in plants, part of starch.
amylopectin	Branched molecules of over a million alpha-d-glucose monomers with 1–6 glycosidic bonds.	Energy storage in plants, part of starch.
glycogen	Compact branching chains of alpha-d-glucose monomers with 1–4 glycosidic linkages. Branches with alpha-1–6 linkages occur approximately every 10 units.	Short-term energy storage in animals, fungi and prokaryotes.
cellulose	Straight chains of beta-D-glucose monomers. Hydrogen bonding occurs between polar hydroxyl groups in the chains.	Structural polysaccharide forming strong rigid fibres of plant cell walls.
chitin	Long chains of sugar molecules derived from glucose by replacing one hydroxyl group with a nitrogen-containing group. Linkages are similar to those	Structural polysaccharide found in cell walls of fungi, exoskeletons of arthropods

	in cellulose so chitin forms linear sheets of strong fibres.	and the scales of fish.

Table 1.4.2: Some important polysaccharides and their roles.

EXTENSION

Structural differences between the glycosidic linkages in starch and cellulose affect an animal's ability to digest plant foods. Enzymes such as amylase break down starch but cannot break down cellulose polymers. Some animals, including cows and termites, digest cellulose by keeping special microorganisms in their digestive systems. These microorganisms produce cellulose-digesting enzymes. Humans and most animals do not make an enzyme capable of digesting cellulose, so cellulose fibres pass undigested through the body and are known as 'fibre' or roughage' (Section 12.1).

1.4.3 Ribose and deoxyribose

Pentose sugars are one component of the nucleotides which make up both DNA and RNA molecules. RNA contains the sugar ribose and DNA contains deoxyribose. The difference between the two molecules is the presence of hydroxyl (–OH) groups on the 2′ carbon of the ribose. Ribose has an -OH group and deoxyribose does not (Figure 1.4.9).

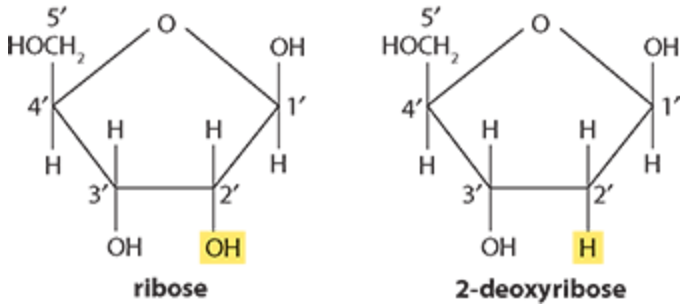


Figure 1.4.9: The structures of ribose and deoxyribose.

KEY POINT

You will notice that each of the numbered carbon atoms has a small dash next to it, for example 2′ or 4′. The dashes are called primes. When spoken out loud they are called ‘2 prime’ or ‘4 prime’. In DNA and RNA the ribose sugars are attached to other molecules that contain ring compounds. The primes help to distinguish from any numbers given to atoms in other rings.

Both DNA and RNA contain nucleotides which have the pentose sugars in a central position with a base attached to the 1′ carbon and a phosphate group attached to the 5′ carbon. The hydroxyl groups on the 3′ and 5′ positions enable the pentose groups to

link to one another via phosphodiester bonds between adjacent molecules.

You can read more about the detail of DNA and RNA structure in Chapter 4.

TEST YOUR UNDERSTANDING

23 Compare the molecular arrangements of cellulose and amylose.

EXTENSION

Using molecular visualisation software

There are many websites that offer molecular visualisation software free of charge. One example is Jmol. Using this software you can look at 3D images of carbohydrates and see how they are assembled. Examine models of glucose and notice how the different atoms are shown. Compare models of unbranched amylose and branched amylopectin. Compare the linkages in unbranched and branched parts of each molecules.

Links

- How does excessive intake of carbohydrate lead to diabetes? (Chapter 8)

1.5 Lipids

LEARNING OBJECTIVES

In this section you will:

- learn that lipids exist in different forms and have a range of properties
- learn how triglycerides are formed by condensation reactions between fatty acids and glycerol
- understand why non-polar lipids are hydrophobic
- learn that phospholipids contain a phosphate group and are amphipathic
- define fatty acids as saturated, monounsaturated or polyunsaturated
- recall that lipids are used as a long-term energy store
- learn that lipids release energy when they are oxidised
- understand that lipids are stored in adipose tissue
- recall that lipids insulate vital organs
- describe steroid hormones as non-polar molecules with a four ring structure and that they can enter cells via cell membranes.

GUIDING QUESTIONS

- How do the structures of different fatty acids and lipids affect their properties and functions?
- Why is it better to store energy as lipid rather than as other organic molecules?
- What are the benefits and risks of lipids in the human diet?

1.5.1 Structure and forms of lipids

Lipids are non-polar and hydrophobic molecules which are insoluble in water but do dissolve in organic solvents. Lipids include fats, waxes, oils, steroids and phospholipids. Fats and oils are part of a sub-group known as triglycerides. Those that are solid are generally referred to as fats, while those that are liquid are known as oils.

KEY POINT

lipids together with carbohydrates and proteins, are the main components of plant and animal cells. Two examples of lipids are triglycerides and cholesterol.

Different forms of lipid have a range of uses; some of these are shown in Table 1.5.1.

Lipid type	Some important uses
fats	• long-term energy reserve and concentrated source of energy for animals • supply essential fatty acids that the body cannot synthesise, and fat-soluble vitamins (A, D, E and K) • constituent of cell membranes
oils	• energy reserve for plants found in seeds such as linseed and olives
waxes	• beeswax is used to construct honeycomb • lanolin is used to protect wool

	• plant waxes control evaporation from leaves
steroids	• form hormones and vitamins
phospholipids	• essential part of membranes and the nervous system

Table 1.5.1: Lipids and their uses in living organisms.

Triglycerides

Triglycerides are the most abundant group of lipids. They are formed by condensation reactions between three fatty acid molecules and one glyceride molecule. They are sometimes called neutral or true fats.

Glycerol + 3 fatty acids → triglyceride lipid + water

Fatty acids consist of long chains of carbon atoms with hydrogen atoms bound to them (Figure 1.5.1).

Lipids have no polar groups and so cannot dissolve in water, although they do dissolve in organic solvents such as chloroform and acetone. Lipids contain about twice as much energy per gram as carbohydrates and proteins (Table 1.5.2), but each type of storage molecule has its own advantages.

Molecule	Approximate energy content per gram/kJ
lipid	39
carbohydrate	17
protein	18

Table 1.5.2: Energy content of different molecules.

EXAM TIP

- Do not confuse the terms glyceride and glycerol.
- Glycerides are formed by the bonding between glycerol molecules and fatty acids.
- Glycerol is a molecule which has three hydroxyl functional groups (Section 1.3) which can be linked to three fatty acids to form triglycerides.

Fatty acids

There are many different fatty acids and they are all long carboxylic acid chains. At one end of the chain is a carboxyl group (-COOH) and at the other a methyl group (CH_3), and there are carbon atoms between them. The number of carbon atoms in a fatty acid is always an even number, most commonly 14–22, although shorter and longer chain fatty acids do exist. If the carbon chain is linked to the maximum number of H atoms with no double bonds we say that it is **saturated** because no more H atoms can be added. If the chain contains a double bond between two of the carbon atoms it is said to be **unsaturated** (Figure 1.5.2). A chain with just one double bond is **monounsaturated**, while one with two or more double bonds is said to be **polyunsaturated**. Polyunsaturated fatty acids tend to be liquids at 20°C and are mainly derived from plant sources. Examples are sunflower oil, corn oil and olive oil.

Unsaturated fatty acids occur as isomers and may be either a ***cis*** or a ***trans*** configuration. If the 'spaces' where additional hydrogen atoms could bond are both on the same side of the hydrocarbon chain, the fatty acid is known as a *cis* fatty acid.

The carbon chain of a *cis* fatty acid is slightly bent. If the spaces are on opposite sides, it is a *trans* fatty acid (Figure 1.5.3).

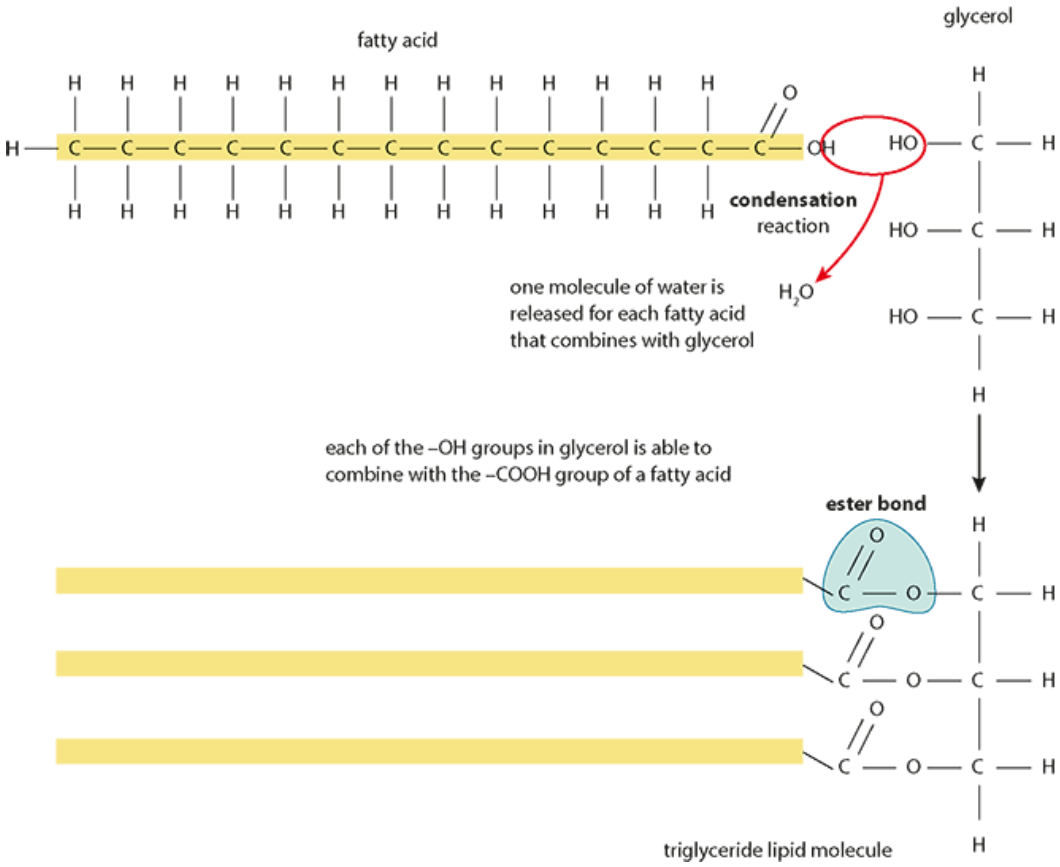


Figure 1.5.1: A triglyceride is formed by condensation reactions between three fatty acids and one glycerol molecule.

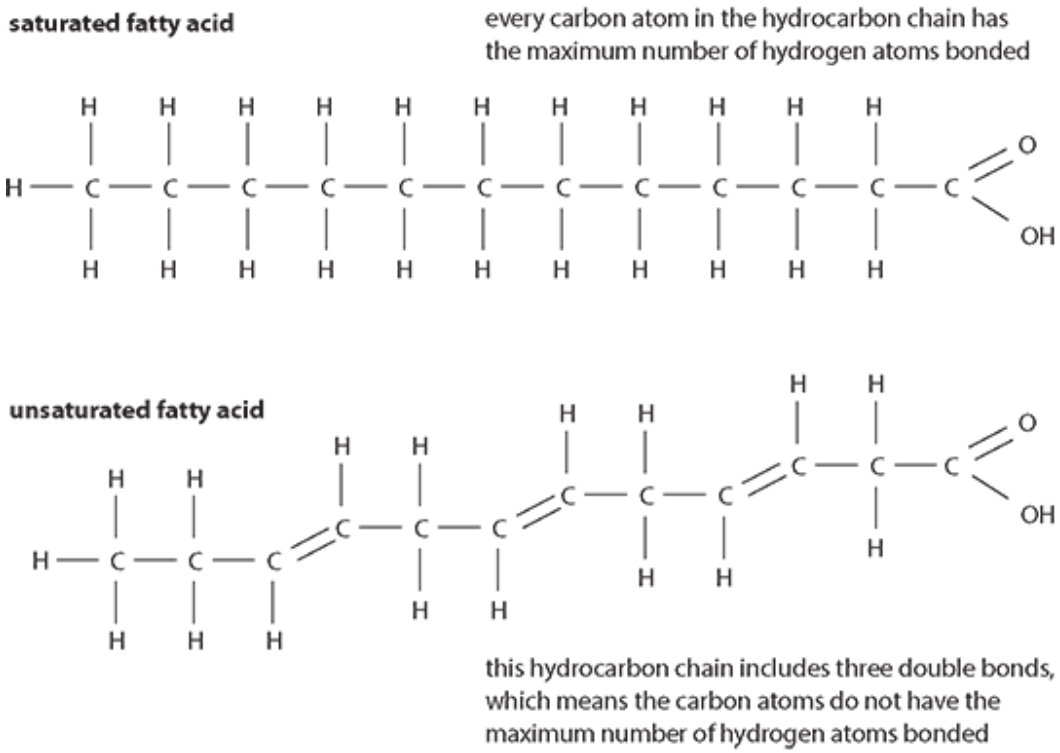


Figure 1.5.2: Saturated and polyunsaturated fatty acids.

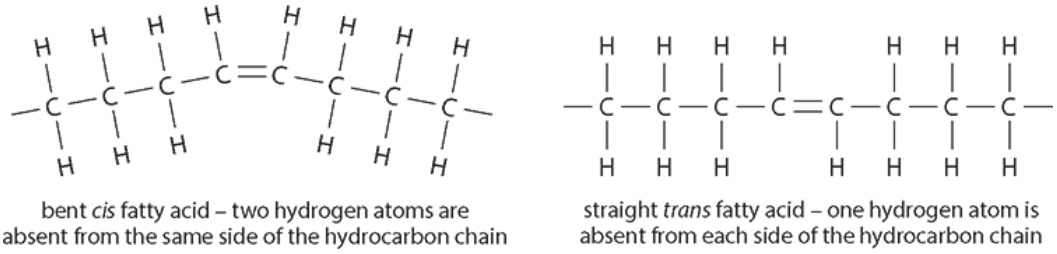


Figure 1.5.3: Structural diagrams of *cis* and *trans* fatty acids.

1.5.2 Saturated and unsaturated fatty acids and health

The relative amounts of different types of fatty acid in a person's diet can, in some cases, be correlated with health issues. Eating a diet that is high in saturated fatty acids is common in some western countries. In research, this type of diet is shown to have a positive correlation with an increased risk of coronary heart disease (CHD). A positive correlation between two measures is where, if one measure increases, so does the other. This does not mean that one causes the other, but it suggests a relationship (see Nature of Science, Correlation and cause).

Saturated fatty acids can be deposited inside the arteries, and if the deposits combine with cholesterol they may lead to atherosclerosis, a condition that reduces the diameter of the arteries and leads to high blood pressure (Section 8.3). Reliable evidence suggests that in countries where the diet is high in saturated fatty acids, and where many high-fat foods, animal products and processed foods are eaten, there is likely to be a high incidence of CHD. Since all fatty acids are high in energy, an excess of these foods in the diet can also lead to obesity, which places a further strain on the heart.

On the other hand, people who eat a Mediterranean-style diet, which is rich in unsaturated fatty acids from olive oil and fresh vegetables, tend to have a low incidence of CHD. These fats do not combine with cholesterol and so arteries tend to remain unblocked and healthy. One type of unsaturated *cis* fatty acid is the omega-3 group. These have a double bond at the third bond from the $-CH_3$ end of the molecule. Omega-3 fatty acids in our diet come from eating fish, such as salmon and pilchards,

walnuts and flax seed. Another group, the omega-6 fatty acids, have a double bond at the sixth position and come from vegetable oils. These two fatty acids are essential for human health and help with the absorption of vitamins and minerals.

THEORY OF KNOWLEDGE

Health issues: *trans* fats or saturated fats?

Trans fatty acids are produced when vegetable oils are hydrogenated (have hydrogen atoms added to their molecules).This changes their properties so they will spread more easily and last longer. *Trans* fatty acids are not often found in nature and have no known health benefits. Hydrogenated fats are used in the manufacture of processed food, such as margarine, baked foods and coffee creamers, although many food manufacturers have reduced the amounts they use. *Trans* fats are metabolised in the liver and released as low-density lipoproteins (LDLs). Increased levels of LDLs in the blood is linked to an increased risk of cardiovascular disease.

Eating a diet that contains high levels of trans fats has been shown to lead to high cholesterol levels, which in turn can lead to CHD and strokes. But most people do not eat large amounts of trans fats. In Western Europe, for example, it has been estimated that most people eat only about half the maximum recommended level of these fats. Most health professionals advise that saturated fats are a greater risk to health because of their contribution to atherosclerosis.

To consider:

1 Do different scientists interpret evidence in the same way?

2 How can we make decisions about the relative harms and benefits of foods in our diet when there are conflicting views?

The evidence is not conclusive about which of the lipids are healthy and which are not. However, all health professionals recommend a balanced diet and recommend reducing the amount of processed food that is eaten.

1.5.3 Lipids and energy storage

Lipids are used to store energy. They are compact molecules and make an efficient long-term energy store. Lipids are high in energy; 1 gram of fat provides twice as much energy as the same mass of carbohydrate (Table 1.5.2). In addition to lipids that are part of the diet, any excess carbohydrate that a person eats is also converted into fats and stored in the body. All mammals store more of their energy reserves as fat than as glycogen. Stores are increased when food is abundant, or if a person eats more than they need for their activities. Fat is released from storage and used when energy is needed. Fatty acids are oxidised in the mitochondria, producing ATP for usable energy and releasing carbon dioxide and water.

A healthy diet is one in which energy intake matches energy used. If a person eats more than they need for their lifestyle, particularly if they do little or no exercise, they will gain weight as fat accumulates in their body. Fatty food contains a lot of energy and it can be easy to consume more kilojoules of energy than the body needs.

EXTENSION

Many small mammals such as hedgehogs, chipmunks, dormice and bats hibernate through the winter months. Their metabolism, heart rate and breathing slow down and their temperatures drop. To prepare for hibernation mammals feed in summer and autumn when food is abundant and they store fat which they use during their winter sleep. If they cannot store sufficient fat animals may die during hibernation, especially if the weather is very severe or they wake up too soon before fresh food is available in spring.

KEY POINT

mass is a measurement of the amount of matter an object contains but weight is a measurement of the pull of gravity on an object. Mass is measured in grams or kilograms whereas weight is measured in Newtons (N).

Non-scientists tend to use the word ‘weight’ when they are really talking about mass.

SCIENCE IN CONTEXT

Body mass index (BMI) is used as a way of assessing if a person’s body mass is healthy. It is calculated using the formula:

$$\text{BMI} = \frac{\text{weight (kg)}}{[\text{height (m)}]^2}$$

Table 1.5.3 shows the ranges of BMI values that are healthy and unhealthy in different groups of people.

BMI (non-Asian) kg/m^2	BMI (Asian) kg/m^2	Weight status
<18.5	<18.5	underweight
18.5–24.9	18.5–22.9	healthy
25.0–29.9	23.0–24.9	overweight
–	>25.0–29.9	pre-obese
>30.0	>30.0	obese

Table 1.5.3: World Health Organization health profiles for different BMI values.

Differences in BMI between individual adults of the same age and sex are usually due to body fat, although there are exceptions to this rule. BMI values will be overestimated for body builders and elite athletes whose bodies have a high proportion of muscle.

BMI values underestimate the amount of body fat for older people and for people with physical disabilities who may have muscle wasting. BMI is not accurate for people with eating disorders such as anorexia nervosa.

Despite these inaccuracies BMI values have been used as a quick and easy way to assess whether a person is carrying too much body fat. But it is important to remember that many other factors must also be taken into account before conclusions about a person's health can be made.

1.5.4 Phospholipids

Phospholipids are important molecules that are found in all cell membranes. A phospholipid molecule has two hydrophobic, fatty acid tails and a phosphate head which is hydrophilic. Phospholipids can form bilayers because they are amphipathic molecules (Section 2.1). Lipid bilayers form when the hydrophobic tails in one layer align to face one another and leave the hydrophilic heads of both layers in contact with water. Phospholipid membranes (Figure 1.5.4) form the outer layer of cells and are important in controlling what can enter and leave a cell. Molecules embedded in the membrane allow cells to communicate with other cells and parts of the body (Section 5.2).

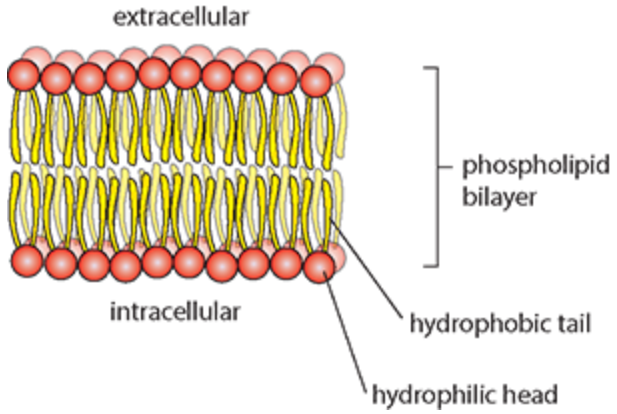


Figure 1.5.4: The phospholipid bilayer of a cell membrane.

Where is energy stored?

Digestion of lipids in our diet begins in the small intestine and releases fatty acids and glycerol that are absorbed across the intestinal membrane.

NATURE OF SCIENCE

Correlation and cause

As we study the occurrence of medical conditions that may be related to diet, it is important to distinguish between correlation and cause. A correlation between two variables, such as a high incidence of CHD and a high intake of saturated fatty acids, does not mean that CHD is caused by the fat intake.

Looking for correlation is one of the most common and useful statistical analysis techniques. Correlation describes the degree of relationship between two variables. For example, in the last 30 years, the number of people taking a holiday each year has increased. In the last 30 years, there has also been an increase in the number of hotels at holiday resorts. This data shows a **positive correlation**.

We could also consider annual deaths from influenza and the number of influenza vaccines given. In this case, there is a **negative correlation**. With these examples, we might feel safe to say that one set of data is linked to the other and that there is a **causal relationship** – because there are more tourists, more hotels have been built; greater use of the influenza vaccine has resulted in fewer deaths from influenza.

However, just because the data shows a trend it does not necessarily mean that there is a causal relationship. If we consider the number of people using mobile phones in the last 10 years against the area of Amazon rainforest cut down there would be a positive correlation. But this does not mean that the use of mobile phones has caused rainforest to be cut down – nor does it mean that a reduction in rainforest area results in more mobile phone use.

Observations without experiments can show a correlation but usually experiments must be used to provide evidence to show the cause of the correlation. To find evidence for correlation between diet and human health is difficult. It is not ethically possible to conduct experiments that restrict the diets of subjects to assess the affects on their health. So observational data or data from **epidemiological** studies is all we have to use. We must think about how the data is gathered and what other variables, such as lifestyle, genetics and family history, are important. There are difficulties in collecting objective data that accounts for all possible variables. It may never be possible to say that one type of diet or fatty acid is 'good' and another is 'bad', because individual subjects vary in so many different ways.

To consider:

1 Can you think of some other examples of two factors that are positively correlated but where one does not cause the other?

2 Why do you think it is very difficult to collect objective data about human diets and health?

TEST YOUR UNDERSTANDING

24 Name the type of reaction that produces a triglyceride.

25 Why are lipids an economical way for animals to store energy?

26 Distinguish between a saturated and unsaturated fatty acid.

27 Describe the difference between a *cis* and a *trans* fatty acid.

28 Calculate the BMI of a person who weighs 90 kg and is 1.6 m tall. What can you say about the health issues this person might face?

Once they have crossed the membrane, the molecules are recombined to form triglycerides that enter the lymphatic system via lacteals in the villi of the intestine. Digested lipids are either transported to the liver, or can be stored in the fat cells (adipocytes) that make up adipose (fat) tissue found throughout the body.

Adipose tissue is formed under the skin where it insulates the body from heat and cold, and around major organs, such as the kidneys and the heart, which it cushions and protects from damage.

Layers of adipose tissue are important insulators against cold environmental conditions for animals such as seals and whales which live on the ice and in freezing arctic waters. These animals are mammals and homeotherms so must maintain a constant body temperature.

KEY POINT

adipose tissue (body fat) formed of adipose cells containing stores of triglycerides found under the skin and around internal organs.

They all have thick layers of fat (blubber) to act as thermal insulation and protect them from the cold.

There are two types of adipose tissue: white adipose tissue which stores energy and brown tissue which generates heat. White tissue stores fat in large droplets whereas brown fat contains smaller droplets and has cells with large numbers of mitochondria.

The hormones leptin and estrogen as well as cytokines are also produced by adipose tissue.

1.5.5 Steroid hormones

Steroid hormones are produced by glands in the adrenal cortex, the testes and the ovaries. All steroid hormones are derived from cholesterol and have similar structures containing four fused rings (Figure 1.5.5).

Steroid hormones are non-polar molecules so they are able to pass through the phospholipid bilayers of cell membranes (Section 6.2). Once inside a cell, the hormones bind with special receptors which carry them to the nucleus of the cell where they influence transcription (Figure 1.5.6).

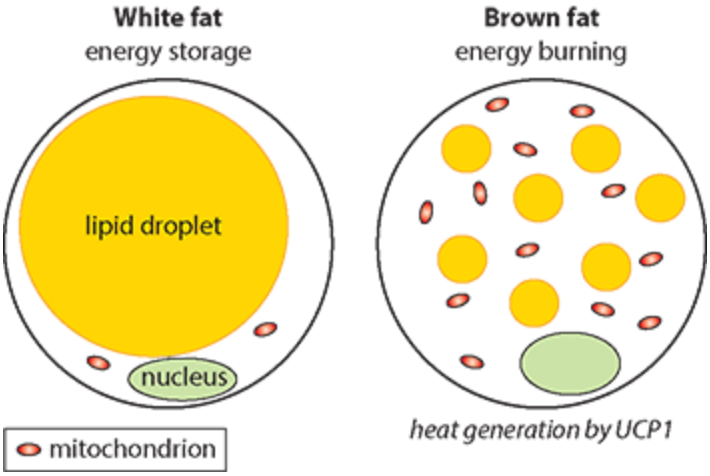


Figure 1.5.5: Comparison of white and brown adipose cells.

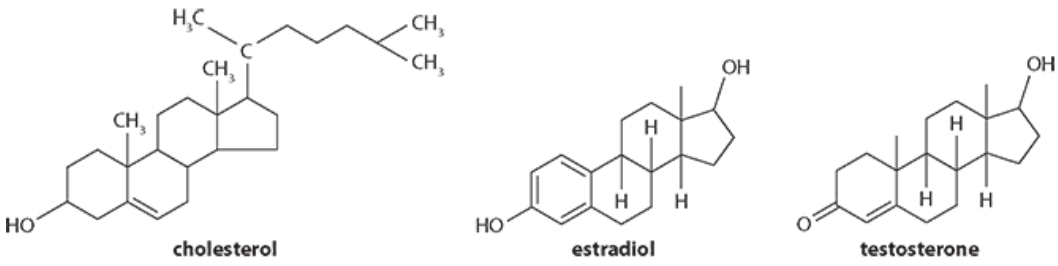


Figure 1.5.6: Structure of cholesterol and two important steroid hormones.

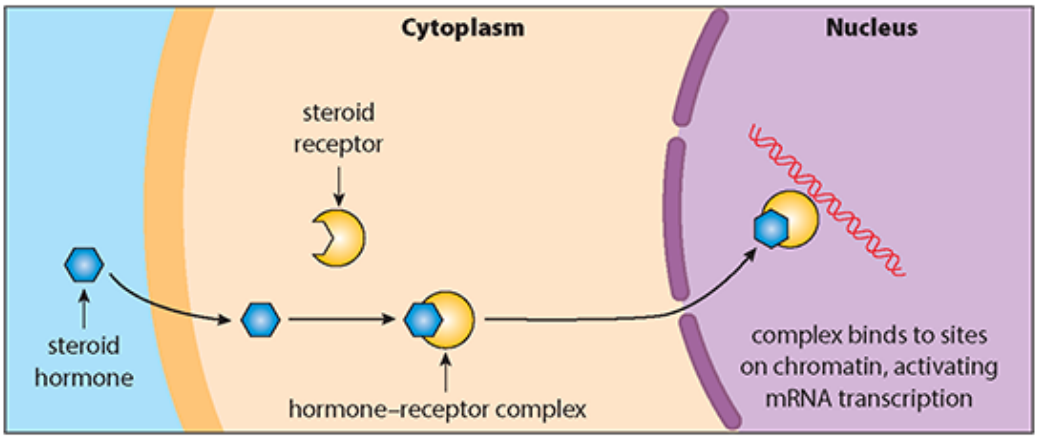


Figure 1.5.7: Steroid hormones enter cells and bind with special receptor molecules.

See Table 1.5.4 for a summary of the key properties of lipids.

energy content	lipids contain more energy per gram than carbohydrates, so lipid stores are lighter than carbohydrates storing an equivalent amount of energy.
density	lipids are less dense than water, so fat stores help large aquatic animals to float.
solubility	lipids are non-polar, insoluble molecules so they do not affect the movement of water in and out of cells by osmosis. Steroid hormones can pass through cell membranes.
insulation	lipids are also important in providing heat insulation. Fat stored under the skin reduces heat loss and is vital for animals such as seals, polar bears and whales, which live in cold conditions.

Table 1.5.4: The important properties of lipids that suit them to particular roles in living organisms.

TEST YOUR UNDERSTANDING

29 Name the type of reaction that releases energy from lipids.

30 State two roles for adipose tissue.

31 Outline the structure of a steroid hormone.

32 Why can steroid hormones pass through cell membranes?

Links

- How are the proportions of C, H and O different in carbohydrates and lipids? (Chapter 1)
- How do the properties of phospholipids contribute to membrane structure and properties? (Chapter 5)
- How do lipids influence temperature regulation in animals? (Chapter 8)

1.6 Proteins

LEARNING OBJECTIVES

In this section you will:

- define the bond linking two amino acids as a peptide bond
- discover that only 20 of the 500 amino acids found in nature are used to build polypeptides
- learn that some amino acids cannot be synthesised in the body and must be obtained from food
- learn that amino acids can be linked in any sequence so living organisms can synthesise many different proteins
- discover that only complex proteins have quaternary structure
- learn how proteins are denatured by changes in pH or temperature

- learn how polar and non-polar amino acids influence the folding of tertiary structure
- learn how tertiary structure is held in place by disulfide bridges
- discover that conjugated proteins are formed as subunits are bound together, often involving a prosthetic group

- Recognise collagen as a non-conjugated protein and hemoglobin as a conjugated protein
- learn that a range of amino acids exist with properties that are determined by their R groups
- learn that there are four levels of structure in proteins: primary, secondary, tertiary and quaternary
- learn that proteins may have a fibrous or globular form

GUIDING QUESTIONS

- How do the components of a protein combine to form a functional protein?
- How is the function of a protein determined by its structural components?

1.6.1 Polypeptides

Polypeptides are built up from amino acid monomers during condensation reactions (Section 1.2 and Figure 1.6.6). Two amino acids are joined with a reaction between the amino ($-NH_2$) group of one amino acid and the carboxyl (-COOH) group of the other, forming a **peptide bond** and producing a dipeptide. If further condensation reactions occur, a series of amino acids can become joined to form a polypeptide. The covalent bonds linking the amino acids produce what is known as the **primary structure** of any protein that will be formed from the polypeptide.

In living cells, polypeptides are synthesised by ribosomes in the cytoplasm. Just 20 different amino acids are used to construct polypeptides. Nine of these amino acids, known as essential amino acids, have to be obtained from our food. The others, known as non-essential amino acids, can be made in the body. Polypeptides can consist of up to 400 amino acids and, because these can be linked together in any sequence, there is a huge range of possible polypeptides. Some amino acids may also be modified once the polypeptide has been incorporated into a protein molecule (Section 1.6.3) so that even more different structures can be formed.

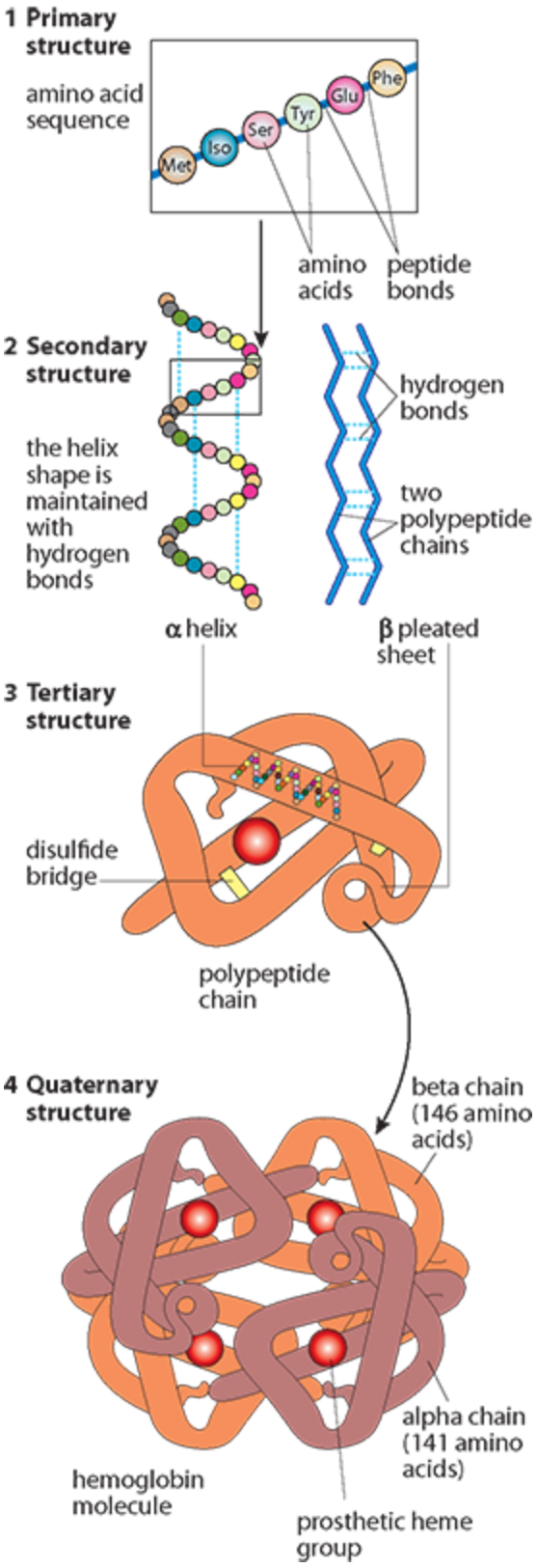

1 Primary structure
amino acid sequence
Met
Iso
Ser
Tyr
Glu
Phe
amino acids
peptide bonds
2 Secondary structure
the helix shape is maintained with hydrogen bonds
hydrogen bonds
two polypeptide chains
α helix
β pleated sheet
3 Tertiary structure
disulfide bridge
polypeptide chain
4 Quaternary structure
beta chain (146 amino acids)
alpha chain (141 amino acids)
hemoglobin molecule
prosthetic heme group

Figure 1.6.1: The structure of hemoglobin. Hydrogen bonds form between the amino acids in the polypeptide chain to form secondary structure and further folding can produce tertiary structure.

EXAM TIP

Make sure you can draw a molecular diagram to show how a peptide bond is formed between two amino acids. Remember that it is a condensation reaction, so water is produced.

The sequence of amino acids in a polypeptide is coded for by an organism's genes. Genes consist of a series of codons, each of which carries the specific code for one amino acid (Chapter 4). The sequence of codons in a gene is used as a template to direct the sequence in which the amino acids will be assembled.

1.6.2 Building a protein

Proteins are large, complex molecules, usually made up of hundreds of amino acid subunits. The way these subunits fit together is highly specific to each type of protein, and is vital to its function. Figure 1.6.1 illustrates the structure of the protein hemoglobin.

The first stage of protein production is the assembly of a sequence of amino acids to form the primary structure of a protein. A protein may either consist of one polypeptide or several linked together.

The basic chain of amino acids in a polypeptide folds and becomes a three-dimensional shape once it is complete. The shape, known as **secondary structure**, results from the formation of hydrogen bonds between different parts of the chain. The secondary structure forms as the polypeptide chain takes up a permanent folded or twisted shape. Some polypeptides coil to produce an α helix; others fold to form β pleated sheets. The most common shape is an α helix, held together by weak hydrogen bonds between the amino acids that lie in the turns of the structure. One section of a polypeptide may become an α helix while another takes up a β pleated form. The final shape will depend on the sequence of amino acids in the polypeptide.

KEY POINTS

peptide bond is a covalent bond between two amino acids.

polypeptide is a chain of amino acids formed by condensation reactions.

primary structure is the sequence of a polypeptide and number of amino acids in the molecule.

secondary structure is the three-dimensional form of sections of a protein. The two most common structural forms are α helices and β pleated sheets.

Living organisms synthesise many different proteins with many different functions (Table 1.6.1). Keratin, a structural protein found in hair, is an α helix. In other proteins, such as silk, polypeptides in parallel chains are linked to form flat, folded shapes known as β pleated sheets (Figure 1.6.1).

Further folding of polypeptide chains occurs in some proteins. This additional folding creates tertiary structure. Interactions occur between the R groups of the amino acids (Figure 1.6.2) and also within the polypeptide chain. The protein takes up a three-dimensional shape, which is held together by ionic bonds between particular R groups, as well as disulfide bridges (covalent bonds) between sulfur atoms of some R groups, and by weaker interactions between hydrophilic and hydrophobic side chains. Figure 1.6.3 shows the different types of bond involved in maintaining the tertiary structure of proteins. Tertiary structure is very important in enzymes (which are protein molecules) because the shape of an enzyme molecule gives it its unique properties and determines which substrates can fit into its active site (Section 3.1).

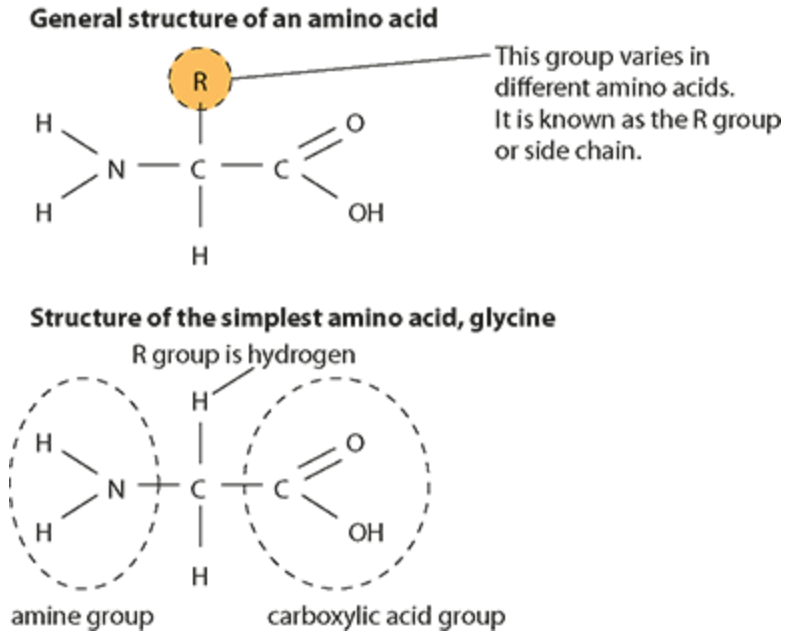


Figure 1.6.2: The general structure of an amino acid and the structure of glycine.

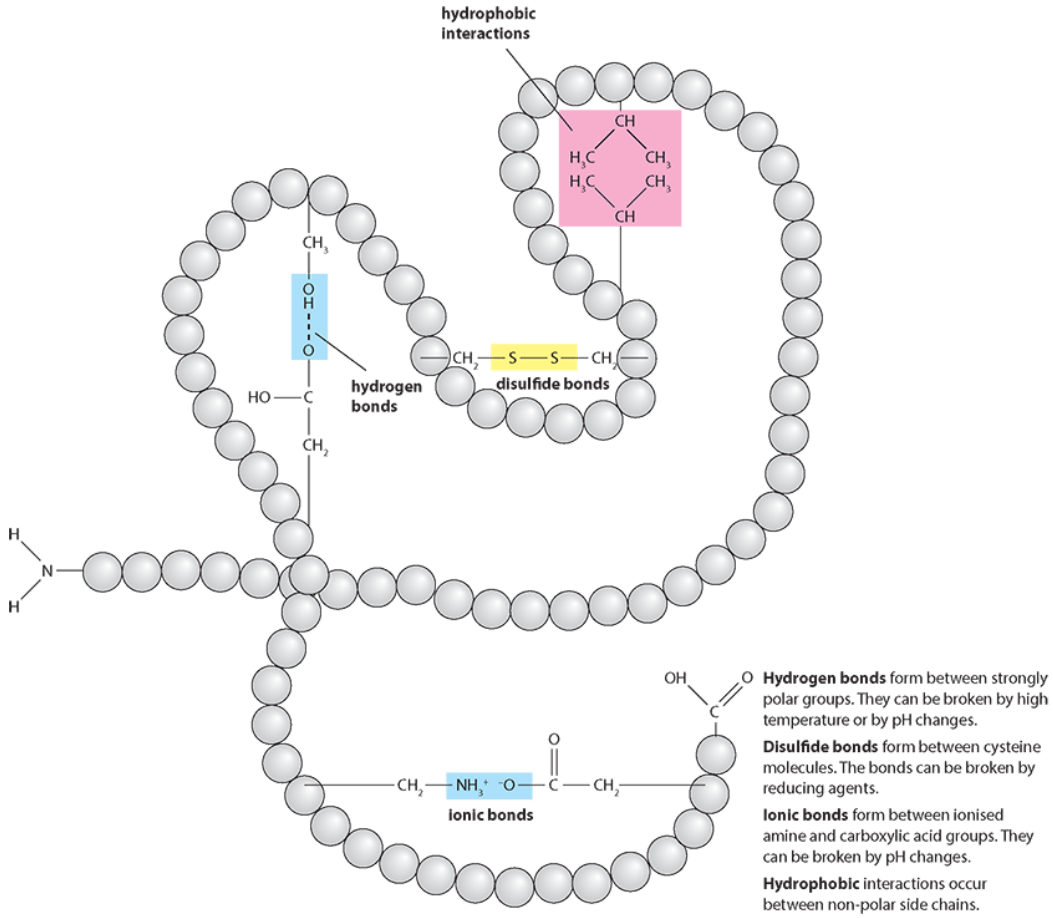


Figure 1.6.3: Types of bond that are important in protein structure.

Some proteins are composed of two or more polypeptides linked together and are said to have the final level of protein structure called quaternary structure. Quaternary structure links the polypeptide chains to form a single, large, complex protein. All the bonds that are important in the previous levels of structure hold the quaternary structure together. Examples of proteins that have quaternary structure are collagen Figure 1.6.4 (which has three polypeptide chains), hemoglobin (which has four), antibodies (which also have four) and myosin (which has six).

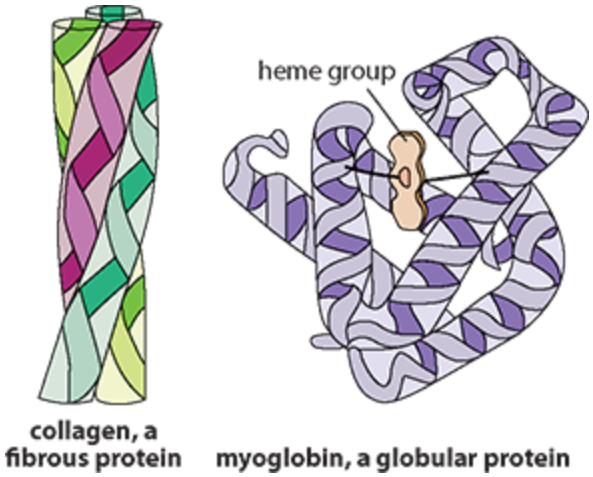


Figure 1.6.4: Structure of a fibrous and a globular protein.

TEST YOUR UNDERSTANDING

33 Draw the structure of a single amino acid. Use R to represent the chemically variable group.

34 Name the type of bond that links adjacent amino acids together.

35 Complete this sentence: Hydrogen bonds between amino acids cause a polypeptide to form a secondary structure and fold into either a ……. or a ……. .

36 What type of proteins are enzymes?

37 Which of the following is not a property of a fibrous protein?

A insoluble

B parallel polypeptide chains

C important in forming enzymes

D tough but may be supple and stretchy

KEY POINT

quaternary structure of proteins made up from the association of several amino acid chains or subunits in a closely packed arrangement.

The pigment hemoglobin found in red blood cells has four subunits and the positioning of these subunits is very important for the role of hemoglobin in carrying oxygen (Section 8.3).

1.6.3 Fibrous and globular proteins

Protein molecules are categorised into two major groups by their shape:

- **Fibrous proteins** have structural roles in the body, they are long molecules and examples include collagen (Table 1.6.1) and keratin. Fibrous proteins are usually insoluble in water and, usually, have secondary structure with many α-helices with hydrogen bonds that give stability to the molecules.
- **Globular proteins** have a more complex rounded, three-dimensional shape and have either tertiary or quaternary structure. They have functional roles in the body. Most globular proteins are soluble in water. Globular proteins include enzymes, such as pepsin, and antibodies as well as two important respiratory proteins, myoglobin and hemoglobin.

Table 1.6.1 summarises the nature, shape and function of some important proteins.

Protein type	Shape	Function
insulin (globular hormone)	Active insulin has two polypeptide chains linked and held in place by disulfide bridges.	Produced by the pancreas, insulin stimulates the liver to take up glucose from the blood and store it as glycogen.

collagen (fibrous, structural)	Individual molecules are bound together by cross-linking covalent bonds. Collagen consists of three polypeptides wound around one another.	Present in connective tissue, builds muscle, tendons, ligaments giving them tensile strength, and gives elasticity to the skin of vertebrates.
immunoglobulin G (IgG) (globular)	Composed of two types of protein chain: two heavy chains and two light chains joined by disulfide bonds. This shape provides two antigen-binding sites so that antibody molecules can cross-link to antigens and hold them securely.	Infection control. Antibody produced by the immune system. Binds antigens from pathogens such as viruses, bacteria and fungi to fight infection.
DNA helicase (globular, enzyme)	Compact globular shape with a hexagonal arrangement of six identical subunits.	Separates double-stranded DNA to single strands allowing each strand to be copied during replication.

Table 1.6.1: The shape and functions of different proteins

1.6.4 Denaturation

Denaturation destroys the complex structure of a protein. Heat or the presence of strong acids or alkalis can all disturb the bonds between the different parts of a protein molecule and disrupt its structure. The primary structure of the protein will remain but secondary, tertiary and quaternary structures are usually lost. A denatured protein will have different properties from the original molecule. Enzymes are easily denatured by extremes of pH or temperature (Section 3.1) and lose the ability to function as catalysts. Some proteins lose their solubility or aggregate to form clumps as they denature and this can be observed during cooking. The heat used to cook meat denatures the proteins found in it so that its texture is changed. Egg becomes hardened as the proteins are denatured by heating.

1.6.5 Polar and non-polar amino acids

Amino acids are divided into two groups according to the chemical properties of their side chains or R groups (Figure 1.6.5). Polar and non-polar amino acids have different properties and their positions in a protein molecule affect the behaviour and function of the whole protein. The **tertiary structure** of proteins is influenced by the formation of disulfide bridges between these **R groups** (Figure 1.6.3).

Amino acids with non-polar side chains are hydrophobic. Those with polar side chains are hydrophilic. Non-polar amino acids are found in parts of proteins that are in hydrophobic areas, while polar amino acids are in areas that are exposed to an aqueous environment such as cytoplasm or blood plasma.

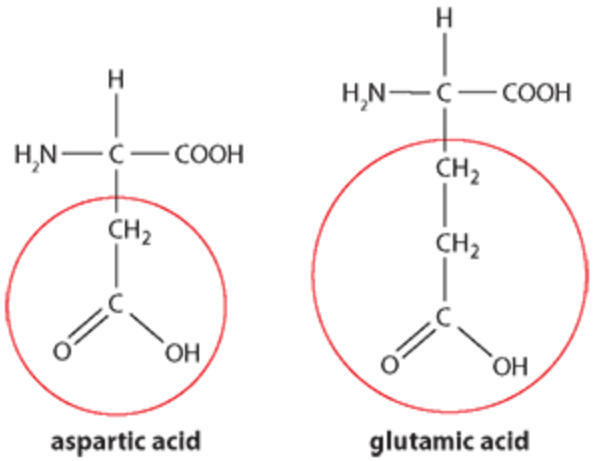


Figure 1.6.5: Aspartic acid and glutamic acid have side chains that are polar.

KEY POINTS

R group is a term used to abbreviate a group attached to a large biological molecule. Six of the 20 amino acids have

hydrocarbon R groups.

tertiary structure is the three-dimensional structure of a protein due to interactions between the R groups of the amino acids.

Proteins are found in cell membranes and it is the polar hydrophilic amino acids that are found on the outer and inner surfaces where they are in contact with the aqueous environment, while the non-polar hydrophobic amino acids are embedded in the core of the membrane in contact with the hydrophobic tails of the phospholipid bilayer (Figure 1.6.6, see Section 6.1). This helps to hold the protein in place in the membrane. Some integral proteins act as channels, and the pore is lined with hydrophilic amino acids to enable polar substances to pass through.

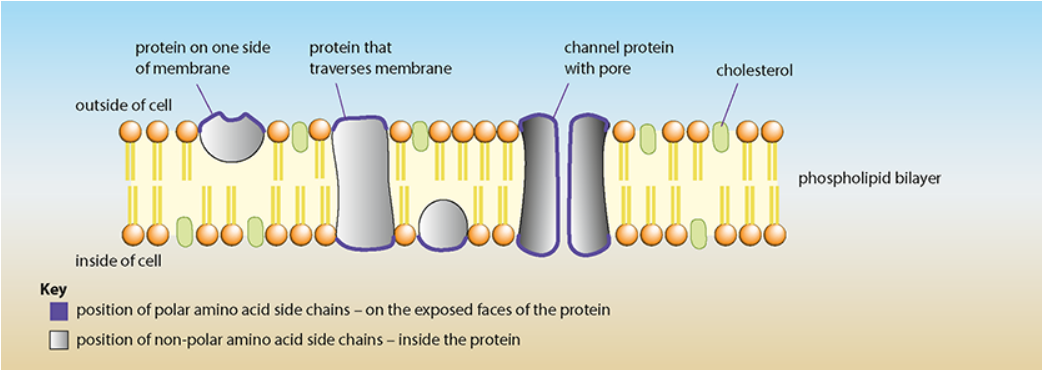


Figure 1.6.6: In membrane proteins, polar (hydrophilic) amino acids are found on the surfaces in contact with the aqueous environment, while non-polar (hydrophobic) amino acids are embedded inside the phospholipid bilayer.

Polar and non-polar amino acids are also important in enzymes, where they assist in the binding of substrates. An enzyme that acts on a polar substance (for example, amylase) has polar

amino acids in its active site, whereas lipases have non-polar amino acids in the active site. These amino acids help to form the temporary bonds between the enzyme and its substrate.

Polar amino acids on the surface of a protein increase its solubility while non-polar amino acids help a protein maintain its structure.

1.6.6 Prosthetic groups

Many proteins contain **prosthetic groups** and those that do are called **conjugated proteins**. Prosthetic groups are non-protein groups that are able to bind to different proteins or parts of them. We can see two examples of prosthetic groups in the respiratory pigments myoglobin (Figure 1.6.4) and hemoglobin which both contain a prosthetic heme group. Hemoglobin consists of four polypeptide chains, each one containing a heme group. The heme group (Figure 1.6.1) consists of a central Fe (iron) atom and a porphyrin ring. The prosthetic heme group is vital to the structure of hemoglobin because the shape of the whole protein is changed as oxygen binds to it. The iron group not only allows oxygen to bind but also holds the compact structure with four subunits in place and allows for progressively easier oxygenation as more oxygen molecules bind to the protein. Proteins with no prosthetic group are called **non-conjugated proteins**. Insulin (Fig 1.6.7) is an example of a non-conjugated protein.

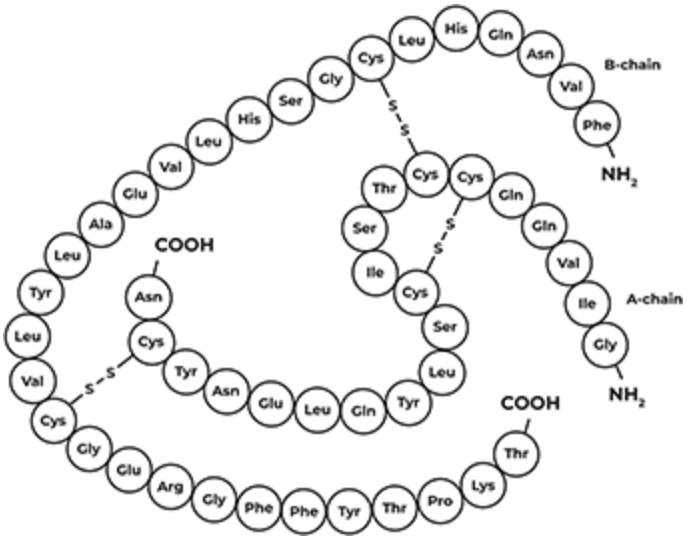


Figure 1.6.7: Insulin is a non conjugated protein

NATURE OF SCIENCE

Looking for trends and discrepancies: do all organisms use only 20 common amino acids in their proteins?

Humans can make 11 of the 20 amino acids we need to build proteins but we do not have the enzymes needed for the biosynthesis of the others. Plants, on the other hand, must be able to make all the amino acids they require.

Researchers have also investigated the trends in amino acid compositions of proteins found in species of the important kingdoms of Archaea, Bacteria and Eukaryotes. International databases ProteomicsDB and SWISS-PROT (which contain information about the structure and composition of proteins) can compare amino acid frequencies for 195 known proteomes and all recorded sequences of proteins. They discovered that the amino acid compositions of proteins do differ substantially for different kingdoms.

In addition to the variations in amino acids in proteins, some microorganisms and plants are able to make so called 'non-standard' amino acids by modifying standard amino acids. Some species are also able to synthesise many uncommon amino acids. For example, some microbes synthesise lanthionine, which is a modified version of the amino acid alanine. Many other proteins are modified after they have been produced. This 'post-translational modification' involves the addition of extra side groups to the amino acids in a protein.

Considering all the evidence, it seems that, although we can observe many similar proteins in different species, we cannot always say that the same amino acids are used in their

construction. The range of amino acids in proteins can vary considerably from species to species.

To consider:

1 What contribution have international databases made to our understanding of protein structure?

2 How can comparing proteomes and amino acids in different organisms help our understanding of evolutionary relationships?

As Table 1.6.1 shows, a wide range of proteins is found in different organisms and each protein has its own structural or biochemical function. Every individual organism has its own unique proteins, which are determined by its unique genome. The proteins found in an organism are known as its proteome, a term derived from a combination of the words 'protein' and 'genome'.

KEY POINT

proteome the complete set of proteins expressed by a genome.

TEST YOUR UNDERSTANDING

38 Define a prosthetic group.

39 Outline the importance of hydrophilic amino acids in the tertiary structure of a protein.

40 State the difference between a conjugated and non-conjugated protein

Links

- How do proteins interact with phospholipid bilayers? (Chapter 6)
- How does quaternary structure enable hemoglobin to collect oxygen? (Chapter 8)
- How do protein hormones influence cell activity? (Chapter 9)

1.7 Nucleic acids

LEARNING OBJECTIVES

In this section you will:

- recall that DNA is the genetic material of living organisms understand the components of a nucleotide
- learn that DNA and RNA are polymers of nucleotides with a sugar phosphate backbone
- understand that the bases in each nucleic acid form the basis of the geneic code
- recall the important differences between DNA and RNA
- understand the role of complementary base pairing in the replication and expression of genetic material
- understand that the genetic code is universal and diversity results from the enormous number of base sequences that are possible

- learn how 5’ to 3’ linkages in the sugar phosphate backbone are significant for transcription and translation
- learn that DNA is a double helix kept stable by purine to pyrimidine bonds
- understand the structure of a nucleosome

- learn that Hershey and Chase provided evidence that DNA is the genetic material
- consider Chargaff's data on the ratio of pyrimidine to purine bases in many organisms.

GUIDING QUESTIONS

- How do the structures of nucleic acids enable them to store genetic information?
- How can molecular modelling permit predictions about a molecule's function?
- How does DNA fit into the small volume of the nucleus?

1.7.1 Structure of DNA and RNA

DNA (deoxyribonucleic acid) molecules make up the genetic material of living organisms. DNA is an extremely long molecule but, like proteins and carbohydrates, it is built up of many monomer subunits. The monomer subunits of DNA are called nucleotides. **RNA (ribonucleic acid)** is also built up of many nucleotides but these differ from DNA nucleotides in the type of pentose sugar they contain and the bases that are attached to them.

Each **nucleotide** consists of three parts – a pentose (five-carbon) sugar (deoxyribose or ribose), a phosphate group and a nitrogenous base (Figure 1.7.1). DNA contains four different bases: adenine, guanine, cytosine and thymine. These are usually known by their initial letters: A, G, C and T (Figure 1.7.2). RNA also contains four bases, but in an RNA molecule thymine is not present and is replaced by uracil (U).

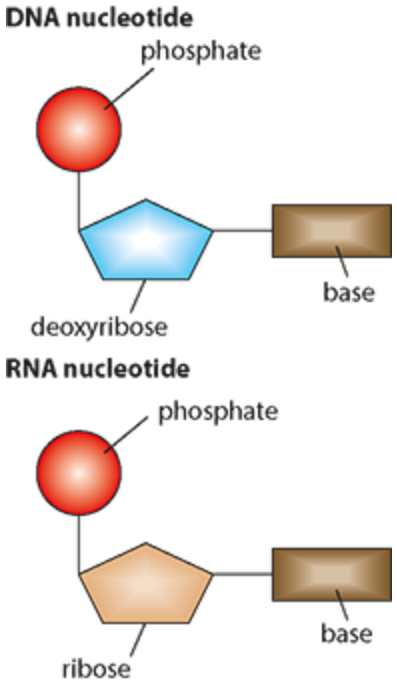


Figure 1.7.1: The general structure of DNA and RNA nucleotides.

To form a DNA molecule, nucleotide monomers are linked together. The phosphate group of one nucleotide links to the deoxyribose of the next molecule to form a chain of nucleotides, as shown in Figure 1.7.2. The sugar and phosphate groups are identical all the way along the chain and form the 'backbone' of the DNA molecule. The sequence of bases in the chain will vary and it is this sequence that forms the genetic code determining the characteristics of an organism.

Different organisms have DNA of different lengths and any base sequence is possible, this gives a DNA molecule a large capacity for storing genetic information.

The genetic code carried in the sequence of bases that make a DNA molecule is said to be **universal** because it is similar across all forms of life.

Two strands of nucleotides are linked by hydrogen bonds that form between the bases and this two-stranded structure makes up the double helix of a complete DNA molecule. Adenine always pairs with thymine and is bonded with two hydrogen bonds, while cytosine is paired with guanine by three hydrogen bonds. The arrangement is known as **complementary base pairing**. Notice that the two DNA chains run in opposite directions and are said to be **antiparallel**.

KEY POINTS

DNA (deoxyribonucleic acid) is the basic material of inheritance, contained in the nucleus in eukaryotes and the cytoplasm of prokaryotes; DNA consists of two strands of

nucleotide subunits containing the bases adenine, thymine, guanine and cytosine.

nucleotide is the basic chemical unit of a nucleic acid – an organic base combined with pentose sugar (either ribose or deoxyribose) and phosphate.

RNA (ribonucleic acid) a nucleic acid that contains the pentose sugar ribose and bases adenine, guanine, cytosine and uracil.

complementary base pairing is pairing of bases A–T and G–C in double-stranded DNA, and of A–U and C–G between DNA and RNA during transcription, and between tRNA and mRNA during translation.

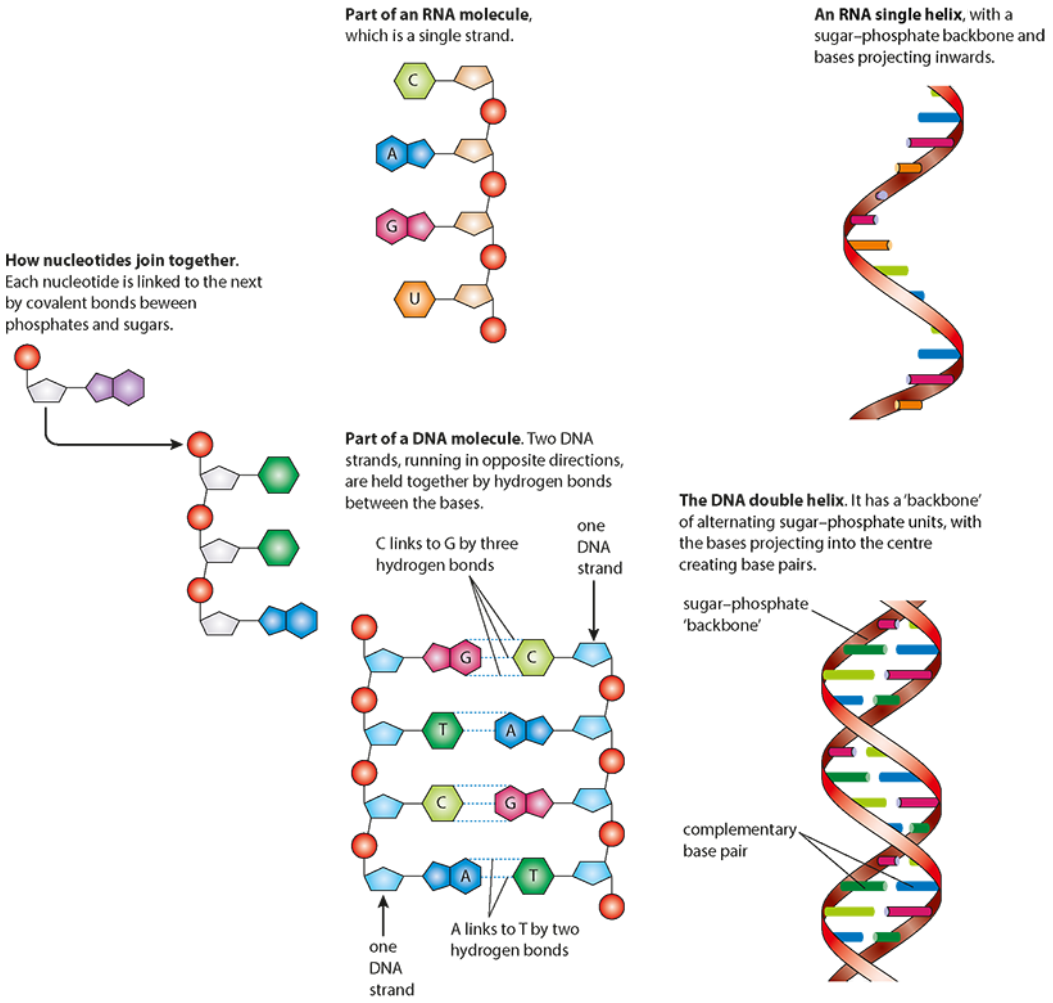


Figure 1.7.2: The structure of DNA and RNA.

KEY POINT

antiparallel means running in opposite direction; the two polynucleotide strands in a DNA molecule are antiparallel.

You can imagine the molecule rather like a rope ladder with the sugar phosphate backbone being the sides of the ladder and the rungs being formed by the hydrogen-bonded base pairs. To form the characteristic double helix of a DNA molecule, the ladder must be twisted to resemble a spiral staircase.

To form a molecule of RNA, nucleotide monomers are linked in a similar way to those of DNA. In the case of RNA the molecule remains single stranded and the bases it contains do not bond with bases in other RNA molecules.

Table 1.7.1 outlines the similarities and differences between the DNA and RNA molecules.

EXAM TIP

Make sure you can draw a simple diagram of DNA and RNA nucleotides using circles, pentagons and rectangles to represent the components.

Always draw DNA structure so that it shows two antiparallel strands and correctly paired bases. You will not be asked to show the helical shape. Try to think of a mnemonic that will help you remember the pairings of bases in DNA, for example Apple–Tart for adenosine–thymine and Chocolate–Gateau for cytosine–guanine.

DNA	RNA
contains the five-carbon sugar deoxyribose	contains the five-carbon sugar ribose
contains the bases adenine, guanine, cytosine and thymine (A, G, C, T)	contains the bases adenine, guanine, cytosine and uracil (instead of thymine) (A, G, C, U)
a double-stranded molecule	a single-stranded molecule

Table 1.7.1: A comparison of the structures of DNA and RNA.

1.7.2 Complementary base pairing and DNA replication

DNA must replicate itself accurately when a cell divides, the genetic code it carries can be passed on to the daughter cells. **DNA replication** copies DNA precisely so that new molecules are produced with exactly the same sequence of bases as the original strands.

As Figure 1.7.3 shows, this process does not occur in a haphazard manner. An enzyme called helicase unzips one region of the DNA molecule and nucleotides are added in a step-by-step process that links them to one another and to their complementary bases with hydrogen bonds in an area known as the replication fork. Details of the process are discussed in Section 3.1.

The two new DNA strands that are produced are absolutely identical to the original strands. Complementary base pairing between the template strand and the new strand ensures that an accurate copy of the original DNA is made every time replication occurs. DNA replication is said to be **semi-conservative replication** because no DNA molecule is ever completely new. Every double helix contains one ‘original’ and one ‘new’ strand.

KEY POINT

DNA replication is the production of two new strands of DNA from one original molecule.

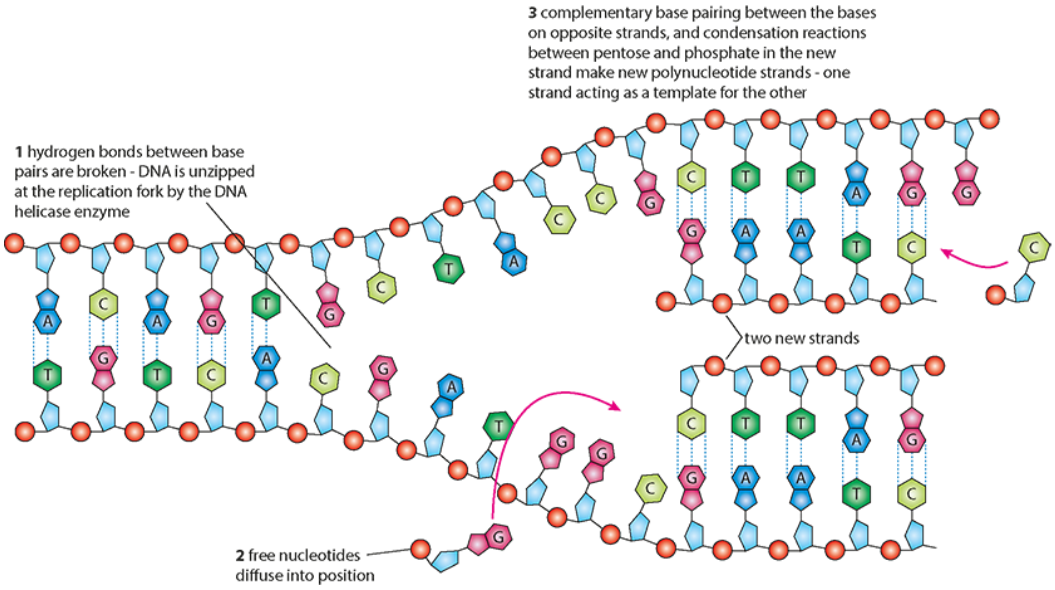


Figure 1.7.3: DNA replication.

KEY POINT

semi-conservative replication is when each of two partner strands of DNA in a double helix acts as a template for a new strand; after replication each double helix consists of one old and one new strand.

THEORY OF KNOWLEDGE

Collaboration versus competition

The story of the discovery of DNA's structure (see Nature of Science, Careful observation: the discovery of DNA) illustrates how important collaboration can be in scientific discovery. Cooperation and competition can both occur between research groups.

To consider:

1 Is keeping research discoveries secret ‘anti-scientific’? Can it ever be justified?

2 How are shared and personal knowledge related in scientific research?

Complementary base pairing is also crucial to the process of gene expression. The genes carried in the DNA genetic code are eventually converted into the polypeptides and proteins that the cell needs. Complementary base pairing between DNA and RNA produces a single strand of messenger RNA (mRNA) in a process known as transcription (Section 4.2). This mRNA strand acts as a template for protein production in the cytoplasm.

1.7.3 DNA packaging in the nucleus

DNA is a very long molecule composed of many thousands of nucleotides. During DNA replication, DNA unwinds so it can be copied. At other times in the cell cycle (Section 6.5), DNA also unwinds so that its instructions can be used to make proteins. But as a cell prepares to divide, each DNA molecule is coiled into a compact chromosome so that it can be passed to new daughter cells.

KEY POINTS

chromosome in eukaryotes, a structure consisting of a long thread of DNA and protein that carries the genetic information of the cell; in bacteria, the DNA molecule that contains the genetic information of the cell.

histone is one of a group of basic proteins that form nucleosomes and act as scaffolding for DNA.

nucleosome refers to a part of a eukaryotic chromosome, made of DNA wrapped around histone molecules and held in place by another histone protein.

In eukaryotes, DNA is found in the nucleus surrounded by a nuclear membrane. Because the cell is very small, and because organisms have many DNA molecules per cell, each DNA molecule must be tightly packaged. This packaged form of the DNA is called a chromosome. DNA is coiled and then coiled again around a structure known as a **nucleosome** (see the Higher Level section on DNA structure and replication). A nucleosome consists of DNA associated with proteins known as histones. Histones have three key functions:

- packaging of DNA
- gene regulation
- supercoiling DNA during cell division.

Prokaryotes do not have **histone** proteins associated with their DNA. It remains free in the cytoplasm and is known as ‘naked’ DNA.

TEST YOUR UNDERSTANDING

41 State two differences between DNA and RNA.

42 Draw the structure of a DNA nucleotide.

43 Outline the importance of nucleosomes in the nuclei of eukaryotes.

1.7.4 DNA structure and replication

Nucleic acids are very large macromolecules composed of a backbone of sugar and phosphate molecules each with a nitrogenous base attached to the sugar. Here, the detailed structure of different nucleic acids is considered.

The 3′–5′ linkage

A DNA nucleotide consists of the sugar deoxyribose to which are attached a phosphate group and a nitrogenous base. The carbons in the sugar are numbered from 1 to 5 in a clockwise direction starting after the oxygen at the apex (Figure 1.7.4).

- The base is attached to carbon 1.
- Carbon 2 has just a hydrogen attached instead of an OH group, this is the reason the sugar is called *deoxy*ribose.
- Carbon 3 is where the next nucleotide attaches in one direction.
- Carbon 5 has a phosphate group attached to it, which is where the next nucleotide attaches in the other direction.

This means that each nucleotide is linked to those on either side of it through carbons 3 and 5. The linkages are called **3′–5′ linkages**.

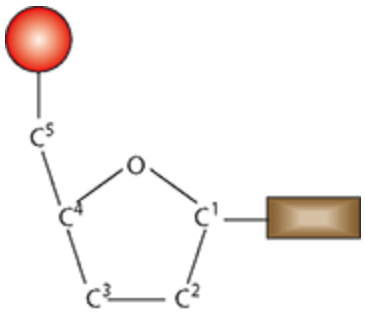


Figure 1.7.4: The structure of a nucleotide.

KEY POINT

3′–5′ linkage bond between the 3′ carbon atom of one sugar molecule and the 5′ carbon atom of another; found in DNA and RNA molecules.

Antiparallel strands

Look back at Figure 1.7.2. This shows part of a DNA molecule in which two polynucleotide strands, running in opposite directions, are held together by hydrogen bonds between pairs of bases. Notice that the deoxyribose molecules are orientated in the opposite directions. Figure 1.7.5 also shows this: at one end of each DNA strand there is a free 3′ carbon and at the other there is a free 5′ carbon. (Ignore the fact that there is a phosphate group attached to this 5′ carbon.) One strand runs in one direction whereas the other runs in the opposite direction. The strands are described as being antiparallel.

DNA bases and hydrogen bonding

The four DNA bases fall into two chemical groups called **pyrimidines** and **purines**. Cytosine and thymine are pyrimidines, and adenine and guanine are purines.

Cytosine pairs with guanine and thymine pairs with adenine, that is, a pyrimidine always pairs with a purine. This is because they are different sizes: pyrimidines are smaller than purines. The pairing of a pyrimidine with a purine ensures that the strands are always the same distance apart (Figure 1.7.5). Understanding that DNA strands are antiparallel and that hydrogen bonds

between the bases can be broken were crucial steps in working out how DNA is replicated.

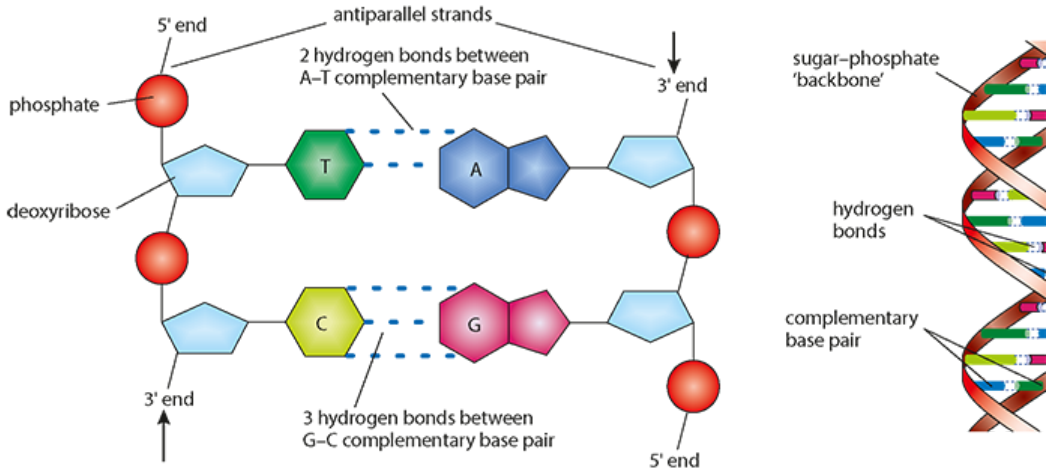


Figure 1.7.5: Hydrogen bonding between antiparallel strands of DNA.

NATURE OF SCIENCE

Careful observation: the discovery of DNA

James Watson and Frances Crick, together with Maurice Wilkins, were awarded the Nobel Prize for Physiology or Medicine in 1962 for their discovery of the structure of DNA (Figure 1.7.6). Watson and Crick put forward their theory for DNA structure in 1953. They based their ideas on the work of an American chemist, Erwin Chargaff, who calculated the proportions of the bases in DNA. Chargaff noticed that DNA varies from species to species and suggested that DNA was genetic material. He also observed that the bases in DNA appear in a 1 : 1 ratio; the number of guanine bases in DNA is equal to the number of cytosine bases, and the number of adenine bases is equal to the number of thymine bases. This suggested the base pair pattern of DNA, and helped Watson and Crick work out how the bases fitted into the double helix.

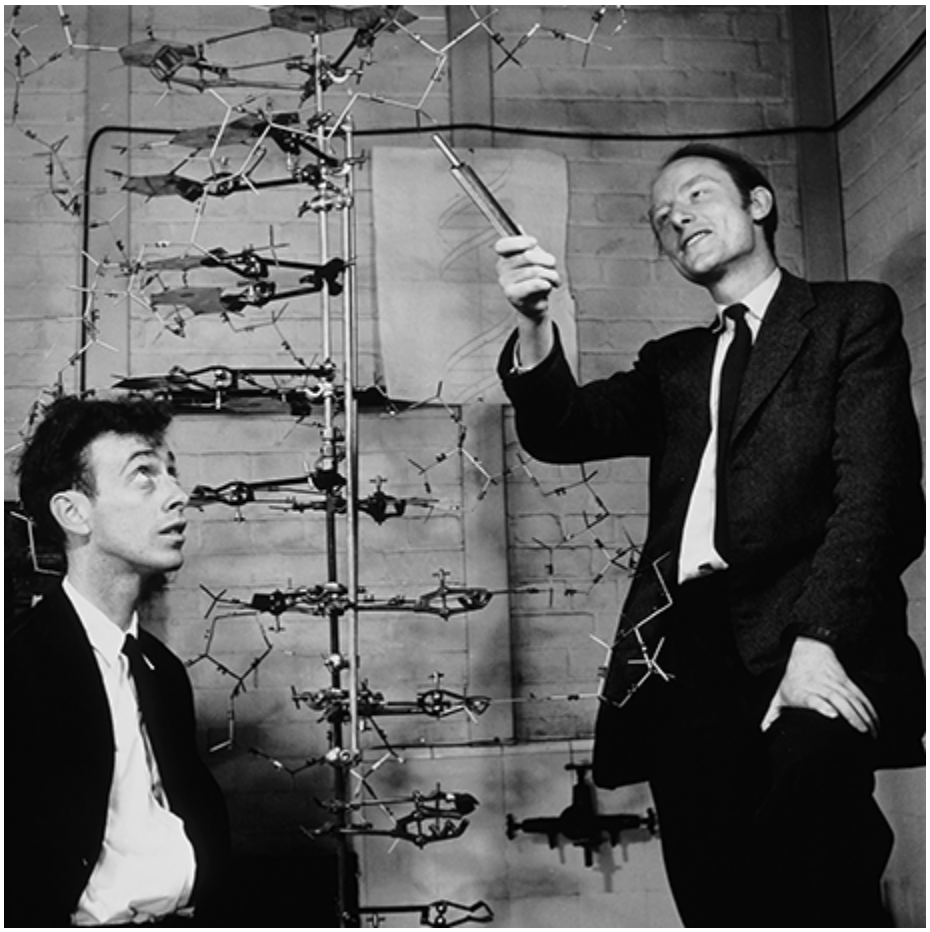


Figure 1.7.6: Watson and Crick built a 3D model to help formulate their proposal for the structure of DNA.

Watson and Crick suggested that DNA was composed of two parallel strands held together by pairs of bases, A pairing with T, and C with G. At the same time, in different laboratories, other researchers were trying to work out DNA's three-dimensional structure using X-ray diffraction. Rosalind Franklin (Figure 1.7.7) and Maurice Wilkins spent many hours trying to interpret photographs of diffraction patterns produced by DNA. From careful observation and calculation of the positions of certain markers on the X-ray photographs, Watson and Crick finally worked out that the two chains of DNA were wrapped into a double helix, linked at regular

intervals by the bases of the nucleotides. Furthermore, they were able to see from a model of DNA that there are 10 nucleotides per turn of the helix. It is important to remember that their achievement would not have been possible without the data, which were provided by the other scientists working on DNA structure at the time.

Figure 1.7.7: Rosalind Franklin was an expert in the field of X-ray crystallography. Her skill and careful observations enabled her to work out that the phosphate groups of DNA are found on the outside of the molecule. She died at the age of 37 before the Nobel Prize was awarded to Watson, Crick and Wilkins. Nobel prizes cannot be awarded posthumously.

To consider:

1 Inference, creativity and imagination are important in scientific discovery. Why was collaboration so important

in the discovery of the structure of DNA?

2 How important are awards and prizes for scientists who made new discoveries?

The accuracy of DNA replication

When Watson and Crick proposed their double-helix model for the structure of DNA in 1953, one of the most striking things they realised was that it immediately suggested a way that DNA could be replicated. If the two strands were unwound each one could provide a template for the synthesis of a new strand.

An essential feature of DNA is that it must be able to replicate itself accurately, so that when a cell divides, the genetic code it carries can be passed on to the daughter cells. In DNA replication, new molecules are produced with identical sequences of bases as the original strands (Figure 1.7.3). DNA replication takes place in the nucleus during interphase of the cell cycle when DNA is not tightly coiled (Section 6.5).

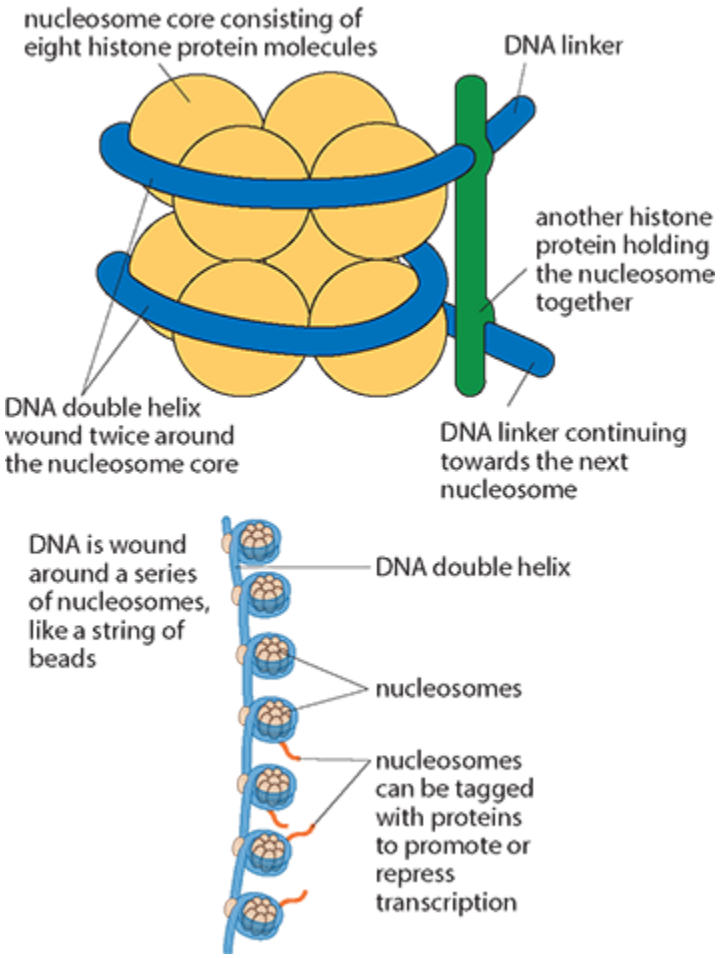


Figure 1.7.8: The structure of a nucleosome.

Nucleosomes

A eukaryotic chromosome is composed of a double strand of DNA combined with proteins. Some of these proteins, called histones, combine together in groups of eight to form a bead-like structure (Figure 1.7.8). The strand of DNA takes two turns around the first bead before continuing on to the next bead. It is held in place on the bead by a ninth histone. The group of nine histones with the DNA is called a nucleosome. The function of nucleosomes is to help supercoil the chromosomes during mitosis and meiosis and also to help regulate transcription

(Section 4.4). Nucleosomes are linked together by a section of linker DNA.

1.7.5 The Hershey and Chase experiments

Hershey and Chase used bacteriophages (viruses that infect bacteria) to provide evidence that DNA is the genetic material.

Although DNA has been well known to science since the 19th century, it is surprising to think that it was not until the middle of the 20th century that scientists discovered its role as the genetic material. Until that time, most people believed that proteins were the molecules responsible for inheritance. Then, in 1952, Alfred Hershey (1908–1997) and Martha Chase (1927–2003) conducted a series of experiments with T2 phage (a type of bacteriophage), which confirmed that DNA was the genetic material. They were able to carry out these investigations thanks to two relatively new techniques: electron microscopy and radioactive labelling.

The structure of the T2 phage had recently been revealed using the electron microscope. The virus injects its DNA into the cell it infects, but leaves behind its protein coat. Hershey and Chase labelled the viral DNA with radioactive ^{32}P. Phosphorus is present in DNA but not in amino acids, so as they followed the transfer of the labelled material into the cytoplasm of the bacterium, Hershey and Chase knew that it was only DNA that was being transferred.

They then labelled viruses with ^{35}S (sulfur is present in amino acids but not in DNA). After these viruses had infected the bacterial cells, Hershey and Chase examined the discarded protein coats and found that they contained radioactive sulfur, but the bacterial cytoplasm did not. These results supported their hypothesis that DNA is the genetic material that infects the

bacteria, and protein (found in the protein coats) was not (Figure 1.7.9).

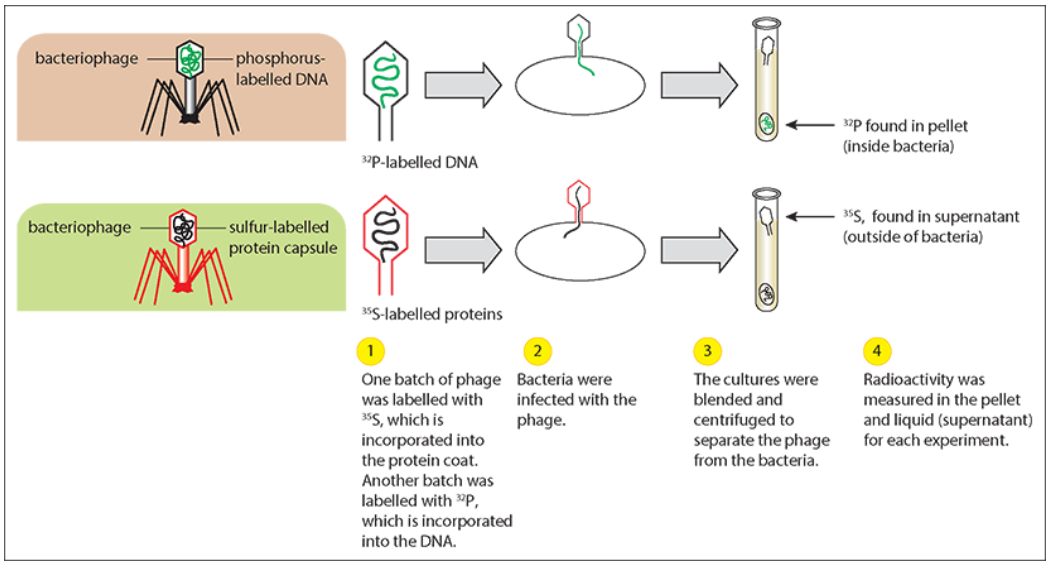


Figure 1.7.9: The Hershey and Chase experiment showed that the genetic material transferred to bacterial cells by infecting T2 phages is DNA, and not protein as previously believed.

EXAM TIP

Check your understanding of the Hershey and Chase experiment. Make sure you can interpret the results correctly. Try answering these questions.

1 What kind of virus was used in the experiment and why?

2 Why were two types of labelling used to identify the phages?

3 Why did Hershey and Chase conclude that DNA was the genetic material?

NATURE OF SCIENCE

Chargaff's rule

In the 1940s Austrian-born chemist Erwin Chargaff (1905 – 2002) discovered that in the double stranded DNA of any organism, the amount of guanine is equal to the amount of cytosine and that the amount of thymine is equal to the amount of adenine.

Chargaff's rule states that in any double stranded DNA molecule A%=T% and G%=C%. Chargaff's work began with the hypothesis that if DNA from different species had different biological activities, there should also be differences in the chemistry of their DNA. With careful experimentation he was able to disprove the tetra nucleotide hypothesis that was current at the time, and which stated that DNA was formed of a large number of repeats of a GACT tetramer. His work was a good example of how a new hypothesis supported by experiments can falsify an earlier hypothesis.

This discovery was crucial in helping Watson and Crick as they worked out the structure of the DNA molecule.

To consider:

Why do scientific discoveries often depend on research that falsifies a widely held theory?

TEST YOUR UNDERSTANDING

44 Draw a nucleotide of DNA and label the carbon atoms 1 to 5.

45 State the evidence that Erwin Chargaff produced that assisted in the elucidation of the structure of DNA.

46 How many histone proteins are present in a nucleosome?

47 Why was the discovery of radioactive isotopes important in proving that DNA is the genetic material?

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
define the terms organic and inorganic molecule	1.1.1			
define an organic molecule	1.1.1			
list the six most common elements in living things	1.1.2			
suggest uses for magnesium and calcium in living things	1.1.2			
list different functions for the element iron in plants and animals	1.1.3			
name an element that is toxic	1.1.4			

explain the terms polar, covalent and hydrogen bond in relation to water	1.2.1			
draw water molecules interacting with each other	1.2.1			
explain how water acts as a solvent for ionic and polar substances	1.2.2			
explain how water and plant sap transport dissolved molecules	1.2.2			
recognise that water is a metabolite and is used and produced during reactions	1.2.2			
define emergent properties	1.2.3			
explain how polarity gives water cohesive properties that support small	1.2.4			

organisms and allow water to travel in tubes				
explain how polarity give water adhesive properties and give an example of why these are important	1.2.4			
outline why a high specific heat capacity makes water temperature resistant to change	1.2.5			
describe how evaporation of water cools organisms	1.2.5			
explain why ice has a lower density than water and how ice helps aquatic organisms survive	1.2.5			
suggest how the transparency of water helps aquatic organisms survive	1.2.5			

outline some adaptations of organisms that live in aquatic habitats	1.2.6			
summarise a hypothesis to explain the origin of water on Earth	1.2.7			
suggest why scientists hunt for water when looking for extraterrestrial life	1.2.7			
outline how carbon atoms can form four covalent bonds to produce a variety of stable organic compounds	1.3.1			
state that carbohydrates, proteins, lipids and nucleic acids are organic compounds	1.3.2			
define a monomer and a polymer	1.3.3			

state what happens in a hydrolysis reaction	1.3.3			
state what happens during a condensation reaction	1.3.3			
identify carboxyl groups, amines and phosphate groups as functional groups found in organic molecules	1.3.4			
recognise how monosaccharide monomers arelinked together by condensation reactions to form disaccharides and polysaccharides	1.4.1			
describe how carbohydrates have a range of different forms which give them their properties	1.4.1			
recognise that large	1.4.2			

carbohydrates are less soluble than small ones				
state glycogen and starch are compact structures that make them suitable storage molecules	1.4.2			
outline the structure of cellulose and how it is related to its structural function	1.4.2			
state that carbohydrates are short-term energy stores which are metabolised to release energy	1.4.2			
explain how alpha-d-glucose and beta-d-glucose arebonded together to produce amylose, amylopectin and cellulose (extension)	1.4.2			
	1.4.2			

list examples of branched polysaccharides and state their functions				
recall that pentose sugars are found in DNA and RNA molecules	1.4.2			
name different forms of lipid and give examples of their properties	1.5.1			
describe the arrangement of molecules in a triglyceride	1.5.1			
state why lipids are hydrophobic	1.5.1			
explain the terms saturated, unsaturated and polyunsaturated	1.5.2			
state why lipids are efficient in storing energy	1.5.3			
describe the importance of adipose tissue	1.5.4.			

name two steroid hormones and outline their roles in the body	1.5.5			
draw the structure of an amino acid	1.6.1			
explain the importance of R groups in amino acids and draw a peptide bond between two amino acids	1.6.1			
define a polypeptide	1.6.1			
outline the four levels of structure found in proteins	1.6.2			
describe how hydrogen bonds and ionic bonds are important in secondary and tertiary structure	1.6.2			
outline the differences between fibrous and globular proteins	1.6.3			
explain how	1.6.4			

denaturation alters protein structure				
explain the difference between polar and non-polar amino acids and how they influence protein folding	1.6.5			
define the terms prosthetic group and conjugated protein using hemoglobin as an example	1.6.6			
draw diagrams of DNA and RNA monomers	1.7.1			
name the bases found in DNA and RNA	1.7.1			
summarise the differences between DNA and RNA	1.7.1			
draw diagrams of DNA and RNA to show the sugar phosphate backbone	1.7.1			

draw a simple diagram of DNA to show the double helix and antiparallel strands	1.7.1			
explain complementary base pairing and why it is important in replication of DNA	1.7.2			
state that there is enormous diversity in the combinations of DNA bases	1.7.3			
label the 1–5 carbon atoms in a nucleotide and explain how they are linked in a sugar phosphate backbone	1.7.4			
outline the structure of nucleosomes and their importance in eukaryotes	1.7.4			
	1.7.5			

outline the Hershey and Chase experiment and its importance.				

REFLECTION

Do you have any questions about water that need further investigation?

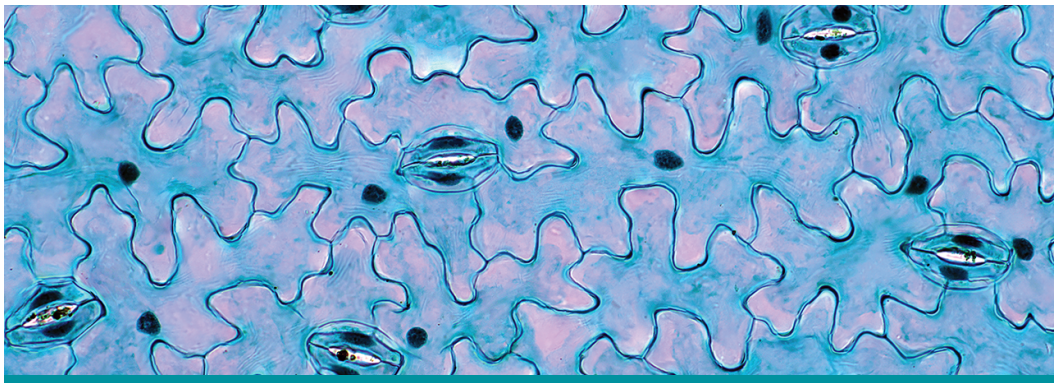
Thinking about the topics covered in the section, are there any areas that you have found particularly difficult?

What were the most challenging aspects of this topic? Why did you find them so?

If you were teaching this subject, what suggestions would you make to your students?

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.



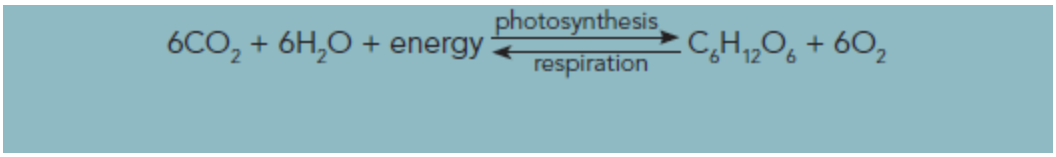
Chapter 2

Metabolism, respiration and photosynthesis

C1.1, C1.2, C1.3

INTRODUCTION

Metabolic reactions are chemical processes that occur in all cells to keep them alive. Respiration and photosynthesis are two key metabolic pathways in ecosystems. Light energy from the Sun is trapped as chemical energy in photosynthesis and then the energy is transferred through food chains and released back to the environment as heat energy from respiration. The two pathways can be simply written as:


6CO2 + 6H2O + energy
photosynthesis
respiration
C6H12O6 + 6O2

2.1 Enzymes and metabolism

LEARNING OBJECTIVES

In this section you will:

- learn that metabolic pathways are made up of chains and cycles of enzyme-catalysed reactions
- understand that metabolic processes can be anabolic or catabolic
- learn that enzymes are globular proteins that act as catalysts
- understand that enzymes have an area on their molecule, known as the active site, to which specific substrates bind
- learn that during enzyme-catalysed reactions molecules move about, and substrate molecules collide with the active sites on enzyme molecules
- discover how the rate of enzyme activity is influenced by temperature, pH and substrate concentration
- learn how enzyme molecules can be denatured

- understand how enzymes lower the activation energy of chemical reactions
- learn that enzyme inhibitors can be competitive or non-competitive

- understand how end-product inhibition can control metabolic pathways
- learn how co-enzymes and co-factors promote enzyme activity.

GUIDING QUESTIONS

- How are molecules transformed by metabolism?
- What is the role of enzymes in metabolic processes?
- What factors affect enzyme activity?

2.1.1 Metabolic pathways

Metabolic pathways consist of chains or cycles of reactions that are catalysed by enzymes. Metabolism includes all the chemical activities that keep organisms alive. Metabolic pathways may be very complex, but most consist of a series of steps, each controlled by an enzyme. Simple linear pathways involve the conversion of substrates to a final product:

$$\text{substrate X} \xrightarrow{\text{(enzyme 1)}} \text{substrate Y} \xrightarrow{\text{(enzyme 2)}} \text{substrate Z} \xrightarrow{\text{(enzyme 3)}} \text{end product}$$

KEY POINTS

anabolism refers to a series of metabolic pathways that build molecules from smaller subunits. These reactions require energy.

catabolism refers to a series of metabolic pathways that break down large molecules into smaller ones.

metabolism is the sum of the chemical reactions that occur within living organisms.

Each arrow represents the specific enzyme needed to catalyse the conversion of one substrate to the next. An example of a linear pathway is the breakdown of glucose in the glycolysis pathway (Section 3.2, Respiration) and the digestion of starch to maltose and glucose in the digestive system:

$$\text{Starch} \xrightarrow{\text{(enzyme 1 amylase)}} \text{maltose} \xrightarrow{\text{(enzyme 2 maltase)}} \text{glucose}$$

Other metabolic pathways, such as photosynthesis or respiration, involve both chains of reactions and cycles of reactions. Two examples of cyclical pathways are the Krebs cycle in aerobic respiration and the Calvin cycle in photosynthesis (Sections 3.2 and 3.3). Both of these cycles have many enzyme-catalysed steps.

Some metabolic reactions, such as protein synthesis, take place inside cells and are said to be intracellular, while others, such as digestion, occur outside cells and are known as extracellular reactions.

There are two types of metabolic process: anabolic reactions which build new molecules and catabolic reactions which break down large molecules.

NATURE OF SCIENCE

How does scientific understanding change and develop over time?

What affects longevity: metabolic rate or size? In 1926 an American biologist, Raymond Pearl, proposed the *rate of living hypothesis*. It suggested that a key factor determining how long a species lives is the speed of their resting metabolism. His evidence came from his observations that bigger animals tend to live longer and have lower heart, breathing and metabolic rates. He proposed that longevity is inversely related to basal (resting) metabolic rate.

The rate of living hypothesis was a well accepted theory for nearly 50 years. But over time, other scientists have made observations that have cast doubt on it. For example, rats and bats have similar metabolic rates, but a bat lives several times longer than a rat. Modern statistical methods that correct for the effects of body size and species group do not support the theory. They show that metabolic rate does not correlate with longevity in either mammal or bird groups.

Another newer model to explore body size, metabolism and ageing looked at how these are linked within the same species. Some of the results support the original rate of living theory and others do not. Scientists allowed some animals to expend more energy than others which were kept inactive and found that the amount of energy used does affect lifespan within a species. But the results were confused by the discovery that in some species smaller individuals with higher rates of metabolism live longer than slower, larger members of the same species.

A *free radical theory* of ageing proposed in the twenty-first century provided a new way of linking metabolism to ageing. Oxygen free radicals are formed during cell respiration in mitochondria. They can damage cells and contribute to ageing. The free radical theory suggested that more or faster respiration could make organisms' lives shorter. But today, scientists believe that free radical damage from metabolism on its own cannot be the cause of ageing. The accumulation of other defects

and imperfections must also be important. It seems that free radical theory has served its purpose in our understanding of the ageing process. More investigations and newer theories are needed to advance our knowledge of how we grow old. Only recently, experimental tools, such as sequencing of DNA and profiling of proteomes and metabolites, have been developed and these may be used to begin assessing the many types of damage that cause cells and bodies to age.

To consider:

1. Scientific knowledge is provisional and theories must be modified in the light of new observations and evidence.
2. Which areas of biology do you think will be important in understanding the ageing process?
3. How has modern technology helped scientists gather data for new theories?

Two examples of anabolic reactions are protein synthesis, when amino acid monomers are linked by condensation reactions (Section 1.6), and photosynthesis (Section 2.3), which builds glucose molecules from carbon dioxide and water.

Catabolic reactions include the hydrolysis of large molecules, such as carbohydrates into glucose monomers during digestion, and respiration, which breaks down glucose to release energy.

2.1.2 Enzymes and active sites

An enzyme is a biological **catalyst**. Like all catalysts, enzymes speed up biochemical reactions, such as digestion and respiration, but they remain unchanged at the end of the process. All enzymes are proteins with long polypeptide chains that are folded into three-dimensional shapes. The arrangement of these shapes is very precise and gives each enzyme the ability to catalyse one specific reaction. If the three-dimensional shape of an enzyme is destroyed or damaged, it can no longer function and is said to have undergone denaturation. Extremes of temperature, heavy metals and, in some cases, pH changes can cause permanent changes in an enzyme.

Some enzyme-controlled reactions, for example respiration and photosynthesis take place inside cells, others such as digestion take place in an extracellular space. Digestion in the gut is an example of extracellular enzyme activity.

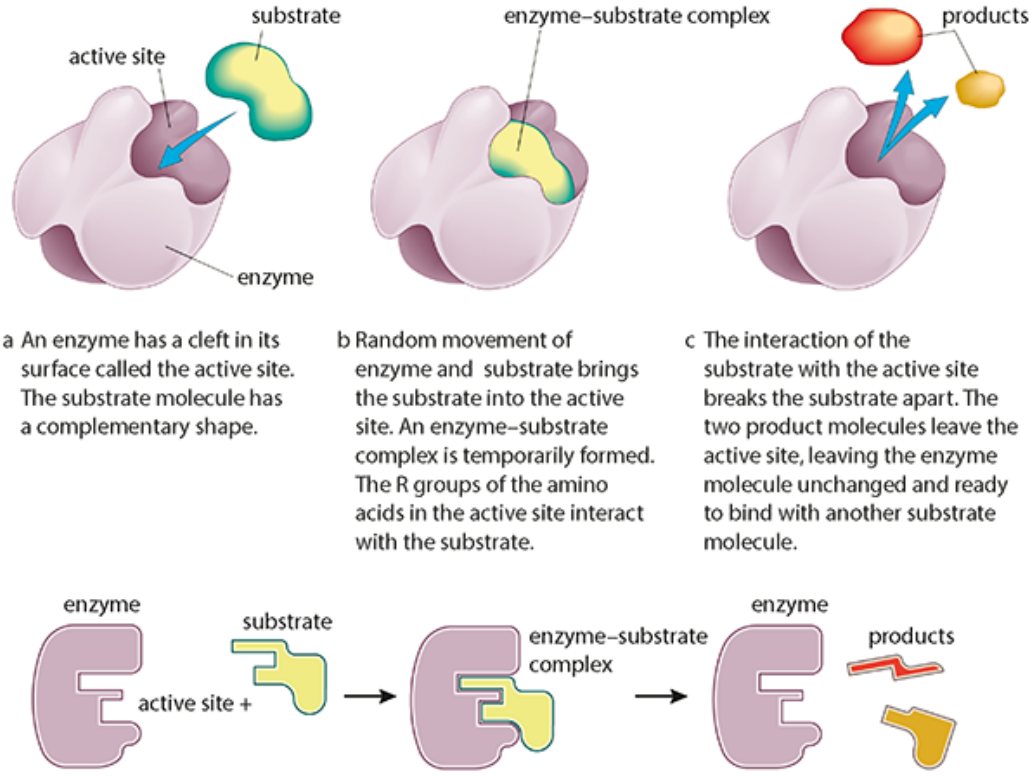


Figure 2.1.1: How an enzyme catalyses the breakdown of a substrate molecule into two product molecules.

KEY POINT

enzyme is a globular protein that functions as a biological catalyst of chemical reactions.

The three-dimensional shape of an enzyme is crucial to the way it works. In the structure of every enzyme is a specially shaped region known as an **active site** (Figure 2.1.1). It is here that the substrate and enzyme bind together. The substrates are the chemicals involved in the reaction catalysed by the enzyme. The shapes of the enzyme and substrate are complementary, so that they fit together perfectly like a key fits into a lock. The 'lock-and-key hypothesis' is a way of explaining how each enzyme can

be so specific. To unlock a door requires just one special key. To catalyse a reaction requires one special enzyme. Just as only one key fits perfectly into the lock, only one substrate fits perfectly into the active site of an enzyme.

Enzyme and substrate molecules move freely in solution and in most cases will eventually collide with one another. When a substrate molecule collides with the active site of an enzyme it will bind with it to form an enzyme–substrate complex. Once in place in an active site, substrates may be bonded together to form a new substance or they may be broken apart in processes such as digestion and respiration. For example, one type of enzyme bonds amino acids together to form a polypeptide, while very different enzymes are involved in digesting them.

Induced-fit model of enzyme action

In 1958, research published by Daniel Koshland (1920–2007) proposed a theory to explain how enzymes and substrates bind together. It is known as the **induced-fit model** of enzyme action. We know that substrates require a specific enzyme to catalyse their reactions and the model explains how only the correct substrate is able to bind to an enzyme.

KEY POINTS

active site is a region on the surface of an enzyme molecule where a substrate molecule binds and which catalyses a reaction involving the substrate.

induced-fit model is a model of enzyme action in which the shape of the active site alters when an enzyme binds to its substrate so that a reaction can take place.

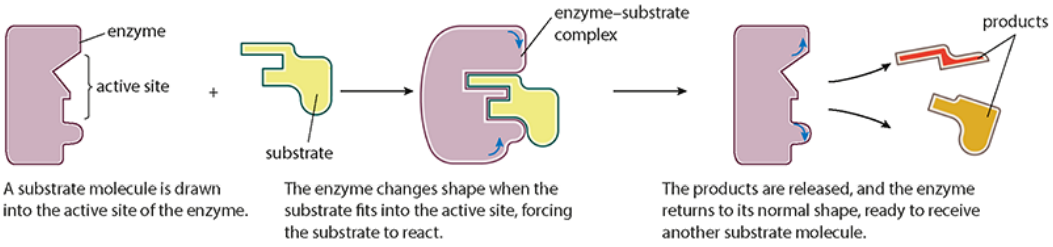


Figure 2.1.2: The induced-fit model of enzyme action.

As a specific substrate approaches an enzyme it induces the correct alignment of substrate and active site so that catalysis can take place. The specificity is a molecular recognition mechanism and it acts so that enzyme and substrate complement each other perfectly. Only the proper substrate is capable of inducing the proper alignment of the active site that will enable the enzyme to perform its catalytic function. The model also suggests that the active site continues to change until the substrate is completely bound to it.

Figure 2.1.2 shows how the substrate causes or induces a slight change in the shape of the active site so it can fit perfectly. As the enzyme changes shape, the substrate molecule is activated so that it can react and the resulting product or products are released. The enzyme is left to return to its normal shape, ready to receive another substrate molecule.

Factors affecting enzyme action

Enzymes work in many different places in living organisms and they require special conditions to work at their greatest, or optimum, efficiency. Temperature, pH and the concentration of the substrates involved all affect the rate at which enzymes operate and produce their products.

Temperature

Enzymes and their substrates usually meet as a result of random collisions between their molecules, which move freely in body fluids or cytoplasm. In the human body, most reactions proceed at their greatest rate at a temperature of about 37 °C and deviations from this optimum temperature affect the reaction rate, as the graph in Figure 2.1.3 shows.

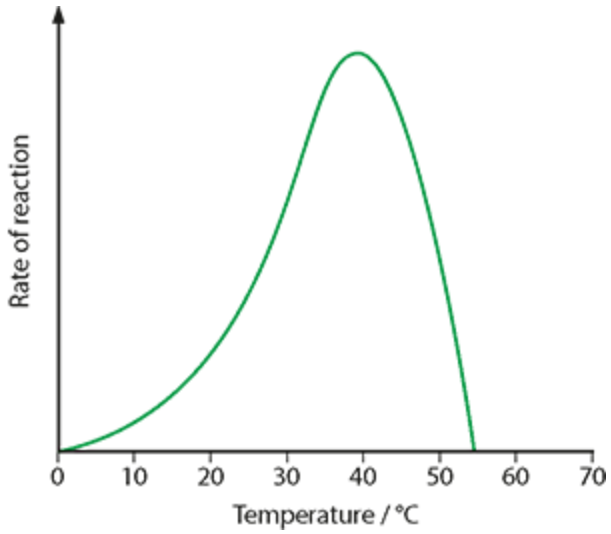


Figure 2.1.3: The effect of temperature on the rate of an enzyme-controlled reaction. An enzyme works most efficiently at its optimum temperature.

At less than 37 °C, molecules in solution move more slowly so the likelihood of collision between them is reduced. This slows down the production of products. At very low temperatures, enzymes hardly work at all and the rate of reaction is very low. As the temperature rises, molecular collisions are more frequent and energetic, and therefore the rate of the enzyme-controlled reaction increases.

As the temperature rises above the optimum, the enzyme and substrate molecules move faster, but atoms in the enzyme molecule itself also move more energetically, straining the bonds

holding it together. Eventually, these bonds may be stressed or broken to such an extent that the enzyme loses its three-dimensional shape and the active site can no longer receive substrate molecules. At these high temperatures, the structure is permanently destroyed: the enzyme is denatured and can no longer catalyse the reaction.

pH

pH is a measure of the relative numbers of H^+ and OH^- ions in a solution. A solution with a low pH value has many free H^+ ions and is acidic, whereas a high pH value indicates more OH^- ions and a basic solution. Pure water is neutral and has a pH value of 7, indicating that the number of OH^- and H^+ ions is equal.

Enzyme action is influenced by pH because the amino acids that make up an enzyme molecule contain many positive and negative regions, some of which are around the active site. An excess of H^+ ions in an acidic solution can lead to bonding between the H^+ ions and negative charges in the active site or other parts of the enzyme. These interactions can inhibit the matching process between the enzyme and its substrate, and slow down or even prevent enzyme activity. A similar effect occurs if a solution becomes too basic: the excess of negative ions upsets the enzyme in the same way. At extremes of pH, the enzyme may even lose its shape and be denatured. The changes are usually, though not always, permanent.

Not all enzymes have the same optimum pH. Proteases (protein-digesting enzymes) in the stomach have an optimum pH of 2 and work well in the acidic conditions there, but proteases in the small intestine have an optimum of pH 8. Most enzymes that work in the cytoplasm of body cells have an optimum pH of

about 7. The graph in Figure 2.1.4 shows how reaction rate varies with pH for this type of enzyme.

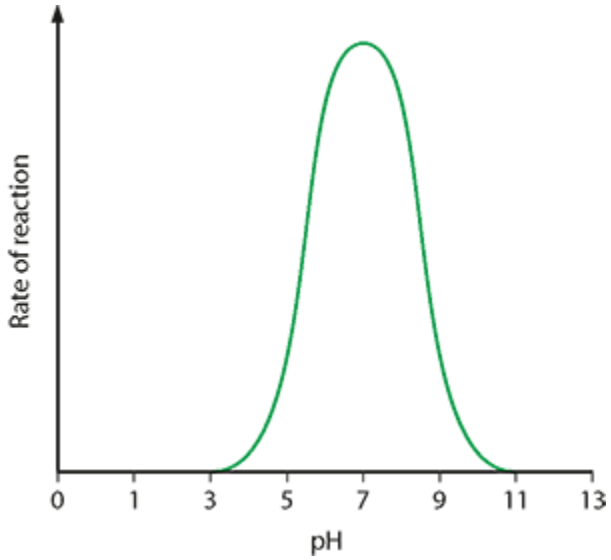


Figure 2.1.4: The effect of pH on the rate of an enzyme-controlled reaction. Changing pH affects the charges on the amino acid molecules in the enzyme. The shape of the enzyme and its active site changes, reducing the rate of reaction

Concentration of substrate

If there is a set concentration of enzyme present in a reaction mixture, and the concentration of substrate increases, the rate of production of the products will increase because of the greater chance of collisions between substrate and enzyme molecules. More collisions mean that the enzyme is able to process or 'turn over' more substrate molecules. But there is a limit to this increase in reaction rate. If the concentration of substrate increases too much, it will exceed the maximum rate at which the enzyme can work. When this happens, at any one moment all the active sites are occupied by substrate or product molecules, and so adding further substrate has no effect. The rate reaches its limit; you can see this as the plateau in the graph in Figure 2.1.5.

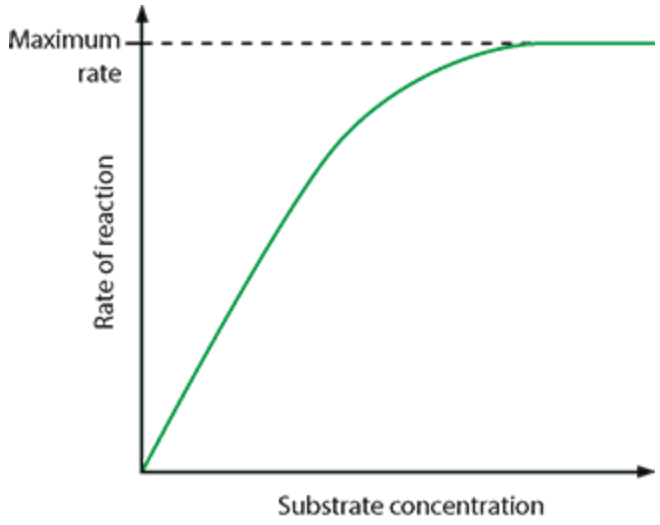


Figure 2.1.5: The effect of substrate concentration on the rate of an enzyme-catalysed reaction.

TEST YOUR UNDERSTANDING

1 Define the term metabolism.

2 Explain the difference between an anabolic and a catabolic reaction.

3 Complete this sentence:

Enzymes are …… proteins that act as …… in metabolic reactions.

4 Describe why temperature tends to speed up the rate of enzyme activity using the terms ‘molecular motion’ and ‘collision’.

5 Why does increasing the substrate concentration in a enzyme-controlled reaction produce a graph that levels off after a certain concentration is reached?

2.1.3 Activation energy

Enzymes work by lowering the **activation energy** of the substrate or substrates. For a metabolic reaction to occur, the substrate has to reach an unstable, high-energy 'transition state' where the chemical bonds are destabilised, and this requires an input of energy, which is called the activation energy. When the substrate reaches this transition stage, it can then immediately form the product. Enzymes can make reactions occur more quickly because they reduce the activation energy of reactions they catalyse to bring about a chemical change (Figure 2.1.6).

KEY POINT

enzymes do not change the quantity of product that is formed, only the rate at which the product is formed.

Metabolic reactions that occur in living organisms have to occur at the body temperature of the organism, which is never high enough to bring substrates to their transition state. The active site of an enzyme is very important because it can lower the amount of energy needed to reach a transition state, so the reaction can occur at the temperature of the organism.

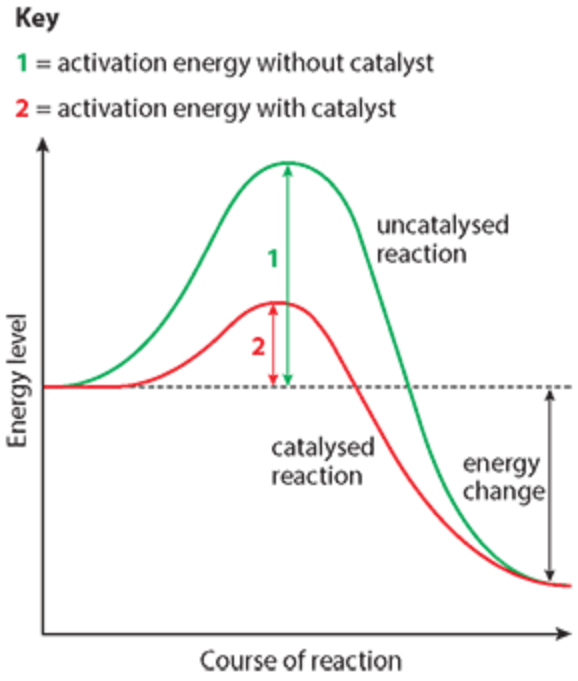


Figure 2.1.6: Graph to show activation energy for an exothermic reaction with and without a catalyst.

2.1.4 Competitive and non-competitive inhibition

Enzyme inhibitors are substances that reduce or prevent an enzyme's activity. Some inhibitors are competitive and others non-competitive.

KEY POINT

enzymes can be affected by the presence of other molecules that temporarily bind to them, either at the active site or at an allosteric site, a region on the surface of an enzyme to which an allosteric, effector molecule binds.

Inhibition by a molecule whose structure is similar to that of the substrate molecule that normally binds to the active site is an example of **competitive inhibition**. Competitive inhibitors compete with the substrate to occupy the active site of the enzyme, and prevent the substrate molecules from binding (Figure 2.1.7, left-hand side). These inhibitors are not affected by the enzyme and do not form products. The rate of reaction of the enzyme is lower because substrate molecules cannot enter the active site of the enzyme molecules that are blocked by an inhibitor. At low concentrations of substrate, competitive inhibitors have a more significant effect than at higher concentrations, when the substrate can outcompete the inhibitor (Figure 2.1.8). A competitive inhibitor occupies the active site temporarily, so the inhibition is reversible.

Permanent binding of an inhibitor to the active site or to another part of an enzyme is known as **non-competitive inhibition**. Inhibitors may bind at part of the enzyme molecule where they

partly block access of the substrate to the active site, or they may cause a change in the shape of the enzyme so that the substrate cannot enter the active site (Figure 2.1.7, right-hand side). Increasing the concentration of substrate in the presence of a non-competitive inhibitor does not overcome inhibition (Figure 2.1.8).

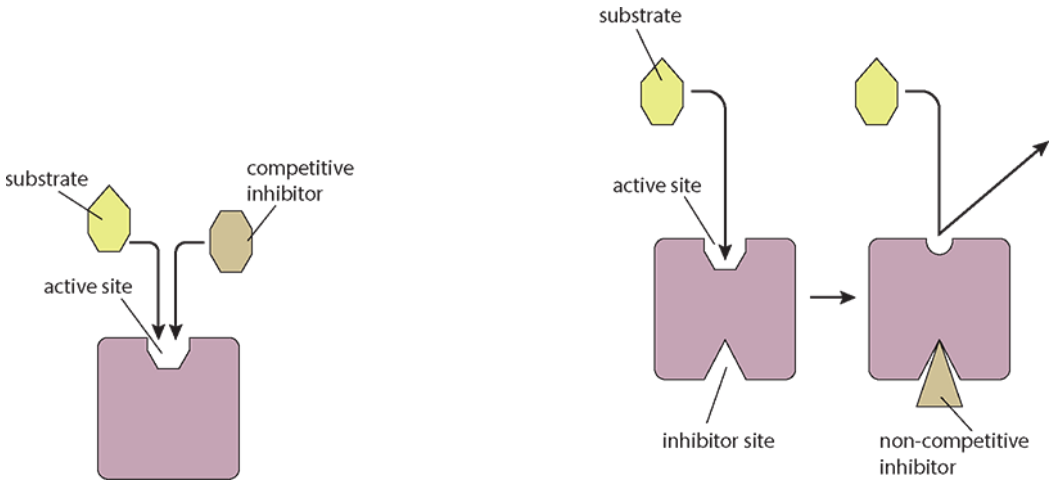


Figure 2.1.7: (Left) Competitive inhibition and (right) non-competitive inhibition.

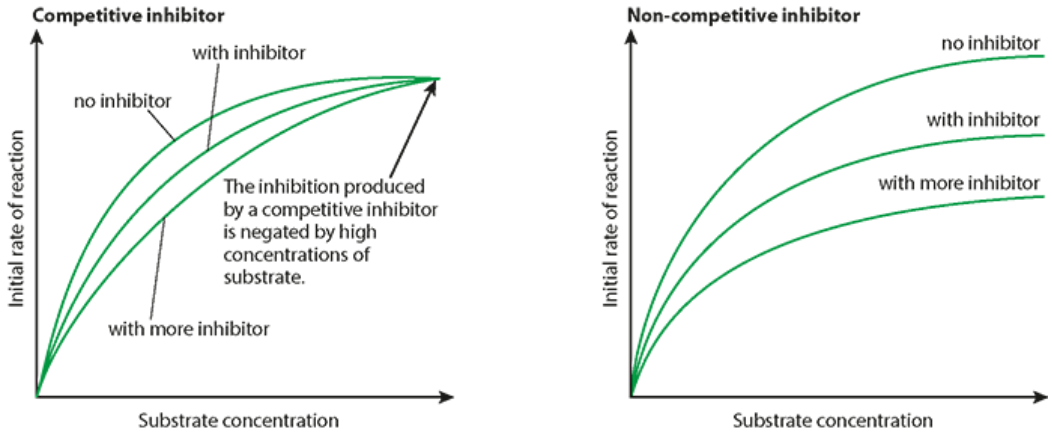


Figure 2.1.8: Graphs to show the effects of competitive and non-competitive inhibitors on reaction rate, as substrate concentration increases.

Table 2.1.1 compares the nature and effects of competitive and non-competitive inhibitors.

Using non-competitive inhibition by penicillin to inhibit bacteria

Penicillin is an antibacterial medicine that works by inhibiting the formation of bacterial cell walls. Penicillin is most effective against Gram-positive bacteria such as staphylococci and streptococci (Chapter 10). These bacteria have thick cell walls built of many linked peptidoglycan molecules. Penicillin is an irreversible inhibitor that covalently binds to the bacterial enzyme transpeptidase. Transpeptidase catalyses the formation of cross-links between the long polymers of peptidoglycan in the bacterial cell wall. The walls are left weakened and, when the bacteria divide, their new walls are not properly formed so that the cells burst and die.

Using competitive inhibition to treat poisoning

Another example of an enzyme inhibitor that is used in medicine is fomepizole. Fomepizole is a competitive inhibitor of the enzyme alcohol dehydrogenase, which usually catalyses the **oxidation** of ethanol (alcohol) to acetaldehyde. Acetaldehyde is converted to harmless products in the liver and so it does not harm the body. But alcohol dehydrogenase also catalyses steps in the metabolism of ethylene glycol (antifreeze) to toxic metabolites that cause severe damage to the kidneys. If ethylene glycol is accidentally ingested, an injection of fomepizole blocks alcohol dehydrogenase so that toxic metabolites are not produced and the kidneys are not harmed.

Competitive inhibitors	Non-competitive inhibitors

structurally similar to the substrate molecule	structurally unlike the substrate molecule
occupy and block the active site	bind at a site away from the active site, reducing access to it
if concentration of inhibitor is low, increasing the concentration of substrate will reduce the inhibition	if concentration of substrate is low, increasing the concentration of substrate has no effect on binding of the inhibitor so inhibition stays high
examples include: • oxygen, which competes with carbon dioxide for the active site of ribulose bisphosphate carboxylase in photosynthesis • disulfiram, which competes with acetaldehyde for the active site of aldehyde dehydrogenase • ethanol, which can be used in preventing antifreeze poisoning because it is a competitive inhibitor of the enzyme alcohol dehydrogenase (refer to section 'Using competitive inhibition to treat poisoning')	examples include: • cyanide and carbon monoxide, which block cytochrome oxidase in aerobic respiration, leading to death • penicillin, which blocks the active site of an enzyme that synthesises the cell walls of some bacteria

Table 2.1.1: Comparing competitive and non-competitive inhibitors.

2.1.5 Controlling metabolic pathways

End-product inhibition

End-product inhibition occurs when an enzyme in a metabolic pathway is inhibited by the product of that pathway. This prevents a cell over-producing a substance it does not need at the time. Many products may be needed by a cell at specific times or in specific amounts and over-production not only wastes energy but may also become toxic if the product accumulates.

In an assembly-line reaction, such as those described in Figure 2.1.9, each step is controlled by a different enzyme. If the end product begins to accumulate because it is not being used, it inhibits an enzyme earlier in the pathway to switch off the assembly line. In most cases, the inhibiting effect is on the first enzyme in a process, but in other cases it can act at a branch point to divert the reaction along another pathway.

When the end product starts to be used up, its inhibiting effect reduces, the inhibited enzyme is reactivated and production begins again. This is an example of **negative feedback**.

KEY POINTS

end-product inhibition is the control of a metabolic pathway by a product in or at the end of that pathway; the product inhibits an enzyme found earlier in the pathway.

negative feedback is a regulating mechanism in which a change in a sensed variable results in a correction that opposes the change.

EXTENSION

Negative feedback is also important in the control of several physiological processes including regulation of blood sugar levels and reproductive cycles.

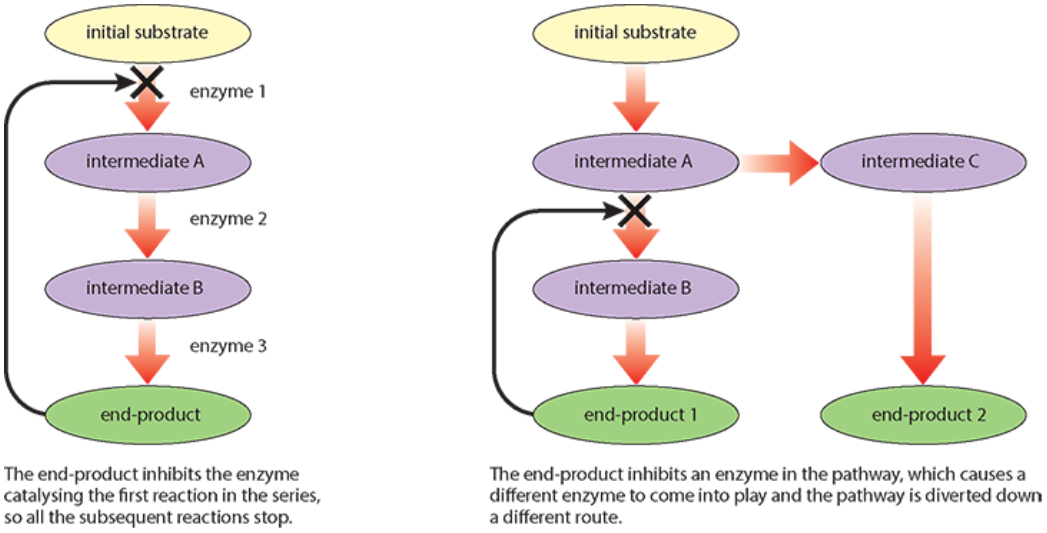


Figure 2.1.9: End-product inhibition.

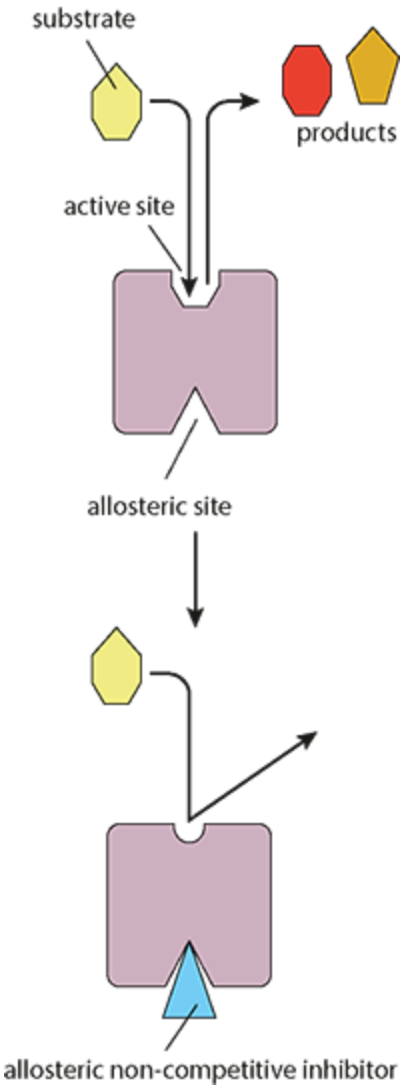


Figure 2.1.10: Allosteric control. Allosteric inhibitors prevent the active site functioning.

End-product inhibition may be competitive or non-competitive. Competitive inhibition will only work if the product is a similar shape to the normal substrate and there can be an induced fit of the product or inhibitor onto the enzyme. In most cases, the product will be a different shape and therefore there has to be non-competitive inhibition. In this case, the enzyme is known as an **allosteric enzyme**, the product is called an **allosteric**

inhibitor and the place where it binds to the enzyme is called the **allosteric site** (Figure 2.1.10).

An example of end-product inhibition

Threonine is converted to isoleucine in series of five enzyme-controlled stages. Isoleucine is important in the immune system and also in the synthesis of proteins including hemoglobin. Isoleucine, as the end product of threonine metabolism, can inhibit threonine deaminase, the first of the five enzymes in the process (Figure 2.1.11). Isoleucine inhibits the enzyme by binding on the molecule at a site away from the active site. When it is attached, the active site of the enzyme is changed so that no further substrate can bind to it. As isoleucine concentration increases, more and more isoleucine molecules attach to this inhibition site on enzyme molecules and therefore inhibit further production of isoleucine. As their concentration falls, isoleucine molecules detach from the threonine deaminase enzyme molecules and are used in the cell. Once the inhibitor has been removed, the active site can bind new substrate and the pathway is reactivated.

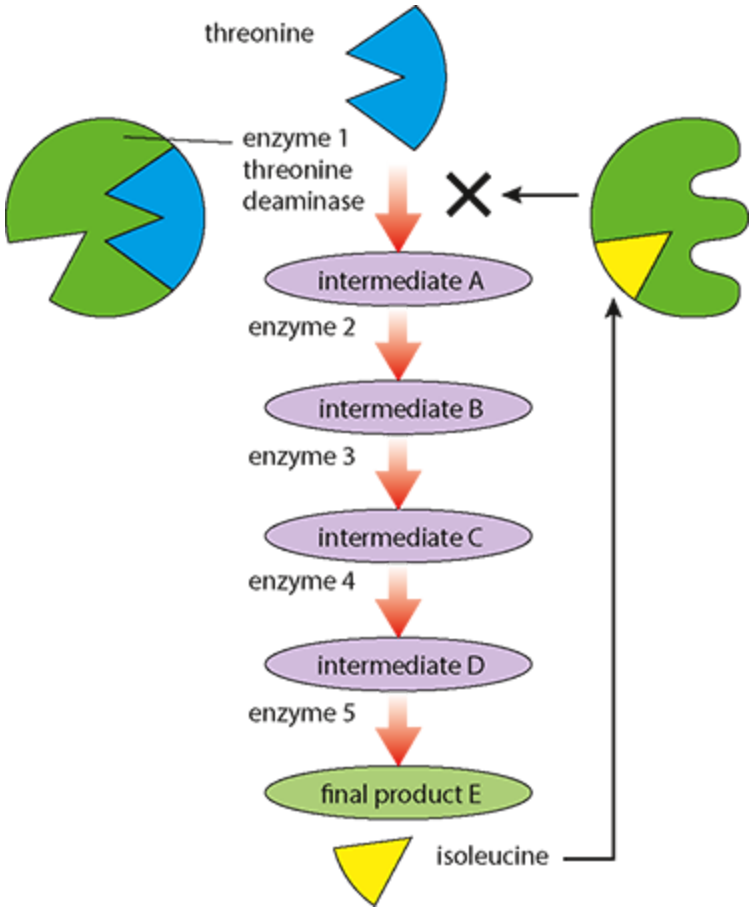


Figure 2.1.11: The pathway that converts threonine to isoleucine – a specific example of end-product inhibition.

This mechanism makes the metabolic pathway self-regulating so that there is always sufficient isoleucine present in the cell.

2.1.6 Co-enzymes and co-factors

A **co-enzyme** is an organic, non-protein molecule that binds to an enzyme to allow it to catalyse a reaction. Co-enzymes cannot work alone, but they can be reused several times with an enzyme. Many co-enzymes are vitamins or are derived from vitamins. For example, pantothenic acid (also called vitamin B5) is needed for the synthesis of co-enzyme A, which is essential for fatty acid metabolism and in the link reaction of aerobic respiration (Section 2.2).

Co-factors are inorganic substances that promote enzyme activity. Many metal ions are co-factors.

KEY POINT

Co-enzymes and co-factors both promote enzyme activity. Co-enzymes are organic molecules; co-factors are inorganic.

For example, cupric ions are co-factors needed to promote cytochrome oxidase activity and zinc is a co-factor in nearly 300 enzymes involved in metabolism. Chloride ions are allosteric activators for human amylases. As one chloride ion binds to amylase, it induces activation of the enzyme and increases the rate of reaction of the hydrolysis of starch.

TEST YOUR UNDERSTANDING

6 Outline what is meant by activation energy.

7 Explain how an enzyme pathway can be switched off by an accumulation of the end product of the pathway.

8 Outline the way in which penicillin leads to the death of bacteria.

9 List three differences between competitive and non-competitive enzyme inhibitors.

THEORY OF KNOWLEDGE

Studying metabolic pathways

Metabolic pathways have been studied for centuries but one of the most significant advances was made by Eduard Buchner (1860–1917), who discovered enzymes at the start of the 20th century. At first, studies of whole animals were made, but more recently it has been possible to analyse metabolic pathways and their component reactions using modern techniques such as chromatography, X-ray diffraction, spectroscopy and radioactive isotopes. In the mid-20th century, the Krebs cycle (often called the citric acid cycle; see Section 2.2) and the glyoxylate cycle were discovered by Hans Krebs (1900–1981) and Hans Kornberg (1928–2019). But metabolic pathways are very elaborate. Many pathways are interrelated and together make up a complex metabolic network in a cell. These pathways are vital to homeostasis and cell function. Some pathways are connected by intermediate products, and products of one pathway may be substrates for another.

To consider:

Most biochemical studies are made using carefully controlled experiments that look at one part of a pathway. To what extent can looking at component parts of a complex system give us knowledge of the whole?

Links

- How do plants and algae convert light energy into organic compounds, or chemical energy, to use in their metabolism? (Section 2.3)
- How does metabolism help to maintain constant internal conditions? (Chapter 8)
- Why is compartmentalisation in cells important for the control of metabolism? (Chapter 5)

2.2 Respiration

LEARNING OBJECTIVES

In this section you will:

- understand that cell respiration is the controlled release of energy from organic substances to produce ATP (adenosine triphosphate)
- learn that cell respiration is a series of complex enzyme-catalysed reactions that involve hydrolytic breakdown of glucose and other molecules
- understand that much of the energy is released to the environment as heat
- discover that ATP from respiration is an immediately available source of energy in the cell
- understand that anaerobic respiration (called fermentation in microorganisms) gives a small yield of ATP from glucose and occurs in the cell cytoplasm
- recall that yeast cells respire anaerobically and are used in baking and brewing
- recognise that humans can respire anaerobically for a short time in the absence of oxygen, which produces lactate
- understand that lactate produced by some bacteria is useful in yoghurt production

- learn that aerobic respiration requires oxygen and gives a much larger yield of ATP from glucose
- learn that aerobic respiration starts in the cytoplasm and requires mitochondria

- learn that during cell respiration electron carriers are oxidised and reduced
- understand that phosphorylation of molecules increases their energy level making them less stable, and decarboxylation generates carbon dioxide
- understand that glycolysis does not use oxygen. It takes place in the cytoplasm and each glucose molecule in converted into two pyruvate molecules with a small net gain of ATP
- learn that during anaerobic respiration reduced NAD from glycolysis is oxidised and reduces pyruvate to avoid it building up in the cytoplasm
- understand how in aerobic respiration pyruvate is decarboxylated and oxidised. In the link reaction it is converted to an acetyl compound and then attached to coenzyme A to form acetyl co-enzyme A
- understand that in the Krebs carboxylic acid cycle the oxidation of acetyl groups is coupled with the reduction of hydrogen carriers and carbon dioxide is released
- learn that energy released during oxidation reactions is transferred by reduced NAD and FAD to the cristae of mitochondria

- learn that the transfer of electrons between carrier molecules in the electron transport chain (ETC) in the membrane of the cristae is coupled to proton pumps
- understand how during chemiosmosis, protons diffuse through ATP synthase to generate ATP
- discover how oxygen binds with free protons to form water, thus maintaining the H^+ (proton) gradient.

GUIDING QUESTIONS

- What is the role of adenosine triphosphate (ATP) in the transfer of energy in cells?
- How do living organisms release the energy on which their cells depend?
- Why does aerobic respiration generate a much larger yield of ATP per molecule of glucose than anaerobic respiration?

2.2.1 Cell respiration and ATP

All living cells need energy to stay alive. The energy is used to power all the activities of life including digestion, protein synthesis and active transport. A cell's energy sources are the sugars and other substances derived from nutrients, which can be metabolised in a series of chemical reactions to release the energy that holds their molecules together. Much of the energy is eventually transferred to the environment as heat.

Cell respiration is the gradual breakdown of nutrient molecules such as glucose and fatty acids in a series of enzyme-controlled metabolic pathways that ultimately release energy in the form of **ATP (or adenosine triphosphate)**.

KEY POINT

ATP (adenosine triphosphate) is the immediately available energy currency of a cell. It is needed for every activity that requires energy. Cells make their own ATP in mitochondria. When energy is used, ATP is broken down to ADP (adenosine diphosphate) and inorganic phosphate. This conversion releases energy for use and a cyclic process reforms the ATP during respiration.

$ADP + Pi \leftrightarrow ATP + H_2O$

ATP (adenosine triphosphate)

ATP is the molecule that acts as a source of energy for all living organisms. It is often called the universal energy carrier. It is an organic molecule that transports chemical energy for use in metabolic processes such as movement, active transport and the

synthesis of new molecules. Metabolic processes are chemical processes that occur in all cells to keep them alive.

ATP consists of three parts: a pentose sugar, a purine base (adenine) and three phosphate groups attached to the pentose sugar (Figure 2.2.1). Energy is released from ATP during a hydrolysis reaction in which the last phosphate group is removed. This leaves ADP (adenosine diphosphate) and an inorganic phosphate (shown as Pi).

ATP → ADP + Pi

ADP is converted back to ATP in the reactions of cell respiration.

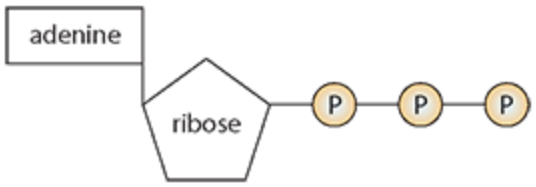


Figure 2.2.1: The structure of a molecule of ATP.

Glucose is the most commonly used source of energy. Enzymes hydrolyse each glucose molecule in a number of stages, which release energy in small amounts as each covalent bond is broken. If there is insufficient glucose available, fatty acids or amino acids can be used instead.

The energy in glucose or other nutrient molecules can be released in a single reaction. This is what happens when glucose burns in the reaction known as combustion. In this case, the energy in the glucose is released as heat. In the series of reactions that occur during respiration, glucose is broken down gradually, with each step catalysed by a different enzyme. This releases energy in small amounts so that it can be used by cells. Nevertheless as energy is used much of it is lost to the surroundings as heat. We

can experience this for ourselves when we use energy during exercise and our bodies become hot.

Glycolysis

The first stage in cell respiration is **glycolysis**. Glucose that is present in the cytoplasm of a cell is broken down by a series of enzymes, to produce two molecules of a simpler compound called pyruvate. As this occurs, there is a net production of two molecules of ATP (Figure 2.2.2).

$$\underset{\text{six-carbon sugar}}{\text{glucose}} \rightarrow \underset{2 \times \text{three-carbon sugar}}{2 \text{ pyruvate}} + 2 \text{ ATP}$$

Glycolysis actually uses two molecules of ATP to get the process under way but produces four molecules of ATP in total, per molecule of glucose. Thus we say there is a net production of two ATPs.

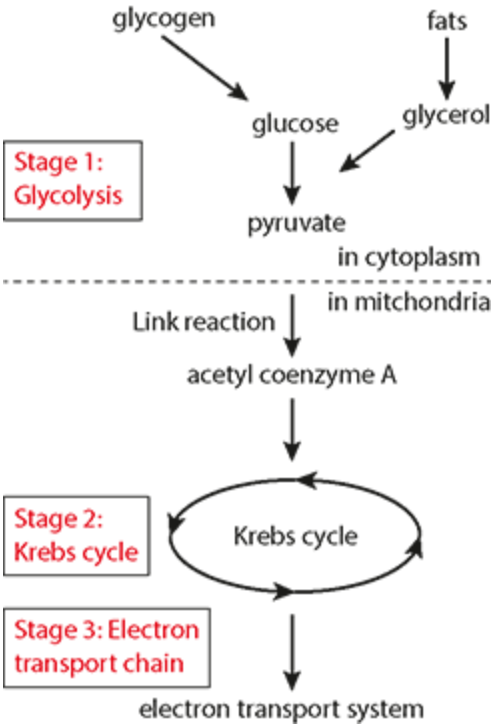


Figure 2.2.2: A summary of the stages in aerobic respiration.

2.2.2 Aerobic and anaerobic respiration

The next stage of cell respiration depends on whether or not oxygen is available. In the presence of oxygen, aerobic respiration can take place; without it, respiration must be anaerobic.

Aerobic respiration is the most efficient way of producing ATP. Aerobic respiration is carried out by cells that have mitochondria and it produces a great deal of ATP. A labelled micrograph of a mitochondrion is shown in Figure 2.2.2. Pyruvate molecules produced by glycolysis enter the mitochondria and are broken down, or oxidised, in a series of reactions that release carbon dioxide and water and produce ATP.

In the first step, two pyruvate molecules are transported into the mitochondria in the **link reaction**, as shown in Figure 2.2.1. Each pyruvate loses a carbon atom which forms carbon dioxide and a hydrogen atom, so that they become two molecules of acetyl CoA. Acetyl CoA then enters a stage called the **Krebs carboxylic acid cycle** (also known as the citric acid cycle) and is modified still further, releasing more carbon dioxide. The Krebs cycle takes place in the matrix of the mitochondria. Finally, on the inner membranes of the mitochondria, products of the cycle react directly with oxygen and the result is the release of large amounts of ATP. The original glucose molecule is completely broken down to carbon dioxide and water so the equation for aerobic respiration is often summarised as:

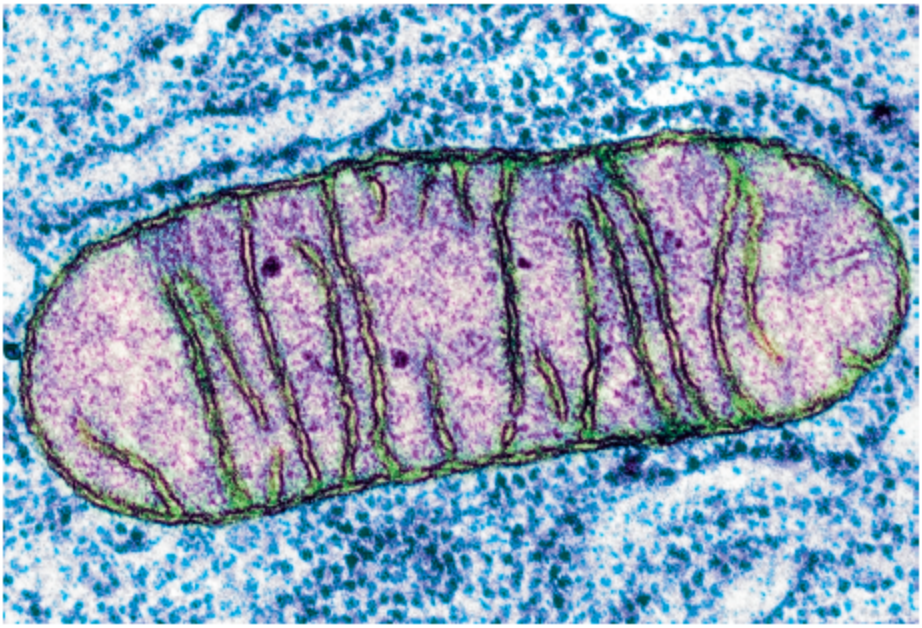


Figure 2.2.3: Coloured electron micrograph of a mitochondrion (× 72 000).

$$\text{glucose} + \text{oxygen} \rightarrow \text{carbon dioxide} + \text{water} + 38\ \text{ATP}$$

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + 38\ \text{ATP}$$

EXAM TIP

You should be familiar with the general structure of a mitochondrion (Figure 2.2.3), you don't need to remember all the details but thinking about where the reactions of respiration take place can help you understand the different stages. You will learn more about mitochondria in Section 5.2.

Anaerobic respiration occurs in the cytoplasm of cells. In animal cells, the pyruvate produced from glucose by glycolysis is converted to lactate (Figure 2.2.4), which is a waste product and is taken out of the cells. In humans, anaerobic respiration occurs if a person is doing vigorous exercise and their cardiovascular

system is unable to supply sufficient oxygen for aerobic respiration to provide ATP at the necessary rate. Although anaerobic respiration releases far less energy per molecule of glucose than aerobic respiration, the extra ATP enables the person to continue exercising for a short period, at a time of great exertion, to maximise power output. One consequence of the build-up of lactate in the muscles that occurs during anaerobic respiration is the sensation of cramp, so this type of respiration cannot be sustained for very long. The word equation for anaerobic respiration in animals is:

Glucose → pyruvate → lactate

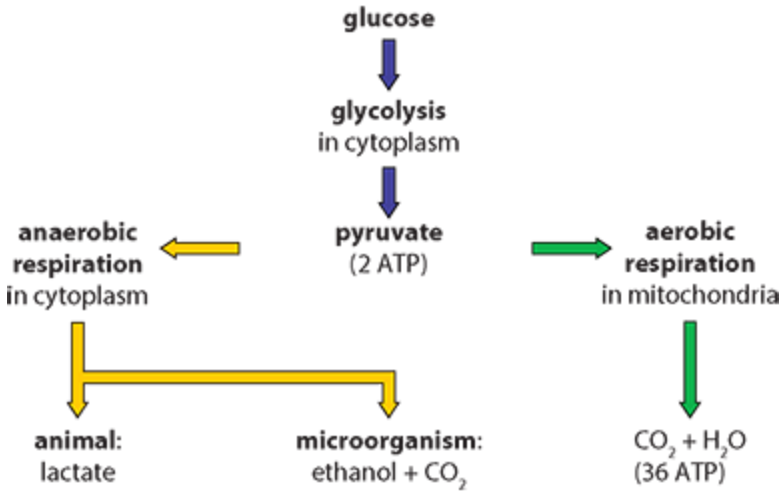


Figure 2.2.4: Simple diagram to show the products of aerobic and anaerobic respiration.

EXTENSION

You'll often see both the terms lactate and lactic acid used when you read about anaerobic respiration, so what is the difference? Like other acids, lactic acid is a substance that is able to donate hydrogen (H^+) ions. When it loses an electron it is called a base and known as lactate:

Lactic acid → lactate + H^+

When lactic acid is produced in muscles it will dissociate into lactate and hydrogen ions so the terms are used interchangeably.

Humans can only respire anaerobically for a short period of time. A sprinter running a 100 metre race can run an entire race anaerobically, but a long-distance runner will use only aerobic respiration for maximum efficiency.

Lactate is carried in the blood from muscles to the liver where it is converted back to pyruvate. Pyruvate may either be converted back to glucose, a process which requires energy, or used as a fuel, producing carbon dioxide and water.

In microorganisms, such as yeast, anaerobic respiration is also known as **fermentation**, and produces a different outcome. The pyruvate molecules from glycolysis are converted to ethanol (alcohol) and carbon dioxide (Figure 2.2.4).

Glucose → pyruvate → ethanol + carbon dioxide

No further ATP is produced by the anaerobic respiration of pyruvate, so this type of respiration gives only a small yield of ATP from glucose. Aerobic and anaerobic respiration are compared in Table 2.2.1

	Aerobic	**Anaerobic**
Requires oxygen	yes	no
Glycolysis occurs	yes	yes
oxidation	complete	incomplete
Glucose completely	yes	no

broken down		
Waste products	CO_2 and water	Lactic acid (animals) or CO_2 and alcohol (yeast)
ATP yield	38	2

Table 2.2.1: Comparing aerobic and anaerobic respiration

2.2.3 Anaerobic respiration in food production

People have benefitted from the anaerobic respiration of yeast in baking and brewing for thousands of years. Today, many different types of yeast are used in the production of bread, wine and beer. The strains of yeast used for baking and brewing are different and each has been selected for its specific characteristics. Baking yeasts feed on sugar and flour in bread dough and grow more quickly than brewing yeasts, which are slow-growing but able to tolerate higher alcohol concentrations. In bread making, the yeast initially respires aerobically, releasing carbon dioxide gas and water into the dough in a very short period of time. Carbon dioxide in the dough causes it to rise as the gas becomes trapped in pockets between gluten fibres in the flour. When oxygen in the dough has been depleted, the yeast continues to respire anaerobically, producing ethanol which evaporates during baking. The yeast cells are also killed by the high temperature of the oven.

Yoghurt production also relies on anaerobic respiration. Two bacteria that are often used to make yoghurt are *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. Milk, either raw or pasteurised, is first heated to denature the whey proteins so that the final yoghurt that is made will form a stable gel and not separate if it is stored. Next, it is cooled to about 40 °C. This is the optimum temperature (Section 2.1) for the bacterial enzymes involved in fermentation. Live bacterial cells are put into the milk and they feed on lactose sugars, converting the sugars into lactic acid as they respire. After about 12 hours, lactic acid causes the milk to thicken and produces the characteristic acidic taste and texture of yoghurt.

INTERNATIONAL MINDEDNESS

Cow's milk is used to produce yoghurt and kefir in many parts of the world, but milk from yaks, buffalo, goats, sheep and camels is also used where it is available. In Central Asia mare's or donkey's milk is used to produce koumiss, a fermented drink similar to thin yoghurt. The flavour and texture of each product depends on the types of milk and bacteria that are used.

NATURE OF SCIENCE

Assessing ethics in science: using invertebrates in a respirometer

A simple respirometer, such as the one shown in Figure 2.2.5, can be used to monitor respiration in small organisms such as woodlice or in germinating seeds. The apparatus can demonstrate that oxygen is used and carbon dioxide is produced during respiration. Test organisms are placed in two large boiling tubes as shown, so that one contains living organisms (tube A) and the other, which acts as a control, contains either dead organisms or is left empty (tube B). Soda lime or another alkali such as potassium hydroxide absorbs carbon dioxide. As oxygen is used by the living things in tube A, the level of liquid rises in the arm of the manometer attached to tube A. If required, measurements of time can be made so that the rate of respiration can be estimated. The temperature in the apparatus is kept constant by immersing the tubes in a water bath. This minimises any change in volume due to temperature change.

To consider:

1 How can we ensure that the invertebrates used in experiments like this are treated ethically?

2 What measures would you use to minimise the distress and disturbance to the organisms and also to the habitat from which they are taken?

3 How can we know whether the organisms are experiencing distress?

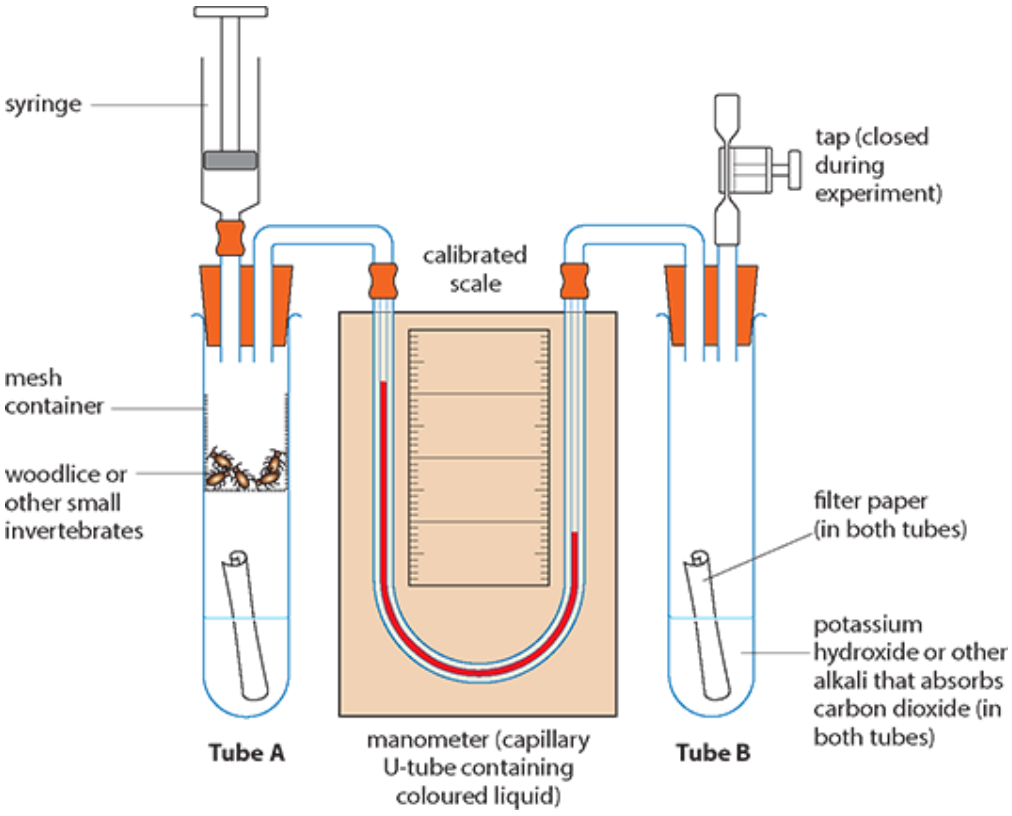


Figure 2.2.5: A simple respirometer.

TEST YOUR UNDERSTANDING

10 Which stage of respiration takes place in both aerobic and anaerobic respiration?

11 Where does aerobic respiration take place in eukaryotic cells?

12 Outline the role of anaerobic respiration in baking.

2.2.4 Biochemistry of cell respiration

Oxidation and reduction

KEY POINT

redox reaction a reaction in which reduction and oxidation occur simultaneously.

Cell respiration involves several oxidation and reduction reactions. Such reactions are common in biochemical pathways. When two molecules react, one of them starts in the oxidised state and becomes reduced, and the other starts in the reduced state and becomes oxidised, as shown in Figure 2.2.6.

Figure 2.2.6: Oxidation and reduction are linked processes – as one molecule is reduced another is oxidised in a redox reaction.

There are three different ways in which a molecule can be oxidised or reduced, as outlined in Table 2.2.2. In biological oxidation reactions, addition of oxygen atoms is an alternative to removal of hydrogen atoms. Since a hydrogen atom consists of an electron and a proton, losing hydrogen atoms (oxidation) involves losing one or more electrons.

Oxidation and reduction occur together in biochemical reactions. As one compound loses electrons, another one gains electrons.

In the simple equation for respiration, glucose is oxidised as hydrogen atoms, and therefore electrons, are gradually removed from it and added to hydrogen acceptors (the oxygen atoms on the left side of the equation), which become reduced.

$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$ + energy

Oxidation	Reduction
loss of electrons	gain of electrons
loss of hydrogen	gain of hydrogen
gain of oxygen	loss of oxygen

Table 2.2.2: Changes involved in oxidation and reduction.

Chemical reactions like this are referred to as redox reactions. In redox reactions, the reduced molecule always has more potential energy than the oxidised form of the molecule. Electrons passing from one molecule to another carry energy with them.

The electron carriers used during cell respiration are **NAD^+** and FAD^+.

KEY POINT

NAD^+ is a hydrogen carrier that accepts hydrogen atoms removed during the reactions of respiration. During glycolysis, two hydrogen atoms are removed and NAD^+ accepts the protons from one of them and the electrons from both of them.

$NAD^+ + 2H \rightarrow NADH + H^+$

Cell respiration is the controlled breakdown of food molecules such as glucose or fat to release energy, which can be stored for later use. The energy is most commonly stored in the molecule adenosine triphosphate, or ATP. The respiration pathway can be divided into four parts:

1 glycolysis

2 link reaction

3 Krebs cycle

4 electron transfer chain and chemiosmosis.

Glycolysis, the link reaction and the Krebs cycle are summarised in Figure 2.2.7, and the electron transfer chain and chemiosmosis are discussed later in this chapter in the section on ‘the electron transport chain, oxidative phosphorylation and chemiosmosis’.

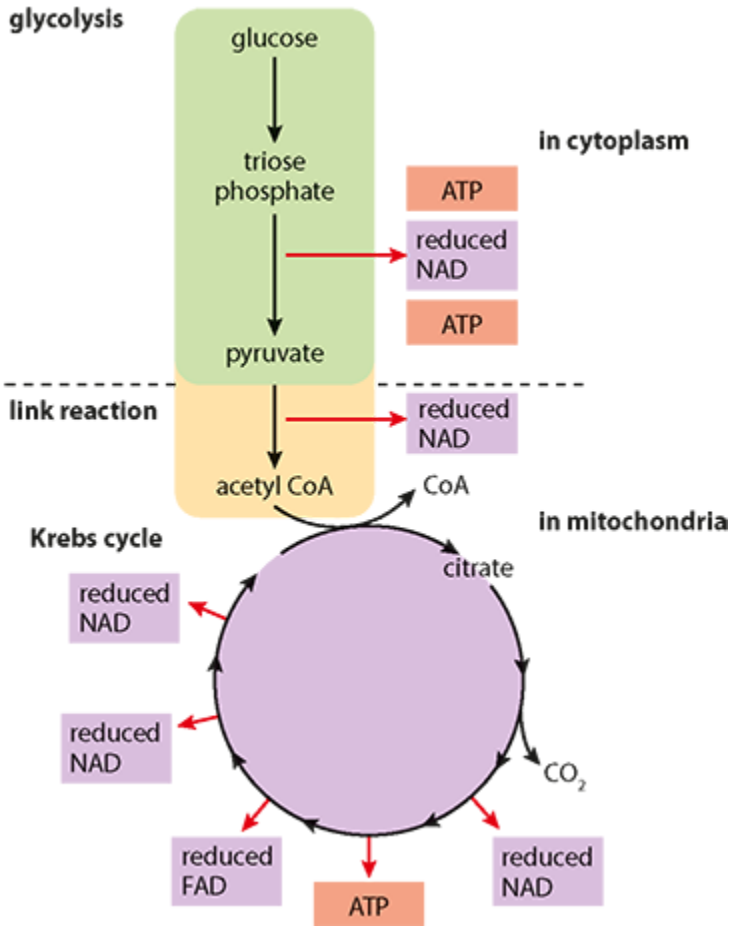


Figure 2.2.7: Summary of glycolysis, the link reaction and the Krebs cycle.

Glycolysis

Glycolysis is the first stage in the series of reactions that make up respiration. It literally means 'breaking apart glucose'. The glycolysis pathway occurs in the cytoplasm of the cell. It is anaerobic (that is, it can proceed in the absence of oxygen) and produces pyruvate and a small amount of ATP. One molecule of the hexose sugar glucose is converted to two molecules of the three-carbon molecule called pyruvate with the net gain of two molecules of ATP and two molecules of NADH + H^+. The process is shown in detail in Figure 2.2.8.

1 The first steps are to add two phosphate groups from ATP, in a process called phosphorylation, which destabilises the glucose molecule. A hexose bisphosphate molecule is produced. (This appears contrary to the purpose of respiration, which is to *make* ATP, but the two lost ATPs are recovered later.)

2 The hexose bisphosphate is now split into two triose phosphates in a reaction called lysis.

3 Now, another phosphorylation takes place but this time an inorganic phosphate ion, P_i, is used and not ATP. Two triose bisphosphates are formed. The energy to add the P_i comes from an oxidation reaction. The triose bisphosphate is oxidised and at the same time NAD^+ is reduced to NADH + H^+.

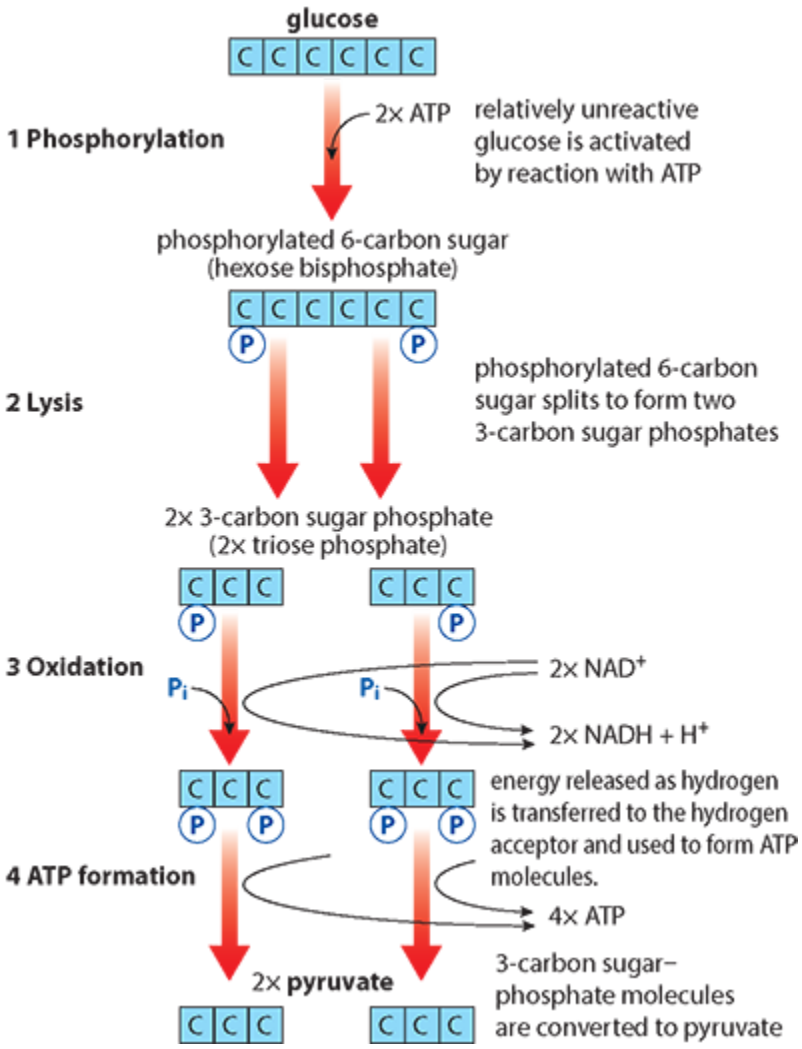


Figure 2.2.8: The stages of glycolysis. Note that for each molecule of glucose, two molecules of ATP are used and four are formed, so there is a net gain of two ATPs.

4 There now follows a series of reactions in which the two phosphate groups from each triose bisphosphate are transferred onto two molecules of ADP, to form two molecules of ATP: this is ATP formation. A pyruvate molecule is also produced for each triose bisphosphate molecule.

EXAM TIP

NADH + H^+ must not be simplified to $NADH_2$.

Four molecules of ATP are formed by converting one molecule of glucose to two molecules of pyruvate. However, two molecules of ATP were required to start the pathway and so there is a net gain of two molecules of ATP per glucose. In addition, two NADH + H^+ are formed.

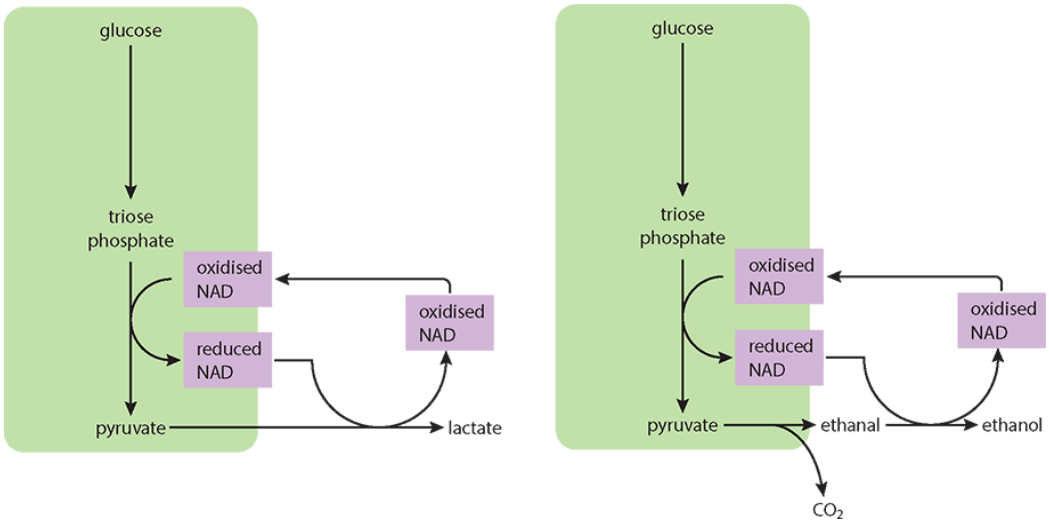


Figure 2.2.9: During anaerobic respiration reduced NAD is used to remove pyruvate from glycolysis so that it does not build up. Left: anaerobic respiration in animal cells; right: anaerobic respiration in yeast cells.

To summarise, the net products of glycolysis per glucose molecule are:

- 2 ATP
- 2 NADH + H^+ (reduced NAD)
- 2 molecules of pyruvate.

EXAM TIP

Try to think of your own acronym, such as People Love Outdoor Activities, to help you recall the steps in glycolysis.

In anaerobic respiration the reduced NAD (NADH + H^+) is oxidised (dehydrogenated) and used to reduce pyruvate. This prevents a build-up of harmful concentrations of pyruvate in the cell cytoplasm. In microorganisms this leads to the production of ethanol and carbon dioxide, while in animals the product is lactate. No further ATP is produced by the anaerobic respiration of pyruvate, so this type of respiration gives only a small yield of ATP from glucose (Figure 2.2.9).

Lipids and Carbohydrates as respiratory substrates

A respiratory substrate is any organic molecule that can be broken down to release energy to synthesise ATP. Glucose, lipids and proteins can all be used as respiratory substrates but they each release different amounts of energy.

KEY POINTS

Phosphorylation of glucose is the first stage in its breakdown and involves the addition of phosphate groups from ATP. This turns glucose into a more unstable phosphorylated compound which can be split to form two three-carbon sugars.

Energy coupling involves a sequence of reactions in which energy from an energy-releasing process is used to drive an energy-requiring process. Phosphorylation is an example of energy coupling: the transport of a phosphate group from ATP to a reactant molecule in the coupled reaction supplies energy for that reaction. The reactant molecule becomes a

phosphorylated intermediate, an unstable molecule compared to the unphosphorylated state.

Most of the energy released in respiration comes from the oxidation of hydrogen to water so the more carbon-hydrogen bonds there are in the structure of a molecule, the greater the amount of energy it can provide. Hydrogen atoms are used to generate ATP in the electron transport chain. Fatty acids have more hydrogen per gram than carbohydrates, so lipids have more oxidisable hydrogen and carbon and a greater energy value (Table 2.2.4)

Respiratory substrate	**Energy value/kj g^{-1}**
Carbohydrate	16
Lipid	39
Protein	17

When lipids are respired, they are broken down into fatty acids and glycerol. The glycerol is converted into triose phosphate and enters the glycolysis stage. The fatty acids are broken down into two carbon acetyl groups that enter the Krebs cycle via acetyl co-enzyme A.

Anaerobic respiration can only take place if carbohydrate is the energy substrate.

TEST YOUR UNDERSTANDING

13 List three ways in which a substance can be reduced.

14 Name the molecule used to phosphorylate glucose at the start of glycolysis.

15 Outline the importance of phosphorylation in glycolysis.

2.2.5 Aerobic respiration

The link reaction and Krebs cycle

If oxygen is present, pyruvate formed during glycolysis moves into the mitochondrial matrix by facilitated diffusion. The structure of a mitochondrion is shown in Figure 2.2.10.

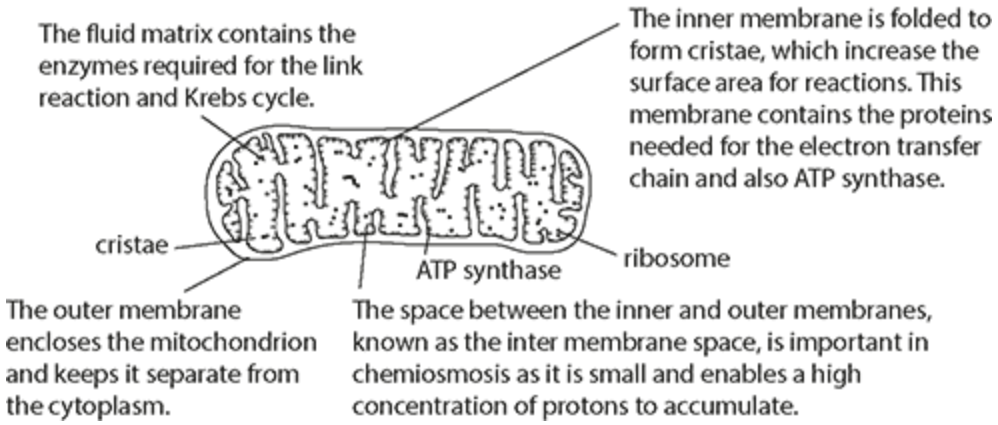


Figure 2.2.10: Diagram of a mitochondrion in longitudinal section.

The link reaction and the Krebs cycle pathways occur in the mitochondrial matrix (Figure 2.2.11).

1 The link reaction converts pyruvate to acetyl CoA using co-enzyme A, and a carbon atom is removed as carbon dioxide. This is called a **decarboxylation reaction**. At the same time as the carbon dioxide is removed, pyruvate is oxidised by the removal of hydrogen. The hydrogen atoms are removed by NAD^+ to form $NADH + H^+$.

2 Acetyl CoA now enters the Krebs cycle to continue the processes of aerobic respiration. Immediately, the co-enzyme A is removed to be recycled. The acetyl component

of the acetyl CoA combines with a four-carbon compound to form the six-carbon compound, citrate.

3, 4 The acetyl (two-carbon) groups are dehydrogenated to release four pairs of hydrogen atoms and decarboxylated to form two molecules of carbon dioxide so that the two carbons that enter with acetyl CoA leave as carbon dioxide.

5 One molecule of ATP is formed.

6 Hydrogen is removed during oxidation reactions to the two hydrogen carriers NAD^+ and FAD^+.

7 Since the Krebs cycle is a cyclic process, what enters must eventually leave so that the cycle begins and ends with the same substances.

EXAM TIP

Reduced FAD is written as $FADH_2$.

Because each molecule of glucose forms two molecules of pyruvate during glycolysis, each glucose molecule requires two link reactions and two rotations of the Krebs cycle. Thus, when working out the products of the cycle we must consider two sets of products. To summarise, the products of the link reaction and Krebs cycle, per glucose molecule, are:

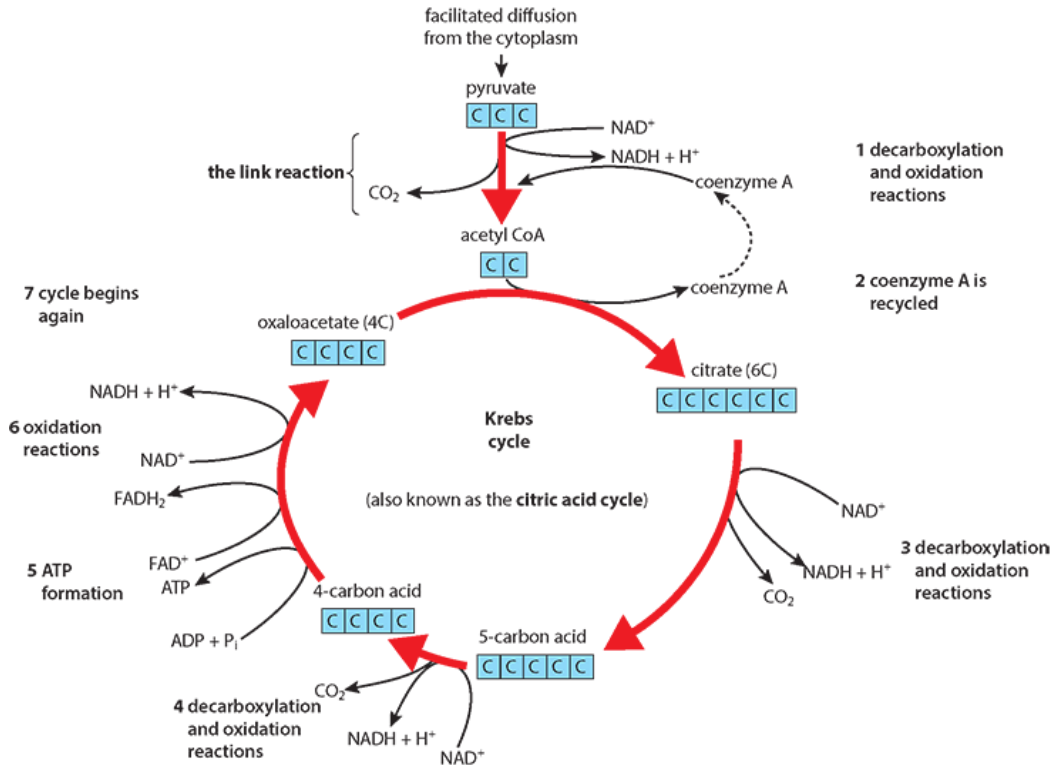


Figure 2.2.11: The link reaction and Krebs cycle.

- 8 molecules of NADH + H^+
- 2 molecules of $FADH_2$
- 2 molecules of ATP
- 6 molecules of CO_2.

The electron transport chain, oxidative phosphorylation and chemiosmosis

Most of the ATP produced from glucose breakdown occurs in the last phase of respiration at the end of the **electron transport chain (ETC)**. Reactions take place on the inner mitochondrial membrane of the cristae and in the intermembrane space between the inner and outer membranes (Figures 2.2.3 and 2.2.10). The inner membrane holds molecules called **electron**

carriers, which pick up electrons and pass them from one to another in a series of oxidations and reduction reactions. The pathway is called the electron transport chain because electrons from hydrogen are moved along it. Just as the inner lining of the small intestine is folded to increase its surface area to absorb food, so the inner mitochondrial membrane is highly folded into cristae to increase its surface area. The cristae provide a large area for the protein molecules used in the electron transport chain. Several protein molecules are electron carriers and the three key ones are shown in Figure 2.2.12.

Electrons from NADH + H^+ are transferred onto the first electron carrier. As they pass through the carrier, they lose energy and this is used to pump a proton (H^+) from the matrix to the intermembrane space, lowering the pH of the space. The electrons are then transferred to two further carriers and the process is repeated. As the electrons from one NADH + H^+ pass along the chain, a total of nine protons are pumped into the intermembrane space. At the end of the chain, the electrons are combined with protons and oxygen atoms to make water, in the oxidative part of **oxidative phosphorylation**. This completes the release of energy from the oxidation of glucose to produce ATP.

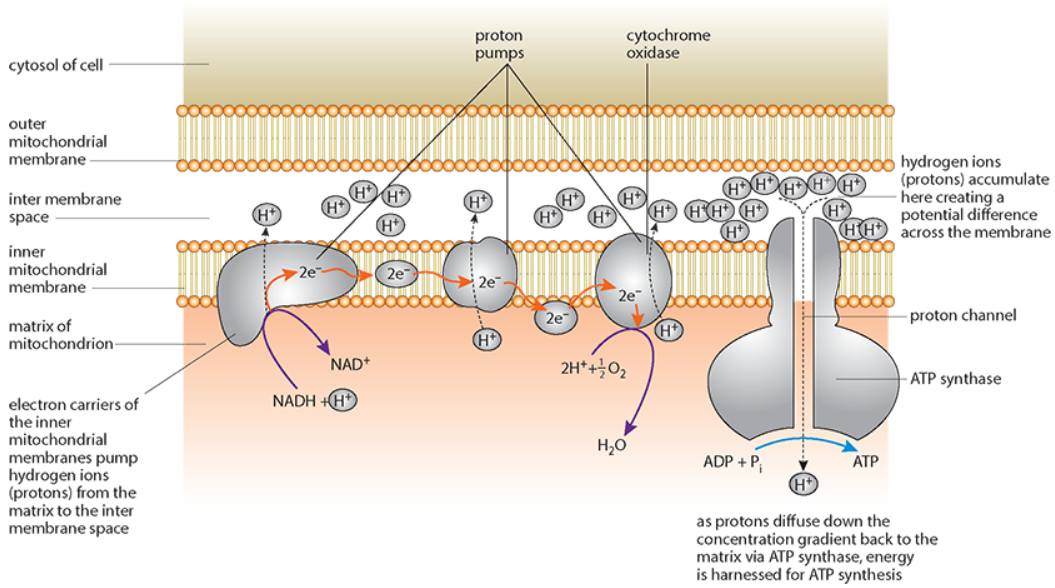


Figure 2.2.12: The electron transport chain showing oxidative phosphorylation and chemiosmosis.

The formation of water ensures that the H^+ gradient is maintained.

The space between the membranes is very narrow and allows for a rapid increase in the concentration of the protons that are pumped into it during the electron transfer reactions. The protons in the intermembrane space create a concentration gradient between the space and the matrix. These protons can now flow passively down this concentration gradient back into the matrix, through a very large integral protein. This is called **chemiosmosis**. The large protein contains the enzyme **ATP synthase**, which joins ADP and P_i to form ATP. Three protons flowing through this enzyme result in one ATP being formed. Since the electrons from one NADH + H^+ pump nine protons into the intermembrane space, each NADH + H^+ results in the formation of three ATP. This is the phosphorylation part of oxidative phosphorylation.

$FADH_2$ also supplies electrons to the electron transport chain but further down the chain than NADH + H^+, missing the first protein pump. $FADH_2$ allows the production of just two ATPs.

Overall ATP production during aerobic respiration

Stage		ATP use	ATP yield
glycolysis	2 ATP used at the start	−2 ATP	
	2 NADH + H^+		+4 ATP
	ATP formation		+4 ATP
link reaction	2 NADH + H^+		+6 ATP
Krebs cycle and electron transport chain	ATP formation		+2 ATP
	6 NADH + H^+		+18 ATP
	2 $FADH_2$		+4 ATP
net energy yield			+36 ATP

Table 2.2.3: Together, glycolysis, the link reaction and the Krebs cycle and the electron transport chain can yield 36 ATP molecules for each molecule of glucose broken down by aerobic respiration.

NATURE OF SCIENCE

Changing views with new discoveries

As Table 2.2.3 shows, the net production of ATP from one molecule of glucose is, in theory, 36. Biochemists have discovered that the actual production is closer to 30 ATPs and propose that this discrepancy occurs because some protons are used to transfer ATP from the matrix to the cytoplasm. There are also losses such as the cost of moving pyruvate, phosphate and ADP (for ATP synthesis) into the mitochondria.

EXTENSION

The reactions of respiration have the potential to release about 36 molecules of ATP from one molecule of glucose. The process of glycolysis produces 2 ATP and all the rest are produced during the electron transport chain.

But the exact number of ATP molecules generated from glucose is not as precise as the theory suggests. For example, the number of hydrogen ions pumped through the electron transport chain varies between species and ATP yield can also be reduced because the intermediate compounds in the respiration reactions are used for other reactions. The ribose sugars that build nucleic acids and some amino acids are made from intermediates of glycolysis which means fewer molecules proceed to the next stages of respiration. The percentage of potential ATP molecules that are actually produced from the catabolism of glucose can be as low as 40%.

NATURE OF SCIENCE

Paradigm shift: the chemiosmosis theory required a significant change of view

A paradigm shift occurs when a new theory radically changes our understanding of key concepts. It can change our view of how the natural world works.

The chemiosmosis hypothesis was proposed in 1961 by Peter Mitchell (1920–1992) to explain how the mitochondria convert ADP to ATP. At the start of the 1960s, scientists did not understand the exact mechanisms by which electron transfer is coupled to ATP synthesis. Various hypotheses at the time proposed a direct chemical relationship between oxidising and phosphorylating enzymes and suggested that a high-energy intermediate compound was formed. Mitchell's theory was completely new and proposed an indirect interaction between these enzymes with no intermediate compound. He suggested that ATP synthesis is driven by a reverse flow of protons down a concentration gradient, the so-called 'chemiosmotic theory'. This theory was first received with scepticism as his work was considered to be radical and outside the popularly held view. Mitchell struggled to persuade his contemporaries to reject the more accepted theories because his theory used a completely different approach. After several years of research, he published detailed evidence to support his theory, both in a pamphlet in 1966 and also in further publications in 1968, which were known as 'the little grey books' because of their bland covers. Eventually in the early 1970s, Mitchell's chemiosmosis theory gained scientific acceptance, and scientists conceded that no high-energy intermediate

compounds were likely to be found. Mitchell was awarded the Nobel Prize for Chemistry in 1978.

To consider:

1 Despite Peter Mitchell's strong evidence for chemiosmosis, which falsified earlier theories, he struggled to have his work accepted.

2 Why is it often difficult for a paradigm shift to gain acceptance?

TEST YOUR UNDERSTANDING

16 State the sites of the link reaction and the reactions of the Krebs carboxylic acid cycle.

17 Name the molecule that enters the Krebs cycle.

18 Where is ATP synthase located?

Links

- What is the importance of ATP for the movement of substances across cell membranes? (Chapter 6)
- How does the structure and function of a mitochondrion compare with that of a chloroplast? (Chapter 6)
- What is the significance of the inefficiency of respiration and heat losses to the environment? (Chapter 12)

2.3 Photosynthesis

LEARNING OBJECTIVES

In this section you will:

- learn that photosynthesis is a series of metabolic pathways carried out by plants, algae and some prokaryotes, uses light energy to make carbon compounds in cells
- recognise that the inputs for photosynthesis are water, carbon dioxide and light and the outputs are oxygen and glucose
- learn that light from the Sun is made up of a range of wavelengths, visible light has wavelengths between 400 nm (violet) and 700 nm (red)
- understand that chlorophylls are the main photosynthetic pigments and they absorb most red and blue light but reflect green more than the other colours. This phenomenon is shown by absorption spectra
- discover that action spectra demonstrate the wavelengths that are most effective for photosynthesis
- learn that photosynthesis consists of a light-dependent and a light-independent stage
- learn that the light-dependent stages take place in the thylakoids of chloroplasts
- learn that enzymes in the stroma of chloroplasts produce glucose using energy from the light-dependent stages

- understand how temperature, light intensity and carbon dioxide concentration affect the rate of photosynthesis and can be limiting factors
- understand why the net uptake or production of carbon dioxide and oxygen depend on both the rate of photosynthesis and cell respiration
- learn that carbon dioxide enrichment in greenhouses can promote plant growth and may be used as a method of predicting the effect of future rates of photosynthesis

- learn that the light-dependent reactions occur in the intermembrane space of the thylakoids and the intergranal lamellae and involve photosystems located in membranes
- understand how the light-dependent reactions lead to the production of reduced NADP
- discover how absorption of light excites electrons which are transferred between electron carriers
- learn that two photosystems are involved; excited electrons from photosystem I are used to reduce NADP and electrons from both photosystems are used to generate a proton gradient
- understand that the photolysis of water generates electrons to replace excited electrons and produce oxygen as a waste product
- learn that light-independent reactions occur in the stroma and are controlled by enzymes

- learn that a carboxylase catalyses the carboxylation of RuBP in the light-independent reactions of the Calvin cycle
- understand that reduced $NADP^+$ and ATP are used to reduce glycerate 3-phosphate to triose phosphate
- understand how ATP synthase in thylakoids generates ATP using the proton gradient
- learn that triose phosphate is used to produce carbohydrates and regenerate RuBP and ATP is also needed
- recognise that a wide range of organic molecules are derived from the glucose produced by photosynthesis.

GUIDING QUESTIONS

- How do plants and algae convert light energy into chemical energy that is stored in organic compounds?
- Why does photosynthesis depend on the presence of pigment molecules?
- Where in cells do the reactions of photosynthesis take place?

2.3.1 Photosynthesis and light

The Sun is the source of energy for almost all life on Earth. Plants, algae and some prokaryotes are able to convert light energy into chemical energy in organic compounds by the process of photosynthesis. Photosynthesis is a complex series of metabolic pathways which use carbon dioxide and water to produce glucose, other organic compounds and oxygen. Oxygen is released as a waste product. The series of reactions that occurs during photosynthesis is summarised as:

$$\text{carbon dioxide} + \text{water} \rightarrow \text{glucose} + \text{oxygen}$$

$$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$$

KEY POINT

photosynthesis ‘making things with light’. Glucose is the molecule most commonly made.

The energy stored in molecules such as glucose provides a source of food for organisms that cannot use light energy directly and, within the photosynthesising organisms, it can be converted into other organic compounds that are required for life.

Visible light is composed of a spectrum of colours, which can be separated using a prism (Figure 2.3.1). A prism bends rays of light and separates the colours because each one has a slightly different wavelength and is refracted (bent) to a slightly different degree. Visible light has a range of wavelengths that are between 400 and 700 nm. Violet light has the shortest wavelength and red the longest, but the most important regions of the spectrum for photosynthesis are red and blue.

The colour of any object is determined by the wavelength of the light that it reflects back into our eyes. A blue shirt appears blue because it reflects blue light, which our eyes can perceive, but the shirt absorbs other wavelengths that fall on it and we do not see those colours. A black object absorbs all wavelengths of light, while something white reflects them all.

Most plants have green leaves and many other photosynthesising organisms also appear green. This tells us that they do not absorb the green part of the spectrum well; green light is reflected and makes a leaf appear green. The leaves of plants (Figure 8.3.11) have cells which contain chloroplasts. Chloroplasts contain green pigments called chlorophylls which gives them their green colour. Chlorophylls are unable to absorb green light, which is reflected, but do absorb other wavelengths well. Red and blue light are absorbed particularly well and provide the energy needed for photosynthesis. The top graph in Figure 2.3.2 shows that the red and blue ends of the visible spectrum are the wavelengths that the photosynthetic pigments in plants absorb most efficiently. The bottom graph, known as an action spectrum, shows that the rate of photosynthesis is highest when plants absorb these wavelengths.

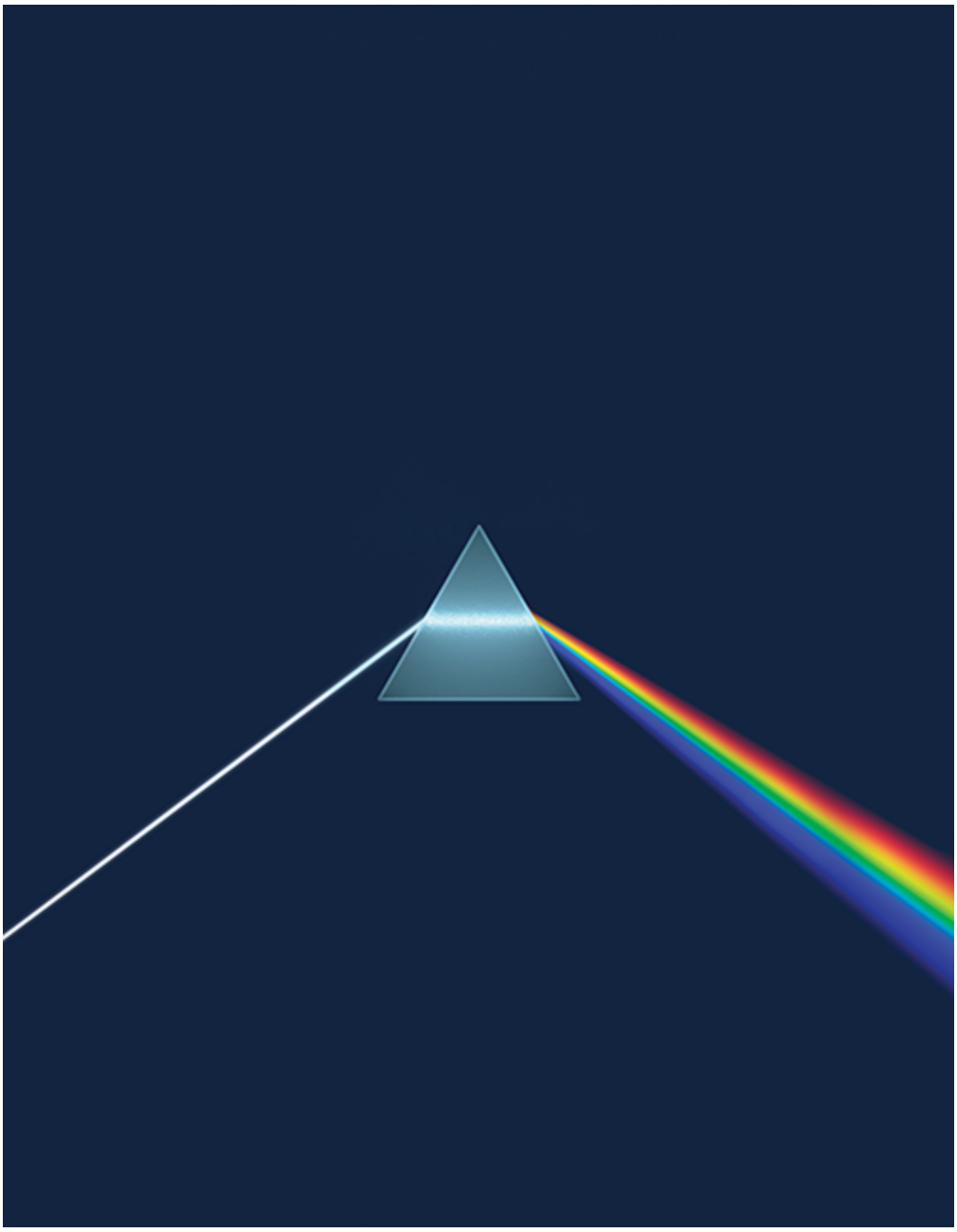


Figure 2.3.1: ‘White light’, such as sunlight, is composed of a range of wavelengths, which become separated as they pass through a glass prism.

KEY POINT

chlorophyll the name for the most important group of photosynthetic pigments of green plants, found in the grana of chloroplasts and responsible for trapping light energy (some bacteria have a chemically different form called bacteriochlorophyll).

Photosynthetic pigments

Chloroplasts contain a number of different pigments that are associated with light absorption. Figure 2.3.2 shows **absorption spectra** and **action spectra** for two types of chlorophyll pigment and carotenoid pigments found in green plants. There is a strong correlation between the absorption spectra (the range of wavelengths of light that a pigment is able to absorb) of the pigments and the action spectrum (showing the rate of photosynthesis at each wavelength). In the figure, both have two peaks, one in the blue region and a smaller one in the red region, and both are lower in the green and yellow areas of the spectrum.

EXAM TIP

Chromatography is a simple technique used to separate different substances in a mixture and it can be used to separate the pigments in extracts from plant leaves.

Two techniques are commonly used: paper chromatography, which uses a special high-grade paper with carefully controlled spaces between the cellulose fibres, and thin-layer chromatography (TLC), which is carried out on a thin plate of glass or plastic coated with a layer of adsorbent material such as silica gel or cellulose (known as the stationary phase).

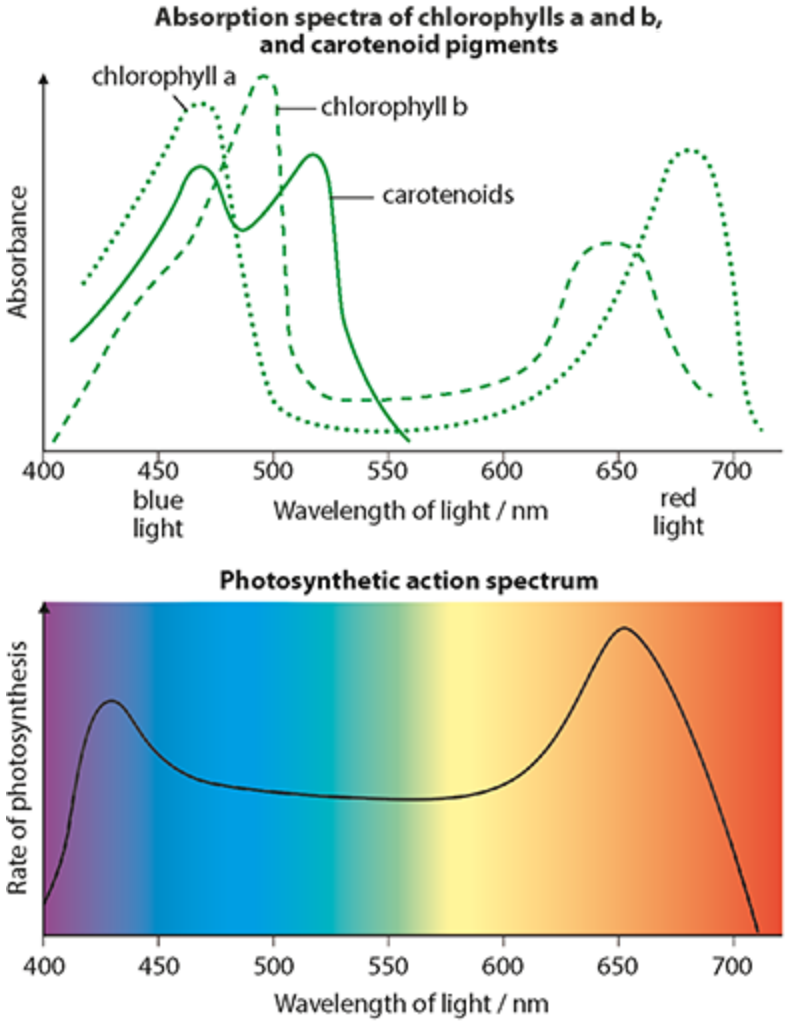


Figure 2.3.2: These graphs show the wavelengths (colours) of light absorbed by plants and the rate of photosynthesis that occurs at each wavelength.

During chromatography a solvent moves up the paper or plate by capillary action and carries pigments with it by mass flow. Smaller molecules are able to move more easily and so can travel further than larger molecules. After a period of time, photosynthetic pigments from chloroplast extracts become separated (Figure 2.3.3) and can be compared and measured.

EXAM TIP

You should be able use experiments like the one shown in Figure 2.3.3 to work out R_f values and identify different pigments.

EXTENSION

The **R_f value** is a ratio that is used to identify components in a mixture from a chromatogram. It is calculated by dividing of the distance travelled by a component of the mixture by the distance travelled by the solvent front from the origin. The R_f value can be used to identify each pigment by comparing its R_f value to that of a known standard at the same temperature using the same type of chromatogram.

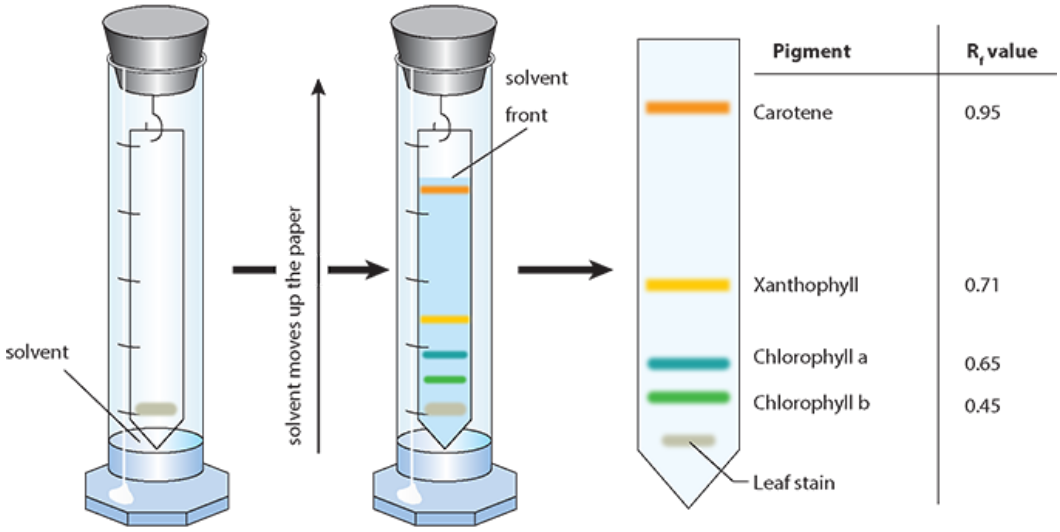


Figure 2.3.3: A chromatogram can be used to identify different photosynthetic pigments. Different pigments are found in different plants and the pigments may vary with the seasons.

NATURE OF SCIENCE

Careful observation: Engelmann's experiment

Theodor Wilhelm Engelmann (1843–1909) was a German botanist who used the filamentous alga *Spirogyra* to demonstrate not only that oxygen is evolved during photosynthesis but also that different wavelengths of light affect the rate of photosynthesis. *Spirogyra* is an alga that has cylindrical cells containing spiral-shaped chloroplasts. Engelmann mounted a sample of *Spirogyra* under a microscope and, after a period of darkness, illuminated it with different colours of light. He carefully watched the movement of motile bacteria (*Pseudomonas*) that he had added to the water and noticed that, after a period of oxygen deprivation, they moved towards areas where there was a higher concentration of oxygen around the alga's chloroplasts (Figure 2.3.4).

TEST YOUR UNDERSTANDING

19 Use these questions to analyse the results of Engelmann's experiments.

a Look at Figure 2.3.4a and explain why the bacteria moved towards certain areas of the *Spirogyra* and not towards others when a spot of red light was used.

b Explain why there are no bacteria between the areas of *Spirogyra* illuminated with light of 500–600 nm (Figure 2.3.4b).

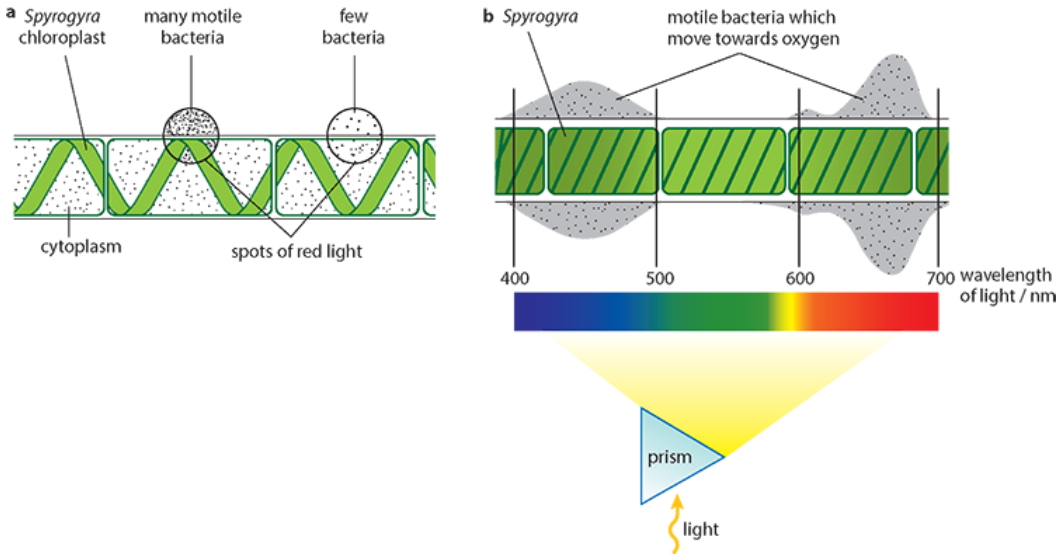


Figure 2.3.4: Engelmann used bacteria in two experiments a and b to measure rates of photosynthesis and determine which wavelengths are most effective for photosynthesis.

2.3.2 The chemistry of photosynthesis

Photosynthesis is a complex series of reactions catalysed by a number of different enzymes. The processes take place in chloroplasts (Figure 2.3.5; see also Section 5.2). To help us understand the reactions, we can consider photosynthesis in two stages: the light-dependent reactions and the light-independent reactions.

Light-dependent reactions

The first stage of photosynthesis is known as the 'light-dependent reactions' because light is essential for them to occur. These take place in the thylakoids of the chloroplast.

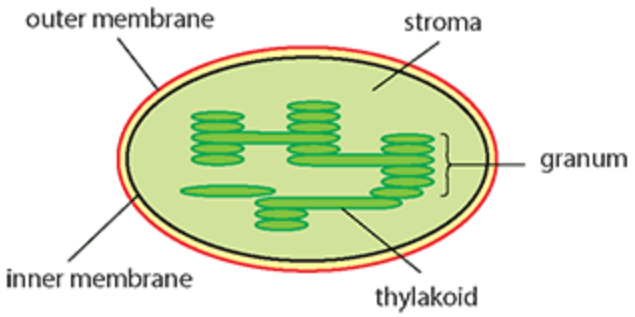


Figure 2.3.5: Structure of a chloroplast.

Chlorophyll in the **thylakoids** absorbs light energy and this energy is used to produce ATP. The energy is also used to split water molecules into hydrogen and oxygen in a process called **photolysis**. Hydrogen ions and electrons (from the hydrogen part of water) and oxygen are released. Oxygen is a waste product of photosynthesis but is vital to sustain the lives of aerobic organisms once it has been released into the atmosphere. The ATP, hydrogen ions and electrons are used in the light-independent reactions.

Light-independent reactions

ATP, hydrogen ions and electrons are used in the second stage of photosynthesis, the 'light-independent reactions', which take place in the stroma.

During the 'light-independent reactions', enzymes in the stroma use carbon dioxide, taken in from the environment, and combine it with hydrogen using energy from ATP. The reactions form mainly glucose, but also a range of other organic molecules for the plant. The conversion of inorganic carbon dioxide to organic molecules such as glucose is known as **carbon fixation**. ATP provides the energy for the process.

The series of reactions that occurs during photosynthesis is summarised as:

$$\text{carbon dioxide} + \text{water} \rightarrow \text{glucose} + \text{oxygen}$$

$$6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$$

Measuring the rate of photosynthesis

The equation for carbon dioxide and water shows that when photosynthesis occurs, carbon dioxide is used and oxygen is released. The mass of the plant (its **biomass**) will also increase as glucose is used to produce other plant materials. Any of these three factors can be used to measure how quickly the reactions of photosynthesis are occurring.

Aquatic plants release bubbles of oxygen as they photosynthesise and if the volume of these bubbles is measured for a period of time, the rate of photosynthesis can be determined directly (Figure 2.3.6).

Aquatic plants also remove carbon dioxide from their environment, causing the pH of the water to rise. Carbon dioxide dissolves in water to form a weak acid, so as it is removed the pH will go up. Therefore, another way of determining the rate of photosynthesis experimentally is to monitor the change in pH of the water surrounding an aquatic plant over a period of time.

Terrestrial plants also remove carbon dioxide from their surroundings but this is difficult to measure. It can be done experimentally by supplying a confined plant with radioactive carbon dioxide, which can be measured as it is taken up and released from the plant.

Figure 2.3.6: The rate of oxygen production can be used as a direct measure of the rate of photosynthesis.

A third method of measuring the rate of photosynthesis in plants is to determine their biomass at different times. This is an indirect method. Samples of the plants can be collected and measured at different times and the rate of increase in their biomass calculated to determine their rate of photosynthesis.

2.3.3 Limits to photosynthesis

The rate at which a plant can photosynthesise depends on factors in the environment that surrounds it. On a warm, sunny afternoon, photosynthesis will be more rapid than on a cool, shady morning.

More oxygen will be produced and more carbon dioxide used. Temperature, light intensity and carbon dioxide concentration are all possible limiting factors on the rate of photosynthesis. But photosynthesis cannot increase beyond certain limits. The effect of light, temperature and carbon dioxide in the environment can be measured experimentally, varying one factor while keeping the others the same, and graphs such as those in Figure 2.3.7 can be drawn.

An increase in light intensity, when all other variables are unchanging, will produce an increase in the rate of photosynthesis that is directly proportional to the increase in light intensity. However, at a certain light intensity, enzymes will be working at their maximum rate, limited by temperature and the availability of carbon dioxide. At very high light intensities, light absorption (and therefore the rate of photosynthesis) reaches its maximum and cannot increase further. At this point, the graph reaches a plateau (Figure 2.3.7a).

Increasing temperature also increases the rate of photosynthesis as the frequency and energy of molecular collision increases (Figure 2.3.7b). Photosynthesis has an optimum temperature above which the rate will decrease sharply as enzymes are denatured, or the plant wilts and is unable to take in carbon dioxide.

An increase in the concentration of carbon dioxide causes the rate of photosynthesis to increase, as carbon dioxide is a vital raw material for the process. At very high concentrations, the rate will plateau as other factors such as light and temperature limit the rate of reaction (Figure 2.3.7c).

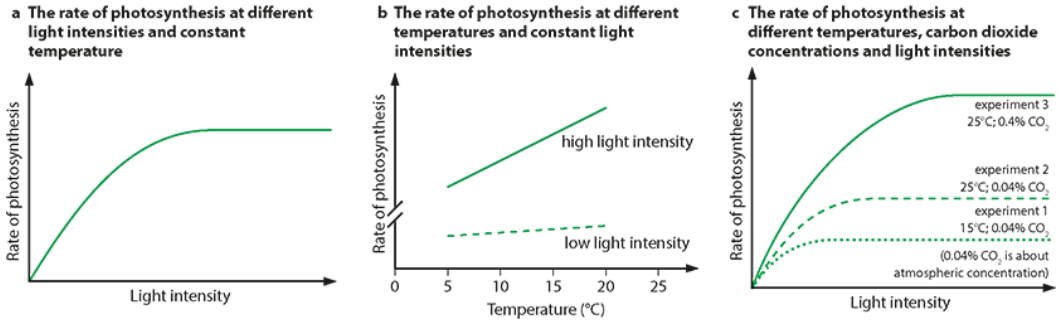


Figure 2.3.7: These graphs show the effects on photosynthesis of varying light intensity, temperature and carbon dioxide concentration.

SCIENCE IN CONTEXT

The effects of temperature, light and carbon dioxide concentration are well known to horticulturalists who grow crops in glasshouses. Commercial producers of cucumbers and tomatoes keep their glasshouses warm and well lit. They may also seal the greenhouse and introduce carbon dioxide to boost photosynthesis to its maximum rate, thereby increasing crop production and profits.

Carbon dioxide enrichment experiments have been used to predict future rates of photosynthesis as the amount of carbon dioxide in the air increases. Some experiments are carried out in sealed greenhouses, others known as free-air carbon dioxide experiments (FACE) are conducted outdoors.

Research these experiments and consider how scientists control the different variables.

Compensation point

The net production of oxygen (or uptake of carbon dioxide) will depend on both the rate of cellular respiration and the rate of photosynthesis. When the two are in equilibrium, the amount of oxygen being produced by photosynthesis will be equal to that which the plant is using in respiration. This is known as the compensation point (Figure 2.3.8).

KEY POINT

compensation point is when the light intensity at which the amount of carbon dioxide released in respiration equals the amount used in photosynthesis, and at which the amount of oxygen used in respiration equals the amount released in photosynthesis.

EXAM TIP

Remember that photosynthesising organisms respire throughout the day and night but can only photosynthesise when light is available.

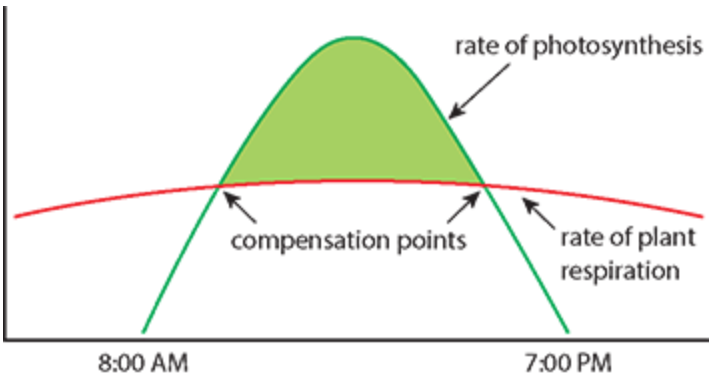


Figure 2.3.8: Graph showing the rate of photosynthesis and respiration for a plant over a period of 24 hours.

NATURE OF SCIENCE

Experimental design: controlling variables

Any investigations involving living organisms must be designed so that all the possible variables are controlled. Consider this diagram of apparatus (Figure 2.3.9) that has been set up in a laboratory to estimate the rate of photosynthesis of a pond plant.

To consider:

1. Which variable is being controlled by the presence of the beaker of water?
2. Why must this variable be controlled?
3. Photosynthesis is a metabolic reaction controlled by enzymes. List the factors that affect enzyme action (Section 3.1).
4. How could each one of these be controlled in this experiment?

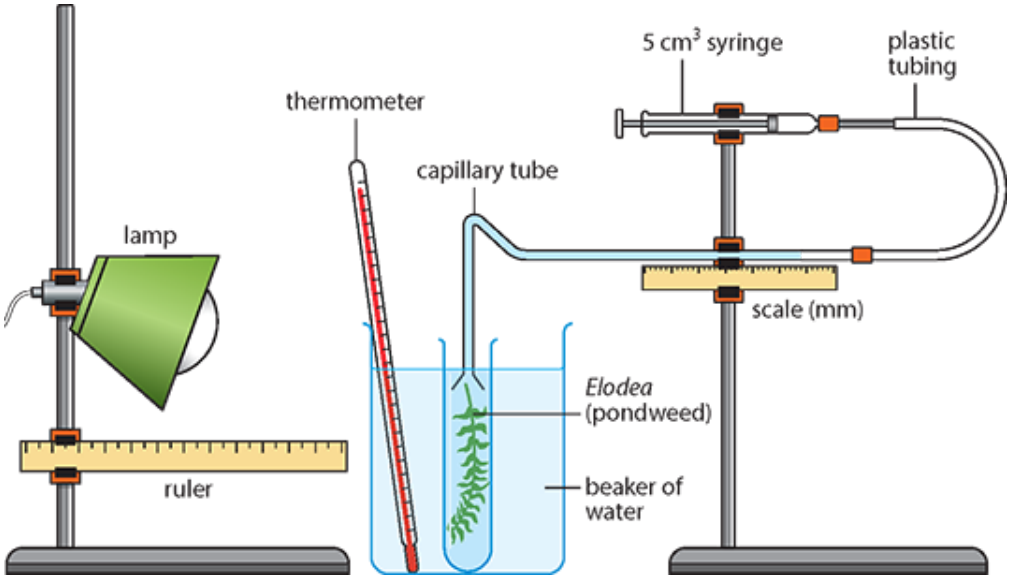


Figure 2.3.9: Diagram of experiment to determine the rate of photosynthesis.

TEST YOUR UNDERSTANDING

20 If you want to make plants grow as efficiently as possible, what colour of light should you shine on them?

21 Suggest what would happen to a plant's growth if it was illuminated by green light.

22 Where in the chloroplast does the light-dependent reactions take place?

23 Outline two ways in which the rate of photosynthesis can be measured.

2.3.4 Advanced photosynthesis

Photosynthesis is the process by which light energy is harvested and stored as chemical energy, primarily in sugars but also in other organic molecules such as lipids. It occurs in green plants, algae and some bacteria. All these organisms are known as **autotrophs**, which means they can make their own food.

Photosynthesis can be divided into two parts:

- the light-dependent reactions
- the light-independent reactions.

The light-dependent reactions produce compounds that are used in the light-independent reactions.

KEY POINTS

light-dependent reactions occur on the thylakoids and produce ATP and NADPH

light-independent reactions series of stages in photosynthesis that take place in the stroma and use the products of the light-dependent reactions to produce carbohydrate.

photosystems arrays of pigment molecules that can generate and emit excited electrons

Both the light-dependent and the light-independent reactions take place in the chloroplasts of plant cells (Figures 2.3.5, 2.3.10 and 2.3.11). The stroma contains the enzymes required for the light-independent reactions and the stacks of thylakoid membranes increase the surface area for the light-dependent reactions.

Both these sets of reactions are part of photosynthesis and can only occur when there is sufficient light. Light-dependent reactions can only take place in light and although light-independent reactions do not require light directly – and can take place when it is dark – they do require the products of the light-dependent reactions.

The light-dependent reactions

The **light-dependent reactions** are a series of stages in photosynthesis that occur on the grana of the chloroplasts. Light is used to split water, and ATP and NADPH + H^+ are produced.

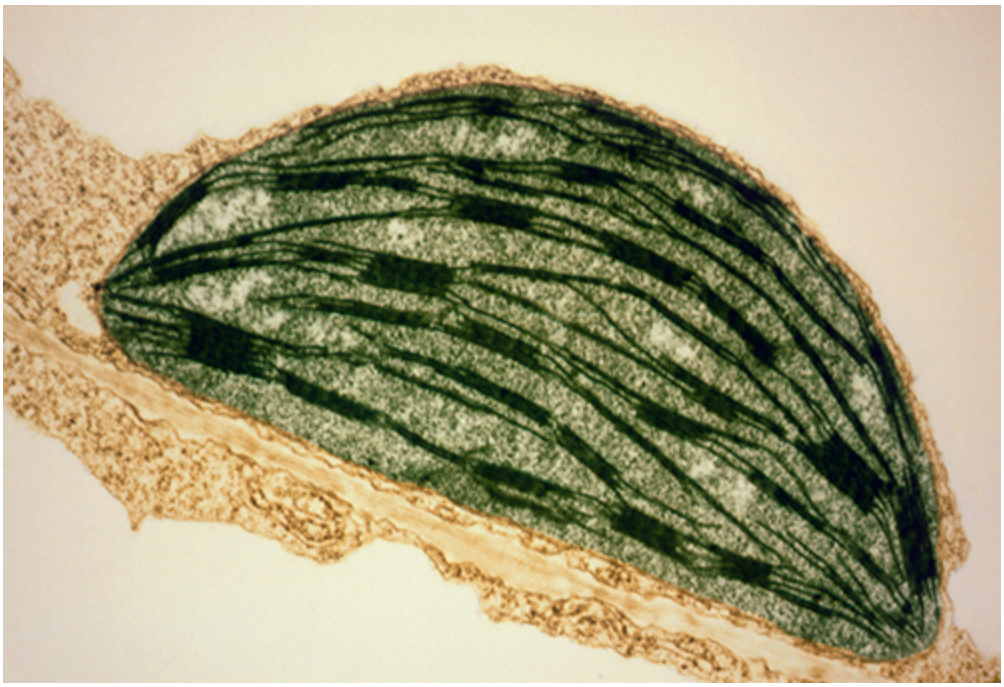


Figure 2.3.10: Coloured electron micrograph of a chloroplast (× 20 000).

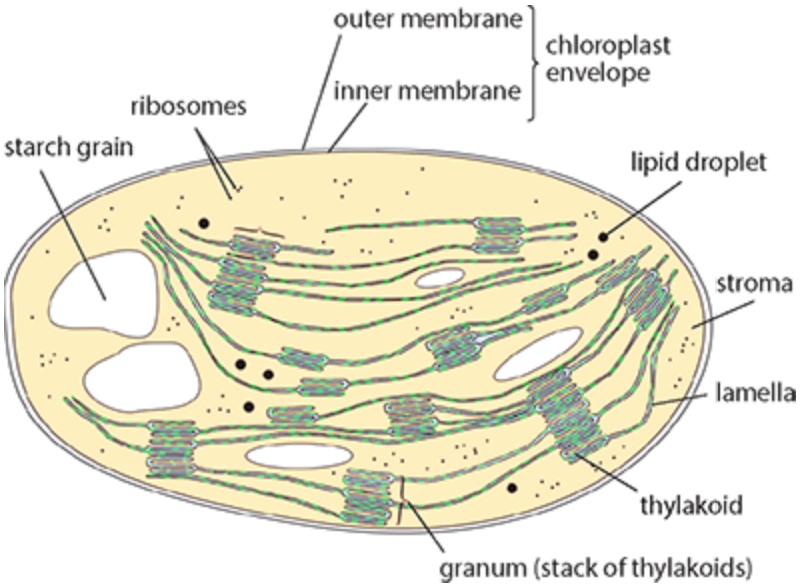


Figure 2.3.11: Diagram of a chloroplast.

The reactions take place on the thylakoid membranes that make up the grana of the chloroplast and are powered by light energy from the Sun. Each thylakoid is a flattened sac so the space in the middle is narrow. The thylakoid membranes form the stacks called **grana**, which may be joined together by intergranal lamellae (membranes). Light is absorbed by photosynthetic pigments, including the chlorophylls, which are found on the granal membranes. There are several pigments found in plants and each one absorbs light of a slightly different wavelength.

The photosynthetic pigments are combined into two complex groups called **photosystems** I and II (PS I and PS II), which absorb the light energy and use this to boost electrons to a higher energy level so that they become 'excited'. The pigments are associated with proteins that are involved in electron transport, proton pumping and chemiosmosis. The arrangement of photosystems in membranes and of different pigment molecules in the photosystems means that the passage of electrons is structured effectively to enable photosynthesis to take place.

Both PS I and PS II have a chlorophyll a molecule at their centre together with accessory pigments, such as chlorophyll b and carotenoids, around them.

At the centre of each chlorophyll a molecule is a magnesium (Mg^{2+}) ion which is essential to its structure and functioning (Section 1.1). Fluctuations in magnesium levels in the chloroplast regulate the activity of key photosynthetic enzymes.

In the light-dependent reactions, electrons are removed from water and passed through PS II and PS I before ending up in NADPH. This process needs light to be absorbed by both photosystems and it also produces ATP. The process is called **photophosphorylation** because it uses light energy to produce ATP from ADP.

The key stages of the light-dependent reactions are as follows:

1 Photoactivation of PS II

Light is absorbed by pigments in PS II and passed to the reaction centre. Here energy boosts an electron in chlorophyll a to a higher energy level. The electron is passed to an acceptor molecule at the start of the electron transport chain.

Lost electrons must be replaced and this is done by taking them from water. Water is split into electrons, protons (hydrogen ions) and an oxygen atom. Since the splitting is brought about by light energy, it is called photolysis. The oxygen is released as a waste product and is the oxygen we breathe.

2 ATP synthesis

Excited, high-energy electrons travel down the electron transport chain into PS I. As they do this, they lose energy which is used to pump protons into the thylakoid interior (in a similar way as occurs in the electron transport chain in the mitochondrion). The thylakoid interior is small and so a proton concentration gradient builds up quickly. The protons then flow out through a large channel protein, almost identical to the one in mitochondria, which contains the enzyme ATP synthase. Ions flow down their gradient and into the stroma, driving ATP production in a process known as chemiosmosis. This time though, the formation of ATP is called photophosphorylation and it occurs between PS II and I (Figure 2.3.12).

3 Light absorption in PS I

Absorption of light energy causes photoactivation in PS I, boosting more electrons to an even higher energy level. The electrons that arrive from PS II replace those that are displaced.

4 NADPH formation

The electrons at the higher energy level continue down the electron transport chain and are combined with protons in the hydrogen carrier $NADP^+$ to form $NADPH + H^+$.

The two products of the light-dependent reaction, ATP and $NADPH + H^+$, are used to drive the light-independent reaction.

The net effect of these steps is to convert light energy into chemical energy in the form of ATP and NADPH. The ATP and NADPH from the light-dependent reactions are used to make sugars in the next stage of photosynthesis, the Calvin cycle.

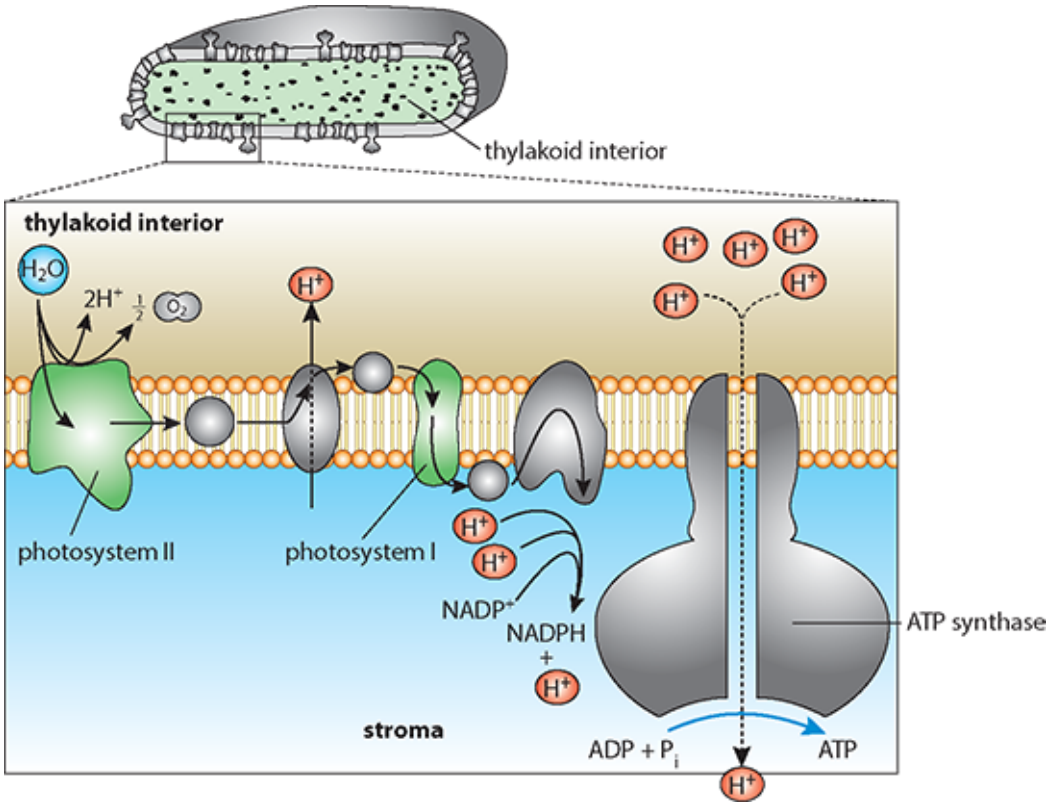


Figure 2.3.12: Chemiosmosis in photosynthesis.

The light-independent reactions

The light-independent reactions occur in the stroma of the chloroplast. These reactions are catalysed by enzymes and therefore are temperature dependent.

The reactions follow a cyclic pathway called the **Calvin cycle** (shown in Figure 2.3.13). ATP and NADPH + H^+ formed during the light-dependent stage supply energy and reducing power for the Calvin cycle. The final product of the cycle is carbohydrate.

KEY POINT

Calvin cycle of light-independent reactions in the stroma of the chloroplast in which carbon dioxide reacts with ribulose

bisphosphate (RuBP), producing glycerate 3-phosphate, triose phosphate and regenerating RuBP.

During each turn of the Calvin cycle, one molecule of carbon dioxide is used, so Figure 2.3.13 shows three cycles combined together. As this is a cycle, what goes in must leave, so three carbons enter in three molecules of carbon dioxide and three carbons leave in one molecule of triose phosphate, which can be used to form glucose or other organic compounds. There are three stages in the cycle, as follows.

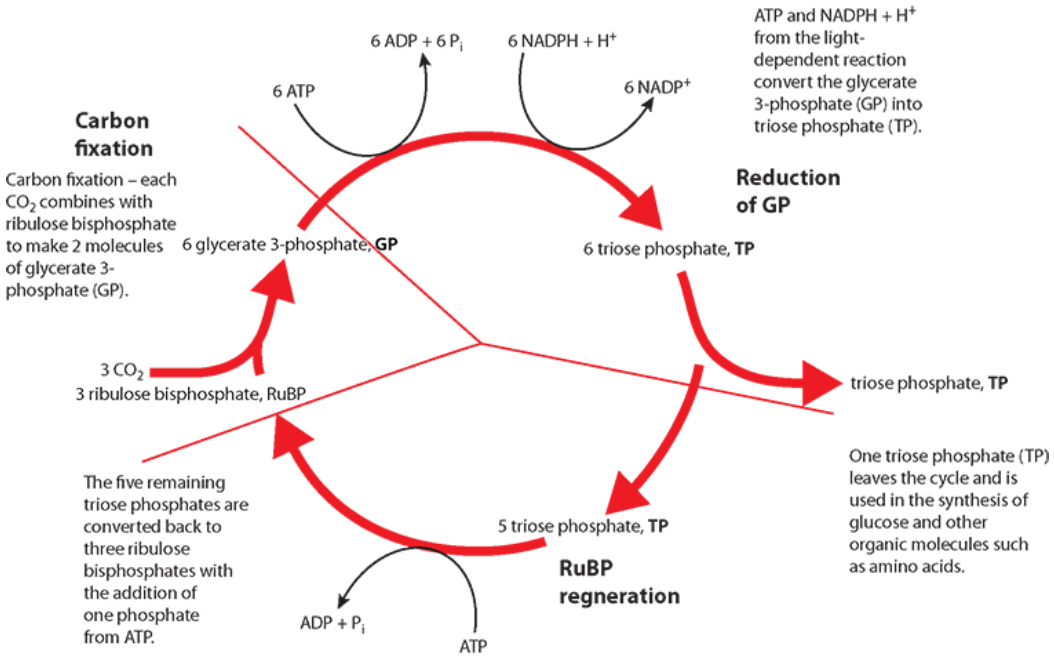


Figure 2.3.13: The light-independent pathway of photosynthesis – the Calvin cycle.

1 Carbon fixation

At the start of the cycle, the acceptor molecule ribulose bisphosphate (RuBP) combines with incoming carbon dioxide from the air to form glycerate 3-phosphate (GP). This reaction is called carbon fixation because it ‘locks’ the

carbon into the cycle. It is catalysed by RuBP carboxylase (this enzyme is sometimes called Rubisco).

2 Reduction of glycerate 3-phosphate

The ATP and NADPH + H^+ from the light-dependent reaction convert the glycerate 3-phosphate into triose phosphate (TP). Glycerate 3-phosphate is reduced to triose phosphate. No more phosphate is added so the only input from ATP is energy.

Six molecules of triose phosphate are produced but only five are needed to reform the ribulose bisphosphate to keep the cycle going. The extra triose phosphate leaves the cycle and is used to synthesise organic molecules such as glucose or amino acids.

3 RuBP regeneration

The triose phosphate that leaves the cycle takes a phosphate with it, so this is replaced in the cycle with a phosphate from ATP, as the five remaining triose phosphates are converted back to three ribulose bisphosphate molecules, and the cycle begins again. This is an example of a cyclic metabolic pathway.

Six 'turns' of the Calvin cycle produce two triose phosphate molecules, which can be combined to form the final product, glucose. Many triose phosphate molecules will be converted immediately to starch. Other triose phosphate molecules are exported to the cytoplasm to be converted to sucrose and transported around the plant. The products of the Calvin cycle are used to make all the other different organic molecules the plant needs, such as cellulose, amino acids, fatty acids and vitamins.

KEY POINT

RuBP carboxylase (Rubisco) is the enzyme that catalyses carbon fixation in the Calvin cycle in the light-independent reactions of photosynthesis; in carbon fixation, the five-carbon acceptor molecule ribulose bisphosphate (RuBP) combines with carbon dioxide to form glycerate 3-phosphate.

TEST YOUR UNDERSTANDING

24 Where do the light-dependent reactions take place?

25 State the colour of light that is not absorbed by green plants and algae.

26 Where in a chloroplast do you find magnesium ions?

27 State the names of the two products of the light-dependent reactions that are needed for the light-independent reactions.

28 State the name of the acceptor molecule that reacts with carbon dioxide in the Calvin cycle.

Links

- How does the structure of chlorophyll compare with that of hemoglobin? (Chapter 1)
- How does the structure of a chloroplast compare with that of a mitochondrion?
- What is the origin of chloroplasts in eukaryotic cells? (Chapter 5)

- What are the similarities and differences between chemiosmosis in respiration and photosynthesis?

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
define metabolism and state that reactions may take place inside or outside cells	2.1.1			
recall that metabolic pathways may be linear or cyclical	2.1.1			
recall that metabolic processes may be anabolic or catabolic	2.1.1			
explain the importance of enzymes and their properties in metabolism	2.1.2			

outline the induced-fit hypothesis of enzyme action	2.1.2			
summarise the importance of temperature, pH and substrate concentration on enzyme action and interpret graphs which show these	2.1.2			
outline how enzymes lower activation energy	2.1.3			
summarise how enzymes can be affected by other molecules that bind to them at their active or allosteric sites	2.1.4			
describe the effects of competitive and non-competitive inhibitors on rates of reaction	2.1.4			
outline end-product inhibition	2.1.5			

define co-enzymes and co-factors and give an example of each	2.1.6			
explain the inputs and products of cellular respiration	2.2.1			
recall that respiration occurs as a series of metabolic pathways catalysed by enzymes	2.2.1			
outline the uses of energy-requiring actions and recall that energy is lost from organisms as heat	2.2.1			
outline the stages of anaerobic respiration and explain its lower energy yield	2.2.2, 2.2.3			
recall that aerobic respiration has a higher yield of energy, takes	2.2.2			

place in mitochondria and requires oxygen to produce carbon dioxide and water				
outline the importance of redox reactions involving NAD and FAD	2.2.4			
define phosphorylation, decarboxylation and explain their importance in cell respiration	2.2.4, 2.2.5			
list the four steps involved in aerobic respiration	2.2.4			
compare the products and energy output of aerobic and anaerobic respiration	2.2.5			
draw a mitochondrion and indicate where reactions of	2.2.5			

respiration take place				
summarise the reactions that occur in the ETC	2.2.5			
define chemiosmosis and explain its importance in generating ATP	2.2.5			
explain the inputs and products of photosynthesis and write an equation that summarises the reactions	2.3.1			
recall the colours of light that are used for photosynthesis	2.3.1			
explain the difference between an absorption spectrum and an action spectrum	2.3.1			
name the two stages of photosynthesis	2.3.2			

and state where they take place				
name three limiting factors on the rate of photosynthesis and sketch graphs to show their effects	2.3.3			
explain how photosynthesis and respiration are related in algae and plants	2.3.3			
outline the location and structure of the two photosystems	2.3.4			
state the precise locations of the light-dependent and light-independent reactions in the chloroplast	2.3.4			
summarise the differences between the light-dependent and light-independent reactions	2.3.4			

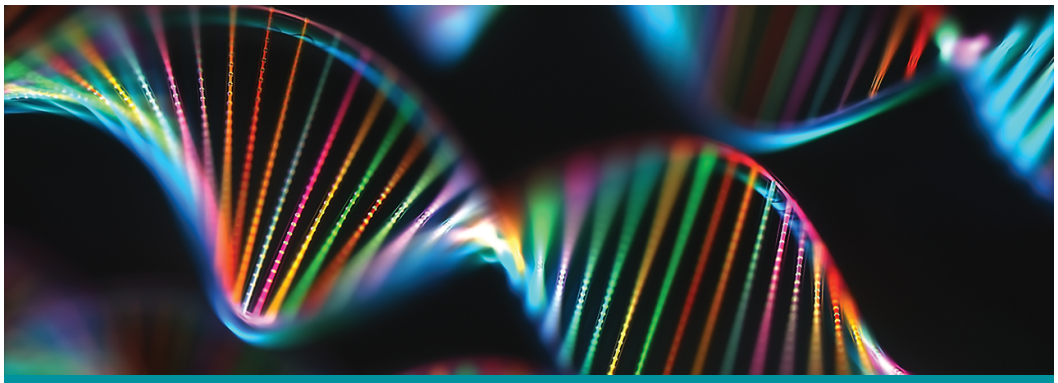
name the products of the light-dependent reactions that enter the light-independent reactions	2.3.4			
summarise the three stages of the Calvin cycle	2.3.4			
outline the importance of carboxylase, RuBP and glycerate 3-phosphate in the Calvin cycle.	2.3.4			

REFLECTION

How well is your understanding of respiration developing? Could you explain the various stages to a fellow student?

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.



Chapter 3

DNA and protein synthesis

D1.1, D1.2, D1.3

INTRODUCTION

Nucleic acids are very large macromolecules composed of a backbone of sugar and phosphate molecules each with a nitrogenous base attached. In Chapter 2 the basic structure of these molecules was considered. Here we will look at the vital role of nucleic acids in producing the proteins we need for life and how the genetic information contained in DNA is passed from one generation to the next.

3.1 DNA replication

LEARNING OBJECTIVES

In this section you will:

- understand that DNA replication is a semi-conservative process that produces two identical new molecules
- learn that the enzyme helicase unwinds the double helix and separates the strands
- understand that the polymerase chain reaction is a laboratory process that amplifies small quantities of DNA
- recognise that gel electrophoresis separates DNA fragments by their charge and size and is useful in paternity and forensic investigations

- recognise that DNA strands are antiparallel and are orientated in opposite directions
- learn that DNA replication is regulated by a series of enzymes: primase, polymerase and ligase
- understand that DNA polymerase can only work in a 5′ to 3′ direction
- discover that DNA polymerases proofread new DNA strands.

GUIDING QUESTIONS

- How is inherited material copied?
- How do laboratory techniques enable us to analyse DNA?

3.1.1 DNA replication

An essential feature of DNA is that it must be able to replicate itself accurately, so that when a cell divides, the genetic code it carries can be passed on to the daughter cells. DNA replication copies DNA precisely so that new molecules are produced with exactly the same sequence of bases as the original strands. DNA replication takes place in the nucleus during interphase of the cell cycle when DNA is not tightly coiled (Section 6.5).

KEY POINTS

DNA replication is copying DNA so that two identical new molecules are produced.

replication fork is the point where the DNA double helix is being separated to expose the two strands as templates for replication.

semi-conservative replication happens when both strands of a DNA double helix are used as templates for replication so that new DNA molecules contain one original and one new strand.

As Figure 3.1.1 shows, DNA replication does not occur in a haphazard manner. An enzyme called DNA helicase unzips one region of the DNA molecule and nucleotides are added in a step-by-step process that links them to one another and to their complementary bases in an area known as the replication fork.

1 The first step in the process is the 'unzipping' of the two strands. **DNA helicase** moves along the double helix, unwinding the two strands, which separate from one another as the relatively weak hydrogen bonds between the bases are broken.

2 The unpaired nucleotides are exposed and each single strand now acts as a template for the formation of a new complementary strand. Free nucleotides move into place: C pairs with G and A pairs with T.

3 The free nucleotide bases form complementary pairs with the bases on the single DNA strands. **DNA polymerase** is the enzyme involved in linking the new nucleotides into place. Finally, the two new DNA molecules are rewound, each one forming a new double helix.

The two new DNA strands that are produced are absolutely identical to the original strands. Complementary base pairing between the template strand and the new strand ensures that an accurate copy of the original DNA is made every time replication occurs. DNA replication is said to be semi-conservative replication because no DNA molecule is ever completely new. Every double helix contains one 'original' and one 'new' strand.

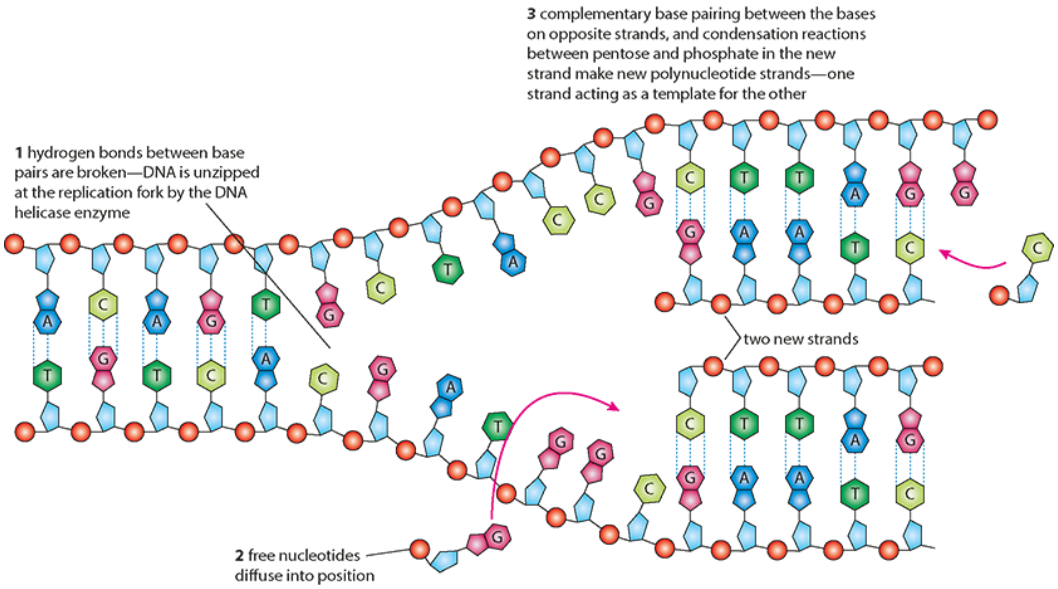


Figure 3.1.1: DNA replication.

NATURE OF SCIENCE

Obtaining evidence: Meselson and Stahl's experiment and semi-conservative replication of DNA

The research of Meselson and Stahl demonstrates the importance of making and testing a hypothesis in science. They investigated the two hypotheses about DNA replication that were current in the 1950s. The first hypothesis proposed that when DNA is replicated the original helix is conserved unchanged and the newly produced helix contains all new material. This conservative hypothesis was in contrast to the semi-conservative hypothesis, which proposed that one of the original DNA strands from a helix would always be found as one-half of the new double helix produced after replication. Meselson and Stahl designed their experiments using *Escherichia coli.* The bacteria were grown on a medium containing nitrogen ^{15}N, which is a heavy **isotope** of the normal ^{14}N. These isotopes were essential to Meselson and Stahl's experiments. After many generations the bacteria have incorporated ^{15}N into their cells, so that their DNA became 'labelled' with the heavy isotope and could be identified easily. The bacteria were then transferred to a new medium containing the lighter isotope ^{14}N, and allowed to grow for a period of time that corresponded to the length of a generation. Figure 3.1.2 shows how the labelled DNA would be distributed among the daughter molecules after one and two replications, according to the semi-conservative theory and the conservative theory. Meselson and Stahl's careful measurements of the amounts of ^{15}N in the daughter molecules after one replication showed that all the helices contained one strand of labelled DNA and one strand of normal DNA. Their results therefore supported the theory of semi-conservative replication.

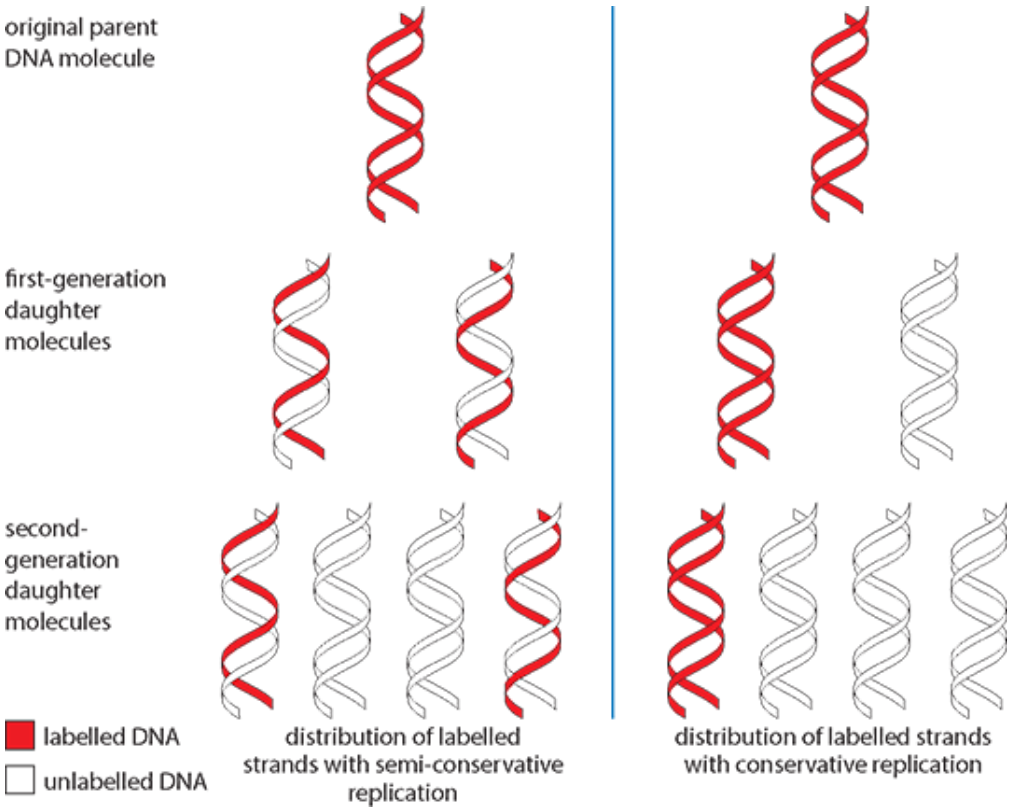


Figure 3.1.2: The distribution of labelled DNA in daughter molecules after replication, according to the semi-conservative theory and the conservative theory of replication.

3.1.2 DNA sequencing

DNA sequencing is a technique that analyses sequences of DNA bases to work out the sequence of individual genes, groups of genes or even entire chromosomes and genomes. Since the advances in technology made during the Human Genome Project at the turn of the century many new and automated methods of sequencing have been developed. Geneticists and forensic scientists use these techniques to analyse and compare sequences in DNA, some of which are common to different species and others that are repeated many times. The results are used in criminal investigations, for establishing family relationships and in medical diagnoses.

KEY POINTS

chromosome in eukaryotes, a structure consisting of a long thread of DNA and protein that carries the genetic information of the cell; in bacteria, the DNA molecule that contains the genetic information of the cell.

genome refers to the whole of the genetic information of an organism.

DNA profiling

At a crime scene, forensic scientists check for fingerprints because a person's fingerprint is unique and can be used to identify them. Forensic scientists also collect samples of hair, skin, blood and other body fluids left at a crime scene because they all contain a person's DNA and that too is a unique record of their presence.

Matching the DNA from a sample to a known individual is called **DNA profiling**. In forensic science, DNA profiles from crime scenes can be used to establish the possibility of guilt or to prove a suspect innocent (Figure 3.1.3). DNA profiling can also be used to determine paternity. For example, a woman might believe that a particular man is the father of her child. By comparing DNA samples from all three individuals – the woman, the man and the child – paternity can be established.

KEY POINT

DNA profiling is the process of producing a specific DNA pattern, called a profile, from a sample of DNA.

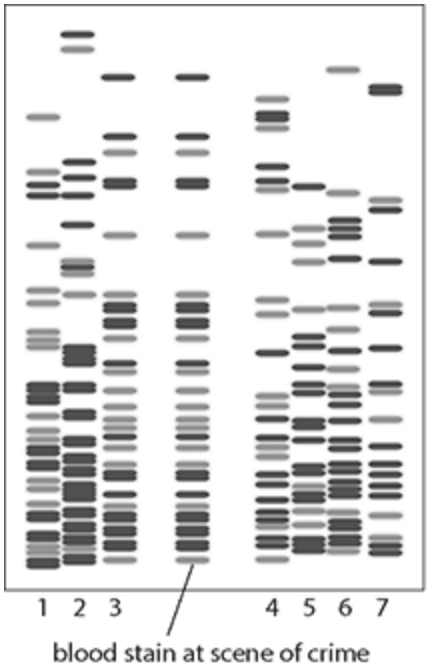


Figure 3.1.3: DNA profile of a blood stain found at the scene of a crime compared with profiles from seven suspects. Which suspect was at the scene of the crime? What is the evidence to support your answer?

The polymerase chain reaction

The polymerase chain reaction (PCR) is an automated method used to amplify (copy) segments of DNA. To study DNA in forensic or genetic analysis large amounts of a sample of DNA are needed. In most cases only small amounts are available, so PCR is a vital tool.

KEY POINT

polymerase chain reaction (PCR) a process in which small quantities of DNA are artificially amplified for research and diagnosis.

Sometimes, at a crime scene or when a body is found after a very long time, only a minute amount can be collected. The PCR can make millions of copies of tiny amounts of DNA so there is a sufficient amount to produce a profile or study a gene of interest. Technicians must take great care when handling the original sample so that it is not contaminated with their own or other DNA. Only the DNA region that is of interest will be amplified. A geneticist might be studying gene function or a forensic scientist could want to match crime scene DNA with that of a suspect. In medicine, the DNA of bacteria or viruses can be used in diagnosis. DNA amplified by PCR may be used for sequencing or to produce DNA profiles using gel electrophoresis (see the following stages).

The stages in the process are:

1 Denaturation – Heat the DNA sample to 95 °C so that the double strands of DNA separate into two single strands.

2 Annealing and extension – Cool to 68 °C and add the enzyme *Taq* polymerase, DNA primers and DNA

nucleotides that are needed to build duplicate copies of the original DNA using the two strands of DNA as templates.

3 Repeat the cycle of separation and synthesis of new DNA 30–40 times so that eventually more than a billion exact copies of the original DNA segment are produced (Figure 3.1.4).

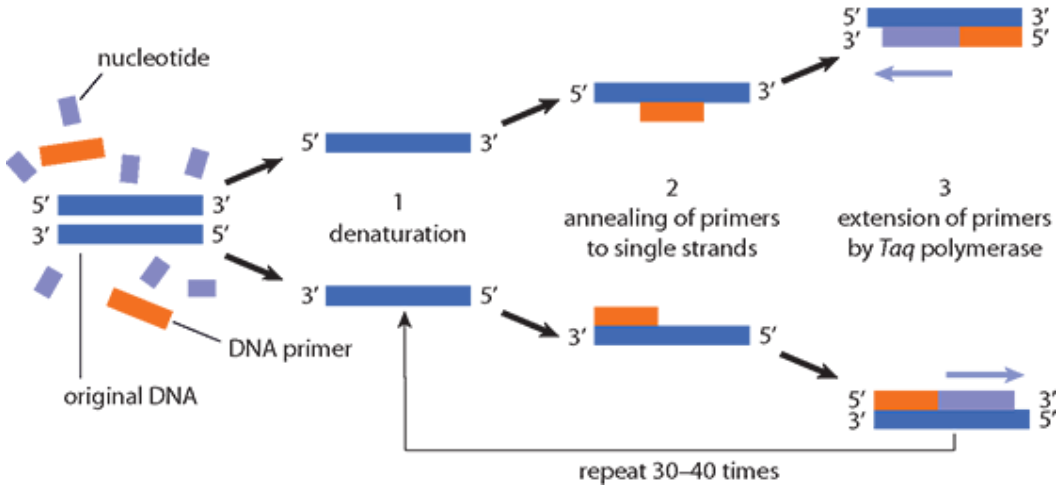


Figure 3.1.4: Stages in the polymerase chain reaction.

PCR is automatically controlled in a machine called a thermocycler that alters the temperature of the reaction every few minutes, firstly to cause DNA separation and then for synthesis of new strands.

Taq polymerase

The PCR needs the enzyme DNA polymerase to build new strands of DNA from the existing template strands, just as a cell does when it copies its DNA. The DNA polymerase used in PCR is *Taq* polymerase. It has been obtained from the heat-tolerant bacterium *Thermus aquaticus*, which is found in thermal vents and in hot springs. *Taq* polymerase is unaffected by high

temperature and is most active around 70 °C, a temperature at which a human DNA polymerase would not work.

KEY POINT

Taq polymerase a heat-stable DNA polymerase named after the microorganism *Thermus aquaticus* used to amplify DNA in the polymerase chain reaction.

PCR primers

Primers are short sequences of nucleotides that provide a starting point for DNA synthesis by *Taq* polymerase. The technician using PCR will add primers to select the region of DNA that is be copied, or amplified. PCR primers are pieces of single-stranded DNA, usually around 20 nucleotides in length.

Two primers are used in each PCR reaction. The sequences bind by complementary base pairing to opposite strands of DNA at the ends of the region to be copied. The *Taq* polymerase will add nucleotides to the primers so that the region between them is copied (Figure 3.1.5).

PCR in diagnosis

PCR can be used in medical diagnosis to detect genetic sequences of bacteria and viruses and thus identify active infections. Using specific primers, PCR can be used to amplify known sequences that only exist in certain viruses or bacteria. If that sequence is not found and there is no infection, then no amplification will take place and no DNA will be produced. Pathogens that are difficult, or take a long time, to culture can be identified quickly. The detection of the presence of bacteria in clinical specimens such as spinal fluid, blood and urine can

enable doctors to make speedy diagnoses and give the correct treatments.

Genetic disorders are caused by mutations (changes in DNA) that can be just a few base sequences, or be changes in large sequences of DNA or, sometimes, whole chromosomes. PCR enables geneticists to study just a small segment of DNA at a specific region of chromosome. The sequences of bases in a gene can be amplified and disorders detected, diagnosed and monitored. In some cases, **gene therapy** (Section 3.3) is available to rectify these disorders, and PCR can be used to monitor the functioning of the relevant genes and gene segments.

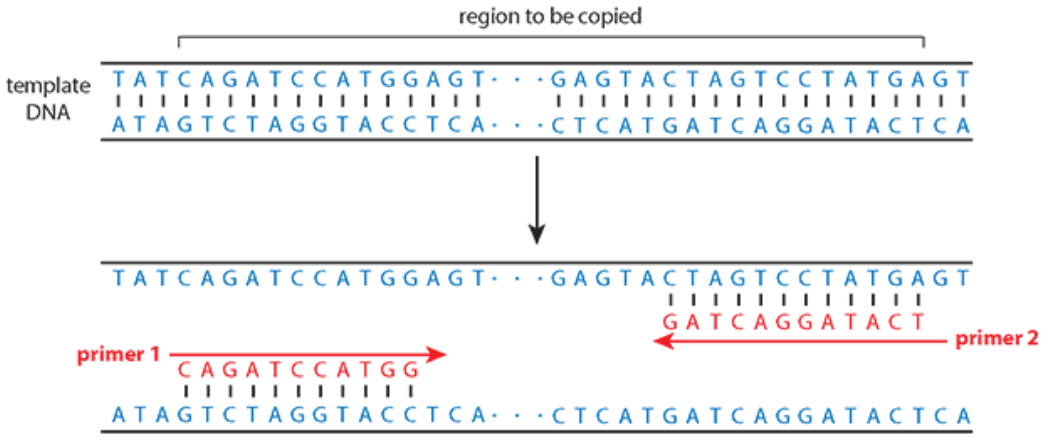


Figure 3.1.5: Positions of primers on single strands of DNA that are to be copied.

SCIENCE IN CONTEXT

Hunting for coronaviruses

There are two main types of coronavirus test: those that can detect the presence of the virus that is active in the body and those that detect a previous response to the virus by the immune system.

The PCR test is used in the first type of test, and looks for evidence that the COVID-19 virus, SARS-CoV-2, is present in a person's body by detecting the presence of its RNA in a swab sample from their nose or throat. The PCR test detects the genetic material from the virus by amplifying tiny amounts that may be present. PCR can only tell us if the virus is currently present in a person's body.

PCR tests involve several stages so errors are possible between sampling and analysis. False negatives (that is, a result that is negative when in fact the patient has the virus) do occur but estimates are that 80–85% of the results are correct.

Gel electrophoresis

Gel electrophoresis is a method used to separate fragments of DNA on the basis of size and the electric charge they carry. It can identify natural variations found in every individual's DNA.

Any DNA sample usually contains long molecules that are too large to be used for profiling so DNA profiling often examines repetitive sequences of so-called 'satellite' DNA that vary in their degree of repetitiveness from person to person. These are called variable number tandem repeats (VNTRs) and short tandem repeats (STRs). These regions have repeated sequences of DNA that are very similar in close relatives but so variable in unrelated people that non-relatives are extremely unlikely to have the same repeated sequences.

The DNA fragments are placed in a well in a plate of gel (a jelly-like material) and one well is reserved for a reference fragment known as a DNA ladder, this is a DNA molecule of a known length, used as a reference to estimate the size of the unknown DNA molecules in the sample.

KEY POINTS

DNA ladder DNA molecules of different lengths used in gel electrophoresis, used as a reference to estimate the size of unknown DNA molecules.

gel electrophoresis a technique which separates DNA fragments according to their size and charge.

An electric field is applied and because each DNA fragment has a small negative charge, it will move in the electric field, through the gel. The distance a fragment can move depends on its size; smaller fragments move most easily through the gel matrix and travel further, while larger fragments are left behind close to their starting point. After the fragments have been separated in the gel, they are stained and produce a unique pattern of bands called a DNA profile (Figures 3.1.3 and 3.1.6).

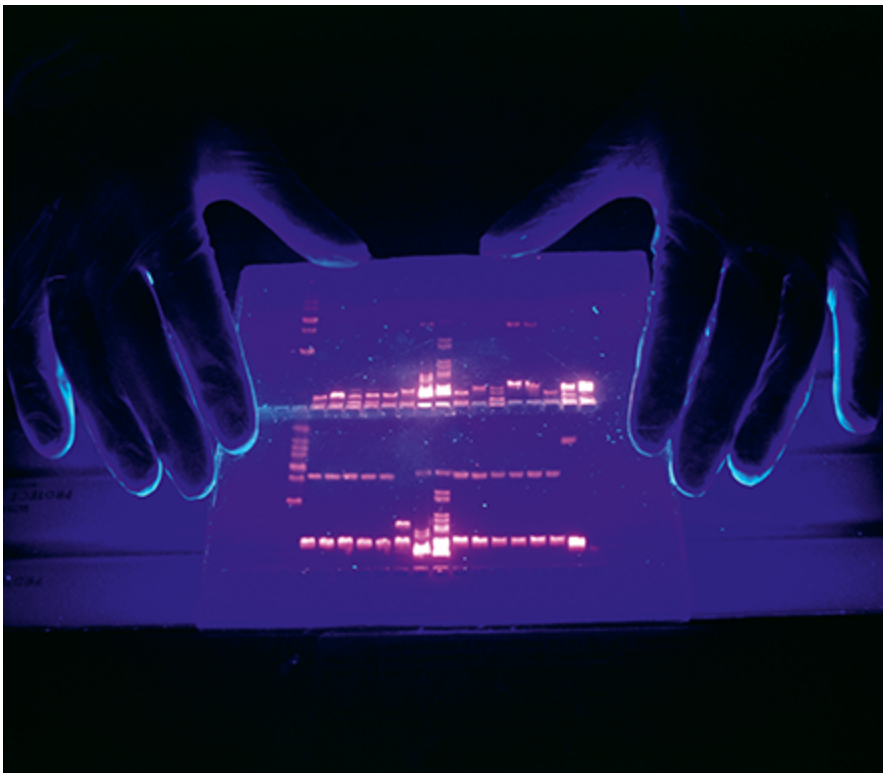


Figure 3.1.6: Scientist examining an agarose electrophoresis gel used to prepare a DNA profile. The sample of DNA is marked with a radioactive substance, so the DNA banding pattern appears pink under ultraviolet light. The pattern is preserved by applying radiographic film to the gel.

SCIENCE IN CONTEXT

Short tandem repeats and tracing our ancestors

Any one STR will be shared by between 5 and 20% of people who are not related. But in forensic science many STRs are examined at the same time; the more STR regions that are examined, the more accurate the test becomes. The pattern of repeats can identify an individual with a high degree of accuracy. In the world of genealogy (tracing family history), DNA profiling and STRs are also used as vital tools. Today, if you want to prove that you are descended from a certain line then you may be able to use genetics to prove it. Genealogists research the ancestry of families, looking for groups of people who share the same STRs and who can be identified as being related to each other over hundreds or even thousands of years.

THEORY OF KNOWLEDGE

DNA profile databases

In the USA, the Federal Bureau of Investigation (FBI) has a national database of DNA profiles from convicted criminals, suspects, missing persons and crime scenes. The data that are held may be used in current investigations and to solve unsolved crimes. There are many commercial laboratories that carry out DNA profiling analysis on behalf of the law

enforcement agencies. Many of them check 13 key STR sequences in DNA samples, which vary considerably between individuals. The FBI has recommended that these should be used because they provide odds of one in one thousand million that two people will have the same results.

CODIS is the acronym for the Combined DNA Index System, a computer software program that operates the national database of DNA profiles. Every US state has a statutory right to establish a DNA database that holds DNA profiles from offenders convicted of particular crimes. CODIS software enables laboratories to compare DNA profiles electronically, linking serial crimes to each other and identifying suspects from profiles of convicted offenders. CODIS has contributed to thousands of cases that have been solved by matching crime scene evidence to known convicted offenders.

To consider:

1. DNA profiles do not show individual base sequences but only identify repeated sequences. How much confidence should be placed on DNA evidence?
2. How secure is DNA profiling?
3. What are the implications for society if the authorities were to hold a DNA profile for every person?
4. What safeguards should be in place to protect the rights of individuals whose DNA profiles have been placed on a database but who have not been convicted of a crime?
5. Is it right to convict a person on DNA evidence alone?

TEST YOUR UNDERSTANDING

1 Outline what is meant by the term semi-conservative replication.

2 State the role of the enzyme helicase.

3 Give two examples of the use of DNA profiles.

3.1.3 The detailed process of DNA replication

DNA replication ensures that exact copies of existing molecules are produced before a cell divides. The process is said to be semi-conservative and each strand of an existing DNA molecule acts as a template for the production of a new strand (see Nature of Science, Obtaining evidence: Meselson and Stahl's experiment and semi-conservative replication of DNA, earlier in this section).

In eukaryotes, replication is controlled through interactions between proteins, including cyclins and CDKs (Section 6.5) and takes up to 24 hours to complete. Each of the original DNA strands acts as a template to build up a new strand (Figure 3.1.7). The DNA double helix is unwound to expose the two strands for replication by the enzyme DNA helicase, at a region known as a replication fork. The action of helicase creates single-stranded regions, which are less stable than the double-stranded molecule. To stabilise these single strands, **single-stranded binding proteins** (SSBs) are needed. SSBs protect the single-stranded DNA and allow other enzymes involved in replication to function effectively upon it.

Replication must occur in the 5′→3′ direction (and also in transcription and translation, described in Section 3.2), because the enzymes involved only work in a 5′→3′ direction (adding new nucleotides to the 3′ end of the newly forming DNA molecule). As the two strands are antiparallel, replication has to proceed in opposite directions on the two strands. However, the replication fork where the double helix unwinds moves along in one direction only. This means that on one of the strands

replication can proceed in a continuous way, following the replication fork along, but on the other strand the process has to happen in short sections, each moving away from the replication fork (Figure 3.1.8). The strand undergoing continuous synthesis is called the **leading strand**. The other strand, in which the new DNA is built up in short sections, is known as the **lagging strand**.

KEY POINTS

lagging strand is the new strand that is synthesised in short fragments in the opposite direction to the movement of the replication fork.

leading strand is the new strand that is synthesised continuously and follows the replication fork.

single-stranded binding protein is the protein which binds to single-stranded regions of DNA to protect them from digestion and remove secondary structure.

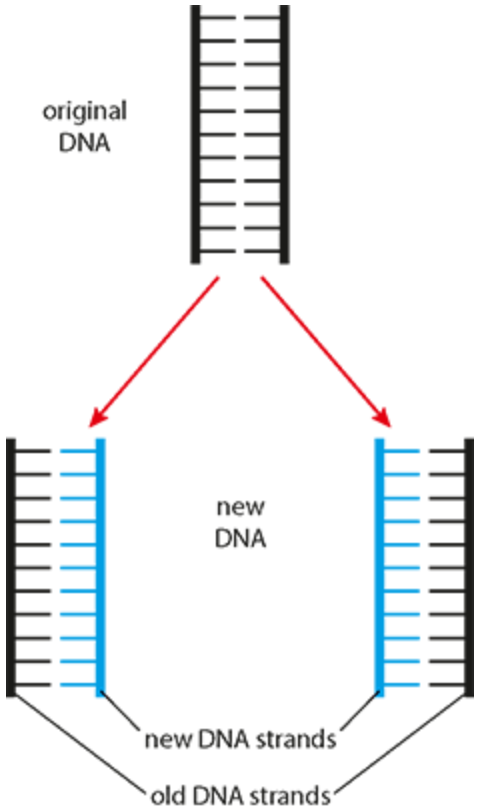


Figure 3.1.7: DNA replication is semi-conservative. As it is copied one original strand becomes paired with one new strand. One of the two strands in each new DNA molecule is conserved, hence 'semi-conservative'.

Copying the leading strand

Replication to produce the leading strand begins at a point on the molecule known as the 'origin of replication' site. First **RNA primase** adds a short length of RNA, attached by complementary base pairing, to the template DNA strand. This acts as a **primer**, allowing the enzyme **DNA polymerase III** to bind. DNA polymerase III adds free 'building units' called **deoxynucleoside triphosphates (dNTPs)** to the 3′ end of the primer and then to the forming strand of DNA. In this way the new molecule grows in a 5′→3′ direction, following the progress of helicase as it moves the replication fork along the DNA

double helix. The RNA primer is later removed by **DNA polymerase I**. In this way, a continuous new DNA strand is built up on the leading strand.

KEY POINTS

primer is short strand of nucleic acid that forms a starting point for DNA synthesis.

RNA primase is an enzyme that catalyses the synthesis of RNA primers as the starting point for DNA synthesis.

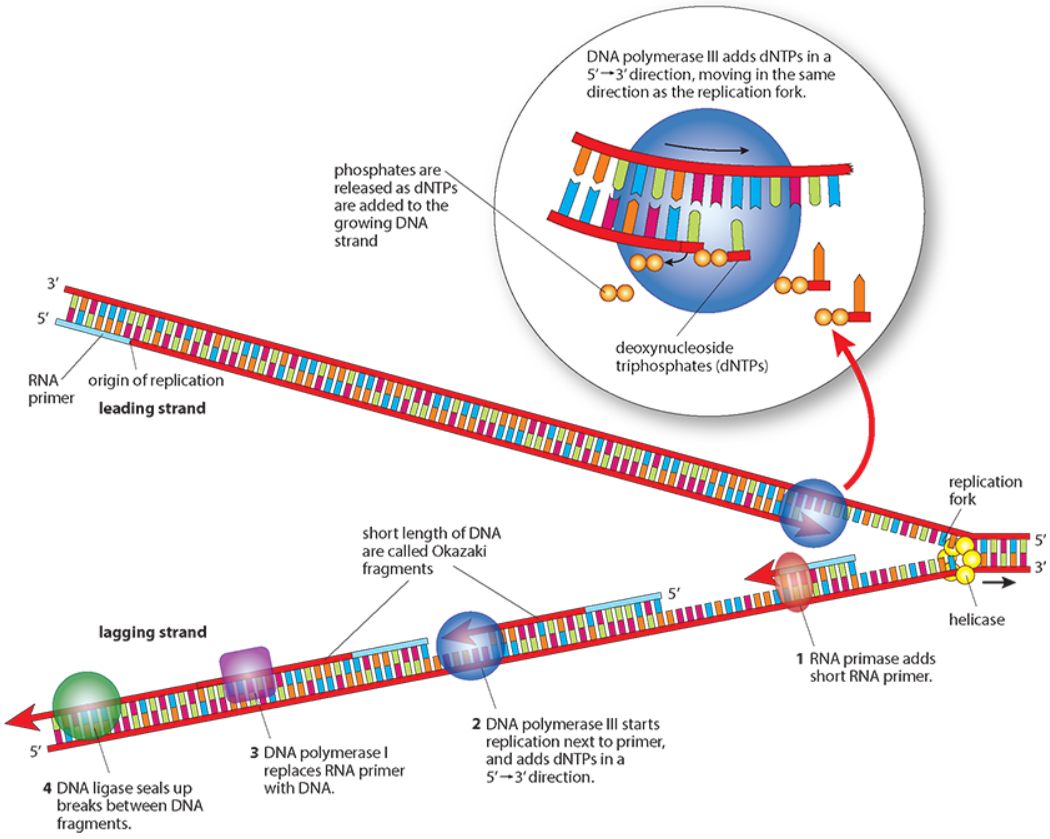


Figure 3.1.8: DNA replication showing the leading and lagging strands, and the direction of DNA synthesis.

KEY POINTS

DNA polymerase III extends the new DNA strand in a 5′ → 3′ direction from the RNA primer.

deoxynucleoside triphosphate (dNTP) is a building block for DNA: deoxyribose, three phosphate groups and one of the four bases.

DNA polymerase I removes the RNA nucleotides of the primers on the lagging strand and replaces them with DNA nucleotides.

EXTENSION

The dNTPs have two extra phosphate groups attached, and are said to be 'activated'. They pair up with their complementary bases on the exposed DNA strand and DNA polymerase III then link together the sugar and the innermost phosphate groups of adjacent nucleotides. The two extra phosphate groups are broken off and released.

Copying the lagging strand

Synthesis of the lagging strand is a little more complicated, as it has to occur in discontinuous sections, which are then joined together.

1 As for the leading strand, **DNA primase** first synthesises a short RNA primer, complementary to the exposed DNA. This happens close to the replication fork.

2 DNA polymerase III starts replication by attaching at the 3′ end of the RNA primer and adding dNTPs in a 5′→3′ direction. As it does so, it moves away from the replication fork on this strand.

3 DNA polymerase I now removes the RNA primer and replaces it with DNA using dNTPs. Short lengths of new DNA called **Okazaki fragments** are formed from each primer. The new fragment grows away from the replication fork until it reaches the next fragment.

4 Finally, **DNA ligase** seals up each break between the Okazaki fragments by making sugar–phosphate bonds so that a continuous strand of new DNA is created.

KEY POINTS

DNA primase a type of RNA polymerase catalyses the production of a short length of RNA called a primer which is base-paired to the parent DNA strand. The primer is removed when replication is complete and replaced by DNA.

DNA ligase joins adjacent Okazaki fragments by forming a covalent bond between adjacent nucleotides.

Okazaki fragments short fragments of a DNA strand formed on the lagging strand.

Proofreading new DNA

DNA polymerases 'proofread' their work as they build up new DNA strands. If the polymerase enzyme detects that an incorrect nucleotide has been added and does not pair up correctly the enzyme will remove and replace it.

EXAM TIP

There are several important enzymes to remember in the process of replication so it is helpful to keep a list of them and their jobs.

TEST YOUR UNDERSTANDING

4 Outline what is meant by antiparallel.

5 State the direction in which DNA replication occurs.

6 Outline the role of DNA primase.

7 Summarise the differences between forming the leading and lagging strands.

REFLECTION

Could you explain DNA profiling to someone who had never heard of it? Reflect on its importance to your understanding of ourselves.

Links

- How is the molecular structure of DNA linked to its function? (Chapter 1)
- Why must the genetic code carried by DNA be copied exactly? (Chapter 3.4)
- How is replication involved in cell division? (Chapter 6)

3.2 Protein synthesis

LEARNING OBJECTIVES

In this section you will:

- define transcription as the synthesis of RNA from a DNA template
- recognise that transcription is needed for the expression of genes
- learn that complementary base pairing between DNA and mRNA ensures that the polypeptides produced function properly
- define translation as the production of polypeptides from mRNA using tRNA
- recognise how complementary base pairing between codons and anticodons ensures accurate translation
- learn that ribosomes are the sites of translation; free ribosomes synthesise proteins for use within the cell, whereas bound ribosomes synthesis proteins for secretion or use in lysosomes

- understand the directional nature of transcription and translation
- recognise that transcription begins at a promoter region
- understand that in prokaryotes, translation occurs immediately after transcription but eukaryotes modify

- mRNA by removing introns to form mature mRNA composed of exons
- recognise that exons can be spliced in different ways to produce different proteins from a single gene
- understand how nucleosomes regulate transcription in eukaryotes
- understand that a large portion of the eukaryotic genome consists of non-coding sequences
- learn that non-coding DNA persists for many generations and has important functions
- recognise that polysomes allow many polypeptides to be made at the same time
- understand that translation does not always result in functional protein and that polypeptides are modified before they can function
- learn that amino acids are recycled in the cell by proteasomes.

GUIDING QUESTIONS

- How does complementary base pairing contribute to the resilience of the genetic code?
- How are enzymes involved in protein production?

3.2.1 Transcription

The main role of DNA is to direct the activities of the cell. It does this by controlling the proteins that the cell produces. Enzymes, hormones and many other important biochemical molecules are the proteins that control what the cell becomes, what it synthesises and how it functions. Protein synthesis can be divided into two sets of reactions: the first is transcription and the second is translation. In eukaryotes, transcription occurs in the nucleus and translation in the cytoplasm.

The sections of DNA that code for particular proteins are known as genes. Genes contain specific sequences of bases in sets of three, called triplets. Some triplets control where transcription begins and ends.

KEY POINTS

gene is a particular section of a DNA strand that codes for a specific polypeptide; a heritable factor that controls a specific characteristic.

transcription means copying a sequence of DNA bases to mRNA.

translation decoding mRNA at a ribosome to produce a polypeptide.

KEY POINTS

triplet a sequence of three bases that code for an amino acid.

messenger RNA (mRNA) is a single-stranded transcript of one strand of DNA, which carries a sequence of codons for

the production of protein.

Copying the DNA message to RNA

The first stage in producing a protein is the production of messenger RNA from a segment of DNA so that the genes that code for the required polypeptides can be moved to the cytoplasm. After this the message is translated and the necessary amino acids are used to build polypeptide chains.

The first stage in the synthesis of a protein is the production of an intermediate molecule that carries the coded message of DNA from the nucleus into the cytoplasm where the protein can be produced. This intermediate molecule is called messenger RNA (mRNA). RNA (or ribonucleic acid) has similarities and differences with DNA and these are shown in Table 2.5.1.

The building blocks for RNA are the RNA nucleotides that are found in the nucleus. Complementary base pairing of RNA to DNA occurs in exactly the same way as in the replication process but this time uracil (U) pairs with adenine since the base thymine (T) is not found in RNA. Transcription results in the copying of one section of the DNA molecule, not its entire length. Figure 3.2.1 describes the process.

1 DNA is unzipped by the enzyme RNA polymerase and the two strands uncoil and separate.

2 Free nucleotides move into place along one of the two strands.

3 The same enzyme, RNA polymerase, assembles the free nucleotides in the correct places using complementary base pairing. As the RNA nucleotides are linked together, a single strand of mRNA is formed. This molecule is much

shorter than the DNA molecule because it is a copy of just one section, a gene. The mRNA separates from the DNA and the DNA double helix is zipped up again by RNA polymerase.

Once an mRNA molecule has been transcribed, it moves via the pores in the nuclear envelope to the cytoplasm where the process of translation can take place. In prokaryotes, translation occurs immediately after transcription because there is no nuclear envelope.

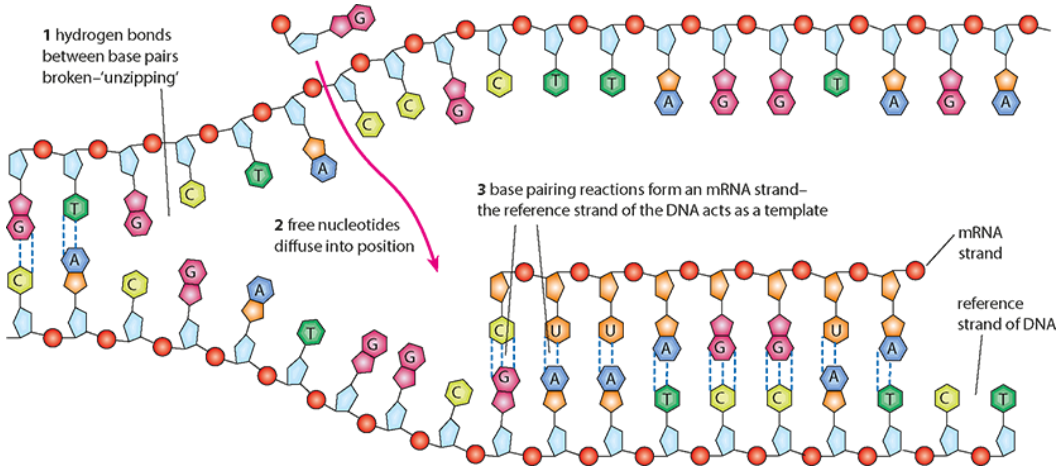


Figure 3.2.1: Transcription.

Initiating transcription

Before transcription begins, RNA polymerase must attach to a promoter region of DNA. This process is different in prokaryotes and eukaryotes. In prokaryotes the RNA polymerase binds directly to a promoter region, close to the region that will be transcribed. Transcription is controlled by another section of DNA called the regulator gene, which can produce a repressor molecule that binds to the operator region and prevents transcription (Figure 3.2.2).

Eukaryotes have several transcription factors that are needed to bind the RNA polymerase to the promoter region. Some of these bind to an enhancer region, away from the promoter. The transcription factors (labelled the transcription-initiation complex in Figure 3.2.3) bring the enhancer region close to the promoter and RNA polymerase can then bind and begin transcription (Figure 3.2.3).

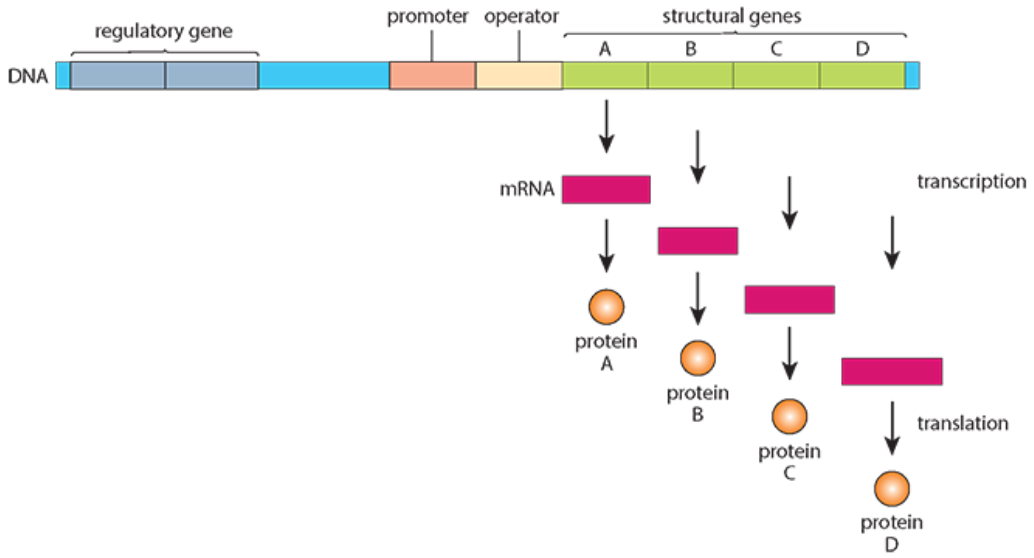


Figure 3.2.2: Controlling transcription in a prokaryote.

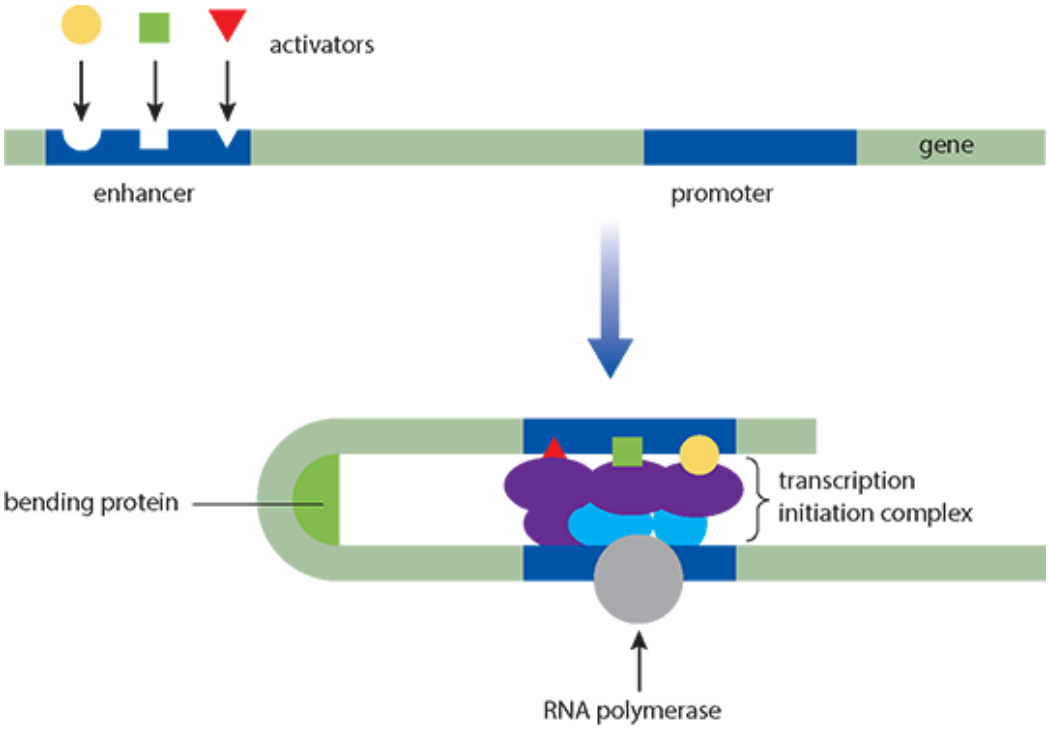


Figure 3.2.3: In eukaryotes transcription is controlled by proteins that bind to specific sequences of DNA. Bending proteins are discussed in the Higher Level Section 3.2.3.

Both prokaryotes and eukaryotes have regions of DNA that do not code for protein. In prokaryotes these include promoter regions that serve regulatory functions.

Regulation of transcription by nucleosomes

DNA in eukaryotes is incorporated into nucleosomes so that the genetic material can be stored in a compact form (Section 1.7.4). Nucleosomes are important because they can either inhibit or allow transcription by controlling whether the necessary molecules can bind to DNA.

KEY POINT

nucleosome a part of a eukaryotic chromosome important in regulating transcription, made up of DNA wrapped around histone molecules and held in place by another histone protein.

In order to transcribe genes, activators and enzymes involved in transcription must be able to gain access to DNA. In all eukaryotic species, the regions of DNA that contain the promoters and regulators, which are the binding sites for RNA polymerase and the starting point for transcription, have fewer nucleosomes than other areas, allowing greater access for binding proteins. Conversely, the regions that are transcribed have a higher density of nucleosomes. This suggests that nucleosomes have an important role in determining which genes are transcribed. This in turn can influence other factors such as cell variation and development.

DNA does not need to be completely released from a nucleosome to be transcribed and, although nucleosomes are very stable protein–DNA complexes, they are not static. They can undergo different structural rearrangements including so-called ‘nucleosome sliding’ and DNA site exposure: if a nucleosome is ‘unwrapped’ there is a significant period of time during which DNA is accessible. They can also be modified by methylation (Section 3.4.2). The new transcript of mRNA before modification is called pre-mRNA and it becomes known as mature mRNA after removal of introns’

3.2.2 Translation

Translation is the process by which the information carried by mRNA is decoded and used to build the sequence of amino acids that eventually forms a protein molecule. During translation, amino acids are joined together in the order dictated by the sequence of codons on the mRNA to form a polypeptide. This polypeptide eventually becomes the protein coded for by the original gene.

Complementary base pairing ensures that the sequence of bases along the mRNA molecule corresponds to the sequence on the original DNA molecule. Each sequence of three mRNA bases is called a **codon** and codes for one specific amino acid, so the order of these codons determines how amino acids will be assembled into polypeptide chains in the cytoplasm. The completed polypeptide chains will be folded to make functioning proteins. Translation is carried out in the cytoplasm by structures called ribosomes and molecules of another type of RNA known as transfer RNA or tRNA.

The mRNA codons that code for each amino acid are shown in Table 3.2.1. From the table you should be able to deduce which amino acid corresponds to any codon.

The genetic code is said to be a **degenerate code** because there are many codons that specify the same amino acid. It is also said to be **universal** because all living things use the same triplet code to specify the same amino acids.

Mutations are changes in sequence of bases that affect protein structure, you can read more about mutations in Section 3.3.

Transfer RNA

The process of translation requires a type of nucleic acid known as transfer RNA (tRNA). tRNA is made of a single strand of nucleotides that is folded and held in place by base pairing and hydrogen bonds (Figure 3.2.4). There are many different tRNA molecules but they all have a characteristic ‘clover leaf’ appearance with some small differences between them.

KEY POINTS

activating enzyme is an enzyme that catalyses the attachment of an amino acid to the appropriate tRNA.

anticodon is a triplet of bases in tRNA that pair with a complementary triplet (codon) in mRNA.

		Second base					

		U		C		A		G			
First base	U	UUU	phenylalanine	UCU	serine	UAU	tyrosine	UGU	cysteine	U	Third base
		UUC		UCC		UAC		UGC		C	
		UUA	leucine	UCA		UAA	‘stop’	UGA	‘stop’	A	
		UUG		UCG		UAG		UGG	tryptophan	G	
	C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U	
		CUC		CCC		CAC		CGC		C	
		CUA		CCA		CAA	glutamine	CGA		A	
		CUG		CCG		CAG		CGG		G	
	A	AUU	isoleucine	ACU	threonine	AAU	asparagine	AGU	serine	U	
		AUC		ACC		AAC		AGC		C	
		AUA		ACA		AAA	lysine	AGA	arginine	A	
		AUG	methionine or ‘start’	ACG		AAG		AGG		G	
	G	GUU	valine	GCU	alanine	GAU	aspartic acid	GGU	glycine	U	
		GUC		GCC		GAC		GGC		C	
		GUA		GCA		GAA	glutamic acid	GGA		A	
		GUG		GCG		GAG		GGG		G	

Table 3.2.1: Amino acids and their associated mRNA codons.

At one position on the molecule is a triplet of bases called the anticodon, which pairs by complementary base pairing with a codon on the mRNA strand. At the 3′ end of the tRNA molecule is a base sequence CCA, which is the attachment site for an amino acid.

An amino acid is attached to the specific tRNA molecule that has its corresponding anticodon, by an activating enzyme. As there are 20 different amino acids, there are also 20 different activating enzymes in the cytoplasm.

KEY POINT

transfer RNA (tRNA) short lengths of RNA that carry specific amino acids to ribosomes during protein synthesis.

Ribosomes

Ribosomes are the site of protein synthesis. Some ribosomes occur free in the cytoplasm and synthesise proteins that will be used within the cell. Others are bound to the endoplasmic reticulum, forming rough endoplasmic reticulum (RER) (Section 5.2), and synthesise proteins that will be secreted from the cell or used within lysosomes.

Ribosomes are composed of two subunits, one large and one small. The subunits are built of protein and ribosomal RNA (rRNA). On the surface of the ribosome are three tRNA-binding sites (the entry site A, the P site and the exit site E), and one mRNA-binding site (Figure 3.2.5). Two tRNA molecules carrying amino acids can bind to a ribosome at one time. Polypeptide chains are built up in the groove between the two subunits.

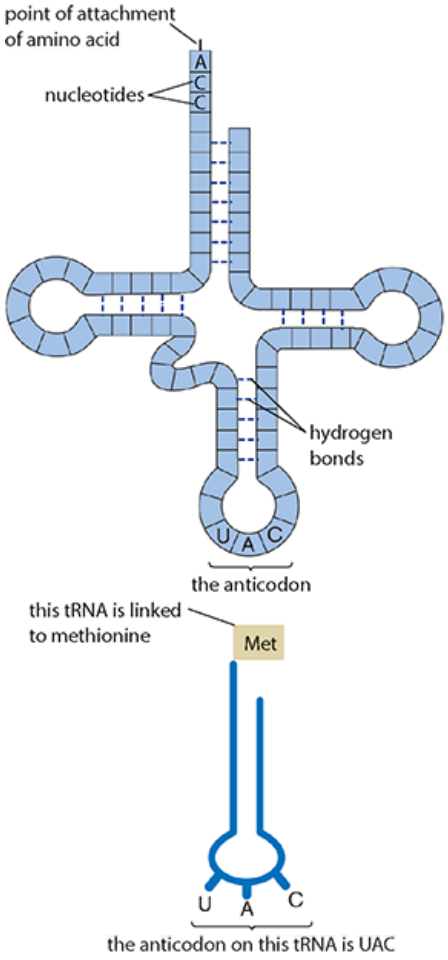


Figure 3.2.4: Transfer RNA (tRNA) has a 'clover leaf' shape.

Building a polypeptide

Ribosomes have binding sites for both the mRNA molecule and tRNA molecules. The ribosome binds to the mRNA and then draws in specific tRNA molecules with

anticodons that match the mRNA codons. Only two tRNA molecules bind to the ribosome at once. Each one carries with it the amino acid specified by its anticodon. The anticodon of the tRNA binds to the complementary codon of the mRNA molecule with hydrogen bonds.

When two tRNA molecules are in place on the ribosome, a peptide bond forms between the two amino acids they carry to form a dipeptide. A peptide bond links the amino group of one amino acid to the carboxyl group of the next.

Once a dipeptide has been formed, the first tRNA molecule detaches from both the amino acid and the ribosome. The ribosome moves along the mRNA one triplet to the next codon.

These processes, shown in Figure 3.2.5, are repeated over and over again until the complete polypeptide is formed. The final codon that is reached is a 'stop' codon, which does not code for an amino acid but tells the ribosome to detach from the mRNA. As it does so, the polypeptide floats freely in the cytoplasm or into the RER.

EXTENSION

A single chromosome contains DNA that codes for many proteins. Most genes are about 1000 nucleotides long, a few are longer and a very small number are less than 100 nucleotides. The size of a gene corresponds to the size of the polypeptide for which it codes.

TEST YOUR UNDERSTANDING

8 Define transcription.

9 Where does transcription take place in a prokaryotic cell and a eukaryotic cell?

10 Outline the structure of a ribosome.

11 Where in the cell are proteins for secretion or use in lysosomes produced?

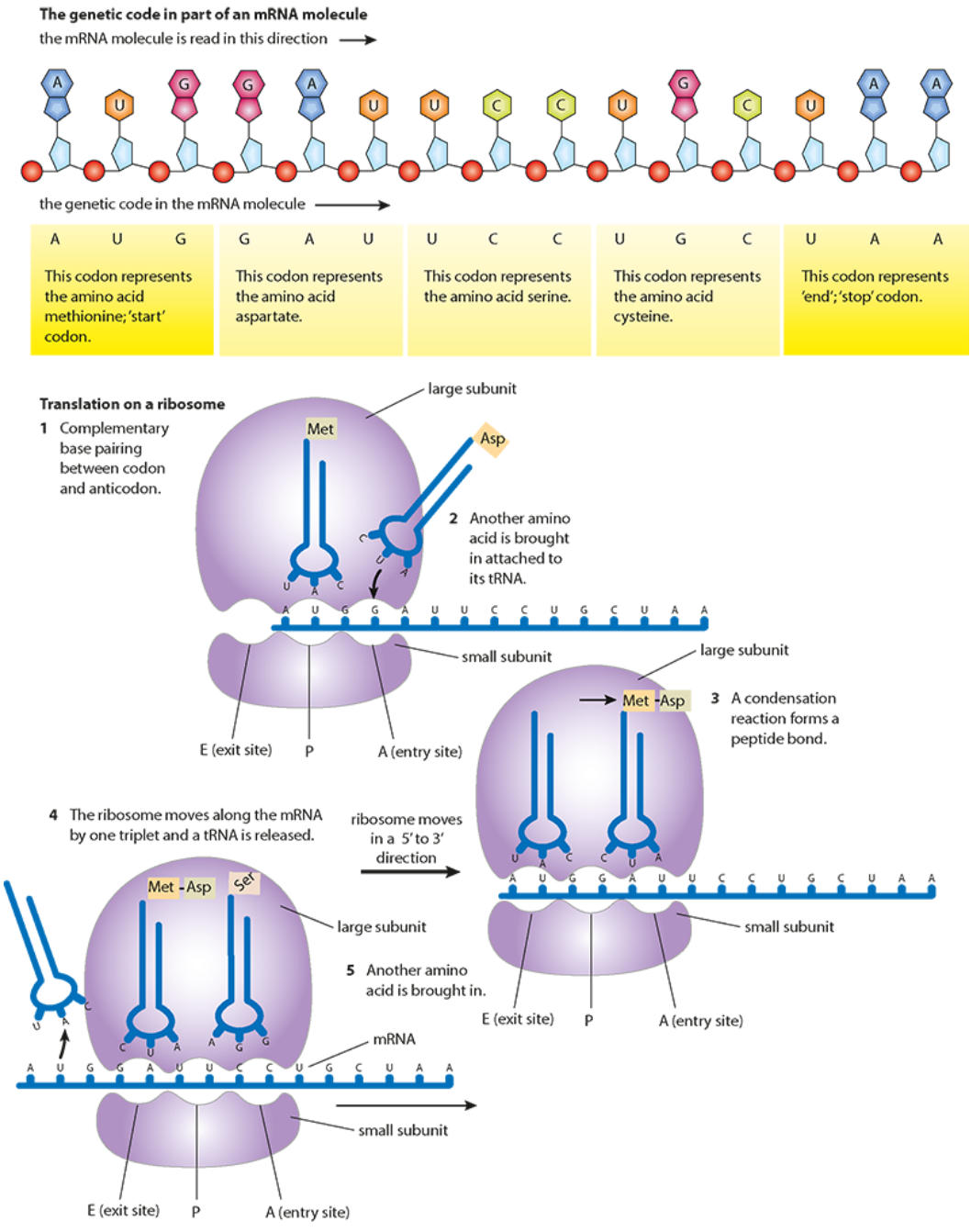


Figure 3.2.5: The stages of translation.

3.2.3 Non-coding regions of DNA

Repeated sequences

DNA molecules are very long but every strand has regions that do not code for proteins. For many years these regions were poorly understood but many of them have now been found to have important functions in the regulation of gene expression and other cell activities. Eukaryotic genomes contain mostly non-coding DNA; in fact, nearly 99% of the human genome is non-coding. About 7% of this DNA is thought to have regulatory functions but the exact proportion is not fully known. You may hear this type of DNA being referred to as 'junk DNA', but it is not a term that is now accepted by the scientific community because at least some of this DNA has a function.

When the highly repetitive sequences of non-coding DNA, called variable number tandem repeats (VNTRs), were first discovered they appeared to have no function but as genomes were mapped and compared it was found that several long repeated sequences in humans, mouse and rat DNA were common to all three species. These repeated non-coding sequences regulate and control the activity of genes and possibly embryo development. Studies of the genomes of many species have shown that non-coding DNA is conserved over hundreds of millions of years, suggesting that these regions have been conserved through evolution. It is likely that they give advantages in preserving certain vital genetic characteristics.

KEY POINT

tandem repeat is a repeated sequence of DNA base pairs where multiple repeats lie side by side on a chromosome.

Tandem repeats are generally associated with non-coding DNA. The number of times the DNA sequence is repeated is variable. Variable tandem repeats are used in DNA profiling because they are very similar in close relatives but very different in unrelated individuals.

Some regions of non-coding DNA act as ‘switches’ that determine when and where genes are expressed by controlling where and when transcription can begin. Others may be essential for chromosome structure and play a role in cell division. Introns are also transcribed but not translated into proteins.

Genes for transfer RNA

In humans, the genes that code for tRNA molecules are found on all chromosomes except 22 and Y. These genes do not code for protein but code either for cytoplasmic tRNA or for mitochondrial tRNA. The number of genes that code for tRNA is related to evolutionary history, so that organisms in the Archaea and Eubacteria domains have fewer than those in the domain Eukarya. This seems to be due to the duplication of the genes over time.

Promoter regions

Promoter regions are DNA sequences that define where transcription of a gene by RNA polymerase begins. They are usually found at the 5′ end of the area where transcription begins.

RNA polymerase requires the presence of a class of proteins known as ‘general transcription factors’ before transcription can begin. Interactions between the transcription factors, RNA polymerase and the promoter region allow the polymerase to

move along the gene so that transcription can occur. Many different transcription factors have been found and each one is able to recognise and bind to a specific nucleotide sequence in DNA. A specific combination of transcription factors is necessary to activate a particular gene.

Other DNA sequences, known as enhancer sequences, are also important and provide a place for regulatory proteins, called activators, to bind.

The role of binding proteins in gene expression is shown in Figure 3.2.3 and discussed in Section 3.4. The proteins bind to the enhancer, which may be some distance from the gene. 'Bending proteins' may then assist in bending the DNA so that the enhancer region is brought close to the promoter. Activators, transcription factors and other proteins attach, so that an 'initiation complex' is formed and transcription can begin. Some activator proteins affect the transcription of multiple genes.

Transcription factors are regulated by signals produced from other molecules such as hormones that are able to activate transcription factors and thus control transcription. Many other molecules in the environment of a cell or an organism can also have an impact on gene expression and protein production.

Telomeres

Telomeres are regions of repeated nucleotide sequences of non-coding DNA at each end of every eukaryotic chromosome. They protect chromosomes from damage and from fusing with adjacent chromosomes. They have been likened to the protection that a plastic tip on the end of a shoelace gives to protect a lace from fraying.

When chromosomes are replicated during cell division, the enzymes cannot copy the sequences at the end of the chromosomes. Telomeres act to protect important genes from being lost, by capping the end sequences. Cells contain enzymes called telomerases, which can replenish the repeated sequences after cell division in stem cells, but telomeres tend to shorten over time as cells replicate. Telomere shortening can block cell division and, by limiting the number of cell divisions, they protect cells from losing genetic information. The average cell will divide between 50 and 70 times before chromosomes are shortened too much and the cell dies.

EXTENSION

You may like to read about the work of Leonard Hayflick (b. 1928), a United States scientist who worked on cloning cells for vaccine production. He discovered that the number of times a cell can divide is not infinite, it is limited to a number called the Hayflick Limit.

Polysomes

Translation occurs at many places along an mRNA molecule at the same time. The electron micrograph in Figure 3.2.6 shows transcription and translation occurring simultaneously in a bacterium. A polysome is a group of ribosomes along one mRNA strand (Figure 3.2.7). Part of the bacterial chromosome can be seen as the fine pink line running horizontally along the top of the micrograph and two growing polypeptide chains are shown forming below it. DNA is being transcribed by RNA polymerase and the newly formed mRNA is being immediately translated by the ribosomes. In eukaryotes, the two processes

occur in the nucleus and cytoplasm, respectively, and so are separated not only in time but also in location.

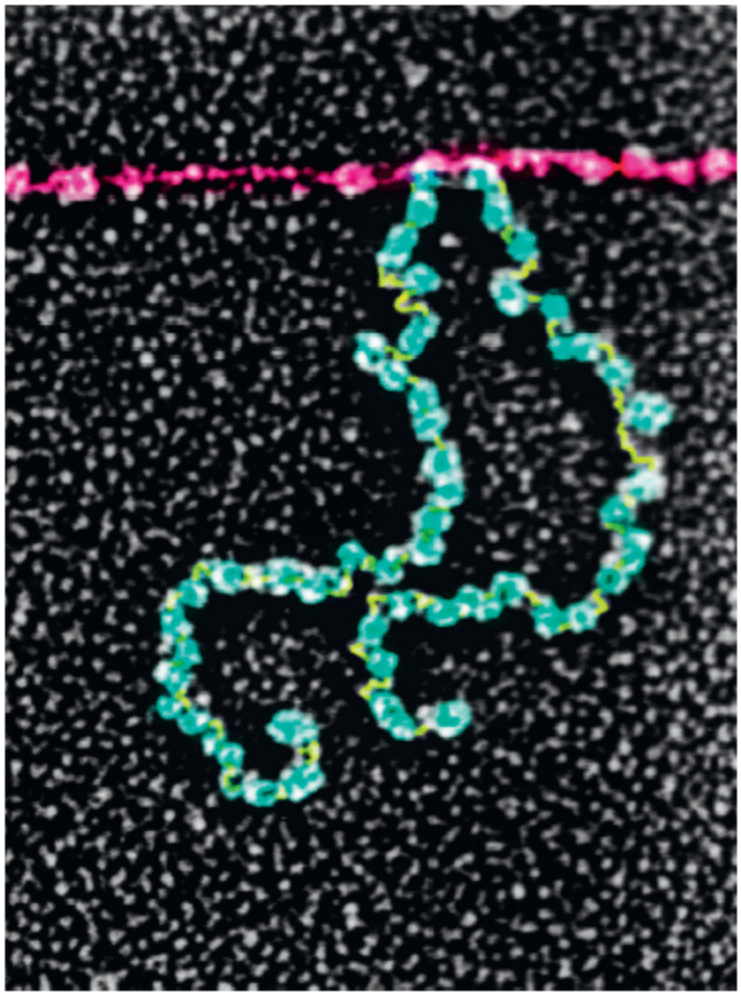


Figure 3.2.6: Electron micrograph of polysomes in a bacterium (× 150 000).

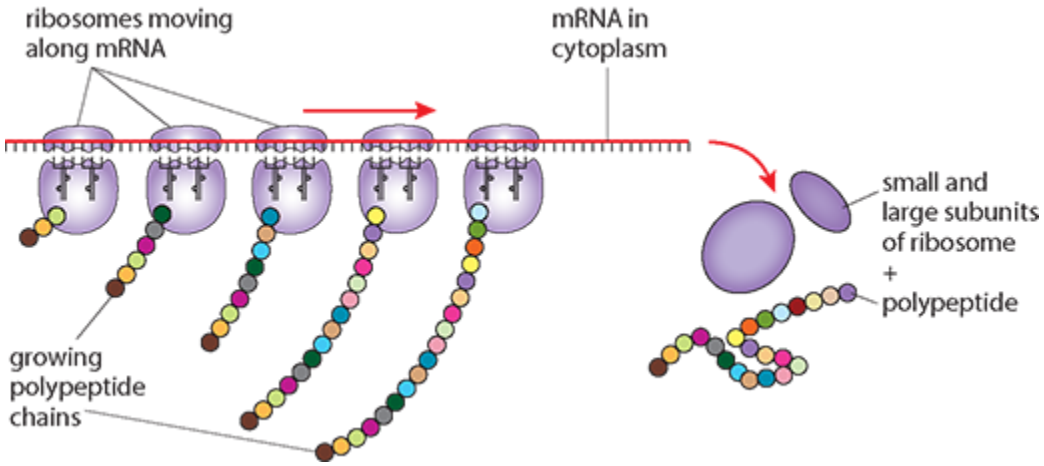


Figure 3.2.7: Diagram of a polysome in a eukaryotic cell. Polysomes appear like beads moving along a string of mRNA.

3.2.4 Post-transcriptional modification

1 Introns and exons

In eukaryotes many genes contain sequences of DNA that are transcribed but not translated. These sequences, known as introns, appear in mRNA but are removed before it is translated. Introns (or intragenic regions) are sequences of nucleotides *within* genes. After transcription of a gene, the introns are removed in a process known as post-transcriptional modification. Introns are removed in the nucleus before the mRNA moves to the cytoplasm for translation. Once introns have gone, the remaining sequences of bases, known as exons, are spliced together to form mature mRNA that will then be translated (Figure 3.2.8). Mature mRNA leaves the nucleus via the nuclear pores and moves to the cytoplasm.

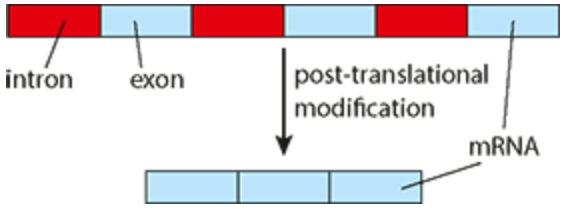


Figure 3.2.8: Introns and exons in mRNA.

Introns occur in many genes in all eukaryotic organisms and the number of introns per gene varies considerably between species. For example, introns are common in humans and mice, where genes almost always contain introns but are rare in other eukaryotes such as the yeast (*Saccharomyces cerevisiae*).

The human genome has been found to have about eight introns per gene but other organisms such as fungi may have fewer than 20 in their entire genome.

It also seems that there are more introns in species with smaller populations, and evolutionary and biological factors are thought to influence this.

Exons may be spliced together in different ways so that a number of different, but similar, protein sequences can be produced from a single gene. This alternative splicing means that a single gene can code for more than one protein. Different exons may be included or excluded from the final mRNA produced and as a result, the proteins produced from alternatively spliced mRNAs will contain different amino acid sequences and can have different biological functions. Mature mRNAs containing various combinations of exons from one original precursor mRNA increases the diversity of the proteins that can be produced. One example of this is in the production of immunoglobulins by B cells. A cell can splice together different exons and produce different immunoglobulins (antibodies) in response to different antigens that are present in the body (Chapter 10).

Splicing is controlled by molecules that respond to signals from both inside and outside the cell. Alternative splicing allows the human genome to synthesise many more proteins from its 20 000 genes than it could if there was no splicing.

2 5’ Caps and 3’ polytails

Two other modifications made to mRNA before translation are:

1 Adding a 5’ cap – a 5’ cap (a modified G nucleotide) is attached to the 5’ end of the mRNA

2 Adding a 3’ poly-A tail – poly-A tail is attached to the 3’ end of the mRNA and consists of a long string of A nucleotides

KEY POINTS

polysome an arrangement of many ribosomes along a molecule of mRNA so that multiple copies of the same polypeptide are produced at the same time.

alternative splicing including different exons in processed mRNA so that a cell can produce different proteins from the same gene.

exons sequences of bases in mRNA that are spliced together and translated after introns have been removed.

intron sequences of bases in mRNA that are removed after transcription.

Both these modifications help new mRNA strands leave the nucleus. They also help to protect the mRNA from damage. In the cytoplasm, the addition of the 5' cap and poly-A tail, help the ribosomes attach to the 5' end of the mRNA.

EXAM TIP

Remember: introns intervene in genes, but only exons are expressed.

3.2.5 Post-translational modification – producing functional proteins

Translation produces polypeptides which are the starting point for forming the working proteins that we need. But proteins are large, complex molecules, usually made up of hundreds of amino acid subunits linked in polypeptides. But it is the folding and linking of polypeptide chains into secondary, tertiary and, in some cases, quaternary structures that leads to the formation of functional proteins (Section 1.6). Folding takes place in the cytoplasm.

Another modification made to polypeptides and proteins after translation is the addition of a prosthetic group. Prosthetic groups are not polypeptides but they bind to different proteins or parts of them to enable them to function. An example is the respiratory pigment hemoglobin, which contains four polypeptide chains, each one containing a prosthetic heme group (Section 1.6).

Prosthetic groups

Many proteins contain **prosthetic groups** and those that do are called **conjugated proteins**. Prosthetic groups are non-protein groups that are able to bind to different proteins or parts of them. We can see two examples of prosthetic groups in the respiratory pigments myoglobin (Figure 1.6.4) and hemoglobin which both contain a prosthetic heme group. Hemoglobin consists of four polypeptide chains, each one containing a heme group. The heme group (Figure 1.6.1) consists of a central Fe (iron) atom and a porphyrin ring. The prosthetic heme group is vital to the structure of hemoglobin because the shape of the whole protein is changed as oxygen binds to it. The iron group not only allows

oxygen to bind but also holds the compact structure with four subunits in place and allows for progressively easier oxygenation as more oxygen molecules bind to the protein.

Protein modification and processing

Almost all proteins are chemically altered after they have been made. Modifications change the activity, life span and location of the protein. There are two main types of alteration that take place:

1. Chemical modification involves making additions to the side groups of amino acids or to the ends of the protein.
2. Processing involves removing peptide segments before the active protein is formed.

The most common chemical modifications are the addition of an acetyl group ($-CH_3CO$) to the amino group ($-NH_3$) at the end of a polypeptide. An estimated 80% of all proteins are modified in this way. Phosphorylation is another common modification and about 30% of proteins are modified by the addition of phosphate. The addition of phosphate regulates the activity of certain enzymes. Other modifications, such as the addition of a carbohydrate, can stabilise a protein and make it fully functional.

Processing involves the removal of segments of a protein. Active peptide hormones are produced in this way. One example is the polypeptide preproinsulin which is the inactive precursor to insulin. Preproinsulin is cut twice (Figure 3.2.9) so that the final active hormone insulin consists of two separate chains linked by disulfide bridges.

Preproinsulin is first converted to proinsulin by peptidase enzymes that remove a peptide from the $-NH_3$ end of the

molecule. Next, proinsulin is converted to active insulin by the removal of a second section of peptide at the C- end. Disulfide bridges link the remaining section of the molecule so that it becomes active insulin.

EXTENSION

The proteins present in human cells are far greater than the number that are coded for by our genes and this is due in part to modifications made to the proteins after they are formed. You can find out more about the proteins that are modified using the SWISS-PROT protein sequence database. Recent studies have identified more than 8000 proteins that are phosphorylated, more than 3000 that have acetyl groups added to them and around 5000 that have carbohydrates added to them.

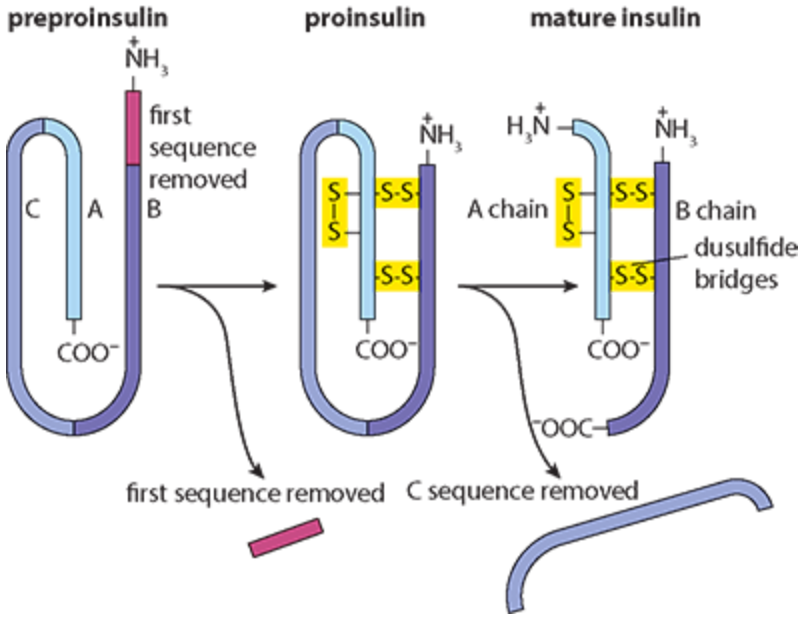


Figure 3.2.9: The formation of active insulin.

3.2.6 Protein transport molecules

Some proteins are able to operate as transport molecules and signalling systems inside cells. The bonds that hold these proteins together can, in some cases, be broken and remade to enable them to do so. For example, when molecular signals are received from outside a cell by receptors on its membrane surface they can be processed and transferred to the nucleus by the modification of transport proteins.

The enzyme protein kinase is important in this process because it can phosphorylate a transport protein. Phosphorylation changes the function of the protein by changing its activity, its location or the way it interacts with other proteins. A protein can be phosphorylated at the cell surface by protein kinase and dephosphorylated later in the cytoplasm. These changes in the protein's structure activate and then deactivate it while it acts as a signalling mechanism because phosphorylated proteins can be moved along many different pathways in the cell.

Recycling amino acids

Proteins synthesis requires a constant supply of amino acids. Some are provided from our diet but many come from recycled amino acids that have formed part of proteins in the body. Proteasomes are protein complexes which degrade unneeded or damaged proteins by proteolysis, a chemical reaction that breaks peptide bonds. Proteasomes are present in the cytoplasm and in the nuclei of all eukaryotic cells. The amino acids that are released can be used to build new polypeptides and proteins.

KEY POINT

protein kinase an enzyme that regulates the biological activity of proteins by phosphorylating specific amino acids using ATP.

NATURE OF SCIENCE

Looking for trends and discrepancies: do all organisms use only 20 common amino acids in their proteins?

Humans can make 10 of the 20 amino acids we need to build proteins but we do not have the enzymes needed for the biosynthesis of the others. Plants, on the other hand, must be able to make all the amino acids they require.

Researchers have also investigated the trends in amino acid compositions of proteins found in species of the important kingdoms of Archaea, Bacteria and Eukaryotes. International databases ProteomicsDB and SWISS-PROT (which contain information about the structure and composition of proteins) can compare amino acid frequencies for 195 known proteomes and all recorded sequences of proteins. They discovered that the amino acid compositions of proteins do differ substantially for different kingdoms.

In addition to the variations in amino acids in proteins, some microorganisms and plants are able to make so called ‘non-standard’ amino acids by modifying standard amino acids. Some species are also able to synthesise many uncommon amino acids. For example, some microbes synthesise lanthionine, which is a modified version of the amino acid alanine. Many other proteins are modified after they have been produced. This ‘post-translational modification’ involves the addition of extra side groups to the amino acids in a protein.

Considering all the evidence, it seems that, although we can observe many similar proteins in different species, we cannot always say that the same amino acids are used in their construction. The range of amino acids in proteins can vary considerably from species to species.

To consider:

1 What contribution have international databases made to our understanding of protein structure?

2 How can comparing proteomes and amino acids in different organisms help our understanding of evolutionary relationships?

TEST YOUR UNDERSTANDING

12 Explain the difference between introns and exons.

13 Name three types of non-coding DNA sequence.

14 Where are telomeres located and what is their role?

15 Why are polysomes important in cells?

REFLECTION

Reflect on the areas of this topic that you found particularly interesting. What was it about them that caught your attention?

Links

- How does the variety of proteins produced contribute to the functioning of a cell? (Chapter 6)
- How does the degenerate genetic code protect a cell against mutations? (Chapter 3.3)

3.3 Mutations

LEARNING OBJECTIVES

In this section you will:

- learn that mutations are structural changes to genes at the molecular level
- learn how new alleles form by mutation; changes may be neutral, harmful or beneficial
- understand that mutations in germ cells can be passed to offspring
- define a mutagen as a substance that can cause genetic change
- recognise that mutations can add, delete or substitute base or bases in genes
- recall examples of insertion and deletion mutations
- learn that the genetic code is degenerate and so it is resistant to some changes caused by mutations

- understand that DNA polymerases can make proofreading errors, and the errors remain permanently
- recognise that mutations do not always cause changes to a protein's function
- understand the technique of 'gene knockout' and its use in investigating gene function

- learn how CRISPR sequences are used in gene editing
- recognise the importance of highly conserved sequences in genes

GUIDING QUESTIONS

- Why are variation and mutation essential for evolution? (Chapter 11)
- Why must the cell cycle be regulated and controlled? (Chapter 6)

3.3.1 Chromosomes, genes and mutations

A DNA molecule comprises a pair of strands, each strand consisting of a linear sequence of nucleotides, with weak hydrogen bonds between the bases holding the two strands together. This linear sequence of bases contains the genetic code in the form of triplets of bases. A gene is a particular section of a DNA strand that, when transcribed and translated, forms a specific polypeptide (Figure 3.3.1). Some of the polypeptides will form structural proteins, while others become enzymes or pigments such as hemoglobin and it is the translation of the genes which gives each individual organism its own specific characteristics. (Transcription and translation are described in Section 3.2.) Each gene is found at a specific position on a chromosome and so, for example, it is possible to say that the gene for human insulin is always found on chromosome number 11.

Organisms that reproduce sexually almost always have pairs of chromosomes, with one of each pair coming from each parent. The members of the pair carry equivalent genes, so that – for example – in humans, both versions of chromosome number 11 carry the insulin gene. But there may be slight differences in the version of the gene on each chromosome. These slightly different forms of the gene are known as alleles. Alleles differ from one another by one or only a few bases and it is these differences in alleles that give rise to the variation we observe in living organisms.

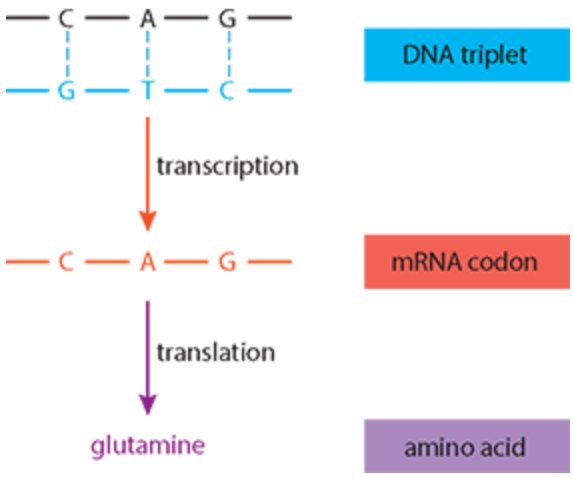


Figure 3.3.1: The base sequence in DNA is decoded via transcription and translation.

What are mutations?

The process of DNA replication is complex and mistakes sometimes occur – a nucleotide may be left out, an extra one may be added or the wrong one inserted. These mistakes are known as gene **mutations**. Mutations may occur spontaneously, as a result of errors in copying DNA, or they can be caused by factors in the environment known as **mutagens**, described in Section 3.3.2. A mutation that occurs in **germ cells** (**gametes**) that will go on to form a new offspring will be passed to every cell in those offspring, including their germ cells. As a result, the offspring may have a genetic condition that is not present in either of their parents. A mutation that occurs in **somatic cells** (body cells) will not be inherited. A mutation involving the change of a single nucleotide is called a **base substitution mutation**. When the DNA containing an incorrect nucleotide is transcribed and translated, errors may occur in the polypeptide that is produced. Errors may be beneficial, neutral or harmful.

Beneficial mutations change DNA and allow the synthesis of new proteins that may work slightly differently. One example of

a new, recently discovered, beneficial mutation has been found in a gene that codes for a receptor protein on the cell surface in the plasma membrane. Only a very few people carry this mutation but the change to their receptor protein gives them total immunity to infection by human immunodeficiency virus (HIV) because the virus cannot bind to their cells.

In 2009, United States researchers located another beneficial mutation in a gene (*SLC30A8*), which affects insulin. They found that subjects who were both overweight and elderly who carried the altered gene had considerable protection from developing type II diabetes. Evolutionary biologists also believe that our ability to discriminate three colours – red, green and blue – is due to a beneficial mutation that occurred in our primate ancestors' DNA millions of years ago.

However, beneficial mutations can be closely associated with harmful ones. Sickle-cell anemia, also called sickle-cell disease (SCD), is caused by a mutation that causes red blood cells to develop a crescent, or sickle shape; this abnormal shape can lead to a number of health problems but also has some advantages, which are described in the section on sickle-cell anemia.

How do mutations occur?

A mutation can involve the addition, deletion or substitution of a base in DNA or the inversion of a section of DNA so that it is turned backwards in the sequence.

Table 3.2.1 shows the amino acids that are specified by different mRNA codons. Most amino acids are coded for by more than one codon and so many substitution mutations have no effect on the final polypeptide that is produced. These are said to be neutral (or silent) mutations. For example, a mutation in the DNA triplet CCA into CCG would change the codon in the

mRNA from GGU to GGC but it would still result in the amino acid glycine being placed in a polypeptide. Other examples of neutral mutations are those that affect non-coding regions of the chromosome, or which result in changes to features such as blood type or eye colour in humans that do not adversely affect a person.

Some substitution mutations, however, do have serious effects. For example, one important human condition that results from a single base substitution is sickle-cell anemia.

Degeneracy in the genetic code

As Table 3.2.1 shows, there are 64 combinations of three-letter nucleotide sequences that can be made from the four nucleotides. Of these, 61 represent amino acids and three are stop signals. Although each codon is specific for only one amino acid or one stop signal, the genetic code is described as degenerate because a single amino acid may be coded for by more than one codon.

Degeneracy in the genetic code makes the code more resistant to changes. It means that single base substitutions can occur without disrupting protein synthesis or functioning of the organism.

EXAM TIP

You may be asked to explain inversions, deletions, substitutions or additions to a DNA sequence. When you do this, first remember to check whether you need to convert the DNA sequences to mRNA sequences before identifying the amino acids. Most tables used to decode the genetic code are shown as tables of mRNA codons.

Sickle-cell anemia: the result of a base substitution mutation

Sickle-cell anemia is a blood disorder in which red blood cells become sickle shaped and cannot carry oxygen properly (Figure 3.3.2). It occurs most frequently in people with African ancestry, about 1% suffer from the condition and between 10 and 40% are carriers of it. Sickle-cell anemia is due to a single base substitution mutation in one of the genes that make hemoglobin, the oxygen-carrying pigment in red blood cells.

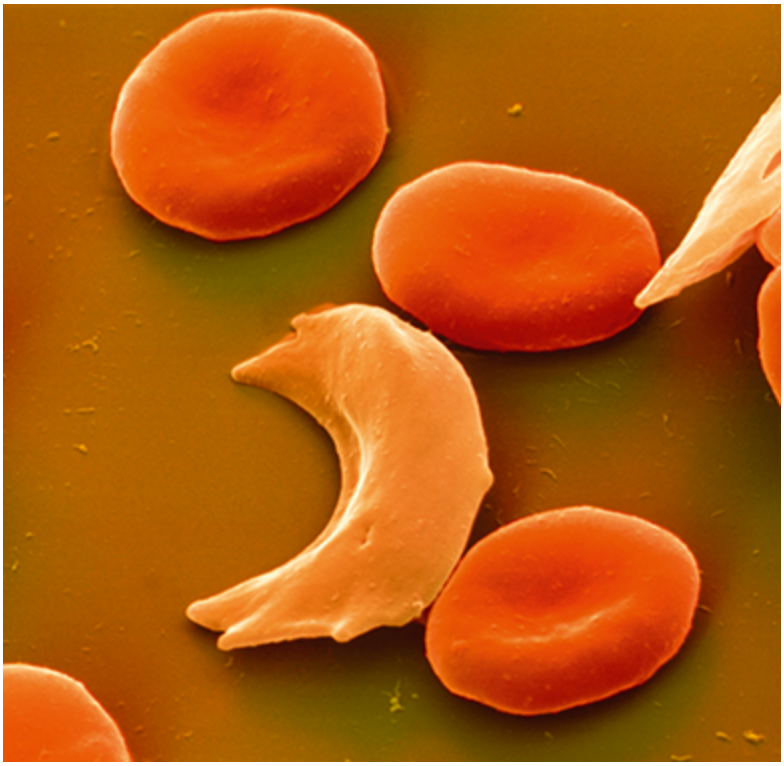


Figure 3.3.2: Coloured scanning electron micrograph showing a sickle-cell and normal red blood cells (× 7400).

Hemoglobin is made up of four subunits, as shown in Figure 3.3.3 – two a-chains and two b-chains. The b-chains are affected by the sickle-cell mutation. To form a normal b-chain, the particular triplet base pairing in the DNA is:

The C–T–C on the coding strand of the DNA (in blue here) is transcribed into the mRNA triplet G–A–G, which in turn is translated to give glutamic acid in the polypeptide chain of the b-chain.

If the sickle-cell mutation occurs, the adenine base A is substituted for thymine base T on the DNA coding strand, so the triplet base pairing becomes:

WORKED EXAMPLE 3.3.1

A mutation can involve the addition, deletion or substitution of a base in DNA. Consider the effect of these changes on this short length of DNA:

CTG GGG GGT GTG AAC

The sequence of amino acids produced by this sequence should be

Leu Gly Gly Val Asn

If the base highlighted in red above is *deleted* the consequence would be:

CTG GGG GGT TGA AC

This would result in the amino acids

Leu Gly Gly STOP

because TGA is a stop codon. The polypeptide produced will be shorter than it should have been.

Question

What type of mutation has occurred in these examples and what are the consequences?

a CTG GGG GGT AGT GAA C

Answer

In this case a base has been *added* to the sequence after the third codon. This will result in serine, coded for by AGT, being added as the fourth amino acid instead of valine. All the subsequent amino acids in the sequence will also be incorrect. This is known as a frameshift mutation because the 'reading' of the DNA in sets of three bases is changed completely. New amino acids will be inserted and produce a different translation of the code from the inserted bases onwards. Deletion of a base can also cause a frameshift mutation.

b CTG GGG GGT GTG CAA

Answer

Here the final triplet has been *inverted* (turned around, so it is backwards) and is CAA instead of AAC. In this case glutamine will be inserted into the amino acid chain instead of asparagine.

c CTG GGG GGG GTG AAC

Answer

In this example a *substitution* of a base has occurred. G has replaced T. In this case there will be no effect on the amino acid sequence that is assembled because GGT and GGG both code for the amino acid glycine. This is an example of a neutral (or silent) mutation.

This type of mutation is known as a **point mutation**. C–A–C on the coding strand of the DNA is now transcribed into the mRNA triplet G–U–G, which in turn is translated to give the amino acid valine. Valine replaces glutamic acid in the b-chain.

Valine has different properties from glutamic acid and so this single change in the amino acid sequence has very serious effects. The resulting hemoglobin molecule is a different shape and it is less soluble and, when in low oxygen concentrations, it deforms the red blood cells to give them a sickle shape. Sickle cells carry less oxygen, which results in anemia. They are also rapidly removed from the circulation, leading to a lack of red blood cells and other symptoms such as jaundice, kidney problems and enlargement of the spleen.

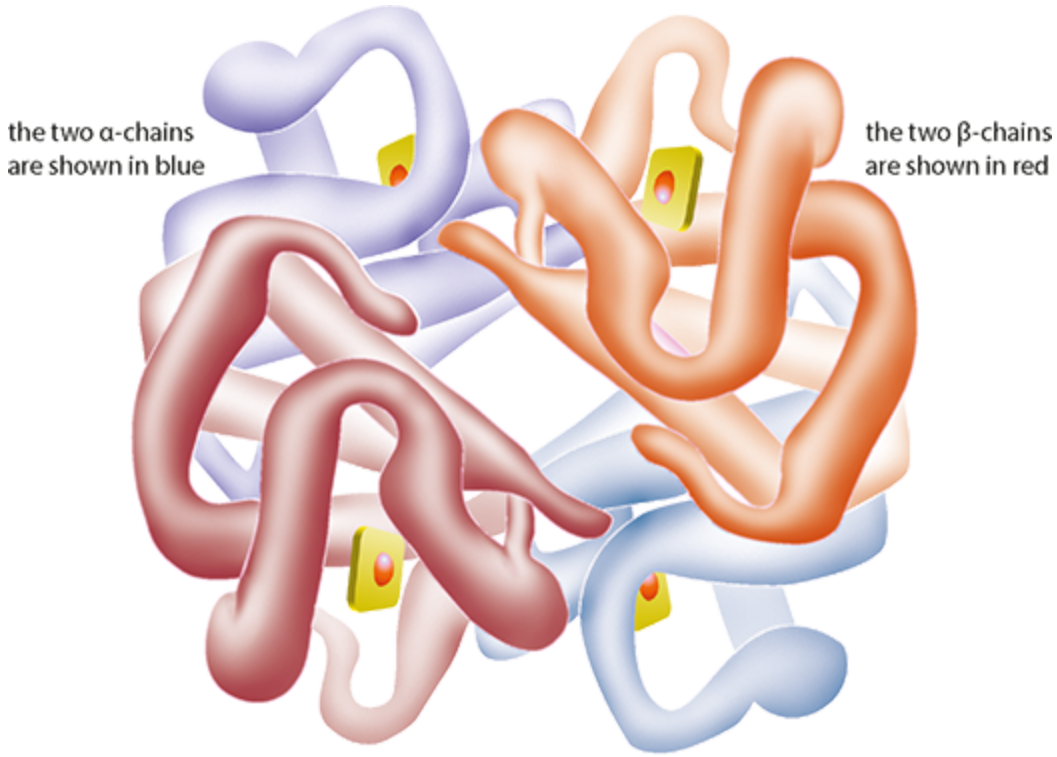


Figure 3.3.3: The structure of a hemoglobin molecule showing the 3D arrangement of the subunits that make it up.

People who have one sickle-cell allele and one normal allele are said to have sickle-cell trait and have some resistance to malaria. This benefit explains why the mutation persists in areas where malaria is endemic. But, in parts of the world where malaria is not a problem, the mutation no longer provides a survival advantage. Instead, it poses the threat of sickle-cell disease, which occurs in the children of carriers who inherit the sickle-cell gene from both their parents.

Insertions and Deletions of bases

Inserted and deleted bases can have serious consequences for protein production in cells. Sometimes changes in bases can cause serious illness but in other cases they lead to benefit. Two

examples are the mutations in the *HTT* gene and in the CCR5 gene.

Huntington's disease: result of insertions - trinucleotide repeats

The *HTT* gene provides instructions for making a protein called huntingtin which plays an important role in nerve cells in the brain. One region of the *HTT* gene contains a particular DNA segment known as a CAG trinucleotide repeat. This segment is made up of a series of three DNA bases (cytosine, adenine, and guanine) that appear multiple times in a row. Normally, the CAG segment is repeated 10 to 35 times within the gene. But if the number of repeats increases to more than 40 the result will be Huntington's disease. If more than 60 repeats are present a more severe form of the disease develops. The greater the number of repeats, the sooner the disease appears.

Huntington's disease stops parts of the brain working properly causing memory loss, difficulty with movement, mood swings and depression. It develops over a period of time and is usually fatal after about 20 years.

HIV resistance: result of a deletion

The human CCR5 gene located on chromosome 3 codes for a protein called CCR5. The protein is found on surface of lymphocytes and other cells of the immune system. The proteins form part of the receptors that are involved in cell signalling and the coordination of immune responses. The CCR5 receptors provide a point of entry for the HIV-1 virus to infect the cells. Some people have inherited a mutation called Delta 32 and have part of the CCR5 gene deleted. The gene changes alter the structure of the CCR5 receptors and make it difficult for the HIV virus to enter cells. Individuals who have this deletion mutation

live normal lives, but if they inherit copies of the mutation from their both parents, they are naturally immune to HIV.

NATURE OF SCIENCE

Tests for genetic diseases

There are many commercially available tests for potential health and disease risks. For example, the most effective and accurate method of testing for Huntington's disease is called a direct genetic test. DNA from a blood sample is taken and analysed for the number of CAG repeats it contains. The presence of 36 or more repeats is an indication that the person has, or will develop Huntington's disease. In a few cases, the test result is not clear and a definite answer is not possible. Any person who decides to take this test must be prepared to face not only the emotional effect it will have on themselves and their family, but also the affect on other aspects of their life. Life insurance and some job opportunities may become unavailable to them if the test is positive.

If any genetic test is taken it is important that the outcomes are interpreted correctly by an expert who can explain the consequences of the results. A negative result can eliminate the need for check-ups while a positive result can help direct a person to monitoring and treatment options. Some tests can help people decide about whether or not to have children and both genetic and non-genetic tests can provide information about a person's health in the future.

To consider:

1 Why do some people think that genetic test results can cause family discord, psychological distress and stigmatisation?

2 Why might some people decide not to have children as a result of a genetic test?

Gene knockout to investigate gene function

Gene knockout is a method that is used to damage or 'knock out' specific genes so that they no longer function and are not expressed. Gene knockout is used with model organisms such as mice and yeast which have specific genes knocked out so they can be used to study how those genes function and investigate what happens when the genes are lost. Researchers can draw inferences from the difference between the knockout organism and normal individuals with a similar genetic background. Knockout organisms are also used in the development of new drugs which target specific biological processes or genetic deficiencies. A library of the genomes of model organisms is available to researchers. The loss of gene activity often causes the phenotype of the model organisms such as mice to change so that living organisms can be used to study gene function. For example, 'Metheuselah' is a knockout model mouse which lives for far longer than an average animal and 'Frantic' is a model mouse which is used for studying anxiety disorders. The loss of the knocked out genes provides valuable information about what the gene normally does. Mice are useful model organisms because humans share many genes with mice.

CRISPR

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. Repeated DNA sequences, called CRISPR, were first noticed in bacteria. They have 'spacer' DNA sequences in between the repeats that exactly match sequences found in viruses. These sequences which contain short repetitions of base sequences are involved in the defence

mechanisms of prokaryotic organisms such as bacteria and archaea to viruses that infect them. The sequences have come from DNA fragments of viruses that have previously infected the prokaryotes and the organisms use them to detect and destroy DNA from similar viruses that might infect them.

CRISPR is now used as a genetic engineering tool that uses repeated sequences of DNA to edit genes. Using CRISPR it is possible to find a specific section of DNA inside a cell so that a gene can either be modified or even turned on or off without altering the DNA sequences. The key to CRISPR is the many variations of 'Cas' proteins found in bacteria. These are the proteins produced by prokaryotes which help defend them against viruses. A protein called Cas9 is the most widely used by scientists. This protein can easily be programmed to find and bind to almost any desired target sequence, simply by giving it a piece of RNA to guide it.

When the CRISPR Cas9 protein is added to a cell along with a piece of guide RNA, the Cas9 protein links to the guide RNA and then moves along the strands of DNA until it finds and binds to a 20-DNA-letter long sequence that matches part of the guide RNA sequence.

One successful use of CRISPR has been in the treatment of human papillomavirus (HPV). This very common virus has more than 100 different strains; some of them affect the skin causing warts and are barely noticed, but others contribute to 99% of cervical cancers. Researchers have been able to turn off two genes in the virus and knock out the production of two viral oncoproteins. A constant supply of these viral proteins is needed to transform normal cells into cancer cells. Without the proteins, cells infected with the virus go into senescence which means that they stop dividing. Targeting the genes for the proteins can

potentially treat HPV-related cancers. CRISPR and Cas9 has also been used in the treatment of hepatitis B, in this case the ends of certain repeated sequences in the Hepatitis B viral genome are targeted. It has also been used experimentally to repair the mutations that cause cataracts in mice.

NATURE OF SCIENCE

Since the human genome was first sequenced, genetic research has expanded rapidly. The cost of sequencing the entire genome of one person has dropped from about US$1 billion to $1,000, and the speed of sequencing has become many times faster. Scientists around the world have a new and powerful way of understanding how genetic variation may affect not only organisms that humans make use of, but also human health and disease. CRISPR technology can be used to edit genes and has the potential to change genomes.

The CRISPR method was first used in 2012 and replaced costly methods of gene editing that had been used in some plants and animals previously. CRISPR has made gene editing cheap and easy. The technique is widely used in research and already has the potential to alter plants and animals on our farms. The technology also has the potential to treat and prevent many diseases.

It could even change the genomes of future generations, although many people think this is unethical. CRISPR is already being used to fingerprint cells and observing what happens inside them, and for directing evolution.

The most common use of CRISPR involves using Cas9 protein to cut the DNA at a target area. When the cut is repaired, mutations can be introduced that disable a gene. But

CRISPR can also be used to make precise changes such as replacing faulty genes. At present this is much more difficult. The knowledge gained from studying human gene knockouts gives scientists a tool to identify potential new targets for medical treatments and better understand safety concerns of treatments that are being developed. But the technique could have the potential to make permanent changes to a person's genome. Scientists around the world are subject to different rules in the use of genome technology. For this reason, there is an ongoing effort to make a system of regulation for all scientists working in this rapidly growing field of research.

Questions

1 Why is it important to have clear rules about what should and should not be attempted using technology such as CRISPR?

2 What are the potential benefits and dangers in gene editing and replacement?

Why are some gene sequences conserved in a species?

Conserved sequences are sequences of DNA (or protein) that are identical or very similar across a species or group of species. A highly conserved sequence is one that has remained relatively unchanged for long periods of evolutionary history. Examples of highly conserved sequences include those for the RNA found in ribosomes that are present in all domains of life, and the homeobox sequence which is a DNA sequence of around 180 base pairs that regulates the early stages of embryonic development. This sequence is found in many eukaryotes. Studies of sequence conservation now form part of

investigations in genomics, proteomics and evolutionary biology.

KEY POINT

conserved sequence a base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained relatively unchanged throughout evolution.

An explanation for the presence of these conserved sequences was put forward in 1965, by Emile Zuckerandl and Linus Pauling who proposed the hypothesis of a molecular clock. Amino acid sequences can be conserved to maintain the structure or function of a protein. Conserved proteins undergo fewer amino acid replacements or are more likely to substitute amino acids with similar biochemical properties. The molecular clock theory went on to suggest that steady rates of amino acid replacement could be used to estimate the time when two organisms diverged in evolution (Section 11.1.3). Many phylogenetic relationships worked out from studies of the fossil record seemed to support this theory but some other genes were found to evolve at different rates. This led to the development of theories of molecular evolution. In 1966 Margaret Dayhoff compared ferrodoxin (small proteins involved in a range of metabolic reactions) sequences in many organisms and proposed that natural selection will act to conserve protein sequences that are essential to life. This hypothesis explains why conserved sequences of DNA are found for many important proteins in many species.

3.3.2 Harmful mutations and mutagens

New cells are needed to replace cells that have died or to allow an organism to grow. The nucleus and cytoplasm of a cell divide in processes known as mitosis and cytokinesis, which are phases in a series of events known as the cell cycle (Section 6.5).

In normal circumstances the cell cycle is strictly controlled with cell division (mitosis) occurring to form new cells to replace damaged or dying cells.

In most cases, mitosis continues until a tissue has grown sufficiently or repairs have been made to damaged areas.

Most normal cells also undergo a programmed form of death known as **apoptosis** as tissues develop. Apoptosis can be caused when a cell experiences stress or if it receives signals that indicate it should die.

But sometimes mitosis does not proceed normally. Cell division may continue unchecked and produce an excess of cells, which clump together. This growth is called a **tumour**. Tumours can be either **benign**, which means they are restricted to that tissue or organ, or **malignant** (cancerous), in which some of the abnormal cells migrate to other tissues or organs and continue to grow further tumours there. If they are allowed to grow without treatment, tumours can cause obstructions in organs or tissues and interfere with their functions.

Mutagens are physical, chemical or biological agents that can cause mutations and modify DNA. Mutagens include ionising radiation – such as X-rays, gamma rays and ultraviolet light – and also chemical compounds, such as those found in tobacco smoke and aflatoxins produced by certain fungi. The DNA

changes caused by mutagens are not all harmful. However, because some of them cause cancer, some mutagens are said to be **carcinogens** (cancer causing). The development of a **primary tumour** can also be caused by mistakes in copying DNA, or a genetic predisposition as a result of inheritance. (You can learn more about the control of cell division and the development of cancer in Section 6.5)

Smoking and cancer

Smoking is a major cause of several types of cancer. There is strong evidence to show that it increases the risk of cancer of the bladder, cervix, kidney, larynx and stomach, and smokers are seven times more likely to die of these cancers than non-smokers. In the UK, approximately 70% of lung cancers in both males and females are related to smoking.

SCIENCE IN CONTEXT

Smoking and lung cancer

All tobacco products contain various amounts of carcinogenic substances. Tobacco smoke contains more than 70 chemicals, including many which are known to initiate or promote cancer. Recently the role of nicotine, the addictive drug in tobacco, has come under scrutiny as more people are turning to e-cigarettes and other non-tobacco sources of nicotine as substitutes for smoking. There is no clear evidence that nicotine is a direct carcinogen, but it seems to act as a promoter and may inhibit anti-tumour immune responses. In experiments, nicotine has been shown to induce breaks in DNA and enhance the growth of existing cancers.

The link between smoking and lung cancer was recognised in the 1940s and 1950s, with evidence from **epidemiology**,

animal experiments, examination of cells and chemical analysis. Cigarette manufacturers disputed the evidence, as part of a campaign to maintain cigarette sales. Their propaganda was successful in the short term and, as late as 1960, only one-third of all US doctors believed that the case against cigarettes had been established.

Today it is understood that the risk of contracting lung cancer increases with the number of cigarettes that a person smokes and the number of years that they continue to smoke. If a person gives up smoking, their risk of developing cancer decreases (Figure 3.3.4).

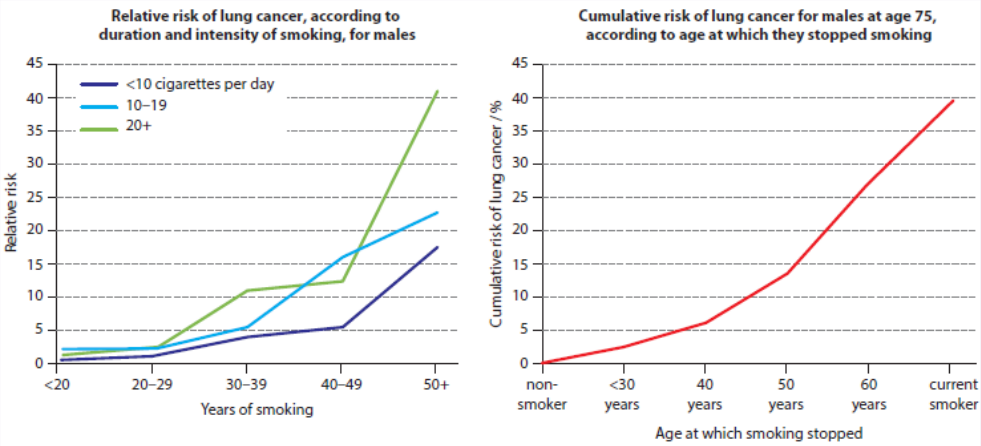


Figure 3.3.4: Graphs to show the relationship between smoking and lung cancer, and the cumulative risk of lung cancer among men in the UK at age 75 according to the age at which they stopped smoking (data from Cancer Research UK).

Lung cancer develops slowly and it takes years before the effects of the carcinogens become obvious. The number of males who suffered lung cancer in the UK was at its highest

levels in the early 1970s. This was as a result of a peak in smoking 20–30 years earlier. Cancer in females increased through the 1970s and 1980s because more females smoked in the 1950s and 1960s. Statisticians predicted that cancer in females would increase to reach the same levels as those in males over the next decade. New government education campaigns have persuaded people give up smoking and new laws have limited smoking in public places, and so the number of deaths have started to decrease. Figure 3.3.5 shows the incidence of lung cancer in the UK 1978–2017 and the incidence of smoking since 1948. As had been predicted, the rate for females increased by 15% but for males rates decreased by 11%.

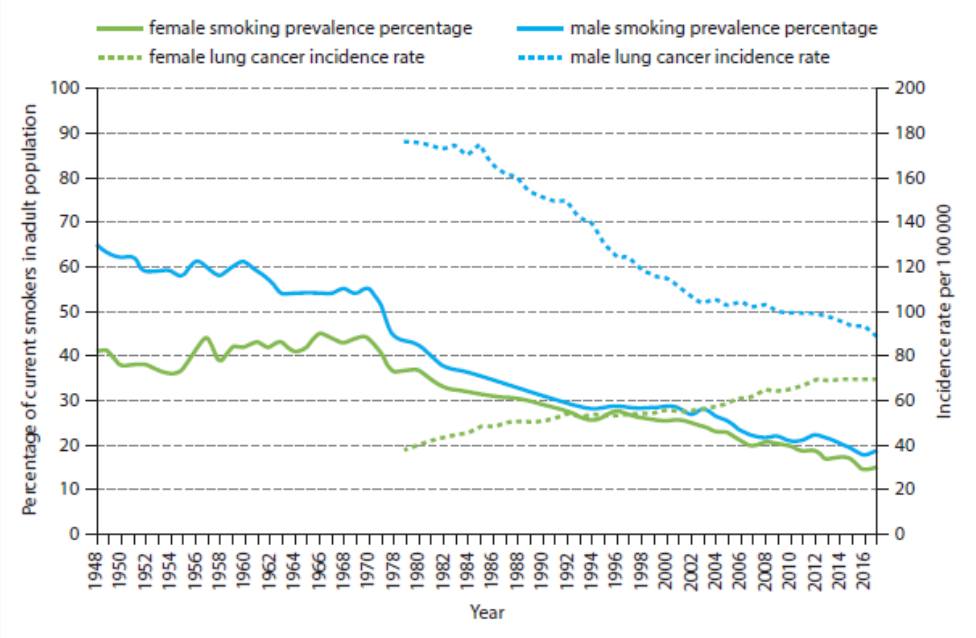


Figure 3.3.5: The incidence of smoking in UK males and females 1948–2017 and the incidence of lung cancer 1978–2017.

To consider:

- Lung cancer affects the economic performance of a country in health care costs and loss of production when people are unwell. Why do you think that governments and officials were reluctant to accept the link between smoking and health in the 1950s?
- How important were laws and restrictions in persuading people to modify their smoking behaviour?

3.4 Epigenetics

LEARNING OBJECTIVES

In this section you will:

- define gene expression as the mechanism in which genetic information affects the phenotype
- define epigenetics as the study of changes to gene activation in differentiated cells
- learn that gene expression is regulated by proteins that bind to base sequences in DNA and that degradation of mRNA can regulate translation
- understand that epigenetic changes modify the activation of certain genes but do not change their base sequences so that the phenotype will change but genotype does not
- learn that DNA methylation inhibits transcription
- understand that epigenetic changes are faster than changes caused by natural selection
- recognise that environment has an impact on gene expression and can trigger heritable changes in epigenetic factors
- learn that most epigenetic tags in gametes are removed from the embryo genome, but some remain and are

inherited leading to the appearance of phenotypic differences in organisms such as lion-tiger hybrids

- learn that external factors such as hormones and nutrients can affect the pattern of gene expression
- understand how environmental effects on DNA methylation can be studied using monozygotic twins.

GUIDING QUESTIONS

- How is DNA modified to influence gene expression?
- How is differentiation of cells brought about by epigenetics?

3.4.1 Epigenetics and gene expression

As our genes are transcribed and translated and proteins are built from the information they carry, an organism's appearance, enzymes and metabolism are all controlled and decided. **Gene expression** is the mechanism by which information carried in DNA has its effects on the phenotype of an organism. The stages of this process include transcription, translation and the function of the protein that is produced. Gene expression is regulated and controlled so that the genes that are expressed match the needs of the organism. Transcription can be regulated by proteins that binds to specific base sequences; these may be promoters, enhancers or transcription factors which either allow or prevent transcription of a gene. Translation can be regulated by the length of time that mRNA is present in a cell and this too is controlled in the cytoplasm. In addition, epigenetic changes can influence patterns of development and differentiation of cells without changing the genotype of the cell.

KEY POINTS

gene a particular section of a DNA strand that codes for a specific polypeptide; a heritable factor that controls a specific characteristic.

gene expression the mechanism by which genetic information affects the phenotype of an organism

genome the entire set of DNA instructions found in a cell

proteome the complete set of proteins expressed by an organism

transcriptome all the mRNA molecules expressed from the genes of an organism

Regulation of transcription by proteins that bind to DNA

Gene expression and binding proteins

Before transcription can begin and mRNA production start, RNA polymerase requires the presence of a class of proteins known as general transcription factors. **Transcription factors (TFs)** are regulatory proteins whose function is to activate transcription of DNA by binding to specific DNA sequences. TFs specifically bind to target sequences which are highly conserved. These sequence specific transcription factors are probably the most important mechanism of gene regulation in cells. In eukaryotes gene expression requires the co-ordinated interaction of several of these proteins. Interactions between the transcription factors, RNA polymerase and the **promoter region** of the DNA molecule allow the RNA polymerase to attach and move along a gene so that transcription can occur. There will be greater transcription of certain genes when specific transcription factors are present so that the proteins that a cell needs can be assembled. Many different transcription factors have been found and each one is able to recognise and bind to a specific nucleotide sequence in DNA.

KEY POINTS

transcription factors (TFs) proteins that bind to particular sites on DNA and activate transcription. Together with RNA polymerase and other proteins (activators), TFs form the

transcription apparatus and have a key role in regulating genes.

promoter a region of DNA to which proteins (RNApolymerase and TFs) bind to initiate transcription of a gene

The role of activators is shown in Figure 3.2.3. These proteins bind to a region of the DNA called the **enhancer**, which may be some distance from the gene. 'Bending proteins' may then assist in bending the DNA, so that the enhancer region is brought close to the promoter. Activators, transcription factors and other proteins attach, so that a 'transcription-initiation complex' is formed and transcription can begin.

Transcription factors are regulated by signals produced from other molecules. For example, hormones are able to activate transcription factors and thus control transcription of certain genes. Many other molecules in the environment of a cell or an organism can also have an impact on gene expression and protein production.

Regulation of translation by mRNA degradation

Once mRNA has reached the cytoplasm, translation can be regulated by mRNA persistence (length of time it is present). mRNA is degraded by nuclease enzymes and the degradation of mRNA and the efficiency with which it is translated are another essential stage in determining gene expression. Individual mRNA molecules can exist in an active state, a silent state or a state that is targeted for decay. In general, RNA is degraded at the end of its useful life, which is very short for introns and spacer fragments, but longer for other sections of mRNA. The time varies between a few minutes to a few days. RNA

molecules with defects in processing or assembly are rapidly identified and degraded by the nuclease enzymes. mRNA lifespan can be shortened if translation is incorrect affected or made more stable if translation elongation or termination are inhibited. mRNA lifespan can also be altered in response to developmental, environmental and metabolic signals.

RNA silencing

RNA silencing is one method of gene silencing that includes several pathways to control and regulate gene expression. Small non-coding strands of RNA (such as microRNAs and RNAi) may either block sections of transcribed mRNA by pairing with it, or degrade the mRNA in the cytoplasm. In both cases the mRNA is not translated and both methods therefore can prevent the translation of some genes. RNA silencing is also used to silence genes in research into the production of medicines to combat cancer and other diseases. This is because RNA silencing is used in the cells of most organisms to fight RNA viruses which are destroyed in the cytoplasm after transcription.

Epigenetics is a relatively new area of investigation in biology. It is the study of how the expression of DNA can be changed without changing the structure of DNA itself. The phenotype (characteristics) of an organism may be changed but its genotype (sequences of DNA) remain the same. Epigenetic changes can affect how cells read their genetic code and a few of the changes can be passed on to the next generation.

Environment and gene expression

The expression of genes can be influenced by the environment - not only the organism’s external environment, but also its internal environment, which is affected by chemicals such as hormones and various products of metabolism. Temperature,

light and chemicals are just some of the environmental factors that can cause some genes to be turned on or off and influence how an organism functions or develops.

KEY POINT

gene silencing interruption or suppression of the expression of a gene either at transcription or translation

3.4.2 Epigenetic changes

The activity of genes can be influenced by certain DNA modifications that do not change an organism's DNA sequence but which do affect which genes are active and which are not. These changes to gene activity will influence the phenotype of an individual but not its genotype. Chemical compounds attached to single genes can produce modifications known as epigenetic changes. When chemical compounds are attached to the genome, it is referred to as an epigenome. The additions or 'tags' are not part of the DNA sequence, but remain attached to DNA as cells divide and, in some cases, can be passed on to offspring. One group of chemical tags are methyl groups which attach to the DNA base cytosine when it is followed by a guanine, and it is their location that influences the expression of the associated gene. Tagging patterns vary from one cell to the next.

KEY POINTS

epigenetics the study of changes to gene activation in differentiated cells.

epigenome all the chemical compounds that have been added to a genome to regulate the expression of all the genes within the genome.

All cells in an organism contain the same DNA, but the many different cell types function differently, so that muscle cells, for example, produce different proteins from cells of the intestine. **Epigenetic changes** help to determine whether genes are turned on or off and determine which proteins are transcribed in each

cell. Cells are different because some of their genes are turned off, while others are turned on. **Epigenetic silencing** turns genes off so that only necessary proteins are produced and enable different cells to behave differently. Environmental influences, such as diet and exposure to pollutants, can also affect the **epigenome**. (See Science in Context, Nutrition and epigenetics). Here we discuss the three most important types of epigenetic change: DNA methylation, histone modification and RNA silencing.

SCIENCE IN CONTEXT

Nutrition and epigenetics

We know that an organism's development is influenced by genes being switched on or off at specific times and there has been much debate about how environmental factors can lead to such epigenetic modifications. The environment's effects can influence human health, and some of these effects can be inherited.

In the early 21st century, Swedish scientists investigated whether nutrition affected the death rate associated with cardiovascular disease and diabetes in a number of Swedish families. They were interested to learn whether the effects were passed from parents to their children and grandchildren.

Researchers examined records of harvests and food prices in Sweden from the 1890s onwards and the medical records of three generations of the families. Their studies revealed that if a father did not have sufficient food in his pre-pubescent years, his sons were less likely to suffer from cardiovascular disease.

For diabetes the picture was different: the children's death rate due to diabetes was unaffected if their father had had a plentiful supply of food at the same critical period. But if the children's grandfather had been well nourished at the same critical period of his life, the incidence of diabetes in his grandchildren was increased.

This suggests that diet can cause epigenetic changes to human genes that affect likelihood of disease. Furthermore, the changes have been passed on through males in a family in a similar way to the inheritance of coat colour in mice described in section 3.4.4.

To consider:

1 Why do you think that researchers used records from the 1890s in their investigations?

2 What other factors might be important in the development cardiovascular disease and diabetes?

DNA methylation

DNA methylation is the most common type of epigenetic modification. It involves attaching methyl groups, consisting of one carbon atom and three hydrogen atoms, to segments of DNA (Figure 3.4.1). When methyl groups are added to a particular gene, that gene is turned off or silenced, and no protein is produced from it.

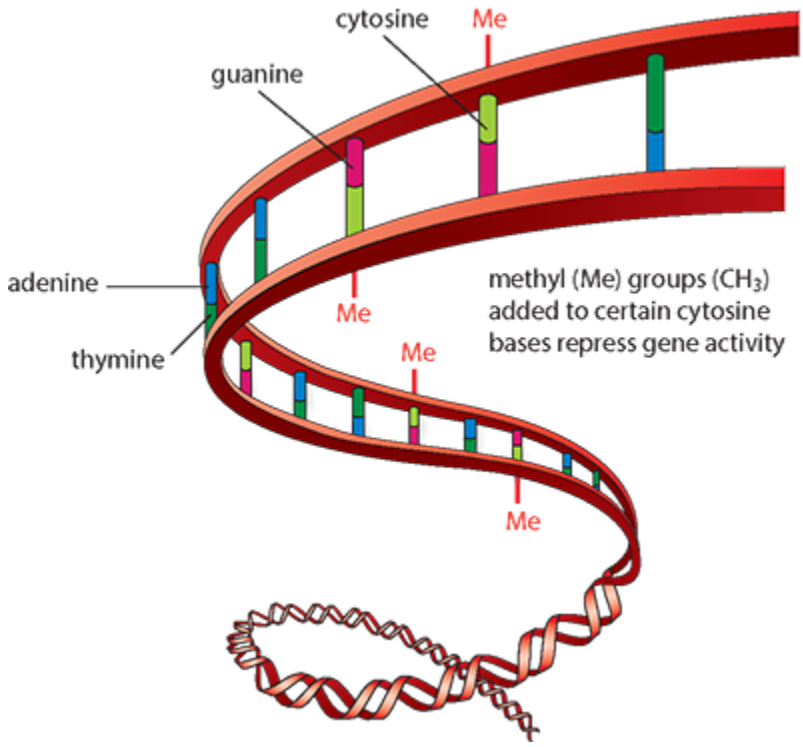


Figure 3.4.1: Methylation of DNA is one epigenetic factor affecting gene expression.

DNA methylation always happens in a region known as a CpG site where a cytosine nucleotide is located next to a guanine nucleotide that is linked by a phosphate. The enzyme DNA methyl transferase adds methylation markers to the base cytosine. Inserting methyl groups here changes the appearance and structure of DNA, and modifies the interactions between transcription factors that determine whether the gene will be expressed. Promoter regions of genes often lie within areas known as 'CpG islands' and if CpG is methylated, the gene will not be expressed.

Special binding proteins can recognise and 'read' these epigenetic markers. If binding protein is missing or a mutation occurs in it, genes which should not be expressed will transcribed.

KEY POINT

DNA methylation a process that adds a methyl group to the DNA base cytosine. It is an example of an epigenetic marker.

EXTENSION

Rett syndrome, a neurological disorder which affects brain development in girls, has been linked to the absence of binding proteins which recognise methylated markers in DNA.

Histone modifications

DNA in eukaryotes is packaged around histones and incorporated into nucleosomes so that the genetic material can be stored in a compact form (Figure 3.4.2) known as chromatin. In order to transcribe genes, enzymes involved in transcription must be able to gain access to DNA. In all eukaryotes, the regions of DNA that contain promoters and regulators have fewer nucleosomes than other areas, allowing greater access for binding proteins, while regions that are transcribed have a higher density of nucleosomes. Nucleosomes have an important role in determining which genes are transcribed and can influence cell variation and development.

KEY POINT

chromatin is an association of histone proteins and DNA which help to package DNA in a compact form in the cell nucleus.

DNA does not need to be completely released from a nucleosome to be transcribed. Nucleosomes are very stable protein–DNA complexes but they are not static. They can undergo structural rearrangements, including ‘nucleosome sliding’ and DNA site exposure. Nucleosomes are important because they can either inhibit or allow transcription by controlling whether binding proteins can access DNA.

Histones can be modified so that they influence the arrangement of chromatin about the chromosome and thus also DNA transcription. If chromatin is compact (condensed) it will prevent DNA transcription but if it is loose (active) DNA can be transcribed. Histones can either be methylated or acetylated by the addition of a methyl or acetyl group to the amino acid lysine in the histone. Acetylation produces active, less condensed chromatin so that proteins involved in transcription can bind to DNA and a gene can be transcribed. Histone methylation can indicate either active or inactive regions of chromatin (Figure 3.4.3).

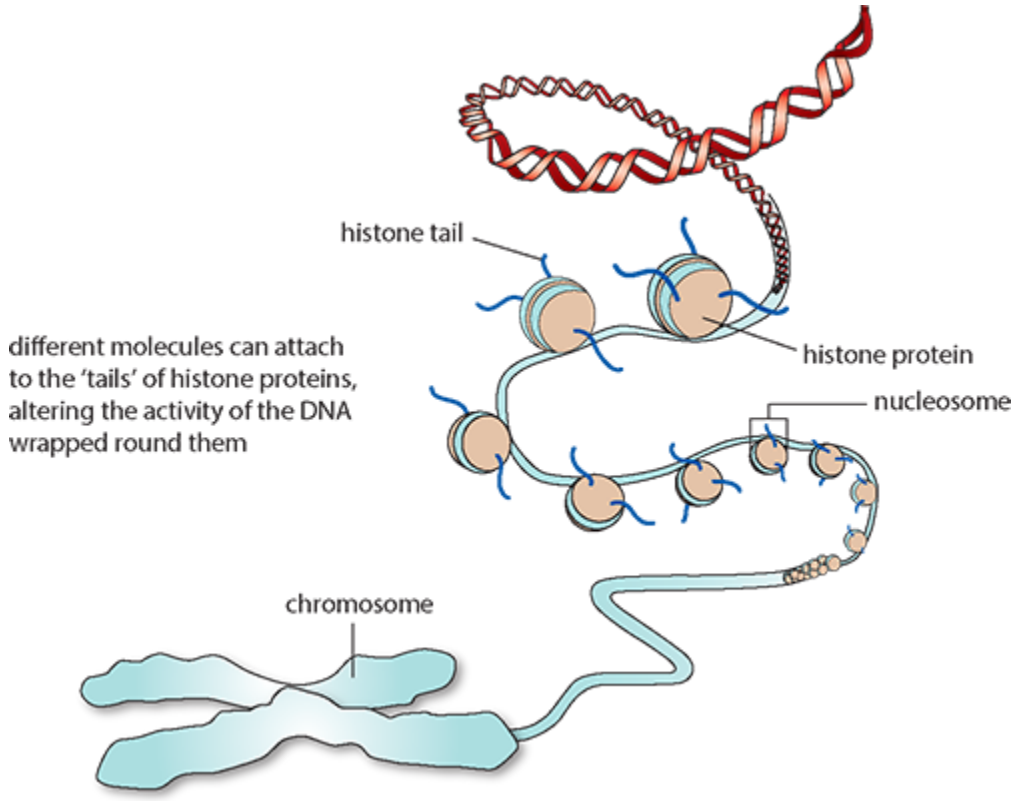


Figure 3.4.2: Histone modification is another epigenetic factor affecting gene expression.

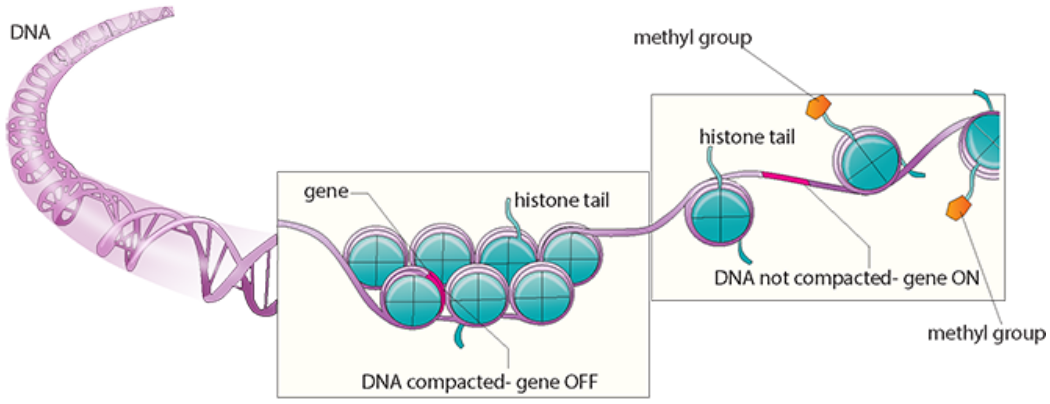


Figure 3.4.3: Effect of methylation on chromatin.

Histone methylation and deacetylation are also important in female mammals who have two X chromosomes. One of the two

chromosomes is inactivated so that females do not produce twice as many X-chromosome gene products as males.

3.4.3 Epigenetic markers and offspring

The genes in egg and sperm cells from the same species contain different epigenetic markers, which cause them to be differentially expressed in the zygote (the cell produced by the fusion two gametes) and developing embryo. These genes are known as imprinted genes. The expression of these genes depends upon which parent contributed them. Most epigenetic markers (tags) found in egg and sperm cells are removed from the epigenome of the embryo and so are not inherited. But in mammals a few do remain and are passed on. One example of this has been investigated in agouti mice. The *Avy* (agouti variable yellow) gene in mice influences the animal's coat colour. Research into the *Avy* gene established that coat colour was related to the degree of methylation of the gene. A high degree of methylation inactivates the gene so the mouse has a dark coat. Without methylation, the gene is active and the coat is yellow. An active gene is also linked to an increased likelihood of obesity and diabetes. Later research showed that if pregnant mice were fed with increased the levels of methylated molecules, such as folic acid and zinc, DNA methylation at the agouti locus was found in their offspring. Baby mice were born with darker coats and leaner bodies (Figure 3.4.4). Furthermore, the mother's diet affected not only characteristics of her own offspring, but also of her daughters' offspring in the next generation.

Scientists had previously believed that methylation markers were always removed from DNA as sperm and egg cells were produced but these experiments suggest that markers must remain, for at least some genes.

Another example of imprinted genes affecting phenotype can be seen in the cat family. Lions and tigers do not normally meet in nature, but in captivity they may mate and sometimes produce hybrid offspring. The offspring look different, depending on which animal is the mother. A male lion and a female tiger produce a liger – the largest of the big cats. A male tiger and a female lion produce a tigon, a cat that is about the same size as its parents. The difference in size and appearance between ligers and tigons is due in part to the parents' differently imprinted genes.

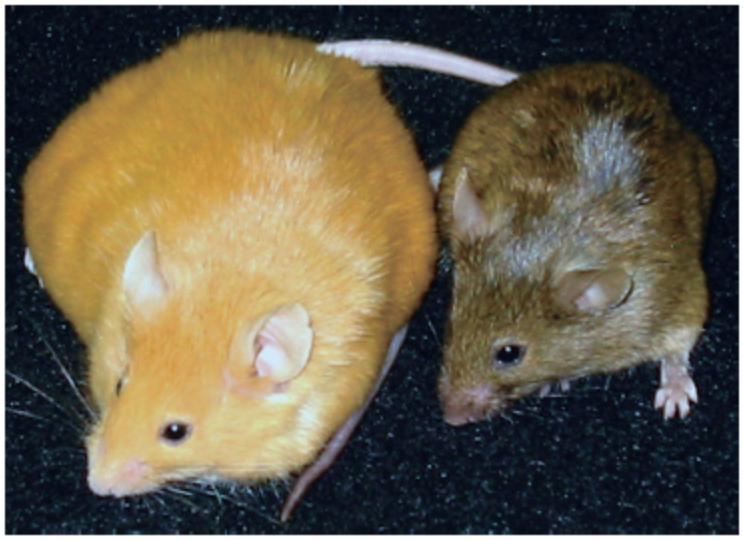


Figure 3.4.4: These mice are genetically identical and the same age. The mother of the left-hand mouse received a normal 'mouse diet' during pregnancy, while the mother of the mouse on the right was fed supplements including folic acid.

Genomic imprinting

Even though both parents contribute equally to the genetic content of their offspring, genomic imprinting sometimes leads to expression of certain genes from only one parent. A marker or imprint can affect particular genes on the maternal and paternal

chromosomes in such a way that only one copy of those genes is expressed in the offspring.

KEY POINT

genomic imprinting inheritance that is not controlled in a Mendelian way. Genes are silenced through DNA methylation but the pattern of gene expression is different depending on whether the gene comes from the father or mother.

Prader–Willi syndrome and Angelman syndrome are human genetic disorders caused by the same mutation on chromosome 15. Prader–Willi syndrome is a disorder that causes behavioural and cognitive problems, deficiencies in sexual development and obesity. It occurs when a mutation is inherited from the child's father and the gene from the mother is imprinted or silenced. Angelman syndrome occurs when the mutated gene from the mother is active. Although the same mutation is involved, Angelman syndrome causes developmental problems, sleep disorders and hyperactivity, but people with the condition have a normal life expectancy and laugh readily.

3.4.4 Rate of epigenetic change

Epigenetics shows us that gene expression can change in a more complex way than simple changes to the DNA sequence. If epigenetic changes occur in sperm or egg cells, the changes are inherited by offspring, as in the case of the *Avy* gene in mice.

Epigenetic changes occur more rapidly than changes due to natural selection (Section 11.2). The rate of changes, such as DNA methylation, is much higher than rates of mutations transmitted genetically and they are also easily reversed. This provides a way for variation within a species to increase quickly, especially if the environment is rapidly changing. This will be the case for both epigenetic effects within a single generation, as well as those which persist into the next generation. Epigenetic changes may also create new heritable variation that will enable organisms to adapt over a longer period of time. If a population is very small and has little genetic variation then epigenetic variation can help organisms adapt to new or changing environments. Many epigenetic effects are caused by the environment (refer to Section 3.4.5) and influence phenotypes in many ways. If the environment is changing slowly, the environment of a parent may serve to predict what their offspring may encounter.

Recent data also suggest that epigenetic patterns may change during the course of life, so that key genes in vital processes may be affected with age.

3.4.5 Pollution, methyl tags and twin studies

Imprinted genes are very sensitive to environmental signals because they have only a single active copy so any epigenetic changes will have a greater impact on gene expression.

Environmental signals can also affect the imprinting process itself. Imprinting happens during egg and sperm formation, when epigenetic tags are added to silence specific genes. Diet, hormones and toxins can all affect this process and the expression of genes in the next generation.

Chemicals in the environment are being linked to many processes that affect DNA methylation and histone modification. Investigations have identified a range of environmental chemicals that affect epigenetic markers and these include the metals cadmium, arsenic and nickel, and air pollutants including particulates and benzene.

Air pollution from car exhausts is an important cause of breathing problems and other respiratory disorders. High levels of the tiny particles (less than 2.5 μm in diameter) found in exhaust fumes not only irritate the lungs but can also enter the bloodstream and cause inflammation in the body. Inhaling these fine particles has been linked to DNA methylation in the T-helper cells of the immune system. These cells have a key role in our response to inflammation.

THEORY OF KNOWLEDGE

Nature or nurture?

Studies of identical and fraternal twins are used to separate the influences of genes and the environment on particular characteristics. If a characteristic is more common in identical twins than fraternal twins, it is likely that genetic factors are at least partly responsible. This is because identical twins have the same genes, whereas fraternal twins are likely to share only 50% of their genes. Twin studies allow scientists to study the influence of 'nature versus nurture', a phrase that was first used by British scientist Francis Galton. Galton came to realise how important studying twins could be and in 1875 he wrote 'The History of Twins' and tried to quantify the relative effects of nature versus nature on human intelligence. He believed strongly that intelligence is largely inherited and so, to improve humanity, the ablest and healthiest people should be encouraged to have more children. But his ideas were used by eugenicists who took them further and proposed that the human species could be improved by preventing the least able or those with 'undesirable' characteristics from having children at all.

In the early 21st century, Eric Turkheimer, a United States professor of psychology, looked again at the inheritance of IQ and studies involving twins. He noticed that most of the studies that reported that IQ is an inherited characteristic involved twins from affluent homes. When he looked at twins from low-income families, he found that the IQ of identical twins varied just as much as the IQ of fraternal twins. From his research he deduced that income can affect a child's natural intelligence. More recently, a much larger study showed that the relationship between income, genetics and IQ is not straightforward and there are many more variables and influencing factors.

Today twin studies are being used to investigate a range of factors including eating disorders, obesity and sexual orientation.

To consider:

1. To what extent should science be used to provide evidence for complex human characteristics such as intelligence?
2. How difficult is it to assess a person's intelligence?
3. How are such assessments influenced by the background and preconceptions of the assessor?

Monozygotic (identical) twins have the same genomes as they develop from a single fertilised ovum. But in many cases monozygotic twins are not identical in their phenotypes and in the diseases and conditions that affect them. When their genomes are investigated, results show that there are differences in methylation patterns in the twins' DNA. In studies, twins have shown substantial differences in the occurrence of schizophrenia, autism and diabetes, and these differences have been show to be related to methylation patterns of the twins' DNA. Other common diseases that may also be linked to epigenetic effects include heart disease and cancers, which are often influenced by environmental factors such as pollution.

Disease-discordant identical twins make good subjects for studying differences in disease linked to pollution and methylation patterns because their genes are matched and many non-genetic effects, such as their early environment, maternal influences and age, are also the same.

3.4.6 External factors affecting the pattern of gene expression

As well as factors such as TFs inside the nucleus, mRNA degradation and epigenetic factors, many external factors also influence the pattern of gene expression. These include hormones and chemicals which are present in the external environment of the cell and influence what happens inside it.

Hormones are factors produced from outside a cell that can affect the cell's gene expression. Steroid hormones such as estrogen influence the transcription of many genes because they interact with receptors inside cells. Estrogen binds to estrogen receptors in the plasma membrane and from there activates signalling pathways. Steroid hormones pass through the plasma membrane of a target cell and bind to intracellular receptors in the cytoplasm or in the nucleus. The cell signalling pathways induced by the steroid hormones regulate specific genes in the cell's DNA. The hormones and receptor complex act as transcription regulators by binding to the promoter region of the gene, this stimulates RNA polymerase binding and gene transcription and thus increasing or decreasing the synthesis of mRNA molecules of specific genes. This, in turn, determines the amount of corresponding protein that is synthesized. (You can read more about the details hormone activity in section 7.3.1)

In bacteria **biochemical factors** affect gene expression. One example is lactose which affects the expression of genes needed for lactose metabolism in *E coli* bacteria. The *lac* operon of *E. coli* contains genes involved in lactose metabolism. The bacteria are able to break down lactose, but if glucose is present, they prefer to use it as an energy source. Glucose metabolism

involves fewer steps and needs less energy to metabolise. But if lactose is the only sugar available, the *E. coli* will use it instead. The lac operon is only expressed if lactose is present, and glucose is absent.

Two regulators turn the operon on and off in response to lactose and glucose levels. The *lac* repressor acts as a lactose sensor. It usually blocks transcription of the operon but stops acting as a repressor when lactose is present. The other regulator is catabolite activator protein (CAP), which acts as a glucose sensor. CAP will bind to specific DNA sites in or near promoter regions and enhance the ability of RNA polymerase to bind and initiate transcription. It activates transcription of the operon, but only when glucose levels are low. (Fig 3.4.5)

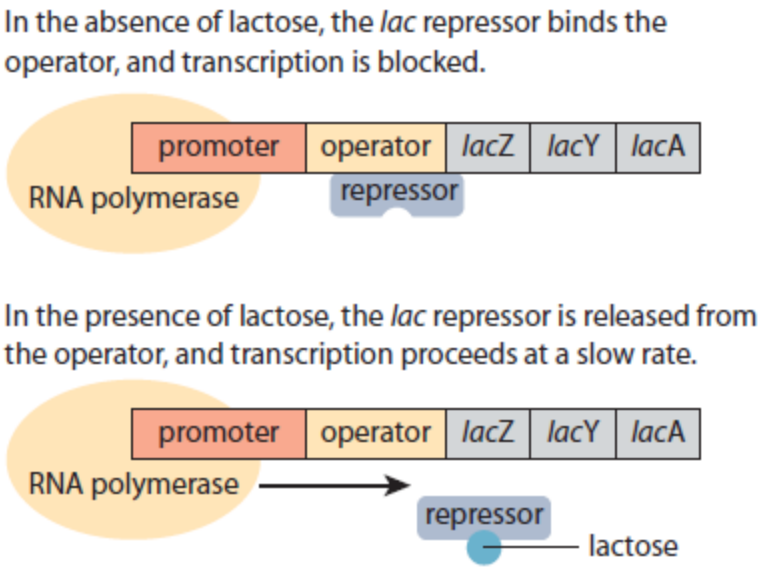


Figure 3.4.5: The *lac* operon.

TEST YOUR UNDERSTANDING

16 Define the term epigenetics.

17 Outline how DNA methylation affects activation of genes.

18 Suggest what effect pollution may have on DNA methylation.

19 Give an example of an epigenetic tag that is not removed from an embryo's epigenome.

20 Why are monozygotic twins useful in epigenetic studies?

REFLECTION

Epigenetics is a new area of scientific research. How much did you know about it before working on this section?

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
explain that DNA replication is semi-conservative and produces two identical molecules	3.1.1			
state the two roles of the enzyme helicase	3.1.1			
outline the function of the PCR and list two examples of its use	3.1.2			
outline the technique of gel electrophoresis and its use	3.1.2			
	3.1.3			

explain the orientation of DNA strands and how DNA polymerases work in a 5′→3′ direction				
distinguish the leading and lagging strands	3.1.3			
describe the process of DNA replication in eukaryotes and the functions of primase, polymerases and ligase	3.1.3			
define and outline the process of transcription	3.2.1			
describe how nucleosomes regulate transcription	3.2.1			
explain the importance of introns and exons	3.2.1			
define translation and the	3.2.2			

importance of complementary base pairing				
explain the role of ribosomes in the formation of polypeptides	3.2.2			
recall that free ribosomes synthesise proteins for use within the cell	3.2.2			
summarise the types and importance of non-coding DNA	3.2.3			
outline the roles of promoter regions and telomeres	3.2.3			
describe a polysome and identify them in electron micrographs	3.2.3			
outline how functional proteins are produced after translation	3.2.3			

state that new alleles are formed by mutation and changes may be harmful, beneficial or neutral	3.3.1			
explain that mutations may add, delete or substitute a base, or invert a section of dna, and that substitution causes sickle-cell disease, addition of repeated sequences leads to Huntington's disease	3.3.1			
describe how the genetic code is degenerate and gives resilience to changes	3.3.1			
state that tumours are groups of cells that grow out of control and may be benign or malignant	3.3.2			

define mutagen and give some examples	3.3.2			
outline the importance of apoptosis	3.3.2			
define epigenetics	3.4.1			
explain how gene expression is regulated by binding proteins	3.4.1			
state that epigenetic changes affect phenotype but not genotype	3.4.2			
describe how epigenetic changes can be due to DNA methylation and modification of histones	3.4.2			
give an example of a heritable epigenetic change	3.4.3			
state that most epigenetic changes are not inherited but	3.4.4			

some can affect the epigenome of the offspring				
recall that epigenetics can cause variation more quickly than natural selection in a changing environment	3.4.4			
outline the importance of pollution in epigenetics	3.4.5			
describe the importance of monozygotic twins in the study of epigenetics.	3.4.5			

REFLECTION

Reflect upon the content of this chapter and identify those areas of strength and weakness in your understanding. How can you improve in those topics you have found difficult?

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Chapter 4

Genetics

D1.3, D2.2, D3.2, D3.3

INTRODUCTION

Genetics is the study of genes and the way in which they are passed from one generation to the next. Every organism has a unique combination of genes making up their own genome. Genes determine all the characteristics of an organism and variations of genes, known as alleles, make each individual different from others members of the species.

4.1 Inheritance

LEARNING OBJECTIVES

In this section you will:

- define the genome is the whole genetic information of an organism
- understand that prokaryotes have one circular chromosome without histones
- learn that eukaryotes have a number of linear chromosomes associated with histones, and are normally enclosed within a nucleus
- recognise that diploid nuclei have pairs of homologous chromosomes, while haploid nuclei have one set of chromosomes
- define a gene as a length of DNA that occupies a specific locus on a chromosome and carries instructions for a specific characteristic
- define an allele as a form of a gene that differs from a corresponding allele by one or a few bases
- learn that homologous chromosomes carry the same sequence of genes but not necessarily the same alleles
- learn that one copy of each pair of homologous chromosomes is inherited from each gamete
- understand that alleles are inherited to form a genotype and interact to form a phenotype

- discover that different eukaryotes have different genome sizes, numbers of chromosomes and numbers of genes. The number of chromosomes is a characteristic of each species
- define an organism's karyotype as the number and type of chromosomes in the nucleus and understand that variations may be the result of whole chromosome mutation
- learn that a karyogram is a picture that shows an organisms' chromosomes
- recognise that most mammals have heteromorphic males (XY) and homomorphic (XX) females. In mammals X inactivation occurs so that one of the X chromosomes becomes and inactive Barr body

GUIDING QUESTIONS

- How is genetic information organised inside cells to carry genetic information for the whole organism?
- How can genes and chromosomes interact to affect the phenotype of an organism?
- What changes to genes and chromosomes cause interactions that lead to the appearance of a new phenotype or the loss of an existing phenotype?

4.1.1 The genome

Chimpanzees are set apart from all other organisms because their parents were chimpanzees and their offspring will also be chimpanzees. Every living organism inherits its own blueprint for life in the chromosomes and genes that are passed to it from its parents. The study of genetics attempts to explain this process of heredity and it also plays a very significant role in the modern world, from plant and animal breeding to human health and disease.

The genome of an organism is defined as the whole of its genetic information and every cell has a complete copy of the organism's genome. Genome analysis is an important field of modern biological research. Comparative genomics is an area that analyses genomes from different species. Genomes are compared to gain a better understanding of how species have evolved and to work out the functions of genes and also of the non-coding regions of the genome. Researchers look at many different features such as sequence similarity, gene location, the length and number of coding regions within genes and the amount of DNA that does not code for proteins. They use computer programs to line up genomes from different organisms and look for regions of similarity. There are many databases that store this information on DNA base sequences (Table 4.1.1). Most are freely available online to scientists and students, so that anyone with internet access can use them to compare DNA and protein sequences. You can find them by typing the database name into any search engine.

Database	Description
ENA	overview of all complete genomes deposited in

Genomes Server	the European Nucleotide Archive
Ensembl	a joint project between the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and the Wellcome Sanger Institute in the UK to produce and maintain automatic annotation of eukaryotic genomes
Ensembl Genomes	provides access to genomes of non-vertebrate species
GenBank (US National Center for Biotechnology Information)	a collection of all publicly available DNA sequences, providing up-to-date and comprehensive DNA sequence information

Table 4.1.1: Some databases that hold information on genomes.

4.1.2 Chromosome structure

Chromosomes are made of DNA molecules and carry the genetic code for each organism but prokaryotic and eukaryotic chromosomes are different in their structure. Prokaryotes have a much simpler chromosome than eukaryotes. Prokaryotes contain a circle of DNA that is often concentrated in one area of the cell, whereas eukaryotes have linear DNA that is associated with histone proteins. Some of these proteins are structural and others regulate the activities of the DNA. Prokaryotes have additional genetic material in the form of small circular structures known as **plasmids**. Prokaryotes are much simpler organisms and so require fewer genes to maintain themselves. The differences between prokaryotic and eukaryotic genetic material are summarised in Table 4.1.2.

Prokaryotic DNA	Eukaryotic DNA
cell contains a circular chromosome, sometimes called a nucleoid	chromosomes are made of linear DNA molecules enclosed in a nucleus, bound by a double membrane
cell contains additional genetic material as small circular plasmids	no plasmids
DNA is ‘naked’ and is not associated with proteins	DNA is associated with histone proteins
cell contains just one circular chromosome	cell contains two or more chromosome types

Table 4.1.2: A comparison of prokaryotic and eukaryotic genetic material.

During the phase of the cell cycle known as interphase, eukaryotic chromosomes are in the form of long, very thin threads, which cannot be seen with a simple microscope. As the nucleus prepares to divide, these threads undergo repeated coiling and become much shorter and thicker (Figure 4.1.1). When stained, they are clearly visible even at low microscope magnifications.

Other aspects of chromosome structure are described in Sections 2.5 and 4.2.

Eukaryotic chromosomes

Eukaryotic species have two or more chromosome types. Their chromosomes form pairs, which are known as homologues. **Homologous** pairs are about the same length and carry the same sequence of genes at the same locations along their length. The form of the genes (alleles) on each of the pair is not necessarily the same because, in sexually reproducing organisms, one chromosome will have been inherited from each of the two parents. So a gene that determines flower colour in a plant would be at the same location on each chromosome in a homologous pair but the allele on the maternal chromosome might not be the same as that on the paternal chromosome.

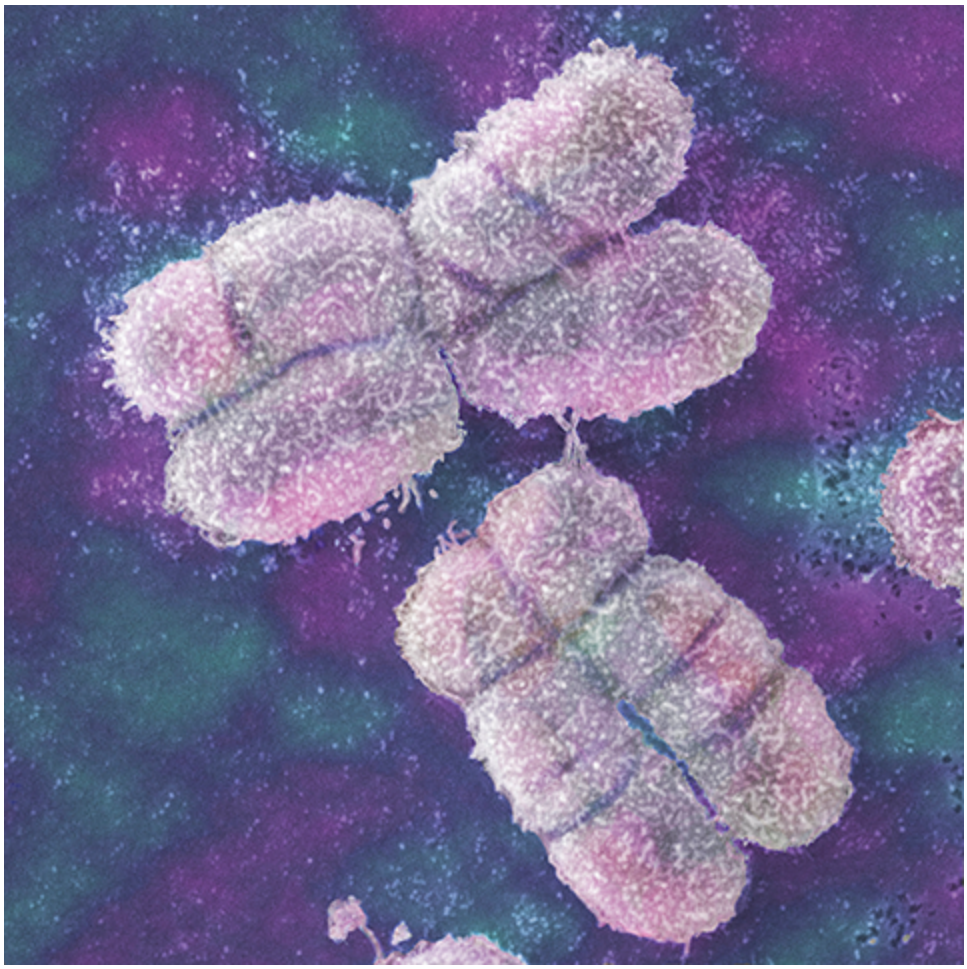


Figure 4.1.1: Coloured scanning electron micrograph of human chromosomes. They have replicated prior to cell division and so consist of two identical copies (chromatids) linked at the centromere (× 7080). The chromatids split at the start of anaphase (Section 6.5) and become individual chromosomes.

Homologous chromosomes are found in the nuclei in the cells of **diploid** organisms. But if each chromosome exists alone with no partner, the cell is said to be **haploid**. Human somatic cells (or body cells) are diploid and contain 46 chromosomes in 23 homologous pairs; human gametes are haploid and contain only 23 chromosomes, one of each pair found in the body cells.

The number of chromosomes found in the cells of an organism is a characteristic feature of that organism. The diploid numbers of chromosomes in some well studied species are shown in Table 4.1.3.

Genome size

Genome size is the total number of nucleotide base pairs in one copy of a single genome. Measurements are often made in numbers of base pairs. It is interesting to note that an organism's complexity is not proportional to its genome size (Table 4.1.4): many organisms have more DNA than humans. Nor is variation in genome size proportional to the number of genes. This is due to the high proportion of non-coding DNA that is found in some organisms.

Organism	**Diploid number of chromosomes**
Canis familiaris (domestic dog)	78
Pan troglodytes (chimpanzee)	48
Homo sapiens (human)	46
Mus muscularis (house mouse)	40
Oryza sativa (rice)	24
Drosophila melanogaster (fruit fly)	8
Parascaris equorum (parasitic worm)	2

Table 4.1.3: The diploid numbers of different species.

Organism	**Genome size**	**Notes**

	in base pairs	
T2 phage (a virus which infects bacteria)	3569	first RNA genome sequenced
Escherichia coli (a bacterium)	4.6×10^6	
Drosophila melanogaster (fruit fly)	130×10^6	
Homo sapiens (human)	3200×10^6	
Protopterus aethiopicus (marbled lungfish)	$130\,000 \times 10^6$	largest known vertebrate genome
Paris japonica (Japanese native pale-petal)	$150\,000 \times 10^6$	largest known plant genome

Table 4.1.4: Genome sizes of some different organisms.

4.1.3 Genes and alleles

The study of genetics aims to explain similarities and differences between parents and their offspring. Genes are lengths of DNA that form part of chromosomes and carry the instructions for the production of a particular protein and the development of specific characteristics. Each gene occupies its own specific position known as a locus, on a chromosome, for example the human insulin gene is located on the short arm of chromosome 11. Chromosome 16 contains more than 90 million DNA base pairs, almost 3% of the total DNA in a human cell. It contains 800–900 genes.

In diploid organisms homologous chromosomes carry the same genes at the same loci, but any gene can occur in one of a number of different forms, known as **alleles**. Alleles differ from one another by a few bases or sometimes a more substantial amount. In sexually reproducing organisms, one of each pair of homologous chromosomes passes to the gametes, which are haploid. At fertilisation alleles are passed on and form part of the genotype of new offspring. As alleles interact they contribute to the phenotype of the new organism. Most characteristics are determined by more than one gene, for example human height and skin colour are determined by many genes.

KEY POINTS

genotype refers to the genetic constitution of an individual, the alleles it contains; each allele is represented by a letter; chromosomes come in pairs and so alleles come in pairs, so a genotype is represented by a pair of letters, such as TT or Tt.

locus is the specific position of a gene on a homologous chromosome; a gene locus is fixed for a species, for example, the insulin gene is always found at the same position on chromosome 11 in humans.

phenotype means the characteristics of an organism that may be physical appearance or biochemical features.

4.1.4 Karyotyping

Chromosomes have unique banding patterns that are revealed if they are stained with specific dyes. These patterns enable us to study the structure and type of chromosomes present in an organism. The technique has been used in prenatal diagnosis and in forensic science. Chromosomes are prepared from the nuclei of dividing cells. If human cells are being studied, a sample of cells may be taken from amniotic fluid, in the case of a fetus, or from a blood sample. Division is halted at the end of prophase when chromosomes are condensed and can be seen (Section 6.5). A **karyogram** is a photograph or diagram of the stained chromosomes. Each chromosome is a characteristic length and each one has a homologous partner. The karyogram image is organised so that each chromosome is separated from the others and they are arranged in order of their size, as shown in Figure 4.1.2.

A karyogram shows the **karyotype** of the cell, that is, the number and types of chromosomes present in its nucleus. It indicates the sex of an individual because it shows the **sex chromosomes** and in prenatal diagnosis it is possible to check for chromosome abnormalities.

KEY POINTS

karyogram a diagram of photograph of the chromosomes from an organism.

karyotype the number and type of chromosomes present in the nucleus.

Some chromosome abnormalities result from non-disjunction, the failure of homologous pairs of chromosomes to separate properly during meiosis (Section 6.5). It results in gametes that contain either one too few or one too many chromosomes. Those with too few seldom survive, but in some cases a gamete with an extra chromosome does survive and after fertilisation produces a zygote with three chromosomes of one type. This is called a **trisomy**.

Trisomy in chromosome 21 results in the human condition known as **Down syndrome** (Figure 4.1.3), a condition in which a child is born with an extra copy of chromosome 21 resulting in physical and mental disabilities. A gamete, usually the female one, receives 24 chromosomes instead of 23 and a baby with 47, instead of the usual 46, chromosomes in each cell is born.

Karyotyping is used when there is concern about potential chromosome abnormalities. Cells from an unborn child are collected in one of two ways: chorionic villus sampling (CVS) or amniocentesis (Section 8.4). The cells are grown in the laboratory and a karyogram is prepared. This is checked for extra or missing chromosomes. The procedure is normally used if the mother is over the age of 35 years, because Down syndrome is more common in babies of older mothers and can be detected using this method (Figure 4.1.3).

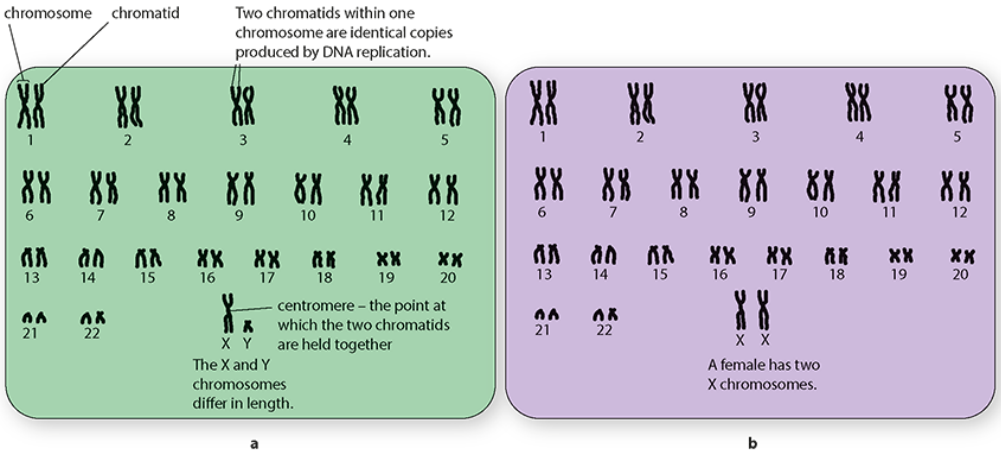


Figure 4.1.2: Karyograms for a human male **a** and a human female **b**.

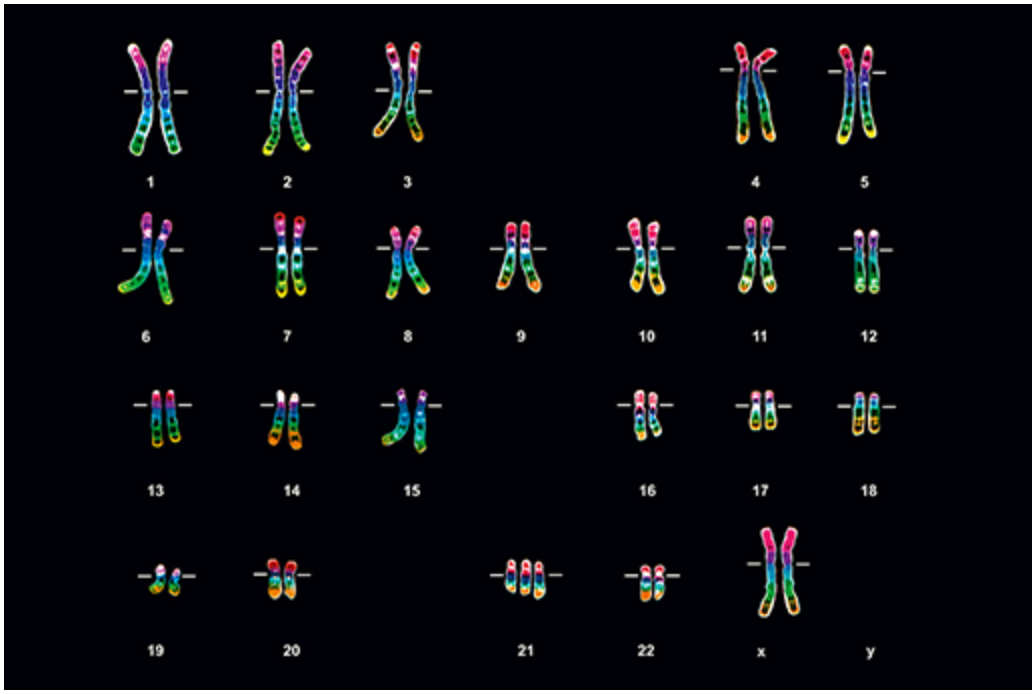


Figure 4.1.3: People with Down syndrome have characteristic physical features. The karyogram, showing chromosomes that are stained and photographed, for a person with Down syndrome shows three copies of chromosome 21 (called trisomy 21).

4.1.5 Determination of sex

Most mammals, including humans, have one pair of chromosomes that determines whether the person or organism is male or female. These sex chromosomes are the last two chromosomes shown in the human karyograms in Figure 4.1.2 (in the centre of the bottom row, in each case). These chromosomes determine the sex of an individual: a human female has two X chromosomes (XX), whereas a male has one X and one Y chromosome (XY). In humans, the X chromosome is longer than the Y and carries more genes (Figure 4.1.4). Sex chromosomes are inherited in the same way as other chromosomes.

The other 44 chromosomes in a human karyotype are known as autosomes and they determine the other characteristics of the individual. (Sometimes, you will see sex chromosomes referred to as the non-autosomal chromosomes.)

THEORY OF KNOWLEDGE

Prenatal screening

Obtaining fetal cells for karyotyping by amniocentesis involves taking a sample of amniotic fluid from the mother between weeks 14 and 16 of her pregnancy. Chorionic villus sampling (CVS) means taking a sample of cells from the chorionic villi, which are the fine projections of the placenta embedded in the lining of the uterus. This can be done at 8–10 weeks into the pregnancy. Both methods carry a small risk of damaging the fetus or even causing a miscarriage. If the results of the test reveal abnormalities, the parents may be offered the option to terminate the pregnancy. The test may

show an abnormality, but it cannot give any indication about the likely severity of the condition.

To consider:

1. Karyotyping is a procedure involving medical and ethical decisions. Who should make the decision to carry out the procedure, the parents or health care staff? How important are legal and religious arguments?
2. Both procedures carry the risk of a miscarriage. How can this potential risk to the unborn child be balanced with the parents' desire for information? What safeguards should be in place when the karyotyping procedure is used?
3. Does the importance of the information that can be obtained from the karyogram outweigh the risk to the unborn child?
4. If the karyogram indicates a genetic abnormality, should the parents be permitted to consider a termination of the pregnancy?

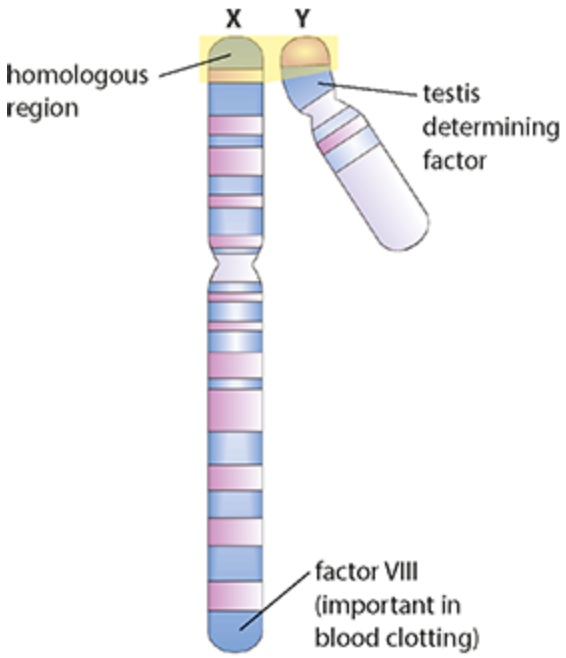


Figure 4.1.4: Human X and Y chromosomes.

In mammals, a process called X inactivation 'switches off' one of the two X chromosomes, which remains as an inactive Barr body. This occurs early in the early development of a female in a random way. A female receives one X chromosome from her male parent and the other from her female parent. If X inactivation occurs at the 100-cell stage of development, 50 cells may have the maternal X chromosome activated while the other 50 have the paternal X chromosome activated.

This pattern of activation will be retained throughout fetal development. Females have a combination of different cells containing either the paternal or maternal X chromosome.

In most cases this does not affect the organism's phenotype, but in tortoiseshell cats it can be clearly seen. A female may have inherited a 'ginger' allele from her mother and a 'non-ginger' allele from her father on her X chromosome. After X inactivation, some of her cells will produce ginger pigment, and

some will produce black. During development all cells specialise and occupy different parts of the cat's body. Some cells will specialise to produce pigment in the skin of the cat. Those carrying the 'ginger' allele will produce orange pigment and those without this allele will produce black pigment. The result is the characteristic appearance of a tortoiseshell cat (Figure 4.1.5).

KEY POINTS

autosome a chromosome that does not determine an organism's sex.

Barr body an inactivated X chromosome found in the cells of human females. Barr bodies appear as darkly stained masses in cells during interphase.

Figure 4.1.5: Tortoiseshell cat fur colour is a result of X chromosome inactivation.

Over evolutionary time non-autosomal chromosomes developed from autosomes to form the sex chromosomes we have today. But different types of non-autosomal chromosomes have evolved in different species. In birds, females are **heteromorphic** and males are **homomorphic**: female birds have ZW and males have ZZ as their sex chromosomes.

KEY POINTS

heteromorphic is a homologous chromosome pair that is not morphologically identical.

homomorphic is a homologous chromosome pair that is morphologically identical.

EXAM TIP

Take time to check you can define similar pairs of terms that appear in the same topic. For example: karyogram and karyotype, heteromorphic and homomorphic, diploid and haploid.

TEST YOUR UNDERSTANDING

1 Define the terms genome, diploid and homologous.

2 Outline the relationship between a locus, a gene and an allele.

3 State two differences between prokaryotic and eukaryotic chromosomes.

4 Suggest how karyograms are useful in prenatal diagnosis.

5 Describe what is meant by X inactivation.

6 Explain the difference between an organism's genotype and its phenotype.

Links

- How does DNA fit into the tiny volume of a nucleus?
- How does chromosome behaviour during cell division lead to variation?
- How does understanding genes help us understand cladistics and classification?

4.2 Genetic inheritance

LEARNING OBJECTIVES

In this section you will:

- learn that the principles of inheritance were first described by Greg Mendel
- recall that gametes are haploid and contain one allele of each gene
- understand that diploid zygotes contain two alleles of each gene
- recognise that alleles represent an organism's genotype; its phenotype is the observable result of the genotype
- learn that dominant alleles mask the effects of recessive alleles. Codominant alleles have combined effects
- recognise that although many genes have multiple alleles but an organism will only inherit one of each from their parents
- learn that polygenic inheritance occurs when two or more genes influence a characteristic. Phenotypes may be continuous or discrete
- understand that due to their location on sex chromosomes, sex-linked genes show different inheritance patterns from autosomal genes. Some sex chromosomes carry genes for genetic disorders

- learn that many human diseases are recessive but some are dominant

  - learn that environment can determine the sex of offspring in some species
  - learn that PKU is a recessive condition caused by a mutation
  - understand that gene loci are linked if they are on the same chromosome
  - learn that unlinked genes segregate independently at meiosis. Linked genes do not.

GUIDING QUESTIONS

- How do recessive conditions contribute to resilience in a population?
- How does a gene pool change over time?

4.2.1 Principles of inheritance

The study of genetics aims to explain similarities and differences between parents and their offspring. Today we discuss genes, chromosomes and DNA but the study of inheritance began in the 19th century with the work of a monk, Gregor Mendel, who lived in what is now Czechia. Mendel knew nothing of DNA or chromosomes, but his studies of plant breeding are crucial to our understanding of inheritance.

Sexual reproduction involves the fusion of two gametes: an egg and a sperm in animals, or pollen and ovule in plants. Each gamete is haploid and contains one allele of each gene of the parent individual. Alleles separate from one another at meiosis (Section 6.5). When two gametes fuse at fertilisation, a diploid zygote with two copies of each allele is produced. The zygote develops to form a new individual, with its own genotype determined by the alleles it has received. The phenotype or appearance of the individual is determined by its genotype.

In the mid-1850s Gregor Mendel began a series of hundreds of painstaking experiments with pea plants, which he grew at his monastery. Mendel chose plants with characteristics he wanted to study – such as those with short or long stems, or with wrinkled or smooth seeds – and he made crosses between them. He pollinated selected plants with characteristics of interest, using pollen from specially chosen plants, and grew new plants from the seeds that were produced. He observed, counted and recorded the different characteristics of all the plants that he grew.

Mendel's early studies involved the inheritance of just one pair of characteristics. He conducted what is now called a **monohybrid cross** using **pure-breeding** plants, which, when

crossed with each other, always produce offspring that resemble their parents. Mendel crossed pure-breeding short pea plants with pure-breeding tall plants and grew seeds that were produced from the crosses (Figure 4.2.1). In the new generation of peas (known as the first filial or **F_1 generation**) he noted that all the plants were tall.

Mendel self-pollinated these tall F_1 plants and grew seeds from the cross to produce a second filial or **F_2 generation**. He discovered that in this generation both tall and short plants were present (Figure 4.2.1). Mendel recorded the results from 1064 plants. He found 787 tall plants and 277 short plants. Approximately three-quarters of the plants were tall or, put another way, the ratio of tall : short plants was 3 : 1.

Mendel carried out other monohybrid crosses with several other characteristics and obtained similar results. These are shown in Table 4.2.1.

The results showed that inheritance did not produce a blending between the characteristics of the parent plants – so, for example, in the first cross there were no plants of medium height – and he described inheritance as being ‘particulate’. We call the inherited particles genes or alleles, but Mendel called them ‘factors’. He understood that factors were transmitted to offspring in gametes.

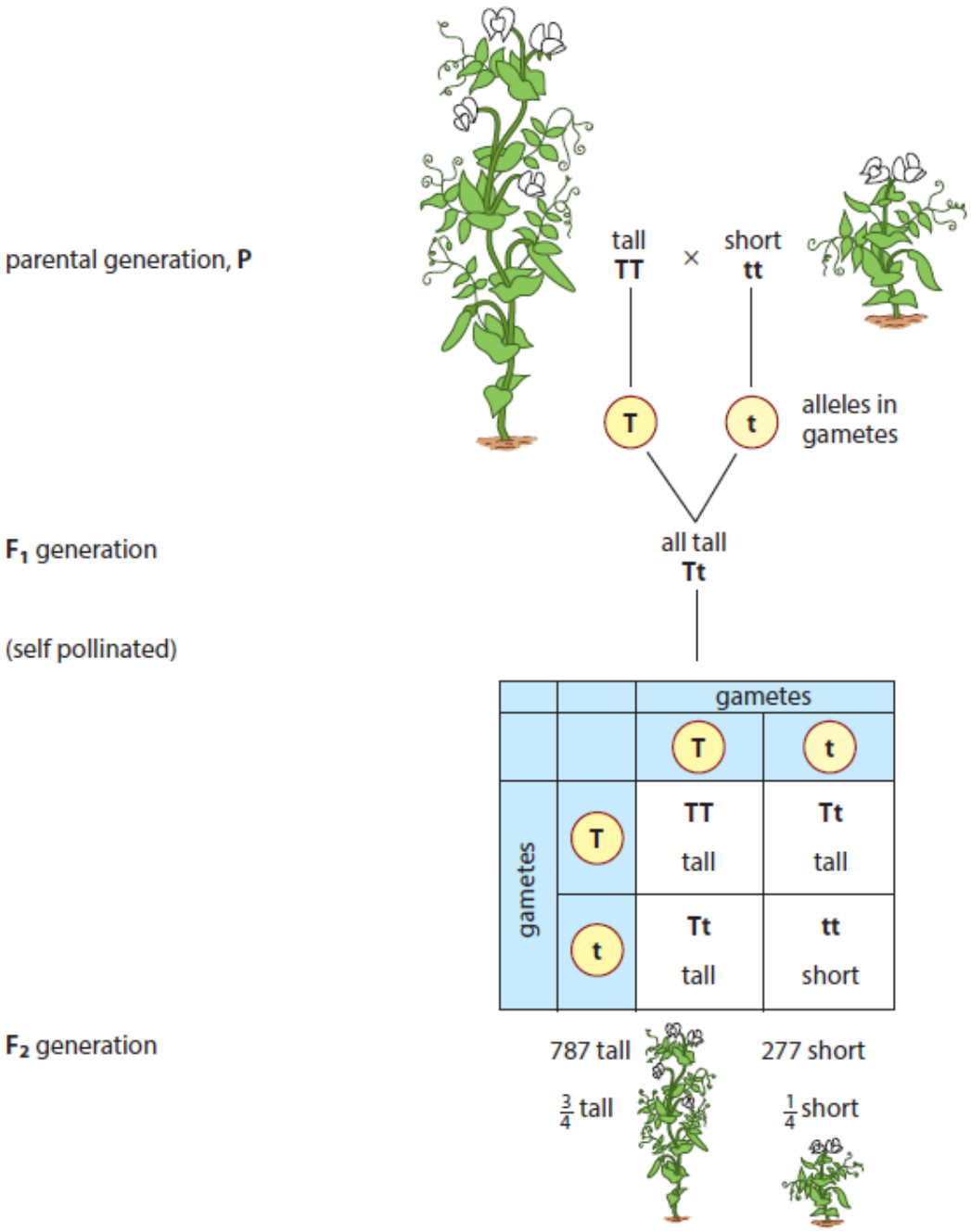


Figure 4.2.1: Diagrams to show Mendel's crosses of tall and short pea plants, with modern knowledge of alleles included.

Characteristic	Cross made	Numbers produced in F_2 generation	Ratio calculated

height of stem	tall × short	787 tall, 277 short	2.84 : 1
petal colour	purple × white	704 purple, 244 white	2.89 : 1
seed shape	smooth × wrinkled	5474 smooth, 1859 wrinkled	2.95 : 1
seed colour	yellow × green	6022 yellow, 2001 green	3.01 : 1

Table 4.2.1: Mendel's experimental results.

Mendel also noted that short plants 'reappeared' in the F_2 generation despite the fact that there were no short plants in the F_1 generation.

Although all the F_1 plants were tall they must contain a 'factor' from their short parent that was 'masked' and did not reappear until the F_2 generation. Today, the factor for tallness is described as being dominant to the factor for shortness, which is said to be recessive. If a dominant and a recessive allele are present together the dominant allele masks the recessive allele.

NATURE OF SCIENCE

Evidence

Mendel used thousands of pea plants in his meticulous experiments. Without this amount of data he would not have been able to establish the ratios he observed with any certainty. If fewer replicates are used in any trial, the degree of uncertainty increases and the results are less reliable. Large numbers of quantitative measurements are more objective

than qualitative observations and today are analysed using statistical methods.

KEY POINTS

dominant an allele that has the same effect on the phenotype when in either the homozygous or heterozygous state; the dominant allele is always given a capital letter, for example, T.

recessive an allele that only has an effect on the phenotype when in the homozygous state; a recessive allele is always given the lower case of the same letter given to the dominant allele, for example, t.

Mendel proposed that factors were transmitted to offspring via gametes and came to the conclusion that gametes contain only one of the factors for height, for example, while the seeds and plants that grow from them contain both. Put another way, we would now say that gametes are haploid and seeds are diploid.

Without knowledge of genes and alleles, Mendel demonstrated that gametes contain one allele of each gene. Nowadays we know that the two alleles of each gene separate into different haploid daughter nuclei during meiosis (Section 6.5). Furthermore, his experiments showed that the fusion of gametes at fertilisation produces diploid zygotes (which in Mendel's experiments developed into seeds) with two alleles of each gene. The reappearance of short plants in the F_2 generation proves that the two alleles may be the same or different. Genetic crosses similar to Mendel's are used by horticulturalists and plant breeders today to develop new varieties of crops and ornamental plants.

4.2.2 Determining allele combinations (genotypes) and characteristics (phenotypes) in genetic crosses

Using a Punnett grid

A genetic diagram called a Punnett grid can be used to work out all the possible combinations of alleles that can be present in the offspring of two parents (the **parental generation**) whose genetic constitutions (genotypes) are known. Punnett grids show the combinations and also help to deduce the probabilities of each one occurring.

When working out a problem, it is helpful to follow a few simple steps.

1 Choose a letter to represent the gene. Choose one that has a distinctly different upper and lower case for the alleles, so for example O, P and W would not be good choices. It is useful to base the letter on the dominant phenotype, so for example R = red could be used for petal colour.

2 Represent the genotype of each parent with a pair of letters. Use a single letter surrounded by a circle to represent the genotype of each gamete.

3 Combine pairs of the letters representing the gametes to give all the possible genotypes of the offspring. A Punnett grid provides a clear way of doing this.

4 From the possible genotypes, work out the possible phenotypes of the offspring.

Worked examples 4.2.1 and 4.2.2 show how to tackle genetics problems using these steps.

Single-nucleotide polymorphisms

Single nucleotide polymorphisms, often called SNPs, are the most common type of human genetic variation. Each SNP represents a difference in a single DNA nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA. SNPs are found throughout a person's genome. They occur approximately once in every 1,000 nucleotides, which means there can be 4 to 5 million SNPs in a person's DNA. The variations are very common and to be classified as a SNP, a variant must be found in at least 1 percent of the population, so far scientists have identified more than 600 million SNPs in populations around the world.

WORKED EXAMPLE 4.2.1

Suppose that fur colour in mice is determined by a single gene. Brown fur is dominant to white. A mouse homozygous for brown fur was crossed with a white mouse. Determine the possible genotypes and phenotypes of the offspring.

Step 1 Choose a letter. Brown is dominant, so let B = brown fur and b = white fur.

Step 2 We are told the brown mouse is homozygous, so its genotype must be BB. As white is recessive, the genotype of the white mouse can only be bb. If a B were present, the mouse would have brown fur.

Step 3 Set out the diagram as shown.

Step 4 The Punnett grid shows that all the offspring will be phenotypically brown and their genotype will be Bb.

Answer

parental phenotypes: brown white

parental genotypes: **BB** **bb**

gametes: (B) (B) (b) (b)

Punnett grid for F1:

		gametes from brown parent	
		(B)	(B)
gametes from white parent	(b)	**Bb** brown	**Bb** brown
	(b)	**Bb** brown	**Bb** brown

KEY POINTS

heterozygous having two different alleles at a gene locus, for example, Tt.

homozygous having two identical alleles at a gene locus; the alleles may both be dominant or both recessive, for example, TT or tt.

WORKED EXAMPLE 4.2.2

Seed shape in the pea plant is controlled by a single gene. Smooth shape is dominant to wrinkled shape. A plant that was heterozygous for smooth seeds was crossed with a plant that had wrinkled seeds. Determine the possible genotypes of the offspring and the phenotype ratio.

Step 1 Choose a letter. Smooth is the dominant trait but S and s are hard to distinguish so use another letter, such as **T**.

Step 2 We are told the smooth plant is heterozygous so its genotype must be **Tt**. Since 'wrinkled' is a recessive trait, the genotype of the wrinkled-seed plant must be **tt**.

Step 3 Set out the diagram as shown, in exactly the same way as before. Notice that, in this case, the smooth-seeded parent produces two different types of gamete because it is heterozygous.

Step 4 Here the Punnett grid shows us that half of the offspring will have smooth seeds with the genotype **Tt** and half will have wrinkled seeds with the genotype **tt**.

Answer

parental phenotypes: smooth wrinkled

parental genotypes: **Tt** **tt**

gametes: T t t t

Punnett grid for F_1:

		gametes from smooth-seed parent	
		(T)	(t)
gametes from wrinkled-seed parent	(t)	**Tt** smooth	**tt** wrinkled
	(t)	**Tt** smooth	**tt** wrinkled

TEST YOUR UNDERSTANDING

7 Define each of these terms:

a genotype

b phenotype

c dominant allele

d recessive allele

e homozygous

8 If red R is dominant to yellow r, state the phenotype of each of these genotypes:

a RR

b Rr

c rr

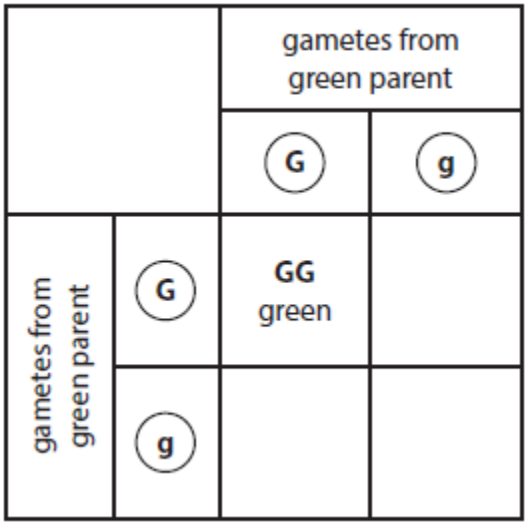
9 State the gametes produced by a parent with each of the following genotypes

a RR

b rr

c Rr

10 Copy and complete this Punnett grid. Green seed colour G is dominant to purple seed colour g. Label the colours of the seeds that are produced.



Most commonly, SNPs are found in between genes. They can act as biological markers, and help scientists to locate genes that are associated with disease. When SNPs occur within a gene or in a regulatory region near a gene, they may affect the gene's function.

The majority of SNPs have no effect on a person's phenotype. But some can help predict an individual's response to certain drugs or the risk of developing diseases. SNPs can also be used to follow the inheritance of some genetic diseases in families.

4.2.3 Codominance and multiple alleles

In Worked examples 4.2.1 and 4.2.2, one of the alleles completely dominates the other, so in a heterozygous genotype the phenotype is determined solely by the dominant allele. In the case of codominance, both alleles from the parents contribute to the phenotype of the offspring. The examples of the mouse coat colour and the smooth and wrinkled peas are both known as monohybrid crosses because they involve just one gene with two alleles: brown **B** and white **b**, or smooth **T** and wrinkled **t**. But in many other cases characteristics are determined by genes which have more than two alleles. Any number of alleles of a gene can be present in the gene pool of a species but any individual can only inherit two of them, one from each parent. One example of this is human blood groups.

The ABO human blood grouping is an example of both **codominance** and multiple alleles. There are three alleles: I^A, I^B and i which determine a person's blood group.

I^A and I^B are both are dominant to i but are codominant to one another. This results in four different phenotypes or blood groups.

KEY POINT

codominance in which a pair of alleles both have an influence on the phenotype when present in the heterozygous state, so that both parental phenotypes are expressed together in their offspring. Codominant alleles are represented in a different way in genetics: a capital letter is chosen to represent the gene and then other (superscript) letters represent the alleles (for

example, in human blood grouping, A and B are codominant alleles and are represented as I^A and I^B).

Genotype	Phenotype or blood group
I^AI^A	A
I^Ai	A
I^BI^B	B
I^Bi	B
I^AI^B	AB
ii	O

Table 4.2.2: Human blood groups and their genotypes.

A person's blood group depends on which combination of alleles they receive. Each person has only two of the three alleles and they are inherited just as though they are alternative alleles of a pair. Table 4.2.2 shows the possible combinations of alleles and the resulting phenotypes.

4.2.4 Incomplete dominance

Incomplete dominance occurs when a dominant allele does not completely mask the effects of a recessive allele and the resulting phenotype is a combination of both alleles. One example is seen in the colour of the flowers of the Marvel of Peru. Flowers with homozygous alleles may be red with alleles RR, or white with the alleles rr. If a red flower is crossed with a white one, neither allele is dominant and the flowers that are produced are pink (Figure 4.2.2).

KEY POINT

incomplete dominance is when the phenotype of heterozygous offspring is a combination of the two homozygous phenotypes, for example, a cross between a red and a white flower producing pink offspring.

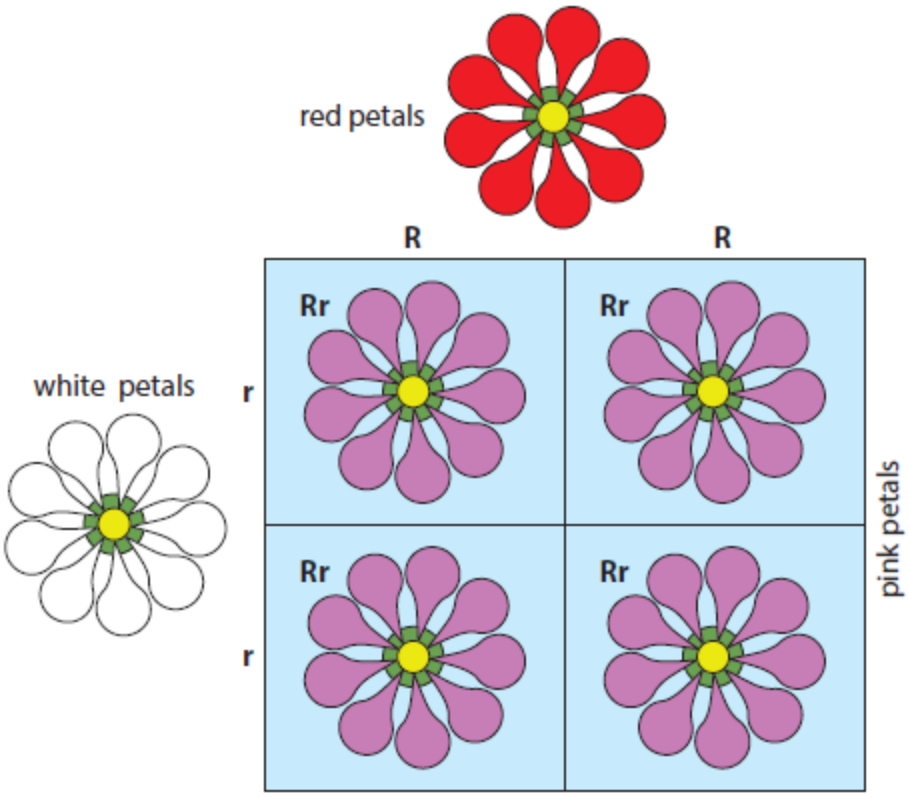


Figure 4.2.2: Red and white petal colour in the Marvel of Peru (*Mirabilis jalapa*) show incomplete dominance so a cross between red and white flowers produces pink offspring.

Genotype or environment?

Sometimes a person's phenotype is determined only by their genotype but in other cases their environment plays a part. For example, a farmer knows the effect that fertiliser, water and pesticides have on the growth of plants even when all the plants have the same genotypes. It is often difficult to assess how much variation is due to genes and how much to environmental factors.

Here are some examples of human variation. Can you identify which are determine by genes, which by the environment and which by a combination of both?

- Blood group

- Eye colour
- Freckles
- Curliness of hair
- Interest in music
- Sense of humour
- Hair colour
- Shape of earlobes
- Mass
- Height

EXAM TIP

Note that incomplete dominance is different from codominance. In incomplete dominance the phenotype is new and somewhere between the two homozygous phenotypes, but in codominance both parent phenotypes are expressed at the same time in offspring.

4.2.5 Sex chromosomes and autosomes

Humans have one pair of chromosomes that determine whether the person is male or female. These chromosomes are called the sex chromosomes. Each person has one pair of sex chromosomes, either XX or XY, along with 22 other pairs known as autosomes. The X chromosome is longer than the Y and carries more genes. Human females have two X chromosomes and males have one X and one Y.

Sex chromosomes are inherited in the same way as other chromosomes.

The ratio of phenotypes female : male is 1 : 1. This means, at fertilisation, there is always a 50% chance that a child will be a boy and 50% that it will be a girl.

Sex chromosomes and genes

The sex chromosomes not only carry the genes that control sex, the X chromosome also carries genes called sex-linked or X-linked genes. These genes occur only on the X chromosome and not on the Y chromosome, which is much shorter. The Y chromosome carries alleles that are mainly concerned with male structures and functions.

Sex linkage has a significant effect on genotypes. Females have two X chromosomes, so they have two alleles for each gene and may be homozygous or heterozygous. In a female, a single recessive allele will be masked by a dominant allele on her other X chromosome. Males only have one allele on their X chromosome with no corresponding allele on the Y chromosome, so a recessive allele will always be expressed in a male.

A female who is heterozygous for a sex-linked recessive characteristic that does not affect her phenotype is called a **carrier**.

KEY POINT

carrier an individual with one copy of a recessive allele that causes a genetic disease in individuals that are homozygous for this allele.

Examples of sex-linked characteristics

Two examples of sex-linked human characteristics are hemophilia and red–green colour blindness.

Hemophilia is a condition in which the blood of an affected person does not clot normally. It is a sex-linked condition because the genes controlling the production of the blood-clotting protein factor VIII are on the X chromosome. A female who is X^HX^h will be a carrier for hemophilia. A male who has the recessive allele X^hY will have hemophilia. Figure 4.2.3 is a pedigree chart showing how a sex-linked condition like hemophilia may be inherited. Notice that hemophilia seldom occurs in females, who would have to be homozygous for the recessive allele X^hX^h. This condition is usually fatal before birth, resulting in a miscarriage. Today, hemophilia is treated by giving the affected person the clotting factor they cannot produce.

EXAM TIP

When you write diagrams for alleles that are on the X chromosome make sure you use the standard format required for your exams. Use a superscript letter on an upper case X, like this: X^BX^b

A person with red–green colour blindness has difficulty distinguishing between red and green. Red–green colour blindness is inherited in a similar way to hemophilia. A female who is X^BX^b is a carrier for colour blindness and a male with just one copy of the recessive allele will be colour blind. Remember that a male cannot be a carrier for a sex-linked gene.

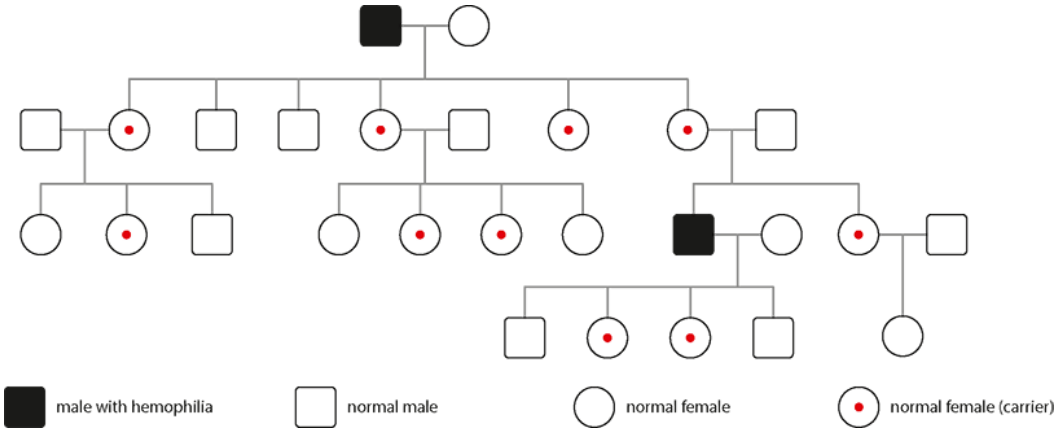


Figure 4.2.3: Pedigree for a sex-linked recessive disease, such as hemophilia.

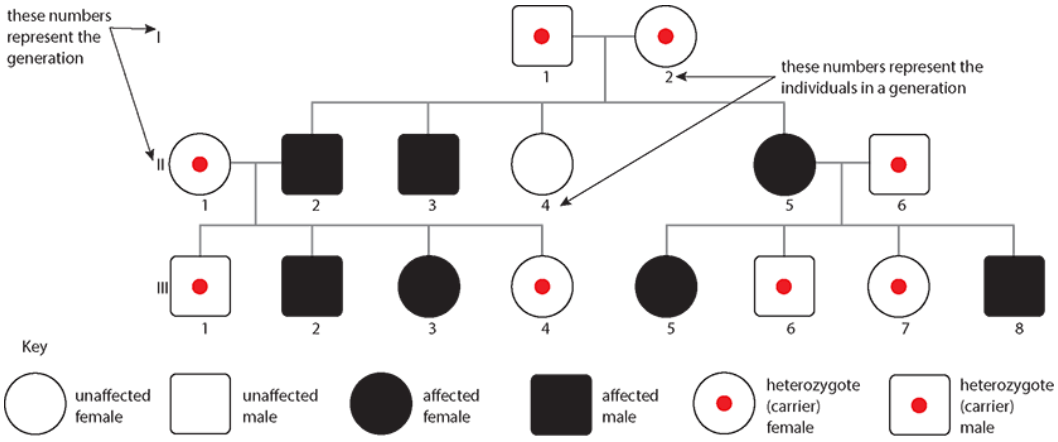


Figure 4.2.4: This pedigree chart shows the occurrence of an autosomal recessive genetic condition cystic fibrosis in a family.

4.2.6 Pedigree charts

Pedigree charts, like the ones shown in Figures 4.2.3 and 4.2.4, are a way of tracing the pattern of inheritance of a genetic condition through a family. Specific symbols are always used, and the chart is set out in a standard way. The horizontal lines linking the male and female in a generation indicate a marriage (or mating) and the vertical lines indicate their offspring. Offspring are shown in the order of their birth. For example, in the family shown in Figure 4.2.4, the oldest individual affected with the genetic condition is II2, who is a male.

4.2.7 Genetic diseases

As well as the inherited genetic conditions that are carried on the sex chromosomes and inherited with the sex of an individual, there are many other inherited disorders that occur due to mutations in single genes on the autosomal chromosomes (numbers 1–22). Most are caused by the presence of two recessive alleles, although some – such as Huntington's disease – are caused by a dominant allele (Table 4.2.3). Genetic 'diseases' are unlike other diseases such as a cold or flu. They cannot be 'caught' from another person but they may be passed through families from parent to child. More than 6000 physiological 'diseases', caused by mutations to single genes, are known. Most are very rare (Table 4.2.3) and there is a wide variation in the occurrence of some genetic disorders between different racial groups and geographic locations of people or their ancestors.

WORKED EXAMPLE 4.2.3

A female who is homozygous for normal vision married a male who is red–green colour blind. Determine the possible types of vision inherited by their two children, one girl and one boy.

Step 1 Standard letters are used for these alleles – normal vision is $\mathbf{X^B}$ and colour blind is $\mathbf{X^b}$. The X is always included.

Step 2 The female is homozygous for normal vision so her genotype must be $\mathbf{X^BX^B}$.

Since the male is colour blind, his genotype must be $\mathbf{X^bY}$.

Step 3 Set out the diagram as shown.

Step 4 The Punnett grid shows that a daughter will have normal vision, but be a carrier for red–green colour blindness. A son will have normal vision.

Answer

parental phenotypes: female male

parental genotypes: X^BX^B X^bY

gametes: X^B X^B X^b Y

Punnett grid for F_1:

		gametes from the male	
		X^b	Y
gametes from female	X^B	X^BX^b girl, normal vision, carrier	X^BY boy, normal vision
	X^B	X^BX^b girl, normal vision, carrier	X^BY boy, normal vision

TEST YOUR UNDERSTANDING

11 A person is blood group A. What are the possible combinations of alleles that this person could have? Use the proper notation for your answer.

12 What is meant by the term ‘sex-linked condition’?

13 Write down the combination of alleles that a female with hemophilia would have.

Phenylketonuria

Phenylketonuria (PKU) is an error in metabolism that leads to decreased metabolism of the amino acid phenylalanine. It is an autosomal recessive condition caused by a mutation in the gene on chromosome 12 that codes for the liver enzyme tyrosine hydroxylase. This enzyme converts phenylalanine into another amino acid, tyrosine. Phenylalanine is essential for normal growth but, if too much builds up the blood, it can cause brain damage and lead to intellectual disability, seizures and other mental disorders. The condition can be treated if it is diagnosed soon after birth. Babies in many parts of the world are given a simple blood test to check for PKU. Babies identified as having the condition must be given a special low-protein diet containing low amounts of phenylalanine. They must avoid many high-protein foods such as milk and dairy products, nuts, fish and meat. PKU only affects children until puberty, after which they can have a non-restricted diet.

Genetic disease or condition	Frequency	Cause	Gene location (chromosome number)
Cystic fibrosis (CF)	Variable. About 1 in 25 people of a white ethnic background are carriers of the disease. In Europe, about 1 in 2000	Autosomal recessive allele; more than 500 different mutations of the CF gene have been found.	7

	babies are born with CF. In the USA, the incidence is 1 in 3500.		
β Thalassemia	Most common in Asia, Middle East and Mediterranean areas where malaria was or is endemic. 1% of people here may be affected. α and β Thalassemias are the world's most common single-gene disorders.	Autosomal recessive allele of the gene that codes for the β chain of hemoglobin.	11
Sickle-cell disease	Most common in people whose ancestors come from sub-Saharan Africa but also in South and Central	Autosomal recessive mutation, which causes substitution of one amino acid in the β chain of hemoglobin.	11

	America, Mediterranean countries and India. In West Africa, 1% of the population are affected.		
Huntington's disease (HD)	Rare disease affecting about 1 in 20 000 people. Males and females are equally affected and the disease occurs in all ethnic and racial groups.	Autosomal dominant mutation of the HD gene, which increases the length of a repeated sequence of CAG.	4
Phenylketonuria (PKU)	Rare condition affecting about 1 in 25,000 people world wide.	Caused by mutation in the gene that codes for the enzyme needed to convert phenylalanine to tyrosine. Phenylalanine can build up in the brain and blood and	12

		lead to brain damage. Can be detected in new born babies by a blood test and treated with special diet.	

Table 4.2.3: Some information on five well documented genetic diseases that affect human metabolism (data from the World Health Organization).

PKU is a recessive condition, which means that a child must inherit a copy of the recessive allele from both parents (Figure 4.2.5).

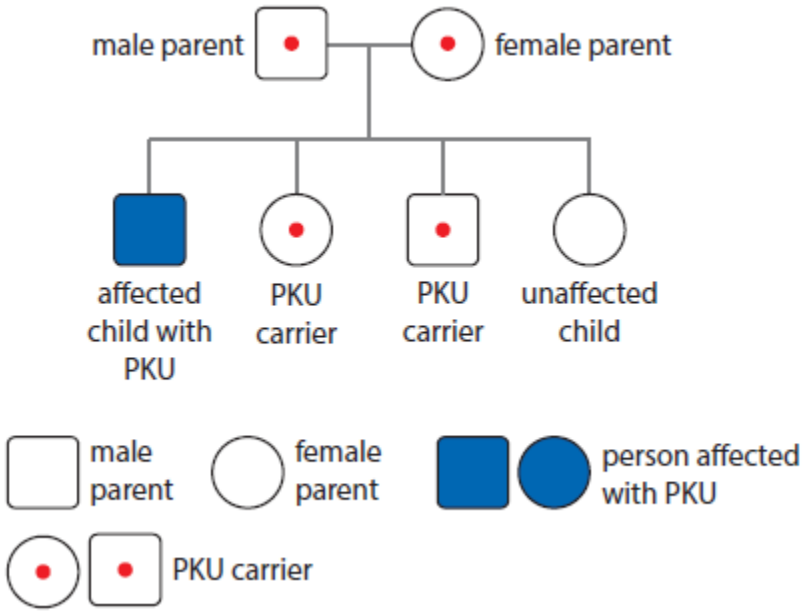


Figure 4.2.5: A person with PKU must inherit a recessive allele from both parents. Their siblings may either be carriers or

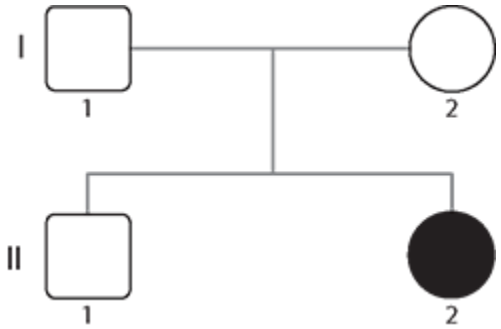
unaffected by the condition.

WORKED EXAMPLE 4.2.4

Cystic fibrosis (CF) is a genetic disorder that causes the excessive production of thick sticky mucus. It is due to a recessive allele that is not sex linked. The pedigree chart shows two generations of a family. A filled-in symbol represents an individual who has cystic fibrosis.

Deduce the genotypes of the parents I1 and I2.

Deduce the probability that II1 is heterozygous.



Step 1 Cystic fibrosis is a recessive disorder. The 'normal' condition, without cystic fibrosis, is dominant. So choose **N** to represent the normal allele.

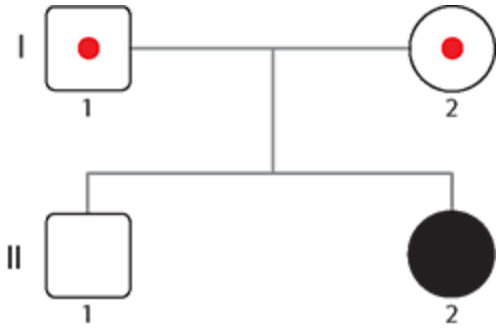
Step 2 Neither of the parents, I1 and I2, have cystic fibrosis so both must have at least one normal allele **N**. As cystic fibrosis is recessive and II2 has the condition, she must have the genotype **nn**.

II2 received one allele from each of her parents so both of them must have passed one n allele to her. Both parents must have one **n** but they do not have cystic fibrosis so their genotype must be heterozygous **Nn**.

The pedigree chart could now be redrawn to show that the parents are heterozygous carriers.

Step 3 Now that both parents are known to be heterozygous, a Punnett grid can be drawn.

Step 4 Person II1 does not have cystic fibrosis and so could have the allele combination shown by any of the shaded boxes in the grid. The probability of person II1 being heterozygous is 2 out of 3, or $\frac{2}{3}$ or 66%.

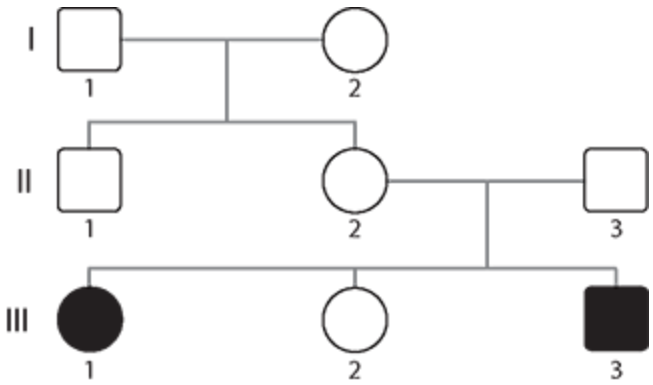


Answer

		gametes from I 1	
		N	n
gametes from I 2	N	**NN** normal	**Nn** normal
	n	**Nn** normal	**nn** II 2 with CF

WORKED EXAMPLE 4.2.5

The pedigree chart shows the family history of a recessive human condition called woolly hair. A filled-in symbol indicates that the person has woolly hair. Deduce whether this condition is sex linked or not.



Step 1 Remember that in a sex-linked condition, the allele occurs only on the X chromosome and males only have one X chromosome.

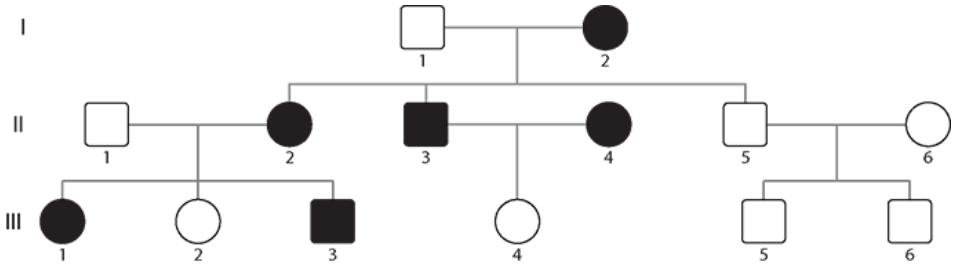
Step 2 Using N to represent the condition, we can see that female III1 must be nn as she has the condition and thus has inherited one n from each parent.

Step 3 If woolly hair is not sex linked, both her parents would be Nn as they have non-woolly hair.

Step 4 If it is sex linked, her mother (II2) would be $\mathbf{X^NX^n}$ and her father (II3) would be $\mathbf{X^nY}$. This would mean he has the recessive allele and no dominant allele. If the condition is sex linked, he would have woolly hair, which he does not. This proves that it is not sex linked.

WORKED EXAMPLE 4.2.6

The pedigree chart in this example shows the inheritance of a particular genetic condition in a family. A filled-in circle or square means that the individual is affected – that is, shows the genetic condition. The condition is not sex linked. Deduce whether the characteristic is dominant or recessive.



We can see that the genetic condition is dominant for the following reasons:

- affected individuals occur in every generation
- every affected individual has at least one affected parent.

THEORY OF KNOWLEDGE

Ethics of genotype identification

Sequencing of the bases of the human genome was completed in 2003 and now the task of identifying all of the genes is on going. As these genes are found, it may be possible for a person to be screened for particular alleles of genes that could affect them, for example, by increasing their susceptibility to cancer or the likelihood that they will develop Alzheimer's disease in later life. Alleles for genetic diseases such as

Huntington's disease, whose onset is typically later in life, can already be identified.

To consider:

1. Does simply knowing the sequence of the 3 billion base pairs of the human chromosomes tell us anything about what it means to be human?
2. Should third parties such as health insurance companies have the right to see genetic test results or demand that a person is screened before offering insurance cover or setting the level of premiums?
3. If treatment is unavailable, is it valuable to inform a person that they carry a genetic condition?
4. Knowledge of an individual's genome has implications for other members of their families. Should their rights be protected?

SCIENCE IN CONTEXT

Why is marriage between close relatives forbidden in many societies?

A marriage between two closely related people, with common grandparents is known as a consanguineous marriage and defined as a union between two individuals who are related as second cousins or closer. Cousin marriages were once common but have declined since the end of the 19th century in Western countries but in other parts of the world many such marriages still take place to preserve family property, cultural values and family structure. If the partners in a marriage are closely related it is more likely that they will carry the same

recessive genes and therefore be more likely to pass them on to their offspring. Even if there is no known genetic disorder in a family, first cousin marriages are generally at double the risk for birth defects in the offspring than the risk in unrelated people.

In some parts of the world including China, North and South Korea and many of the states of the USA cousin marriage is prohibited. Different religions have different rules on marriage; in Roman Catholicism marriages more distant than first-cousins are permitted, whereas protestant churches and Islam permit cousin marriage. Cousin marriage increases the burden of genetic diseases in a society, especially if cousin marriages have taken place over several generations. Genetic education and genetic counselling programmes to make people aware of potential problems are some of the ways in which societies can help to reduce the number of children born with harmful genetic conditions.

To consider:

How does a society balance the freedom of individuals to select their marriage partners with the potential health care needs of their children?

KEY POINTS

continuous variation refers to variation that does not fall into categories, there is a range of measures from one extreme to another. An example is human height.

discrete variation (discontinuous variation) is variation that is clearly distinguishable, categories of a feature can be

identified and for which there are no intermedites. Examples include blood groups.

4.2.8 Polygenes

In the genetic examples considered so far, a particular characteristic is controlled by one gene, which can have different alleles at a specific locus on a pair of chromosomes. There is a clear difference between organisms with different alleles. An organism either has the characteristic or it does not – there are no intermediate forms. These phenotypes are examples of discrete (or discontinuous) variation.

But very few characteristics are controlled by single genes; most are controlled by groups of genes, which together are known as polygenes. The genes that form polygenes are often located on different chromosomes and known as unlinked genes. When two or more genes, each with multiple alleles, are responsible for a characteristic the number of possible phenotypes is greatly increased. Each gene separately may have little impact but their combined effect produces a whole variety of phenotypes. Unlinked polygenes result in a range of degrees of the characteristic from one extreme to another – that is, continuous variation. Human height and skin colour are two examples of continuously varying characteristics. The average heights of males and females will be different in different countries and in different populations but all of them will show continuous variation as shown in Fig 4.2.6.

Human skin colour

Human skin colour is determined by the amount of the pigment melanin that is produced in the skin. Melanin synthesis is controlled by genes. The degree of pigmentation can range from the very dark skin of people originating from regions such as

Namibia in southern Africa, through to the very pale skin of native Scandinavian people.

Melanin protects the skin from the harmful UV rays from the Sun. In parts of the world close to the equator, the Sun's rays are particularly intense so people need more protection from sunburn. Dark-skinned people have a high concentration of melanin, which protects them, while fair-skinned people have much less. Although skin colour is genetically determined, environmental factors also influence it. Fair-skinned people who are exposed to sunlight produce extra melanin and develop a protective suntan. Exposure to sunlight also allows vitamin D to be produced in the skin.

Several genes are involved in determining skin colour and they produce the almost continuous variation that can be seen in the global human population. In Figure 4.2.6 only three genes – A, B and C – are shown. Each gene has two alleles: Aa, Bb and Cc. The Punnett grid shows the possible combinations of skin colour in children from two parents, both heterozygous for all three genes. The parents' phenotype is light brown skin (3).

In this simplified example there are only seven categories of pigmentation but you can see that if the frequencies of the different skin colours in the Punnett grid are plotted on a histogram, as in Figure 4.2.6, it produces a normal distribution. In the case of human skin colour, it is known that more than three genes are involved and the number of categories exceeds seven. The result is a wider distribution curve and more 'continuous' variation.

Other environmental factors can also affect human skin colour. Paler-skinned people will naturally produce more melanin in their skin to protect them from UV rays when they are exposed

to intense sunlight for a period of time. But this is a temporary effect and their skin colour will return to its genetically determined colour when they are no longer exposed to the sun.

Human height and body mass are also examples of polygenic characteristics, but both can be influenced by environmental factors, such as nutrients in a person's diet or the quantity of food that they consume.

4.2.9 Variation in phenotypes without change to genotype

Phenotypic plasticity

Phenotypic plasticity is defined as the ability of individual genotypes to produce different phenotypes when exposed to different environmental conditions. Polyphenism is a type of phenotypic plasticity by which some species adapt to drastic and recurrent changes in the environment such as seasonal alternation in temperate and tropical regions. Polyphenism is the process by which two or more different phenotypes are produced from the same genotype. Seasonal polyphenism in butterflies allows many species can change the colour of their larvae, pupae or adult forms in response to changes in temperature, day length or humidity, for example the squinting bush brown butterfly (*Bicyclus anynana*) of eastern Africa are dull brown in colour during the dry season but in the wet season butterflies with large eyespots develop. The eyespots help to defend the insects from bird predators. Another example is caste polyphenism in social insects which allows species such as ants to develop into either a reproductive queens or a sterile worker ants. In this case females with highly similar genomes look and behave very differently. Many other insect species divide their life stages between larval feeding and growing stages and adults which reproduce and disperse. Among the vertebrates, the sex of offspring in reptiles such as crocodiles is determined by the temperature at which eggs are incubated.

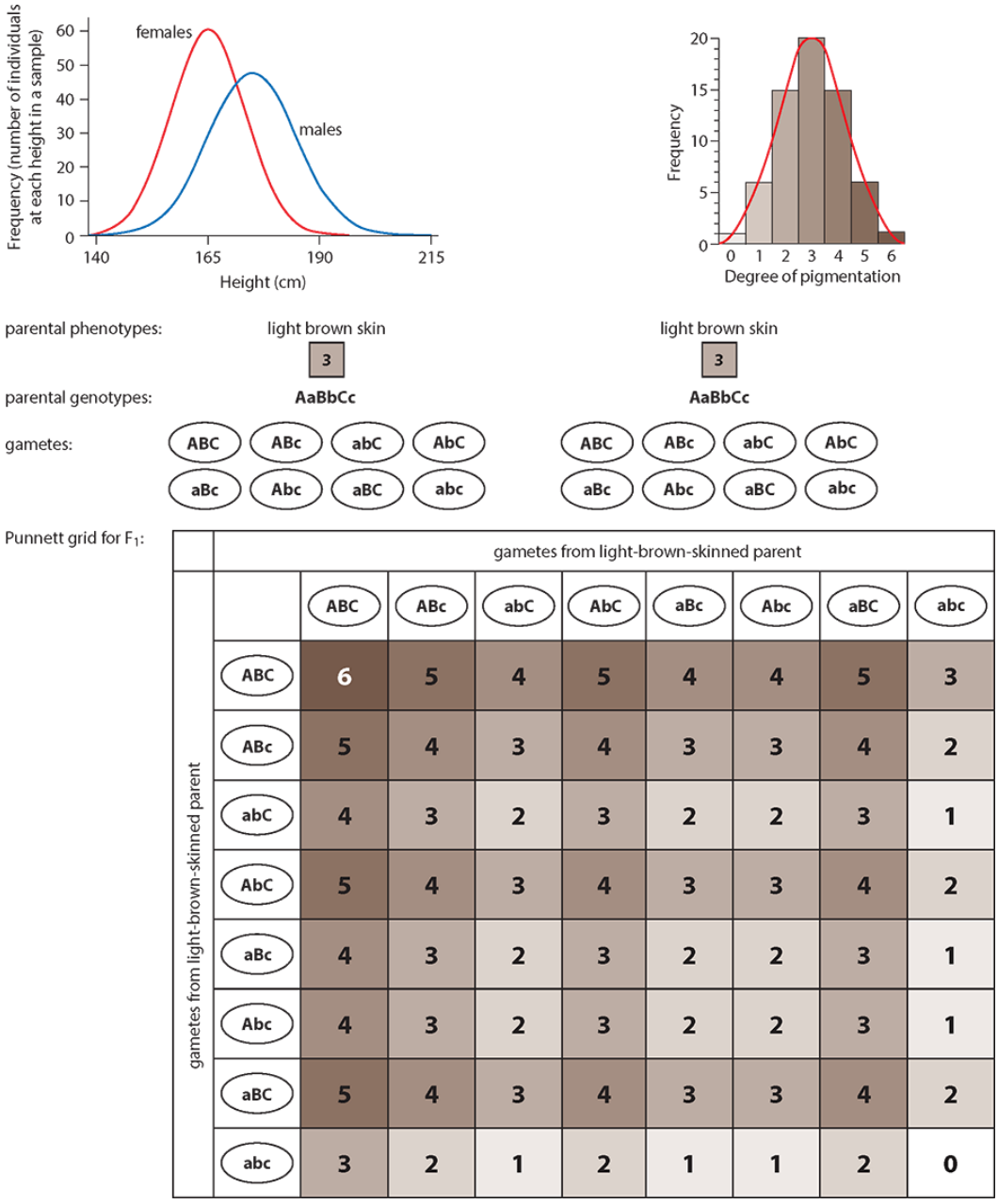


Figure 4.2.6: Top, human height demonstrates continuous variation – when frequency is plotted against height, a normal distribution is obtained. Bottom, frequency of skin variation shown on a bar graph and Punnett grid show the possible combinations of skin colour in children from two heterozygous parents.

Factors affecting sex determination

In some species, sex is influenced not only by genotype at conception but also by the environment that offspring experience during their early development. **Environmental sex determination (ESD)** probably provides a means for the species to adapt when seasonal variations in environmental conditions give one sex an advantage over the other. Temperature, location, nutrient availability and **photoperiodism** are four environmental factors that have been shown to affect the sex of offspring in a number of species.

Temperature

The sex of all crocodiles and most turtle species is determined by the temperature that their eggs experience after fertilisation. The temperature at a critical point in their development is crucial and small changes can cause a dramatic difference in the sex of offspring that hatch. In many species males and females only hatch from the same clutch of eggs over a small range of temperatures. Figure 4.2.7 shows the percentage of male animals produced for three species, the pond slider turtle (*Trachemys scripta*), the freshwater alligator snapping turtle (*Macroclemys temminckii*) and the American alligator (*Alligator mississippiensis*). The alligator snapping turtle produces a higher proportion of males in a clutch up to a temperature of 26 °C but the percentage decreases as the temperature rises beyond this. The American alligator's eggs all develop into male animals when the temperature is 33 °C, whereas the pond slider turtle eggs all hatch as male between 26 and 28 °C.

KEY POINTS

environmental sex determination (ESD) refers to a mechanism by which sex is determined in some species, not only by genotype, but also by the environment that offspring experience in early development.

photoperiodism is the reaction of organisms to changes in day length.

polygenic inheritance is the inheritance of a characteristic controlled by two or more genes.

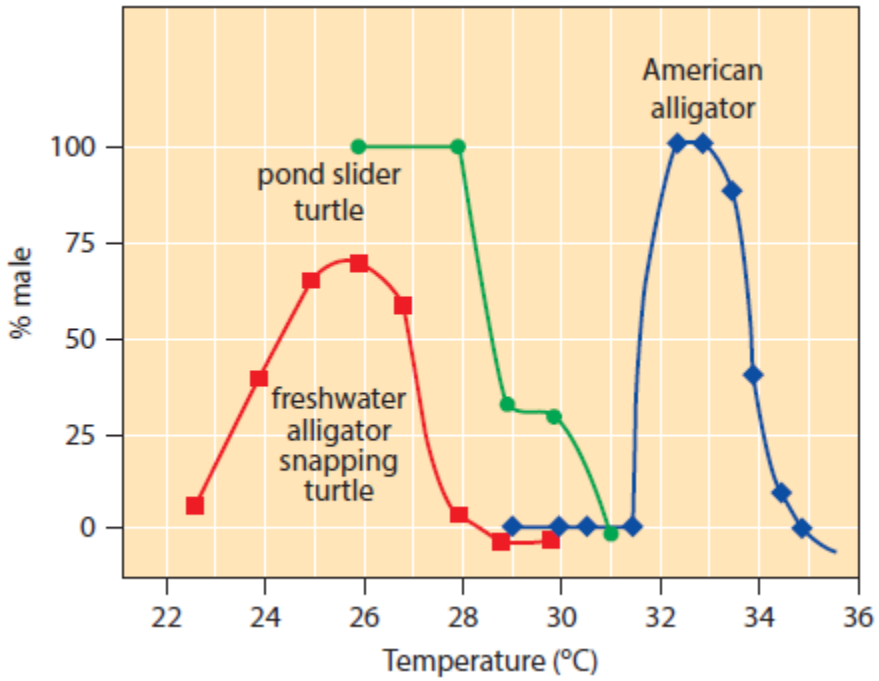


Figure 4.2.7: Temperature-dependent sex determination in three reptile species.

Location

Location is also a factor in determining the sex of the common slipper limpet *Crepidula fornicata*. In this species, individuals pile up on top of one another to form a small tower (Figure

4.2.8). Young individuals are always male but their reproductive systems degenerate after the early stages of their development. In the next phase of their life the individuals can become either male or female, and this will depend on their position in the pile. If the snail is attached to a female, it will become male. But if a snail is removed from its attachment, it will transform into a female. If there are too many males present, some will become females, but once an individual has become female it will not change again.

Figure 4.2.8: Common slipper limpets attach to one another to form piles which determine their sex.

4.2.10 Dihybrid crosses and linked genes

Single genes and monohybrid (single gene) genetic crosses produce variation, while multiple alleles, such as those that control blood groups, increase the possible variety still further. But other dihybrid crosses involve more than one gene. Single genes alone control very few characteristics and, when several genes control a characteristic, the phenotype is determined by their combined effect.

During meiosis (Section 6.5) the two alleles of every gene separate in a process called segregation. Genes found on separate chromosomes segregate independently of the alleles of other genes. Genes found on the same chromosomes are linked genes and so do not segregate independently. In dihybrid crosses the inheritance of two genes is investigated together.

Mendel's law of independent assortment

Gregor Mendel formulated a 'law of independent assortment' that states:

- When gametes are formed, the separation of one pair of alleles into the new cells is independent of the separation of any other pair of alleles.

Or:

- Either of a pair of alleles is equally likely to be inherited with either of another pair.

The dihybrid cross

A **dihybrid cross** involves two pairs of genes on different chromosomes instead of just one pair on one chromosome, but

the principles of setting out a genetic cross diagram to predict the offspring that will be produced are exactly the same as that for the monohybrid crosses. The genetic diagrams should include parental phenotypes, parental genotypes, gametes in circles and a Punnett grid for the F_1 or F_2 generation.

KEY POINTS

dihybrid cross is a cross involving two pairs of genes on different chromosomes.

linked genes genes which have gene loci on the same chromosome.

linkage group in genetics, the genes carried on one chromosome that do not show random or independent assortment.

EXAM TIP

Remind yourself of the correct way of writing dominant, recessive and codominant alleles in genetics problems.

Mendel carried out dihybrid crosses and an example is shown as one of the following Worked examples.

Linkage

The human genome contains between 25 000 and 30 000 genes, but there are only 23 pairs of chromosomes. This means that each chromosome must carry very many genes. Chromosome 1 contains over 3000 genes but the much smaller chromosome 21 contains only around 400 genes.

Any two genes with loci on the same chromosome are said to be linked. Linked genes are usually passed on together. The genes on any chromosome form a linkage group, so a human has 23 linkage groups. The difference between unlinked and linked genes is shown in Figure 4.2.9.

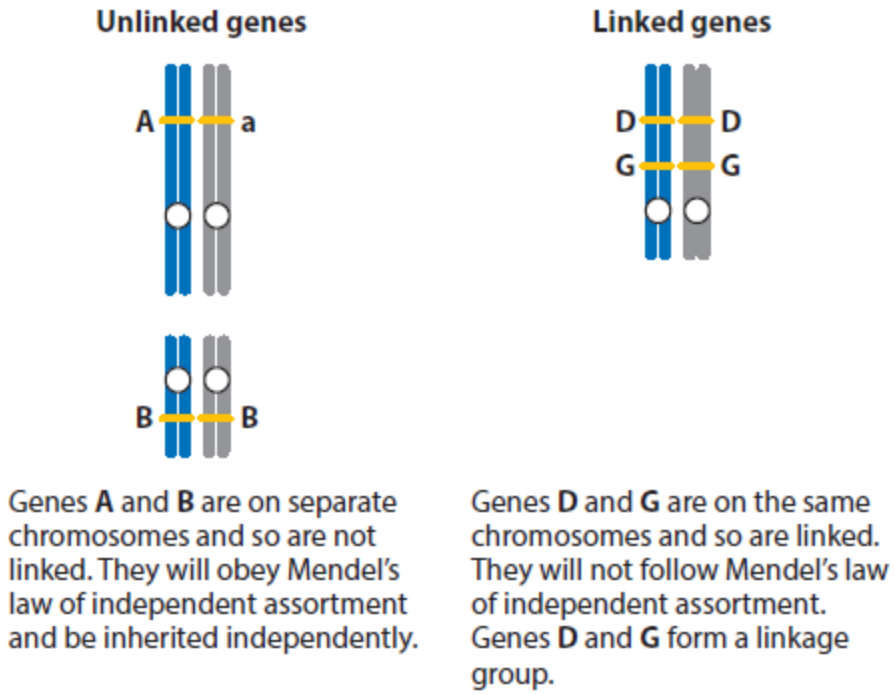


Figure 4.2.9: The difference between unlinked and linked genes.

WORKED EXAMPLE 4.2.7

Fur colour in mice is determined by a single gene. Brown fur is dominant to white. Ear size is also determined by a single gene. Rounded ears are dominant to pointed ears.

A mouse homozygous for brown fur and rounded ears was crossed with a white mouse with pointed ears. Determine the possible phenotypes and genotypes of the offspring.

Step 1 Choose suitable letters to represent the alleles. Brown fur is dominant, so let **B** = brown fur and **b** = white fur. Rounded ears is dominant, so let **R** = rounded ears and **r** = pointed ears.

Step 2 The brown mouse with rounded ears is homozygous so its genotype must be **BBRR**.

As white and pointed are recessive, the genotype of the white mouse with pointed ears must be **bbrr**.

Step 3 Set out the genetic diagram as shown.

Step 4 All the F_1 mice have brown fur and rounded ears.

Answer

parental phenotypes:	brown fur, rounded ears	white fur, pointed ears
parental genotypes:	**BBRR**	**bbrr**
gametes:	all (**BR**)	all (**br**)

Punnett grid for F_1:

	gametes from brown, round-eared parent	
		(**BR**)
gametes from white, pointed-eared parent	(**br**)	**BbRr** brown, rounded ears

WORKED EXAMPLE 4.2.8

Mendel carried out genetic studies with the garden pea. Tall plants are dominant to short, and green seed pods are dominant to yellow. Mendel crossed a homozygous tall plant with yellow seed pods with a short plant homozygous for green seed pods. Determine the possible genotypes and phenotypes in the offspring.

Step 1 Tall is dominant to short, so T = tall and t = short. Green pod is dominant to yellow so G = green and g = yellow.

Step 2 Each parent has one dominant and one recessive characteristic but we are told the dominant characteristic is homozygous. The tall plant with yellow seed pods therefore has genotype TTgg, and the short, green-podded plant is ttGG.

Step 3 Set out the genetic diagram as shown.

Step 4 All the offspring are tall plants with green seed pods.

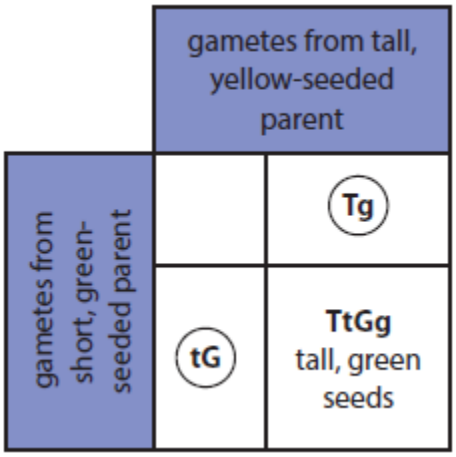
It was fortunate that Mendel chose these two characteristics for his crosses. Seed pod colour and height are unlinked genes on different chromosomes. Had they been on the same chromosome and linked, the results would have been different.

Answer

parental phenotypes:	tall, yellow seed pods	short, green seed pods
parental genotypes:	**TTgg**	**ttGG**
gametes:		

all (Tg) all (tG)

Punnett grid for F_1:



WORKED EXAMPLE 4.2.9

One of the heterozygous F_1 mice with brown fur and rounded ears from the cross in Worked example 4.2.7 was crossed with a mouse with white fur and rounded ears. Some of the offspring had pointed ears. Deduce the genotype of the second mouse and state the phenotype ratio of the offspring.

Step 1 Use the same letters as in Worked example 4.6.7: **B** = brown fur and **b** = white fur, **R** = rounded ears and **r** = pointed ears.

Step 2 The first mouse has the genotype **BbRr**.

We are told the second mouse is white, so it must have the alleles **rr**. It has rounded ears but we are not told if this is homozygous or heterozygous, so the alleles could be **RR** or **Rr**.

Reading on, we find that there are some offspring with pointed ears, so they must have the genotype **rr**. This means that the unknown parent genotype must have been heterozygous, **Rr**. If the parent was **RR**, no recessive allele would have been present so that a homozygous genotype could not occur in the offspring, and none of them would have had pointed ears.

Step 3 Having written down your reasoning, as in Step 2, now set out the usual genetic diagram.

Answer

parental phenotypes:	brown rounded ears	white rounded ears
parental genotypes:	**BbRr**	**bbRr**
gametes:	(BR) (Br) (bR) (br)	(bR) (br)

Punnett grid for F_1:

		gametes from brown, rounded-eared parent			
		(BR)	(Br)	(bR)	(br)
gametes from white, rounded-eared parent	(bR)	**BbRR** brown, rounded ears	**BbRr** brown, rounded ears	**bbRR** white, rounded ears	**bbRr** white, rounded ears
	(br)	**BbRr** brown, rounded ears	**Bbrr** brown, pointed ears	**bbRr** white, rounded ears	**bbrr** white, pointed ears

Step 4 The phenotypes produced are:

3 brown fur, rounded ears

3 white fur, rounded ears

1 brown fur, pointed ears

1 white fur, pointed ears.

This produces a ratio of phenotypes of 3 : 3 : 1 : 1, which is an important Mendelian ratio.

Notice that the individuals with white fur and pointed ears and brown fur and pointed ears are recombinant offspring because they have characteristics that differ from both their parents. Recombinants are always present in lower numbers than non-recombinant individuals.

KEY POINT

recombinant offspring is an offspring with characteristics that are different from both their parents. They are always present in lower numbers than other offspring.

Linkage and inheritance

If alleles are linked together on a chromosome, then it follows that they will be inherited together because during meiosis they will move together as the cell divides. In genetics problems, dihybrid crosses involving linked genes do not produce Mendelian ratios.

Linked genes do not follow Mendel's law of independent assortment, they are not inherited independently and can give a variety of different ratios.

Writing a linkage genotype

In the dihybrid crosses considered so far, genotypes have been written in the form AABB. With linked genes, a different notation has to be used because, although there are still four alleles to be considered, they are found on only one pair of

chromosomes. The genotype is therefore always written as shown in Figure 4.2.12. The horizontal lines signify that the two genes occur on the same chromosome.

EXAM TIP

In monohybrid crosses, two ratios for the offspring of a genetic cross are possible.

The first is 1 : 1 if a heterozygous individual **(Aa)** and a homozygous recessive individual (**aa**) are crossed. The second is 3 : 1 when two heterozygous individuals are crossed (**Aa** × **Aa**). These are called Mendelian ratios.

Mendelian ratios also occur in dihybrid crosses, but with more gametes there are more possibilities. A heterozygous individual (**AaBb**) crossed with a homozygous recessive (**aabb**) produces a 1 : 1 : 1 : 1 ratio and the ratio produced by crossing two double heterozygous (**AaBb**) individuals is 9 : 3 : 3 : 1.

The 3 : 3 : 1 : 1 ratio in Worked example 4.6.9 is another Mendelian ratio. It is helpful to be familiar with these ratios.

NATURE OF SCIENCE

Looking for trends and discrepancies: the work of T.H. Morgan

Thomas Hunt Morgan (1866–1945) was a pioneering geneticist who studied the inheritance of mutations in the fruit fly *Drosophila melanogaster*. After the rediscovery of Mendelian genetics in 1900, Morgan worked to show that genes are carried on chromosomes and provide the basis for inheritance. He induced mutations in his flies using chemicals

and radiation and began cross-breeding experiments to find mutations that were inherited. Despite the difficulty of spotting mutations in the tiny flies, he eventually noticed a white-eyed mutant male among the typical 'wild type' red-eyed flies. He bred white-eyed male flies with red-eyed females and all the offspring were red-eyed. The F_2 (second generation) cross produced white-eyed males so Morgan concluded that the white-eye mutation was a sex-linked recessive trait (Figure 4.2.10). He also discovered a pink-eyed mutant which was not sex linked.

Morgan reported his discoveries of sex linkage and autosomal inheritance in the journal *Science* in 1911. Morgan's team discovered more mutants amongst thousands of flies they studied and also identified flies with multiple mutations. They studied more complex patterns of inheritance, finding more examples of crosses that did not fit the pattern of simple Mendelian ratios. To explain these discrepancies, Morgan went on to suggest that genes could be linked and inherited together.

Morgan proposed the hypothesis of crossing over in his book and the relationship between crossing over and linked genes. He suggested that cross-over frequency gave a measure of the distance separating genes on a chromosome.

His book *The Mechanism of Mendelian Heredity*, published in 1915, was a foundation for modern genetics. (Morgan, Thomas Hunt, *The Mechanism of Mendelian Heredity.* New York: Henry Holt and Company, 1915. Electronic reproduction. New York, N.Y. : Columbia University Libraries, 2007.)

To consider:

1 In what ways do scientists improve the quality and quantity of evidence they collect?

2 *Drosophila* flies are known as 'model' organisms and are used in many genetic experiments. Investigate other model organisms and their importance to science.

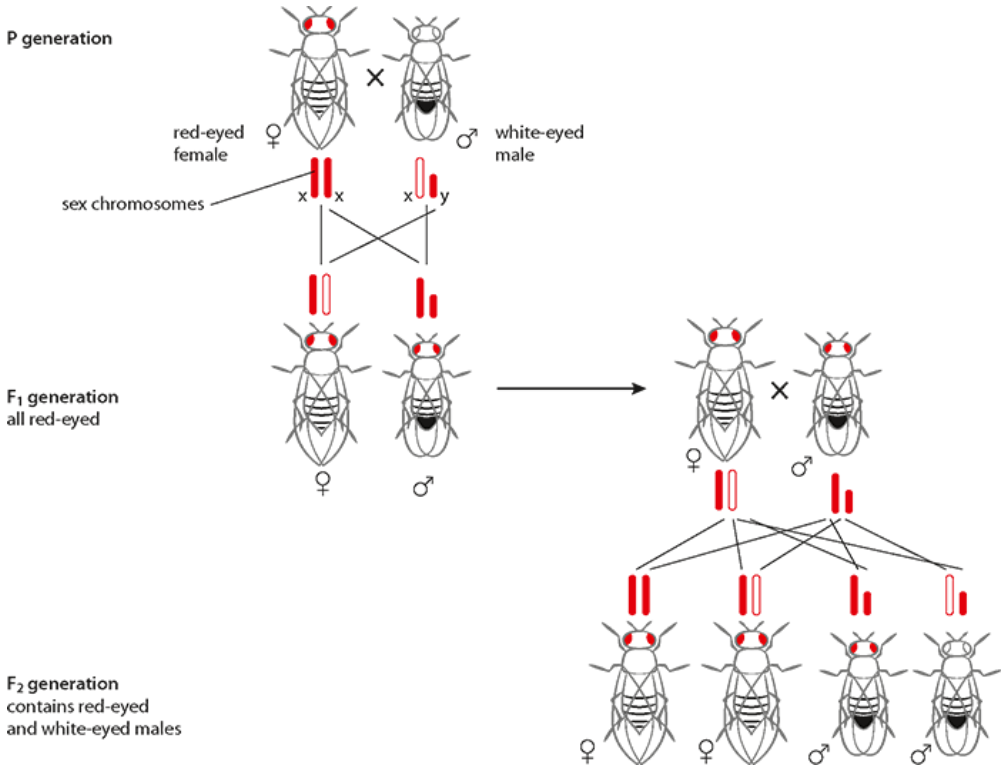


Figure 4.2.10: Inheritance of a sex-linked characteristic in Drosophila.

Linkage and crossing over

In Chapter 6, Figure 6.5.9 you can study how crossing over creates genetic variety by exchanging parts of the maternal and paternal chromosomes during meiosis. Figure 4.2.13 shows what

happens to two closely linked alleles after a cross-over occurs between them.

Look back at the left-hand example in Figure 4.2.11. Without crossing over, the parental gametes formed will be **DG** and **dg**. If a cross-over does take place (as shown by the red cross in Figure 4.2.12) then additional recombinant gametes **Dg** and **dG** will be formed.

Four types of gamete – **DG**, **dg**, **Dg** and **dG** – are possible, but there is a very significant difference in the numbers of each type that are formed. The chance of a chiasma forming between the two loci, which are close together, is very small. So the chance of forming the gametes **Dg** and **dG** is also very small. The majority of gametes therefore carry the alleles **DG** and **dg** and they will form in equal numbers. If a cross-over does takes place, for every **Dg** gamete there will be a **dG** gamete. The numbers of these two gametes will also be equal but very small. The geneticist T.H. Morgan made these observations in his work (see Nature of Science, Looking for trends and discrepancies: the work of T.H. Morgan).

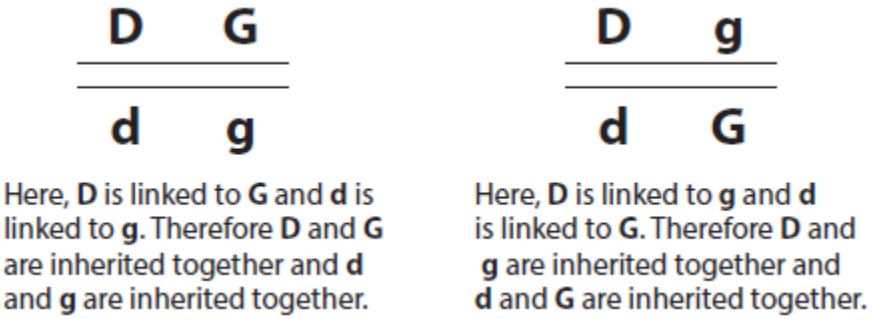


Figure 4.2.11: The two possible linkage patterns for the four alleles. The difference between the two linkage patterns makes a very big difference in the ratios of the phenotypes in the offspring of a cross.

Figure 4.2.12: If a cross over (shown by the red cross) takes place then the recombinant gametes Dg and dG will be formed.

Now look at the right-hand example in Figure 4.2.13. What will be the allele combinations in gametes where crossing over has not taken place? What will be the allele combinations in gametes where crossing over has taken place, and which combinations will be present in greater numbers?

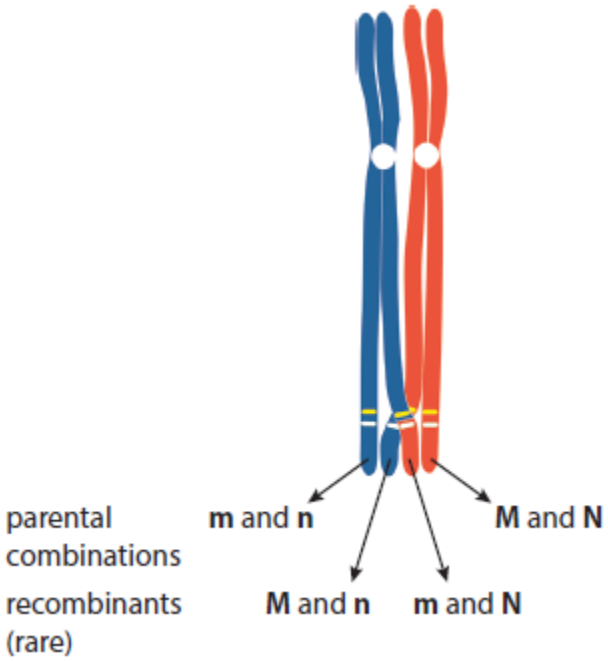


Figure 4.2.13: A single chiasma (point of contact) has formed between two chromatids so crossing over of the alleles can take place to form recombinant gametes. No crossing over has taken place with the other chromatids and so these will retain the parental combination of alleles.

WORKED EXAMPLE 4.2.10

In the fruit fly, *Drosophila*, red eye colour is dominant to purple eyes and long wings is dominant to dumpy wings. These genes are linked on chromosome 2. A fly that was homozygous for red eyes and long wings was crossed with a fly that had purple eyes and dumpy wings. Determine the ratios of genotypes and phenotypes of the F_2 offspring by using a full genetic diagram.

Step 1 Red eye colour is dominant so **R** = red eye and **r** = purple eye. Long wings is dominant so **N** = long wings and **n** = dumpy wings.

Step 2 The fly with the dominant characteristics is homozygous and the other fly shows both recessive characteristics. The parental genotypes are:

$$\frac{RN}{RN} \text{ and } \frac{rn}{rn}$$

Step 3 Set out the genetic diagram as below.

Answer

parental phenotypes:	red eyes, long wings	and	purple eyes, dumpy wings
parental genotypes:	$\frac{R\ N}{R\ N}$	and	$\frac{r\ n}{r\ n}$
gametes:	all (RN)		all (rn)

Punnett grid for F_1:

	gametes from red-eyed, long-winged parent	
gametes from purple-eyed, dumpy-winged parent		RN
	rn	$\frac{R\ N}{r\ n}$ red eyes, long wings

All the F_1 have red eyes and long wings.

Now, the F_2 generation is obtained by crossing two offspring from theF_1 generation.

parental phenotypes: red eyes, long wings and red eyes, long wings

parental genotypes: $\frac{R\ N}{r\ n}$ and $\frac{R\ N}{r\ n}$

gametes: RN rn RN rn

Punnett grid for F_2:

		gametes from red-eyed, long-winged parent	
		RN	rn
gametes from red-eyed, long-winged parent	RN	$\frac{R\ N}{R\ N}$ red eyes, long wings	$\frac{R\ N}{r\ n}$ red eyes, long wings
	rn	$\frac{R\ N}{r\ n}$ red eyes, long wings	$\frac{r\ n}{r\ n}$ purple eyes, dumpy wings

Step 4 In the F_2 generation, the ratio of phenotypes is 3 red eye, long wing : 1 purple eye, dumpy wing. Note that

this 3 : 1 ratio is what you would expect in a monohybrid cross. The reason for this is that the two genes are linked and so there is only one pair of chromosomes involved, as in monohybrid crosses.

WORKED EXAMPLE 4.2.11

Grey body and red eyes are dominant to stripe body and cardinal eye in *Drosophila*. They are autosomal, linked genes on chromosome 3. Homozygous grey flies with red eyes were crossed with stripe flies with cardinal eyes. No crossing over occurred.

Then the F_1 flies were crossed with stripe, cardinal flies.

If no crossing over occurs between the two loci, what phenotypes would be expected in the offspring of this second cross?

If crossing over did occur between the loci, what phenotypes would be expected this time?

Answer

First cross, with no crossing over:

Step 1 Grey body is dominant, so **G** = grey body and **g** = stripe body. Red eye is dominant so **R** = red eye and **r** = cardinal eye.

Step 2 The fly with the dominant characteristics is homozygous and the other fly shows both recessive characteristics. The parental genotypes are:

$\frac{GR}{GR}$ and $\frac{gr}{gr}$.

Step 3 Set out the diagram as shown.

parental phenotypes: grey body, red eyes stripe body, cardinal eyes

parental genotypes: $\frac{G\ R}{G\ R}$ $\frac{g\ r}{g\ r}$.

gametes: all (GR) all (gr)

Punnett grid for F_1:

gametes from stripe-bodied, cardinal-eyed parent	gametes from grey-bodied, red-eyed parent: (GR)	gametes from grey-bodied, red-eyed parent: (gr)
	(GR)	
(gr)		$\frac{G\ R}{g\ r}$ grey body, red eyes

Step 4 All the F_1 flies have a grey body and red eyes.

Second cross, with no crossing over:

parental phenotypes: grey body red eyes stripe body, cardinal eyes

parental genotypes: $\frac{G\ R}{g\ r}$ $\frac{g\ r}{g\ r}$.

gametes: (GR) (gr) all (gr)

Punnett grid for F_1:

		gametes from grey-bodied, red-eyed parent	
		(GR)	(gr)
gametes from stripe-bodied, cardinal-eyed parent	(gr)	$\frac{G\ R}{g\ r}$ grey body, red eyes	$\frac{g\ r}{g\ r}$ stripe body, cardinal eyes

Second cross, with crossing over:

parental phenotypes: grey body red eyes / stripe body cardinal eyes

parental genotypes: $\frac{G\ R}{g\ r}$ / $\frac{g\ r}{g\ r}$

gametes produced by crossing-over: (GR) (gr) (Gr)+ (gR)+ all (gr)

Punnett grid for F_1:

		gametes from grey-bodied, red-eyed parent			
		(GR)	(gr)	(Gr)	(gR)
gametes from stripe-bodied cardinal-eyed parent	(gr)	$\frac{G\ R}{g\ r}$ grey body, red eyes	$\frac{g\ r}{g\ r}$ stripe body, cardinal eyes	$\frac{G\ r}{g\ r}$ grey body, cardinal eyes	$\frac{g\ R}{g\ r}$ stripe body, red eyes

Step 4 The four F_1 phenotypes are:

- grey body and red eyes
- stripe body and cardinal eyes
- grey body and cardinal eyes
- stripe body and red eyes

The 'grey, cardinal' and 'stripe, red' (shown in red type) flies are recombinants as they have a phenotype that is

different from the parental phenotypes. These recombinant phenotypes will occur in approximately equal numbers. The parental phenotypes (shown in black type) will also be in approximately equal numbers among the offspring, but the recombinant phenotypes will be very few in number compared to the parental phenotypes.

NATURE OF SCIENCE

Careful observation: were Mendel's results 'too good to be true'?

In Mendel's time, statistical analyses, such as the chi-squared test, were not routinely used to test the validity of scientific results. Analysis of the ratios that Mendel published in 1866 suggests that they may be too close to the expected 3 : 1 ratio. Some have suggested that Mendel selected the data to present or that, having obtained a 3 : 1 ratio in some experiments, he persisted with counting until he achieved his expectations in subsequent experiments.

Mendel considered seven genes in his monohybrid experimental crosses, and also carried out dihybrid crosses involving pairs of these same genes. Crosses involving a pair of genes on different chromosomes produce the ratios Mendel reported, but he would have obtained unexpected results had he used linked genes, that is, genes that occur on the same chromosome. It is not likely that the seven genes Mendel investigated would each occur on a different chromosome by chance (the pea plant only has seven pairs of chromosomes) so perhaps some of his crosses did involve linked genes. He would not have been able to explain the results from such crosses as he did not know about chromosomes, so it may be

that he only published results from crosses that met his expectations.

Later work by T.H. Morgan involving meticulous observation and record keeping revealed anomalous data in dihybrid crosses similar to those that Mendel had performed. As a result, Morgan developed his ideas of gene linkage to explain the results.

To consider:

1 How do expectations and personal bias affect the results that scientists collect and present?

2 How is our understanding of science improved by reviewing results from experiments conducted a long time ago?

3 How important are statistics to modern science?

THEORY OF KNOWLEDGE

Breaking the law

Mendel's law of independent assortment was found to have exceptions that geneticist T.H. Morgan explained by using the idea of linked genes that occur on the same chromosome

To consider:

1 Is it correct to call Mendel's proposals a 'law'?

2 What is the difference between a law and a theory in science?

4.2.11 The chi-squared test and dihybrid crosses

The chi-squared (χ^2) test is a statistical test used to check if the results of an experiment support a theory. It can be used in cases in which variation is discrete. In genetic or ecological investigations, the chi-squared test is useful to compare your observed results with the results that you would expect if your theory about how the system works is correct. The test tells you whether or not any difference between your observed results and your expected results is significant. If the difference is significant that means the results do not fit well with your theory, so you may need to revise your theory. If the difference is not significant – that is, it is so small that it could have occurred by chance – then you can say that your results do support your theory.

In the case of genetics the test is used to check if the results of genetic crosses (the Observed results) match predictions made from Mendelian ratios (the Expected results). The formula for calculating the chi-squared value is:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where χ^2 is the test statistic, Σ means 'the sum of', O is the observed frequencies and E is the expected frequencies.

The greater the value of chi-squared you calculate, the greater the difference between your observed and expected results. To find out whether the difference is significant or not, you must compare your chi-squared value with a table of 'critical values' like the one in Table 4.2.4. The **null hypothesis** states that there is no significant difference between the observed and expected results, that is, that the results fit the expected pattern and therefore support your theory.

The chi-squared test can be used to test the outcome of monohybrid or dihybrid crosses, to see if the observed ratios fit the expected pattern. The worked examples that follow use data that were produced as a result of dihybrid crosses.

The null hypothesis predicts the ratio of offspring of different types from Mendelian ratios for the cross, assuming that there is no linkage. A

significant difference from the predicted ratio can indicate that alleles are linked.

WORKED EXAMPLE 4.2.12

Theory tells us that wing length in *Drosophila* is controlled by a single pair of genes on their second chromosome. Wings may be normal or vestigial. Flies with vestigial wings have a recessive gene and cannot fly. If two heterozygous parents are crossed, a ratio of 3 : 1 normal to vestigial wings is the expected result. We put forward a null hypothesis that there is no significant difference between the observed and expected results. If the chi-squared test shows that there is no difference between observations and expectations we can accept the null hypothesis that wing length is controlled by monohybrid inheritance.

In a cross between heterozygous parents, 320 offspring were counted.

Answer

Step 1 Set the null hypothesis; this predicts the ratio of offspring of each phenotype according to Mendelian ratios for the cross. For a monohybrid cross, the expected ratio of phenotypes in the F1 generation is 3 : 1. So here, with 320 flies, our expected result is that we should see a ratio of 3 : 1 normal : vestigial wings,

¼ × 320 = 80 flies should have vestigial wings and

¾ × 320 = 240 flies should have normal wings

The actual numbers of offspring were counted and recorded:

Phenotype	Ratio	Expected result	Observed result
Normal wings	3	240	232
Vestigial wings	1	80	88

Step 2 We must now use the chi-squared equation to test whether the observed numbers differ significantly from our expectations. Calculate the value of chi-squared. This is the sum of the differences between

each pair of observed and expected values, squared, and divided by the expected value:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Phenotype	Ratio	Expected result	Observed result	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Normal wings	3	240	232	−8	64	0.266
Vestigial wings	1	80	88	8	64	0.80

Using the formula, $\chi^2 = 0.266 + 0.80 = 0.376$.

Step 3 Select the appropriate row in a table of critical values of chi-squared, like the one in Table 4.6.4.To do this, we must calculate the 'degrees of freedom', which is the number of categories among your results, minus 1. In this case there are two categories (normal wings and vestigial wings) so the degrees of freedom 2 − 1 = 1. Look along the row of the table that corresponds to 1 degree of freedom.

Step 4 Find the critical chi-squared value at the 5% (0.05) significance level. In biology, the 5% significance level is used. This level means that the probability for rejecting the null hypothesis is 5%, that is, there is a 5% probability that the difference between the observed results and the expected values occurred purely by chance, and is not significant. If our calculated chi-squared value is less than the critical value at the 5% level, then the probability of obtaining the difference we observed by chance alone is greater than 5%, so we can accept the null hypothesis and have no reason to think that our results differ significantly from our expected values.

In this case, the chi-squared (0.376) value is lower than the value in the table (3.841) so we accept the null hypothesis that there is no significant difference between the observed and expected results. In other words, our results support the theory on which the expected ratio of 3 : 1 was based. We can accept the theory that wing length in these flies is due to monohybrid inheritance.

Degrees of freedom	*p*									
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Table 4.2.4: A chi-squared table. Critical values of the chi-squared distribution, showing how to read across the appropriate 'degrees of freedom' row to find the critical chi-squared value at the 0.05 level (*p*).

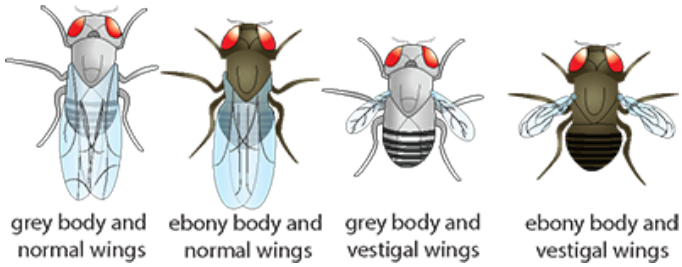


Figure 4.2.14: Diagram to show some inherited characteristics of Drosophila sp.

A cross between *Drosophila* involved four characteristics: ebony body, grey body, long wings and vestigial wings (Figure 4.2.14). Assuming that there is no linkage of alleles our expected result should be a 1 : 1 : 1 : 1 ratio.

After the cross, 800 flies were collected and the numbers of offspring of each type were recorded:

Grey body, normal wings 196	Ebony body, normal wings 204
Grey body, vestigial wings 176	Ebony body, vestigial wings 224

Phenotype	**Ratio**	**Expected result**	**Observed result**	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Grey, normal wings	1	200	196	−4	16	0.08
Grey, vestigial wings	1	200	176	−24	576	2.88
Ebony, normal wings	1	200	204	4	16	0.08
Ebony, vestigial wings	1	200	224	24	576	2.88

Using the formula, $\chi^2 = 0.08 + 2.88 + 0.08 + 2.28 = 6.52$

From Table 4.2.4, the critical value at the 5% (0.05) level, for 3 degrees of freedom = 7.815

Our calculated value is less than the critical value so we can accept the null hypothesis that there is no significant difference between our observations and expectations. We can accept the conclusion that a 1 : 1 : 1 : 1 ratio is the result of this type of dihybrid cross.

TEST YOUR UNDERSTANDING

14 List three environmental factors that can determine the sex of offspring of certain organisms.

15 Define the term linked gene loci.

16 How do we work out the ‘degrees of freedom’ in a chi-squared test?

EXAM TIP

If the expected ratio in a Mendelian cross is 9 : 3 : 3 : 1 the expected values (E) must be calculated accordingly and the totals divided by 16 (9 + 3 + 3 + 1).

Links

- How do variations that are inherited contribute to evolution? (Chapter 11)
- How do genetic conditions caused by dominant alleles remain in the population? (Chapter 11)
- To what extent does sexual reproduction contribute to variation? (Chapter 6 and 8)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
define genome	4.1.1			
summarise the differences between prokaryotic and eukaryotic chromosomes	4.1.2			
differentiate between diploid and haploid nuclei and a chromosome and a chromatid	4.1.2			
recall that genome sizes are different in different species but that size does not relate to	4.1.2			

complexity of the organism				
define a gene and explain how a gene may have different alleles at the same locus	4.1.3			
recall that chromosome number is a characteristic of a species	4.1.4			
summarise the how karyograms are made and their uses, including benefits and disadvantages	4.1.4			
outline the importance of X inactivation and why Barr bodies are produced	4.1.5			
summarise the importance of Mendel's work to genetics	4.2.1			
state that gametes are haploid, and that fusion of	4.2.1			

gametes produces a diploid zygote				
work out genotypes and phenotypes from genetic crosses using Punnett grids	4.2.2			
define the terms genotype, phenotype, homozygous and heterozygous, incomplete dominance and codominance	4.2.3, 4.2.4			
describe the ABO blood system as an example of the inheritance of a characteristic with multiple alleles	4.2.3			
outline how sex chromosomes are inherited and explain how sex-linked disorders such as hemophilia are inherited	4.2.5			
give examples of	4.2.7			

recessive conditions caused by inherited alleles				
explain polygenic inheritance and the difference between discrete and continuous variation	4.2.7			
outline the genetic condition PKU caused by a mutation	4.2.7			
define phenotypic plasticity	4.2.8			
give examples of species in which environmental conditions determine the sex of organisms	4.2.9			
use a Punnett square to demonstrate the inheritance of unlinked dihybrid characteristics	4.2.10			
explain the difference	4.2.10			

between linked and unlinked alleles and identify recombinants in crosses involving two genes				
use the chi-squared test on data from dihybrid crosses to check for the significance of results.	4.2.11			

REFLECTION

Reflect upon the content of this chapter and identify those areas of strength and weakness in your understanding. How can you improve in those topics you have found difficult?

EXAM-STYLE QUESTIONS

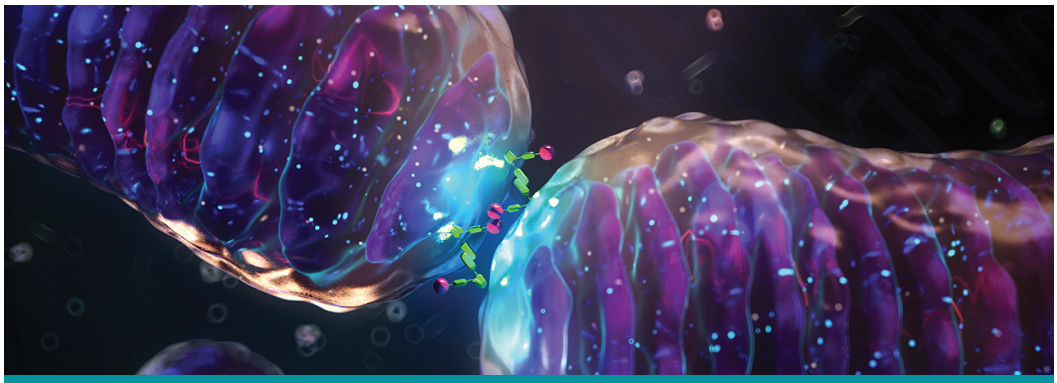
You can find questions in the style of IB exams in the digital coursebook.

Unit 2
Cellular organisation

INTRODUCTION

Cells as we know them today originated millions of years ago. Most scientists agree that life arose from organic molecules that were present on the early Earth. These molecules provided the building blocks of the cells that make up all living organisms. All cells have similarities, but also many differences. Structures such as membranes and mitochondria are organised in the same ways in different cells, but cells can express different genes and develop their own unique properties. Some cells become neurones, while others grow into bone or brain cells. In multicellular organisms cells interact to form tissues and organs to carry out their life processes.

All cells are limited in size and must remain small because substances must enter and leave through their plasma membrane. As cells grow larger they divide to replace parts of the body or to enable an organism to grow larger. Cells communicate with one another by cell signalling. Different parts of multicellular organisms communicate as nerve impulses and chemical signals that pass between their cells.



Chapter 5

Cell structure

A2.1, A2.2, A2.3

INTRODUCTION

In the middle of the 17th century, one of the pioneers of microscopy, Robert Hooke (1635–1703), decided to examine a piece of cork tissue with his home-built microscope. He saw numerous box-shaped structures that he thought resembled monks' cells or rooms in a monastery, so he called them 'cells'. As microscopes became more sophisticated, other scientists observed cells and found that they occurred in every organism. No organism has yet been discovered that does not have at least one cell. Living things may vary in shape and size but scientists agree that they are all composed of cells. The

study of cells has enabled us to learn more about how whole organisms function.

5.1 Origins of life

LEARNING OBJECTIVES

In this section you will:

- recognise how conditions on the early Earth led to the formation of organic molecules and the origins of life
- recognise that boundaries separate cells from their environments
- learn that cell theory states that cells are the smallest functional unit of life
- recognise that all cells arise from pre-existing cells
- discover that the Miller–Urey experiments provided evidence on the origin of organic compounds
- understand that the deep-sea vent hypothesis suggested how energy was provided for primitive life forms
- learn that RNA has properties that make it a likely part of primitive life
- learn that lipids have properties that make them important in protocell growth
- understand that the Last Universal Common Ancestor is proposed as the link between the abiotic and biotic phases of the early Earth

- learn that the evidence for LUCA comes from the genetic code shared across all organisms and fossilised evidence of life from ancient seafloor hydrothermal vents.

GUIDING QUESTIONS

1. What plausible hypothesis could account for the origin of life?
2. What is the evidence that supports the theory that life arose from organic molecules?

5.1.1 Forming organic molecules in the early Earth

Most scientists agree that evidence gathered from many sources indicates that the Earth formed about 4.5 billion years ago. Within the first billion years, the first signs of life appeared in the oceans and Earth's atmosphere began to change. Living organisms emerged, changed and have evolved, but how did the first molecules and organisms arise?

Abiogenesis, the origin of life, was a natural process. Scientists think that life arose from non-living, simple molecules and organic compounds. This probably did not happen all at once but in stages, so that life gradually became more complex and produced molecules that could self-replicate, assemble themselves and eventually produce a cell membrane that separated them from the environment. We say that life is an emergent property that has evolved over a long period of time.

KEY POINTS

abiogenesis the process by which life has arisen from non-living matter.

emergent properties of a complex system that arise from simple interactions of individual component parts. In the case of water, these properties are due to interactions between individual molecules.

In 1952 an experiment was carried out by Stanley Miller and Harold Urey which attempted to show that the conditions that existed on Earth millions of years ago were suitable for the synthesis of complex organic molecules from simple ones. The

experiment used molecules that were present on the early Earth: water, methane, ammonia and hydrogen. In a sealed flask (Figure 5.1.2) the water was heated and electrical sparks used to simulate lightning in the water vapour and mixture of gases. The experiment showed that organic compounds, amino acids, did form from the simple ingredients. The experiment is described in more detail in Section 5.1.3 The Miller–Urey experiments.

The next step in evolution must have been the formation of macromolecules. Monomers of macromolecules can polymerise spontaneously in the conditions that existed billions of years ago. For example, amino acids can polymerise to form polypeptides. But the most important feature of the macromolecules from which life evolved must have been the ability to replicate themselves. Only a macromolecule able to synthesise new copies of itself would have been able to reproduce and evolve.

We can never recreate the conditions that were present at the time, but the first cell is presumed to have developed when self-replicating RNA was enclosed in a membrane of phospholipids. Phospholipids are the main component of biological membranes in both prokaryotic and eukaryotic cells (Section 5.2). Because they are amphipathic molecules they can form a bilayer boundary that separates the interior of a cell from the external environment.

Primitive **protocells** containing RNA and enclosed in a membrane would have been able to evolve further and eventually code for, and produce, their own proteins.

KEY POINT

protocells structures formed from the aggregation of abiotic components but which have some similarities to living cells.

5.1.2 Cell theory

Today, scientists agree that the cell is the fundamental unit of all life forms. **Cell theory** proposes that all organisms are composed of one or more cells and, furthermore, that cells are the smallest units of life. An individual cell can perform all the functions of life – it must have a metabolism and the ability to replicate – and anything that is not made of cells, such as viruses, cannot be considered living (Section 5.3).

KEY POINT

Key principles of cell theory

- Living organisms are composed of cells.
- Cells are the smallest units of life.
- All cells, apart from the first cells, come from pre-existing cells.

As one of the key life processes of all living organisms is reproduction, one of the first principles of cell theory is that cells can only come from pre-existing cells. Louis Pasteur (1822–1895) carried out experiments that provided evidence for this. He showed that bacteria could not grow in a sealed, sterilised container of chicken broth. Only when living bacteria were introduced would more cells appear in the broth. Figure 5.1.1 summarises Pasteur's experiment.

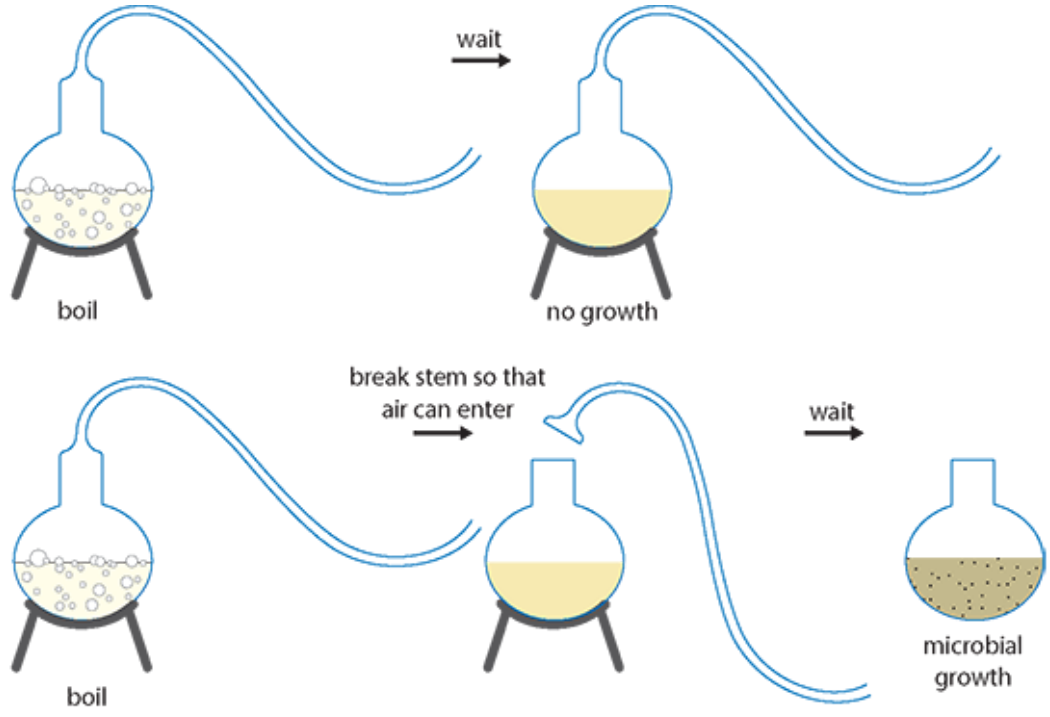


Figure 5.1.1: Pasteur demonstrated that living cells cannot 'spontaneously generate' but must originate from pre-existing cells.

TEST YOUR UNDERSTANDING

1 Name two substances present on the early Earth that may have contributed to the formation of the first organic molecules

2 Define emergent property.

3 List the fundamental features of a living cell.

5.1.3 The Miller-Urey experiments

In 1952 Stanley Miller and Harold Urey, working at the University of Chicago, simulated the conditions that they thought existed on Earth at the time life originated. The gases they used were enclosed in a sealed glass flask connected to a source of water vapour (Figure 5.1.2). Water vapour evaporated into the larger flask and electrical sparks were used to imitate lightning. Liquid from the flask was cooled and flowed down to the trap at the bottom of the apparatus. After a week the solution in the trap was removed and tested. It was found to contain at least five amino acids.

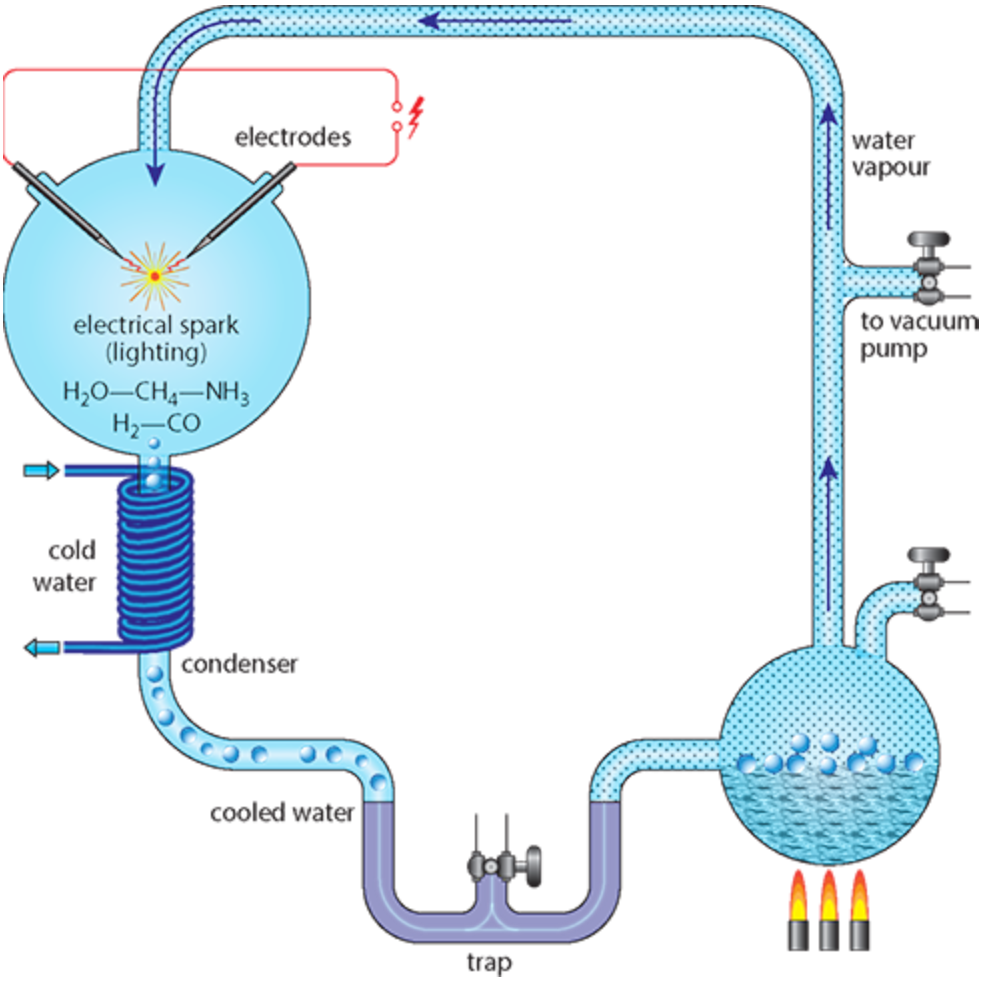


Figure 5.1.2: Diagram of the apparatus used in the Miller–Urey experiment.

In the early 21st century sealed flasks that had been kept from the original experiment were reopened and found to contain more than 20 amino acids. These results provide strong evidence that it is possible to create complex molecules from simple substances that were present on the prebiotic Earth.

More recent evidence has been gathered about the atmosphere on the early Earth and it suggests that the gas mixture Miller and Urey used was not exactly right. Nevertheless, similar experiments using different proportions of gases have been able to convert simple substances to complex compounds.

Today there are several different hypotheses about how life might have arisen on Earth.

5.1.4 The deep-sea vent hypothesis and a source of energy for primitive life

We do not know where life on Earth started and we cannot be sure what the environmental conditions were like 3–4 billion years ago. However, since their discovery in 1977, deep-sea hydrothermal vents under the oceans have been investigated as a possible place where life started.

The first vents to be discovered were ‘black smokers’, which give out plumes of hot water at temperatures of up to 400 °C. This water contains high levels of sulfides that produce the black smoke that is seen as they come into contact with the cold ocean water. In 2000, a new type of alkaline hydrothermal vent was discovered, the first one, located on the seabed in an area of the mid-Atlantic is known as the Lost City.

In 1993, an American geochemist, Michael Russell, first suggested a theoretical mechanism to explain how alkaline vents might have been important to the development of early life. He suggested that the energy gradients that exist when alkaline vent water mixes with more acidic seawater could fuel the formation of organic molecules.

In some ways this is similar to the way in which cells harness energy. Cells maintain a proton gradient by pumping protons across their membranes to create a charge difference from inside to outside (Chapter 3). When protons are allowed to pass back through the membrane they phosphorylate adenosine diphosphate (ADP) and make ATP, an energy source that can be used by the cell. Russell’s theory suggests that pores in the hydrothermal vent chimneys worked in a similar way and provided energy that fuelled the reduction of carbon dioxide and

the production of organic molecules. In time, this could have led to self-replicating molecules, and eventually true cells with their own membranes.

Scientists are working on this theory in laboratories in the USA and the UK. They use small-scale models of hydrothermal vents (Figure 5.1.3) and seed them with chemicals that are present around them in the deep oceans in an attempt to produce organic molecules. One way that RNA might have first formed is with the help of minerals, such as iron and sulfur, that are found at alkaline hydrothermal vents on the seabed. These minerals act as catalysts and help to build organic compounds from inorganic molecules.

Figure 5.1.3: Hydrothermal volcanic vents deep under the ocean. Minerals dissolve in their hot waters and concentrations of molecules may have played a role in the origin of life.

5.1.5 RNA and the origin of life

Living things need molecules that have the ability to catalyse reactions that lead to the production of more molecules like themselves so that life can continue.

RNA has both the ability to replicate itself and to act as an enzyme and so one widely held view is that an RNA-dominated world existed on Earth before modern cells formed. This RNA hypothesis suggests that RNA both stored genetic information and catalysed chemical reactions in primitive 'cells'. DNA took over as the genetic material and proteins became important catalysts much later in evolutionary time.

Support for this view comes from the fact that RNA still catalyses several important reactions in modern-day cells. Perhaps these reactions are the molecular equivalent of fossils from long ago. RNA is a likely molecule to form the basis for a self-replicating set of catalysts but it probably was not the first kind of molecule to do so.

But RNA molecules are difficult to synthesise and build into nucleotides without enzymes, so it has been suggested that the earliest molecules to have both catalytic and information-transferring abilities were chemically simpler 'RNA-like' polymers.

Protein catalysts that we know today have a surface with uniquely shaped active site where a given set of substrates can react. In exactly the same way, RNA molecules with appropriately folded shapes can serve as enzymes. For example ribozymes (ribonucleic acid enzymes) are RNA molecules that have the ability to catalyse certain biochemical reactions,

including RNA splicing as genes are expressed and peptide bond formation during the synthesis of protein. Like some proteins, many of these RNA enzymes work by positioning metal ions at their active sites. This feature increases the range of catalytic activities they can perform and explains how they can carry out more reactions than can be accounted for than by chemical arrangements of the polynucleotide chain alone.

5.1.6 Micelles

Before cells as we know them today came into existence, they needed to separate themselves and their chemical reactions from the outside environment. To do this they developed a membrane bound compartment to allow the reactions inside the compartment to be different from those outside it. **Micelles** are aggregates formed by self-assembly of amphipathic molecules. Amphipathic structures contain a hydrophilic/polar region (head) and a hydrophobic/non-polar region (tail). In water, micelles form so that the polar region faces the outside surface of the micelle and the non-polar region forms the core.

Modern-day cells have membranes formed of phospholipid bilayers which separate cells from their environment and control what enters and leaves the cell (Section 6.5). Phospholipids placed in water form micelles, **vesicles** and bilayers.

Early life forms also needed to separate themselves from their environment and the first membrane-like barriers may have formed from fatty acids. Fatty acids are thought to have formed in low concentrations near hydrothermal vents. Fatty acids have hydrophilic heads and hydrophobic tails and they can arrange themselves into micelles. These arrangements may well have formed in the prebiotic environment and been part of the transition from chemical evolution to biological evolution on the early Earth.

KEY POINT

vesicles are structures that consist of a lipid bilayer and form within or outside cells. They form naturally and transport substances inside cells.

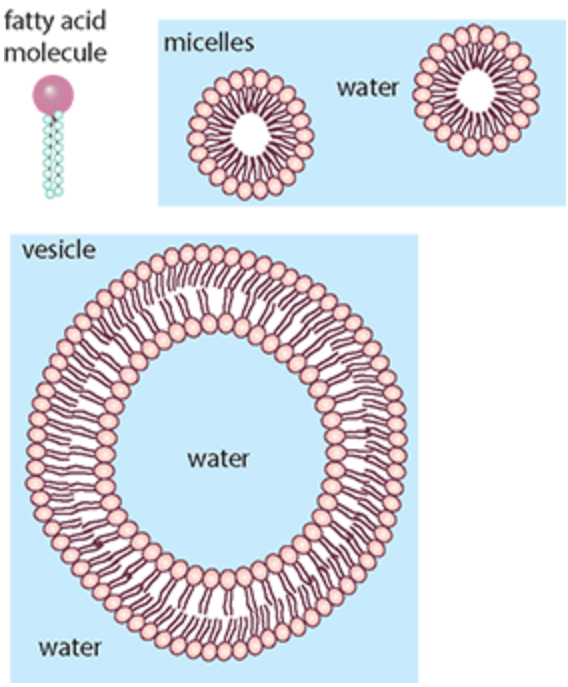


Figure 5.1.4: Micelles form from amphipathic molecules with hydrophobic tails at the centre and hydrophilic head around the outside.

Jack Szostak (b. 1952) is a Canadian American biologist and Nobel laureate who has studied the chemical and physical processes that may have led to this evolution.

The Szostak laboratory in Massachusetts, USA has investigated the formation of micelles and vesicles and discovered that, at low concentrations of fatty acids, micelles form, while at higher concentrations and the correct pH, vesicles will form.

Fatty acid barriers would have been important in an RNA world, especially to separate RNA sequences with different functions from one another. Movement across the fatty acid vesicles would have been possible because fatty acids can enter the vesicle membrane and small molecules including nucleotides can also pass inside. Experiments have shown that vesicles form and

grow rapidly in the presence of micelles. Fatty acid vesicles are very stable and do not appear to change, although their molecules are mobile and will enter and leave the vesicle layer and ‘flip’ between layers. Flipping would have enabled small molecules such as RNA nucleotides to enter the vesicle. Once incorporated into polymers, RNA strands would have been too large to leave through the fatty acid layers.

The membrane may also have been important in the early cell’s ability to store energy by creating a chemical gradient to create ATP in a similar way to the membranes in mitochondria and chloroplasts (Chapter 3).

Membrane boundaries enable cells to stabilise their internal environment by limiting what can enter and leave the cell. They also allow certain molecules to be accumulated to concentrations that allow biochemical reactions to take place.

5.1.7 Comets

It is unlikely that the early Earth had sufficient organic molecules needed to build all the complex molecules of life. But comets, which arrive regularly on the Earth's surface from space, are a rich source of organic materials.

The core of a comet contains simple molecules such as carbon dioxide and methanol but the Rosetta experiment, which landed a probe on the comet 67P in 2014, discovered that 16 different organic molecules were present. In addition, scientists from NASA discovered the amino acid glycine present in dust from another comet, Wild 2.

It is possible that comets brought these molecules, and others like them, to Earth during a period known as the Late Heavy Bombardment. This period, when many comets landed on Earth, is thought to have occurred between 4.1 and 3.8 billion years ago. Laboratory-based experiments have simulated the high-energy collisions between comets and Earth and discovered that the simple organic molecules found in comets will combine to form a number of amino acids. When amino acids were added to the chemicals from the comet and the experiment repeated, peptide chains were formed. This has led to speculation that comet collisions were key to the formation of organic molecules and ultimately life on Earth.

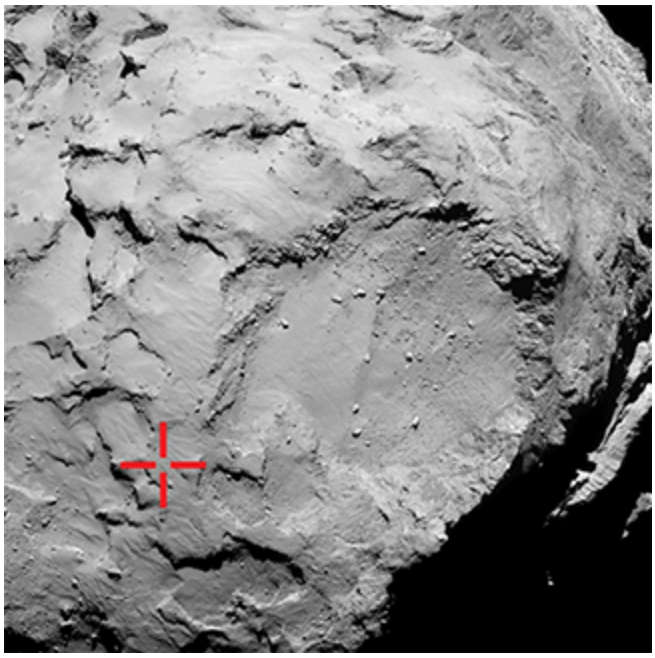


Figure 5.1.5: The comet 67P contained 16 different organic molecules.

NATURE OF SCIENCE

Establishing new theories and gathering evidence

Theories and models have been developed to try to explain how life might have begun on Earth. None of the theories discussed here can ever be proven because we can never observe what happened billions of years ago. Theories such as the clay hypothesis and the deep-sea vent hypothesis are scientists' attempts to explain processes and phenomena using data that are available and experiments that model possible events. Evidence to support a theory can be collected from laboratory experiments that model scientists' ideas and data from other sources. All of these may be helpful to support one theory over another. Any theory can be falsified when evidence does not support it. Science is based on and limited by evidence.

To consider:

1. How difficult is it for a scientist to develop a new theory or hypothesis when there is evidence to support an earlier one?
2. How do new discoveries, such as the locations of thermal vents, lead to changes in theories?
3. How important are new technologies, such as space travel, in learning about the past?

5.1.8 Last universal common ancestor

The last universal common ancestor (LUCA) is the name given to an evolutionary intermediate between the abiotic phase of Earth's history and the first traces of microbes found in rocks that are around 4 billion years old. The idea of a last universal common ancestor of all cells is central to the study of early evolution and the origin of life (Fig 5.1.6), but we can never have definite information about how and where LUCA lived. We can only gather evidence from experiments, the fossil record and studies of present-day genomes. There is evidence that LUCA could have lived in deep underground in iron-sulfur rich hydrothermal vents. LUCA would have been anaerobic and autotrophic, able to produce food in the dark, metal-rich environment around it. Its metabolism probably depended upon hydrogen, carbon dioxide and nitrogen and it would have produced organic compounds from them.

Evidence for LUCA

Evidence for LUCA's existence comes from the universal genetic code and shared genes that exist in all organisms. It is probable that other life forms also evolved but became extinct due to competition from LUCA and its descendants. A team of researchers based in Germany have identified a few hundred genes that definitely belonged to LUCA and can tell us something about how LUCA lived. Geneticists have checked the sequences of nearly 2000 genomes of modern microbes to search for these shared genes and using knowledge of what the genes code for today have suggested that LUCA lived without oxygen and used hydrogen. This provides support for the theory that LUCA was a thermophile (heat-loving) microbe that lived

in and around hydrothermal vents like those found near undersea volcanoes. Other researchers have noticed that LUCA shares this way of life with two groups of present-day microbes: Clostridium sp. (a genus of anaerobic bacteria) and methanogens, a group of Archaea which use hydrogen in their metabolism.

Evidence from the types of proteins that have been identified suggest that LUCA harvested chemicals across a gradient between the hot vent water and colder sea water to make ATP. If this is true, primitive life forms would have had to stay close to hydrothermal vents.

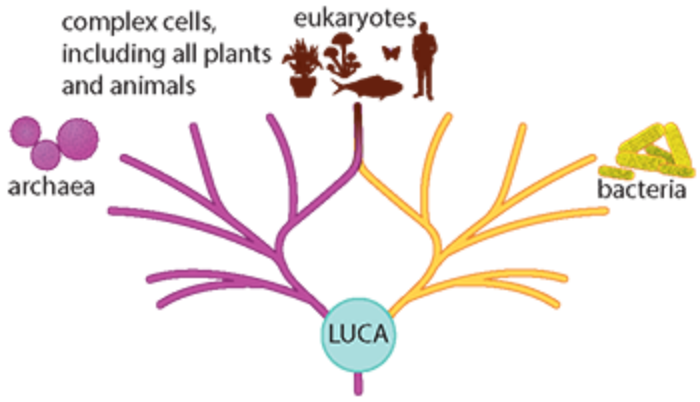


Figure 5.1.6: Scientists propose that LUCA has given rise to all organisms that we know today.

They would have been unable to survive anywhere else because they could not pump ions across a membrane to make ATP, as all organisms can do today. No genes have yet been found that could make amino acids, so the first organisms probably relied on amino acids that form naturally at hydrothermal vents.

EXTENSION

The exact date that LUCA appeared cannot be worked out precisely from the evidence that has been found and analysed. Early microbes used to exchange genes in a process known as **horizontal gene transfer** and some of these genes may have been misinterpreted as coming from the time of LUCA.

TEST YOUR UNDERSTANDING

4 State the properties of phospholipids that enable them to form a boundary between cell contents and the outside environment.

5 What are the key principles of cell theory?

6 Outline the inputs and products of the Miller–Urey experiment

7 Why is RNA regarded as being important part of the origin of life?

8 What is the main source of evidence for LUCA?

Links

- What features do all cells have in common (Chapter 5.2)
- Boundaries separate cells from their environment, but to what extent must they permit gases, nutrients and waste to cross them? (Chapter 6)

5.2 Cell structure

LEARNING OBJECTIVES

In this section you will:

- recall that living organisms are composed of one or more cells
- understand that unicellular organisms carry out all the functions of life
- recognise that multicellular organisms contain specialised tissues which develop as a result of differentiation
- discover that in multicellular organisms not all cells carry out all the functions of life
- learn that prokaryotic cells have a different structure from eukaryotic cells
- recognise that developments in microscopy have increased our understanding of cell structure
- learn that some cells have an atypical structure.

- understand that the endosymbiotic theory explains the origin of eukaryotic cells
- recall that differentiation is the process for developing specialised tissues

- understand that multicellularity has advantages for body size and specialisation

GUIDING QUESTIONS

- How are all cells similar and how do they differ?
- How do compartments in different cells differ between cells?

5.2.1 Cells and their structure

Today, scientists agree that the cell is the fundamental unit of all life forms. An individual cell can perform all the functions of life and anything that is not made of cells, such as viruses, cannot be considered living. Cell theory proposes that all organisms are composed of one or more cells and, furthermore, that cells are the smallest units of life.

Unicellular organisms

By definition, a living organism comprising just one cell has to perform all the necessary functions for survival.

The functions of life are:

1 metabolism – all the biochemical reactions of life

2 growth – increase in size or number of cells

3 response (or sensitivity) – the ability to react to external conditions

4 homeostasis – maintenance of a constant internal environment

5 nutrition – feeding or making substances needed for metabolism

6 reproduction – producing new individuals

7 excretion – removal of waste productions of metabolism.

A unicellular organism such as *Paramecium* (Figure 5.2.1) needs to **metabolise** organic materials in order to make the chemicals needed to sustain life. It must also be able to **excrete** waste

produced during metabolism and dispose of it. It must be able to detect changes in its environment, so it can respond to more favourable or less favourable conditions. Some unicellular organisms photosynthesise and they have a light spot that enables them to move to a brighter environment to maximise photosynthesis. A unicellular organism must also be able to control its internal environment (**homeostasis**), as large changes in water or salt concentrations may have a detrimental effect on metabolism and other cellular functions. It must also obtain food, whether produced from simple inorganic substances through photosynthesis (as in *Chlorella*; Figure 5.2.2) or ingested as complex organic materials from outside as a source of nutrition. If the species is to survive, an organism must be able to reproduce. This could be either asexual or sexual reproduction.

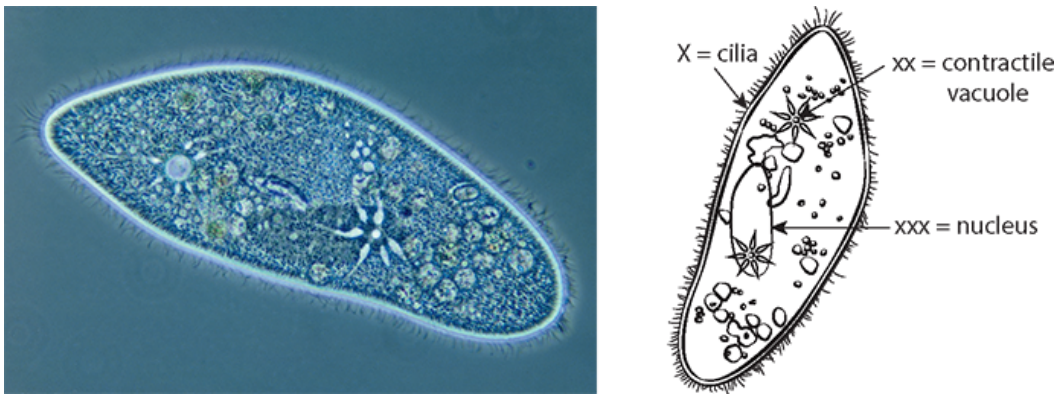


Figure 5.2.1: *Paramecium* carries out all the life functions within its single cell. A paramecium is about 0.5 mm long (x 100 magnification).

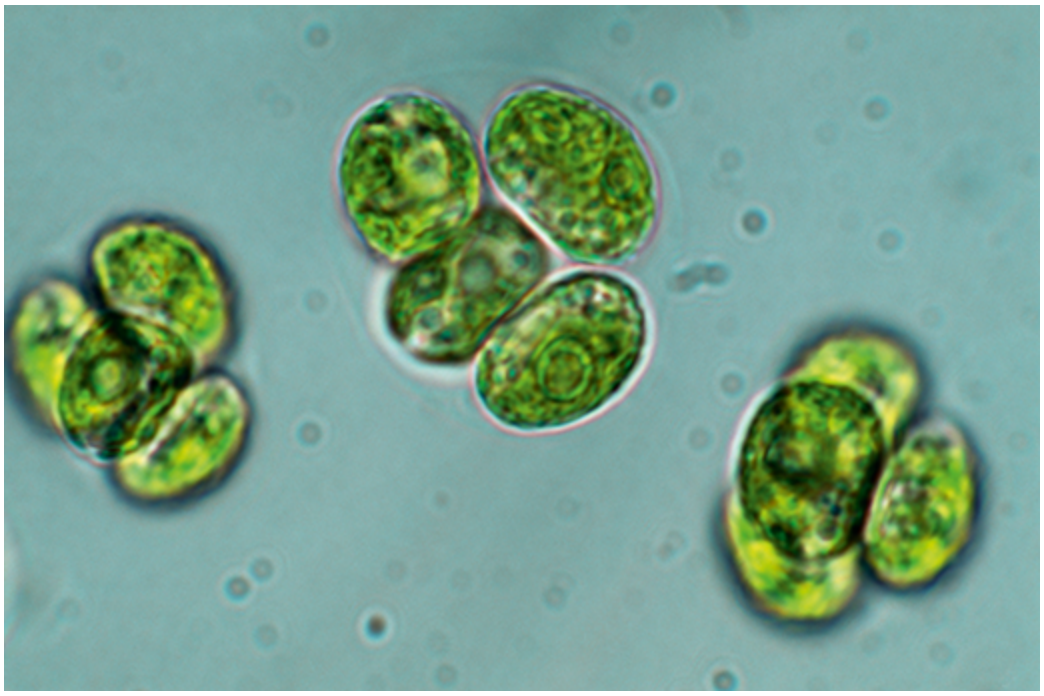


Figure 5.2.2: *Chlorella* is a unicellular organism containing a chloroplast (× 1200 magnification).

Multicellular organisms and differentiation

Multicellular organisms such as animals, plants and fungi are made of many cells and can be much larger and more complex than simple unicellular organisms. Becoming multicellular has enormous advantages. An organism can grow in size and its cells can undergo **differentiation**, that is, each cell can take on specific functions, so the organism can grow in complexity as well as size. Differentiation involves the expression of some genes from the organism's genome in the cell, but not others.

A multicellular organism may have specialised nerve cells for communication and interaction with the outside, and muscle cells for movement. It may also have special reproductive cells and secretory cells that produce enzymes for digestion. New properties emerge as a result of differentiation. Different cell

types interact with each other to allow more complex functions to take place. For example, nerve cells may interact with muscle cells to stimulate movement.

How do cells in the same organism behave in different ways when they all arose from the same parent cell and so have the same genome (genetic make-up)? In a particular organism, nerve cells and muscle cells all have the same genes, but look and behave very differently. The logical answer is that in some cells certain genes are expressed that are not expressed in other cells, and vice versa. For example, a pancreatic cell will express genes for the production of digestive enzymes or insulin, but a skin cell will not. Differentiation involves the expression of some genes from the organism's genome in the cell, but not others (Chapter 4).

Prokaryotic cells

Living things are divided into two types – prokaryotes and eukaryotes – according to the structure of their cells. Prokaryotic cells are usually much smaller than eukaryotic cells and have a much simpler structure. Prokaryotes are thought to be the first cells to have evolved. Bacteria are all prokaryotic cells.

Prokaryotic cells are so called because they have no nucleus ('prokaryote' comes from the Greek, meaning 'before the nucleus') and their cell functions do not take place in separate compartments in the cytoplasm. From the mid-20th century, when the electron microscope was developed, it became possible to study the internal detail of cells. Figures 5.2.3 and 5.2.4 show the main features of a typical prokaryotic cell.

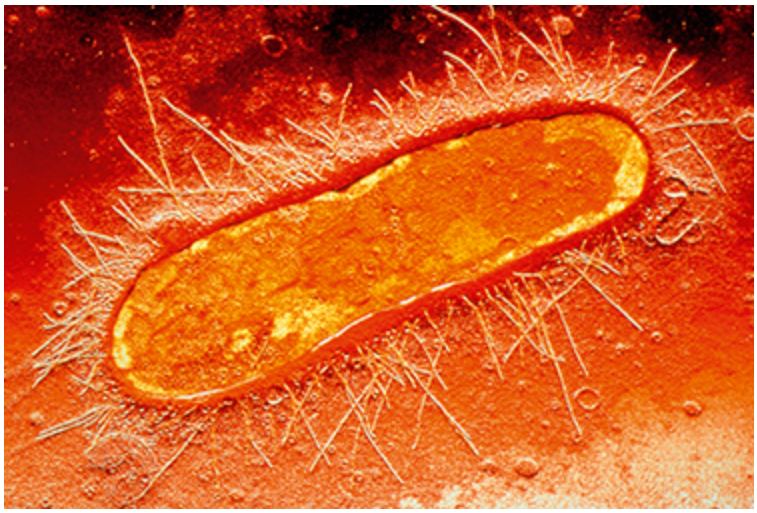


Figure 5.2.3: The bacterium *Escherichia coli* is a typical prokaryotic cell. (Coloured transmission electron micrograph, × 30 000 magnification.)

- The **cell wall** surrounds the cell. It prevents the cell from bursting and is composed of peptidoglycan, which is a mixture of carbohydrate and amino acids.
- The **plasma membrane** controls the movement of materials into and out of the cell. Some substances are pumped in and out using active transport.
- **Cytoplasm** inside the membrane contains all the enzymes for the chemical reactions of the cell. It also contains the genetic material.
- The chromosome is found in a region of the cytoplasm called the nucleoid. The DNA is not contained in a nuclear envelope and it is also ‘naked’, that is, not associated with any proteins. Bacteria also contain additional small circles of DNA called plasmids. Plasmids replicate independently and may be passed from one cell to another. They also may transmit resistance to antibiotics (Chapter 10).

- The **cytoskeleton** is a network of protein fibres which form a scaffolding to give the cell shape and allow substances to be directed through the cell.
- Ribosomes are found in all prokaryotic cells, and they synthesise proteins. They can be seen in very large numbers in cells that are actively producing protein. Prokaryotes have 70S ribosomes, which are smaller than those found in eukaryotes.
- A **flagellum** is present in some prokaryotic cells. A flagellum, which projects from the cell wall, enables a cell to move.
- Some bacteria have **pili** (singular **pilus**).These structures, found on the cell wall, can connect to other bacterial cells, drawing them together so that genetic material can be exchanged between them.

EXTENSION

Ribosome sizes

The 'S' (Svedberg) unit is used to define the size of a ribosome. It is a measure of the behaviour of particles during sedimentation. 70S and 80S ribosomes are different in size and so take different times to sediment when they are centrifuged. They are said to have different sedimentation coefficients.

Prokaryotic cells are usually much smaller in volume than more complex cells because they have no nucleus. Their means of division is also simple. As they grow, their DNA replicates and separates into two different areas of the cytoplasm, which then

divides into two. This is called binary fission. It differs slightly from mitosis in eukaryotic cells (Chapter 6).

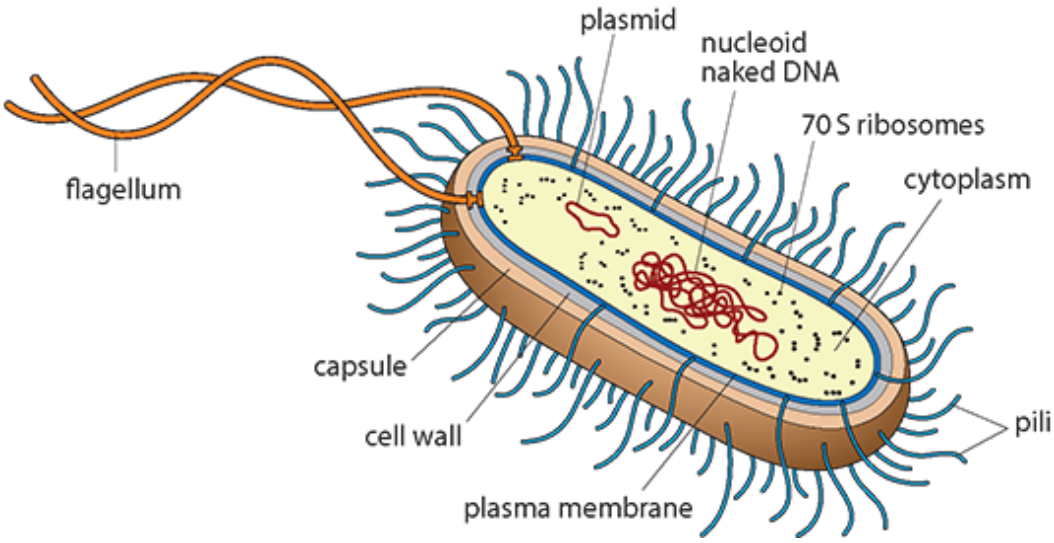


Figure 5.2.4: The structure of a typical prokaryotic cell.

Eukaryotic cells

Eukaryotic organisms have cells that contain a nucleus. Animals, plants, fungi and protoctista all have eukaryotic cells.

The complexity of a eukaryotic cell cannot be fully appreciated using a compound light microscope. But in images made using an electron microscope, which has a much higher resolution, the fine details of many different organelles are visible (for more on microscopy, Section 5.2.3). Figure 5.2.5 shows what can be seen of animal and plant cells using a light microscope; compare these images with the electron micrographs and interpretive drawings in Figures 5.2.11–5.2.14.

Electron micrographs like the ones on p. 200 reveal far more detail. Eukaryotic cells contain structures called **organelles**, each of which forms a ‘compartment’ in which specific functions take place. This compartmentalisation enables a eukaryotic cell to

carry out various chemical reactions or processes in separate parts of the cell, which all form part of the same system. Different types of cell have different organelles in different proportions, depending on the role of the cell.

The largest and most obvious structure in a eukaryotic cell is the **nucleus**, which contains the cell's chromosomes. Chromosomes are composed of DNA combined with histone protein, forming a material known as chromatin. The nucleus is surrounded by a double-layered membrane, the **nuclear envelope**. Small gaps in the envelope, called nuclear pores, are visible and it is through these that material passes between the nucleus and the rest of the cell. A distinctive feature of the nucleus is the darkly staining **nucleolus**. This is the site of ribosome production.

Associated with the nuclear envelope is a series of membranes known as the **endoplasmic reticulum** (ER). Ribosomes attach to this network to form rough endoplasmic reticulum (rER), the site of protein synthesis. As proteins are produced, they collect in the spaces between the membranes, known as the **cisternae**. From here they can be transported in vesicles to other parts of the cell such as the Golgi apparatus. ER that has no ribosomes attached is known as smooth endoplasmic reticulum (sER). The membranes of sER have many enzymes on their surfaces. Smooth ER has different roles in different types of cell: in liver cells, it is where toxins are broken down; in the ovaries, it is the site of estrogen production. Smooth ER also produces phospholipids for the construction of membranes and lipids for use in the cell.

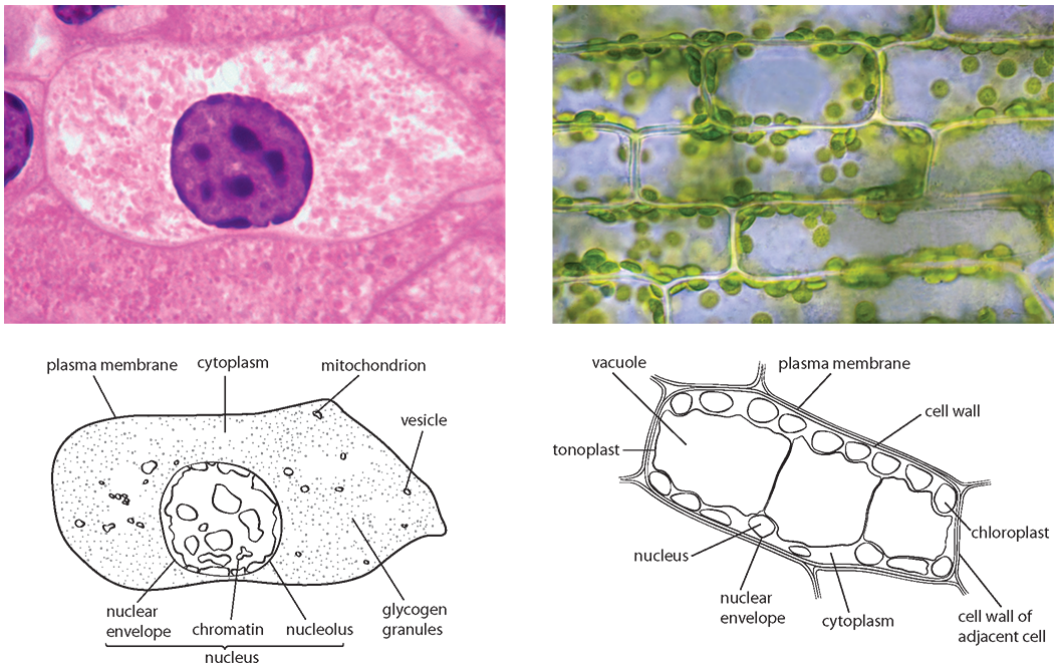


Figure 5.2.5: Photographs and interpretive drawings to show typical animal and plant cells as they appear using a light microscope.

The **Golgi apparatus** is similar in appearance to the sER, composed of stacks of flattened, folded membranes. It processes proteins made in the rER, collecting, packaging and modifying them, and then releasing them in vesicles for transport to various parts of the cell or for secretion from the cell. The pancreas contains many secretory cells, which have large areas of Golgi apparatus (Figure 5.2.14).

Eukaryotic cells also contain **mitochondria** (singular **mitochondrion**). These are elongated structures surrounded by a double membrane that are found throughout the cytoplasm. Mitochondria are known as the cell's 'powerhouses' because they are the site of aerobic respiration. The inner membrane is folded to form **cristae**, which greatly increase the surface area

for the production of ATP in the cell. Cells that respire rapidly, such as muscle cells, have numerous mitochondria.

Lysosomes are spherical organelles with little internal structure, which are made by the Golgi apparatus. They contain hydrolytic enzymes for breaking down components of cells. They are important in cell death, in breaking down old organelles and, in white blood cells, digesting bacteria that have been engulfed by phagocytosis. Plant cells do not normally contain lysosomes.

Centrioles are tube-like structures composed of the protein tubulin. They are involved in organising microtubules in the cytoplasm (Figure 5.2.6) and the location of the centriole controls the position of the nucleus and other structures in the cell. Centrioles are also involved in the formation of the spindle during cell division and in the completion of cytokinesis (Section 6.5)

The cytoskeleton is a network of protein filaments which connect different areas of the cell (Figure 5.2.6). Its primary function is to support the cell and maintain its shape. It is also involved in cell signalling (Section 7.1), and serves as a highway for the transport of materials and movement of organelles inside the cell. Three types of filament are present, actin microfilaments, intermediate filaments and microtubules made of polymers of tubulin.

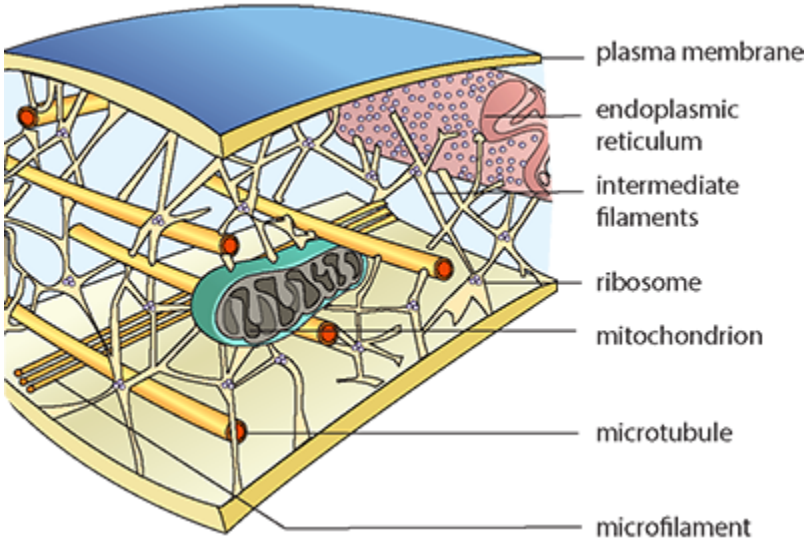


Figure 5.2.6: The cytoskeleton supports cell shape and enables transport within the cell.

Ribosomes are the site of protein synthesis in cells. They may be free in the cytoplasm or attached to the rER. They are made of RNA and protein but they do not have a membrane around them. Eukaryotic cells contain 80S ribosomes, which are larger than those found in prokaryotes.

As in prokaryotic cells, the plasma membrane controls the movement of materials into and out of the cell, and the gel-like cytoplasm, which fills much of the volume of the cell, provides a medium for many metabolic reactions.

Plant cells have three additional structures. All plant cells have an outer cellulose cell wall and most have a large central **vacuole** which contains sap. Some plant cells, such as palisade mesophyll cells (see Figures 5.2.16 and 5.2.17 in the section on electron microscopy), contain chloroplasts. The **chloroplasts** are found in cells exposed to the light, as they are the sites of photosynthesis. Chloroplasts have a double membrane and are about the same size as bacteria. Both chloroplasts and mitochondria have their

own DNA and ribosomes and are able to reproduce independently of the cell.

Fungi are also eukaryotes and they also have cell walls but these are built of different materials. Fungal cell walls are made of a matrix of chitin, glucans and protein. Fungal cells are linked together to form threads called hypha.

The large central vacuole contains water and salts. The membrane that surrounds it is under pressure from within and exerts a force on the cytoplasm, which in turn exerts a force on the cell wall, making the cell turgid and firm. The outer cell wall is composed of cellulose and other carbohydrates such as lignin and pectin, giving plant cells further support and a more rigid structure than animal cells. The cell walls and turgidity of plant cells give strength and support to tissues like leaves, holding them in the optimum position to catch the energy from sunlight for photosynthesis.

Although they are both types of eukaryotic cell, there are several key differences between animal and plant cells. These are summarised in Table 5.2.1.

Differences between prokaryotic and eukaryotic cells

Comparisons of images of prokaryotic and eukaryotic cells show numerous differences between them. These are summarised in Table 5.2.2. Note the difference in size of ribosomes between prokaryotic and eukaryotic cells and the presence of compartments in eukaryotes.

EXAM TIP

If you are asked to make comparisons in an exam question, use a table to organise your information clearly and neatly.

Animal cells	Plant cells	Fungal cells
cell wall absent	cellulose cell wall present	cell wall made of chitin present
small vacuoles sometimes present	large central vacuole present in mature cells	large central vacuole
no chloroplasts	chloroplasts often present	no chloroplasts
cholesterol in plasma membrane	no cholesterol in plasma membrane	no cholesterol
centrioles present to aid mitosis	centrioles absent	centrioles absent
store food reserves as glycogen	store food reserves as starch	store food reserves as glycogen

Table 5.2.1: Differences between animal, plant and fungal cells.

Structure	Eukaryotic cell	Prokaryotic cell
nucleus	usually present, surrounded by a nuclear envelope and containing chromosomes and a nucleolus	no nucleus, and therefore no nuclear envelope or nucleolus
mitochondria	usually present	never present
chloroplasts	present in some plant cells	never present
endoplasmic reticulum	usually present	never present

cytoskeleton	contains three components: microtubules, intermediate filaments and microfilaments	present but contains slightly different proteins to those found in eukaryotes
ribosomes	relatively large, about 30 nm in diameter, or 80S	relatively small, about 20 nm in diameter, or 70S
chromosomes	DNA arranged in long strands, associated with histone proteins	DNA present, not associated with proteins, circular plasmids may also be present
cell wall	always present in plant cells, made of cellulose, never present in animal cells	always present, made of peptidoglycan
flagella	sometimes present	some have flagella, but these have a different structure from those in eukaryotic cells

Table 5.2.2: Differences between prokaryotic and eukaryotic cells. The unit 'S' is a Svedberg unit, used to compare sizes of cell organelles.

NATURE OF SCIENCE

New techniques enable new discoveries to be made

Until the 1990s it was believed that prokaryotic cells did not contain a cytoskeleton but information gathered from bioinformatics, structural data and new methods of cell imaging have provided more convincing evidence that both bacteria and archaea have active cytoskeletons with proteins that are similar to tubulin and actin found in eukaryotes.

You can read more about this in the *Journal of Cell Biology*. (Wickstead, B., & Gull, K. (2011). The evolution of the cytoskeleton. *The Journal of Cell Biology*, 194(4), 513–525).

Atypical cells

Extensive examination of many organisms and millions of different types of cell has found that most eukaryotic cells have a similar structure and pattern. But a few cell types have some differences that make them atypical. One example is fungi, whose structures consist of long threads called hyphae (Figure 5.2.7), which have many nuclei but are not divided into separate cells by cell walls.

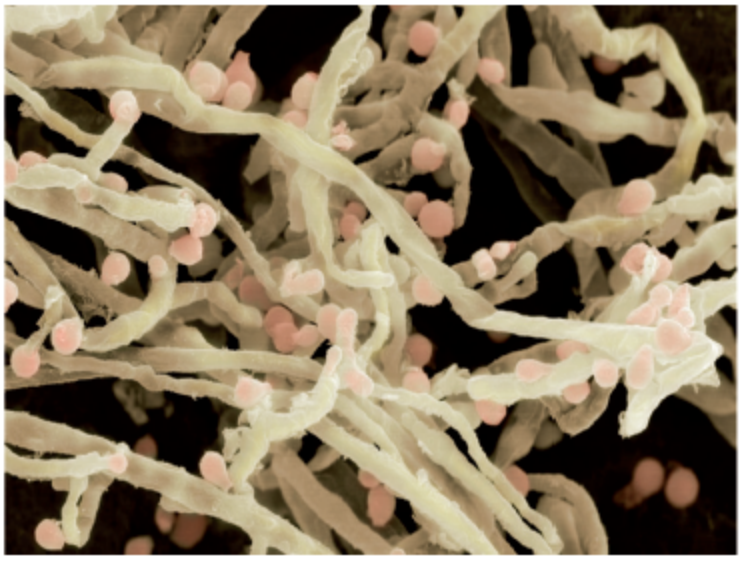


Figure 5.2.7: Fungal hyphae grow through material that nourishes the fungus. In this scanning electron micrograph the thread like structures are the hyphae and the pale pink spheres are the reproductive spores (× 2000).

Another example is skeletal muscle, which is composed of muscle fibres that are much larger than a single cell and contain several hundred nuclei (Chapter 9). Bone cells are also unusual because they have a matrix of extracellular material around them that seems to be greater than the cells themselves. Mammalian erythrocytes (red blood cells) are also atypical as they do not contain nuclei once they have matured and have been released into the bloodstream, which means that they cannot carry out all the functions of life at this stage of their life cycle.

Xylem and phloem

Cells in some organisms have a structure which is atypical but which allows them to perform their functions within the organism. Xylem and phloem are the two components of a plant's transport system. They are found in vascular bundles in stems and leaves. Although they both develop from living, dividing cells they do not have a typical cell structure and once they are mature they cannot divide. Their development begins in the apical meristems, which are specific parts of a plant in buds and root tips where cell growth and differentiation can occur. **Meristem** cells are totipotent and can produce all the cells and organs in a mature plant. Xylem cells are arranged end to end, forming tubes of tissue that transport water and minerals from the roots to other parts of the plant. Mature xylem is dead and consists simply of cell walls, thickened with lignin, with no separations between adjacent cells, and no cell contents or

cytoplasm (Figure 5.2.8). These cells can neither divide nor metabolise.

Phloem carries dissolved food molecules, mainly in the form of sugars, around the plant. Phloem is living tissue and like xylem is part of the vascular bundles. The main part of the phloem has tubular cells, known as sieve tubes, which have lost most of their organelles but remain alive (Figure 5.2.9).

Beside each one is a companion cell that is responsible for keeping the sieve element alive. The ends of each sieve tube cell are perforated so that solutes can pass between adjacent cells.

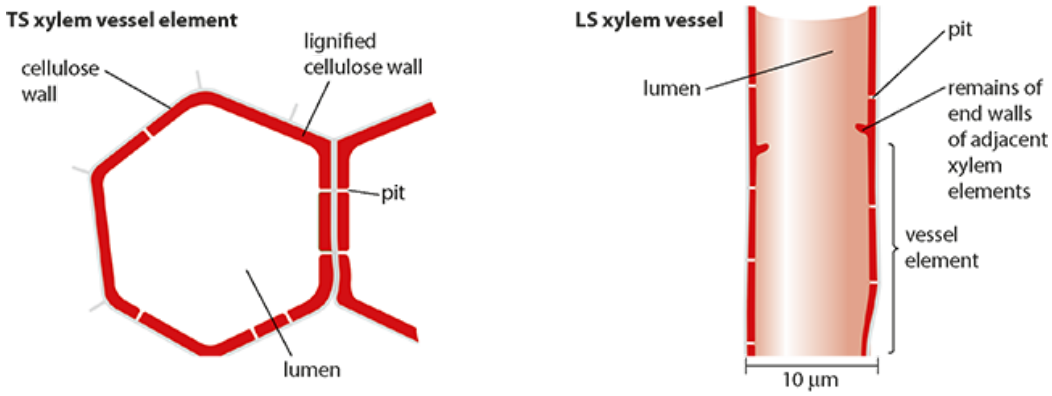


Figure 5.2.8: Cross-sections and longitudinal sections of xylem tubes. LS, longitudinal section; TS, transverse section.

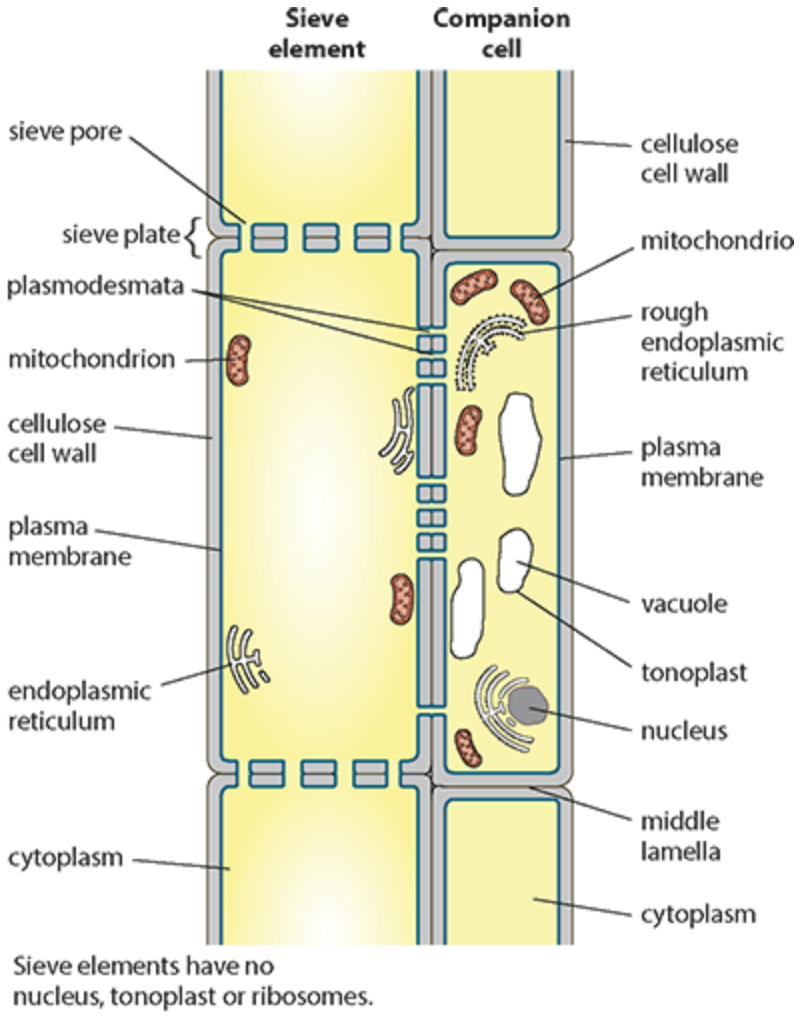


Figure 5.2.9: A phloem sieve tube element and its companion cell.

The features of xylem and phloem are outlined in Table 5.2.3.

Xylem	Phloem
composed of a column of dead cells, once mature, cell end walls removed	composed of a column of living cells with perforated walls between them
continuous tube of cells enables an unbroken column of water (held together by	living cells enable substances to be loaded by active transport

cohesive forces) to move inside the xylem	
thickened with lignin to withstand negative pressure as water vapour is lost in transpiration	associated with companion cells which carry out cell functions and supply energy for active transport into the phloem
transports water and minerals passively from roots to leaves	transports sugars, amino acids, hormones to all parts of the plant by mass flow

Table 5.2.3: The features of xylem and phloem.

5.2.2 The endosymbiosis theory

The theory of endosymbiosis explains how eukaryotic cells could have developed from a simple cell or prokaryote. The theory suggests that some organelles found inside eukaryotes were once free-living prokaryotes, and that the relationship is a form of endosymbiosis. There is evidence to suggest that some prokaryotes were engulfed by larger cells, and were retained inside their membranes where they provided some advantages to the larger cell (Figure 5.2.10).

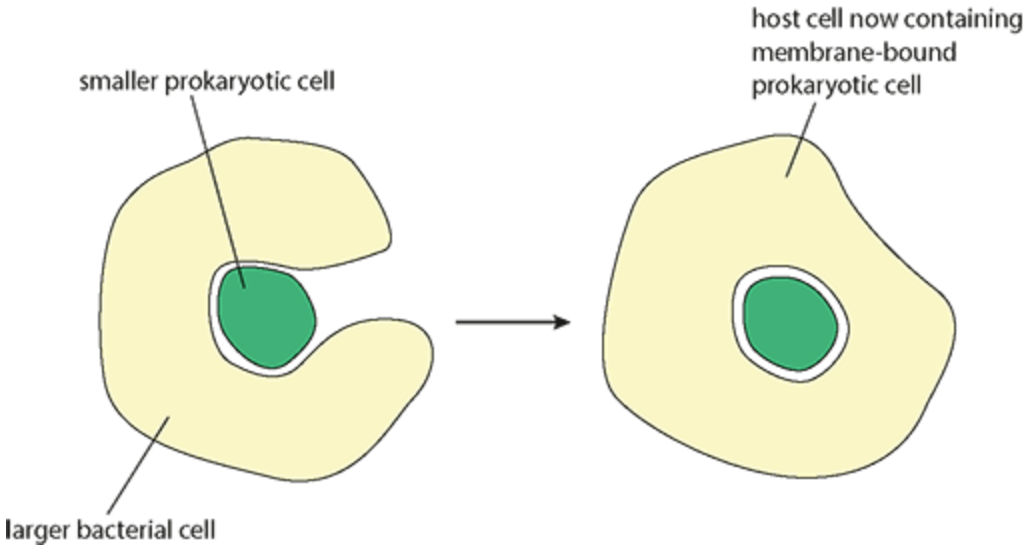


Figure 5.2.10: Organelles such as chloroplasts may have originated from free-living prokaryotes that were engulfed by larger cells.

Evidence for this theory includes the fact that two important organelles, mitochondria and chloroplasts, share many characteristics with prokaryotic cells. Both chloroplasts and mitochondria:

- contain ribosomes that are smaller than those found in other parts of eukaryotic cells but are identical in size to those

found in bacteria

- contain small circular pieces of DNA resembling bacterial plasmids in their basic structure
- have their own envelope surrounding them, on the inner membrane of which are proteins synthesised inside the organelle, suggesting that they may have used this ability long ago when they were independent organisms
- can replicate themselves by binary fission.

This evidence supports the theory that these organelles are modified bacteria that were taken in by phagocytosis, early in the evolution of eukaryotic cells. Here they became useful inclusions. The double outer envelope of chloroplasts and mitochondria may have originated from the bacterial plasma membrane together with the membrane of an engulfing phagocytic vesicle. Perhaps some of the enclosed bacteria carried pigment molecules on their membranes and used light energy to make organic molecules and release oxygen; these may have become chloroplasts. It may be that others became efficient at using the oxygen molecules for aerobic energy production, and these became mitochondria.

KEY POINT

endosymbiosis symbiosis means ‘life together’ and endo means ‘inside’ so endosymbiosis describes a relationship taking place inside a cell.

TEST YOUR UNDERSTANDING

9 List three differences between prokaryotic and eukaryotic cells.

10 State one advantage a cell gains from having organelles.

11 Outline the function of the endoplasmic reticulum.

12 What is the evidence for the theory of endosymbiosis?

Cell differentiation and multicellularity

Multicellular organisms have cells which perform different functions. Cells which differentiate to carry out the same function form groups called **tissues**. Tissues contains cells which have changed and developed to carry out their roles, for example muscle tissue develops the enzymes and contractile proteins needed for contraction. Tissues enable multicellular organisms to function more efficiently than they would if all their cells carried out the same jobs and specialised tissues develop by cell differentiation.

Cell differentiation is the process in which a cell undergoes changes in gene expression to become a more specific type of cell. Differentiation allows multi-cellular organisms to produce different cell types and body shapes. The process of cell differentiation is controlled by genes, and the interaction of genes with the environment.

NATURE OF SCIENCE

How strong and reliable is a theory?

A theory is a proposed explanation for observations of the natural world. Theories are made using scientific methodology and usually bring together several facts and hypotheses. The theory of endosymbiosis will always remain a theory because we cannot go back in time and observe whether our theory is true. Nevertheless, some theories can be

regarded as better and more reliable than others. Scientists collect evidence and make observations to explain and support their theories. The more evidence that is gathered to support the theory, the more likely it becomes to be accepted as reliable. Scientists have gathered a lot of reliable evidence to support the theory of endosymbiosis and today it is generally accepted as a good explanation of how eukaryotic cells could have arisen.

To consider:

1 Think about other theories in biology. Are they strong and reliable or could they easily be disproved?

All the cells in a multicellular organism have the same set of genes, even though the cells have many differences in their structure and their jobs. There are about 25,000 genes in a human cell but not all these genes are switched on, or expressed, in every cell. There are just over 200 cell types in a human body and, as a cell develops into a liver cell or a nerve cell, different genes are switched on or off. As different genes are expressed, cells differentiate into all the different types that we can observe. Controlling gene expression is crucial to development. Cell differentiation is triggered by environmental factors that affect gene expression. These factors may be in the internal environment, for example as a cell develops in an embryo its position will determine which genes are expressed. In other cases, hormones can influence what form a cell will take. Or the influences may be in the external environment where available nutrients, salinity, and temperature affect differentiation. For example, in Himalayan rabbits, the genes that code for fur colour are turned on and off depending on temperature. Cold conditions lead to the expression of darker pigmentation. The colder areas

of the rabbits' bodies, their paws, nose and ears are black while their body fur is white.

Cells which have not differentiated and retain the ability to develop into many different cell types are known as stem cells (Section 8.1.2).

Evolution of Multicellularity

As Fig 5.1.5 shows multicellular organisms have evolved and developed along many different lines. Many fungi, eukaryotic algae and all plants and animals are multicellular which gives these organisms many advantages over single-celled organisms. A multicellular organism can grow to a larger size than a single celled organism because it is not limited by diffusion of substances into its cells. A single-celled organism is limited by its surface area to volume ratio. A large single cell cannot absorb and transport nutrients efficiently because, as it grows larger, its surface area to volume ratio decreases (Section 6.4.1). Multicellular organisms with many different cell types can become more complex as different cells take on different jobs. They can also have longer life spans because if one cell dies the whole organism does not die.

5.2.3 Developments in microscopy

Light microscopes

Light microscopes, also called optical microscopes, enable us to see magnified images of objects placed on the microscope stage (Figure 5.2.11). Visible light is passed through the specimen and a system of lenses produces magnified images which can be seen by looking through the eyepiece. One of the few cells large enough to be visible to the unaided eye is the mature human ovum, which has a diameter of approximately 150 μm. However, most cells are much smaller than this, and can only be seen using a microscope. The most powerful light microscopes can magnify up to 2000 times and reveal some internal structures such as the nucleus (Figure 5.2.1), but to see greater detail more powerful microscopes such as the electron microscope, which magnifies cell structures up to 500 000 times must be used.

The advantage of light microscopes is that they can be used to examine whole specimens or sections of them, and either living or dead material. There are many staining and lighting techniques that can be used to show specimens in colour so that physical and biological features can be seen. Images can be photographed or videoed or computerised for later examination. Light microscopes are small and relatively inexpensive.

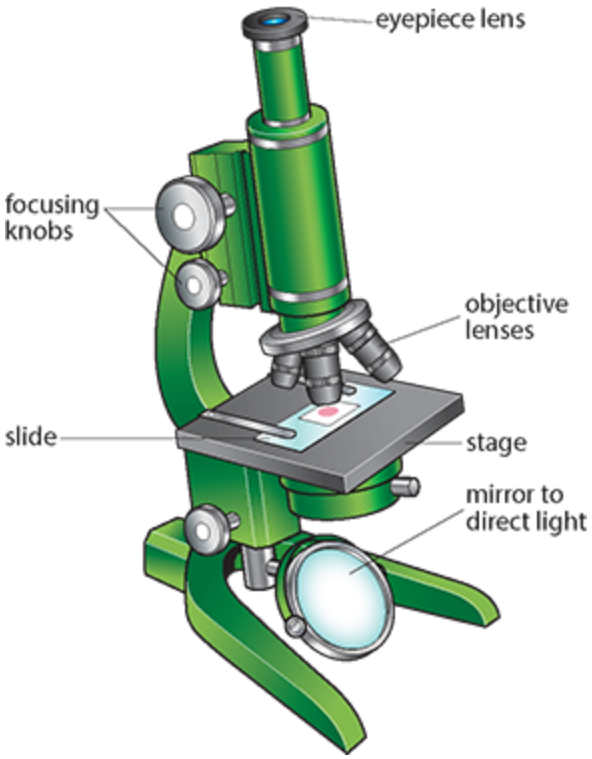


Figure 5.2.11: Typical compound light microscope.

Magnification and scale

Knowing the sizes of objects viewed under the microscope can be very useful, for example, a scientist might want to compare the relative sizes of pollen grains from plants in the same genus to help identify different species or to identify the range of sizes in a population of unicellular organisms.

Magnification is defined as the ratio of the size of the image to the size of the object:

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

KEY POINT

magnification refers to how many times larger an object appears compared with its actual size. It is calculated from the ratio of length of image to the length of the object.

With a compound microscope (as shown in Figure 5.2.11), the magnification is the product of both lenses, so if a microscope has a ×10 eyepiece and ×40 objective, the total magnification is ×400.

Printed images of structures seen with a microscope usually show a scale bar or give the magnification, so that the size of an object can be calculated. For example, the magnification of the micrograph in Figure 5.2.12 is given as ×165.

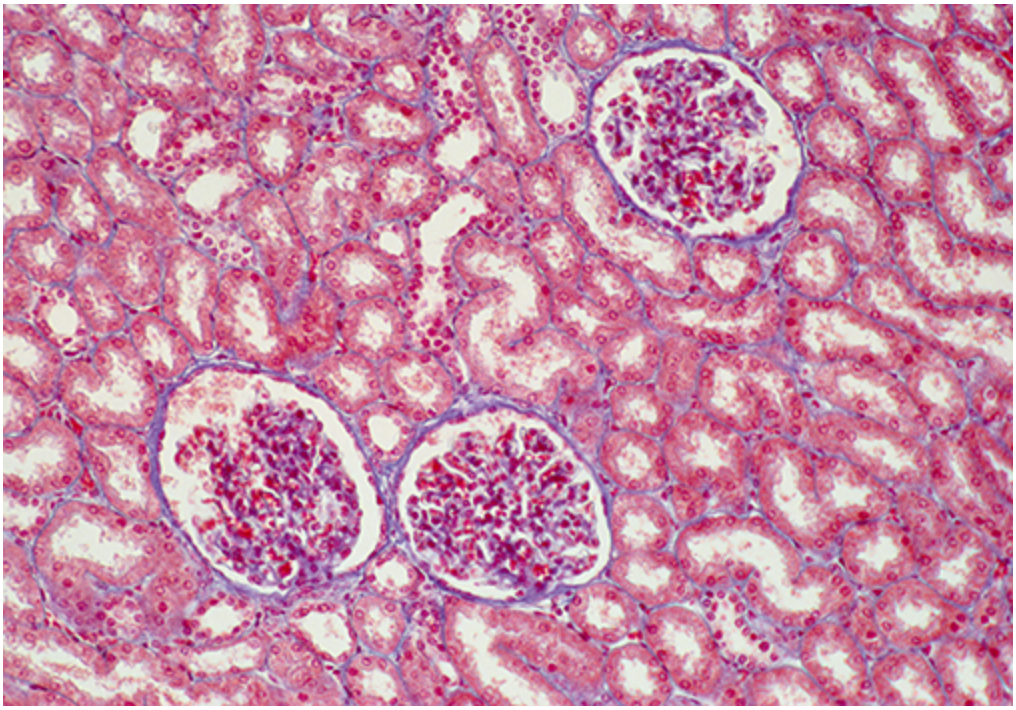


Figure 5.2.12: Coloured light micrograph of a section through the cortex of a kidney (×165).

In Figure 5.2.12, there are three spherical glomeruli present. In the image, each one is approximately 25 mm across. You can check this using a ruler. Thus:

$$\text{actual size of glomerulus} = \frac{\text{size of image}}{\text{magnification}}$$
$$= \frac{25\ \text{mm}}{165}$$
$$= 0.15\ \text{mm}$$

In photographs produced by an electron microscope, most measurements are expressed in micrometres. A micrometre (μm) is 10^{-3} mm, so 1 mm is 1000 μm.

So, the diameter of the glomerulus = 0.15 × 1000 = 150 μm.

WORKED EXAMPLE 5.2.1

The image below represents a red blood cell. The scale bar shows 2 μm. From this, you can calculate both the size of the cell and the magnification of the image.

Size of the cell

Step 1 Use a ruler to measure the length of the cell (its diameter in this case). This is 30 mm.

Step 2 Use a ruler to measure the length of the scale bar. This is 9 mm.

Step 3 Use the ratio of these two values to work out the actual length of the cell.

$$\frac{2\ \mu\text{m}}{9000\ \mu\text{m}} = \frac{\text{actual length of cell}}{30\ 000\ \mu\text{m}}$$

(Remember to convert all the units to μm. 1 mm = 1000 μm.)

Answer

Rearranging the equation:

$$\text{actual length of cell} = 2\ \mu\text{m} \times \frac{30\ 000\ \mu\text{m}}{9000\ \mu\text{m}}$$
$$= 6.7\ \mu\text{m}$$

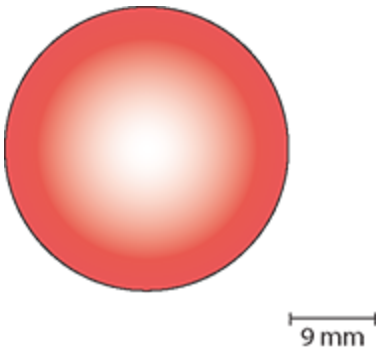
Magnification of the image

Use the formula:

$$\text{magnification} = \frac{\text{measured lenght of the cell}}{\text{actual lenght of the cell}}$$

So in the case:

$$\text{magnification} = \frac{30\ 000\ \mu\text{m}}{6.7\ \mu\text{m}}$$
$$= \times 4500$$



If you are given a value for the magnification, you can measure the length of the object in the image and then rearrange the equation to work out the actual length of the object.

KEY POINT

SI units – International System

1 metre (m) = 1 m

1 millimetre (mm) = 10^{-3} m

1 micrometre (μm) = 10^{-6} m

1 nanometre (nm) = 10^{-9} m

1 centimetre cubed = 1 cm^3

1 decimetre cubed = 1 dm^3

1 second = 1 s

1 minute = 1 min

1 hour = 1 h

concentration is measured in mol dm^{-3}

New techniques in light microscopy

Fluorescent and phosphorescent stains are non-protein molecules that absorb light at a specific wavelength and re-emit it at a longer wavelength from the visible part of the spectrum. In this way they produce coloured images. The technique of fluorescence microscopy has become an essential tool in biology and the biomedical sciences. Biomolecules such as proteins, antibodies and peptides can have fluorescent molecules added to them and these labelled molecules can be seen in the microscope (Figure 5.2.13). Different stains are used so that different parts of the same cell can be distinguished easily. In Figure 5.2.13 DNA in the cell nuclei can be distinguished from other cell structures.

KEY POINTS

fluorescent able to emit light immediately; usually visible under a light source such as UV light.

phosphorescent refers to a molecule that can store absorbed light for some time and release it later.

Phosphorescence is also used to monitor the delivery of medicine and drugs. Medications carrying fluorescent markers can be tracked as they move to specific tissues in the body.

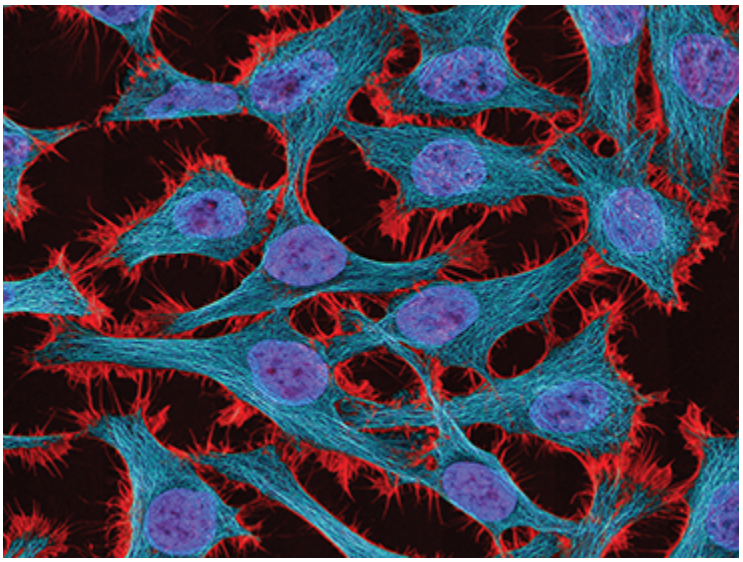


Figure 5.2.13: Fluorescence image of HeLa cells. Actin molecules are stained in red and the cell nuclei are stained blue. Microtubules in the cell are stained with cyan.

Electron microscopy

Electron microscopes use a beam of electrons, instead of light, to produce an image. The **resolution (resolving power)** of an electron microscope is much better than that of a light microscope because of the shorter wavelength of electrons

compared to light. Resolving power is the ability of the microscope to separate objects that are close together so that more detail can be seen. Figure 5.2.5 shows us the appearance of an animal cell and a plant cell. Different organelles and structures can be seen clearly and we can recognise the cell membranes and the plant cell wall. The diagrams (Figures 5.2.15 and 5.2.17) made from the electron micrographs help us to identify the various parts of the cells.

Only non-living material can be observed in an electron microscope and specimens must be specially prepared with heavy metals or coated with carbon or gold. Viruses can only be seen with these microscopes, so the structure of viruses was unknown until the invention of these microscopes in the 20th century (Section 5.3).

EXAM TIP

You should be able to recognise the organelles and structures present in cells and label diagrams like those shown in figures 5.2.14 and 5.2.16 with their names.

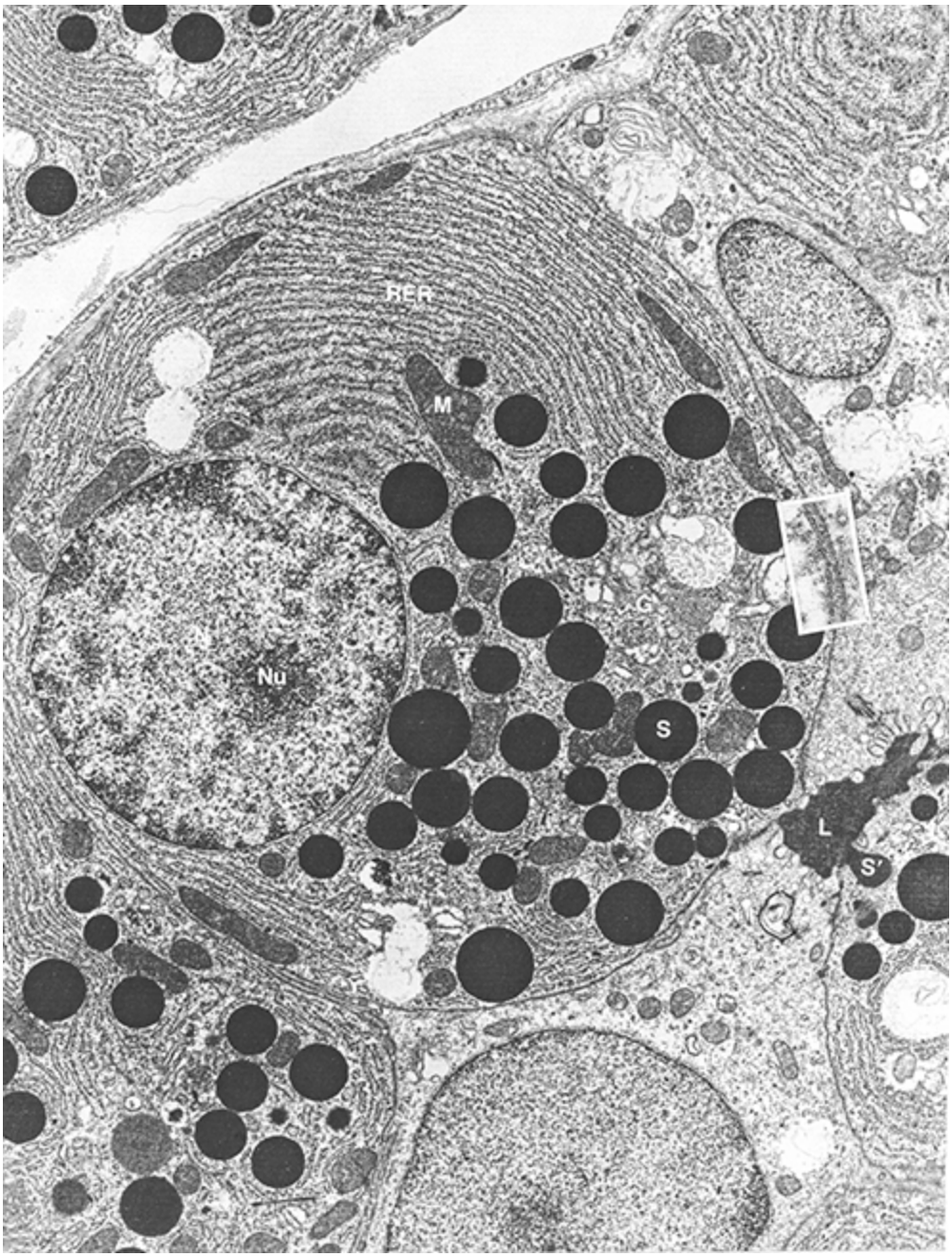


Figure 5.2.14: Electron micrograph of an exocrine cell from the pancreas (×12 000).

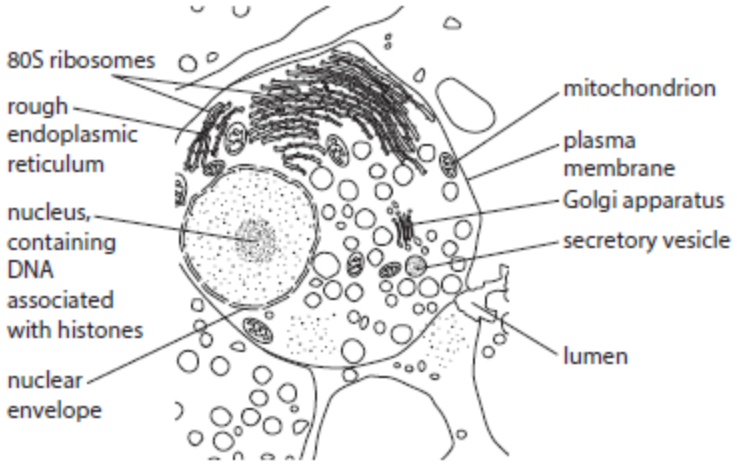


Figure 5.2.15: Interpretive drawing of some of the cell structures visible in Figure 5.2.14.

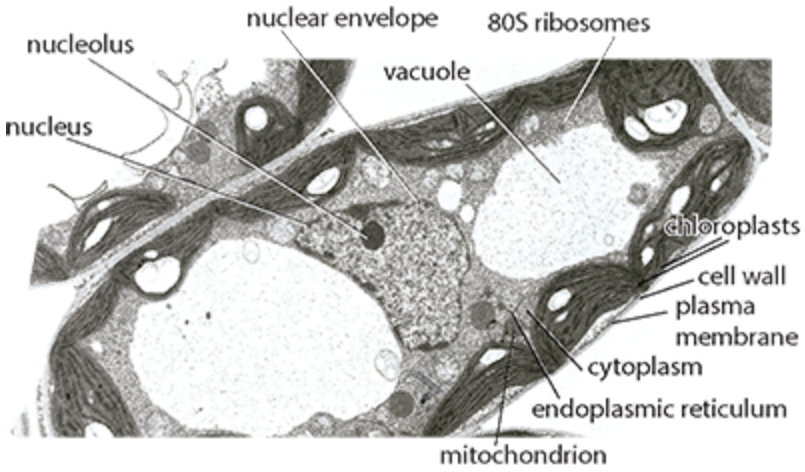


Figure 5.2.16: Electron micrograph of a palisade mesophyll plant cell (×5600).

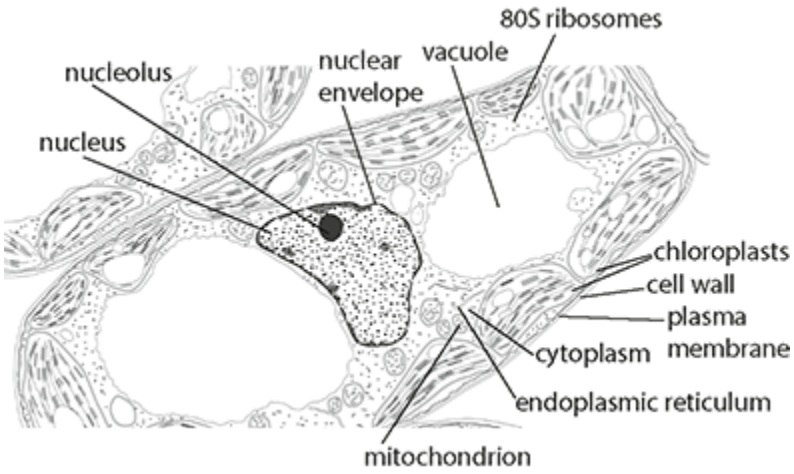


Figure 5.2.17: Drawing of a palisade mesophyll plant cell made from the electron micrograph in Figure 5.2.16.

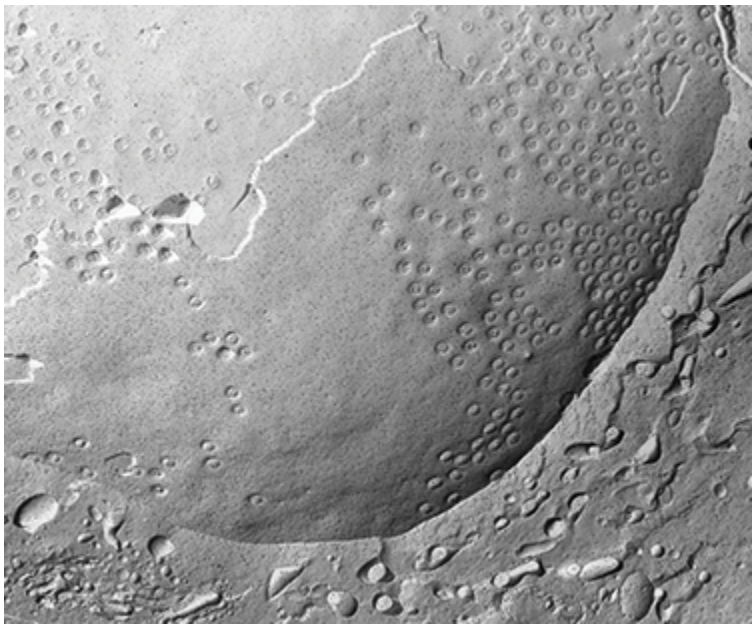


Figure 5.2.18: Freeze-fracture image of cell nuclear membrane.

There are two main types of electron microscope: the transmission electron microscope (TEM) and the scanning electron microscope (SEM). A TEM was the first electron microscope to be built. It produces clear images of very thin sections of material. A beam of electrons passes through a

specimen and is scattered, producing an image that can be viewed on a screen. A SEM directs a focused beam of electrons across a specimen and, as electrons are bounced off the surface, detailed images of the external shape and composition of the specimen appear. A SEM can create images of larger specimens and give a good idea of their real shape.

Electron microscopes produce black and white images that are often artificially coloured so that certain features can be seen more clearly, and different techniques are used to study different aspects of the living world.

Freeze-fracture electron microscopy is a technique that was developed in the 1960s. It has helped us to understand the structure of membranes more clearly. The technique involves rapidly freezing a specimen and then cracking it along a line through the tissues. Specimens will fracture along their weakest parts, usually the membranes or the surfaces of cell organelles. If the broken surfaces are shadowed with a film of platinum, a replica of the surface can be made and viewed in a TEM (Figure 5.2.18).

Until recently, even electron microscopes could not distinguish individual molecules and X-ray crystallography was always used to work out the structure and arrangement of atoms in molecules. But now, a newer technique, **cryogenic electron microscopy** (cryo-EM), is proving so successful that the majority of new molecular structures are worked out using it. Cryogenic electron microscopy allows scientists to work out the structures of biomolecules, especially proteins, without the need to produce crystals of them first. The technique was developed in the 1970s, but has become more useful as new software and algorithms have been developed and used to work out the structure of thousands of biomolecules. This type of microscope is used on

samples that are cooled to cryogenic temperatures (temperatures at which molecules cease moving) and embedded in amorphous ice, which has no crystalline structure of its own. In 2020, the Electron Microscopy Data Bank (EMDB), a database for molecular structures, added its 10 000th entry, and the vast majority of structures had been worked out using cryo-EM.

Table 5.2.4 compares the main features of different types of microscope.

	Light microscope	**Transmission electron microscope**	**Scanning electron microscope**
	uses light to produce images	uses electron beams to produce images	uses electron beams to produce images
Maximum resolution	200 nm	1 nm	1 nm
Maximum magnification	×2000	up to ×1 000 000	×200 000
Preparation of material	thin sections of material mounted on slides, living organisms can be examined	very thin sections of material supported on metal grids, living organisms cannot be examined	very thin sections of material supported on metal grids, living organisms cannot be examined
Stain used	coloured	heavy metals	carbon or gold

	dyes		coating
Image	viewed directly through eyepiece lens	viewed on a screen or photographic plate	viewed on a screen or photographic plate

Table 5.2.4: Comparison of light microscopes with the transmission electron microscope (TEM) and scanning electron microscope (SEM).

TEST YOUR UNDERSTANDING

13 Calculate how many cells of 100 μm will fit along a 1 mm line.

14 What property of fluorescent stains makes them useful in microscopy?

NATURE OF SCIENCE

Scientific advance follows technical innovation: the electron microscope

A typical animal cell is 10–20 μm in diameter, which is about one-fifth the size of the smallest particle visible to the naked eye. Robert Hooke was the first scientist to see and describe cells, although he didn't know what they were. Later, Anton van Leeuwenhoek, who built one of the first microscopes in 1674, was able to see living cells of *Spirogyra* and bacteria.

It was not until good light microscopes became available in the early part of the 19th century that plant and animal tissues were seen as groups of individual cells and Schleiden and

Schwann in 1838 were able to see sufficient structure to propose cell theory, which incorporated the work of their predecessors.

Animal cells are tiny and colourless so it was not until the end of the 19th century, when staining techniques were first used, that it was possible to see a little more detail of cell contents. In the early 1940s, far more powerful electron microscopes were used for the first time and organelles and greater complexity of cell structure could be studied. Developments proceeded more rapidly in the 20th century, because international communication allowed for more efficient collaboration, not only in the designing and building of microscopes but also in the discussion and understanding of what could be observed.

A light microscope can resolve (view separately) cell details that are about 0.2 μm apart. Resolution is limited by the wavelength of light so that bacteria and mitochondria (500 nm or 0.5 μm) are the smallest objects that can be seen. An electron microscope uses a beam of electrons to probe specimens and in theory it should be able to resolve structures that are 0.002 nm apart (a resolution 10 000 times that of a light microscope). For biological material this reduces to about 0.2 nm but, even so, an electron microscope allows a resolution which is 100 times better than a light microscope.

New techniques in microscopy such as electron microscopy such as cryo-EM are improving our understanding of cells at the molecular level and new labelling techniques using phosphorescent pigments and light microscopes have improved our knowledge of how cell structures are organised.

Technology has improved the range and precision of human senses in examining cells and led to a greater understanding of cells and how they function.

To consider:

1 Discuss other areas of science where technology has been important in improving human understanding of the natural world.

2 Why has international communication and collaboration been so important in technological advances?

THEORY OF KNOWLEDGE

Perception and interpretation: can we believe what we see?

Our own perception is a crucial source of knowledge. The way we see things depends on the interpretation of the information that our sense organs relay to our mind. What we perceive may be a selective interpretation.

When examining microscope images we must remember that preparing cells by staining them and taking sections through them will alter their appearance. Images we view through microscopes or on screens will be influenced by these methods as well as our own expectations about what we are looking at.

To consider:

1 Look at the shapes of the mitochondria in the electron micrographs in Figures 5.2.16 and 5.2.13. Why do some appear cylindrical and others circular?

2 Plant cells have a single central vacuole. Study the plant cell in Figure 5.2.17. How many vacuoles can you see? How can you explain this?

Links

- Why do cells have different structures? (Chapter 6)
- What advantages does compartmentalisation give to eukaryotes? (Chapter 6)
- How does cell structure help with classification and building evolutionary relationships? (Chapter 11)

5.3 Viruses

LEARNING OBJECTIVES

In this section you will:

- learn that viruses are acellular, have no metabolism and need to be inside a host cell to replicate themselves
- discover that viruses are much smaller than prokaryotic cells
- learn that viruses contain RNA or DNA, have a protein coat and that some have a viral envelope
- understand that viruses have different structures and modes of infection
- recognise lytic and a lysogenic cycles of different viruses
- recognise that the wide diversity of viruses suggests they have several origins, although all are obligate parasites
- learn that there are several hypotheses to explain the origins of viruses
- Understand that viruses such as HIV and influenza virus evolve very rapidly.

GUIDING QUESTIONS

- How are viruses similar to and different from living organisms?
- How diverse are viruses?

5.3.1 The structure of viruses

Viruses are not considered to be living things. They are said to be **acellular** because they have no metabolism and can only replicate inside a host cell. A virus consists of DNA or RNA enclosed inside a protective protein coat known as a **capsid**. Viruses vary in their structure, and the shape of the capsid varies between different types of virus. Capsid shape is one method used to classify viruses. Genes that are present in viral genetic material code for the capsid. The simplest viruses contain just enough DNA or RNA to code for four proteins, whereas the most complex can code for up to 200 proteins. Viral proteins self-assemble to form a capsid (Figure 5.3.1).

Some viruses are able to surround themselves with an envelope which is derived from their host cell's plasma membrane, endoplasmic reticulum or nuclear membrane. This **viral envelope** contains proteins coded for by both the virus genome and the host cell genome and can give the virus better protection from the host's immune system. Proteins in the envelope may include **glycoproteins**. These are proteins that have carbohydrate groups covalently bonded to their amino acid chains. Glycoproteins act as receptor molecules and can make it easier for viruses to bind to and enter host cells. Many enveloped viruses depend on their envelope to infect new cells and to enable them to survive for longer periods outside a host cell.

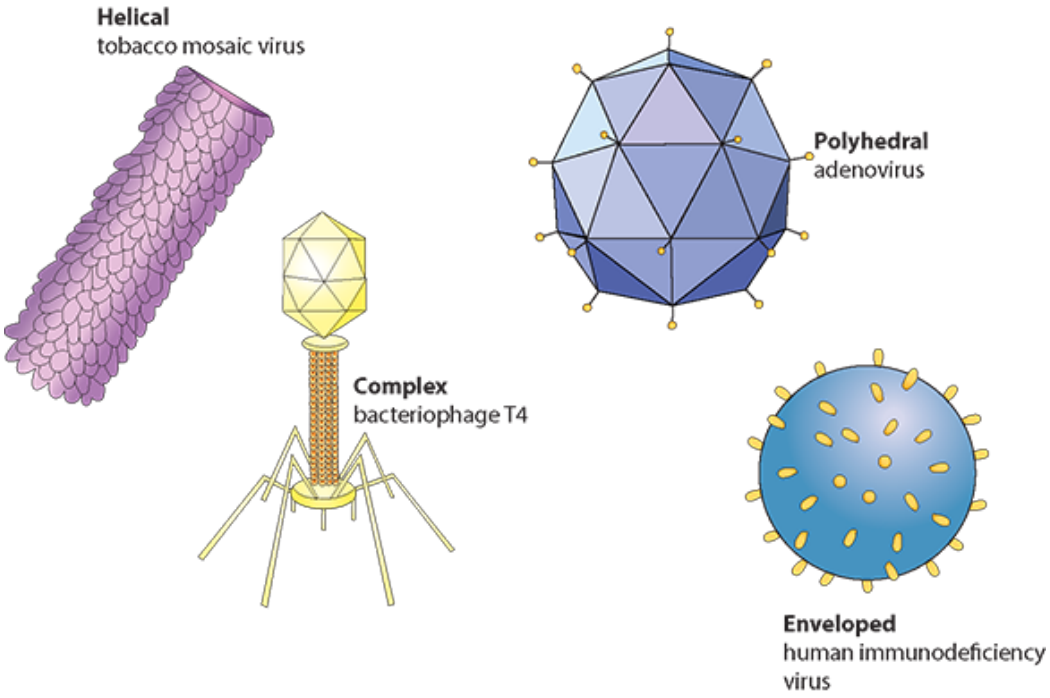


Figure 5.3.1: Virus capsids may be helical, polyhedral, complex or enveloped.

SCIENCE IN CONTEXT

Viruses and disease

Viruses are thought to cause approximately 60% of human infections, some like the common cold are mild while others such as Ebola and polio are serious and potentially fatal. Many thousands of viruses are known and more are emerging every year. The rise of zoonoses, diseases that pass from animals to humans, are causing particular concern.

The most common viral illnesses are infections of the digestive and respiratory systems. In developed countries, viruses are responsible for about 30% of all infectious gastroenteritis and it is estimated that adults have between two and five colds per year, while young children may have twice as many.

Viruses pass from an infected person directly to a new host by physical contact or in a cough or sneeze, or indirectly when a person picks up a virus from a contaminated surface. Viruses can be transmitted indirectly, because many can survive outside the body for short periods of time. Their survival time depends on the virus and the environment. Temperature, humidity and level of UV light can all cause viruses to degrade. But some retain their power to infect for several days or even longer on surfaces such as clothes, utensils or furniture. Viruses can be transferred from a contaminated surface via hands to food or into a person's mouth.

Respiratory viruses that cause influenza and the common cold are often found on surfaces that have been touched by infected people. Viruses may survive for up to 18 hours on places such as door handles, light switches, keyboards and phones and be transferred to the hands of a new host. During the coronavirus pandemic public health authorities advised strict hygiene measures to prevent the spread of viruses. These included masks, to prevent droplets being coughed or sneezed out, frequent hand washing to remove any viruses picked up from contaminated surfaces and sanitisation of surfaces likely to be touched by infected people.

To consider:

1 What effect has international travel had on the spread of virus diseases across the world?

2 How effective are public health campaigns in raising people's awareness of the need for hygiene to prevent disease?

KEY POINTS

lytic cycle reproduction by viruses that use a host cell to manufacture more viruses; the viruses then burst out of the cell e.g. bacteriophage lambda

lysogenic cycle incorporation of the viral genome into the host cell genome and infection from inside the cell e.g. HIV

Infection and replication

Viruses are much smaller than prokaryotic cells (Figure 5.3.2). They infect both prokaryotic and eukaryotic cells in a similar way. A viral surface protein binds with a receptor on the host cell surface and the virus enters the host cell and releases its nucleic acid from the capsid or envelope. Viruses then use the host cell machinery for their reproduction. They use host cell ribosomes to produce their proteins and many viruses modify host cell transcription and translation mechanisms to favour the production of viruses over normal cell activities.

The ways in which two important viruses, **bacteriophage lambda**, a virus which infects bacteria, and **human immunodeficiency virus** (HIV) which infects cells of the human immune system, enter their host cells are described shown in Figures 5.3.3 and 5.3.4 .

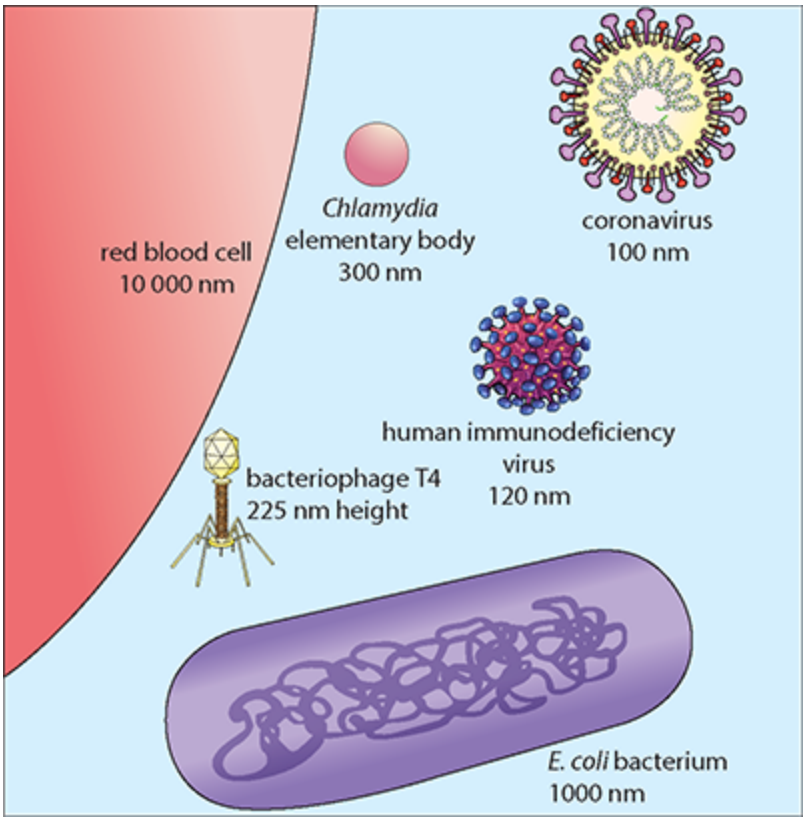


Figure 5.3.2: Relative sizes of prokaryotic cells, eukaryotic cells and viruses.

Bacteriophage lambda is a virus that infects *E. coli* bacteria. It has a lytic lifecycle. It is a complex virus with an icosahedral head, a tail and tail fibres and double-stranded DNA as its genetic material (Figure 5.3.1).

The stages of the life cycle of bacteriophage λ (Figure 5.3.3) include absorption and penetration into the bacterium, followed by synthesis of proteins and replication of viral DNA using the bacterial enzymes. The viral genetic material is enclosed in new capsids and finally the host cell wall breaks, releasing new viruses that go on to infect more bacterial cells.

Human immunodeficiency virus infects cells of the human immune system that have specific CD4 protein receptors on their

membranes. The cells that have these receptors are called the helper T cells, which communicate with other cells involved in antibody production to fight infections (Chapter 10). Without treatment, the numbers of these cells fall to such low levels that the body is unable to resist infections and may succumb to acquired immunodeficiency syndrome (AIDS). HIV is an enveloped retrovirus (Figure 5.3.1) containing a capsid, inside which are single-stranded RNA and viral enzymes. The envelope contains glycoproteins that attach to particular CD4 receptors on the host cell plasma membrane. The envelope fuses with the plasma membrane and the HIV contents are released into the host cell. The virus RNA is released along with viral enzymes. One of these is **reverse transcriptase**, an enzyme that converts the single-stranded RNA into double-stranded DNA. Another viral enzyme, integrase, then integrates the double-stranded DNA into the host cell genome. This virus has a lysogenic lifecycle. The host cell transcribes the viral genes and new viruses are produced and assembled. New viruses bud out through the plasma membrane of the host cell (Figure 5.3.4) and are released to infect more helper T cells.

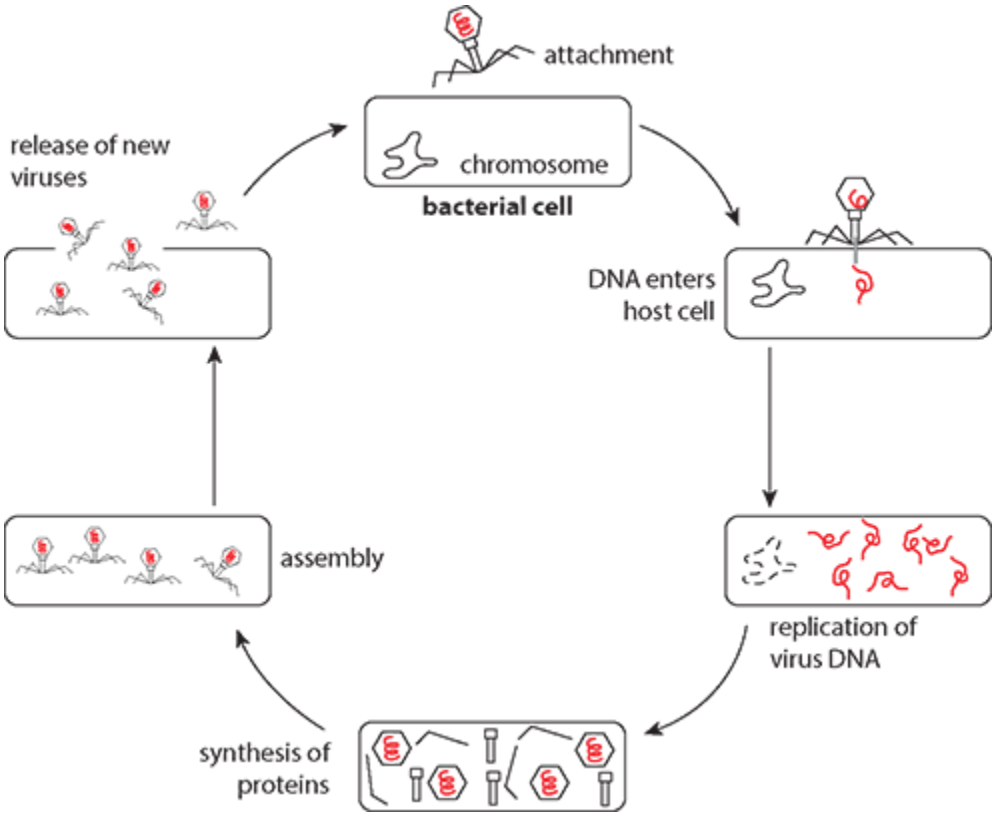


Figure 5.3.3: Stages in the lifecycle of bacteriophage lambda.

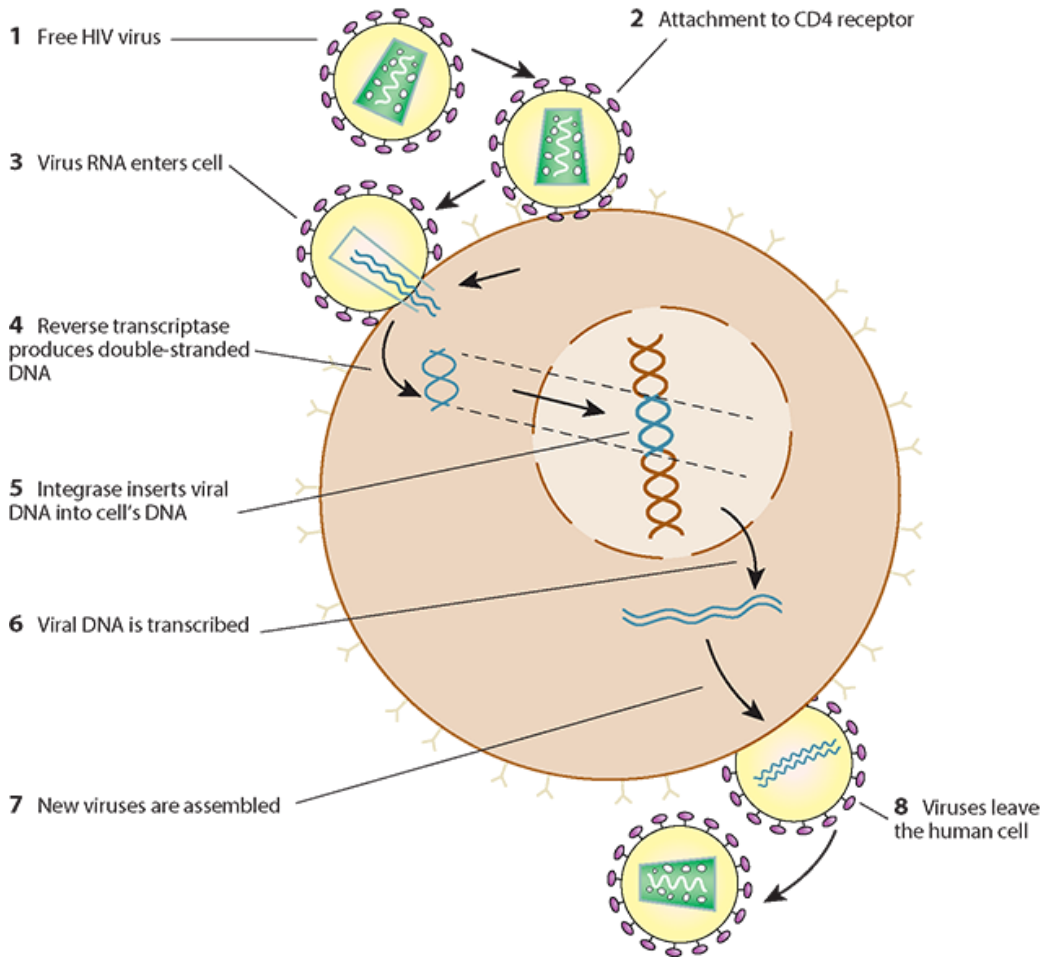


Figure 5.3.4: Life cycle of HIV.

Similar immunodeficiency viruses also affect other species including monkeys, apes and cats, but none of these viruses can cross the species barrier and infect humans, nor can HIV infect any other species that do not have the correct receptors on their plasma membranes.

TEST YOUR UNDERSTANDING

15 Explain why most biologists do not consider viruses to be living organisms.

16 Outline the difference between a capsid and a viral envelope.

17 Suggest one advantage of a viral envelope.

5.3.2 Diversity and origins of viruses

Viruses are a diverse group: some have RNA as their genetic material while others have DNA; some have single-stranded genetic material and some doubled-stranded genetic material, as we can see by comparing T4 phage and HIV. Their replication strategies are also very different. But all viruses have some key similarities: they are very small, with few being larger than 200 nm, no virus has ribosomes, so none can produce their own proteins and all viruses must replicate themselves inside a host cell. Table 5.3.1 summarises the features of some common viruses.

Virus	Size	Genetic material	Envelope	Host
Poliovirus	30 nm	single-stranded RNA	no	humans are the only natural host
Influenza	90 nm	single-stranded RNA	yes	mammals: each species has its own specific virus
Herpes simplex	150 nm	double-stranded DNA	yes	primarily human cells but can infect a range of different species
HIV	120 nm	single-stranded RNA	yes	humans, other types of immunodeficiency viruses infect other species

Coronavirus	100 nm	single-stranded RNA	yes	different coronaviruses infect humans, bats, rats, cattle and birds
Bacteriophage T4	225 nm tall	single-stranded DNA	no	*E. coli* bacteria

Table 5.3.1: Features of some common viruses.

Viruses are a very diverse group and this suggests that they may have several possible origins. Viruses all share a very extreme form of obligate parasitism as their mode of life and it is possible that this developed as a result of convergent evolution. But their genetic code is the same as living organisms. These differences and similarities have led to several hypotheses about the origins of viruses, as follows.

The **virus first hypothesis** proposes that viruses existed before cells and may have been the first 'entities' capable of replicating themselves. These very early viruses may have become more complex over time and eventually developed the ability to synthesise membranes and cell walls, so that cells could form from them. It is generally accepted that RNA was the first replicating material so perhaps today's single-stranded viruses arose from early RNA molecules? Some researchers have proposed that the nucleus of eukaryotic cells evolved from a DNA virus that became engulfed by a developing cell. Support for this theory comes from the fact that viruses can infect cells from all three domains: the Archaea, Bacteria and Eukarya.

The progressive hypothesis or cellular origin hypothesis suggests that viruses originated through a progressive process.

The hypothesis describes how pieces of genetic material that were able to move within a genome became able to leave one cell and enter another. Retroviruses, the group that includes HIV, contain the enzymes reverse transcriptase and integrase that enable them to enter a host cell's genome and then leave the cell when new viruses have been produced. This behaviour is very similar to that of a component of the human genome known as a retrotransposon. Some retrotransposons can code for reverse transcriptase and integrase. Within a cell these elements can move to new locations within a human genome. It may be that if additional structural protein genes were acquired by these elements in the past, they could have become able to leave one cell and 'infect' another. Support for this theory comes from the fact that viruses are assembled within host cells.

The **regressive hypothesis** suggests that viruses may have originated by a regressive or reductive process. Support for this theory includes the fact that some bacteria have evolved in this way and are now also obligate intracellular parasites. One example is *Chlamydia* spp. This bacterium has evolved from a free-living ancestor and has its own metabolism, but can only reproduce inside a host cell. Researchers propose that existing viruses may have evolved over time and that once-complex viruses lost the genes for independent life and kept only those which were required for a parasitic mode of existence. This theory is supported by one particular group of very large DNA viruses that includes the smallpox virus and a recently discovered mimivirus, which infects amoebae. These viruses have complex genomes: pox viruses have 200 000 base pairs and mimivirus 1.2 million base pairs. This is far more than other viruses such as poliovirus, which contains only 7500 bases. These large viruses are less dependent on their hosts for replication and can produce their own mRNA within the host

cell cytoplasm. The regressive hypothesis suggests that similar large viruses were once free living but developed first a symbiotic relationship with their hosts and eventually lost genes that were previously present.

KEY POINTS

progressive hypothesis (or cellular origin hypothesis) is a theory that proposes viruses may have evolved from DNA or RNA fragments from the genes of another organism.

regressive hypothesis is a theory proposing that viruses are derived from fragments of cellular organisms via a reductive process.

virus first hypothesis is a theory proposing that viruses existed before cells and may have been the first 'entities' capable of replicating themselves.

INTERNATIONAL MINDEDNESS

Fighting a new virus: COVID-19 research

In 2019 a new strain of coronavirus emerged, which spread rapidly across the world in 2020. Called SARS-CoV-2, the virus causes lung infections with a high temperature, a continuous cough and a loss or change to the sense of smell. The virus spreads primarily through droplets when an infected person coughs or sneezes. The COVID-19 pandemic has led to loss of human life worldwide and presents an unprecedented challenge to public health.

Scientists around the world mobilised to use their skills to help fight COVID-19 disease. While many politicians used almost military terms to describe the fight against the virus as

a ‘biotechnology arms race’, scientists responded in a completely different way. Politicians may have closed national borders but, in science, expert researchers from many countries worked together and focused on COVID-19 research with an urgency never seen before.

Many other research projects were put on hold and features that are typical of research, such as academic credit, were set aside. The results of studies from all over the world were quickly posted online and became available months ahead of published reports in scientific journals. Researchers identified and shared hundreds of genome sequences from the virus. More than 200 clinical trials were organised to bring together laboratories and medical facilities around the world.

Dr Francesco Perrone, an Italian scientist, said ‘I never hear scientists – true scientists, good quality scientists – speak in terms of nationality. My nation, your nation. My language, your language. My geographic location, your geographic location. This is something that is really distant from true top-level scientists.’

By August 2020, vaccines were being trialled and more than 150 countries were engaged in discussions to establish a global initiative to work with vaccine manufacturers and provide countries all over the world equal access to safe and effective vaccines. ‘Equal access to a COVID-19 vaccine is the key to beating the virus and paving the way for recovery from the pandemic,’ said Stefan Löfven, Prime Minister of Sweden. ‘This cannot be a race with a few winners, making sure all countries can benefit from fair and equitable distribution of vaccine doses is vital.’

Vaccine trials took place all over the world and the first mass vaccination programmes began in early 2021.

To consider:

1. Institutional and international cooperation has been key in the work to combat COVID-19. Is it possible that the landscape of global research in the search for medical treatments has changed forever?
2. Discuss the importance of international collaboration in research into SARS-CoV-2 and other disease-causing viruses.

While there is some evidence to support all three theories there is no definite answer to explain the origin of viruses. It may be that modern viruses arose in different ways by many different mechanisms. Research in the fields of structural biology, genomics and virology continues to try to gather more evidence and no single hypothesis may be the correct one.

5.3.3 Rapid evolution in viruses

Viruses evolve by natural selection in the same way as cellular organisms. Natural selection and evolution occur when there are genetic differences in a population and those individuals with favourable characteristics become more likely to inherit them and survive and reproduce (Chapter 11). Variation and evolution of viruses happens rapidly and regularly, for example new strains of influenza virus appear each year.

Viruses evolve in two main ways:

1 By recombination of genetic material

If two viruses infect a cell at the same time, they can exchange genetic material to produce new viruses with different properties, new strains of influenza virus arise like this.

2 As mutations occur in their DNA or RNA sequences

RNA viruses have a very high rate of mutation and this can lead to the evolution of drug-resistant strains.

Viral evolution is usually rapid because of their fast rate of reproduction. Viruses that infect animals are able to evolve much more quickly than their hosts. Some viruses, for example HIV, have a very high mutation rate so they can produce a more variable population very quickly. They also produce very large populations of virus and they have a very short lifecycle. Large populations tend to have more individuals with random mutations and if these are useful to the virus, for example if they make the virus more infectious or resistant to drug treatments, natural selection will enable these to survive.

Evolution of influenza viruses

Viruses exchange genetic material by recombination. If two viruses infect the same cell at the same time many viral genomes will be produced inside the cell at the same time.

If this happens, recombination can occur, either by similar regions of the genomes breaking and reconnecting with other fragments or by a process called reassortment when viruses exchange segments of DNA or RNA. (Fig 5.3.5.)

Influenza viruses evolve rapidly by **reassortment**. They have eight RNA segments and if two different viruses, perhaps strain X and strain Y, infect a cell together, new viruses may have a mixture of segments from both X and Y (Figure 5.3.5). Various combinations can be made. The two strains may originally have infected different organisms, for example H1N1 strain of swine flu that caused a pandemic in 2009 was found to have RNA segments that originated from human, bird and pig viruses from two continents. Reassortments had occurred naturally over many years and in a series of stages to produce this virus and several other strains.

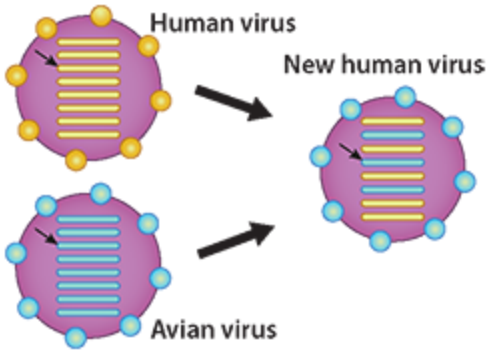


Figure 5.3.5: Influenza viruses evolve by reassortment of their RNA segments to produce new strains which may originate in different species.

KEY POINT

reassortment only occurs in viruses that have their genetic material divided into segments. It is the exchange of intact genes within the entire segment, which occurs when a cell is infected by more than one strain at the same time. It is not to be confused with the reassortment that occurs in the genetics of living organisms (Chapter 4).

Evolution of HIV viruses

HIV is an RNA virus and RNA viruses tend to have higher rates of mutation than DNA viruses. The reason is that RNA viruses use RNA polymerases to copy their genetic material, unlike DNA viruses that use the host cell's DNA polymerase to do so. DNA polymerases proofread new DNA that is produced and edit out any mistakes. RNA polymerases do not do this so any mutations in the new RNA remain and while some will be harmful, others will be of benefit to the virus.

Some HIV viruses have developed a mutation that can provide resistance to drug treatments that are used to treat the disease. One treatment that is used is a drug called nevirapine that blocks reverse transcriptase, the enzyme that viruses use to copy their RNA genomes to DNA. Without this enzyme the virus cannot reproduce and cannot permanently infect a cell. Most HIV viruses are blocked by nevirapine but a few have a mutation in the genes for reverse transcriptase which alters the enzyme's binding site so that the drug cannot inhibit it. These resistant viruses will survive and reproduce and can re-establish a resistant population of HIV.

Today a new method of treatment for HIV is used. It is known as HAART or highly active antiretroviral therapy. It uses a combination of three or more drugs because it is less likely that a

population of HIV viruses will carry mutations to all three drugs at the same time. Multi-drug resistant strains will eventually evolve but this treatment will slow down the rate of evolution of these viruses.

TEST YOUR UNDERSTANDING

18 How does the bacterium *Chlamydia* support the regressive hypothesis of viral origin?

19 Outline the 'virus first' hypothesis of viral origin.

20 Summarise the reasons for the rapid evolution of viruses.

Links

- Why are viruses used as vectors in genetic modification? (Chapter 4)
- How does the Hershey–Chase experiment support the idea that DNA is the hereditary material? (Chapter 4)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
explain the importance of organic molecules and boundaries in the origin of life	5.1.1			
outline the principles of cell theory	5.1.2			
outline the Miller–Urey experiment and its importance	5.1.3			
describe how the deep-sea vent hypothesis provided evidence for early molecules and their energy sources	5.1.4			

outline the properties of RNA that make it important in the origin of life	5.1.5			
describe how micelles form and suggest how protocells may have originated	5.1.6			
outline the endosymbiosis theory	5.2.2			
state that living organisms are composed of cells and list all the functions of life	5.2.1			
state that in multicellular organisms cells are differentiated and that all cells do not carry out all the functions of life	5.2.1			
describe the atypical structure of xylem and phloem	5.2.1			

describe the similarities and differences between prokaryotic and eukaryotic cells	5.2.1			
identify the components of different types of cell in micrographs	5.2.1, 5.2.3			
summarise how developments in microscopy have helped us understand cell structure	5.2.3			
describe the key features of a virus	5.3.1			
state that all viruses contain genetic material and are protected in a capsid	5.3.1			
describe how some viruses have a viral envelope and outline the benefits this gives	5.3.1			
	5.3.1			

describe the structure of some different viruses and outline the lytic and lysogenic life cycles of bacteriophage lamda and HIV				
explain the similarities and differences between viruses that suggest several origins	5.3.2			
outline the hypotheses which try to explain the origin of viruses	5.3.2			
summarise the reasons for the rapid evolution of viruses.	5.3.3			

REFLECTION

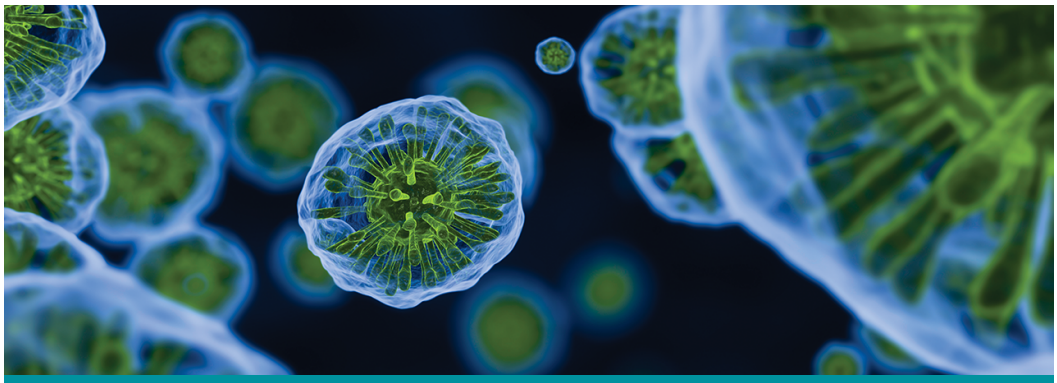
Can I explain the importance of looking at life at the microscopic level to someone else? If not, what do I need to review?

What are your first thoughts about this chapter? Are they positive or negative?

How do you feel this topic relates to real-life situations and problems?

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.



Chapter 6

Cell function

B2.1, B2.2, B2.3

INTRODUCTION

Membranes enclose the contents of the cell. There is much activity at membrane surfaces, especially at the plasma membrane where it contacts the extracellular space. Membranes surround organelles inside eukaryotic cells and are important in separating different substances and reactions that take place in organelles. Some substances can cross membranes while others cannot, and the cell controls the amount of water that it contains by regulating what can cross its membranes by diffusion and active transport. Cells must retain a suitable surface area to volume ratio if these

mechanisms are to operate. So, after cells have grown, cell division is needed to increase the surface area of the cell.

6.1 Membranes and organelles

LEARNING OBJECTIVES

In this section you will:

- learn that all cells are surrounded by a phospholipid bilayer that forms due to amphipathic properties of the molecules
- understand that the phospholipid layer contains other molecules including proteins, glycoproteins and integral proteins
- discover that integral proteins act as receptors and enzymes and are involved in transport across membranes, while glycoproteins are used for cellular recognition or attachment
- learn that cholesterol in animal cells and sterols in plant cell membranes reduce membrane fluidity and permeability
- recognise that organelles are membrane-bound structures that form compartments with specific functions and that separation confers advantages such as isolating metabolic processes
- learn that prokaryotic cells do not contain organelles

> recognise the relationship between fatty acid and cholesterol content of bilayers and their fluidity

- understand that membrane fluidity affects endocytosis and exocytosis
- understand that organelles interact due to the fluid nature of the phospholipid bilayer
- learn that rough endoplasmic reticulum supports ribosomes and packages proteins
- distinguish between the function of attached and free ribosomes
- recognise the role of the Golgi apparatus
- recognise that coated vesicles aid transport between different parts of the cell
- recall that mitochondria have two membranes and are adapted for aerobic respiration
- discover that compartmentalisation permits organelles, including mitochondria and chloroplasts to isolate enzyme systems, generate high quantities of protons and synthesise ATP
- learn that chloroplasts have three membranes that provide compartmentalisation to isolate enzymes and synthesise ATP. Membranes of the chloroplast also support photosynthetic pigments.

GUIDING QUESTIONS

- Why do phospholipids form bilayers?

- How do proteins integrate into membranes?
- Why is compartmentalisation within cells important?

6.1.1 Membrane structure

All membranes, wherever they occur in cells, have the same basic structure. Membranes are usually between 7 and 10 nm thick, and are composed of two layers of phospholipid, which form a bilayer. **Phospholipids** are made up of a polar, hydrophilic area containing a phosphate group bonded to glycerol, and a non-polar, lipophilic area containing fatty acids. In the bilayer, the lipophilic or hydrophobic (water-hating) parts all point towards each other, and the hydrophilic (water-loving) areas point outwards. Phospholipids are amphipathic molecules as they contain hydrophilic and hydrophobic parts as Figure 6.1.1 shows.

It is the different properties of each end of the phospholipid molecule that cause the phospholipids to arrange themselves in this way. The hydrophilic 'heads' of the molecules always appear on the outside of the membrane where water is present, while the hydrophobic 'tails' orientate inside the double layer, away from water.

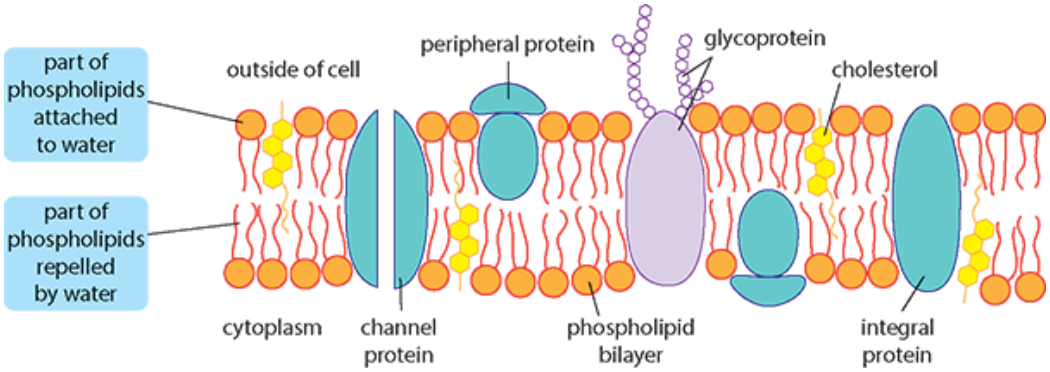


Figure 6.1.1: A plasma membrane contains structures which produce a dynamic matrix.

The structure of a membrane is called a ‘mosaic’ because, just as a mosaic picture is made up of many small, separate pieces, so the surface of the membrane is composed of the heads of many separate phospholipid molecules. The whole structure is flexible or ‘fluid’ because the phospholipids can float into a position anywhere in the membrane. Cell membranes are dynamic, fluid structures because most of their molecules are able to move about in the plane of the membrane. Research using radioactively labelled phospholipids shows that these molecules move not only within their own layer, but also between the two layers of the membrane.

The phospholipid bilayer contains other molecules that form a part of the dynamic matrix. Embedded in the membrane bilayer are different molecules that contribute to the functions of membranes (Figure 6.1.1).

- Cholesterol is often present in animal cells and is most commonly found in the plasma membrane. Plant cell membranes have similar molecules called **sterols** which serve the same function. One end of the cholesterol or sterol molecule associates with the polar heads of phospholipid molecules while other parts of it are embedded in the membrane next to the non-polar fatty acid chains. This interaction makes the membrane less ‘fluid’, more rigid and less permeable to water-soluble molecules.

There are also different types of protein in the bilayer. Membrane proteins have many different structures and positions in the bilayer. They also have different functions:

- **Integral proteins** are embedded in the bilayer. Part of their surface is hydrophobic and remains embedded in the centre of the membrane. Some extend through the membrane and have hydrophilic parts which project between the phosphate

heads. Some integral proteins are enzymes immobilised within the membrane structure and perfectly placed to carry out sequences of metabolic reactions.

- **Peripheral proteins** are hydrophilic and are attached to the membrane's surface. Most of them are attached to the surface of an integral protein. Many of the proteins on the outer surface are glycoproteins, that is, they have carbohydrate groups attached to them. Some of them serve as hormone-binding sites and have special shapes to recognise the specific hormones or antigens to which the cell will respond. Others are important in cell to cell communication and adhesion.
- **Channel proteins** are integral proteins that span the bilayer, acting as channels for ions and molecules to pass by passive transport, or forming pumps that use active transport to move molecules into or out of the cell.
- **Aquaporins** are a special class of channel proteins that act as channels for the transfer of water, and in some cases, small solutes across the membrane.
- **Glycolipids** are lipids which are attached to a carbohydrate by a glycosidic bond. They maintain the stability of the cell membrane and allow cells to recognise other cells. This is vital to the immune response and in allowing cells to make connections to one another to form tissues

EXAM TIP

You must be able to draw and label a simple diagram like the one shown in Figure 6.1.1 to show the range of membrane proteins, including glycoproteins, and cholesterol or sterol.

6.1.2 Organelles

All cells contain a number of different structures that enable them to stay alive and function (see also Section 5.2). Organelles are special structures that have their own specific roles in the cell. Membrane-bound organelles, such as mitochondria and chloroplasts, form compartments so that metabolic processes can be separated from one another, material can be stored and certain molecules can be concentrated. The cell wall, cell membrane, cytoskeleton and cytoplasm are not organelles, but vesicles such as the Golgi apparatus and lysosomes are considered to be organelles because of their contents.

Membrane-bound organelles are only found in eukaryotic cells. Prokaryotes do not contain membrane-bound organelles, but they do have ribosomes that carry out protein synthesis and a plasma membrane that encloses their cytoplasm.

Advantages of separating the cell into compartments

The nucleus and cytoplasm are separated by the nuclear envelope, the membrane which encloses the nucleus. This separation means that gene transcription (copying DNA to mRNA) can occur in the nucleus and is separated from translation (production of polypeptides from mRNA) that occurs in the cytoplasm. Modifications of mRNA such as the removal of introns can take place before the mRNA meets ribosomes and the translation process begins. The translation of modified mRNA produces the correct amino acid sequences needed to make a functional protein. Prokaryotes which do not have a separate nucleus have mRNA that meets ribosomes as soon as it is produced.

The cytoplasm contains many organelles which separate the functions of the cell into compartments. Within a compartment metabolites and enzymes can be concentrated so that enzyme-controlled reactions such as respiration can take place more quickly and efficiently than if the reactants were spread through the cytoplasm. Certain biochemical processes such as the digestion of unwanted materials can take place within an enclosed vacuole without affecting other parts of the cell.

Lysosomes are small spherical organelles that contain hydrolytic (digestive) enzymes, including proteases, amylases, nucleases and lipases.. The enzymes must be kept separated from the contents of the cytoplasm, because the pH inside the lysosome is different from that of the cytoplasm and would damage cell structures. The function of lysosomes is to engulf and break down unwanted macromolecules and respond to and destroy invading particles, such as bacteria or viruses, that enter the cell. They are also capable of destroying a cell if they burst and release their enzymes; they are sometimes called the cell's 'suicide packets'.

NATURE OF SCIENCE

Using models to represent the real world: alternative models of membrane structure

The fluid mosaic model that we accept today was not the first scientifically accepted explanation of membrane structure. In 1935 Hugh Davson and James Danielli proposed the first model that attempted to describe the structure of the bilayer and the proteins in it (Figure 6.1.2a). They proposed that the membrane consisted of two layers of protein that enclosed a phospholipid bilayer. The model was known as a 'lipoprotein sandwich'. But the model wrongly assumed that all

membranes were the same, with a constant lipid to protein ratio. This is not true. The model also could not explain how certain hydrophilic substances could pass through the hydrophobic centre of the membrane.

Later research found that membrane proteins are insoluble and thus have hydrophobic surfaces. This means that they could not to form a continuous layer around the outside of a membrane. More recent fluorescent tagging of membrane proteins also showed that they are mobile and can occupy different positions in a membrane and do not form a static layer. In addition, freeze fracturing, a technique to cut open and expose the inner surfaces of a biological specimens, was used to split open the membrane and showed that there are irregular rough surfaces inside the membrane. These rough surfaces are interpreted as being trans-membrane proteins and indicate that proteins are not only found on the outside of membranes.

The discoveries led to a new model being proposed by Singer and Nicolson in 1972. According to this model, proteins were embedded within the lipid bilayer and do not exist as separate layers. This model is known as the fluid mosaic model and it is still the model that is most widely accepted today (Figure 6.1.2b).

To consider:

1 Models are used to explain patterns that are not directly observable. They cannot be proven but can be falsified when their predictions do not agree with current evidence. How important are the two models of membranes to our understanding of their structure?

2 Why were the new techniques of electron microscopy, fluorescent marking and freeze fracturing crucial to the development of the fluid mosaic model of membranes?

3 Could a new model eventually replace the fluid mosaic model?

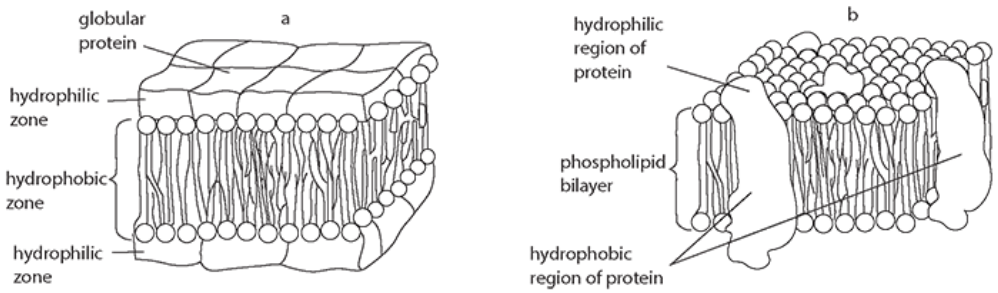


Figure 6.1.2: Diagrams showing **a** the Davson Danielli 'sandwich' model of membrane structure and **b** the later fluid mosaic model that is accepted by scientists today.

Membrane fluidity

Fatty acids in lipid bilayers are the main component of cellular membranes. There are many different fatty acids which vary in structure, and the distribution of different lipids in membranes varies between organisms, cell types, organelles and membranes. Lipid composition affects the physical properties of a membrane, for example, unsaturated fatty acids in phospholipids reduce membrane rigidity and affect processes that take place as the membrane changes shape.

Membrane fluidity is affected by the type of fatty acids that are present; saturated fatty acids make the membrane less fluid, while unsaturated fatty acids make it more fluid. (Fig 6.1.3). Saturated fatty acids have higher melting points than unsaturated

molecules and make membranes stronger at higher temperatures. They have no double bonds in their hydrocarbon chains (Chapter 1) and this results in a strong, tight membrane. Polyunsaturated fatty acids have more than one double bond in their hydrocarbon chains and this creates a bend in the molecule which increases fluidity. The correct ratio of saturated to unsaturated fatty acids will keep the membrane fluid at different temperatures. For example, microorganisms can adjust the fatty acid composition of their membranes in response to heat stress in their environment.

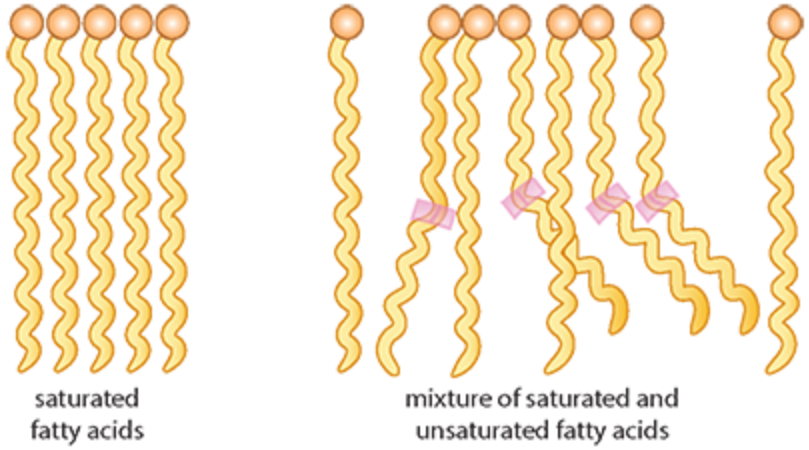


Figure 6.1.3: Saturated fatty acids produce a stable, stronger membrane at high temperatures. Unsaturated fatty acids increase the fluidity of the membrane.

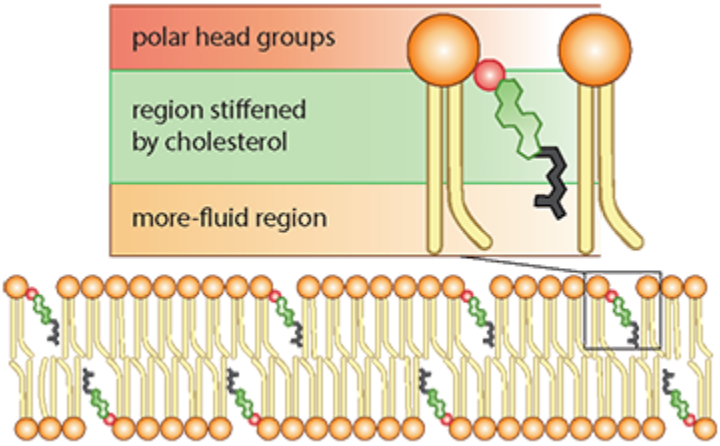


Figure 6.1.4: Cholesterol between phospholipid molecules stiffens the membrane at higher temperatures and prevents it from melting.

Membrane fluidity also affects the functioning of other molecules in the membrane. Binding to some peripheral proteins and movement of enzymes within the membrane depend on its fluidity. Functions such as cell signalling and endocytosis and exocytosis (described here) can be regulated by the fluidity of the cell membrane.

Cholesterol makes up 30% of the lipids in the membranes of animal cells and it also affects their fluidity. Its presence stabilises and stiffens membranes at higher temperatures and increases their melting points. At low temperatures it is positioned between phospholipids and prevents them aggregating and becoming stiff at lower temperatures. (Figure 6.1.4)

Membranes and the protein channels which are present in them also have important functions in the transmission of nerve impulses. You can read more about gated ion channels for

neurotransmitters and about sodium potassium pumps which maintain membrane potentials in Chapter 6.2.3 and in Chapter 7.

Importance of membranes in cell adhesion

A certain group of surface glycoproteins are known as cell adhesion molecules (CAMs). They are involved in helping cells stick to each other and to their surroundings. They are very important in maintaining the structure and function of tissues.

Cell adhesion molecules are found on a cell's surface and form different types of bonds and junctions so that they can join:

- Cells to cells
- Cells to extracellular matrix (chains of sugar and protein molecules with other structural proteins such as collagen which surround cells and make tissues more stable)
- Extracellular matrix to the cell cytoskeleton

CAMS help in processes including:

- The adhesion of cells to one another to provide organized tissue structure
- The transmission of cues and signals from outside the cell across the membrane
- The movements of certain cells buy regulating the cell that they adhere to.

As well as acting as a molecular glue, CAMS have important roles in growth, contact inhibition and apoptosis.

There are four different forms of CAM,

- integrins that connect cells to the extracellular matrix to give integration,
- immunoglobulins CAMS which act as adhesion molecules in the nervous system,
- cadherins responsible for cell to cell adhesion
- selectins which bind to cell surface carbohydrates and are involved in the way cells respond to inflammation

CAMS may bind with the same CAMS (homophilic binding) or the CAM of one cell may bind with different CAMS (heterophilic binding) on another cell.

Mitochondria and chloroplasts

Two of the most important organelles found in cells are mitochondria and chloroplasts.

Mitochondria are found in all eukaryotic cells, but only organisms that photosynthesise contain chloroplasts. Figure 6.1.5 shows diagrams of the two organelles, drawn from electron micrographs (see also Section 2.3).

Both organelles are enclosed in two layers of membrane which separate their contents from the cytoplasm of the cell. Both have important roles in transforming energy and generating ATP, and both have their own ribosomes, DNA and enzymes. Chloroplasts and mitochondria can also move, change their shape and divide independently of the cell cycle by simple fission. Both these organelles are thought to have originated by endosymbiosis (Section 5.2) and they have retained some independence from the cell which contains them.

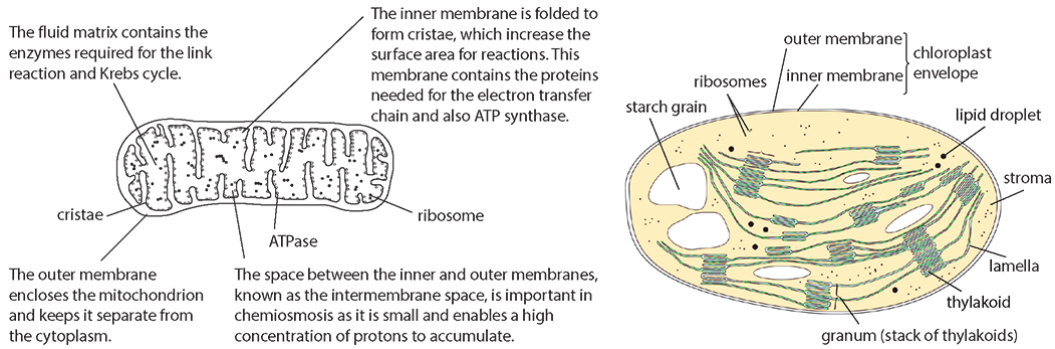


Figure 6.1.5: Diagrams to show the detailed structures of a mitochondrion and a chloroplast.

WORKED EXAMPLE 6.1

Recognising and identifying cells

1 Diagrams of cells

You must be able to recognise prokaryotic and eukaryotic cells from electron micrographs and draw diagrams of eukaryotic cells to show their structure and organelles. Look at Figure 6.1.6, which is a diagram of a cell.

a To identify the type of cell, first check: is it a prokaryote or a eukaryote?

- If the cell has organelles it is a eukaryote. In this case we can see a nucleus with a membrane, as well as mitochondria and other organelles.

b Is the cell a plant or an animal cell?

- If the cell has a cell wall, large vacuole and chloroplasts then it is a plant cell.

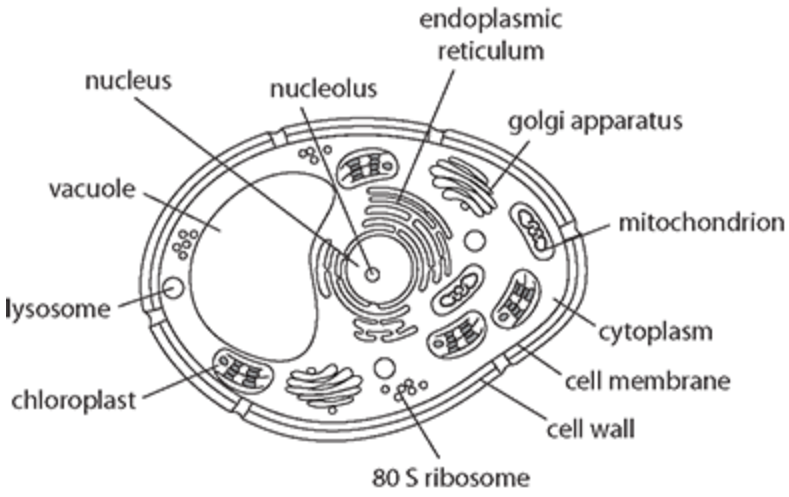


Figure 6.1.6: Diagram of a cell.

2 Electron micrographs of cells

In electron micrographs such as Figure 6.1.7 you must be able to identify organelles including mitochondria, chloroplasts, the Golgi apparatus, rough and smooth endoplasmic reticula and ribosomes.

a Can you identify the cell membrane in this photograph?

- It appears at the top left-hand corner of the image.

b Notice the mitochondria (the structures coloured purple), they are not all the 'sausage shape' that you see in diagrams. Think about why this is.

c This cell contains a large amount of endoplasmic reticula (the structures coloured green), is it rough and grainy?

- If it is, this indicates that the cell is manufacturing protein and that ribosomes are attached. If it is smooth there are no ribosomes attached.

d There are pale blue lipid droplets present in the cell.

- The cell is storing lipids in droplets to be used in the production of testosterone. The lipids are separated from the rest of the cytoplasm inside a membrane.

e Look at the nucleus. Notice the nuclear membrane that surrounds it. What can you see inside it?

- The darker areas are uncoiled chromosomes and are called chromatin.
- At higher magnifications it is possible to see that the nuclear membrane is double and has pores through it.

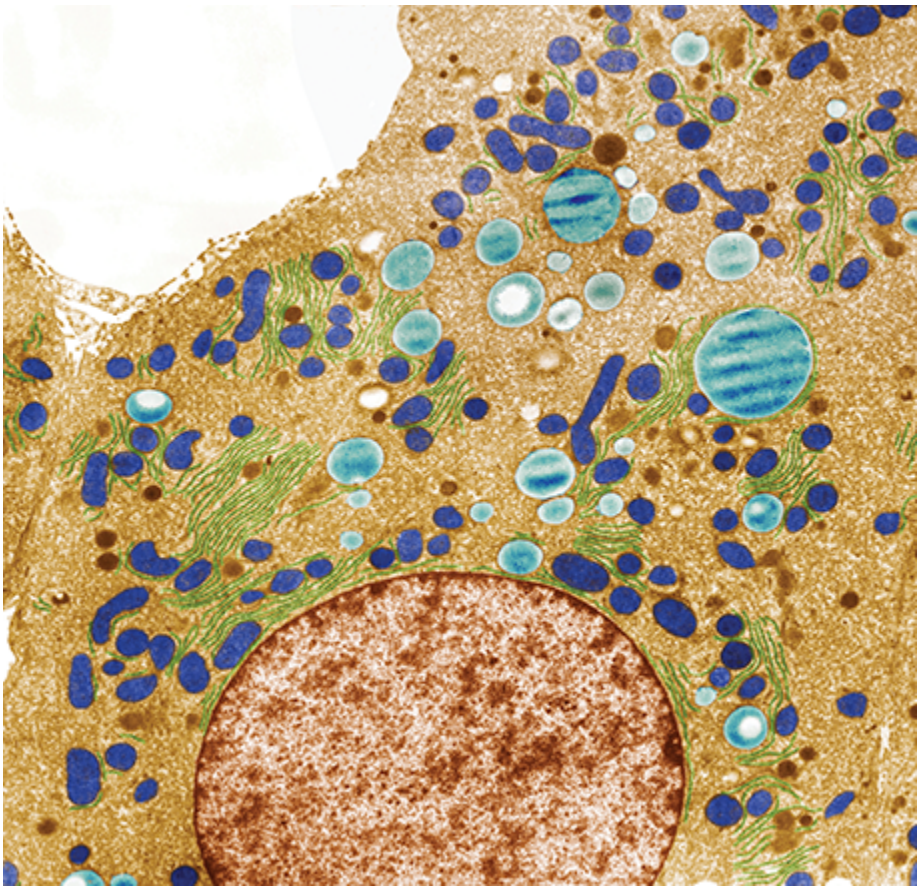


Figure 6.1.7: Coloured electron micrograph of a testis cell.

3 Prokaryotic cells

You should be able to draw a diagram of a prokaryotic cell (Figure 6.1.8) of a rod-shaped bacterium and label the structures it contains.

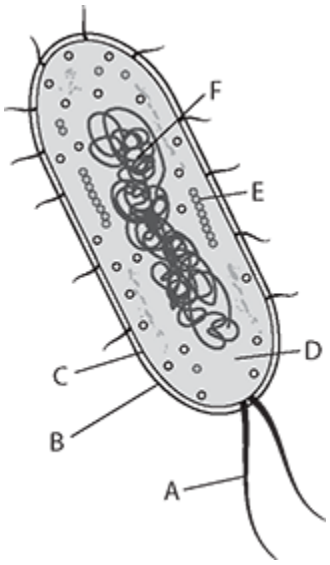


Figure 6.1.8: A rod-shaped bacterium.

Use these names to identify the structures A–F.

naked DNA; 70S ribosomes; cell wall; cell membrane; flagellum; cytoplasm

TEST YOUR UNDERSTANDING

1 State the difference between hydrophilic and hydrophobic molecules.

2 Name three organelles found in a eukaryotic cell.

3 State two roles of integral proteins.

6.1.3 Organelles and interactions between them

The phospholipid molecules in a membrane bilayer are free to move between layers and within a layer. This fluid nature of the membrane allows membranes from different organelles, or the same organelle, to stick to or fuse with one another.

The endoplasmic reticulum (ER) is a membrane that makes up about half of the total amount of membrane in an animal cell. It is composed of folded, flat, sealed sacs that are joined to the nuclear membrane. It occurs throughout the cell but is more concentrated close to the nucleus and Golgi apparatus. The ER is called rough endoplasmic reticulum (RER) if ribosomes are attached to it. Attached ribosomes produce the characteristic grainy appearance seen in electron micrographs. Both ER and RER are present in cells at the same time and are parts of the same organelle. If a cell is actively synthesising protein it will contain more RER. (A liver cell contains more than 10 million ribosomes.) Many of the substances produced by the RER are exported to other organelles. Ribosomes synthesise proteins but there are two separate groups of ribosomes in the cytoplasm. Membrane-bound ribosomes, attached to the ER synthesise proteins that are produced for export from the cell whereas free ribosomes synthesise all the other proteins that are encoded by the nuclear genome and are to be used within the cell. Membrane-bound and free ribosomes are structurally and functionally identical. They differ only in the proteins they are making at any given time. Many ribosomes can bind to a single mRNA molecule forming a polyribosome and this becomes attached to the ER membrane.

The Golgi apparatus is also made of membranes and is closely linked to the ER. Some substances produced by the RER pass directly between the two organelles, whereas others are packaged as droplets in vesicles that carry them through the cytoplasm. The Golgi apparatus processes proteins which it receives from the ER and sorts them for transport to their final destinations which may be lysosomes, the plasma membrane or for secretion out of the cell.

Coated vesicles are vesicles that are coated with a protein complex called clathrin. These structures act as a link between different parts of the cell and they transport proteins between the plasma membrane, Golgi network and lysosomes. Coated vesicles also bud from the plasma membrane, where the addition of the clathrin coat gives them a polyhedral pattern (Figure 6.1.9). The combination of proteins in the coat of a coated vesicle determines the vesicle's destination in the cell.

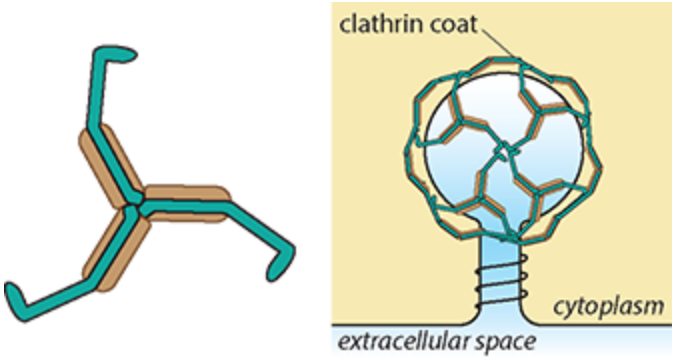


Figure 6.1.9: A clathrin-coated vesicle is formed when clathrin triskelion proteins assemble over the surface of a membrane to form a structural scaffold.

Lysosomes are small spherical organelles that contain hydrolytic enzymes, including proteases, amylases, nucleases and lipases. Lysosomes bud off from the Golgi apparatus and the enzymes

they contain are formed in the endoplasmic reticulum. The enzymes must be kept separated from the contents of the cytoplasm, where they would damage cell structures because the pH inside the lysosome is different from that of the cytoplasm. The function of lysosomes is to engulf and break down unwanted macromolecules and respond to and destroy invading particles, such as bacteria or viruses, that enter the cell. They are also capable of destroying a cell if they burst and release their enzymes; they are sometimes called the cell's 'suicide packets'.

Membranes in mitochondria

Mitochondria have two membranes and chloroplasts have three (Figure 6.1.5). The inner membrane of the mitochondrion and the thylakoid membrane of the chloroplast are both highly folded and both provide a large surface area for the reactions of respiration and photosynthesis (Sections 2.2 and 2.3).

The membranes of mitochondria divide the organelle and form compartments in which the reactions of aerobic respiration take place (Section 2.2). In the centre, the inner membrane contains the compartment known as the matrix which separates the enzymes needed in the reactions of the Krebs cycle from the contents of the inner membrane space. The inner membrane space, between the outer membrane and the folded cristae of the inner membrane, has a much smaller volume than the matrix. A high concentration of H^+ ions from the Krebs cycle can build up between the two membranes and so the pH is much lower than in the matrix. Membranes of the cristae, which have a large surface area due to their folding, contain ATP synthase molecules through which H^+ ions flow to produce ATP.

Membranes in chloroplasts

Chloroplasts and their membranes provide compartments in which the reactions of photosynthesis can take place (Figure 6.1.5). The interior of the chloroplast, inside the two outer membranes, is filled with liquid stroma that contains the enzymes for the Calvin cycle, the light-independent reactions of photosynthesis (Section 2.3). These reactions are separated from the light-dependent reactions which occur on the thylakoids. The thylakoid membranes are folded to produce grana and lamellae and have a large surface area with a small volume of intramembrane space. The thylakoid membrane supports the chlorophyll molecules of the two photosystems and contains protein pumps through which H^+ ions pass into the thylakoid space. This inner thylakoid space is a compartment of the chloroplast that has a low pH. During photosynthesis, H^+ ions are pumped through ATP synthase that is embedded in the thylakoid membrane and back into the stroma to produce ATP.

Table 6.1.1 summarises the similarities in the membranes of mitochondria and chloroplasts.

Mitochondria	**Chloroplasts**
two layers of membrane	three layers of membrane
enzymes of the Krebs cycle are contained in the matrix	enzymes of the Calvin cycle are contained in the stroma
cristae have a large surface area and a small intramembrane volume at a low pH	thylakoid membranes and intergranal lamellae have a large surface area and small intramembrane volume with a low pH
membranes of the	membranes of the thylakoid contain

cristae support channel proteins and ATP synthase molecules	channel proteins, ATP synthase and photosynthetic pigments of photosystems 1 and 2

Table 6.1.1: The similarities in the membranes of mitochondria and chloroplasts.

Membranes of the nucleus

The nucleus has a double membrane known as the nuclear envelope surrounding it. It consists of an outer and an inner phospholipid bilayer. The thin space between the layers connects with the space (lumen) between the layers of the rough endoplasmic reticulum and the outer membrane links to the RER so that mRNA from the nucleus can be passed directly to the site of protein synthesis. The inner membrane has a protein lining which binds to chromatin and other contents of the nucleus and provides the structural framework of the nucleus. The nuclear envelope separates the contents of the nucleus from the cytoplasm and act as barriers that prevent the free movement of molecules between the nucleus and the cytoplasm. The nuclear envelope contains many nuclear pores which do allow the nucleotide DNAand RNA, as well as adenosine-triphosphate, which provides the energy for synthesizing genetic material, to enter. Histones and other large proteins can also pass through the pores. The pores also regulate the export of the mRNA and subunits of ribosome from the nucleus to the cytoplasm.

EXAM TIP

You must be able to label and annotate diagrams of mitochondria and chloroplasts to show the adaptations of these two organelles to their functions. You should be able to

identify the structures that are shown in Table 6.1.1 and summarise their importance in respiration and photosynthesis.

TEST YOUR UNDERSTANDING

4 Why are membranes important in separating compartments in chloroplasts?

5 What is the function of the rough endoplasmic reticulum?

Links

- How is evidence for endosymbiosis provided by chloroplasts and mitochondria? (Chapter 5)
- How does the solubility of a substance in water affect its movement across phospholipid membranes? (Section 6.2)
- How do the hormones that do not cross membranes influence cell activities? (Chapter 7)

6.2 Movement across membranes

LEARNING OBJECTIVES

In this section you will:

- learn that the plasma membrane is semi-permeable
- discover that ions and molecules move across membranes by simple diffusion, facilitated diffusion, osmosis and active transport
- recall that only water moves by osmosis
- recognise that proteins in the membrane act as pores or pumps for transport
- learn that channel proteins allow facilitated diffusion
- discover that active transport via protein pumps allows substances to move by active transport against a concentration gradient
- recognise that active transport requires energy in the form of ATP
- learn that the fluid nature of membranes allows endocytosis and exocytosis to transport substances in and out of the cell by active transport

- recognise that neurons are very specialised cells that receive and transmit impulses

- learn that gated ion channels allow ions to move in and out of nerve fibre membranes by passive transport
- recognise that the sodium–potassium pump in the neuron membrane allows active transport of Na^+ and K^+ ions in both directions
- discover that saltatory conduction of nerve impulses is possible due to myelination of nerve fibres.

GUIDING QUESTIONS

How do cells control how substances enter and leave?

6.2.1 Diffusion, facilitated diffusion and osmosis

Many molecules pass across the plasma membrane. Particles may move by simple diffusion, facilitated diffusion or osmosis, which are passive processes, or by active transport, which involves the use of energy. Water, oxygen, carbon dioxide, excretory products, nutrients and ions are continuously exchanged. Many cells also secrete products such as hormones and enzymes through the plasma membrane. The different methods that enable each substance to move ensure that these substances travel across the plasma membrane quickly and efficiently. The membrane is permeable to some molecules but not others, for this reason it is said to be selectively permeable.

Simple diffusion

The simplest way for molecules to move into or out of a cell is by simple **diffusion** through the plasma membrane. Diffusion is a passive process, which takes place as molecules move randomly. No energy input is required, and movement occurs via the **concentration gradient**. Molecules move from an area of high concentration to an area of lower concentration. A concentration gradient is a difference in concentration of a substance between two regions. Diffusion will always occur where such a gradient exists until particles of the substances are evenly distributed and equilibrium is reached.

Gas molecules move in and out of cells by simple diffusion, which means that diffusion has an important role in cell respiration. Oxygen is needed continuously because, as a cell respires, the oxygen concentration inside it decreases. As the oxygen concentration inside the cell becomes less than the

concentration outside, oxygen molecules diffuse in. In a similar way, as carbon dioxide forms during respiration, its concentration builds up inside the cell and it diffuses out through the plasma membrane to an area where the concentration is lower. Simple diffusion occurs where the membrane is fully permeable to the substance or where channel proteins in the membrane are large enough for the substance to pass through.

Diffusion via channel proteins

Large molecules, and charged particles such as chloride ions (Cl^-) and potassium ions (K^+), cannot pass through the membrane by simple diffusion. So, certain protein channels in the membrane form pores which provide a route for them to travel through. As with simple diffusion, no energy is used by the cell because the transport relies on the kinetic energy (the energy due to motion) of the particles moving down their concentration gradient. Channel proteins have an interior which is hydrophilic (Figure 6.2.1) so water-soluble materials can pass though them. In addition, they are specific: that is, they only allow one particular substance to move through.

Some channels are permanently open, whereas others are gated and only open to allow certain ions to pass when they are stimulated to do so. For example, gated channels in the axons of nerve cells open when there is a change in the voltage (potential difference) across the membrane. Gated potassium channels only allow K^+ ions to pass out through the membrane after a nerve impulse has passed along the axon. You can read more about nerve impulses in Chapter 7.

Facilitated diffusion

Substances such as glucose and amino acids, which are polar, cannot diffuse though the lipid layer of the membrane. They are transported across membranes by **facilitated diffusion** through special channel proteins. A carrier protein combines with the diffusing molecules on one side of the membrane, carries them through the channel protein and then releases them on the other side (Figure 6.2.1). Facilitated diffusion allows a faster diffusion rate for molecules that an individual type of cell needs; for example, the diffusion of glucose into active muscle cells. No energy input is required because the molecules move down their concentration gradient.

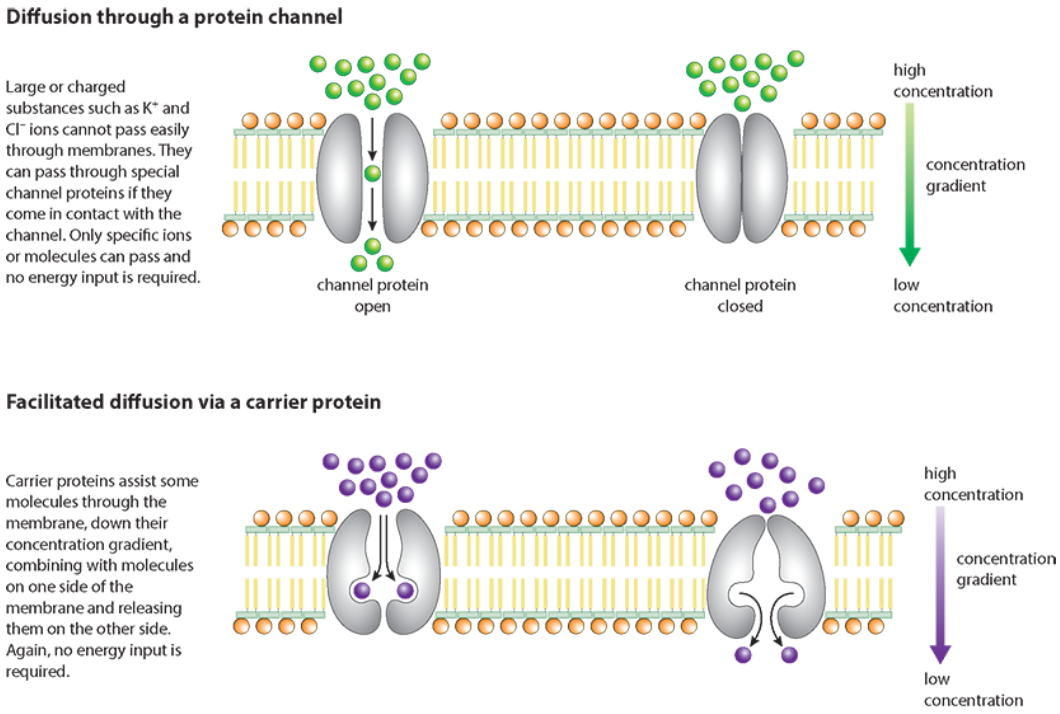


Figure 6.2.1: Some large or charged ions and molecules pass through the membrane via special channel proteins.

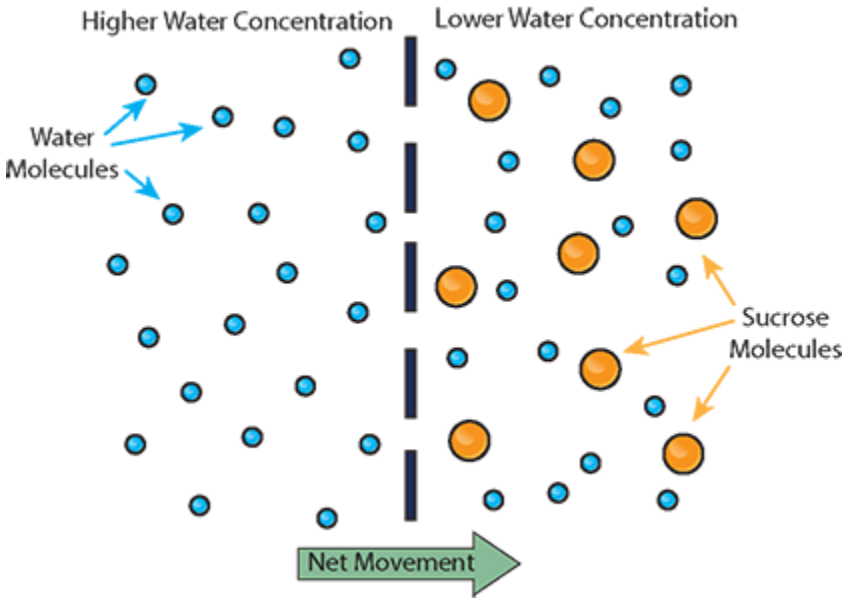


Figure 6.2.2: Solute molecules cannot move across the semi permeable membrane but water molecules can move in either direction.The net movement of water is shown by the arrow. Water moves from an area of lower solute cocentration of an area of higher solute concentration.

Osmosis

Osmosis is a special case of diffusion that moves water in and out of cells (Figure 6.2.2). Osmosis is the **passive movement** of water across the **semi-permeable** plasma membrane from an area of lower solute concentration to an area of higher solute concentration. Water moves through special protein channels known as **aquaporins**.

KEY POINTS

osmosis is the passive movement of water molecules across a membrane from a region of lower solute concentration, where there is a high concentration of water molecules, to a region of

higher solute concentration, where the concentration of water molecules is lower.

passive movement is the movement of substances down a concentration gradient from an area of high concentration to an area of lower concentration without the need for energy to be used.

when the plasma membrane is semi-permeable; it is permeable to some molecules but not other usually larger particles. It allows certain molecules or ions to pass through it by diffusion, while other pass by facilitated diffusion, passive transport or active transport.

When the solute concentrations inside and outside a cell are the same, the same number of water molecules will pass across the membrane into the cell as those that leave. You can read more about water potential and its effect on cells in Section 6.3.

Water enters the root tissue of a plant by osmosis, while ions enter by diffusion and active transport. Water enters through root hair cells, tiny extensions of the root cells, which increase the area for absorption (Figure 6.2.3). Water passes through the root to the xylem and from there is drawn up the plant in the **transpiration** stream as water evaporates from leaves.

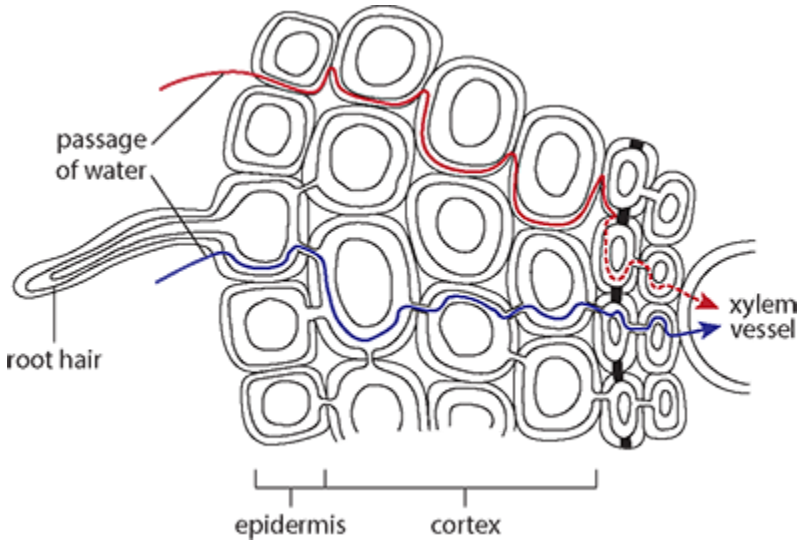


Figure 6.2.3: Cross-section of a root. Water enters the root hair cells by osmosis and passes to the xylem vessel. We need to extend the caption to explain the red and blue lines. "90% of the water moves through the cellulose cell walls (red line) and some passes through the cytoplasm through cell connections known as plasmodesmata (blue line)

SCIENCE IN CONTEXT

In medical procedures, tissues and organs are bathed in a solution of 'normal saline', which has exactly the same osmolarity (a measure of the solute concentration in a solution) as human cell cytoplasm. The saline is said to be **isotonic** with the cytoplasm. The saline solution ensures that osmosis does not occur, water does not enter or leave body cells and the cells are not damaged.

6.2.2 Active transport

Many of the substances a cell needs occur in low concentrations in the surroundings outside the plasma membrane. For example, plants must take in nitrate ions from very dilute solutions in the soil to build their proteins, and muscle cells actively take in calcium ions to enable them to contract. To move these substances into the cell against a concentration gradient, the cell must use metabolic energy released from the breakdown of ATP. This is called **active transport** (Figure 6.2.4). Specific proteins in the plasma membrane act as transporters or 'carriers' to move substances through. Many of the carrier proteins are specific to particular molecules or ions so that they can be selected for transport into the cell.

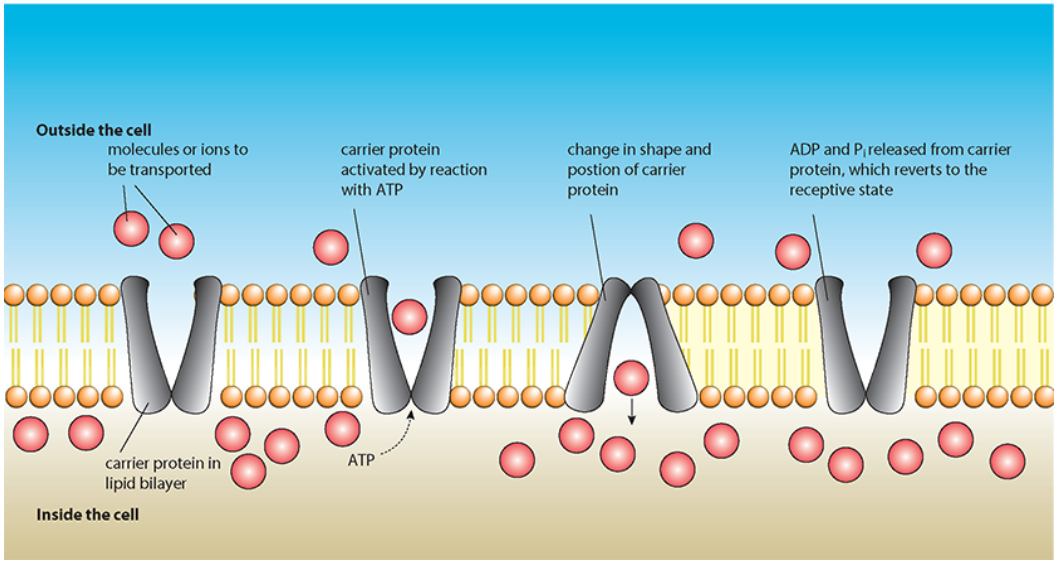


Figure 6.2.4: Active transport of a single substance.

The sodium–potassium pump found in nerve cell fibre membranes (Figure 6.2.5) is important in keeping the correct proportions of ions on the inside and outside of the nerve cell. Active transport is used to move these ions against their

concentration gradients. The sodium–potassium pump maintains the concentration of sodium and potassium ions in the cells and extracellular fluid. Cells are able to exchange sodium ions for potassium ions against concentration gradients using energy provided by ATP. Figure 6.2.5 shows this very important example of active transport. Chapter 7 covers nerve impulses in more detail.

KEY POINTS

endocytosis is the movement of liquids or solids into a cell, by the indentation of the plasma membrane to form vesicles containing the substance; endocytosis is an active process requiring ATP.

exocytosis the movement of liquids or solids out of a cell by the fusion of vesicles containing the substance with the plasma membrane; exocytosis is an active process requiring ATP.

Exocytosis and endocytosis

Cells often have to transport large chemical molecules or large materials across the plasma membrane. Neither diffusion nor active transport are suitable for this. Instead, cells can release or take in such materials in vesicles, as shown in Figure 6.2.6. Uptake into a cell is called **endocytosis** and export from a cell is **exocytosis**. Both require energy from ATP and both rely on the fluid nature of the plasma membrane.

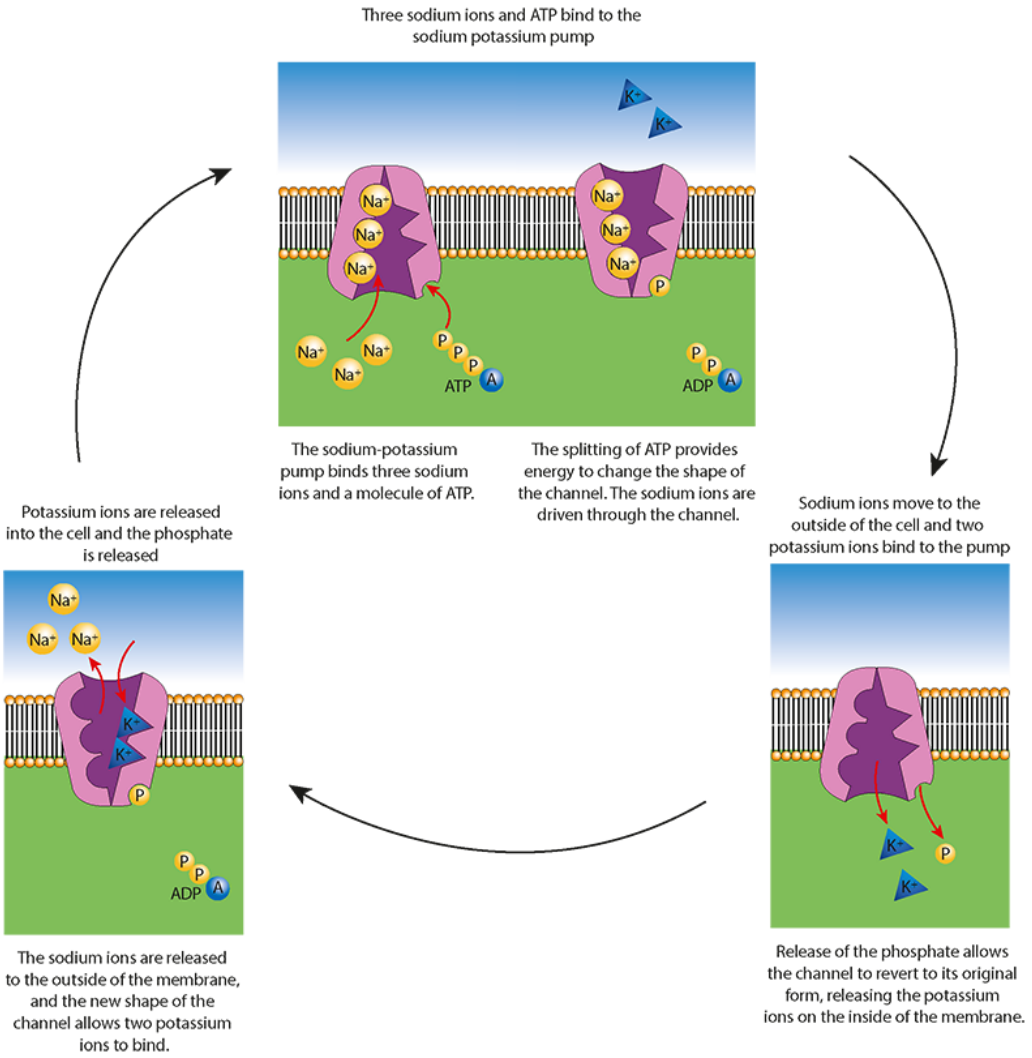
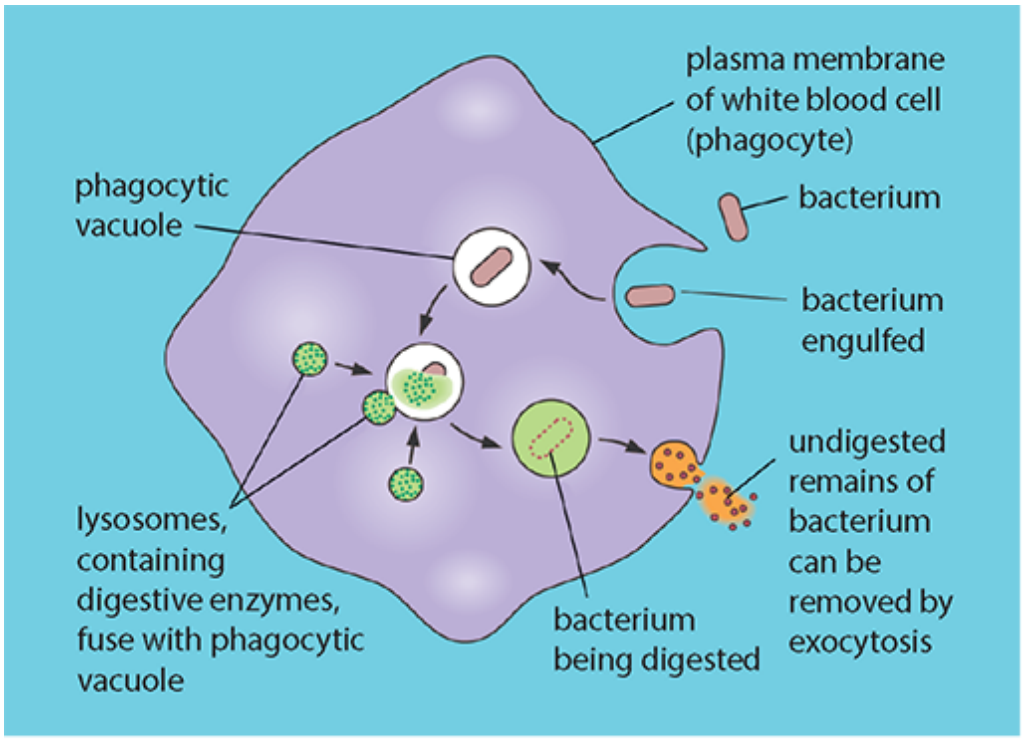


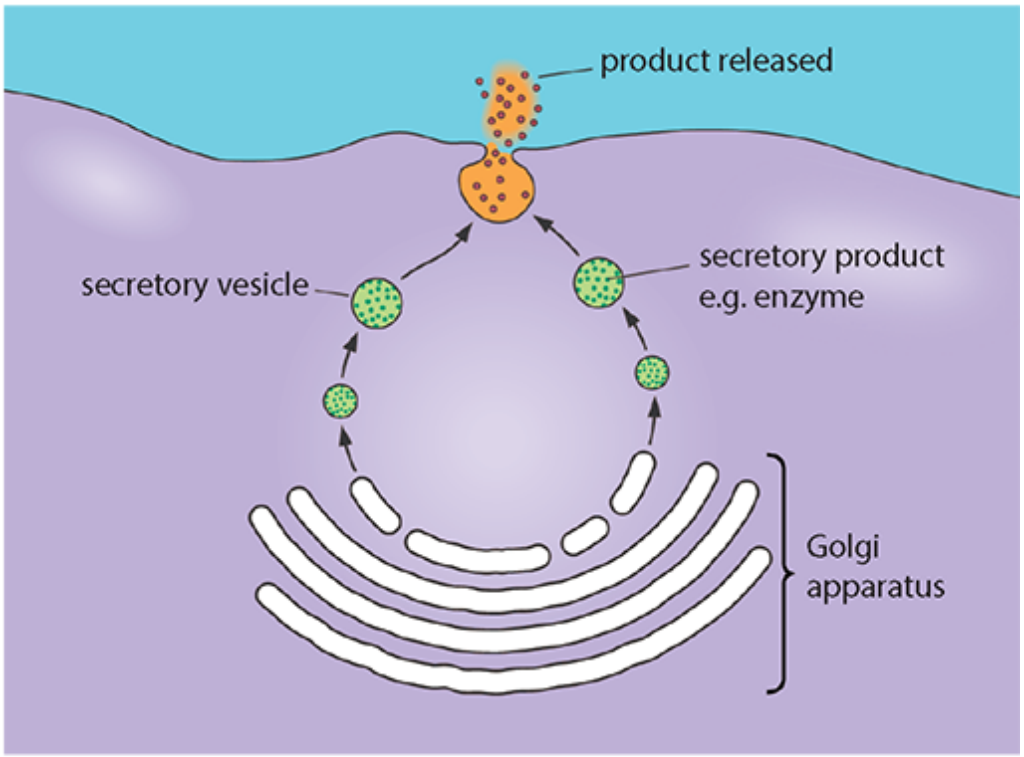
Figure 6.2.5: Sodium–potassium pump. At rest, sodium ions are pumped out of a nerve cell fibre and potassium ions are pumped in, to establish the resting potential.

During endocytosis, part of the plasma membrane is pulled inward and surrounds the liquid or solid that is to be moved from the extracellular space into the cell. The material becomes enclosed in a vesicle, which pinches off from the plasma membrane and is drawn into the cell. This is how white blood cells take in bacteria (Figure 6.2.6).

Materials for export, such as digestive enzymes, are made in the RER and then transported to the Golgi apparatus to be processed. From here they are enclosed within a membrane-bound vesicle, and moved to the plasma membrane along microtubules (Chapter 5). The arrangement of phospholipid molecules in the membrane of a vesicle is very similar to that in the plasma membrane. As a vesicle approaches the plasma membrane, it is able to fuse with it and in doing so release its contents to the outside. The flexibility and fluidity of the plasma membrane are essential to enable both endocytosis and exocytosis to happen. Vesicles also help to transfer and organise substances in the cell. They are involved in metabolism, transport and enzyme storage and some chemical reactions also occur inside them, separated from the cytoplasm in their own compartment.



Phagocytosis of a bacterium by a white blood cell – an example of endocytosis.



Exocytosis in a secretory cell. If the product is a protein, the Golgi apparatus is often involved in chemically modifying the protein before it is secreted, as in the secretion of digestive enzymes by the pancreas.

Figure 6.2.6: Examples of endocytosis and exocytosis.

There are two types of endocytosis. If the substances being taken in are particles, such as bacteria, the process is called phagocytosis. If the substances are in solution, such as the end products of digestion, then it is called pinocytosis.

EXTENSION

Nerve impulses are able to pass across synapses (the tiny gaps between one nerve cell and the next) due to exocytosis and endocytosis. Neurotransmitters are secreted at the end of a nerve cell fibre by exocytosis. They stimulate the adjacent nerve and are then reabsorbed by endocytosis to be recycled and reused. You can find out more about the transmission of nerve impulses in Section 7.2.

SCIENCE IN CONTEXT

Artificial membranes

As well as natural semi-permeable membranes, artificial membranes can be made using cellulose and they are used in both laboratory experiments and medicine. You may have used Visking tubing (also known as dialysis tubing) in your laboratory experiments to demonstrate diffusion and osmosis. The tubing can be produced in a variety of different forms with pores that have sizes between 1 and 10 nm, so that different types of membrane can restrict the passage of molecules with different molecular sizes. In your experiments you may have discovered that large molecules such as starch cannot pass through the tubing but smaller molecules such as glucose or maltose can do so. Unlike the membranes of living

cells dialysis tubing is not semi-permeable based on the charge of molecules, which means that ions can move freely across it.

Artificial semi-permeable membranes are also used to separate molecules from blood or DNA samples and are very important in kidney dialysis machines. If a person's kidneys fail, a dialysis machine can be used to clean their blood and remove excess salts, urea and water, a process usually done by the kidneys. A kidney dialysis machine contains an artificial semi-permeable membrane that uses the same selective absorption process as plasma membrane. The patient's blood passes through the machine and the dialysis membrane uses differential diffusion to restrict the entry of certain materials, such as glucose, which the body needs, but allows other waste products, such as urea, to diffuse through it. A concentration gradient on the different sides of the membrane is maintained by refreshing the dialysis fluid so that only unwanted substances diffuse across the membrane and the correct levels of other substances remain in the blood.

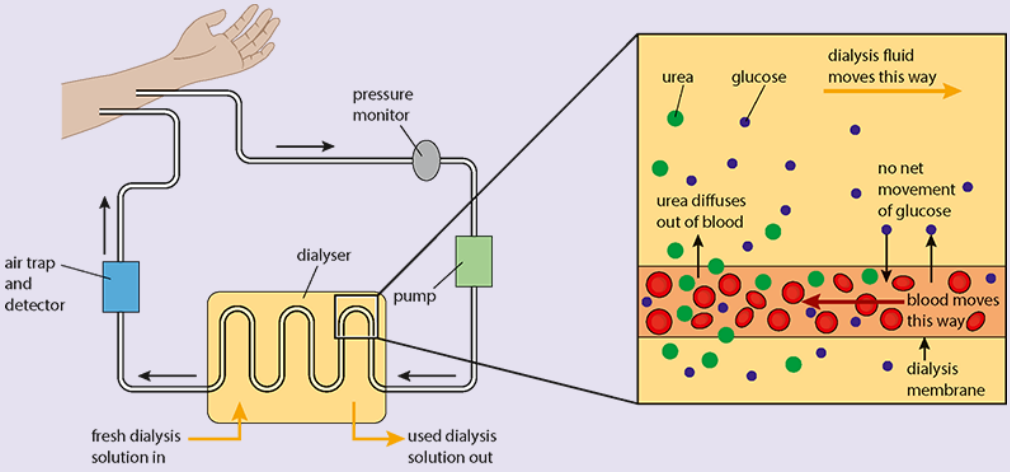


Figure 6.2.7: A dialysis machine.

TEST YOUR UNDERSTANDING

6 Outline the difference between simple diffusion and facilitated diffusion.

7 List three ways that substances move from one side of a membrane to the other.

8 State one transport mechanism across a membrane that requires energy from ATP and one that does not.

9 State one difference and one similarity between exocytosis and endocytosis.

6.2.3 Membranes and transmission of nerve impulses

Movement of ions across plasma membranes is essential for the conduction of nerve impulses. Neurons (nerve cells) are specialised cells which receive and transmit impulses throughout the body. A neuron consists of a cell body with two types of extensions that carry messages to and from the cell body. Long extensions called axons carry messages away from the cell body. Shorter, branched extensions called dendrites receive messages from other cells and carry them towards the cell body (Section 7.2). Active transport and the sodium–potassium pump not only maintain the sodium/potassium balance in these cells but also are key to the way nerve impulses are passed along the nerve cell fibres.

Gated ion channels are a special group of membrane protein channels that are activated by changes in **membrane potential** close to them (membrane potential is the difference in electrical potential between the inside and outside of the cell). Gated ion channels are important in nerve fibres because they allow ions to pass quickly in and out of a neuron when they are opened. When a neuron is not active, sodium ions are pumped out of the cell and potassium ions are pumped into the cell to establish a 'resting potential'. At rest, the inside of the nerve fibre is negatively charged with respect to the outside (Figure 6.2.8). The resting potential is about −70 mV.

When a neuron is stimulated the distribution of charge across its membrane is reversed for a millisecond. This change in membrane potential changes the shape of the sodium channel proteins and triggers a rapid **depolarisation** (the distribution of

charge is reversed from negative to positive). Sodium ions diffuse into the neuron down the concentration gradient into the nerve fibre. Nerve impulses pass along the neuron because as charge is reversed in one area of a nerve fibre, the adjacent ion channel is affected so that the depolarisation spreads rapidly along the axon. As this happens, a nerve impulse is passed along the nerve fibre.

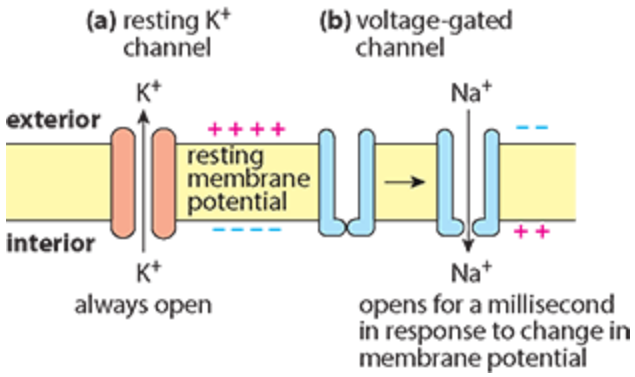


Figure 6.2.8: When a neuron is depolarised, voltage-gated protein channels open to allow sodium ions to rush in.

After a nerve impulse has passed, the gated sodium channels close and the resting potential is re-established. This happens as gated potassium channels open to allow potassium ions to flow out, in a process known as repolarisation.

Speed of transmission

A nerve fibre with a simple structure carries impulses at a speed of about 1 metre per second. Nerve fibres with a larger diameter conduct impulses faster than smaller ones and part of the reason for this is that most large fibres are enclosed in a fatty covering called a myelin sheath (Figure 7.2.3). The sheath is produced by special cells known as Schwann cells which wrap themselves many times around the nerve fibre, producing several membrane

layers around it. The myelin prevents the flow of ions across the membrane for most of its length but at intervals there are gaps in the sheath known as nodes of Ranvier. Impulses can 'jump' from node to node in a process called **saltatory conduction** which speeds up the transmission of the nerve impulse (action potential). Changes in the membrane potential (the action potential) only occur at the nodes of Ranvier which makes the process much faster than if the process occurred continuously along the whole nerve fibre. The speed can reach up to 100 metres per second in some neurons with a myelin sheath. Clusters of sodium-gated channels are found at the nodes of Ranvier. The transmission of nerve impulses is explained in more detail in Section 7.2.

NATURE OF SCIENCE

What role do chance meetings and unexpected discoveries play in scientific progress?

Theodor Schwann discovered the existence of the myelin covering around nerve fibres that now bears his name. But, by chance, in the 1830s Schwann met another scientist Matthias Schleiden and the two men spoke about a very different subject, the nuclei of plant and animal cells. Schleiden said that he thought that new plant cells formed from the nuclei of existing cells and Schwann recalled seeing similar things occurring in animal cells. After the conversation, Schwann put forward his idea that 'All living things are composed of cells and cell products'. By the 1860s his ideas were accepted and formed the basis of what we know today as Cell theory. Schwann's theory and observations formed the foundation of modern cell biology.

TEST YOUR UNDERSTANDING

10 What stimulates a gated ion channel to open?

11 What effect does a myelin covering have on the transmission of a nerve impulse?

REFLECTION

Reflect upon the most interesting new aspects of biology that you have learned from this chapter.

Links

- Why is phagocytosis important for leucocytes? (Chapter 10)
- Why is membrane transport important at synapses? (Chapter 7.2)
- How is the uptake of glucose into cells regulated by insulin? (Chapter 8.4)

6.3 Water potential

LEARNING OBJECTIVES

In this section you will:

- recall (from Chapter 1) that water is a good solvent because of the interaction between solute and water molecules
- learn that water potential defines whether water will enter or leave a solution. It is determined by solute potential and pressure potential of the solution
- learn that the more negative the solute potential, the more solutes are dissolved in the solution
- recognise that a higher pressure potential means more pressure is exerted on the solution
- discover that the movement of water into a plant cell produces turgor pressure but that water moving into an animal call can cause it to burst
- learn that in a hypertonic solution water leaves cells, resulting in cell shrinkage, in a hypotonic solution water enters cells and that in an isotonic solution there is no net movement of water
- recognise that isotonic solutions have important medical uses
- define osmoregulation as the maintenance of a consistent osmotic pressure in an organism

- appreciate that water moves from a higher to lower water potential
- understand that solute potential and pressure potential contribute to the water potential of plant cells
- explain in terms of solute and pressure potentials the changes that occur when plant cells are placed in hypotonic and hypertonic solutions

GUIDING QUESTIONS

- Why is the regulation of water content in living organisms important?
- How is internal water content regulated?

6.3.1 Water potential in plants and animals

An understanding of the processes of osmosis is important for learning how the human kidney works and how plants absorb water from the soil. Water potential is the term used to explain whether water will enter or leave a solution or a cell. It is affected by the concentration of solute molecules present and by the pressure of water molecules in the solution.

Water molecules, enclosed inside a semi-permeable membrane, hit the membrane as they move about and generate pressure. This pressure is called the water potential. The more molecules that hit the membrane in a certain time, the higher the pressure and higher the water potential.

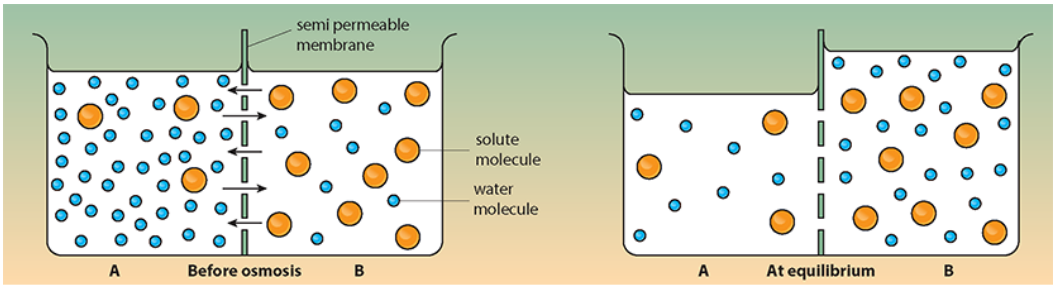
Water potential is represented by the Greek letter psi (ψ) and is measured in kilopascals (kPa). The water potential of pure water is 0 kPa. Water potential determines whether water will move into or out of a solution.

The presence of solutes (dissolved substances) increases the tendency of water to move from one solution into a more concentrated solution by osmosis.

Water potential = pressure potential + solute potential

$$\psi = \psi_P + \psi_S$$

Osmosis occurs when there is a net movement of water molecules from a region where their concentration is higher to one where their concentration is lower. This occurs through a semi-permeable membrane such as the plasma membrane (Figure 6.3.1).



Two solutions are separated by a semi-permeable membrane. B has a higher solute concentration than A. The solute molecules are too large to pass through the pores in the membrane but the water molecules are small enough.

As the arrows in the left diagram indicate, more water molecules moved from A to B than from B to A, so the net movement has been from A to B, raising the level to the solution in B and Lowering it at A. The water potentials (the tendency of water molecules to move in each direction) in A and B are now the same.

Figure 6.3.1: Osmosis and water potential.

When the solute concentrations inside and outside a cell are the same, the same number of water molecules will pass across the membrane into the cell as those that leave.

An animal cell that is placed in pure water will take in water by osmosis until eventually it may burst (Figure 6.3.2). This is because the cell contains many dissolved substances in its cytoplasm and water enters freely through the plasma membrane. If the cell is placed in a solution with a very high concentration of solutes, the cell will shrink as water leaves the cell by osmosis. This shrinkage is known as **crenation**. In either situation, animal cells will not function properly and their metabolism will be affected.

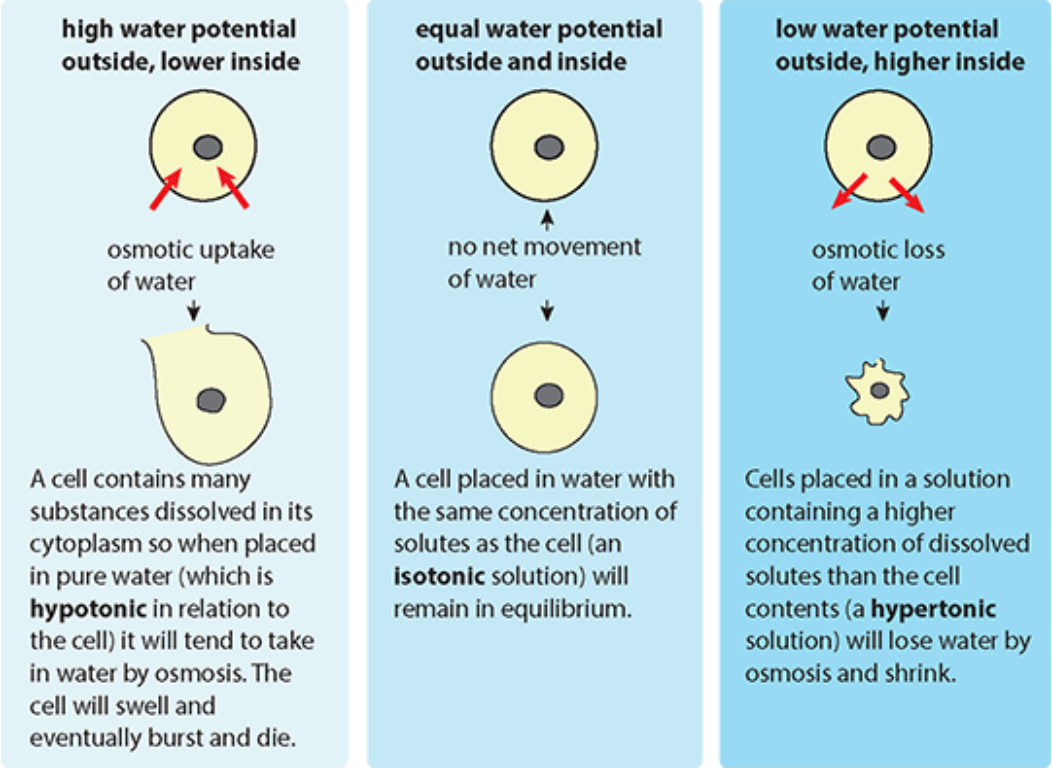


Figure 6.3.2: Responses of animal cells to solutions containing different concentrations of solutes.

KEY POINTS

water potential defines of the tendency of water to move from one area to another due to concentration of solute molecules present and the pressure of water molecules in the solution. Water potential is determined by these two factors, the solute potential and the pressure potential of a solution.

pressure potential the component of water potential due to the hydrostatic pressure that is exerted on water in a cell; in a turgid plant cell it has a positive value, water enters the vacuole and the cell membrane is pushed up against the cell wall.

KEY POINT

solute potential also known as osmotic potential; pressure needed to be applied to a solution to prevent the inward flow of water across a semi-permeable membrane.

Plant cells are also affected by the movement of water into and out of their cells but the presence of a cell wall prevents plant cells being damaged or bursting. If a plant cell is put into water that is hypotonic to the content of the plant cell, that is containing a lower concentration of solute than the inside of the cell, water will enter by osmosis. However, the plant cell wall resists the entry of further water once the cell is full. A plant cell that is full becomes firm and rigid, a condition known as turgor. The force within a cell that pushes the plasma membrane against the cell wall is known as turgor pressure.

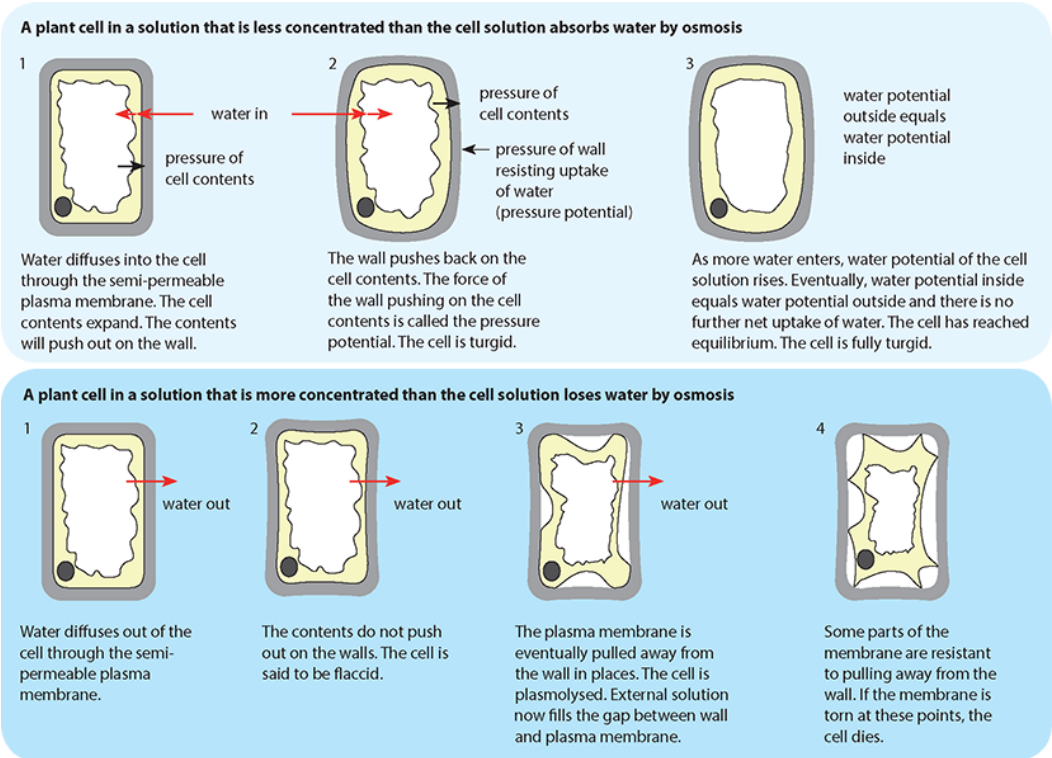


Figure 6.3.3: The effect of hypotonic and hypertonic solutions on a plant cell. Plant cells are not damaged when water enters by osmosis because their cell wall protects them.

If a plant cell is placed in an isotonic solution, water will enter and leave the cell in equal amounts. But, if the cell is placed in a hypertonic solution, containing a higher concentration of solute than the inside of the cell, water will move out of the cell (Figure 6.3.3). The cell inside the cell wall becomes flaccid and the plasma membrane is pulled away from the cell wall. This is called **plasmolysis**.

TEST YOUR UNDERSTANDING

12 Look at the graph in worked example 6.2. Describe and explain the change in mass in 0.0 M (distilled water) and 1.0 M sucrose solution.

13 What is the term used to describe the appearance of the cells in 0.0 M distilled water?

14 Why is it important to use three samples for each molarity of solute?

15 Why is the percentage change in mass used in the experiment?

KEY POINTS

hypotonic is when a solution is having a lower concentration of solutes (a more negative water potential) than the cell contents; this causes water to enter the cell making it swell.

isotonic is having the same osmotic concentration of the same water potential as another solution; there is no net movement

of water.

turgor pressure is the force within a cell that pushes the plasma membrane against the cell wall.

hypertonic having a higher solute concentration (less negative water potential) than inside the cell; this causes water to leave the cell, making it shrivel.

Isotonic solutions are used in hospitals to replace lost bodily fluids if a person is dehydrated due to burns or intestinal disorders (Figure 6.3.5). They are also used to surround organs that are being prepared for transplantation. The intracellular and extracellular environments have the same osmotic pressure. This means that there is no net movement of water into or out of the patient's cells or the organ for transplant, so the cells are kept alive and healthy.

WORKED EXAMPLE 6.2

Estimating the water potential of potato tissue

To estimate the water potential of potato cells we can bathe potato samples in hypotonic and hypertonic solutions to establish when there is no net movement of water between the intracellular environment and the extracellular (external) environment of the cells. You may have carried out these experiments yourself. When the pressure inside the cell becomes large enough, no additional water will enter the cell due to the presence of the cell wall. The movement of water cannot be predicted from the relative solute concentrations inside and outside the plant cell wall. Instead, we must consider the water

potential to predict the direction in which water will diffuse.

1 Equally sized samples of potato are cut, blotted to remove excess water and carefully weighed. Their mass is recorded.

2 The samples are placed in sucrose or salt solutions of different molarity. Three samples are used in each solution.

3 After a period of time the samples are removed, blotted and reweighed.

4 The percentage change in mass for each sample is plotted against the molarity of the solutions (Figure 6.3.4).

5 The graph shows that in this experiment the concentration of sucrose causing no change in mass has a molarity of 0.3 M. At this point we can say that there is no net movement of water molecules and the solutions inside and outside the cells are isotonic.

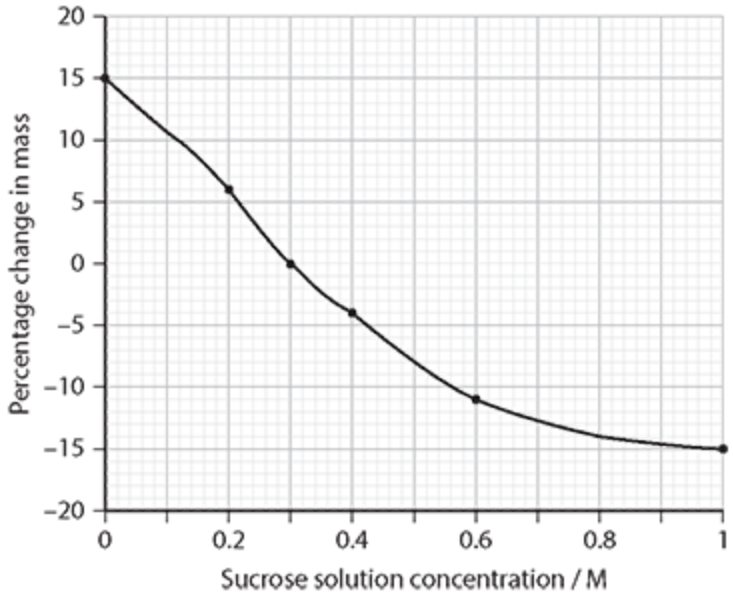


Figure 6.3.4: Graph to show the percentage change in mass of potato samples placed in different molarities of sucrose solution.

Answer

Pure water has a water potential of 0, which is higher than the water potential inside the cell. In these conditions there will be a net movement of water into the cell. The pressure potential inside the cell will increase as water enters until the cell reaches a state of equilibrium. In different molarities of sucrose solution, water will either enter or leave the potato cells. Where the line on the graph crosses the x-axis is the point at which the molar concentration of sucrose has a water potential that is equal to the water potential of potato tissue.

Figure 6.3.5: This person is being given intravenous fluid from an isotonic saline drip to replace lost body fluids quickly.

SCIENCE IN CONTEXT

Sports drinks and exercise

If you take exercise you may have bought a sports drink to refresh yourself afterwards. Manufacturers advertise these drinks to help athletes hydrate before, during and after exercise to improve their sports performance and minimise fatigue. Rehydration is important during exercise as performance deteriorates rapidly when a person is dehydrated. Sports drinks supply energy, replace salts such as sodium and potassium, which are lost through sweat, and replace water.

There are several types of sports drinks available and all of them contain various levels of fluid, electrolytes (salts) and carbohydrate. Hypotonic drinks have low carbohydrate

content and a lower concentration of salt and sugar than human cells. They are designed to replace fluids rapidly. Athletes, such as gymnasts, who require water but not carbohydrate during their performances, may use them. Isotonic drinks have a similar concentration of salts and sugar as human cells and also replace fluids quickly. These supply carbohydrates and most athletes who are involved in middle- or long-distance running would chose them. Popular sports drinks are a compromise designed to meet the needs of most people in many different situations. No one formula will suit everyone because individuals vary. Check the labels and ingredients of drinks you buy and ask yourself:

- Are you paying for a product you don't need?
- Could you drink water instead of a commercially made sports drink?

6.3.2 Advanced water potential

Water potential is a measure of the potential energy in water as well as the difference between the potential in a given sample and pure water. Potential energy is the energy that is stored in water but this is impossible to measure precisely. For this reason, values relative to pure water at atmospheric pressure and 20°C are used. Water potential is measured in kilopascals (kPa)

The plasma membrane of a plant cell is differentially permeable and a plant cell functions as an osmotic system. Water can pass through the plant membrane and so can some solutes, but not other. The cell wall is permeable to practically all solutes and not to water. This means that there is a relationship between the water potential of the cell contents and the turgor pressure of the cell and the water potential of the cell as a whole. Water molecules and mineral ions enter plants cells by physical processes. Water will always move from regions of higher water potential to regions of lower water potential, that is from a region of higher potential energy to a region where the potential energy is lower.

$\psi = \psi_p + \psi_s$

- Where ψ is the water potential
- ψ_p is the pressure potential
- ψ_s is the solute potential

Solute potential and pressure potential both contribute to the water potential within the walls of plants. Solute potential (ψ_s) is negative in a plant cell and zero in distilled water while pressure potentials are usually positive inside cells. If a plant cell is

immersed in water it will becomes fully turgid as water enters it. If it is immersed in a solution with a water potential of –0.5MPas it will soften and lose its turgor as water leaves the cell, if it is immersed in a solution of –1.0MPas it will lose all turgor.

If some plant cells were examined they might have the following values:

$$\begin{array}{ccccc} \psi & = & \psi_p & + & \psi_s \\ -0.5\ \text{MPa} & = & +0.5\ \text{MPa} & + & (-1.0\ \text{MPa}) \end{array}$$

If this cell were place into a **hypotonic** solution of pure water with a water potential of 0.0 MPa water would flow into it with a force equal to the difference between the water potential of the cell (ψ) and the zero potential of the water, that is a force of 0.5 MPa. This would eventually raise the pressure potential (ψ_p) of the cell until it reached its maximum of 1.0 MPa and full turgor. At this point the cell would have a water potential of zero.

$$\begin{array}{ccccc} \psi & = & \psi_p & + & \psi_s \\ 0.0. & = & +1.0. & + & (-1.0) \end{array}$$

If a fully turgid cell is placed in a **hypertonic** solution of salt with a solute potential of –0.5 MPa. Then the cell with a ψ of 0 would be at a higher water potential than the surrounding solution. Water would flow out of the cell and reduce its turgor until a new equilibrium was reached and pressure potentials inside and outside the cell are equal. This can be shown in this equation

$$\begin{array}{ccccc} \psi_s & = & \psi_p & + & \psi_s \\ \text{External pressure} & = & \psi_p & + & \text{Internal pressure} \\ -0.5 & = & +0.5 & + & (-1.0) \end{array}$$

REFLECTION

Reflect upon the most interesting discoveries you made when working on this chapter. How does your experience compare with that of your classmates?

TEST YOUR UNDERSTANDING

16 Describe the effect of a hypotonic solution on a plant cell and on an animal cell.

17 Outline the reasons for plasmolysis in a plant cell and crenation in an animal cell.

18 Why are isotonic solutions important in medical treatments?

Links

- How are concentration gradients important in moving materials in and out of cells? (Chapter 6.2)
- Which are the most useful properties of water to living organisms? (Chapter 1)

6.4 Limitations to cell size

LEARNING OBJECTIVES

In this section you will:

- understand that cells obtain substrates for metabolism from their environment and remove wastes through the plasma membrane
- learn that a large surface area to volume ratio ensures that this process is efficient
- recognise that as cells grow larger they require more resources, but the rate of exchange of materials depends on the surface area of the cell
- understand that as cells grow the surface area to volume ratio decreases
- discover that structures, such as folds, or flattening of a cell can increase the surface area to volume ratio
- learn that cells maintain their volume but increase their surface area by dividing in two

- understand that the size of cells in a tissue depends on rate of growth and rate of division
- learn that cells are influenced by extracellular signals including growth factors and mitogens.

GUIDING QUESTIONS

- Why are surface area and volume important for growing cells?
- How do cells maintain a high surface area to volume ratio?

6.4.1 Surface area to volume ratio

Cells are very small, no matter what the size of the organism that they are part of. Cells do not and cannot grow to be very large and this is important in the way living organisms are built and function. The volume of a cell determines the level of metabolic activity that takes place within it. The surface area of a cell determines the rate of exchange of materials with the outside environment. As the volume of a cell increases, so does its surface area, but not in the same proportion. Table 6.4.1 shows this for a theoretical cube-shaped cell. As a cell grows larger, it has proportionately less surface area to obtain the materials it needs and to dispose of waste.

As a cell grows in size, the rate of exchange of materials across the outer membrane may become limiting and will not keep up with the cell's requirements. The distance from the plasma membrane to the interior of the cell also increases. This makes diffusion to the centre of the cell slower, so that that cell cannot function efficiently.

Side of cube/mm	Surface area/mm^2	Volume/mm^3	Ratio of surface area : volume
1	6	1	6 : 1
2	24	8	3 : 1
3	54	27	2 : 1

Table 6.4.1: Surface area to volume ratios for a cube.

Increasing surface area

Some cells have specialised structures, such as folds and microvilli, to provide a larger surface area relative to their volume (Figure 6.4.1). Many multicellular structures, such as the digestive system and the lungs (Section 8.3), also have folded surfaces to increase their surface areas for the exchange of materials.

TEST YOUR UNDERSTANDING

19 Many cells are roughly spherical in shape. The volume of a sphere is πr^3 and its surface area is $4\pi r^2$, where r is the radius. Make a table similar to Table 6.4.1, this time for a sphere using different radii as a starting point. Describe the relationship between surface area and volume in this case.

20 Take a 2 cm cube of modelling clay. Change its shape so that it becomes a cuboid, a thin cylinder or a sphere. Calculate its surface area each time. Try creating folds in the surface. Which shape produces the greatest surface area?

A flattened or elongated shape also increases a cell's surface area to volume ratio (Figure 6.4.2).

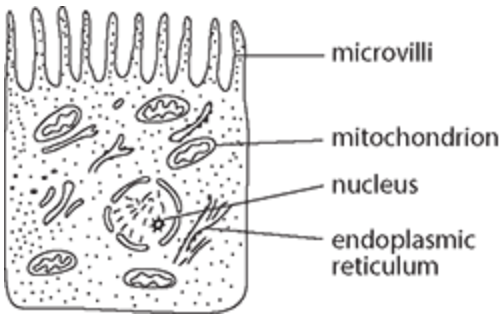


Figure 6.4.1: Microvilli are folds that increase the surface area of cells so that materials can be exchanged. Cells like these are

found in the intestine and in the lining of the trachea.

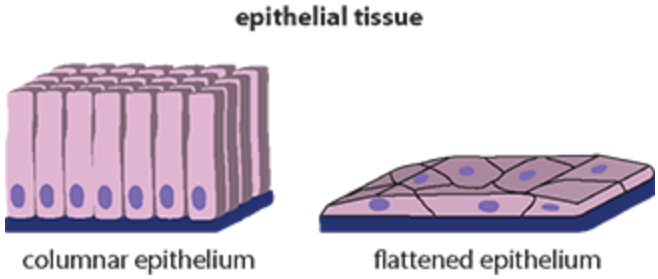


Figure 6.4.2: Flattening and elongation are methods of increasing the surface area to volume ratio of cells.

But, whatever the modifications a cell adopts, there will always be a limit to the size of a single cell. When a cell becomes too large and its surface area to volume ratio is insufficient to supply its needs, it must divide. Division maintains the same total volume in the two new cells that form, but each has its own surface to provide an optimum rate of movement across the plasma membrane (Section 6.5 for more on cell division).

6.4.2 Cell growth and division

The size of cells in the body varies depending on the rate of growth of the cells and their rate of division. Table 6.4.2 shows the volumes of some human cells. Their sizes and shapes are related to the function of each type of cell. Red blood cells are very small, biconcave discs (round cells which are concave on both sides). This is a shape which maximises the surface area to volume ratio available for the exchange of oxygen and carbon dioxide. Epithelial cells are bound together and form layers called epithelium which line the inside of the intestine and other hollow structures in the body. These cells are small with a large surface area which is very important for absorption and secretion of materials.

Cell type	**Average volume / µm^3**
red blood cell	100
columnar epithelium in the intestine	1400
fibroblast (cells which secrete collagen and form a framework for tissues)	2000
heart cell	30 000
fat storage cell	600 000
oocyte	4 000 000

Table 6.4.2: The average volume of some cell types.

Cells such as fat cells and oocytes have very large volumes because they store materials and oocytes do not divide. In contrast, cells of the intestine are many times smaller and divide

frequently as they are damaged by digesting food materials flowing past them. Neurons can be over 1 metre long to extend from the spinal cord to the legs but are only about 10 µm wide.

Both bacterial and eukaryotic cells have characteristic sizes but cell size is not rigidly fixed and varies slightly in response to external factors, such as nutrient levels. The average size of a cell can change when the conditions for its growth change. The size a cell eventually becomes when it is ready to divide is determined by a balance between cell growth (the increase in mass or volume) and the timing of division.

Cell division is influenced by signals from outside the cell. Cells increase in size under the stimulation of **growth factors**. Growth factors are protein molecules made by the body. They regulate cell growth and cell survival. Cell growth is a separate process to cell division: some cells can grow without dividing (fat cells, muscle fibres and neurons do this), or divide without growing (fertilised egg cells divide without growing as their cytoplasm is shared between many small cells in the early stages of development).

Most cells must reach a specific minimum size to progress in the cell cycle (Section 6.5) and this progress is controlled by another groups of proteins called mitogens. A **mitogen** is a peptide or small protein that causes a cell to enter mitosis. The proteins bind to receptors on the plasma membrane and activate cell division. Some stimulate division in many cell types while other are specific to just one type of cell.

NATURE OF SCIENCE

Models can help our understanding

Models are simplified versions of things that can be very complex. Think about the figures and calculations about surface area and volume shown in Table 6.4.1. We can make model cubes with different side lengths as an easy way to investigate surface area to volume ratios. We can use them to study what happens to the surface area of a cell as it becomes larger. The cubes do not have the same shape as real cells but the scale factor (as you enlarge the shape and each side is multiplied by the same number) will work in exactly the same way as that for cells.

TEST YOUR UNDERSTANDING

21 Explain the importance of surface area to volume ratio to a cell.

22 Outline two ways in which a cell can increase its surface area.

23 Name two substances that enter and one that leaves a cell through its plasma membrane.

24 Describe the difference between a growth factor and a mitogen.

Links

- How did the evolution of eukaryotes allow cells to become larger? (Chapter 6.5)
- How does a cell wall influence the size of a plant cell? (Chapter 6.3)

- Is there a relationship between cytoplasmic volume and nuclear volume? (Chapter 6.5)

6.5 Cell division

LEARNING OBJECTIVES

In this section you will:

- recall that new cells are produced by cell division
- learn that single-celled organisms reproduce asexually by binary fission
- discover that the cell cycle consists of interphase, mitosis and cytokinesis
- discover that cytokinesis can be asymmetric in some cases
- understand that metabolic reactions and DNA replication occur during interphase.
- define mitosis as the division of the nucleus
- learn that DNA assisted by histones, supercoils and condenses for mitosis
- define cytokinesis as the division of the cell which occurs after mitosis
- learn that cell division is controlled for appropriate growth, development and repair
- define meiosis as cell division that produces gametes
- understand that non-disjunction occurs if homologous pairs of chromosomes fail to separate during anaphase 1

- recognise that chromosome behaviour in metaphase I demonstrates the patterns of inheritance expressed in Mendel's second law
- learn that crossing over and random orientation of chromosomes at meiosis increases genetic variation

- recall that cell proliferation is needed for growth and repair
- understand the phases of the cell cycle and that cylins are responsible for controlling stages of the cycle
- learn that mutations in genes that control the cell cycle can lead to the development of tumours
- recognise that malignant tumour cells have faster rate of cell division than normal cells and the capacity for metastasis
- define the terms benign, malignant, primary tumour in relation to tumour growth and cancers
- understand the use of mitotic index to monitor cell division

GUIDING QUESTIONS

- How does cell division make a species more resilient if environmental conditions change?
- Why must the cell cycle be efficiently controlled?

6.5.1 Binary fission in single-celled organisms

Simple single-celled organisms are usually much smaller in volume than more complex cells. Their means of reproduction is also simple. As they grow, their DNA replicates and separates into two different areas of the cytoplasm. The cytoplasm then divides in two and forms two identical new daughter cells. This is called **binary fission** (Figure 6.5.1). Many organisms reproduce themselves using binary fission; examples include the unicellular organisms such as amoebae, *Paramecium* spp. and yeast. Reproducing in this way is known as **asexual reproduction** as no gametes are involved and the offspring are genetically identical to the parent.

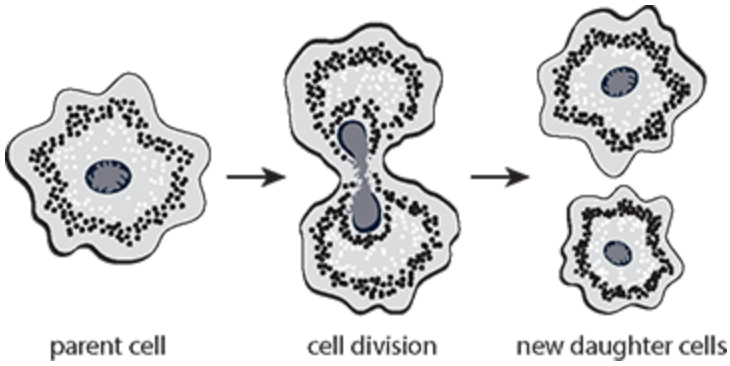


Figure 6.5.1: Binary fission in amoebae.

6.5.2 The cell cycle

In larger organisms, new cells are needed to replace cells that have died and to allow an organism to grow in size. The nucleus and cytoplasm of a cell divide in processes known as **mitosis** and **cytokinesis**, which are phases in a series of events known as the **cell cycle**. Mitosis occurs in tissues such as plant meristems and animal embryos where rapid growth is taking place. It also occurs in tissues that have been damaged and need repair, such as when the skin is cut or scratched and a wound needs healing.

The cycle of a cell's life can be divided into three stages, as shown in Figure 6.5.2:

1 interphase

2 mitosis (division of the nucleus)

3 cytokinesis (division of the cytoplasm).

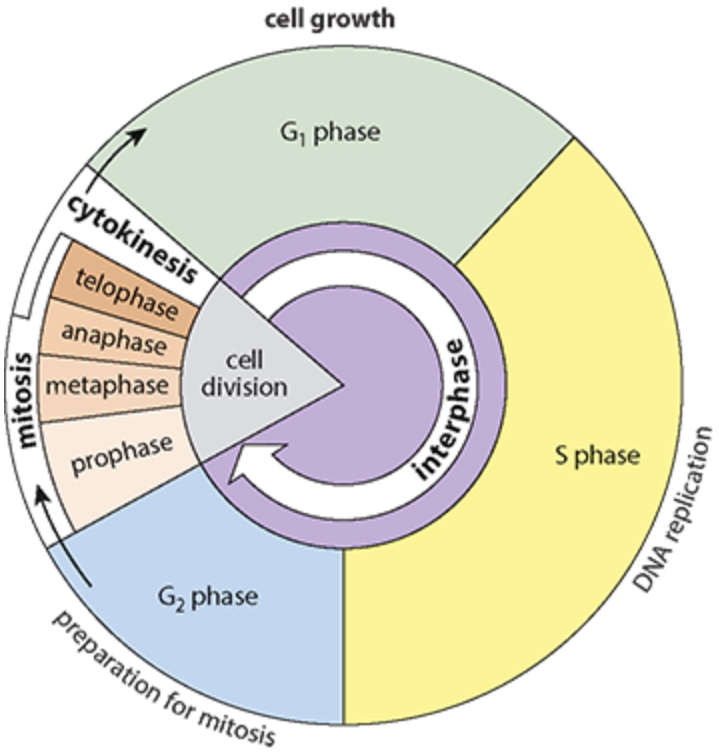


Figure 6.5.2: Summary of the cell cycle.

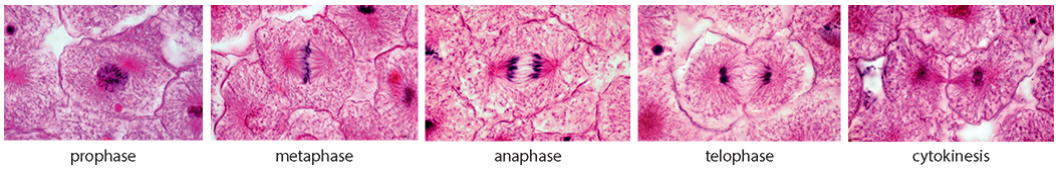


Figure 6.5.3: Stages of mitosis and cytokinesis in stained onion cells, as seen in a root squash preparation (×900 magnification).

G_1, S and G_2 are the three stages of the part of the cell cycle known as interphase, which is described in the next section. Mitosis and cytokinesis are the phases when the cell nucleus and cytoplasm divide. The phases of the cell cycle are summarised in Table 6.5.1.

Phase of the cell cycle		**Activities in the cell**
interphase	G_1	cell growth

		DNA transcription protein synthesis
	S phase	DNA replication
	G_2 phase	cell prepares for division
mitosis		cell nucleus divides
cytokinesis		cytoplasm divides

Table 6.5.1: Summary of the phases and events of the cell cycle.

Interphase

During most of the life of a cell, it performs the task for which it has been pre-programmed during differentiation. This period is called **interphase**. Part of interphase is spent in preparation for cell division (the **G_2 phase**) and part of it is the period immediately after division (the **G_1 phase**). The two stages of cell division are the separation and division of the chromosomes (mitosis), and the division of the cell into two daughter cells (cytokinesis).

If you look at a cell during interphase using a light microscope, not much appears to be happening, but this is a very active phase of the cell cycle when the cell carries out its normal activities and also prepares itself for mitosis. In the nucleus, the DNA in the chromosomes is replicated, double-stranded DNA molecules are copied (**S phase**), so that after cell division there will be exactly the same number of chromosomes in the two daughter cells. During interphase many proteins necessary for the division need to be synthesised at the ribosomes in the cytoplasm. The number of mitochondria increases so that the respiratory rate can be rapid enough to provide energy for cell division. In the case of plant

cells with chloroplasts, the number of chloroplasts increases so there will be enough for each daughter cell.

KEY POINTS

anaphase is the stage in cell division in which homologous chromosomes (in meiosis I) or chromatids (meiosis II and mitosis) separate and move to opposite poles.

centriole refers to a cylindrical structure in an animal cell that forms and organises the spindle microtubules in cell division.

centromere is the region where sister chromatids are joined and where the spindle microtubule attaches during cell division.

interphase refers to the period between successive nuclear divisions when the chromosomes are uncoiled and the cell is actively transcribing and translating genetic material.

metaphase is stage in nuclear division at which chromosomes become arranged on the equator of the spindle.

prophase is the first stage in cell division by meiosis or mitosis.

sister chromatids are two joined copies of a chromosome after it has replicated and before the centromeres separate at anaphase.

spindle is structure formed of microtubules to which centromeres attach during meiosis and mitosis.

supercoiling refers to winding of two complementary strands of DNA around one another and around a common axis.

telophase is the final phase of mitosis when duplicated genetic material is separated into the nuclei of two identical daughter cells

Mitosis

The two new cells that form after mitosis and cytokinesis are genetically identical. These processes allow an organism to grow more cells, or to repair injured tissue by replacing damaged cells, or to make new cells to replace old ones. Mitosis is also the way in which an embryo grows from a fertilised egg during development.

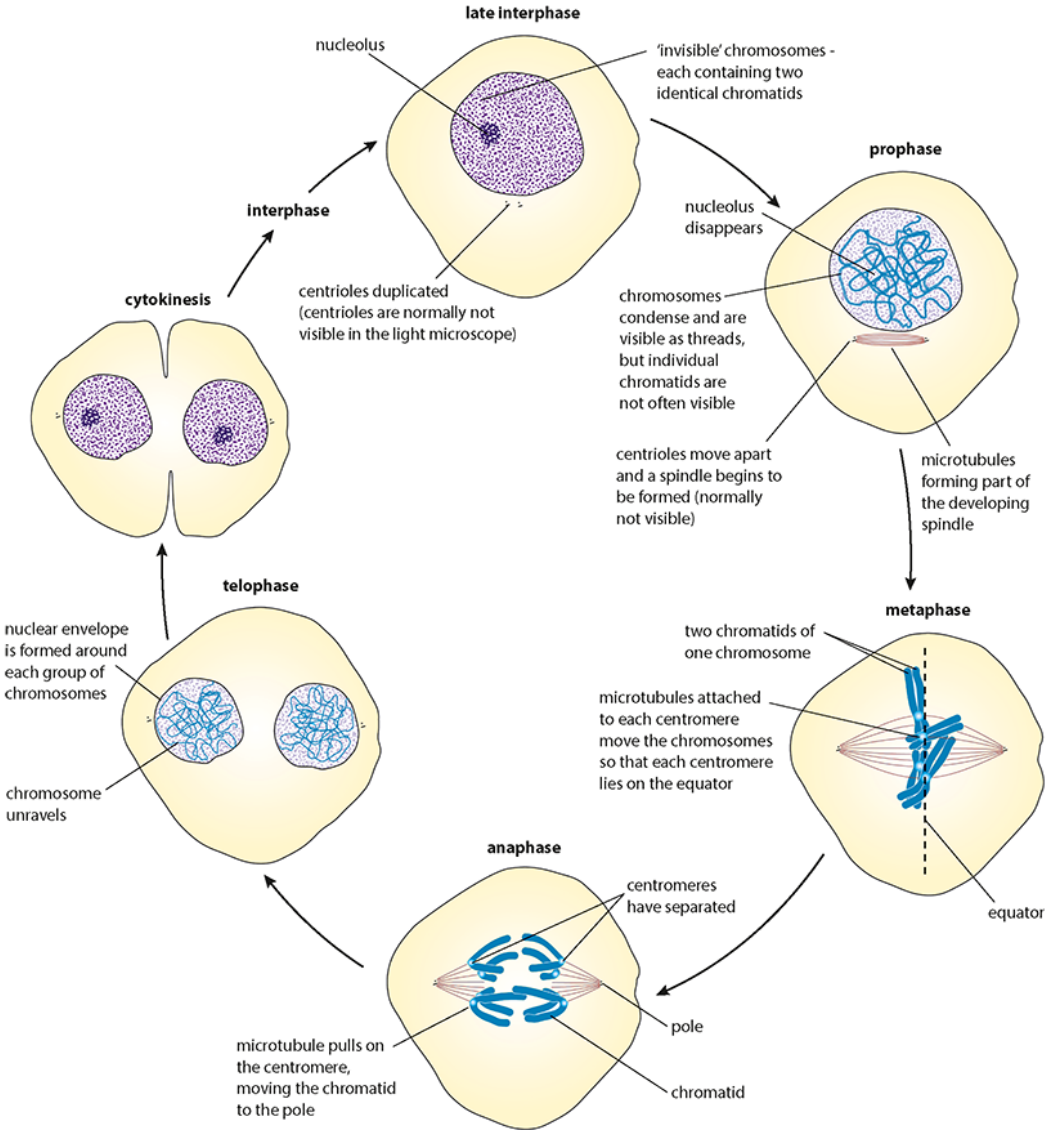


Figure 6.5.4: The stages of the cell cycle, including mitosis. Note that the cells are shown with just four chromosomes here, to make it easier to understand the process.

There are four distinct stages in mitosis, although the process is continuous, with each stage running into the next. There are no intervals in between the stages. Figures 6.5.3 and 6.5.4 show the stages of mitosis in detail.

Prophase

During prophase, chromosomes become visible using a microscope. During interphase they have been drawn out into long threads, giving the cellular machinery access to the genes. But now, the chromosomes coil round histone proteins and around themselves several times to produce a supercoil (Figure 6.5.5). Supercoiling prevents transcription and makes the chromosomes shorter and thicker.

It also reduces the space that they take up and enables them to take part in the processes that follow.

We can follow the stages of mitosis because supercoiled chromosomes can be seen using a microscope. DNA was replicated during interphase so at this stage each chromosome consists of two identical copies. These two copies are called the sister chromatids and are attached to each other at a place called the centromere. Also visible at this time are structures known as centrioles, which move to opposite sides of the cell as microtubules form between them. This microtubule structure is called the spindle. As prophase draws to a close, the nuclear envelope breaks down.

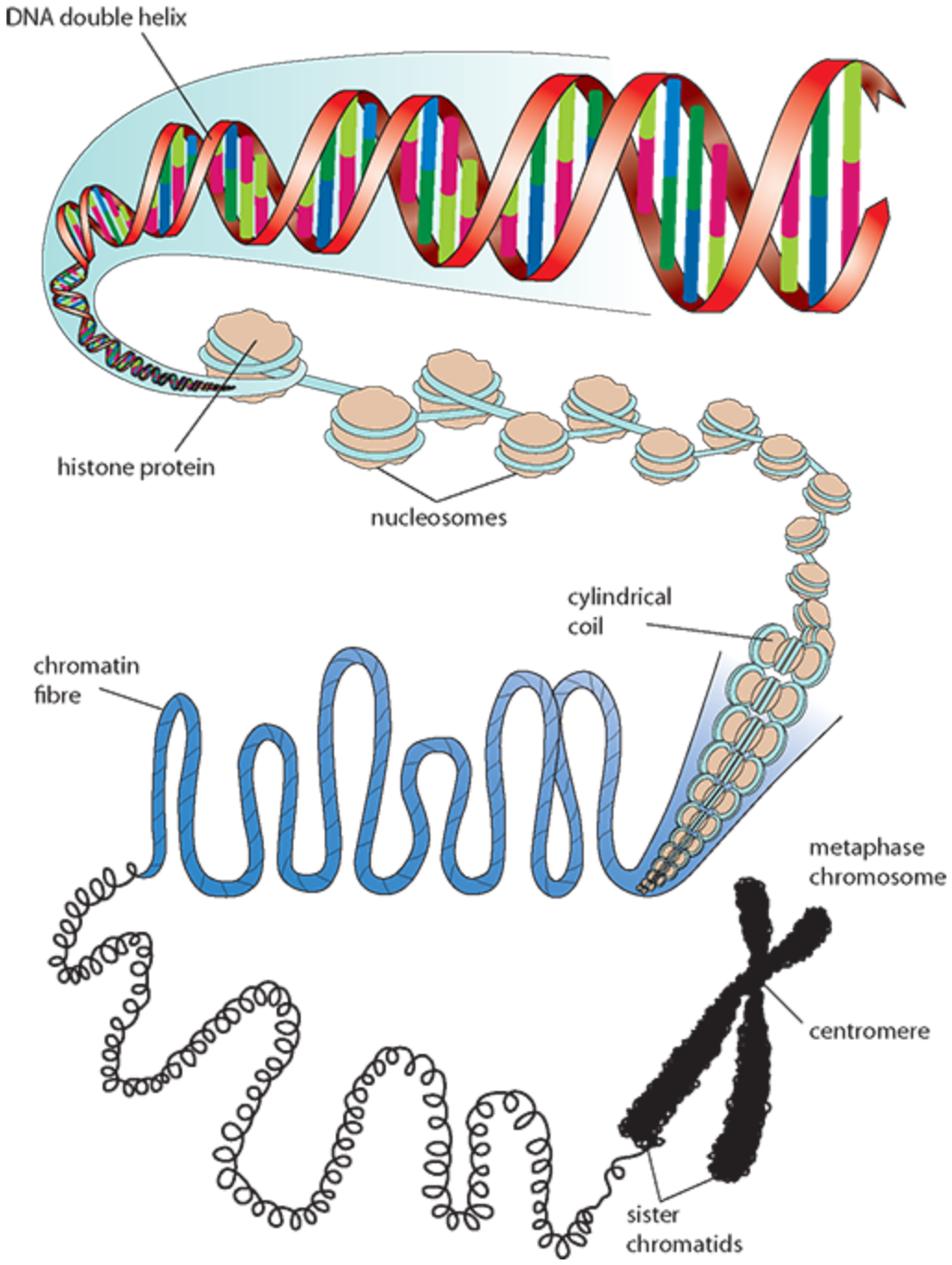


Figure 6.5.5: Supercoiling produces condensed, compact chromosomes in preparation for the next stages of mitosis.

Metaphase

Metaphase begins when the nuclear envelope has broken down. As it disappears, more space is created so that the chromosomes can move into position during their division. The sister chromatids align themselves on the microtubules in the middle, or equator, of the spindle and are attached by their centromeres.

Anaphase

During anaphase, the centromeres split and the sister chromatids are pulled apart. The chromatids move towards the centrioles at opposite sides (poles) of the cell as the spindle fibres shorten. Each sister chromatid is now called a chromosome again.

Telophase

Once the two sets of chromosomes reach opposite poles, the spindle fibres break down and a nuclear envelope forms around each set of chromosomes. At the same time, the chromosomes uncoil and become invisible through a light microscope.

Following telophase, in animal cells, the plasma membrane pinches in and the two new nuclei become separated as the cell enters cytokinesis.

Cytokinesis

During cytokinesis, the two sides of the plasma membrane meet and two completely new cells are formed. Each has a complete set of chromosomes, cytoplasm, organelles and a centriole. In animal cells a ring of contractile actin and myosin proteins pinch the cell membrane together to split the cytoplasm.

In plant cells, the cytoplasm divides in a slightly different way. Firstly, a cell plate forms along the centre of the cell, separating the cytoplasm into two regions. Vesicles accumulate at the edges of the cell plate and release cellulose and pectins, which are

needed to form a new cell wall. Gradually, a cell wall builds up along the cell plate separating the two nuclei and dividing the cytoplasm to form two new cells.

In most cases the division of cytoplasm between two new daughter cells is equal; both cells must receive at least one mitochondrion and other organelles that are made by dividing pre-existing structures. But in a few cases division of the cytoplasm is unequal. Oogenesis (formation of human ova) in humans is asymmetric and produces one large cell which receives all the cytoplasm while smaller polar bodies receive none. This unequal cytokinesis provides the ovum with a much greater amount of stored food than if an equal division were to occur. Another example of unequal cytokinesis occurs in yeast which reproduce by budding new cells from the original parent cell. The daughter cell is smaller and has a longer subsequent cell cycle than the parent cell which produced it.

EXAM TIP

Try to think of an acronym to help you remember the four stages of mitosis: PMAT.

Checkpoints in the cell cycle: cyclins and kinases

The rate and timing of cell division must match an organism's need for growth, development and repair. Cells in the intestine divide twice a day and those in the liver do so once a year. Nerve and brain cells divide very rarely. Understanding the factors that control the cell cycle is important in the study of cancer, which occurs when cell division is disrupted.

Checkpoints regulate the cell cycle so that division only takes place in favourable conditions when the cell is the correct size

and its DNA has been copied properly. There are three checkpoints during the cell cycle. At each one a set of conditions must be met before the cell can proceed to the next stage (Table 6.5.2).

Checkpoint	Conditions to be met
G_1	cell has received signals from other cells cell is large enough cell has sufficient nutrients
G_2	cell is large enough chromosomes have been duplicated
Metaphase	chromosomes are attached to the spindle

Table 6.5.2: Checkpoints in the cell cycle and cytokinesis.

At each stage of the cell cycle proteins called **cyclin-dependent kinases (CDKs)** are involved in control and regulation. CDKs are enzymes which modify various proteins needed for the cell cycle. To become active, CDKs require the presence of another group of proteins called **cyclins**. Cyclins have no enzymatic activity of their own but activate CDKs when they bind to them.

The interaction of CDKs and cyclins forms enzymes that direct cells through the cell cycle and control specific events such as microtubule formation and chromatid alignment. The levels of cyclins and kinases fluctuate at different stages of the cell cycle and allow the cell to move to the next stage of mitosis.

Cyclins are divided into four types based on observations from vertebrate and yeast cells (Figure 6.5.6) but some cyclins have different functions in different types of cell.

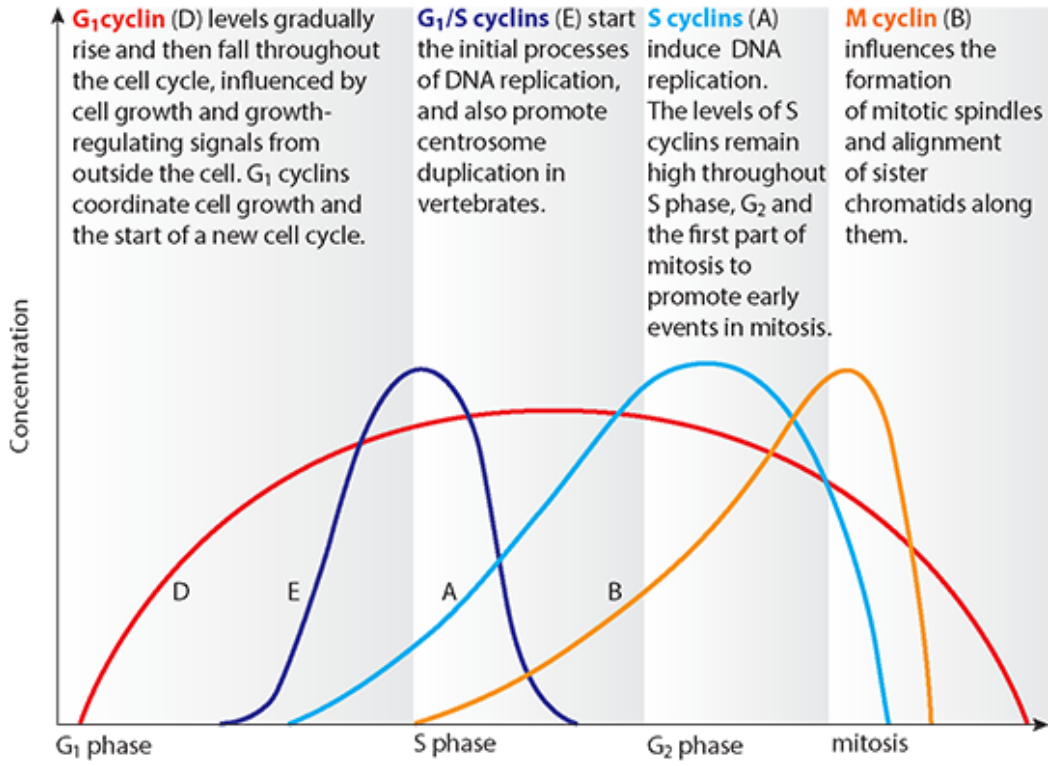


Figure 6.5.6: Cyclins can be divided into four types, which are important at different stages of the cell cycle: G_1 cyclins (D), G_1/S cyclins (E), S cyclins (A), M cyclins (B)

NATURE OF SCIENCE

Serendipity in science: the discovery of cyclins

Serendipity is a term derived from an old name for Sri Lanka. It is said to come from a Persian fairy tale 'The Three Princes of Serendip', about princes who made discoveries by accident. It has come to describe the role of chance in science and indicate how unexpected discoveries are sometimes made. Working scientifically, researchers often benefit from serendipity or 'happy accidents' as new discoveries are made by chance or from apparently unrelated findings.

The discovery of cyclins is one example of a serendipitous discovery. Timothy Hunt and two other scientists, Lee Hartwell and Paul Nurse, were all working on separate areas of the cell cycle and with different organisms. By chance the three strands of their work coincided. Hartwell worked with baker's yeast in the 1970s and discovered 'checkpoint' genes, which seemed to start the cell cycle. In the 1980s Nurse, worked with a different species of yeast. He found a gene that, if it became mutated, stopped the cell cycle or initiated early cell division, and he identified CDK. In 1982 Hunt, who worked with sea urchin eggs, discovered the other key factor that drives the cell cycle, the protein cyclin. Cyclin regulates the function of the CDK molecule and increases and decreases as cell division occurs.

In 2001, Hunt, Hartwell and Nurse were awarded the Nobel Prize in Physiology or Medicine for their work.

- You can read more about their discoveries on the Nobel Prize website by visiting the website (nobelprize.org) and searching for 'cyclins'.

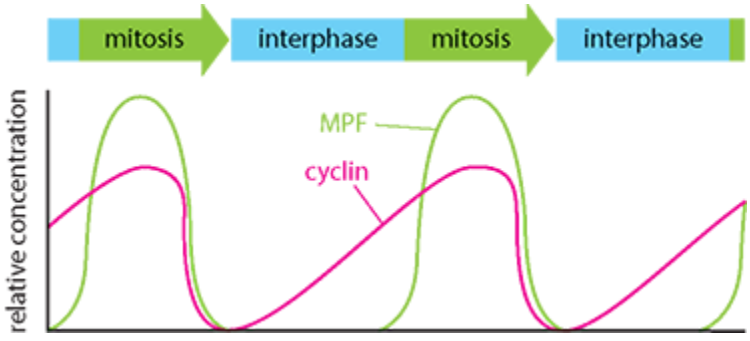


Figure 6.5.7: Levels of MPF rise as a cell enters mitosis, reach a peak and fall during anaphase.

Oncogenes and tumour suppressor genes

Tumour suppressor genes and oncogenes are two important types of gene that have important roles in controlling the cell cycle. Errors in the way they work can lead to the development of tumours and cancer.

Proto-oncogenes are functioning genes that help to regulate normal cell growth by providing signals that initiate cell division or regulate apoptosis. More than 40 different human proto-oncogenes have been identified. Many of them are important in development of the embryo and are switched off once the processes they control are complete. However, if a mutation occurs in these genes they can become permanently overactive, reactivated later in life or switched off. Activation can occur by a mutation or gene amplification. Most cancers are caused when proto-oncogenes are activated and become oncogenes. Activation can happen when chromosomes rearrangements alter the positions of genes in relation to one another, or when gene duplication produces extra copies of a gene and an excess of certain protein is produced.

Many tumours are caused by activated oncogenes. Activated oncogenes can cause cells that should die during apoptosis to survive and divide instead. Most oncogenes become active as a result of some additional process such as mutation in another gene (often those which regulate cell growth or differentiation), direct exposure to a mutagen or another environmental factor such as a viral infection. Because of their importance in human cancers, oncogenes are specifically targeted in many new cancer treatments that are being developed in laboratories all over the world.

KEY POINTS

oncogene a gene that has the potential to cause cancer. In tumour cells, oncogenes can have mutations or be expressed at high levels.

tumour suppressor gene a gene that regulates cell division and replication.

Tumour suppressor genes are functioning genes that slow down cell division, repair DNA or control apoptosis. Without them cells can grow out of control and tumours can develop. The important difference between oncogenes and tumour suppressor genes is that oncogenes cause problems when they are activated but tumour suppressor genes cause cancer when they are inactivated. Some inherited abnormalities in tumour suppressor genes cause certain types of cancer to occur in several family members, but most tumour suppressor gene mutations are acquired during a person's lifetime, not inherited.

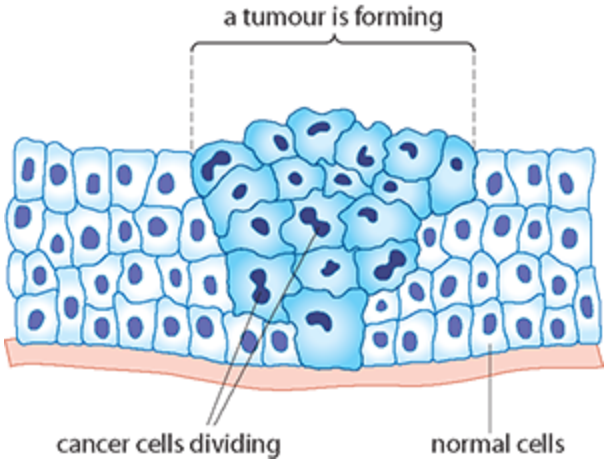


Figure 6.5.8: Formation of a primary tumour. If cells from a primary tumour become detached and form a new tumour in another part of the body, then the cells are said to be cancerous.

Cancer occurs when cells from a **primary tumour** (Figure 6.5.8) migrate to other tissues and form new **secondary tumours** in a process known as **metastasis**. Cancer is caused by damage to genetic material, producing cells that undergo uncontrolled, abnormal mitosis, but it cannot be thought of as a single disease. Cancer can take different forms in different tissues and the DNA damage that leads to cancer can be caused by a range of factors.

Apoptosis

Apoptosis is an orderly, controlled process. It is quite different from cell death caused by damage or injury. Apoptosis has two key functions: it enables cells to be removed during development (Figure 6.5.9) as structures form and grow, and it also removes cancerous cells or cells which are infected with viruses. Both cancerous cells and cells that are infected with viruses can be harmful to the rest of the organism so it is important that they are eliminated.

Figure 6.5.9: Cells between forming fingers are removed by apoptosis as the human hand develops.

If a cell is programmed to die, the contents of the plasma membrane will be packaged into their own membranes and removed by phagocytic cells of the immune system. During apoptosis a cell will shrink and develop 'blebs', which are bubble-like extensions on its surface. DNA is broken up and organelles are fragmented and packaged within a membrane. Signals on the outer plasma membrane attract phagocytic macrophages (Chapter 10) which engulf the cell fragments.

If a cell's DNA is damaged, the cell will try to repair the damage using DNA polymerases. If this is not possible, the cell will cause their own death so that damaged DNA is not passed on. Damaged cells which fail to die may become cancerous. Pre-cancerous cells that do not cause their own death are often noticed by the cells of the immune system, which recognise external markers on their plasma membranes. If the cells are not removed they will be able to divide uncontrollably and form tumours.

Mitotic Index

In a population of cells, the ratio of the number of cells undergoing mitosis to the number of cells not undergoing mitosis is known as the **mitotic index**. A higher than normal mitotic index is an indication that cells are dividing more rapidly than usual and can be an indicator of cancerous cells. You should be able to work out mitotic indices from photographs of dividing cells.

TEST YOUR UNDERSTANDING

25 State two reasons why a cell may need to divide.

26 During which stage of the cell cycle does chromosome replication take place?

27 Outline the events of cytokinesis in an animal and a plant cell.

6.5.3 Meiosis

Meiosis is a type of cell division that produces **gametes** (sex cells). Meiosis is called a reduction division because in any organism, each cell that is produced as a result of meiosis has half the number of chromosomes of other cells in the body.

Eukaryotic body cells have a **diploid** nucleus ($2n$), which contains two copies of each chromosome, in **homologous** pairs (Section 4.1). Humans have a diploid number of 46 chromosomes in 23 pairs, whereas mangos and soybean both have 40 chromosomes in 20 pairs, and camels have 70 chromosomes in 35 pairs.

During sexual reproduction, two gametes fuse together so, in order to keep the chromosome number correct in the offspring that are produced, each gamete must contain only one of each chromosome pair. A gamete must contain half the diploid number of chromosomes, which is a number called the **haploid** (n) number. During gamete formation, meiosis reduces the diploid number to the haploid number and for this reason, meiosis is called a reduction division. At the moment of fertilisation, the normal diploid number is restored as two gametes fuse. For example, in the camel, the haploid sperm (35 chromosomes) and haploid egg (35 chromosomes) fuse at fertilisation to form the diploid zygote, with 70 chromosomes.

The process of meiosis

Meiosis occurs in a series of stages, as illustrated in Figure 6.5.10, which result in the production of four cells. Mitosis is achieved with one cell division but meiosis involves two divisions: the first reduces the number of chromosomes by half

and the second produces four gametes each containing the haploid number of chromosomes. Exactly the same terms are used for the names of the stages, but since meiosis involves two divisions, the phases are numbered I and II.

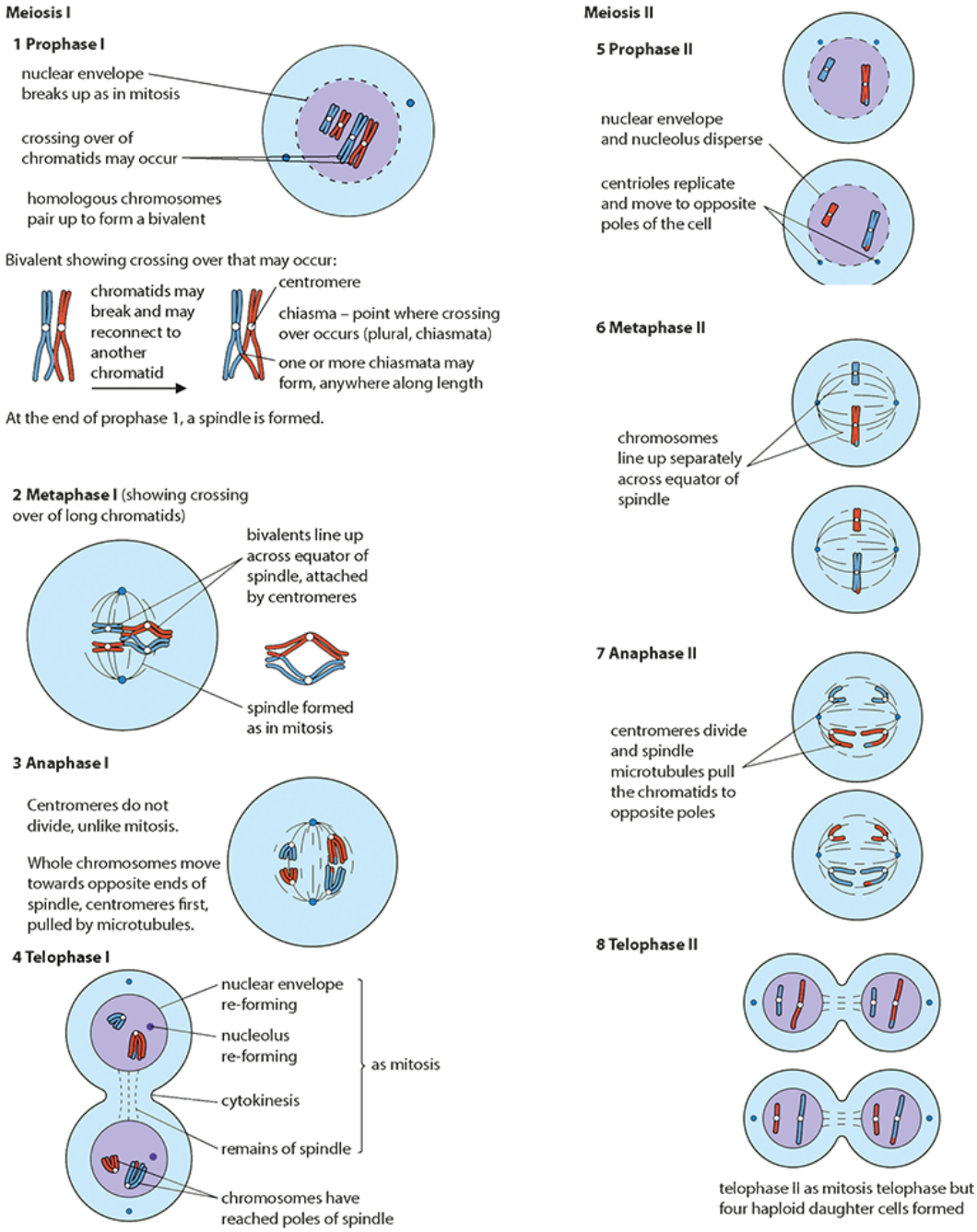


Figure 6.5.10: The stages of meiosis in an animal cell. Note that the cells are shown with just two homologous pairs of chromosomes to make it easier to understand the process.

The first division is very similar to mitosis and the second division is exactly the same as mitosis.

Prophase I

During interphase, before the start of prophase, chromosomes are replicated and consist of two identical sister chromatids joined at the centromere. In prophase I these chromosomes supercoil and the homologous pairs of chromosomes line up side by side. These are called **bivalents**. A bivalent has two chromosomes and four **chromatids**, with one chromosome coming from each parent.

Although the genes carried by each chromosome pair are identical, the alleles may not be. Exchange of genetic material between the pair can occur at this point. Chromatids may become entangled, break and re-join so that alleles are exchanged between homologous chromosomes during a process called **crossing over**. New combinations of alleles are formed and genetic variety among the resulting gametes increases.

The final step in prophase I is the formation of spindle microtubules and the breakdown of the nuclear envelope.

Metaphase I

Chromosomes line up on the equator at the centre of the cell. Each one attaches by a centromere to the spindle microtubules. The alignment of the chromosomes is random so that maternal and paternal chromosomes can appear on either side of one another on the equator. This means that either chromosome from

a pair may move into each daughter cell during the first division at anaphase I, which results in increased genetic variety among the gametes.

Anaphase I

During anaphase I the microtubules contract towards opposite poles. The pairs of sister chromatids remain together but the homologous pairs are separated. This is the **reduction division** where the chromosome number is halved from diploid to haploid.

Telophase I

During telophase I spindles break down and a new nuclear envelope forms around each new nucleus. Cytokinesis follows and the cell splits into two cells, each containing only one chromosome of each homologous pair. Each chromosome, however, still consists of two sister chromatids at this point.

The second division of meiosis now follows to separate the two sister chromatids.

Prophase II

In each of the two cells resulting from meiosis I, new spindle microtubules start to form, the chromosomes re-coil and the nuclear envelope begins to break down.

Metaphase II

The nuclear envelope is broken down and individual chromosomes line up on the equator of each cell. Spindle fibres from opposite ends of the cell attach to each chromatid at the centromere.

Anaphase II

Sister chromatids are separated as the centromere splits and spindle fibres pull the chromatids to opposite ends of the cell.

Telophase II

Nuclear envelopes form around the four new haploid nuclei and the chromosomes now uncoil. A second cytokinesis occurs, resulting in four cells.

Meiosis and variation: crossing over and random orientation of chromosomes

Meiosis not only halves the chromosome number, it also promotes genetic variation in the gametes and individuals that are produced from them. There are several reasons for this:

1 Each chromosome of the homologous pair carries different genetic information so that the gametes formed are genetically different.

2 Random orientation: different homologous pairs arrange themselves independently on the spindle and also separate independently so that gametes contain different combinations of each chromosome pair (Figure 6.5.11).

3 Crossing over during prophase I (Figure 6.5.12) means that genetic material is exchanged between homologous chromosomes. This produces entirely new combinations. In each homologous pair, one chromosome is a maternal chromosome and the other a paternal chromosome. Homologous chromosomes contain the same genes, but since they came from different parents, they can have different alleles. As they line up together, the non-sister chromatids may touch and break. The two segments may

then re-join at the corresponding position on the other chromatid. In this way, chromatids are formed that are a mixture of paternal and maternal alleles. After crossing over, the chromatids recombine to produce new and unique combinations of alleles, different from both the maternal and the paternal arrangements. This is called recombination. The region where this happens is called a chiasma (plural: chiasmata) (Figure 6.5.12).

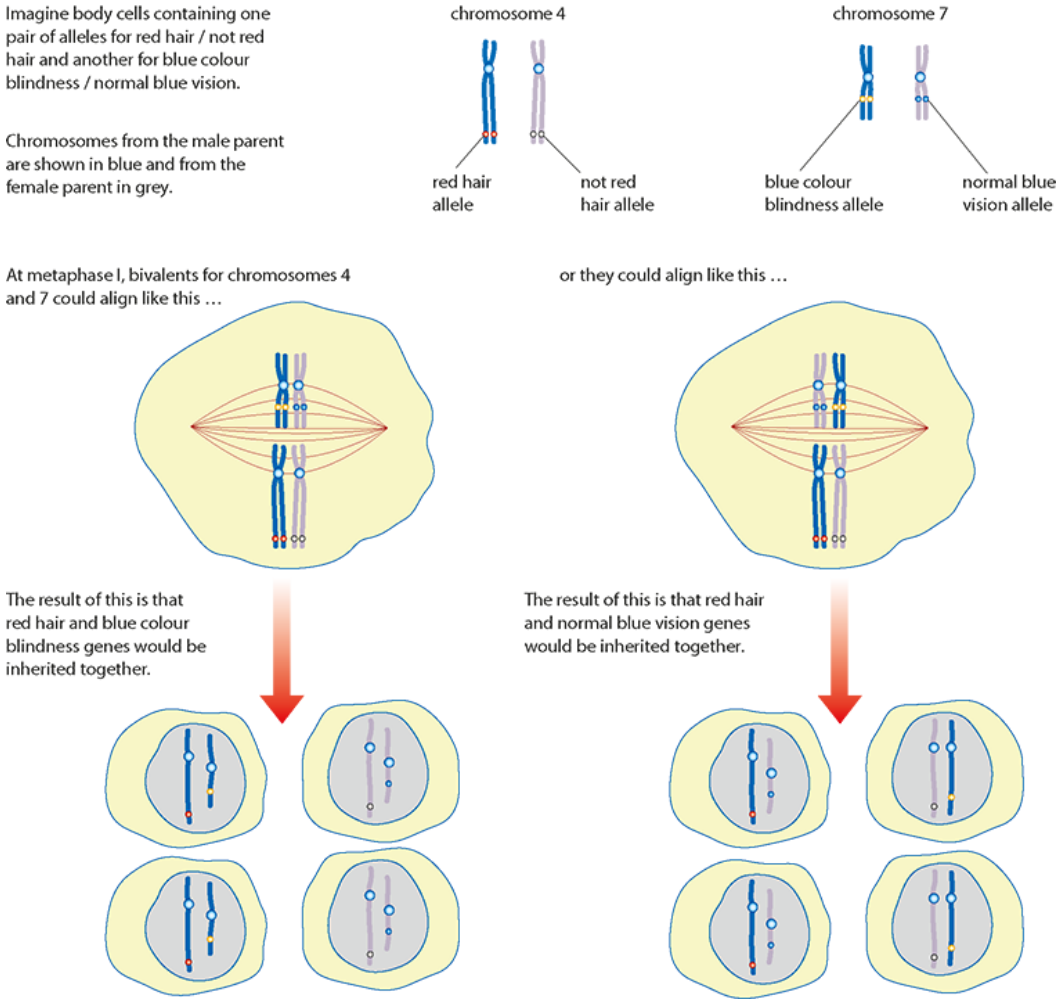


Figure 6.5.11: How random orientation produces variation.

Crossing over does not occur between the X and Y chromosomes. This is because the two chromosomes are

very different sizes and therefore do not sit alongside each other for their full length in the same way as other homologous pairs. The fact that there is no crossing over between the sex chromosomes is advantageous because it means that genes that determine sex remain on the appropriate chromosome.

4 At fertilisation, gametes from different parents fuse together. This promotes yet more genetic variation among the offspring produced.

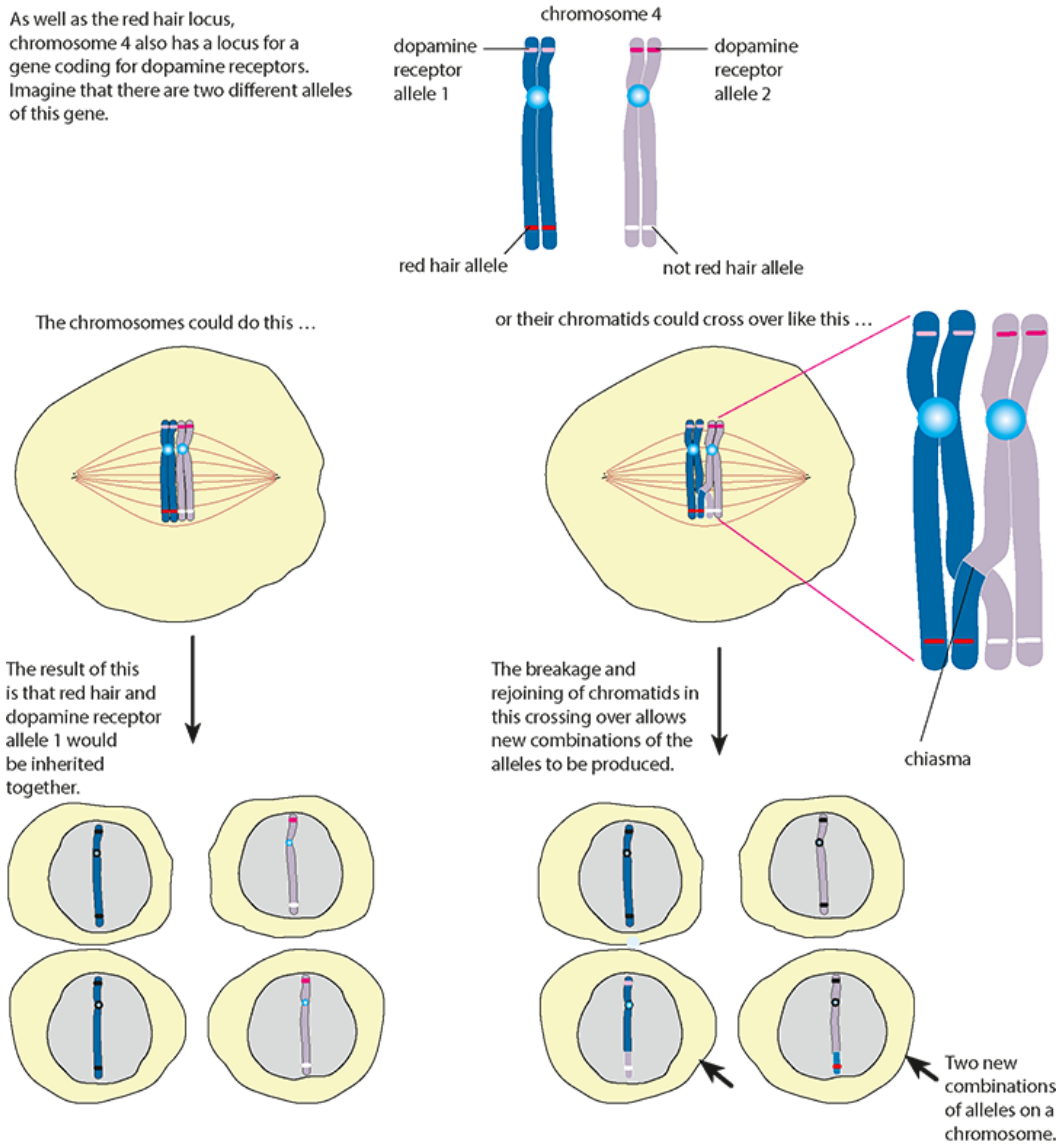


Figure 6.5.12: How crossing over produces variation. If a single cross-over occurs in a pair of chromosomes, four different daughter chromatids are produced instead of two.

TEST YOUR UNDERSTANDING

28 Why is meiosis often called a reduction division?

29 State the number of cells formed when meiosis in one parent cell is complete.

30 Describe what is meant by 'random orientation' of chromosomes during meiosis.

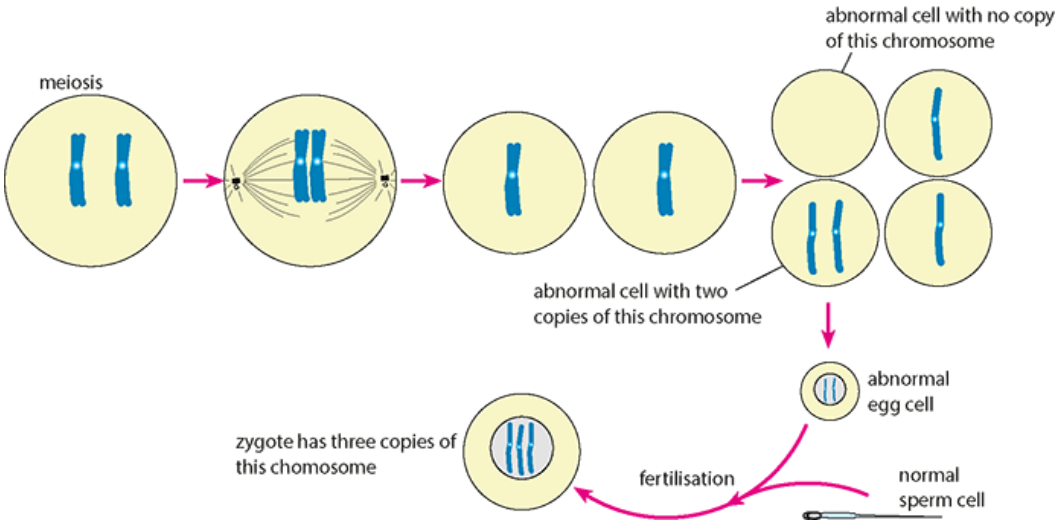


Figure 6.5.13: Non-disjunction at anaphase II of meiosis. Non-disjunction can also occur at anaphase I.

6.5.4 Non-disjunction

Non-disjunction is a failure of homologous pairs of chromosomes to separate properly during meiosis. It results in gametes that contain either one too few or one too many chromosomes. Those with too few rarely survive, but in some cases a gamete with an extra chromosome does survive and after fertilisation produces a zygote with three chromosomes of one type, as shown in Figure 6.5.13. This is called a trisomy.

Trisomy in chromosome 21 results in the human condition known as Down syndrome (Figure 4.1.3). A gamete, usually the female one, receives 24 chromosomes instead of 23 and a baby with 47 instead of the usual 46 chromosomes in each cell is born.

Karyotyping is used when there is concern about potential chromosome abnormalities. Cells from an unborn child are collected in one of two ways: chorionic villus sampling (CVS) or amniocentesis. The cells are grown in the laboratory and a karyogram is prepared. This is checked for extra or missing chromosomes (Section 4.1).

6.5.5 Chromosome behaviour and Mendel's laws

Mendel investigated inheritance in pea plants (Section 4.2) and when crossing purple and white plants he discovered that offspring were purple, indicating that one colour was dominant over the other. Mendel had no knowledge of chromosomes and genes, but nevertheless stated that each individual has two factors (alleles) for each characteristic, one from each parent.

His first law the law of segregation, states that individuals possess two alleles and that a parent passes only one allele to their offspring. One allele is given by each parent (Figure 6.5.14).

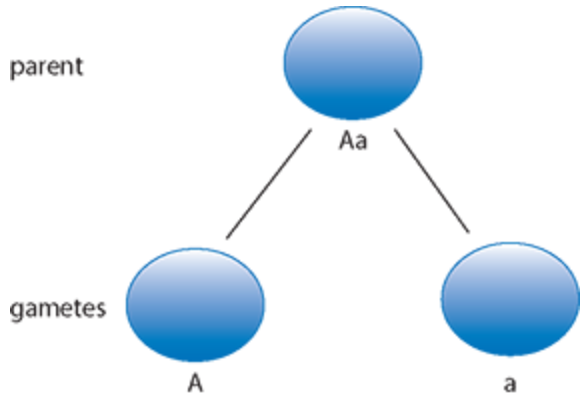


Figure 6.5.14: During meiosis one allele from each pair is passed from parent to gamete.

Mendel's second law – the law of independent assortment – states that during gamete formation the segregation of the alleles of one gene is independent of the segregation of the alleles of other pairs. When we examine the stages of meiosis and look closely at the alignment of chromosomes at metaphase I we can understand how this segregation and independent assortment of

alleles takes place. Chromosomes can line up in different combinations and produce different combinations in the resulting gametes.

KEY POINTS

independent assortment alleles of different genes are sorted into gametes independently of one another.

segregation separation of pairs of alleles at meiosis and their independent transmission in separate gametes to offspring.

Two possibilities are shown for two pairs of chromosomes in Figure 6.5.11.

During metaphase I, the bivalents line up on the equator and spindle microtubules become attached to their centromeres. However, the way in which they line up is random. This is shown in Figure 6.5.11, which illustrates the possibilities for just two chromosomes, 4 and 7.

- The paternal chromosomes could both line up together on one side of the equator with the maternal ones on the other side, as shown on the left. Two of the gametes that are produced, after the sister chromatids separate in meiosis II, they contain just paternal chromosomes while the other two contain just maternal chromosomes.
- Another possibility is that the chromosomes line up as shown on the right, with maternal and paternal chromosomes on both sides of the equator. The end result here is that all four gametes contain a mixture of paternal and maternal chromosomes.

Independent assortment of alleles at metaphase I of meiosis results in the production of different combinations in gametes.

One gene does not influence any other gene, because alleles are sorted of into gametes: every possible combination of alleles for every gene is equally likely to occur. Independent assortment of genes is important to produce new genetic combinations that increase genetic variation in a population.

TEST YOUR UNDERSTANDING

31 Outline the result of non-disjunction of chromosomes.

REFLECTION

Reflect upon the level of difficulty of the ideas in this section. Is it one of the topics you find easier to tackle?

Links

- How does variation produced from sexual reproduction contribute to evolution? (Chapter 11)
- Why is cell division important for multicellular organisms? (Chapter 8)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
draw and describe the structure of a membrane including phospholipid bilayer, membrane proteins and cholesterol and sterol	6.1.1			
outline why phospholipids form a bilayer and the importance of the fluid mosaic arrangement to cells	6.1.1			
define the term organelles	6.1.2			
explain why compartments that store, isolate and concentrate materials are important to cells	6.1.2			

state that prokaryotes do not contain organelles	6.1.2			
outline the functions of RER, Golgi apparatus, coated vesicles and lysosomes	6.1.3			
describe the importance of compartmentalisation in chloroplasts and mitochondria and how membranes form their compartments	6.1.3			
annotate diagrams of mitochondria and chloroplasts to link their structures and functions	6.1.3			
define: semi-permeable, diffusion, facilitated diffusion osmosis and active transport	6.2.1			
outline how gases move in and out of cells	6.2.1			
outline how glucose	6.2.1			

is moved into cells by facilitated diffusion				
state that membrane-bound proteins act as pores or pumps for transport of molecules	6.2.1 and 6.2.2			
understand that active transport requires energy in the form of ATP	6.2.2			
explain how exocytosis and endocytosis result from the fluid nature of the plasma membrane	6.2.2			
describe the structure of a neuron and its adaptations to carry impulses	6.2.3			
outline the importance of gated ion channels in the transmission of a nerve impulse	6.2.3			
outline how the sodium–potassium pump permits active	6.2.3			

transport in both directions				
summarise how myelinated nerve fibres allow for salutatory conduction at the nodes of Ranvier	6.2.3			
state the factors that determine whether water will move into or out of a cell	6.3.1			
define hypotonic, isotonic and hypertonic	6.3.1			
explain crenation and plasmolysis	6.3.1			
define turgor in plants	6.3.1			
outline an experiment to estimate the water potential of a plant cell	6.3.1			
outline how water moves from a higher to lower water potential	6.3.2			
recognise the contributions of	6.3.2			

solute and pressure potential to the water potential of plant cells				
explain the changes that occur when plant cells are placed in hypotonic and hypertonic solutions in terms of solute and pressure potentials	6.3.2			
state that cells interact with their environments to obtain substrates for their metabolism	6.4.1			
recall that cells require a maximum surface area to volume ratio to ensure substances can move across the membrane at optimum rate	6.4.1			
understand that larger cells have a smaller surface area to volume ratio	6.4.1			
outline how cells increase their surface area by becoming	6.4.1			

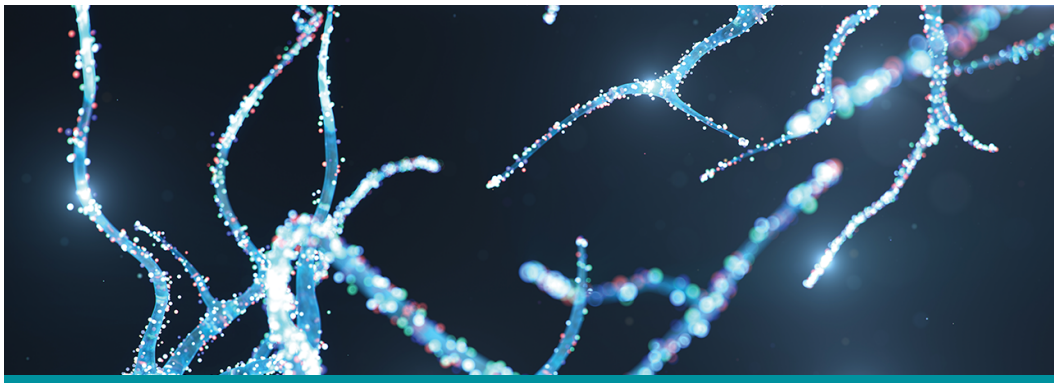
flattened or elongated, while some have villi and microvilli				
state that a cell conserves its volume but increases its surface area by dividing in two	6.4.1			
recall that the size of cells in a tissue depends on their rate of growth and rate of division	6.4.2			
define binary fission	6.5.1			
describe the stages of the cell cycle	6.5.2			
name the phases of mitosis and describe the events that occur during each	6.5.2			
understand the phases of the cell cycle and that cylins are responsible for controlling stages of the cycle	6.5.2			
recall that mutations in oncogenes and tumour suppressor	6.5.2			

genes can lead to uncontrolled division				
explain the difference between benign, malignant, primary and secondary tumours	6.5.2			
determine the mitotic index of dividing cells	6.5.2			
describe the events of meiosis and outline the importance of meiosis to sexual reproduction	6.5.3			
summarise crossing over and random orientation and how each increases genetic variation	6.5.3			
define non-disjunction and describe the consequence of a trisomy in chromosome 21 in humans	6.5.4			
outline how cell proliferation is needed for growth,	6.5.5			

replacement and repair in plant meristems, embryos and in wound healing.				

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.



Chapter 7

Cell control and communication

C2.1, C2.2

INTRODUCTION

Communication between cells is essential in both single-celled and multicellular organisms. It allows organisms to carry out more complex activities and coordinate the functions of cells, tissues and organs. Cells use different methods of communication to send, receive and process electrical and chemical messages.

7.1 Principles of cell signalling

LEARNING OBJECTIVES

In this section you will:

- learn that cell signalling is essential in single-celled and multicellular organisms
- learn that cells have protein receptors with binding sites for specific signalling chemicals (ligands)
- discover that cell signalling allows complex emergent properties to occur because cells, tissues and organs can work together
- recognise quorum sensing as an example of signalling in bacteria
- understand that a range of substances including ions, neurotransmitters, hormones and cytokines move inside and between cells to send and receive electrical and chemical messages

- discover that some molecules operate in a localised way and others have distant effects
- recognise that there are differences between transmembrane receptors and intra cellular receptors
- understand that as signalling chemicals bind to receptors a sequence of responses within cells takes place

- understand the modes of action of acetylcholine, G proteins, epinephrine receptors and receptors with kinase activity
- understand how steroid hormones affect gene expression
- learn how cell signalling pathways are regulated by positive and negative feedback

GUIDING QUESTIONS

- How does cell signalling allow cells to interact effectively?
- How does the nervous system of animals rely on electrical and chemical signals?

KEY POINTS

receptors proteins with binding sites for specific proteins

hormone a chemical substance produced by an endocrine gland, which is transported in the blood and which affects the physiology or biochemistry of specific target cells.

ligand a signalling chemical which interacts with receptors in or on target cells

neurotransmitter a substance produced and released by a neuron, which passes across a synapse and affects a post-synaptic membrane.

cell surface receptors are found in the cell membrane

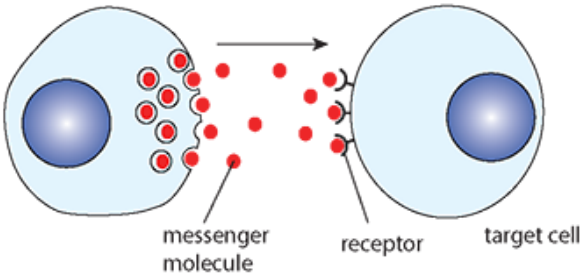
intracellular receptors are protein receptors found inside cells

7.1.1 Principles of cell signalling and cell interaction

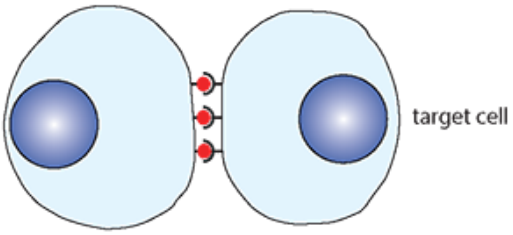
Cells signalling allows cells to communicate and is important in both single-celled and multicellular organisms.

Most cells communicate with each other using chemicals; proteins or other molecules, such as ions, that leave one cell, enter the extracellular space and pass to other cells that can receive the signals they carry. From the extracellular space signals interact with other different cells. Not all cells will be able to receive the signal that has been sent. Only cells with the correct receptors on their surfaces will be able to accept the molecules and respond to them. Once the signal message has bound to a receptor on the surface of another cell, it will trigger a change in the cell. For example, it may change the activity of a particular gene, or stimulate the cell to divide.

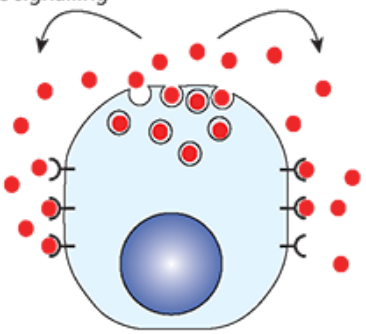
1. Paracrine signalling



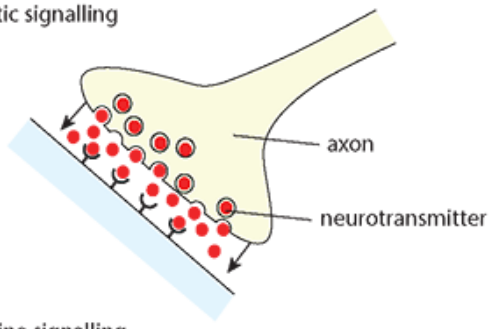
2. Contact-dependent signalling



3. Autocrine signalling



4. Synaptic signalling



5. Endocrine signalling

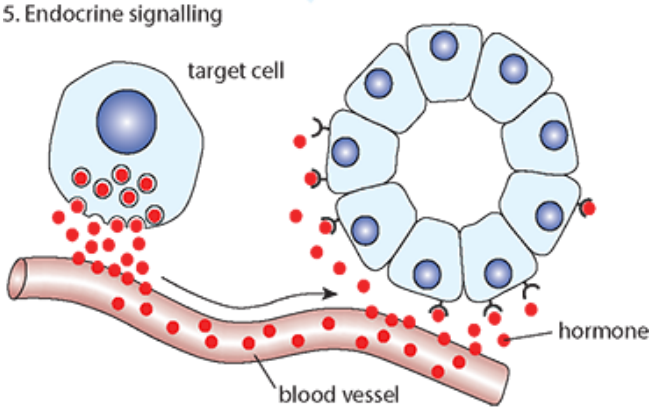


Figure 7.1.1: Cells can pass signals to other cells in a number of different ways.

There are several different forms of cell signalling (Figure 7.1.1):

1 Paracrine signalling allows cell that are close to one another to communicate over short distances to coordinate activities. This type of signalling is important during development and differentiation. Cells can tell nearby cells what type of cell to become.

2 Contact-dependent cell to cell signalling allows small molecules and ions to pass directly to adjacent cells that touch each other. Plants cells have microscopic channels, called plasmodesmata, through their cell walls and these allow communication and transport between every cell and its neighbours.

3 Autocrine signalling is a signal sent by a cell to itself. This is important in development when cells are differentiating into a specific cell type. It has also been implicated in the spread of cancer cells from one part of the body to another.

4 Synaptic signalling is used by neurons when an impulse triggers the release of a neurotransmitter at the end of a nerve fibre.

5 Endocrine signalling sends signals as hormones to target cells that may be a long way from the cell that produces them.

7.1.2 Cell signalling in unicellular organisms

Many types of bacteria use a type of signalling called quorum sensing. Bacteria regulate their activities and physiological processes by releasing, sensing and responding to small diffusible signal molecules. The signals allow them to assess the density of the local population and if the signals reach a certain level all the bacteria in the population will change their behaviour or gene expression at the same time. This increases their chances of survival.

Bioluminescence in the marine bacterium *Vibrio fischeri* is a good example of quorum sensing which the bacteria use to communicate with each other. *Vibrio fischeri* is a symbiotic bacterium found in the light organ of the bobtail squid (*Euprymna scolopes*). The bacteria only produce light (or bioluminescence) when they have multiplied to a high cell density. *V. fischeri* use proteins coded for by a set of genes known as the lux operon to produce light. The mechanism is controlled by the excretion of an inducer which interacts with a regulator and activates transcription of the lux operon. When cell density is high, all the bacteria bioluminesce together.

SCIENCE IN CONTEXT

Quorum sensing, biofilms and human health

Pseudomonas aeruginosa is a human pathogen; it is bacterium that causes infections including pneumonia and urinary tract infections. Individual bacteria use quorum sensing to communicate with each other. Together they are able to form closely bonded groups called biofilms. As part of

a biofilm, bacteria stick to one another and to the underlying surface and keep in contact with other bacteria using quorum sensing. As part of a biofilm, bacteria act together so they can resist the body's immune system and antibiotics. If the bacteria accumulate in large numbers, for example at the site of a wound, they communicate and can become more mobile and more virulent.

Another species of bacteria that can form biofilms is *Staphylococcus aureus* (Figure 7.1.2). These bacterial biofilms can colonise equipment such as catheters in hospitals and cause serious infections.

Biofilms can contain more than one species of bacteria that live together. Quorum sensing plays an important role in the formation, maintenance and breakdown of biofilms like these.

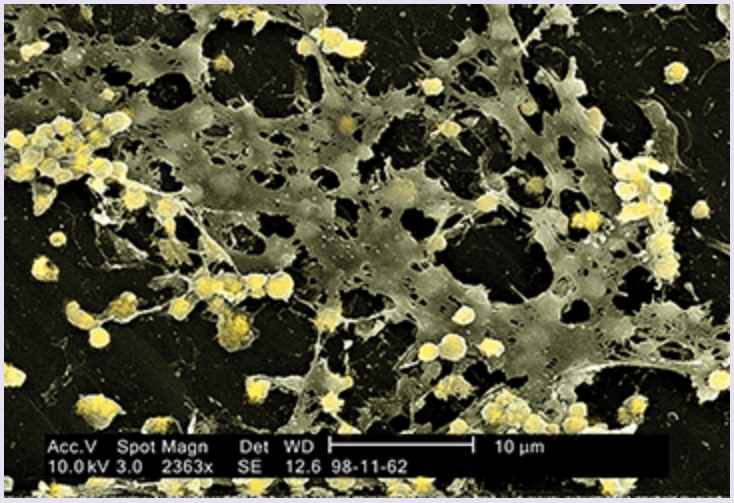


Figure 7.1.2: S. aureus biofilm.

7.1.3 Cell signalling in multicellular organisms

Cell signalling is vital to multicellular organisms. It means that they can coordinate the activities of their cells, tissues and organs so that the body works efficiently. Movement, growth, reproduction and homeostasis all happen because of cell signalling as one cell or group of cells influences the activities of others.

Emergent properties

One person playing the flute can produce a simple, recognisable tune but if several musicians with other instruments join in and play together as a group, they produce a wide variety of sounds and many different effects. New properties emerge in the cells of multicellular organisms in a similar way. Their cellular components interact so that the organism can carry out a range of more complicated functions. One cell can function on its own, but with other cells in a group, it can produce tissues and organs that carry out a range of roles in the organism. For example, lungs are made of many cells it is only when all these cells work as a unit that the lungs are able to perform their function. Cells form tissues, tissues form organs, organs form organ systems and organ systems work in synergy so that the whole organism can carry out a complex range of tasks and is greater than the composition of its parts.

Methods of communication

Cells use different substances including ions, neurotransmitters and hormones to communicate with one another (Table 7.1.1). These molecules move down concentration gradients either

within a cell or between one cell and another (Figure 7.1.1). Once a cell has received it, a messenger molecule is processed and the appropriate response can take place. Hormones are a diverse group of molecules that act as chemical messengers for cell signalling. Animal hormones circulate in the blood to reach specific target cells. Ions such as sodium and potassium are important in the transmission of nerve impulses and move in and out of nerve cells. Most animals have nervous systems consisting of nerve cells that transmit electrical impulses when ions cross their membranes. Neurotransmitter molecules enable nerve cells to communicate with other nearby nerve cells. You can read more about the transmission of nerve impulses in Section 7.2 and more about the action of hormones in Section 7.3.

Substance	**Example**	**Method of movement**	**Action in communication**
neurotransmitter	acetylcholine	diffuses between adjacent cells	enables synaptic signalling between different nerve cells
ions	sodium and potassium ions	diffuse down concentration gradients or are pumped into cells	maintains the action potential in nerve cells so that electrical messages can be transmitted
hormone	insulin, glucagon	carried in the bloodstream from the site	enables endocrine signalling

		of production to target cells	between cells that are not close together
cytokines	chemokines, interferon	small peptides that cannot cross the membrane and act through cell surface receptors	regulate humoral and cell-based immune responses, and maturation and growth of some cells

Table 7.1.1: Different molecules enable cells to communicate using electrical and chemical signalling.

THEORY OF KNOWLEDGE

The systems approach

A system is defined as an assemblage of parts and the relationships between them. This makeup enables a system to work together as a functioning whole. The systems approach has long been used in engineering but for many years natural systems were examined from a reductionist point of view. We can see how the two approaches differ if we consider the study of a pond. A reductionist study of the pond would describe the organisms found there in terms of their features and characteristics; for example, whether they are vertebrates or invertebrates, plant or animal. But a reductionist study would not try to consider how the pond worked as a dynamic system.

A systems approach would take a holistic view of the pond that considers interrelationships such as food chains and nutrient cycling that occur between the various components of the pond. In this way a picture of the interdependence of the different parts of the pond – that is, the system's structure – could be built up.

In a study of cells and their components, the systems approach would consider a single cell in terms of the flows of energy and materials between the various structures within in it. On a larger scale, groups of cells, an organ or even a whole organism can be studied using the systems approach so that the parts and the interactions between them can be viewed as a complete functioning entity. Emergent properties in any system can only be studied by means of a systems approach.

To consider:

1 What are the advantages and disadvantages of the systems approach compared with the reductionist approach to the study of cells and cell signalling?

2 In science, the reductionist and systems approach may use similar methods of study. What is the most important difference between the philosophies of the two approaches?

Signal transduction pathways

As we have seen, cells respond to signals from their environment through a process known as cell signalling. A signal is transmitted when a signalling molecule called a ligand binds to a receptor protein in the target cell.

This signal leads to a specific cellular response such as cell division, release or absorption of a substance or even cell death.

As signalling chemicals (ligands), bind to receptor sites, there are several possible outcomes which depend on the type of receptor and type of signalling molecule that is involved. The receptors involved in cell signalling can be broadly classified into two types: **cell-surface receptors** and **intracellular receptors**.

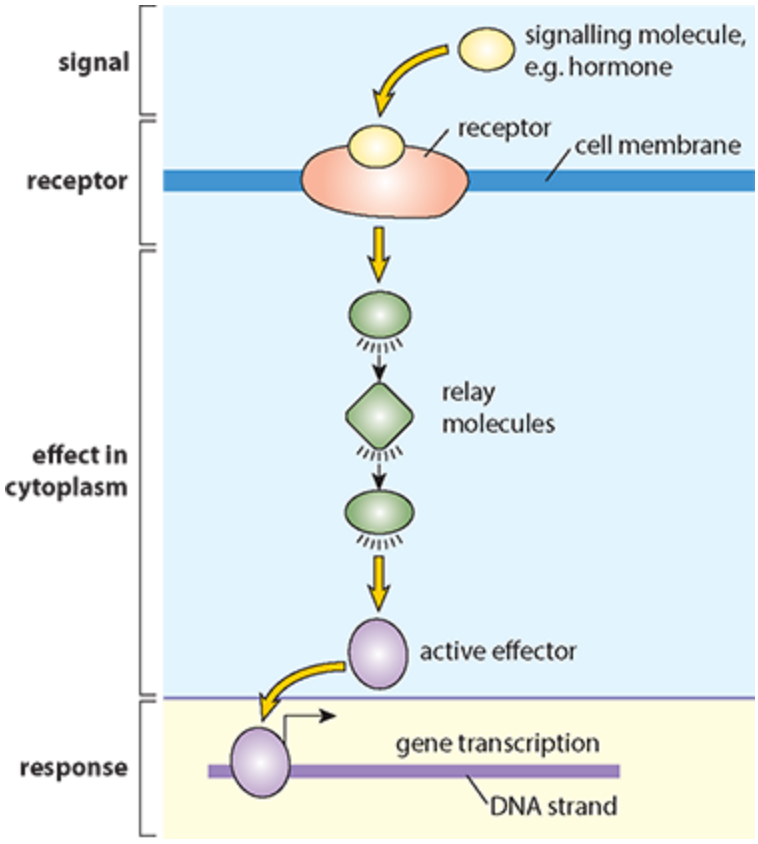


Figure 7.1.3: Some signalling molecules do not enter cells but stimulate reactions inside them.

Cell-surface receptors (transmembrane receptors) are integral to the plasma membrane with each receptor having an extracellular, transmembrane region and a part of their molecule in the

cytoplasm. When a ligand binds to a cell-surface receptor, extracellular signals are transformed into intracellular signals in a process called **signal transduction**.

Intracellular receptors (also known as internal receptors) are globular proteins that are found inside the cell not in the cell membrane.

If receptors are located on the cell membrane, the signal must be passed on through chains of molecules that make up **intracellular signal transduction pathways**.

As a signalling molecule binds to a cell-surface receptor, the receptor changes shape. This may let it bond to other molecules or give it enzyme activity. Changes in the receptor set off a series of signals inside the cell. For example, the receptor may turn on another signalling molecule in the cell which can then activate its own target. The chain of reactions can lead to changes in cell activity or characteristics (Figures 7.1.3, 7.3.5 and 6).

Transmembrane protein receptors and their effects include:

- **G protein coupled receptors**. These are integral membrane proteins that are used by cells to convert extracellular signals into intracellular responses, including responses to some hormones, neurotransmitters, as well as responses to vision, olfaction and taste signals. One example is the receptor for the neurotransmitter acetylcholine, a membrane bound protein that changes shape to regulate the opening of the ion channels that allow positively charged ions to diffuse into nerve cells (Section 7.2)

- **G protein and cyclic AMP**. cAMP (cyclic adenosine monophosphate) is a second messenger inside cells and this system is co-ordinates the body's response to the hormone epinephrine (adrenaline). cAMP is a small molecule made from ATP. In response to signals, an enzyme converts ATP into cAMP which can then activate an enzyme called protein kinase. Kinase goes on to phosphorylate target molecules and pass along the signal. Protein kinase is found in a variety of types of cells, and it has different target proteins in each. This allows cAMP second messenger to produce different responses in different cells.

- **Receptors with tyrosine kinase activity**. These receptors mediate the cells' responses to the protein hormone insulin, as well as many other hormones. As insulin binds to a receptor, tyrosine inside the cell is phosphorylated. Phosphate groups are linked to this amino acid which has a hydroxyl (-OH) group in its side chains. The transfer of the phosphate group is catalysed by an enzyme called a kinase. Phosphorylation switches on a series of reactions which cause vesicles containing glucose transporters (carriers) to move to the plasma membrane.

Intracellular receptors and their effects

Intracellular receptors are activated by signals (ligands) that can pass through the plasma membrane. Ligands are usually small, hydrophobic and non polar. The group includes receptors for steroid hormones such as oestradiol, which stimulates the hypothalamus to secrete GNrh, progesterone that targets cells of the endometrium (Ch 8), as well as testosterone and fat-soluble vitamins such as vitamin D. Steroid hormone receptors are structurally similar and work in similar ways. Unlike cell-surface receptors, ligands that bind to intracellular receptors can

easily diffuse across the plasma membrane so they do not need to transmit the signal to other receptors or messengers. They directly activate the intracellular receptor and as the receptors are stimulated, they bind to DNA and affect the transcription of certain genes.

Intracellular receptors share similar features because they work in similar ways. Most have three areas on their molecules: a ligand-binding area, a DNA-binding area, and a transcription-activating area.

Figure 7.1.3 shows how a peptide hormone affects the actions of its target cell in animals. This type of hormone does not enter cells but causes a response. It stimulates receptor molecules on the membrane that trigger a response inside the cell. Other steroid hormones do enter cells to bring about a response. You can read more about hormones in Section 7.3.

Regulating signalling pathways

Cell signalling pathways may be regulated either by a positive feedback mechanism or by negative feedback. Positive feedback occurs to increase the change that takes place so that it happens more quickly. Negative feedback occurs to reduce the change, so the result of a reaction is reduced and the system returns to a stable state.

As you read about the hormone insulin (Section 7.3), notice that it works by negative feedback. When blood glucose levels rise, more insulin is released, and works to lower blood glucose, as this happens and glucose levels return to normal, insulin release is switched off. Thermoregulation (control of body temperature) and osmoregulation (control of body water levels) are also controlled by negative feedback.

Positive feedback is less common in biological systems but two examples of processes that have positive feedback are childbirth and lactation. During childbirth stretching of uterus walls causes contractions that stimulate the release of the hormone oxytocin. This hormone stimulates further contraction of the walls until the baby is born. During lactation, the hormone prolactin stimulates milk production and as the baby suckles more prolactin is released and causes further milk production. This continues until the baby is weaned and stops feeding on milk. You can read about these processes in Chapter 8.

Phytohormones (plant growth regulators)

Plants produce growth regulator molecules, which act as signalling substances. These are sometimes called plant hormones or phytohormones, but they are different from animal hormones in structure and the way they work. Plant growth regulators are produced in very low concentrations but they control all aspects of plant growth and development, from embryogenesis, the sizes of different tissues and defence against disease to reproduction. You can read more about these substances in Section 7.3.

TEST YOUR UNDERSTANDING

1 List three substances that may act as cell signalling molecules.

2 Explain what is meant by the term ‘emergent properties’.

3 Outline the importance of cell signalling to multicellular organisms.

4 Give an example of a negative feedback effect in cell signalling.

5 Name a ligand that binds to an intracellular receptor.

Link

- What are other advantages of being multicellular? (Chapter 8.1)

7.2 Neural signalling

LEARNING OBJECTIVES

In this section you will:

- recognise that animals have a nervous system consisting of cells called neurons that transmit electrical impulses
- understand that the brain integrates and co-ordinates information
- recognise that the spinal cord integrates unconscious processes
- describe the structure of a typical sensory neuron and a typical motor neuron
- learn that many nerve fibres are covered with myelin sheaths
- learn that sodium and potassium ions are pumped across the membranes of neurons to generate a resting potential
- understand that an action potential consists of the rapid depolarisation and repolarisation of the neuron membrane
- recognise that nerve impulses are action potentials that are propagated along the axons of neurons
- discover that a nerve impulse is transmitted as a result of local currents that cause each successive part of the axon to reach the threshold potential

- define a synapse as the junction between two neurons or between neurons and receptor or effector cells
- understand that pre-synaptic neurons release a neurotransmitter into the synapse when they are depolarised
- learn that cholinergic synapses use acetylcholine as their neurotransmitter, which is the most common neurotransmitter in vertebrates and invertebrates

- recognise that neonicotinoid pesticides block synaptic transmission in insects
- discover that the speed of transmission of action potentials is increased by myelination, saltatory conduction and increased diameter of axons
- recognise that depolarisation and repolarisation occur during action potentials
- understand that local currents propagate an action potential along a nerve fibre
- learn that myelinated fibres have gated and non-gated ion channels at the nodes of Ranvier
- discover that neurotransmitters can be excitatory or inhibitory and that initiation or inhibition of an action potential results from summation of the two
- learn that drugs and other chemicals can stimulate or depress post-synaptic transmission

- understand that free nerve endings in the skin send pain messages to the brain
- discover that consciousness is a consequence of nerve interactions in the brain

GUIDING QUESTIONS

How does the nervous system of animals rely on electrical and chemical signals?

Most animals have nervous systems consisting of nerve cells called neurons. They transmit signals in the form of nerve impulses from one part of the body to another. In the human body there are about 85 billion neurons but a simple animal such as *Hydra* has only about 6000. Nervous systems have evolved to receive sensory inputs, integrate them and send out motor signals. A nervous system allows animals to perceive signals from its environment and to respond to them in a number of different ways such as moving in an organised way.

7.2.1 The structure of nervous systems

Most animals have nervous systems consisting of nerve cells called neurons that transmit electrical nerve impulses through their bodies. The simplest, sedentary multicellular animals, the sponges (Porifera), do not have nerves but simple cnidarians such as *Hydra* have a nerve net that runs throughout their body (Figure 7.2.1). More active invertebrates such as arthropods have a concentration of neurons in the head and these coordinate signals from the animal's body so that it can move and respond to stimuli in an organised way.

The human nervous system is more complex: it consists of a central nervous system, or CNS, is made up of the neurons of the brain and the spinal cord. The CNS receives information from sensory receptors all over the body. Information is processed and interpreted before the CNS initiates suitable responses.

The peripheral nerves are the networks of neurons that carry information to and from the CNS. Peripheral nerves include sensory neurons, which carry information to the CNS, and motor neurons, which transmit impulses from the CNS to muscles and glands that then cause a response.

The brain processes information that it receives from all the peripheral nerves. Different parts of the brain are responsible for different functions, the cerebral hemispheres co-ordinate learning, memory, language, speech and reasoning, while the cerebellum co-ordinates movement posture and balance. The hypothalamus in the centre of the brain (Figure 7.2.3) co-ordinates the endocrine and nervous system by regulating the secretions of the pituitary gland which lies just underneath it. The medulla oblongata (brain stem) controls automatic and

homeostatic activity such as breathing swallowing, digestion and heart rate (see Chapter 9).

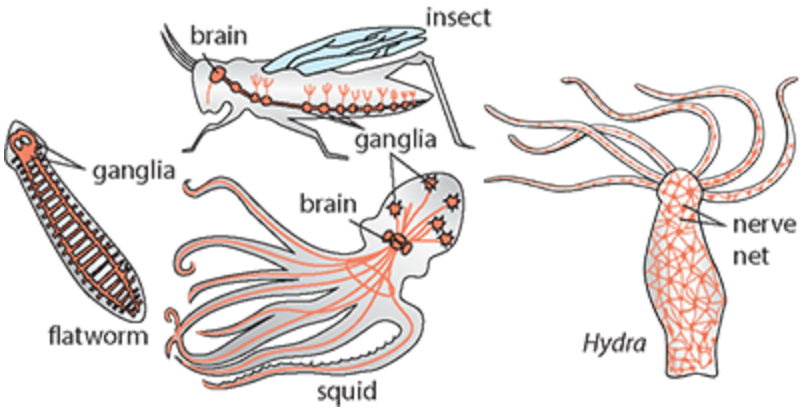


Figure 7.2.1: Examples of invertebrates with nervous systems include simple cnidarians such as *Hydra*, molluscs such as squid, and arthropods such as insects.

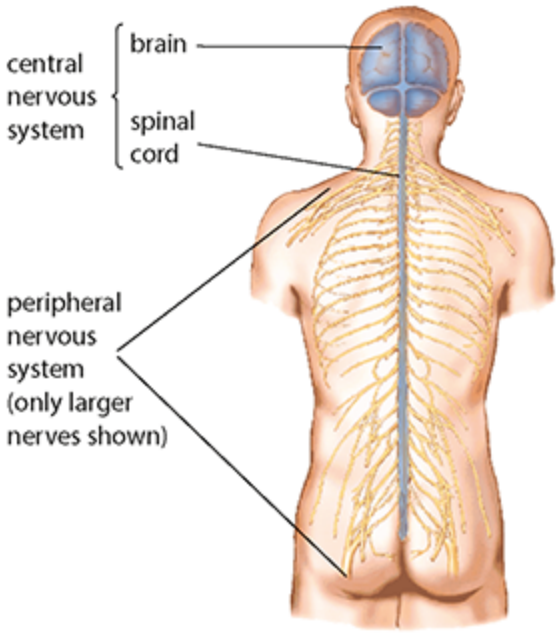


Figure 7.2.2: The human nervous system.

The spinal cord integrates unconscious processes which are those that take place without us thinking about them. Reflex actions such as the knee jerk reflex that contracts the muscles in the leg when the tendon below the knee cap is tapped, are co-ordinated by the spinal cord which receives and transmits messages to make the leg kick out.

Conscious actions are those which we are aware of, they require us to think or consider our actions and to make intentional responses to a situation. Conscious actions are coordinated by the brain.

Consciousness is a feature of the human brain which results from the interaction of many individual neurons in the brain. It is another example of an emergent property which is a consequence of interactions of different cells and chemicals in the brain.

You can read more about voluntary and involuntary control of actions in Chapter 9. Fig 9.1.5 shows the relationship between the different parts of the nervous system that are involved.

Three types of neuron are found in the nervous system. **Sensory neurons** and **motor neurons** transmit information to and from the CNS, while **relay neurons** within the CNS form connections between them.

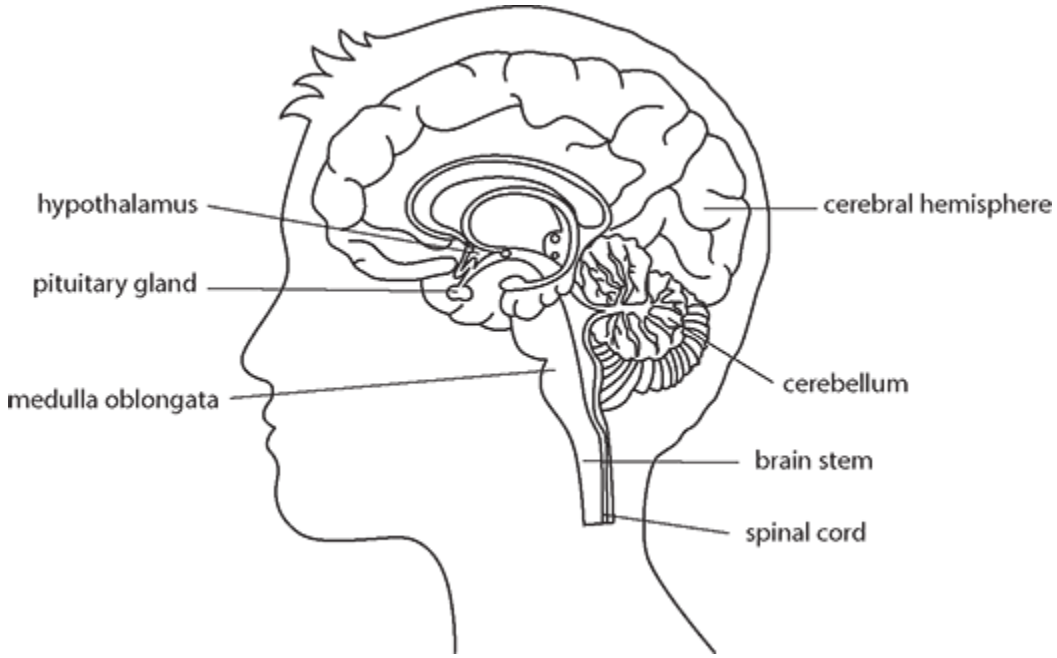


Figure 7.2.3: The human brain

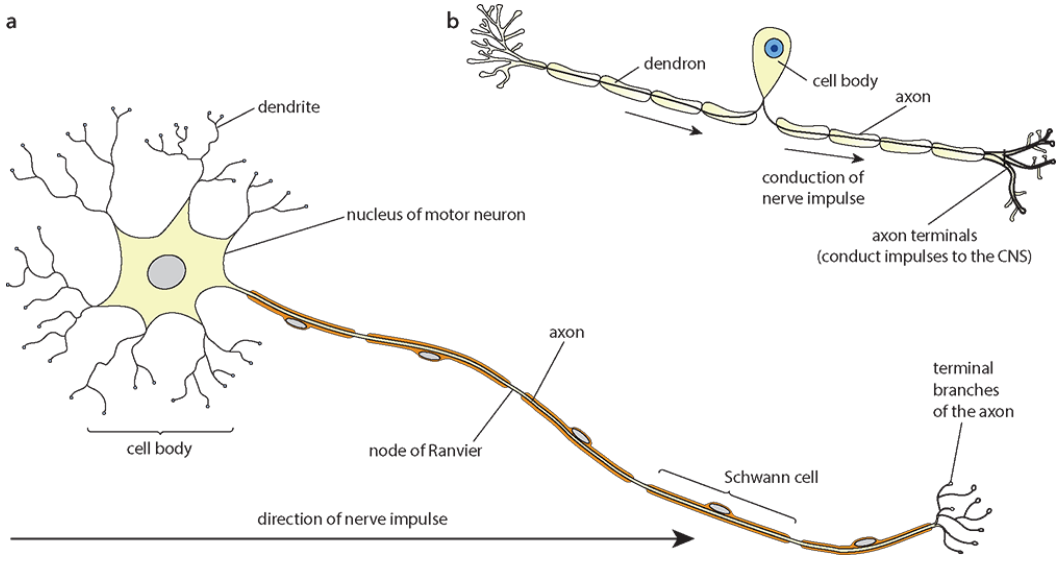


Figure 7.2.4: a A motor neuron. b A sensory neuron.

The structure of a typical motor neuron is shown in Figure 7.2.4a and a typical sensory neuron in Figure 7.2.4b. In motor neuron many small dendrites receive information from relay neurons and transmit the impulses to the cell body. One long axon then carries

impulses away. The cell body contains the nucleus and most of the cytoplasm of the cell. The axon is covered by a myelin sheath formed from Schwann cells, which wrap themselves around it. Myelin has a high lipid content and forms an electrical insulation layer that speeds the transmission of impulses along the axon.

7.2.2 Transmission of nerve impulses

Resting potential

Neurons transmit information in the form of impulses, which are short-lived changes in electrical potential across the membrane of a neuron. All neurons contain sodium (Na^+) and potassium (K^+) ions. Impulses occur as these important ions move in and out through the plasma membrane.

When a neuron is not transmitting an impulse, it is said to be at its **resting potential**. The resting potential is the potential difference across the plasma membrane when it is not being stimulated: for most neurons, this potential is –70 mV. The inside of the axon is negatively charged with respect to the outside (Figure 7.2.5).

As a nerve impulse occurs, the distribution of charge across the membrane is reversed. For a millisecond, the membrane is said to be **depolarised**. As charge is reversed in one area of the axon, local currents depolarise the next region so that the impulse spreads along the axon (Figure 7.2.6). An impulse that travels in this way is known as an action potential.

Action potential

KEY POINT

action potential is the reversal (depolarisation) and restoration (repolarisation) of the resting potential across the plasma membrane of a neuron as an electrical impulse passes along it.

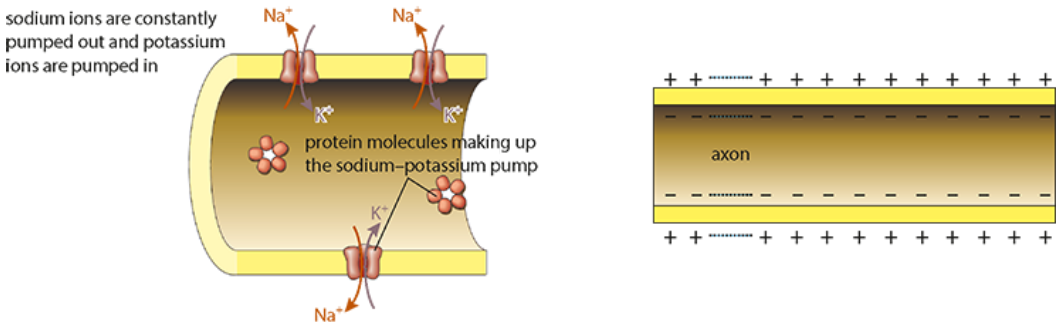


Figure 7.2.5: At rest, sodium ions are pumped out of the neuron and potassium ions are pumped in, to establish and maintain the resting potential. (You can read more about the sodium–potassium pump in Section 6.2.) Inside the neuron is negatively charged because of the presence of chloride and other negative ions.

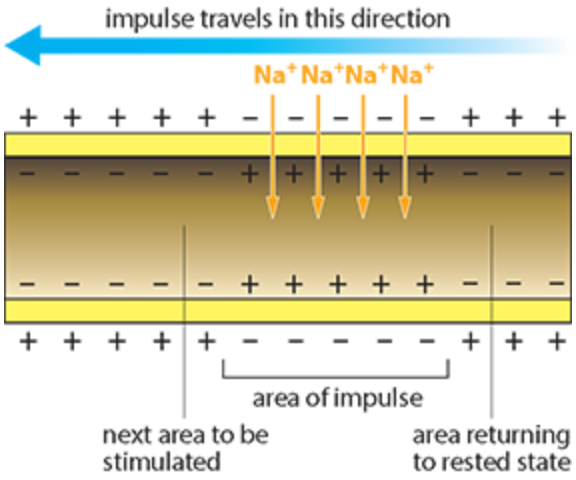


Figure 7.2.6: When an impulse passes along the neuron, sodium ions diffuse via ion channels and the potential is reversed. This process is called an action potential (only a small part of a long neuron is shown).

Figure 7.2.7 shows what is happening at the plasma membrane of the neuron as an action potential is generated. Nerve impulses are action potentials that travel along neurons due to sodium and

potassium ions moving down electrochemical gradients to create local currents.

1 When a neuron is stimulated, gated sodium channels in the membrane open and sodium ions (Na^+) from the outside flow in. They follow both the electrical gradient and the concentration gradient, together known as the electrochemical gradient, to move into the cell. The neuron is now said to be depolarised.

2 For a very brief period of time, the inside of the axon becomes positively charged with respect to the outside as sodium ions enter. At this point, the sodium channels close.

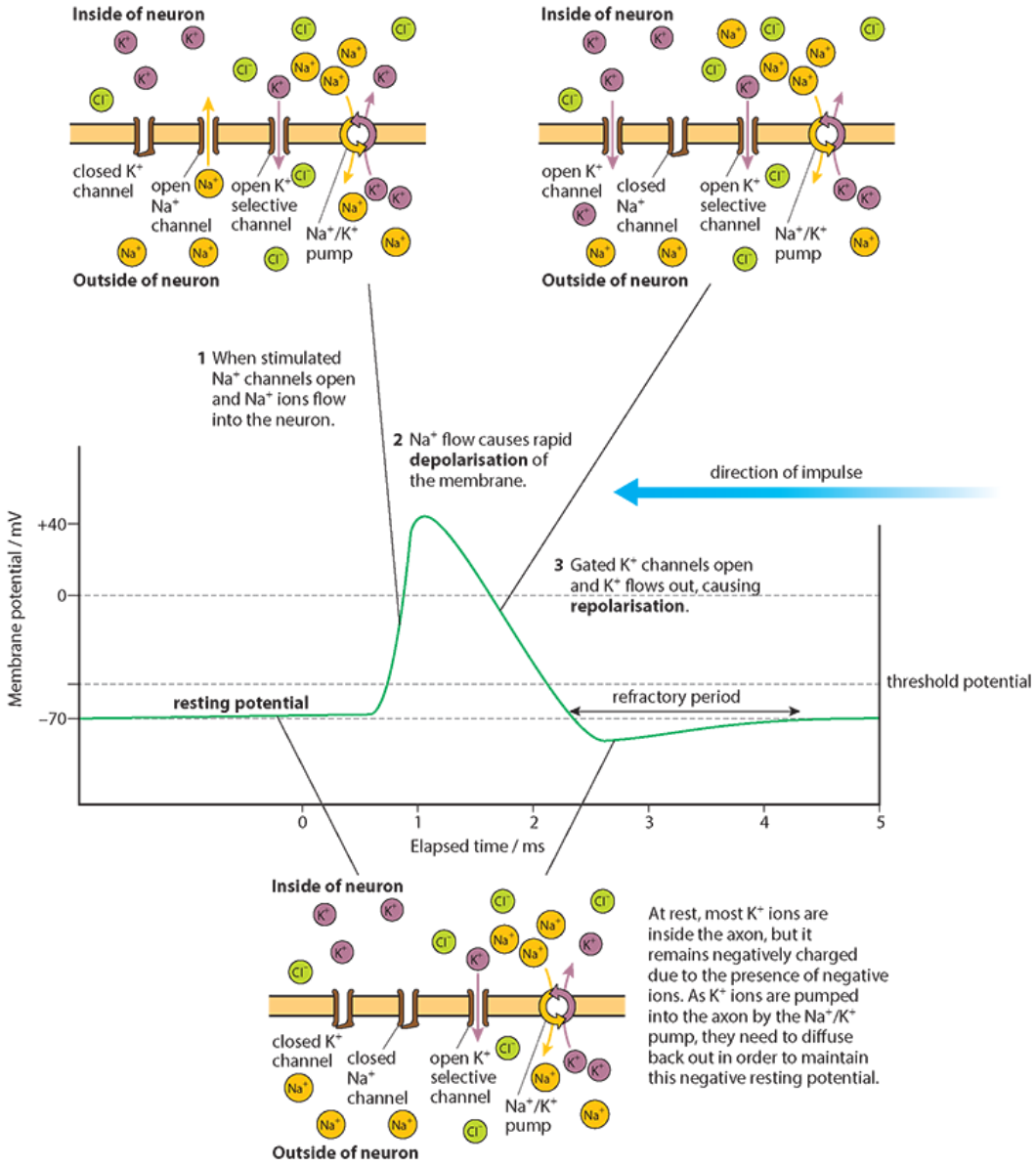


Figure 7.2.7: The action potential.

3 Now, gated potassium channels open and potassium ions (K^+) begin to leave the axon, moving down their electrochemical gradient to re-establish the resting potential, a process known as **repolarisation**.

4 Because so many potassium ions start to move, the potential difference falls below the resting potential. At this point,

both sodium and potassium channels close. The resting potential is re-established by the action of sodium–potassium pumps, which move ions back across the membrane.

An action potential in one part of an axon causes the **depolarisation** of the adjacent section of the axon. It occurs because local currents are set up between adjacent regions and these cause ion channels to open, allowing sodium ions in and potassium ions out of the axon. When this happens, each successive part of the axon to reaches its threshold potential and becomes depolarised.

KEY POINT

threshold potential the electrical potential across the plasma membrane of a neuron that is required to trigger an action potential.

The action potential travels along the neuron in a cascade, rather like a 'Mexican wave'. The impulse can only pass in one direction because the region behind it is still in the recovery phase of the action potential and is temporarily unable to generate a new action potential. The recovery phase is known as the **refractory period**.

The speed of conduction along an axon is affected by the diameter of the axon. A larger diameter means faster conduction. Larger axons are myelinated while smaller ones are not. At intervals along myelinated axons are gaps between the myelin covering known as nodes of Ranvier (Figure 7.2.4a). The sheath prevents the flow of ions across the membrane so the current must jump from node to node that speeds up the transmission of the nerve impulse.

Analysing oscilloscope traces

The potential of membranes in neurons can be measured using electrodes on each side of the membrane. The potentials can be shown as a graph or trace on an oscilloscope (a device that displays varying voltages as graphs) as shown in Figure 7.2.8. At rest, the trace shows a potential of −70 mV but as an action potential occurs the trace has a spike. The rising and falling of the line indicates the depolarisation and repolarisation of the membrane. The repolarisation does not return the membrane to its original resting potential immediately: there is a short delay, or refractory period, before it finally returns to −70 mV. Note that the resting potential is not −70 mV for all neurons in all species.

Figure 7.2.8 shows a trace taken from a mouse neuron after it had been stimulated with a small pulse of current.

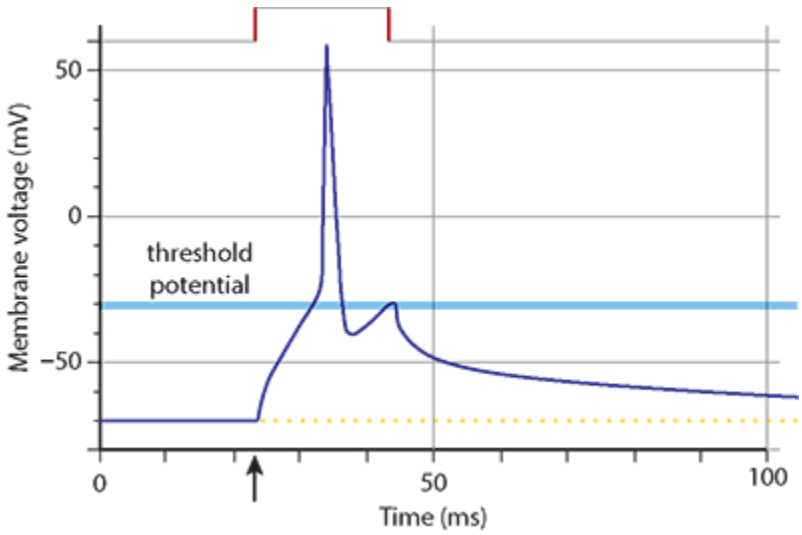


Figure 7.2.8: Action potential of a mouse neuron.

TEST YOUR UNDERSTANDING

6 State the resting potential of this neuron.

7 Work out the time taken for depolarisation and repolarisation shown on this trace.

8 Why does the membrane potential rise for a short time at the end of the repolarisation phase?

7.2.3 Synapses and synaptic transmission

A **synapse** is the place where two neurons meet or where a neuron meets a receptor cell (such as a touch receptor in the skin) or effector cell (such as a muscle cell). Neurons do not touch one another and the tiny gap of about 20 nm between them is known as the **synaptic cleft**. Action potentials must be transmitted across this gap for the impulse to pass on its way. This is achieved by the presence of chemicals known as **neurotransmitters**. Neurotransmitters are held in vesicles in the pre-synaptic cell until an action potential arrives. They are then released into the synaptic cleft, and diffuse across to the post-synaptic membrane. There they can cause another action potential to be produced and a nerve impulse to be initiated, provided the threshold potential is reached.

SCIENCE IN CONTEXT

Multiple sclerosis

Multiple sclerosis (or MS) is a neurological condition that affects the brain and spinal cord. It causes problems with vision, movement and balance. It is an autoimmune condition, which means that the body's own immune system begins to attack a healthy part of the body. In MS, the immune system attacks the myelin covering of neurons. The myelin becomes scarred and damaged so that messages travelling along the neurons are slowed down or disrupted (Figure 7.2.9). MS is usually diagnosed when people are in their 20s and 30s and it is twice as common in females than males. The causes of MS are not fully understood, but are probably a combination of genetic and environmental factors.

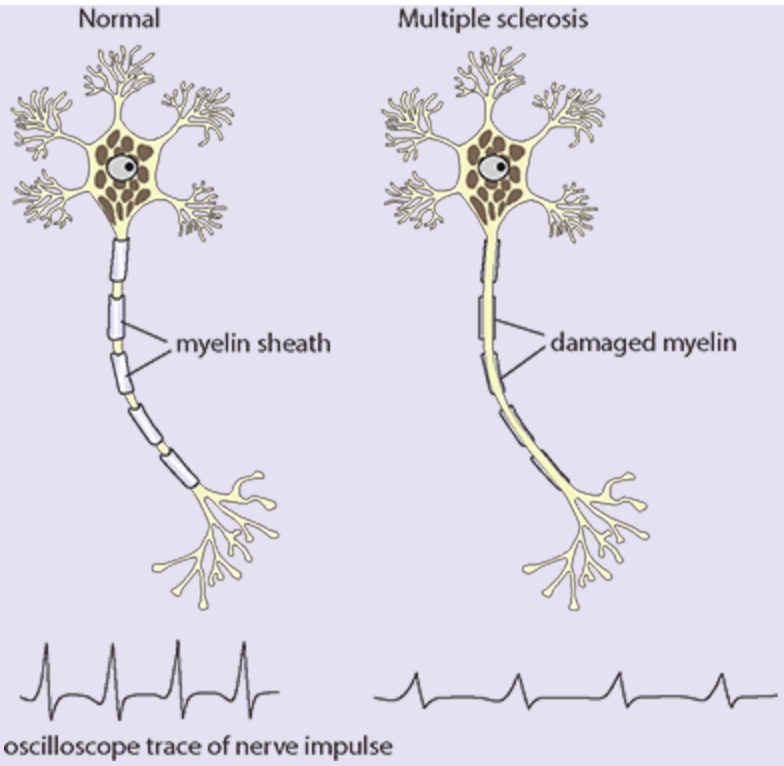


Figure 7.2.9: A normal neuron and its oscilloscope trace compared with a neuron from a person with MS.

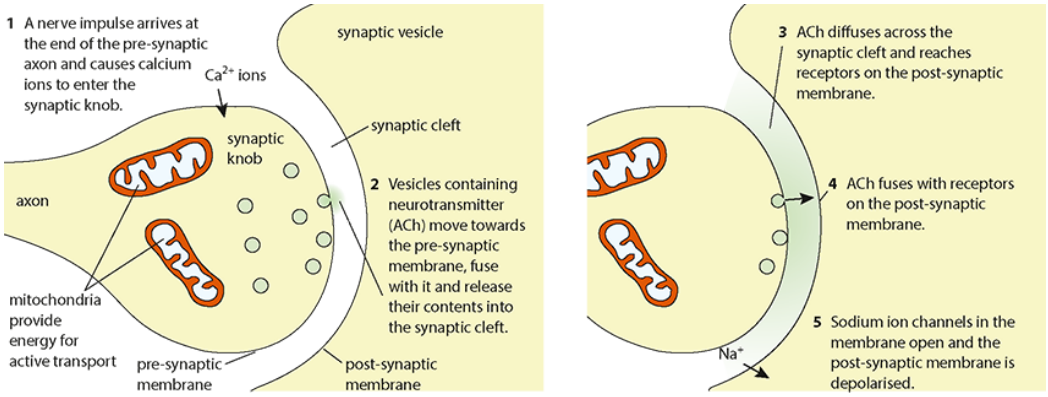


Figure 7.2.10: A cholinergic synapse. The whole sequence of events of transmission at a synapse takes 5–10 ms.

The synapse shown in Figure 7.2.10 uses the neurotransmitter acetylcholine (ACh) and so is known as a cholinergic synapse.

ACh binds to receptors and causes depolarisation of the post-synaptic membrane and the initiation of an action potential. Once an action potential is generated in the post-synaptic membrane, ACh in the synaptic cleft is deactivated by acetylcholinesterase enzymes and the products are reabsorbed by the pre-synaptic membrane to be remade and repackaged in vesicles.

Acetylcholine is the most common neurotransmitter in both invertebrates and vertebrates. It is used at many synapses including those between neurons and muscle fibres. In the human body there are more than 40 different neurotransmitters. Acetylcholine and norepinephrine (also called noradrenaline) are found throughout the nervous system, others (for example, dopamine) are found only in the brain.

Neonicotinoids are chemical pesticides used in insecticides. They are similar in structure to nicotine and block transmission at the synapses of insects by binding to acetylcholine receptors. Neonicotinoid pesticides have been linked to the decline of bee populations throughout the world. In 2013, the European Food Safety Authority declared that these insecticides pose an unacceptably high risk to bees and the European Commission imposed restrictions on their use. In 2018 three main types of the pesticide were banned for all outdoor uses. In the USA the Environmental Protection Agency (US EPA) has also taken steps to reduce the use of neonicotinoids. In 2019 it revoked permissions for the use of several pesticides containing them.

SCIENCE IN CONTEXT

Botox

A neurotoxin produced by the bacterium *Clostridium botulinum* is one of the most toxic substances known.

Nevertheless it has been used in recent years for cosmetic and medical procedures under the trade name Botox. In the early 1980s, it was used to treat muscle spasm, strabismus (squint) and uncontrollable blinking. More recently it has been used to induce temporary muscle paralysis and conceal the appearance of wrinkles.

TEST YOUR UNDERSTANDING

9 Explain how sodium and potassium ions establish a resting potential.

10 Define the term 'action potential'.

11 Name the substance released by a pre-synaptic membrane.

12 List the key events of synaptic transmission.

THEORY OF KNOWLEDGE

How can we test the accuracy of scientific knowledge?

If you wanted to carry out an experiment to verify the information you have been given about neural transmission you could plan to carry out an experiment to gather evidence. It might seem a good idea to observe results first hand for yourself. But this is not easy. You would need special equipment, you would need to take precautions to make sure any experiment was reliable and ethical. You would also need to be completely familiar with concepts such as potential difference and depolarisation.

Gathering knowledge depends on methods that have been used before. Science uses methods to collect information

using observations that limit the number of variables. This allows us to draw conclusions that are as clear as possible based on the abilities we have to see, recognise and understand them.

We can gain, evaluate and produce knowledge for ourselves but we should be aware that other people have the same ‘tools’ for understanding as we do and have used them in the past; so sharing knowledge and understanding is possible and valid. We do not have to repeat every experiment to be sure of its validity.

7.2.4 Myelination of nerve fibres

Speed of transmission of impulses

Action potentials travel along nerve fibres as ions cross the neural membrane. Positive sodium ions enter the neuron and trigger the opening of channels at the start of an axon; this lets in additional positive ions. These positive ions trigger the channels next to them, which let in further positive ions and so on, creating the action potential. The speed of transmission of an impulse is affected by both the diameter of the axon and whether or not it is myelinated.

Axons with a larger diameter can conduct impulses faster than smaller ones because of their lower surface area to volume ratio. Leakage of ions through the membrane is less of a problem in larger axons.

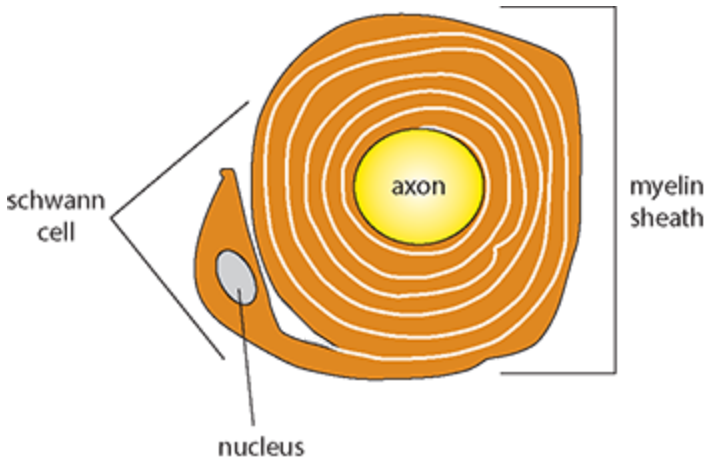


Figure 7.2.11: Schwann cells wrap axons in an insulating layer of myelin.

The rate of transmission of action potentials in an axon is also speeded up if the axon is wrapped in myelin, a fatty material

found in Schwann cell membranes (Figures 7.2.4 and 7.2.11). Schwann cells wrap around axons and create an insulating layer. The presence of myelin prevents ions being lost across cell membranes and maintains the action potential so that the speed of transmission of an impulse is about ten times faster in a myelinated axon than an unmyelinated one.

The reason for the increase in speed is that the process of generating an action potential only takes place at specific points, the nodes of Ranvier (Figure 7.2.4a), where there are gaps in the myelin covering. The gaps between nodes are each about 1 μm long. At each node there is a cluster of voltage-gated sodium ion channels (Figure 7.2.6) as well as non-gated sodium and potassium ion channels. Positive ions gather at the nodes of Ranvier to balance negative ions that maintain the resting potential. At the nodes there is no covering to prevent the ions being close to the neuron membrane. When the action potential arrives at a node sodium ions move into the neuron as sodium channels open (Figures 7.2.7 and 7.2.12). The potential then jumps to the next node. Because it misses out the myelin-covered sections of the neuron, progress of the action potential is very rapid. This type of conduction in a myelinated neuron is called **saltatory conduction**.

In non-myelinated axons, the action potential must propagate continuously along the plasma membrane that makes the process much slower.

KEY POINT

saltatory conduction impulse conduction ‘in jumps’ along myelinated neurons, between nodes of Ranvier.

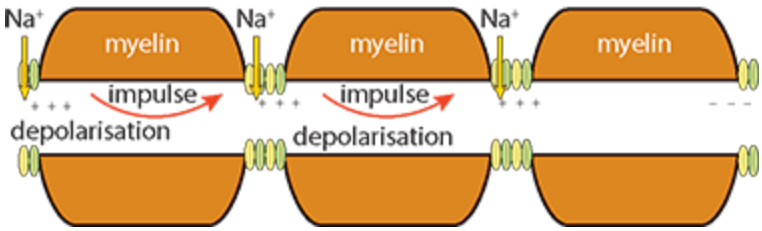


Figure 7.2.12: Saltatory conduction. Depolarisation only occurs at the nodes of Ranvier and the nerve impulse jumps from node to node.

Comparing rates of transmission of action potentials

Myelinated human axons vary in size, but most are between 13 and 20 µm in diameter and in these an action potential can travel at a velocity of 80–120 $m \cdot s^{-1}$. Smaller myelinated axons in the range 6–12 µm conduct at 33–75 $m \cdot s^{-1}$ (Figure 7.2.13). The velocity of an impulse in an unmyelinated axon is much lower, as the graph shows.

Invertebrate animals have unmyelinated axons that can be very large. The giant squid (*Architeuthis dux*) has giant axons that transmit messages to the muscles in its mantle. When stimulated these muscles can contract instantly to expel water from inside the animal so it can move rapidly away if it needs to escape a predator. Squid axons are unmyelinated and have diameters between 500 and 1000 µm (0.5–1 mm). Despite their size, these axons can only conduct an impulse at a velocity of about 25 $m \cdot s^{-1}$.

Temperature is also a factor affect speed of transmission of impulses because it affects diffusion. Sodium and potassium ions diffuse into and out of the axon: cold temperatures slow down the process of diffusion. The speed of transmission of impulses is faster in homeotherms, animals such as mammals and birds that maintain a constant body temperature, than in poikilotherms like

the giant squid, which do not. Velocity of travel of action potentials is directly correlated with increasing temperature.

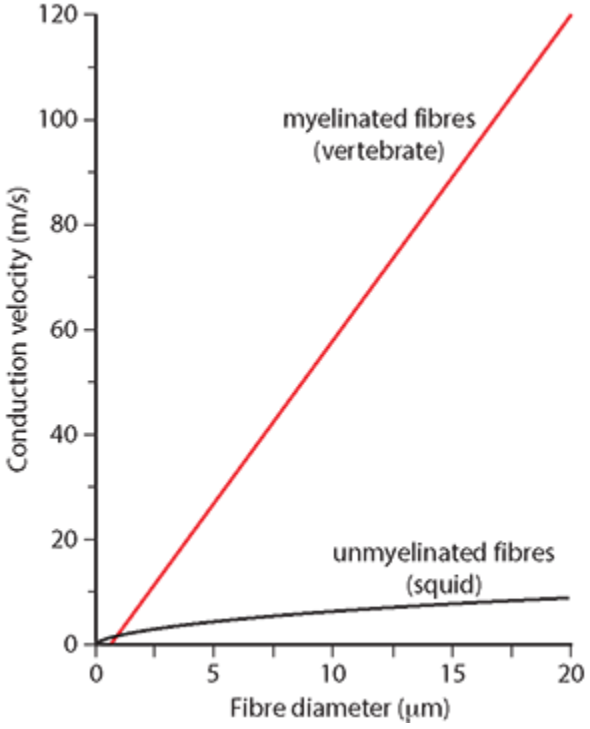


Figure 7.2.13: The conduction of action potentials in myelinated and unmyelinated fibres.

7.2.5 Excitatory and inhibitory neurotransmitters

At synapses, action potentials are passed from one neuron to the next. Some post-synaptic neurons are stimulated by many different pre-synaptic neurons, some of which release excitatory neurotransmitters and others inhibitory neurotransmitters (Figure 7.2.14).

Acetylcholine is an excitatory neurotransmitter. It increases the permeability of the post-synaptic membrane to sodium ions and depolarises the neuron.

KEY POINTS

excitatory neurotransmitter is one that activates receptors on the post-synaptic membrane and increases the likelihood that it will initiate an action potential.

inhibitory neurotransmitter is one that hyperpolarises the post-synaptic membrane and decreases the likelihood that an action potential will be triggered.

summation is the combination of stimuli received by the post-synaptic membrane from excitatory and inhibitory stimuli. If the two types of stimuli cancel each other out there will be no action potential generated.

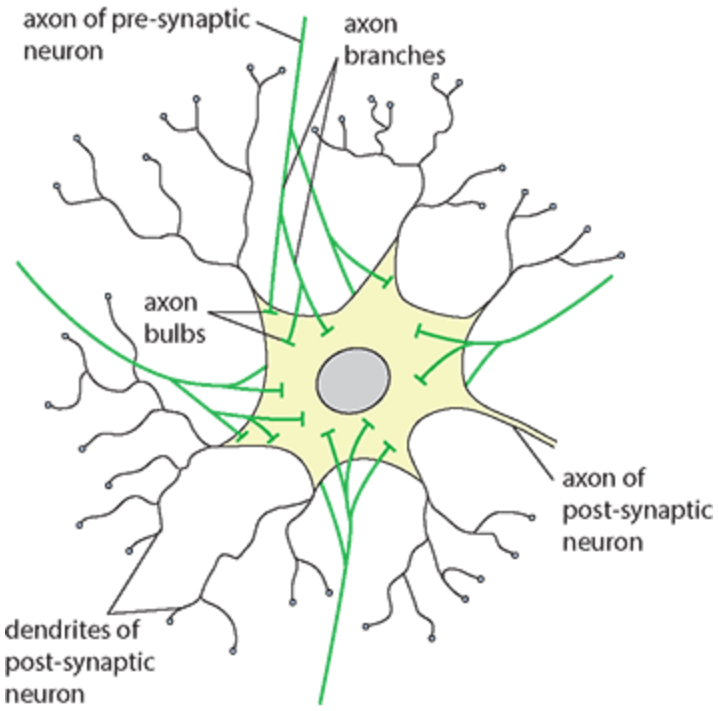


Figure 7.2.14: Some of the neurons that form synapses with the post-synaptic neuron are inhibitory producing inhibitory post-synaptic potentials (IPSPs) and prevent an action potential being stimulated. Others stimulate the propagation of the impulse producing excitatory post-synaptic potentials (EPSPs).

GABA (gamma-aminobutyric acid) is the most important inhibitory neurotransmitter in the brain. It causes the post-synaptic membrane to be **hyperpolarised**, so the inside of the neuron becomes more negative with respect to the outside. This makes it more difficult for an action potential to be generated. Inhibitory neurotransmitters may either cause positively charged potassium ions to move out of the post-synaptic cell or negatively charged chloride ions to move into it.

The balance of stimuli from these many pre-synaptic neurons can either excite or inhibit the post-synaptic neuron, giving a range of possible outcomes. The neuron may receive more stimulatory impulses overall so that it fires an action potential,

or it may receive mainly inhibitory impulses so that it does not. Summation of the excitatory and inhibitory neurotransmitters received from a pre-synaptic neuron will result in an action potential if the excitatory stimuli outnumber the inhibitory ones (Figure 7.2.15). Table 7.2.1 outlines the properties of some important neurotransmitters.

Summation

The effect of a neurotransmitter at a synapse depends on a number of factors, including how much neurotransmitter is released and the action of the neurotransmitter on the post-synaptic neuron. Receptors on the post-synaptic membrane influence ion channels that open when neurotransmitter molecules bind to them, producing a post-synaptic potential. These post-synaptic potentials may be excitatory (EPSPs) or inhibitory (IPSPs), depending on whether they bring the post-synaptic neuron closer to, or further away from, the threshold required to produce an action potential (Figure 7.2.15). A post-synaptic potential is approximately 1–2 mV. When a neuron receives both excitatory and inhibitory inputs together, the response will depend on whether the number of excitatory inputs overall are sufficiently great. If so, it will result in the generation of an action potential in the post-synaptic neuron and the transmission of a nerve impulse.

Neurotransmitter	Effects
acetylcholine	excitatory neurotransmitter that is found throughout the body. It triggers muscle contraction and stimulates the release of certain hormones. It is involved in wakefulness, attentiveness, anger and

	aggression. In some cases it can act as an inhibitory neurotransmitter
GABA (gamma-aminobutyric acid)	inhibitory neurotransmitter found in most inhibitory synapses in almost every part of the brain. many sedative drugs act by enhancing the effects of GABA. GABA contributes to motor control, vision and regulates feelings of anxiety
serotonin	an inhibitory neurotransmitter that is needed maintain stable moods and to balance stimulatory neurotransmitters in the brain
glutamate	excitatory neurotransmitter found in the brain and spinal cord. It is also used at modifiable synapses that are capable of increasing or decreasing the strength of an impulse and influencing memory
endorphins(opioid peptides)	a group of neurotransmitters found in pain pathways and emotional centres of the brain. They are produced by the pituitary gland and hypothalamus during exercise or excitement or when pain is felt. They are known as ‘natural painkillers’ because some have an analgesic effect

Table 7.2.1: The properties of some important neurotransmitters.

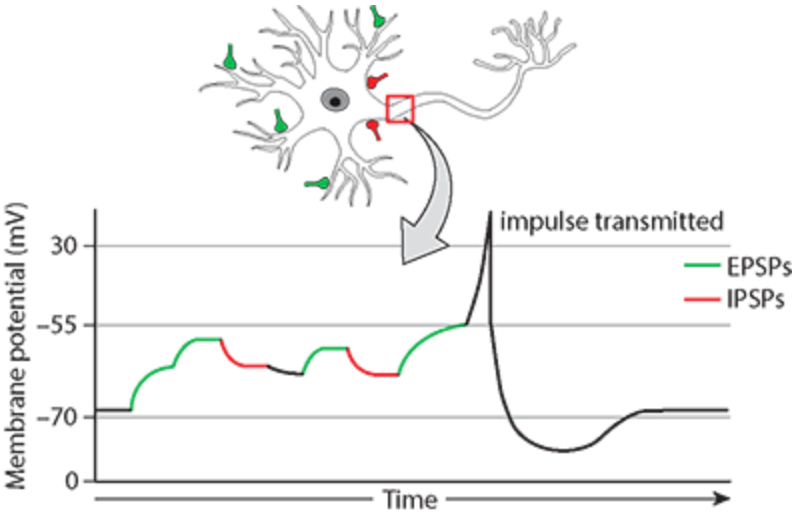


Figure 7.2.15: A neuron both receives EPSPs and IPSPs. The neuron is only able to transmit an impulse when the combination of these incoming impulses reaches the threshold potential of −55 mV.

7.2.6 Effects of chemicals on synaptic transmission

Psychoactive drugs and other chemicals can affect the way that the impulses are transmitted at synapses. They are capable of altering the functioning of the brain and a person's personality.

Chemicals and drugs act in different ways:

- Some have similar structures to neurotransmitters and so either block receptors, preventing a response, or have the same effect as the neurotransmitter, but are not removed, so that the response is prolonged.
- Some prevent neurotransmitters being released.
- Some increase the release of neurotransmitters.
- Some prevent neurotransmitters being broken down and so prolong their effects.

Excitatory substances (stimulants)

Some drugs and chemicals are excitatory, that is, they promote the transmission of impulses at excitatory synapses or inhibit transmission at inhibitory synapses. Examples of excitatory drugs include nicotine, found in tobacco products and cocaine. An example of a chemical with a similar action is neonicotinoid insecticide. Their action and effects are summarised in Table 7.2.2.

Excitatory drug	Mode of action	Effects
nicotine and	acts at synapses that use the	nicotine

neonicotinoid insecticides	neurotransmitter acetylcholine, but is not broken down by acetylcholinesterase, so it remains in the synapse, binding to the same receptors on the post-synaptic membrane as acetylcholine so that synaptic transmission is blocked	produces feelings of pleasure but is strongly addictive
cocaine	stimulates transmission of impulses at brain synapses that use dopamine (a neurotransmitter than is part of the 'reward' pathway) so it leads to a build-up of dopamine in the synapse because the re-uptake of the neurotransmitter is blocked	can produce feelings of excitement and confidence. Raises heart rate and body temperature. Strongly addictive

Table 7.2.2: Action and effects of some excitatory drugs.

Inhibitory drug	Mode of action	Effects
alcohol	increases the binding of GABA to receptors in post-synaptic membranes and causes hyperpolarisation of post-synaptic membranes; decreases the action of the neurotransmitter	in small quantities, affects behaviour by reducing inhibitions in larger quantities, can cause a lack of coordination, slurred speech, loss of balance and, in

	glutamate, which stimulates post-synaptic neurons	some cases, aggressive behaviour
THC (the most important psychoactive substance in cannabis)	affects receptors in cells in the cerebellum and cerebral hemispheres of the brain and causes hyperpolarisation of post-synaptic membranes so that they are more difficult to stimulate	induces feelings of relaxation and affects coordination but causes panic and paranoia in some users; can interfere with short-term memory and learning
benzodiazepines	bind to the same post-synaptic receptors as GABA, the main neurotransmitter at inhibitory synapses cause hyperpolarisation of post-synaptic membranes so that they are more difficult to stimulate	reduce anxiety, cause relaxation and can induce sleep used therapeutically to treat anxiety, insomnia and seizures

Table 7.2.3: Action and effects of some inhibitory drugs.

Inhibitory substances (depressants)

These drugs increase transmission at inhibitory synapses or suppress transmission at excitatory synapses. This class of drugs is also known as sedatives and includes anaesthetics used in medicine. Examples of inhibitory drugs include alcohol, THC and benzodiazepines. Inhibitory substances cause the

postsynaptic membrane to become hyperpolarised. Their effects are summarised in Table 7.2.3.

7.2.7 Perception of pain and consciousness

Pain receptors, also called **free nerve endings** are found in the skin at the base of hair follicles and close to the surface of the skin. Some types of pain receptor respond to pinching or cutting, others respond to extreme heat and other types respond to chemicals such as acids or the chemical capsaicin found in chilli peppers. The receptors respond when a stimulus causes or may cause tissue damage. If cells are damaged or substances synthesised at the site of an injury the receptors have channels for positively charged ions which enter the nerve endings and cause the threshold potential to be reached. This in turn initiates an action potential in the pain nerve fibres. The action potential is transmitted to the spinal cord and onward to the brain where the feeling of pain is perceived.

INTERNATIONAL MINDEDNESS

Human life expectancy is getting longer all over the world and the number of people affected by dementia and reduced brain function is also increasing. Dementia is an age-related neurodegenerative disease. The symptoms include decline in memory, attention span, language and problem-solving abilities. The World Health Organization (WHO) estimates that the number of individuals affected by dementia will triple by 2050 in populations globally. There is no cure for dementia, but treatments for the early stages of the disease include acetylcholinesterase inhibitors and other drugs that prevent the breakdown of acetylcholine and help nerve cells to communicate with each other. Scientists continue to seek ways to prevent the disease. Encouraging new research has

found that *statins*, a common medication given to people who suffer from high blood pressure, may offer potential benefits to prevent dementia in the future.

TEST YOUR UNDERSTANDING

13 List three factors that affect the speed of neural transmission.

14 Outline the difference between an excitatory and an inhibitory neurotransmitter.

15 Where would you expect to find free nerve endings?

Consciousness is loosely defined as the state of being aware of and responsive to one's surroundings, but scientists and philosophers have argued for many years about what being conscious really means. Nevertheless, we can say that interactions between the millions of neurons gives us the ability to speak, think, remember, feel pain and make decisions. Consciousness and these emergent properties are all consequences of the interactions that take place between the individual neurons in the brain.

REFLECTION

After studying this section can you say that you met your goals in understanding methods of cell signalling?

Link

- How is cell signalling important in animal communication?

7.3 Chemical signalling in animals and plants

LEARNING OBJECTIVES

In this section you will:

- learn that hormones include peptides, steroids and other organic molecules that act as chemical messengers
- recall that endocrine glands secrete hormones into the bloodstream that carries them to target organs
- discover that hormones act as first messengers to cause a response in target cells.
- learn how blood sugar level is controlled by insulin and glucagon
- understand that failure to produce enough or failure to respond to insulin can cause diabetes
- recognise that animal hormones or chemicals with similar effects are used in medicine

- learn that steroid hormones enter cells and bind to receptors to form receptor–hormone complexes that promote transcription. Examples include estrogen and testosterone
- learn that peptide hormones bind to receptors on cell surfaces to activate second messengers inside cells

- understand the role of G protein and (cAMP in the action of adrenaline
- understand the mode of action of the hormone insulin involving phosphorylation of tyrosine
- recognise that phototropism is a growth response to the plant growth regulator auxin
- understand that auxin affects gene expression and makes it easier for cell walls to expand so that shoots grow towards a light source
- learn that auxin efflux carriers maintain concentration gradients of phytohormone
- recognise that auxin and cytokinin interact to regulate root and shoot growth
- understand that positive feedback is involved in fruit ripening by ethylene.

7.3.1 Hormones in animals

The chemical structures of hormones

Hormones are chemical substances that are secreted directly into the bloodstream from **endocrine glands** found throughout the body (Figure 7.3.1). Each endocrine gland secretes a low concentration of hormone that travels in the blood to cells and tissues that are elsewhere in the body. Since hormones circulate in the bloodstream, they come into contact with many cells in the body but only cells that have specific, genetically determined receptors will respond.

These target cells have receptors on their plasma membranes that recognise and bind to the hormone.

Different hormones have different chemical structures and can be divided into three categories, as shown in Table 7.3.1.

KEY POINT

target cell a specific cell in the body that can respond to a particular hormone.

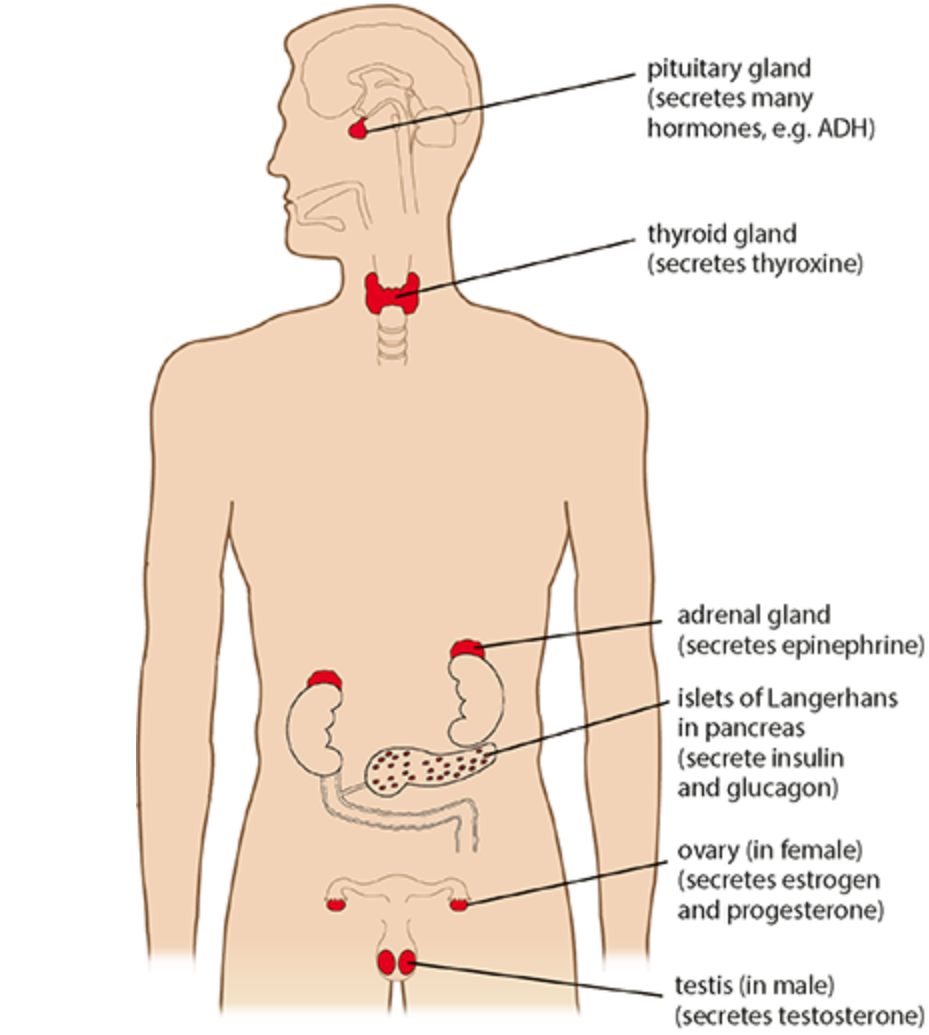


Figure 7.3.1: The location of the endocrine glands and the hormones they produce.

Chemical category of hormone	**Examples**
steroids derived from cholesterol	testosterone, progesterone
peptides (hormones consisting of chains of amino acids)	follicle-stimulating hormone (FSH), luteinising hormone (LH)

tyrosine derivatives (hormones derived from the amino acid tyrosine)	thyroxine (each thyroxine molecule contains four iodine atoms), epinephrine (adrenaline)

Table 7.3.1: Hormones are divided into three categories based on their chemical structure.

Hormones act as ‘first messengers’ to cause a response in a target cell. First messengers may or may not cross the plasma membrane but they initiate changes in the cell or tissue that is their target. Steroid hormones enter their target cells, but amino acid-derived hormones are hydrophilic molecules and do not enter the cell. Instead, they only bind to the surface of a target cell.

KEY POINT

first messenger an extracellular factor, often a hormone or neurotransmitter, that initiates changes within a cell even though it may not physically enter the cell.

Feature	Hormonal signalling	Nervous signalling
nature of the signal	chemicals, travelling in the blood	electrical impulses travelling along neurons
speed of action	can be slow, controlling long-term processes	very rapid
duration of response	minutes to days or years; effects may continue after the signal has been removed	milliseconds

area of response	signal affects many organs, more widespread response	response in one area only
example of control	growth, reproduction	movement, reflexes such as blinking

Table 7.3.2: Comparing hormonal and nervous signalling.

Comparing nervous and hormonal cell signalling

Both nerves and hormones send signals to cells, but their effects are quite different. Table 7.3.2 compares their two modes of action.

7.3.2 Insulin and glucagon, and control of blood glucose

Blood glucose level is the concentration of glucose dissolved in blood plasma. It is expressed as millimoles per decimetre cubed ($mmol \cdot dm^{-3}$). Normally blood glucose level stays within narrow limits, between 4 $mmol \cdot dm^{-3}$ and 8 $mmol \cdot dm^{-3}$, so that the osmotic balance (Section 6.3) of the blood remains constant and body cells receive sufficient glucose for respiration. Levels are higher after meals as glucose is absorbed into the blood from the intestine. They are usually lowest in the morning because food has not been eaten overnight.

Glucose levels are monitored by cells in the pancreas. If the level is too high or too low, cells in regions of the pancreas known as the islets of Langerhans produce hormones that turn on control mechanisms to correct it. β cells produce the hormone insulin and α cells in the pancreas produce the hormone glucagon.

Table 7.3.3 summarises these responses.

	Responses to a rise in blood glucose above normal	**Responses to a fall in blood glucose below normal**
Pancreas	β cells in the pancreas produce the hormone insulin	α cells in the pancreas produce the hormone glucagon
Glucose uptake or release	insulin stimulates cells in the liver and muscles to take in glucose and convert it to glycogen and fat, which can be	glucagon stimulates the hydrolysis of glycogen to glucose in liver cells:

	stored inside the cells: blood glucose levels fall	glucose is released into the blood

Table 7.3.3: The body's responses to changes in blood glucose.

Diabetes

Diabetes is the inability of the body to control blood glucose level. A person with untreated diabetes will experience wide fluctuations in their blood glucose above and below the normal limits (Figure 7.3.2). There are two types of diabetes: Type I is an autoimmune condition, while Type II is related to lifestyle factors.

In **Type I diabetes**, the β cells in the pancreas do not produce insulin. Without insulin, glucose is not taken up by body cells so blood levels remain high, a condition known as **hyperglycemia** (Figure 7.3.3). Excess glucose is excreted in urine and its presence is used to diagnose diabetes. About 10% of people with diabetes have Type I diabetes. Symptoms usually begin in childhood, which is why Type I diabetes is sometimes known as 'early-onset' diabetes.

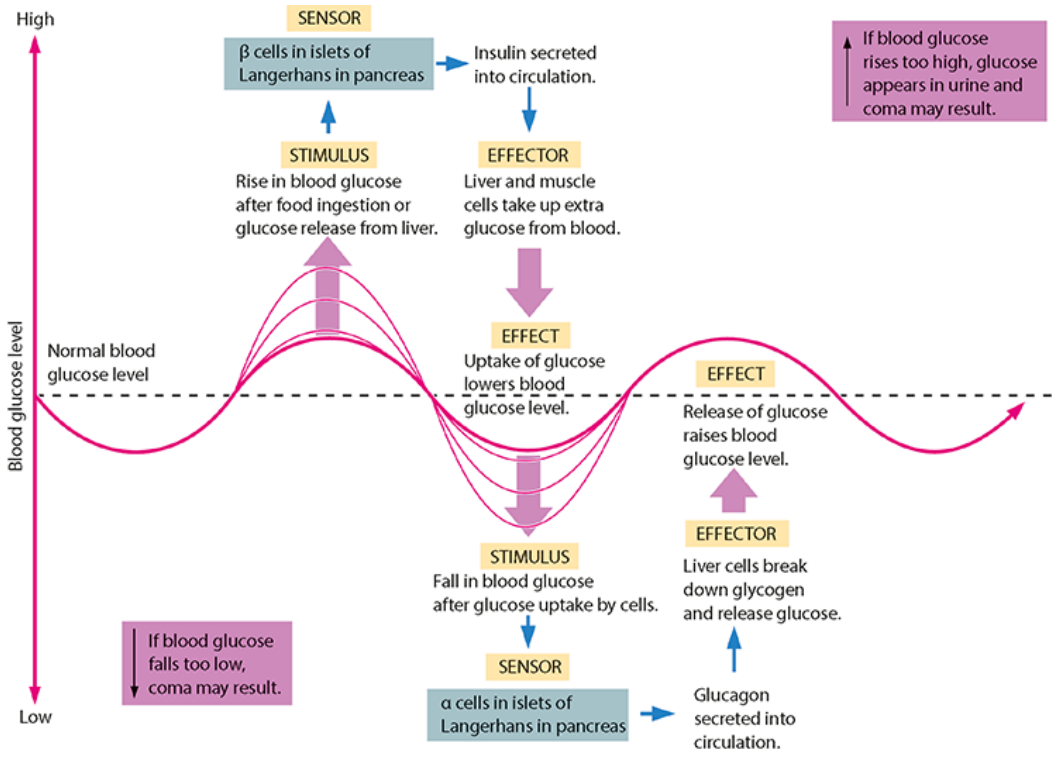


Figure 7.3.2: The control mechanism for blood glucose.

Type II diabetes is the most common form of diabetes, accounting for nine out of ten cases worldwide. The pancreas does produce insulin although levels may fall as the disease progresses. Type II diabetes occurs when body cells fail to respond to the insulin that is produced. Again, the result is that blood glucose levels remain too high. This form of diabetes is also known as late-onset diabetes or non-insulin-dependent diabetes mellitus. Individuals who have the condition develop insulin resistance, which means that the receptor cells that normally respond to insulin fail to be stimulated by it, even though the β cells in the pancreas still produce insulin. This type of diabetes is often associated with obesity, age, lack of exercise and genetic factors.

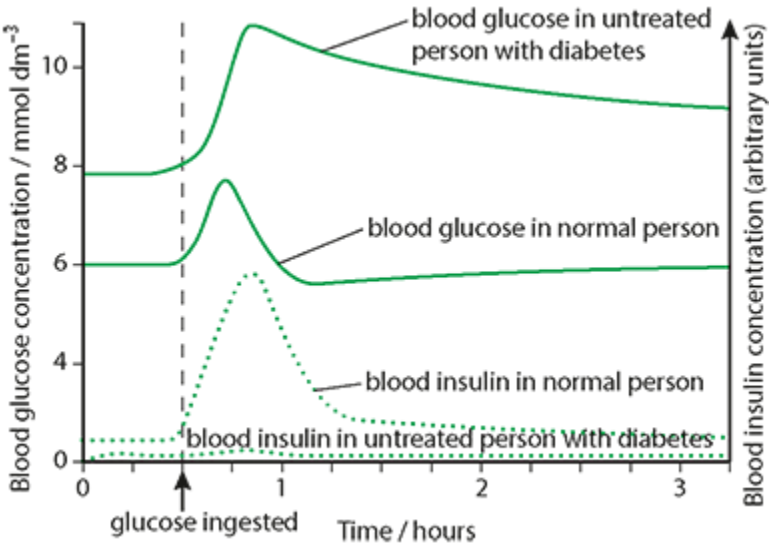


Figure 7.3.3: Blood glucose and insulin levels following intake of glucose in a normal person and in a person with untreated Type I diabetes.

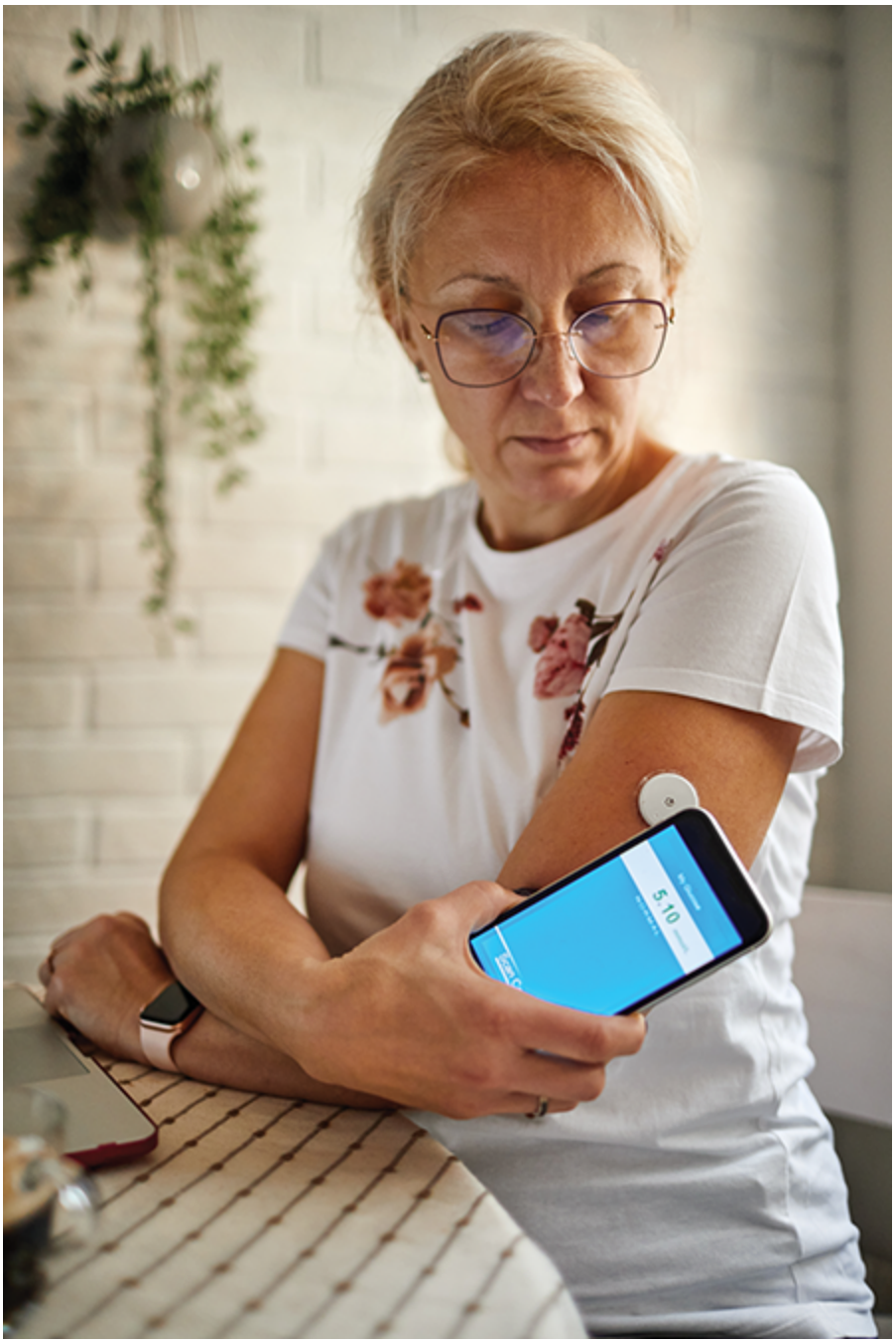


Figure 7.3.4: This woman has diabetes. She is wearing a glucose sensor on her arm and checking her blood sugar level using a smartphone app.

People with diabetes must monitor their blood glucose level carefully so that they can control it, as the body's internal control

mechanism is not working properly. This may involve taking tiny blood samples throughout the day and checking the amount of glucose that is present. Or, the person may be able to wear a glucose monitor that can be read using a smartphone app (Figure 7.3.4).

Causes and symptoms of diabetes

Type I diabetes is caused when the β cells in the pancreas do not produce insulin. This can be a result of autoimmune disease in which the body's immune system destroys its own β cells.

The causes of Type II diabetes are not fully understood but the risk of developing the disease has a strong correlation with weight and diet. High levels of fatty acids in the blood may be a factor causing the condition. People whose diets are high in fat but low in fibre seem to be most at risk. Obesity, associated with a lack of exercise or a genetic makeup that influences fat metabolism, is a key risk factor. The condition is more common in older people but there are an increasing number of cases in overweight children.

Studies of ethnic groups worldwide have shown that some are more likely to develop Type II diabetes, and this provides evidence for a genetic predisposition to the condition. The indigenous Aborigine population in Australia, people of Asian and Afro-Caribbean origin, the indigenous United States population and Polynesian Ma¯ori peoples all have a higher incidence of diabetes than would occur by chance.

The symptoms of diabetes include:

- high glucose levels in the blood
- glucose in the urine

- frequent need to urinate, which leads to dehydration and increased thirst
- tiredness and fatigue
- some loss of weight.

In Type II diabetes, these symptoms tend to develop slowly.

Treatment of diabetes

Type I diabetes must be controlled by regular insulin injections, but many people who have Type II diabetes are advised to control their blood sugar levels by following a healthy diet, taking exercise and losing weight. They are advised to eat foods that are low in saturated fat and salt but high in fibre and complex (slowly absorbed) carbohydrates, such as wholegrain cereals, pulses, beans and lentils. These foods, especially if they are taken at regular intervals during the day, help to keep blood sugar levels steady. Foods that should be avoided include sugary snack foods and drinks, and food with a high level of saturated fat. These foods cause a rapid rise in blood sugar level that the person with diabetes is unable to deal with.

If left untreated, Type II diabetes can lead to long-term health problems such as kidney disease, retinal damage, high blood pressure, stroke and heart attack.

7.3.3 Using hormones in medical treatments

Animal hormones or their artificially produced equivalents can be used to correct medical problems in many cases. The hormone insulin is needed to treat millions of people with diabetes all over the world. Doctors prescribe the hormone to protect their patients from harm caused by excessive blood sugar. Insulin is produced in recombinant bacteria or yeast by biotechnology companies. It is delivered to the patient with insulin pens or pumps that deliver it directly to the bloodstream.

Human growth hormone is produced by the pituitary gland and regulates growth in children and adolescents. Growth hormone deficiency is the most common cause of short stature in children, with about 1 in 5000 children being affected. Synthetic human growth hormone, first produced in 1985, is used to treat children with genetically inherited growth hormone deficiency or conditions such as Turner's syndrome, a genetic disorder that affects girls' development, and Prader–Willi syndrome. Growth of such children is increased so that they reach a more average height.

Two other widely used synthetic hormones are estrogen and progesterone, which are two female sex hormones that are used in oral contraceptive pills and in hormone replacement therapy treatments.

NATURE OF SCIENCE

Cause and correlation

A correlation between two variables does not guarantee a causal relationship, where one factor causes another. It is often difficult to work out whether one factor causes another or whether the two are simply correlated. In hormone research, many scientists have noticed an association between a person's testosterone levels and the risk of some diseases, including Type II diabetes. It was not clear whether testosterone level caused the risk of disease or was simply correlated with it.

In 2020 a new study involving nearly half a million European males and females analysed data from the UK Biobank and was able to show that many genetic variations influence testosterone levels and also that levels of testosterone are causally related to the risk of several diseases. The study showed that testosterone levels in males and females are inherited and influenced by many genetic variants and genes. Research continues to establish how testosterone influences diabetes and also how it is different between females and males.

SCIENCE IN CONTEXT

Human growth hormone in athletics

Human growth hormone (HGH) has been used by some athletes to promote muscle growth in an attempt to improve performance, particularly in power sports such as bodybuilding, swimming and weight lifting. The Olympic Committee declared the hormone a banned substance in 1989, but there is evidence that it is still used illegally (a practice called 'doping'). In the USA, HGH is available with a doctor's prescription. Scientific evidence about the effect of HGH on muscles is mixed. The hormone seems to reduce

body fat and increase lean body mass, but does not increase the strength of muscles.

It builds up connective tissue in muscles, making them appear larger, and it helps muscles resist injury and repair themselves. But these effects do not make the muscle stronger.

In 2010 researchers funded by the World Anti-Doping Agency in Sydney, Australia, reported that there was no evidence that HGH gave an increase in power, strength or endurance. Despite the results of studies like this, and HGH's side effects, which include muscle and joint pain and swelling, some athletes still use the hormone illegally. Part of its attraction may be that, because it is a naturally occurring substance, it is hard to detect in anti-doping tests.

TEST YOUR UNDERSTANDING

16 Define the term hormone.

17 Name three types of chemical molecules that act as hormones.

18 State the effect of insulin on blood sugar levels.

19 Where is glucagon produced?

20 Summarise the difference between Type I and Type II diabetes.

7.3.4 Mode of action of steroid and amino acid-derived hormones (ligands)

Amino acid-derived hormones and steroid hormones control their target cells in different ways (Figure 7.3.5).

Steroid hormones, such as estrogen and testosterone, enter target cells because they are lipophilic and can easily diffuse through the plasma membrane. They bind to specific receptor proteins in the cytoplasm, forming a receptor–hormone complex, which is transported through a nuclear pore into the nucleus. Here, the hormone regulates the process of transcription of one or more specific genes (Figure 7.3.6).

Peptide hormones bind to a surface receptor, very often a glycoprotein, but do not enter the cell. Instead, the binding process triggers the release of a second messenger, a chemical that cascades from the cytoplasmic side of the plasma membrane and this messenger controls the activity of the cell. The second messenger may regulate the activity of a specific enzyme in the cell, either activating it or inhibiting it or trigger changes such as cell division, differentiation and apoptosis.

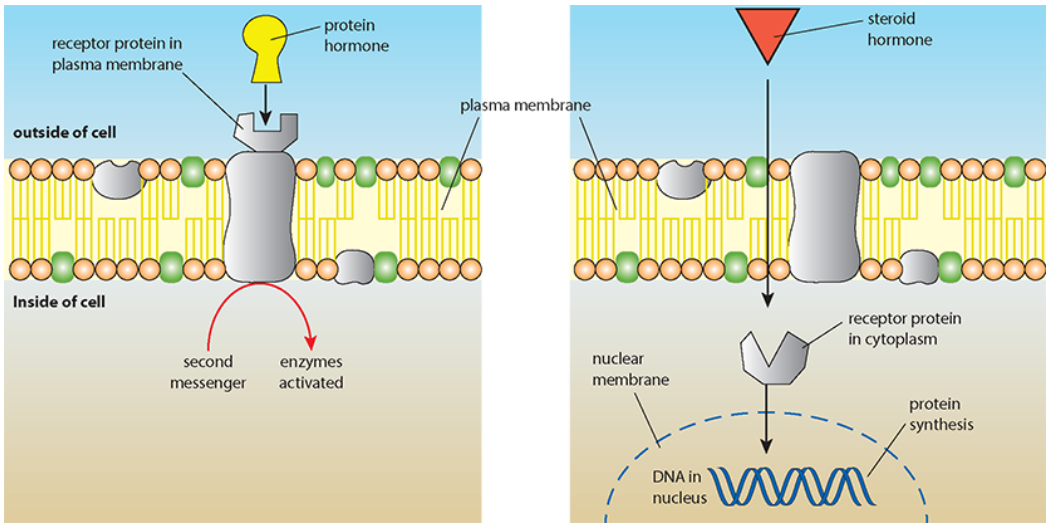


Figure 7.3.5: Ligands which are protein hormones do not enter cells but bind to receptors in the membrane. Steroid hormones enter cells and bind to receptors inside.

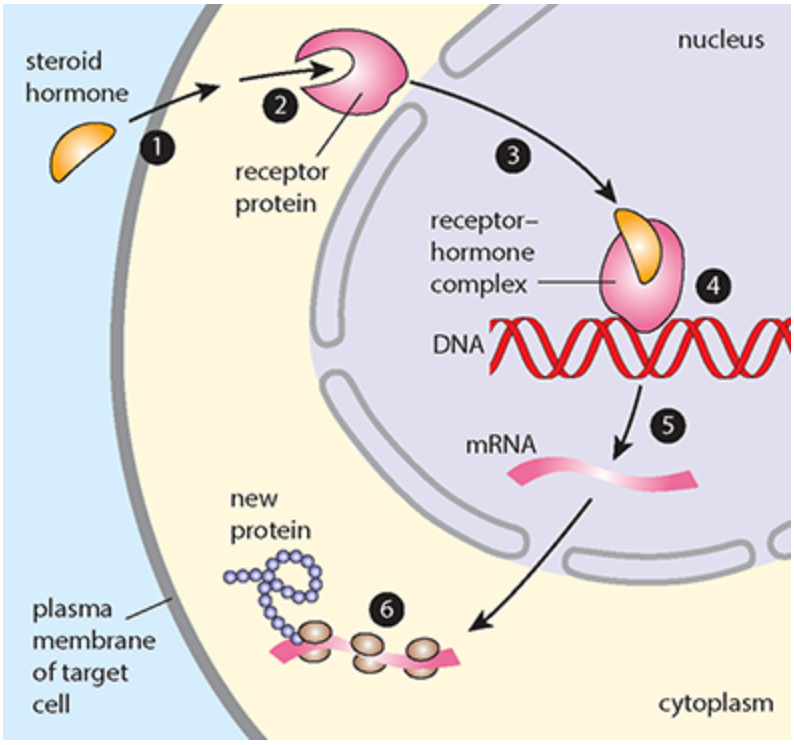


Figure 7.3.6: Mode of action of steroid hormones such as estrogen.

NATURE OF SCIENCE

Understanding key terms

International naming conventions are used in science so that all scientists and students can benefit and discuss their work. Adrenaline and epinephrine are two terms used for the same hormone. Both are based on the fact that the hormone is produced by the adrenal glands situated just above the kidneys. Adrenaline comes from Latin Ad = at and ren = kidney, while epinephrine is derived from Greek. Epi= above and nephros = kidney. Both these names are used interchangeably in different parts of the world.

Estrogen is a category of sex hormones that control the development and regulation of the female reproductive system and secondary sex characteristics. There are three major types of estrogen: estrone, estradiol, and estriol. Estradiol is present in women of reproductive age, estrone is predominant in women after menopause and estriol is at the highest level in the blood during pregnancy. You will probably see the term estrogen used most of the time although medical professionals are more likely to distinguish between the three types of estrogen.

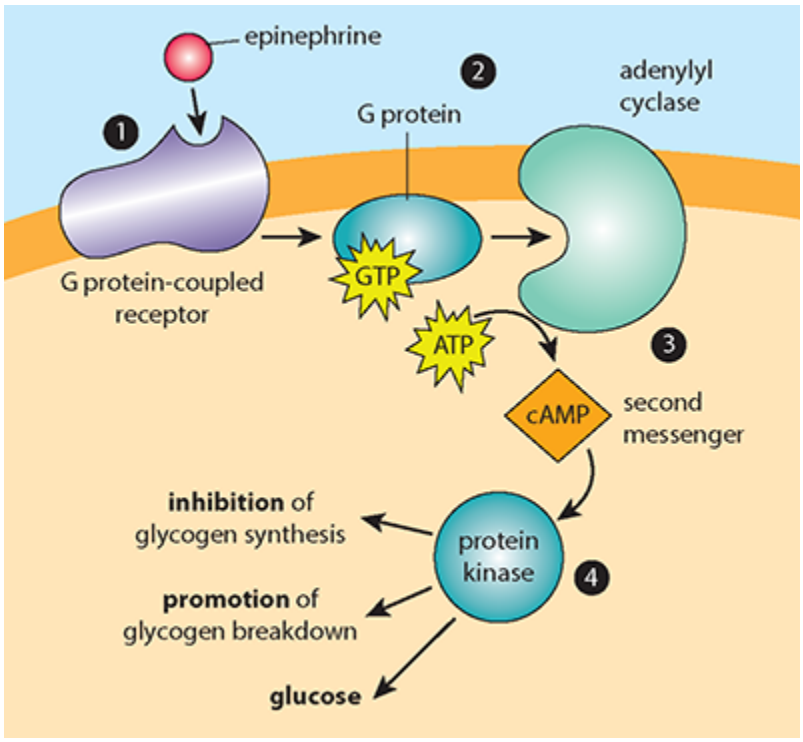


Figure 7.3.7: Mode of action of the amino acid derived hormone epinephrine. **1** Epinephrine binds to a G-protein-coupled receptor. **2** When stimulated, this receptor activates a G protein. **3** This initiates production of cAMP. **4** Cell signalling pathway acts to break down glycogen to glucose in the liver.

Insulin works in this way. Insulin binds directly to a receptor in the plasma membrane and causes the phosphorylation of tyrosine inside a cell. Insulin receptors are present on all cells, but their density is greatest on liver and adipose (fat) cells. A sequence of events follows:

1 Insulin binds to the receptor in the plasma membrane

2 Tyrosine inside the cell is phosphorylated

3 A cascade of phosphorylation causes vesicles containing glucose transporters to move to the plasma membrane

4 Glucose can now enter through the membrane and be metabolised or stored.

Epinephrine (adrenaline) is an example of a hormone derived from amino acids. It is involved in the ‘fight or flight’ response that occurs when a person feels under threat. Epinephrine is released by the adrenal glands and travels to its target cells in the blood. Epinephrine and glucagon together stimulate the release of glycogen stores in the liver to release energy for action. Epinephrine binds to alpha-1 adrenergic receptors on the outside of cells, causing them to change their shape and starting a cascade of reactions inside the cell. As a result of the change in shape of the receptors, a G protein is activated and this enables adenylyl cyclase and ATP to become active as well. Adenylyl cyclase breaks down ATP into a second messenger molecule called cyclic AMP, usually called cAMP. This second messenger causes protein kinase to inhibit glycogen synthesis and promote the breakdown of glycogen so that glucose can be released. (Figure 7.3.7).

7.3.5 Effects of phytohormones (plant growth regulators)

Plants produce signalling chemicals called phytohormones which control their growth, development and response to stimuli.

Charles Darwin noted the existence of these growth substances in 1880 in a report of his experiments on plant shoots. He observed that oat seedlings grew towards light because of some 'influence', which he thought was transmitted from the shoot tip to the area immediately below. We now know that the substance that causes shoots to bend towards the light is **auxin** and the response it causes is called **phototropism**. Auxin is found in the embryos of seeds and in apical meristems, where it controls several growth responses, or tropisms. Positive phototropism is a directional growth response of shoots towards a source of light.

Phototropism involves auxin and **phototropins**, which are photoreceptor proteins that respond to the blue wavelengths of visible light. These proteins are found in shoot cell plasma membranes and throughout a plant. Auxin which controls the growth of stems and roots can diffuse freely into plant cells but not out of them. It is moved from cell to cell, and its flow though the plant is redirected continuously as a plant is developing. Auxin is pumped out of cells by proteins in the cell membrane called **auxin efflux carriers**.

Phototropins become phosphorylated and activated on the side of a shoot that receives more light. Auxin is then redistributed to the side of a shoot tip that is away from a light source. The uneven distribution of auxin allows cell elongation on the

shaded side of a shoot, which in turn causes bending towards light (Figure 7.3.8).

In an evenly lit shoot, auxin promotes the elongation of cells at the shoot apex so that the shoot grows straight upward.

When auxin is present, a high concentration of hydrogen ions in the cell walls builds up due to the transport of ions though the plasma membrane. This reduces the pH and enables enzymes to break the cross links between cellulose molecules in the cell wall. As bonds are loosened between cellulose fibres, their structure becomes more flexible so that cells can expand. Auxins also interact with other plant hormones such as cytokinins to influence the general pattern of development. They influence and coordinate the development of plants by influencing the expression of genes. Auxin may activate certain genes, leading to a rapid response, or it may inhibit other sets of genes. Auxin affects both cell division and cell enlargement.

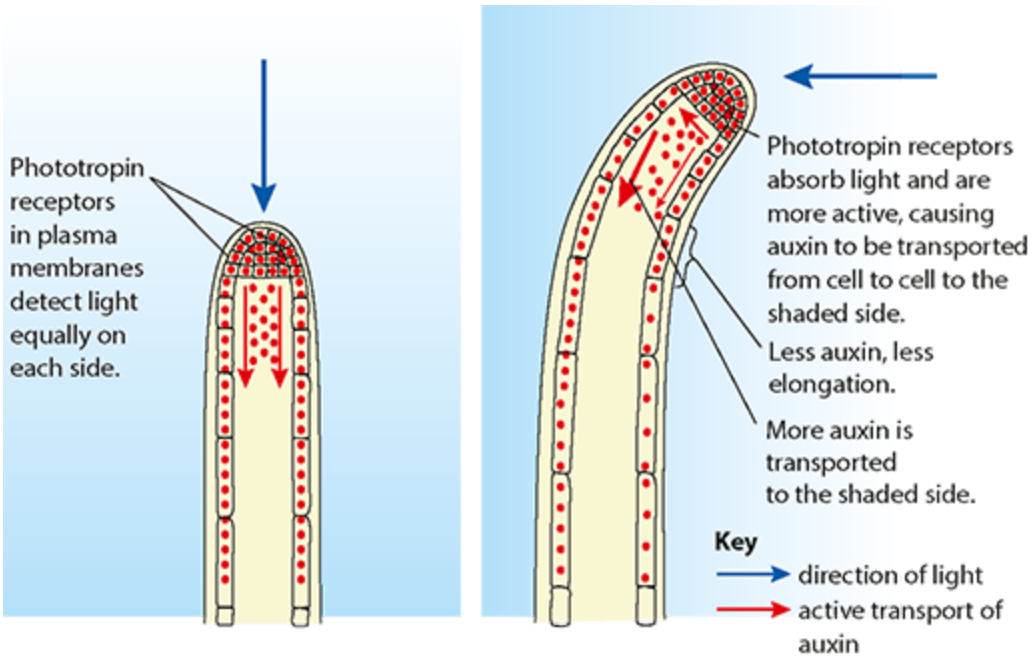


Figure 7.3.8: Some proteins (phototropins) in the plasma membranes of certain cells in plant shoots are sensitive to light. When light falls on them, they cause auxin to be transported to the shaded side of the shoot, which in turn causes the shoot to bend towards the light.

EXAM TIP

You should be able to make observations and measurement of tropic responses from experiments using seedlings.

You should also be able to annotate diagrams such as the one shown in Figure 7.3.8 to show the reactions of seedling you have grown.

There are five types of **phytohormone** with very different chemical structures: they are auxin, gibberellins, cytokinins, abscisic acid and the gas ethylene. Based on the effects they have, they can be divided into two main groups:

1 Growth promoters that promote cell division, cell enlargement, flowering, formation of fruits and seeds. Examples are auxins, gibberellins and cytokinins.

2 Growth inhibitors that inhibit growth and promote **dormancy** and abscission (leaf fall). An example is abscisic acid that regulates seed dormancy, and helps the plant respond to environmental stress.

Animal hormones	**Phytohormones (plant growth regulators (PGRs))**
produced in special endocrine glands	produced by most plant cells

target cells may be a long way from the site of hormone production	target cells are usually cells close to the cells that produce PGRs
two main types of hormone are peptides and steroids, which are complex organic molecules	five different types of growth substances are small simple organic molecules
hormones are secreted into the blood and are carried to target cells in the bloodstream	PGRs diffuse to nearby cells but may travel in the phloem
influence body growth and development	control all aspects of plant growth and development

Table 7.3.4: Comparison of animal hormones and plant growth regulators.

Ethylene is usually a growth inhibitor although in some circumstances it can become a promoter.

In most cases plant growth substances influence cells that are close to the cells that produce them. They are said to have a local effect. They travel to the target cells by diffusion. One example is auxin, which is produced in the stem, buds and root tips of plants and influences stem elongation and inhibits the growth of lateral (side) buds. Unlike animals, plants have no special organs to produce or store hormones and their growth regulators usually diffuse only short distances, although sometimes they can be carried in the phloem to target cells further away. Auxin and cytokinin work together to regulate root and shoot growth. Root tips produce cytokinin which is transported to shoots, while shoot tips produce auxin which is transported to roots. Interactions between these two growth substances helps to

ensure that root and shoot growth proceed at the correct rates. Table 7.3.4 shows some differences between plant and animal hormones.

SCIENCE IN CONTEXT

Ethylene and ripe bananas!

Bananas produce the plant growth regulator, ethylene (C_2H_4). Ethylene is a gas that influences the ripening of bananas and other fruits including apples, pears, peaches and melons, which all produce ethylene. Commercial fruit companies use ethylene to speed up the ripening process. Ethylene works using positive feedback. It stimulates the start of the ripening process and ripening also stimulates increased production of ethylene. Positive feedback ensures that fruits ripen rapidly and that ripening is synchronised. Tropical fruits such as bananas are picked when they are green and unripe. They may have to travel long distances to consumers and ripe fruits would be past their best after a long journey. At their destination bananas are ripened artificially by exposing them to ethylene in special 'ripening rooms' before they are sent to the shops. The rooms are kept at a high level of humidity so that the fruit does not crack or dry out and this has the added advantage that it extends the shelf life of the fruit.

TEST YOUR UNDERSTANDING

21 Summarise how steroid hormones promote the transcription of specific genes.

22 Outline how amino acid-derived hormones activate a second messenger inside α cell.

23 Name an example of a plant growth regulator.

24 What is a positive phototropism?

REFLECTION

After studying this section can I say that I met my goals in understanding methods of cell signalling?

Links

- How do hormonal and nervous communication vary between different classes of organisms? (Chapter 9)
- How do the structures of cell membranes help cell signalling? (Chapter 6.1)
- How are hormones involved in homeostasis? (Chapter 8.5)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
describe the importance of cell signalling to unicellular and multicellular organisms	7.1.1, 7.1.2			
outline how emergent properties can develop as a result of cell signalling	7.1.3			
state that ions, neurotransmitters and hormones act as signals to cells	7.1.3			
summarise the differences between animal hormones and	7.1.3			

plant growth regulators				
summarise the effects of ligands on receptors in the plasma membrane and intracellular receptors	7.1.3			
recognise the differences in receptors and signal transduction involved in signals from adrenaline, acetylcholine, insulin and steroid hormones	7.1.3			
describe the structure of a typical neuron	7.2.1			
explain the importance of sodium and potassium ions in generating a nerve impulse	7.2.2			
define the terms resting potential,	7.2.2			

threshold potential, action potential, depolarisation and repolarisation and interpret traces of action potentials				
describe the events that occur at a synapse as an action potential arrives	7.2.3			
state that the brain processes information from the nerves while the spinal cord processes unconscious actions	7.2.3			
summarise the importance of myelin sheaths and saltatory conduction in speeding up transmission of impulses	7.2.4			
describe the events of saltatory conduction	7.2.4			

including gated and non-gated sodium and potassium channels at the nodes of Ranvier				
explain the difference between excitatory and inhibitory neurotransmitters and the importance of summation at a post-synaptic neuron	7.2.5			
outline the effects of chemicals such as nicotine and alcohol on the brain and their addictive nature	7.2.6			
recall that free nerve endings in the skin send pain messages to the brain	7.2.7			
state that consciousness is a consequence of	7.2.7			

nerve interactions in the brain				
identify three types of hormone molecule that act in cell signalling	7.3.1			
explain that hormones are produced by endocrine glands that secrete directly into the bloodstream	7.3.1			
state that hormones act as first messengers	7.3.1			
recall that insulin and glucagon are produced by α and β cells in the pancreas and control blood sugar levels	7.3.2			
explain how failure to produce or respond to insulin can lead to diabetes	7.3.2			
animal hormones can be produced	7.3.3			

artificially and used to treat medical problems				
recall that steroid and amino acid hormones work in different ways	7.3.4			
understand that steroid hormones enter cells and form receptor–hormone complexes that affect transcription	7.3.4			
understand that amino acid-derived hormones (insulin and adrenaline) do not enter cells but they stimulate release of second messenger molecules	7.3.4			
explain that positive phototropism is a directional growth response in plants controlled by the	7.3.5			

phytohormone auxin				
describe how movement of auxin controls cell elongation	7.3.5			
describe how auxin and cytokinn regulate root and shoot growth	7.3.5			
explain positive feedback in fruit ripening by ethylene.	7.3.5			

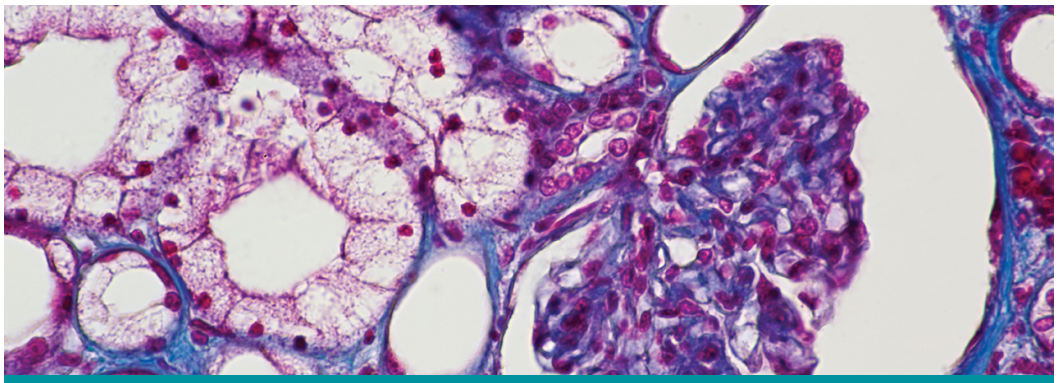
EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Unit 3
Organ systems and integration

INTRODUCTION

In multicellular organisms, cells are organised into tissues and organs that have their own jobs to do in the body. In complex organisms, processes such as transport of materials and gas exchange are carried out by specialised organ systems. For example, the heart and the blood system are responsible for transporting substances to and from cells. Plants exchange gases through their leaves and absorb nutrients through their roots. Animals exchange oxygen and carbon dioxide at surfaces such as the alveoli of the lungs. The tissues in a mulitcellular organism communicate with each other via nerves, hormones and cell signalling. Homeostasis monitors and keeps the levels of different substances constant so that the organism can always function at its optimum level. Animals also develop immune systems to protect their bodies from disease.



Chapter 8

Physiology

B3.1, B.3.2, D3.1, D3.3

INTRODUCTION

Multicellular organisms have cells that work together and are organised into tissues and organs that must function in harmony to carry out the functions of life. Physiology is the study of the organs and organ systems of the body and how they interact to keep complex organisms alive. Physiologists examine both the physical structure of organs and investigate the biochemical processes that operate to keep them working efficiently. Hormonal and nervous signalling systems integrate the actions of organs while transport systems carry materials between them.

8.1 Physiology - organ systems and integration

LEARNING OBJECTIVES

In this section you will:

- learn that multicellular systems show cell adhesion, communication and interdependence
- discover that multicellular organisms have emergent properties and differentiation of cells involves the expression of specific genes
- understand that complex multicellular organisms have cells organised in tissues
- recognise that differentiation involves programmed cell death and cell division
- recall that stem cells retain the ability to differentiate into many other cells
- recognise the location of stem cells in adult humans
- understand the differences between totipotent, pluripotent and multipotent stem cells
- recall that different tissues work together to become organs and that organs are organised to form organ systems
- discover that multicellularity has problems and advantages for organisms

- recognise that stem cells persist in animal bodies to replace cells.

GUIDING QUESTIONS

- How are multicellular organisms organised?
- What are the advantages of multicellularity?
- How do cells specialise in a multicellular organism?

8.1.1 Multicellular organisms

Multicellular organisms are composed of more than one cell. All multicellular organisms, even the simplest, such as *Hydra* (Figure 8.1.1), have cells that are held together by cell adhesions that form connections between adjacent cells. Cells are linked together either by special proteins that are part of the cytoskeleton (Section 5.2) or through tight junctions or desmosomes, both of which are formed by proteins embedded in the plasma membranes of adjacent cells. The connections made by cell adhesion are important in enabling cells to communicate and regulate cell activities, as well as in the development and maintenance of tissues. Cell communication in a multicellular organism means that cells are interdependent and enables different cells to carry out different specialised tasks.

Hydra is a small pond-dwelling organism that has different cells that form a two-layered body cavity, tentacles, stinging cells (called nematocysts) and a nerve net.

Cells in a multicellular organism form a system that carries out all the functions that are needed for life: metabolism, reproduction, response to the environment, nutrition, homeostasis, excretion and growth (Section 5.1).

KEY POINT

cell adhesion the process by which cells form connections with each other.

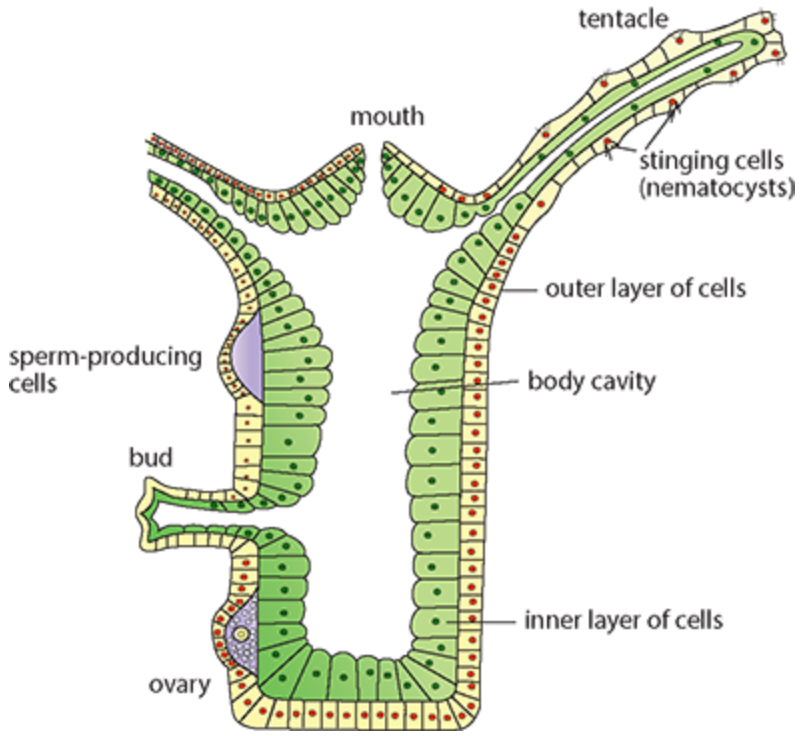


Figure 8.1.1: *Hydra* is a small aquatic multicellular animal.

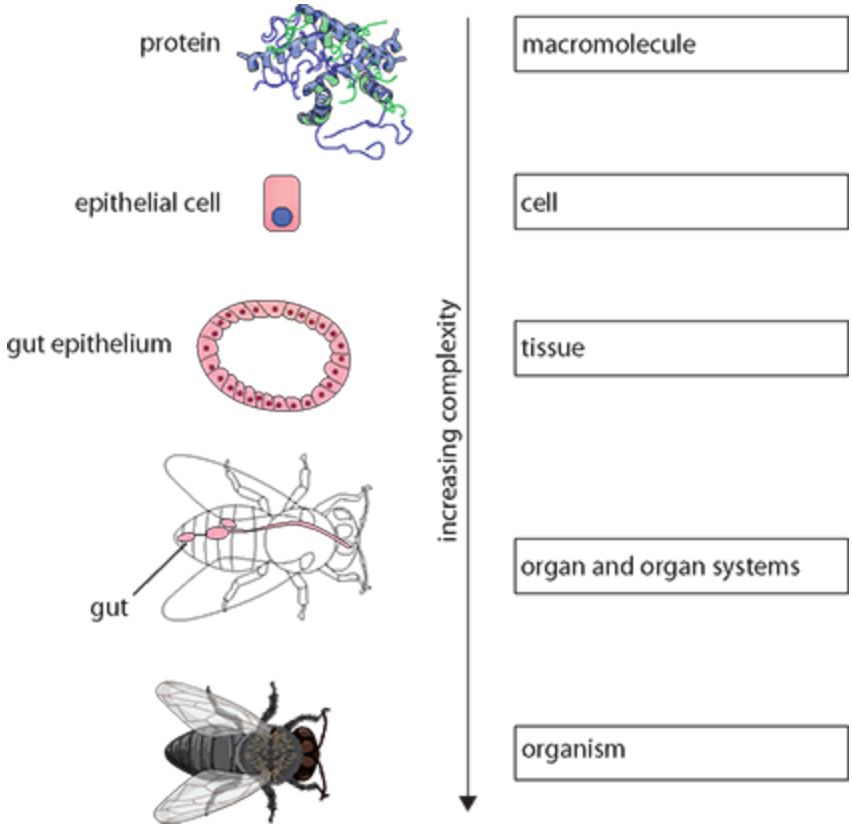


Figure 8.1.2: Multicellular organisms have properties that emerge from the interaction of their components.

Multicellular organisms like hydra and more complex plants and animals have emergent properties that arise from the interaction of their cells. But we cannot predict what these new properties will be from studying the individual component parts of the organism. Even though we can separate the parts of a living cell into the macromolecules that build it, and we have information about macromolecules, we could not have predicted the characteristics that emerge in a cell when the macromolecules are combined.

The cells built of macromolecules are organised into **tissues**, which are groups of similar cells with the same function that work together. Tissues form **organs**, which work with other organs to form organ systems, such as the heart and circulatory systems. A complex multicellular organism will contain many organ systems with different jobs to do in the body (Figure 8.1.2).

8.1.2 Differentiation

How do cells in the same organism behave in different ways when they all arose from the same parent cell and so have the same genome? In our own bodies, nerve cells and muscle cells all have the same sets of genes but look and behave very differently. The logical answer is that in some cells some genes are expressed that are not expressed in other cells, and that each type of cell expresses a slightly different set of genes. For example, a human pancreatic cell will express genes for the production of digestive enzymes or insulin, but a skin cell will not. **Differentiation** of cells into different types involves the expression of some genes from the organism's genome in a cell, while others are not expressed.

KEY POINT

programmed cell death (apoptosis) removal of cells during development to eliminated unwanted cells.

As well as gene expression, programmed cell death (apoptosis) plays an important role in forming organs and structures during embryo development. Apoptosis refers to changes that occur inside cells when they are no longer needed. These changes cause them to die or cause their own death. Apoptosis is called programmed cell death because it occurs at a specific time in the growth or development of an organism. One example of the role of apoptosis can be seen in the formation of the human hands as a baby grows during fetal development. Cells, which are present between individual fingers in the early stages of development, are removed through apoptosis so that the structure of the hand can be formed. Another example can be seen in frogs and toads

that develop from tadpoles. Cells in the tail of a tadpole are removed by apoptosis as the tadpole becomes an adult. See Figure 8.1.3.

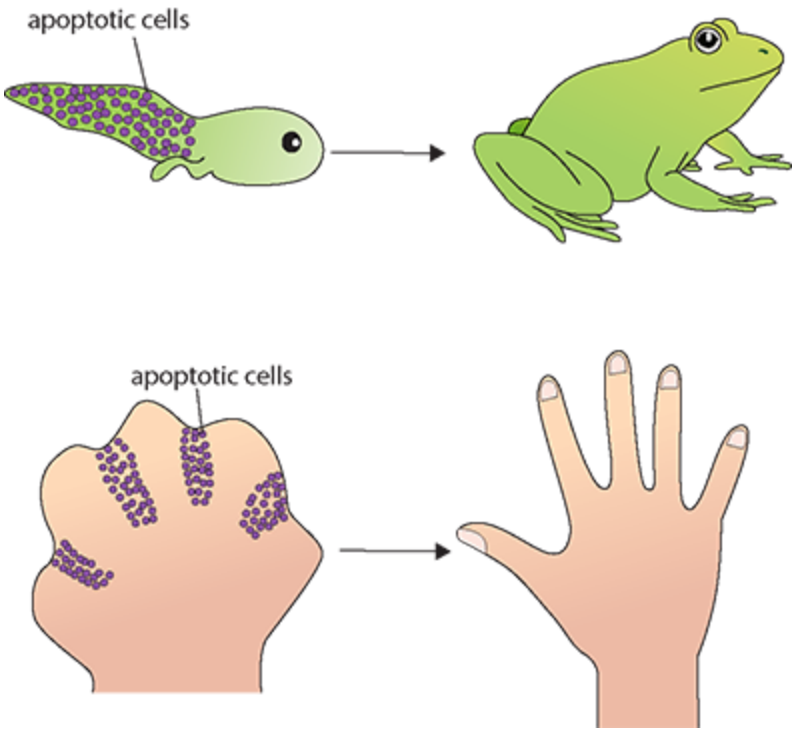


Figure 8.1.3: Apoptosis contributes to development in multicellular organisms as specific cells are programmed to die.

8.1.3 Stem cells

A stem cell is defined as any cell type that has two important properties. Firstly, it can divide and renew itself endlessly to make more cells with the same properties and, secondly, it can differentiate to make other cell types, with other biological functions.

Stem cells differ from most other cells in the following ways:

- They are unspecialised.
- They can divide repeatedly to make large numbers of new cells.
- They can differentiate into several types of cell.
- They have a large nucleus relative to the volume of the cytoplasm.

But not all stem cells have the same **potency**; that is the ability to give rise to similar cell types. Stem cells may be totipotent, pluripotent or multipotent. Stem cells found in bone marrow are an example of **multipotent** cells. They can produce many kinds of cells, but only types of blood cell. (Figure 8.1.4) In contrast, **pluripotent** stem cells can make multiple types of cell from all three layers of an embryo (known as embryonic germ layers and shown in Fig 8.1.5) but not form extra-embryonic tissue that becomes the amnion. Finally, **totipotent** stem cells can make all three embryonic germ layers and the extra-embryonic tissue. The only truly totipotent cell is the zygote (fertilised ovum).

KEY POINTS

stem cells retain the ability to develop into specialised cells to replace cells that are damaged or diseased.

totipotent cells have the potential to divide until an entire organism is formed

pluripotent cells can divide into most cell types in an organism but cannot form an entire organism

multipotent cells are able to self-renew by dividing and develop into several specialised cell types in a specific tissue or organ. Most adult stem cells are multipotent stem cells

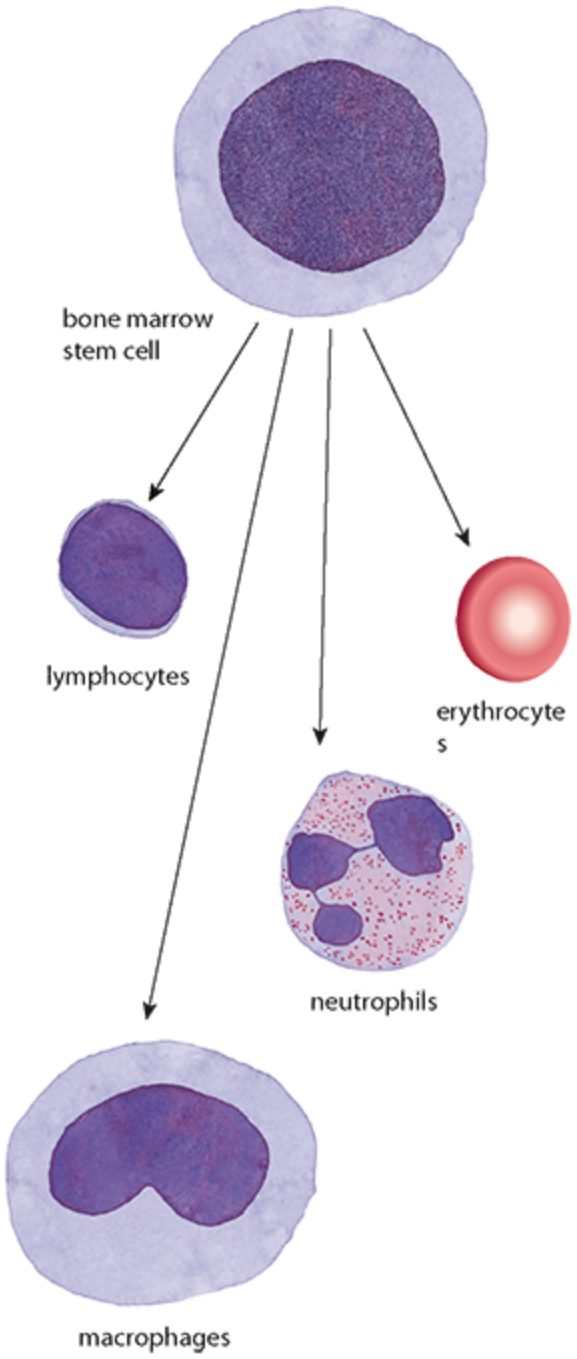


Figure 8.1.4: Bone marrow cells differentiate into the different types of blood cell.

Stem cells are found in embryos but also persist in adults. Examples of adult stem cells are found in the bone marrow and hair follicles, both are multipotent. Bone marrow stem cells only

produce cells which will become part of the immune system and red blood cells while hair follicle stem cells have the capacity and ability to generate various tissues. They are found in an area called a niche close to the base of a hair follicle sebaceous gland. They generate all cell types of the skin, including keratinocytes, that make up the structure of the skin and hair.

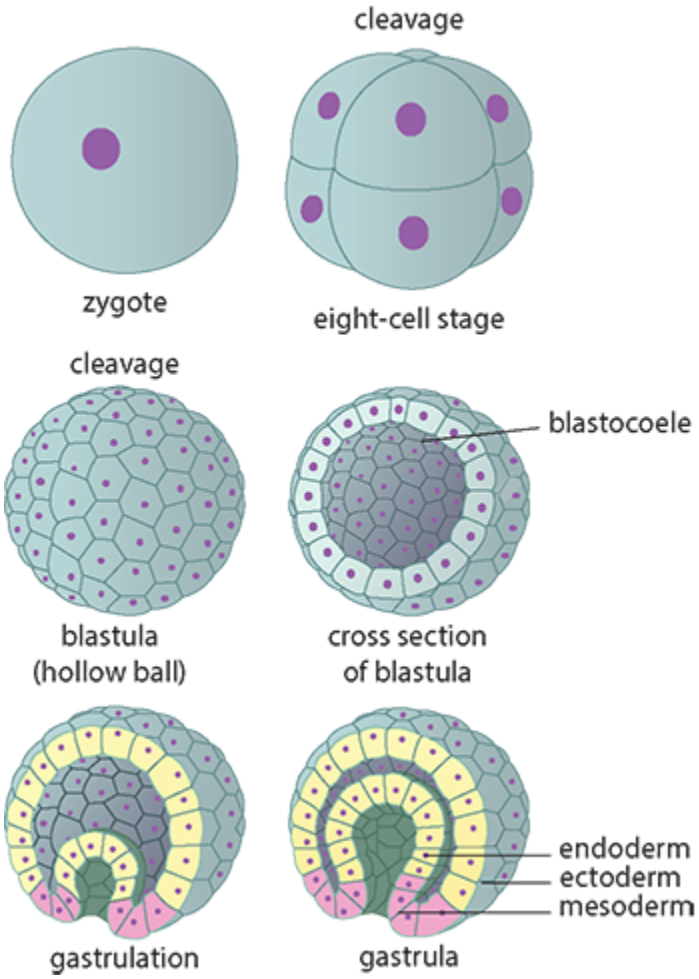


Figure 8.1.5: Development of the three layers in an embryo.

Early cell specialisation

The fertilised egg of any organism contains all the information needed to enable that single cell to develop into a complex organism consisting of many different types of cell. The genes

inherited from the maternal and paternal DNA carry this information. A fertilised egg which is totipotent divides rapidly and produces a ball of cells called a blastocyst in which all the cells are alike (Figure 8.1.5). Gradually the cells in the blastocyst differentiate and begin to develop to form specialised tissues such as muscle or liver. The process of differentiation produces cells for specific purposes: muscle cells for contraction, liver cells for metabolism of toxins, and so on. Once differentiation has happened, it cannot be reversed. Cells in the blastocyst have the potential to turn into a great many different cell types: they are said to be **pluripotent** and are known as embryonic stem cells. Pluripotent stem cells can divide into most, or all, cell types but cannot develop into an entire organism on their own.

In human development, the first eight weeks after fertilisation of an egg are when stem cells in the blastocyst grow and differentiate. During week three, the process of gastrulation occurs, which folds layers of cells until there are three distinct cell layers: the mesoderm, endoderm, and ectoderm. These are layers are known as the primary germ cell layers from which organs are formed. The position of a cell will determine which genes are expressed and what each cell will become.

- The endoderm forms organs of the intestine and respiratory systems, as well as the thymus, parathyroid, bladder, and urethra.
- The ectoderm forms the skin, the nervous system, and parts of sensory organs.
- The mesoderm forms the circulatory system and blood, lymphatic system, bone, cartilage, muscles, and many internal organs such as the kidneys.

After eight weeks of development, organ systems have developed and by week nine, the embryo becomes known as a fetus. Growth and differentiation continue until birth.

As the embryo continues to develop, individual cells continue to differentiate from what were pluripotent **embryonic stem cells**. Cells are guided towards particular developmental pathways, creating different cell types. This is possible because the cells can control which genes are expressed and translated into proteins. Different signals cause embryonic cells to select specific parts of their DNA that are used to synthesise the proteins needed to form different cell types. This differentiation is brought about by factors inside the cells as well as factors that act on the cell from the outside. The exact molecular interactions that control cellular differentiation in the developing embryo are still not fully understood.

You can read more about the development of the human embryo and fetus in Section 8.4.

Plants also contain stem cells. These are found in the meristems just behind the tips of growing roots and stems. These cells can only differentiate to become various tissues of the root and stem.

TEST YOUR UNDERSTANDING

1 Define the terms tissue and organ.

2 Outline the special properties of stem cells.

3 Describe how cells differentiate to carry out different functions.

4 What is the function of stem cells in the human body?

SCIENCE IN CONTEXT

Therapeutic use of stem cells

Another source of stem cells, which has been successfully used in medical treatments, is the blood in the umbilical cord of a newborn baby (Figure 8.1.6). These stem cells can divide and become any type of blood cell. Cord blood can be used to treat certain types of leukemia, a cancer that causes overproduction of white blood cells in the bone marrow. Cells from the cord blood are collected and their tissue type is determined. After chemotherapy to destroy the patient's own bone marrow cells, stem cells that are the correct match to the patient's tissue, are given by transfusion. If the treatment is successful, the stem cells become established in the person's bone marrow and start producing blood cells as normal.

This treatment can work well in young children, but there are not enough cells in a single cord to meet the needs of an adult patient. Scientists have been looking for ways to either combine the cells from more than one baby, or to use laboratory techniques to increase the number of cells. Allowing the stem cells to divide in the laboratory produces many blood cells, but not more stem cells. In 2010, scientists in Seattle, USA, managed to alter a signalling pathway in the stem cells so that the cells could increase in number without losing their stem cell properties. As a result of this process, known as therapeutic cloning, umbilical cord blood may prove to be an even more valuable source of stem cells in the future.

Figure 8.1.6: This technician is collecting blood from an umbilical cord. This blood is a rich source of stem cells.

Transplantation of stem cells is a new technology. So far it has also been successfully used in the treatment of Type I diabetes, and research is continuing into therapies to treat a range of conditions involving neurological damage, such as multiple sclerosis and Alzheimer's disease.

NATURE OF SCIENCE

Ethics and new medical treatments

Although scientists from many countries have cooperated on stem cell research, the laws governing research vary from place to place. In the European Union, research using human embryos is permitted in some countries but is illegal in others such as Germany, Ireland, Italy, Portugal and Austria. In the USA some states fund the research, while others ban it. Australia permits research and New Zealand restricts it. Laws also differ in their application to embryonic stem cell research and in stem cells taken from adult tissues. Because laws depend on political and religious viewpoints, they may change as new information and research develops.

REFLECTION

Could I teach this topic to someone else? Which are the most difficult concepts to communicate?

Links

- Why are prokaryotic cells in a biofilm not considered to be multicellular? (Chapter 7)
- How does epigenetics contribute to differentiation? (Chapter 4)
- How can cells in a multicellular organism specialise when they all contain the same genes? (Chapter 4.2)

8.2 Transport in animals and plants

LEARNING OBJECTIVES

In this section you will:

- learn that circulatory systems are mass flow systems for transport, communication and defence
- learn that the heart, arteries, veins and capillaries make up a circulatory system
- understand that the circulatory system adapts to the changing needs of the body by alterations in blood pressure and heart rate
- compare the structures of arteries, veins and capillaries and relate the differences to their functions
- recall that materials are exchanged between cells and blood in capillaries
- understand that water is transported from roots to leaves during transpiration
- recognise the adaptations of the xylem
- draw a diagram of a section of a plant stem

> Describe and explain stages of the cardiac cycle including interpretation of graphs

- compare the single circulatory system of fish with the double circulation of mammals
- understand the function of the lymphatic system
- understand that blood flow to different parts of the body varies with its needs and activity levels
- understand how root pressure causes water movement
- recognise the adaptation of phloem for the transport of substances up a stem.

GUIDING QUESTIONS

- How are organs adapted to their functions?
- How do the different organs in an organ system work together?

Multicellular organisms need to co-ordinate the different parts of their bodies so that they work together and enable the organs to perform their functions. Cells work with other cells to produce tissues and organs and organs are integrated to form effective body systems.

In complex organisms the nervous system and endocrine system send nerve and hormonal messages to ensure effective co-ordination (Chapter 7). In our own bodies, the blood system transports materials and chemical messages between vital organs while in plants the xylem and phloem perform similar integrating functions.

8.2.1 Circulatory systems

Circulatory systems are needed in multicellular animals to ensure that all cells receive the oxygen and nutrients they require and so that waste can be removed and disposed of. These circulatory systems are **mass flow** systems, which contain blood or other fluids that transport materials. They enable different sections of the body to communicate with one another, for example through hormones that travel in the blood. Blood flows down a pressure gradient with the highest blood pressure being produced as blood is expelled from the heart. Blood also carries cells and antibodies that defend the body from infection (Chapter 10) and it has an important role in homeostasis (Section 8.5).

KEY POINTS

circulatory system refers to an organ system that enables blood to flow round the body and transport nutrients, oxygen, carbon dioxide and hormones.

mass flow is the movement of fluids down a pressure gradient, in the case of the circulatory system, flow from the heart to other parts of the body.

Different sorts of circulatory system

Different organisms have circulatory systems that have evolved to suit to their needs. Unicellular organisms exchange materials through their cell surfaces but all multicellular organisms have some form of transport and circulatory system to deliver oxygen and food and to remove metabolic waste. Very simple organisms such as marine sponges use the seawater that is drawn in and out of their bodies for transport. The cnidarian *Hydra* exchanges

materials by diffusion through its body cells. But these simple systems cannot supply the needs of large animals and so the organisms that use them are limited in size.

More complex organisms such as arthropods and molluscs have open circulatory systems with no blood vessels. They have a simple heart that pumps a fluid known as hemolymph, which is similar to blood, around their bodies. Hemolymph enters blood spaces called hemocoels, which surround body tissues to exchange materials. From here hemolymph must diffuse back to the heart. Flow in these simple **open systems** is slow. Vertebrates have larger and more complex bodies and they have evolved **closed circulatory systems** with blood that is kept inside blood vessels.

The heart

In the human circulatory system, blood is kept on the move by the pumping action of the powerful heart muscle. It has been estimated that a normal human heart beats more than 2.5×10^9 times in a lifetime, sending a total of more than 1.5 million dm^3 of blood from each ventricle.

A human heart is about the size of a clenched fist. It is a double pump with two separate sides (Figure 8.2.1). The right-hand side receives deoxygenated blood from all over the body and pumps it to the lungs via the pulmonary artery to pick up more oxygen. The left-hand side receives oxygenated blood from the lungs via the pulmonary vein and pumps it to cells all over the body where the oxygen is unloaded. This arrangement means that humans, like all mammals, have a double circulation: a pulmonary circulation between the heart and lungs and a larger circulation that carries blood from the heart to the rest of the body and back

again (Figure 8.2.7). On any complete journey round the body, blood passes through the heart twice.

Type of organism	**Internal body structure**	**Circulatory system**	**Organ that controls blood flow**
cnidarians (*Hydra*, jellyfish and sea anemones) and unsegmented flatworms (e.g. *Planaria*)	one internal cavity that is used for both digestion and circulation	no blood system	none
molluscs and arthropods	digestive and circulatory systems are separate	most have open systems containing hemolymph that is in direct contact with the cells of the body	simple heart
annelid (segmented) worms (e.g. earthworms) and vertebrates	digestive and circulatory systems are separate	closed system; blood is always enclosed inside vessels	annelids: specialised blood vessels that can contract vertebrates: heart with more than

			one chamber

Table 8.2.7: As organisms have evolved increasing complexity in their body structure, the need for an efficient circulatory system has increased. Simple organisms have no circulatory system, small multicellular organisms have open systems and larger more complex organisms have evolved an enclosed circulatory system.

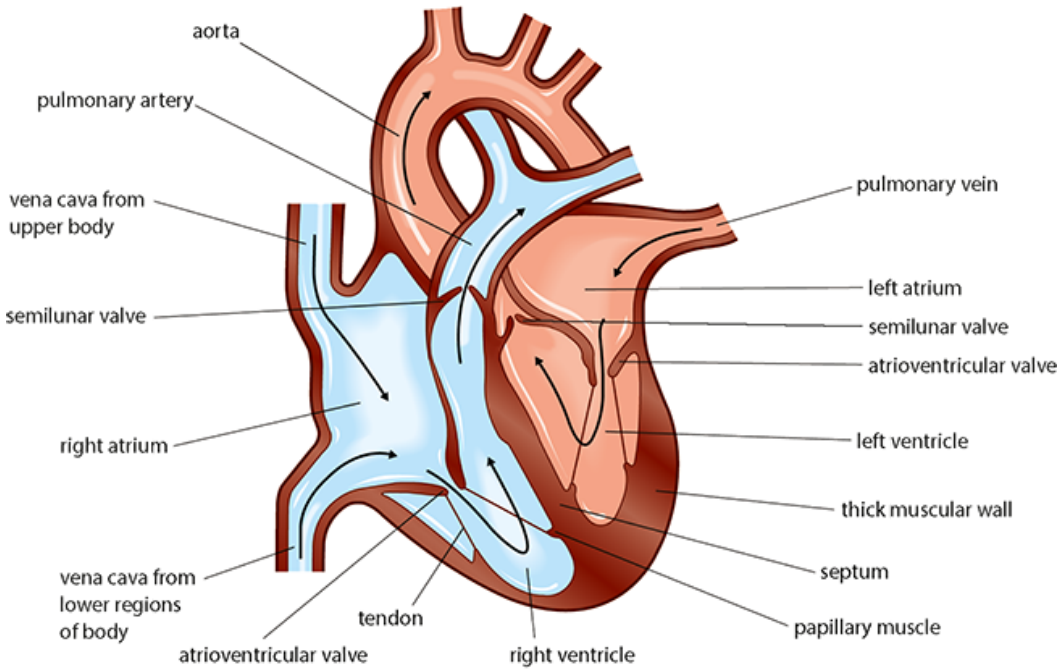


Figure 8.2.1: Diagram of the human heart, in longitudinal section, showing the direction of blood flow.

The heart has four chambers: two smaller atria (singular **atrium**) at the top and two larger ventricles below. The right-hand and left-hand sides are completely separated from one another. Atria have thin walls as the blood they receive from the veins is under relatively low pressure. Ventricles are stronger and more muscular as their job is to pump blood out of the heart. Both

ventricles hold the same volume of blood, but the left ventricle wall is thicker than the right as it must generate enough pressure to pump blood all round the body. The right ventricle pumps blood a much shorter distance to the lungs.

Atria are separated from ventricles by atrioventricular valves, which prevent the blood flowing backwards into the atria. A second set of valves in the aorta and pulmonary arteries – the semilunar valves – prevent backflow into the ventricles as they relax after a contraction.

KEY POINT

blood pressure the pressure of circulating blood against the walls of blood vessels.

Heart muscle works continuously, beating about 75 times per minute when a person is resting, and so it has a large demand for oxygen. Coronary arteries extend over the surface of the heart and penetrate deep into the muscle fibres to supply oxygen and nutrients for this constant activity (Figure 8.2.2).

The pulse rate, measured at the wrist or neck, can give an indication of heart rate in beats per minute (Figure 8.2.3). Electronic devices that measure blood pressure (Figure 8.2.4) also provide a pulse rate reading.

Blood pressure

As the heart pumps harder and faster, blood pressure also rises. Blood pressure is the pressure of circulating blood against the walls of blood vessels, which is produced by the heart pumping. In medicine ‘blood pressure’ usually refers to blood pressure in the large arteries and it is measured using a sphygmomanometer.

Older devices to measure blood pressure contained mercury and blood pressure is still recorded in millimetres of mercury (mmHg) even though modern devices do not contain mercury (Figure 8.2.4).

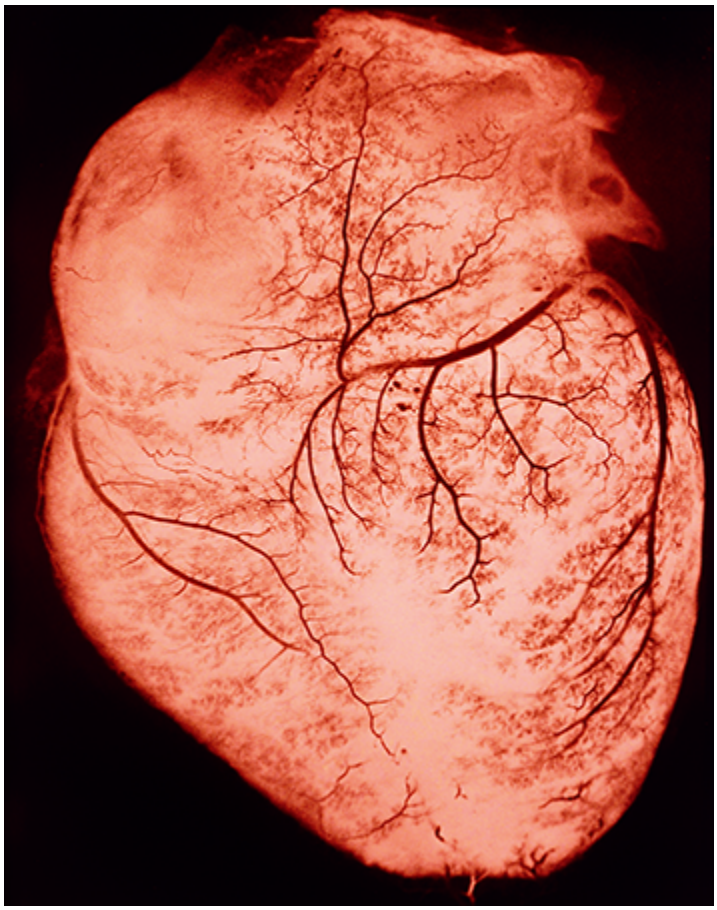


Figure 8.2.2: A human heart. Clearly visible are the coronary arteries, which supply oxygen to the heart muscle.

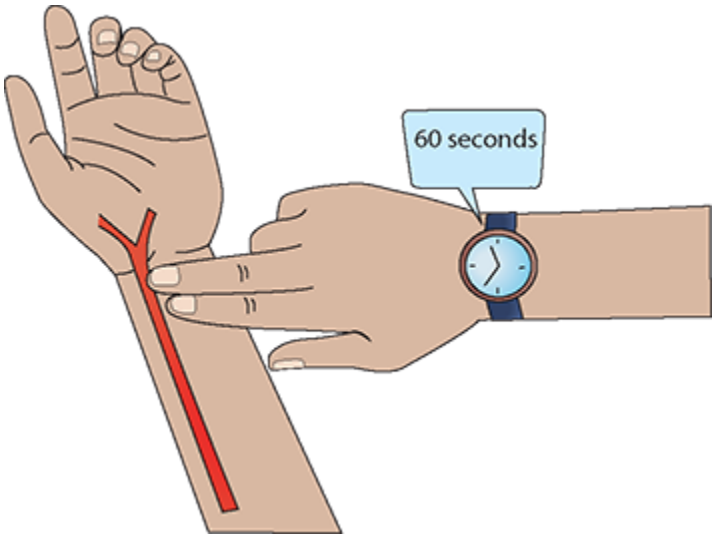


Figure 8.2.3: Pulse measurements are taken at small arteries in the wrist or neck.

A blood pressure reading consists of two parts. **Diastolic pressure** is the pressure on the blood vessels when the heart muscle relaxes. The diastolic pressure is always lower than the systolic pressure, which is recorded as the ventricles contract.

Figure 8.2.4: Taking a pulse and blood pressure reading.

Normal blood pressure for a healthy adult at rest should be under 140 mmHg for systolic pressure and under 90 mmHg for diastolic pressure.

KEY POINT

systolic pressure pressure generated in the arteries as the heart contracts.

EXAM TIP

You should be able to interpret measurements of pulse and blood pressure readings.

Blood vessels

Arteries are blood vessels that carry blood away from the ventricles of the heart. They branch and divide many times, forming small arteries called **arterioles** and eventually the tiny **capillaries** that reach all our tissues. Arteries have thick outer walls of collagen and elastic fibres (Figure 8.2.5), which withstand high blood pressure and prevent vessels becoming overstretched or bursting. Just beneath the outer covering is a ring of circular smooth muscle that contracts with each heart beat to maintain blood pressure and keep blood moving along. The central space inside an artery, called the lumen, is narrow to keep blood pressure high. The lumen's lining of smooth epithelial cells reduces friction and keeps blood flowing smoothly. Each heartbeat sends a pulse of blood through the arteries, which expand slightly and then recoil. This produces the pulse that we can monitor at the wrist or neck.

NATURE OF SCIENCE

'Margin of error' is a term that is used to describe the likelihood and extent of error in quantities that we measure in experiments. Every time we repeat a reading with an instrument such as a thermometer or timing device such as a stopwatch, we may get a slightly different result. To reduce errors like this, it is important to take a series of readings and calculate an average.

If we use statistics in calculations, errors are said to be of two types:

- Systematic error or bias always occurs with the same value when we use an instrument in the same way. This can be minimised by using the correct measuring instrument and the same technique each time we take a reading. Examples include using a stopwatch not an alarm clock for timing something, and holding a thermometer in a liquid for exactly the same period of time in an experiment.
- Random errors may vary from reading to reading and are due to factors that we cannot control.

When you carry out practical work you will be expected to show the margin of error in your readings, using either error bars on a graph or ± in your table of data. This shows you understand the degree of error that your data may contain.

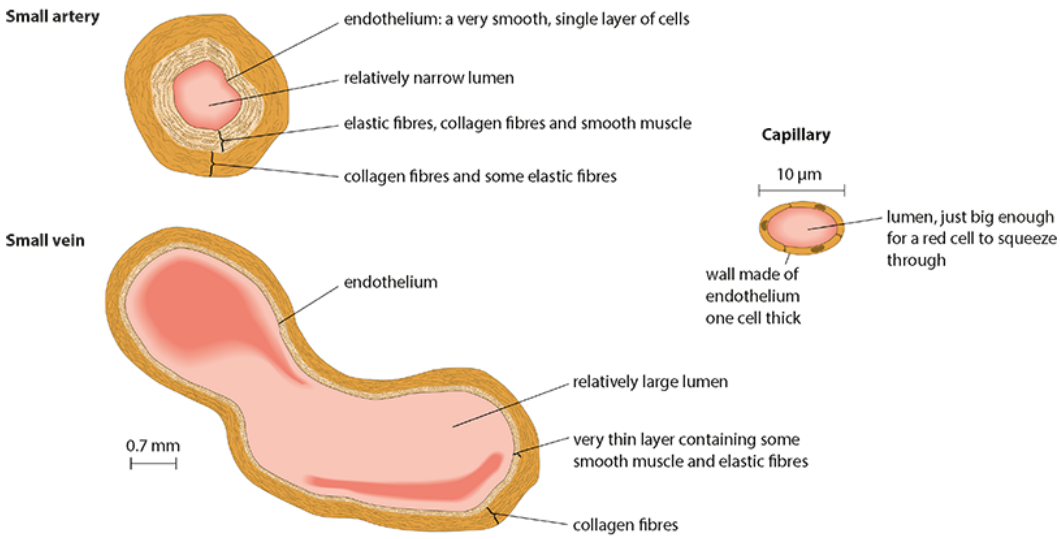


Figure 8.2.5: Diagrams of transverse sections of an artery, a vein and a capillary.

Capillaries are the smallest vessels. The lumen of a capillary is only about 10 µm in diameter and some are so small that red blood cells must fold up in order to pass along. Networks of these tiny capillaries reach almost every cell in the body. Blood flow here is very slow, at less than 1 mm per second, and capillary walls are only one cell thick so the distance for diffusion of materials in and out of them is as small as possible. Some capillary walls have spaces between their cells, enabling plasma and phagocytes (white blood cells) to leak out into the tissues.

Veins carry blood back towards the atria of the heart from body tissues. Small veins called **venules** join up to form large veins, which can be distinguished from arteries by their much thinner walls, which contain few elastic and muscle fibres. Blood inside a vein is not under high pressure and does not pulse along and the lumen is large to hold the slow-moving flow. The relatively thin walls can be compressed by adjacent muscles and this helps to squeeze blood along and keep it moving. Many veins contain

valves to prevent blood flowing backwards, a problem that can arise if flow is sluggish.

Table 8.2.1 summarises some differences and similarities between the three types of blood vessel.

Artery	Vein	Capillary
thick walls	thin walls	walls one cell thick
no valves (except in aorta and pulmonary artery)	valves sometimes present	no valves
blood pressure high	blood pressure low	blood pressure low
carries blood from the heart	carries blood to the heart	links small arteries to small veins

Table 8.2.1: Comparing arteries, veins and capillaries.

Coronary heart disease

Three large coronary arteries branch from the aorta and supply heart muscle with oxygen-rich blood (Figure 8.2.2). If any of these arteries is blocked, an area of the heart will receive less oxygen and cells in that region may stop contracting or even die. A blockage in a coronary artery is known as an **occlusion** and can lead to a heart attack.

One serious cause of coronary heart disease (CHD) is **atherosclerosis**, a slow degeneration of the arteries caused by a build-up of material known as **plaque** inside them. Plaque becomes attached to the smooth endothelium lining where it can

accumulate. Over time, the diameter of the artery becomes restricted so that blood cannot flow along it properly, and it loses elasticity (Figure 8.2.6). As the rate of flow slows down, blood may clot in the artery, further restricting the movement of blood along it. Clots may also break free and travel to block another smaller artery elsewhere in the body. If this artery is in the brain, the clot may cause a stroke.

Feedback control of heart rate

Heart rate, the number of contractions per minute, must be controlled to ensure the body receives the appropriate amounts of nutrients and oxygen. The amount of blood that leaves the left ventricle per minute is called the **cardiac output**. It is determined by the heart rate, number of beats per minute and the stroke volume, the volume of blood that is pumped out with each contraction. Heart rate is regulated by the cardiovascular control centre in the medulla oblongata in the brain stem and is controlled by the sino atrial node (SAN) or pacemaker in the left atrium. Two types of receptors, **baroreceptors** (pressure receptors) and **chemoreceptors** (chemical receptors) are responsible for detecting stimuli in the blood and signalling to the medulla oblongata to adjust our heart rate. Baroreceptors detect changes in blood pressure and are found in the **aortic** and **carotid bodies**. Aortic bodies are located in the aortic arch and carotid bodies in the carotid arteries in the neck. Chemoreceptors detect the concentration of oxygen in the blood and are also sensitive to changes in pH caused by the carbon dioxide dissolved in the blood. Chemoreceptors are also located in the aortic and carotid bodies.

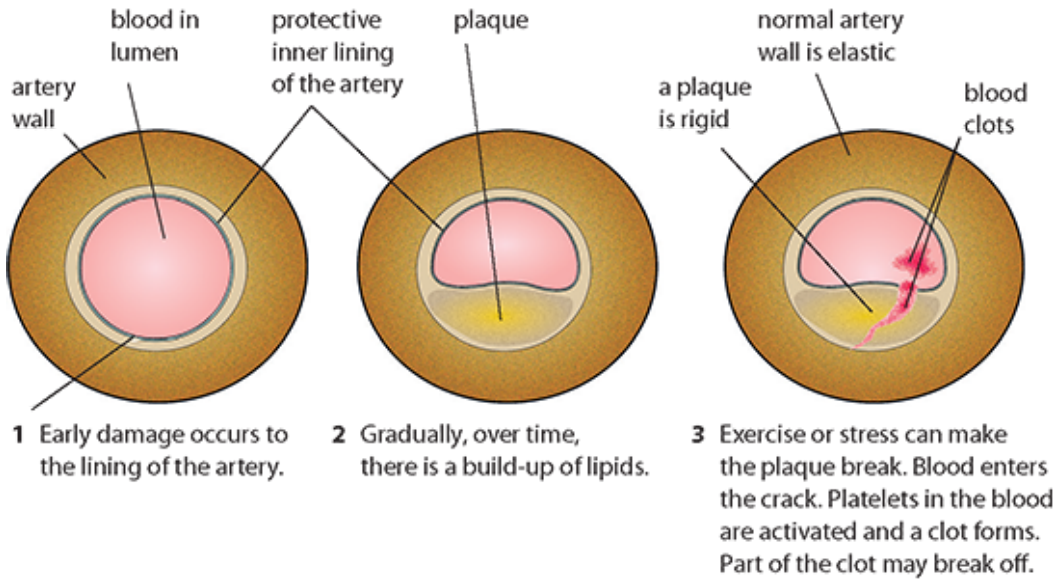


Figure 8.2.6: The development of atherosclerosis.

The medulla sends impulses via the sympathetic or parasympathetic neurons which stimulate the heart pacemaker (SAN) to slow down or speed up the heart rate. If blood pressure or oxygen levels are low the sympathetic nervous system increases heart rate by releasing noradrenaline which binds to the pacemaker and increases heart rate. If blood pressure or oxygen levels are high, the parasympathetic nervous system releases acetylcholine which slows down the heart rate.

You can read more about how blood pressure and heart rate are controlled in Section 8.5.

8.2.2 Single and double circulations

Bony fish are vertebrates that have a single circulation. This means that blood flows through the heart only once on a complete journey around the fish's body. Humans and all mammals, have a double circulation so that blood passes through the heart twice on any journey around the body (Figure 8.2.7).

Fish have a lower metabolic rate than mammals. So, in fish a single circulation is able to deliver sufficient oxygen to their tissues at a fairly low blood pressure. Mammals have a high metabolic rate and, unlike fish, are able to maintain body temperatures at the optimum rate for enzyme activity. This is only possible because of their double circulation that keeps their blood pressure high enough to deliver oxygen to the tissues that need it.

The main function of a circulatory system is to link exchange surfaces such as the alveoli in the lungs, where oxygen enters the body, with the muscles that need the oxygen to work. You can read more about exchange surfaces in Section 8.3. Figure 8.2.7 shows how the gills of a fish and lungs of a human are linked to the body cells by the circulatory system. Oxygen and other substances transported in the blood move across exchange surfaces by diffusion. Oxygen diffuses into the blood in the gills and lungs and diffuses out in the body tissues that require it. Blood must keep flowing past the different surfaces to maintain a concentration gradient. As blood carries oxygen away and more deoxygenated blood arrives, the difference in oxygen concentration between the alveoli or gills and the blood is kept at a suitable level to ensure that diffusion is efficient (you can read more about concentration gradients in Section 6.2). The movement of blood is maintained by the heart, which pumps to

keep blood moving in a continual one-way flow through the body The three essential components of a closed circulatory system are:

1. blood, a transport fluid
2. the heart, a pump to maintain flow and pressure gradient
3. vessels to contain the blood.

Adaptations of the human heart to deliver pressurised blood to the arteries

The heart has four chambers – two smaller **atria** (singular **atrium**) at the top and two larger **ventricles** below. The right- and left-hand sides are completely separated from one another. Atria have thin walls as the blood they receive from the veins is under relatively low pressure. Ventricles are stronger and more muscular as their job is to pump blood out of the heart. Both ventricles hold the same volume of blood but the left ventricle wall is thicker than the right as it must generate enough pressure to pump blood all round the body. The right ventricle pumps blood a much shorter distance to the lungs.

Humans have a double circulation: a pulmonary circulation between the heart and lungs and a circulation which carried blood from the heart to the rest of the body and back again. The pulmonary artery carries deoxygenated blood from the right ventricle of the heart to the lungs and the pulmonary vein returns oxygenated blood to the left atrium.

Atria are separated from ventricles by **atrioventricular valves**, which prevent the blood flowing backwards into the atria. A second set of valves in the aorta and pulmonary arteries – the

semilunar valves – prevent backflow into the ventricles as they relax after a contraction.

Heart muscle works continuously, so it has a large demand for oxygen. Coronary arteries extend over the surface of the heart and penetrate deep into the muscle fibres to supply oxygen and nutrients for this unremitting activity (Figure 8.2.8).

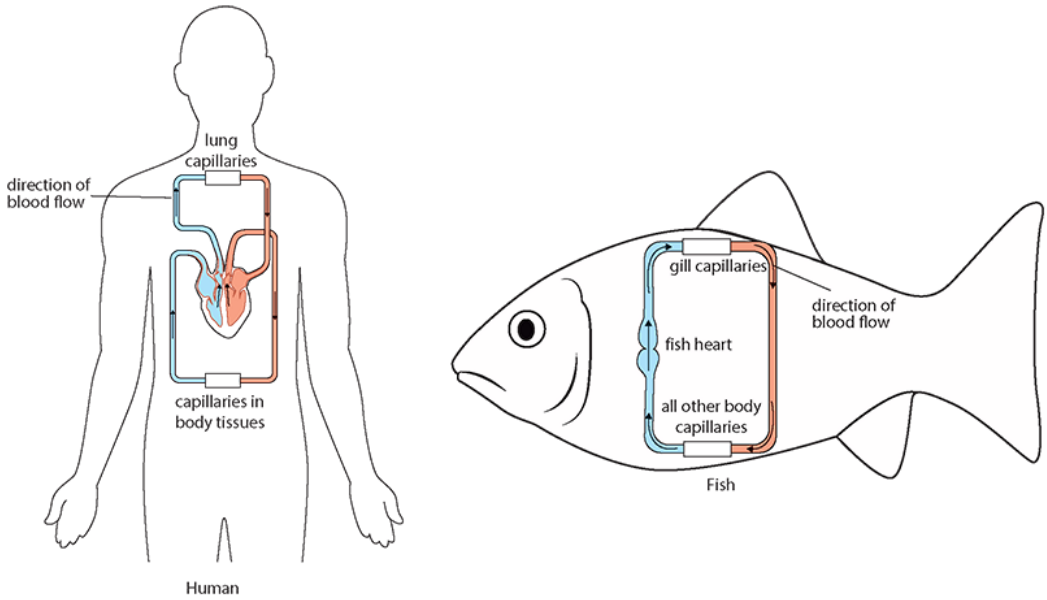


Figure 8.2.7: Single and double circulations in a fish and a human.

Control of blood flow and heart beat

Heart tissue is made of a special type of muscle that is different from other muscles in our bodies. **Cardiac muscle** is unique because it contracts and relaxes without stimulation from the nervous system. It is said to be myogenic. Natural **myogenic** contractions are initiated at an inbuilt pacemaker, which keeps cardiac muscle working in a coordinated, controlled sequence. The pacemaker, or **sinoatrial node** (SAN) is a special region of muscle cells in the right atrium that sets the basic pace of the

heart. The rate set by the SAN is also influenced by stimulation from the nervous system and by hormones.

The natural rhythm of the pacemaker is modulated by the nervous system so that the heart rate is adjusted to our activity levels. It speeds up when we are exercising and need extra oxygen and nutrients and slows down as we sleep. Changes to our heart rate are not under our conscious control but result from impulses sent from a control centre in the part of the brain stem known as the medulla. Impulses to speed up the heart pass along the sympathetic nerve, which stimulates the pacemaker to increase its rate. Impulses sent along the parasympathetic (vagus) nerve cause the heart rate to slow down. The medulla monitors blood pressure and carbon dioxide levels using information it receives from receptors in arteries (Section 8.5).

Emotions such as stress, as well as increases in activity level, can cause an increase in heart rate. During periods of excitement, fear or stress the adrenal glands release the hormone **epinephrine** (adrenaline), which travels in the blood to the pacemaker and stimulates it to increase the heart rate.

The cardiac cycle and circulation

In the human circulatory system, blood is kept on the move by the pumping action of the powerful heart muscle. It has been estimated that a normal human heart beats more than 2.5×10^9 times in a lifetime, sending a total of more than 1.5 million litres of blood from each ventricle.

The **cardiac cycle** is the sequence of events that takes place during one heart beat (Figure 8.2.9). As the heart's chambers contract, blood inside them is forced on its way. Valves in the heart and arteries stop the blood flowing backwards.

The pressure and volume of blood in each of the chambers of the heart change during the cardiac cycle. Fig 8.2.10 shows these changes for one complete cycle.

Control of the heart beat

At the start of every heart beat, the SAN produces an impulse that stimulates both atria to contract. A second structure, the **atrioventricular node** (AVN) at the base of the right atrium, is also stimulated. It delays the impulse briefly until the atrial contraction finishes and then transmits it on down a bundle of modified muscle fibres – the bundle of His and Purkinje fibres – to the base of the ventricles. Impulses radiate up through the ventricles, which contract simultaneously about 0.1 seconds after the atria. (Fig 8.2.9)

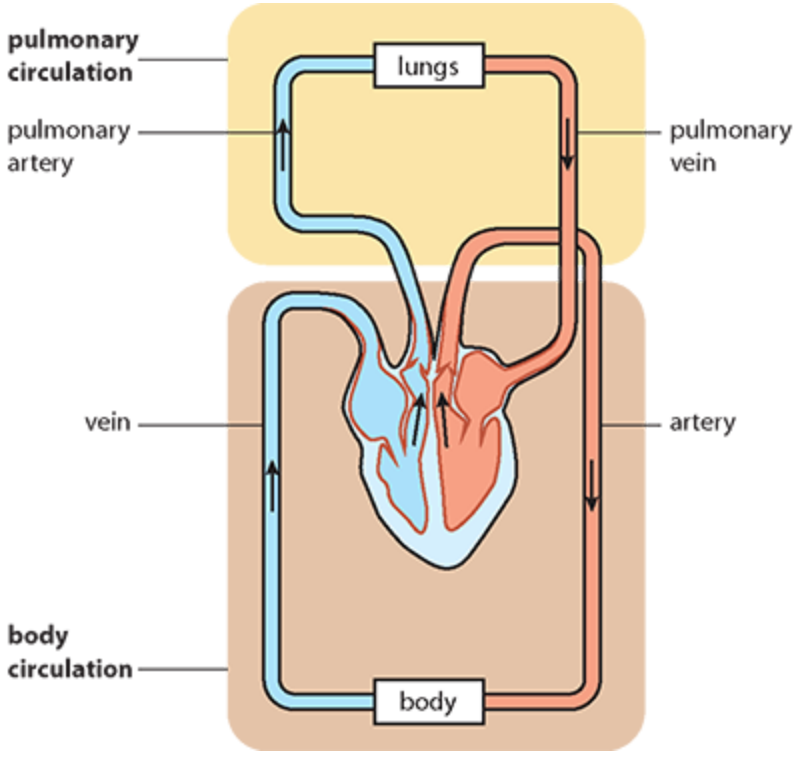


Figure 8.2.8: Diagram to show the double circulation of blood through the heart.

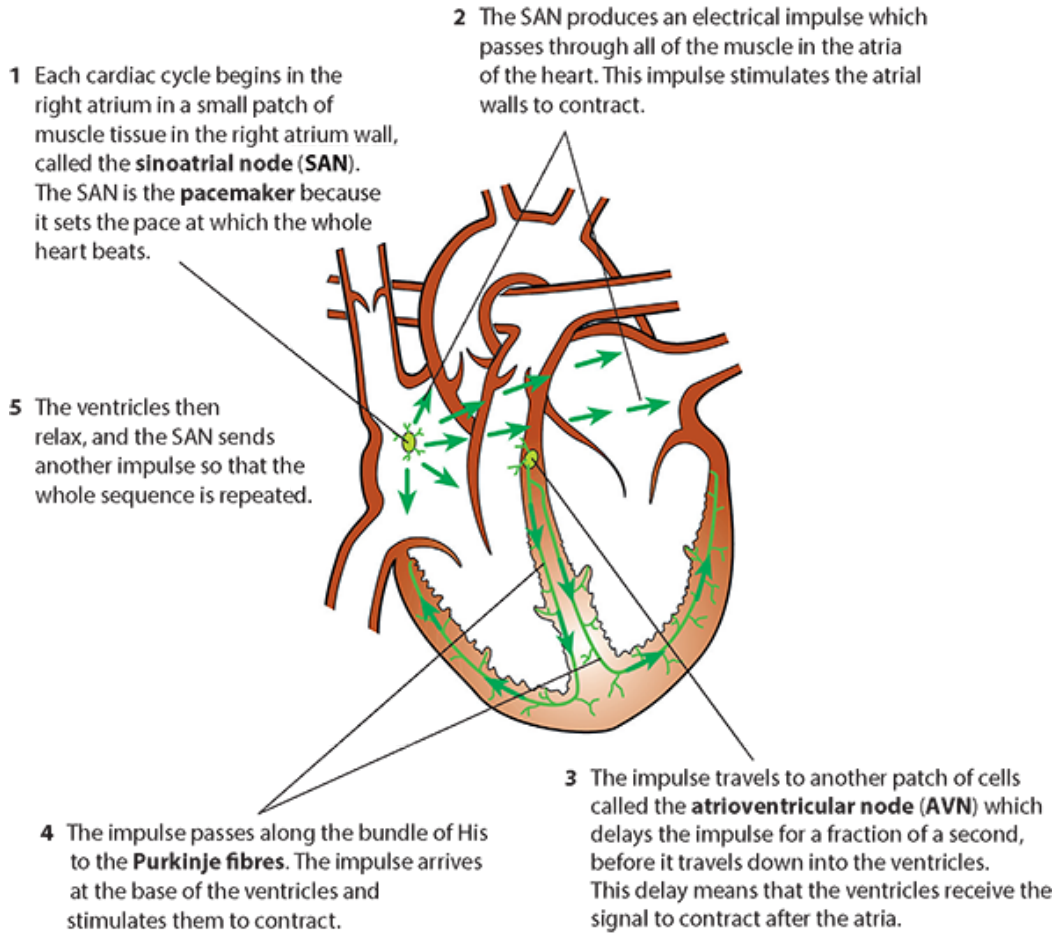


Figure 8.2.9: How electrical impulses move through the heart.

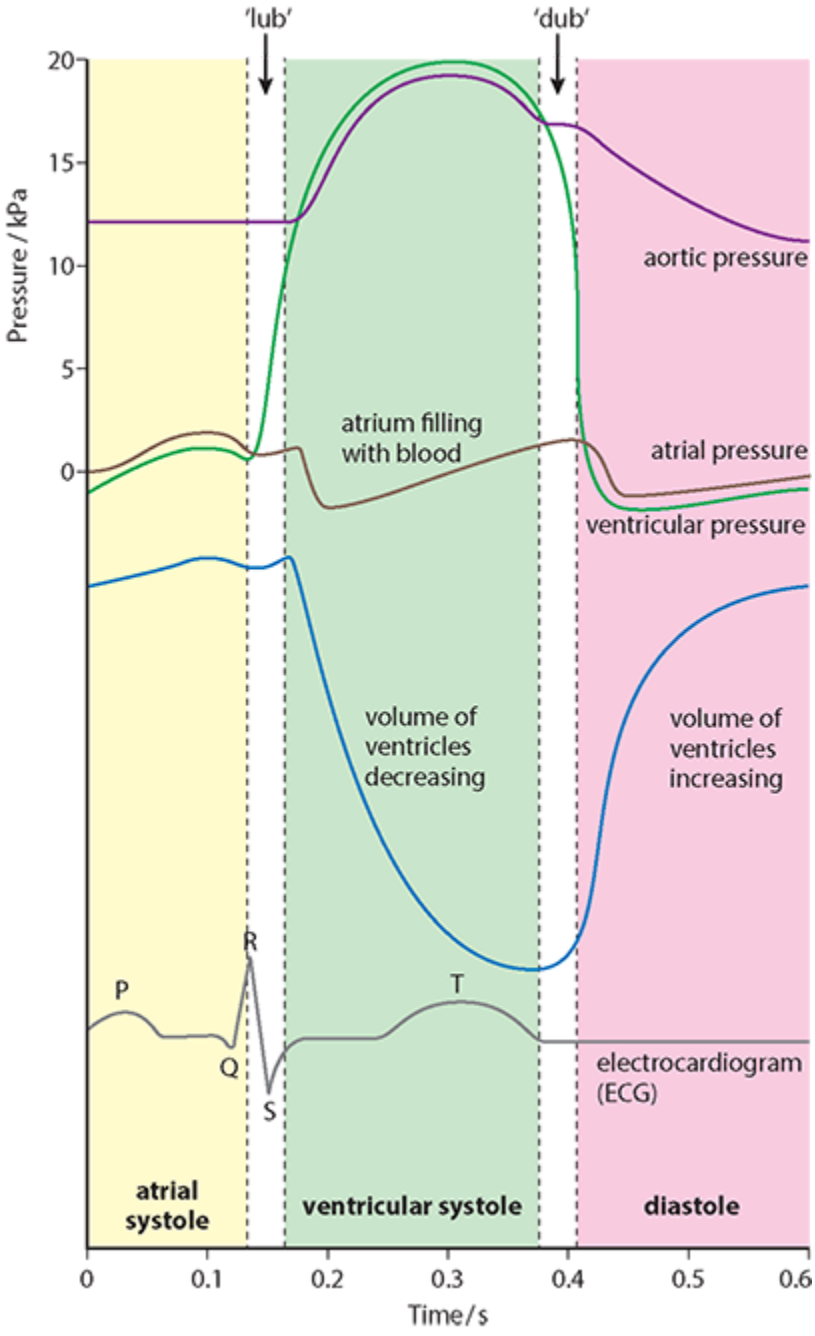


Figure 8.2.10: Pressure and volume changes in the heart during the cardiac cycle.

SCIENCE IN CONTEXT

The maximum heart rate that an individual can achieve is estimated from a simple calculation:

$$\text{Maximum heart rate} = \frac{220 \text{ beats}}{\text{min}} - \text{age in years}$$

A 20-year-old person will have a maximum heart rate of about 200 beats per minute but for a 50-year-old this will decrease to about 170 beats per minute. This rate cannot be modified by exercise, training or other factors because it is genetically determined for each individual.

8.2.3 Blood distribution

When we are resting about 25% of the output of blood from the heart travels to the muscles and the heart itself. If we exercise vigorously this will change to about 95%. Cardiac output is the volume of blood pumped out of the heart per minute. At rest it is about 5 litres per minute but this will increase five fold to about 25 litres per minute during heavy exercise.

Table 8.2.3 shows how blood is distributed to the organs of the body depending on their needs and how the distribution changes in response to activity.

Blood flow to the brain seldom changes as its function is vital but blood flow to the heart, muscles and skin increases with exercise to supply more oxygen and allow heat to be lost from the skin. Blood flow to the digestive system, liver and kidneys decreases during activity so food absorption and excretion also decrease.

Organ	Blood flow % when body is at rest	% blood flow during strenuous activity
brain	15%	12%
skin	5%	7%
heart	5%	8%
muscles	20%	70%
kidney	20%	1%
Liver and intestines	30%	1%
Other parts	5%	1%

of the body		

Table 8.2.3: Distribution of blood to different parts of the body at rest and during activity.

8.2.4 Lymphatic system

Blood that acts as the means of transport for nutrients, oxygen and many other materials is composed of cells and a liquid called plasma. Blood plasma is a pale yellow watery liquid that makes up 50–60% of our blood volume. Plasma contains dissolved proteins, nutrients, gases and waste substances. Suspended in plasma are three important groups of cells: erythrocytes (red blood cells), whose job is to carry oxygen, leucocytes (white blood cells), which fight disease, and platelets (cell fragments), which are needed for blood clotting.

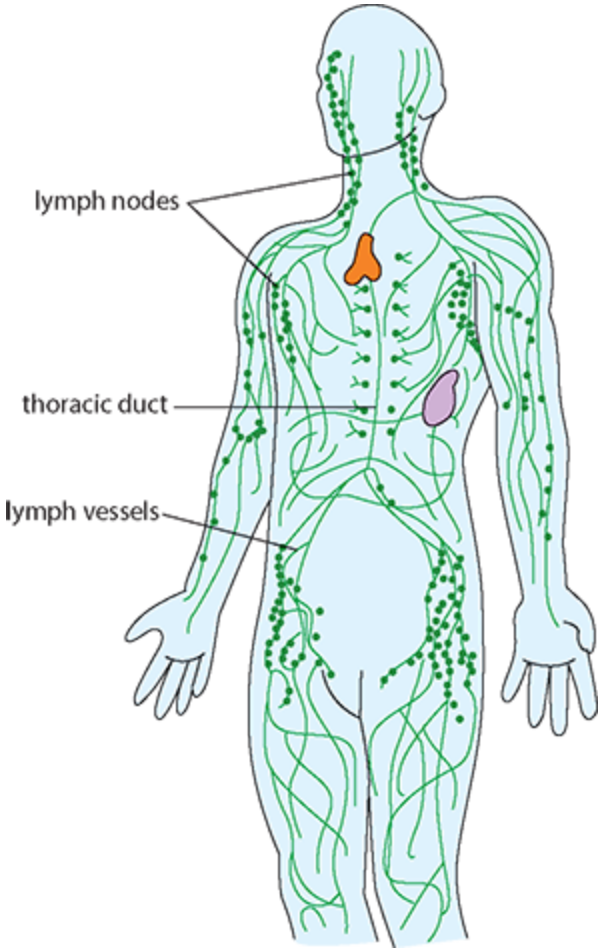


Figure 8.2.11: The lymph system. Lymph nodes are distributed throughout the body. The lymphatic system is an organ system that is part of both the circulatory system and immune systems. It is made up of a large network of lymphatic vessels that contain the clear fluid called lymph.

As blood is pumped around the body, plasma leaks out of the many tiny capillaries and bathes the nearby tissues to supply oxygen and nutrients to the cells. Once outside the capillary, the fluid is known as tissue fluid. This fluid must be collected up and returned to the circulation and this is done by the lymphatic system (Figure 8.2.11 and Figure 8.3.1).

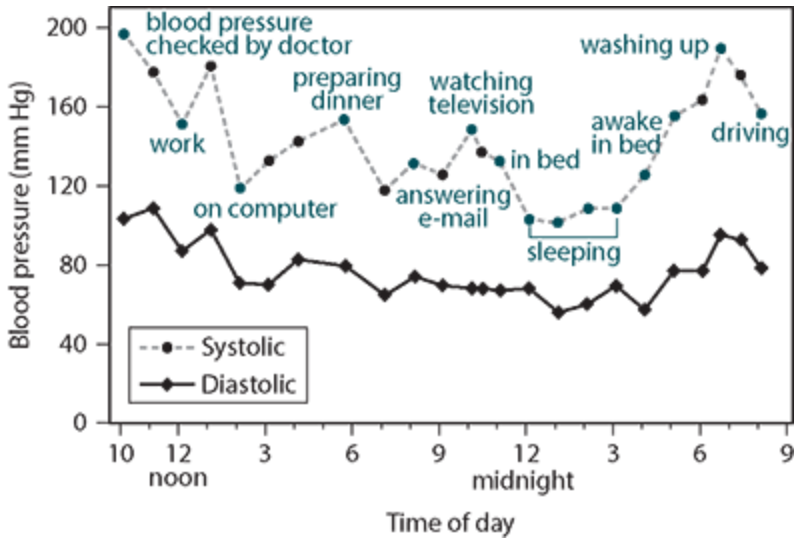
Much of the tissue fluid that leaves the capillaries is reabsorbed by the capillaries, but the remainder enters lymph vessels and is taken back to the subclavian vein close to the heart. Fluid in the lymph system contains waste products, bacteria and cell fragments. These are filtered out by the lymph nodes. Clusters of lymph nodes in the groin, neck and armpits are also responsible for releasing lymphocytes that are part of the body's immune response to infection. These help to fight bacteria, viruses and any other pathogens that cause infection.

SCIENCE IN CONTEXT

Many people call lymph nodes 'glands'. Swollen lymph nodes are usually a sign of infection and tend to go down when you recover. Lymph nodes are usually the size of a pea but they can swell to a few centimetres in response to disease or infection. Many different infections, such as a cold, tonsillitis or glandular fever, cause swollen glands. The glands in the affected area may become tender or painful as the production of lymphocytes increases to fight the infection.

TEST YOUR UNDERSTANDING

Study the graph in the figure and answer the following questions.

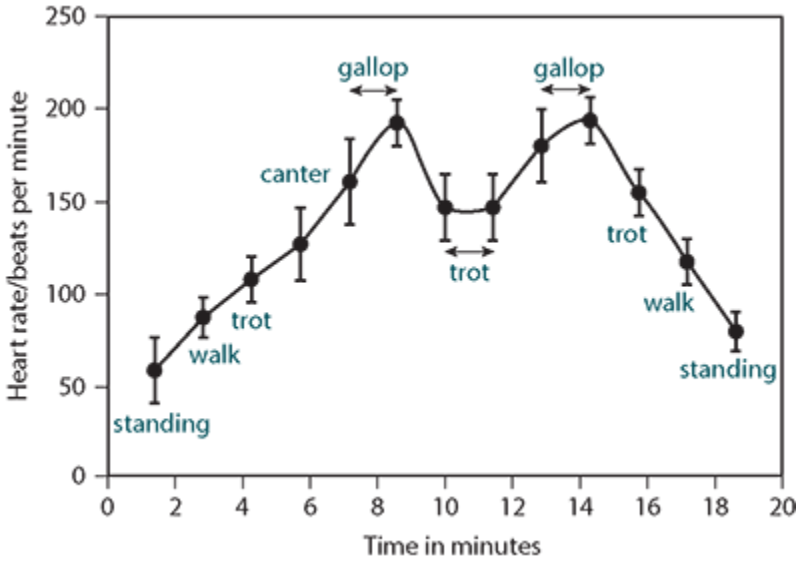


5 Suggest a reason for the high systolic reading at 1 p.m.

6 Describe the effect of sleeping on blood pressure.

7 State whether the graph indicates that this person has blood pressure within the normal range.

Study the graph in the figure, which shows the heart rate of a horse.



8 What is the resting heart rate for this animal?

9 Suggest reasons for the different values for heart rate during the three periods of trotting shown on the graph.

10 Why is the final heart rate when the horse was standing higher than the original value?

8.2.5 Transport in plants

Transpiration is the loss of water vapour from the leaves and stems of plants. Water is absorbed into the roots, travels up the stem in the xylem vessels in the vascular bundles to the leaves, and is lost by evaporation through stomata, which open to allow the exchange of oxygen and carbon dioxide in the leaf (Figure 8.2.12).

Most plants grow in areas where the amount of water in the air, the humidity, is less than in the leaves. During the day, water vapour leaves the air spaces in the spongy mesophyll and evaporates through open stomata in the lower epidermis of the leaf and in the stem. The evaporating water is drawn from the xylem in the vascular bundles in the leaf and stem. The vascular bundles are continuous with those in the roots so a column of water is formed, connecting the roots, stem, leaves and air spaces. This is known as the transpiration stream. Water molecules form a continuous column due to the cohesive forces between water molecules and the adhesive forces between water molecules and the walls of the xylem (Section 1.2).

Transpiration also carries minerals through the plant, and serves to cool the plant.

Pairs of modified epidermis cells, known as **guard cells**, that surround each stoma regulate transpiration. Guard cells have unevenly shaped cell walls with more cellulose on the side next to the **stoma**. This inner part of the cell wall is less elastic, so that when guard cells take up water and become turgid, they take on a sausage-like shape and an opening – the stoma – is formed between the two guard cells (Figure 8.2.12). When the guard cells lose water, the cell walls relax and the stoma closes.

The opening and closing of stomata is controlled by the concentration of potassium ions in the plant's cells. In darkness, these ions move out of the guard cells into surrounding cells. In light conditions, potassium ions are actively pumped into the vacuoles of guard cells. This creates an increased solute concentration so that water enters by osmosis, making the cells turgid and opening the stomata. A plant growth factor (or plant hormone) called abscisic acid, produced in the roots during times of drought, affects potassium ion movement in guard cells. When abscisic acid is present, potassium ions leak out and water follows by osmosis. This means that the guard cells lose turgor and stomata close, thus conserving water.

Distribution of tissues in the stem and root

Water is absorbed by the roots of a plant and travels up the stem in the xylem vessels in the vascular bundles. Evaporating water is drawn from the vascular bundles in the leaf and stem. Vascular bundles are continuous from the roots, through the stem and into the leaves so a column of water is formed connecting the roots, stem, leaves and air spaces.

Vascular bundles in the roots, stems and leaves of plants are made up of both xylem and phloem. Cells in the xylem become long series of cells joined end to end in a continuous fine tube which forms once the cells have stopped growing and their end walls have been removed. The phloem on the other hand is composed of living cells with perforated end walls known as sieve plates. When viewed under a microscope xylem and phloem can easily be distinguished in both cross sections and longitudinal sections through stems. (Figure 8.2.14)

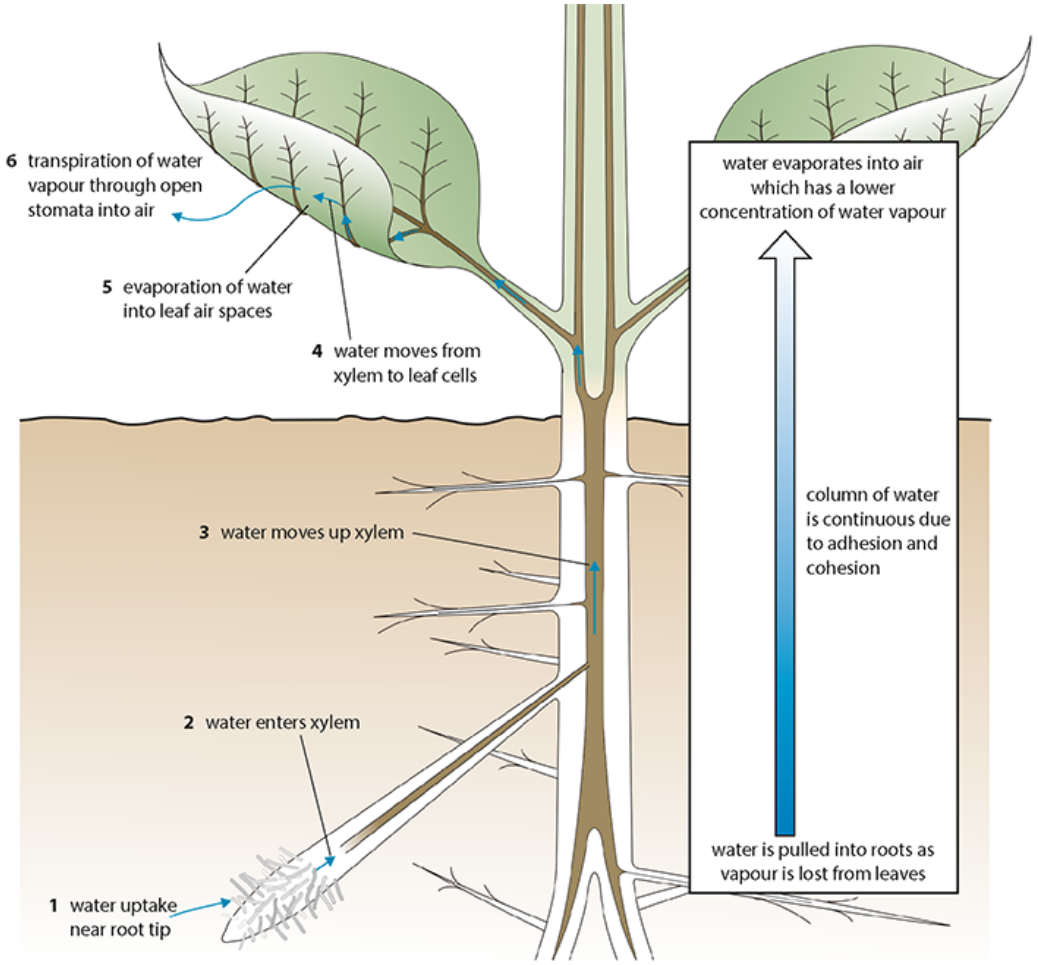


Figure 8.2.12: The movements of water through a plant: overall, water moves from the soil to the air (from where there is more water to where there is less water).

EXAM TIP

You should be able to draw diagrams of cross sections of stems and identify xylem phloem, cortex and epidermis

Adaptations of the xylem

The movement of water in the xylem can be explained by the cohesion– tension theory. A strong tension is produced in the

xylem as water is lost. The xylem is strengthened so that xylem vessels do not collapse.

- Loss of water vapour from the stomata in the leaves results in ‘tension’ or negative pressure in the xylem vessels.
- Water vapour re-enters the air spaces in the leaf from the xylem vessels.
- Continuous columns of water are drawn up the xylem due to cohesion between water molecules in the xylem and forces of adhesion between the water molecules and the xylem vessel walls. **Cohesion** is due to hydrogen bonding between water molecules and **adhesion** is caused by the hydrogen bonds between water molecules and molecules in the walls of the xylem vessels.

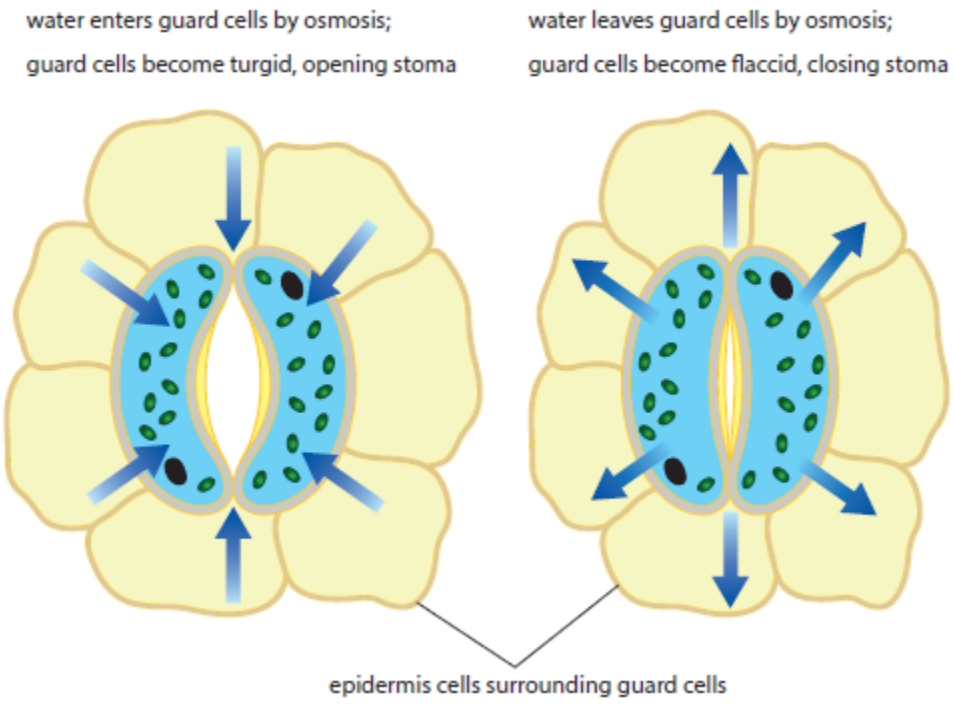


Figure 8.2.13: The opening and closing of stomata. Gases can diffuse in and out of open stomata. When stomata are closed, water loss is minimised.

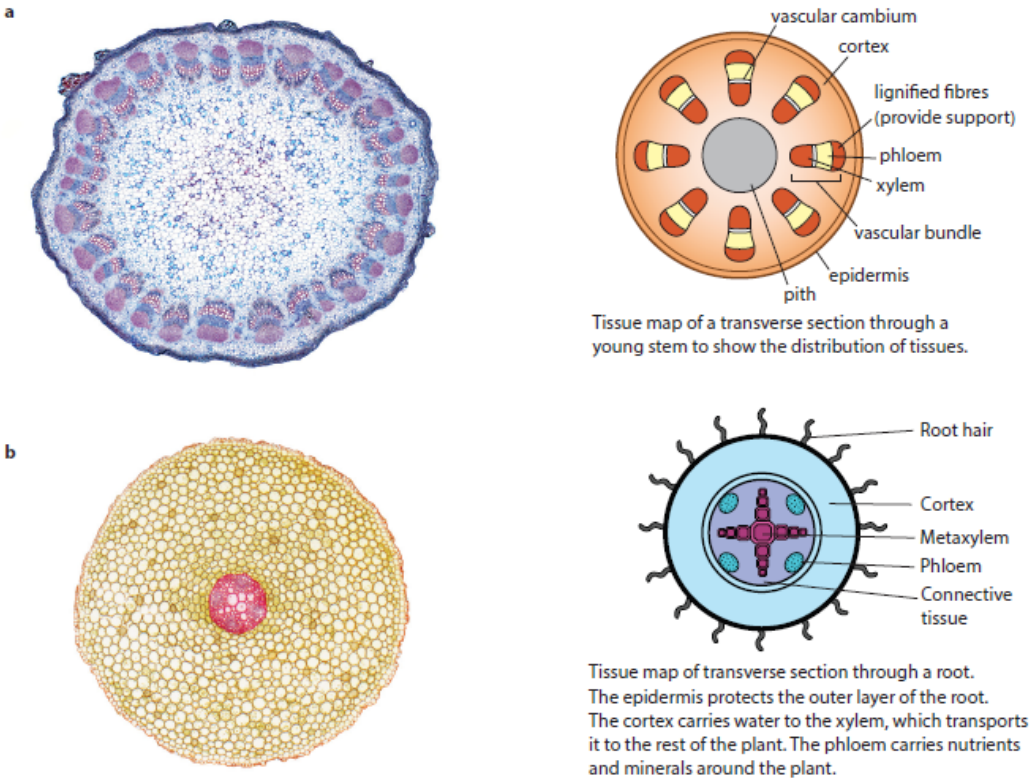


Figure 8.2.14: a transverse section through a stem **b** transverse section through a root.

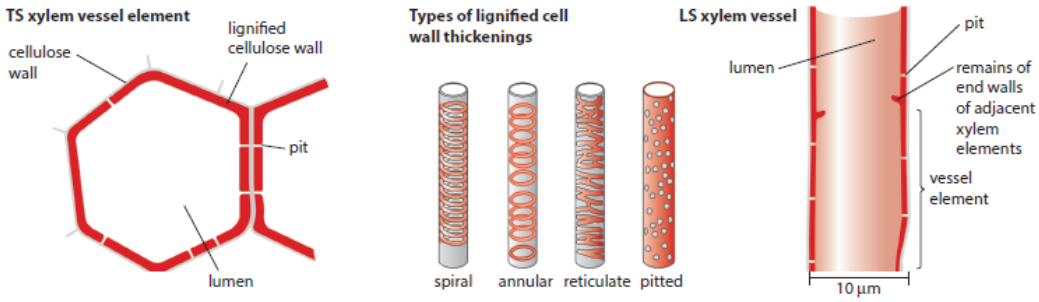


Figure 8.2.15: Xylem vessels are not alive and have no plasma membrane, so water can easily move in and out of them.

- The tension in the xylem is strong due to loss of water and there would be a tendency for xylem vessels to collapse

inwards. The thickening provided by lignin prevents this happening.

- Water is drawn in from the cortex in the roots to replace water that is lost in transpiration.
- The tension caused by transpiration also causes water to be drawn into the roots from the soil.

Factors affecting transpiration

Several abiotic environmental factors (notably light, temperature, humidity and wind speed) influence the rate of transpiration in plants.

- Light affects transpiration directly by controlling the opening and closing of stomata. As light intensity increases, stomata open, speeding up the rate of transpiration. In darkness, stomata close, thus restricting transpiration.
- Temperature affects transpiration because heat energy is needed for the evaporation of water. As the temperature rises, the rate of transpiration also rises as water evaporates from the air spaces in the spongy mesophyll and diffuses out of the stomata.
- An increase in atmospheric humidity reduces the rate of transpiration. Air in the mesophyll air spaces tends to be saturated with water vapour so if atmospheric air becomes more humid, the concentration gradient between the air space and the atmosphere is reduced and transpiration is slowed down.
- An increase in wind speed increases the rate of transpiration because it blows away the air just outside the stomata, which is saturated with water vapour. Reduced

humidity near the stomata enables water vapour to diffuse more readily from the spongy mesophyll, where the air is very humid, to the air just outside the leaf, which has lower humidity. The concentration gradient between the air space and the atmosphere is increased and transpiration speeds up.

Root pressure in the xylem

Roots are responsible for absorbing water and mineral ions from the soil. Many plants develop an extensive, branching root system in order to increase the surface area of root in contact with the soil. In addition, as new roots grow, numerous root hairs develop to increase the surface area even more (Figure 8.2.16). Root hairs are temporary and die away to be replaced by new ones near the growing tip.

Plants require a number of minerals to make a variety of substances necessary for growth. A few of these are listed in Table 8.2.4.

Minerals are present in the soil as salts – for example, calcium occurs in the form of carbonates. These dissolve in soil water and the dissolved ions can move into root cells in different ways.

KEY POINT

root pressure is a force that helps to drive water upwards in the xylem

Figure 8.2.16: A root of a young radish showing the root hairs.

Mineral ion	Importance
calcium	constituent of cell walls
magnesium	needed to make chlorophyll
iron	required as a cofactor for many enzymes

Table 8.2.4: How plants use some important mineral ions.

- Dissolved minerals may move into the root by **mass flow** of water carrying the ions, or by **facilitated diffusion** of ions from the soil water into root hairs, down their concentration gradient (Figure 8.2.17). Both these processes are passive – that is, they do not require energy in the form of ATP.
- Where the concentration of a mineral is lower in the soil water than in plant cells, **active transport** is needed to take it up. Potassium, nitrate and phosphate are usually absorbed by active transport. Root hair cells contain mitochondria to provide ATP and most roots can only take in minerals if oxygen is available for aerobic respiration, to provide sufficient ATP. Experiments have shown that potassium ions stop moving into root cells from the soil when potassium cyanide is added. Cyanide is a potent blocker of respiration as it inhibits enzyme action, and so it prevents active transport.

Active uptake of minerals into roots leads to an increase in the solute concentration inside root cells. This in turn causes the absorption of water by osmosis. The water then travels to the leaves in the transpiration stream. When transpiration is high, water in the xylem is usually under tension, not under pressure, as transpiration pulls water upward. But if transpiration is low, for example if the humidity is high or if soil moisture levels are high, water can enter the xylem and move up a plant as a result of **root pressure**.

KEY POINT

root pressure is osmotic pressure generated in the cells of a plant's root system that causes water and minerals to travel up the xylem

Root pressure can transport water and dissolved mineral nutrients from roots through the xylem to the tops of short plants even when transpiration is low or even zero. It occurs in trees in spring before the leaves of deciduous species have developed.

Root pressure is caused by the active transport of mineral ions into the root xylem. If there is no transpiration to carry the ions up the stem, they accumulate in the root xylem and lower the water potential. Water then diffuses from the soil into the root xylem due to osmosis. Root pressure is caused by this accumulated water in the xylem pushing on the rigid cells. Root pressure provides a force, which pushes water up the stem.

You can observe root pressure by cutting the stem of a plant close to the soil. Fluid will exude from the cut xylem for several hours due to root pressure.

Adaptations for translocation in the phloem

Translocation is the movement of organic molecules through the phloem tissue of plants. The phloem consists of two types of living cell: **sieve tube cells**, which are perforated to allow the movement of solutes through them, and **companion cells**, which are connected to the sieve tube cells as shown in Figure 8.2.18.

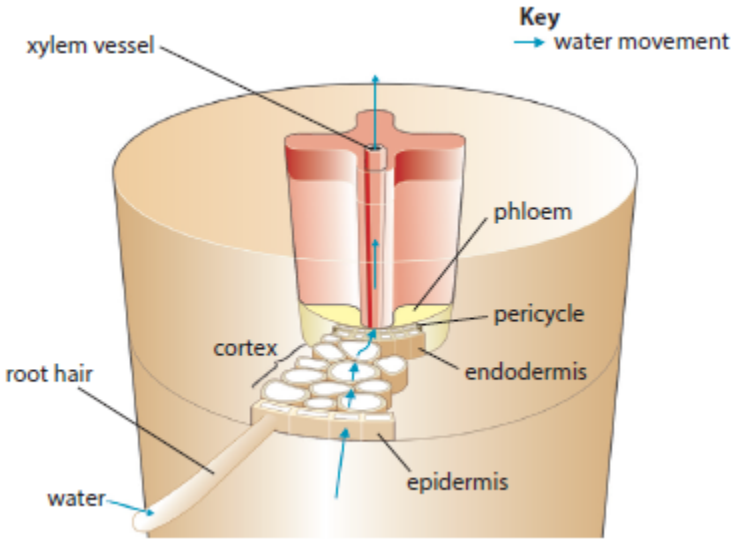


Figure 8.2.17: The pathway of water movement from root hair to xylem. The water may carry dissolved mineral ions.

Whereas the xylem carries water and mineral salts only in an upward direction, the phloem can transport materials either up or down the plant. Translocation moves materials from a **source**, where they are made or stored, to a **sink**, where they are used, as shown in Figure 8.2.19.

Storage structures such as seeds and bulbs are sinks during the growing season but may also act as sources when they begin to sprout.

The products of photosynthesis, including sugars and amino acids, move from leaf cells, which are a source, into the phloem. Once in the phloem, they are translocated to sink regions, such as growing tissue in the meristems of roots, buds and stems, or storage organs like fruits and seeds.

All the materials that are moved by translocation are dissolved in water to form a solution called ‘sap’ which also carries plant hormones. Sugar is usually carried as sucrose, which enters and

leaves the phloem by active transport using energy provided by the companion cells. High concentrations of solutes in the phloem, at sources, such as the leaves, leads to the uptake of water which enters by osmosis. Once materials have entered the phloem, they move passively throughout the plant towards 'sinks' as a result of raised hydrostatic pressure in the phloem.

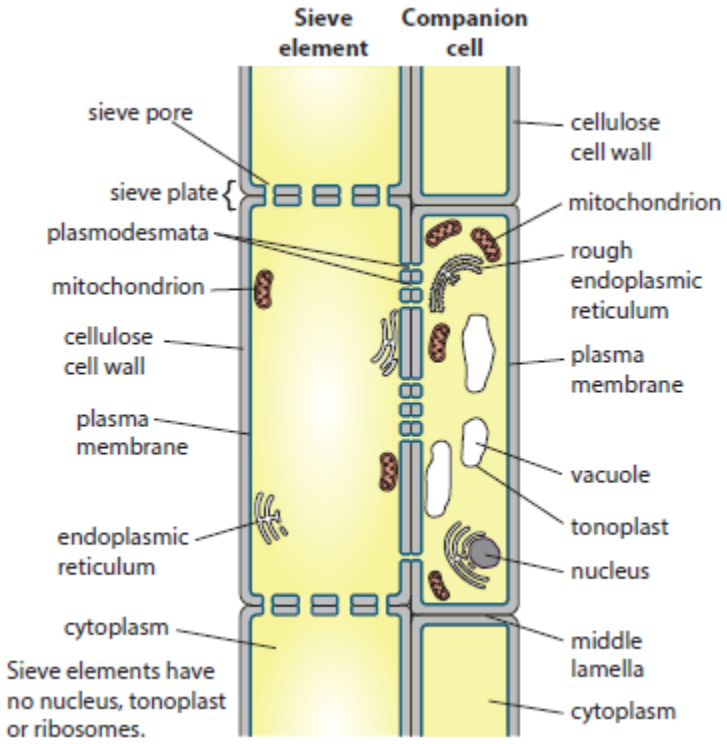


Figure 8.2.18: A phloem sieve tube element and its companion cell.

Hydrostatic pressure is defined as the pressure exerted by a liquid. It depends on the height of a column of liquid and gravity. In the case of plants, the contents of the phloem are at a hydrostatic pressure which increases lower down the plant because the hydrostatic pressure in a volume of liquid increases with depth as the fluid above exerts a downward force. This change in pressure with depth is known as the hydrostatic pressure gradient. In the leaves, a hydrostatic pressure gradient

is formed and water enters the phloem along with sugar and both will be moved down the phloem to other parts of the plant.

At the roots a negative pressure gradient is induced across the root cortex by transpiration. Soil has a higher solute concentration and water enters along with minerals and both are transported in the xylem as a result of the negative hydrostatic pressure gradient.

Table 8.2.5 Compares the structure and function of the xylem and phloem

Xylem	**Phloem**
Composed of a column of dead cells (when mature) – cell end walls removed	Composed of a column of living cells with perforated walls between them
Continuous tube of cells enables an unbroken column of water (held by cohesive forces) to move inside the xylem	Living cells enable substances to be loaded by active transport
Thickened with lignin to withstand negative pressure as water vapour is lost in transpiration	Associated with companion cells which carry out cell functions and supply energy for active transport into the phloem
Transports water and minerals passively from roots to leaves	Transports sugars, amino acids, hormones to all parts of the plant by mass flow

Table 8.2.5: Structure and function in the xylem and phloem.

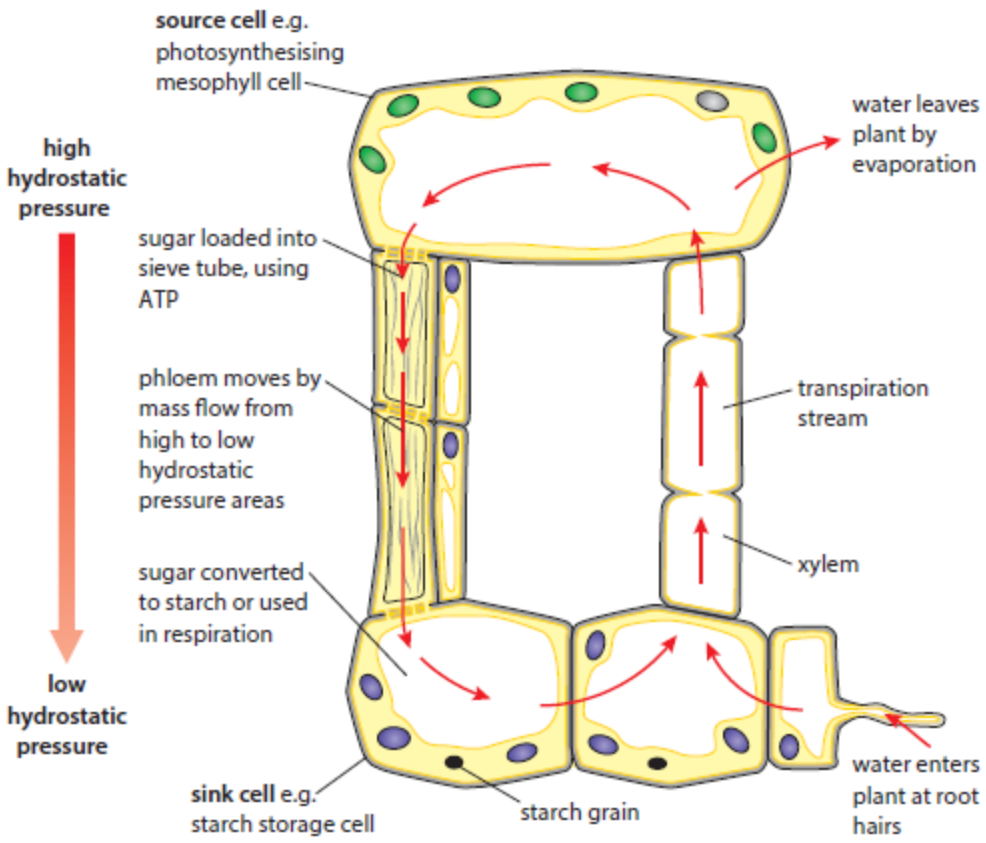


Figure 8.2.19: Sources, sinks and mass flow in phloem.

TEST YOUR UNDERSTANDING

11 Draw a plan of the structure of a typical stem to show the arrangement of tissue inside.

12 Compare the structure of xylem and phloem.

13 State the substances that are carried in the xylem

14 List the factors that affect the rate of transpiration.

15 Explain why hydrostatic pressure is important to translocation.

Links

- How does the circulatory system contribute to homeostasis? (Chapter 8.5)

8.3 Gas exchange

LEARNING OBJECTIVES

In this section you will:

- recognise that multicellular organisms need to exchange material within their bodies and with the external environment
- understand that exchange surfaces are permeable, moist and have a large surface area to allow materials to cross them easily
- recognise that circulatory surfaces maintain concentration gradients across exchange surfaces
- recall that materials are exchanged between blood and tissues via capillaries.
- learn that blood pressure influences the rate of exchange in capillaries
- understand that lungs contain alveoli that provide a large surface for the exchange of gases
- recognise that ventilation rate and depth influence the rate of exchange of gases in the lungs
- Recognise the adaptations of type 1 and type II pneumocytes in alveoli

> understand that the affinity of hemoglobin for oxygen at different partial pressures of oxygen can be shown in

a graph called a dissociation curve

- learn that carbon dioxide is carried in the blood both in solution and bound to hemoglobin and that carbon dioxide is converted to hydrogen carbonate ions in the red blood cells
- learn that the increased release of oxygen by hemoglobin in respiring tissues can be explained by the Bohr shift
- recall that chemoreceptors are sensitive to pH changes in the blood
- learn that the respiratory centre in the medulla oblongata controls the ventilation rate
- understand that ventilation rate increases in response to the amount of carbon dioxide in the blood during exercise
- recognise that fetal hemoglobin differs from adult hemoglobin, which means that oxygen can be transferred across the placenta to the fetus
- understand that terrestrial plants have adaptations to their structure to enable them to survive on land
- recognise that hydrophytes and xerophytes have different leaf structures for gas exchange.

GUIDING QUESTIONS

- How are exchange surfaces adapted to carry out their roles?
- Why are these surfaces needed?
- Why are exchange surfaces closely linked to a circulatory system?

8.3.1 General features of exchange surfaces

Large organisms must supply nutrients and oxygen to all their cells and carry waste metabolic products away from them. To do this they need surfaces that enable them to exchange materials quickly and efficiently. Some materials enter from the environment outside the organism and are absorbed through a digestive system or respiratory surface. Others, such as hormones, are produced inside the body and must enter cells from the internal environment. All exchange surfaces have features in common that allow these processes to take place efficiently and speedily. Exchange surfaces must:

- be permeable to the substances that must pass across them.
- be thin, so that there is a short distance for exchange by diffusion or other means.
- be moist, so that materials can dissolve if necessary.
- have a large surface area so there is maximum area for exchange.
- have a means of maintaining a concentration gradient so that substances can flow down a gradient to where they are needed.

Maintaining a concentration gradient

In animals, a concentration gradient at exchange surfaces is usually maintained by blood flow through capillaries (Figure 8.3.1), which have all the features needed to make them good for efficient transfer of materials. Tiny capillaries have a very large

surface area and as blood enters a capillary network (also called a capillary bed) its flow slows down because the capillaries are so narrow. The slow flow means that materials can be exchanged with the nearby cells. The walls of the capillaries are only one cell thick so the diffusion distance across them is very small.

The concentration of substances, such as oxygen and glucose, in the cells around the capillary is lower than the concentration in the blood arriving from arterioles (small arteries) that branch into capillaries. At the arterial end of the capillary bed blood pressure is much higher and plasma (the liquid part of blood), containing dissolved glucose, oxygen, amino acids and salts, leaves the capillary through gaps between the cells of the capillary walls. These substances move down the concentration gradient from the liquid (which is known as tissue fluid once it has left the capillary) into cells. Carbon dioxide and wastes diffuse into the fluid in the opposite direction down their concentration gradients. Oxygen, glucose and other materials can pass by diffusion into the cells. The narrow capillaries produce a drop in blood pressure at the venous end of the capillary bed so that water in tissue fluid re-enters by osmosis. With each pump of the heart, a fresh supply of blood arrives in the capillaries to maintain a high concentration gradient and pressure at the arterial end.

Blood pressure is responsible for the efficient exchange process and is maintained by:

- Pumping of the heart: keeps pressure high and supplies fresh materials in the blood
- Blood volume in the blood vessels: produces pressure on the capillaries
- Diameter of the blood vessels: narrow diameter increases blood pressure.

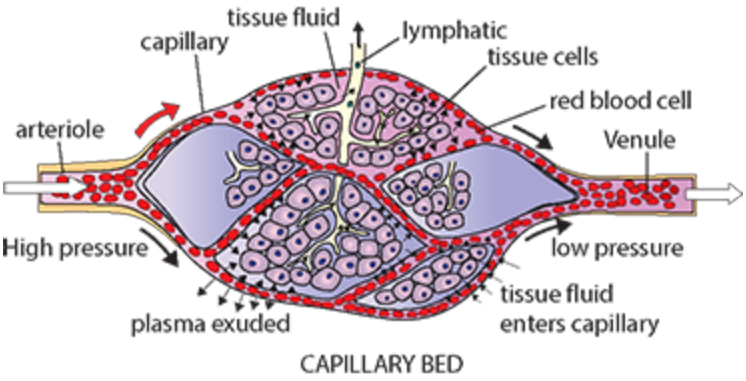


Figure 8.3.1: In a capillary network, blood pressure is higher at the arteriolar end and lower at the venous end where blood returns to the veins via venules.

8.3.2 Gas exchange in the lungs

Exchange and absorption of gases in the lungs

Oxygen enters an animal's body from the external environment, the water or air surrounding it. In simple animals, oxygen can be absorbed by the entire exposed body surface, but in more complex animals special respiratory surfaces such as gills or lungs are needed. Carbon dioxide is usually eliminated through the same surface.

An effective respiratory surface has a large surface area, a rich capillary network, a thin layer of cells separating air or water from the blood vessels and, in land animals, a moist surface. Animals also have a means of renewing the air or water in contact with the respiratory surface.

Gas exchange in land animals such as humans occurs in the **alveoli** of the lungs. Alveoli are tiny air sacs that form the exchange surface in the lungs. Oxygen from the air diffuses into blood capillaries, and carbon dioxide passes in the opposite direction through the walls of the alveoli. Gases are also exchanged in the tissues where oxygen diffuses into respiring cells and is exchanged for carbon dioxide.

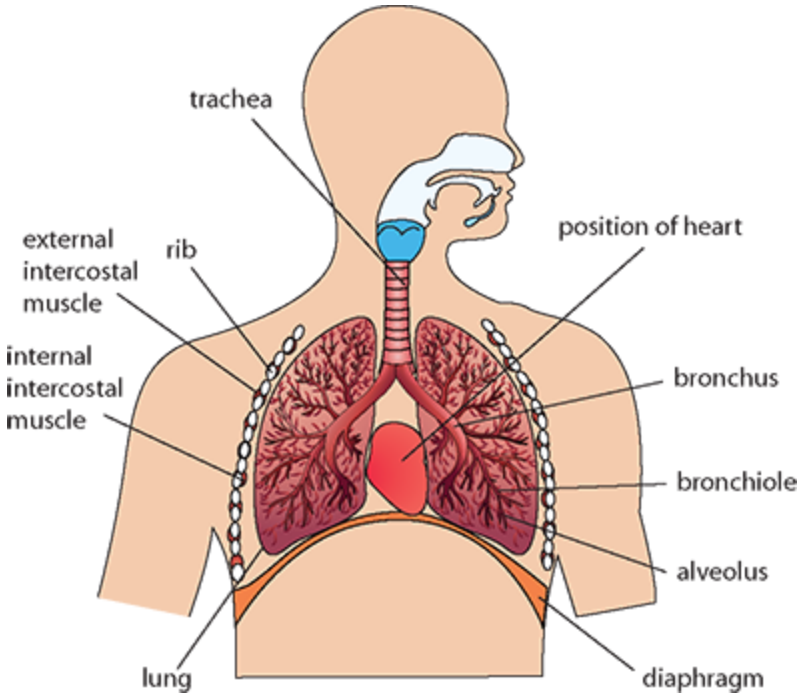


Figure 8.3.2: The human respiratory system.

Whenever diffusion occurs, there must always be a concentration gradient with a higher level of the diffusing substance in one area than in another. Air inside the alveoli contains a higher concentration of oxygen than the blood, so oxygen diffuses into the blood. Blood contains a higher level of carbon dioxide than inhaled air, so carbon dioxide diffuses into the alveoli.

For gas exchange to continue, these concentration gradients must be maintained. As oxygen diffuses out of the alveoli, the level of oxygen inside them gradually falls and the level of carbon dioxide rises. Stale air with high levels of carbon dioxide and low levels of oxygen must be expelled regularly and replaced with a fresh supply to restore the concentration gradients of the two gases. This is achieved by breathing in and out, a process known as **ventilation**.

The human respiratory system

The human respiratory system consists of two lungs protected inside an airtight cavity formed by the ribs and diaphragm. Air is drawn in through the nose, down the trachea to the two bronchi that connect to the two lungs. Bronchi divide into smaller tubes called bronchioles that end in tiny air sacs called alveoli where gas exchange occurs.

Importance of alveoli

Alveoli are the body's gas exchange surfaces. Formed in clusters at the ends of the smallest bronchioles, more than 300 million alveoli in each lung together provide a surface area of about 75 m^2. Alveoli are roughly spherical in shape and are made of cells less than 5 μm thick. The capillaries that wrap around them also have thin walls of single epithelial cells. These two thin layers make the distance for diffusion of gases as small as possible. Oxygen diffuses through the alveolus wall and capillary into the blood and carbon dioxide diffuses in the opposite direction (Figure 8.3.3). As long as the diffusion gradient is maintained by regular breathing, diffusion will continue.

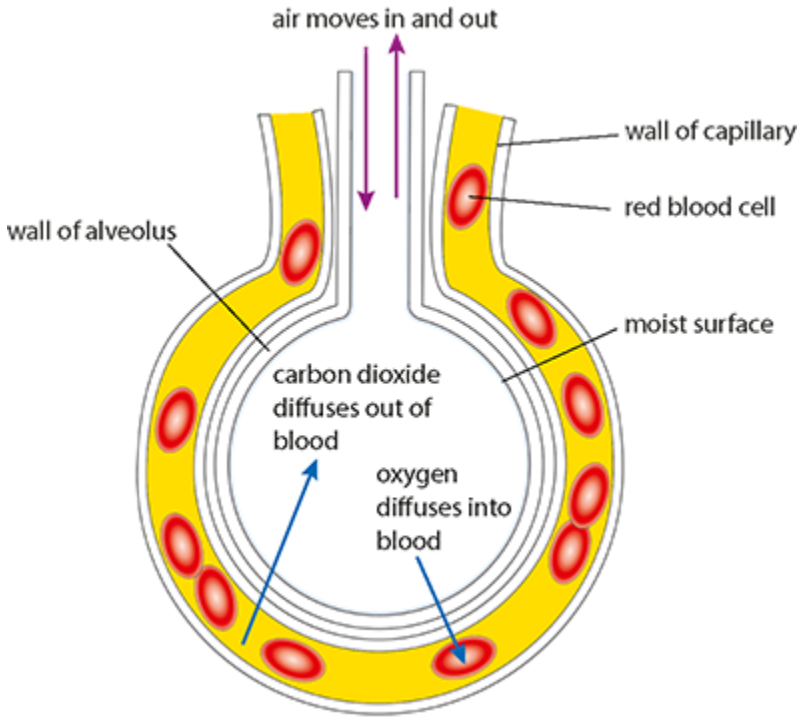


Figure 8.3.3: Gas exchange in the alveolus.

Two types of special cells called **pneumocytes** line the alveoli. Type 1 pneumocytes are very thin so that gases can diffuse easily while type II pneumocytes secrete a surfactant into the alveolus. The surfactant reduces surface tension and prevents the sides of the alveolus sticking to one another.

Table 8.3.1 summarises ways in which the alveoli are well adapted for their role in gas exchange.

Ventilation

Ventilation is essential to bring a fresh supply of oxygen-rich air into the alveoli and to remove carbon dioxide containing air from the alveoli.

Lungs have no muscles and cannot move by themselves. Breathing is brought about by two sets of intercostal muscles

between the ribs, and by the diaphragm, the sheet of muscle separating the thorax from the abdomen (Figure 8.3.4).

Feature of alveoli	Importance
many small, spherical alveoli	provide a large area for gas exchange
thin walls of flattened single cells type I pneumocytes	short diffusion distance
type II pneumocytes	secrete surfactant
rich blood supply from capillaries	maintains concentration gradient and carries absorbed gases away rapidly

Table 8.3.1: Adaptations of alveoli for gas exchange.

During **inhalation**, contraction of the external intercostal muscles raises the ribs and contraction of the diaphragm lowers the floor of the thorax. These movements increase the volume of the chest cavity. The pressure in the lungs becomes lower than that of the air outside. As a result, air is drawn down the trachea to fill the lungs.

Gentle **exhalation** occurs as the intercostal and diaphragm muscles relax, reducing the volume of the chest cavity. Elastic fibres around the alveoli return to their original length and pressure forces air out of the lungs.

Long or forced exhalations involve the internal intercostal muscles, which contract to lower the ribs. Muscles in the abdominal wall also contract and push the relaxed diaphragm

upward. Pressure inside the chest cavity increases and air is forced out of the lungs.

The volume and frequency of breathing are matched to a person's activities. During vigorous exercise the rate and depth of breathing increases, but as we sleep our breathing rate slows to only 12–20 breaths per minute.

Measuring changes in lung volume

A spirometer (Figure 8.3.5) is used to measure the amount of air that is exchanged during breathing. It can also measure the rate of breathing, for example, when a person is at rest or during or after exercise. A simple spirometer has a chamber filled with oxygen or air, which is inverted over a container of water. The subject is connected to the chamber via a tube and mouthpiece (the nostrils are closed with a nose clip). As the subject inhales and exhales, a trace is produced on a rotating drum or computer monitor. Inhalation causes the chamber to fall, producing a falling line on the trace. Exhalation causes the chamber to rise and produces a rising line. The trace is called a spirogram and various volume measurements can be made from it (Figure 8.3.6). Usually soda lime is used to absorb carbon dioxide that is exhaled.

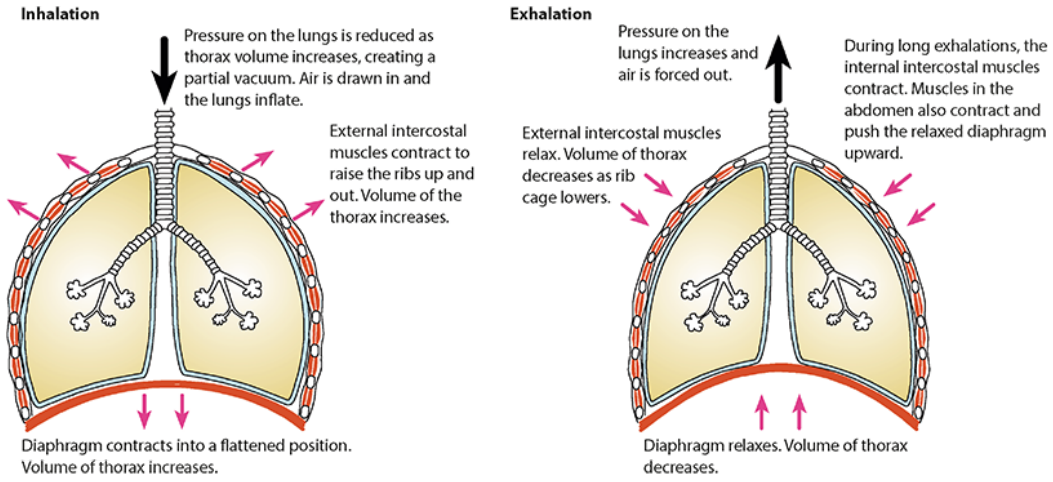


Figure 8.3.4: The mechanism of ventilation.

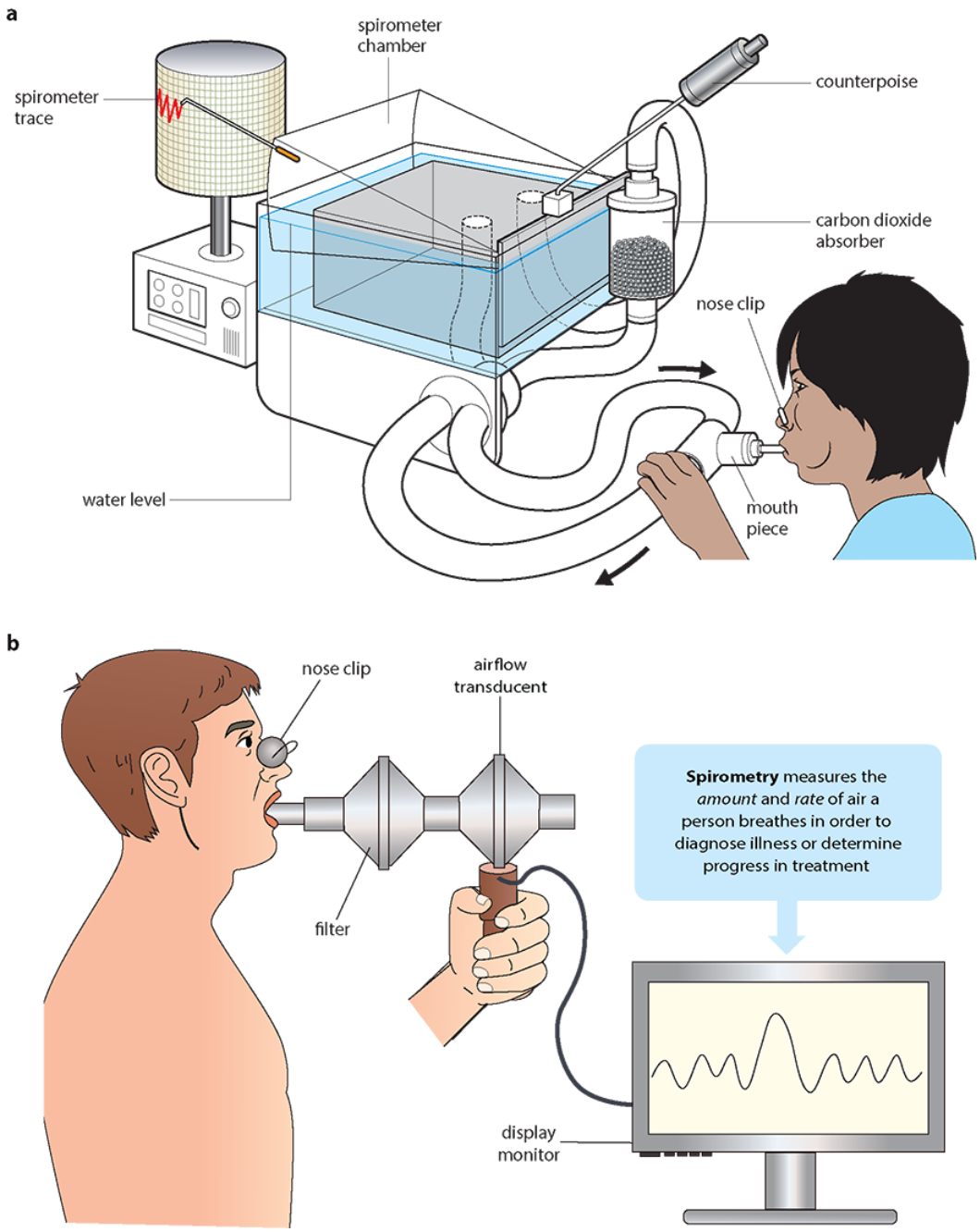


Figure 8.3.5: Simple spirometers can be connected to **a** a rotating drum or **b** a computer interface to record a trace.

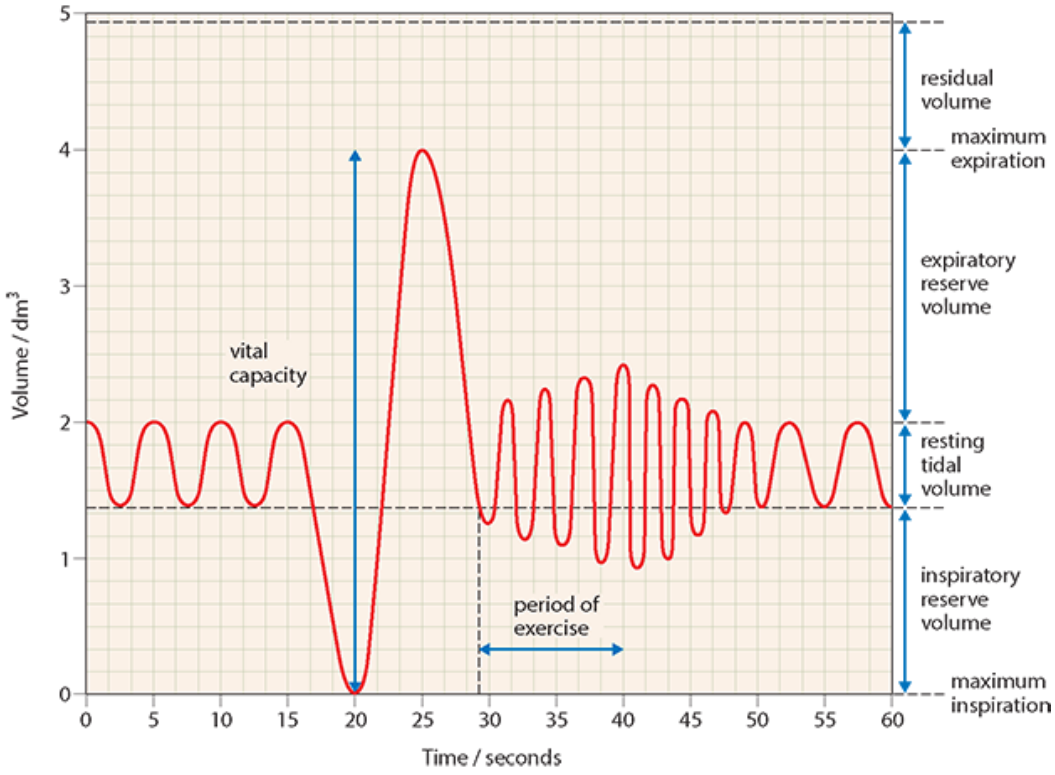


Figure 8.3.6: A spirometer produces a trace that can be used to measure various aspects of a person's breathing. Tidal volume is the volume of air breathed in and out in a single breath. Inspiratory reserve is the volume breathed in by a maximum inhalation at the end of a normal inhalation and expiratory reserve is the volume exhaled by a maximum effort after a normal exhalation. Residual volume is the volume of air remaining in the lungs at the end of a maximum exhalation.

EXAM TIP

You should be able to interpret spirometer traces like the one shown in Figure 8.3.6 and comment on the rate and depth of breathing that they show.

TEST YOUR UNDERSTANDING

16 State three features you would expect to find in an efficient exchange surface.

17 Outline how a concentration gradient is maintained in the alveoli.

18 What is the function of the surfactant in the lungs?

8.3.3 Transport of respiratory gases

Oxygen is transported from the lungs to respiring tissues bound to the hemoglobin molecules that are contained in all red blood cells. A hemoglobin molecule can bind four oxygen molecules via the iron in the heme groups at the centre of each molecule (Figure 8.3.7). When hemoglobin comes into contact with normal air, containing approximately 21% oxygen, it binds readily with oxygen molecules. It holds on to as many as it can so that it becomes almost 100% saturated.

Oxygen dissociation curves

The oxygen content of air is measured as a partial pressure. In a mixture of gases, each component gas exerts a pressure (the partial pressure) in proportion to its percentage in the mixture. It is calculated as follows. For normal dry air at sea level, atmospheric pressure is 101.3 kilopascals (kPa); a Pascal (Pa) is the SI unit of pressure. The partial pressure of oxygen, which makes up 21% of the air, is:

$$\frac{21}{100} \times 101.3\ \text{kPa} = 21.3\ \text{kPa}$$

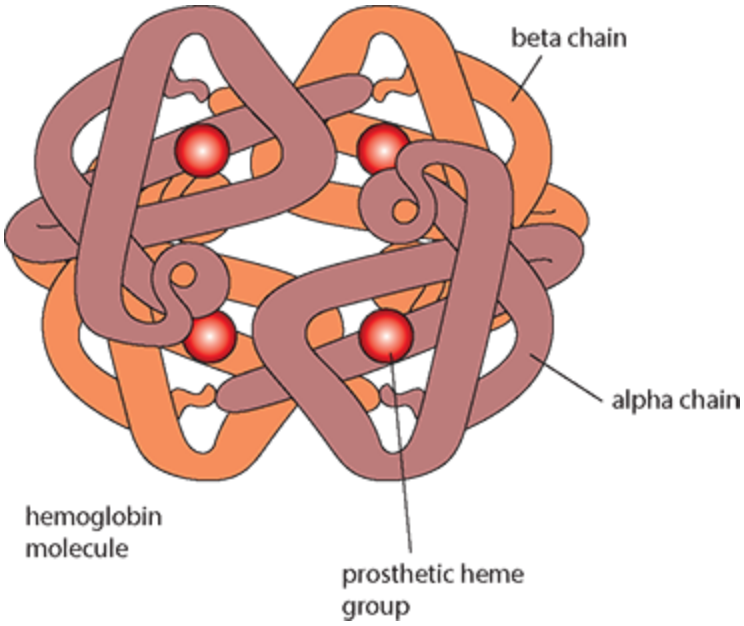


Figure 8.3.7: Hemoglobin is a protein that has quaternary structure. It consists of four subunits bound together. There are two alpha chains and two beta chains, and each of these includes an iron-containing heme group.

KEY POINTS

partial pressure the proportion of the total pressure that is due to one component of a mixture of gases.

Pascal (Pa) the SI unit of pressure; a measure of force per unit area, defined as 1 newton per square metre; 1000 Pa = 1 kilopascal (kPa).

The partial pressures of other gases in dry air at sea level are shown in Table 8.3.2.

In an area of the body where there is a lot of oxygen (a high partial pressure), such as the lungs, most hemoglobin molecules will be carrying the maximum amount of oxygen and will be fully saturated. However, in areas where the oxygen level is

lower, fewer hemoglobin molecules carry their maximum complement of oxygen and the hemoglobin may be only 50% saturated. As blood travels from the lungs to actively respiring tissues, the amount of oxygen bound to hemoglobin changes as the partial pressure of oxygen decreases. Hemoglobin readily releases oxygen where the partial pressure is lower, so it acts as an oxygen delivery service for respiring cells. Oxygen is said to dissociate from (meaning detach from) the hemoglobin molecules that carry it when partial pressure is lower. Figure 8.3.8 shows the percentage saturation of hemoglobin at different partial pressures of oxygen and is known as an oxygen dissociation curve. The steep S-shape of the dissociation curve shows how the affinity of adult hemoglobin changes at different partial pressures of oxygen. As each heme group accepts oxygen, it becomes easier for the next heme group of the molecule to pick up oxygen and hemoglobin is said to have a higher affinity for oxygen.

Gas	Approximate percentage composition / %	Partial pressure / kPa
oxygen	21	21.3
carbon dioxide	0.0035	negligible
nitrogen	79	80.0

Table 8.3.2: The partial pressures of gases in dry air at sea level. At high altitude, the pressure of air falls but the percentage of oxygen in the air remains approximately the same. At 5000 m, the partial pressure of oxygen is 11.5 kPa; at 10 000 m, it falls to just 5.5 kPa.

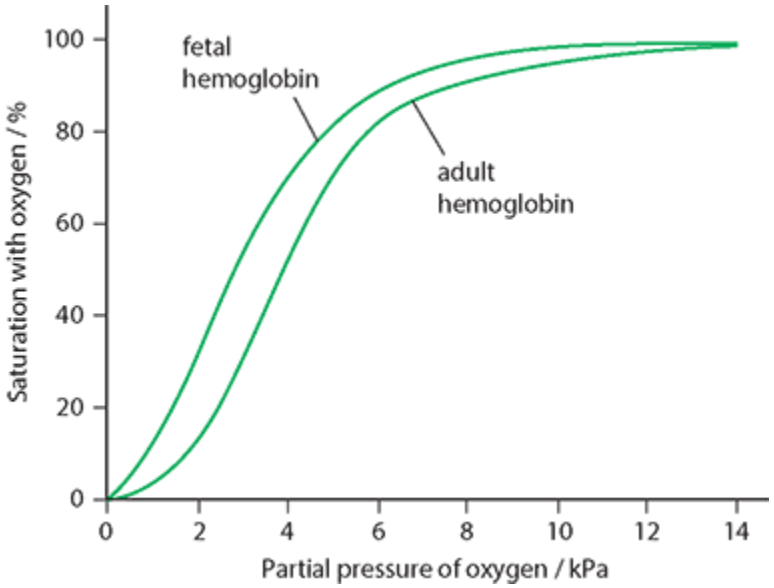


Figure 8.3.8: Dissociation curves for adult hemoglobin and fetal hemoglobin. The curves are constructed using the normal range (at sea level) of partial pressure of oxygen in the body. The partial pressure of oxygen in alveolar air is about 14 kPa due to the presence of water vapour, which forms about 6% of alveolar air.

KEY POINT

oxygen dissociation curve a graph showing the percentage saturation of hemoglobin with oxygen at different partial pressures.

At a partial pressure of 10 kPa, which might be found in the lungs, hemoglobin is 95% saturated. At a partial pressure of 4 kPa, found in the tissues, hemoglobin does not bind with oxygen and will release it, so saturation falls to only about 50%. About half of the oxygen collected by hemoglobin in the lungs is released at this low partial pressure to supply the needs of actively respiring cells.

Fetal hemoglobin

The molecular structure of hemoglobin in the blood of a fetus is different from that of an adult. The dissociation curve for fetal hemoglobin lies to the left of the adult curve for all partial pressures of oxygen (Figure 8.3.8). This tells us that fetal hemoglobin has a higher affinity for oxygen than maternal (adult) hemoglobin, whatever the concentration of oxygen. In the capillaries of the placenta, the partial pressure of oxygen is low. Here the mother's adult hemoglobin releases oxygen, which is easily picked up and bound to fetal hemoglobin. At a partial pressure of 4 kPa, the mother's hemoglobin is only 50% saturated, but fetal hemoglobin becomes approximately 70% saturated. The fetal hemoglobin carries the oxygen to the baby's body and releases it into the respiring fetal tissues.

Transport of carbon dioxide in the blood

Carbon dioxide produced during aerobic respiration is carried back to the lungs by the blood. It diffuses into capillaries close to respiring cells and is transported in one of three ways.

1 About 70% of carbon dioxide enters red blood cells and is converted to HCO_3^- (hydrogen carbonate) ions.

2 About 7% remains in the blood and is transported dissolved in plasma.

3 The remainder is bound to hemoglobin.

Carbon dioxide reacts with water to form carbonic acid, which dissociates to form hydrogen carbonate ions and hydrogen ions:

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

As carbon dioxide dissolves in the blood it lowers the pH, making the blood more acidic. As hydrogen ions and carbon dioxide bind to hemoglobin, they cause the Bohr shift, which is described next.

The Bohr shift

The affinity of hemoglobin for oxygen is not only affected by the partial pressure of oxygen, but it is also reduced in the presence of high carbon dioxide concentrations. As the partial pressure of carbon dioxide in the blood rises, the ability of hemoglobin to combine with oxygen falls and so the dissociation curve moves to the right. This effect is known as the **Bohr shift**. It is caused when hydrogen ions produced from carbonic acid combine with hemoglobin and the pH of the blood decreases. Figure 8.3.9 shows the effect of two different partial pressures of carbon dioxide on the oxygen dissociation curve. In an environment where the partial pressure of carbon dioxide is high, such as in actively respiring tissue, the curve moves to the right. This means that, at any given oxygen partial pressure, oxygen is more likely to dissociate from hemoglobin if the partial pressure of carbon dioxide is high. This effect promotes the release of oxygen in active tissues where respiration is producing high levels of carbon dioxide, so cells receive the oxygen they need.

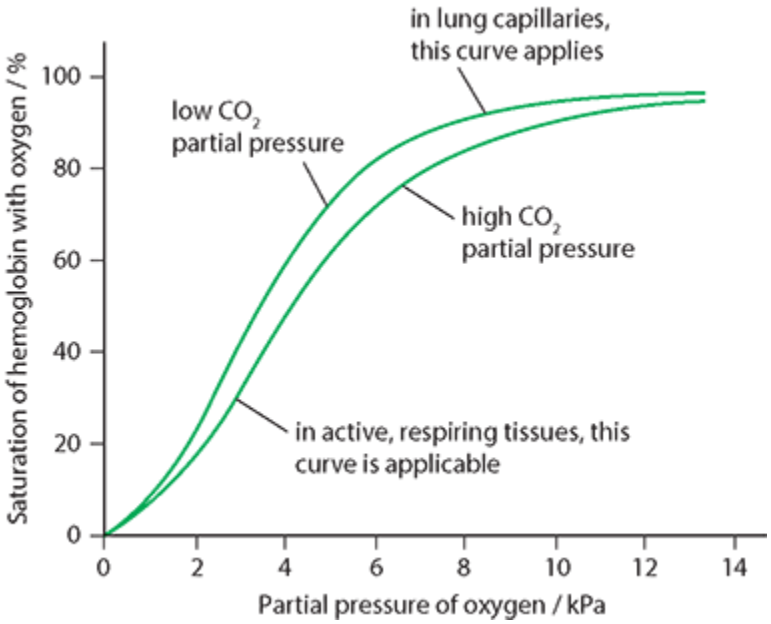


Figure 8.3.9: The effect of carbon dioxide concentration on hemoglobin saturation: the Bohr shift.

Ventilation rate and exercise

When a person exercises, their ventilation rate and tidal volume (depth of breathing) increase. Muscles need oxygen for aerobic respiration and so as the rate of exercise increases so does the rate of oxygen consumption. Blood returning to the lungs also has a higher level of carbon dioxide, produced as a result of the increased activity. An increase in ventilation rate and tidal volume draws in more fresh air to maintain the concentration gradient between the alveolar air and the blood. Thus oxygen can be absorbed at a faster rate and the body can get rid of the additional carbon dioxide produced. These changes in ventilation are adjusted to match the body's metabolic needs.

Ventilation rate is controlled by the breathing centre of the medulla oblongata in the brain stem, which receives nerve impulses from sensory cells in different parts of the body. The breathing centre responds to match ventilation rate to activity

levels (Figure 8.3.10). Chemoreceptors in the inner wall of the aorta and carotid arteries respond to an increase in carbon dioxide in the blood. This excess carbon dioxide forms carbonic acid and so the pH of the blood falls. Impulses are passed from the chemoreceptors to the medulla. The medulla increases the ventilation rate by sending motor impulses to the intercostal muscles and diaphragm to increase their rate of contraction. The breathing centre also contains similar chemoreceptors, which respond to deviations of blood pH from the normal level. An increase in ventilation rate causes carbon dioxide to be removed from the body at a faster rate and blood pH returns to its normal level of between 7.35 and 7.45. After exercise, as the level of carbon dioxide in the blood falls, ventilation rate decreases.

KEY POINT

medulla oblongata area of the brain stem that controls ventilation rate and heart rate.

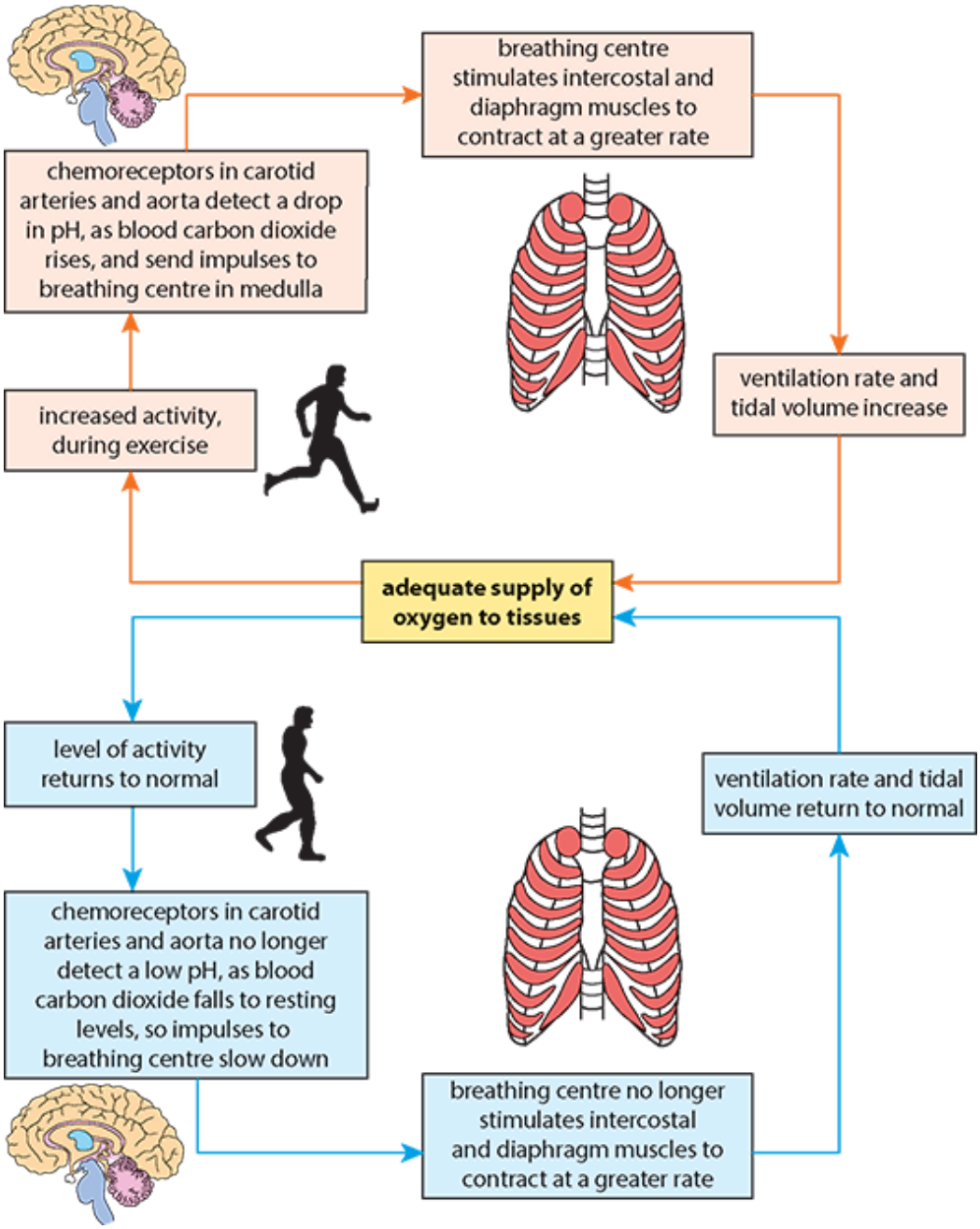


Figure 8.3.10: The control of breathing.

8.3.4 Gas exchange in plants

Terrestrial plants have many adaptations to their exchange surfaces, which enable them to take in the sunlight, gases, water and solutes that they need (Figure 8.3.11).

Leaves of flowering plants are adapted to capture light energy from the sun and exchange the carbon dioxide and oxygen needed for photosynthesis and respiration.

The stomata (pores) are openings in the epidermis that are more abundant on the underside of a leaf that is usually shaded and cooler. Stomata are formed between two guard cells that can open and close the stomata according the environmental conditions. When there is sufficient light intensity for photosynthesis, the guard cells open the stomata to allow gases to enter and leave. But if water is scarce, the stomata close so that water is conserved in the plant cells (Figure 8.3.12).

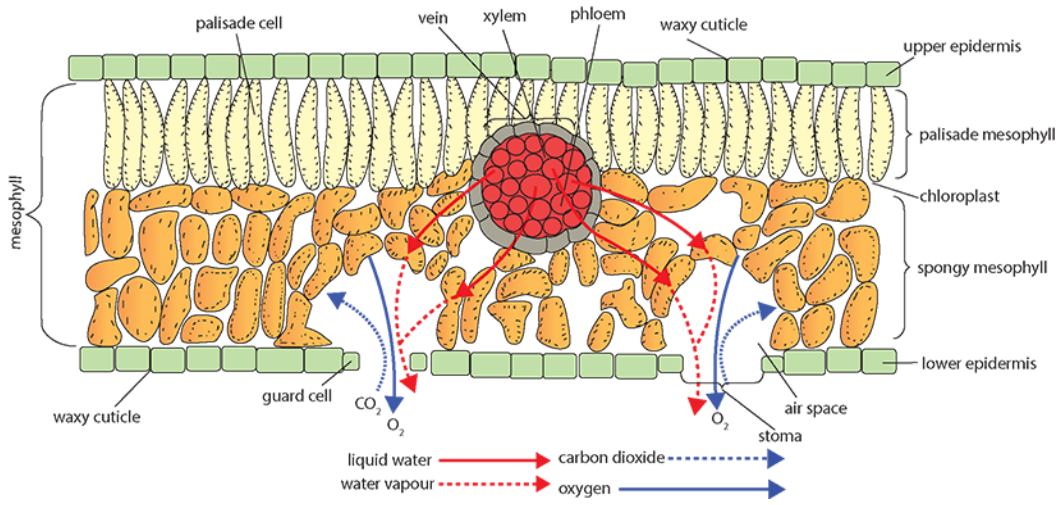


Figure 8.3.11: Cross-section showing the structure of a typical leaf of a terrestrial plant.

Part of leaf	Adaptation and functions
waxy cuticle covering epidermal cells	a waterproof layer that prevents evaporation of water from the leaf surfaces but allows light through
epidermis	single layer of transparent cells that allow light through but prevent the entry of pathogens
palisade mesophyll	elongated cells at the top of the leaf that contain many chloroplasts for photosynthesis
spongy mesophyll	cells with large air spaces to allow the diffusion of gases for photosynthesis and respiration
stomatal guard cells	can open and close pores to allow the passage of gases or to conserve water
veins (xylem and phloem)	contain xylem and phloem to bring water to the leaf and carry away products of photosynthesis

Table 8.3.3: Adaptations found in the leaf of a terrestrial plant.

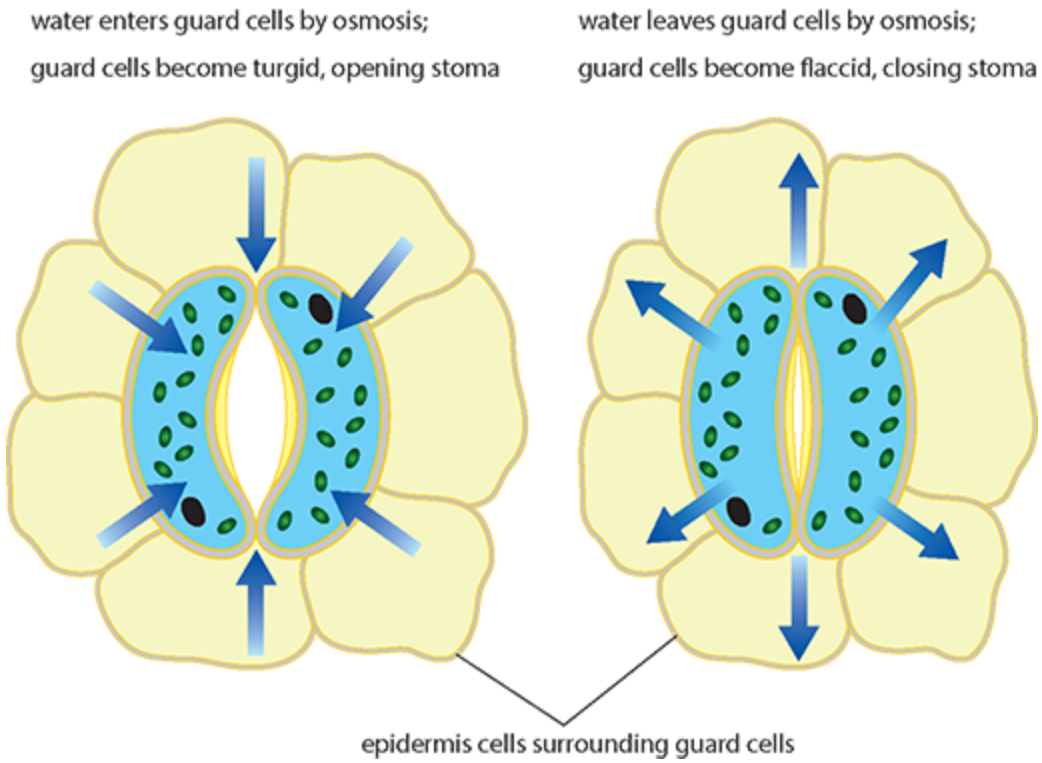


Figure 8.3.12: The opening and closing of stomata. Gases can diffuse in and out of open stomata. When stomata are closed, water loss is minimised.

EXAM TIP

Ensure that you can draw and label a plan (two-dimensional) diagram of a section through a leaf. Check that you can list the functions of each type of cell.

Stomatal density

The number of stomata in a given area of a leaf varies between plant species, and between the underside and upper surfaces of the leaves on an individual plant. The variation in size of stomata and their density is due to genetic factors and also environmental conditions. Stomatal density can vary due to factors such as light, humidity, water availability and concentration of carbon dioxide

in the atmosphere. Stomatal density gives a measure of the potential surface area for the movement of gases in and out of the leaf, but it is the opening and closing of the stomata that controls the final amount of gas exchange.

A simple way to estimate the number of stomata in a leaf surface is to use clear nail varnish to prepare an impression of the leaf. The leaf surface being investigated is coated with a thin layer of nail varnish and left to dry. A piece of clear sticky tape can be used to cover and peel off the layer of nail varnish that can be stuck onto a glass microscope slide. The number of stomata in a number of sample areas can be counted using an eyepiece graticule (Figure 8.3.13) and an average value can then be calculated.

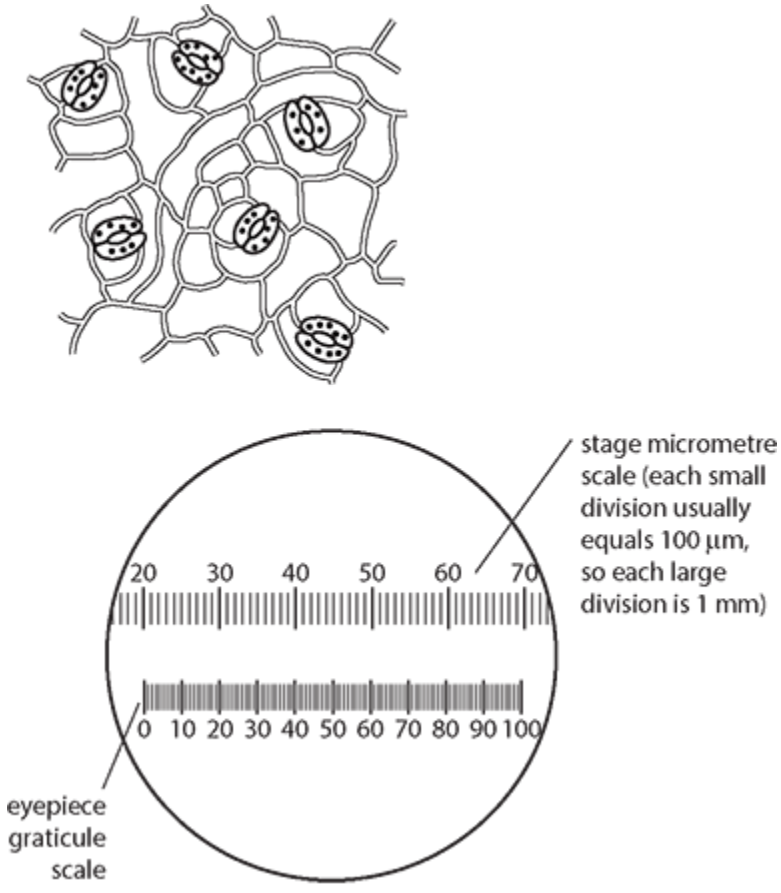


Figure 8.3.13: Stomata in an impression of the lower surface of a leaf and a graticule used with a microscope to observe and count stomata in a sample area.

INTERNATIONAL MINDEDNESS

Research from many parts of the world involving many species of trees and shrubs has shown that there has been a decrease in stomatal density over the last 100 years. This decrease is correlated with increased carbon dioxide levels in the atmosphere due to the use of fossil fuels.

To consider:

Why should an increase in carbon dioxide in the atmosphere lead to a reduction in stomatal density?

TEST YOUR UNDERSTANDING

19 Define 'partial pressure'.

20 Explain what is meant by the term 'Bohr shift' and why it is important in supplying oxygen to respiring tissues.

21 Describe how fetal hemoglobin differs from adult hemoglobin in its affinity for oxygen.

22 Why is a waxy cuticle important to the leaves of terrestrial plants?

23 How could a scientist obtain an accurate measurement of stomatal density?

Links

- What strategies are used by cells and organs to increase surface area? (Chapter 6)
- What selection pressures in an environment may lead to differences in gas exchange systems in fish? (Chapter 11)

8.4 Reproduction

LEARNING OBJECTIVES

In this section you will:

- understand that sexual and asexual reproduction enables species to survive
- asexual reproduction is fast but unless mutations occur, it does not produce genetic variation
- recognise that sexual reproduction involves fusion of gametes and results in genetic variation
- learn that the human menstrual cycle is controlled by hormones
- recall that the menstrual cycle involves hormones from the ovaries and pituitary gland
- understand that both hormone therapy and IVF assist people to conceive
- learn that angiosperms produce seeds inside ovaries that develop into fruits
- recognise that pollination may be carried out by the wind or an animal pollinator
- discover that to increase variation many plants have self-incompatibility mechanisms
- recall that seeds of plants contain the embryo plant and energy stores

- recognise that the onset of puberty is controlled by GnRH and LH and FSH from the hypothalamus and pituitary glands
- recognise adaptations of sperm and ova and difference in spermatogenesis and oogenesis
- learn that fertilisation involves mechanisms to ensure that only one sperm fertilises an ovum
- learn that the human menstrual cycle involves negative and positive feedback
- understand that the blastocyst implants in the endometrium during early pregnancy
- recognise that human chorionic gonadotropin maintains the placenta early in pregnancy and is used in pregnancy tests
- understand the role of the placenta in fetal development
- understand that birth is controlled by positive feedback involving oxytocin and estrogen
- consider the risks and benefits of hormone replacement therapy.

GUIDING QUESTIONS

- What are the advantages of sexual reproduction compared with asexual reproduction?

- Why are hormones important in controlling sexual cycles in animals?
- How do external factors assist in plant reproduction?

8.4.1 Asexual reproduction

Asexual reproduction does not involve the combination of genetic material from different individuals. A single organism produces copies of itself that are almost identical and are said to be a clone of one another. All the offspring will be genetically identical unless a mutation occurs. Asexual reproduction can take place in a number of different ways in different species.

- Parthenogenesis is the development of new individuals from an unfertilised egg (Section 6.5).
- **Budding** is a method of reproduction in which an individual develops a bud from part of its body. The bud then detaches and develops separately. *Hydra* reproduces in this way (Figure 8.4.1).
- Binary fission occurs when one individual divides into two equal parts (Section 6.5, Figure 6.5.1). Bacteria use this method of reproduction, and when conditions are favourable they may divide every 20 minutes. Fission can produce exponential growth and one bacterium dividing to produce 2 cells, and then 4, 8, 16, 32, and so on, could have the potential to produce 4×10^{21} cells in 24 hours, which explains how a pathogenic bacterium can cause illness in a very short time.

KEY POINTS

asexual reproduction reproduction that does not involve gametes or fertilisation; it involves only one organism in which there is no combination of genetic material from different individuals.

clone a population of genetically identical cells produced asexually produced by a single cell or organism.

Figure 8.4.1: *Hydra* reproduces asexually by budding.

Figure 8.4.2: Bread mould (*Rhizopus nigricans*) produces spores in round, black structures called sporangia.

- **Spores** are used by organisms, such as fungi, to reproduce asexually (Figure 8.4.2). Spores are small light structures that are produced in capsules called sporangia. They are dispersed easily to a new location where they germinate and start a new colony.

8.4.2 Sexual reproduction

Sexual reproduction involves the fusion of two **gametes** in a process called fertilisation. Gametes, which include sperm and eggs in animals, or pollen and ovules in plants, are haploid and contain the genetic material of only one parent. Sperm and eggs are found in many different animal species including corals, jellyfish, insects and vertebrates.

Gametes are produced by meiotic division. During the stages of meiosis, crossing over and the independent assortment of alleles increases genetic variation in the gametes. In addition, the combination of parental alleles during fertilisation ensures that sexual reproduction provides further genetic variation in the offspring. (These processes are discussed in Section 4.6 on inheritance and in Section 6.5 on cell division.)

In humans, the gametes that must come together to begin a new life are produced in the ovaries and testes. The male and female reproductive systems enable the gametes to meet and the female reproductive system provides a suitable place for fertilisation to occur and an embryo to develop (Figures 8.4.3 and 8.4.4). The ovaries and testes also produce hormones that regulate sexual development and reproduction.

EXAM TIP

You must be able to label diagrams such as those of the male and female reproductive systems shown in Figures 8.4.3 and 8.4.4. You should be able to annotate the important parts with information about their functions. The functions are summarised in Tables 8.4.1 and 8.4.2.

Roles of testosterone

The hormone **testosterone**, which is produced by the testes, has important roles in the sexual development and reproductive behaviour in males.

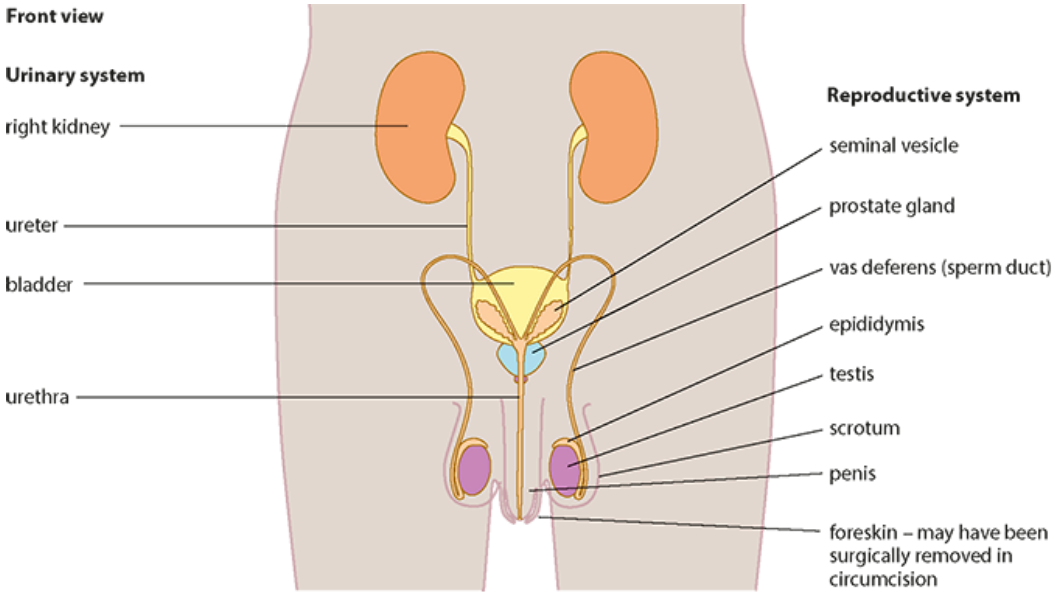


Figure 8.4.3: The male reproductive system. (The diagram also shows the organs of the urinary system.)

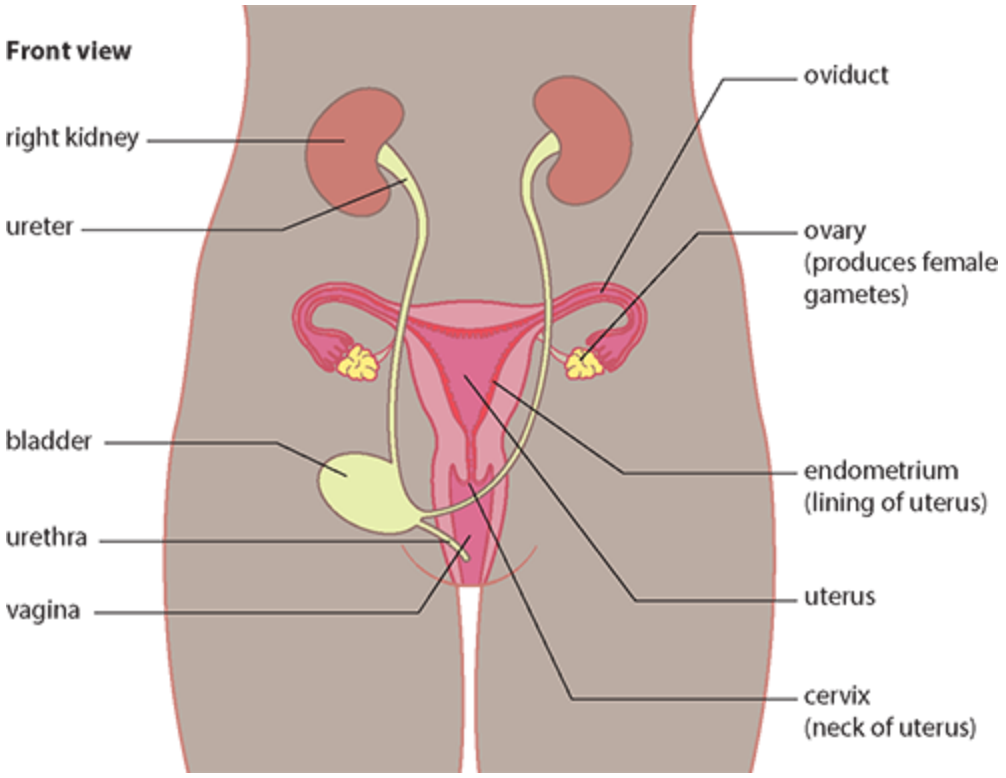


Figure 8.4.4: The female reproductive system. (The diagram also shows the organs of the urinary system – the bladder has been drawn to one side, to reveal the uterus.)

Structure	Function in male reproductive system
testes	sperm production and production of testosterone
epididymis and sperm duct	store male gametes (sperm) and carry them to the urethra
seminal vesicle, prostate gland	produce semen, the fluid in which sperm travel into the female body
urethra	tube that carries semen out of the body; separately, it carries urine out of the body

penis	enters the female body during intercourse to deliver semen to the vagina; contains erectile tissue to enable it to do this

Table 8.4.1: Summary of the functions of the structures of the male reproductive system.

- During fetal development, testosterone causes the development of the male genitalia.
- At puberty, levels of testosterone rise and cause the development of male secondary sexual characteristics including growth of muscle, deepening of the voice, enlargement of the penis and growth of body hair.

Structure	Function in female reproductive system
ovary	produces and releases female gametes (oocytes) and the hormones estrogen and progesterone
oviduct	site of fertilisation, carries oocytes to the uterus
uterus (womb)	place inside the female body where a fetus grows and develops
endometrium	lining of the uterus that receives a fertilised egg or is shed during menstruation
cervix	the neck of the womb, which remains closed during pregnancy
vagina	birth canal and area where semen are deposited during intercourse

Table 8.4.2: Summary of the functions of the structures of the female reproductive system.

- Testosterone stimulates the continuous production of sperm and behaviour associated with the sex drive.

Female hormones and the menstrual cycle

Humans and other primates secrete hormones that control the production and release of gametes in a cycle known as the menstrual cycle. The menstrual cycle involves changes to the ovaries and uterus to prepare for fertilisation.

Ovaries produce two hormones, **estrogen** and **progesterone**. These hormones stimulate the development of female genitalia before birth. At puberty they are responsible for the development of female secondary sexual characteristics including the onset of the menstrual cycle, development of breasts, growth of body hair and widening of the hips. The two hormones also influence the changes in the uterus lining during the menstrual cycle and pregnancy. The pituitary gland in the brain produces two further hormones: **luteinising hormone (LH)** and **follicle-stimulating hormone (FSH)**. FSH stimulates the development of immature follicles in the ovary, one of which will come to contain a mature egg cell. LH stimulates the follicle to release the egg and subsequently to form the **corpus luteum**.

Production of female gametes is a cyclical process, which lasts approximately 28 days. During the first half of this menstrual cycle the egg cell is produced and in the second half the uterus lining thickens to prepare for implantation of a fertilised egg. The cycle involves hormones that are released by the ovaries and the pituitary gland.

The sequence of events begins at the start of **menstruation**, which is often called a period (Figure 8.4.5). During the first 4 or 5 days of the cycle, the endometrium (lining) of the uterus is

shed and leaves the body through the vagina. This indicates that fertilisation has not occurred during the previous month.

In this early part of the cycle, the pituitary gland secretes FSH, which stimulates the development of an immature follicle in the ovary. The follicle then secretes estrogen, which enhances the follicle's response to FSH. Increasing levels of estrogen cause an increase in the level of FSH released by the pituitary gland. As the level of estrogen rises, it also stimulates the repair of the uterus lining.

As the follicle grows, estrogen levels rise to a peak at around day 12, when they stimulate the release of LH from the pituitary gland. As LH levels reach their highest point, ovulation – the release of the egg cell from the follicle – takes place. Ovulation usually occurs at around the day 14 of the cycle. Immediately afterwards, LH stimulates the empty follicle to form the corpus luteum. Levels of estrogen fall and as a result FSH and LH levels fall.

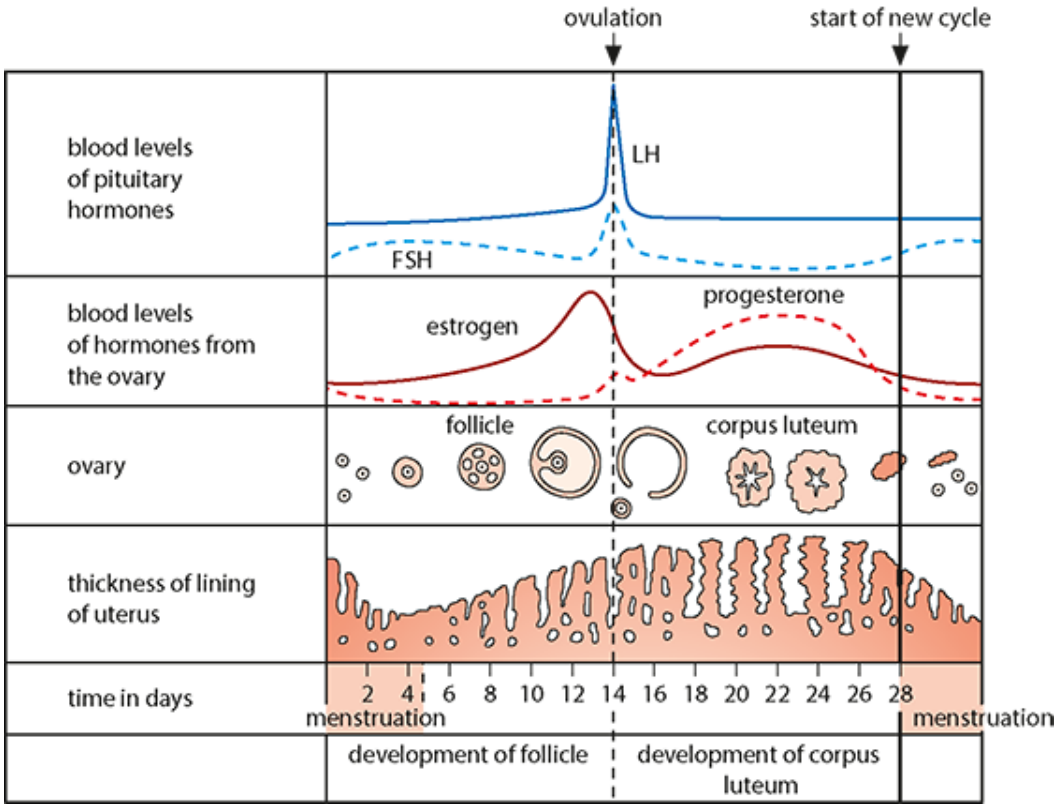


Figure 8.4.5: The menstrual cycle lasts an average of 28 days and involves changes in hormone levels that influence the follicles and lining of the uterus.

The corpus luteum secretes progesterone, which stimulates the thickening of the endometrium and prepares the uterus to receive an embryo. It also inhibits the production of FSH and LH.

If the egg cell is not fertilised, the corpus luteum degenerates and progesterone and estrogen levels fall. The fall in progesterone stimulates the breakdown of the uterus lining. FSH is no longer inhibited, so a new follicle is stimulated and the cycle begins again.

Human fertilisation

Fertilisation usually occurs in one of the oviducts and is the moment when one sperm cell fuses with the secondary oocyte to form a **zygote**. The sequence of events that occur at fertilisation is summarised in Figure 8.4.6.

During sexual intercourse, millions of sperm cells are ejaculated into the vagina and some of them make their way through the cervix and uterus towards the oviducts. Only a very small number of the ejaculated sperm will complete the journey, which is a considerable distance for the tiny cells.

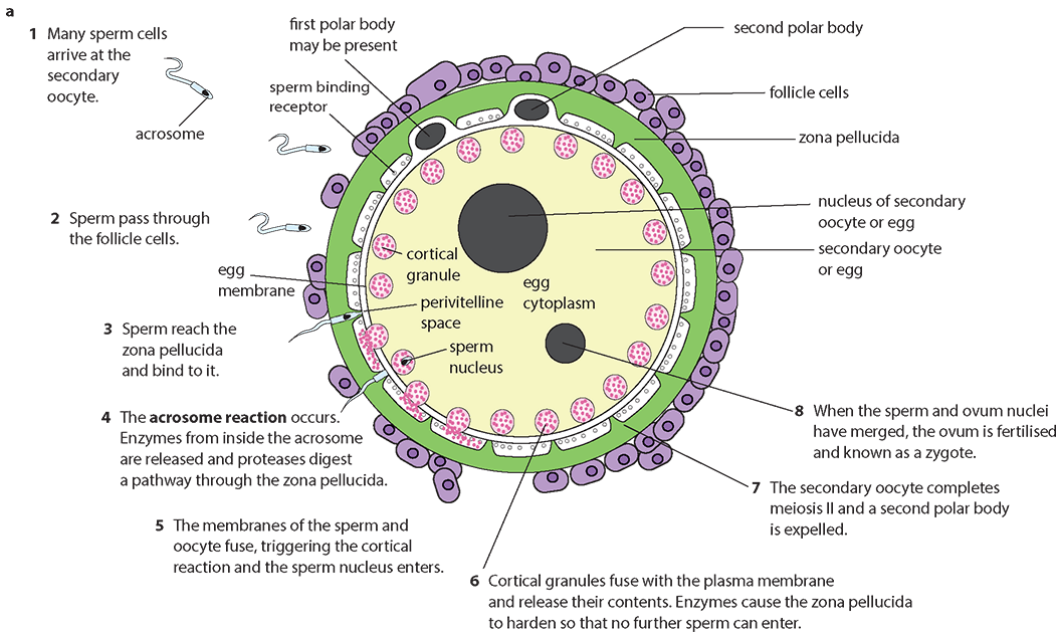


Figure 8.4.6: a The stages of fertilisation. Notice the difference in the sizes of sperm and oocyte.

As sperm cells approach the **zona pellucida**, the thick layer of glycoprotein that surrounds the egg, they go through a process known as the **acrosome reaction**. The contents of their acrosomes, which include many enzymes, are released to penetrate the outer layers of follicle cells covering the secondary oocyte and allow the sperm cells through. For fertilisation to be

successful, many sperm cells must be present to release the contents of their acrosomes, but only one sperm cell will eventually break through and reach the plasma membrane, where the membrane of its empty acrosome fuses with it.

After fusion has occurred, changes known as the **cortical reaction** take place in the membrane of the oocyte, which modify the zona pellucida and also prevent **polyspermy** (the entry of more than one sperm cell nucleus). Cortical granules, the enzyme-containing vesicles found just inside the oocyte, fuse with the plasma membrane in a cascade, away from the point of fusion of the sperm, and release their contents. Some of the enzymes from the cortical granules digest away the sperm cell receptor proteins on the oocyte plasma membrane so that no more sperm cells can attach and fuse.

SCIENCE IN CONTEXT

Oocyte, secondary oocyte or ovum?

Female gametes form in the ovaries during the fetal development of a baby girl. But all the eggs remain at prophase 1 of the first division of meiosis until puberty, and are known as **primary oocytes**. After puberty these primary oocytes continue their development and are released during the menstrual cycle. The immature ovum just after ovulation is known as a **secondary oocyte**. If it is fertilised it completes the final stage of meiosis and becomes a **mature ovum**.

During meiosis, when sperm cells are formed, four mature sperm that are equal in size are produced. But the formation of the ovum involves unequal divisions of the cytoplasm during meiosis. One large cell retains all the cytoplasm from the two divisions of meiosis. The cytoplasm contains materials to nourish a developing embryo and the unwanted chromosomes

are ejected. As the secondary oocyte is fertilised it ejects a polar body containing the final set of chromosomes but no cytoplasm.

8.4.3 Using hormones to treat infertility: *in vitro* fertilisation

***In vitro* fertilisation (IVF)** is a technique used to help couples who have been unable to conceive naturally. *In vitro* are Latin words that mean 'in glass' and the process of fertilisation usually takes place in a small glass Petri dish. There are many reasons for infertility. Males may have a low sperm count, blocked or damaged sperm ducts or be unable to achieve an erection. Females may fail to ovulate or have blocked or damaged oviducts, or produce antibodies in cervical mucus that destroy sperm.

The first step in IVF treatment is an assessment of whether the couple are suitable for treatment. If they are, the woman may first be injected with hormones to suppress her natural hormones before being given injections of FSH for about 10 days. In some treatments, FSH may be given alone. This hormone causes a number of egg cells to mature, all at the same time, in her ovaries (which is called superovulation). Just before the egg cells are released from the follicles, they are collected using a laparoscope (a thin tubular instrument that is inserted through an incision, or cut, in the abdominal wall). The egg cells are 'matured' in culture medium for up to 24 hours before sperm cells are added to fertilise them. Fertilised egg cells are incubated for about 3 days until they have divided to form a ball of cells. These embryos are checked to make sure they are healthy and developing normally. Usually two will be selected and placed into the woman's uterus for implantation. The pregnancy is then allowed to continue in the normal way. Any remaining embryos can be frozen and stored for use later. Figure 8.4.7 summarises the stages in IVF treatment.

THEORY OF KNOWLEDGE

Ethical issues associated with IVF treatment

IVF has enabled people who would naturally be infertile to have children, but it has also produced some serious ethical issues. Some of these are outlined in Table 8.4.3.

To consider:

1 Discuss these issues, which every society needs to think about.

2 Is it ever reasonable to deny treatments such as IVF to individuals who wish to have them?

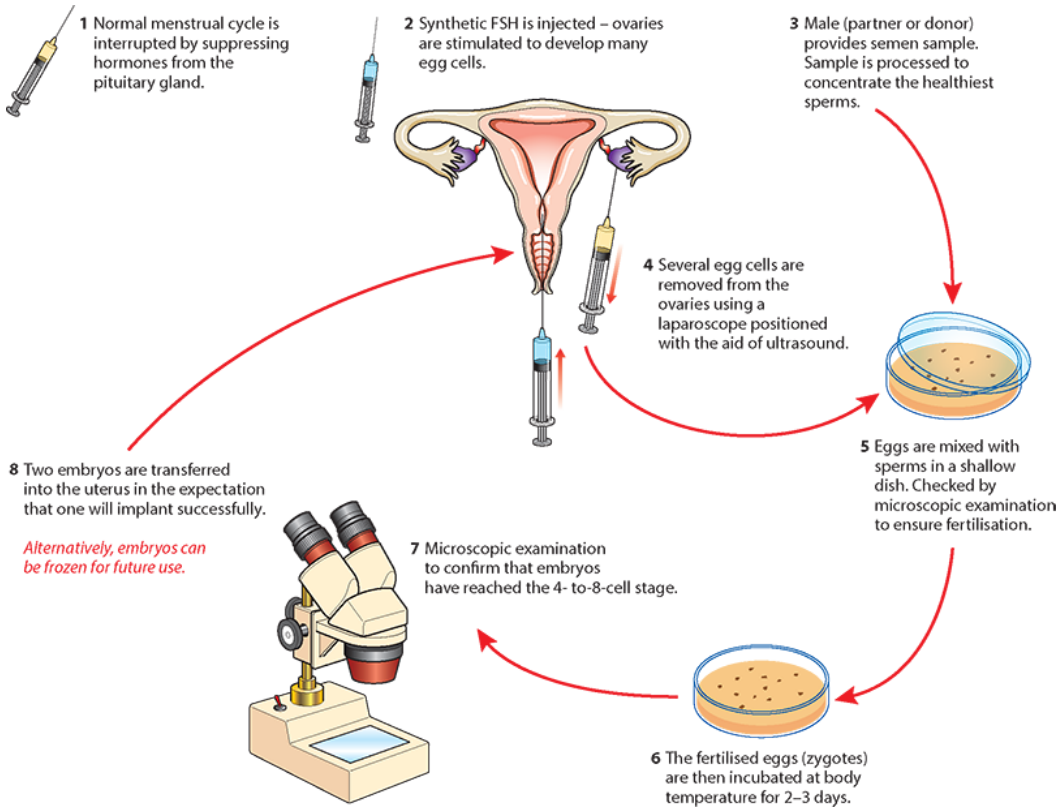


Figure 8.4.7: The stages of IVF treatment.

Arguments in favour of IVF	Arguments against IVF
enables infertile couples to have a family	unused embryos produced by IVF are frozen for a limited period and then destroyed
couples willing to undergo IVF treatment must have determination to become parents	multiple births often result from IVF and this increases the risks to mother and babies
embryos used in IVF treatment can be screened to ensure they are healthy and do not have certain genetic conditions that would be inherited	infertility is natural, whereas IVF is not: some religions object to it on this basis
IVF techniques have led to further understanding of human reproductive biology	some causes of infertility are due to genetic conditions, which may be passed on to children born as a result of IVF
	there may be risks to the health of women who are treated with hormones during IVF

Table 8.4.3: Arguments for and against IVF treatment.

TEST YOUR UNDERSTANDING

24 State two differences between sexual and asexual reproduction.

25 Name the four hormones that control the human menstrual cycle.

8.4.4 Hormonal control of developmental changes (puberty)

Puberty is the time when a boy or girl becomes sexually mature and capable of reproduction. The onset of puberty begins when the hypothalamus, a gland in the brain secretes gonadotropin-releasing hormone (GnRH).

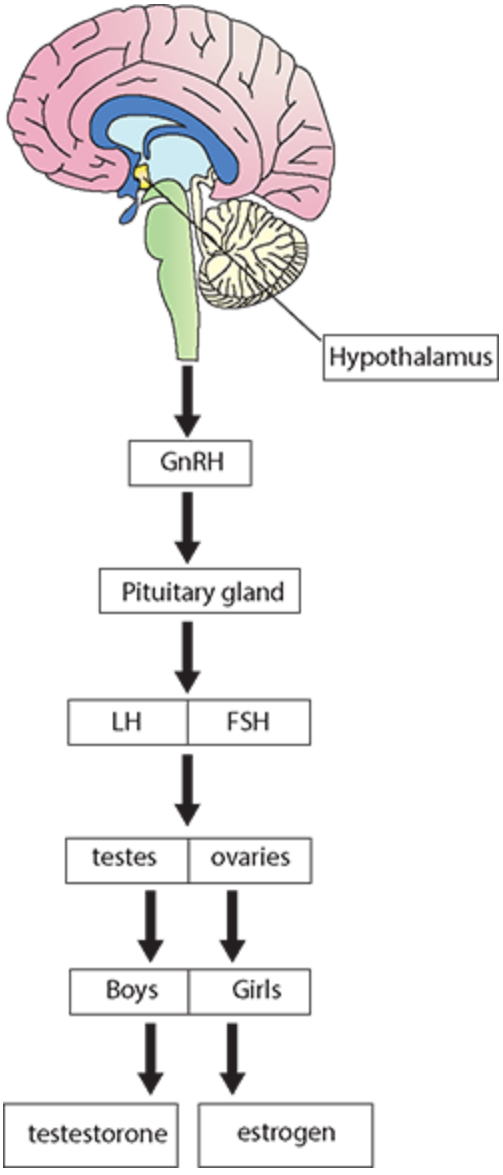


Figure 8.4.8: Hormonal control of puberty.

GnRH stimulates the release of **follicle-stimulating hormone** (**FSH**) and **luteinising hormone** (LH) from the anterior pituitary gland (APT). Before puberty, FSH and LH levels in the body are low. About one year before the onset of puberty, inhibition of GnRH by the nervous system is reduced so that there is an increase in the release of FSH and LH.

FSH and LH act on the ovaries and testes to stimulate the production and release of sex hormones estrogen and progesterone, and testosterone. Sex hormones have a negative feedback effect on the hypothalamus and pituitary gland to ensure circulating levels remain stable. A rise in FSH stimulates an increase in estrogen synthesis and oogenesis in females and the onset of sperm production in males.

Hormonal changes caused by rises in FSH and LH allow for the physical changes of puberty to begin. For males these include enlargement of the genitals and the growth of body hair and beard, while for females they include development of breasts and widening of the hips.

Spermatogenesis

Spermatogenesis is the production of mature sperm cells (spermatozoa) in the testis. More than 100 million sperm cells are produced each day in a process that takes place in the narrow seminiferous tubules making up each testis (Figures 8.4.9).

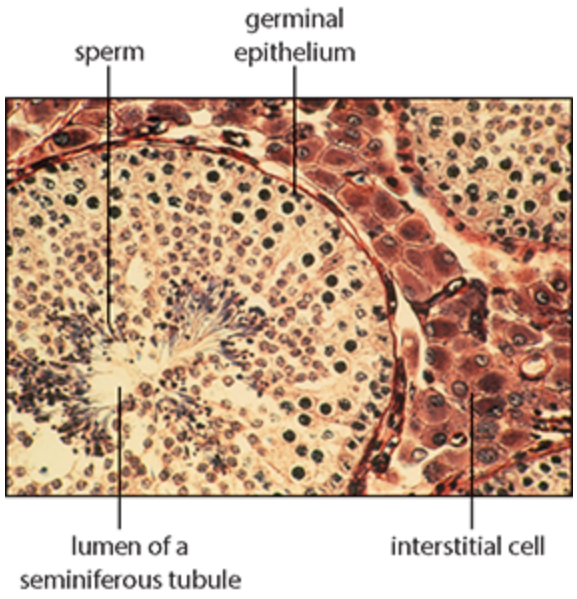


Figure 8.4.9: Coloured light micrograph of transverse section of a testis showing seminiferous tubules with interstitial cells between them (×170).

Sperm production and development takes place from the outer part of the **seminiferous tubules** towards the central lumen, where sperm cells are eventually released. Each tubule is enclosed in a basement membrane beneath which is an outer layer of germinal epithelium cells. These diploid cells ($2n$) divide regularly by mitosis to produce more diploid cells, which enlarge and are known as **primary spermatocytes**.

Primary spermatocytes divide by meiosis and their first division produces two haploid (n) cells called secondary spermatocytes. The second division of these two cells results in four **spermatids** (n).

Developing sperm are attached to **Sertoli cells** (Figure 8.4.10), which are also called nurse cells. These large cells assist the differentiation of immature spermatids into spermatozoa and provide nourishment for them.

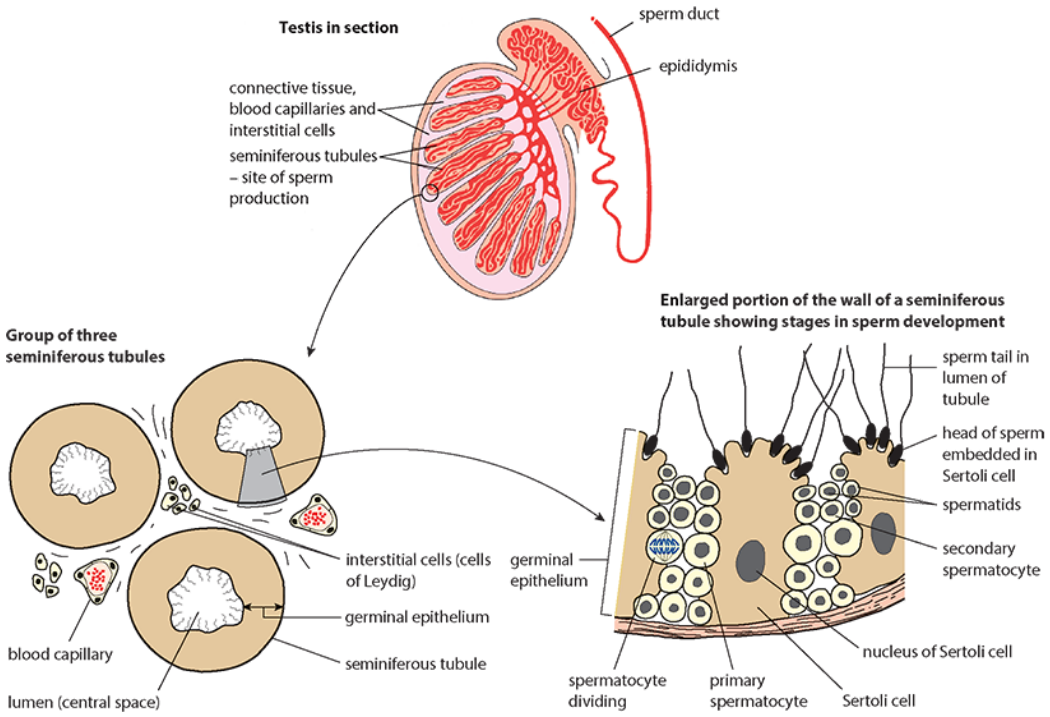


Figure 8.4.10: Structure of the testis.

Spermatozoa that have developed their tails (Figure 8.4.11) detach from the Sertoli cells and are carried down the lumen of the tubule to the epididymis of the testis.

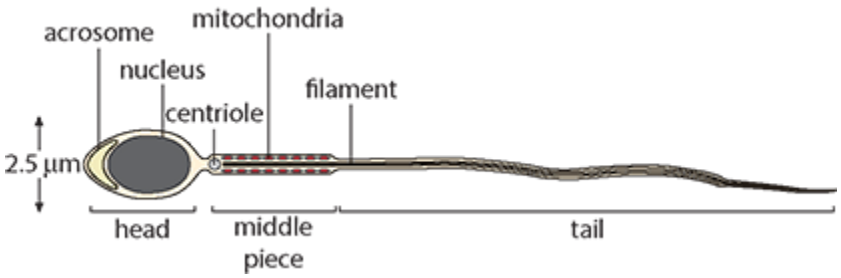


Figure 8.4.11: Structure of a human sperm cell. Total length is 60 µm.

Hormones and sperm production

Sperm production is controlled by three hormones – follicle-stimulating hormone (FSH) and luteinising hormone (LH) from

the pituitary gland, and testosterone produced by the testes.

- **FSH** stimulates meiosis in spermatocytes, to produce haploid cells.
- Testosterone stimulates the maturation of secondary spermatocytes into mature sperm cells.
- LH stimulates the secretion of testosterone by the testis.

Epididymis, seminal vesicles and semen production

Sperm cells are stored and mature in the epididymis, where they also develop the ability to swim. Sperm cells are released at ejaculation in a nutrient-rich fluid known as semen. Semen is produced by two seminal vesicles and the prostate gland. It is mixed with the sperm cells as they leave the epididymis and move along the vas deferens (sperm duct). Fluid from the seminal vesicles makes up about 70% of semen. It is rich in fructose, which provides energy for the sperm cells to swim, and it also contains protective mucus. The prostate gland produces an alkaline fluid that helps the sperm cells to survive in the acidic conditions of the vagina.

Oogenesis

Oogenesis produces female gametes, the ova. Unlike spermatogenesis, which takes place in an adult male, oogenesis begins in the ovaries of a female when she is still a fetus. **Oogonia**, the germinal epithelial cells within the ovaries of the female fetus, divide by mitosis to produce more diploid ($2n$) cells.These enlarge to form primary oocytes, which are also diploid Primary oocytes undergo the first stages of meiosis but this stops during prophase I, leaving the primary oocyte surrounded by a layer of follicle cells in a structure known as the

primary follicle. Development now ceases but the ovaries of a baby girl contain around 300000 primary follicles at birth. The remaining stages of oogenesis are shown in Figure 8.4.12.

At puberty, development of the primary follicles continues. During each menstrual cycle, a few follicles proceed to complete the first division of meiosis, although usually just one will complete its development. Two haploid cells (*n*) are produced but the cytoplasm divides unequally so that one cell is much larger than the other. The larger cell is known as the secondary oocyte (*n*) and the smaller cell is the polar body (*n*). The polar body degenerates and does not develop further.

The secondary oocyte, protected within its follicle, begins meiosis II but stops in prophase II. At the same time, the follicle cells divide and produce a fluid that causes the follicle to swell. At the point of **ovulation**, the follicle bursts, releasing the secondary oocyte, which floats towards the oviduct. Although ovulation is often described as the release of the ovum, the cell that is released is in fact still a secondary oocyte. The detailed structure of a secondary oocyte is shown in Figure 8.4.13, and Figure 8.4.14 shows secondary oocytes in a rabbit ovary in section.

After fertilisation, the secondary oocyte completes meiosis II, becoming a mature ovum, and expels a second polar body, which degenerates. The empty follicle in the ovary develops to become the **corpus luteum**, or ‘yellow body’, which produces the hormone progesterone.

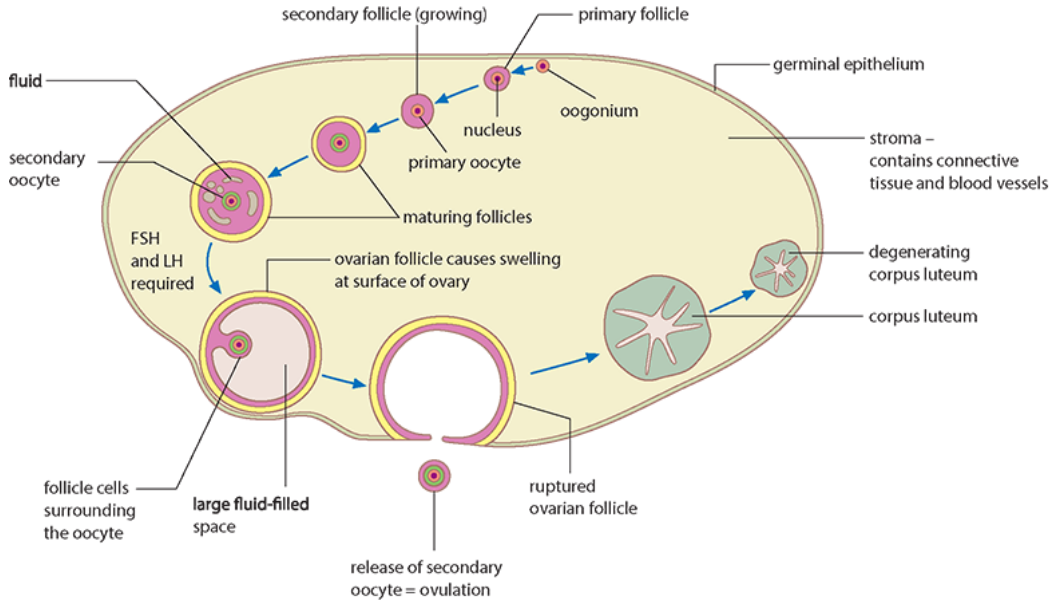


Figure 8.4.12: Stages in the development of one follicle in a human ovary. The arrows show the sequence of events.

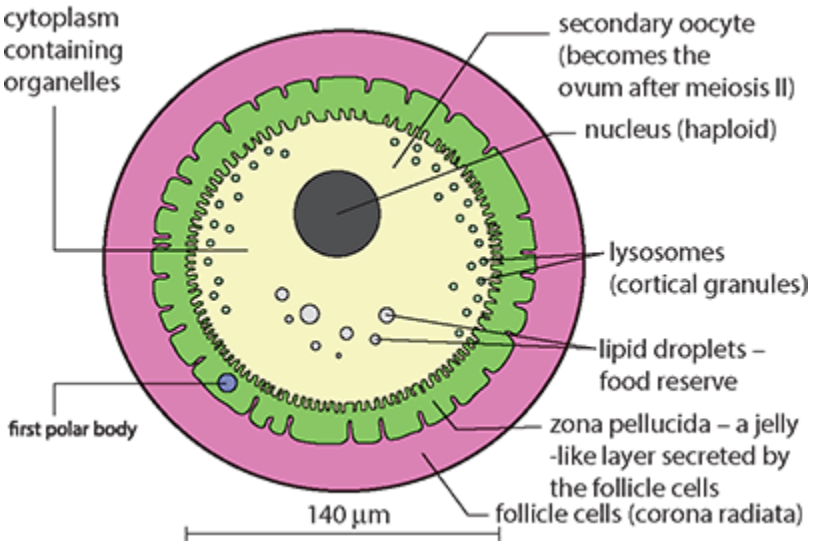


Figure 8.4.13: Structure of the secondary oocyte and ovulation surrounding structures at ovulation.

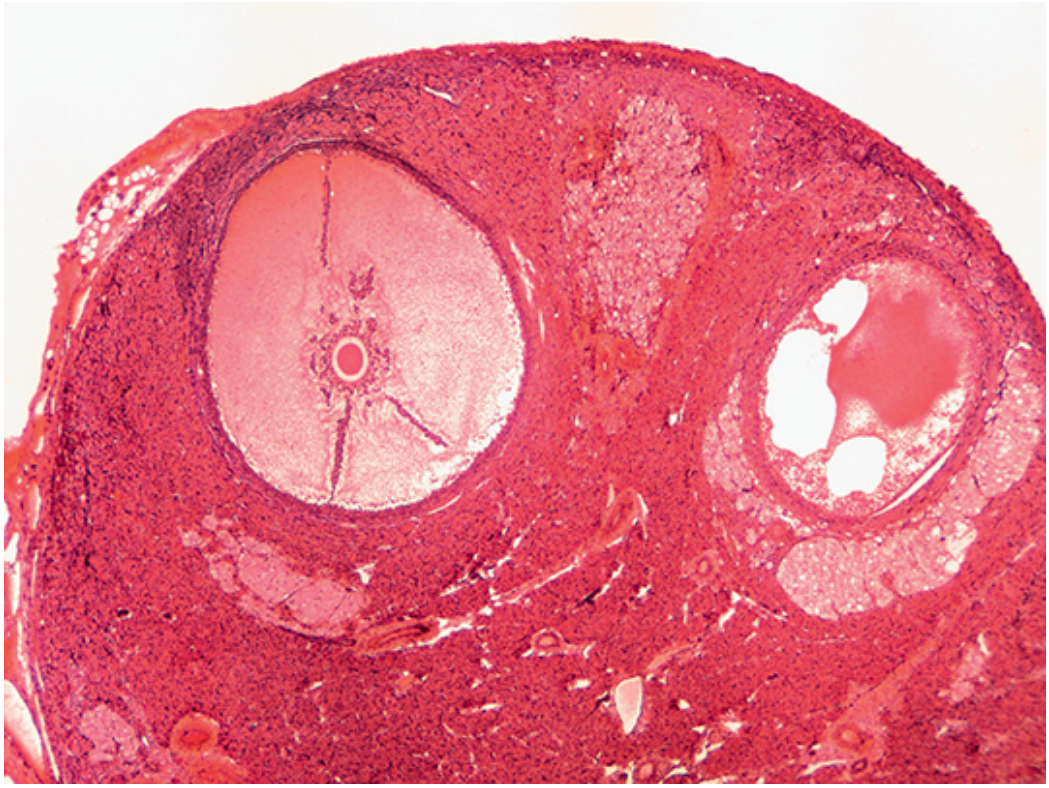


Figure 8.4.14: Longitudinal section of the ovary of a rabbit showing a mature follicle (×22.5).

Comparing spermatogenesis and oogenesis

There are a number of similarities and also several differences between the processes of spermatogenesis and oogenesis, as shown in Figure 8.4.15. Both involve the division of cells in the germinal epithelium by mitosis, and the growth of cells before they undergo meiosis and differentiation.

In both cases, meiosis produces haploid gametes. Table 8.4.4 summarises the differences and similarities in the two processes.

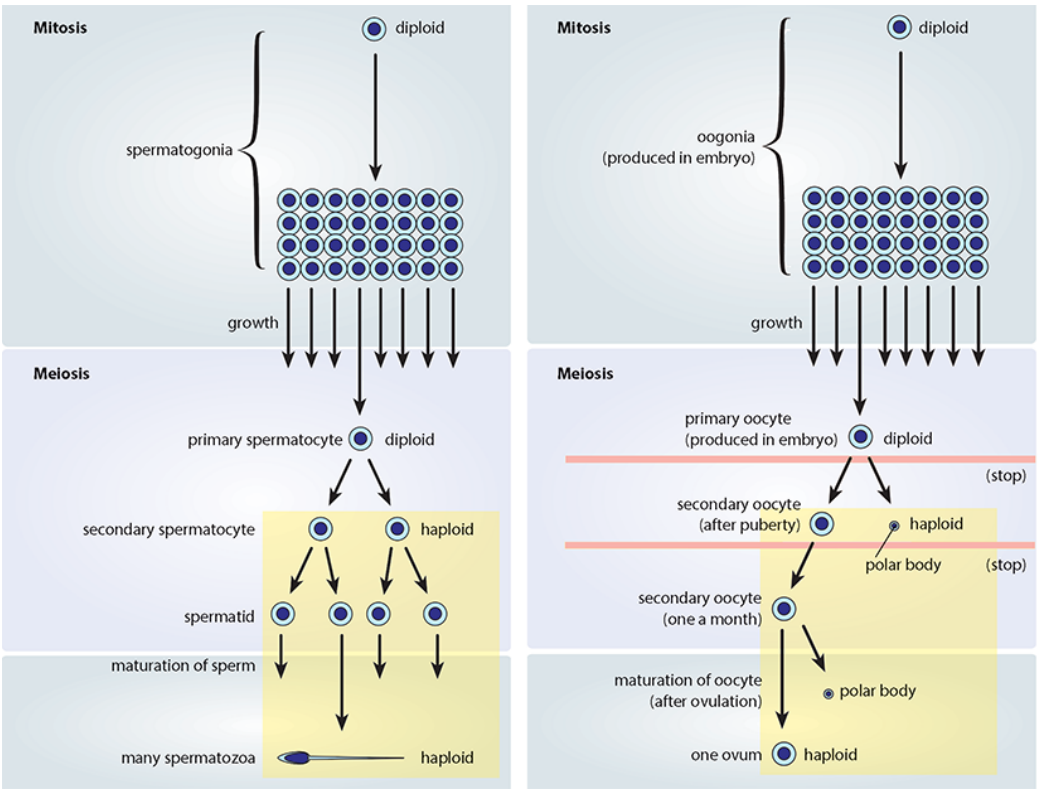


Figure 8.4.15: A comparison of spermatogenesis and oogenesis.

	Oogenesis	**Spermatogenesis**
Similarities	Both begin with production of cells by mitosis	
	In both cells grow before maturation	
	In both two divisions of meiosis produce haploid gametes	
Differences	Usually only one secondary oocyte is produced per menstrual cycle	Millions of sperm cells are produced continuously
	Only one large gamete is produced per meiosis	Four small gametes are produced per meiosis

	Occurs in ovaries which tend to alternate oocyte production	Occurs in tests which both produce sperm cells
	Early stages occur during fetal development	Process begins at puberty
	Ova released at ovulation during the menstrual cycle	Sperms cells released at ejaculation
	Ceases at menopause	Continues throughout an adult male’s life

Table 8.4.4: Oogenesis and spermatogenesis compared.

8.4.5 Pregnancy and prenatal development

Approximately 24 hours after fertilisation, the zygote begins to divide by mitosis. Mitosis continues and, after about 5 days of division, it produces ball of around 100 cells known as a blastocyst, as shown in Figure 8.4.16. As these divisions are occurring, the ball of cells is moved down the oviduct towards the uterus. After about 7 days it reaches the uterus and settles in the endometrium lining, where it implants itself and continues to divide and develop into an embryo.

Once the blastocyst has become established in the endometrium, it begins to secrete the hormone human chorionic gonadotropin (hCG). hCG travels in the bloodstream to the ovary, where its role is to maintain the corpus luteum, the mass of cells that developed from the empty follicle. The corpus luteum produces progesterone and estrogen, which in a non-pregnant woman maintain the endometrium until the end of the menstrual cycle, when the corpus luteum degenerates. During pregnancy, it is important that the lining remains in place. hCG stimulates the corpus luteum so that it grows and continues to produce its hormones for the first 3 months (the first trimester) of pregnancy. Thereafter, the placenta is fully formed and produces placental progesterone and estrogen, so the corpus luteum degenerates.

The embryo grows and develops. After about 1 month, the embryo is only 5 mm long but has a beating heart and the beginnings of a nervous system. From 2 months onwards it is known as a fetus. The fetus at this stage is 30–40 mm long and has recognisable limbs with developing bones. The uterus lining

provides nourishment for the early embryo but the placenta soon forms from the endometrium and fetal membranes and by about 12 weeks it is fully functioning. The fetus is connected to the placenta by the umbilical cord and is surrounded by a fluid-filled sac called the amnion, which contains amniotic fluid. The fetus is supported in this fluid throughout its development and is protected by it from bumps and knocks, as the fluid is an effective shock absorber. Amniotic fluid also enables the growing fetus to move and develop its muscles and skeleton. A human fetus will grow and develop for 40 weeks, or approximately 9 months, before it is born. This time between conception and birth is known as the gestation period.

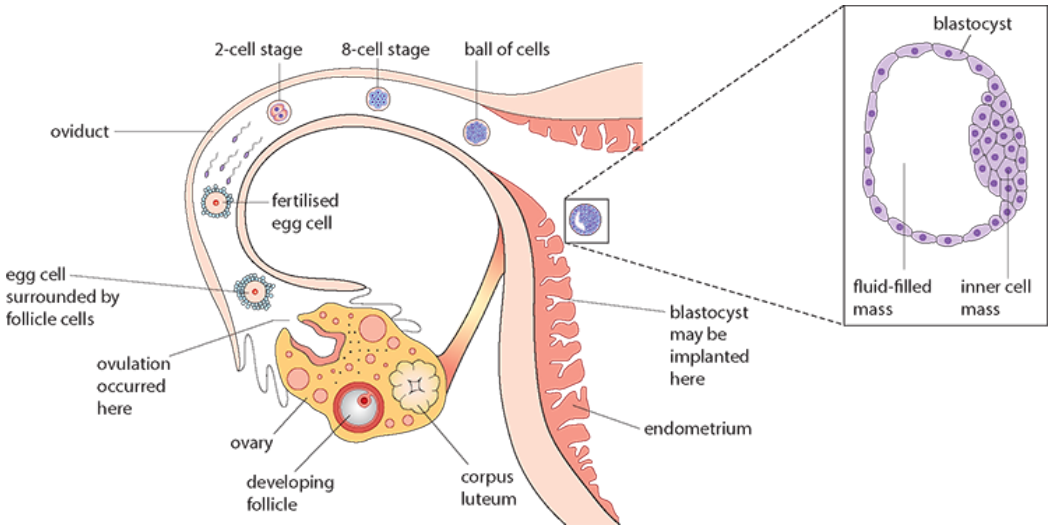


Figure 8.4.16: The blastocyst consists of an outer layer of cells enclosing an inner cell mass and a fluid-filled space. The outer layer forms part of the placenta and the inner cell mass develops to become the body of the embryo.

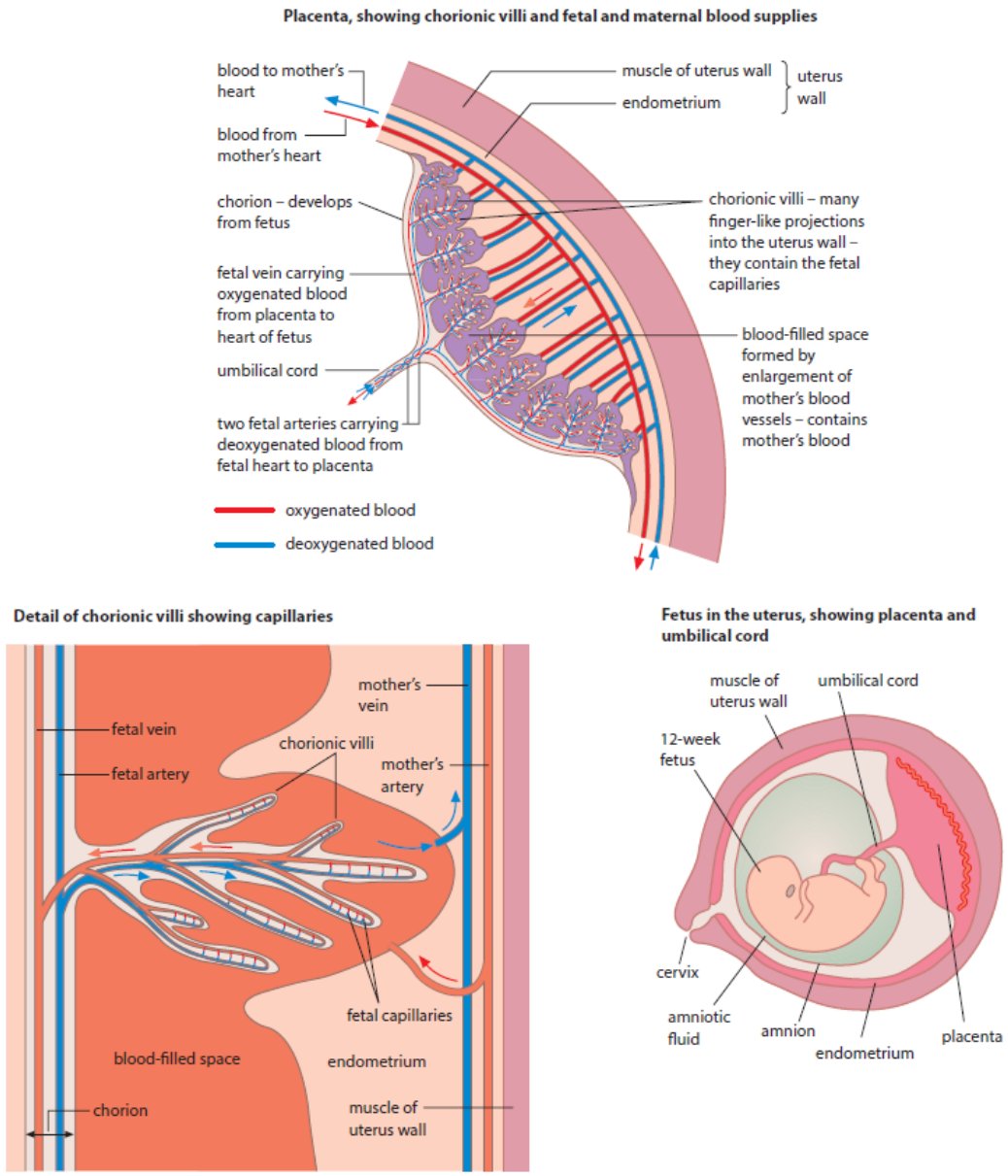


Figure 8.4.17: The position and structure of the placenta.

KEY POINTS

blastocyst ball of approximately 100 cells formed from the fertilised ovum.

fetus an unborn offspring that develops from an embryo; a developing human baby about 2 months after fertilisation.

human chorionic gonadotropin (hCG) hormone secreted by the uterus lining to maintain the early pregnancy.

placenta structure in the uterus formed of maternal and fetal tissue that provides oxygen and nutrients to and removes waste products from a developing fetus.

Pregnancy testing

hCG is excreted in the urine of a pregnant woman and it is this hormone that is detected in a home pregnancy test. Test kits contain monoclonal antibodies that bind to the hCG proteins. Kits produce a result as a colour change or other result that can be seen in just a few minutes.

The placenta

The developing fetus depends on its mother for all its nutrients and oxygen and for the disposal of its waste carbon dioxide and urea. The **placenta** allows these materials to be exchanged between the mother and the fetus and also acts as an endocrine gland, producing estrogen, progesterone and other hormones that maintain the pregnancy.

The placenta is a disc-shaped structure, about 180 mm in diameter and weighing about 1 kg when it is fully developed. It is made up of the maternal endometrium and small projections, or villi, from the outer layers of the **chorion**, which surrounds the embryo. These **chorionic villi**, which are rich in capillaries, grow out into the endometrium to produce a very large surface area for the exchange of gases and other materials (Figure 8.4.17). Fetal blood remains inside these capillaries, which penetrate the endometrium tissue until they are surrounded by maternal blood flowing into blood sinuses (spaces) around them.

In this way, the mother's blood is brought as close as possible to the fetal blood to allow for efficient diffusion without the two ever mixing.

Nutrients and oxygen from the mother's blood diffuse into the fetal capillaries and are carried to the fetus in a single umbilical vein. Waste products and carbon dioxide are carried to the placenta in the two umbilical arteries and diffuse into the mother's blood.

Many materials pass to the fetus from its mother. Some of these – such as drugs (both prescription and illegal), nicotine and alcohol – have the potential to seriously harm the fetus, which is why pregnant people are encouraged not to smoke or drink alcohol during pregnancy and to be careful with any medicines they may take.

8.4.6 Feedback mechanisms in the menstrual cycle and birth

Positive and negative feedback are both important in controlling the menstrual cycle and also the processes of birth.

Negative feedback is the most common type of feedback in living systems. In most cases the body will try to maintain a stable internal state. For example, negative feedback is important in maintaining blood sugar and body temperature at the correct levels to ensure the body functions properly (Section 8.5). During most of the menstrual cycle, the hormones estrogen and progesterone influence, or feed back to, the pituitary gland and regulate the levels of FSH and LH that are released. But during days 12–14 of the cycle a greater amount of these hormones is required to stimulate ovulation (the release of the egg from the ovary), so at this time estrogen has a positive feedback influence on the pituitary gland, so that more LH and FSH are released to stimulate the ovary (Figure 8.4.18). Positive feedback increases the deviation from the stable, steady state to produce a reaction, in this case ovulation.

In mammals, birth is mediated by a positive feedback relationship between the hormone oxytocin from the pituitary gland and contraction of the uterus. After about 9 months in humans, as the end of pregnancy approaches, the levels of progesterone and estrogen produced by the placenta fall (Figure 8.4.19) and this signals the onset of the uterine contractions known as labour. At this time, the endometrium secretes a group of hormones known as prostaglandins, which initiate the contraction of the uterus. Stretch receptors in the cervix then stimulate the hypothalamus, which triggers the release of the

hormone oxytocin, secreted by the posterior lobe of the pituitary gland in the brain. Oxytocin stimulates the uterus muscles to continue their contractions. At first, the contractions are mild and infrequent but oxytocin is a hormone that is controlled by positive feedback. A small contraction of the uterus muscle stimulates the release of further oxytocin, which in turn stimulates more and stronger contractions. As the uterus contracts, the cervix widens and the amniotic sac breaks, releasing the amniotic fluid. Contractions continue for several hours and the baby is pushed through the cervix and out of the mother's body down the vagina. Gentle contractions continue until the placenta, now known as the afterbirth, is also expelled from the uterus.

KEY POINT

negative feedback occurs when a deviation from the normal level is detected and corrective mechanisms in the body act to return the system to normal. Many homeostatic mechanisms, such as the control of body temperature, work by negative feedback.

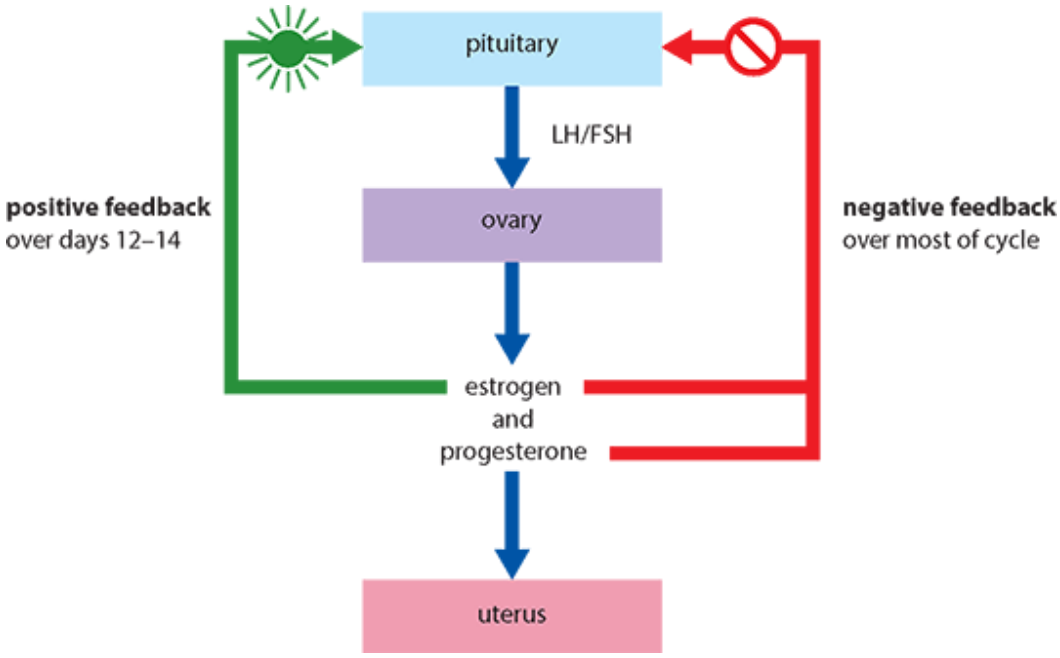


Figure 8.4.18: Negative and positive feedback in the menstrual cycle.

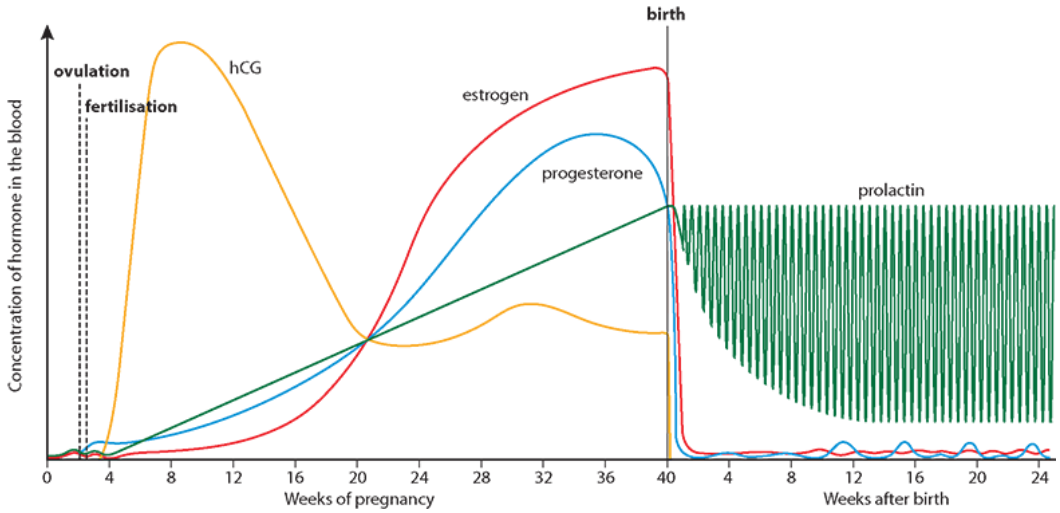


Figure 8.4.19: Changes in the levels of hormones during pregnancy and after the birth.

KEY POINTS

oxytocin hormone released by the pituitary gland in the brain that stimulates contraction of the uterus at birth.

positive feedback occurs when a deviation from the normal level in a system causes the change to increase, so that a reaction is amplified to make it stronger or occur more quickly; for example the contractions of the uterus during the birth of a baby. Positive feedback mechanisms are unusual in living organisms.

prolactin hormone stimulates milk production by the mammary glands.

After birth, blood levels of the hormone prolactin, from the anterior pituitary gland, increase. This hormone stimulates milk production by the mammary glands. As a baby suckles, prolactin secretion is maintained and oxytocin is also released from the posterior pituitary gland. Oxytocin causes milk to be released from milk ducts. This is another example of a positive feedback system. The more milk a baby drinks, the more prolactin is produced and thus more milk is produced. When a baby is weaned and stops suckling, milk, prolactin and oxytocin production cease.

NATURE OF SCIENCE

Hormone replacement therapy (HRT) - risks and benefits

Hormone replacement therapy is a treatment to relieve the symptoms of the menopause, the time when a woman's hormone levels decline and she stops menstruating. HRT replaces hormones that are at a lower level at the menopause. Some treatments contain combinations of estrogen and progesterone, others are estrogen only. HRT not only helps

with menopausal symptoms such as hot flushes but also helps prevent the thinning of the bones (osteoporosis) that is much more common after the menopause.

Early epidemiological studies published over the last 15 years suggested that HRT led to a reduced risk of coronary heart disease (CHD) and this was thought to be a cause and effect relationship. But later studies using randomised, controlled samples show that HRT actually leads to a slight but not significant increase in CHD and is not a cause and effect relationship. Most people who use HRT are in a higher socioeconomic group and this status does have a causal relationship with lower CHD risk. Recent evidence says that risks of CHD and strokes are small and are usually outweighed by benefits especially if treatment begins before 60 years of age.

SCIENCE IN CONTEXT

Diagnosing genetic diseases and congenital disorders before and after birth

Several genetic diseases can be diagnosed while a baby is still in the womb, so that parents will be aware of any conditions their child will have when it is born.

Amniocentesis is one of a number of techniques used in prenatal testing to check human fetuses for abnormalities. A fine needle is inserted through the mother's abdomen into the amniotic sac and a small sample of amniotic fluid is taken. The fluid contains fetal cells, which can be cultured for 3–4 weeks until the cells divide and chromosomes become visible. The chromosomes are stained to produce a karyogram (Section 4.5), which can be checked for mutations.

Karyograms can also be carried out after birth using cells from a child or adult (Section 4.5).

An alternative procedure, which can be done earlier in the pregnancy, is chorionic villus sampling (CVS). In this case, a sample of cells is taken for examination from the chorionic villi, via the cervix. More fetal cells are obtained in this way and the results are produced more quickly. However, CVS does have a greater risk of inducing a miscarriage than amniocentesis. Analysis of fetal genetic material can enable doctors to detect conditions including cystic fibrosis, Duchenne muscular dystrophy, sickle cell disease, Tay–Sachs disease and thalassemia.

Ultrasound scans that take place at several stages during pregnancy check that a baby is developing normally, but a scan can also reveal conditions such as spina bifida, heart defects and cleft palate if, for example, heart or kidney abnormalities or spinal problems are seen (Figure 8.4.20).

Ultrasound waves pass through body tissues and when they bounce back from different structures, they create echoes that are turned into a moving image of the developing baby. The image is displayed on a monitor while the scan is carried out.

As well as amniocentesis and ultrasound scans, pregnant people may also be offered a blood test, the maternal serum alpha-fetoprotein (MSAFP) test, between the 15th and 20th weeks of their pregnancy. The MSAFP test measures the level of alpha-fetoprotein, a protein produced by the fetus, in the mother’s blood. If levels are not normal it may indicate the possibility of Down syndrome or a neural tube defect such as spina bifida, which can then be confirmed by ultrasound or amniocentesis.

After a baby is born it can also be given a heel prick test. This involves taking a tiny sample of blood from the baby's heel. The blood is checked for inherited conditions including sickle cell disease, cystic fibrosis and phenylketonuria (an inability to break down the amino acid phenylalanine) and hypothyroidism (shortage of the hormone thyroxine). If these conditions are diagnosed early in the child's life they are easier to treat before serious problems arise.

Figure 8.4.20: An ultrasound probe produces high-frequency sound waves.

To consider:

1 Discuss the advantages and disadvantages of prenatal screening.

2 Should parents always be told about the gender of their unborn child?

TEST YOUR UNDERSTANDING

26 Outline the difference between positive and negative feedback in a living organism.

27 What is the role of hCG in early pregnancy?

28 Name two maternal hormones whose levels fall at birth and one whose level rises.

8.4.7 Sexual reproduction in plants

Angiosperms (flowering plants) produce haploid gametes in the male and female parts of their flowers (Figure 8.4.21), but the dominant generation of an angiosperm lifecycle is diploid.

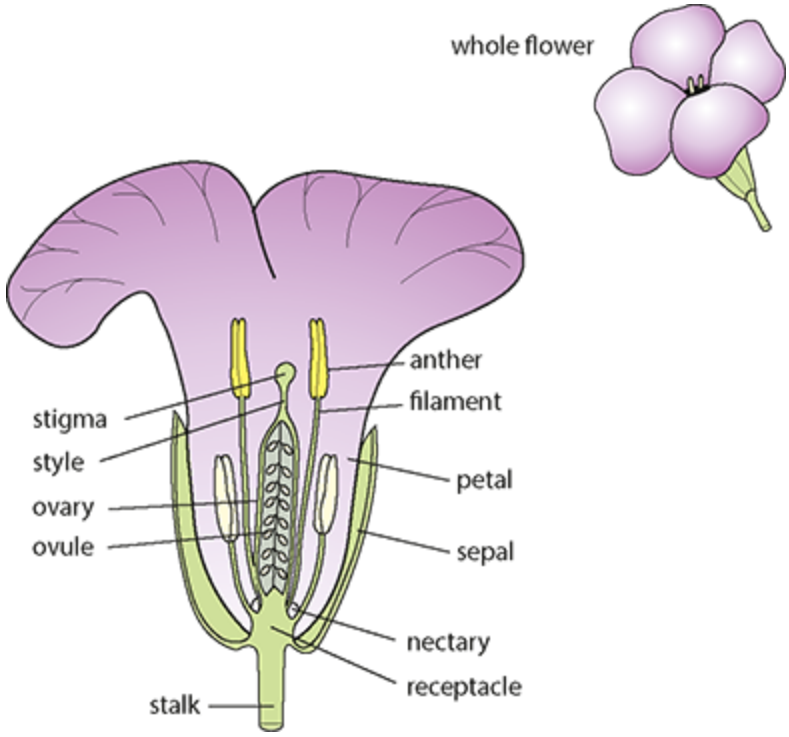


Figure 8.4.21: Half-flower of wallflower (*Cheiranthus cheiri*). The flower is about 2.5 cm in diameter. It is pollinated by bees and hoverflies. Its petals are usually brightly coloured and fragrant.

Male gametes are produced in anthers inside pollen sacs. Pollen grains are not independently mobile and must be carried by the wind or by an insect, bird or other animal to the female parts of a plant before fertilisation can take place. Female gametes are produced in the ovary.

Plants have adaptations to ensure that their pollen is transmitted to the female gametes (refer to Table 8.4.4). Petals of animal-pollinated flowers are often brightly coloured to attract insects or other animals that may visit. Many have nectar guides, which are markings on the petals that tempt pollinators deep into the flower.

Pollen, containing the male gametes, is produced in the anthers, which are held up on long filaments in many flowers, so that as pollinators enter they brush past the anthers and are dusted with pollen.

The female organs are the stigma, style and ovary. The stigma receives pollen grains, which arrive with pollinators as they delve into a flower to obtain nectar. The sticky stigma has sugars present on its surface that cause pollen grains to germinate.

Pollination and fertilisation

Pollination is the transfer of pollen from the anther to the stigma. Pollination occurs when pollen grains land on the stigma. Wind-pollinated plants rely on gusts of wind to transfer their pollen and are structurally different from animal-pollinated plants. They must produce a large amount of pollen because much of it is wasted and does not reach the stigma of another plant. It may be carried in the wrong direction or fall to the ground if there is no wind. Animal-pollinated plants must attract their animal pollinators to their flowers when the anthers are producing pollen. They have features such as scent, colour or a reward of nectar that make them attractive to their pollinators. Some of the differences between wind- and animal-pollinated flowers are summarised in Table 8.4.5.

Feature	Wind-pollinated	Animal-pollinated plant

	plant	
petals	small, inconspicuous	large and colourful, may be arranged to form a landing platform for visiting insects, or a tube for small birds
odour	not present	attracts specific pollinators, *e.g.* rotting smell attracts blow flies
nectar	not present	nectar produced inside the petals ensures pollinators partly enter the flower
stamens and pollen grains	long feathery stamens produce large amounts of light pollen	enclosed stamens produce sticky pollen grains
stigma	hang outside the flower and are often feathery to catch pollen	enclosed in the flower with a flat, sticky stigma

Table 8.4.5: Comparison of the features of wind- and animal-pollinated flowers.

If pollen travels from the anther of one plant to the stigma of another plant, the process is known as cross-pollination. If pollen is deposited on the stigma of the same plant that produced it, self-pollination occurs. Self-pollination produces less genetic variation than cross-pollination and different plants have different strategies to prevent it.

Preventing self-pollination

1 Many plants have **self-incompatibility** mechanisms to prevent inbreeding and increases genetic variation. If pollen grains from a plant land on the stigma of the same plant, or another plant with a very similar genetic make-up or genotype, incompatibility mechanisms prevent fertilisation. The germination of pollen may be prevented, or pollen-tube growth may be blocked, so that fertilisation and embryo development is prevented and no seeds are produced.

2 Other plants time the release of their pollen and the maturation of their stigmas so that they are not mature at the same time in the same flower. When pollen is released, the stigma will not be receptive so pollen will not stick to its own flower.

3 Some plants have separate male and female flowers, or male and female flowers on separate plants. Dioecious plants such as holly, papaya and kiwi have their male and female flowers on separate plants. Monoecious plants have separate male and female flowers on the same plant. Examples include squash and cucumber.

4 In some flowers, the positions of the stamens and stigmas prevent self pollination. The stamens are situated below the stigma and ovary so that pollen does not fall on its own flower. Examples include flowers of tulips, antirrhinum and tomato plants.

KEY POINTS

pollination the transfer of pollen from the anther to the stigma of a flower.

cross-pollination the transfer of pollen from the anther of one flower to the stigma of another flower on a different plant of

the same species.

dicotyledonous a plant that has a seed that contains two embryonic leaves or cotyledons.

incompatibility mechanisms methods used by a plant to reject pollen from the same plant or one that is genetically closely related.

self-pollination the transfer of pollen from the anther to the stigma of the same flower, or another flower on the same plant.

Most flowering plants use a mutualistic relationship with a pollinator to enable them to reproduce. Both the plant and pollinator benefit in this type of relationship. For example, a bee that visits a flower benefits by receiving food in the form of nectar, while the flower benefits as it receives pollen to fertilise its ovules. The pollen is carried on the bee's body from flowers it has already visited.

Fertilisation occurs when male and female gametes fuse to form a zygote. This occurs in the ovule of the flower. When pollen grains from a plant of the right species arrive on the stigma, they germinate and each produce a **pollen tube**, which grows down the style to the ovary (Figure 8.4.22). The tube enters the ovary and a pollen nucleus passes down the tube to fuse with and fertilise the nucleus of the female gamete in the ovule.

Seed structure and dispersal

Fertilised ovules develop over time into seeds, which protect the developing embryo inside. Seeds are held within a seedpod, fruit or nut, which can be dispersed to new locations so that when they

germinate the new plants that develop do not compete with their parents.

Seeds of flowering and non-flowering plants have all the necessary components to ensure successful germination and the growth of a new plant. Within every seed is an embryo root and shoot, ready to develop when the time is right. Once a seed has been formed in the ovary, it loses water so that it can enter a dormant phase and not develop further until conditions for growth are favourable.

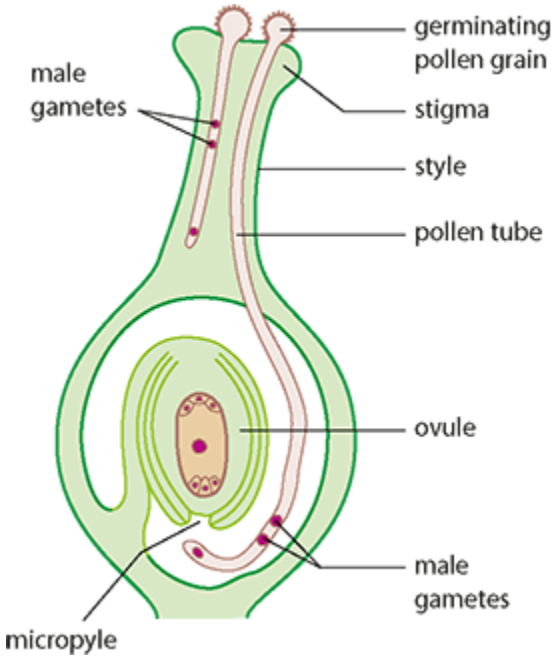


Figure 8.4.22: Fertilisation of an ovule in the ovary of a plant.

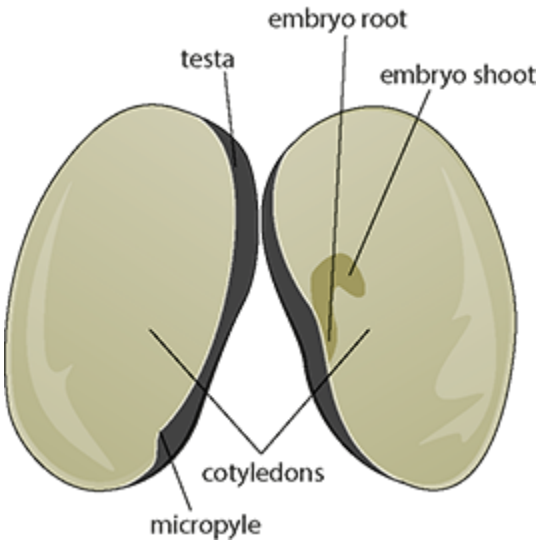


Figure 8.4.23: Two halves of a broad bean seed showing the main parts of a dicotyledonous plant.

Figure 8.4.24: Dandelion seeds can be dispersed over long distances by the wind.

Figure 8.4.25: Squirrels carry away acorns from oak trees and bury them in the ground.

Most plants have mechanism to ensure that they are dispersed at some distance from their parent plant. This reduces competition between parent and seedlings and prevents overcrowding of plants. Most seeds are dispersed by the wind or animals. Wind dispersed seeds have wings or parachutes to help them travel Fig 8.4.24. Animals may carry seeds away stuck to their fur, while some animals eat seeds but cannot digest them so they are deposited in faces some distance away. Squirrels carry nuts away and bury them to eat later but some are forgotten and are left to germinate. Fig 8.4.25.

One group of plants, the **dicotyledonous** plants have two embryonic seed leaves, or cotyledons, inside their seeds. These cotyledons store the food reserves needed for germination (Figure 8.4.23). Other plants such as grasses, bamboo and corn have only one cotyledon. The cotyledons are surrounded by a

hard protective seed coat called the testa. Many seeds have to endure quite harsh environmental conditions, so the testa protects the delicate tissues inside. In the wall of the testa is a pore called the micropyle through which water is absorbed to begin the process of germination.

Germination

Germination is the development of the seed into a new plant (Figure 8.4.26).A dormant seed needs three vital factors to be in place for germination to occur.

Temperature: A suitable temperature is essential for the enzymes in a seed to become active. They cannot work in cold conditions, and very high temperatures also inhibit their activity. Many seeds remain dormant until the temperature is at a particular level so that they germinate when the seedling will have the best chance of survival.

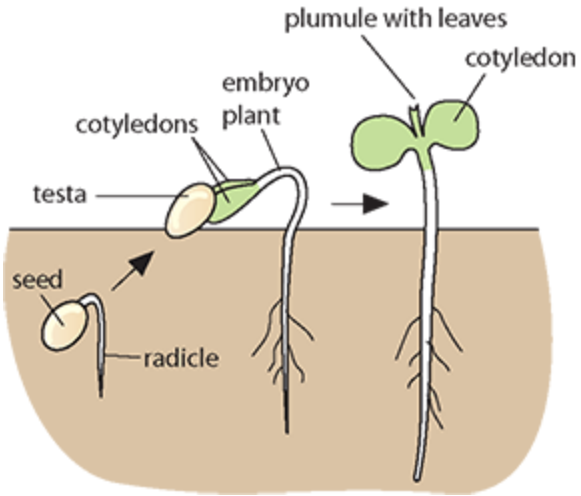


Figure 8.4.26: Germination and early growth in a dicotyledonous plant.

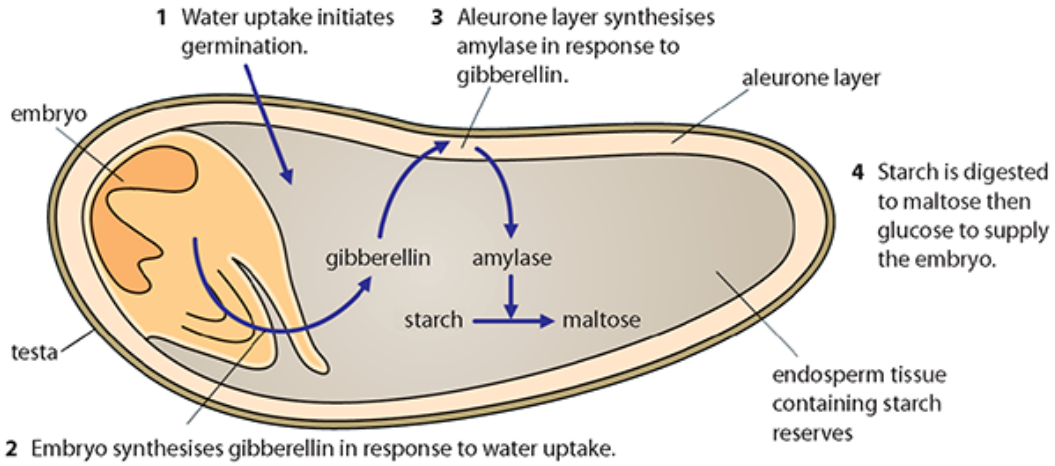


Figure 8.4.27: Longitudinal section through a barley seed, showing how secretion of gibberellin by the embryo results in the mobilisation of starch reserves during germination.

Water: Most seeds contain only about 10% water, so water must be taken in to start the germination process. Water rehydrates the seed and the enzymes contained within it. The enzymes break down food stores to provide energy for the emerging root and stem.

Oxygen: This is essential to provide energy for aerobic respiration.

Germination begins as water is absorbed by the seed in a process known as **imbibition**. Water enters through the micropyle of the testa.

Water rehydrates stored food reserves in the seed and, in a starchy seed such as a barley grain, it triggers the embryo plant to release a plant growth hormone called **gibberellin** (Figure 8.4.27).The gibberellin in turn stimulates the synthesis of **amylase** by the cells in the outer **aleurone layer** of the seed. The amylase hydrolyses starch molecules in the **endosperm** (food store), converting them to soluble maltose molecules. These are

converted to glucose and are transported to the embryo, providing a source of carbohydrate that can be respired to provide energy as the **radicle** (embryo root) and **plumule** (embryo shoot) begin to grow, or used to produce other materials needed for growth, such as cellulose.

Absorption of water by the seed splits the testa, so that the radicle and plumule can emerge and grow. When the leaves of the seedling have grown above ground, they can begin to photosynthesise and take over from the food store in the seed in supplying the needs of the growing plant.

Seeds contain the embryo plant and energy reserves to fuel germination. Once the seed has grown, the plant's reproductive cycle is complete.

TEST YOUR UNDERSTANDING

29 Name the female reproductive structures of an angiosperm (a flowering plant).

30 What is the advantage to a plant of a self-incompatibility mechanism?

REFLECTION

How well do I contribute to discussions about ethical issues that arise as a result of studying this topic?

Link

How do relationships between different species help in the reproductive strategies of some organisms? (Chapter 12)

8.5 Homeostasis

LEARNING OBJECTIVES

In this section you will:

- define homeostasis as the maintenance of a constant internal environment
- learn that temperature, blood pH and blood glucose levels are all maintained within certain limits
- recognise that feedback mechanisms allow organisms to detect, process and respond to changes to regulate internal conditions. Both nervous and hormonal control is used in feedback systems
- recognise that negative feedback returns a condition to the normal level but positive feedback make changes away from the normal
- understand the methods used by the human body to regulate body temperature

- understand the role of the kidney in osmoregulation and excretion
- recognise the functions of the glomerulus, Bowman's capsule, and loop of Henle in excretion and osmoregulation
- recall that, in animals, excess carbon dioxide leads to an increase in respiratory rate

- recall that, in plants, carbon dioxide and oxygen are regulated at the stomata
- recognise that low blood pressure is controlled by epinephrine
- learn that many diseases such as diabetes are a result of failed homeostasis.

GUIDING QUESTIONS

- How do organisms maintain constant internal conditions?
- How are changes to the external environment detected?
- How essential is homeostasis for survival?

8.5.1 Homeostasis

The internal environment of the bodies of most animals remains constant, within certain limits, despite changes that occur in the external environment. The control process that maintains conditions within these limits is known as homeostasis. The factors that are controlled include water balance, blood glucose concentration, blood pH, carbon dioxide concentration and body temperature. Each of these has a 'normal' or set point, although they may vary slightly above or below it. Both the nervous system and the endocrine system are involved in homeostasis. The **endocrine system** consists of ductless endocrine glands, which release different hormones (Figure 8.5.1). **Hormones** circulate in the bloodstream but each one is a chemical messenger that only affects the metabolism of specific target cells.

Feedback mechanisms

Feedback mechanisms allow organisms to regulate internal conditions and maintain equilibrium in response to changes in their surroundings or in their bodies. Any feedback system has three components:

1 receptors that detect a change from a normal set point

2 processing the information by a control centre, for example the brain

3 response to the change by effectors that receive signals from the control centre to return the system to the normal level (set point) or, in the case of positive feedback, to increase the deviation from the set point.

A feedback loop such as the one that regulates body temperature is a typical example of negative feedback. During temperature regulation, temperature receptors in the skin detect any changes and communicate information to the brain (the control centre) that processes the information and sends messages to the effectors, which are the blood vessels and sweat glands in the skin. These adjustments are made continuously to keep body temperature at the set point of 37 °C as the internal and external environment changes. Another example of a feedback loop is the regulation of blood glucose, shown in Figure 8.5.2 and discussed in detail in section 7.3.2.

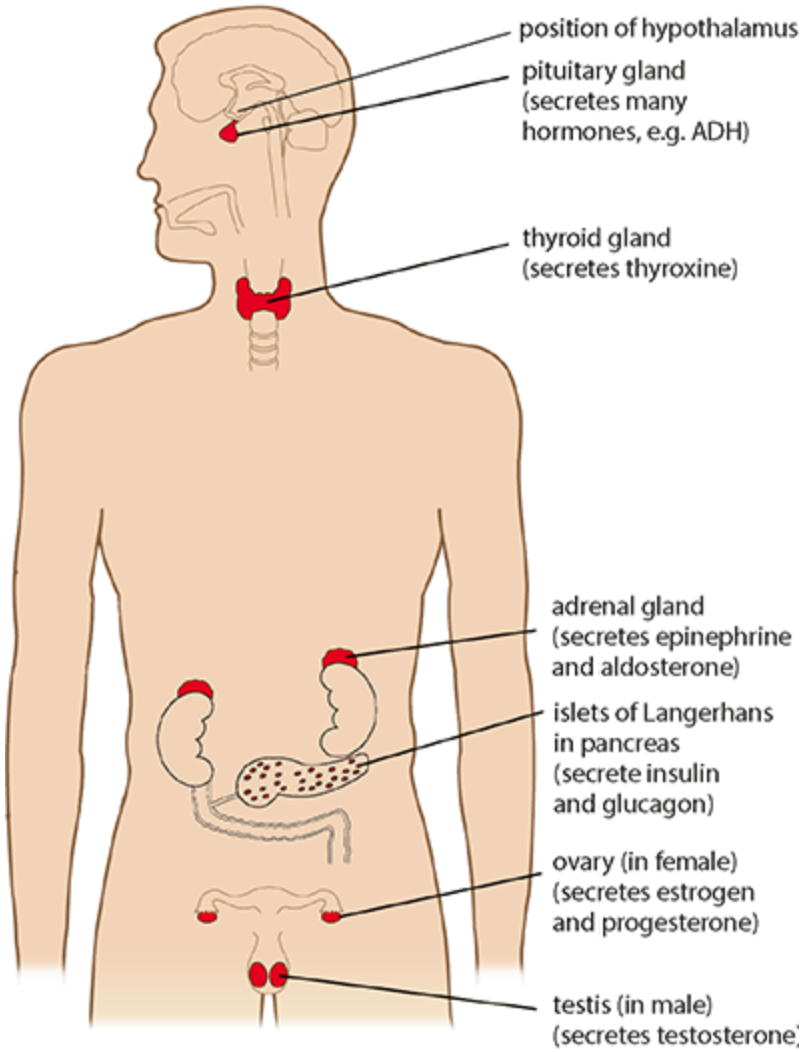


Figure 8.5.1: The positions of some endocrine glands in the human body. Endocrine glands have no ducts and secrete hormones directly into the bloodstream, which carries them to target cells.

KEY POINTS

endotherm animals that maintain a constant body temperature, independent of the environment.

homeostasis maintenance of internal environment of the body within constant limits, independent of the external conditions.

thermoregulation the ability to keep body temperature within certain limits when the surrounding temperature is different.

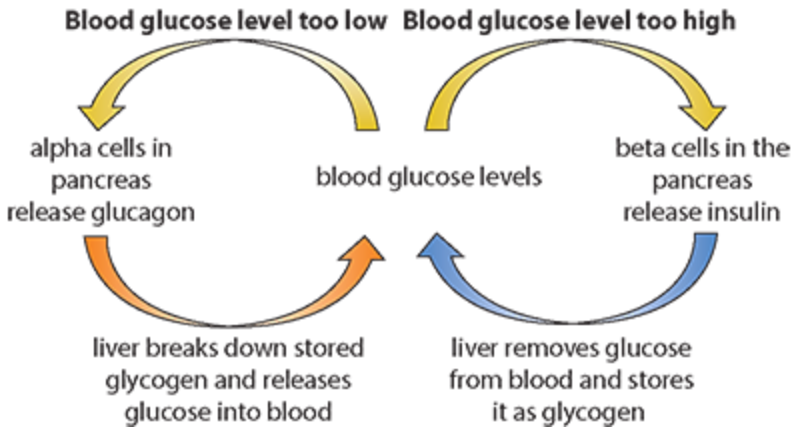


Figure 8.5.2: Control of blood glucose levels is achieved by negative feedback. Insulin produced by beta cells, and glucagon produced by alpha cells, of the islets of Langerhans control the storage or release of glycogen in the liver to keep blood glucose levels constant.

Positive feedback is less common in biological systems. One example of a positive feedback system is the one that controls the

contractions of the uterus when a baby mammal is born (Section 8.4). In this case there is a positive feedback relationship between the hormone oxytocin from the pituitary gland and contraction of the uterus. When a baby is large enough and ready to be born it creates pressure on the cervix (the neck of the uterus). Stretch receptors in the cervix stimulate the hypothalamus and this triggers the release of the hormone oxytocin from the pituitary gland. Oxytocin stimulates the uterus muscles so that they contract. At first, the contractions are mild and infrequent, but a small contraction of the uterus muscle stimulates the release of more oxytocin, which in turn stimulates more and stronger contractions. This cycle continues until the baby is born and contractions and oxytocin release cease.

Thermoregulation: control of body temperature

Endotherms can maintain a constant body temperature by both physiological and behavioural means. This is known as thermoregulation. In humans, body temperature is monitored and controlled by the hypothalamus in the brain. The ‘set point’ for human body temperature is 37 °C, which is optimum for the efficient functioning of the body’s enzymes and cell processes. The hypothalamus responds to nerve impulses from receptors in the skin and also to changes in the body’s core temperature. If body temperature fluctuates above or below the set point, the hypothalamus coordinates responses to bring it back to normal (Table 8.5.2). Body temperature can also be influenced by the hormone thyroxin which influences metabolism and how much blood vessels dilate and thus affects how much heat can be lost from the body. Cold temperatures also activate brown adipose (fat) tissue which breaks down blood sugar and fat molecules to create heat and help maintain body temperature. This is an example of negative feedback. Nerve messages are carried from

the hypothalamus to organs that bring about warming or cooling of the body. Table 8.5.1 lists some of the body's responses to changes in temperature.

	Responses to a rise in body temperature	Responses to a fall in body temperature
Arterioles in the skin	dilate (widen) so that more blood flows to skin capillaries and excess heat is lost from the skin, a process called vasodilation	narrow to restrict flow of warm blood to the skin capillaries: heat is retained in the body, a process called vasoconstriction
Sweat glands	produce more sweat, which evaporates from the skin surface to cool it	cease production of sweat
Muscles	remain relaxed	muscular activity such as shivering generates heat
Metabolic rate	may decrease to minimise heat production	thyroxin increases metabolic rate
Hair in the skin	hairs lie flat to reduce layer of insulation	hairs stand erect and create an insulating layer between skin and the air

Table 8.5.1: The body's responses to changes in core temperature.

Most mammals can sweat and use vasodilation (widening of the small blood vessels in the skin) to cool their bodies but there are many different behavioural strategies among different species that also assist in maintaining a constant temperature. Some animals, such as the jerboa, a desert rodent found throughout the Arabian Peninsula and northern Africa, live in burrows to avoid the heat of the day, only emerging when temperatures are cooler at night.

KEY POINTS

hypothalamus small area of the brain above the pituitary gland that releases neurohormones to stimulate or inhibit the pituitary gland and link the nervous and endocrine systems.

receptor an organ that senses changes in the environment and sends a signal to a control centre, usually the brain, which generates a response.

Most birds have a body temperature of 40 °C, higher than the set point of mammals, and many birds are able to withstand higher temperatures if they are exposed to the heat of the sun. In warm climates, many birds roost out in the open where they are exposed to high temperatures for long periods. They avoid becoming overheated by using evaporation to cool their bodies. These birds dissipate heat by gular fluttering (Figure 8.5.3). They open their beaks and increase the blood flow to the tissues of their mouth, then vibrate the moist gular (throat) area rapidly.

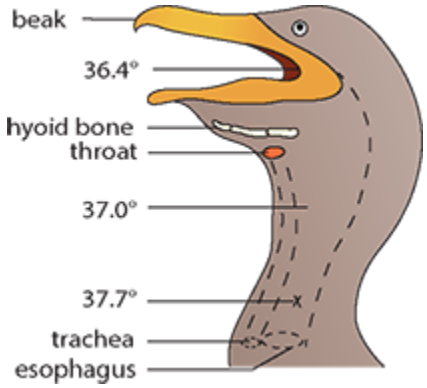


Figure 8.5.3: Gular fluttering can reduce a bird's temperature as water evaporates from the throat.

Camels also withstand extremely high temperatures in their desert habitats but they are still able to store fat that many other animals use as an insulating layer to keep them warm. Camels do this by storing fat in one place: in their humps. This uneven distribution of fat enables them to retain a store of energy that does not enclose and insulate their body against heat loss. Elephants use a different strategy to keep cool; their large ears have a huge surface area through which heat can be lost. Their ears are very thin and have rich supply of blood brings heat to the surface efficiently. Large ears can also act as cooling fans.

Another strategy involves losing heat by evaporation from moist surfaces. Animals such as dogs pant with their mouths open and tongues out when they are hot, to increase this evaporation.

In cold climates animals must work to conserve heat and protect themselves from the cold. Warm blood is retained inside the body as small blood vessels in the skin become narrower and restrict flow to the surface. This is known as vasoconstriction. Avoiding the cold can also help. Animals, such as the arctic hare, burrow under the snow where they are protected from strong winds and insulated from the cold. Humans wear clothes to keep

warm but birds grow soft thick layers of down feathers close to their skin to provide an insulating layer to their bodies. Feathers can be fluffed up to trap a layer of air between the bird's body and cold outside air. Mammals, such as snow leopards and polar bears, grow thick coats of fur that have the same role. Aquatic mammals such as seals and whales that live in cold oceans have thick layers of fat called blubber that enclose and insulate their bodies against the freezing water.

Mammals have a high metabolic rate, which helps to keep their body temperatures constant. However, it means that many mammals need a way of reducing their energy expenditure in cold winter months. Smaller mammals, such as bats and mice, have a large surface area to volume ratio and lose heat very quickly. This makes them especially vulnerable to cold weather so that they may need to adjust their behaviour and metabolism. One way in which small mammals overcome the problem of heat loss is to become less active. Some mammals let their body temperatures fall, either by going into hibernation or by entering a state of torpor. Torpor involves a short-term drop in body temperature that reduces the animal's metabolism whereas hibernation is a much longer period of inactivity.

EXAM TIP

Make sure that you are able to distinguish between physiological and behavioural methods that animals use to maintain their body temperatures.

Melatonin and control of circadian rhythms

Melatonin is a hormone produced by the pineal gland, a pea-sized gland located just above the middle of the brain. One of the key influences of melatonin is to maintain the body's circadian

rhythms and especially sleep–wake cycles. A vital factor in human sleep regulation is the exposure to light or darkness. Exposure to light stimulates a nerve pathway from the retina in the eye to the hypothalamus. Cells in the hypothalamus (the supra-chiasmatic nucleus, SCN) send signals to parts of the brain that control hormones, body temperature and other functions that have a role in our feelings of sleep or wakefulness. The SCN produces a signal that can keep the body on an approximately 24-hour cycle of activity. But the 'internal clock' is not exactly set to 24 hours and environmental clues, the most important of which is light, are needed to reset the clock each morning and keep a person in step with their external environment.

During the day the pineal gland is inactive, but during the hours of darkness, it is 'turned on' by the SCN and begins to produce melatonin, which is released into the blood. Rising levels of melatonin cause our feelings of sleepiness. The level of melatonin remains high for about 12 hours until the following morning when light causes it fall to a minimal level in the blood.

Even if the pineal gland is stimulated, it will not produce melatonin unless a person is in a dimly lit environment – even artificial indoor lighting can be bright enough to prevent the release of melatonin. The amount of melatonin released at night also varies between individuals, but on average children secrete more melatonin than adults.

Jet lag

Jet lag is caused by the disruption of the body's day–night (circadian) rhythms caused by long-distance travel and arrival in a new time zone. These rhythms indicate the right times for eating and sleeping, and also regulate hormone production. After a long journey the day–night patterns may no longer correspond

to the new environment or time zone and some people need many days to adjust.

Jet lag upsets the body clock because the expected patterns of light and darkness are out of alignment. Light is the strongest stimulus for the sleep–wake pattern so jet lag can be controlled by avoiding bright light so that the body clock is reset. Melatonin tablets are sometimes used to adjust a person's body clock but the effectiveness of these treatments is not proven. The effect of melatonin may be very short term and the correct doses and times to take the hormone are not easy to determine. Melatonin is not approved for sale in some countries. In the USA is can be bought in pharmacies but in other countries it is only available with a doctor's prescription.

Control of blood glucose levels

Blood glucose level is the concentration of glucose dissolved in blood plasma. It is expressed as millimoles per decimetre cubed (mmol dm^{-3}). Normally blood glucose levels stay within narrow limits, between 4 and 8 mmol dm^{-3}. This ensures that the osmotic balance of the blood remains constant and body cells receive sufficient glucose for respiration. Levels are higher after meals because glucose is absorbed into the blood from the intestine. They are usually lowest in the morning, as people do not eat overnight when they are asleep.

Glucose levels are monitored by cells in the pancreas. If the level is too high or too low, cells in regions of the pancreas known as the islets of Langerhans produce two hormones, insulin and glucagon, that turn on control mechanisms to correct it. This is another example of negative feedback (Figure 8.5.2). The regulation of blood glucose and how insulin influences cells is described in Section 7.3.2.

SCIENCE IN CONTEXT

The international standard measure for blood glucose levels is millimoles per decimetre cubed ($mmol \cdot dm^{-3}$). For humans the normal range is between 4.4 and 6.1 $mmol \cdot dm^{-3}$. In the United States and some other countries it is usually measured in milligrams per decimetre cubed ($mg \cdot dm^{-3}$), making the normal range for blood glucose 80–110 $mg \cdot dm^{-3}$. The molecular weight of glucose is 180 and so the difference between the two units is a factor of 18. Thus, 1 $mmol \cdot dm^{-3}$ of glucose is equivalent to 18 $mg \cdot dm^{-3}$.

Range of tolerances

Although many organisms are able to control the internal conditions of their bodies, different organisms have different ranges within which they can survive.

The body temperatures of different mammals and birds vary, as shown in Table 8.5.2. Each species survives best at its own set point.

Organism	Normal average body temperature / °C
human	37.0
bird	40.0
horse	38.0
sheep	39.0
elephant	36.5

Table 8.5.2: Average body temperatures for different species.

These ranges of tolerance cannot be exceeded if the animal is to survive. For example, humans cannot survive for long if body temperature falls to less than 35 °C. We suffer from hypothermia because our metabolism slows down too much to allow chemical reactions to take place at this temperature. And if our body temperature rises to over 40 °C for more than a very short time, humans can suffer from hyperthermia (heatstroke), as enzymes cannot function properly. Untreated hyperthermia causes many deaths each year.

Blood glucose levels are also different in different organisms. For example, a horse has a glucose level of between 3.3 and 6 $mmol \cdot dm^{-3}$, but a chicken has a normal level of between 7.2 and 14.8 $mmol \cdot dm^{-3}$. Excessively high or low blood glucose levels in humans can cause tiredness, a coma or weakness and blurred vision. In the most serious cases a diabetic coma can be fatal. Other animals are affected in a similar way.

THEORY OF KNOWLEDGE

Defining key terminology is generally seen as the starting point for academic enquiry. How could knowledge be gained without definition of key terms?

1 Why do you think it is important to define key terms before examining an issue or topic in more detail?

2 What problems might arise if terms were not clearly defined?

TEST YOUR UNDERSTANDING

31 Define homeostasis.

32 List three ways in which a human can regulate body temperature.

33 Draw a simple diagram to show how a feedback loop controls blood glucose levels.

34 Name a process that involves a positive feedback loop.

8.5.2 The role of the kidneys in osmosregulation and excretion

KEY POINTS

Osmoregulation is the control of the water potential of body fluids by the regulation of water and salt content, usually measured in osmoles per litre (osmol L–1)

Excretion is the removal from the body of waste products of metabolic pathways via the skin, kidneys and lungs

Humans, like all mammals, use kidneys to regulate the water content of our bodies. We have two kidneys, situated in the lower back, one on either side of the spine. Each receives a blood supply from a renal artery, which is a branch of the main artery from the heart, the aorta. After passing through the kidney, blood returns to the circulation via a renal vein that joins the vena cava. Because of the processes occurring in the kidney, the composition of the blood in the renal vein is quite different from that in the renal artery. Urea, water content and salt levels are adjusted by the kidney so that they are at the correct levels as blood leaves the kidney, but glucose, protein and the cellular content of blood remain unchanged. Figure 8.5.4 shows a kidney in longitudinal section. Three regions are visible – the outer cortex, the central medulla and the inner renal pelvis. Urine produced by the kidney collects in the renal pelvis and is carried down to the bladder in the ureter.

Each kidney is made up of more than 1 million tiny structures called nephrons. These are the functional units of the kidney, selectively filtering and reabsorbing substances from the blood.

Figure 8.5.4 shows the structure of a **nephron**, which consists of a filtering unit (a complex of capillaries called a **glomerulus** surrounded by a **Bowman's capsule**) together with a tube that extends from the filtering unit to the renal pelvis. This tube is divided into four regions – the proximal convoluted tubule, the **loop of Henle**, the distal convoluted tubule and finally a collecting duct. Each of these regions has a specific role to play in urine formation.

Kidneys filter blood continuously. The blood that comes into the kidney enters millions of tiny kidney tubules called nephrons. At the start of each nephron is the glomerulus, or a tiny knot of capillaries, where blood is filtered so that only plasma, minus large proteins, enters the tubule. Each nephron has several sections: the proximal convoluted tubule, the U-shaped loop of Henle with a thin descending and a thick ascending limb, and the distal convoluted tubule, which winds and twists back up again, before emptying into the collecting duct, which collects the final urine (Figure 8.5.5).

Osmoregulation

The final adjustments to the water content of urine produced by the kidneys are made in the collecting duct, as described earlier in this section. But the process of osmoregulation also occurs in other parts of the kidneys to ensure that the solute and water balance of the body is kept constant. Each kidney is made up of more than 1 million tiny structures called nephrons. These are the functional units of the kidney, selectively filtering and reabsorbing substances from the blood. Figure 8.5.5 shows the structure of a nephron, which consists of a filtering unit (a complex of capillaries called a glomerulus surrounded by a Bowman's capsule) together with a tube that extends from the

filtering unit to the renal pelvis. This tube is divided into four regions:

1. the proximal convoluted tubule,
2. the loop of Henle,
3. the distal convoluted tubule and finally
4. a collecting duct.

Each of these regions has a specific role to play in urine formation.

Bowmans capsule and proximal convoluted tubule

Ultrafiltration occurs in the glomerulus where blood plasma is filtered. Only small molecules can pass through the Bowman's capsule into the nephron; water, salts, glucose and amino acid enter the nephron and the amounts of each are unregulated. This means that the filtrate contains many useful substances which the body needs and these are reabsorbed in the proximal convoluted tubule. Cells in this region of the nephron have many mitochondria. They fuel active transport through membrane pumps that selectively reabsorb ions and glucose from the tubular fluid. All the glucose in the filtrate is actively reabsorbed together with almost 80% of sodium (Na^+), potassium (K^+), magnesium (Mg^{2+}) and calcium (Ca^{2+}) ions. Chloride ions (Cl^-) are absorbed passively and water follows by osmosis as the solute concentration of the cells rises due to the active uptake of ions and glucose. The remaining filtrate now moves into the loop of Henle.

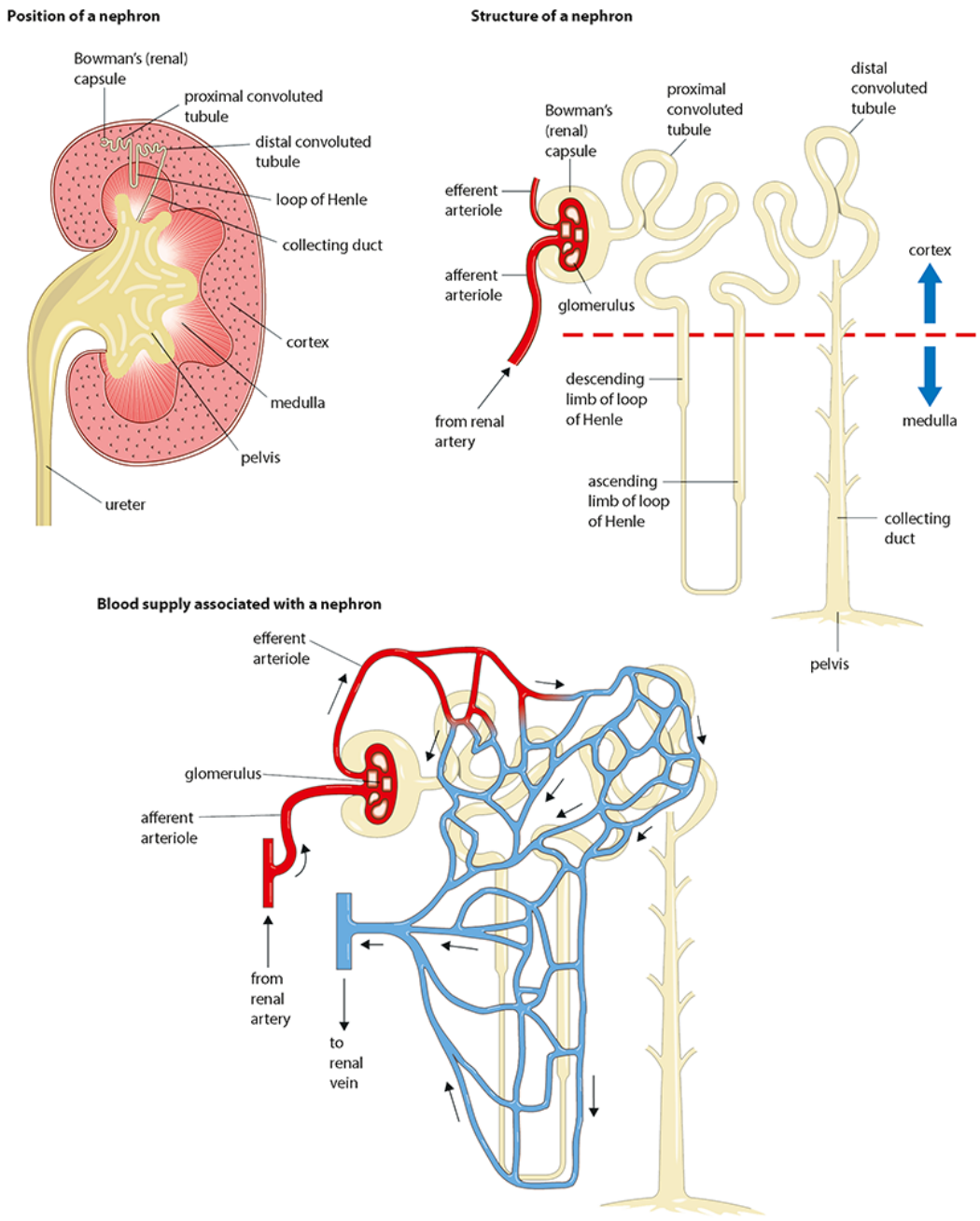


Figure 8.5.4: The structure of the kidney and nephron.

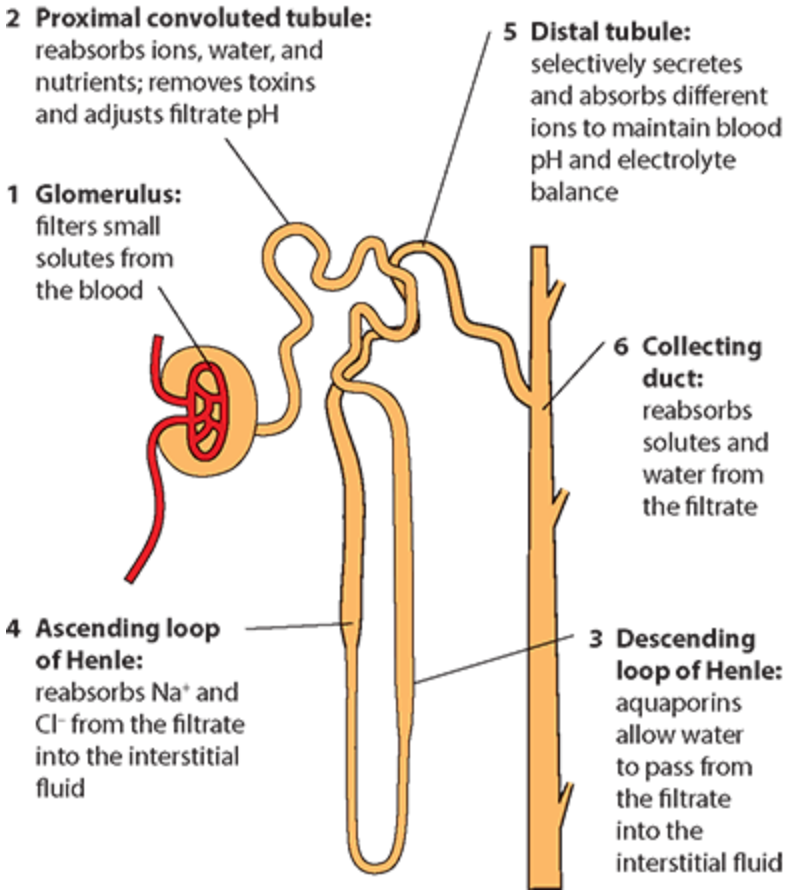


Figure 8.5.5: The structure of a nephron and the functions of the different parts of the tubule.

The loop of Henle

The filtrate that enters the loop of Henle still contains much of the water that was filtered from the blood. The wall of the descending limb of the loop is permeable to water but relatively impermeable to salts, whereas the ascending limb is impermeable to water, but allows salt to be passed through its walls. Na^+ and Cl^- ions move by active transport out of the ascending limb into the tissue fluid of the medulla, creating hypertonic conditions (a high salt concentration) in this region. This means that as the descending limb of the loop of Henle passes down into the medulla, water leaves passively by osmosis

and enters the surrounding blood capillaries (Figure 8.5.6). The hypertonic environment in the medulla of the kidney produced by the loop of Henle is also essential for the fine-tuning of the water content of the blood by the collecting duct.

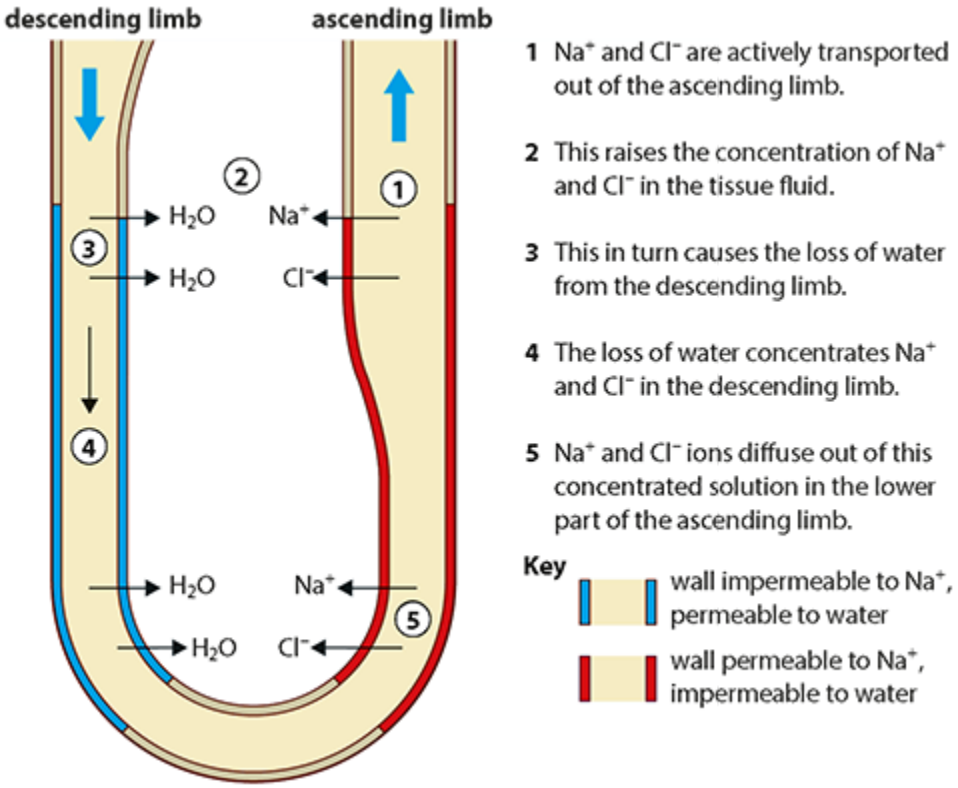


Figure 8.5.6: The counter-current mechanism in the loop of Henle builds up a high Na^+ ion and Cl^- ion concentrations in the tissue fluid of the medulla.

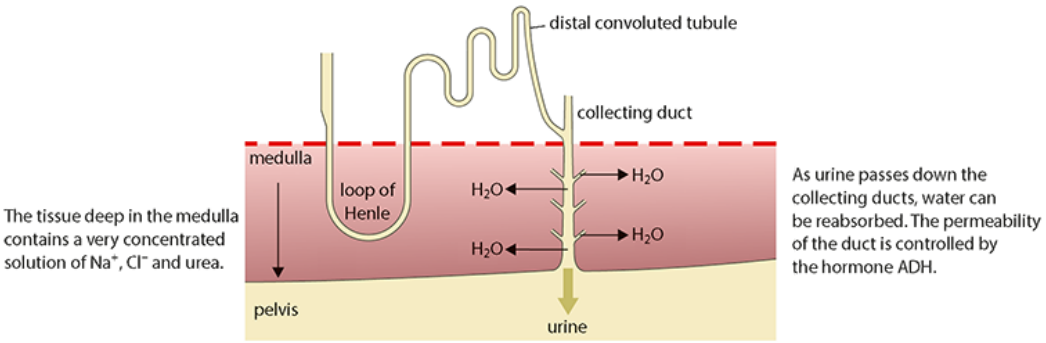


Figure 8.5.7: Water can be drawn out of the collecting duct by the high salt concentration in the surrounding tissue fluid of the medulla.

The length of the loop of Henle is different in different species. Its length is related to an animal's need to conserve water. Terrestrial animals, such as camels, that live in dry environments and need to conserve water, produce small volumes of very concentrated urine. These animals have a longer loop of Henle relative to their size than species such as otters and beavers, which live in places where dehydration is not a problem. Animals that live in wet environments tend to have very short loops of Henle and excrete dilute urine.

The distal convoluted tubule and the collecting duct

Despite the loss of water from the loop of Henle, the filtrate that enters the next section of the tubule still has a high water content. Ions are exchanged between the filtrate and the blood in the distal convoluted tubule. Na^+, Cl^- and Ca^{2+} ions are reabsorbed into the blood while H+ and K+ ions may be actively pumped into the tubule.

The last portion of the nephron is the collecting duct where the final adjustment of water is made (Figure 8.5.7). The permeability of the duct depends on the presence or absence of **antidiuretic hormone** (ADH). More ADH means that more water is reabsorbed into the blood and the urine produced is more concentrated. Less ADH makes the duct impermeable so that more dilute urine is produced.

The last portion of each nephron is the collecting duct where the final adjustments to water content in urine are made (Figure 8.5.4). These occur in response to changes in concentration of

the blood which are detected by osmoreceptors (receptors that detect changes in osmotic pressure) in the hypothalamus in the brain. If blood volume is low and there is insufficient water present, the hypothalamus stimulates the pituitary gland just beneath it to release the hormone antidiuretic hormone (ADH), (this hormone is also called vasopressin). This hormone controls the permeability of the collecting duct and regulates the amount of water that is excreted in urine. If ADH is released, membrane channels, called aquaporins, found in the walls of the collecting duct, will open. The duct becomes more permeable and water leaves the tubule and is taken back into the blood. If the water content of the blood is high, ADH is not produced so the aquaporin channels remain closed and the duct is impermeable to water. The water remains inside the nephron, so that more dilute urine is produced and excess water leaves the body.

SCIENCE IN CONTEXT

Caffeine, alcohol and cold conditions suppress ADH production and can lead to dehydration if too much water is lost in urine. This explains why a person is more likely to need to urinate in cold weather.

Stress and nicotine increase ADH production, producing the opposite effect.

Although the kidney can conserve water already present in the body, only intake of water by drinking or in foods can replace water that has already been lost.

THEORY OF KNOWLEDGE

Ethical issues and kidney donation

Today many organs can be transplanted between well-matched human donors and recipients. Kidneys, corneas, bone marrow and skin are all transplanted regularly for certain medical conditions. Donors usually carry a card or express the wish that they will donate their organs should they die, for example, in an accident. Successful kidney transplants and the drugs needed to prevent rejection of a donated kidney cost less than keeping a person alive using renal dialysis. The quality of life of the recipient is also better. Nevertheless, there is still a shortage of people willing to donate a kidney after their death.

It is possible to live a full and active life with only one functioning kidney, so a person who is in good health can donate a kidney to help someone who desperately needs a kidney transplant. Kidneys are the most commonly donated organ by people who are still alive. Donating a single kidney to a close relative, partner or good friend, is called directed altruistic donation. But some people decide to donate to someone with whom they have no previous connection.

To consider:

1 Some governments adopt policies that make organ donation compulsory rather than voluntary unless an individual has strong moral objection. Is this right?

2 What are the ethical issues associated with altruistic donation? Should a person be allowed to choose who they donate to?

TEST YOUR UNDERSTANDING

35 Define the term ‘osmoregulation’.

36 Outline the role of the loop of Henle in regulating water content of urine.

37 Outline the role of ADH in controlling the water content of urine.

8.5.3 Further examples of homeostasis

KEY POINTS

baroreceptor receptors that monitor blood pressure

chemoreceptor a sensory neurone that responds to pH or the concentration of a chemical, such as carbon dioxide, in the body.

Carbon dioxide levels in the blood

In Section 8.3 you can study the effect of excess carbon dioxide levels in the blood and how they are controlled by an increase in gas exchange at the alveoli. This is triggered by chemoreceptors in the medulla of the brain stem.

EXAM TIP

Recall that regulation of carbon dioxide and oxygen levels in plants is controlled by the stomata in the leaves.

Blood sugar levels

In Section 7.3 you can study the control of blood sugar levels and how a failure of the homeostatic processes can lead to either Type I or Type II diabetes.

Blood pressure and feedback control of heart rate

Another important factor that the body must control is blood pressure. Several mechanisms regulate the flow and pressure of blood throughout the body. Changes in blood pressure are made

so that the correct levels of nutrients, including glucose and oxygen, reach all parts of the body at the right time and that wastes such as carbon dioxide are removed. For example, when we exercise muscles require additional oxygen and glucose, and carbon dioxide must be carried away, but the supply of blood to the intestines can be reduced for a short while. Changes to blood pressure can be made by altering our heart rate or the amount of blood that the heart pumps out with each beat. Blood pressure is also affected by changing the diameter of blood vessels that alters the resistance to blood flow. A narrow vessel will resist the flow of blood and so the pressure inside it will increase. A combination of increased blood flow or volume and narrowed vessels raises blood pressure.

There are several ways in which blood pressure is monitored and regulated, as follows.

1 Chemoreceptors in the cardiovascular centre in the medulla (brain stem) (Figure 8.5.8) send nerve impulses to the sinoatrial node (SAN) in the heart if heart rate needs to increase. These impulses stimulate the SAN to raise or lower the heart rate and output of the heart to supply the body's needs.

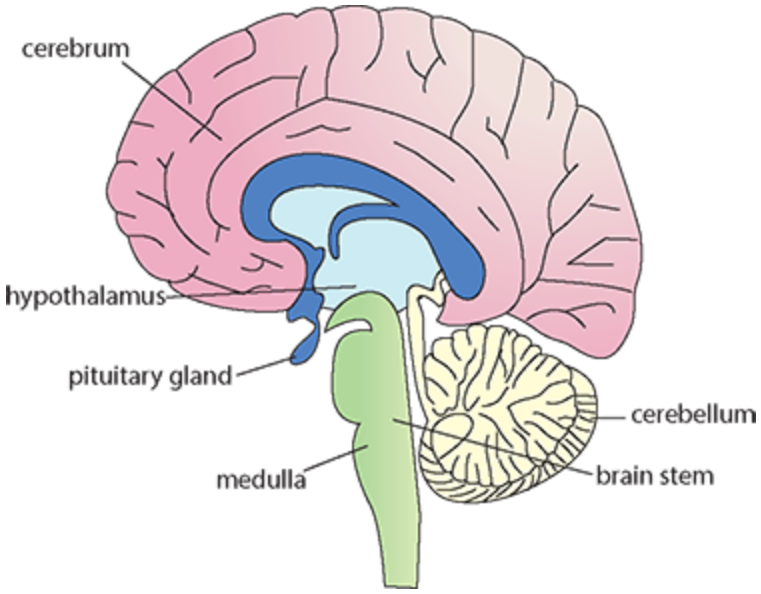


Figure 8.5.8: The medulla is located in the brain stem and controls automatic functions such as breathing, heart rate and reflexes such as swallowing.

2 Baroreceptors in the carotid arteries and aortic arch monitor blood pressure and send messages to the medulla if blood pressure is too high. Low pressure receptors are present in the atria and ventricles as well the blood vessels of the lungs.

3 Blood pressure and the distribution of blood can also be regulated by the hormones epinephrine (adrenaline) and norepinephrine secreted by cells of the adrenal medulla in the adrenal glands (Figure 8.5.1). Both hormones are part of the 'fight-or-flight response', which prepares the body for danger or to respond in a frightening or dangerous situation. The two hormones raise blood pressure by increasing heart rate and by causing vasoconstriction of arteries, arterioles and veins. Adrenaline also causes air passages to dilate to provide the muscles with the oxygen they need to either fight danger or flee. It triggers the blood vessels to contract

to re-direct blood toward major muscle groups, including the heart and lungs. (Section 8.2).

4 Chemoreceptors in the aortic and carotid arteries, which respond to levels of CO_2, send message to the medulla oblongata to increase heart and breathing rates (Section 8.3). This increases blood pressure and ensures that the respiratory and circulatory systems work in unison (Figure 8.5.9).

5 The kidneys also regulate blood pressure under the influence of hormones that control blood volume. When blood pressure is too high, a reduction in blood volume can reduce it to a normal level.

Water reabsorption under the control of ADH

The walls of the second convoluted tubule and the collecting duct are influenced by antidiuretic hormone (ADH). This hormone is produced in the hypothalamus but secreted by the pituitary gland. ADH is released when there is too little water in the blood so that its concentration rises and blood pressure falls. ADH increases the permeability of the kidney tubule to water so that more water is returned to the blood from the filtrate in the tubule and more concentrated urine is produced. More water remaining in the blood will, in turn, raise blood pressure to the normal level.

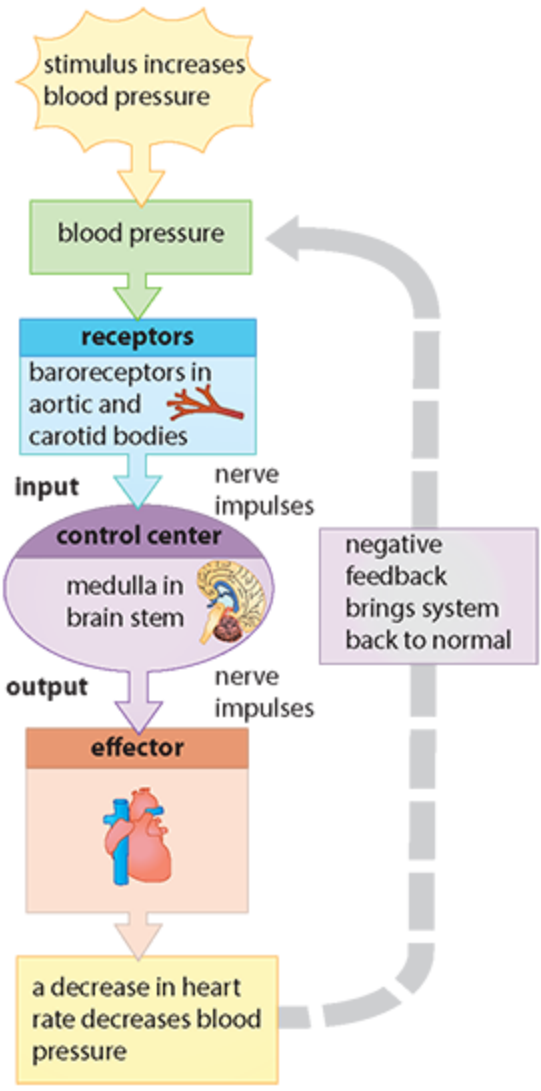


Figure 8.5.9: Blood pressure is also controlled by a feedback system.

SCIENCE IN CONTEXT

Untreated high blood pressure in older people can cause many health problems. Blood vessels can be damaged becoming narrow or leaking and high blood pressure can also cause blood clots to form in the arteries to the brain, blocking blood flow and potentially causing a stroke. If blood clots form in the arteries that supply the heart, they can cause a heart attack.

People who suffer from high blood pressure are encouraged not to smoke. This is because nicotine in tobacco raises blood pressure by increasing vasoconstriction and by stimulating the adrenal medulla to increase secretion of epinephrine and norepinephrine.

Other substances also affect blood pressure. One of these is alcohol, which lowers blood pressure by causing vasodilation and by inhibiting the release of ADH so that more water is lost in urine and blood volume decreases. This also explains why drinking alcohol leads to dehydration and causes a headache known as a hangover.

If blood pressure rises due to excess water in the blood and, as a result, blood concentration falls, the release of ADH is inhibited. The release of ADH is controlled by a negative feedback loop (Figure 8.5.10).

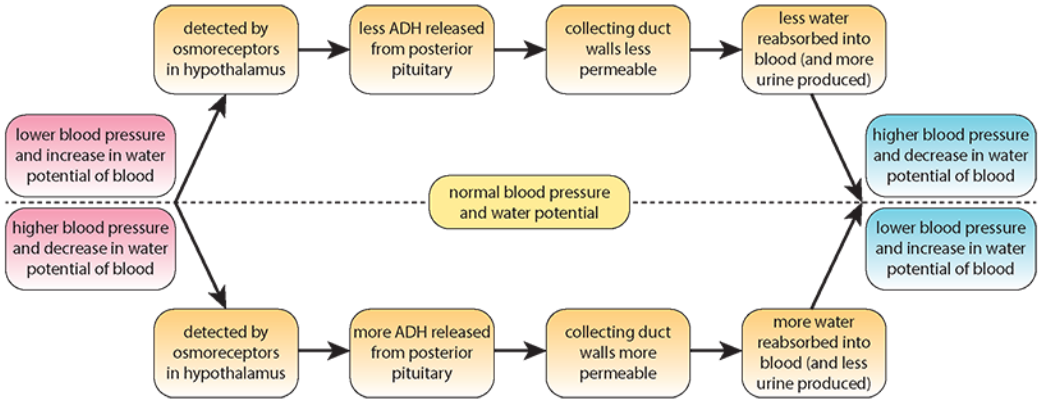


Figure 8.5.10: Control of water content of the blood by negative feedback.

Adrenaline – the fight or flee hormone

Adrenaline is an important hormone which causes changes in many parts of the body to prepare us for action or get us ready to

react if we are threatened or in danger. A sudden release of adrenaline can feel like anxiousness, nervousness, or pure excitement as your body gets ready for an event or activity. There are certain activities like skydiving that will cause a rush of adrenaline but even watching a scary movie or waiting to take an important exam can produce a similar effect. The body's responses ensure we can see, breathe and move quickly and that we are not distracted by functions such as digestion that can be paused for a while. The effects of adrenaline include:

- increasing the rate of heartbeat
- increasing blood pressure and respiration rate
- dilating (widening) bronchi
- increasing metabolic rate
- increasing the release of glucose in the liver and muscles
- dilating the pupils of the eye
- inhibiting peristalsis and digestion
- inhibiting bladder contraction.

TEST YOUR UNDERSTANDING

38 What effect does constriction of arterioles have on blood pressure?

REFLECTION

Could I explain the key points of this topic to someone else? What is the most difficult aspect?

Links

- How do hormones play a role in homeostasis? (Chapter 7)
- Can homeostasis influence evolution? (Chapter 11)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
state that multicellular organisms are composed of cells that form tissues and organs	8.1.1			
describe how emergent properties occur as a result of cell interactions in multicellular organisms	8.1.1			
outline the importance of stem cells in the formation of a multicellular organism	8.1.2			
identify the three	8.1.2			

types of stem cell and their properties				
state that stem cell niches are found in human bone marrow and hair follicles	8.1.2			
explain the importance of a circulatory system to a multicellular organism	8.2.1			
identify the features of arteries, veins and capillaries	8.2.1			
understand how to take a pulse measurement, explain how and why pulse rate and blood pressure change with activity	8.2.1			
recognise the causes and consequences of blockages in arteries	8.2.1			

distinguish between a single and double circulation	8.2.1			
draw a diagram of the heart showing valves that separate the four chambers	8.2.1			
recognise the stages in the cardiac cycle	8.2.1			
outline the structure and function of the lymphatic system	8.2.2			
describe how blood pressure is maintained and how it is important in capillary exchange	8.2.2			
describe how water is transported from roots to leaves during transpiration	8.2.3			
summarise the	8.2.3			

features of xylem needed for transport of water				
draw distribution of tissues in transverse sections of roots and stems	8.2.3			
explain how root pressure and active transport cause water movement in roots	8.2.3			
summarise the adaptations of phloem sieve tubes for translocation	8.2.3			
explain why transpiration is a consequence of gas exchange in a leaf	8.2.5			
explain the importance of gas exchange surfaces and their properties	8.3.1			
describe how concentration	8.3.1			

gradients are maintained for gas exchange				
describe the adaptation of mammalian lungs for gas exchange	8.3.1			
describe how ventilation rate is adapted to meet the need for gas exchange	8.3.2			
describe the adaptations of pneumocytes in alveoli	8.3.2			
describe how oxygen is carried by hemoglobin	8.3.3			
draw a dissociation curve and explain its shape	8.3.3			
describe the dissociation curve of the Bohr shift and for fetal hemoglobin	8.3.3			
draw a labelled diagram of a leaf	8.3.4			

and describe how it is adapted for gas exchange and capturing light				
summarise the differences between asexual and sexual reproduction	8.4.1, 8.4.2			
outline the role of meiosis and fusion of gametes in sexual reproduction	8.4.2			
outline the stages of the menstrual cycle and the hormones involved	8.4.2			
annotate diagrams of male and female reproductive systems to explain their functions	8.4.2			
summarise the stages of fertilisation	8.4.2			
describe the stages in an IVF	8.4.3			

treatment				
outline the control of changes at puberty by GnRH	8.4.4			
describe spermatogenesis and oogenesis and adaptations of human sperm and egg cells	8.4.4			
summarise the mechanisms that prevent polyspermy	8.4.4			
outline the formation of the blastocyst and the role of hCG in pregnancy and in pregnancy testing	8.4.4			
describe the development of a blastocyst and implantation and the role of the placenta	8.4.4			
explain the hormonal control of pregnancy and childbirth	8.4.5			

understand the effects of HRT	8.4.5			
describe positive and negative feedback in the human menstrual cycle and mammalian birth	8.4.5			
outline the sexual reproductive strategies of plants	8.4.6			
recognise the features of an insect-pollinated flower	8.4.6			
describe how angiosperms are pollinated and how they prevent self-pollination	8.4.6			
recall that dispersal and germination of seeds complete a plant's life cycle	8.4.6			
define homeostasis and explain why it is important	8.5.1			

define thermoregulation and explain physiological and behaviour strategies to maintain body temperature	8.5.1			
recall that blood glucose regulation, blood pH, heart rate and ventilation rate regulation are examples of homeostasic processes	8.5.1			
describe a negative feedback system and explain how it differs from a positive feedback system	8.5.1			
recall that individuals have a range of tolerance for internal conditions	8.5.1			
describe the role of the kidney in	8.5.1			

osmoregulation and excretion				
explain the functions of the different sections of the kidney tubule in osmoregulation and excretion	8.5.1			
describe osmoregulation by the collecing ducts and ADH	8.5.1			
outline feedback control of heart rate following sensory input from baroreceptors and chemoreceptors	8.5.2			
recall the roles of the hypothalamus and pituitary gland in homeostasis	8.5.2			
describe how sleep patterns are modulated by melatonin	8.5.2			
list the effects of	8.5.2			

adrenaline on the body.				

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Chapter 9

Coordination, muscles and motility

C3.1, B3.3

INTRODUCTION

The brain and spinal cord act to integrate the organs of the body that enable us to move. Some of these movements are involuntary and we cannot control by thinking about them, while other movements we can think about and control voluntarily. Adaptations for movement are a feature of almost all living things which move to escape, search for food or search for mates.

9.1 Coordination and muscle contraction

LEARNING OBJECTIVES

In this section you will:

- understand that inputs to the spinal cord and cerebral hemispheres arrive through sensory neurones
- recognise that outputs from the cerebral hemispheres via motor neurones enable muscles to contract.
- understand that nerves are bundles of both sensory and motor neurones
- discover that reflex arcs lead to involuntary responses with skeletal muscle as the effector
- learn that the cerebellum coordinates skeletal muscle contraction and balance
- understand that peristalsis in the digestive system is controlled by both voluntary and involuntary muscle contraction

9.1.1 Stimulus and response in the nervous system

KEY POINTS

central nervous system (CNS) the brain and spinal cord

peripheral nervous system the sensory and motor nerves outside the CNS

reflex rapid unconscious response to a stimulus

response reaction or movement as a result of a stimulus

stimulus change in the environment that is detected by a receptor and causes a response

Receptors are parts of the nervous system that detect a stimulus and initiate a nerve impulse. There are many different receptors, for example in the skin there are pain, temperature and pressure receptors and the retina of the eye contains light receptors. Nerve impulses are carried by neurones to effectors which may be muscles or glands. The effectors carry out the response.

The pathway between receptor and effector involves the central nervous system (Section 7.2) which comprises the brain and spinal cord (Figure 7.2.1). The type of neurone that carries the message towards the CNS is a sensory neurone and the one that carries the impulse from the CNS to the effector is called the motor neurone. These sensory and motor neurones make up the peripheral nerves and within the CNS interconnecting relay neurones connect the sensory and motor neurones via synapses. (Figure 9.1.2(a))

Inputs from sensory neurones are carried from receptor cells to the spinal cord and cerebral hemispheres of the brain. The spinal cord co-ordinates involuntary responses while the cerebellum of the brain (Figure 7.2.3) is primarily responsible for muscle control, including balance and movement. The cerebellum plays a major role in adapting and fine-tuning responses so that we can make accurate movements such as catching a ball by practising or through trial and error.

The cerebellum controls voluntary movements such as walking, posture, balance, coordination, eye movements and speech. All these movements require sensory inputs from peripheral nerves and motor outputs from the brain to the muscles that are involved. Outputs from the spinal cord and brain stimulate muscles to contract and cause movement.

Reflex actions

Autonomic and involuntary responses are together known as reflex actions. A reflex is a specific reaction that is always produced in response to a particular stimulus and which does not require prior learning. Sometimes a rapid response to a stimulus is vital for an animal's survival and reflex actions take place quickly and automatically. Human reflexes include the pupil reflex which reduces the diameter of the pupil in very bright light to prevent damage to the retina and the coughing reflex, which occurs when a piece of food enters the trachea.

The pain withdrawal reflex takes place if you touch something that causes pain. For example, if you touch a very hot object or are stung by a bee, you pull your hand away quickly, without thinking about it at all. The pain withdrawal reflex is an example of a reflex action which uses a rapid and simple neural pathway called a reflex arc (Figure 9.1.1).

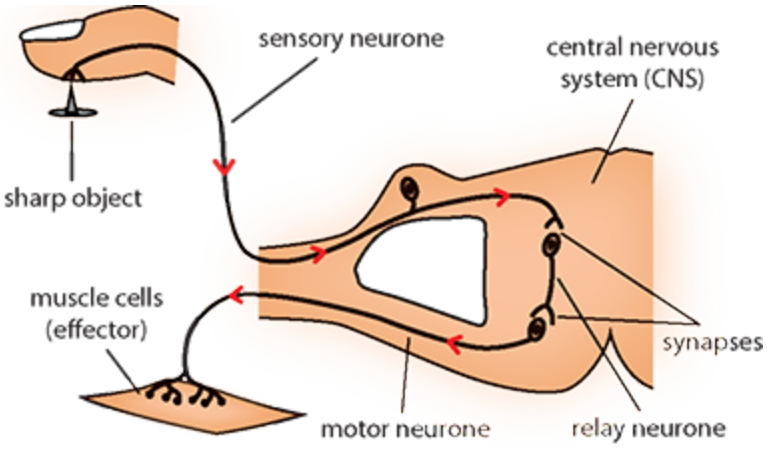


Figure 9.1.1: A spinal reflex arc for a pain withdrawal reflex

The reflex arc involves a receptor in the fingertip, a sensory neurone, a relay neurone in the spinal cord and a motor neurone that stimulates the effector, in this case a muscle in the arm, to contract and draw your hand away.

Relay neurones also connect to neurones going up and down the spinal cord. These ascending and descending neurones carry information to and from the brain. So if you do touch something that causes pain, not only do you remove your hand immediately but information is sent to the cerebral hemispheres of the brain so that you can remember what happened and not do it again.

The pathway of a reflex arc is genetically determined so that appropriate responses to different stimuli occur. There are several different reflexes that are controlled by the spinal cord such as the pain withdrawal reflex and the knee jerk reflex. The brain also controls some reflex actions such as the blinking reflex which happens if something touches the surface of the eye.

EXAM TIP

Review Chapter 7.2 and remind yourself about the structure of the nervous system and the importance of myelinated neurons in transmission of nerve impulses.

Structure of nerves and muscles

Nerves

Nerves that make up the peripheral nervous system throughout the body are made of bundles of different types of neurone (Figure 9.1.2(b)). Some of the fibres are dendrites of sensory neurones that carry impulses towards the CNS, others are axons which carry impulses away from the CNS to effectors. Some nerve fibres are enclosed in a myelin covering which speeds up the transmission of impulses, others are unmyelinated. A layer of connective tissue called the perineurium surrounds the groups of fibres which are known as fascicles and an outer layer called the epineurium encloses several groups of these together.

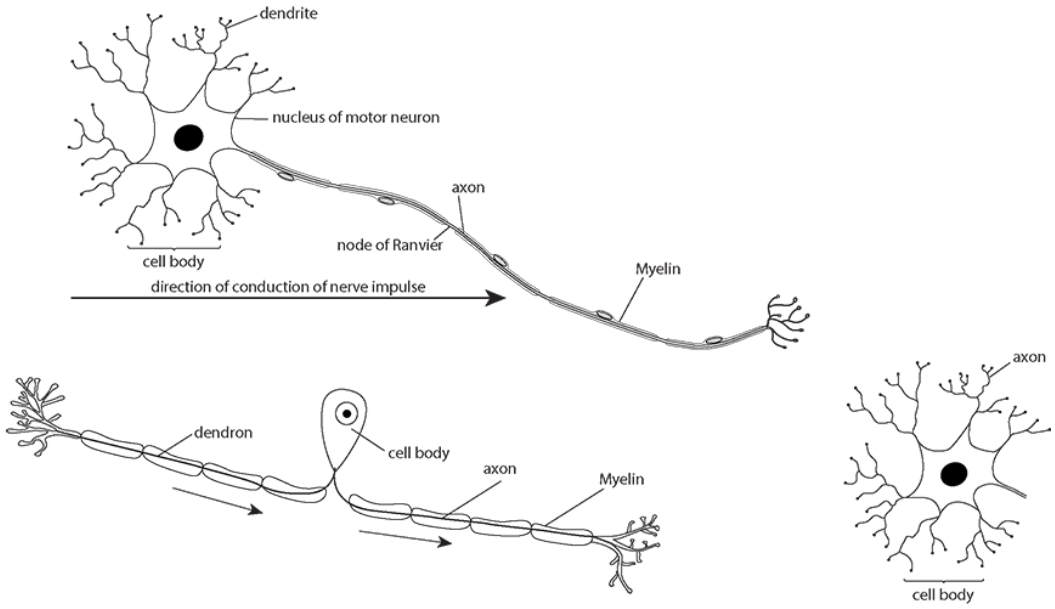


Figure 9.1.2 a: Sensory neurones carry messages to the CNS but their cell bodies are contained in ganglia outside the spinal cord. Motor neurones carry messages away from the CNS, their cell bodies make up the grey matter in the spinal cord. Intermediate neurones form connections between sensory and motor neurones in the CNS.

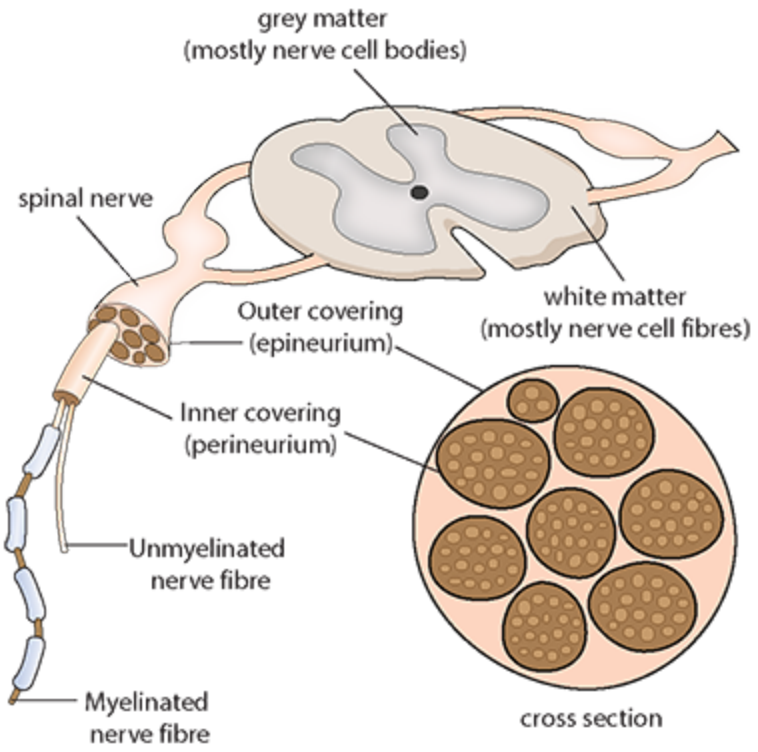


Figure 9.1.2 b: Peripheral nerves consist of bundles of sensory and motor neurones enclosed in a protective sheath

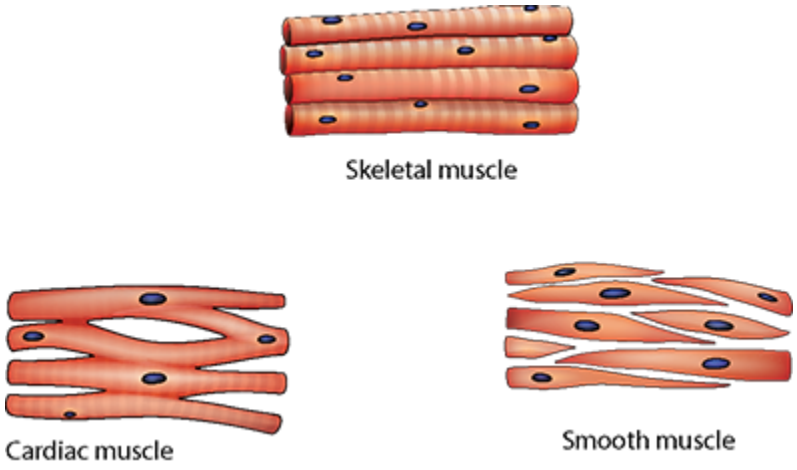


Figure 9.1.3: Three different types of muscle which cause movement in the body

Muscle tissue

There are three different types of muscle tissue in the human body (Figure 9.1.3)

Skeletal or striated muscle is the muscle that causes movement of our joints. It is sometimes called voluntary muscle. Under the light microscope it has a striped appearance, and it is made of cells with many nuclei known as muscle fibres (see Figure 9.2.9). Surrounding each fibre is a plasma membrane called the sarcolemma.

Heart muscle has a unique composition that adapts it for the conduction of waves of excitation from fibre to fibre. It is made up of short, striped muscles fibres which branch and are also joined together at their ends by linking structures known as intercalated discs. (see Figure 9.2.9). This arrangement of linkages between cells allows action potentials to spread rapidly and enables the heart muscle fibres to act together and produce a more powerful beat as they contract simultaneously.

Smooth muscle fibres are found in organs such as the liver, bladder, and intestines. Smooth muscle cells are spindle-shaped and work involuntarily. We have no control over the contraction of our involuntary muscles.

Control of peristalsis in the digestive system

KEY POINTS

Enteric nervous system part of the autonomic nervous system that controls smooth muscle in the digestive system

peristalsis muscle contraction that moves food along the digestive system

ingestion taking in food, eating

egestion removal of waste from the body during defecation

The central nervous system plays a key role in the movement of food along the digestive system. Both voluntary and involuntary control are involved. We take in and chew solid food in our mouths and swallow it so that it passes down the esophagus (food pipe) to the stomach. The acts of chewing and swallowing are voluntary. We can choose to chew and when to swallow and these actions are controlled and coordinated by the CNS.

Once food has entered the digestive system it is moved along the intestine by a sequence of muscle contractions known as **peristalsis**. Peristalsis is under involuntary control and is regulated by the **enteric nervous system** (Figure 9.1.4). Peristalsis involves two layers of involuntary muscle which make up the intestine wall. Longitudinal muscles run along the length of the intestine (Figure 9.1.5) while circular muscles encircle the intestine. Contractions of the bands of circular muscle squeeze

the area of the intestine behind a portion of food while longitudinal muscle relaxes and extends to accommodate it. The two sets of muscles then relax and contract respectively so that food is gradually pushed along the intestine in waves. The action of the enteric nervous system ensures that the movement of food is coordinated. As food moves along it is mixed with digestive enzymes, useful substances are absorbed, and eventually only waste and undigested material remain in the large intestine. This waste material forms feces which are egested through the anus. Egestion of feces is under voluntary control.

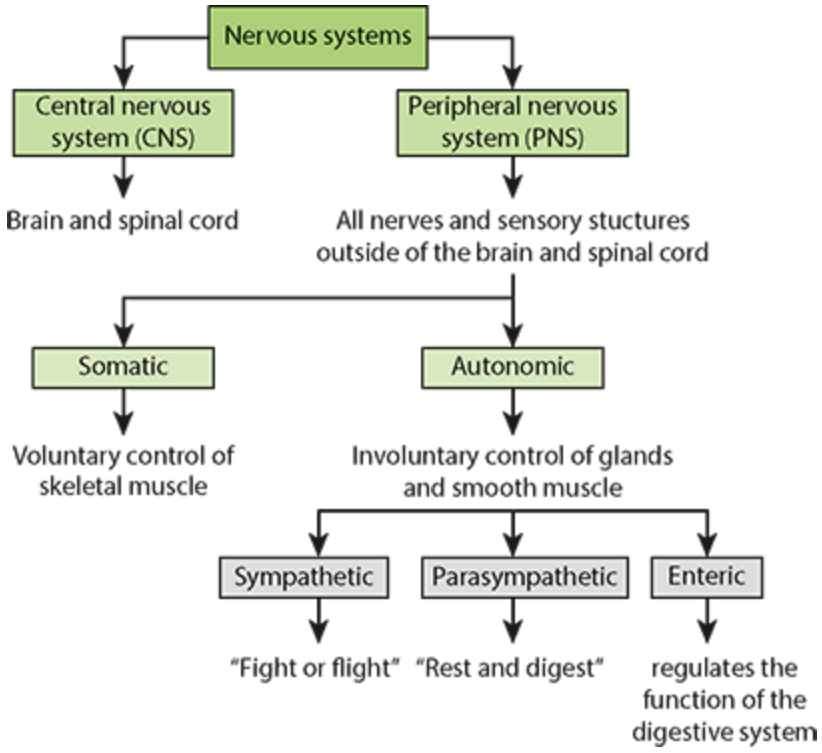


Figure 9.1.4: This diagram shows the relationship between the different parts of the nervous system

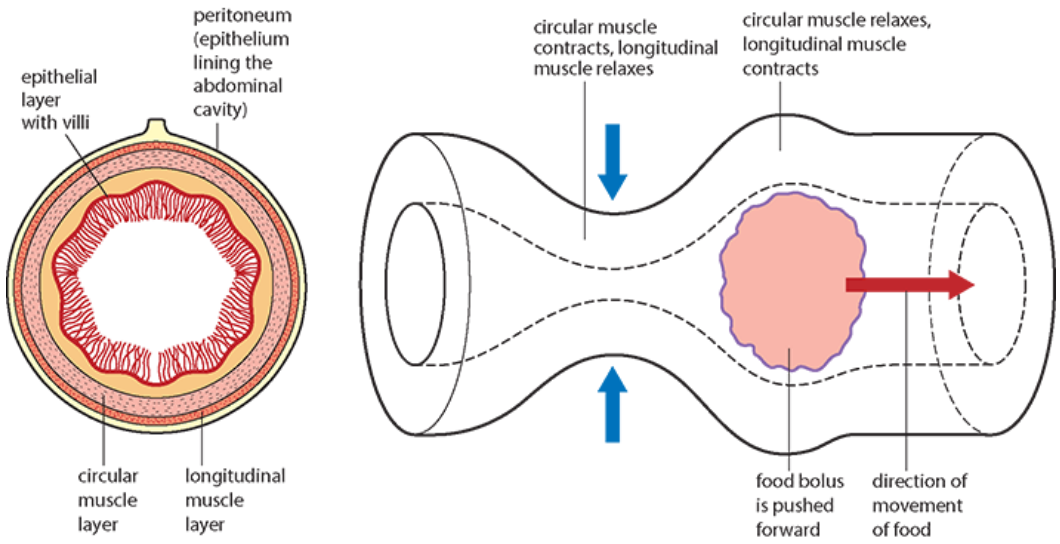


Figure 9.1.5: The actions of longitudinal and circular muscles in the intestine move food along in waves of peristalsis coordinated by the enteric (non-voluntary) nervous system

TEST YOUR UNDERSTANDING

1 List the components of a reflex arc

2 Distinguish between the central nervous system and peripheral nervous system

3 Where would you find each of these muscle types **i** striated muscle **ii** cardiac muscle **iii** smooth muscle

9.2 Muscles and motility

LEARNING OBJECTIVES

In this section you will:

- learn that adaptations for movement are found in all living organisms
- understand how actin and myosin are involved in muscle contraction
- learn that the protein titin and antagonistic muscles are needed for muscle relaxation
- recognise the structure and function of motor units in skeletal muscle
- distinguish between exoskeletons and endoskeletons as anchorage points for muscles
- understand the structure of a synovial joint and the range of movement of different joints.
- recognise an example of antagonistic muscles in breathing
- distinguish between movement and locomotion and understand the reason for locomotion
- recognise the adaptations of marine mammals for swimming.

9.2.1 Types of movement

Movement is the ability of an organism to change its position or place. Most organisms can carry out simple movements and this is a characteristic of life but these movements may not cause a change of location.

Plants turn their leaves to face the sun and adjust their positions so that leaves do not overlap one another, and many flowers open and close their petals during the day and night. These movements are known as heliotropisms, a form of tropism. Movements may occur daily or with the seasons in response to the position of the sun.

KEY POINTS

movement the ability to change position, not always resulting in a change of location.

tropism a growth response of plants in which the direction of growth is determined by the direction of a stimulus.

Plants bend towards light by growth (phototropism) but we can observe many other faster movements in the plant kingdom. Some movements are related to defence. The sensitive plant, *Mimosa pudica*, has leaves that droop almost immediately when they are touched or hit by a drop of rain (Figure 9.2.1). Others are food-related movements. For example, the Venus flytrap has hinged leaves that snap shut when two of its sensitive hairs are touched at the same time by a fly walking near them, while some underwater pitcher plants have a flap on the pitcher that suddenly opens when a sensitive hair is touched so that an insect can be swept in by the current of water.

Figure 9.2.1: *Mimosa pudica*, the sensitive plant, will fold its leaves instantly if they are touched

Sedentary (sessile) animals

In the animal kingdom, some organisms stay in the same place throughout their lives. They are known as sedentary or sessile. Some of these sedentary organisms such as sea anemones (Figure 9.2.2) do move parts of their bodies for example in response to threats, contracting them inward and downward or to find food when they extend their tentacles and reach out into the water.

Figure 9.2.2: Sea anemones extend their tentacles to feed

Amphibians, such as frogs and toads, are not sedentary and can move from place to place, but they can remain motionless, waiting for prey to approach, before extending only their long tongue to capture a worm or slug.

Walking and running

Four-legged animals walk or run and can vary the speed of their motion to suit the situation. As Figure 9.2.3 shows, not all the animal's feet are in contact with the ground all the time but by balancing their weight, the animal remains stable. Legs support the animal's body. The gallop or run is the fastest movement a horse can achieve and this can average at between 40 and 48 kilometres per hour.

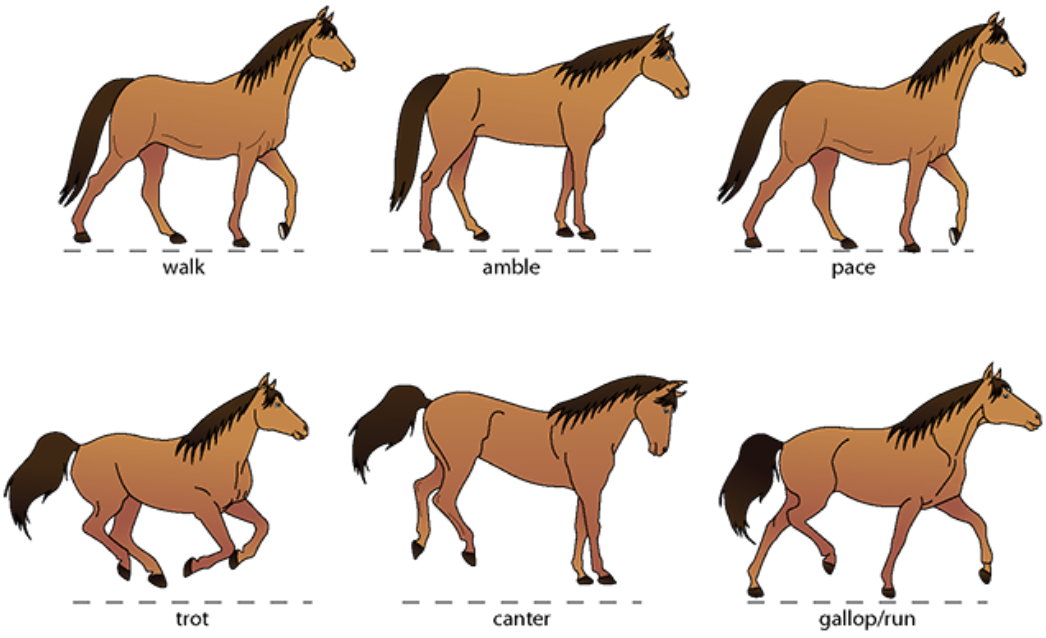


Figure 9.2.3: A horse adapts the movement of its limbs to change its speed of locomotion.

9.2.2 Skeletons and joints

Exoskeletons

Exoskeletons are hard coverings or cuticles that form a protective layer over the bodies of arthropods such as insects and crustaceans. This exoskeleton is outside the body and the muscles that move it are on the inside. Figure 9.2.4 shows the inside of an insect's leg. There are two muscles: one flexor, which bends the leg at the joint, and one extensor, which extends the leg. At a joint an exoskeleton has a flexible membrane instead of a cuticle so that the limb can bend easily.

Exoskeletons allow arthropods to move.

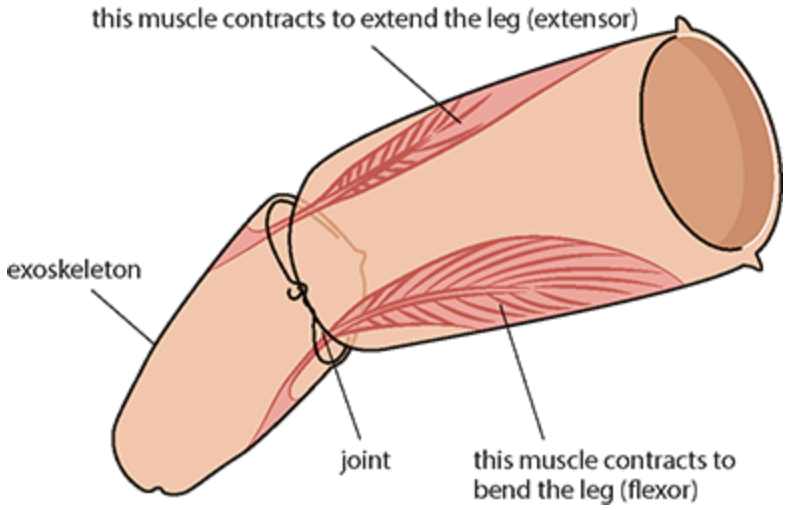


Figure 9.2.4: Part of an insect limb showing exoskeleton and muscle attachment.

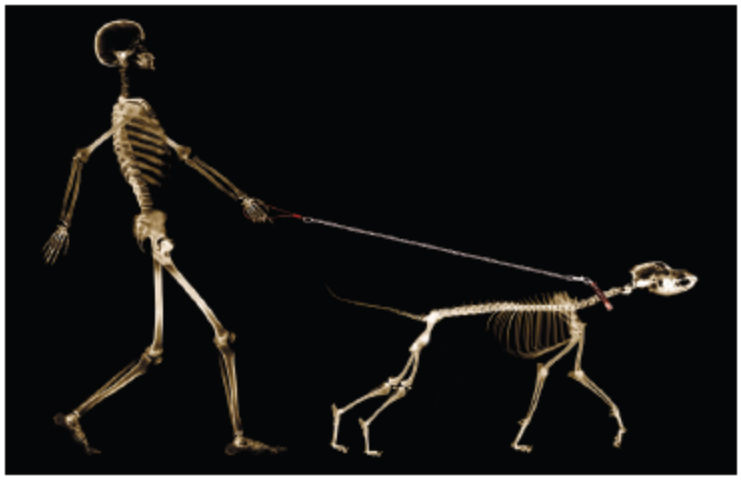


Figure 9.2.5: The skeleton supports the bodies of vertebrates.

Endoskeletons

Vertebrates have **endoskeletons** composed of firm, hard bones that do not change their shape. An internal skeleton allows an animal to move, and to move from place to place. It also provides support and, together with muscles, holds the body up and maintains its shape (Figure 9.2.5). The skeleton encloses and protects the soft parts of the body. For example, the rib cage encloses the lungs and heart and the skull protects the brain.

Bones meet at joints. Some of these joints, such as those between the vertebrae, permit very little movement. Others, such as those between bones in the arms and the legs, allow animals to bend into new positions to move their bodies. Muscles are attached to the bones of the skeleton by tendons. When a muscle contracts, it pulls a bone into a new position. Bones act as levers when they are moved by muscles. Levers are moved at a fixed point called the pivot.

Joints

A joint is a place where two or more bones meet. Joints between bones in the human body, together with the muscles that are attached to them, enable us to move and support the body. Most joints involve bones, muscles, cartilage, tendons, ligaments and nerves.

- Bones provide a framework that supports the body. They protect vital organs such as the brain and the lungs. Blood cells are formed within bones which contain bone marrow. Bones also act as a site for the storage of calcium and phosphate.

- Ligaments attach bones to one another at a joint. Some strap joints together while others form a protective capsule around a joint. They are tough and fibrous and provide strength and support so that joints are not dislocated.

- Tendons attach muscles to bones. They are formed of tough bands of connective tissue made of collagen fibres and are capable of withstanding tension as muscles contract.

- Muscles provide the force needed for movement. They can contract in length and as they do so they move the joint into new positions. Muscles only cause movement by contraction so they occur in antagonistic pairs – one muscle of the pair causes a movement in one direction while the other returns it to its original position.

- Motor neurones stimulate muscle contraction (Section 9.1) Sensory neurones transmit information from proprioceptors (position sensors) in the muscles so that movements can be coordinated and monitored.

The hip joint

The hip is a ball and socket joint. It is called this because it consists of the ball-shaped head of the femur (thigh bone) which fits into a socket in the hip. Ball and socket joints allow movement in more than one direction and allow us to make rotational movements. (Figure 9.2.6). It is an example of a synovial joint. The capsule that seals the joint is lined by a membrane that secretes lubricating synovial fluid so that the bones move smoothly against one another and friction is reduced. Smooth cartilage covers the ends of the bones at the joint and helps to reduce friction and absorb pressure as the joint moves.

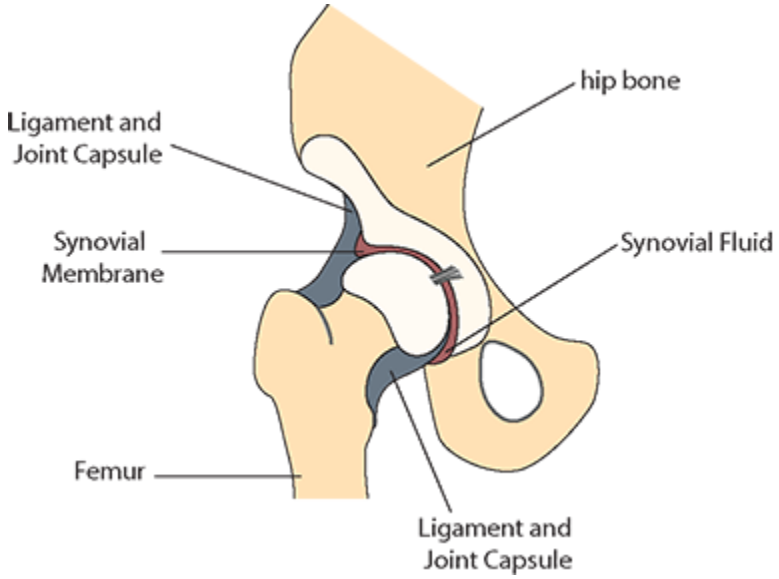


Figure 9.2.6: Diagram to show the structure of the hip joint.

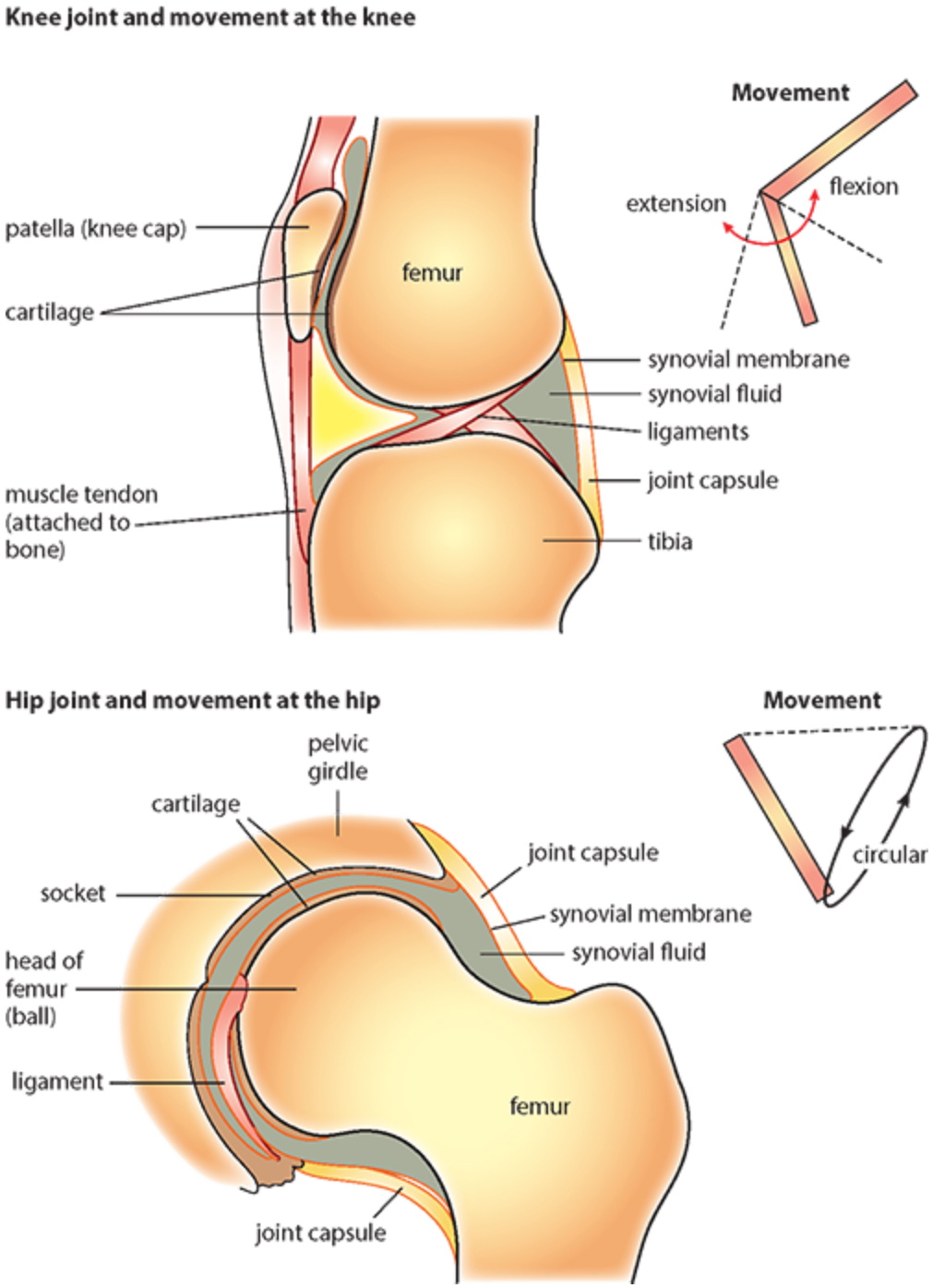


Figure 9.2.7: Longitudinal sections of the knee and hip joints, and the degree of movement they allow.

The hip joint is formed of the thigh bone (femur) of the leg and bones of the pelvis. Tendons attach sets of muscles to these

bones. As they contract the leg can be rotated or moved up and down. Pairs of muscles permit movement in opposite directions and are an example of antagonistic muscles.

The knee and elbow joints

The knee and elbow joints are examples of hinge joints and move to allow movement in only one direction. These joints are so-called because they move in a similar way to the opening and closing of a door hinge. Like the hip joint, both are synovial joints. Figure 9.2.7 shows the range of movement of a hinge joint.

9.2.3 Antagonistic muscles

Muscles can only cause a movement when they contract so they can only cause movement in one direction. To achieve movement in two directions most muscles work in antagonistic pairs which means that they work in opposition to each other. We can see an example of antagonistic muscles in the way we breathe; two sets of antagonistic muscles, the external and internal intercostal muscles of the ribcage together with the diaphragm cause us to breathe in and out. When the inspiratory external intercostal muscles contract, the expiratory internal intercostal muscles relax and we breathe in and vice versa.

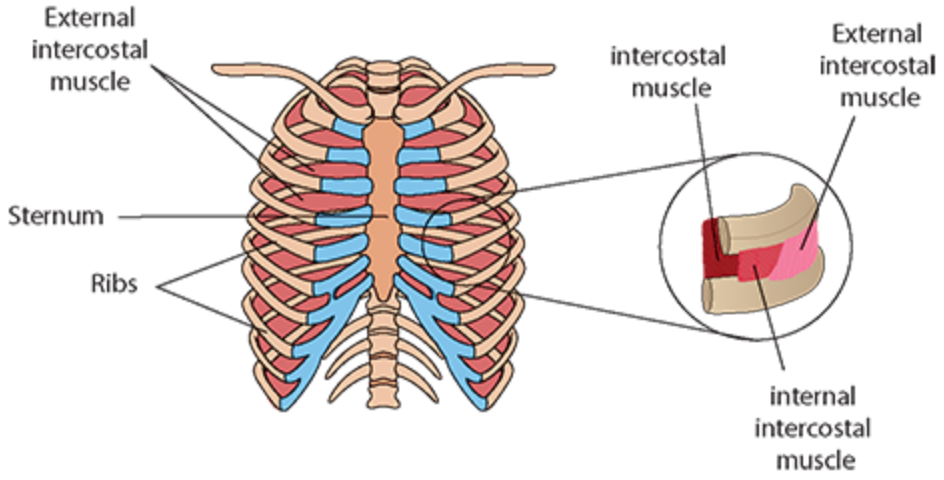


Figure 9.2.8: Arrangement of antagonistic muscles in the rib cage

The intercostal muscles of the rib cage are arranged in layers with fibres that run in opposite directions (Figure 9.2.8). The external intercostal muscles contract to expand the ribcage and move it upwards and increase the volume inside the chest cavity so that air is drawn in. As these muscles contract, they stretch the internal intercostal muscles which store potential energy in

the sarcomere protein titin (section 9.2.1) ready for exhalation (breathing out). During a strong exhalation the antagonistic internal intercostals and muscles of the abdomen contract to compress the chest cavity so that air is forced out. As this happens the external intercostal muscles will be relaxed.

Sliding filament model of muscle contraction

If skeletal muscle is examined with an electron microscope, it is possible to see that surrounding each myofibril is a system of membranes called the sarcoplasmic reticulum (which resembles smooth endoplasmic reticulum) and between the closely packed myofibrils are many mitochondria (Figure 9.2.9). Myofibrils are made up of repeating subunits called sarcomeres, which produce the striped appearance of a muscle fibre and are responsible for muscle contraction. The ends of a sarcomere are called the Z lines.

There are two types of filament that form the striped pattern of a muscle. These filaments are formed from the contractile proteins actin and myosin. The narrow filaments of actin are attached to the Z lines and extend into the sarcomere. Thicker filaments of myosin run between them. Where myosin is present, the myofibril has a dark appearance and a light band is seen where only actin is present. Myosin filaments have ‘heads’ which protrude from their molecules and are able to bind to special sites on the actin filaments.

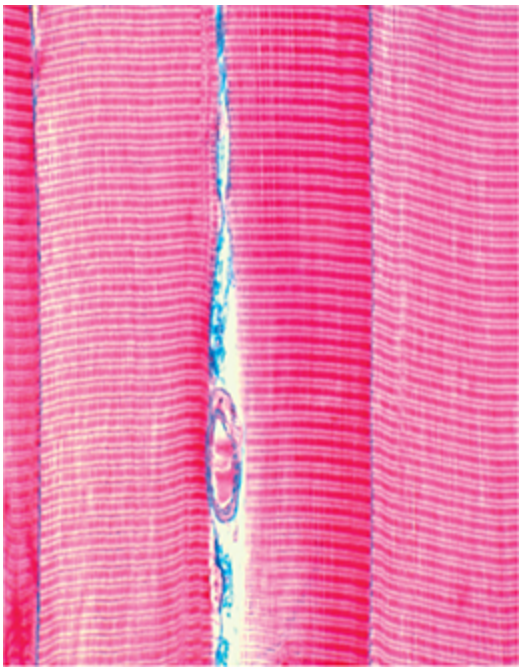


Figure 9.2.9: Light micrograph of striated muscle, stained to show the banding in muscle fibres.

Muscle tone

Contraction of a muscle causes shortening, and this in turn moves bones into a new position. If only a few fibres in a muscle contract, the muscle tightens but does not cause movement. Partial contraction produces muscle tone, which is important in maintaining posture and body shape.

Muscle contraction – the sliding filament theory

Muscle contraction is explained by the 'sliding filament theory', which describes how actin and myosin filaments slide over one another to shorten the muscle. Contraction is initiated by the arrival of a nerve impulse from a motor neurone, which stimulates the sarcolemma of the muscle fibre. This, in turn, causes the release of calcium ions (Ca^{2+}) from the sarcoplasmic reticulum and begins the process that causes actin filaments to

slide inward towards the centre of the sarcomere. The series of events is shown in Figure 9.2.11.

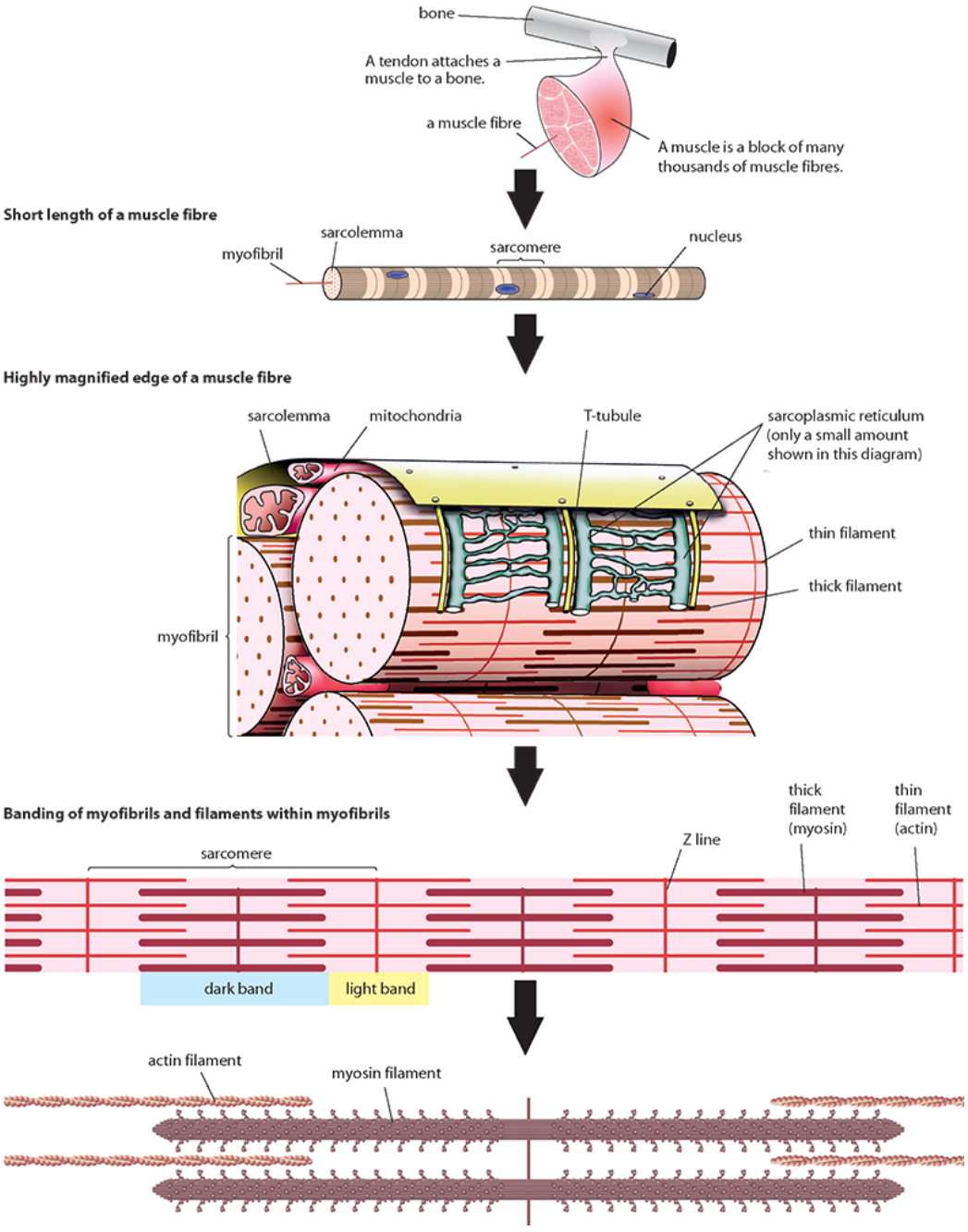


Figure 9.2.10: The structure of skeletal muscle.

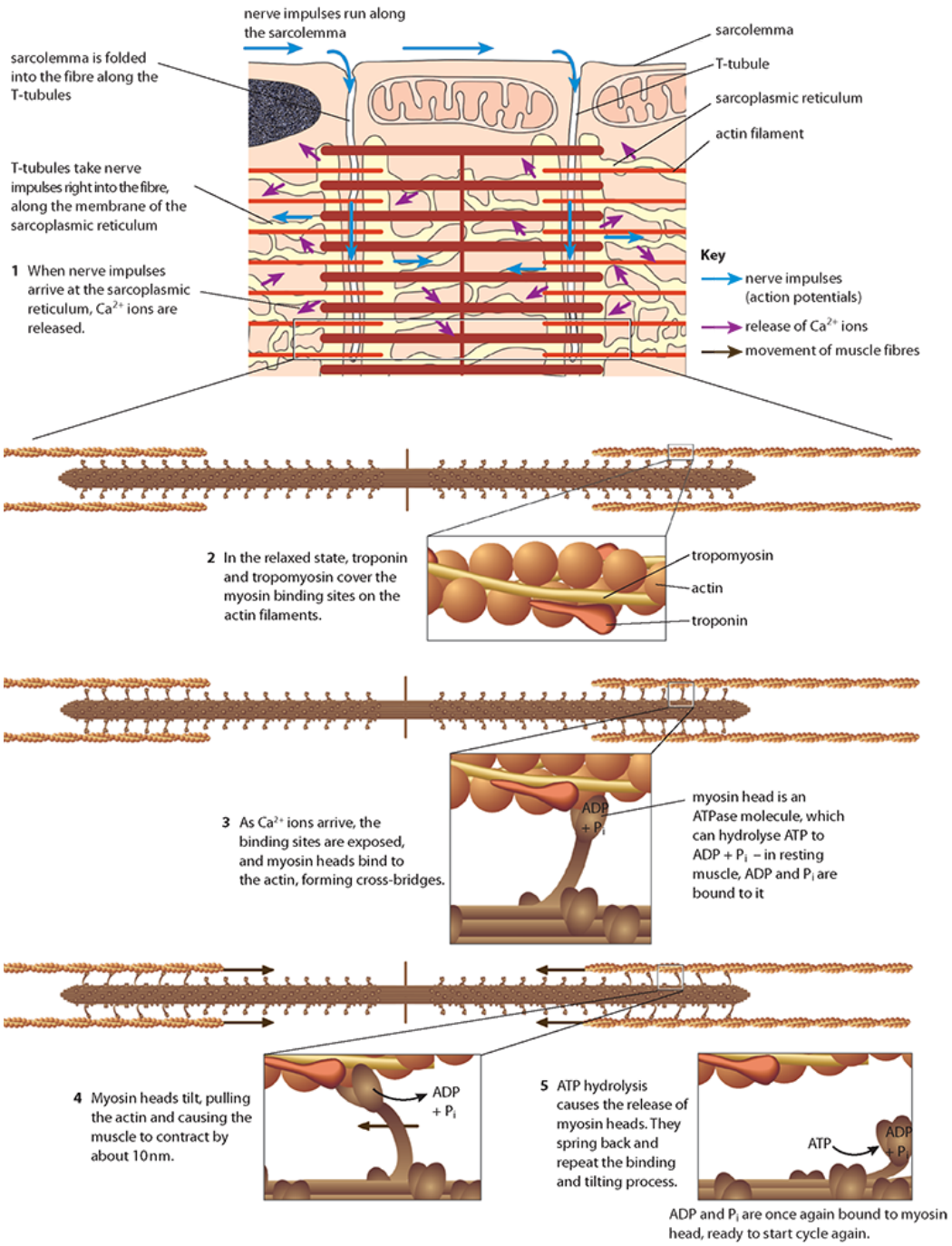


Figure 9.2.11: Muscle contraction.

1 Nerve impulses (action potentials) travel along the muscle fibre membrane, or sarcolemma, and are carried down into the fibre through infoldings called T-tubules. The impulses

then spread along the membrane of the sarcoplasmic reticulum, causing Ca^{2+} ions to be released.

2 Before contraction, binding sites for myosin heads on the actin filaments are covered by two molecules, troponin and tropomyosin. The myosin heads are prepared in an erect position as ATP binds to them.

3 Now Ca^{2+} ions bind to the actin filaments, causing the troponin and tropomyosin to change shape and expose the myosin binding sites. The myosin heads bind to the actin filaments at the exposed binding sites, forming cross-bridges.

4 This causes inorganic phosphate (Pi) to be released and, as each cross-bridge forms, ADP is also released. The myosin heads bend towards the centre of the sarcomere, pulling the actin filaments inward past the myosin filaments, by about 10 nm. This produces a ‘power stroke’.

5 New ATP molecules bind to the myosin heads, breaking the cross-bridges and detaching them from the actin filaments. ATP is used and the myosin heads return to the start position. If the muscle receives further stimulation, the process is repeated and the myosin heads attach further along the actin filaments.

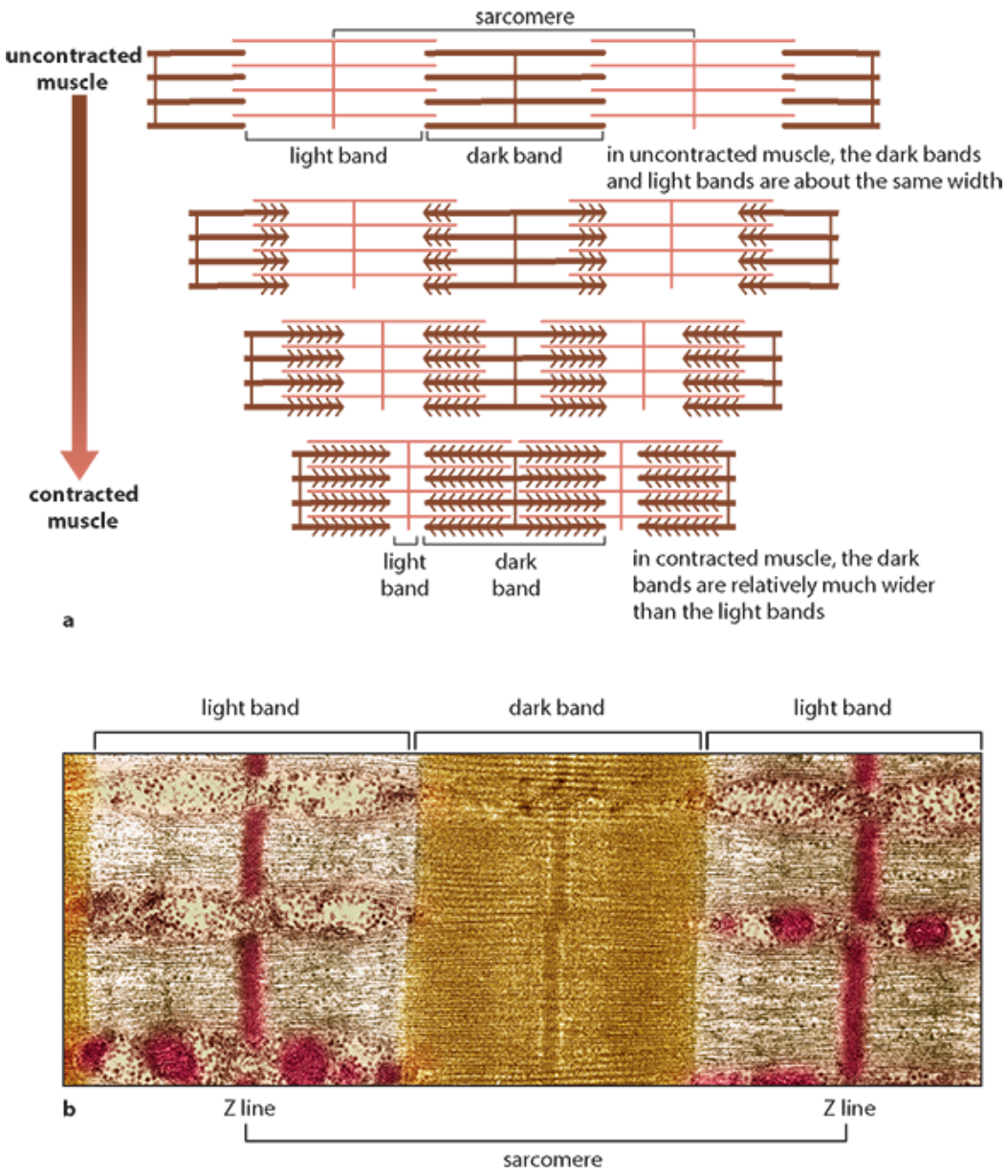


Figure 9.2.12: (a) When muscle contracts, the interleaved fibres slide inward, past each other. This makes the light bands appear narrower, but the dark bands remain the same width. (b). Coloured electron micrograph of a longitudinal section through striated muscle

Although the actin and myosin filaments do not change in length when a muscle contracts, the appearance of the banding patterns

in the sarcomere is changed. The light bands become reduced, and as the overall length of the sarcomere decreases the dark bands take up a greater proportion of the length (Figure 9.2.12).

The role of titin

Titin is the largest known protein, it is more than 1 μm in length. One of its main jobs is to provide structure, flexibility, and stability to muscle cell structures. Titin interacts with actin and myosin, to keep the components of sarcomeres in place as muscles contract and relax. Titin molecules are fixed in the Z-disc and extend to the centre of the sarcomere. Titin acts as a sort of molecular spring which maintains the precise arrangement of thick and thin filaments and keeps muscles firm.

The most important molecule for contraction in striated muscle and heart muscles is the thick filament of myosin. Titin helps to keep this thick filament at a precisely controlled length which determines the force that muscles generate and how this force varies with muscle length. Titin helps sarcomeres to recoil after they have been stretched and prevents muscles overstretching. It also stores potential energy in preparation for the action of an antagonistic muscle.

Motor units in skeletal muscle

A single motor neuron can stimulate many muscle fibres. Each muscle fibre has one neuromuscular junction which is a synaptic connection between the end of the motor nerve and the muscle. It is the site where an action potential is transmitted from nerve to the muscle. A **motor unit** consists of a neurone plus the muscle fibres it supplies. There are many motor units that together provide stimulation to a muscle to generate a contraction. Muscle fibres are stimulated to contract by the release of acetylcholine from the end of a motor neurone. Each

muscle fibre receives a single connection to the end of a neurone in a region known as the motor end plate. The arrival of an action potential triggers depolarisation that is transmitted along the sarcolemma and into the muscle fibre. Transverse or T tubules conductaction potentials that stimulate opening of voltage-gated Ca^{2+} channels. The Ca^{2+} ions diffuse out of the sarcoplasmic reticulum to the myofibrils to stimulate contraction.

KEY POINTS

motor unit a motor neurone plus the muscle fibres it supplies

neuromuscular junction a synaptic connection between motor nerve and muscle

At the end plate are thousands of T tubules which carry the stimulus through the membrane and cause contraction. But the number of muscle fibres stimulated by a single motor neurone axon varies from a few to many because the axon divides and has many end connections. The combination of a motor neurone axon and all the muscle fibres it stimulates is called a motor unit. (Figure 9.2.13). There are many motor units in a nerve that collectively provide electrical stimuli to a muscle to generate a muscle contraction.

The number of muscle fibres connected to each unit can vary within a particular muscle and from muscle to muscle; the largest muscles have motor units with connection to more muscle fibres, whereas smaller muscles have fewer muscle fibres in each motor unit. For example, thigh muscles can have a thousand fibres connected in each unit, while muscles that control the eye might have ten. Not all motor units are always activated at the same time. The more motor units that are active,

the larger the number of muscle fibres that contract, and the greater the muscle contraction. Muscles which possess more motor units have greater individual motor neurone stimuli and can control the output of force more finely.

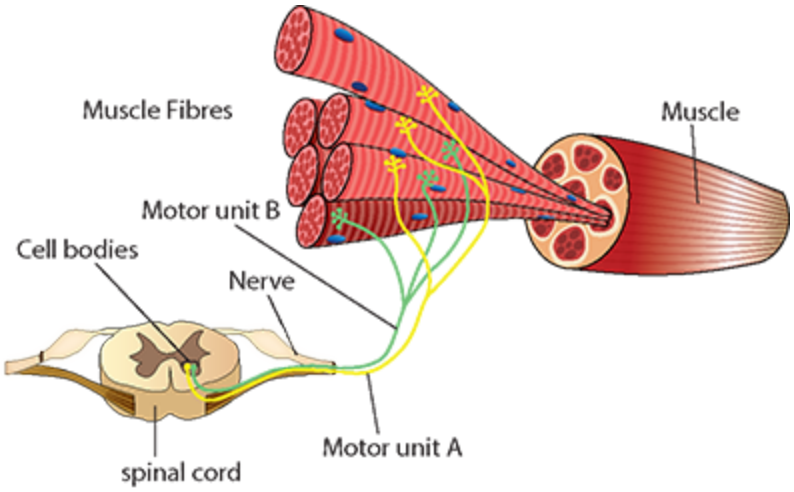


Figure 9.2.13: A motor unit consist of a motor neurone and the muscle cells it supplies

TEST YOUR UNDERSTANDING

4 Explain why muscles occur in antagonistic pairs.

5 Outline the functions of cartilage and synovial fluid in the elbow joint.

6 Compare the movement of a hinge joint and a ball-and-socket joint.

7 Explain how actin and myosin filaments produce the striped appearance of skeletal muscle.

8 Describe the role of ATP in muscle contraction.

9.2.4 Locomotion

Locomotion is defined as the ability of an organism to move from one place to another. Animals move and change their locations for a variety of reasons, including finding food, seeking out a suitable habitat or avoiding an unsuitable one, escaping predators or pursuing a mate. The structures that animals use for movement include cilia, legs, wings, arms, fins and tails. Vertebrate locomotion results from coordinated patterns of movement involving activation of hundreds of muscles under the control of the spinal cord. Neurones of the brain stem initiate locomotion via the spinal cord and motor neurones linked to muscles carry out the movement.

Locomotion by all animals requires three important factors:

1 Propulsion to move the animal in the right direction with sufficient force.

2 Stability to ensure that if a limb is lifted from the ground the animal remains sufficiently stable.

3 Support to ensure the animal is held up by its body (or other medium) as it moves.

Figure 9.2.14: The satin bowerbird builds an elaborate bower decorated with blue objects to attract females.

Advantages of locomotion

Every type of locomotion uses a lot of energy, but this is balanced by the advantages it gives to animals that can move about. Feeding and searching for food, as well as escaping from predators, are made easier for animals that can move from place to place.

1 Searching for and enticing a partner to mate require locomotion. Many fish, such as the three-spined stickleback (*Gasterosteus aculeatus*), perform mating rituals involving specific movements. Stickleback males build a nest and dance erratic, zigzag dances to encourage egg-carrying females to come into the nest where he can fertilise the eggs. Among the birds, the satin bowerbird

(*Ptilonorhynchus violaceus*) has one of the most complex and colourful mating rituals. Male birds build elaborate bowers, which resemble piles of sticks but are carefully decorated exclusively with blue objects (Figure 9.2.14). If a female bird is impressed by a bower and comes to view the male, he must then perform a dance to ensure she will mate with him.

Figure 9.2.15: Dolphins have streamlined bodies to reduce friction as they swim

2 Some species move huge distances in regular journeys called migrations. This enables them to find the most favourable conditions for feeding or breeding. Swallows migrate a distance of about 5000 km from Europe to spend the summer in South Africa. They take about 10 days for their journey, flying through the day and roosting in huge flocks in reed beds at night. Swallows feed on flying insects on their journey, catching their food as they travel. The

birds arrive at their destination in January, leaving behind the winter in Europe and arriving at their summer feeding areas when the temperature is warm, around 25 °C.

3 Foraging means searching for and collecting food resources. Most animals must forage for food and herbivores such as rabbits, horses and may spend most of their active day moving about as they search for and eat plant food. Small birds such as the blue tit will forage for insects and larger birds such as terns and gulls skim the surface of water to seek out fish. All these types of foraging behaviour rely on locomotion on land, sea or in the air.

4 Escaping danger – all species need to escape from their predators and there are a huge range of strategies that animals use. Running, jumping, dropping from trees to the ground as well as climbing into trees or flying away are strategies used by mammals all over the world. For all these strategies, locomotion is essential to move away from danger.

Adaptations for movement

Swimming

Fish and aquatic birds and marine mammals swim and are propelled by their fins or flippers. Water supports their bodies and maintains their stability so that for short periods of time these animals can dive or move forwards without moving their limbs (Figure 9.2.15). Movements of fins and flippers propel the animals downward as they dive for food.

Dolphins are aquatic mammals. Their bodies are perfectly adapted for swimming. They have a streamilined shape which makes it easy to move through water. Like all vertebrates their

limbs have a pentadactyl pattern (Chapter 11) but they are adapted to form flippers. A dolphin's tail has a flattened shape which forms a fluke which moves up and down to propel the animal forward or downward as they dive.

As mammals, dolphins breathe air and must return to the water surface to inhale. They breathe through nostrils, called a blowhole, located on top of their heads. When they come to the surface, the blowhole opens, allowing them to take air into their lungs. When they dive back under water, the blowhole is tightly closed by a layer of muscle. Dolphins cannot take in air through their mouths, they only breathe through their blowholes. In this way, breathing and eating are kept entirely separate so that dophins can capture prey in their mouths and swallow it without the risk of water entering their lungs.

TEST YOUR UNDERSTANDING

9 What is the difference between movement and locomotion?

10 Why is locomotion important to animals

11 Suggest three adaptations for swimming that marine mammals have for swimming.

Links

- How is energy needed for movement made available? (Section 3.2)
- How have movement and locomotion led to analogous structures in different classes of organisms? (Chapter 11)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

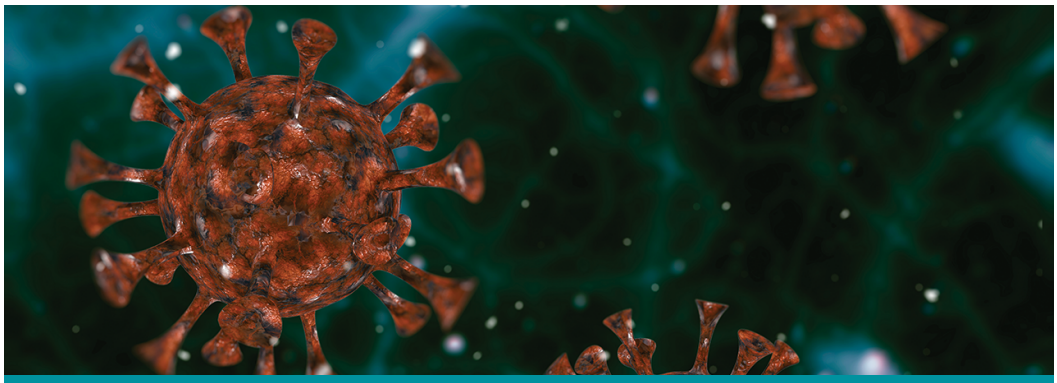
I can...	Subsection	Needs more work	Nearly there	Confident to move on
recall that inputs to the spinal cord and brain are carried by sensory neurones	9.1.1			
understand that muscles are stimulated by outputs from the brain via motor neurones	9.1.1			
recall that a nerve contains bundles of fibres of both sensory and motor neurones	9.1.1			
summarise the stage in a reflex arc	9.1.1			
identify receptors and effectors	9.1.1			

understand the roles of the cerebrum and cerebellum in controlling movement	9.1.1			
distinguish between voluntary and involuntary movement in the digestive system	9.1.1			
recognise that all living things can move but that some are sedentary	9.2.1			
outline the functions of a skeleton and the differences between an exoskeleton and an endoskeleton	9.2.2			
identify the parts of a synovial joint and the range of movements of different joints	9.2.2			
describe the action of antagonistic	9.2.3			

muscles in breathing				
outline the sliding filament theory of muscle action and the importance of the protein titin	9.2.3			
outline the structure of motor units in skeletal muscle	9.2.3			
distinguish between movement and locomotion and summarise reasons for locomotion	9.2.4			
summarise the adaptations of marine mammals that enable locomotion in water.	9.2.4			

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.



Chapter 10

Defence against disease

C3.2

INTRODUCTION

A pathogen is a living organism or virus that invades the body and causes infectious disease. Most pathogens are bacteria and viruses, but protoctists, parasitic worms and fungi can also be pathogenic. Living organisms defend themselves against the pathogens that cause infectious diseases.

10.1 Defence against disease

LEARNING OBJECTIVES

In this section you will:

- learn that living organisms defend themselves against pathogens which cause diseases
- distinguish between the innate and adaptive immune systems
- understand that the first lines of defence are the skin and mucous membranes
- recognise that blood clotting is a series of reactions that seals the skin if it is cut
- learn that phagocytes are a type of blood cell that recognise and remove pathogens in a series of stages that result in phagocytosis
- learn that animals have complex immune systems that attack pathogens by ‘challenge and response’
- define antigens as recognition molecules on cell surfaces
- recognise that antigens on red blood cells stimulate antibody production in a person with a different blood group
- learn that lymphocytes are blood cells that produce antibodies in the specific humoral responses to infection

- recognise that some lymphocytes act as memory cells and confer immunity
- understand that cell-mediated response to infection involves cytotoxic T lymphocytes, B lymphocytes and helper T-cells
- learn that activated B-cells form clones of plasma cells and memory cells as part of the humoral response and that memory cells provide immunity if a pathogen is encountered again
- discover that antibiotics can block prokaryotic cell process but some strains of bacteria have evolved resistance to them
- learn that viral diseases cannot be treated with antibiotics
- discover that scientists are seeking new sources of antibiotics to treat bacterial infections
- understand that human immunodeficiency virus (HIV) is a virus that infects the immune system and can lead to AIDS
- understand that pathogens can transfer from one species to another and are called zoonoses
- discover how vaccines stimulate an immune response and how smallpox was eradicated by vaccination
- recognise how epidemics are prevented by herd immunity.

GUIDING QUESTIONS

- How do organisms defend themselves against pathogens that cause disease?
- What prevents animals from destroying their own cells and tissues?

KEY POINTS

bacteria are prokaryotic microorganisms some of which can cause disease.

infectious diseases are caused by pathogens, such as bacteria, viruses, parasites or fungi that can be spread from one person to another.

pathogens are biological agents that can cause infectious disease.

clotting factors are proteins found in blood that work with platelets to help the blood clot if a blood vessel is broken.

KEY POINTS

leucocytes are immune cells that circulate in the blood and in the lymphatic system. There are five types of leucocyte.

mucous membranes line cavities in the body and cover the surface of internal organs.

platelets are cell fragments in the blood that release clotting factors.

10.1.1 Infection and response

In our daily lives we are exposed to many different disease-causing agents. Any organism or virus that can cause disease is known as a pathogen. Many different organisms can infect the human body and cause infectious disease; these include bacteria, viruses, fungi and protoctists (microscopic single-celled organisms such as *amoebae*). Relatively few bacteria and fungi are pathogens: most are free-living and useful in the environment (Chapter 12) but some cause serious illness and even death. Cholera, leprosy, tuberculosis and syphilis are all caused by bacterial infection. Fungal infections cause athlete's foot, ringworm and yeast infections such as thrush. Most protoctist diseases in humans are caused by protozoa, which are human parasites. *Trypanosoma* protozoa cause Chagas disease and sleeping sickness, *Giardia* protozoa cause giardiasis and *Plasmodium* protozoa cause malaria.

All viruses have the potential to be pathogenic, because no virus can function outside the cell of its host organism. A virus takes over the mechanisms of its host's cells and directs them to make more viruses (Section 5.3). Examples of viral infections include measles, rubella, chickenpox, shingles, influenza and COVID-19.

The body's first line of defence

Despite the fact that we come into contact with many pathogens every day, we are not often ill. This is due to our effective immune system, which both prevents pathogens entering the body and also deals with any that do. The body has a series of responses to disease-causing organisms.

The first line of defence against infection is our skin. Skin acts as a physical and chemical barrier to pathogens. Unbroken skin is a tough barrier to any potential invaders. It is waterproof and its secretions repel bacteria. Sebum from sebaceous glands has antibacterial properties and sweat, which is slightly acidic, also inhibits bacterial growth. Openings in the skin, such as eyes and nose, can provide entry points for pathogens but these are protected by mucous membranes, which line the respiratory, urinary, reproductive and intestinal tracts. Secretions such as tears, mucus and saliva contain the enzymes lysozyme and phospholipase. Lysozymes kill bacteria by catalysing the hydrolysis of bonds in their cell walls, while phospholipases destroy phospholipids in bacterial membranes. In addition, if pathogens are swallowed in food or water, the acidic environment of the stomach helps to kill them.

Blood clotting

If the protective layer of our skin is broken or cut and blood vessels are broken, pathogens have a route into the bloodstream. To prevent blood loss and the entry of pathogens, any blood that escapes from a damaged vessel quickly forms a clot, which plugs the gap.

Platelets, erythrocytes (red blood cells) and leucocytes (a type of white blood cell) are all important in the clotting process. Platelets are small cell fragments that form in the bone marrow and circulate in the bloodstream. Also important are two plasma proteins, which are present in the blood in their inactive forms until they are activated when needed (Figure 10.1.1). These two inactive proteins are **prothrombin**, a glycoprotein, and **fibrinogen**, a plasma protein produced by the liver.

If a small blood vessel is damaged, injured cells or platelets release clotting factors, which cause platelets to stick to the area. These factors activate prothrombin, which is converted to its active form, thrombin. Thrombin in turn activates the soluble protein fibrinogen, converting it to active fibrin, which is insoluble and forms long threads. This cascade of reactions ensures a speedy response to any damage. Fibrin forms a mesh of fibres that covers the damaged area and traps passing blood cells, forming a soft clot (Figure 10.1.2). If a clot is exposed to air, it dries and forms a scab, which will protect the area until the tissue beneath has been repaired.

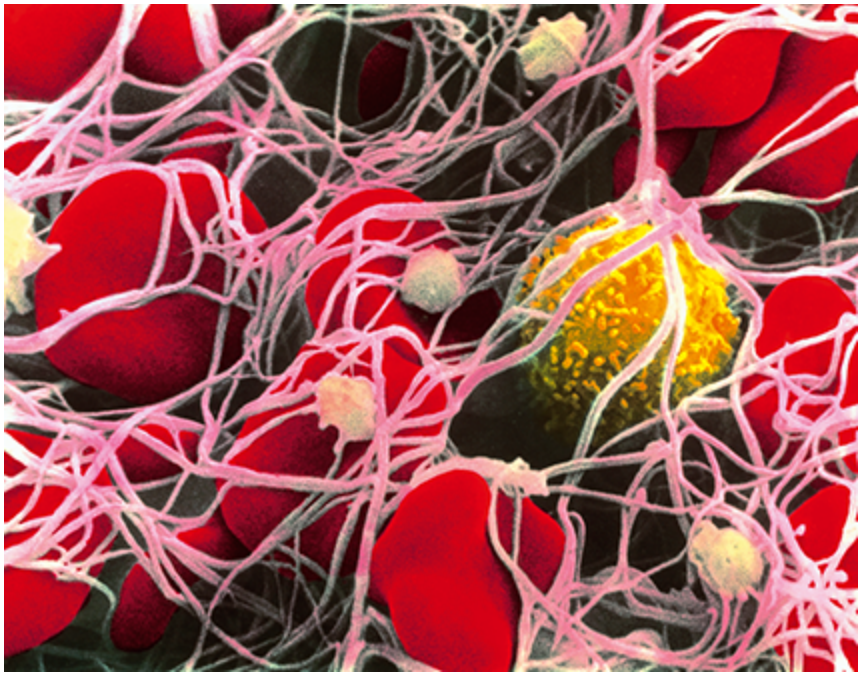


Figure 10.1.1: False-colour transmission electron micrograph showing red blood cells and threads of fibrin forming a clot (× 3600).

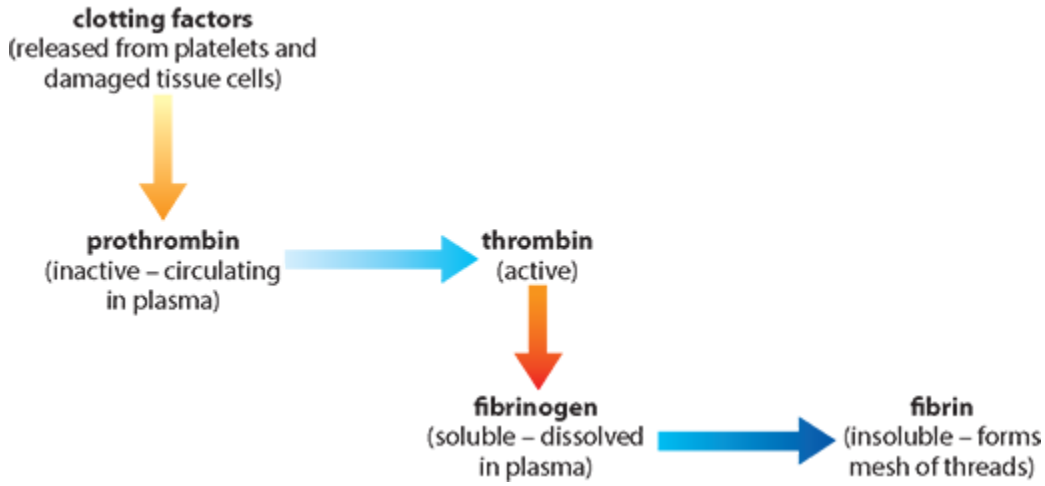


Figure 10.1.2: The sequence of reactions in the blood-clotting cascade.

KEY POINTS

an antigen is a substance that stimulates the production of antibody.

non-specific immunity refers to response of the body to any pathogen.

phagocytes blood cells are capable of engulfing bacteria and other small cells by phagocytosis.

phagocytosis is the process of modifying the shape of a phagocytic cell so that it can engulf bacteria or other particles.

Innate immune responses and adaptive immunity

The **innate immune system** is the body's first line of defence against pathogens that enter the body. It responds in the same way to all invaders and foreign substances. It is often called the non-specific immune system. It acts quickly so that if pathogens enter the skin through a small wound, they will be detected and

destroyed within hours. But the innate immune system has limited ability to prevent pathogens spreading.

The innate immune system consists of:

- protection by the skin and mucous membranes
- protection by phagocytes and proteins.

The innate immune system does not change during an organism's lifetime.

The **adaptive immune system** takes over if the innate immune system is unable to destroy a pathogen. It is very specific to the pathogen that is causing an infection. The response is slower than the innate immune system because the pathogen must be identified before a response can occur. After this the adaptive immune system is more accurate and specific in its response. It can also remember pathogens, so that if the same pathogen enters the body again, the adaptive immune system can respond faster.

This memory is also the reason why the body can become immune to certain infections. It may take a few days for the adaptive immune system to respond the first time, but the next time the body encounters the pathogen it can react immediately. The second infection is likely to be milder.

The adaptive immune system is made up of:

- T lymphocytes in the tissue between the body's cells
- B lymphocytes, also found in the tissue between the body's cells
- antibodies in the blood and other bodily fluids.

Non-specific innate immunity

Pathogens that do enter the body are soon recognised by phagocytic leucocytes, which form a vital part of the body's innate immune system. These specialised white blood cells circulate in the blood system and, because they are easily able to change their shape, can also squeeze in and out of capillaries. Phagocytic leucocytes respond to invaders by engulfing and destroying them in a process called phagocytosis (Figure 10.1.3).

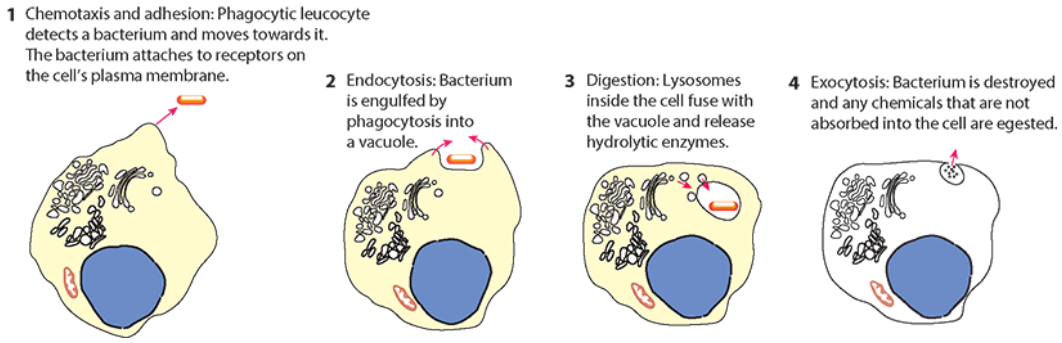


Figure 10.1.3: Phagocytosis of a pathogen.

This type of response provides non-specific immunity, which is so called because the phagocytes respond in the same way no matter what pathogen they meet. Phagocytes are able to move by amoeboid action, changing the shape of their cells to propel themselves along to where they are needed. Phagocytes are able to distinguish between invaders, foreign bodies that are not part of the individual's own body (non-self), and those which form part of the body (self). When phagocytes come into contact with bacteria, the receptors on the phagocyte's surface will bind to them. If the body is infected, chemical signals attract phagocytes to places where the pathogen has invaded the body in a process known as chemotaxis. These signals may come from pathogens or from other phagocytes already in the area. As Figure 10.1.3 shows, the phagocyte adheres to the pathogen, engulfs it by endocytosis and once the pathogen is inside the phagocyte, releases hydrolytic enzymes that destroy it. Any waste products

leave the phagocyte by exocytosis. Phagocytes do not attack the body's own cells and tissues but they may remove dead tissues that result from apoptosis.

KEY POINTS

an antibody is one of millions of blood proteins produced by plasma cells in response to specific antigens, which are then neutralised or destroyed.

immunity is resistance to the onset of a disease after infection by the agent causing the disease.

lymphocyte is a type of leucocyte that produces antibodies. Some act as memory cells.

memory cells are lymphocytes that enable the body to respond quickly to an antigen it has already encountered.

Adaptive (specific) immunity and antibody production

Many animals have complex immune systems that attack different pathogens in very specific ways known as challenge and response. This response can be summarised as follows:

- If the body is challenged by a pathogen it responds with both non-specific and specific immune reactions.
- All the body's own cells have molecular markers on their cell plasma membranes that identify each cell as 'self' and belonging to the body. The body can recognise invading pathogens because they do not have these markers on their surface.
- Non-specific immune cells (macrophages) ingest and put the foreign antigens on their surface membranes. The

macrophages present these antigens to lymphocytes as examples of ‘non-self’.

- Lymphocytes can then respond with the production of specific antibodies to destroy the non-self invaders. This process is known as the humoral response.

Antigens (anti body-generating substances) are glycoproteins and proteins found embedded in the plasma membranes or cell walls of bacteria or in the protein coat of a virus. These antigens enable the body to recognise a pathogen as being ‘not self’ – that is, not a part of the body – and they give a clear signal to switch on the immune response, with the rapid production of antibodies.

Antibodies are protein molecules that are produced by lymphocytes, a type of leucocyte found in the blood and lymph nodes, in response to any antigen that enters the body. There are millions of different antibodies and each one is specific to an antigen. For example, the antibodies that lymphocytes produce in response to infection by an influenza virus are quite different from those produced by different lymphocytes in response to a tuberculosis bacterium. Even fragments of pathogens, or their toxins, can stimulate the release of antibodies. After an infection has passed, some of the lymphocytes giving rise to antibodies specific to the infecting antigen remain in the bloodstream as memory cells. This means that the immune system can respond quickly if the same antigen enters the body again later, by producing antibodies and preventing a widespread infection. The person is said to have acquired immunity to the antigen. Figure 10.1.4 explains how antibodies are made.

Each antibody molecule has a basic Y shape but the arrangement of molecules at the top of the Y shape forms specific binding sites that give every antibody its own unique properties (Figure

10.1.5). These specific binding sites attach to the corresponding antigen site on the surface of the pathogen or its toxin. Once an antibody has bound to an antigen, it can destroy it in one of a number of ways. Some cause bacterial cells to clump together, making the job of phagocytes easier. Others cause cell walls to rupture, deactivate toxins or act as recognition signals for phagocytes, giving a clear indication that action is needed (Figure 10.1.5).

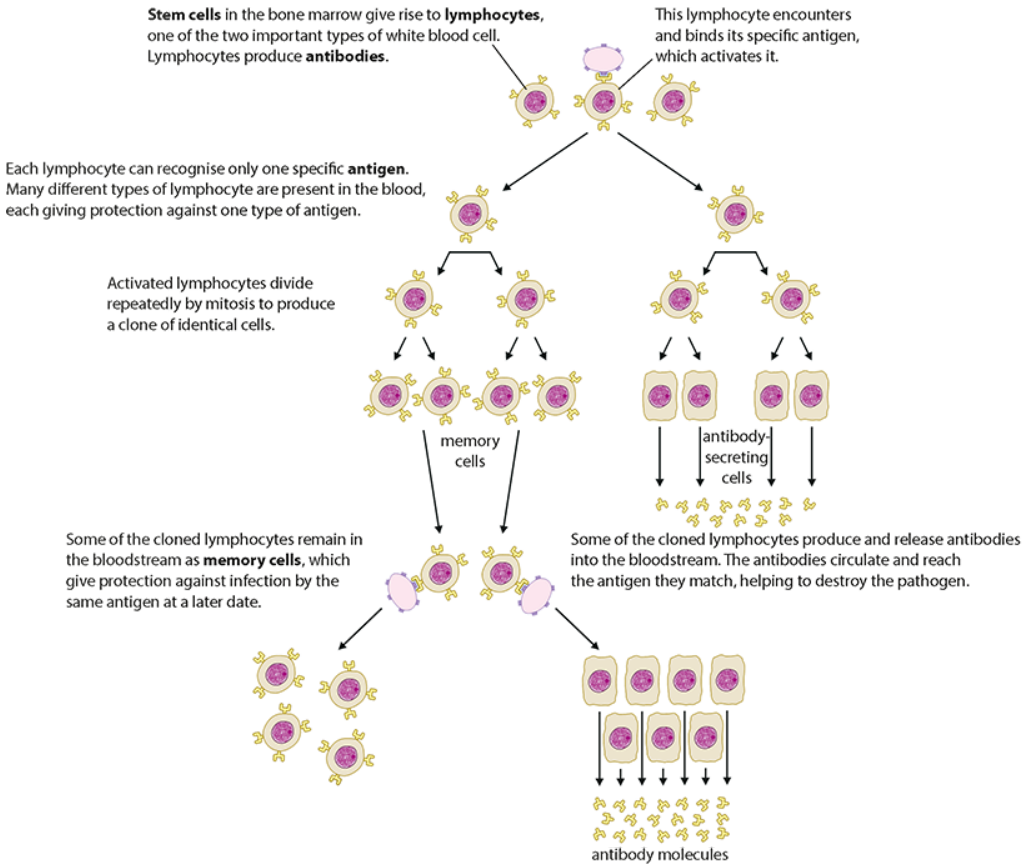


Figure 10.1.4: Antibody production involves a series of stages that also result in memory cells being produced.

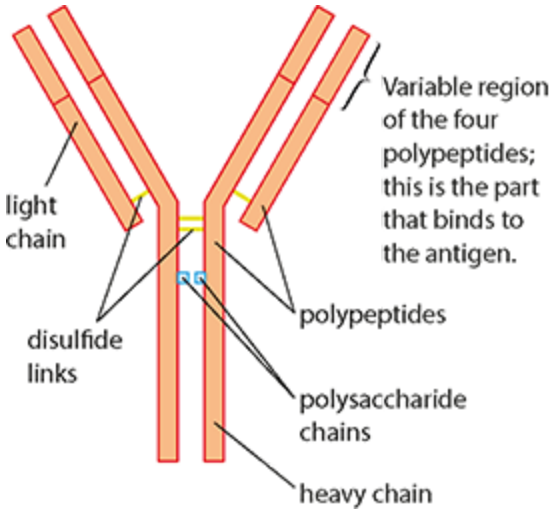


Figure 10.1.5: The basic structure of an antibody molecule.

A vaccination protects us from a specific disease by stimulating our immune system to produce antibodies against antigens carried by the disease-causing agent. Memory cells then remain in our bloodstream, so that if the actual disease-causing agent is encountered later, antibodies can be released quickly and an infection avoided. (You can read more about vaccination in the Higher Level section of this chapter.)

The importance of antigens in blood groups and antigens

There are four main human blood groups and these are classified by the ABO system according to the antigens that red blood cells have on their surfaces and the antibodies found in blood plasma (Section 4.6).

Antigens are protein molecules found on the surface of red blood cells.

- blood group A – has A antigens on the red blood cells with anti-B antibodies in the plasma

- blood group B – has B antigens with anti-A antibodies in the plasma
- blood group O – has no antigens, but both anti-A and anti-B antibodies in the plasma
- blood group AB – has both A and B antigens, but no antibodies.

Blood transfusions are given in medical procedures or to save the life of a person who has lost a lot of blood, but receiving blood from the wrong group can be life-threatening. For example, if a person with group B blood is given group A blood, their anti-A antibodies will attack the group A cells and destroy them. So group A blood must never be given to someone who has group B blood and vice versa. As group O red blood cells do not have any A or B antigens, group O blood can safely be given to a person who as any other blood group.

10.1.2 Cell-mediated and humoral responses

Immunity: challenge and response

Resistance to an infection is known as immunity. Immunity is acquired from infancy onwards as the body is exposed to, and learns to recognise, many different types of pathogen that have the potential to cause disease. We become able to distinguish between cells that are our own ‘self’ and those that are ‘non-self’ and are therefore likely to be pathogens or cause harm. Cells are recognised by the proteins on their plasma membranes.

Certain leucocytes (a type of white blood cell) are able to recognise ‘non-self’ proteins, or antigens. Antigens may be on the surface of a pathogen, or may be part of a toxin secreted by a pathogen. Antigens are also likely to be present on the cell surfaces of transplanted tissues or organs and on the surfaces of some cancer cells.

If a pathogen enters the body, the immune system is stimulated to respond. As it is ‘challenged’ by the pathogen, it ‘responds’ by setting in motion processes that will destroy it. The first line of defence is phagocytic leucocytes (macrophages). These cells respond in a non-specific way and will consume bacteria, viruses and other pathogens, as well as dead cells and cell fragments.

Our second line of defence is a specific response to antigens. Healthy cells have ‘self-antigens’ on the surface of their membranes. If a cell is infected with a virus, it has pieces of viral antigens on its surface. This is a signal for the immune system that lets it know this is a cell that must be destroyed.

One important group of cells involved in destroying both pathogens and cancer cells are cytotoxic T lymphocytes or killer T-cells. Killer T-cells find and destroy infected or mutated cells. To do this they recognise the difference between the infected cells and healthy cells from the antigen markers on their surface membranes. Killer T-cells are able to find cancerous cells or cells that contain viruses and destroy them.

The second line of defence also involves the production of antibodies.

There are two main parts of our immune system that allow us to protect ourselves from disease by producing antibodies. They are the cell-mediated response and the humoral response. These systems work together and separately. The cell-mediated response involves the production of specialised lymphocytes called T-cells, while the humoral response is associated with the blood serum (the non-cellular part of the blood) and involves antibodies secreted by B-cell lymphocytes.

KEY POINTS

cell-mediated response production of specialised T-cells in the immune system.

humoral response a series of immune reactions in the blood that lead to the production of antibodies, specifically the action of B-cells in the non-cellular (serum) part of the blood.

Antibody proteins are vital to the body's immune response, and producing them effectively requires interaction between three types of cell:

1 macrophages

2 B lymphocytes (B-cells)

3 helper T lymphocytes (T-cells).

The immune response takes several days to become fully active and in the meantime we may become ill. Sometimes symptoms are mild, such as with the common cold, but sometimes they are severe, leading to permanent disability or even death.

Clonal selection: humoral response

B-cells are antibody-producing lymphocytes, but each B-cell can only produce one particular type of antibody. Since the antibody–antigen response is highly specific, there must be a great many types of B-cell to be able to respond to all the possible types of antigen. At any time, there can only be a few of each type of B-cell in the bloodstream because most of the blood volume is taken up with red cells.

B-cells recognise and bind antigens, including bacteria, viruses and toxins (free antigens) outside body cells, and each B-cell recognises one specific antigen. When a pathogen enters the bloodstream, its surface antigen molecules are exposed to the antibodies attached to different B-cells in the blood. If there is a match between an antigen and an antibody, the B-cell with the matching antibody becomes ‘selected’, while all the other B-cells are rejected. The selected B-cell is stimulated to divide and produces a clone of antibody-secreting cells, in a process known as clonal selection.

It is likely that any pathogen will have many different antigenic molecules on its surface so several different types of B-cell will probably be selected. Each of these will result in clone of antibody-secreting B-cells. This is therefore called a polyclonal response and it will result in a more efficient destruction of the

pathogen as the antibodies neutralise or inactivate the antigens. B-cells are also important in the cell-mediated response.

Antibody production and the cell-mediated response

The response to pathogens is more complex than simple clonal selection and it involves two types of lymphocyte: B-cells and T-cells. There are several different types of T-cells that respond to antigen molecules that have been processed and presented by infected cells or phagocytic cells in the cell-mediated response.

The stages of antibody production are:

1 When a pathogen enters the bloodstream it is consumed by a macrophage, partly digested and antigen proteins from it are placed on the outer surface of the macrophage. This is called antigen presentation because the proteins are being 'presented' to other cells.

2 Helper T-cells with receptors matching the presented antigens bind to the macrophages and are activated.

3 Activated helper T-cells then start dividing into two types of clones of cells. One clone type is active helper T-cells, which are required for the next step in the process, and the other clone type is memory cells, which will be used if the same pathogen ever invades the body again.

4 B-cells with the matching antibody also take in and process antigen proteins from the pathogen and place them on their outer surface.

5 Active helper T-cells bind to these B-cells and, in turn, activate them.

6 Just like the T-cells, the B-cells now divide into two clones of cells. One is made up of active B-cells, or plasma cells, which secrete huge quantities of antibodies into the bloodstream. The second clone is made up of memory cells.

KEY POINT

clonal selection exposure to antigen results in activation of selected T-cell or B-cell clones, producing an immune response.

7 Antibodies in the bloodstream destroy pathogens and also help the macrophages to detect and consume more pathogens.

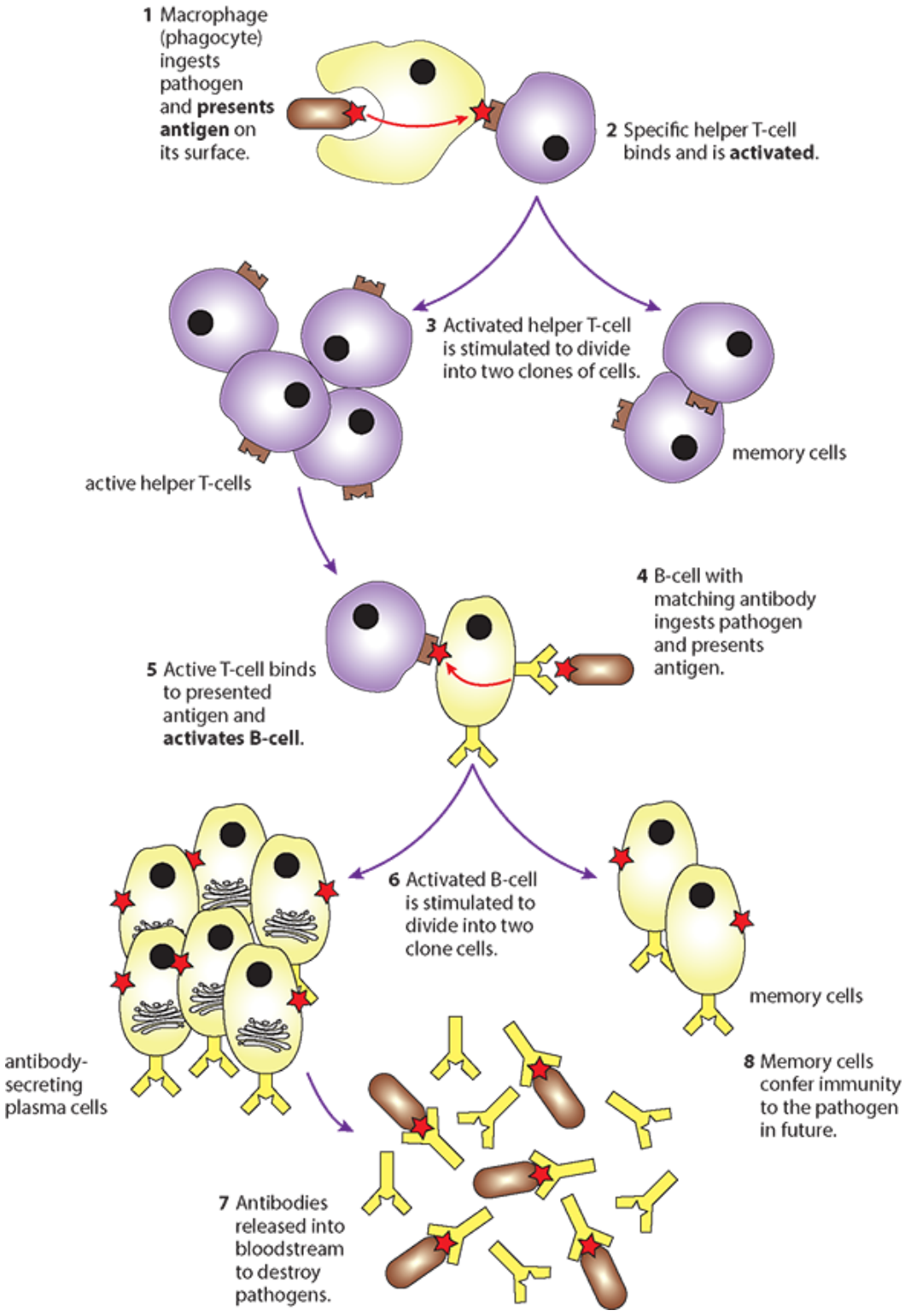


Figure 10.1.6: Summary of the process of antibody production in the cell-mediated response.

Type of cell	Function of cell
macrophages (phagocytic leucocytes)	engulf any pathogen or dead cells and present their antigens on their cell surface membrane
cytotoxic T lymphocytes (killer T-cells)	destroy cells that are ‘non-self’ and have antigen on their surface membrane
B lymphocyte (B-cells)	B-cells are cloned to become activated plasma cells to produce antibodies, or to become memory cells in the humoral response
T lymphocytes (T-cells), helper T-cells	bind to macrophages and become activated; T-cells are cloned to produce more activated T-cells and memory cells
plasma cells	antibody-producing cells derived from T-cells and B-cells; part of the humoral response
memory cells	cells produced by B-cells and T-cells which remain in the bloodstream after infection to prolong resistance to the antigen

Table 10.1: Cells involved in the immune response.

8 Memory cells remain and allow the body to make a large and rapid response should the same pathogen invade again. It is the persistence of memory cells that gives the organism immunity to that pathogen in the future.

Figure 10.1.6 The production of antibodies in the cell-mediated response involves macrophages, helper T-cells and B-cells.

Plasma cells contain large amounts of rough endoplasmic reticulum to synthesise antibodies.

Table 10.1 summarises the types of cell involved in the immune response.

10.1.3 HIV and AIDS

Human immunodeficiency virus (HIV; Figure 10.1.7), first identified in the early 1980s, causes the series of symptoms together known as acquired immune deficiency syndrome, or AIDS. HIV infects only the helper T-cells, a type of lymphocyte that is important in maintaining communication between cells of the immune system. After a latent period of months or years, helper T-cells are gradually destroyed and, as their numbers fall, so does the body's ability to fight infection. Helper T-cells instruct other lymphocytes to clone and generate antibodies, and without them an infected person can no longer fight off pathogens. Secondary infections result and the person is said to be suffering from AIDS.

HIV is a retrovirus, which means it can insert its DNA into that of a host cell using a protein called reverse transcriptase. Even if all the viruses in the body could be removed, the infected T-cells would continue to make new viruses.

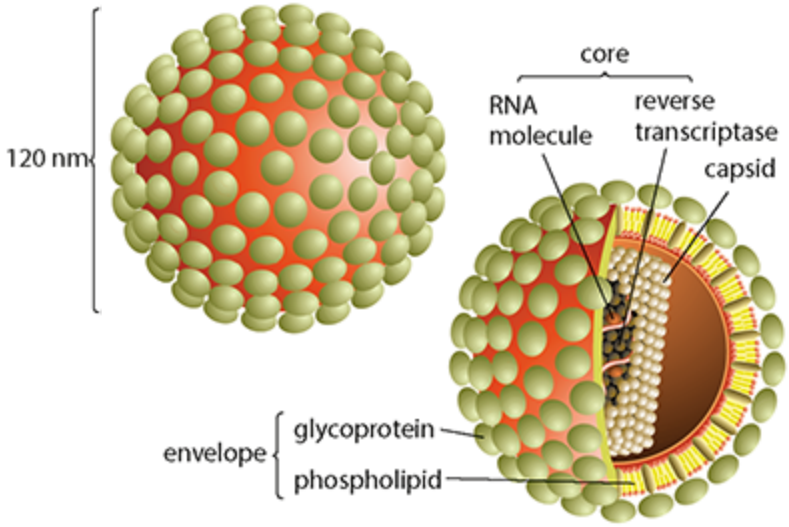


Figure 10.1.7: HIV viruses consist of a spherical glycoprotein and lipid coat enclosing two strands of RNA. The virus is 60 times smaller than a red blood cell.

Cause and consequences of AIDS

HIV is transmitted in blood, vaginal secretions, semen, breast milk and sometimes across the placenta. In some countries, HIV has been transmitted in blood transfusions, but in most places with medical care facilities, blood for transfusion is now screened for the virus. The virus is most frequently passed from person to person in bodily fluids during sex and also when non-sterile syringe needles are used to administer either legal or illegal drugs.

AIDS is the end stage of an HIV infection. It is caused by a severe failure of the immune system as the HIV virus selectively infects helper T-cells. Some infected individuals have no symptoms in the early stages of the disease while others may be slightly unwell when first infected. Symptoms of AIDS develop as the number of active helper T-cells decreases. The symptoms occur as a result of secondary infections caused by bacteria, fungi and viruses. The body is unable to resist these infections due to its compromised immune system (Figure 10.1.8).

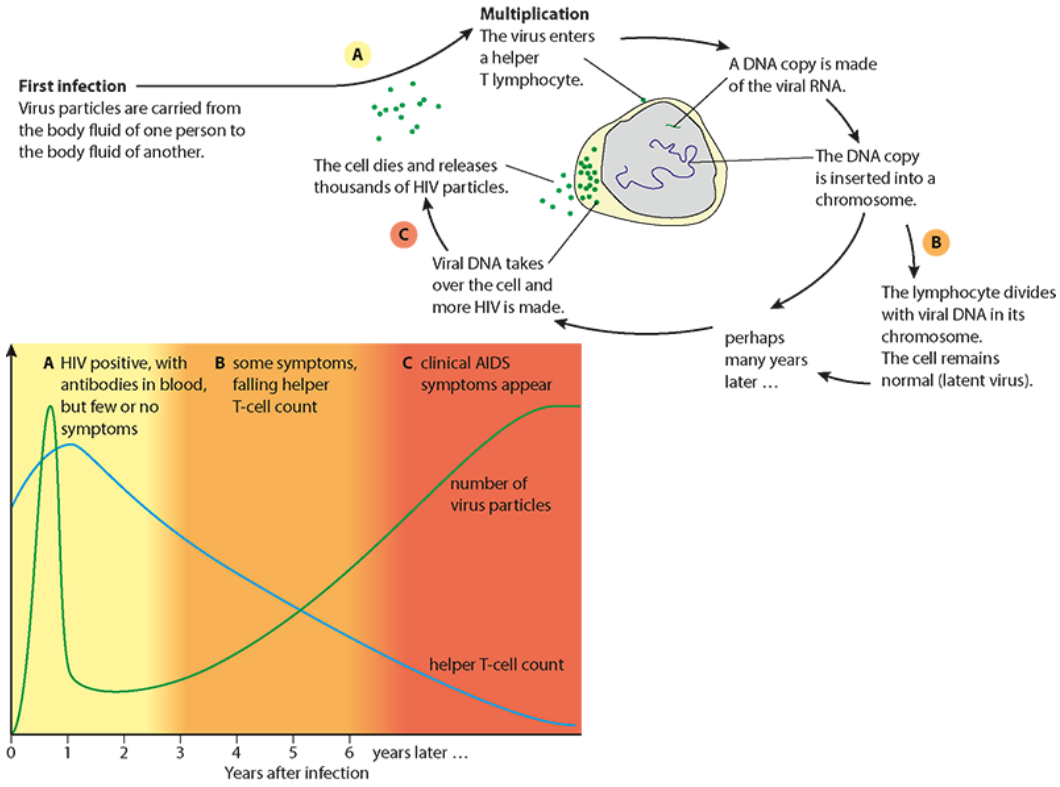


Figure 10.1.8: HIV infection proceeds through three stages: **A** HIV-positive with few symptoms, **B** some symptoms and low helper T-cell count and **C** clinical AIDS with associated symptoms.

10.1.4 Antibiotics

If the body's natural defences are unable to deal with an infection, medical intervention may be needed. Most bacterial infections can be treated with **antibiotics**. Antibiotics are natural substances that slow the growth of bacteria. Since the discovery of penicillin in 1928, many antibiotics have been isolated and about 50 are now manufactured for medical use. These antibiotics work in different ways but are effective because prokaryotic and eukaryotic cells have different metabolic pathways. So, eukaryotic cells are not affected by antibiotics. Some antibiotics block the protein synthesis mechanism in bacteria while not affecting the process in human cells (Figure 10.1.9). Others interfere with the formation of the bacterial cell wall and prevent bacteria growing and dividing.

Viruses are not living and have no metabolic pathways of their own. Since they use their human host's metabolism to build new viruses, antibiotics have no effect against viral infections.

INTERNATIONAL MINDEDNESS

International aspects of disease

AIDS is a worldwide pandemic but some regions and some age groups are more seriously affected than others. There were approximately 38 million people across the world with HIV/AIDS in 2019. Of these, 36.2 million were adults and 1.8 million were children under the age of 15 years. The vast majority of people with HIV are in low-income and middle-income countries. In 2019, there were 20 million people with HIV in eastern and southern Africa, 5 million in western and

central Africa, 6 million in Asia and the Pacific and 2 million in Western and Central Europe and North America.

Nevertheless, education campaigns and new treatments have improved the situation and there has been a decline in new HIV infections of almost 25% since 2010.

In 2019 nearly 70% of people infected with HIV (about 25 million) were receiving antiretroviral therapy (ART). People who know they have HIV and take ART can remain healthy and are unlikely to pass the virus on. In addition, 85% of pregnant people with HIV received ART to prevent transmission of HIV to their babies during pregnancy and childbirth. AIDS-related deaths have been reduced significantly since the peak in 2004. In 2019, around 690 000 people died from AIDS-related illnesses worldwide, compared to 1.1 million in 2010.

The spread of the HIV virus, and other pathogens such as avian influenza (bird flu) and SARS (COVID-19) are problems for the whole world. International travel means that pathogens can travel further and faster than ever before. Governments and health authorities must work together to co-ordinate their responses.

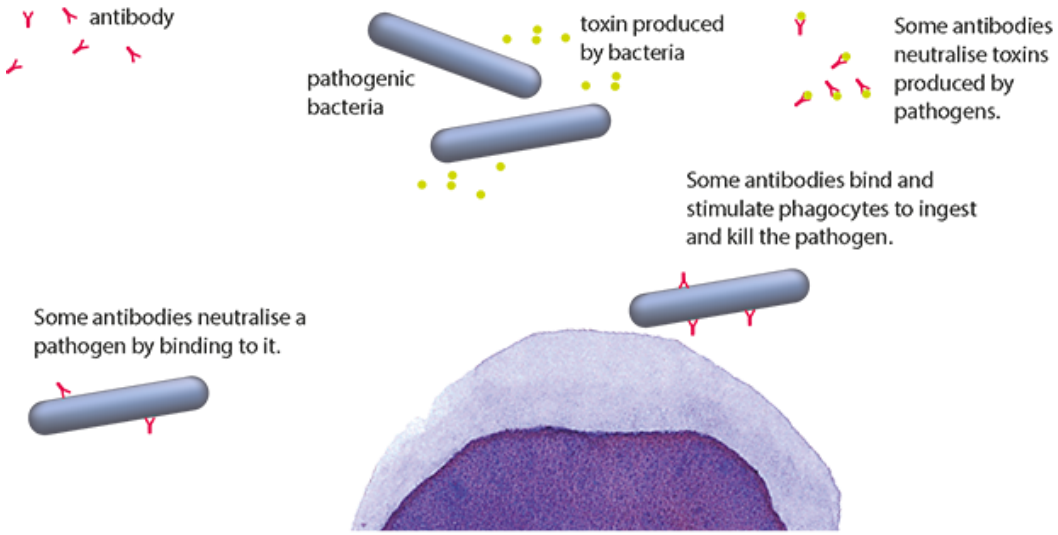


Figure 10.1.9: The various ways in which different antibodies can destroy bacteria or their toxins.

Antibiotic resistance

Antibiotics kill or block the growth of bacteria but not all bacteria are susceptible to them. In any population of bacteria some individuals will have a natural resistance to the antibiotic used to kill them. This resistance may arise spontaneously by mutations. Resistant strains multiply along with susceptible strains but, if antibiotics are used, only the sensitive bacteria will be killed while the resistant ones survive. Resistant bacteria are also able to pass on their resistance to other bacteria via their plasmids. Plasmids are loops of DNA, separate from the bacterial chromosome, that can be passed from one bacterium to another. Treating a disease caused by resistant strains of bacteria becomes very difficult. Doctors may have to prescribe stronger doses of antibiotic or try different antibiotics to kill the resistant bacteria.

NATURE OF SCIENCE

Assessing risk in science: collaboration, safety and new medicines

The discovery of antibiotics began by accident. On 3 September 1928, Alexander Fleming was examining a batch of culture plates on which he had grown *Staphylococcus* bacteria. He noticed that one of the plates had a green mould growing on it. The mould was *Penicillium notatum*. The mould growth was circular in shape, and the area around it seemed to be free of *Staphylococcus*. On other areas of the plate, the bacteria were continuing to grow well. Fleming deduced that the bacteria around the circular mould had been killed by a substance produced by the mould.

Fleming discovered that the mould could kill other bacteria and that it could be given to small animals without any harmful effects. However, he then moved on to other research and it was not until 10 years later that Howard Florey and Ernst Chain, working at Oxford University, isolated the bacteria-killing substance, penicillin, produced by the mould. Chain was a German chemist and Florey an Australian pathologist. It was Chain who isolated and purified penicillin and Florey who tested its safety to use on animals and humans. The first tests the team carried out on mice in 1940 would not have met the stringent standards for testing on animals today. Eight mice were given lethal doses of *Streptococcus* bacteria. Half the mice were then given injections of penicillin. The following day all the untreated mice were found to be dead but those that had been given penicillin survived.

One of the first uses of penicillin was in 1941, when Dr Charles Fletcher gave it to patient at a hospital in Oxford who was near to death as a result of bacterial infection in a wound.

Fletcher used some penicillin on the patient and the wound made a spectacular recovery. Unfortunately, Fletcher did not have sufficient penicillin to clear the patient's body of bacteria and he died a few weeks later as the pathogen regained a hold.

Figure 10.1.10: The discovery of penicillin is commemorated with a plaque on the wall of St Mary's Hospital in London where Alexander Fleming carried out his work.

An American brewing company began mass production of penicillin and soon sufficient quantities were available to treat all the bacterial infections among the troops fighting in World War II. Penicillin was nicknamed 'the wonder drug' and in 1945 Fleming, Chain and Florey shared the Nobel Prize in Physiology or Medicine for its isolation and development.

To consider:

1 Why did the discovery of penicillin have such a profound effect on people at the time?

2 Why was the collaboration of three scientists vital to the discovery of penicillin?

3 What are the ethical issues involved in using a new drug for the first time? Was Fletcher right to use penicillin on his patient?

4 The test on the safety of penicillin used by Florey would not be accepted today. What are the risks associated with the development of new medicines?

The more often antibiotics are used, and the more different types that are used, the greater the risk that resistance will develop. So, over-use and the improper use of antibiotics are thought to have contributed to the development of resistance. Bacteriologists are concerned that some diseases will become untreatable with currently available antibiotics. The so-called superbug meticillin-resistant *Staphylococcus aureus* (MRSA) now has multiple resistance to many antibiotics and recently strains of the bacteria that cause tuberculosis and the sexually transmitted disease gonorrhoea have been found to be resistant to *all* the antibiotics that have been used to treat them.

SCIENCE IN CONTEXT

Sources of new antibiotics

Bacterial diseases cause social and economic problems for countries all over the world and modern travel has enabled infections to spread rapidly. The problem of drug-resistant bacteria is serious and growing and is likely to become one of the biggest threats to human and animal health in the 21st century. Most people are aware of the ‘superbugs’, such as MRSA and *Clostridium difficile*, which are a leading cause of

infections in hospitals. But there are other resistant bacteria, such as drug-resistant *Mycobacterium tuberculosis*, the organism that causes tuberculosis, which is resistant to several antibiotics and can take up to 2 years to treat.

There have been few significant discoveries of new antibiotics for many years. Existing antibiotics are becoming less effective because some of the bacteria that are not killed by current antibiotics pass on antibiotic resistance to other species of bacteria. Traditional sources of antibiotics are fungi and bacteria which produce them to defend themselves against bacteria. Penicillin (described in the section on Assessing risk in science: collaboration, safety and new medicines) is one example of an antibiotic derived from a fungus. Scientists continue to seek potential new sources of antibiotics in the natural world in the soil, in plants and animals such as corals.

There are thousands of species of bacteria present in the soil but they are hard to grow and study in the laboratory. In the early 21st century a team of scientists from several countries worked together and devised a new method to find those that have antibiotic properties. The team designed a device called an ichip (isolation chip) which cultures different species of bacteria in their soil environment and has enabled the researchers to study many new microorganisms. They discovered that the soil bacterium *Eleftheria terrae* produces an antibiotic called teixobactin that has been shown to be effective against many drug-resistant strains. Teixobactin was discovered using the ichip in 2015. Tests revealed teixobactin was effective against Gram-positive bacteria including MRSA and the bacteria that cause tuberculosis. However, it was not effective against Gram-negative bacteria such as *Escherichia coli*. Gram-negative bacteria have an outer lipid layer, which Gram-positive bacteria do not.

Teixobactin inhibits bacteria in a way in which cells are unlikely to develop resistance. Penicillin inhibits the production of new cell walls, but teixobactin does the opposite. It prevents cell walls being broken down. For cells to grow and then divide their walls must be able to expand. If cell breakdown is blocked the cell is trapped and cannot expand or grow and so it will die.

Another new method in antibiotic research has been used at the Massachusetts Institute of Technology in the USA. Researchers used a machine-learning algorithm to seek and identify new antibiotics. The researchers designed a system to look for chemical features that make molecules effective at killing *E. coli*. They trained a computer to check 2500 molecules, including about 1700 already approved drugs and 800 natural substances with a wide range of different structures and biological activities. The computer identified a substance which the team called halicin after the fictional computer Hal in the film *2001: A Space Odyssey*. Halicin was tested against different resistant bacterial strains grown in the laboratory and others isolated from patients. It was able to kill many of them, including *C. difficile* and *M. tuberculosis.* With further development, halicin may become an antibiotic of the future.

TEST YOUR UNDERSTANDING

1 Define the term ‘pathogen’.

2 Describe what is meant by the term ‘antigen’.

3 State why antibiotics are not effective in treating viral diseases.

4 Distinguish between the roles of a ‘phagocyte’ and a ‘lymphocyte’.

10.1.5 Zoonoses - pathogens and species specificity

Pathogens that cause disease in one species do not always affect other species. For example, the pathogens responsible for syphilis, gonorrhoea, measles and polio infect humans, whereas canine distemper virus does not. *Shigella*, a bacterium that causes dysentery in humans and baboons, does not affect chimpanzees. The exact reasons for these differences are not fully understood, but it may be that cells in non-susceptible species do not have suitable receptors on their plasma membranes for the pathogens to bind to them. The temperature of the host organism may also be important: birds cannot be infected with mammalian tuberculosis that affects humans, deer, goats, pigs, cats, dogs and badgers because the bacteria that cause the disease cannot survive at the higher core temperatures of birds' bodies. Similarly, frogs are unaffected by anthrax-causing bacteria because their body temperatures are too low.

Occasionally, however, a disease does cross the species barrier, and a disease that does this is called a zoonosis. Zoonotic pathogens may be bacterial, viral or parasitic and can spread to humans through direct contact or through food, water or the environment. The malaria parasite is one example of a parasite that can infect humans but also infects birds, bats and lizards. It is transmitted via a mosquito that carries the juvenile stages of the parasite. Other zoonotic diseases include influenza, which has many variants that affect different species and is transmitted through the air from one species to another. Examples are: rabies caused by a virus which affects dogs, skunks, bats and foxes and is transmitted in the saliva when an infected animal bites another; anthrax, which affects hoofed animals such as deer, goats and

cattle as well as humans and is transmitted by contact with infected animals; and avian or bird flu, which affects wild and domestic birds and mammals who may become infected by saliva, mucus or feces of infected birds. Bird flu infections among people are very rare, but can happen if sufficient numbers of virus particles enter the eyes, nose, or mouth. Another serious zoonotic disease is Japanese encephalitis, a viral brain infection that is spread to humans through mosquito bites. It is very rare in humans, but the virus is found in pigs and birds in south east Asia. It is passed to mosquitoes when they bite an infected animal. All these diseases have serious social and economic consequences and cause loss of life and income.

KEY POINT

zoonosis is a disease that can cross species barriers and affect a different species.

Most newly emerging diseases that have crossed the barrier from animal to human are caused by viruses. Crossing to a new species is a rare occurrence, but viruses that do so can cause severe outbreaks of disease, especially if they develop the ability to pass from human to human, rather than just from animal to human. The spread of the COVID-19 virus through human populations has shown us how a virus can spread rapidly all over the world and how international travel and trade increase the spread of this, and other viruses.

For a virus to infect a new species, genetic adaptations must occur within the virus. Avian flu arose in this way. The expansion of both human and farm populations has made close interaction between birds and humans more common and enabled the virus to transfer from infected birds to humans via a

bird's saliva, feces or nasal secretions. Many strains of bird flu have emerged, but one of the most widely publicised is the H5N1 influenza virus, a highly pathogenic Asian strain that caused a pandemic in birds in 2003. This virus can cause severe illness in humans who are infected by direct contact with infected birds and it has a high mortality rate. It does not appear that H5N1 can be spread by human-to-human contact at present. However, because viruses can adapt and change quickly, it may evolve this ability at some point and would then have the potential to cause a human pandemic. Health agencies have begun the preparation of vaccines in case this should happen.

INTERNATIONAL MINDEDNESS

COVID-19

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus is thought to have originated in bats, but this has not been confirmed conclusively. Coronaviruses are not rare in nature and previously caused the human epidemic of SARS that began in 2003. The first known case of COVID-19 was identified in Wuhan, China, in December 2019 where the virus was isolated from three people with pneumonia. It is a new variant of the SARS-CoV-2 coronavirus and since it was identified many new variants have developed. All structural features of the new virus occur in related coronaviruses in nature, but mutations of this virus made it highly infectious to humans. COVID-19 disease was declared a pandemic by the World Health Organization on 11 March 2020, mainly due to the speed and scale of the transmission of the disease. The virus is a single-stranded RNA virus, and its genome encodes for 29 proteins involved in the infection, replication and virus

assembly process. Like other coronaviruses it has crown-like spikes on its surface. In less than one year, with international collaboration among the scientific community, the virus genome was sequenced, methods of transmission were identified, treatments were investigated and new vaccines were developed to combat it. The pandemic has caused loss of life and damage to economies throughout the world. It is possible that other viruses will do the same in the future.

Figure 10.1.11: Newspaper headlines around the world summarised the seriousness of the COVID-19 pandemic.

To consider:

1 How did the COVID-19 pandemic affect life where you live?

2 As the pandemic developed, new variants of the COVID-19 appeared. Research how these variants arose and how they affected the spread of the virus.

SCIENCE IN CONTEXT

Allergic responses

Sometimes when microorganisms enter the body, or if the skin is injured, the immune system may trigger an inflammatory response. Inflammation is either a general response to an injury or a reaction that occurs in an area where phagocytes are destroying pathogens. The inflammatory response is brought about by two types of cell. These are basophils, which are a type of leucocyte (white blood cell), and mast cells, found beneath the skin and around blood vessels. Both types of cell can be stimulated to release a substance known as histamine into the affected area. Histamine relaxes the muscle in the walls of arterioles so that blood flow to the affected area is increased and it also loosens the cells in the capillary walls so that they become 'leaky'. Plasma can then escape from capillaries into the surrounding tissue causing swelling (known as oedema) as well as a slight increase in temperature. Histamine also stimulates sensory neurons leading to pain or itching (Figure 10.1.12).

In some cases, histamine release can lead to an excessive immune response known as an allergy. An allergy is an immune response to an antigen (known as an allergen) to which most people show no reaction. Asthma, eczema and hayfever are common allergic disorders. Allergens include substances such as pollen grains, animal fur, house dust and certain foods. These substances have proteins on their surfaces which act as antigens and stimulate plasma cells to produce antibodies called reagins. Unlike normal antibodies, reagins circulate in the blood and bind to cells that contain histamine, especially the mast cells in the skin and mucus

membranes in the respiratory system. (These tissues are said to be hypersensitive.) Reagins cause the mast cells to release histamine, which binds to receptors on cells nearby and leads to inflammation and other symptoms of an allergy (Figure 10.1.12). In the bronchi, inflammation can lead to constriction and breathing difficulties as well as the secretion of excess mucus.

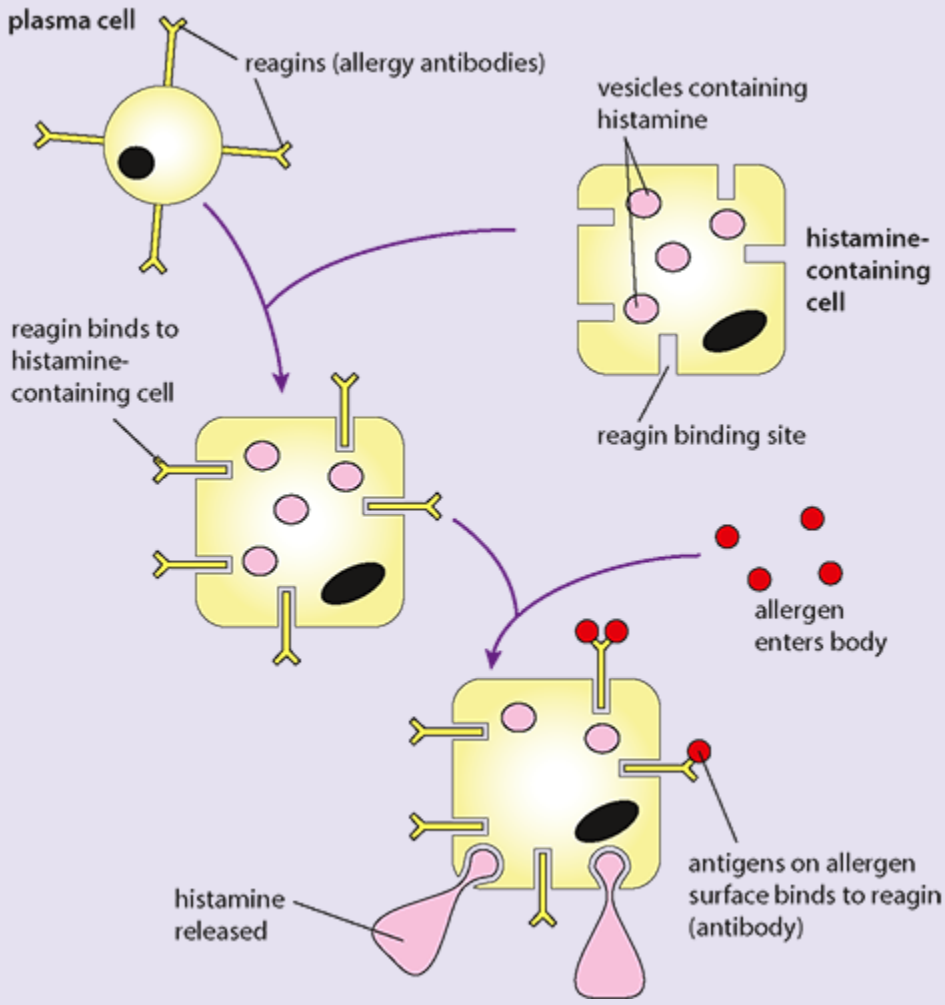


Figure 10.1.12: An allergic reaction occurs when plasma cells release antibodies against an allergen, which leads to the release of histamine.

KEY POINTS

allergy is an excessive immune response to an antigen.

histamine is a chemical produced by some white blood cells (leucocytes) that causes inflammation.

inflammatory response (inflammation) occurs when tissues are injured by bacteria or toxins, causing the release of histamine and other chemicals that leak into tissues, which leads to heat and swelling.

10.1.6 Vaccines and immunisation

Immunity develops when a person has been exposed to a pathogen. For most mild illnesses, such as the common cold or tonsillitis, this happens naturally as a person comes into contact with the viruses or bacteria that cause them. But some pathogens cause diseases that have dangerous or life-threatening symptoms. For these diseases, which include tetanus, tuberculosis, cholera, poliomyelitis, measles and COVID-19, vaccines have been developed to provide a safe first exposure so that a vaccinated person will develop immunity but not the disease.

KEY POINT

vaccine is a modified form of a disease-causing pathogen or their antigens that stimulate immunity without causing illness.

Vaccines are modified forms of the disease-causing pathogens. A vaccine may contain either weakened (attenuated) or dead pathogens carrying antigens, or their toxins. These vaccines are often produced by treating pathogens with heat or chemicals. The newest vaccines use sections of viral mRNA or viral vectors to stimulate an immune response.

Most vaccines are injected into a person's body, although some, such as polio vaccine, can be taken orally. Antigens produced as a result of vaccination stimulate the immune response and the formation of sufficient memory cells to produce antibodies very quickly if the person is infected with the real pathogen later on.

A first vaccination produces a primary response but many vaccinations are followed up with another some time later. The second or 'booster' dose of vaccine causes a greater and faster

production of antibodies and memory cells, known as a secondary response (Figure 10.1.13), and provides long-term protection. The length of time that antibodies and memory cells persist depends on the disease. Rubella vaccination can provide protection for up to 20 years, while vaccinations for tetanus should be repeated every 10 years. Vaccines do not prevent infection by pathogens but they do enable the body to respond quickly to them and prevent serious illness.

KEY POINT

secondary response refers to response to the second exposure to an antigen in a vaccine (or to the same pathogen) that causes a more rapid production of antibodies due to the presence of memory cells.

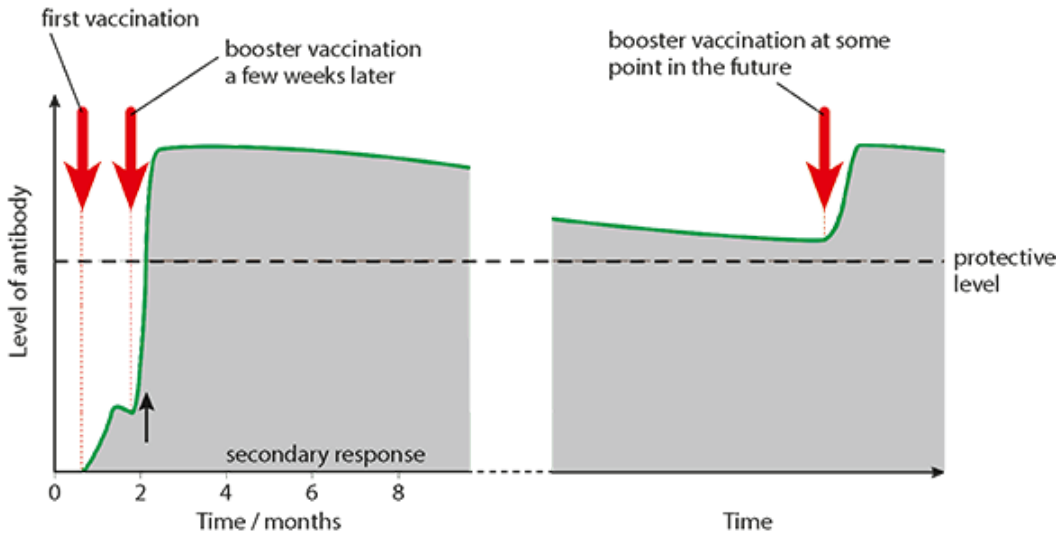


Figure 10.1.13: Antibody levels after vaccination. The persistence of antibodies varies and depends on the vaccine used.

SCIENCE IN CONTEXT

Vaccines against COVID-19

mRNA vaccines are a new type of vaccine to protect against infectious diseases and they have been developed rapidly to vaccinate people against COVID-19. An RNA vaccine consists of an mRNA strand that codes for a disease-specific antigen. Once the mRNA strand in the vaccine is inside the body's cells, the cells use the genetic information to produce the antigen. This antigen is then displayed on the cell surface, where it is recognised by the immune system, which responds by producing antibodies.

Viral vector vaccines use a modified version of a different virus (not the pathogen) to enter our cells and deliver instructions them. The vector virus stimulates them to produce a spike protein from surface of the COVID-19 virus. This protein triggers the production of antibodies and leads to immunity.

NATURE OF SCIENCE

Assessing ethics in science: the case of smallpox

Smallpox was a serious disease caused by the variola virus that killed thousands of people every year. It was transmitted by droplet infection. Edward Jenner (1749–1823) was a British scientist who developed a vaccine to protect people against the disease. He knew that dairymaids who had suffered from a similar but mild disease, called cowpox, were protected against smallpox. In 1796, Jenner isolated pus from the lesions of a dairymaid with cowpox, and applied it to a cut in the skin of an 8 year old boy, James Phipps. The boy caught cowpox and recovered. Next, Jenner inoculated the boy with smallpox viruses and discovered that James had developed immunity to the disease after his exposure to cowpox. Jenner

also tested his ideas on himself and his family. He named the procedure 'vaccination' from the Latin word *vacca* (cow).

Jenner was criticised and ridiculed for his ideas but Louis Pasteur (1822–1895) later supported him and went on to investigate the use of vaccines for other diseases.

In the 20th century, smallpox became the first infectious disease of humans to have been eradicated by vaccination. A campaign to eradicate it was started by the World Health Organization in 1977. Ten years later there were no new cases of the disease in the world. The WHO is now working on a programme to eradicate polio worldwide.

To consider:

1 Why do you think Jenner was criticised for his ideas and experiments?

2 What obstacles today are likely to delay a world vaccination programme like the one that was carried out on smallpox?

SCIENCE IN CONTEXT

Producing and testing new vaccines

Experiments like the ones that Jenner carried out would not be possible now. Today national and international laws require that all new medicines and vaccines are tested on animals before they can be used for humans. Vaccines cannot be tested on humans until 'preclinical trials' have been carried out. Preclinical trials are conducted to ensure the vaccine is safe and effective. These are first carried out using cells and tissue cultures in the laboratory (*in vitro*), and then the vaccine is

studied in more detail using animals, often mice, to confirm that they are effective in living organisms. Mice are important model organisms because, on average, the protein-coding regions of mouse and human genomes are 85% identical and some mouse genes are 99% similar to those in the human genome. But there are many uncertainties involved in using animals to test vaccines. There is considerable variety in the immune responses between different species, even those that are closely related. Also many viruses are species specific, which means that using non-human species for testing may not produce the same results.

Preclinical trials are used to understand the toxicity of the vaccine, which means how well it does its job without causing harm to other tissues. Next, testing in animals helps researchers understand the type of immune response that will be generated by a vaccine, and if it is likely be enough to protect a person from developing the disease.

Following this, the vaccine moves on to its first human trials, known as Phase I clinical trials. Small numbers of healthy, human volunteers are vaccinated to see if the vaccine acts in the same way in humans and to work out the most effective dose. Next, Phase II trials check that the vaccine works consistently in a much larger group of adults and researchers check for any side effects. Later a Phase III trial is carried out on a much larger number of people, usually several thousands, so that statistical data can be collected to check that the vaccine is reliable and produces a level of immunity that would prevent disease, and reduce the number of cases of disease. Only after this can an application be made to license a new vaccine.

Herd immunity and preventing epidemics

Herd immunity occurs when a large portion of a population becomes immune to a disease, either because they have had the disease or been vaccinated against it. The spread of disease from person to person becomes less likely when herd immunity is achieved because the disease-causing pathogen has few people to infect. As a result, the whole population, not just those who are immune, becomes protected to some extent as transmission of infection is greatly reduced.

When a high percentage of the population is vaccinated, it is difficult for infectious diseases to spread, because there are not many people who can be infected. For example, if someone with measles is part of a community of people who are vaccinated against measles, the disease cannot easily be passed on to anyone, and it will quickly disappear. This type of herd immunity (also called community immunity) gives protection to vulnerable members of the community such as newborn babies, elderly people and those who are too sick to be vaccinated for example people undergoing chemotherapy treatment or others with weakened immune systems.

Herd immunity does not protect against all diseases that can be prevented by vaccines, only diseases that are transmitted from person to person. One example of this is tetanus, which is caught from bacteria in the environment, not from other people who have the disease. Even if everyone around you is vaccinated against tetanus you will not be protected unless you are vaccinated too.

Herd immunity only works if a sufficient percentage of people in a population are vaccinated; for example, 19 out of every 20 people need to be vaccinated against measles to protect people

who are not vaccinated. Unlike vaccination, herd immunity does not give a high level of individual protection, and so it is not a good alternative to having a vaccination.

INTERNATIONAL MINDEDNESS

The World Health Organization (WHO)

Vaccination gives protection against many bacterial and viral diseases. Bacterial infections can be treated with antibiotics but virus diseases cannot. This means that vaccination against viral infection is very important. The WHO began a worldwide programme to vaccinate people against smallpox in 1967. They monitored outbreaks of the disease and contained them with vaccinations in many countries. Smallpox only affects humans and other animals do not carry it. With high rates of vaccination it is difficult for a disease to spread as there are few people for the virus to infect. The last known case of smallpox was in Somalia in 1977 and the world was declared free of smallpox.

The WHO is now working on a new challenge to eradicate polio, a disease that can cause paralysis and seriously affects children. In 2020, the WHO was able to certify its African region as free of polio after 4 years without a case. This achievement meant that over 90% of the world's population are now free of the wild poliovirus, and the WHO moved closer to its target of eradicating polio everywhere. Only two countries, Pakistan and Afghanistan, still have poliovirus transmission.

The Director General of the WHO at that time, Dr Tedros Adhanom Ghebreyesus, said, "Ending wild polio virus in Africa is one of the greatest public health achievements of our

time and provides powerful inspiration for all of us to finish the job of eradicating polio globally."

He attributed the success of the vaccination programme to strong leadership and innovation in the region, along with successful co-ordination of effort between countries in Africa. Major challenges in immunising the region's children included high levels of population movement, conflict and insecurity that restricted access to health services, as well as the ability of the virus to spread quickly and cross national borders.

EXTENSION

Model organisms

A model organism is a species used by researchers to study specific processes in biology, especially in investigations in genetics, developmental biology and human disease. Organisms chosen as models are non-human species that can show us how the biological processes of humans or other organisms work.

Ideal characteristics of model organisms include having a short life span, producing a large number of offspring, being easy to observe and having a genome which has been sequenced and is understood. A relatively small number of species are used as models and different organisms are used in different areas of research. Some examples are included in this book. They include the yeast (*Saccharomyces cerevisiae*) used in the study of the cell cycle, the fruit fly (*Drosophila melanogaster*) used to study genetics because it has only four pairs of chromosomes, breeds readily and matures very quickly, and the mouse (*Mus musculus*) used in studies of the

immune system. Mice share many human diseases because they have lived near people for thousands of years.

When studying disease, or testing vaccines, mice are usually chosen because of their similarity to humans in their genetics, anatomy and physiology. But, although all vertebrates share a common ancestor, and have similar developmental and metabolic pathways, information from studies of model organisms must be considered carefully, as there may be important differences as well as similarities.

TEST YOUR UNDERSTANDING

5 Define the term zoonosis and give an example.

6 Why can a person who has blood group O donate blood to any other person?

7 Summarise the importance of memory cells in the immune response.

Links

- How are pathogens classified? (Chapter 11)
- How have some bacteria become resistant to many antibiotics? (Chapter 11)
- How do animals protect themselves from other threats such as predators? (Chapter 12)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
define a pathogen and give examples of organisms that are pathogenic	10.1.1			
summarise the properties of parts of the body that make up the first line of defence	10.1.1			
describe the process of blood clotting	10.1.1			
outline the process of phagocytosis	10.1.1			
describe the humoral immune response including role of lymphocytes and production of antibodies	10.1.1			

summarise the effects of HIV on the body's immune system	10.1.3			
outline how antibiotics kill bacteria and why they are ineffective against viruses	10.1.4			
describe how bacteria can become resistant to antibiotics	10.1.4			
define a zoonosis and give some examples of diseases and their transmission	10.1.5			
outline the differences in antibodies on red blood cells and the categorisation of ABO blood groups	10.1.1			
describe the cell-mediated response and the humoral response	10.1.2			

summarise the roles of cytotoxic T-cells, B-cells and T-cells in the immune response	10.1.2			
describe the importance of memory cells	10.1.2			
outline the action of vaccines and recall that smallpox was eradiated by vaccination	10.1.6			
describe what is meant by herd immunity and how it can prevent epidemics.	10.1.6			

REFLECTION

Reflect on how your knowledge of this subject is important to real-world situations and problems.

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Unit 4
Organisation in ecosystems

INTRODUCTION

In any ecosystem living organisms interact and depend on one another in many different ways. All living things need a place to live, a source of food and resources for reproduction. Some species become food for others, some provide a home for others and some co-operate with members of their own species or different species. If an ecosystem is undisturbed it will reach a balance or equilibrium where species that live there coexist in a sustainable way.

To study ecosystems we classify organisms into different groups and observe their adaptations and interactions. We can measure the ways in which energy is transferred from one organism to another as they feed and how the ecological relationships we see in food chains and webs make different systems resilient to change. Over thousands of years, evolution has occurred as different organisms have become adapted to live in different conditions. In a new or changing environment the best adapted will survive to reproduce and occupy new niches.

Chapter 11

Evolution, speciation and ecosystems

A3.1, A3.2, A4.1, B4.1, B4.2, D4.1

INTRODUCTION

Over long periods of time and many generations, species change as they become adapted to new surroundings or altered conditions. One result of these changes may be the evolution of new varieties and species. There is strong evidence for the evolution of life on Earth, both from the fossil record and from organisms that are alive today. Natural biological classification attempts to arrange living organisms into groups that enable

them to be identified and studied easily. In this way we can monitor diversity and show evolutionary links between different groups.

11.1 Classification

LEARNING OBJECTIVES

In this section you will:

- Learn that organisms have binomial names consisting of their genus and species

- understand that natural classification helps identify and predict characteristics of species in a group
- learn that taxonomists classify organisms using a hierarchy of eight taxa: domain, kingdom, phylum, class, order, family, genus and species but this does not always reflect divergence caused by evolution
- define a clade as a group that contains all the organisms that have evolved from a common ancestor
- learn that organisms are now classified into three domains based on rRNA evidence
- draw clades as branches on phylogenetic diagrams
- understand that modern cladistics uses genetic sequencing and amino acids to group organisms whereas traditional evidence was drawn from appearance
- learn that sequence differences accumulate gradually and that there is a positive correlation between the

number of differences between two species and their divergence from a common ancestor.

GUIDING QUESTIONS

- Why does classification help scientists organise the diverse life on Earth?
- How can similarities and differences be used to identify or classify organisms?

11.1.1 The binomial system of classification

Natural classification (also called biological classification) attempts to arrange living organisms into groups so that we can identify them easily and show the evolutionary links between the groups. The system of classification we use today has its origins in a method devised by the Swedish scientist Carolus Linnaeus (1707–1778).

KEY POINT

natural classification is classification according to relationships based on descent from a common ancestor.

The classification of living organisms is simply a method of organising them into groups to show similarities and differences between them. More than 2000 years ago, the Greek philosopher Aristotle (384–322 BCE) classified organisms into two groups: plants and animals. This was useful as a starting point, but as the two main groups were subdivided, problems started to appear. At that time, organisms were seen to be unchanging, so there was no understanding of evolutionary relationships. Many organisms discovered later did not fit into the scheme very well.

Figure 11.1.1: Carolus Linnaeus, also known as Carl Linnaeus, was a Swedish botanist, physician and zoologist, who laid the foundations for the modern scheme of binomial nomenclature.

KEY POINTS

homology is the similarity due to shared ancestry between a pair of structures in different taxonomic groups.

taxon (plural taxa) is a group of one or more populations of an organism or organisms identified by taxonomists as forming a rank such as family or species.

taxonomy is the science of identifying, naming and grouping organisms.

Birds were separated into a group defined as ‘Feathered animals that can fly’ so no place could be found for the flightless cormorant, a bird that does not fly. Bacteria, which were unknown at the time, were not included at all.

In 1735, Carolus Linnaeus (Figure 11.1.1) adapted Aristotle's work, and his system forms the foundation of modern taxonomy. Taxonomy is the science of identifying, naming and grouping organisms.

Linnaeus gave each organism a Latin name, made up of two words. The first part of the name is a **genus** name and the second part is a **species** name. Thus the binomial, or two-part, name for the American grizzly bear is *Ursus americanus* whereas a polar bear is *Ursus maritimus.* Linnaeus used Latin for his names. Latin has long been the language of medicine and science, and it is unchanging. If *Ursus maritimus* is mentioned anywhere in the world, scientists will know that polar bears are being discussed.

- The genus part of the name indicates a group of species that are very closely related and share a common ancestor.
- The species is usually defined as a group of individuals that are capable of interbreeding to produce fertile offspring.

Linnaeus developed structure in his classification system. For example, he grouped birds into birds of prey, wading birds and perching birds.

Although it is possible to group living things in many different ways, over the last 200 years a hierarchical classification system has emerged through agreement at many scientific meetings. This system is now used by biologists everywhere. Modern taxonomists all over the world classify species using a hierarchy of groups called taxa (singular: taxon). There are eight levels to the hierarchy:

1 domain

2 kingdom

3 phylum (plural: phyla)

4 class

5 order

6 family

7 genus (plural: genera)

8 species.

Two examples of how species are classified are shown in Table 11.1.1.

	Polar bear	**Lemon tree**
Domain	Eukarya	Eukarya
Kingdom	Animalia	Plantae
Phylum	Chordata	Angiospermata
Class	Mammalia	Dicotyledoneae
Order	Carnivora	Geraniales
Family	Ursidae	Rutaceae
Genus	*Ursus*	*Citrus*
Species	*maritimus*	*limon*

Table 11.1.1: The taxonomic hierarchy for a plant species and for an animal species.

NATURE OF SCIENCE

By convention, the genus name starts with a capital letter, while the species does not. Both are written in italic or

underlined. Once an organism has been referred to by its full Latin name in a piece of text, further references abbreviate the genus to the first letter only – for example, *U. maritimus* (Table 11.1.1). Organisms from the same genus can be referred to with the genus name followed by sp. For example, one bear whose species we do not know could be *Ursus* sp. while several bears from the same genus would be written as *Ursus* spp.

Aristotle's original grouping of organisms into just two kingdoms has also been refined. Today the most widely accepted method of classification uses three Domains:

- Archaea – cells which do not contain a nucleus but have a different cell wall from bacteria
- Eubacteria – cells which do not have a nucleus
- Eukarya – organisms whose cells do contain a nucleus.

The Three Domain System was proposed in 1977. It is based on differences in the sequences of nucleotides in a cell's ribosomal RNAs (rRNA), as well as the organisms' membrane structure and sensitivity to antibiotics. Comparing rRNA structure is used because rRNA molecules in all living things have the same function and their structure has remained constant over time. This means that similarities and differences in rRNA nucleotide sequences are a good indicator of how closely related different organisms are.

It is generally thought that all cells came from a common ancestor cell termed the last universal common ancestor (LUCA). These LUCAs eventually evolved into three different cell types, each representing a domain. (You can read more about LUCA in Chapter 6).

- Archaea:
    1. Kingdom Archaebacteria (ancient bacteria) – methanogens, halophiles and thermoacidophiles
- Eubacteria:
    2. Kingdom Eubacteria (true bacteria) – bacteria and cyanobacteria
- Eukarya:
    3. Kingdom Plantae – plants
    4. Kingdom Animalia – animals
    5. Kingdom Fungi – fungi
    6. Kingdom Protista – red algae and dinoflagellates

The problems of classifying asexual organisms

Scientists have devised classification schemes to identify species in an evolutionary or phylogenetic way. This is very difficult for bacteria and other microorganisms which reproduce asexually. The classic definition of a species as a group of organisms that can interbreed and produce fertile offspring cannot be used.

KEY POINTS

homologous structures are anatomical features showing similarities in shape, though not necessarily in function, in different organisms.

phylogenetic classification is the classification system based on evolutionary relationships.

Bacteria are organised into five groups according to their basic shapes: spherical (cocci), rod (bacilli), spiral (spirilla), comma (vibrios) or corkscrew (spirochaetes)(Chapter 10). They do also have many biochemical differences in both their metabolism and cell structure, and this has been useful in classifying some groups.

Why is classification important?

Natural classifications group together organisms that share many of the same characteristics. This means that natural classifications are predictive, so that by studying the characteristics of an organism it is possible to predict the natural group it belongs to.

Phylogenetic classifications are natural classifications that attempt to identify the evolutionary history of species. The role of taxonomy is to group species that are related by common ancestry.

A natural classification such as the one devised by Linnaeus is based on identification of homologous structures that indicate a common evolutionary history. A homology is similarity in structures that are found in different taxonomic groups and which are due to shared (common) ancestry.

If homologies, such as the arrangement of bones in vertebrate limbs, described in Section 11.4 Niches and adaptation, are shared by organisms, then it is likely that those organisms are related. So the binomial system can be both a natural and a phylogenetic classification.

In summary, there are four main reasons why organisms need to be classified:

1 to impose order and organisation on our knowledge

2 to give each species a unique and universal name, because the local names that people use for the same organism vary from place to place around the world

3 to identify evolutionary relationships; if two organisms share particular characteristics then it is likely that they are related to each other, and the more characteristics they share then the closer the relationship

4 to predict characteristics; if members of a particular group share characteristics then it is likely that other, newly discovered members of that group will have at least some of those same characteristics.

Classifications change

New species are being found and new discoveries are being made about existing species all the time. Such discoveries may force us to rethink the way we classify living things. For example, in the past, the name 'bacteria' was given to all microscopic single-celled prokaryotes. But recent molecular studies have shown that prokaryotes can be divided into two separate domains, the Eubacteria and the Archaea, which evolved independently from a common ancestor. Molecular biology, the study of DNA, RNA and protein sequences and studies of cell ultrastructure have shown that the Archaea and Eukarya (eukaryotes) are in fact more closely related to one another than either is to the Eubacteria. Similar principles are applied to all levels of classification. Taxonomists reclassify organisms when new evidence shows that they have evolved from different ancestral species.

11.1.2 Using a dichotomous key

A dichotomous key is a series of steps, each involving a decision (Figure 11.1.2), which can be used to identify unknown organisms. The key prompts us to decide, through careful observation, whether or not organisms we are studying have particular visible features. The key allows us to distinguish between different organisms on this basis.

When a key is made to identify organisms such as those shown in Figure 11.1.3, each specimen is examined carefully, and a characteristic is chosen that is present in about half of the individuals and absent in the others.

KEY POINT

dichotomous key a key in which organisms are separated into pairs of smaller and smaller groups by observation of their characteristics.

For example, in this key for invertebrates, the presence of wings is the first distinguishing characteristic, which divides the specimens into two smaller groups.

Now, for each group, another diagnostic feature is chosen to divide the specimens into two further groups, and so on. The branching tree diagram shown in Figure 11.1.3 progressively divides the specimens into smaller and smaller groups, until at the end of each branch a single individual is identified.

1	Wings present	go to 2
	No wings	go to 5

2	Two pairs of wings	go to 3
	One pair of wings	**fly**
3	Legs all approximately the same length	go to 4
	Hind pair of legs much longer than front two pairs	**locust**
4	Wings covered in scales	**butterfly**
	Wings transparent, not covered in scales	**dragonfly**
5	Four pairs of legs	**spider**
	More than four pairs of legs	go to 6
6	Pair of claws present	**crab**
	No claws	go to 7
7	Body clearly divided into equal-sized segments	**centipede**
	Body in two regions, segments only clear on hind region	**prawn**

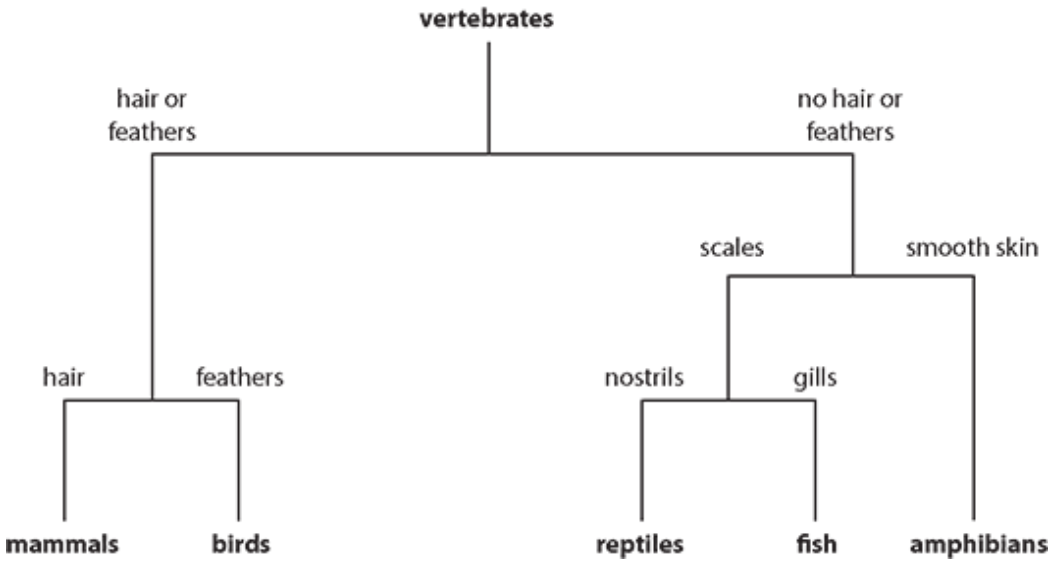


Figure 11.1.2: A simple key like this can be used to identify the vertebrate groups.

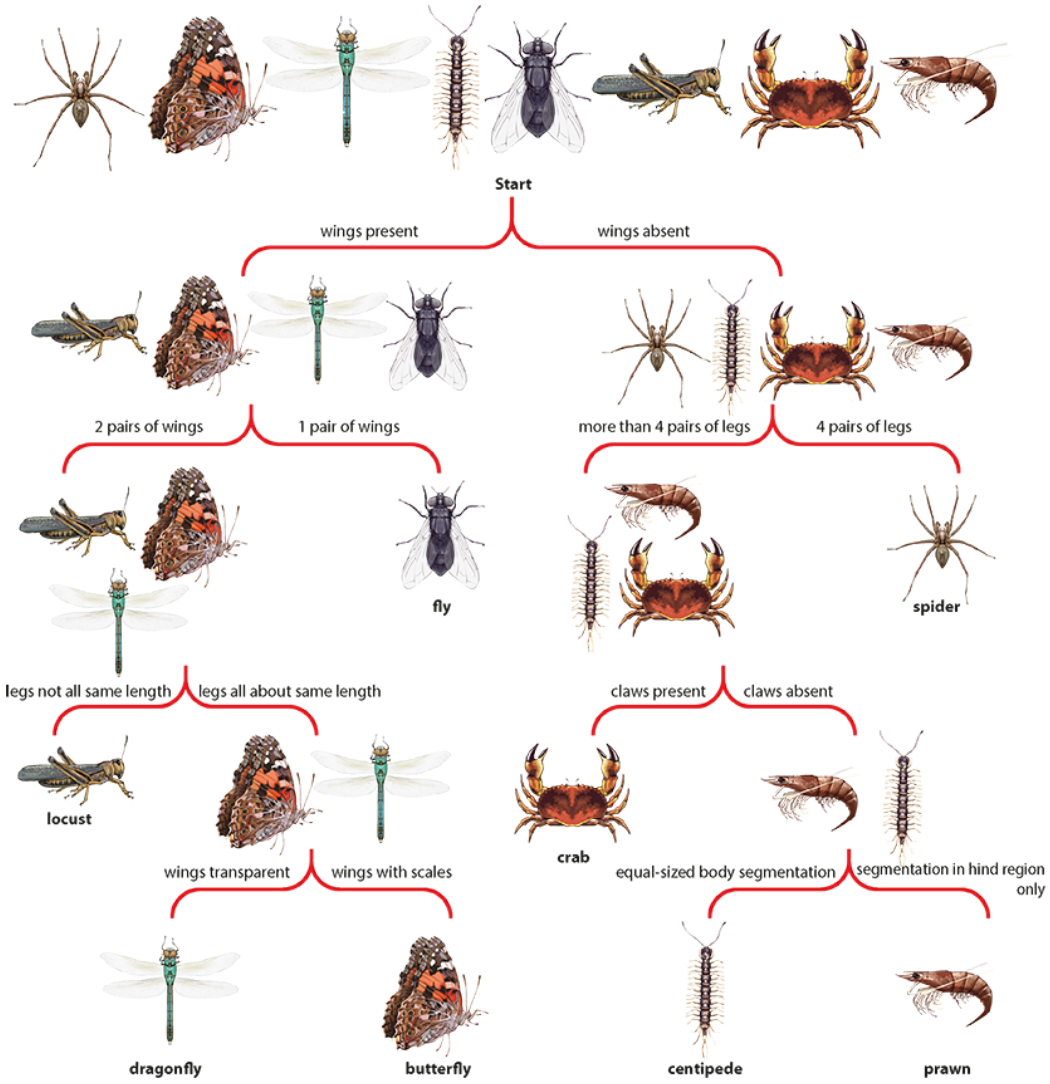


Figure 11.1.3: A dichotomous tree diagram distinguishing eight organisms.

The tree diagram can be ‘translated’ into a written key so that the branch points in the diagram are instead expressed as alternative statements. Each alternative either names the identified specimen or leads to another pair of statements, until an identification is reached. A well written key is composed of a

series of questions or steps, such that an organism that is being studied can only be placed in one of two groups.

NATURE OF SCIENCE

Cooperation and collaboration: an international naming system

Local names for different species can cause confusion. People in different parts of the world or even different parts of the same country use different common names for the same species. What do you think is being described here: armadillo bug, cafner, wood bug, butchy boy, gamersow, chiggley pig, sow bug, chuggypig and pill bug?

All these terms are local names for the woodlouse *Porcellio scaber*, or its relative *Armadillidium vulgare*, and are used in different parts of Europe and North America.

Cooperation and collaboration between international scientists provided an agreed binomial name for the woodlouse so that wherever they are studied, information about them can be attributed to the correct species.

TEST YOUR UNDERSTANDING

1 List, in order, the levels in the hierarchy of taxa.

2 State the two names from the hierarchy of taxa that are used in the binomial system.

3 If you were making a dichotomous key to identify leaves, explain why the question ‘Is the leaf large?’ would not be useful.

11.1.3 Cladistics

The universality of DNA and protein structures

Despite the incredible complexity of life, the building components of living organisms are not only simple in structure but are also universal.

- All living organisms use DNA built from the same four bases to store their genetic information and most use the same triplet code during translation. The few exceptions include mitochondria, chloroplasts and a group of bacteria which all have slight variations in the triplet code they contain.
- Proteins are built up from amino acids and living organisms all make use of the same 20 amino acids. In most cases, if a gene from one organism is transferred into another, it will produce the same polypeptide (if the introns have been removed from it; see Chapter 3).

These facts indicate a common origin of life and provide evidence to support the view that all organisms have evolved from a common ancestor. Study of the genetic code and amino acids of an organism can provide evidence that links it to its close relatives. This enables us to build up diagrams called cladograms, which show how species are related to one another in clades.

Clades and cladistics

Cladistics is a method of classification that groups organisms together according to the characteristics that have evolved most

recently. Diagrams called cladograms divide groups into separate branches known as clades (Figure 11.1.4 and 11.1.6).

Each branch ends in a group that has characteristics the other group does not share. A clade contains the most recent common ancestor of the group and its descendants, so a clade contains all the organisms that have evolved from a common ancestor.

Figure 11.1.4 shows five organisms forming part of an evolutionary tree.

KEY POINTS

clade a group of organisms, both living and extinct, that includes an ancestor and all the descendants of that ancestor.

cladistics a method of classifying organisms using cladograms to analyse a range of their characteristics.

cladogram a diagram that shows species' evolutionary relationships to one another.

root the ancestral population from which all other species originate

a branching point from the ancestral population is a node

the end of each branch, labelled with name of the taxonomic group is called a terminal branch

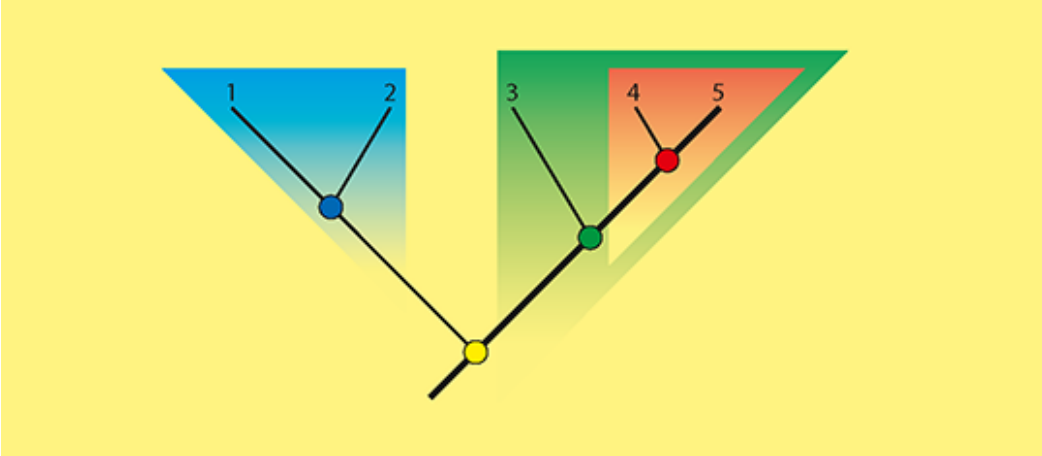


Figure 11.1.4: A cladogram with four clades.

- Organisms 1, 2, 3, 4 and 5 belong to the yellow clade.
- Organisms 1 and 2 belong to the blue clade.
- Organisms 3, 4 and 5 belong to the green clade.
- Organisms 4 and 5 belong to the red clade.
- The common ancestor for each clade is shown by the coloured spot at the branch point, or node.

Why do biologists need cladistics?

There are three important reasons for using cladistics to organise and discuss organisms.

1 Cladistics is useful for creating systems of classification so that biologists can communicate their ideas about species and the history of life.

2 Cladograms are used to predict the properties of organisms. A cladogram is a model that not only describes what has been observed but also predicts what might not yet have been observed.

3 Cladistics can help to explain and clarify mechanisms of evolution by looking at similarities between the DNA and proteins of different species.

11.1.4 Finding evidence for clades and constructing cladograms

Phylogenetics is the study of how closely related organisms are and it is used to establish clades and construct cladograms. Historically, the evidence used to construct a phylogeny was based on visible characteristics but molecular phylogenetics is the modern approach. Molecular phylogenetics examines the sequences of DNA bases or of amino acids in the polypeptides of different species to establish the evolutionary history of a group of organisms. Species that are the most genetically similar are likely to be more closely related. Genetic changes are brought about by mutation and, provided a mutation is not harmful, it will be retained within the genome. Differences in DNA accumulate over time at an approximately even rate so that the number of differences between genomes (or the polypeptides that they specify) can be used as an approximate evolutionary clock. This information can tell us how far back in time species split from their common ancestor. A greater number of differences in a polypeptide indicates that there has been more time for DNA mutations to accumulate than if the number is smaller. There is a positive correlation between the number of differences between two species and the time they evolved from a common ancestor.

KEY POINT

phylogeny a diagram showing relationships between different organisms that represents their evolutionary history.

Evidence from amino acids

We can expect that related organisms will have the same molecules carrying out particular functions and that these molecules will have similar structures. So by comparing proteins in different groups of organisms, and checking them for similarities in amino acid sequences, it is possible to trace their ancestry. Chlorophyll, hemoglobin, insulin and cytochrome *c*, which are found in many different species, have all been studied in this way. Cytochrome *c* is found in the electron transport chain in mitochondria, where it plays a vital role in cell respiration. Its primary structure contains between 100 and 111 amino acids and the sequence has been determined for a great many plants and animals.

Table 11.1.2 shows the number of amino acid differences in cytochrome *c* between humans and four other species.

From these data, it is possible to construct a cladogram showing the relationships between these five organisms, as shown in Figure 11.1.5. There are no differences between rabbit and mouse cytochrome *c*, so they have to be drawn together at the end of a branch, and the same applies to the chimpanzee and human. Rhesus monkey differs from chimpanzee and human by only one amino acid and so the branch point must be one unit from the end. Rabbit and mouse differ by nine amino acids and so the branch point must be nine units further down. Biochemical analysis of other molecules or comparison of DNA sequences would be needed to complete the separation of rabbit from mouse and human from chimpanzee.

Organism	**Number of amino acid differences in cytochrome *c* compared with human**
human	–
chimpanzee	0

rhesus monkey	1
rabbit	9
mouse	9

Table 11.1.2: Table comparing cytochrome *c* in five species.

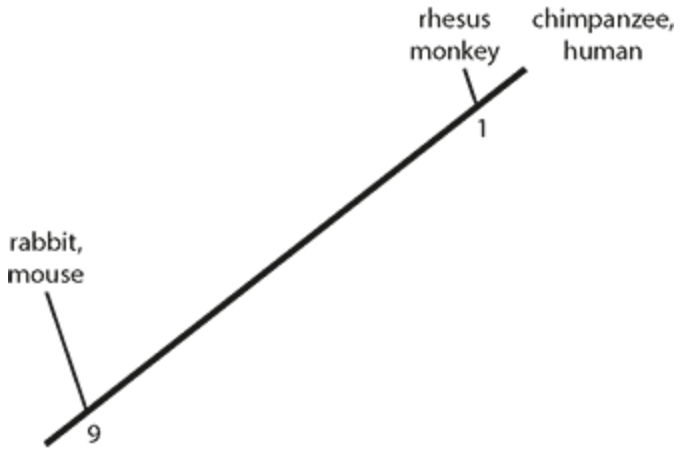


Figure 11.1.5: A cladogram for five mammal species based on the amino acid differences in cytochrome *c*.

There is only one difference between the human cytochrome *c* and that of a rhesus monkey but rabbits and mice have nine differences when compared with humans, which indicates they are less closely related. This biochemical evidence supports the classification of the animals that has been made from morphological observations.

Sequence differences accumulate gradually so there is a positive correlation between the number of differences between two species and the time since they diverged from a common ancestor.

11.1.5 The shapes of cladograms

Cladograms can be drawn in one of two ways, as shown in Figure 11.1.6, which shows two formats for a cladogram of living vertebrate animals. By looking at the left-hand diagram we can see that the organism with the greatest number of differences from mammals branches off first. These organisms are the least closely related to mammals.

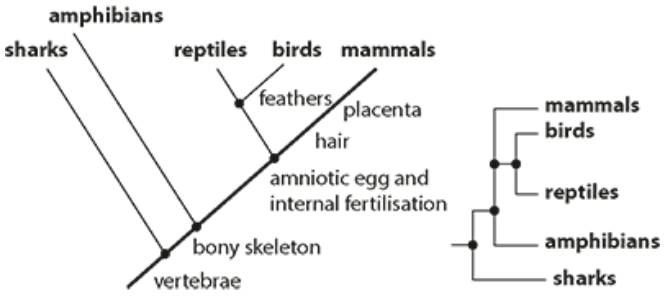


Figure 11.1.6: A cladogram shown in two different formats.

The relationship between reptiles and birds has been the subject of much debate amongst scientists. Some reptiles (e.g. crocodiles and dinosaurs) are more closely related to birds than other reptiles (e.g. lizards and turtles). Cladists tell us that 'reptiles' is not a clade and that a better grouping would be Archosauria (crocodiles, dinosaurs and birds), Lepidosauria (snakes and lizards) and Testudines (turtles and tortoises).

NATURE OF SCIENCE

How does our scientific understanding develop over time?

Our understanding of terms such as birds, reptiles and dinosaurs has changed as more information has become available. Today we can say that birds are not only the descendants of dinosaurs, biologically they are dinosaurs. Molecular evidence tells us that during the Triassic period (251–199 million years ago) the major groups of what we now call the reptiles evolved. These animals were the ancestors of both crocodiles and dinosaurs.

About 65 million years ago a massive extinction event wiped out all but one group of small, feathered dinosaurs. These feathered dinosaurs developed over time to become what today we call birds. Despite their shared evolutionary history and close relationship to other reptiles, birds are not normally grouped with reptiles. So why is this?

The Linnaean system of classification divides animal into groups based on their physical similarities. In this system, reptiles are organisms that are ectothermic (do not produce their own body heat) and have scales, so birds do not fit into this group. But using phylogeny and grouping the same organisms based on genetic similarities we discover that birds (and all dinosaurs), lizards, turtles, snakes and crocodilians were all descended from the original reptile ancestor. So why are birds still separated from the group we call reptiles?

The answer is probably for convenience. When a scientist studies features such as flight or behaviour which are unique to birds, it is easier to separate birds from the 'non-avian reptiles' and to consider them as a separate group.

We have two systems of classification for reasons of history but both are useful. The Linnean system is more useful for understanding how organisms live while the phylogenetic system and cladograms are more useful for understanding relationships between organisms.

To consider:

1 How has phylogeny helped in our understanding of evolution?

2 What new problems arise from new knowledge?

WORKED EXAMPLE 11.1

Which apes are the closest living relatives of *Homo sapiens*?

Answer

Gibbons, orangutans, gorillas and chimpanzees have many physical similarities to humans. For example, humans, chimpanzees and gorillas all have a cavity in the skull, just above the eyes, known as a frontal sinus. Gibbons, orangutans and other primates do not have this, so the physical evidence suggests that chimpanzees and gorillas are more closely related to humans than gibbons and orangutans. Evidence from the analysis of blood proteins also suggests that orangutans are more closely related to humans than gibbons. This evidence can be shown as in Figure 11.1.7.

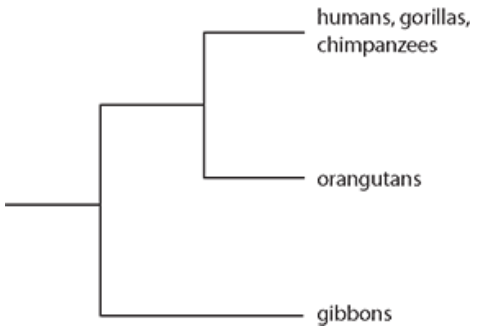


Figure 11.1.7: A cladogram showing the relationship between five apes.

Chimpanzees and gorillas are more closely related to humans than other living animals are but which are our closest living relatives?

To sort out the relationships between human beings, chimpanzees and gorillas, we must assess the evidence and check which features are shared. We can construct a table to summarise the evidence (Table 5.3).

Consider which of the three cladograms shown in Figure 11.1.8 is supported by most evidence.

Characteristic	Other primates such as baboons	Gorillas	Humans	Chimpanzees	Cladogram supported by evidence (Figure 11.1.8)
DNA evidence:					
number of chromosomes	42 or more	48	46	48	C
structure of chromosomes 5 and 12		different from other primates	like other primates	different from other primates	C
chromosome Y and 13		same as human beings	same as gorillas	like other primates	A
% genetic	orangutan	1.6%	–	1.2%	B

difference from humans	3.1%rhesus monkey 7%				
Molecular evidence:					
α-hemoglobin compared with human	several differences	one amino acid difference	–	identical to humans	B
protein factor in blood	not variable	not variable	same variability as chimpanzees	same variability as humans	B

Table 11.1.3: Summary of molecular and DNA evidence for the relatedness of primates.

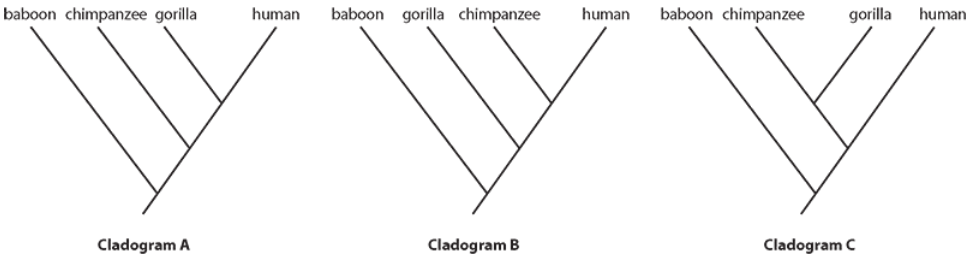


Figure 11.1.8: Three possible cladograms to show the relationship between humans, chimpanzees and gorillas.

None of the cladograms can be proved to be correct from this evidence but cladogram B is the best supported based on the data. It is therefore hypothesised to best reflect our current understanding of the evolutionary relationships of human beings. If further evidence is collected in future, the hypothesis may be changed.

NATURE OF SCIENCE

Re-classifying the figwort family

Using DNA sequences has shown that classifications of some groups which were based on visible characteristics do not correspond with the evolutionary origins of the group or species. One example of this is the figwort. Figwort is a family of flowering plants that includes black root, Culver's root and the mulleins (*Verbascum* spp.). In the past the family was

named Scrophulariaceae and contained about 275 genera and 5000 different species. Since the early 1990s, research into the DNA sequences of three chloroplast genes has revealed at least five separate clades. The traditional group Scrophulariaceae was found to be an unnatural group, which was not made up of clades. There were few, clear distinguishing physical characteristics to separate members of the old group which meant that taxonomists were unable to identify clades until molecular evidence became available. The new classification now places some genera into completely different families as the molecular studies have shown that they are not closely related to the figwort family after all.

To consider:

- How important is molecular evidence in modern classification?
- Will molecular evidence supersede observations of physical characteristics in the future?

TEST YOUR UNDERSTANDING

4 Define the term ‘clade’.

5 Suggest why DNA is so useful in establishing evolutionary relationships.

6 Explain how evidence from cladistics can lead to new classifications.

REFLECTION

Can I use my new knowledge to explain to others in my class why classification is important in understanding of the living world?

Links

- Why is it better to use genetic sequences and not anatomical features when classifying organisms? (Chapter 4)
- How does classification help our understanding of evolution? (Chapter 11.3)

11.2 Selection

LEARNING OBJECTIVES

In this section you will:

- learn that natural selection is the mechanism driving evolution
- learn that variation between individuals is required for natural selection to occur
- discover that variation can be caused by mutation, meiosis and sexual reproduction
- understand that adaptations are features that make an individual better suited to its environment and way of life
- recognise that more offspring tend to be produced than the environment can support, which means there is a struggle for survival between individuals
- discover that better-adapted individuals tend to survive and produce more offspring while less well adapted individuals tend to die or produce fewer offspring
- recall that when individuals reproduce, they pass on their heritable characteristics to their offspring
- understand that natural selection increases the frequency of well adapted individuals in a population and thus leads to changes in the species

- recognise that human activity can provide evidence for natural selection if populations are exposed to pesticides, pollution or antibiotics
- recognise that selective breeding of domesticated species shows that artificial selection can cause evolution
- recognise that sexual selection is natural selection for mating success

- discover that there are three modes of selection: directional, disruptive and stabilising
- learn that gene and allele frequencies change over time leading to evolutionary change

GUIDING QUESTIONS

- How can a species adapt to changing environmental conditions?
- What is the mechanism for natural selection?
- What is the importance of variation in selection and evolution?

11.2.1 A mechanism for evolution

The theory of evolution by means of natural selection was proposed by Charles Darwin (1809–1882) and Alfred Wallace (1823–1913). Darwin explained his ideas in a book called *On the Origin of Species by Means of Natural Selection,* published in 1859. The explanation remains a theory because it can never be completely proved but there is an abundance of evidence to support the key ideas, which are based on the following observations and deductions. Some terms we use now were not used by Darwin, who had no knowledge of genes or alleles. However, the fundamental basis of his argument was the same as outlined here.

1 Populations are generally stable despite large numbers of offspring

Organisms are potentially capable of producing large numbers of offspring, far more than the environment can support. Trees can produce thousands of seeds and fish produce hundreds of eggs. Yet few of these survive to maturity and we rarely see population explosions in an ecosystem.

2 Better-adapted individuals have a competitive advantage

Both plants and animals in a growing population will compete for resources. A resource may be food, territory or even the opportunity to find a mate. In addition, predators and disease will remove members of a population. This competition will bring about a struggle for survival between the members of a population. Organisms that have adaptations that make them well suited to the conditions will be good at competing. They will tend to survive long enough to reproduce, passing on heritable traits to their offspring. Others will die.

3 There is heritable variation within species

Different members of the same species are all slightly different and this variation is due to the mechanism of sexual reproduction. The process of meiosis (Section 6.5) produces haploid gametes. Furthermore, the genes in the gametes that an individual produces may be present in different forms or alleles. When an egg is fertilised, the zygote contains a unique combination of genetic material from its two parents. Sexual reproduction gives an enormous source of genetic diversity, which gives rise to a wide variation among the individuals of a species.

4 Advantageous heritable traits become more frequent over generations

As a result of variation, some members of a population may be better suited (better adapted) to their surroundings than others. They may have keener eyesight, or have better camouflage to avoid predators. These individuals will out-compete others; they will survive better, live longer and pass on the genes for their advantageous characteristics to more offspring. Gradually, as the process is repeated generation after generation, the proportion of the advantageous genes in the population as a whole increases. This is called natural selection, and it occurs as the 'fittest' (best adapted) survive to reproduce.

KEY POINTS

adaptations characteristics that make an organism suited to its environment.

evolution is the cumulative change in the heritable characteristics of a population.

heritable trait is a characteristic in an offspring that resembles its parents' corresponding characteristic more than that in any other individual in the population.

natural selection is a mechanism for evolution in which various genetic types in a population make different contributions to the next generation.

Sexual reproduction promotes variation

Mutations in genes create new variations, but sexual reproduction also increases variation in a population by forming new combinations of alleles.

- During meiosis, crossing over at prophase I and random assortment in metaphase I produce genetically different gametes (Section 6.5).
- Different alleles are also brought together at fertilisation, promoting more variation.

In species that reproduce asexually, variation can arise only by mutation.

11.2.2 Natural selection and the evidence for evolution

Once species have evolved to become well adapted to conditions in a stable environment, natural selection tends to keep things much the same. However, if the environment changes, a population will only survive if some individuals have heritable traits that suit them to the new conditions, and these then become more frequent in the population, because of natural selection. Three examples of how this can happen in a relatively short period of time are the beak adaptations of Galápagos finches after a change in food availability, the response of a moth population to pollution, and the emergence of new strains of bacteria following the introduction of antibiotics.

How human actions affect selection

Human actions cause changes to the variation in some populations. These influences can provide evidence for the action of selection and for evolution. Several species of bacteria have become resistant to antibiotics, and pesticides and pollution have also led to evolutionary changes in populations of plants and animals that humans have attempted to control.

Figure 11.2.1: Light and melanic forms of peppered moths on light and dark tree bark.

Industrial melanism – the influence of pollution

The peppered moth (*Biston betularia*) is a night-flying moth that rests during the day on the bark of trees, particularly on branches that are covered with grey–green lichen. It is a light speckled grey, and relies on camouflage against the tree branches to protect it from predatory birds.

In Britain in the mid-19th century, a black form of the moth was noticed (Figure 11.2.1). The appearance of this new colour coincided with the period of the Industrial Revolution when many factories were built and contributed to growing pollution in the atmosphere. This pollution killed the lichens that grow on the bark of trees, and the tree bark became blackened with particles of soot.

The colour of the moth is due to a single gene, which can be present in two forms. The common, recessive form gives rise to a

light speckled colour. The much less common dominant form gives rise to the black, melanic moth.

In the polluted areas of Britain, the speckled form was no longer camouflaged on the blackened tree bark, and was easily seen by birds that ate speckled moths. The black moths were better suited to the changed environment because they were camouflaged. Black moths survived and bred, and the proportion of black moths with the dominant allele grew in the population.

In 1956, the Clean Air Act became law in Britain and restricted air pollution. Lichen grew back on trees and their bark became lighter. As a consequence, the speckled form of the peppered moth has increased in numbers again in many areas, and the black form has become less frequent.

Antibiotic resistance

Antibiotics are drugs that kill or inhibit bacterial growth. Usually, treating a bacterial infection with an antibiotic kills every invading cell. But, because of variation within the population, there may be a few bacterial cells that can resist the antibiotic. These individuals will survive and reproduce. Because they reproduce asexually, all offspring of a resistant bacterium are also resistant, and will survive in the presence of the antibiotic. In these conditions, the resistant bacteria have enormous selective advantage over the normal susceptible strain, and quickly out-compete them.

Treating a disease caused by resistant strains of bacteria becomes very difficult. Doctors may have to prescribe stronger doses of antibiotic or try different antibiotics to kill the resistant bacteria.

The problem of antibiotic resistance is made more complex because bacteria frequently contain additional genetic

information in the form of plasmids, which they can transfer or exchange with other bacteria, even those from different species. Genes for enzymes that can inactivate antibiotics are often found on plasmids, so potentially dangerous bacteria can become resistant to antibiotics by receiving a plasmid from a relatively harmless species. Many bacteria are now resistant to several antibiotics (Figure 11.2.2), so pharmaceutical companies are constantly trying to develop new antibiotics to treat infections caused by these multiply resistant forms of bacteria.

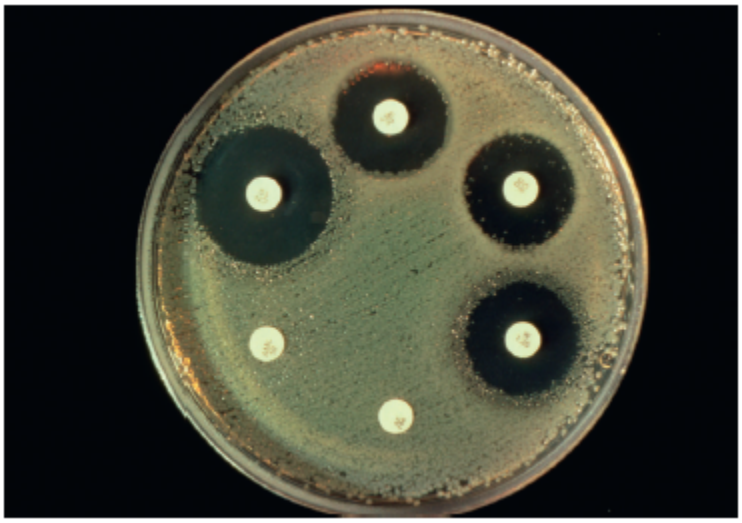


Figure 11.2.2: The grey–green areas on the agar jelly in this Petri dish are colonies of the bacterium *Escherichia coli*. The white card discs are impregnated with different antibiotics. This strain of *E. coli* is resistant to the antibiotics at the bottom left and has been able to grow right up to the discs. The other discs have a 'zone of inhibition' around them because those antibiotics kill the *E. coli* cells.

Resistance to pesticides

Pests including rodents, weeds and insects can develop resistance to pesticides that humans have developed to control them. Pest

species develop resistance as a result of natural selection as the individuals that are most resistant to a pesticide survive and pass on their resistant genes to their offspring. Many pest species produce large numbers of offspring and this increases the probability of mutations and the rapid spread of resistant populations. Pesticide resistance is a serious problem for agriculture all over the world. Pests consume between 10 and 20% of all global crops either while they are growing or when they are stored.

Some examples of species that have become resistant to pesticides include rats which have developed resistance to the rodenticide warfarin. In Europe some resistant rat populations can still survive after eating five times more of this pesticide than normal rats. New pesticides now have to be used to kill them.

Plant pest species include many of the weeds that grow in maize, cotton and soybean farms. In the USA several species of these weeds are now resistant to the herbicide glyphosate. Resistant weeds can often survive application of herbicide at rates that are much greater than the recommended rate. While in Australia over 25 weed species now have populations that are resistant to at least one of the herbicides that are used there.

Insecticide resistance has also occurred as pesticides have been applied over large areas and resistant genes have spread rapidly through insect populations. One example of resistant insect species is the Colorado potato beetle which has developed populations resistant to all the commercially produced insecticides that have been used against them.

11.2.3 Artificial selection

People have altered certain domesticated species by breeding selected individuals in a process called artificial selection. Plants or animals with favourable characteristics are chosen and bred together, to increase the number of offspring in the next generation that have the favourable characteristics. Individuals that do not have the desired features are not allowed to breed. People have domesticated and bred plants and animals in this way for thousands of years and, over many generations, this has resulted in the evolution of numerous breeds and varieties, which differ from each other and from the original wild ancestors.

Modern varieties of wheat, barley, rice and potatoes produce higher yields and are more resistant to pests and disease than ever before. Wheat and rice plants are shorter and stronger than varieties from 100 years ago, so that they are less likely to be damaged by wind and rain and are easier to harvest. The plants of 100 years ago were also very different from the original grasses that wheat was bred from 10 000 years ago. Many plants are bred for their appearance, and ornamental varieties have different petal shapes and colours from the original parent stock.

Animals are chosen and bred by farmers and animal breeders for special characteristics such as high milk yield in a cow, or good-quality wool in a sheep. Individuals with these characteristics are selected to breed, so that more of the next generation have these useful features than if the parents had not been artificially selected (Figure 11.2.3).

Although the driving force for artificial selection is human intervention, which is quite different from natural evolution,

selective or artificial breeding does show that species can change over generations.

But selective breeding can remove some genes from populations and lead to a limited gene pool (Sections 11.2.4 and 11.3). A population of selected plants or animals may have reduced genetic diversity and may become less healthy as a result.

We can see how this can happen in the breeding of certain animals where there have been negative effects on the health of animals. Pedigree dogs, like animals used in farming have particular characteristics such as height or type of fur that people select when breeding them. But some dog breeds such as bulldogs and labradors have a higher incidence of conditions such as heart disease or abnormal hip joints. Artificial selection has led to some genes being lost from the dogs' gene pool resulting in less healthy animals.

KEY POINTS

artificial selection selective breeding of domesticated organisms by humans.

gene pool all the genes and their alleles present in an interbreeding population.

Figure 11.2.3: Selective breeding of cows over many centuries has produced many breeds including the Guernsey, bred for the production of large quantities of fat-rich milk. Other breeds have been produced with flat backs to make giving birth easier, or thick fur to survive in colder regions.

11.2.4 Gene pools

Processes that increase or decrease variation in a population alter the proportions of alleles in a gene pool, which is all the genes and their alleles present in a population.

Natural populations can gain alleles when then are introduced from other gene pools, for example when new individuals immigrate, or move into an area. In a similar way, alleles can be lost from a gene pool if individuals leave, or emigrate, or if certain alleles are artificially bred out of a population. The examples of domestic farm animals and pedigree dogs show us that selective breeding for chosen characteristics runs the risk of losing some genes from the gene pool altogether and once they have been lost from a population, the loss cannot be reversed unless organisms from other populations are reintroduced. This can happen in natural populations as well as artificially bred plants and animals and lead to a problem known as inbreeding depression which reduces a population due to a lack of healthy mates and low genetic diversity.

KEY POINT

inbreeding depression reduction in the fitness of a population due to a lack of healthy mates and low genetic diversity.

One example of this was seen in Sweden where populations of adders became isolated from others due to farms that separated their territories. The snakes could only breed with members of their own small populations. As a result, the snakes not only gave birth to offspring with defects, but also suffered a decrease in breeding success. A loss of genes which leaves a limited gene

pool can result in populations with many individuals with similar genes as they only breed with others who also have them. In Sweden new adders were taken to breed with the isolated snakes and the populations recovered and flourished. But for endangered species this solution cannot work because there are not enough individuals to replenish the population.

You can read more about the importance of changes in gene pools and their effect on evolution in Section 11.3.

NATURE OF SCIENCE

Scientific theories and evidence

Natural selection is a theory. In science, the term 'theory' has a very specific meaning. Scientific theories require a hypothesis that can be tested by gathering evidence. If any piece of evidence does not fit in with the theory, a new hypothesis must be put forward and more scientific evidence gathered. It is important to recognise the difference between a theory and a dogma, which is a statement of beliefs that are not subject to scientific tests. Since Darwin's time, a large amount of evidence has been collected to support his theory. With technologies such as DNA profiling and carbon dating, further evidence continues to accumulate and must be tested against the theory.

To consider:

1. How much evidence is needed to support a theory?
2. What kind of evidence is needed to refute a theory?
3. Will it ever be possible to prove that evolution has taken place?

TEST YOUR UNDERSTANDING

7 Explain why sexual reproduction is important for evolution.

8 Individuals in a population are often said to be ‘struggling for survival’. Name the key factor that causes this struggle.

9 If an environment changes, individuals with particular combinations of genes are more likely to survive. State the name given to this phenomenon.

10 Define the term ‘speciation’.

11.2.5 Types of selection

Allele frequency and evolution

Evolution is defined as the cumulative change in the heritable characteristics of a population. The 'heritable characteristics' referred to in this definition are all the alleles in the gene pool of a population. So, if the frequencies of these alleles in the gene pool do not change, then the population is not evolving. Allele frequencies in a population are always fluctuating, in fact, because they depend on the reproductive success of individuals. But if just a single allele shows a change in frequency over a prolonged period of time, then we can say that the population has evolved. Any process that allows favourable alleles to be passed on or prevents the inheritance of unfavourable alleles can contribute to evolution.

Modes of natural selection and change

In any population, an individual can have any combination of the alleles present in the gene pool. This gives rise to variation in the population. In most cases, a population will be well adapted to its environment and so the same alleles will be selected, maintaining a stable population. This is known as **stabilising selection** (Figure 11.2.4).

If the environment, changes the population may also change. Some individuals may have alleles that are more favourable in the new conditions and these alleles will provide an advantage to those individuals, making them more likely to survive and reproduce successfully. This will result in a change in the population, known as **directional selection** and lead to the prevalence of new forms.

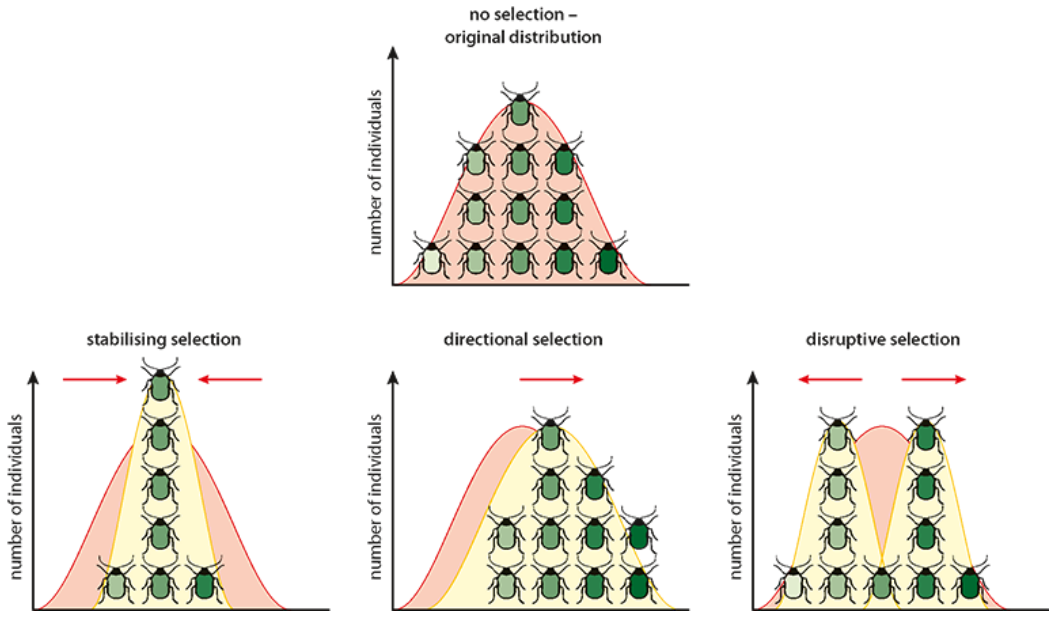


Figure 11.2.4: In these examples three kinds of selection change the colour of beetles – a feature that is controlled by several genes. Stabilising selection eliminates extreme forms, directional selection favours dark coloured beetles and disruptive selection eliminates the 'average' form and leads to two distinct forms.

A third possibility is that natural selection results in the formation of two new forms from a single existing population. This is known as **disruptive selection**.

Sexual selection

Sexual selection is a special example of natural selection. It affects an organism's ability to obtain a mate. As individuals select mates with characteristics that favour survival, the overall resilience of a population, that its ability to recover quickly from changes in the environment, will also be increased.

Many organisms have extreme features or behaviours to do this. Examples include the elaborate tails of male peacocks, battles between males over territories and growth of extremely large

horns or antlers. But sexual selection can sometimes produce features that are harmful to an individual. For example, brightly coloured feathers may attract predators as well as mates.

KEY POINTS

allele frequency the percentage of a population that carries an allele at a particular locus. It is also sometimes referred to as gene frequency.

resilience the ability of a population to recover quickly from changes in the environment.

Sexual selection can come about because a characteristic such as the antlers of a stag increases its abilities in competition with members of the same sex. Stags, rams and bulls use antlers or horns in contests of strength and a male that is successful in battle usually secures more female mates.

The elaborate tail of the peacock (*Pavo cristatus*) has fascinated biologists for many years (Figure 11.2.5). Why did such an impractical structure evolve? Research in the 1990s revealed that females prefer not only males with long tails but also those with more eye spots on their tails. Females probably prefer these males because they are likely to produce attractive and therefore reproductively successful male offspring, or because a large tail is a sign of good health. Evidence also suggests that the number of eye spots on a tail can be correlated with the numbers of B and T cells the male produces, which is a sign of a well functioning immune system.

Figure 11.2.5: The tail of a male peacock produces a colourful display to attract female birds.

It seems that females choose mates for their appearance and that a large unwieldy tail does not reduce the male's survival chances.

Factors that affect allele frequencies in populations

Over a period of time, directional selection will lead to an increase the frequency of a favoured allele. For example, if three genotypes – **BB**, **Bb** and **bb** – vary so that **BB** individuals produce more offspring than the other genotypes, the **B allele** would become more common with each generation. The rate at which an advantageous allele approaches the point where it is 'fixed' and is the only variant of the allele in the population can be shown in a graph (Figure 11.2.6). The initial increase in frequency of an advantageous, dominant allele that is rare at first is more rapid than that of a rare, but advantageous, recessive allele. A recessive allele cannot become fixed until it is frequent

enough to occur in homozygous organisms but a new dominant allele has an immediate effect on heterozygous individuals.

Another important factor in determining allele frequency is genetic drift. Genetic drift is defined as a change in allele frequency in a gene pool due to chance events.

KEY POINTS

fixed an allele that is the only variant of that gene that is present in a population. It is homozygous in all members of the population.

genetic drift a change in allele frequency in a gene pool due to chance events.

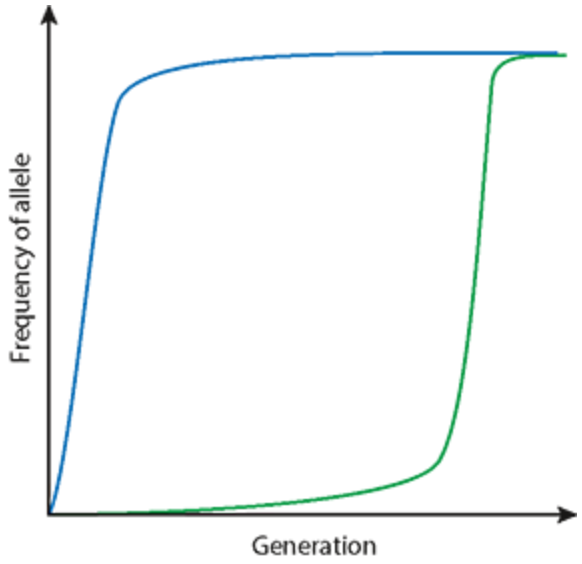


Figure 11.2.6: Graph to show how allele frequencies change under directional selection that favours a dominant advantageous allele (blue curve) and that which favours a recessive advantageous allele (green curve).

It can cause gene pools of two isolated populations to become dissimilar, as some alleles are lost and others become fixed. Genetic drift can occur when a small number of ‘founders’ (new colonisers) separate from a larger group and establish a new population. Founding individuals carry only a small part of the total diversity of original gene pool, and the alleles that they have is determined by chance alone, which means that rare alleles, or rare combinations of alleles, may occur in higher frequencies in the new isolated population than in the general population. This is called the ‘founder effect’.

Many island populations of species, such as Darwin’s finches on the Galápagos Islands, display founder effects and may have allele frequencies that are very different from those of the original population.

The finches living on the Galápagos Islands (Figure 11.2.7), about 900 km off the coast of Ecuador, were important in shaping Darwin’s ideas about natural selection. Studies of the birds, now known as Darwin’s finches, continue to this day and modern DNA analysis indicates that all 13 species now found on the islands probably evolved from a small flock of about 30 birds that became established there around 1 million years ago.

When the birds first arrived, the Galápagos Islands were probably free of predators and initially the resources were sufficient for all the individuals. As the population grew, the finches started to adapt their feeding habits to avoid competition and as each group selected different foods, they developed differently as gene and allele frequencies changed. Eventually a number of separate species were established. Today we recognise 13 different species including the cactus finch which has a long beak that reaches into blossoms, the ground finch with a short stubby beak adapted for eating seeds buried under

the soil, and the tree finch with a parrot-shaped beak suited for stripping bark to find insects (Figure 11.2.8).

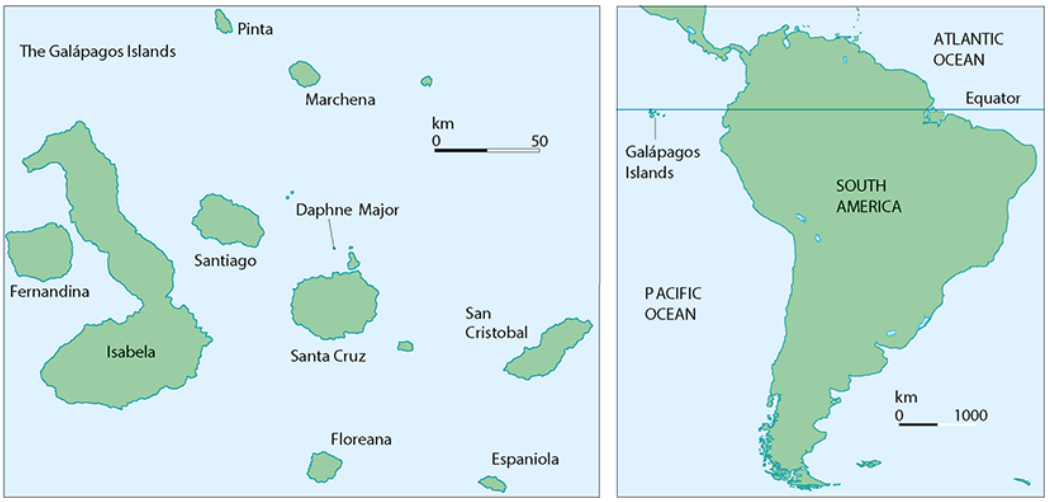


Figure 11.2.7: The Galápagos islands.

Figure 11.2.8: The tree finch, the ground finch and the cactus finch all have very different alleles from the original population of finches that arrived on the Galápagos Islands about 1 million years ago.

One human example of the founder effect is that of the German immigrants who first arrived in present-day Pennsylvania, USA, from Europe in the 16th century, and formed the Amish community. This group of new colonisers did not have all the

genetic variation of the human species, or even of the European population, so the allele frequencies in their gene pool were different from those in the wider human population. For instance, a condition called Ellis–Van Creveld syndrome (a form of dwarfism with polydactyly, or additional fingers and toes) can be traced back to Samuel King and his wife, who arrived in Pennsylvania in 1744. Today, the gene causing the syndrome is many times more common among the Amish people, who tend to marry within their own community, and demonstrates how the genes of the ‘founders’ of a new community are disproportionately frequent in the population from that point on.

11.2.6 The Hardy–Weinberg principle

Godfrey Hardy, an English mathematician, and Wilhelm Weinberg, a German doctor, were two scientists who worked at the start of the 20th century and produced mathematical models of the frequencies of alleles in the gene pool. They concluded that if allele frequencies were stable, they could be calculated with a simple equation based on Mendelian genetics. As a species changes, the frequency of alleles in the population changes. The Hardy–Weinberg equation is used to calculate and model allele frequencies.

The formula is
$p^2 + 2pq + q^2 = 1$

where p = frequency of the dominant allele and q = frequency of recessive allele. If the frequencies of the alleles at a specific locus are considered, then $p + q = 1$ because chromosomes that do not have the dominant allele must have the recessive one.

Using the Hardy–Weinberg equation

If a gene has two alleles, **A** and **a**, the frequency of the dominant allele **A** is represented by p and the frequency of the recessive allele a is represented by q.

In a population of 100 individuals there will be 200 alleles. If 170 of these are the dominant allele then the remainder must be the recessive allele.

$$= \frac{170}{200} \times 100 = 85\% \text{ and } q = \frac{30}{200} \times 100 = 15\%$$

Populations are always represented by 1 so in this example p would be 0.85 and q would be 0.15.

In a randomly mating population, the possible combinations of allels can be determined from a Punnett grid:

		Male allele	
		A	**a**
Female allele	**A**	**AA**	**Aa**
	a	**Aa**	**aa**

If the alleles are replaced with their frequencies:

		Male allele	
		Frequency of dominant allele **A** (p)	Frequency of recessive allele **a** (q)
Female allele	Frequency of dominant allele **A** (p)	p^2	pq
	Frequency of recessive allele **a** (q)	pq	q^2

The four genotypes represent the total population, which is always 1. So the equation becomes:
$p^2 + 2pq + q^2 = 1$

Conditions for using the Hardy–Weinberg equation

Mathematical models like the Hardy–Weinberg equation are used to predict what happens in nature. Any model can only work if certain assumptions are made. This is because all natural

systems are very complex and have many factors that affect them. A model is a simplified version of a complex system.

Evolution is the change in frequency of an allele in a gene pool, so if the frequency remains constant then the population is not evolving. But populations do evolve as a result of selection pressures on them so Hardy and Weinberg outlined seven assumptions. They allow the equation to be used and applied to allele frequencies in a population that is not evolving:

1 There is no mutation that affects allele frequencies.

2 There is no natural selection: the frequency of alleles over time should not vary.

3 The population is large, because the equation is based on proportions and percentages, and larger numbers increase the reliability of the model.

4 All members of the population breed.

5 Mating is random between individuals who have the alleles and the equation does not work for alleles on the sex chromosomes.

6 Each mating produces the same number of offspring.

7 There is no immigration or emigration to alter allele frequencies.

WORKED EXAMPLE 11.2.1

Phenylthiocarbamide (PTC) is a substance that to some people tastes very bitter and to others is tasteless. The ability to taste PTC is controlled by a dominant allele (T).

Two populations of people were sampled: indigenous Aborigine population in Australia and the Quechua people of Peru. Both sample sizes were 500. In the Australian sample, 245 people were non-tasters and in the Quechua sample 20 people were non-tasters. Calculate the allele frequencies in these two groups.

Answer

Step 1 Non-tasters must be homozygous recessives (**tt**) and are therefore represented by q^2.

Step 2 For the Australian sample the proportion of people who are non-tasters is calculated as:

$$\frac{245}{500} = 0.49$$

So, $q^2 = 0.49$

Therefore, $q = 0.7$

The frequency of the non-tasting allele among the indigenous Aborigine population in Australia is 0.7.

Step 3 Since $p + q = 1$

$p = 0.3$

The frequency of the tasting allele in the Australian sample is 0.3.

Step 4 In the sample of the Quechua people, 20 were non-tasters so the proportion of Quechua people who are non-tasters is:

$$\frac{20}{500} = 0.04$$

So, $q^2 = 0.04$

Therefore, $q = 0.2$

The frequency of the non-tasting allele among the Quechua people is 0.2.

Step 5 Since $p + q = 1$

$p = 0.8$

The frequency of the tasting allele among the Quechua people is 0.8.

TEST YOUR UNDERSTANDING

11 Suggest why a changing environment is likely to favour disruptive selection.

12 Outline an example of sexual selection and how it benefits the species that you name.

REFLECTION

Reflect on the most important things you have learnt in this section.

Links

- How can climate change influence the survival of species? (Chapter 12.6)

11.3 Evolution

LEARNING OBJECTIVES

In this section you will:

- define a species as a group of organisms that share common characteristics and can interbreed to produce fertile offspring
- learn that all species have originated from a common ancestor by evolution
- define evolution as cumulative change in the heritable characteristics of a population
- describe the evidence for evolution from fossils, observation of phenotypes, selective breeding and comparisons of DNA sequences
- learn that population can diverge into separate species and define speciation as the process which forms new species
- learn that speciation occurs when two populations of a species are isolated by geography, behaviour or time and evolve differently
- recognise how infertile hybrids isolate evolving populations and prevent mixing of alleles between them

- define a gene pool as all the genes and alleles in an interbreeding population

- learn that evolution requires allele frequencies to change with time in populations
- learn that reproductive isolation can be temporal, behavioural or geographic
- discover that speciation of isolated populations can be gradual or occur abruptly.

GUIDING QUESTIONS

- How is evolution driven by natural selection?
- How do new species arise?

11.3.1 What is evolution?

Over long periods of time and many generations, the genetic make-up of species may change as they become adapted to new surroundings or altered conditions. One result of these changes may be the evolution of new varieties and species. There is strong evidence for the evolution of life of Earth, both from fossils and from organisms that are alive today. Natural selection provides an explanation of how evolution might have occurred.

KEY POINTS

evolution cumulative change in the heritable characteristics of a population.

species a group of organisms that share common characteristics and can interbreed to produce fertile offspring.

Life on Earth is always changing. Just by looking at any group of individuals of any species – whether humans, cats or sunflowers – you can see that individuals are not all the same. For example, the people in Figure 11.3.1 vary in height, hair colour, skin tone and many other ways. How do the differences within a species occur? How do different species arise?

A species is defined as a group of organisms that share common characteristics and can interbreed to produce fertile offspring. Variation within a species is a result of both genetic and environmental factors. We say that selection pressures act on individuals and because of variation, some may be better suited to their environment than others. These are likely to survive longer and have more offspring.

Figure 11.3.1: Most of the variation between humans is continuous variation and is influenced by the environment as well as genes.

The characteristics of a species that are determined by the genes are inherited and passed on to succeeding generations (Section 4.3). The cumulative change in these heritable characteristics over generations is called evolution. If we go back in time, then existing species must have evolved by divergence from pre-existing ones. All life forms can therefore be said to be linked in one vast family tree with a common origin that has changed and developed over time. You can read more about convergent and divergent evolution of the forms and functions of organisms in section 11.4.3.

11.3.2 Evidence for evolution

Evidence for evolution can be gathered from a number of different sources including evidence from fossils, observations of organisms alive today and studies of DNA base sequences. Evidence like this can enable us to build up a picture of how the organisms we see today have evolved.

The fossil record

Fossils, such as the one shown in Figure 11.3.2, are the preserved remains of organisms that lived a long time ago. They are often formed from the hard parts of organisms, such as shell, bone or wood. Minerals seep into these tissues and become hardened over time. As the living tissue decays, the minerals form a replica that remains behind. Soft tissue can sometimes be preserved in the same way, as can footprints and animal droppings. Most fossils become damaged over time or are crushed through land or sea movement, but some are discovered remarkably well preserved. The earliest fossils found date from over 3 billion years ago, so the time scale of the fossil record is immense. Most fossils are of species that died out long ago, because they did not adapt to new environmental conditions.

KEY POINT

fossil record a history of life found in the fossil remains of organisms from long ago preserved in sedimentary rock.

The study of fossils is called palaeontology. Palaeontologists have been collecting and classifying fossils for over 200 years, but they have only been able to work out how old they are since the 1940s. Scientists can estimate when a fossilised organism

might have lived by studying the amount of natural radioactivity in the fossils they find. This process is called radiometric dating. Over time the amount of radioactivity decreases because radioactive elements decay. The rate of decay is fixed for each element, so it is possible to date a fossil by measuring the amount of radioactivity present in it and so determine the time that has passed since it was formed. Carbon-14 is used to study material up to 60 000 years old. For older material, other elements are used.

Figure 11.3.2: A fossil of *Archaeopteryx*, which is seen as an evolutionary link between reptiles and birds. It looked like a small dinosaur, but it had feathers and could fly.

Although the fossil record is incomplete and fossils are very rare, it is possible to show how modern plants and animals might have evolved from previous species that existed hundreds or thousands of millions of years ago. It is important to recognise

that we can never say that ‘this species evolved into that species’, based on a fossil sequence, even when we have many fossils. All that we can say is that species appear to be related, that they probably share a common ancestor. Other species may have existed too, for which no fossils have ever been found.

A few organisms seem to have changed very little. The horseshoe crab we see today is very similar to fossil specimens that are 1 million years old. This suggests that there has been little selection pressure on these crabs.

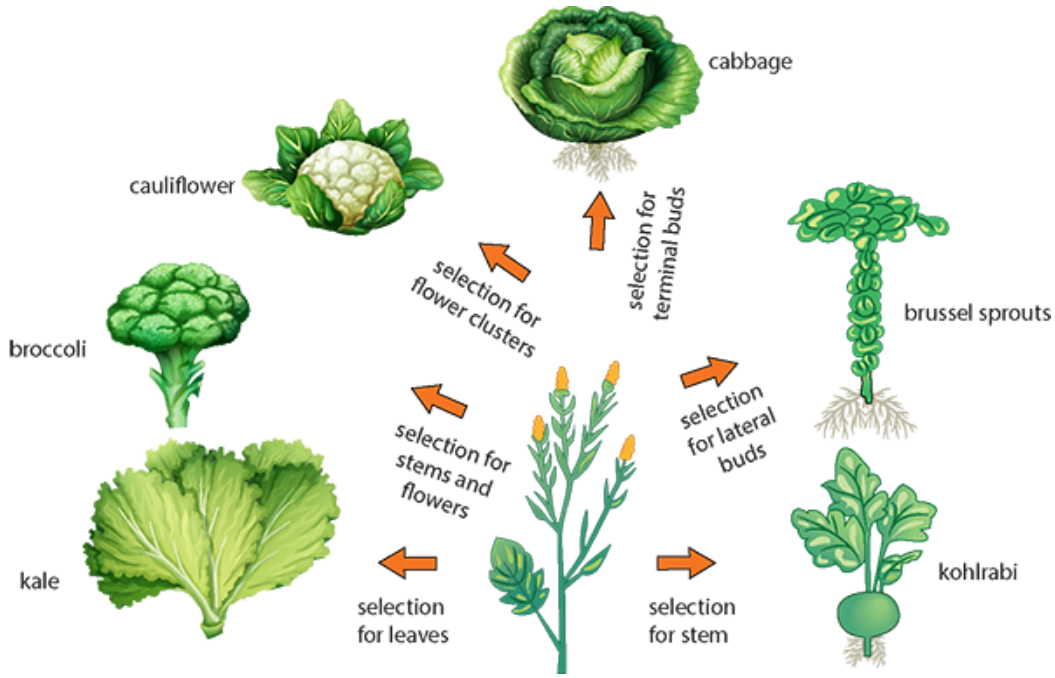


Figure 11.3.3: Selective breeding of the wild mustard plant *Brassica oleracea* has produced a variety of different food plants.

Observations of fossils provide evidence that life on Earth changes and that many of the changes occur over millions of years.

Selective breeding

Further evidence for the way evolution occurs comes from observations of selective breeding. People have altered the balance of alleles and the characteristics of domesticated species by breeding selected individuals in a process called artificial selection (Section 11.2). Plants or animals with desirable characteristics are chosen and bred together so that the numbers of offspring in the next generation with these characteristics increase. Individuals that do not have the desired features are not allowed to breed. People have selected plants and animals in this way for thousands of years and, over many generations, this has resulted in the evolution of numerous breeds and varieties, which differ from each other and from their original wild ancestors (Figure 11.3.3).

Direct observations of phenotypes: the work of John Endler

Canadian biologist John Endler studied wild guppies (*Poecilia reticulata*) in Trinidad in the 1970s. He noticed that there was a wide variation in colour among guppies from different streams and even among guppies living in different parts of the same stream. Males from one area could be bright blue with orange patches along their sides, while those further downstream were less brightly coloured with fewer patches. Endler recorded differences in the distribution of guppies and their predators, and also the background colours of the gravel in the different streams. He discovered a clear correlation between location and colouration in the fish and carried out experiments to test his hypothesis. His experiments show how selection can work to increase desirable characteristics and remove those which are not favourable for survival.

Figure 11.3.4: *Poecilia reticulata* is a species of guppy that is found in Trinidad. Male coloration depends on the presence or absence of predators.

cow	ATG---ACTAACATTCGAAAGTCCCACCCACTAATAAAAATTGTAAAC
sheep	ATG---ATCAACATCCGAAAAACCCACCCACTAATAAAAATTGTAAAC
goat	ATG---ACCAACATCCGAAAGACCCACCCATTAATAAAAATTGTAAAC
horse	ATG---ACAAACATCCGGAAATCTCACCCACTAATTAAAATCATCAAT
donkey	ATG---ACAAACATCCGAAAATCCCACCCGCTAATTAAAATCATCAAT
ostrich	ATGGCCCCCAACATTCGAAAATCGCACCCCCTGCTCAAAATTATCAAC
emu	ATGGCCCCTAACATCCGAAAATCCCACCCTCTACTCAAAATCATCAAC
turkey	ATGGCACCCAATATCCGAAAATCACACCCCCTATTAAAAACAATCAAC

Figure 11.3.5: DNA base sequences from eight vertebrate organisms.

Endler's hypothesis was that female guppies prefer colourful males for mating. But predators can also spot colourful fish more easily. In locations where the numbers of predators are low, males are more colourful. If there are more predators, males are less colourful.

Endler transferred predatory fish to the areas with brightly coloured male guppies and he found that, after about 10 years, selection had acted to produce a population of duller males.

Brighter males had been eaten and their genes removed from the population. His experiments show that variation within a population provides the basis for rapid evolution when the environmental conditions change and that we can directly observe changes taking place.

Comparing DNA base sequences

Over the course of millions of years, mutations accumulate in any given segment of DNA. The number of differences between comparable base sequences can be used to demonstrate the degree of evolutionary divergence. Figure 11.3.5 shows sequences of bases in DNA of eight vertebrate species. The degree of similarity indicates how closely related the species are and how they may have evolved. Non-coding regions of DNA (Chapter 4) provide the best way to compare DNA sequences because mutations that occur in them are likely to remain over many generations. This is because selection pressure on these sequences is less than on sequences that code for proteins (exons). From information carried in DNA we can construct cladograms that show the relatedness of different species (Section 11.1). DNA sequences that make up specific genes and code for proteins mutate at a similar rate to non-coding regions, but changes in these areas are selected against more strongly. This is because changes to base sequences in genes can have detrimental effects on protein structure.

Species can also be compared for relatedness using similarities in their proteins, based on amino acid sequences. But this method will demonstrate a slower rate of change because degeneracy in the genetic code means that different **codons** can be used to insert the same amino acid into a protein. So some mutations in the DNA will not lead to a change in amino acid.

TEST YOUR UNDERSTANDING

13 The table shows the sequence of amino acids in part of a hemoglobin molecule of five different species. Use the information to work out the most closely related species in the list.

Species	Amino acid sequence in the same section of hemoglobin molecules
human	Lys–Glu–His–Iso
horse	Arg–Lys–His–Lys
gorilla	Lys–Glu–His–Lys
chimpanzee	Lys–Glu–His–Iso
zebra	Arg–Lys–His–Arg

14 State why the analysis of repeated DNA sequences is a useful way of searching for evolutionary relationships.

SCIENCE IN CONTEXT

Using mitochondrial DNA to investigate human evolution

DNA found in mitochondria is also used for examining evolutionary relationships *within* one species. A fertilised egg cell contains mitochondria from an offspring's mother but none from its father as only the sperm nucleus enters the egg at fertilisation. This means that mitochondrial DNA is only inherited from the mother of an offspring and can give a more direct indication of an evolutionary relationship.
Recombination does not occur in mitochondrial DNA and it also mutates more quickly than nuclear DNA so that changes

can easily be spotted. Every cell in the body has many mitochondria so it is also easier to collect more DNA for analysis. This type of analysis is most useful for looking at relationships between organisms that have evolved relatively recently in evolutionary history. Analysis of human mitochondrial DNA has been used to trace the origins of present-day human populations. All humans are thought to be related to one woman, known as ‘mitochondrial Eve’. This woman lived between 100 000 and 200 000 years ago in southern Africa. She was not the first human, but other female lines disappeared or had no female offspring and did not pass on their mitochondrial DNA. As a result, all humans today can trace their mitochondrial DNA back to Eve. In the years since she was alive, different populations of humans have drifted apart physically and genetically to form the different ethnic groups we see today.

11.3.3 How new species arise

Speciation

Speciation is the formation of new **species** from an existing population. Once a species has evolved to become well adapted to conditions in a stable environment, natural selection tends to keep things much the same. But if the environment changes or part of the population becomes separated from another, populations can gradually diverge into separate species by natural selection. Members of the same species and, therefore, the same gene pool can fail to reproduce as a result of a barrier that separates them. New species appear as a result of the population of a single species splitting into two or more new ones, each with its own gene pool.

KEY POINTS

population a group of organisms of the same species that live in the same area at the same time.

speciation the evolution of a new species.

Speciation can only occur if there is a barrier dividing the population (Figure 11.3.6). The barrier may take different forms, such as:

- geographical separation, such as a river, mountain range or a road
- temporal differences (meaning differences in time), when two groups mate at different times so that two populations never meet to mate and exchange genetic material

- behavioural differences, such as mating rituals or songs becoming different and incompatible so that two groups are no longer able or interested in reproducing.

When one part of the divided population is isolated from the other, mutation and selection can occur independently in the two populations so that each has the potential to become a new species.

Speciation is said to be sympatric or allopatric.

- **Allopatric speciation** occurs in different geographical areas.
- **Sympatric speciation** occurs in the same geographical area.

The differences between the two are summarised in Table 11.3.1.

Geographical separation, reproductive isolation and speciation

Bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*) are two species of ape which diverged from a common ancestor between 1 and 2 million years ago The apes are different in their behaviour, body shape and even their emotions and understanding.

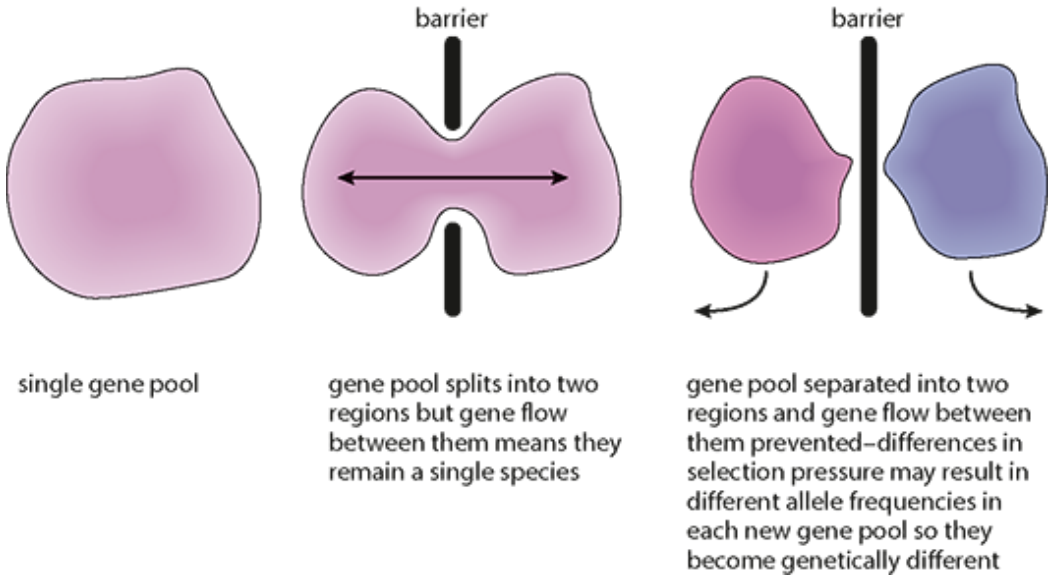


Figure 11.3.6: If a population is separated into two groups by a barrier the two groups may gradually diverge into separate species.

Sympatric speciation	Allopatric speciation
a new species arises from an existing species that is living in the same area	a new species arises because a geographic barrier separates it from other members of an existing species
temporal or behavioural isolation can produce significant changes in the genetic make-up within a species so that a new species is formed	geographic barriers may include mountain ranges, valleys or bodies of water, or human-made features such as roads, canals or built-up areas

Table 11.3.1: Comparison of sympatric and allopatric speciation.

Chimpanzees are larger and have stocky, muscular bodies and can be aggressive and confrontational even with members of their own species. Bonobos have smaller, more slender bodies and tend to be gentler and more cooperative, they are also more likely to walk on two legs (Fig 11.3.7). The common ancestor of both species was separated into two groups after the formation of the Congo River which divided one population of apes into two. Neither of the species can swim so there was no interaction between the separated groups. Once separated, the two groups faced different environments. To the north the group had to compete with gorillas for food and as a result there was also competition between individuals for resources. To the south, there were no gorillas, food was more available and so there was less fighting for food and mates. The northern group eventually became chimpanzees who are more aggressive than the group to the south of the river who became bonobos. The two groups developed into the two species we recognise today.

Infertile hybrids

Reproductive isolation prevents alleles in separated populations, which have evolved differences, from mixing. But sometimes individuals from different species may meet and attempt to breed, forming a hybrid. Hybrids between different species are known as interspecific hybrids and are formed when two species mate, or are bred by humans. The offspring of these crosses are usually **sterile** and are not able to produce fertile offspring. Hybrid sterility shows us how speciation can occur. It prevents genes moving from one species to another so both species are kept separate even if geographic, temporal or even behavioural barriers are removed.

KEY POINT

hybrid an offspring produced by interbreeding of two organisms of different species, or of genetically distinct populations within a species. Such offspring often possess combinations of new characteristics.

Hybrid offspring are often sterile because their parents have different numbers of chromosomes. One example is the mating of a donkey and a horse, which produces a mule or a hinny. Mules (Figure 11.3.8), which are produced by mating a male donkey and a female horse, and hinnies, from mating a female donkey and a male horse, develop normally and reach sexual maturity, but they cannot reproduce. Horses have 64 chromosomes while donkeys have 62. A hybrid between the two species has 63 chromosomes and cannot produce viable gametes because the odd chromosome cannot form homologous pairs at meiosis. Other examples of infertile hybrids include ligers, a cross between a lion and a tiger, and zonkeys which are hybrids between a zebra and a donkey.

Most organisms are diploid and have cells with two sets of chromosomes ($2n$), one from each parent, but **polyploid** organisms have cells that contain more than two sets.

Figure 11.3.8: A mule is produced by mating a male donkey and a female horse. Mules have a short, thick head, long ears, small hooves, and short mane.

Those containing three sets of chromosomes are said to be triploid (3*n*), those with four sets tetraploid (4*n*), five sets (5*n*) pentaploid, six sets (6*n*) hexaploid, and so on.

Polyploidy is widespread in plants but rare in animals. It happens when sets of chromosomes are not completely separated into individual nuclei during cell division so that one cell ends up with additional chromosomes. It is estimated that more than 50% of flowering plants have undergone polyploidy at some time during their evolution.

In many cases extra sets of chromosomes can produce plants that have improved 'vigour', such as a greater resistance to disease or larger fruits (Figure 11.3.9). Polyploidy happens naturally but can be induced by treatment with chemicals and is used in agriculture to produce higher yields. Bread wheat is a hexaploid plant and many members of the Brassica family (cabbages and mustard plants) are tetraploids.

Figure 11.3.9: Diploid (2*n*) and tetraploid (4*n*) grapes.

A tetraploid plant will have chromosomes that each have a matching pair and will be able to undergo meiosis to form fertile gametes. A tetraploid can cross with another tetraploid to form fertile offspring in just the same way as normal plants. But a cross between a tetraploid and a diploid plant would produce triploid plants that would be sterile and unable to form gametes. In this case polyploidy acts as a barrier between the diploid and tetraploid species. The two populations may become so different that they develop into new species.

Polyploidy in plants, which increases the number of sets of chromosomes in plant species, can result in the formation of new plant species without isolation from the parent plants. This 'instant speciation' occurs because the new polyploid plants cannot interbred with the diploid parent plants.

One example of possible 'instant speciation' can be seen in Japanese knotweed (*Fallopia japonica*). It was introduced into North America about 100 years ago and is now known as one of the world's top 10 invasive species. In England, Japanese knotweed is sterile and is genetically a single, large female clone. It spreads by fragmentation (small pieces breaking off and starting new plants). But in other parts of Europe and North America, where there are no male plants, female Japanese knotweed plants are reproducing with male plants of other species. Female plants are being fertilized by pollen from giant knotweed (*Fallopia sachalinensis*), and other Fallopia species. A fertile hybrid Bohemian knotweed (*Fallopia bohemica*) is produced and is a special case of polyploidy called allopolyploidy which occurs when two different species interbreed and produce fertile hybrid offspring. Japanese knotweed is octoploid (88 chromosomes); giant knotweed is

tetraploid (44 chromosomes); and many hybrid plants have varying numbers of chromosomes, the most common is sexaploidy (66 chromosomes). So the hybrid *F. bohemica* seems to be a new species, developed without geographic isolation from the parent species. This is an example of sympatric speciation and we may see more speciation within genus Fallopia as hybrids interbreed in future.

SCIENCE IN CONTEXT

Hybridisation is used commercially to develop new plants with desirable qualities such as disease resistance or the size of plants, flowers or fruits. Many plants that gardeners and horticulturalists buy are hybrids. A hybrid plant is produced by cross-pollinating two different plant species or varieties by hand and growing the seeds that the cross produces. Producing hybrid plants can take many years. Seeds from the first crosses are grown the following year and the plants they produce are evaluated. If they have the right features, the cross will be repeated and the seeds sold. If they are not what the grower wants, the hybridisation process must be repeated again.

Seeds for popular hybrids, such as varieties of tomatoes called Sungold (Figure 11.3.10), have to be crossed, harvested and saved every year. These plants are called F_1 (or first generation) hybrids because they are the direct product of a cross. The plant breeder who creates a hybrid owns the rights to it, which is why they are more expensive than non-hybrid plants.

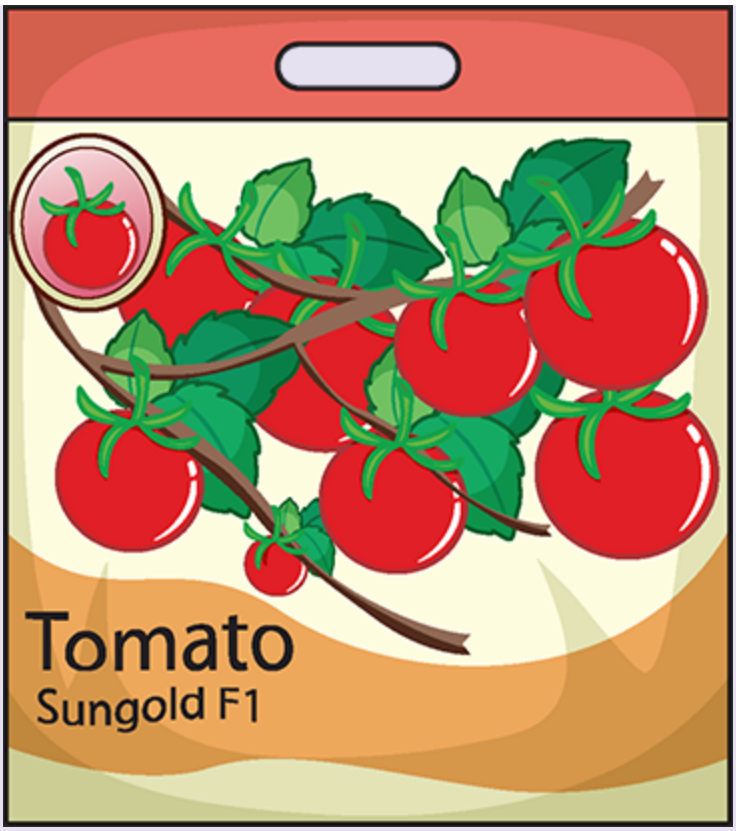


Figure 11.3.10: F_1 hybrid plants tend to be more expensive.

TEST YOUR UNDERSTANDING

15 Define a species.

16 List three sources of evidence that evolution has taken place.

17 Why is isolation of part of a population essential for speciation to take place?

18 Why are many hybrids infertile?

11.3.4 Effects of isolation on the gene pool

The gene pool is all the genetic information in a population that is reproducing at a given time. Allele frequency gives us an indication of the proportion of a specific gene variant in a population. Gene pools do not change greatly over time but new alleles can be introduced as a result of mutations, and other alleles may disappear from a population if the last individual that has it leaves or dies. Genetic drift can change the allele frequency in a gene pool due to chance events. It can cause gene pools of two isolated populations to become dissimilar, as some alleles are lost and others become fixed. Genetic drift can also occur when a small number of individuals separate from a larger group and establish a new population.

KEY POINTS

gene pool is when all the genes and their alleles are present in an interbreeding population.

genetic drift is a change in allele frequency in a gene pool due to chance events.

Alleles which are not favourable will be passed on to fewer offspring and so their frequency will tend to decrease. In Section 11.2 you can read about how over a period of time, selection will lead to an increase the frequency of a favoured allele in the gene pool. The initial increase in frequency of an advantageous, dominant allele that is rare at first is more rapid than that of a rare, but advantageous, recessive allele. A recessive allele cannot become fixed until it is frequent enough to occur in

homozygous organisms, but a new dominant allele has an immediate effect on heterozygous individuals. If new individuals enter the population due to immigration there will be a change in allele frequencies as a result. Similarly if individuals emigrate the gene pool will be changed. If we can see that gene pools and allele frequencies have changed, we can deduce that some evolution of the population has happened.

Reproductive isolation

Reproductive isolation keeps two species separate so that there is no gene flow between populations. Three types of isolating mechanism operate in newly separated populations.

1 Geographical isolation describes isolation of a gene pool by a physical barrier such as a river, ocean, mountain range or desert. For example, many Galápagos Island species such as iguanas and finches are now separate from the South American mainland species that they arose from. The island populations have evolved to be different from their mainland ancestors (Section 11.2).

2 Temporal isolation occurs when individuals from different species do not breed because they are active at different seasons or times of the day. Flowers such as orchids in different species of the genus *Dendrobium* open at different times and can only be pollinated by pollen from their own species. Other members of the same genus living nearby cannot pollinate them because their pollen is not produced at the correct time.

3 Behavioural isolation occurs when different populations or species develop different courtship behaviours. Members of the same species are only attracted by the specific calls, songs or displays of their own kind and ignore or reject

displays of other species. For example, the blue-footed booby (*Sula nebouxii*) lives in the same habitat as other species of the genus *Sula*. Even though they are similar, the blue-footed booby never mates with them. The female blue-footed booby selects a partner after watching a courtship ritual that is unique to its own species.

SCIENCE IN CONTEXT

Lactase persistence (LP) and allele frequency around the world

Lactase is an enzyme needed for the digestion of lactose in milk. Without the enzyme, drinking milk can cause the symptoms of lactose intolerance: bloating, flatulence, pain and nausea. Production of lactase in adult life, known as lactase persistence (LP), is a genetically determined characteristic and is common in people of European origin and some African, Middle Eastern and Southern Asian populations, but rare in other places.

Lactase persistence is an example of natural selection taking place in humans. About 65% of people stop or reduce the production of the enzyme lactase after they stop drinking milk as infants (at weaning). But five different variants of the lactase gene are found at different frequencies in different parts of the world. The persistence of lactase production is thought to be an example of the co-evolution of genes and culture. Analysis of ancient DNA has shown that in Spain, Germany and Sweden the alleles were present between 4000 and 5000 years ago. One of these alleles is fixed in some parts of Europe, while others are found at variable frequencies in the Middle East and Africa. Present-day frequencies of the alleles tell us that there is positive selection for lactase

persistence. Nucleotide analysis shows that lactase persistence seems to have evolved independently in at least four parts of the world and several different mutations have been found. Lactase persistence can provide an evolutionary advantage. The ability to digest lactose in adulthood means that milk can be consumed to provide increased nutritional benefits as well as being a source of water. So natural selection works to retain the LP genes and alleles at high frequencies.

To consider:

1 Milk can be treated with lactase so that lactose-intolerant people can drink it. What happens to the lactose?

2 Why is this treatment important to people who do not have lactase persistence?

NATURE OF SCIENCE

Paradigm shift

A paradigm shift is a change in the core beliefs or assumptions of an accepted scientific theory. It occurs when scientists are faced with anomalies that cannot be explained by the accepted paradigm. In modern science, a number of paradigm shifts have taken place in recent times. These include the acceptance of plate tectonics to explain large-scale changes in the continents and the replacement of Newtonian mechanics with quantum mechanics. In biology, 'pangenesis' (Darwin's provisional theory that a reproductive cell contained 'gemmules' from every part of an organism in order to produce a new individual) was replaced with an acceptance of Mendelian genetics.

Two competing theories – gradualism and punctuated equilibrium – attempt to explain the appearance of new species and the absence of intermediate forms in the fossil record. The first view assumes that species gradually change over long periods of time, while the theory of punctuated equilibrium proposes short periods of rapid evolution interspersed with long periods of equilibrium.

Some scientists reject punctuated equilibrium, but as it can be explained in terms of natural selection it is possible that both processes may have occurred. On the other hand, new evidence and analysis of gene sequences has given support to the gradualism viewpoint. For a long time scientists thought that the **mass extinction** of the dinosaurs 65 million years ago led to the rise of the mammals. This view seemed to be supported by the fossil evidence. But recent genetic analysis using the GenBank database has indicated that early mammals were present at least 100 million years ago, 35 million years before the extinction of the dinosaurs. Furthermore, their evolution followed a gradualism path, not a pattern of punctuated equilibrium.

To consider:

1 How does a paradigm shift take place in science?

2 What factors are involved in the acceptance of a paradigm shift?

TEST YOUR UNDERSTANDING

19 Give an example of a feature that can cause geographic isolation.

20 Outline the differences between gradual and punctuated evolution.

21 Define a gene pool.

REFLECTION

Reflect on how our understanding of genetics has increased our understanding of evolution and changes in populations. Could you explain our knowledge of DNA, genes and evolution to a scientist from 150 years ago?

Links

- What aspects of inheritance help us to explain evolution? (Chapter 4)

11.4 Ecological niches, adaptations and evolution

LEARNING OBJECTIVES

In this section you will:

- learn that environments contain a variety of different niches
- discover that every organism has a specific niche, defined by habitat, tolerance limits and an organism's function in the habitat
- learn that niches can vary in size and diversity
- understand the difference between fundamental and realised niches
- distinguish between convergent and divergent evolution
- understand the evolution of homologous structures by adaptive radiation to explain similarities in structure when there are differences in function
- recognise that homologous structures have evolved from a common ancestor
- understand that adaptive radiation increases the biodiversity of a community and use the Simpson's reciprocal index of diversity to analyse community
- distinguish between homologous and analogous structures

- Understand that two species cannot occupy the same niche in a habitat

GUIDING QUESTIONS

- What factors determine where a species is likely to be found?
- How are vacant niches filled in ecosystems?
- How do species evolve to occupy a niche?

11.4.1 Niches and community structure

Habitats and niches

In all communities, each species has a unique role. This role is determined by the species' place in the habitat and the interactions that it has with other species. A **habitat** is an area offering living space to a number of different types of organism, and includes all the physical and **abiotic** (non-living) factors such as climate or soil type in the environment. An example might be a woodland habitat, whose community includes a huge variety of species, from burrowing invertebrates at ground level to nesting birds in the tree canopy. Every organism occupies its own space in an ecosystem, which is known as its spatial habitat. The surroundings are changed by the presence of the organisms; for example, woodpeckers live in woods and forests and make their nests within hollows in trees, adapting them to provide shelter for eggs and chicks, while a rabbit burrowing underground affects the soil and plant species growing there.

A **niche** is the particular environment and 'lifestyle' that is adopted by a species. It includes the place where the organism lives and breeds – its spatial habitat – as well as its food and feeding method, and its interactions with other species. As an organism feeds within its niche, it affects the other organisms that are present. For example, an owl feeding on mice in woodland helps to keep the population of mice at a stable level, and rock limpets grazing on small algae on a rocky shore control the degree of algal cover. A habitat comprises a number of niches, each of which is unique to its particular species because it offers the exact conditions that the species needs or has become adapted to.

KEY POINTS

abiotic refers to non-living features of an environment such as climate, soil type or temperature.

habitat is the features that describe the environment where a species normally lives.

niche is a concept that describes where an organism lives (its spatial habitat), what and how it eats, and its interactions with other species.

11.4.2 Adaptations to environment

Tolerance limits

The organisms present in a community depend on the other organisms living there, as well as on the non-living, abiotic aspects, such as soil or climate. These abiotic aspects define a species' tolerance limits, those factors that affect a species' survival and distribution. Some of these abiotic factors that influence the distribution of plants and animals in communities are outlined here.

KEY POINTS

tolerance limits the abiotic factors in an environment that limit the survival and distribution of a species.

spatial habitat is the space within an ecosystem where an organism lives.

Plants

- Temperature – No plant can survive extremely cold conditions for very long, because to grow and reproduce plants must carry out chemical reactions in their cells. These reactions require enzymes. In cold climates plant growth is very slow, but it increases when the temperature rises.
- Water – All plants require water. It is the universal solvent in their cells, the substrate for photosynthesis and their transport medium. However, plants have evolved a variety of mechanisms to survive periods of drought. Some species remain dormant, some (such as cacti and succulent plants)

store water and others complete their life cycle in a brief rainy season.

- Light – Plants need light for photosynthesis. Many plants use the changing day lengths of the different seasons to trigger flowering.
- pH, salinity and nutrients – Most plants prefer a pH of 6.5–7.0 because nutrients are easily available in this range. Saline (salty) soils present a particular problem to plants because salt makes it difficult for plants to take up water and minerals. A few plants, such as marram grass and lyme grass, can survive in saline conditions. Soils that are rich in minerals can support a diverse community of plant species, including trees and shrubs. Plants that survive in mineral-poor soils often have special adaptations to supplement their needs.

Animals

The distribution of animals is also affected by the abiotic factors in their environment. If any factor required by an animal is in short supply or is unsuitable for survival, the distribution of the species will be limited by that factor.

- Temperature – Animal enzymes are influenced by temperature in much the same way as those of plants. However, animals have the advantage that they can move to avoid the harshest of conditions and some use homeostasis to maintain their body temperature. Animals can also hibernate during cold months, or hide when temperatures are extreme (Section 8.5).
- Water – Most animals need to drink water to survive: very few have evolved to be independent of liquid water. Lack of

water in certain seasons may change the distribution of animals. Herds of wildebeest and zebra in Africa undertake huge migrations to find new supplies of water and vegetation (Section 8.5).

- Breeding sites – Animals need to find appropriate sites to find a mate and perform their mating behaviours and then rear young. These sites may be chosen for safety away from predators, or because they provide rich feeding grounds so that the young may benefit.
- Food supply – All animals need a source of food and this will depend on the abiotic factors in the environment. Herbivores need plants and carnivores need other animals to feed on. The availability of food will determine the distribution of different types of animal.
- Territory – Territories provide sources of food and breeding sites. Herbivores, such as wildebeest, graze on large areas of grassland and, when the dry season arrives, migrate to find fresh grass. Some birds, such as the European robin, defend their territories vigorously because they contain food and a nesting area. Carnivores that live in packs, such as wolves, require a large area in which to hunt and they mark their territories with scent and defend it from other packs.

Coral reefs – a marine ecosystem

Corals are animals in the same taxonomic group as sea anemones, but unlike anemones coral secretes an exoskeleton of calcium carbonate over its body. Coral polyps do not photosynthesise but have a symbiotic relationship with microscopic algae called zooxanthellae (see section 12.3). These organisms live inside the polyps and provide organic nutrients such as glucose and amino acids A coral reef is a sensitive

underwater ecosystem which is formed of colonies of coral polyps held together by calcium carbonate Most coral reefs occur in shallow water near shore and are very sensitive to water conditions As a result, they are particularly vulnerable to the effects of human activities either on nearby land or in the wider world.

A coral reef can only form and survive in the right abiotic conditions. They require:

- oxygen, and carbon dioxide
- water at the correct salinity, temperature, pH and clarity
- light: corals need a moderate amount of sunlight
- depth of water, corals grow at depths where sunlight can reach them.

Coral reefs have declined by about 50% in the last 70 years. Coral reefs are threatened by sediment which makes the water around them cloudy. Light cannot penetrate and the sediment can smother corals so they and the zooxanthellae cannot feed and grow. Pollution and toxins can also kill the coral. Ocean warming and acidification, caused by global warming, increase both levels of carbon dioxide and sea temperatures to a level which coral cannot tolerate.

Adaptations and tolerance limits of mangrove trees

Mangrove trees grow on tropical shores at the edge of the sea. Mangrove forests grow at tropical and subtropical latitudes near the equator where the sea surface temperatures never fall below 16°C. Mangrove trees line about two-thirds of the coastlines in tropical areas of the world. As a tropical species they survive at air temperatures above 19°C and cannot tolerate any change in

temperature which is more than 10°C or below freezing for more than a short period of time.

Mangrove trees have two important adaptations that allow them to survive in the extreme conditions of shores and estuaries. They can survive in waterlogged and anoxic (no oxygen) soil, and they can tolerate salty water. Some mangroves remove salt from water using ultra-filtration in their roots, other species have glands on their leaves that actively secrete salt, so that salt crystals appear on the upper surface of the leaves. All mangrove trees have spreading roots and vertical anchor roots to hold them in place in the shallow soil as the tides come and go. In areas where their roots become flooded at high tides, mangrove trees grow aerial root called pneumatophores which absorb oxygen from the air to supply the roots. Mangrove trees can survive in water which is 100 times saltier than most plants can tolerate and withstand twice-daily flooding by the ocean tide that would kill other trees. The tides bring nutrients and carry waste products away from the mangroves. Tides also distribute tree seedlings which reduces interspecific competition for space. If the tides transport salt water into estuaries where it mixes with fresh water, mangrove trees can extend their range and grow further inland.

Figure 11.4.1: Buttress roots stabilise mangrove trees and help them resist tidal waters.

Obligate anaerobes, facultative anaerobes and aerobes

Plants and animals require oxygen for aerobic respiration to release energy from their food but several other groups of organisms do not. **Obligate anaerobes** are microbes that are killed by the concentrations of oxygen found in the air. Some bacteria and protozoa can only live in environments such as the intestines of animals, the deepest ocean and waterlogged soils where there is no oxygen present. Examples include the bacteria that cause tetanus and methanogenic bacteria found in hot thermal vents. Archaea are obligate anaerobes, living in extreme environments.

Some anaerobic organisms can use oxygen if it is available, but can also respire anaerobically, these organisms are called **facultative anaerobes** and include yeast which respires

aerobically if oxygen is present but will change to fermentation (anaerobic respiration) when it is not.

Distribution of species and abiotic variables

When ecologists want to understand the distribution of a species and relate it to an abiotic factor, or to compare the distribution of one species with another in a different location, it is usually impossible to do so by a direct counting method. There are a number of sampling methods used to collect data. Two commonly used methods are **quadrats** and **transects**. They can show not only which species are present, but also how many individuals of each species there are. They provide a method of systematic sampling.

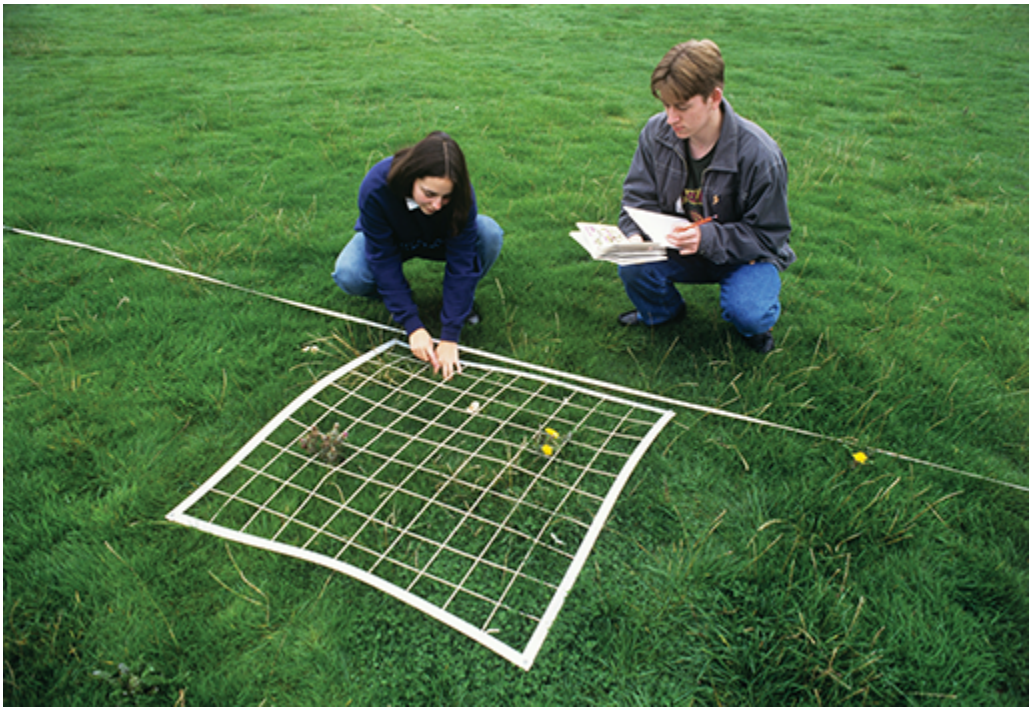


Figure 11.4.2a: These students are using a transect line to survey the plants in a grassy area. A quadrat is placed at measured intervals along the transect line and the plants at each location

are counted and recorded. In this way, the plant population can be estimated from a series of samples in a few areas.

A quadrat is a portable square with a fixed area. It is placed on the ground and the species inside the square are counted and recorded (See Figure 11.4.2a). A transect consists of a tape or rope stretched from a fixed point across an area of interest, where the abiotic factors may be changing and the distribution of organisms may be different. Transects can be used to sample the distributions of plants along a beach or through a field, or to study the vegetation as soil or moisture changes along a line of interest. Samples are taken by placing a quadrat down at intervals along the tape so that the organisms within it can be accurately counted (Fig 11.4.2b). Relevant abiotic factors such as temperature, light, salinity or soil pH can also be measured at each location. In this way a transect can reveal the distribution of a species in relation to a particular abiotic factor or it can give an idea of successions or changes in communities of organisms across a habitat (Figure 11.4.2c).

The type of transect that is used will depend on the terrain and on the organisms present. Sometimes organisms are recorded at specific sampling points at intervals along the tape. Or, a continuous ‘belt’ transect can be used, where all species in a 1 m zone along the transect are recorded. This can be helpful in providing a detailed picture of the area.

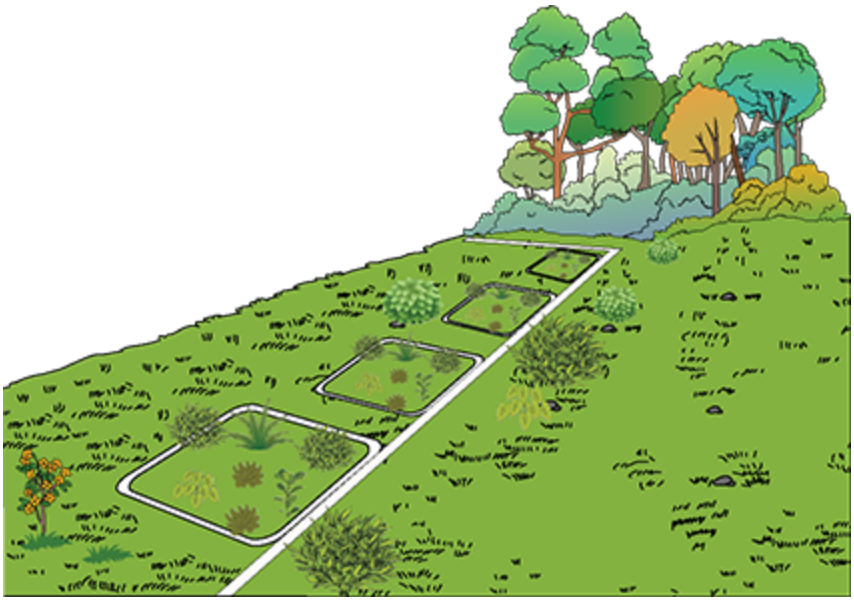


Figure 11.4.2b: Sampling along a transect like this one and measuring light intensity at each sample point enables us to relate distribution of plants to light intensity.

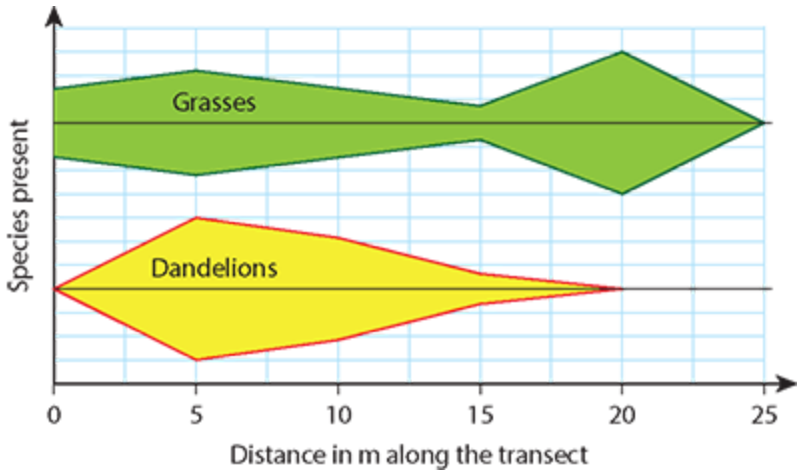


Figure 11.4.2c: This kite diagram of data from grassland shows how kite diagrams are used to present data. It shows that dandelions are more abundant near the starting point of the transect while grasses are present in greater numbers 20 metres along the transect.

NATURE OF SCIENCE

Using models to study the real world: limits of tolerance graphs

Ecologists often use models to predict events in the natural world. Graphs such as the one shown in Figure 11.4.3 can be drawn to indicate the likely ranges of different species in different situations. Information about the stresses, such as temperature, desiccation (drying out through lack of water) or availability of nutrients, that apply to different species can help ecologists to predict whether a species might be able to survive in a habitat or whether the species being studied could survive in a different location with different pressures.

Consider the graph and try to identify the environmental conditions that apply in the case of:

1 The common limpet, a mollusc which lives in an area of the rocky seashore between the high and low tide marks. This area is covered and uncovered twice a day by the tide. Limpets are adapted to being exposed to air and immersed in sea water but their range is limited by their tolerance of several of the conditions higher up on the seashore.

2 The bristlecone pine (*Pinus aristata*) is a conifer native to the USA. It grows in Colorado and other states at altitudes of between 2000 and 4000 metres. It is able to live on exposed, cold, dry rocky slopes and high mountain ridges but its limit of tolerance is 4000 metres, above which environmental conditions are too extreme for it to survive.

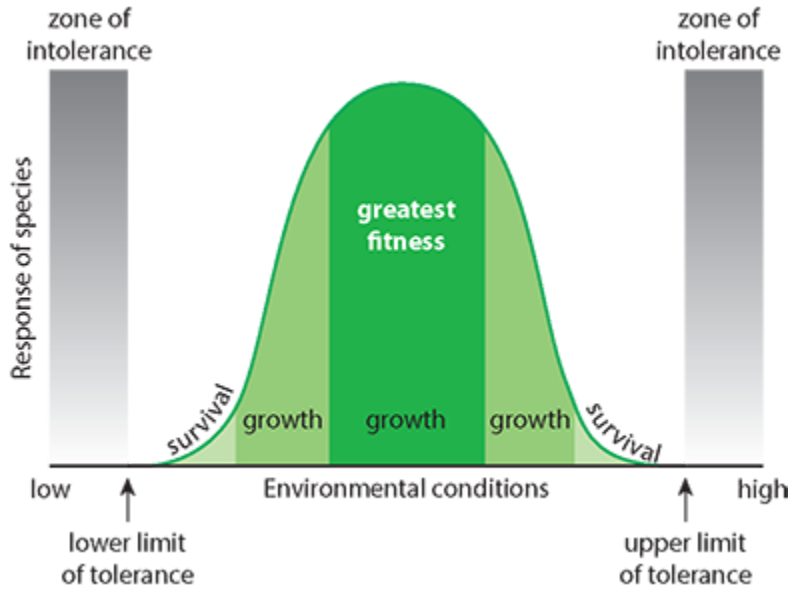


Figure 11.4.3: The graph shows how the range of a species is limited at the upper and lower environmental extremes by zones of intolerance in which the organism cannot survive.

3 For each species, try to describe what conditions might be like in the 'zone of intolerance' where the species are absent.

To consider:

Choose an organism that lives in your home area and try to identify limiting factors that affect its survival. What are the limits of tolerance for the species you have chosen?

11.4.3 Niches and the effects of competition

Organisms interact with other organisms living in the same community. The interactions include competition, different methods of feeding such as being herbivore, a carnivore or a parasite. Almost all organisms influence the lives of others and their interactions all have different effects.

Ecosystems have complex community structures involving interactions between all the species that live in them and are affected by the abiotic factors in the area. We can say that ecosystems show emergent properties. Ecosystems have properties that you would not find in any of the individual species on their own and make up the community structure we can observe.

Competition and competitive exclusion

Competition occurs when two organisms require the same limited resource. For example, if a pride of lions kills an antelope, they must protect this source of food from scavenging hyenas and vultures that will compete with them for the prey. In most cases, competition will lead to the exclusion of one species by another. As one species uses the resource, less is available to the other, so that the less successful species may have to adapt to use a different resource to enable it to survive.

Plants also compete for resources such as light and space. Fast-growing birch trees quickly become established in areas of cleared land, but they require high light levels. Slower-growing species such as oak begin to grow up around them and, for a

while, they form a mixed woodland. Eventually the birch trees are over-shadowed and outcompeted by the more dominant oaks.

Loss of habitat, often caused by human activities such as farming or deforestation, severely limits vital resources such as food, water and breeding sites for the species that live there. When two different species require the same limited resources in the same area, they may find themselves in competition for the same niche. If they are prey species, they may become susceptible to the same predators as well. The principle of competitive exclusion states that no two species can occupy the same niche. The species cannot exist together because one will come to dominate and exclude the other. The oak and birch trees are examples of competitive exclusion. Both compete for soil resources and light but eventually the oak trees block the light from the birches and the birches die out.

KEY POINT

competitive exclusion no two species can occupy the same niche because one will come to dominate and exclude the other.

Fundamental and realised niches

We have described a niche as the special space and 'lifestyle' inhabited by a particular species. This is the **fundamental niche** for a species. It is the potential way of life of the species, given its adaptations. Often the environment will change through natural events, competition or human intervention. So a species may find that its niche becomes more restricted or begins to overlap with that of another species. This more restricted life pattern is known as the **realised niche**.

KEY POINTS

fundamental niche is the potential mode of existence of a species given its adaptations.

realised niche is the actual mode of existence of a species resulting from its adaptations and competition with other species.

The realised niche is the actual mode of existence of a species resulting from its adaptations as well as from competition with other species. A realised niche can only be the same size as or smaller than the fundamental niche.

Gause's study with *Paramecium* (described in the Nature of Science box) showed that the fundamental niche of both *P. aurelia* and *P. caudatum* was the tank in which they grew alone. However, in a tank together each occupied a more restricted, realised niche where *P. caudatum* was outcompeted and failed to thrive as it became limited by *P. aurelia*.

In natural and urban situations, we can observe animals occupying realised niches. For example, normally wild animals such raccoons and foxes have fundamental niches living in open countryside and hunting small mammals, amphibians and other small prey. As humans have encroached on forest and open countryside, turning areas into roads or farmland, the fundamental niche of the fox has reduced. Some prey items have disappeared and the animals find themselves in competition with other species. The new smaller niche is the realised niche. In some areas foxes compete with coyotes, and both racoons and foxes may occupy a realised niche in which they scavenge on the waste left by humans.

NATURE OF SCIENCE

Evidence for competition and niches

In 1934, a famous study on competition was conducted by G.F. Gause (1910–1986), a Russian ecologist. He experimented with two species of *Paramecium*, a large protozoan that is common in fresh water: *P. aurelia* and *P. caudatum*. If the two species were allowed to grow in separate cultures on a food source of bacteria, both species grew well. When the two species were cultured together with an identical food source, *P. aurelia* survived while *P. caudatum* died out (Figure 11.4.4). Both species had similar needs in the culture but *P. aurelia* had an advantage that enabled it to outgrow *P. caudatum*.

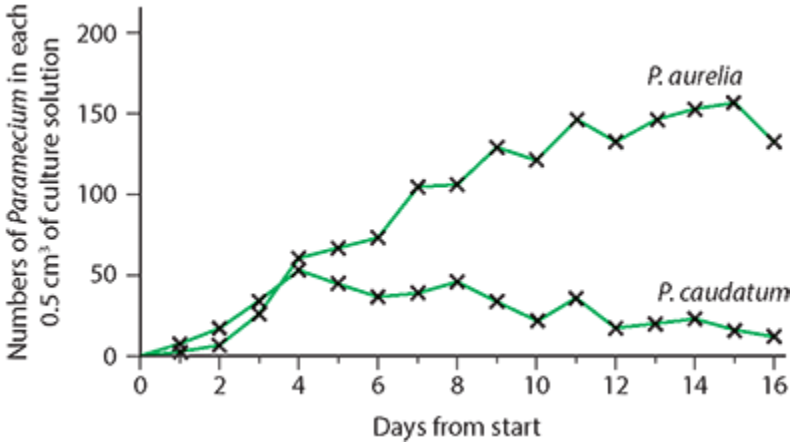


Figure 11.4.4: Over the 16-day culture period, the population of P. aurelia increased while P. caudatum declined. P. caudatum was competitively excluded by P. aurelia.

Another example is a sparrow population living in woodland where the birds feed on berries that grow on bushes. The fundamental niche of the sparrows is the area where there are

berries, and includes the bushes as well as the forest floor where berries fall to the ground. But mice also live in the woodland, where they eat berries that fall to the ground. The presence of the mice causes competition between the two species and means that there are fewer berries on the woodland floor for birds to eat. The fundamental niche of the sparrows is partly occupied, so they fill a smaller, realised niche, which is just the area on the bushes.

11.4.4 Convergent and divergent evolution and changes in structure

Organisms have adapted to fill available niches. In some cases, they have done this through either convergent or divergent evolution. A population that moves to a new location may diverge from its ancestors and form new species. If two or more different species are formed from an original group in different habitats then a process known as divergent evolution takes place. If a large number of new species form from an original ancestor and occupy new niches the process is known as adaptive radiation. Convergent evolution takes place when different organisms that are not closely related ancestors independently evolve similar characteristics as they adapt to similar environments or niches. Table 11.4.1 compares the two types of evolution.

Divergent evolution	Convergent evolution
new species arise from a common ancestor	different ancestors
species diverge and produce homologous structures	species converge and have analogous structures
species appearance becomes more different over time	species appearance becomes more similar over time
species are closely related and have genes in common	species are genetically different and unrelated
examples include the pentadactyl limbs of vertebrates, and Galápagos	examples include wings in birds, bats and insects

finches	

Table 11.4.1: The two types of evolution.

Divergent evolution

One of the most famous examples of divergent evolution was observed by Charles Darwin on the Galápagos Islands (Figure 11.2.7). Darwin noticed that each of the islands had a population of finches that belonged to the same family but that individual bird populations on each island had beaks of different shapes and sizes. Darwin suggested that each species had originally belonged to a single common ancestor species, which had diverged and undergone modifications of its features based on the type of food available on each island. For example, the birds that fed on seeds evolved a short stubby beak, those that fed on nuts evolved a large crushing beak, while cactus eaters developed a longer beak, and insect eaters evolved a finer beak to pick insects out of holes in trees.

KEY POINTS

adaptive radiation divergence of organisms into a range of new forms from a common ancestor under the influence of selection pressure is adaptive radiation.

convergent evolution is when unrelated organisms evolving similar characteristics occupy similar niches.

divergent evolution is diversification of new species from a common ancestor to occupy available niches.

DNA analysis indicates that all 13 species now found on the islands probably evolved from a small flock of about 30 birds that became established there around 1 million years ago. When

the ancestral form of finches colonised the islands, each group contained some individuals who were able to adapt to the conditions and the available food source more readily than others. As the population of birds grew and competition increased, the individuals with favourable characteristics survived and reproduced. Each species now occupies its own niche exploiting its own source of food on each island and this is an example of adaptive radiation (Figure 11.4.5).

Convergent evolution

When convergent evolution takes place, similar phenotypes evolve independently in unrelated species. For example, flight has evolved in both bats and insects; they both have wings, which are adaptations to flight, but the two groups evolved this ability independently and the wings of bats and insects have evolved from very different original structures. Another example of convergent evolution is shark and dolphin body shape. Sharks are fish and dolphins are mammals but over time both populations have been exposed to the same selection pressures. Changes in body shape to make swimming more efficient have been favoured in each group. Structures like wings and body shapes that develop as a result of convergent evolution are known as analogous structures.

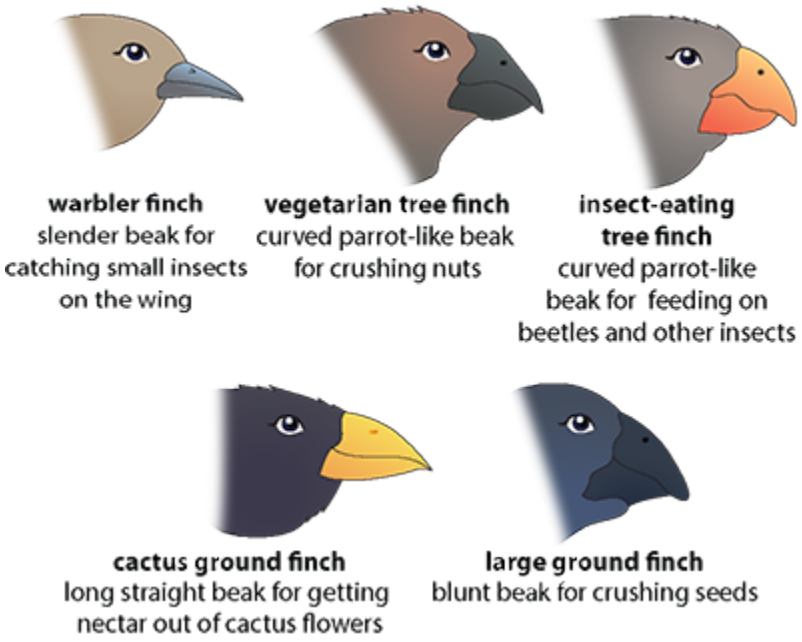


Figure 11.4.5: These diagrams show how the beaks of each of the finches have become adapted so that each bird now occupies its own niche, exploiting its own source of food on the different islands.

KEY POINT

analogous structures are structures that are similar because they have evolved to have the same function not because they are inherited from a common ancestor.

Homologous structures

Homologous structures are anatomical features showing similarities in shape, though not necessarily in function, in different organisms. The evolution of homologous structures by adaptive radiation explains similarities in structure and suggests that the species which have them are closely related and derived from a common ancestor. A good example is the vertebrate pentadactyl limb. This is found in a large range of animals

including bats, whales and humans, as shown in Figure 11.4.6. In each group, limbs have the same general structure and arrangement of bones but each one is adapted for different uses in the different environments that the organisms inhabit. Bird wings and reptile limbs are also homologous structures. Even though a bird uses its wings for flying and reptiles use their limbs for walking, they share a common arrangement of bones.

In many plants, homologous structures are made by modifications of primary leaves, stems or roots to form structures for a range of functions. Leaves are modified to form the insect-trapping pitchers of pitcher plants, the insect traps of the Venus flytrap and the spines of cacti. Modified leaves also form tendrils to grip on to surrounding objects and enable plants to climb upward towards the light. Figure 11.4.7 shows some homologous structures found in plants.

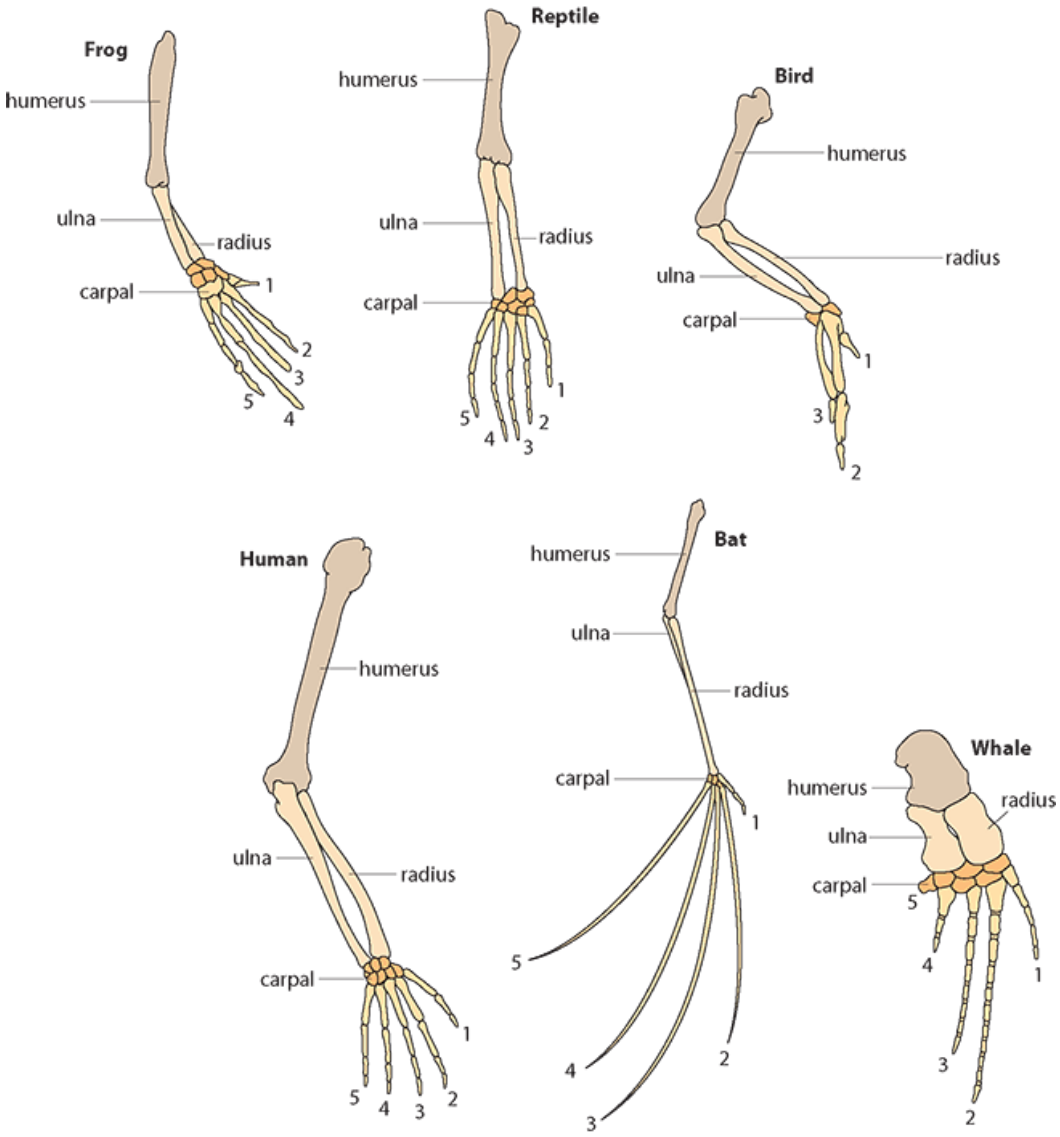


Figure 11.4.6: Homologous structures: the forelimbs of animals with pentadactyl limbs all have a clearly visible humerus, radius, ulna and carpal bones.

Adaptive radiation is a term used to explain how organisms diverge into a range of new forms from a single common ancestor. It can occur if the environment changes and new sources of food or new habitats become available. The pentadactyl limb demonstrates adaptive radiation in the vertebrates, and Darwin's finches are an example of how one

original species adapted to exploit new resources and fill available niches.

Analogous structures

Analogous structures evolve by convergent evolution to fulfil the same functions in very different species. The vertebrate eye is very similar in structure to the eye of the octopus but the octopus and vertebrates are not closely related.

Figure 11.4.7: Homologous structures derived from leaves occur in many plant species.

One key difference is that the octopus eye does not have a blind spot (fovea). Also, unlike the vertebrate eye, an octopus eye is focused by the lens moving backwards and forwards, in a similar way to focusing the lens of a camera. In contrast, a vertebrate eye focuses by changing the shape of the lens using ciliary muscles.

In vertebrate eyes, the nerve fibres to the optic nerve line the inside of the retina and create a blind spot where they pass through the retina. In octopus eyes, the nerve fibres run to the optic nerve from behind the retina, and so do not block light and cause a blind spot (Figure 11.4.8). These differences are all due to the very different ways in which the two types of eye are formed during development. Octopus and human eyes have evolved from different structures, the two organisms do not share a common ancestral structure.

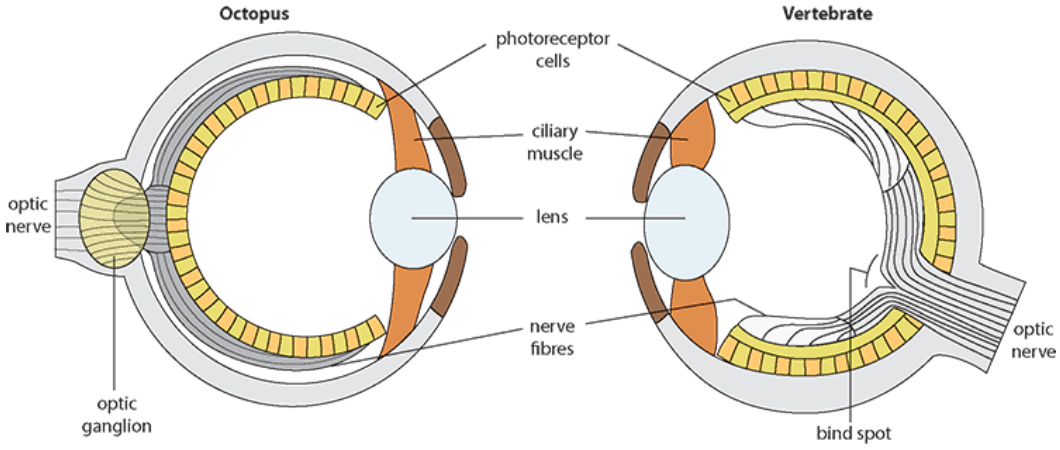


Figure 11.4.8: Analogous structures: the octopus and vertebrate eye have many similarities but have evolved in completely different ways.

Analogous structures are also present in plants and we can see an example of this in the adaptations of the North American cacti and African euphorbias. Both these species have structures that enable them to survive in dry environments. Their features make

the species appear similar but they have evolved from very different plant families. Both plants were typical of the species that lived many millions of years ago on the early Earth; both would have had slim stems and large, wide leaves. As the supercontinents, Pangaea and Gondwana, were separated by continental drift, from about 180 million years ago, plants became isolated on continents that we now know as Africa, America and Australia. Plants on these different continents began to evolve as the changing climate caused the development of arid, desert environments. Long ago the cactus and euphorbia families evolved and adapted to new conditions in order to survive. Despite the fact that they live on separate continents, they have converged to have similar forms and metabolisms because they were exposed to similar environments. Their present-day similarities include branching stems with ribs that run along their length, small leaves and short spines, a spreading shallow root system and succulent stems that can store water (Figure 11.4.9). Table 11.4.2 summarises the differences in the origins of the adaptations in the two plant types.

Cacti	**Euphorbia**
have condensed growth nodes called areoles that produce spines and flowers	no areoles present
spines are modified leaves produced by areoles	spines are modified shoots that grow in pairs from the stem
flowers grow from areoles and have visible petals and stamens. they are often colourful	tiny flowers occur inside a cuplike structure, known as a cyathium
most cacti contain watery sap	euphorbias contain thick,

	milky sap, known as latex

Table 11.4.2: Comparison of the origins of structures in cacti and euphorbias.

Figure 11.4.9: Cacti and euphorbias have evolved by convergent evolution to have similar appearances but are from different plant families.

Table 11.4.3 summarises the differences between homologous and analogous structures in vertebrate limbs.

Homologous structures **Similar structure, different function**		**Analogous structures** **Different structure, similar function**	
lizard forelimb	digging	whale flipper	swimming
bird wing	flying	turtle forelimb	swimming
whale flipper	swimming	fish fin	swimming

human arm	grasping	penguin flipper	swimming

Table 11.4.3: Comparison of homologous and analogous structures in vertebrate limbs.

11.4.5 Evolution and biodiversity

As organism diverge and evolve into new forms from common ancestors, the biodiversity in a community or ecosystem increases. ‘Biodiversity’ simply means the ‘variety of life on Earth’. Adaptive radiation increases the biodiversity in a community or ecosystem ‘biodiversity’ is a term that simply means ‘the variety of life on Earth’. One of the best ways to assess the health of an ecosystem is to measure its biodiversity: the variety of species living in communities there. This can be done using Simpson’s diversity index. This index allows us to quantify the biodiversity of a habitat by taking into account both the number of different species present (the species ‘richness’ of the habitat) and the abundance, or number, of each species.

Diversity can also be assessed on a much smaller scale at the level of genes present in a species. A high genetic variation indicates a large and stable population which has a large pool of genes and maintains the ability to respond to selection pressures if conditions change.

On a much larger scale, diversity can be assessed by looking at the types of ecosystems in an area of the biosphere. Ecosystem diversity considers both biodiversity and abiotic factors that are present. Examples of ecological diversity are variations in ecosystems, such as deserts, wetlands, oceans or forests. This type of diversity is important because it maintains the biological health of the planet. For example, diversity in ecosystems increases oxygen production and absorption of carbon dioxide by photosynthesis. Diversity in aquatic habitats maintains water quality. You can read more about biodiversity and the effects that humans have had on biodiversity in different areas of the world in Chapter 12.5.

KEY POINT

biodiversity a measure of variation at the genetic, species, and ecosystem level.

NATURE OF SCIENCE

Measurement of Biodiversity using Simpson's diversity index

Simpson's diversity index allows us to quantify the biodiversity of a habitat. It takes into account both the number of different species present and the abundance of each species. If a habitat has similar population sizes for each species present, the habitat is said to have 'evenness'.

Simpson's diversity index gives us a measure of both richness and evenness. It is calculated using the formula:

$$D = \frac{N(N-1)}{\Sigma n(n-1)}$$

where:

- D is the diversity index
- N is the total number of organisms in the habitat
- n is the number of individuals of each species
- Σ is a Greek letter that means total, or sum of.

Calculating Simpson's diversity index

The value of Simpson's diversity index is best illustrated by comparing two habitats. Two ponds might contain species of invertebrates in the numbers shown in Table 11.4.4.

Although there are fewer organisms in pond B, the individual populations are more even, so the community is not dominated by one or two species. We conclude that pond B is more biodiverse.

If we want to predict the effect of changes in an ecosystem, we can alter some of the figures we have collected. For example, by increasing the number of water spiders, we could get an idea of what effect an increase in this species might have on the value of *D* and the diversity in the pond.

An advantage of the index is that you do not need to know the name of every different species: it must simply be distinguished as a separate species.

Calculating Simpson's diversity index at intervals over time can give a good indication of the health of an ecosystem and whether conservation measures in some circumstances might be valuable.

	Species					**Total number of organisms**
	Water boatmen	Water measurers	Pond skaters	Whirligig beetles	Water spiders	
number of organisms in pond A	43	18	38	3	1	103
number of organisms in pond B	26	18	29	11	5	89

Table 11.4.4: The numbers of invertebrates in two ponds, A and B.

WORKED EXAMPLE 11.4.1

Using the formula, calculate the diversity index for pond A and pond B:

Answer

For pond A:

Simpson's diversity index D

$$= \frac{(103 \times 102)}{43(43-1) + 18(18-1) + 38(38-1) + 3(3-1) + 1(1-1)}$$
$$= \frac{10\ 506}{3524}$$
$$= 2.98$$

For pond B:

Simpson's diversity index D

$$= \frac{(89 \times 88)}{26(26-1) + 18(18-1) + 29(29-1) + 11(11-1) + 5(5-1)}$$
$$= \frac{7832}{1898}$$
$$= 4.13$$

TEST YOUR UNDERSTANDING

22 Define the term niche.

23 Distinguish between a fundamental and a realised niche.

24 Distinguish between convergent and divergent evolution.

25 Give an example of homologous structures.

INTERNATIONAL MINDEDNESS

International cooperation and collaboration

Conserving biodiversity requires international cooperation between scientists, organisations and politicians. In the last 50 years, the importance of biodiversity has come to the forefront of science. Species are not evenly distributed on Earth. Biodiversity is far richer around the tropics and areas containing rainforest are among the most diverse on the planet. People have come to realise that there are many compelling reasons for conserving the biodiversity of habitats such as the rainforests, where as-yet-undiscovered species may provide valuable medicines and other resources for future generations. Conservation in one part of the world may depend on cooperation and collaboration in another. International organisations such as the World Wide Fund for Nature (WWF) and the United Nations Environment Programme (UNEP) coordinate such work in many countries.

The key objective of all conservation organisations is to preserve species and their habitats. Some work at a local level while others are global. Some organisations, such as UNEP, are funded by governments while others, such as WWF, are non-governmental organisations (or NGOs), which are funded by individuals or groups. Organisations

such as WWF work with businesses, governments and local communities to create solutions that take account of the needs of both people and nature. Conservation programmes must

select which species are to be protected, but it is often difficult to decide which species most merit conservation efforts.

To consider:

1. On what basis should one species be chosen over another? For example, is a large mammal such as a tiger or panda more important than a small, seemingly insignificant mollusc? An endearing mammal may encourage people to support a conservation programme but smaller, less appealing species may be more important and play a pivotal role in an ecosystem.
2. Should endangered animals be given priority over other species whose numbers are not yet so low?
3. How can international cooperation help less wealthy countries conserve the biodiversity in their regions?

11.4.6 Competition in identical niches

The competitive exclusion principle explains that two species cannot have exactly the same niche and coexist because species with identical niches also have identical needs and compete for all the same resources. G.F. Gause's experiments with *Paramecium* show us that two species cannot survive indefinitely in the same habitat if their niches are exactly the same. There will be competition between them. Competition occurs when two organisms require the same limited resource. For example, if a pride of lions kills an antelope, they must protect this source of food from scavenging hyenas and vultures that will compete with them for the prey. In most cases, competition will lead to the exclusion of one species by another. As one uses the resource, less is available to the other, so that the less successful species may have to adapt to use some other resource if it is going to survive. Species such as lions and hyenas have adapted so that their niches only partly overlap. They may be able to coexist if food or other resources are plentiful. But over long periods of time, species may adapt to make use of more different, or less overlapping resources.

In the 1960s the American ecologist Joseph Connell (1923–2020) tested Gause's ideas on competition by studying barnacles (shelled marine organisms) that live on rocks along European coastlines and need similar resources. He found that two barnacle species, *Balanus* and *Chthamalus*, can coexist because of two important differences between them. Firstly they grow at different rates, and secondly they have a different tolerance to dry conditions. *Balanus* grows rapidly, which allows it to grow over and cover the slower-growing *Chthamalus* if both species are present. But *Balanus* dies in areas close to shore because it

cannot tolerate the dry conditions that occur at low tide. *Chthamalus* tolerates these dry conditions well. Although *Balanus* is a better competitor for space, the two species of barnacle coexist in the same areas because *Chthamalus* can survive in dry conditions where *Balanus* cannot (Figure 11.4.10). This example supports the competitive exclusion principle. Species can only coexist if they have different niches.

When ecologist Joseph Connell did this he found that when *Chthamalus* was alone, the species occupied all the rocks between the high and low tide marks (its fundamental niche). When the two species were present together, both occupied their smaller, realised niches.

Loss of habitat, often caused by human activities such as farming or deforestation, may limit vital resources such as food, water and breeding sites for the species that live there. As we have seen, when two different species require the same limited resources in the same area, they may find themselves in direct competition for the same niche. If they are prey species, they may become susceptible to the same predators as well. Since the principle of competitive exclusion states that no two species can occupy the same niche, some species avoid competition and predators by adapting to extreme niches.

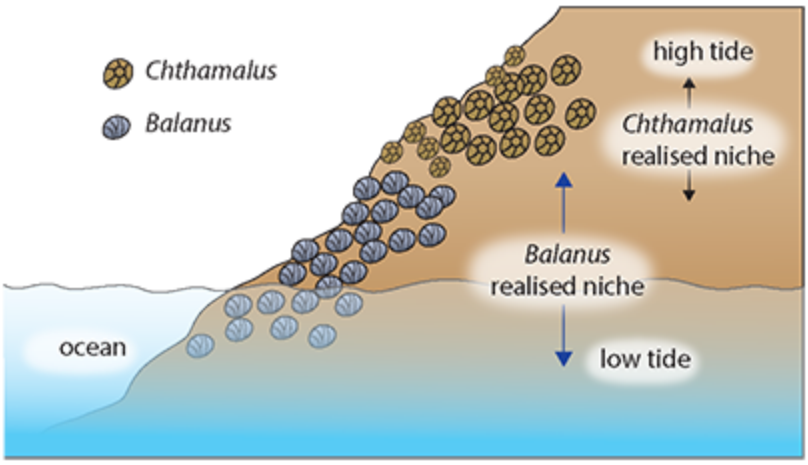


Figure 11.4.10: Connell experimented by removing Balanus barnacles from rocks.

11.4.7 Adaptations to different niches

All organisms are adapted to the conditions in their niche. Organisms that can adapt to extreme niches encounter less stress from competition and predation because few organisms live in these difficult environments. With fewer competitors the adapted species can thrive. You can learn more about the abiotic conditions in different biomes in Section 12.4.

Hot deserts

Organisms that can survive in hot deserts have adaptations that enable them to live where temperatures are very high and water is scarce.

Desert plants such as cacti and other succulent plants have adaptations that we can observe are:

- The plants appear swollen, spiny and have no leaves.
- They store water in fleshy stems or roots.
- They can absorb large quantities of water in the very short periods when desert rain falls.
- Most have extensive, shallow root systems to absorb water close to the surface.
- Waxy cuticles prevent water loss when stomata are closed.

Desert animals such as the camel also have a range of adaptations:

- Camels have few or no sweat glands.
- They produce tiny amounts of very concentrated urine.

- They store fat reserves in humps rather than as an insulating layer around their body.
- They can drink up to 100 litres of water when it is available.
- Broad lips allow them to feed on dry thorny vegetation.
- Large flat feet spread their weight on soft sand and protect them from the heat.
- Their nostrils can close and their eyelashes are long to prevent sand blowing into them.

Tropical rainforests

In tropical rainforests conditions are hot and humid but rainfall is very high. It is an environment which has many species but in which there is great competition for the natural resources. Soil in the forest is shallow and not very fertile as nutrients cycle rapidly and are quickly eroded in the heavy rain. Plants compete for light and nutrients.

Adaptations that allow plants such as the woody vine lianas to survive include:

- roots in the shallow rainforest soil
- long woody stems that climb up trees to reach sunlight for photosynthesis
- leaves and flowers only in the high canopy layers.

Epiphytes are plants which live high in the canopy. They have no contact with the ground and get all their nutrients from the air and water, not from the soil.

Trees are also adapted to the conditions. They have tall smooth trunks which grow rapidly to reach the light. Many have smooth trunks which allow heavy rain to run quickly down to the tree roots. Large buttress roots support tall trees and extend into the shallow soil. (Fig 11.4.11)

Although it covers extensive areas, the tropical rainforest contains very few large animals because its understory (the layer above the forest floor) is so dense that it makes it hard for them to move around. Animals avoid predators using camouflage and hiding from view. For example, the sloth has a colour that matches the tree trunks and branches, and it moves extremely slowly so that predators do not notice it. Much of the food that is in the forest grows high up in the branches so rainforest monkeys have long, strong limbs to move easily between the trees, while tree frogs have webbed hands and feet so that they can glide between trees. Birds can fly up into the trees, but the branches may be too weak to support their weight. Toucans have long strong beaks to reach and cut fruit from thin branches.

Figure 11.4.11: Buttresses support rainforest trees

Some species develop physiological, morphological or behavioural adaptations to extreme environments that other species cannot inhabit. Two examples of such species are emperor penguins (*Aptenodytes fosteri*), which live on the ice in the frozen waters of the Antarctic, and marram grass (*Ammophila arenaria*), which grows on coastal sand dunes where water is difficult to obtain.

Freezing conditions (*Emperor penguins*)

Emperor penguins are the largest of the Earth's penguins and are 115 cm tall.

The birds have physiological adaptations and cooperative behaviour strategies to survive in an environment where the temperature can reach −60°C.

- They huddle together in groups to protect themselves from the wind and conserve heat. Individuals take turns moving to the centre of the huddle. When they have warmed their bodies they change places with other birds and move to the edge of the group.
- Emperor penguins breed on the open ice. A females lays one egg and then leaves it with the male bird while they go to hunt in the ocean for period of up to 2 months. They may need to travel 40 km across the ice to reach the ocean where they feed on fish and squid.
- The male bird remains on the ice and incubates the egg, standing upright and keeping the egg on his feet. A brood pouch made of feather-covered skin encloses the egg. A male does not feed at all for the 2 months that he incubates the egg and guards the chick (Figure 11.4.12).

Figure 11.4.12: The male emperor penguin cares for its chick for up to 2 months until the female returns to feed it.

- The female returns to feed their newly hatched chick and the male leaves to feed. Another adaptation is the penguins' ability to dive down to 500 metres, deeper than any other birds, and stay underwater for up to 20 minutes.

- In the Antarctic summer (December), when the ice begins to melt, the chicks are able to reach open water and swim and fish for themselves.

Dry conditions (*Marram grass*)

Marram grass grows on mobile, coastal sand dunes and is one of the few species that can survive there. The plant is known as a xerophyte and has physiological adaptations to survive shortages of water. The conditions on a sand dune are harsh, with strong winds which carry salt spray from the sea. Sand dunes drain quickly and bare sand has few nutrients. The terrain is composed of calcium carbonate from seashells and rotting seaweed on the sand adds a few nutrients. Few species can survive here but marram grass has adaptations to the harsh conditions.

- Marram has deep, matted roots which reach the water table and bind the sand together.
- Marram has reduced rates of transpiration compared to other plants. They have protected stomata situated deep inside the plant's waxy leaves. The leaves are rolled up to prevent evaporation from the surface, with tiny hairs that minimise air flow, trap water vapour and prevent water being carried away.

You can read more about the detailed structure and adaptations of marram grass in Section 6.3.

Figure 11.4.13: Marram grass is a species which stabilises the shifting sand of a dune.

SCIENCE IN CONTEXT

Marram grass was once harvested and used to weave mats for barn roofs, nets for fishing and even shoes. Long ago, a family that lived near the sea would have its own sand dune and the whole village would often be involved in collecting the grass. Today it is not harvested because of its the importance in stabilising fragile sand dune habitats.

TEST YOUR UNDERSTANDING

26 Why can two species with identical niches not survive indefinitely?

27 Outline the behavioural adaptations of the emperor penguin that enable it to survive in an extreme habitat.

28 Why do organisms in extreme niches have less stress from competition?

Links

- What are the mechanisms that allow species to adapt to their niches? (Section 11.3)
- What interspecific interactions limit a species to its realised niche? (Section 11.2)
- What are the roles of heterotrophs in recycling nutrients in an ecosystem (Section 12.3)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
list the hierarchy of taxa used for classification	11.1.1			
explain why scientists use a binomial name in the identification of species	11.1.1			
recall that organisms are classified into three domains based on rRNA evidence	11.1.2			
design a key to identify different organisms	11.1.2			
define a clade and understand	11.1.3			

diagrams of phylogeny to explain evolutionary relationships				
outline the reasons for using DNA or amino acid sequences to establish relationships	11.1.4			
draw and interpret cladograms	11.1.5			
outline reasons for a positive correlation between differences in two species and the time they diverged from a common ancestor	11.1.4, 11.1.5			
describe how natural selection can occur and how variation, adaptations, reproduction and sexual selection are important	11.2.1			
summarise how	11.2.2			

human impact provides evidence for evolution by exposing populations to pesticides, pollutants and antibiotics				
explain how selective breeding provides evidence for evolution	11.2.3			
define a gene pool	11.2.4			
outline directional, disruptive and stabilising selection	11.2.5			
use the Hardy–Weinberg theory to compare allele frequencies	11.2.6			
define a species and state that all species have originated from a common ancestor by evolution	11.3.1			
define evolution	11.3.1			
summarise how	11.3.2			

the evidence for evolution is gathered from fossils, observation and DNA analysis				
explain how populations can diverge into new species by natural selection	11.3.3			
explain the importance of isolation for speciation and identify three types of isolation	11.3.3			
outline the importance of sterile hybrids to understanding reproductive isolation	11.3.3			
define a gene pool	11.3.4			
outline why evolution requires a change in allele frequencies in populations	11.3.4			
	11.3.4			

give examples of the mechanisms of geographic, temporal and behavioural isolation				
outline how changes in allele frequency in gene pool can arise as a result of natural selection	11.3.4			
summarise how speciation can be gradual or abrupt	11.3.4			
define a niche and a habitat	11.4.1			
recognise that all species have tolerance limits that define their niche	11.4.2			
outline how species distribution data and abiotic factors can be correlated using a transect	11.4.2, 11.4.7			
distinguish between a	11.4.3			

fundamental and realised niche				
distinguish between divergent and convergent evolution and name examples of each	11.4.4			
distinguish between homologous and analogous structures and give examples	11.4.4			
state that adaptive radiation increases biodiversity in a community	11.4.5			
describe why two species cannot survive together in identical niches.	11.4.6			

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Chapter 12

Ecological relationships

C4.1, C4.2, A4.2, D4.2, D4.3

INTRODUCTION

Almost the entire surface of the Earth – the land, rivers, lakes, seas and oceans – is home to organisms of one kind or another. It has been estimated that there are as many as 10 million different species on Earth and understanding where and how they live and interact is a branch of biology known as ecology. Humans are not the most numerous species on Earth: there are many more bacteria and insects. But humankind is having a disproportionate effect on the world's ecosystems as damage is caused by pollution, rainforest destruction and global warming.

12.1 Modes of nutrition

LEARNING OBJECTIVES

In this section you will:

- recall that organisms need energy to drive their metabolism
- understand that organisms may be obligate anaerobes, facultative anaerobes or obligate aerobes
- distinguish between autotrophs that obtain energy from inorganic nutrients and heterotrophs that obtain energy and nutrients from other organisms
- learn that a few organisms obtain nutrition in both ways
- distinguish between consumers, detritivores and saprotrophs
- define the term trophic level as a way of classifying organisms by their feeding relationships
- learn that consumers may be primary, secondary or tertiary consumers or omnivores
- recognise that trophic feeding relationships simplify the complexity of an ecosystem
- learn that some autotrophs are chemosynthetic rather than photosynthetic
- understand that omnivores are species which feed at more than one trophic level

- identify some adaptations that herbivores, carnivores and plants have to feed

GUIDING QUESTIONS

- How do living organisms obtain the energy they need?
- How do modes of nutrition affect the interactions of organisms in ecosystems?

12.1.1 Feeding groups

All organisms need to acquire energy to drive their metabolism. This enables them to grow, reproduce and function in their environment. Species are divided into groups that are defined by their method of obtaining food. Different organisms feed in different ways and feeding interactions influence the growth, survival and reproduction of all species. A niche is the role a species has in an ecosystem, it includes all the biotic and abiotic interactions that influence a species (Chapter 11.4).

Autotrophs and heterotrophs

Autotrophs are species that are able to make their own food from basic inorganic materials. This group includes all plants that can photosynthesise, as well as mosses, ferns, seaweed, unicellular algae and purple and blue-green bacteria. Autotrophs (which means 'self feeding') use light energy to synthesise sugars, amino acids, lipids and vitamins, using simple inorganic substances such as water, carbon dioxide and minerals. Heterotrophs are consumer species that obtain their food from organic matter. **Heterotrophs** obtain both energy and nutrients such as minerals and vitamins from other organisms.

KEY POINTS

autotroph is an organism that produces complex organic compounds from simple inorganic molecules, usually by photosynthesis.

a heterotroph that feeds on living organisms by ingestion is a consumer.

a heterotroph that feeds on organic nutrients from dead organisms by internal digestion is a detritivore.

ecology is the study of the relationships between living organisms and their environment, including both the physical environment and the other organisms that live in it.

heterotroph is an organism that obtains energy and nutrients from other organisms.

a heterotroph that feeds on organic nutrients from dead organisms by external digestion and absorption is a saprotroph. Saprotrophs are also known as decomposers.

This group includes herbivorous and carnivorous animals, which feed on living organisms. Methods of feeding are used to explain the relationships of organisms within an ecosystem, as shown in Figure 12.1.1.

Two important groups of heterotrophs are detritivores and saprotrophs, which feed on dead organic matter. These organisms are vital to the well-being of any ecosystem because of their recycling role. When an organism dies, the remains of its body provide nutrients for detritivores and saprotrophs, which feed on them in different ways.

The two groups of heterotrophs are distinguished by their methods of digestion.

Consumers feed by **holozoic** nutrition which means their food is ingested (eaten), digested inside their bodies then absorbed, and assimilated. Undigested material is egested and leaves from the end of the gut.

Saprotrophic nutrition is the digestion of food outside the body followed by the absorption of already digested materials. Fungi and bacteria, together known as decomposers use this method to feed on dead and waste material. These organisms are important for the recycling of nutrients in an ecosystem.

Mixotrophic organisms such as *Euglena* (Figure 12.1.1) are both heterotrophic and autotrophic. *Euglena* has chloroplasts and photosynthesises when there is sufficient light but can also feed on small organisms by endocytosis. *Euglena* cannot be classified as a plant or an animal and is placed in the Kingdom Protista.

- Detritivores are organisms such as earthworms, woodlice and millipedes that ingest dead organic matter such as fallen leaves or the bodies of dead animals (Figure 12.1.2). They digest the organic matter inside their bodies.

- Saprotrophs – which include bacteria and fungi – secrete digestive enzymes onto organic matter and then absorb their nutrients in a digested form (Figure 12.1.3). Saprotrophs are responsible for the decomposition of organic matter and are sometimes referred to as decomposers. Saprotrophic bacteria and fungi are vital for most ecosystems and are crucial to the recycling of inorganic nutrients such as nitrogen compounds. Recycled inorganic compounds can be re-used over and over again by autotrophs, which in turn continue to grow and provide food for heterotrophic consumers.

If there is enough sunlight to provide energy for the autotrophs to photosynthesise, and the community itself can maintain the recycling of inorganic materials within the abiotic environment, an ecosystem like this has the potential to remain stable and self-sustaining for a long period of time. This happens for as long as there are no adverse interferences from outside. Catastrophic

natural events and human interference are two factors that can disrupt otherwise stable ecosystems. It is important to remember that sustainable, stable systems are vital for the continued survival of all species, including our own.

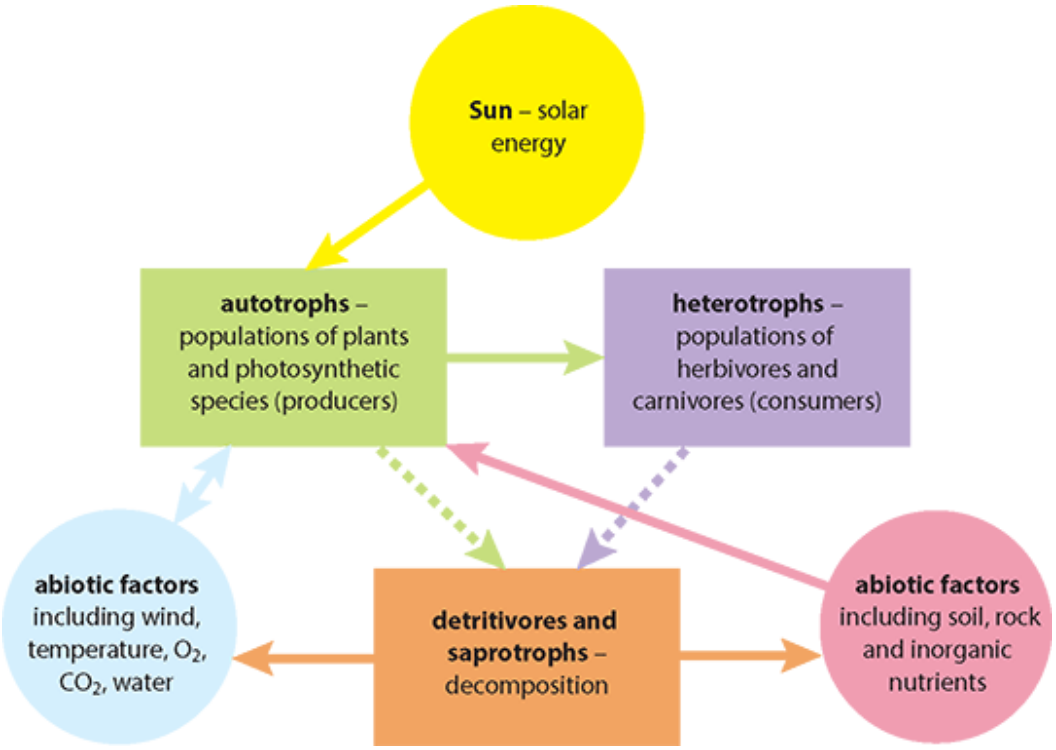


Figure 12.1.1: The feeding relationships in an ecosystem. A few organisms are both autotrophs and heterotrophs. One example is the aquatic species *Euglena*.

NATURE OF SCIENCE

Looking for trends and discrepancies: are all plants and algae autotrophic?

Autotrophs or 'self-feeders' are organisms that produce complex organic compounds from simple substances in their environment. In the majority of cases, autotrophs are plants and algae, which use light from the Sun as a source of energy.

One organism that is neither plant nor alga, but which was once classified with the plants because it can photosynthesise, is *Euglena*. *Euglena* is a unicellular organism with chloroplasts in its cell that enable it to feed autotrophically, like a plant. But *Euglena* can also feed heterotrophically, taking in organic materials as animals do. When *Euglena* was first discovered, organisms were classified into just two kingdoms – the plant kingdom and the animal kingdom – and *Euglena* was impossible to place. Then, in the 19th century, Ernst Haeckel added a third kingdom, which he called the Protista, to accommodate organisms like *Euglena* that display characteristics of both plants and animals.

Figure 12.1.2: Millipedes feed on dead leaves on the forest floor and recycle the nutrients they contain via their feces. They are detritivores.

Chemosynthetic autotrophs

Some autotrophs are chemosynthetic rather than photosynthetic. Like all autotrophs chemoautotrophs make their own food but need a source of energy to do so. They get this energy not from

the sun, but from the chemical bonds of inorganic molecules. *Nitrosomonas* is a nitrogen-fixing bacterium found in soil. It obtains energy from the oxidation of ammonia. Using this energy, it can synthesise glucose from carbon dioxide. Many microorganisms in dark regions of the oceans, where light is unavailable, also use chemosynthesis to produce biomass from carbon molecules. In most cases the energy for chemosynthesis comes from the oxidation of hydrogen sulfide or ammonia.

Figure 12.1.3: Fungi secrete enzymes onto dead material and absorb digested material into their cells. They are saprotrophs.

Diversity of nutrition in Archaea

Archaea is a domain of single-celled organisms which have no cell nuclei and so are prokaryotes but they do share some characteristics with eukaryotes. Archaea were initially classified as bacteria but analysis of genetic material has identified that they are a separate domain. Archaea are metabolically diverse; some can photosynthesise but do not use chlorophyll to capture light. Others are chemoautotrophs and obtain their energy for ATP production by breaking down molecules in their environment. Archaea can live on a huge range of energy sources: ammonia, metal ions, even hydrogen gas. Some salt-tolerant types found in salt lakes use sunlight as an energy source, and others can fix carbon from the atmosphere. Some are adapted to life in hot springs and thermal vents. Table 12.1.1 summarises the range of nutritional types in this domain.

KEY POINT

chemoautotroph refers to an organism which uses energy from chemical reactions to generate ATP and produce organic compounds from inorganic substances.

Type of nutrition	Source of energy	Source of carbon
Phototrophs	Light	Organic compounds
Chemotrophs (Lithotrophs)	Oxidation of inorganic compounds, e.g. sulphur, ammonia, methane, iron	Organic compounds or carbon fixation
Heterotrophs	Oxidation of organic compounds from other organisms	Organic compounds or carbon fixation

Table 12.1.1: Modes of nutrition in Archaea.

Trophic levels and feeding relationships

Every organism needs food to survive but eventually it too will be eaten. In any ecosystem, there is a hierarchy of feeding relationships that influences how nutrients and energy pass through it. The sequence of organisms that provide food for one another is known as a **food chain**.

Green plants are autotrophs, which start food chains because they are able to capture light energy from the Sun. Plants are called sometimes called producers because they 'produce' organic compounds by photosynthesis. These organic compounds contain chemical energy that has been converted from light energy by the process of photosynthesis. Every other organism in a food chain obtains organic compounds from its food and so is called a heterotroph or consumer.

Every ecosystem has a structure that biologists divide into categories on the basis of the food sources of the different organisms. These categories are known as trophic levels. Trophic means 'feeding' and every organism in a food chain can be placed in a particular feeding category or trophic level.

Green plants are producers and are placed at the first trophic level. Primary production is the accumulation of carbon compounds in biomass by autotrophs. After them come all the consumer levels. The first consumers, or primary consumers, are always herbivores. Any organism that feeds on herbivores will be a carnivore and can be listed as secondary consumer, tertiary consumer and so on depending on its food source. Secondary production is the accumulation of biomass by heterotrophs who

are secondary consumers. A food chain can therefore be summarised as shown in Figure 12.1.4.

Figure 12.1.4: A food chain consists of a producer and consumers.

Every organism fits somewhere in a food chain, and although the organisms that make up the food chain will vary from place to place, almost every food chain starts with a green plant. It may be any part of the plant, for example the leaves, roots, stems, fruits, flowers or nectar.

Figure 12.1.5 shows three examples of food chains from different ecosystems. Notice that the arrows in a food chain always point in the direction in which the energy and nutrients flow. Each of these food chains contains an autotroph at the start of the chain, followed by primary, secondary and tertiary consumers.

KEY POINTS

primary consumer is an organism that feeds on autotrophs (plants).

secondary consumer is a heterotroph that feeds on primary consumers.

tertiary consumer is a heterotroph that feeds on secondary consumers.

trophic level is the position of an organism in a food chain.

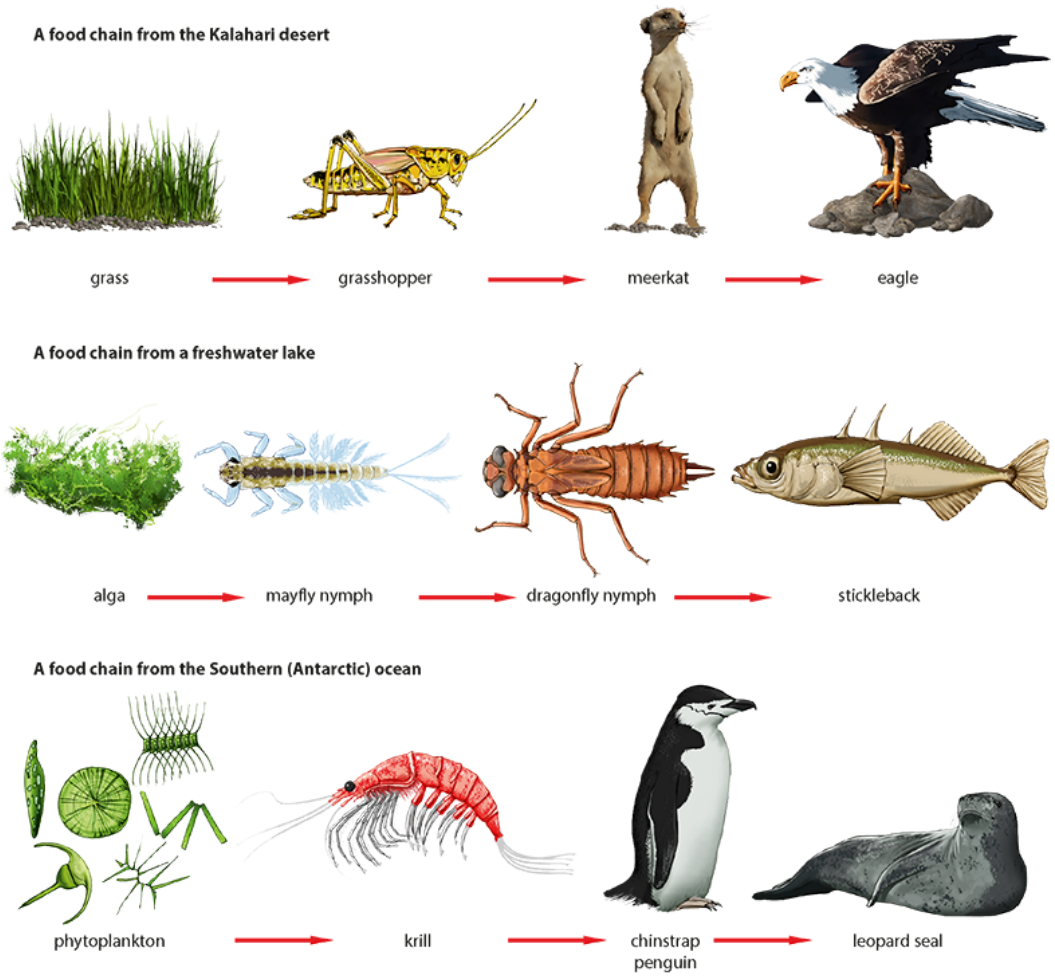


Figure 12.1.5: Grass, algae and phytoplankton are all examples of photosynthesising autotrophs, which use light as their source of energy. Almost all food chains start with light as the initial source of energy. Grasshoppers, mayfly nymphs and krill are primary consumers. Meerkats, dragonfly nymphs and chinstrap penguins are secondary consumers.

When these organisms die, they will provide nutrients for a series of detritivores and saprotrophs, which will return inorganic nutrients to the abiotic environment, so that they can be re-used by autotrophic organisms.

The food chains shown in Figure 12.1.5 are taken from different ecosystems and continents but you can construct some food chains using organisms in your local area. You might use your school campus, a garden or local park or nature reserve. Try to separate the organisms you can see into groups. Identify the autotrophs and heterotrophs and classify the consumers into primary, secondary or tertiary categories.

SCIENCE IN CONTEXT

Spraying with DDT insecticide: unintended consequences

In the 1950s the World Health Organization (WHO) sprayed parts of an island in Borneo with DDT insecticide to eradicate the malarial mosquito. The treatment was successful and a malaria epidemic was halted. But there were other results that no one had predicted.

The huts in the villages had thatched roofs which were eaten by a particular species of caterpillar. The caterpillar was not affected by DDT but the wasp that was its natural predator was wiped out. Soon the thatched roofs began to collapse. Insect-eating geckos and the local gecko-eating cats also died of DDT poisoning. The roofs of the huts were replaced with corrugated iron but there was soon a big increase in the number of rats which threatened to spread diseases in the village. The WHO solution was to airlift in new cats and drop them into the village by parachute.

To consider:

Use your knowledge of food chains to explain these events on the island.

1 Why did the thatched roof collapse?

2 DDT is an insecticide so what caused the death of the cats?

3 Why did the rat population increase and how did the new cats help the villagers?

4 You can read more about the effects of DDT on food chains in Section 12.4.

TEST YOUR UNDERSTANDING

1 Distinguish between autotrophs and heterotrophs.

2 Describe the feeding methods of a detritivore and a saprotroph.

3 How is the trophic level of an organism decided?

4 Divide these organisms into producers and consumers.

leaves snails grass pond algae

mice owl earthworm

5 Look at this diagram which shows feeding relationships on a seashore.

How many primary consumers are there?

How many secondary consumers are there?

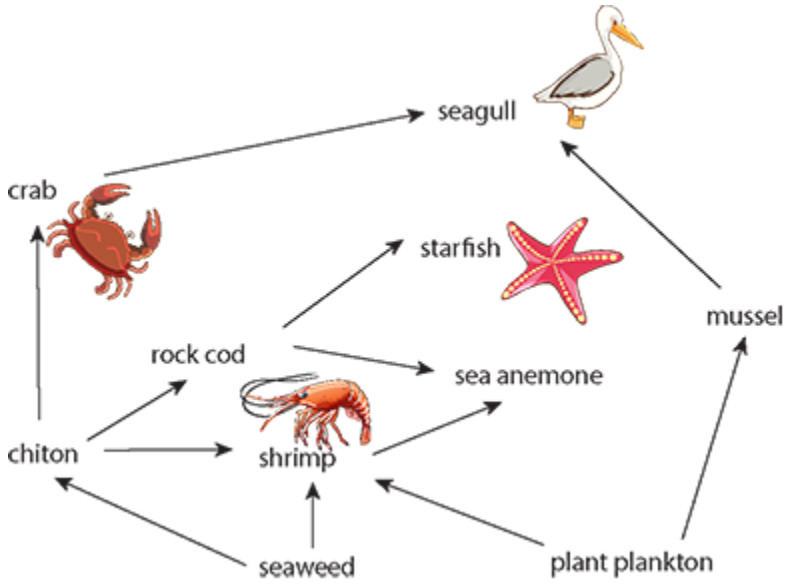


Figure 12.1.6: Feeding relationships on a seashore.

12.1.2 Complexities in feeding relationships

The concept of trophic levels is a quick and easy way to get an idea of relationships in an ecosystem but few consumers feed on only one source of food. For example, this food chain describes one set of feeding relationships:

grass → beetle → robin → sparrowhawk

But beetles eat a wide range of plants, robins eat other types of insect and sparrowhawks eat other birds. So this food chain could be interlinked with many others. A **food web** like the one shown in Figure 12.1.7 shows a much more realistic picture of the feeding relations of the organisms in a habitat. Notice how organisms change trophic levels depending on what they are eating at any particular time. In Figure 12.1.7, the fox is a primary consumer when it is eating a crab apple but a secondary consumer when it is eating a wood mouse.

KEY POINTS

omnivore an animal that is able to eat both plants and animals.

niche is the role a species has in an ecosystem, it includes all the biotic and abiotic interactions that influence a species.

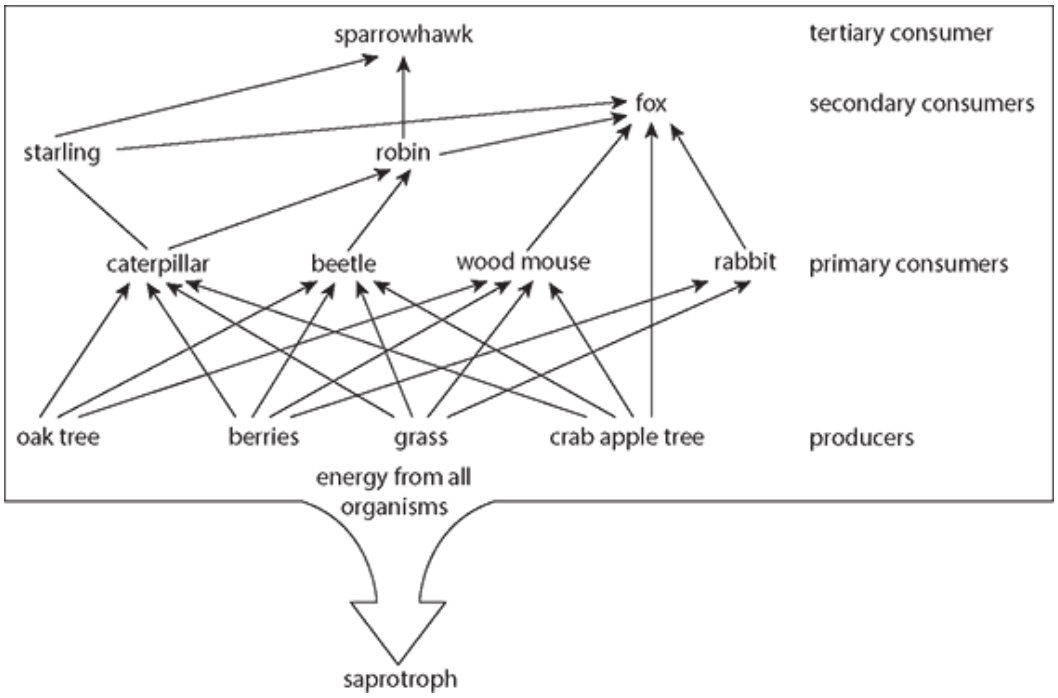


Figure 12.1.7: Many species can be defined as omnivores because they feed at more than one trophic level.

SCIENCE IN CONTEXT

Generalists and specialists

Classifying a species as a generalist or a specialist is a way to identify what kinds of food and habitat resources it relies on to survive. Generalists can eat a variety of foods and thrive in a range of habitats. Specialists, conversely, have a limited diet and habitat requirements.

Raccoons (*Procyon lotor*) are a good example of a generalist species. They live in a range of different environments in North America, including forests, mountains and large cities. They are omnivores and can eat fruit and nuts, insects, frogs, eggs and waste left by humans, and can adapt to whatever food source is available.

In contrast, the Australian koala (*Phascolarctos cinereus*) Figure 12.1.9 is a specialist. It is a herbivorous marsupial (pouched mammal) that feeds only on the leaves of the eucalyptus tree and so can only live in habitats where eucalyptus trees grow.

Figure 12.1.8: Koalas are specialists and feed exclusively on the leaves of eucalyptus trees.

Trophic feeding relationships do not take account of other complexities in ecosystems and they simplify interactions between species. An organism may feed at more than one trophic level, change its diet through its lifecycle or its interactions may be affected by abiotic factors or interference. All food chains and webs have two or three trophic levels but most do not have more than four.

The Canada lynx (*Lynx canadensis*), is a specialist secondary consumer which preys almost exclusively on snowshoe hares. Both these animals live in the forested, mountainous areas and tundra where lynx are very well-adapted to survive. Their padded paws enable them to walk in deep, soft snow as they hunt.

Specialist species like koalas and the lynx have evolved to fit a very specific **niche** and can easily be affected by environmental disturbances, such as climate change or habitat loss due to forest fires. Specialists are more likely to be affected by these disruptions because they cannot use other food sources or habitats in the way that generalists can. Some specialist species are in decline due to human activity and interference whereas the number of generalist species is increasing.

Other feeding relationships

Parasitism is a symbiotic feeding relationship between species, where one organism, the **parasite**, lives on or inside another organism, the **host**, causing it some harm, and is adapted structurally to this way of life.

Examples of internal parasites are tapeworms and *Plasmodium*.

- Tapeworms are segmented flatworms that attach themselves to the insides of the intestines of animals such as cows, pigs and humans. They get food by eating the host's partly digested food, depriving the host of nutrients.
- *Plasmodium* is a unicellular eukaryote that causes malaria in humans. It reproduces in liver cells and red blood cells. Part of the parasite's life cycle takes place in the *Anopheles* mosquito, which is the vector that transmits the parasite from one person to another as it feeds on blood.

Examples of external parasites are fleas and leeches.

- Fleas (Figure 12.1.9) feed on the blood of mammals and birds. Their bodies are adapted and flattened so they can move through fur and feathers of their hosts and claws prevent them being removed. They can be vectors of other

diseases and cause localised infections where their bites cause irritation to their host.

Figure 12.1.9: A flea in the fur of a dog.

- Leeches live in ponds and feed on the blood of mammals. They puncture the skin of their host and secrete an enzyme that stops the blood from clotting. They may ingest more than their body weight in blood and leave bleeding wounds on their hosts' bodies.

12.1.3 Adaptations for feeding

Organisms are adapted for their modes of nutrition and their adaptations may either be for capturing their food, or for ingesting and digesting it. We can draw inferences about the diets of extinct species or newly discovered organisms by looking for these adaptations.

1 Adaptations in the teeth of Hominidae

Hominidae is a group which includes all modern and extinct great apes which are modern humans, chimpanzees, gorillas and orang-utans plus all their ancestors. Some members of the group are herbivores while others are omnivores and have some meat in their diet. The teeth and skulls of living hominids, including ourselves and our extinct relatives, can give us clues to the diets that they ate. The teeth of our ancestors can also show us how our own diet has evolved. Herbivores tend to have large, flat grinding teeth to break apart plant material. Omnivores have several different kinds of teeth to tear at meat and grind plant food.

Different teeth are needed for herbivorous and carnivorous diets. (Table 12.1.2).

Hard, brittle foods such as seeds can be crushed between teeth with rounded cusps and shallow jaws. Tough foods, such as raw meat or leaves, need to be sliced or sheared by teeth with thinner, blade-like crests. The shape of teeth reveals what these primates are capable of eating. Their diets probably varied with seasons and availability of food. You can use digital images of skulls to examine other Hominid teeth and deduce what they may have eaten.

Hominid	***Pan troglodytes* Chimpanzee**	***Paranthropus robustus* (An ape that lived between 1 and 2 million years ago in South Africa)**	***Homo floriensis* (an extinct species of small human that lived in Indonesia about 50,000 years ago)**	***Homo sapiens* Human**
Cusps on molar teeth – provide more surface area for chewing. They help teeth grind and break down fibrous material	yes	yes	yes	yes
Enamel thickness – thick	thin	Very thick	Very thick	Very thick

enamel prevents teeth breaking as animals chew				
Size of canines – canine teeth may not be used for hunting but for social interaction	Very large	medium	medium	medium
Jaw and muscles – strong muscles and jaws indicate a low energy plant diet for at least part of the time.	Incisors cut and rip food from trees. Molars and premolars grind and crush.	Thick bones and powerful muscles, strong molars for grinding	Smaller premolars, similar to humans, not capable of forceful biting	Teeth that are similar to H. floriensis but are not always needed for the food we eat.
Diet	Primarily plant material:	Tree and shrub food: Nuts seeds	Probably plants and	High energy food and

	fruits, seeds, nuts, leaves and insects, Sometimes meat.	and hard fruits	uncooked meat.	a mixed diet which includes meat and plant sources

Table 12.1.2: The teeth and diets of some hominids.

2 Adaptations of herbivores

Herbivores range in size from tiny ants, termites and caterpillars to koalas, wildebeest, goats, horses and rhinos. Some herbivores such as flying foxes and fruit bats only eat fruit while others such as caterpillars and koalas only feed on leaves and termites only eat wood.

Small leaf eating insects have piercing and chewing mouthparts; aphids have a tube-like stylet which they use to pierce stems and drink plant sap, caterpillars have biting and cutting mouthparts which saw through leaves. (Figure 12.1.10).

Herbivores adapted to a high **cellulose** diet such as ruminants (goats and sheep) have specialised teeth. They do not have **incisors**, (the flat front teeth which carnivores use to tear flesh), instead, they have a dental pad that helps chew plants. Herbivores have broad, flat **molars** (back teeth) with rough surfaces, which are used for grinding up tough plant tissues. Many herbivores (like rabbits and squirrels) have chisel-like front teeth used for gnawing through wood or hard seeds. These teeth grow continually as they are worn down with use. Ruminants have a digestive system with a four chambered

stomach. Their saliva is alkaline as they don't begin to digest food as soon as it enters their mouths.

Figure 12.1.10: The sawfly cuts circular holes in the leaves as it feeds

3 Adaptations of predators

Predators also have adaptations which help them catch and kill their prey. These may be physical, chemical or behavioural. Physical adaptations include good eyesight and hearing for locating prey. Sharp teeth and claws for holding and biting are also important. As well as physical adaptations, predators may use chemical signals to locate their prey. Great white sharks use sensory organs in their snouts to detect the tiny electrical impulses produced by their prey. Snakes and spiders produce venom and toxins to poison and kill their prey. Snake venom is toxic saliva that is injected by fangs into prey and spiders also

produce venom which can immobilise prey before it is eaten. Predators may either be solitary hunters, like the lynx (Section 12.3) or hunt in packs or family groups like wolves and lions. Behavioural adaptations such as calls and scents ensure that the group works together.

4 Adaptations of plants

You can read about the adaptations plants use to transport and gather the substances they need for photosynthesis in Chapter 8.3 and about their adaptations to life in very dry or very wet environments. Some species in dense forests grow upward to reach sunlight which you can read about in Chapter 11.4 Other species, known as shade-tolerant plants, are adapted to living in the dimmer light on the forest floor. These plants can photosynthesise using far red light more efficiently than other plants. Red light tends to be absorbed in the forest canopy while far red light (730 nm) penetrates to the lower levels of the forest. Shade tolerant plants grow broader, thinner leaves to catch the maximum sunlight and they use nutrients more efficiently than other species. Some plants move their leaves to capture light at different times of day and others move their chloroplasts to the top of their palisade cells when most light is available.

5 Adaptations to avoid being eaten

Plants and animals have many strategies to protect themselves from being eaten or damaged by organisms which feed on them. These are discussed in section 12.3.3

Mutualism is a relationship where both species involved benefit from their interactions. In some cases, the species are entirely dependent on each other and in others, they both benefit but could survive without each other.

The relationship between aphids (small sap-sucking insects) and some species of ants is an example of mutualism. Aphids secrete sugary honeydew, a liquid that is the waste product of their diet. Ants feed on the honeydew and in return some ants will protect the aphids from predators and parasites., even carrying aphid eggs and young into their nests.

Coral and the zooxanthellae they contain are another example of a mutualistic relationship. Corals begin life as a tiny, free-swimming larvae which settle and develop into polyps. Polyps replicate to form a colonies of polyps, growing on top of each other and they produce a calcified skeleton around themselves. As corals grow, they take in single-celled zooxanthellae from the environment around them. The coral provides shelter and nutrients for the zooxanthellae to use for their photosynthesis, while the zooxanthellae produce sugars, and other substances, as well as oxygen as a by-product which the coral uses for respiration.

TEST YOUR UNDERSTANDING

6 List three adaptations of predatory species which assist with capturing prey.

7 How can a study of dentition (teeth) help identify an animal's diet?

8 Outline two examples of a parasitic relationship.

REFLECTION

Can I explain the importance of categorising organisms in an ecosystem even though the categories are only approximate?

12.2 Transfer of energy and matter

LEARNING OBJECTIVES

In this section you will:

- learn that most ecosystems rely on light energy from the sun
- recall that photosynthesis converts light energy into carbon compounds for metabolism, to increase biomass for growth and reproduction
- understand that energy flows through food chains as organisms feed and as it is used it is converted to heat
- learn that organisms cannot convert heat energy to other forms of energy
- recognise that energy is lost from trophic levels as dead organic material
- understand that because energy is lost at each trophic level, the length of food chains and accumulation of biomass is limited
- draw food webs to represent feeding relationships in a community and use them to construct pyramids of energy
- learn that organisms also acquire nutrients such as nitrogen as they feed and that nutrients are eventually recycled to the environment
- summarise the stages in the carbon cycle

- recognise that energy flows in a food web can be quantified and represented in diagrams.

GUIDING QUESTIONS

- How do feeding relationships between organisms show the flow of energy in an ecosystem?
- What patterns can we identify in the complex structure of an ecosystem?

12.2.1 Energy flow

Food chains

In Section 12.1 we discovered that every organism needs food to survive but eventually it too is eaten. There is a hierarchy of feeding relationships that influences how nutrients and energy pass through it producing a sequence of organisms known as a food chain.

Most ecosystems rely on a supply of energy from the Sun. This can be converted into chemical energy in carbon compounds, such as glucose, by photosynthesis. Only a few food chains begin with chemoautotrophs (Section 12.1), which do not need sunlight. All green plants are autotrophs and are able to capture light energy from the Sun and form the first link in different food chains.

Carbon compounds that plants produce are used

- as a source of energy in respiration
- to increase biomass in growth or reproduction.

As the chemical energy in stored carbon compounds passes from one organism to the next, the energy is used for similar processes in each organism.

autotroph → primary consumer → secondary consumer → tertiary consumer

TEST YOUR UNDERSTANDING

9 Use organisms from where you live to make a food chain which includes the four categories of feeding listed in the

section on Food chains.

EXAM TIP

In an exam, do not say that plants 'make' energy for food chains. Remember that they only convert energy from one form to another.

Energy and food chains

Arrows in a food chain show the direction of flow of both the energy and nutrients that keep organisms alive. Energy flow through an ecosystem can be quantified and analysed. Studies reveal that, at each step in the food chain, energy is lost from the chain in various ways:

1 some energy is not consumed,

2 some energy leaves the food chain as waste, or when an organism dies, and

3 some is used by living organisms as they respire.

In all three cases, the lost energy cannot be passed to the next trophic level.

Consider an area of African savannah where grass, antelopes and cheetahs form a simple food chain.

grass → antelope → cheetah

- Energy loss 1: food not consumed. The grass stores energy from photosynthesis but the antelopes only eat some parts of the grass. So, they do not consume all the energy it has stored.

- Energy loss 2: not assimilated, or lost through death. The grass that is eaten passes through the digestive system of the antelope but not all of it is digested and absorbed, so some passes out in the feces. If an antelope dies and is not eaten by a predator, but decays and is eaten by detritivores and saprotrophs, the energy in its body is lost to this food chain.
- Energy loss 3: cell respiration. Antelopes use energy to move and to keep their body temperature constant. As a result, some energy is lost to the environment as heat.

Organisms cannot convert heat energy into any other form of energy and so it is lost from the ecosystem.

The assimilated energy that remains after respiration goes into building the antelope's body and this energy becomes available to the cheetah when it eats the antelope.

Food webs

Different food chains in an ecosystem are interconnected because it is unusual for an organism to have a single source of energy and nutrients. The organisms in the simple food chain in the 'food web below' will eat and be eaten by different organisms. For example, the antelope could be eaten by a lion or a wild dog, and the grass eaten by zebra and wildebeest (Figure 12.2.1). Linked food chains form a web of interactions known as a food web.

Pyramids of energy

We can represent the transfer of energy between trophic levels as an energy pyramid. Each trophic level is drawn as a rectangle and the width of each layer in the pyramid is proportional to the amount of energy it represents.

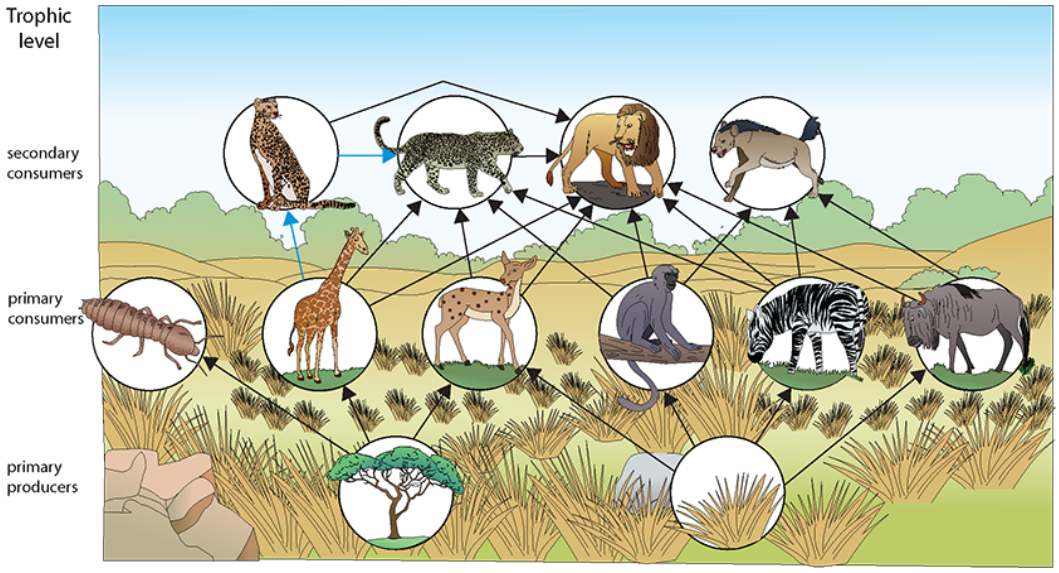


Figure 12.2.1: A food web from a grassland ecosystem.

NATURE OF SCIENCE

Using theories to explain natural phenomena: the concept of energy flow and food chains

Plants are the primary source of energy in nearly all ecosystems, but energy capture and photosynthesis are not efficient processes. So, not all of the energy of the sunlight is used. When herbivores feed, the energy transferred from plant to herbivore is also not 100% efficient. Not all of the plant material is eaten and absorbed, and some energy is lost in movement and respiration. The same is true for carnivores eating prey animals. Only about 10% of the energy in producers is passed to herbivores and a similar low percentage of energy is passed from herbivores to carnivores (Figure 12.2.2).

All along a food chain or food web, energy is lost at each trophic level through respiration and waste. This is why ecosystems rarely contain more than four or five trophic

levels. There is simply not enough energy to support another level and understanding the concept of energy flow explains why food chains are limited in length.

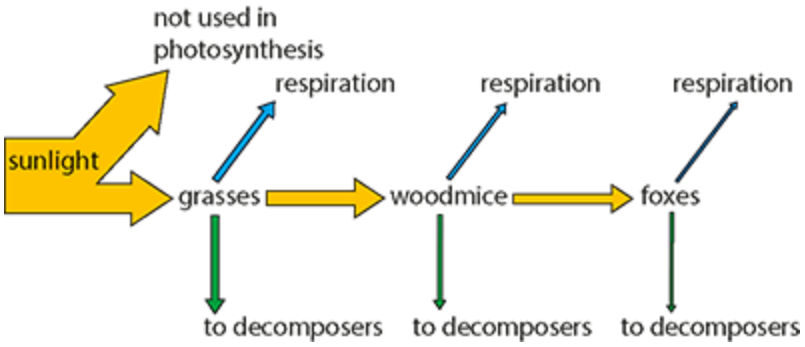


Figure 12.2.2: Energy losses in a food chain (not to scale). Energy is measured in kilojoules per square metre per year ($kJ \cdot m^{-2} \cdot year^{-1}$). Only a small percentage of the energy in each level is transferred to the next.

To consider:

1 How useful are diagrams such as food chains and pyramids of energy in helping us understand an ecosystem?

2 How can ecologists improve the quality of the data they collect in an ecosystem?

So the antelope → cheetah energy transfer would appear as in Figure 12.2.3. This section of an energy pyramid shows that only about 10% of the energy from the antelope passes to the cheetah and that about 90% is lost.

Energy losses of up to 90% occur at every step in a food chain, between one trophic level and the next.

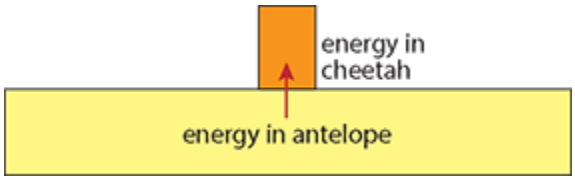


Figure 12.2.3: A simple energy pyramid for a single energy transfer.

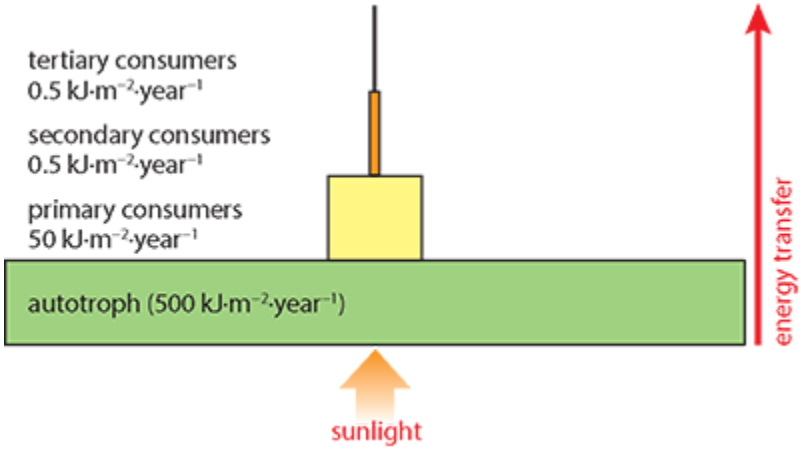


Figure 12.2.4: A generalised energy pyramid. It shows only 10% of available energy passing to the next trophic level.

The energy content of each trophic level can be calculated and used to construct an energy pyramid for an ecosystem (Figure 12.2.4).

Pyramids of energy show the amount of energy at each trophic level.

One consequence of the energy losses between one level and the next is that the quantity of biomass (biological material) at each level decreases due to the loss of waste products at each transfer of energy. Every link in the chain results in losses, so that eventually there will be insufficient energy to support any further trophic levels. Most food chains commonly contain between three and five species, and seldom more than six. The energy that

enters an ecosystem as light is converted to stored chemical energy and finally lost as heat.

TEST YOUR UNDERSTANDING

10 Explain what the arrows represent in a food chain.

11 State the initial source of energy for most food chains.

12 List the three ways in which energy is lost when moving from one trophic level to the next.

INTERNATIONAL MINDEDNESS

Feeding the world

We can also draw pyramids of energy for different types of farming and food production and compare their trophic levels and efficiency of energy conservation. As we grow arable crops or farm animals, energy is used in the growth of biomass and production of offspring. In a terrestrial ecosystem, food is usually harvested at trophic levels 1 and 2 (autotrophs and herbivores). From an energy point of view, farming systems that produce crops are much more efficient than those that produce animals because the energy has passed through only one trophic level.

One hectare of land can produce approximately 7.5 tonnes of wheat (equivalent to 11 000 loaves of bread) or 0.3 tonnes of beef. Many more people can be provided with food from the wheat rather than the beef.

A steadily rising world population has increased the demand for food and led to environmental damage, as more land is needed for farming. Forests have been cut down and soil

degraded. Also, as newly emerging economies worldwide have seen incomes rise, the demand for meat has also increased.

To consider:

1. Why is arable farming more efficient in energy terms than farming animals?
2. How important are other products that we get from farmed animals, for example milk, wool and leather?
3. Do you think that people should be encouraged to reduce their consumption of meat and to eat more vegetables?

12.2.2 Nutrient recycling

All the organic matter from an organism, including living or dead material and waste, is eventually consumed by other organisms, which include detritivores and saprotrophs (Section 12.1). All these organisms respire and release energy as heat, which is no longer available for use by living things.

Nutrients are continually recycled and the supply of nutrients is limited and finite. Nutrient cycles transfer chemical elements such as carbon, nitrogen and phosphorus through the biotic and abiotic parts of the ecosystem from one organism to the next. Organisms usually take in nutrients as ions and use them to construct cells or for respiration.

For example, nitrogen is absorbed as nitrate or ammonium ions. It is used to construct proteins and nucleic acids. Phosphorus may be absorbed as phosphate to construct nucleic acids or lipid molecules.

A nitrogen atom may be absorbed as nitrate by plant roots and used to make an amino acid. The amino acid may pass into an animal when the plant material is eaten, and then pass out of the animal's body in urine as urea during excretion. Soil bacteria may convert urea in the excreted material back into nitrate and the cycle begins again.

In the soil, the saprotrophic bacteria and fungi are essential for the recycling of nutrients. Relationships like this form part of biogeochemical cycles.

The carbon cycle

Carbon is one of the most important elements that is recycled in an ecosystem (Figure 12.2.5). Inorganic carbon dioxide in the atmosphere is trapped or ‘fixed’ as organic carbon compounds during photosynthesis. Carbon dioxide gas needed for photosynthesis passes by diffusion from the atmosphere into land-dwelling autotrophs and dissolved carbon dioxide diffuses from water into aquatic organisms. Some of this carbon is soon returned to the atmosphere or the water as the plants respire. The other steps in the carbon cycle follow the same path as food chains. As herbivores eat plants, and carnivores eat herbivores, the carbon compounds move from plants to animals. Respiration by any organism in this sequence returns carbon to the environment as carbon dioxide. When a plant or animal dies, carbon compounds in their bodies provide nutrition for detritivores and saprotrophs and may also be respired, returning carbon dioxide to the atmosphere. Aerobic respiration depends on the release of oxygen from photosynthesis and photosynthesis depends on the release of carbon dioxide from respiration.

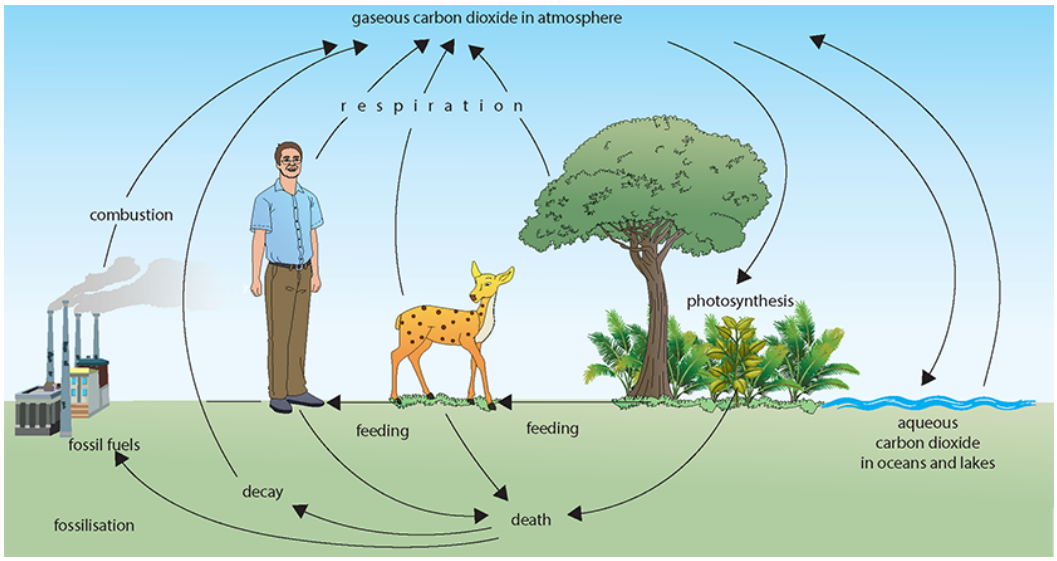


Figure 12.2.5: The carbon cycle.

This relationship forms a major interaction between autotrophs and heterotrophs.

EXAM TIP

Recall that energy *flows*, but that nutrients *cycle*.

Methane, peat and fossil fuels

In some conditions, plants and animals do not decay completely when they die and organic carbon in their bodies may not be released directly as carbon dioxide as they decompose. In wetlands, where the soil is waterlogged and the concentration of oxygen is very low, organic material is broken down by anaerobic **methanogenic** bacteria (Achaeans), which produce **methane** as a byproduct. Peat bogs represent the largest natural source of atmospheric methane, a greenhouse gas many times more harmful than carbon dioxide. A large proportion of this methane is recycled by methane-eating bacteria (methanotrophs) in the peat bog, which form a symbiotic relationship with peat-bog moss. The bacteria supply additional carbon dioxide to the moss, which in turn provides the bacteria with a habitat.

Methane is oxidised to carbon dioxide and water released into the atmosphere. Thus, eventually, carbon dioxide rejoins the carbon cycle and contributes to the carbon dioxide in the atmosphere, although some remains in the ground.

Peat is also produced in wetlands where partly decayed vegetation accumulates and flooding prevents oxygen flow in the soil, so that anaerobic and sometimes acidic conditions persist. For peat to form, the production of biomass must be greater than its chemical breakdown. The first stage in peat formation takes place when water levels are low and aerobic micro-organisms act

on decaying vegetation in surface layers. As this layer becomes covered by further layers of vegetation and thus subjected to anaerobic conditions in wetter, deeper layers, it becomes preserved and changes very little with time. Depending on the local conditions, the types of peat from different areas can be quite different in their degree of decomposition. Areas where peat is found also have specific kinds of plants, such as heather, sphagnum moss and sedges. Since organic matter accumulates over thousands of years, peat deposits also provide records of past vegetation and climates.

KEY POINTS

methanogenic bacteria respire anaerobically and release methane

methane (CH_4) is a greenhouse gas released by animals as they digest food, by rice fields and some bacteria

Peat is harvested as a fuel in some parts of the world as it can be compressed into bricks and unlike coal, burns without producing smoke (Figure 12.2.6). But peat is not classified as a renewable energy resource because its rate of extraction is far greater than its very slow rate of formation which can take between 1000 and 5000 years.

Figure 12.2.6: Peat is cut to be used as a fuel. It can also be used as an additive improve soil structure.

Some partially decomposed organic matter from past geological eras has become compressed and fossilised in a process that has taken millions of years. Conditions on Earth at the time prevented decomposers from feeding on this material and over time it was fossilised to become **fossil fuels**. Vast coal, oil and natural gas deposits have been formed deep under the ground where they contain reserves of carbon that are locked in and excluded from the carbon cycle for very long periods of time. The carbon trapped in these fuels cannot return to the atmosphere unless the fuels are burned, when the carbon in them combines with oxygen in the air to form carbon dioxide. Over a very long period of geological time, fossil fuel formation has gradually lowered the carbon dioxide level of the Earth's atmosphere, but in more recent times this balance has been upset. As people burn wood, peat, coal, oil or gas, the carbon molecules locked up in them for thousands or millions of years are released back into the atmosphere as carbon dioxide. As carbon dioxide re-enters the carbon cycle it has the potential to cause global warming (Section 12.6).

NATURE OF SCIENCE

Obtaining evidence – how is reliable data on carbon dioxide and methane concentrations obtained?

The level of carbon dioxide in the atmosphere has been increasing since the end of the nineteenth century. The first measurements of the gas were made by Charles Keeling (1928–2005) who took careful measurement of samples from Mauna Loa in Hawaii and the Antarctica in 1957 and 1958. Since then records of the gas have been kept continuously and the Keeling curve (named after Charles Keeling) and additions to it are well recognised (Figure **12.2.7**). Today, many different types of data are collected about the Earth's climate (Figure **12.2.8**) and the concentrations of both carbon dioxide and methane are recorded.

Carbon dioxide has been measured using a weather satellite instrument called AIRS (Atmospheric Infrared Sounder) developed by a NASA scientist Moustafa Chahine (1935–2011). AIRS works by measuring the infrared light emitted by carbon dioxide molecules. Gas molecules absorb infrared rays emitted by the Earth's surface and then re-radiate them at a slightly lower energy level. The exact frequencies of the emissions depend on temperatures. About three million measurements are taken in this way every day and a computer processes the information that is collected. AIRS focuses on a section of the atmosphere known as the middle troposphere but a second satellite, NASA's Orbiting Carbon Observatory (OCO), collects data from the entire atmospheric column.

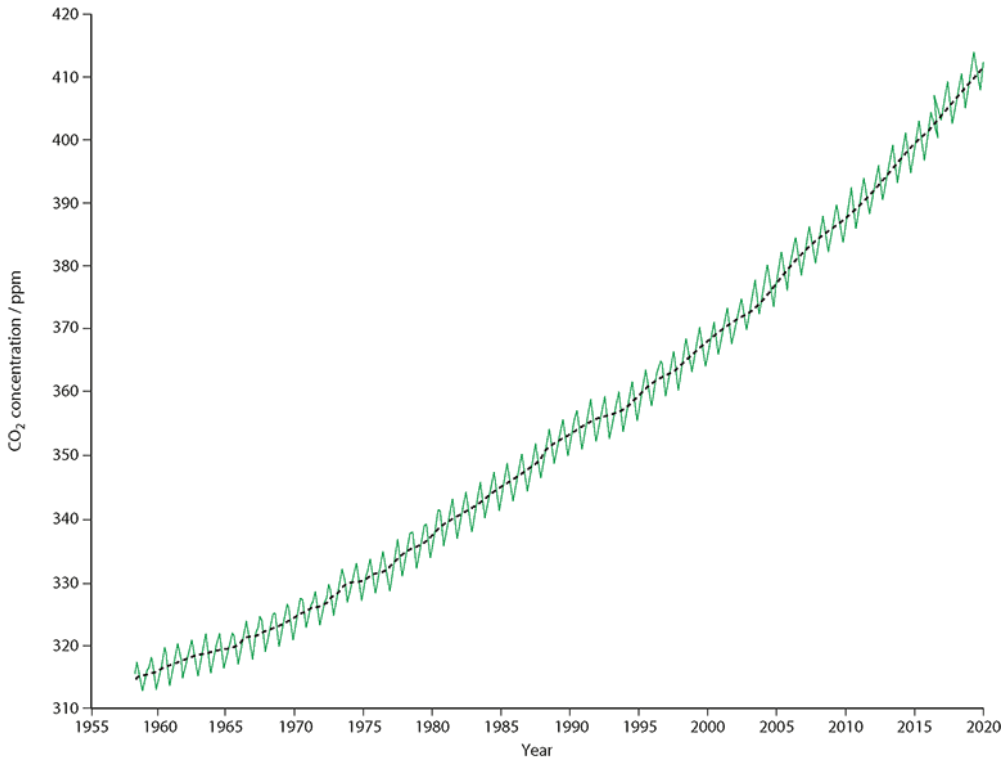


Figure 12.2.7: Atmospheric carbon dioxide concentrations measured at monthly intervals in Hawaii.

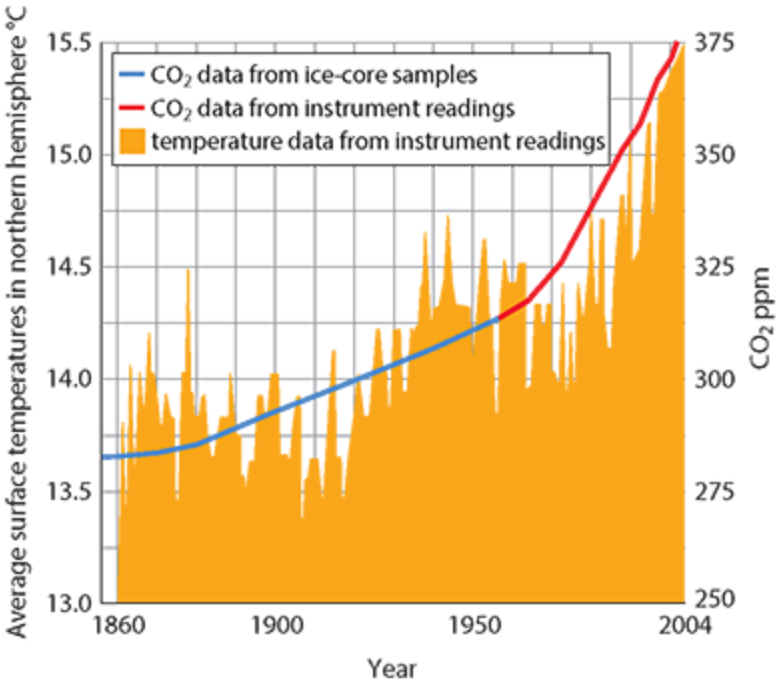


Figure 12.2.8: Graph showing carbon dioxide data from different sources and evidence of warming global temperatures.

The OCO uses optical properties of carbon dioxide to measure its presence using an AIRS. Carbon dioxide molecules vibrate at certain frequencies of light and this instrument uses the sunlight as a light source to monitor the vibrations. The satellite circles the Earth and collects light that has passed through the atmosphere and into three spectrometers, two of which respond to carbon dioxide. The spectra produced resemble bar codes, which can be analysed precisely back on Earth and provide data on areas as small as a square kilometre.

Question to consider

- Look at the curve shown in Figure **12.2.7**. The dotted line shows the trend of increasing carbon dioxide in the atmosphere but what do the peaks and troughs along the line indicate?

TEST YOUR UNDERSTANDING

13 The leaves of a tree store 20 000 $J \cdot m^{-2} \cdot year^{-1}$ of energy. Estimate the amount of energy that may be stored in caterpillars that feed on the leaves.

14 State the difference between energy movement and nutrient movement in an ecosystem.

15 Why is it important to understand the flow of energy in an ecosystem?

REFLECTION

Could I communicate the importance of understanding feeding relationships to someone else? Which ideas would I focus on?

Links

- Photosynthesis is essential to provide energy for organisms' metabolisms. What do organisms use this energy for? (Chapter 2.3)
- How does understanding ecosystem stability help us understand the importance of biogeochemical cycles? (Chapter 12.4)

12.3 Ecological relationships and populations

LEARNING OBJECTIVES

In this section you will:

- define the terms populations and community in an ecosystem
- discover that organisms and populations in a community interact in many ways
- recognise that when overall effects are limited, organisms can coexist and biodiversity can increase
- recall that ecological relationships can be intraspecific and interspecific and can affect the distribution and population sizes of species
- learn that population size can be estimated using random sampling methods which include quadrats and the Lincoln index
- understand that competition occurs when different organisms require the same resources, and may be inter- or intraspecific
- discover that the chi-squared test can be used to investigate evidence for interspecific competition
- investigate predation including quantifying predator–prey relationships and adaptations that prey animals have for defence

- consider herbivory and the defence mechanisms plants use to protect themselves
- understand the importance of keystone species in an ecosystem

- understand the interspecific interaction mutualism
- understand that cooperative intraspecific relationships can have advantages and some disadvantages
- learn how new populations in suitable environments grow in size, and define the terms lag phase, log phase and stationary phases
- define the carrying capacity of an ecosystem
- understand how numbers in a population may fluctuate around the carrying capacity due to density-dependent and -independent factors
- learn that in some cases a population may enter a phase of decline
- recall that plants, fungi and bacteria may release chemicals that are detrimental to other species close to them.

GUIDING QUESTIONS

- What interactions occur between organisms and populations?

- Why are organisms and population in communities interdependent?
- What effect do ecological relationships have on the population size of different species?

12.3.1 Interactions between populations

A **population** of organisms is a group of the same species living in the same place at the same time and interacting with each other. The *number* of individuals is known as the population size and is only one aspect of any population.

A **community** is the biotic (living) component of an ecosystem. It consists of populations of different species that live in the same area and interact with one another. In any ecosystem there will be communities of organisms – that is, groups of different populations of all the different species – that are present that **coexist** and interact with each other. Some of these species will have overlapping habitats and niches (Section 11.4). Interactions in communities are important factors in natural selection. The interactions help shape the ecosystem and evolution of the interacting species within it. Three major types of community interactions are predation, competition and symbiosis. **Symbiosis** means 'living together' and includes **commensalism**, mutualism and parasitism. These interactions are explained in Section 12.2.

Interactions, population size and distribution

The size of a population and the distribution of organisms in an ecosystem is dependent upon biotic and abiotic factors. Every organism has an ideal living space that it requires, its fundamental niche, but the space it actually inhabits is its realised niche (Section 11.4). Abiotic factors such as soil type, shelter, weather and climatic conditions are key to determining an organism's distribution but the size of a population and its distribution are also influenced by interactions between members of its own species and of other species. A distribution pattern may be explained if one species excludes another through

competition or limits the range of a species by feeding on it, or affecting its health.

The availability of food is a major factor in how many organisms can live in an ecosystem. Rainforests with rich food supplies can support more species that are more evenly distributed than deserts and the polar regions where there is less food. Both interspecific and intraspecific competition influence the number of organisms that can survive. In most situations predators and prey have evolved together and a balance is achieved but the arrival of new predators can upset this balance, for example the red fox was introduced to Australia in the mid 1800s and their numbers increased rapidly. Foxes have been blamed for the decline in native species such as quokkas, numbats and bettongs. They also reduce the food available to native predators whose populations are reduced.

Pathogens and parasites also influence the abundance of a species, and interactions with new pathogens can limit population size and distribution. For example, a new pathogen, the fungus *Hymenoscyphus pseudoalbidus* which causes ash die back disease has spread across Europe since the start of this century. This chronic fungal disease of ash trees causes loss of leaves and death of upper parts of infected trees. The fungus has killed many ash trees and the number and distribution of ash trees has been severely reduced in many parts of Europe.

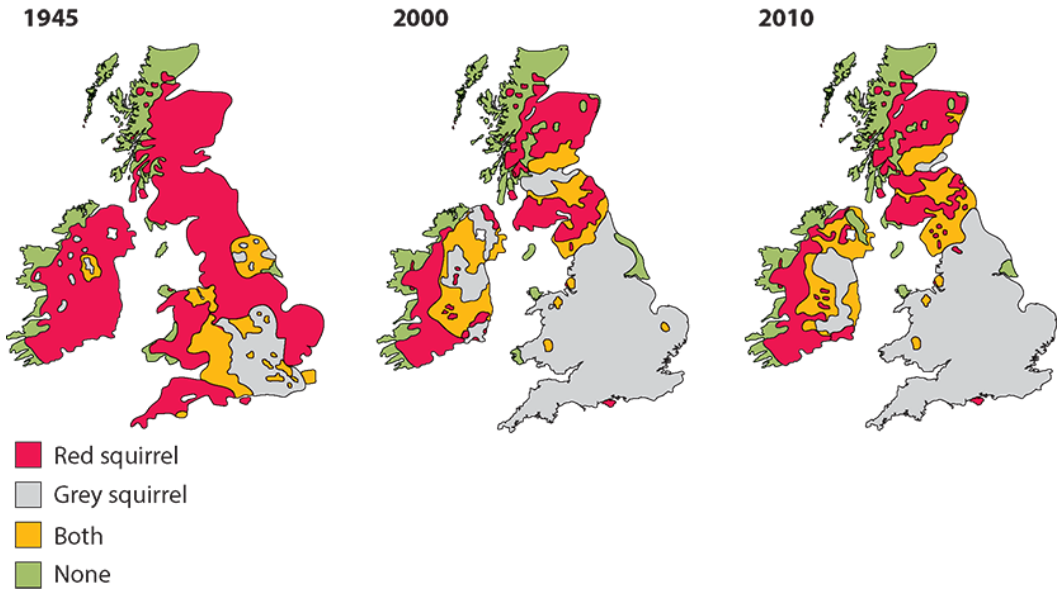


Figure 12.3.1: The distribution of red and grey squirrels in the United Kingdom and Ireland over a 65 year period.

Competition between species can lead to one being excluded by another. This can occur between existing species or because a new species has arrived or been introduced. Examples of out-competition of native species by newly introduced species include the Canada goose in Europe. In the 1990s there were only a few hundred Canada geese (*Branta canadensis*) in France but now their numbers have increased so much that they are limiting the population size and distribution of ducks, coots and other goose species. The cane toad in Australia (Section 11.4) and the invasive Himalayan balsam, now growing in many countries, are also affecting the populations and distribution of native species that they compete with.

Figure 12.3.1 shows how the distribution and population size of squirrels in the United Kingdom and Ireland has changed over a period of 65 years. Red squirrels (*Sciurus vulgaris*) are the native species in the British Isles. The grey squirrel (*Sciurus carolinensis*) species was introduced around 100 years ago from

America. The grey squirrel is larger and heavier, with a typical weight of 550 g, compared to 300 g for native red squirrels. Greys compete with reds for food and also carry a virus known as squirrelpox. Grey squirrels are immune to the disease but they transmit it to red squirrels, for whom it is fatal. Populations of red squirrels are now only found in places where greys are rare or absent. Grey squirrels can also affect the composition of native woodland by removing bark from trees bark and selecting the seeds of certain trees to eat. The population size and distribution of the red squirrel has reduced as a result of this interspecific interaction while those of grey squirrels have increased.

The abundance and distribution of species are estimated using techniques such as transects and quadrats described in Section 12.3.2. Distribution–abundance relationships give an insight into the relationship between the local abundance of species and the size of their ranges within a region and these can be related to both abiotic factors and interactions between species.

TEST YOUR UNDERSTANDING

16 Suggest two abiotic factors that might affect the distribution of a species.

17 Why can a newly arrived species reduce the population of an existing species in a habitat?

18 Why is the distribution of species more even and plentiful in a rainforest than a desert?

12.3.2 Estimating population sizes

Investigating distribution of species: random sampling

A population is a group of individuals of the same species that live in the same area.

Population numbers can and do change over time and are affected by a number of factors in the environment. When ecologists want to understand the distribution of a species, or to compare the distribution of one species with that of another in a different location, it is usually impossible to do so by a direct counting method. In most cases, ecologists take a sample of the population and, if the sample is chosen at random, it should provide a good representation of the whole population. Random sampling is used if the area under investigation is large or if time is limited, and it assumes that every organism has an equal chance of being sampled (that is, of being selected as part of the sample). There are a number of sampling methods used by ecologists to collect data on the distribution of species in relation to one another and to abiotic factors in their environment.

Two common methods used for stationary organisms are quadrats and transects (transects are described in Section 11.4.2). Quadrats can show the population number (abundance), the distribution of organisms, their density and the frequency of occurrence. Mobile organisms are estimated using the capture–mark–release–recapture method.

Population density is calculated using the equation:

$$\text{Density} = \frac{\text{total number in all quadrats}}{\text{number of quadrat samples} \times \text{area of quadrat}}$$

KEY POINT

random sampling a technique in which each sample has an equal probability of being chosen to provide an unbiased representation of the total population.

Quadrats: to sample sessile (non-moving) species

One of the simplest and easiest sampling techniques involves using a quadrat (Figure 12.3.2). A quadrat is a square made of metal or wood divided into smaller squares. The quadrat is placed on the ground so that the organisms inside the square can be counted. The size of the quadrat will largely be decided by what is being measured. To estimate the number of different trees in a wood may require quadrats of 10 m by 10 m, but a 1 m quadrat would be the best size for studying wild flowers in grassland. Very small 10 cm quadrats might be used for sampling lichens on walls or tree trunks.

Figure 12.3.2: Using a quadrat to sample an area of grassland.

To use quadrat sampling you should:

- Use the correct number quadrats to get enough data to get a representation of the population.
- Identify the species in the quadrats so they can be distinguished from one another.
- Use a random number generator or table to select the squares to sample.

THEORY OF KNOWLEDGE

Sampling bias

Placing quadrats to collect data on the distribution of a species may not always be a truly random process. Researchers can introduce personal bias, even without meaning to, by placing a quadrat in a spot that they think will be more interesting or easier to work in. To ensure that the samples within a survey area are made completely randomly, a numbered grid of the area may be drawn up, and random number tables or generators used to select squares on the grid where a quadrat should be placed (Figure 12.3.3). Random number tables and random number generators are lists of numbers selected by a computer without any human bias.

To consider:

1 Is random sampling a useful tool for scientists?

2 How significant is the potential for sampling bias and can this ever be completely avoided?

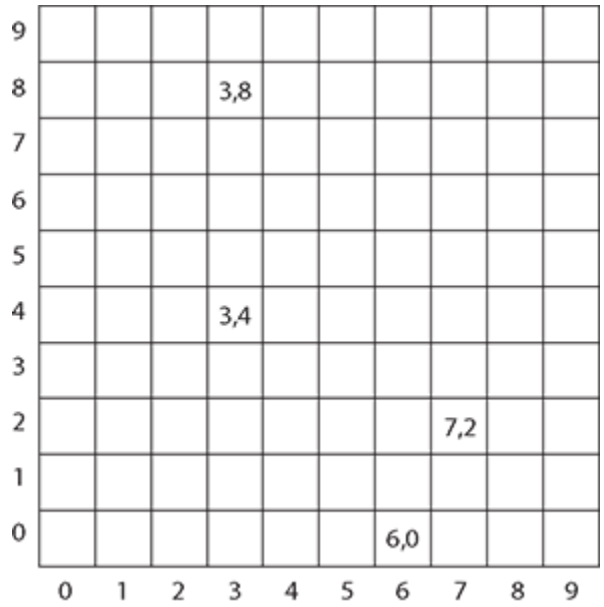


Figure 12.3.3: To select a part of an area to sample with a quadrat, divide the area into a grid of squares and then select a row and a column using randomly generated numbers.

Mark–release–recapture method: to estimate numbers of a motile (moving) species

The most common method of estimating population size of animals that can move is the capture–mark–release–recapture technique (Figure 12.3.4). It is used for populations where individuals are mobile and move freely in their habitat.

1 A sample of the population is collected by netting, trapping or another suitable method. The sample must be as large as possible and the trapping method must not harm the animals.

2 The number of organisms in the sample is counted and recorded.

3 Each of the captured animals is inconspicuously marked in some way; for example, with non-toxic paint for invertebrates or by trimming a concealed area of fur for small mammals.

4 The animals are returned to the wild and left for long enough to interact freely with the rest of their population.

5 A second sample of the population is collected after this time.

6 The number of marked and unmarked individuals in the second sample is counted.

The population size is calculated using the Lincoln index formula:

$$P = \frac{M \times N}{R}$$

where:

P is the total population

M is the number of organisms caught and marked originally

N is the number caught in the second sample (with and without marks)

R is the number of marked individuals in the second sample.

This method depends on a number of factors, which need to be taken into account.

- Marking the organisms must not harm them or cause them to be conspicuous to predators. That is, the marking itself must have no effect on the population size.

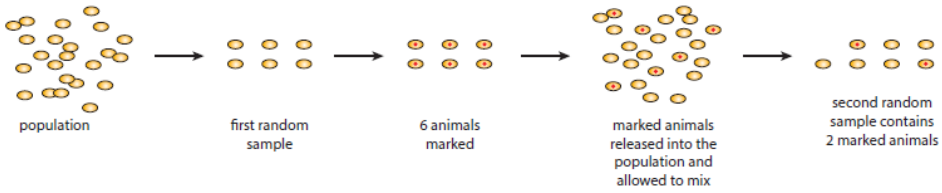
- There should be minimal immigration into or emigration from the population.
- The measurements must be conducted within a single life cycle, so there are no significant changes to the population through births or deaths.

The capture–mark–release–recapture technique is most appropriate for invertebrates such as woodlice, snails and ladybirds or small mammals, such as mice, with a limited territory. Sampling organisms with a large territory, or those where the population is small, is not accurate using this method

WORKED EXAMPLE 12.3.1

Here is a calculation using the Lincoln index for a theoretical population of small organisms that mix freely and live in the same area.

Answer



$$\text{estimated population size} = \frac{\text{number in first sample} \times \text{number in second sample}}{\text{number of marked animals in second sample}}$$

$$\text{estimated population size} = \frac{6 \times 7}{2}$$

$$\text{estimated population size} = 21$$

Note : This method only produces results of acceptable accuracy if the numbers in the samples are larger than shown here. At least 20 animals should be sampled.

Figure 12.3.4: Capture–mark–release–recapture technique for estimating population size.

Testing for the association between two species: the chi-squared test

Interspecific competition occurs when two different species need the same resource and one species is more successful when the other is not present.

You can read about examples of this type of competition in the example in which lions, hyenas and wild dogs are competing for a prey species, the zebra. Laboratory experiments can show us more about the effects of interspecific competition on a smaller scale. The work of G.F. Gause (1910–1986) on two species of single-celled *Paramecium* is described in Section 11.4.3.

We can use the chi-squared (χ^2) test to compare sets of data and evaluate if the differences between them are statistically significant or due to chance. To use this test in ecology we must:

- Use precise categorical data, for example species and counts of the numbers present.
- The test can be used to test for the association between species in an ecosystem. If the presence or absence of two species is recorded in quadrats during sampling it is possible to test whether:
    - the two species are distributed independently of one another (the null hypothesis), or
    - the two species are associated either positively or negatively, so they tend to occur either together or apart.

1 Draw up a table of results like this and enter the observed numbers of the two organisms in each quadrat:

	Species A present	Species A absent	Total for rows
Species B present			
Species B absent			
Total for columns			

2 Calculate the expected values, assuming the two species are independently distributed:

$$\text{Expected frequency} = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

3 Calculate the value of chi-squared (χ^2):

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where χ^2 is the test statistic, Σ means 'the sum of', O is the observed frequencies and E is the expected frequencies.

4 Calculate the degrees of freedom:

$$\begin{aligned}\text{The degrees of freedom} &= (\text{rows} - 1) \times (\text{columns} - 1) \\ &= (2 - 1) \times (2 - 1) \\ &= 1\end{aligned}$$

5 Use a chi-squared table to find the critical value at 1 degree of freedom for the chi-squared value you calculate. If the value you calculated (step 3) is greater than the chi-squared value at the 5% level then the null hypothesis can be rejected and we can accept that the two species are associated.

If the calculated value is less than the chi-squared value we accept the null hypothesis that the two species are distributed independently of one another (check the table of chi-squared values in Table 4.2.2).

EXAM TIP

Biologists use statistics when testing hypotheses. In biology we set the significance level (the cut off) for the probability of rejecting the null hypothesis at the 5% level. Put another way, we can reject a null hypothesis with 95% certainty of being correct.

TEST YOUR UNDERSTANDING

19 List three conditions that must be met to make the capture–mark–release–recapture method suitable to measure population size.

20 Which sampling method is used to measure populations of sessile organisms?

WORKED EXAMPLE 12.3.2

Two species of plantain, *Plantago major* and *Plantago minor*, grow in the same field. The field was surveyed using quadrats and random sampling and the following data are collected.

The null hypothesis is that there is no association between the two species and no significant difference between the observed and expected frequencies.

Answer

Step 1 Tabulate the data collected.

	***P. major* present**	***P. major* absent**	**TOTAL**
***P. minor* present**	22	46	68
***P. minor* absent**	33	49	82
TOTAL	55	95	150

Step 2 Calculate the expected frequencies.

	***P. major* present**	***P. major* absent**	**TOTAL**
***P. minor* present**	$(68 \times 55) \div 150$ $= 24.9$	$(68 \times 95) \div 150$ $= 43.1$	68
***P. minor* absent**	$(82 \times 55) \div 150$ $= 30.1$	$(82 \times 95) \div 150$ $= 51.9$	82
TOTAL	55	95	150

		***P. major* present**	***P. major* absent**
***P. minor* present**	*O*	22	46
	E	24.9	43.1
		0.34	0.19

	$\frac{(O-E)^2}{E}$		
***P. minor* absent**	O	33	49
	E	24.9	51.9
	$\frac{(O-E)^2}{E}$	2.6	0.16

Step 3 Calculate the value of chi-squared.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 0.34 + 0.19 + 2.6 + 0.16 = 3.29$$

Step 4 Check the chi-squared table on page 176.

Degrees of freedom = (2 − 1) = 1 as there are two species in the sample.

Compare the calculated value of chi-squared against the critical value at the confidence level of 95% (or $p = 0.05$) at 1 degree of freedom. As the chi-squared value is less than the critical value, we *accept* the null hypothesis. There is no significant difference between the observed and expected values for the two species of plantain.

NATURE OF SCIENCE

Assessing the risks and benefits of scientific research

As ecologists study organisms in their natural habitats they make observations and measurements to assess the numbers and distributions of species that are present. Sometimes it is necessary to disturb or even remove organisms from their habitats in order to discover the relationships that they have

with other species. Joseph Connell (1923–2020) who worked on two species of barnacle (Figure 11.4.9) removed one species from an area of rocky shore as he investigated the niches that the two species occupied. When carrying out surveys using the capture–mark–release–recapture method of estimating populations, organisms are trapped, counted and released. In both cases, ecologists inevitably cause some damage to the ecosystem that is being studied.

In the same way, when you carry out field studies you will trample and disturb the areas you are studying and remove organisms to investigate. It is important to carry out studies of the natural world so that we can understand relationships that may be vital to conservation of ecosystems or species. It is also important to keep in mind that we must do all we can to prevent harm to organisms as we study them and ensure that we minimise the damage and disturbance we cause.

To consider:

1 Which of these approaches to the study of competition do you think would cause least harm to organisms being investigated?

- **a** Field observations of interspecific competition between two predators.
- **b** Manipulation of animals in the natural world, for example removing one species to assess its effect on another.
- **c** Laboratory experiments, for example in an artificial ecosystem such as a tank of fish.

2 How have you tried to minimise disturbance in field studies you have carried out?

12.3.3 Growth of new populations

New populations can start to establish themselves when a few individuals of a species enter an unoccupied area. Perhaps a few rabbits arrive on an uninhabited island covered by lush grassland, or some fish are washed into a newly established pond, or bacteria land on a fresh plate of agar jelly. Assuming there is enough food and there are few predators, the newcomers will reproduce and the population will increase rapidly. After a time, when there are large numbers of individuals, the food supply will start to be used up faster than it can be replaced. The population will be unable to increase any further and the population numbers will stabilise. We can observe these changes taking place and plot the numbers present in the population as a graph.

The typical pattern of growth for a newly established population is shown in Figure 12.3.5.

It is possible to follow the development of a population as it grows and changes over a period of time. All new populations produce similar growth curves as their numbers increase as a result of reproduction, then stabilise and, in some cases, later decline. The phases of population growth can be identified and described as follows:

1 During the lag phase, the new population settles into its new environment and slowly begins to reproduce so that numbers increase.

2 The log phase (or exponential growth phase) occurs as more individuals reproduce and the population shows exponential growth (the steepest part of the curve). At this time, the population inhabits an ideal, unlimited environment: there is

abundant food, little competition for space and the effects of predation and disease are minimal.

3 After a time, the exponential phase ceases as one or more resources that individuals need become limited. The shape of the curve at this stage resembles a letter S, and it is called the S-phase. It shows that the rate of population growth is slowing down. The population is said to be in transition. Individuals must compete with one another for resources such as space, light, food, nutrients and water. This intraspecific competition increases as population numbers increase. When the rate of demand for a particular resource is greater than the rate of supply we say that the resource has become a limiting factor. Predation, disease and, in some cases, the accumulation of toxic wastes products such as carbon dioxide can also limit some populations.

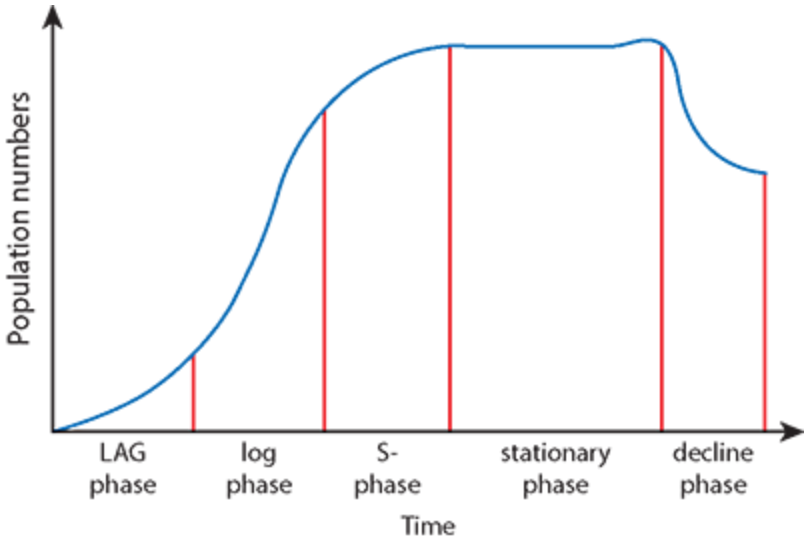


Figure 12.3.5: This graph shows the rate of growth for a new population as it becomes established and matures.

4 In the stationary phase population numbers become more or less constant and the curve on the graph levels off. The

ecosystem has reached its carrying capacity, which is the number of individuals in a population that the resources in the environment can support for an extended period of time. At this time the population growth rate will slow down, either because organisms die through lack of an essential resource, or because they fail to breed and their birth rate falls. The number of individuals in the population fluctuates around the carrying capacity due to density-dependent factors and density-independent factors. The population remains more or less stable because rates of both natality (birth) and mortality (death) are balanced, and so are rates of **emigration** and **immigration**.

5 In some species a population may enter a phase of decline. This is caused by events such as disease or starvation. There may also be longer-term effects such as low birth rates or low fertility in the reproducing population. There may also be high mortality rates and high emigration rates as individuals move away from a difficult environment.

Why do population sizes vary?

We have identified a number of important reasons why a population may change in size:

- natality – the birth rate may change (the number of new individuals joining the population due to reproduction)

KEY POINTS

carrying capacity is the maximum average number of individuals of a species that an ecosystem can support.

lag phase is a period of time when a new population arrives in a new environment and begins to reproduce.

log phase (or exponential growth phase) is period of maximum population growth.

limiting factor is in populations, a resource that prevents the growth of a population above a certain level.

phase of decline is period of time during which a population size decreases.

stationary phase is period of time during which a population size remains unchanged.

- mortality – the number of deaths may change
- emigration – members of the population may move away to new habitats
- immigration – new members of the species may arrive from elsewhere.

Some factors limit an increase in population numbers, no matter what species is considered. These include:

- availability of key resources such as food, water, oxygen, light, space,
- mates and shelter
- levels of waste products, such as carbon dioxide or nitrogenous waste
- disease
- predation or herbivory.

Density-dependent and density-independent factors

Density-dependent factors affect populations and their effects are related to the size and density of the population. Disease, migration and availability of resources are all density-dependent. If the population is small, individuals are less likely to compete for resources such as food, and are less likely to move away if conditions are favourable.

Figure 12.3.6: The common orange lichen (*Xanthoria parietina*) is an association between a fungus and an alga.

Density-independent factors also affect populations, but they are not related to population size or density. Density-independent factors include environmental factors such as light, temperature and rainfall, which may vary with the seasons.

KEY POINTS

density-dependent factor is a factor that influences a population and varies with population size.

density-independent factor is a factor such as light intensity or temperature which affects a population, whatever its size.

12.3.4 Competition

Competition is an interaction between organisms that require the same limited resources. The resources might be food, mates, nesting sites or territory. Two different types of competition are intraspecific and interspecific competition.

Intraspecific competition occurs between members of the same species. One example is competition between two male animals for mates (Figure 12.3.7). Competition has negative effects on both and leads to a reduction in fitness for both individuals, but the fitter individual survives and is able to reproduce. Intraspecific competition is a necessary factor in natural selection. It leads to adaptive changes in a species through time.

Plants, which cannot move, use the same resources as other members of their species that grow in the same area. For example young oak trees must compete for light, much of which will be blocked out by taller trees. Young oaks find themselves in competition with larger members of their own species. Acorns which germinate close to the parent plant are likely to be outcompeted and the young trees may die. This is one of the reasons why seeds of most plants are dispersed as far as possible from the parent plant so that competition is avoided.

Figure 12.3.7: Male bighorn sheep fighting for the right to mate.

Interspecific competition occurs between members of different species.

For example, two predator species might compete for the same prey. Interspecific competition takes place in communities of interacting species (Figure 12.3.8).

Interactions in communities and populations

Ecological interactions by one organism or population on another can have positive , negative or limited effects. Examples of these interactions are outlined in the next sections.

Mutualism: both organisms benefit

Mutualism is an interaction in which both species benefit. Lichen (Figure 12.3.6) is an example of a mutualistic relationship. A lichen is an association between a fungus and an alga. Lichens such as common orange lichen (*Xanthoria parietina*), which grows on rocks and walls, twigs and branches, consist of fungal filaments that surround cells of a green alga. The fungus absorbs water from the air and minerals from rocks or soil and protects the alga from intense sunlight and drying out. The alga uses the

water and minerals, and photosynthesises to makes carbohydrate for itself and the fungus. Both gain from this mutualistic relationship.

KEY POINT

mutualism a cooperative interspecific interaction; sometimes called a mutual symbiosis.

Figure 12.3.8: Lions, hyenas and vultures compete for the carcass of a zebra.

Other mutualistic relationships include;

- the Egyptian plover (*Pluvianus aegyptius*), and the Nile crocodile. The bird feeds on parasites and food particles left around the crocodile's mouth, keeping its teeth clean and healthy. The crocodile openly invites the birds to hunt on its body, even allowing them to enter its mouth.
- coral polyps and the zooxanthellae which live inside them. Zooxanthellae photosynthesise to provide the organic

products to the coral, while the coral provides a sheltered location and access to sunshine.(Section 12.6)

- Root nodules found on the roots of legumes (peas, beans and clover). Nodules contain Rhizobium bacteria which fix nitrogen from the air and supply nitrogen compounds to the plants. The bacteria receive nutrients in return.

Predation: one species benefits, the other does not

When a lion, a hyena or a vulture feeds on a zebra they will benefit but the zebra population does not. **Predation** benefits the predator but not the prey species. Adaptations of predators and their prey are discussed later in this chapter.

Herbivory: one species benefits, the other does not

Herbivory is the process of an animal feeding on a plant. Different herbivores may feed on the same plant species at different times. For example, deer or moose may browse on the branches or bark of a tree in the autumn and winter but insect caterpillars attack the leaves in spring. In both cases the effect on the tree is a negative one.

KEY POINTS

herbivory refers to consumption of plant tissue for the purposes of nutrition.

predation is the process of an animal catching killing and eating a prey species.

Parasitism and pathogenic relationships: one species benefits, the other is harmed

Parasitism is a symbiotic interaction in which the parasite benefits while the host is harmed. Some parasites, such as fleas and lice, live on the surface of their host, while others live inside their host's body. The roundworm (*Ascaris* spp.) is an internal parasite that lives and feeds in the human intestine. The worms produce huge numbers of eggs, which pass out in the host's feces to the environment. People may be infected by swallowing the eggs in contaminated food or water. Some parasites eventually kill their host but most do not, as it would also result in the death of the parasite. More examples of parasites are described in Section 12.1.

Pathogens are any biological agent that causes disease or illness to its host. Human pathogens include strains of bacteria like *Salmonella*, *Listeria* and *Escherichia* coli which cause infections of the digestive system. Pathogens reproduce and thrive in their hosts' bodies but disrupt the normal physiology of the animal or plant they infect, causing them harm.

12.3.5 Chemical inhibition: allelopathy

Allelopathy is the chemical inhibition of one species by another and is important to plants, fungi and bacteria, which use it as a competitive strategy.

Plants release chemicals which inhibit the growth of other plants or prevent their seeds from germinating.

One example of this is the black walnut tree and butternut tree. The roots of these trees release a chemical called juglone which affects the roots and stems of other plants attempting to grow nearby. It also inhibits the other plants' seed germination (Section 12.1). The chemical is also found in all green and growing parts of the trees and in the unripe outer shells of their nuts. Juglone has been shown to inhibit respiration in plants that are sensitive to it and these affected plants cannot obtain energy for their metabolic activity.

Many plants also have germination-inhibiting substances in a wide variety of parts including fruits, leaves, bulbs and roots. These are usually non-specific in their effects and prevent the growth of their own seeds until conditions are favourable and may affect other seeds. For example, there is evidence that *Rhododendron ponticum* inhibits germination and growth of seeds and seedlings nearby. The main inhibitors are hydrogen cyanide, ammonia and ethylene, although the sap or extracts of some plants affect the local pH of the soil and inhibit seed germination.

Fungi and bacteria also release substances that kill or restrict the growth of other microorganisms. We use many of these substances as antibiotics to treat human bacterial infections

because they are chemicals that kill or inhibit the growth of bacteria. All such chemicals are produced naturally by soil bacteria and fungi to enable them to compete with other microbes in their habitats. Penicillin is produced by a fungus *Penicillium*, and the antibiotics streptomycin, chloramphenicol and tetracycline are all derived from substances produced by soil organisms (Figure 12.3.9).

Figure 12.3.9: Four different antibiotic tablets have been placed on a petri dish in which bacteria are growing. Each has established a zone of inhibition around it as it releases antibiotic and kills bacteria.

Actinomycetes are fungi found in the soil that produce streptomycin and many of the antibiotics used in medicine today. Fungi also produce important inhibitors including cephalosporins, a class of antibiotics originally derived from the fungus *Acremonium*, and griseofulvin, which is produced by *Penicillium griseofulvum* to kill competing fungi. We use griseofulvin to treat fungal infections such as ringworm and athlete's foot.

No overall effect

Sometimes organisms or populations have no overall effect on one another. If there are no interactions or if direct influence is avoided then the organisms will coexist. This type of interaction enables the biodiversity of the ecosystem to increase.

Ecological relationships like these all affect the way organisms are distributed in an ecosystem and how many organisms can survive.

12.3.6 Features of relationships between predators, prey and plants

Predators catch, kill and eat other animals, known as their prey, to obtain energy and nutrients. The number of predators and their prey in an ecosystem will fluctuate (that is, increase and decrease) in relation to each other. If there are large numbers of predators the prey species will decrease in number. Likewise, if there are too few predators the population of the prey species may grow in size and could have the potential to damage or overgraze other areas of the ecosystem.

The relationship between the numbers of predators and their prey can be modelled mathematically. Data can also be collected from observations of predators and their prey in their habitats over long periods of time.

Predator–prey relationships

A well studied example of the effects of predation is that of the Canadian lynx, which feeds on the Arctic hare. Data were collected over a period of almost 100 years in the undisturbed cold, northern forests of Canada. The numbers of predator and prey fluctuated over the years with changes in the hare population being followed by corresponding changes in the numbers of lynx (Figure 12.3.10).

Snowshoe hare is the primary food of the lynx so the two population cycles are closely linked. When the hare population is high, lynx will eat about two hares every 3 days. Hare populations across the forest fluctuate in a cycle that ecologists have followed and which lasts 8–11 years. When their numbers are highest, snowshoe hares can reach a density of up to 1500

animals per km^2 and the habitat cannot support this many animals. With more hares in the forest, more are eaten as predation increases and starvation sets in as the hares overgraze the area. As this happens the hare population starts to decline.

When the hare population reaches its lowest level, it will remain stable for several years. The food plants slowly recover and the hare population starts to increase again. Since hares have several litters each year, the population increases rapidly. After 1 or 2 years at high densities, the hare population cycle repeats itself.

When the snowshoe hare population declines, the lynx population also declines after a lag of 1 or 2 years. At first as hare numbers start to decline, lynx continue to feed well because they can easily catch the starving hares.

But when hares become scarce, lynx numbers fall. Lynx do not have big reserves of fat and they starve in the freezing temperatures. Malnourishment significantly affects lynx reproduction and population levels. Females in poor condition are less likely to breed and produce litters.

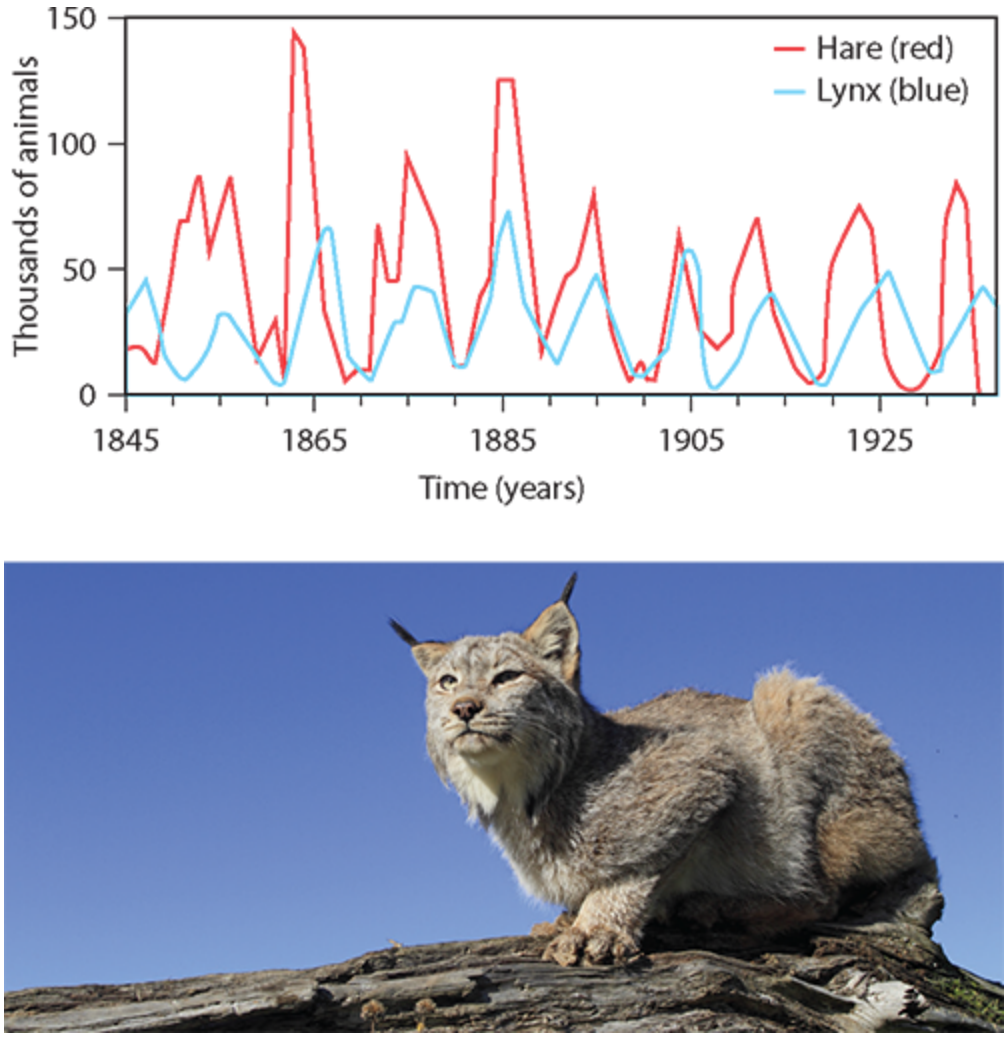


Figure 12.3.10: Changes in the populations of the Canadian lynx and the Arctic hare over time.

Their litters are smaller and many kittens die soon after birth. For a period of up to 5 years, few kittens survive to adulthood.

Avoiding predation

Prey animals do all they can to avoid being eaten and have a range of different defence mechanisms to protect themselves. Some of these defences are physical and some are chemical. Others are behavioural, such as hiding or running away.

Some examples of physical defences are outlined here**.**

- A porcupine's quills are used to deter leopards or other large predators (Figure 12.3.11).
- The crown-of-thorns starfish has venomous spines which give it physical and chemical protection.
- The shells of snails and other molluscs protect their soft bodies inside.
- Camouflage allows prey species to conceal themselves from passing predators.

Chemical defences are equally variable.

- The slow loris (Figure 12.3.12) is too slow to avoid predators by moving, so it coats its fur in poison which tastes unpleasant and discourages any predator that tries to eat it.

Figure 12.3.11: A porcupine can raise its quills to fend off large predators.

Figure 12.3.12: The slow-moving slow loris coats its own fur in a foul-tasting poison to protect itself from predators.

Figure 12.3.13: The bombardier beetle shoots a spray of hot chemicals into the path of approaching predators.

- Bombardier beetles (Figure 12.3.13) release a noxious spray of hydrogen peroxide and enzymes from their anus in an explosive burst. This is a mixture of hydrogen peroxide and enzymes that reaches a temperature of 100°C.

- Skunks release foul-smelling, foamy spray from their anal glands to drive away any animal that might threaten them.
- Millipedes can release hydrochloric acid, which can cause burns and irritation and has a bad odour.

Avoiding herbivory

Herbivory is the process of feeding on plants. Plants are food providers and parts of a plant will be lost as herbivores feed on them to grow and thrive. A single plant may provide leaves for browsing animals, fruits and seeds for birds, and roots for burrowing animals. Some plants have defence mechanisms against herbivores to protect themselves and stop or deter herbivores that try to eat them. These defences include spines, thorns and thick woody tissue. Many plants produce toxins that have a bad taste or contain substances that will harm a herbivore.

- Bougainvillea and roses both have thorns, which are modified stems.
- Spines are modified leaves or parts of leaves, such as extensions of leaf veins, and protect many types of cacti.
- Trichomes are much smaller structures formed from the outer layers of plant epidermis. Examples include the stinging structures of nettles which defend them against small insects and people.
- Some of the strategies which are used by coniferous trees are shown in Figure 12.3.14.

EXAM TIP

Remember that if an examination question asks you to ‘discuss’, it is important to present alternative points of view.

If you are asked to ‘explain’ a concept or situation, remember to include the steps in the process and write about them in some detail.

TEST YOUR UNDERSTANDING

21 List three different prey species and outline their methods of avoiding predation.

22 Define herbivory.

23 Suggest a defence strategy a plant might use to avoid being eaten.

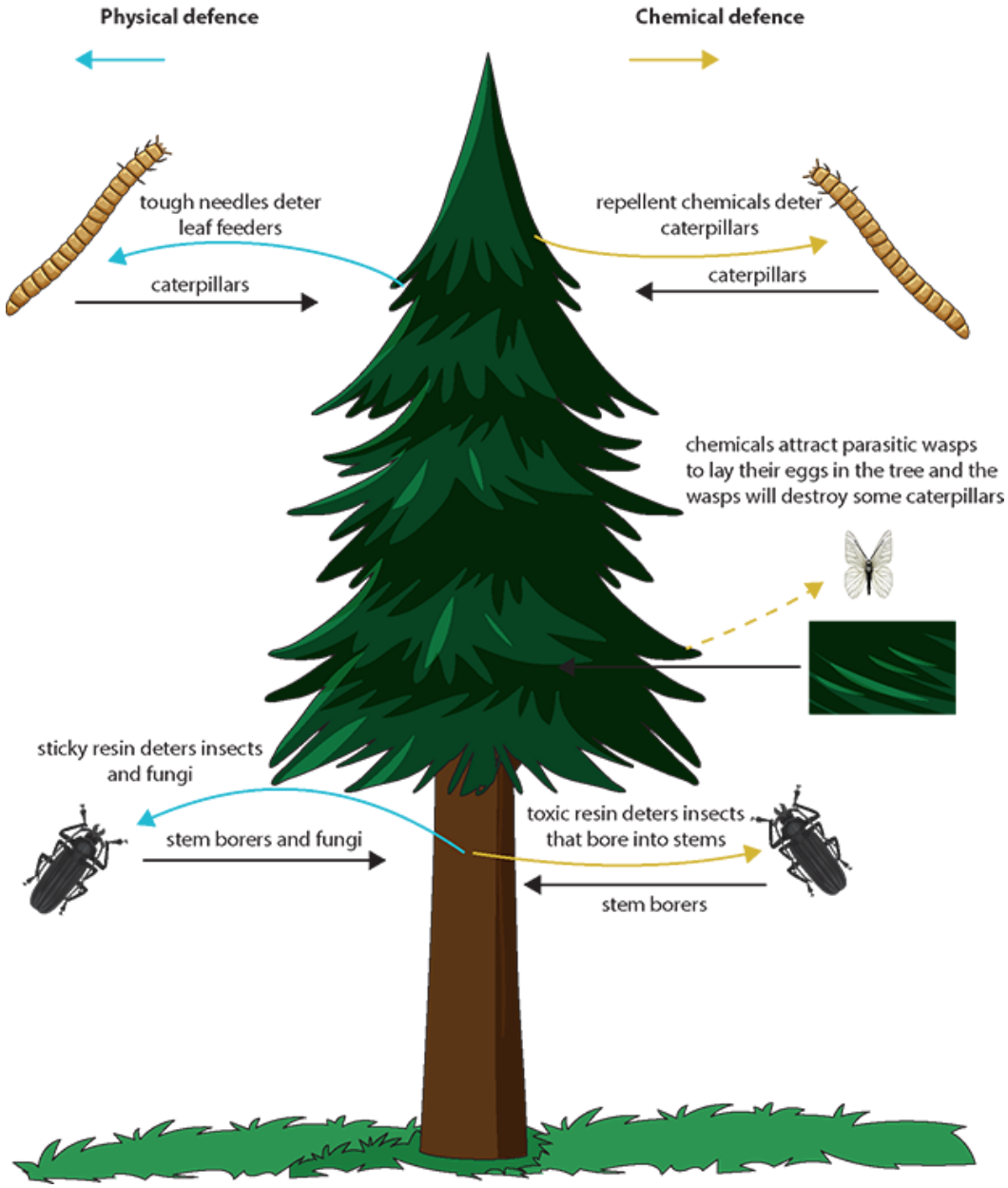


Figure 12.3.14: Some of the many strategies that plants, such as this pine tree, may have to deter herbivores that feed on them.

12.3.7 Cooperative interactions

Interspecific interactions

One type of cooperation between species is cooperative interspecific interaction, usually referred to as mutualism or mutualistic symbiosis. Earlier in this chapter we considered the cases of lichens and of plovers which feed on parasites on the body of crocodiles. Mutualism benefits both partners in an interaction, and there are many more examples of this type of relationship.

Plants form mutualistic relationships with fungi, forming **mycorrhizae**. A fungus will live either in or around the plant's root system and form part of that system (Figure 12.3.15). Mycorrhizae are important in providing the plant with nutrition and also affect the condition of the soil. As the plant photosynthesises it supplies the fungus with organic molecules such as sugars, while the fungus provides the plant with water and minerals from the soil.

Around 70% of plants have mycorrhizae with fungal hyphae (long branching filaments of the fungus, collectively called a mycelium) that penetrate the cells of the plants' roots. Inside the cells the hyphae branch to form arbuscules where nutrient exchange can take place (see Figure 12.3.15, right-hand side). In other plant species the mycorrhizae form a mycelium over the surface of the root and penetrate between the root cells (see Figure 12.3.15, left-hand side).

Both the plant and the fungus benefit from these arrangements. Plants with mycorrhizae can resist pathogens in the soil more effectively and some mycorrhizae release enzymes that combat soil pests such as nematode worms.

Bees and other pollinators of flowering plants also have mutualistic relationships in which both species benefit. Flowers provide bees with nectar and pollen, which worker bees collect to feed their colonies. Bees provide flowers with the means to reproduce by pollinating them as they spread pollen from flower to flower.

The majority of pollinators are insects – bees, butterflies, moths, flies, wasps and beetles – but many birds and bats are also important for some species of plant.

Plants always receive pollination services through the interaction, but the benefits received by the animal pollinators vary. Most receive food in the form of pollen or nectar, but some bees also use waxes and resins from flowers to build their hives.

Intraspecific interactions

Cooperative intraspecific relationships form between members of the same species. This may mean living together in small groups, such as lions living in families (prides), or in larger groups, such as zebra or wildebeest living together in a herd.

These relationships have both advantages and disadvantages for species that form such groups.

Advantages

There are many advantages to living with members of your own species. As well as social interactions, being a member of a herd or group allows protection from predators and the environment. It also gives access to information from other members of the group.

KEY POINT

cooperative interspecific interaction refers to the interaction between two species in which both species benefit, also known as mutualism.

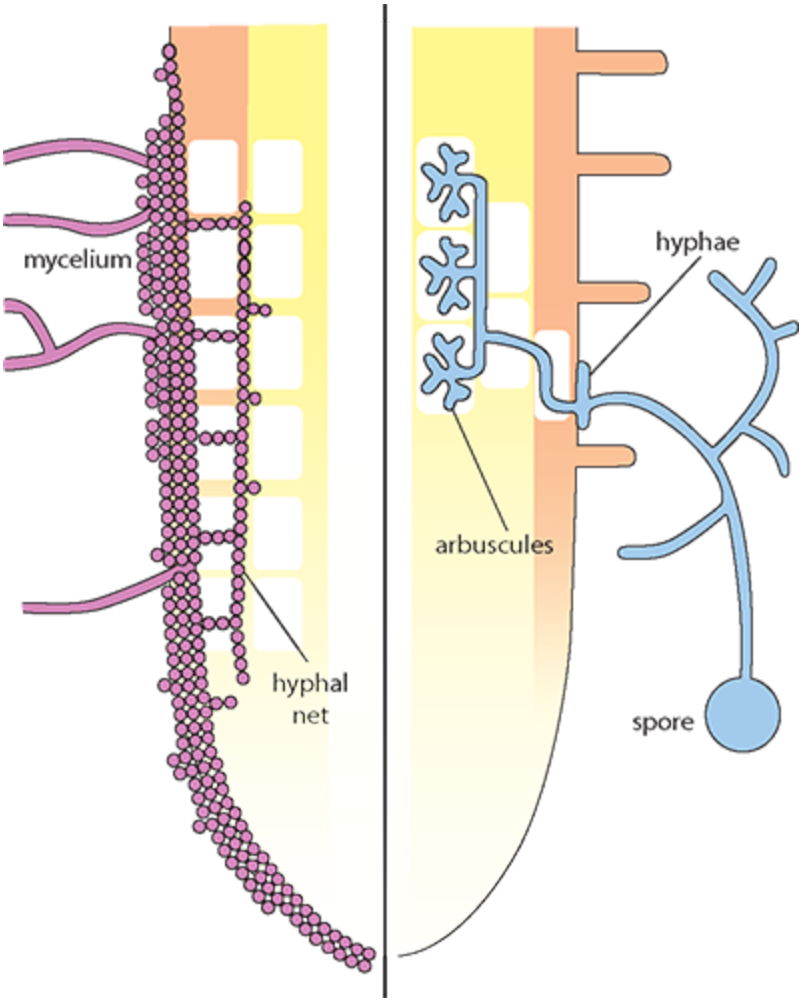


Figure 12.3.15: This diagram shows two ways in which fungi can form mycorrhizae. Some types of fungi have hyphae which penetrate root cells (right-hand side) others grow between the cells (left-hand side).

All these factors can increase the chances of survival for an individual in the group.

- **Protection from predation:** an individual in a group may be better protected from predation. Some members of the group can act as lookouts and watch for approaching predators. Herd animals such as zebra run together and confuse approaching predators by scattering in different directions. Some bird species use similar tactics. For example, a murmuration is a large group of starlings which form flocks and fly together in complex formations. In the largest, most densely packed groups, starlings are well protected from attacks by predatory hawks. This effect is known as the confusion effect and predators are less likely to succeed in catching prey as the prey group size increases.
- **Protection from extreme environments:** groups of emperor penguins huddle together in freezing conditions, protecting the individuals at the centre of the group. Penguins take turns on the outer edges of the huddle before moving to take their place in the warmer central area (Section 11.4).
- **Reproduction advantages:** animals living in groups are always closer to other members of their species, which are likely to be suitable mates for breeding. Once young are born other members of the group can assist with caring for them. Bats will suckle the young of other females from their roost if mothers are absent, and female lions often live in groups with their sisters who collectively guard and feed all the cubs in the pride.

Disadvantages

Living with members of your own species also has some disadvantages. There will be more competition for food and greater vulnerability to diseases that spread rapidly between

organisms that are close together. Some of these factors can be detrimental to an individual's survival.

- **Availability of resources:** if many members of the same species live together they will all require similar food, space and access to mates. Intraspecific competition will occur and decrease the availability of these resources. For example, in the case of the snowshoe hare, described in the section on predator – prey relationships, there will be insufficient food for all members of the group if population numbers rise too much.
- **Risk of infections:** living closely together will increase the likelihood that disease can pass quickly between individuals. Pneumonia can spread among populations of bighorn sheep as well as animals such as goats. Tuberculosis occurs and spreads through herds of buffalo, bison and deer, and many other intraspecific interactions cause the spread of parasites such as tapeworms, fleas and ticks.
- **Attracting consumers:** herbivores are more likely to feed where there is an abundance of their preferred food plants growing together. For example, small birds are more likely to visit bushes that are covered in berries, if the bushes are growing in dense clumps. Similarly predatory animals will be attracted to feed where there are many of their prey species living closely together.

Limiting factors and populations

Top-down limiting factors

Top-down limiting factors are those that involve an organism higher up the food chain limiting the numbers of a species at a

lower trophic level, usually through predation or herbivory. One example of this is the control of the small algae (Fucus spp.) on a rocky shore by grazing limpets (Section 11.4 Nature of Science). Another example is the control of kelp forests due to the impact of sea otters. Otters feed on sea urchins, which use kelp as a source of food so if sea otter numbers fall, the sea urchin populations expand and reduce the kelp forest. Ecosystems such as those where sea otters and limpets are found are not controlled by the productivity of the primary producer but rather by a top predator or major herbivore acting as a keystone species.

Bottom-up limiting factors

Bottom-up control by limiting factors occurs where the nutrient supply and productivity of primary producers (plants and phytoplankton) control the structure of the ecosystem. In marine coastal ecosystems, plankton populations depend on and are controlled by the availability of nutrients. Phytoplankton populations increase so that large growths known as algal blooms appear when nutrients are abundant. This happens when sea currents cause upwelling, which brings nutrients to the surface where they are accessible to phytoplankton. The abundant growth of phytoplankton is then controlled by top-down control by herbivores, which use it as food. Algal blooms are also controlled by bottom-up control at times when nutrients are in short supply or in places where currents do not bring nutrients to the surface.

Limiting factors and ecosystem stability

Bottom-up and top-down control tends to keep a stable population at the carrying capacity of the ecosystem. The bottom-up resources set the limit for the maximum sustainable population, while top-down control removes individuals from a

large population, with the result that resources are not over-exploited. The concept of internal control of populations by interactions between them is a key argument for the conservation of ecosystems.

12.3.8 Keystone species

A keystone species is one which has a disproportionate impact on an ecosystem and strongly affects community structure. Examples of keystone species include the sea otter, elephants and the prairie dog.

Keystone species are an example of top-down limiting factors. These factors involve an organism higher up the food chain limiting the numbers of a species at a lower trophic level, usually through predation or herbivory. One example is the control of kelp forests by the feeding activities of sea otters. Kelp forests are underwater areas found around many coastal areas. A high density of kelp plants results in a productive and dynamic ecosystem that is an important source of nourishment and shelter for fish and other sea animals. Kelp is eaten by sea urchins and sea otters that then feed on sea urchins.

But sea otters have thick, rich fur and became a target for hunters who, by the 1900 s, had brought the animal close to extinction. In the end, an international ban on sea otter hunting was imposed in 1911, saving the animal from complete extinction. Researchers discovered that around coastal islands that lacked sea otters, sea urchins had increased in size and in numbers with devastating consequences. The forests of kelp had disappeared, and large sea urchins were found on the empty sea floor, having consumed all the kelp plants in the area. Close to islands where sea otters had survived, or had been reintroduced, the kelp forests were alive and healthy.

Figure 12.3.16: Sea otters (*Enhydra lutris*) feed on sea urchins and are a keystone species

Sea otters are keystone species because if sea otter numbers fall, the sea urchin populations expand and reduce the kelp forest. The ecosystem is not controlled by the productivity of the autotrophic organisms but rather by a top predator acting as a keystone species. By almost wiping out sea otters, humans upset vital feeding relationships. Sea otters remain an endangered species but now we realise that high otter numbers mean fewer sea urchins, which in turn mean abundant healthy kelp forests, and a more diverse ecosystem.

INTERNATIONAL MINDEDNESS

The human population

The graph of human population growth has not followed the pattern shown in Figure 12.3.5. Since the evolution of humans, no plateau has been reached, and instead the global population continues to rise exponentially. The natural carrying capacity of Earth has been manipulated as humans have found ever more technologically advanced ways to produce food and extract resources from the environment. Humans have also overcome other limiting factors, like disease, through improved medicine, and colonised almost every part of the planet. But over-population has led to poor living conditions and environmental degradation in many parts of the world.

EXAM TIP

You should be able to draw and interpret population growth curves. The curves can be drawn for organisms such as yeast cells in a flask or larger organisms which colonise or inhabit new areas.

TEST YOUR UNDERSTANDING

24 Define carrying capacity.

25 Give an example of a density-independent factor.

26 List two advantages and two disadvantages for a species which lives in a large group.

27 Sketch a growth curve for a population of yeast cells that are growing in a vessel of nutrient fluid.

REFLECTION

Could I summarise the importance of ecological relationships in the survival of organisms? Which ideas would I describe first to a classmate?

Links

- What roles do predators and herbivores have in the transfer of energy and matter? (Chapter 12.2)
- How important is competition in ecological succession? (Chapter 12.4)

12.4 Stability, change and succession in ecosystems

LEARNING OBJECTIVES

In this section you will:

- recognise the factors that maintain stability in an ecosystem
- understand how human actions can lead to a tipping point in ecosystem stability
- define a mesocosm and learn how it can used to study ecosystem sustainability
- learn that a sustainable ecosystem can support itself and recover from outside influences
- understand how farming leads to increased land clearance and increased carbon emissions and loss of biodiversity
- understand the effect of human activity on biochemical cycles and that the release of carbon dioxide is linked to increased global temperatures
- learn that a sustainable ecosystem has high biodiversity
- recognise examples of sustainability in fishing and forestry

define ecological succession as a predictable change in species structure over time

- understand that a climax community is a state of equilibrium in a stable and functioning community
- learn that primary succession begins in a previously unoccupied area
- discover that human influence can interrupt a succession and lead to a plagioclimax
- Learn that in some communities there is a cyclical succession of communities.

GUIDING QUESTIONS

- How do human activities threaten the sustainability of ecosystems?
- Why is it important for nutrients such as carbon to be recycled?
- To what extent is increased biodiversity important?

12.4.1 Stability, change and succession

Ecosystems will remain stable if they have a sufficient supply of energy, nutrients, genetic diversity and the climate remains stable. A stable ecosystem needs to be resilient to change and this may mean either it has:

- inertia: resistance to change
- resilience: the ability to recover from change
- succession: the replacement of species by another or others.

Ocean ecosystems are very stable because they do not have much variation in productivity from year to year, they are resistant to natural and human disturbance and are resistant to invasion by alien species. Ocean ecosystems have had a stable structure for tens of millions of years. Rainforest ecosystems are also very stable when undisturbed, some are ancient and have been present in their current form for more than 70 million years.

Human interference is threatening some of these ecosystems by upsetting their ability to recover from change. The Amazon rainforest is under threat and it is feared that the removal of trees will lead to a tipping point beyond which the forest cannot recover.

The rapid growth of animal agriculture is the leading cause of deforestation. It is estimated that 70% of the Amazon rainforest has already been destroyed and is now used for pasture and to grow crops. One of the crops grown in the rainforest is soybeans used for animal feed. Tropical deforestation and forest clearing contribute to climate change as carbon emissions are increased. They also cause loss of biodiversity, flooding and soil

degradation. Forests are often cleared by burning and this also increases carbon emissions and as well as reducing the capacity of an area to absorb carbon dioxide during photosynthesis.

Some scientists have suggested that to preserve maximum species diversity, humans should change from animal-based agriculture to a greater reliance on plant-based agriculture. Across the world, more land is used to produce feed for farmed animals than to produce plants for human consumption. It is estimated that it takes 20 times less land to feed a person whose diet is mainly plant-based than to feed meat-eaters.

Tipping points in ecosystem stability

Deforestation in the Amazon rainforest has led researchers and scientists to suggest that the forest is losing its resilience and the ability to recover from the effects of droughts, deforestation and climate change. Surveys carried out over more than 30 years using up to 200 satellite images each year show that there are signs of poor growth in more than 75% of the forest and the fear is that trees will die in large areas. The Amazon rainforest holds between 90 and 140 billion tons of carbon and traps carbon dioxide so that it does not contribute to global warming. Some recent studies have shown that parts of it are now emitting more carbon dioxide than can be absorbed. Forest trees also release water vapour into the air as they transpire and this water affects air flow, rainfall and the temperature of the atmosphere. If the cycle of damage continues a tipping point could be reached when the forest is no longer able to recover at all. When this might happen will depend on how we work to prevent it. No one is certain how much rainforest will be sufficient to maintain suitable levels of carbon capture and transpiration but about one fifth of the forest has already been lost compared to pre-industrial times so stopping the removal of trees and preventing

deforestation would help. If the critical tipping point is reached the effect on climate change and biodiversity will be very serious. It could be that in less than 50 years the Amazon might become a savannah ecosystem of grassland and trees.

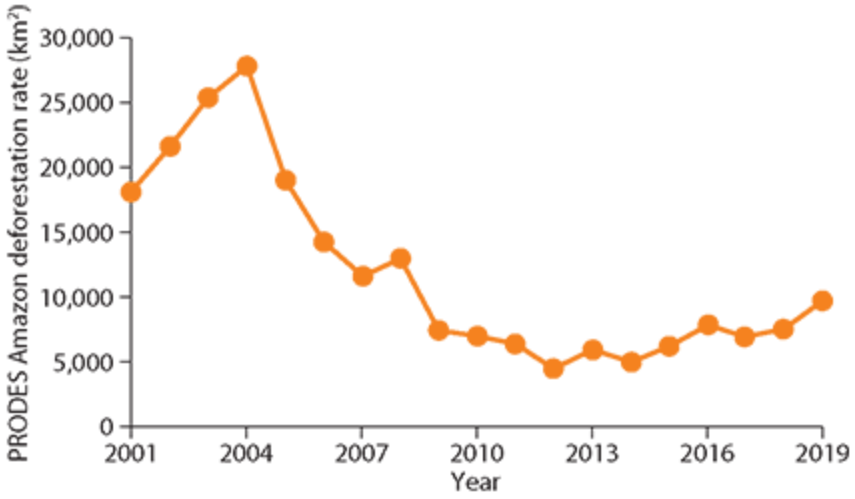


Figure 12.4.1: The deforestation of the Amazon rainforest this century is causing concern

Mesocosms

A **mesocosm** is a small-scale, self-sustaining natural system that can be set up in a laboratory or outdoors to mirror conditions that may occur on a larger scale. We can use a mesocosm to study part of an ecosystem under controlled conditions and draw inferences about how the ecosystem works in the natural environment. Indoor mesocosms, such as simple growth chambers (Figure 12.4.2), provide experimenters with a way to control variables such as temperature and light. In this way the effect of environmental change on species and communities can be predicted.

Open mesocosms have been used outdoors to study the feeding habits of fish when in the presence of different proportions of

plankton and competitors.

KEY POINTS

Tipping point is the point beyond which an ecosystem becomes unstable and can no longer recover from change

mesocosms are small-scale enclosed environments that allow part of a natural environment to be observed under controlled conditions.

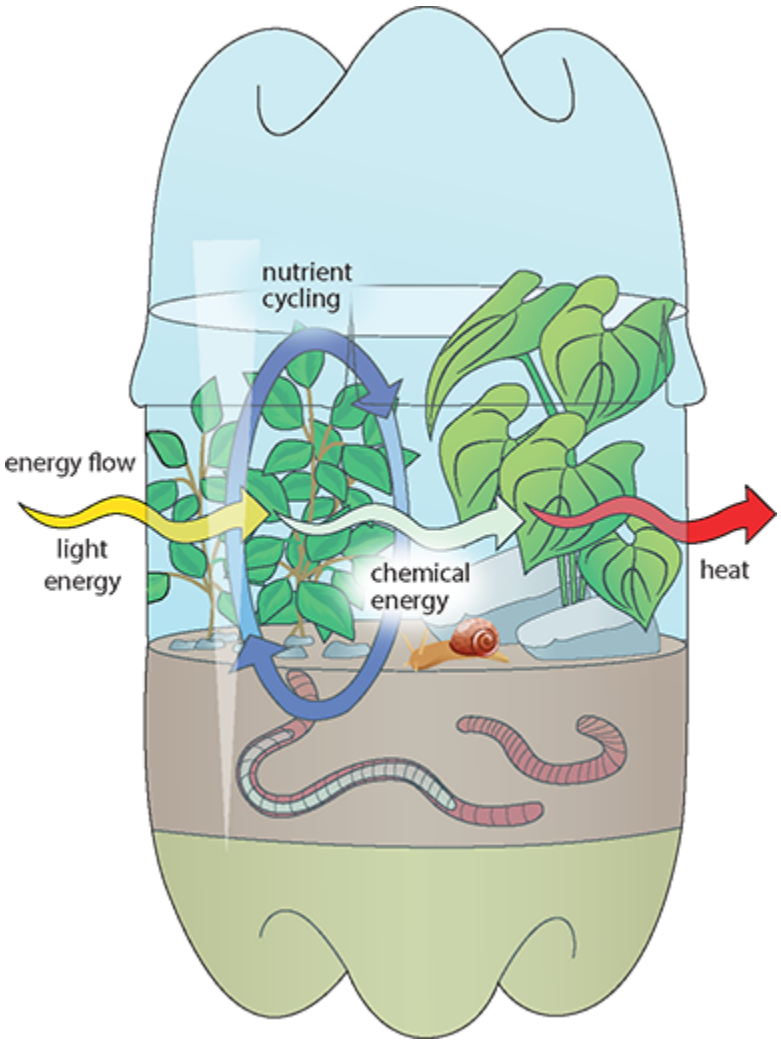


Figure 12.4.2: A simple mesocosm can be set up in a school laboratory.

These mesocosms are usually small enough to allow researchers to monitor important variables and can take account of daily cycles or tides. In shallow ponds, cylinder-shaped mesocosms can be used to assess the effect of temperature change or carbon dioxide levels. Using data from such systems we can attempt to predict the possible effects of environmental change on organisms or communities.

12.4.2 Impact of agriculture

Farming animals and growing crops to feed them uses land. As the human population increases there is an increased need for food. Therefore, more land must be cleared for agriculture. Grassland and fields containing only one crop (monocultures) reduce the natural biodiversity of an ecosystem.

Assessing the sustainability of harvesting resources from natural ecosystems

KEY POINT

sustainability or sustainable yield, depends on the rate of harvesting of a plant or animal being lower that the rate of its replacement.

In forestry, sustainable harvesting is a way of harvesting from a forest that provides a constant supply of wood for human use, but which also protects trees so there are sufficient for harvesting in the future. Sustainably-managed forests meet the needs of wildlife and many support other important features of an ecosystem, such as carbon storage and reducing the risk of flooding.

Here are a few examples of methods used in sustainable **forestry**:

1 **Selective logging** – only harvesting certain trees from the forest instead of removing all of the trees at once. Selective logging helps preserve the balance of the forest and maintain its valuable assets.

2 **Maturity** – All trees have value but that value increases as a tree matures. Allowing young trees time to mature helps protect the long-term value of the forest.

3 **Replanting** – When a tree is harvested a new sapling is planted in its place. This simple method helps ensure the forest is being restored to its original state.

Some examples of the most sustainable hardwood trees are black cherry, willow, aspen, elm, cottonwood, and soft maple trees. These species are used for timber and are fast-growing and absorb large amounts of carbon dioxide.

In **fishing** sustainable yield is the amount of fish that can be caught on a regular basis but still leave sufficient fish of the species to reproduce and maintain the population. Fish stocks need to renew themselves due to losses from both fishing and natural causes. It is important that small fish are allowed to reach maturity and reproduce.

Methods used in sustainable fishing include:

1 Reducing overfishing by legislation and only allow a certain quota of endangered fish species to be caught.

2 Avoiding damage to marine habitats. Some fishing methods – like bottom trawling and dredging – involve scraping heavy machinery along the seafloor and longlines used to catch bluefin tuna can trap birds, turtles and swordfish. These methods can be very destructive to marine habitats.

3 Avoiding bycatch – during fishing, other animals are accidentally caught along with the target species these are known as bycatch. These animals can include dolphins, turtles, sharks and whales, as well as small young fish. In many parts of the world, bycatch are usually thrown back

into the sea either dead or dying. Approximately 10% of fish caught worldwide is bycatch.

An example of sustainable fishing involves the Chilean seabass. This fish became a popular restaurant fish in the 1990s. The species lives in the South Pacific and South Atlantic oceans and although fishing in the area is regulated by international agreements, illegal fishing still takes place. At that time, too many fish were caught and the average size of fish declined as poaching and severe overfishing led the species close to extinction. Now, the Marine Stewardship Council (MSC), an international environmental agency based in London, has certified as sustainable a small Chilean sea bass fishery in the South Georgia and South Sandwich Islands near Antarctica. In addition, some sea bass are now farmed in Europe and other areas. Shops and restaurants are encouraged to check that the fish they buy and sell bears the mark of the MSC.

Sustainable agriculture

The aim of sustainable agriculture is to provide the food and plants people need now without damaging the ability of future generations to do the same. For many decades people have used industrial methods of farming with large farms growing monocultures of crops year after year. These farms require large areas of land, often cleared forests, as well as large amounts of fertiliser and pesticides that lead to damage to the soil, water, air and climate. This style of farming degrades the resources it depends on.

To keep farming sustainable we must take care of the land and natural systems that farms rely on. Some of the important things that must be done include:

1 Keeping the soil healthy and preventing erosion

2 Using water resources carefully

3 Keeping water and air pollution to a minimum

4 Storing carbon on farms

5 Protecting farms from extreme weather

6 Encouraging biodiversity

You can read more about aspects of agriculture including **eutrophication**, **biomagnification** and pollution in Section 12.5, as well as human efforts to repair some of the damage that has been caused.

TEST YOUR UNDERSTANDING

28 List 3 factors that are needed to keep an ecosystem stable

29 How can we use the resources of a woodland sustainably?

30 What do the term 'industrial' farming and monoculture mean?

KEY POINT

eutrophication is natural or artificial addition of nitrates and phosphate to water, resulting in depletion of the oxygen content.

12.4.3 Impact on biogeochemical cycles

Humans also have important effects on biogeochemical cycles. Figure 12.2.3 in Section 12.2 shows the details of the carbon cycle. Humans interfere with the carbon cycle as we use non-renewable resources such as coal, oil and gas. These non-renewable resources are natural resources that cannot be replaced at a rate that could keep up with human consumption. Non-renewable sources of energy will eventually run out because it takes millions of years for them to be formed. Using these carbon-containing resources also contributes to the release of pollutants such as oxides of sulphur and carbon into the atmosphere.

Renewable resources such as wind energy, solar energy and thermal energy from the ground are abundant and cannot be exhausted as coal and oil will be. Renewable resources are natural resources which are replenished by natural processes, or in recurring cycles, at a pace that keeps up with the rate at which humans use them.

KEY POINT

biogeochemical cycle the way in which an element moves between its biotic and abiotic forms and locations in the biosphere.

Carbon flux

Carbon flux is defined as the flow of carbon from one ‘carbon pool’ to another. The net difference between the carbon removal and the carbon addition can help us understand how we are affecting the carbon cycle. For the Earth’s atmosphere, carbon is

removed by plant growth, mineral formation and dissolving in the ocean, while it is added through respiration, burning of fossil fuels and volcanic activity. Scientists monitor carbon flux in order to build up a picture of changes and disturbances in the balance of the carbon cycle (Figure 12.4.3).

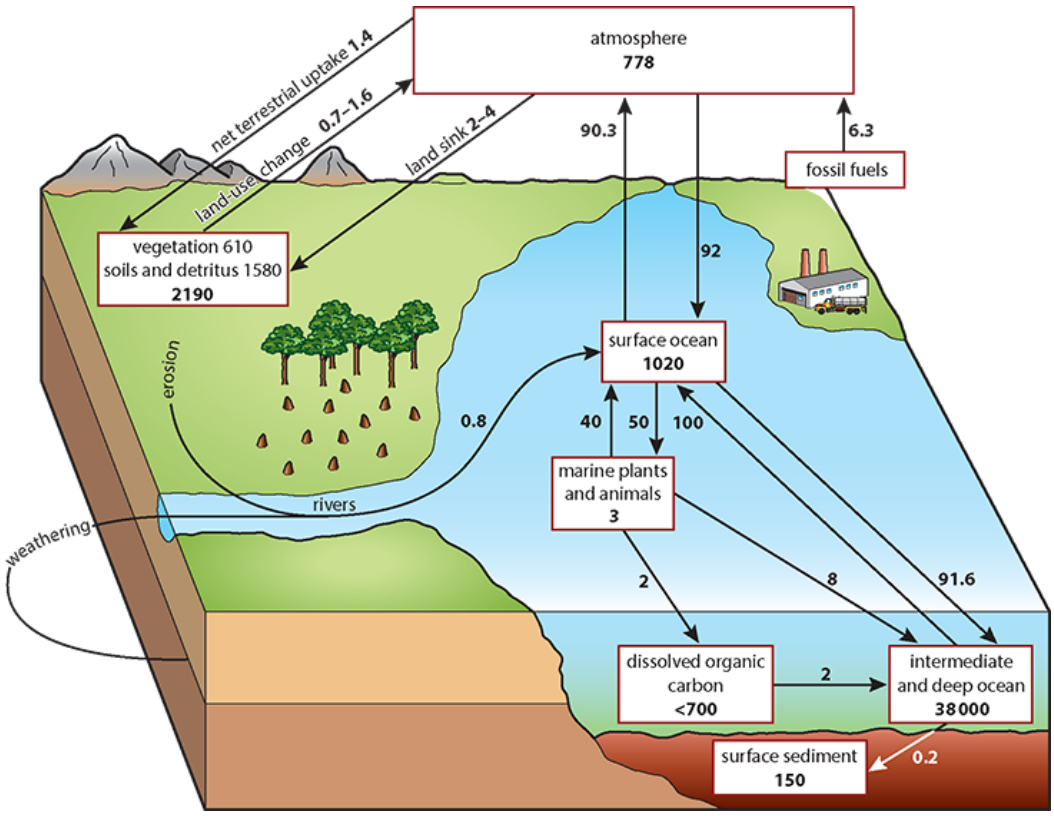


Figure 12.4.3: Diagram showing carbon fluxes in a natural system. Carbon stores (boxes) are shown in gigatonnes (Gt) and carbon fluxes (arrows) are shown in gigatonnes of carbon per year (Gt C·year−1^{-1}).

Carbon dioxide currently forms only 0.04% of the atmospheric gases but it plays a significant part in the **greenhouse effect**. Other greenhouse gases include water vapour, methane, oxides of nitrogen and fluorocarbons.

The human population has increased dramatically in recent history, with a consequent increase in demand for energy in industry, transport and homes. Most of this energy demand has been met by burning fossil fuels, mainly oil, coal and gas. Burning fossil fuels releases both carbon dioxide and oxides of nitrogen. This activity has raised the concentration of carbon dioxide in the Earth's atmosphere significantly since the mid-1800s, a period which has coincided with increasing industrialisation (Figure 12.4.4). Carbon dioxide concentration has risen by more than 20% since 1959 (Figures 12.4.5 and 12.4.6).

In the tropical regions of the world huge rainforests trap carbon dioxide through photosynthesis and have been important in maintaining the low level of atmospheric carbon dioxide. Humans have upset this balance by removing vast areas of forest for agriculture and timber production.

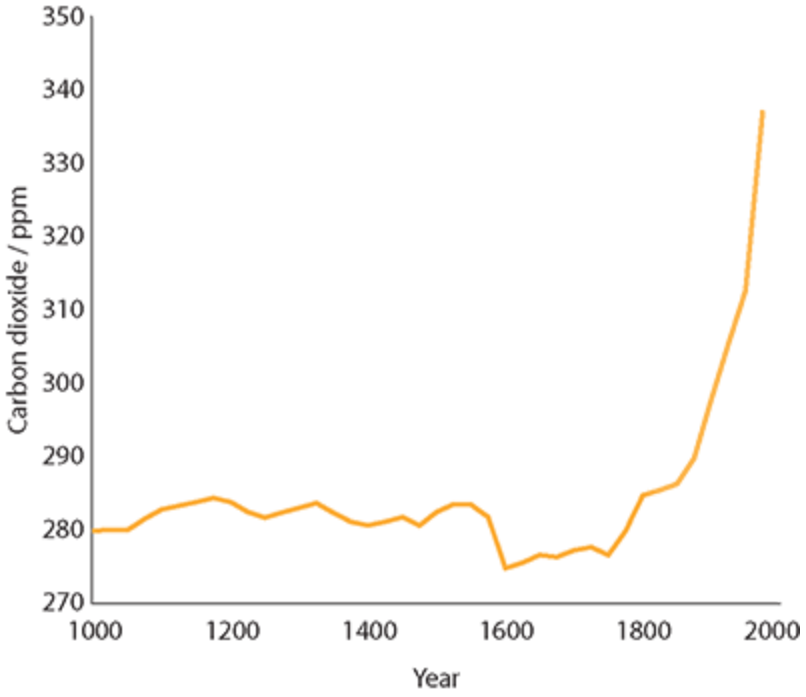


Figure 12.4.4: Graph to show the increase in carbon dioxide concentration in the atmosphere since the Industrial Revolution in the mid-1800s.

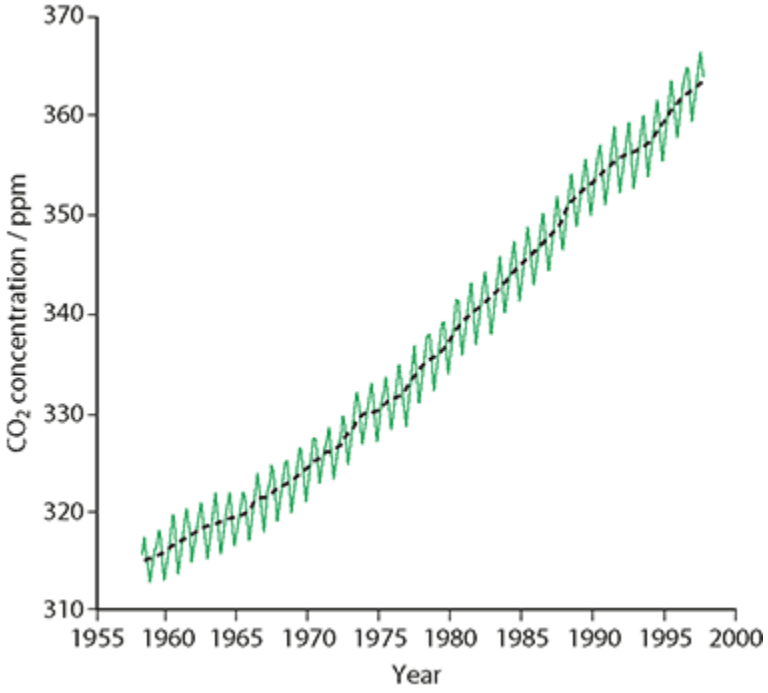


Figure 12.4.5: Atmospheric carbon dioxide concentration measured at monthly intervals in Hawaii.

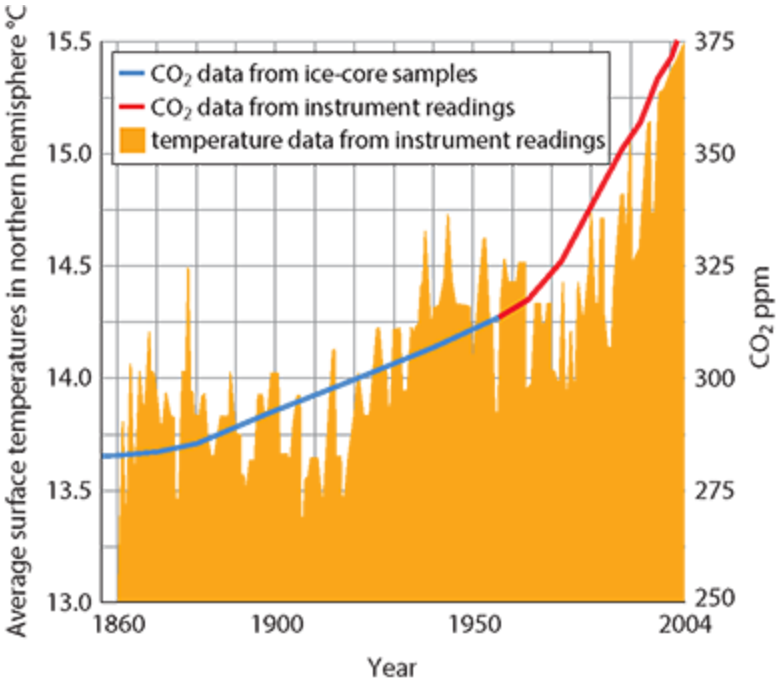


Figure 12.4.6: Graph showing carbon dioxide and temperature data from different sources. There is a correlation between rising carbon dioxide levels and global temperature which is evidence of global warming.

Forest destruction has multiple effects, but the most important ones for the atmosphere are the loss of carbon dioxide uptake by photosynthesis and the increase in carbon dioxide released from the rotting or burned vegetation.

The increase in carbon dioxide in the atmosphere has been linked to increased global temperatures. The Earth is kept warmer by the presence of carbon dioxide, water vapour and other gases in the atmosphere. These gases are known as greenhouse gases. Radiation of longer wavelengths from the Sun strikes the Earth and is reflected from its surface. Up to 85% of the radiation is reabsorbed by greenhouse gases before it has passed back into space. Energy is re-emitted towards the Earth so that the greenhouse gases keep the heat in the atmosphere. There is a

strong correlation between the concentration of carbon dioxide in the atmosphere and global temperature (Figure 12.4.6). Data on the carbon dioxide concentration in the atmosphere are collected from instruments at monitoring points all over the world. They are also taken from the study of ice cores removed from deep below the Earth's surface. Ice cores have been used since the 1950s to study our atmosphere and climate. Ice cores have been drilled from ice sheets worldwide, but especially from those in Greenland and Antarctica. Bubbles in the frozen ice preserve samples of the world's ancient atmosphere. These can be analysed to show how atmospheric carbon dioxide levels in the atmosphere have changed over the last 10 000 years (Section 12.5).

The size of the circles and arrows is proportional to the amount of nutrients stored or transferred.

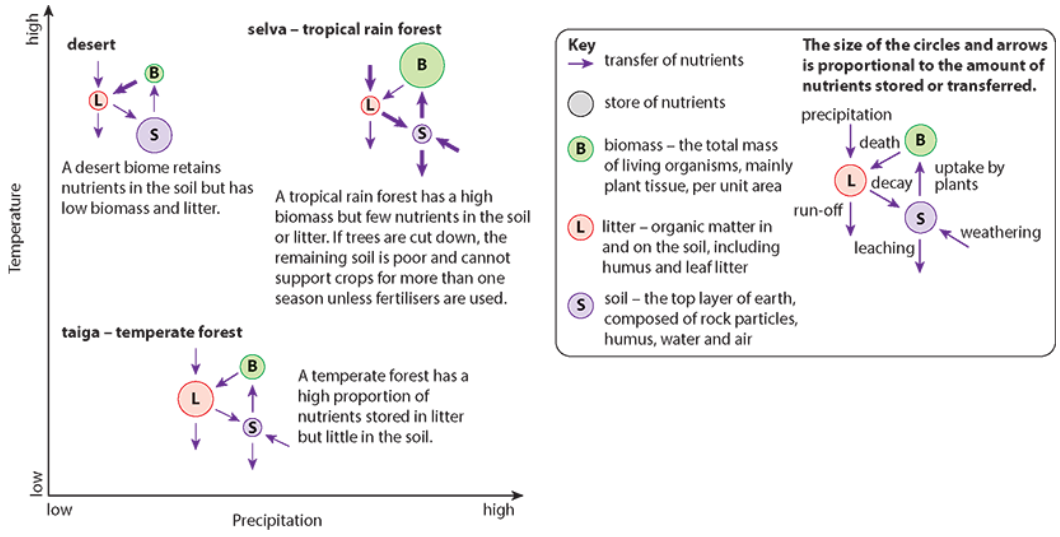


Figure 12.4.7: The model of nutrient cycling developed by Gersmehl.

NATURE OF SCIENCE

Using models to represent the natural world – Gersmehl diagrams

The United Nations Intergovernmental Science-In 1976, the geographer and scientist P.F. Gersmehl developed a model of nutrient cycling to highlight differences between ecosystems. His diagrams (Figure 12.4.7) show how nutrients are transferred and stored between three different parts of an ecosystem: the leaf litter (dead and decomposing leaves), biomass of organisms and the soil. As nutrients cycle in an ecosystem, there are interactions between the atmosphere and soil, and many food chains are involved. Nutrient cycles are different in different ecosystems and the rate of nutrient transfer is dependent on the amount of moisture, heat and vegetation, and the length of the growing season. Diagrams can be drawn for different ecosystems and provide insights into systems that have high levels of nutrients in the soil or a large biomass of organisms.

- A desert biome retains nutrients in the soil but has low biomass and litter.
- A tropical rain forest has a high biomass but few nutrients in the soil or litter. If trees are cut down, the remaining soil is poor and cannot support crops for more than one season unless fertilisers are used.
- A temperate forest has a high proportion of nutrients stored in litter but little in the soil.

12.4.4 The processes of succession

Ecological succession is the process of change that occurs in communities in a particular area over a period of time. Eventually the appearance of the whole area develops and changes. Succession involves interactions between both the biotic (living) and abiotic (non-living) components of the area. Abiotic factors limit the distribution of organisms and the organisms have an effect on the abiotic factors.

If an area of land is left bare as a result of an event such as a fire or land clearance, early 'pioneer' communities modify the physical environment, which, in turn, modifies the biotic community. This enables more species to move in and modify the physical environment still more, and so on until a stable situation is reached. The different stages of succession are known as seral stages and the final stable community, which remains unless there is further disturbance, is called a climax community.

KEY POINT

ecological succession is the predictable sequence of different communities that appear in a given habitat over a period of time.

Figure 12.4.8: Surtsey Island was formed in 1963 and has been studied since that time. A primary succession began as lichens and mosses began to colonise the bare volcanic rock.

A primary succession begins when an area of bare ground or rock, with no existing soil, is colonised for the first time. It may take place where a glacier is retreating and leaving a place where no organisms have existed before. Two examples of primary succession have been studied over many years by ecologists. These are an area on the Indonesian island of Krakatau that was left bare when the volcano Krakatoa erupted in 1883, and the newly formed volcanic island of Surtsey off the coast of Iceland, which formed in 1963 (Figure 12.4.8).

The first organisms to colonise bare rock are bacteria, lichens and mosses, which can settle on the rock surface. Lichens gradually break up the rocks and use dissolved minerals for

growth. As lichens die they decompose, leaving debris, which begins the formation of humus (fragments of organic matter) and soil. Low-growing lichens and mosses modify the environment sufficiently for seeds of grasses and small shrubs to start growing and these plants modify the land still further. A deeper layer of soil develops as plants die and decompose, and this soil can hold more moisture and contains more organic matter. Later, fast-growing trees will begin to grow and, as they extend their roots, the soil is bound together and protected from erosion. Eventually these trees will be replaced by slower-growing species, which form a climax community, usually after a period of about 100–200 years.

Secondary succession occurs where an ecosystem has already existed. In this case succession is initiated by a change in conditions, perhaps a land clearance, a fire or a landslip. An ecosystem has been established in the past but is replaced as conditions have changed. Soil is already present so secondary succession is usually much quicker than primary succession and a variety of plants such as annual grasses and low-growing perennials (plants that live for more than 2 years) can colonise rapidly. Shrubs and trees follow and eventually a stable climax community is established (Figure 12.4.9). Changes still take place when a climax community is present but they are slow and the system is more stable and resistant to change. In the later stages of succession, there are more consumers present, with more complex feeding interactions and food webs developing.

Stability and climax communities

A **climax community** can tolerate the existing environmental conditions and it has a wide diversity of different species. This type of sustainable ecosystem can support itself and provide requirements such as inorganic nutrients without external

influences. A climax community has the potential to remain stable and self-sustaining for a long period of time. There needs to be enough sunlight to provide energy for the autotrophs to photosynthesise, and the community itself must be able to maintain the recycling of inorganic materials within the abiotic environment. The climax community will remain stable as long as there are no adverse interferences from outside.

KEY POINTS

climax community is the stable stage at the end of a succession of communities in an ecosystem.

primary succession is colonisation of an area of bare ground or rock.

secondary succession is colonisation of an area of land that has been cleared (e.g. by a fire or landslide) and where soil is already present.

A typical succession in the northern hemisphere might be:

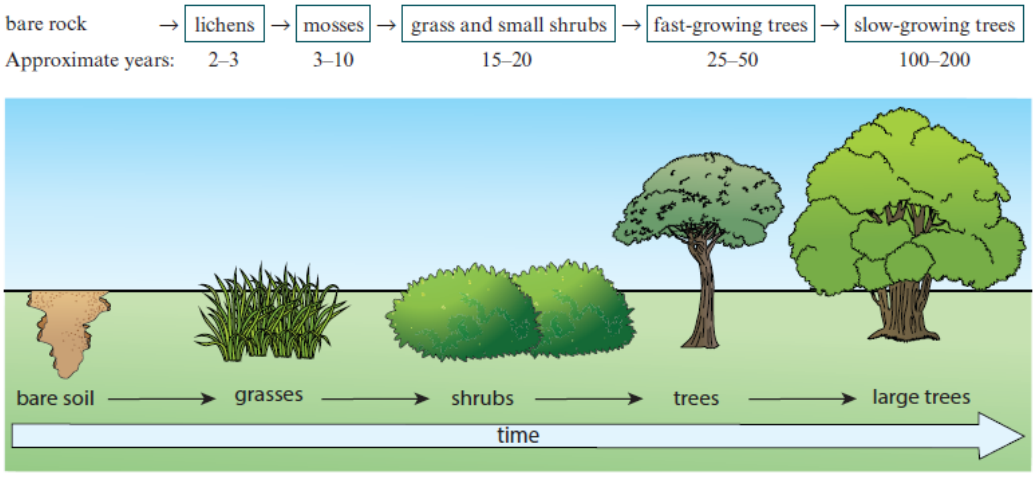


Figure 12.4.9: Undisturbed, bare land will gradually change from grassland to shrubland and then become home to small

short-lived trees. Eventually larger, slow-growing trees will grow, which form mature woodland.

Climax communities are said to be in a state of equilibrium or a ‘steady state’ because the organisms that live there have adapted to their environment and succession is no longer taking place. The community has achieved a high level of biodiversity and is not changing. Catastrophic natural events, such as flooding, wildfires, hurricanes or earthquakes, and human interference are two factors that can disrupt otherwise stable ecosystems. It is important to remember that sustainable, stable systems are vital for the continued survival of all species, including our own.

The type of stable climax community that emerges following a succession depends on the local climate. Rainfall and temperature are the two most important factors that determine the communities in a stable ecosystem. Figure 12.4.10 shows a climograph, which is a diagram that relates the type of ecosystem to conditions of temperature and rainfall in the region. A climograph can be used to predict the type of stable ecosystem that will emerge in an area.

Cyclical succession

Cyclical succession is a pattern that can been seen in a stable area where there are no specific disturbances. Vegetation may change as a small number of plant species replace one another in a regular cycle (Figure 12.4.11) as the resources available change over time.

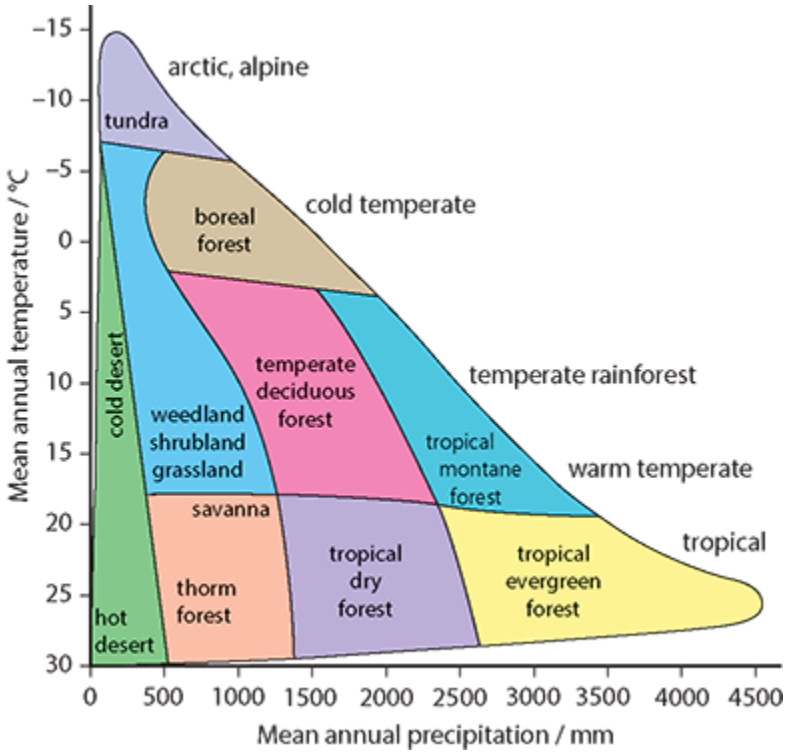


Figure 12.4.10: A climograph shows the differences in vegetation, and therefore type of ecosystem, in different climatic regions.

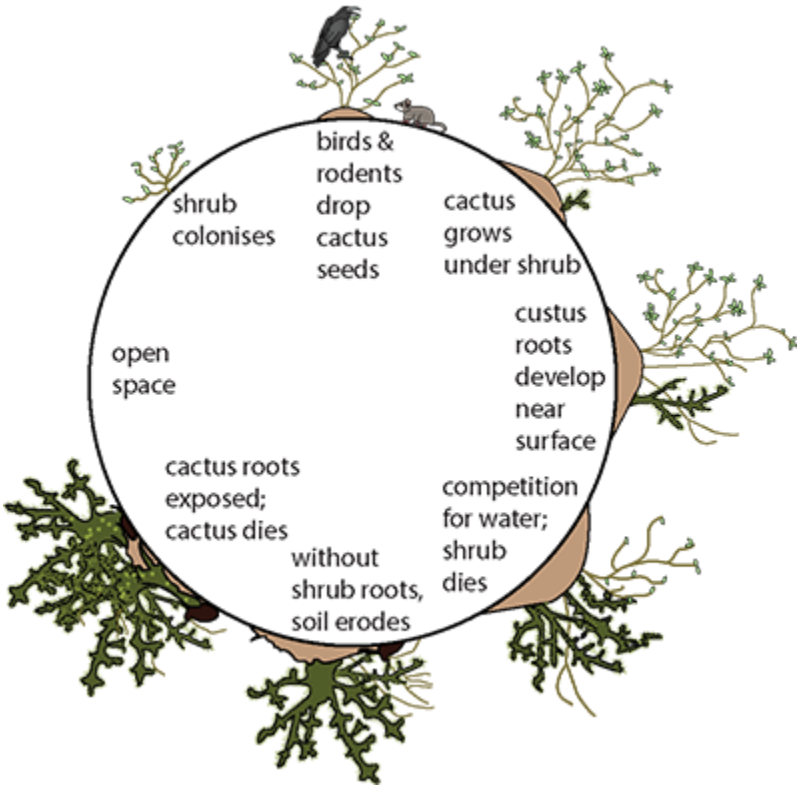


Figure 12.4.11: Cactus and shrub replace each other in an example of cyclical succession.

THEORY OF KNOWLEDGE

Complexity theory and communities

Complexity theory is a theory that tries to describe and explain the properties of complicated systems, not only in the biological world, but also in our social world. Examples of such systems include neural networks, community ecosystems and business organisations. Systems like these have emergent properties that are a result of the ability of individual parts of the system to organise themselves. For example, consider the millions of people who interact to create a society, or the thousands of species that interact to make up a climax community in an ecosystem. If you examine a single animal

or a single plant, or just one person, you could not tell whether it would be able to do things and operate with others to create the community in which they all exist. Each system or community is a product of the interactions between the individuals that live in it. Without those individuals there could not be a complex community. Complexity theory takes a holistic approach to the natural world and does not separate the components and processes in a system. This is unlike the reductionist approach, which proposes that we explain a system by breaking it down into its individual parts.

We cannot predict with any certainty how a complex system will evolve, and this unpredictability underlines the complexity of life.

To consider:

1 To what extent do you think a climax community provides an example to support complexity theory?

2 Is it ever possible to predict the effect of changes to the climate or other abiotic factors on a climax community?

TEST YOUR UNDERSTANDING

31 List three ways in which human activities can lead to the formation of a plagioclimax.

32 Outline the problems caused by the introduction of an alien species.

33 Summarise the problems caused by microplastics in the ocean.

12.5 The biodiversity crisis

LEARNING OBJECTIVES

In this section you will:

- recall that biodiversity is the variety of life in all forms and can relate to ecosystems, species or genetics
- understand that there are more species on earth today than in the past
- learn that human causes are responsible for the extinction of many species today
- recognise that diversity is lost due to loss of habitats, overexploitation of resources and agriculture. recall the reasons for the loss of species including giant moas, caribbean monk seals.
- recognise the causes of ecosystem loss due to agriculture, including southeast asian forests understand the evidence for biodiversity crisis and its causes.
- recognise that conservation of biodiversity requires many different measures including nature reserves, rewilding and reclamation of degraded ecosystems
- understand why some species are selected for conservation efforts while others are not in the EDGE of existence programme
- recognise that human activities disrupt the nitrogen cycle understand the impact of biomagnification.

- recognise that non-renewable resources are finite while renewable resources are abundant
- learn that introduced species can become invasive and reduce biodiversity
- recognise that release of pollutants can increase levels of these substances in ecosystems

GUIDING QUESTIONS

- What are the factors that are causing the 6th mass extinction?
- How can we minimize the loss of biodiversity?

12.5.1 Conservation of biodiversity

The importance of diversity

Ecosystem diversity is the number of ecological **niches** or range of different habitats that are present per unit area of a **biome**, **ecosystem** or **community**. If habitat diversity is conserved, this usually leads to the conservation of both species and genetic material.

There are three main ways to study diversity:

Ecosystem diversity

The range of different habits in an ecosystem is one of the most important factors to consider in a study of the conservation of biodiversity. To assess habitat diversity, ecologists study the variety of niches that a habitat contains. A rainforest has a high diversity of habitats which include the canopy, the soil and pools of standing water, so that there are many ecological niches present. On the other hand, a desert has little habitat diversity, simply a sandy terrain and a few plants, so it provides few ecological niches. An increase in habitat diversity is always likely to lead to an increase in both species diversity and genetic diversity.

Species diversity

Species diversity is a measure of the variety of species in a given area; it is quantified by measuring both the number of species present (species richness) and their relative abundance. (You can read about the Simpson diversity index which is one of the ways used to quantify diversity in Chapter 11.) A complex ecosystem

like a rainforest contains a wide variety of species which are likely to be abundant, so the species diversity is high.

Genetic Diversity

Genetic diversity is the range of genetic material present in a population. It follows that where more species and more individuals of a species are present, the range of genetic material present will also be greater. Genetic diversity is recognised by a large gene pool. The gene pool is the number of variations of the same gene that are present in the DNA of a particular species. A large gene pool is the sign of a healthy population with high genetic variability. A small gene pool will indicate low genetic diversity.

Comparing current and past levels of species and diversity

Scientists have estimated that there are around 8.7 million species of plants and animals in existence today. However, only around 1.2 million species have been identified and described so far, most of these are insects.

There is no accurate figure for the number of species alive on Earth today. Organisms which are found are described by scientists and recorded or stored in institutions such as the Natural History Museum in London and the collections of other research organisations such as **CITES** (**Convention on International Trade in Endangered Species of Wild Fauna and Flora**) which records endangered species. Estimates of the number of species that have never been found or named vary widely. An accepted view is that over 1.75 million species have been described but that there may be more than 10 million actually alive (Table 12.5.1).

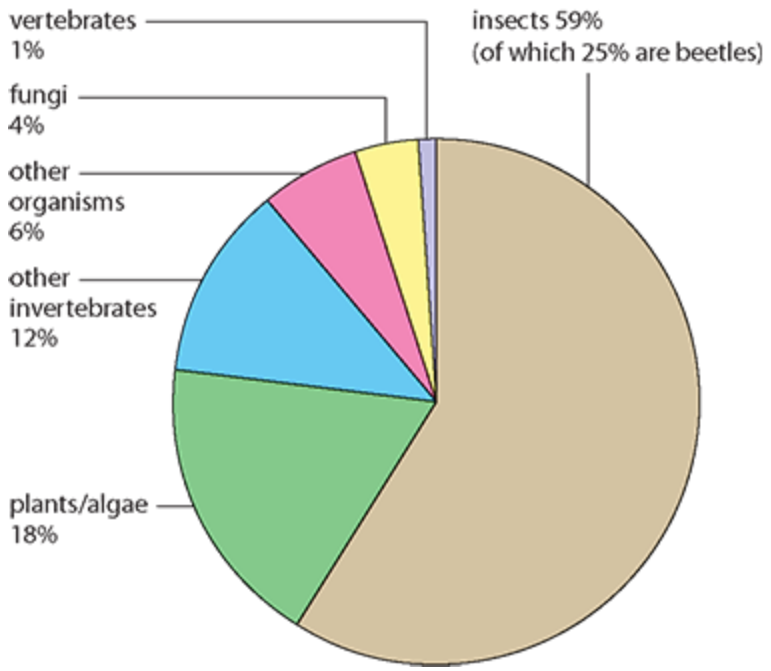


Figure 12.5.1: The largest group of organisms that have been identified is the insects and 25% of the named species are beetles.

Kingdom	Number of species
bacteria	4 000
fungi	72 000
protoctists (algae and protozoa)	80 000
plants	270 000
animals (vertebrates)	52 000
animals (invertebrates)	1 272 000
total number described	1 750 000
possible number of unknown species	14 000 000

Table 12.5.1: ***Estimates of total numbers of species***. Some scientists think that the figure may be substantially higher than

this.

Adapted from: UNEP–WCMC (2000). Global Biodiversity: Earth's living resources in the 21st century. *Cambridge, World Conservation Press*

Most estimates of the number of species are based on mathematical models. These models depend on the amount of data which is available to be put into them. Many habitats and groups of species are under-recorded because it is difficult to reach them. There may also be insufficient funding for expeditions and scientific research and, in some cases, there is disagreement about the classification of certain groups.

Species extinction caused by humans

KEY POINT

mass extinctions are periods in the Earth's history when very large numbers (more than 75%) of species die out simultaneously or within a very short period of geological time.

Mass extinctions result in an **Earth system succession**, which occurs when significant global environmental or biotic change takes place and pushes the biosphere and geosphere out of equilibrium. Species and ecosystems do evolve again after a mass extinction. As they do so they change the world's biogeochemical cycles, as well as species and ecosystems, in a process that takes many thousands or millions of years. Earth system succession goes some way to explaining the patterns of evolutionary and environmental change and the organisms that are present in the fossil record.

The most severe extinction occurred at the end of the Permian period 250 million years ago when 96% of all species were wiped out. There are about 12 million different species present on Earth today but this represents only about 1% of the total number of species that have lived. Since the beginning of life on Earth there have been several mass extinctions.

Movements of the continents, huge volcanic eruptions, drought and ice ages, and also the impact of huge meteorites on the surface of the Earth, have all caused mass extinctions. Events like these cause such massive changes in climate or physical features of the Earth that, as well as destroying species, they present new challenges to the survivors and lead to new evolutionary paths. An increase in biodiversity may be the long-term consequence of mass extinctions.

Causes of mass extinctions

The average time between mass extinctions has been about 100 million years. Palaeontologists, who study fossils, have recorded five mass extinction events throughout Earth's history (Figure 12.5.2).

1 **Drop in sea level** – The first great mass extinction event took place about 440 million years ago (mya). Extinction was probably caused by a drop in sea levels as glaciers formed, followed by rising sea levels as they melted.

2 **Cause unknown** – About 360 mya, Earth experienced the second mass extinction event. Around 20% of all families (50% of all genera) became extinct.

3 **Asteroid or volcano** – Around 250 mya, a third mass extinction occurred. It could have been due to either an asteroid colliding with the Earth or a flood of volcanic

material escaping from an area in what is now eastern Russia.

4 **Undersea volcanic eruption** – About 200 mya. An extinction was probably caused by a flood of lava escaping from a volcano in the Atlantic Ocean.

5 **Asteroid or comet collision** – The mass extinction of 65 million years ago is famously associated with the extinction of the dinosaurs. Almost no large land animals survived. Details of the five mass extinctions are summarised in Table 12.5.2.

Human activities: the sixth mass extinction

Today, populations of many species are in decline and some scientists have warned that Earth is on the brink of another mass extinction. Unlike the five previous mass extinctions, the sixth is related to human activities and is happening at a much faster rate. Current species extinction rates are hundreds or thousands of times faster than the normal background rates that occurred in the last tens of millions of years. A recent United Nations report on biodiversity and ecosystems estimated that a quarter of all species on Earth now face extinction. The loss of a species is permanent and each one has a role in the living systems on which we all depend. The huge losses that we are observing are being caused, directly or indirectly, by human activities. The two main reasons for this are the way that humans have spread and occupied territories all over the world, and the development of agriculture in the last 10 000 years. At that time there were about 1 million people in the world but today there are 7.7 billion. The human population is still growing and with more people, more land has been taken and more species exploited. Species do not have the chance to move to new areas, and pollution and climate

change have added to the destruction of their ecosystems. Species have been unable to adapt because of the speed of changes on Earth, largely caused by humans. Organisms do not have the chance to move to new areas, and pollution and climate change (Section 12.6) have added to the destruction of their ecosystems.

Extinction	Approximatetime/mya	Geological era	Loss of species (estimate)
1	439	Ordovician	25% families
2	364	Late Devonian	20% families
3	251	Permian–Triassic	55% families
4	199–214	End-Triassic	25% families
5	65	Cretaceous–Tertiary	17% families (all dinosaurs)

Table 12.5.2: The five mass extinctions that have been identified.

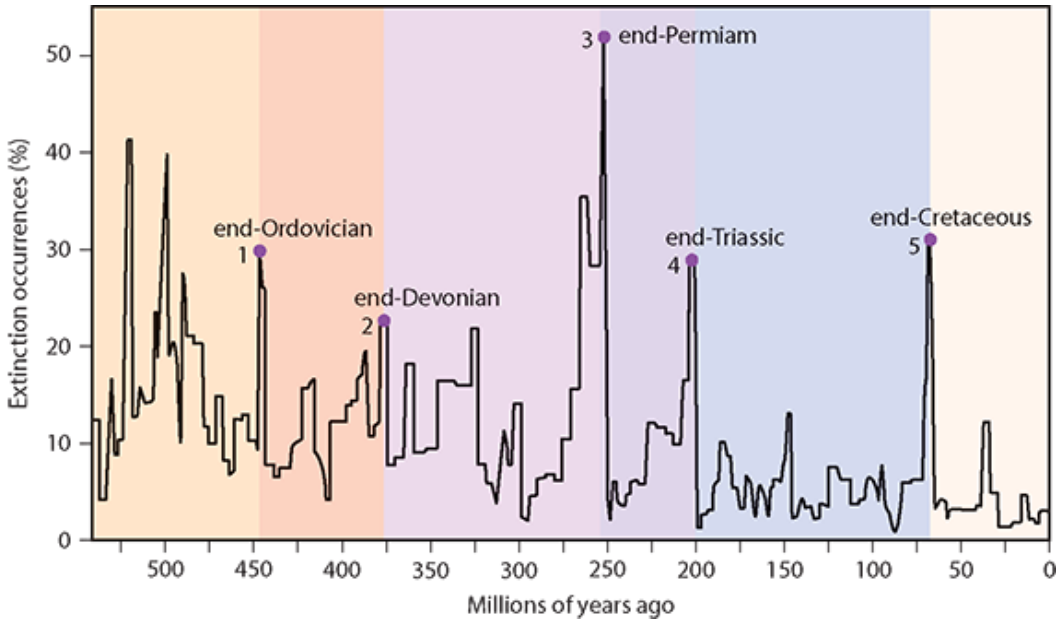


Figure 12.5.2: The five mass extinctions on Earth.

Extinction is defined as the point when a species ceases to exist or the last known individual of the species dies. Palaeontologists characterise mass extinctions as times when the Earth loses more than three-quarters of its species in a geologically short interval. The background extinction rate is the natural extinction rate of all species. Scientists estimate that it should be about one species per million per year or up to 100 species per year. Just over 5000 mammal species are known to be alive today, so the background extinction rate should be one per 200 years, but in fact about 90 mammals have become extinct in the last 400 years and 170 are listed as critically endangered. This evidence suggests that the current extinction rate is far higher than it should be and, in many cases, it is humans who are causing the extinctions. But since the total number of species known to science is only a small fraction of the estimated total, estimates of extinction rates can also vary. The current rates are probably between 100 and 10,000 times greater than the background rates. Humans have already wiped

out many species including large mammals and flightless birds such as the dodo.

Causes and consequences of the current sixth mass extinction

The key factors responsible for the current mass extinction are:

- loss of habitat to agriculture, cities, roads and industry
- overexploitation of resources such as timber and fish, and in hunting and agriculture
- pollution and climate change
- introduction of alien species as humans move species from one continent to another.

The north island giant moa (*Dinomis novaezealandiae*) is an example of a species which has become extinct due to human activities. For millions of years, nine species of large, flightless moas thrived in New Zealand (Figure 12.5.3). Then, about 600 years ago, they abruptly became extinct. Their disappearance coincided with the arrival of the first humans on the islands in the late 13th century. Scientists have speculated that the moas were hunted to extinction by these Polynesian settlers who killed them for food. Recent DNA studies of the bones of 281 moas from four species determined that the birds were healthy and that their populations were increasing before the arrival of settlers. Their genes did not indicate that they were likely to face extinction. The birds disappeared shortly after humans arrived in their habitats and were probably exterminated by overhunting as humans took what they needed to survive.

The extinction of many other species of megafauna (giant animals including mammoths, mastodons, and moas) began

between 9000 and 13,000 years ago, when human populations began to spread around the world.

Figure 12.5.3: The Giant moa can only be seen in old drawings and engravings

Another more recent extinction directly linked to human activity is the Caribbean monk seal (*Neomonachus tropicalis*) which originally inhabited the beaches and reefs of the Caribbean. The last seal was seen in 1952 at Serranilla Bank, between Nicaragua

and Jamaica. In the early 17th century there were at least 13 breeding colonies with a total population of more than a quarter of a million individuals. When early explorers arrived in the region, the seals were a curiosity and a source of food. Records show that Christopher Columbus ordered 8 ‘sea-wolves’ to be killed in 1494 and the Spanish explorer de Leon killed 14 seals in 1512. European hunters often visited the area in the 17th century, and the scientist Hans Sloane in 1707 wrote that local fishermen would kill 100 in one night to fuel their oil lamps. Later settlers who had established plantations sent hunters to kill hundreds of seals for oil to grease their machinery. The Caribbean monk seal is the only seal species which has become extinct in modern times. The reason for its extinction is due to human actions. Seals were slaughtered in large numbers not only by European hunters, local fishermen and plantation settlers but also by scientists from the 17th to the 19th centuries.

Today, overhunting, climate change, pollution and loss of habitat are the main factors which are threatening populations of species all over the world. Some examples of large vertebrate species that are at risk include the Amur leopard, black rhino, cross river gorilla, hawksbill turtle, Javan rhino, leatherback turtle, mountain gorilla, and south China tiger, but many smaller species and invertebrates are at risk too.

In the high altitudes of South America, the golden toad (Fig 12.5.4) has become extinct because of pollution, global warming and fungal infections and many other species of amphibians have been lost. They are very temperature sensitive and are also affected by the toxins in their water sources. and the reduction of insect populations due to pollution also reduces the food available for the toads.

Figure 12.5.4: This is the last photograph of the Golden toad before it became extinct. It was taken in Costa Rica in 1978

Causes of ecosystem loss

Human activity has had a significant effect on natural ecosystems throughout the world. Habitat degradation and fragmentation reduces the number of places for organisms to live. Human activities have changed land with building projects and land clearance. Agriculture, particularly intensive farming reduces diversity because monocultures replace more complex systems or remove them completely. One clear example of the effect of intensive farming can be seen in the mixed dipterocarp forest Southeast Asia (Figure 12.5.5). This forest used to be the dominant forest type, it has a complex but uniform structure with a dense, multi-storied trees with an uneven canopy.

Southeast Asia is a hotspot of biodiversity with many unique species, but the region is also one of the most threatened. Ecosystems are threatened by factors which are likely to lead to the extinction of many species. Deforestation rates in Southeast Asia are some of the highest in the world, it also has the highest

rate of mining in the tropics and the greatest number of hydropower dams under construction.

Tree-plantations and deforestation are two of the most immediate threats, and some countries such as the Philippines and parts of Indonesia have already lost over half their native forests. Predictions suggest that as much as 98% could be lost in some places in the next 10 years. Mining is also a threat, as the region exports large quantities of limestone and minerals. Biodiversity suffers directly because land is used for mines, but indirectly as roads are built through habitats and land is polluted by heavy metals.

Figure 12.5.5: Unspoilt mixed dipterocarp forest in Borneo

Figure 12.5.6: Oil palm plantations in Borneo appear as bright green areas with only small areas the natural forest remaining

Large-scale tree plantations have had the greatest impact on the forests. Oil-palm trees, rubber trees and trees for wood pulp and biofuels have replaced native forests. Oil-palm was first planted in around 1970, yet by 2011, exports had reached 30 million tonnes annually from Indonesia and 20 million tonnes from Malaysia. These two countries export almost 86% of global oil-palm.

As tropical forests are cleared carbon is released into the atmosphere as carbon dioxide; estimates suggest that tropical deforestation accounts for about 10 percent of total global warming emissions. But the biggest impact of palm oil production is on biodiversity and the large-scale devastation of habitats for endangered species such as Asian rhinos, elephants, tigers and orangutans.

There are many other examples of human interference causing loss of ecosystems. You might like to study the effect of major building projects such as the Yellow River Dam in China or the threat to biodiversity as a result of the introduction of non-native

fish to the rivers and lakes in Madagascar which have led to serious reductions in the biodiversity on the island.

NATURE OF SCIENCE

Evidence for a biodiversity crisis

The United Nations Intergovernmental Science-Policy Platform on biodiversity and ecosystems (IPBES) carries out regular assessments of biodiversity and ecosystems throughout the world. They have published eight assessments so far and most suggest that biodiversity is being badly affected in most regions. Their estimates are that one million plant and animal species face extinction. Such surveys must take into account not only the number of species present but also species richness and evenness in a given ecosystem.

The International Union for Conservation of Nature and Natural Resources (IUCN) is an organisation dedicated to nature conservation. It has assessed those species which are most at risk based on data relating to critically endangered, endangered and vulnerable species. The IUCN estimations are that 30% of amphibians and 28% of reptiles are seriously under threat and many smaller, invertebrate organisms may become extinct before they have even been discovered and given names.

Surveys by international organisations are invaluable as the data is collected using standardized methods and can be checked by other scientists and experts. And repeated at regular intervals to monitor changing situations. But the input of ‘citizen scientists’ who follow the appearance of bird species in the local area or the diversity of insects on their land is also needed. Data bases can collect such information and feed it into larger nationwide surveys.

To consider:

1 How are the methods used by international organization likely to differ from those you might use to report on local biodiversity?

2 Why is it important that scientific evidence is checked and published?

12.5.2 Causes of the Biodiversity crisis

The current biodiversity crisis has largely been caused by humans. Our population has increased from around 3 billion in 1960 to 8 billion in 2022. More people require more land to live on and to grow food so habitat destruction and species exploitation have both increased. In section 12.5.1 we considered the effect of habitat destruction for agriculture but there are many other factors that have caused the current biodiversity crisis.

Pollutants released by humans

People release many kinds of pollution into the environment. Pesticides, industrial chemicals, waste from mining and agriculture, and gases from combustion are some examples. All of these pollutants find their way into ecosystems that may be thousands of kilometres away. One serious and non-biodegradable pollutant is plastic.

Over recent years the production and use of plastics has increased enormously. It is estimated that over 250 million tonnes of plastic are used annually and that making plastic items uses approximately 8% of the world's annual oil production. Plastic litter degrades very slowly: a plastic bottle will take 450 years to break down and so plastic waste builds up on land and in the oceans. Plastic makes up between 60 and 80% of marine debris and as much as 90% of floating debris.

Macroplastic debris is defined as plastic fragments which are greater than 1 mm across, while microplastic debris has fragments that are less than 1 mm. Macroplastics include items such as plastic bottles and bags, detergent containers and food wrapping. These items accumulate in marine habitats all over the

world and may remain for centuries. Microplastics from PVC, polyester, acrylic and polyamide account for more than 65% of marine debris. Researchers have traced much of the microplastic back to synthetic clothes, which can release up to 2000 tiny fibres per garment every time they are washed.

Both types of plastic are ingested by marine organisms, which mistake them for food. This plastic may enter the food chain as residues of the plastic accumulate in organisms' cells. An animal whose stomach is full of plastic fragments feels full and may stop feeding so that it starves to death.

Figure 12.5.7: Dead Laysan albatross chick with a stomach full of plastic debris.

The Laysan albatross lives on Midway Atoll in the North Pacific Ocean, thousands of kilometres from both mainland Asia and North America. Albatrosses skim the water surface to feed and pick up plastic as they do so. Adults feed the plastic to their chicks and while adults are able to regurgitate some plastic, the chicks cannot and can be killed by its effects. As well as making the chick feel falsely full, sharp plastic pieces can cut through the

stomach and cause infections and death. One tragic example is shown in Figure 12.5.7.

Human introduction of alien species

An **alien species** is one that is not native to the region in which it is found. There have been many occasions throughout history when an organism has been introduced from one ecosystem to another, whether it is:

- accidentally
- deliberately
- or for biological control of a pest organism.

Alien species can become invasive if they spread and modify their new environment and cause ecological damage. Invasive species can lead to a reduction in biodiversity in the areas they are introduced into, especially if they outcompete local species or have few natural predators.

Accidental introduction

The zebra mussel (*Dreissena polymorpha*) is a small freshwater species, originally native to lakes in southeast Russia. It has been accidentally released in many other areas, probably carried in ballast water of cargo ships. It has become an invasive species in many different countries. Zebra mussels are now found in the Great Lakes of the USA where they grow on docks and boats. They have spread into streams and rivers and block water pipes and interfere with water supplies (Figure 12.5.8). In some areas, they have outcompeted all other freshwater mussels because they grow in dense clumps. Zebra mussels are also believed to be the source of deadly avian botulism poisoning that has killed tens of

thousands of birds in the Great Lakes since the start of the 21st century.

Figure 12.5.8: Masked workers use a water jet to clear zebra mussels clogging the pump room of Detroit Edison's power station in Michigan, USA. Zebra mussels also encrust water pipes and excrete a corrosive chemical.

Deliberate introduction

Many plants, collected in one part of the world, have been deliberately introduced to domestic gardens far away because of their attractive flowers or exotic foliage. Orchids, bamboos and rhododendrons are now seen all over the world but most were introduced following plant-collecting expeditions in the 19th and 20th centuries. In some cases, an introduced species finds the new conditions so advantageous that it becomes invasive. It grows rapidly and becomes a threat to endemic (native) species, which it outcompetes and eventually eliminates. One such example is Japanese knotweed (*Fallopia japonica*), which was deliberately introduced into European gardens in the 19th century for its attractive flowers. It reproduces vegetatively and even short sections of root can regrow to become whole new

plants. This plant now covers huge areas of land in Europe. It can be controlled with herbicides, but there is a problem using these chemicals near rivers, as the herbicide gets into the waterway and upsets its ecological balance, harming plant and animal life.

No two species can occupy the same niche indefinitely because one will come to dominate and exclude the other. Both *Rhododendron* and Japanese knotweed have outcompeted and excluded native species and reduced biodiversity in the places where they grow.

12.5.3 Approaches to conservation of Biodiversity

People have come to realise that without concerted action we will lose the biodiversity of the Earth's habitats. There are many different approaches to conservation and many different organizations that are promoting conservation. No one approach will be suitable for every area and it is also difficult to quantify the value of a species or a habitat. Conservation arguments struggle to balance the ethical and aesthetic values with commercial considerations.

- economic arguments cite the valuation of ecotourism, genetic resources and natural capital
- ecological arguments centre on preservation of the ecosystem
- ethical arguments include the intrinsic value of a species.

Intergovernmental and non-governmental conservation organisations

The key objective of conservation organisations is to preserve species and their habitats throughout the world. Some organisations work at a local level while others are global. Different organisations are categorised according to the way they are set up and funded. Governmental organisations follow the policies of one or more governments and are funded by them. Non-governmental organisations (NGOs) are funded by individuals or independent groups. The effectiveness of the different organisations varies due to the different strategies they adopted in their work.

UNEP and WWF

United Nations Environment Programme (UNEP) is a governmental organisation which coordinates United Nations work on the environment and helps LEDCs to implement environmentally sound policies. UNEP was founded in 1972 and has its headquarters in Kenya. Its stated objectives are ‘to provide leadership and encourage partnership in caring for the environment by inspiring, informing and enabling nations and peoples to improve their quality of life without compromising that of future generations’.

UNEP gathers, collates and verifies data on biodiversity and ecosystems from many sources. This can be used as a reliable source of information. UNEP also promotes global and regional cooperation and develops environmental laws covering a range of issues from the atmosphere, marine and terrestrial ecosystems, to the green economy. Like many governmental organisations, it also works with NGOs to implement its policies.

The World Wide Fund for Nature (WWF) is an NGO and one of the best-known international conservation organisations. Since 1961 it has campaigned for the natural world and worked to ease pressure on the world’s natural resources. WWF is an independent organisation but around the world, it works with businesses, governments and local communities to create sustainable solutions that take account of the needs of both people and nature.

The organisation’s conservation work focuses on safeguarding wildlife and places it considers to be of global importance. It also lobbies governments and runs campaigns to change legislation and policy to protect the environment and biodiversity. Major campaigns have focused on climate change, energy, housing and

the protection of the marine environment. It states that its ultimate goal has always been ‘people living in harmony with nature’ and finding ways to share the Earth’s resources fairly.

Some approaches to conservation include:

- conserving and protecting a natural habitat as a nature reserve
- *in situ* conservation
- *ex situ* conservation.

Conserving and protecting a natural habitat as a nature reserve should benefit all species. However, if population numbers are very low and a species are at risk, more active intervention may be required. Each nature reserve will have its own unique solutions to conservation problems. At Belsize Wood Nature Reserve, a small woodland reserve near the centre of London in the UK, nesting boxes for birds and bats have been put in place, because the number of mature trees providing suitable natural nesting sites is low. In a wetland nature reserve, nesting platforms that float on lakes can be beneficial and offer some protection against predators for nesting birds. At Sungei Buloh Wetland Reserve in Singapore, sluice management allows the control of water levels in the ponds. At any one time, the water level in at least one pond is kept low to expose the mudflats for shorebirds to feed and roost.

Figure 12.5.9: Walkways like this one enable people to observe wildlife without disturbing it.

Local people may not support the funding and existence of a nature reserve if they are not allowed into it. But, as the more people that visit a nature reserve, the more chance there is of habitats being damaged or destroyed. On the other hand, visitor access can have positive outcomes, if public awareness and knowledge of wildlife is improved. Usually, special trails or walkways are built at reserves to ensure that observers can visit safely without risk of damage to the surrounding habitats (Figure **12.5.9**). Legislation can also protect nature reserves from development and industrial activities.

In situ conservation

***In situ* conservation** protects species within their normal habitat. This makes sense because each species has evolved to adapt to a particular environment. In situ conservation protects species in their own habitats by maintaining the environment, often within nature reserves or national parks. Protecting turtle nests using cages and fencing off an area of beach so that hatchlings can reach the sea is one example.

In situ conservation work may involve removal of invasive species: one example has taken place on Montague Island, one of the offshore islands in New South Wales, Australia. The island is home to several seabird species but because of the growth of large areas of non-native kikuyu grass (*Pennisetum clandestinum*) habitat that used to be used by nesting birds has been lost. The little penguin (*Eudyptula minor*), burrowing short tailed shearwaters (*Ardenna tenuirostris*), wedge-tailed shearwaters (*A. pacifica*) and sooty shearwaters (*A. grisea*) were suffering and their populations were declining. The Seabird Habitat Restoration Project aimed to remove the introduced grass which had smothered native vegetation, so that large expanses of seabird breeding habitat would be available and native birds could resume breeding there.

The project has proved highly successful in controlling and reducing the spread of kikuyu grass and restoring degraded seabird habitat so that the birds have been protected and their numbers can increase.

Ex situ conservation

***Ex situ* conservation** involves preserving a species whose numbers are very low in a **captive-breeding programme** in a zoo or botanic garden to prevent it dying out.

In situations where *in situ* conservation is difficult or inadequate, *ex situ* conservation must be used. This is not ideal, because an organism behaves differently outside its natural habitat. However, it does give rise to the opportunity for captive breeding using scientific knowledge and modern technology. Techniques such as artificial insemination and embryo transfer may be used if animals fail to breed normally, and embryos can be preserved

for later use. Difficult pregnancies can be monitored and the young cared for by staff.

An *ex situ* breeding programme has proved invaluable for the Arabian oryx (Fig 12.5.10). This animal, once almost extinct in the wild, has been successfully bred in a number of zoos in the USA and Europe.The DNA from the few remaining animals was compared and animals specially selected for breeding so that genetic diversity was maintained as far as possible.

Figure 12.5.10: Arabian oryx have been saved from extinction by ex situ breeding programmes

EXAM TIP

You do not have to remember all the Latin species names for organisms that you may write about in your exams.

Studying the behaviour of captive animals is key to breeding programmes. Some species with complex behaviours such as the giant panda from China are highly challenging to breed in captivity, but the centre at Chengdu in China has been very successful.

Plants are more straightforward to maintain in an *ex situ* situation. Botanic gardens can supply the correct environmental conditions for different plants and computer-controlled glasshouses can maintain the temperature and humidity that each requires. Many countries maintain 'national collections' of a variety of species including endemic plants, exotic genera and important food plants. There are also **seed banks** for many of the world's staple crops such as rice and maize. These preserve varieties of important crops, called **landraces**, which may be useful in the future to produce new varieties of food plants. At the Millennium Seed Bank at Wakehurst Place in England, seeds are kept in cool, dark conditions, which prevent germination, and can be stored for many decades. The Svalbard Global SeedVault, on the Norwegian island of Spitsbergen, holds duplicate samples of seeds held in gene banks worldwide, in an underground cavern.

Rewilding and reclamation

Rewilding is a means of conservation that aims to restore and protect natural systems and wilderness areas. Rewilding is a form of ecological restoration that tries to recreate an area's natural uncultivated state. We can help rewilding by creating the conditions an area needs to re-establish a stable, self-sustaining ecosystem with levels of diversity that existed before humans affected them. One way to summarise the important factors of rewilding is the three. Cs: Conservation of Cores, Corridors between areas and Carnivores. Carnivores are often keystone species (Section 12.3.8) and are essential in any balanced ecosystem.

Examples of rewilding include, protecting, enlarging and connecting ancient woodland to allow a range of wildlife to establish, breed and disperse. This may require a reduction in the

number of grazing animals to help young trees and other vegetation grow. A larger area of woodland also increases the amount of carbon storage. Other ways of creating the right conditions for rewilding include removing dams or dykes to release the flow of water, reintroducing species that have disappeared and allowing natural forest to regenerate.

SCIENCE IN CONTEXT

A rewilding success story

Hinewai is an ecological restoration project occupying 1500 hectares on the South Island of New Zealand. Initially 109 hectares were privately purchased by the Maurice White Native Forest Trust in 1987 and botanist Dr Hugh Wilson began a project to allow introduced gorse to grow as a low canopy to restore farmland into native forest. The aim was to allow natural regeneration of native vegetation and wildlife with minimal interference and after 30 years of natural succession the vegetation and forest looks similar to the one that was present before clearance by humans; first Polynesian settlers from about 700 years ago and secondly by European settlers from about 1850. The native forest has regenerated, and birds and wildlife have returned. Dr Wilson has proved that nature can recover when given time and the opportunity.

Fools & Dreamers: Regenerating a Native Forest is a documentary film from ***HappenFilms*** about Hinewai Nature Reserve.

Figure 12.5.11: These biologists from the Rewilding Foundation evaluate a turtle inside the turtle reintroduction centre at El Impenetrable National Park, Chaco province, Argentina.

Reclamation is the practice of renewing land that has been damaged or degraded by agriculture or industries such as mining. Common disturbances include logging, damming rivers, intense grazing, hurricanes, floods, and fires. Reclamation activities try to replicate a pre-disturbance ecosystem or to create a new ecosystem through human intervention. They may include activities such as controlling soil erosion and waterways, reforestation by tree planting, removal of weeds and non-native species and replanting species that have been lost or replacing them with similar species that can survive in the reclaimed land.

The EDGE of Existence programm

KEY POINT

EDGE species are Evolutionarily Distinct and Globally Engandered

The **EDGE of Existence programme** is a research and conservation organisation developed by the Zoological Society of London. It aims to draw attention to the species that are thought to be the most Evolutionarily Distinct and Globally Endangered. Research and conservation plans of the programme are trying to halt the decline of species and take action to conserve them by training conservationists (called EDGE Fellows) to protect them.

The organisation has an interactive website which lists the top 100 EDGE mammals, reptiles, birds and amphibians and the top 25 corals that need urgent conservation. Each species is rated according to the rarity of the species and the current conservation efforts that are underway. Distinctive species do not have many closely related species and EDGE species are often the only surviving member of their genus. If species like these become extinct it would cause a significant loss of biodiversity and evolutionary history.

Examples of EDGE species include some well-known animals such as elephants and pandas but others such as the echidna, the bumblebee bat, the world's smallest mammal which weighs only 2 g, and the vaquita, the smallest of the porpoises that is found in the Gulf of California, are highly endangered, poorly understood and often ignored by other organisations.

NATURE OF SCIENCE

Cooperation and collaboration – conserving biodiversity requires international cooperation between scientists, organisations and politicians

In the last 50 years, the importance of biodiversity has come to the forefront of science. Species are not evenly distributed on Earth. Biodiversity is far richer around the tropics and

areas containing rainforest are among the most diverse on the planet. People have come to realise that there are many compelling reasons for conserving the biodiversity of habitats such as the rainforests. Undiscovered species may provide valuable medicines and other resources for future generations. Conservation in one part of the world may depend on cooperation and collaboration in another and international organisations such as Worldwide Fund for Nature (WWF) and the United Nations Environment Programme (UNEP) coordinate such work in many countries. The key objective of all conservation organisations is to preserve species and their habitats. Some work at a local level while others are global. Some organisations, such as UNEP, are funded by governments while others, such as WWF, are non-governmental organisations (NGOs), which are funded by individuals or groups. Organisations such as WWF work with businesses, governments and local communities to create solutions that take account of the needs of both people and nature.

Conservation programmes must select which species are to be protected, but it is often difficult to decide which species most merit conservation efforts. On what basis should one species be chosen over another? For example, is a large mammal such as a tiger or panda more important than a small, seemingly insignificant mollusc? A striking or endearing mammal may encourage people to support a conservation programme but smaller, less appealing species may, in fact, be more important and play a pivotal role in an ecosystem. Should endangered animals be given priority over other species whose numbers are not yet so low?

The choice of species for ex situ conservation can also be difficult, and many factors must be considered. For example,

when zoos select animals for captive breeding programmes, certain animals with aesthetic appeal are likely to increase visitor numbers and therefore raise public awareness and attract greater financial support for conservation. If these animals are returned to the wild, they may engage local people who could benefit from ecotourism. On the other hand, choosing a species for ecological reasons is more likely to benefit a whole ecosystem – assuming the programme does not fail through lack of funding and support.

Science can support conservation efforts by providing the expertise needed to ensure breeding programmes are successful. Different zoos have different areas of expertise and are likely to be more successful at ex situ conservation with some species than with others, so this factor too will influence the organisms whose preservation is prioritised.

Linking questions

How does variation contribute to maintaining stability in an ecosystem?

INTERNATIONAL MINDEDNESS

The IUCN Red List

The International Union for Conservation of Nature and Natural Resources (IUCN) is an organisation dedicated to nature conservation in all countries of the world. It has assessed which species are most at risk, based on data about critically endangered, endangered and vulnerable species. The IUCN Red List of Threatened Species is the world's most comprehensive information source on the global conservation status of animal, fungi and plant species. The Red List is an indicator of the health of the world's biodiversity. The IUCN

estimates that 40% of amphibians and 26% of mammals are under serious threat and that many smaller, invertebrate organisms may become extinct before they have even been discovered and given names.

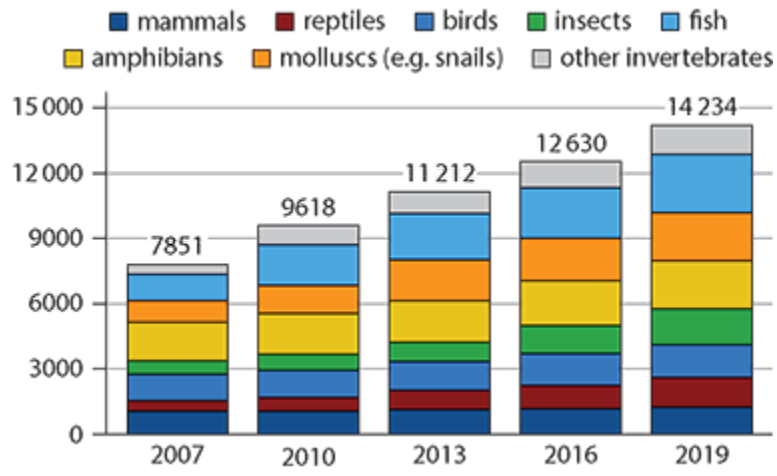


Figure 12.5.12: The number of endangered animal species, from the IUCN Red List.

12.5.4 Eutrophication - human activities and the nitrogen cycle

The nitrogen cycle

Nitrogen is a vital element for the formation of proteins and nucleic acids in the bodies of plants and animals. However, although almost 80% of the Earth's atmosphere is nitrogen gas, it is so stable that it cannot be used directly by living organisms, and nitrogen is often in short supply as a nutrient. It is recycled in ecosystems through the actions of many organisms including nitrogen fixing bacteria, nitrifying bacteria and denitifrying bacteria.

Plants obtain the nitrogen they need to grow in the form of nitrates, which they absorb from fertile soil through their roots. All nitrates are soluble and can be absorbed by plants.

The following are good for plant growth:

- nitrogen fixation, which converts nitrogen gas to useful nitrates
- nitrification, which converts ammonia to useful nitrates.

But the following is bad for plant growth:

- denitrification, which converts useful nitrates into nitrogen gas that plants cannot use. This takes place in anaerobic conditions in compacted or waterlogged soil.

Eutrophication

Human activities such as agriculture have interfered with the natural cycling of nitrogen. Nitrates and phosphates are added to

the soil in the form of fertilisers, animal manure and sewage, which all contribute to eutrophication. Eutrophication is defined as the natural or artificial addition of nutrients (especially nitrates and phosphates) to water, which leads to a reduction or depletion of the oxygen content of the water.

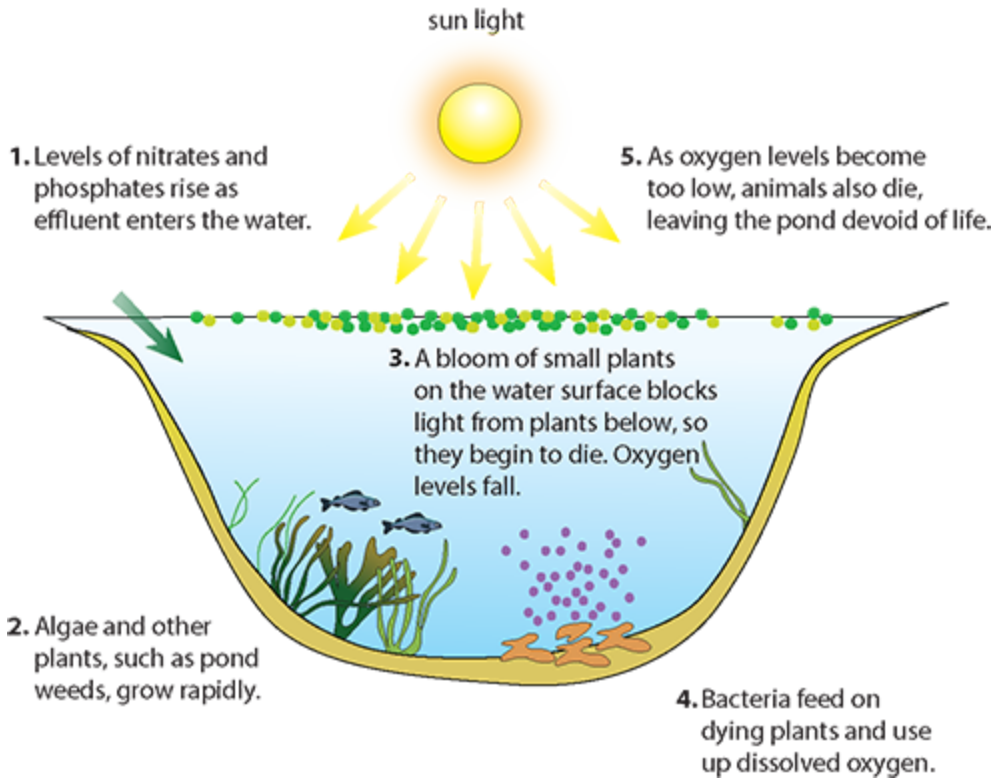


Figure 12.5.13: Eutrophication in a pond ecosystem.

KEY POINT

biochemical oxygen demand (BOD) is the amount of dissolved oxygen needed by aerobic organisms to break down organic material in water, at a certain temperature, over a certain period of time.

Excess fertiliser can run off the land, particularly in areas where large numbers of livestock are kept or where slurry (produced

from animal waste and other organic matter) is used as a fertiliser. Soil erosion also deposits both manure and artificial fertilisers in waterways, and the problem is worse where forests have been cut down and leaching (washing out) of soluble minerals from the soil by rainwater is increased. Excess nitrate and phosphate flowing into waterways and ponds can cause ecological problems and eventually lead to eutrophication. The process is summarised in Figure 12.5.13. In many countries, fertiliser use is controlled, and in modern farming the requirements of crop plants are closely monitored because of the high cost of fertiliser.

Aquatic systems produce and consume oxygen (Figure 12.5.13). The number of aerobic microorganisms and their rate of respiration at any particular location in a river or stream determine the **biochemical oxygen demand (BOD)**. The greater the BOD the more rapidly oxygen is depleted, resulting in less oxygen being available to other forms of aquatic life. When this happens aquatic organisms can become stressed, suffocate and die. BOD is used as an indirect measure of the amount of organic matter in a water sample. Oxygen demand can be increased by excess growth of algae and by other oxygen-demanding factors such as decaying plants and animals, animal waste and effluent from wastewater treatment plants.

12.5.5 Biomagnification

Some chemicals used in the environment as pesticides are taken into the bodies of non-pest species where they remain and accumulate because the organism cannot break them down and excrete. Insecticides such as DDT and dieldrin are well studied examples of the way toxic chemicals can accumulate in the environment by **biomagnification**.

KEY POINT

biomagnification is the process by which chemical substances accumulate to progressively more concentrated levels at each trophic level in a food chain.

Small quantities of these insecticides, used to control pests, are sprayed into the environment where they may be absorbed by plants, or fall on the surface of their leaves. The plants may be unaffected, but when primary consumers feed on the sprayed plants they take in a far greater quantity of the toxin. The chemical remains in the bodies of the primary consumers and if a secondary consumer feeds on a number of these animals, it accumulates an even greater amount of the chemical.

DDT is an organochlorine insecticide that was widely used to kill mosquitoes that carry the malarial parasite. It is stored in the fatty tissues of animals that have ingested it. We now know that it is not readily biodegradable and can remain in the environment for up to 15 years. The first signs of the damage it can do were noticed in a survey of peregrine falcons in Europe in the 1960s, which showed that the peregrine population was declining in number. Their bodies contained high levels of DDT, which caused the shells of the birds' eggs to be thinner than normal. As

females tried to sit on their eggs to incubate them, the eggs broke.

The effects of DDT were reported in many other parts of the world in a variety of wild bird populations. Even penguins in Antarctic regions were found to have the chemical in their bodies.

Although the original concentration of DDT used in insecticide sprays was low, at about 3 parts per million, the chemical ran into waterways and was taken up by microscopic plants in rivers and lakes. When microscopic animals ate these plants, the DDT became more concentrated in their bodies. Small fish feeding on the microscopic animals accumulated about 0.5 ppm in their body fat and fish-eating water birds, such as the osprey, had about 25 ppm of DDT in their bodies (Figure 12.5.14).

DDT was a successful insecticide because it remained effective for a long time without breaking down, but its damage to the environment was considerable. Since the 1970s, it has been banned in most countries, although it is still legal to use it in some parts of South America, Africa and Asia. Slowly, wild bird populations have recovered from its effects.

SCIENCE IN CONTEXT

Minamata Bay and mercury poisoning

Minamata is located on the coast of Japan's westernmost island. The city and the adjacent Minamata Bay form a relatively closed ecosystem. The bay used to be a source of fish for the residents of the city, until the mid-1950s when people started to fall ill with an unknown neurological condition. They suffered increasing loss of motor control when walking or carrying out simple everyday tasks.

Sometimes they became partly paralysed and were unable to see or speak properly. It took more than 5 years for medical researchers to identify the cause as mercury poisoning from eating the fish and shellfish from Minamata Bay. The source of the problem was the town's Chisso Corporation factory which had been manufacturing acetaldehyde to produce plastics since the 1930s. Mercury from the production process spilled into the bay and entered the food chain. At that time, Minamata residents relied almost exclusively on fish and shellfish from the bay as their source of protein.

Direct evidence that mercury was responsible was not confirmed until nearly 100 people had been identified as suffering from poisoning and more than 20 had died. Mercury is still present in the sediment of the bay, where fishing remains prohibited. Mercury accumulated in the food chain because animals that lived in the bay and fed on contaminated food were unable to break it down. When people in the area ate fish from the bay, they ingested mercury with each meal until the amount in their bodies reached dangerous levels.

To consider:

1. How do you think incidents such as Minamata Bay have influenced present-day thinking about environmental issues?

2. Suggest two reasons why the problem of mercury poisoning was not identified and dealt with until 1956, even though the factory had been operating since the 1930s.

3. Research methods of removing mercury and other poisonous metals from ecosystems that are polluted with them.

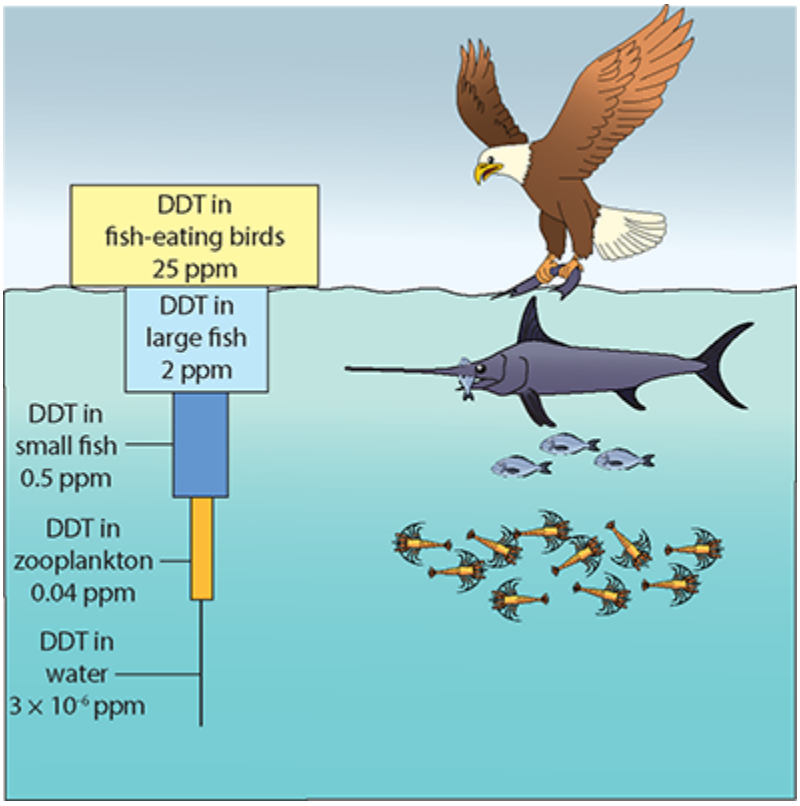


Figure 12.5.14: An example of how DDT concentrations increase up the trophic levels of an estuarine food chain.

Heavy metals and industrial chemicals such as polychlorinated biphenyls (PCBs) that are also released into the environment remain a problem for living organisms. This is because these chemicals accumulate in a similar way.

TEST YOUR UNDERSTANDING

34 Define a keystone species.

35 Summarise how nitrate fertilisers disrupt the nitrogen cycle.

REFLECTION

Could I summarise the effect that humans have had on ecological relationships and the survival of organisms? Which factors would I describe first to a classmate?

Links

- How do food chains and webs transfer matter and energy in ecosystems? (Chapter 12.2)
- How do human activities affect the evolution of species? (Chapter 11.3)

12.6 Climate change

LEARNING OBJECTIVES

In this section you will:

- recall the ways in which humans have caused climate change
- understand that there is positive feedback in global warming
- recall that forest ecosystems may reach a tipping point if changes continue
- learn that habitats will change as sea ice melts
- learn that changes in ocean currents will change nutrient upwelling
- recall the threat to coral reefs of climate change
- understand some measures that are being taken to restore ecosystems

- recognise some factors that influence the timing of events such as flowering and migration
- understand that changes to these factors may disrupt the growth reproduction and development of some species
- recognise that evolution is one consequence of climate change

GUIDING QUESTIONS

- What are the main factors causing climate change?
- How does climate change affect ecosystems?

12.6.1 Human causes of climate change

The greenhouse effect

Certain gases, the most important of which are carbon dioxide and water vapour, enable the atmosphere to retain heat. Without these gases in the atmosphere, the Earth's temperature would be too low to support life. The warming effect of these gases is known as the **greenhouse effect** because it is caused in a similar way to the warming of a greenhouse.

A greenhouse is made of glass, which allows shortwave radiation from the Sun to pass through it. As the sunlight passes through the glass, the radiation is absorbed, changed into heat – which has a longer wavelength – and re-radiated. Glass is less transparent to these long wavelengths and heat is trapped in the greenhouse, making it warmer inside. So-called 'greenhouse gases' in the Earth's atmosphere (such as carbon dioxide, methane and water vapour) act in a similar way to the greenhouse glass. They trap heat that is radiated from the Earth's surface and keep the Earth at a comfortable temperature for life to exist (Figure 12.6.1).

Carbon dioxide currently forms only 0.04% of the atmospheric gases but it plays a significant part in the greenhouse effect. Other greenhouse gases include water vapour, methane, oxides of nitrogen and fluorocarbons (FCs). Methane is estimated to have more than 80 times more warming power than carbon dioxide, even though the effect of carbon dioxide in the atmosphere lasts longer. Estimates suggest that about 25% of global warming is currently due to methane from human activities

The human population has increased dramatically in recent history and the demand for energy in industry, transport and

homes has also increased. Most of this energy demand has been met by burning fossil fuels, mainly oil, coal and gas. Burning fossil fuels releases both carbon dioxide and oxides of nitrogen.This activity has raised the concentration of carbon dioxide in the Earth's atmosphere significantly since the mid-1800s, a period which has coincided with increasing industrialisation (Figure 12.6.2).

The influence of increased concentrations of greenhouse gases has produced changes in global temperatures and climate patterns. Rising levels of greenhouse gases are believed to be causing an enhancement of the natural greenhouse effect. Scientists have shown that the Earthis experiencing a rise in average global temperature, known as **global warming** (Figure 12.4.6) which is thought to be happening because of this enhanced greenhouse effect. Climatologists are concerned that, as a result of all this activity, humans are upsetting the balance of the carbon cycle and adversely affecting our atmosphere. (You can review the stages of the carbon cycle in Section 12.2).

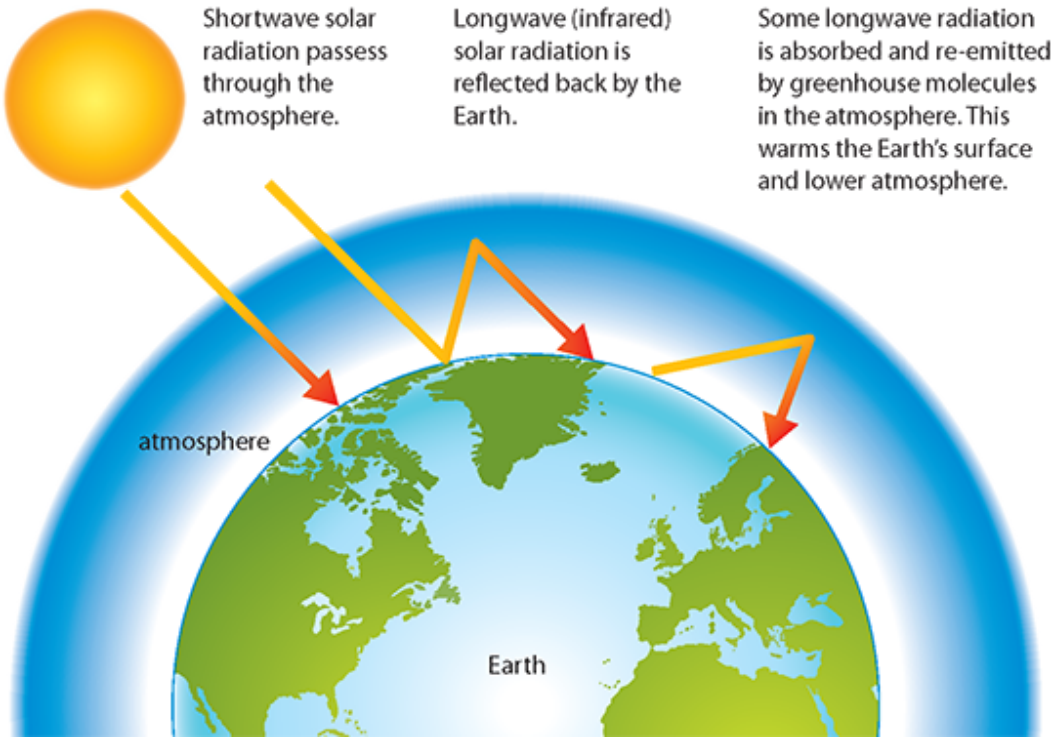


Figure 12.6.1: Greenhouse gases trap heat, warming the atmosphere.

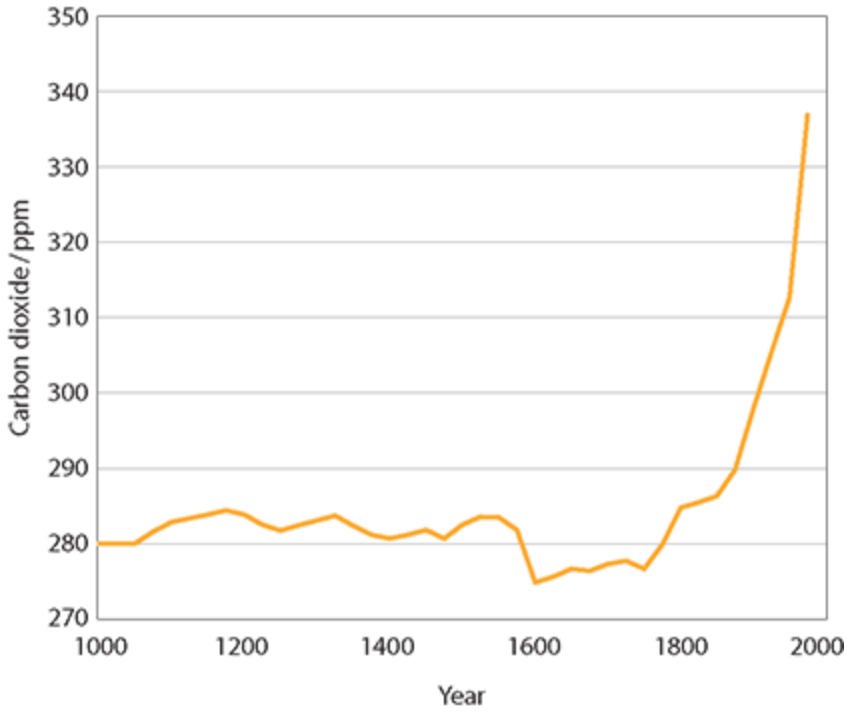


Figure 12.6.2: Graph to show increase in carbon dioxide concentration in the atmosphere since the start of industrialisation

The consequences of climate change are wide ranging and varied. Some results of global warming are likely to be:

- melting of ice caps and glaciers
- a rise in sea levels, causing flooding to low-lying areas
- changes in the pattern of the climate, winds and ocean currents
- changes in ecosystems and the distributions of plants and animals
- increases in photosynthesis as plants receive more carbon dioxide.

As the atmosphere warms, it causes sea ice to melt. Ice is white and reflects solar radiation, while the darker ocean surface absorbs more solar radiation. As the oceans absorb more heat, this causes more ice to melt and warms the Earth further. This effect is known as the ice-albedo feedback and it is a **positive feedback** system. More warming leads to more melting which leads to more warming. Deposits of frozen methane and carbon dioxide are also locked up in the biomass and decomposed biomass that is contained in the permafrost in polar regions. Permafrost is ground that remains below 0°C for more than two years and it covers about a quarter of the Northern Hemisphere. As the environment warms and the permafrost thaws, methane and carbon dioxide can be released into the atmosphere and contribute to the positive feedback effect of global warming. Carbon dioxide is naturally stored in the ocean either as a

dissolved gas or in the form of carbonates. Global warming also causes more of this carbon dioxide to be released.

- As land and sea ice melt habitats for animals which live on the ice are lost. In the Antarctic the breeding grounds of the emperor penguin may be lost while in the northern hemisphere, the sea ice habitat of walruses and polar bears could be lost.

- Warmer temperatures and less snowfall in the cold boreal forests which extend across the northern USA, Canada, Scandinavia and Russia may lead to drought in these areas of taiga biome. Without water the trees turn brown and primary production falls. Warmer temperatures cause the peatland to dry out and hotter drier conditions are leading to more wildfires which in turn speed up the melt of permafrost. Peatlands store twice as much carbon as all the world's forests so a change from net carbon accumulation to net carbon loss could prove to be a tipping point in ecosystem stability, like that seen in boreal forest and rainforest.

- Ocean currents transport warm water and rain from the equator toward the poles and cold water from the poles to the tropics, rather like a global conveyor belt. Ocean currents regulate the global climate and distribute solar radiation. Ocean currents are found both at the ocean surface and in deep water. Water moves not only horizontally but vertically. In upwelling currents vertical movements bring cold nutrient rich water to the surface and send denser cold water downward. If global warming heats the surface waters, nutrient upwelling may be prevented so that primary production in the oceans and energy flow in marine food chains is reduced.

- Effect on coral reefs – Coral reefs are fragile ecosystems, which are very sensitive to rising carbon dioxide levels. They are sometimes known as the ‘rainforests of the sea’ because of their extensive biodiversity which includes large numbers of fish and other species. Coral polyps extract calcium from seawater and use it to build the elaborate limestone skeletons that make up a coral reef. Most reef-building corals contain photosynthetic algae, called zooxanthellae, which live within their tissues. The corals and algae have a symbiotic, mutualistic relationship: the coral provides the algae with a protected environment and inorganic nutrients, while the algae produce oxygen and help the coral to remove waste products. Zooxanthellae also supply the coral with essential compounds from photosynthesis and produce the wide range of colours that are seen in corals. Corals respond to the stress of higher sea temperatures which result from global warming by expelling these algae – an event known as coral bleaching (Figure 12.6.3). Some coral can recover from bleaching but often the coral dies, and the entire reef ecosystem disappears. Coral reefs also suffer as a result of ocean acidification which occurs when higher CO_2 levels in the atmosphere cause the oceans to absorb more carbon dioxide. Acidic oceans may inhibit the growth of producers and affect food chains generally, but in a more acidic ocean, calcium dissolves more easily and corals cannot form their skeletons properly so their growth slows down. It has been estimated that a doubling of atmospheric carbon dioxide will reduce calcification in some corals by as much as 50%. Predicted rises in sea level caused by melting sea ice could also affect some reefs by making the water too deep to allow adequate sunlight to reach them.

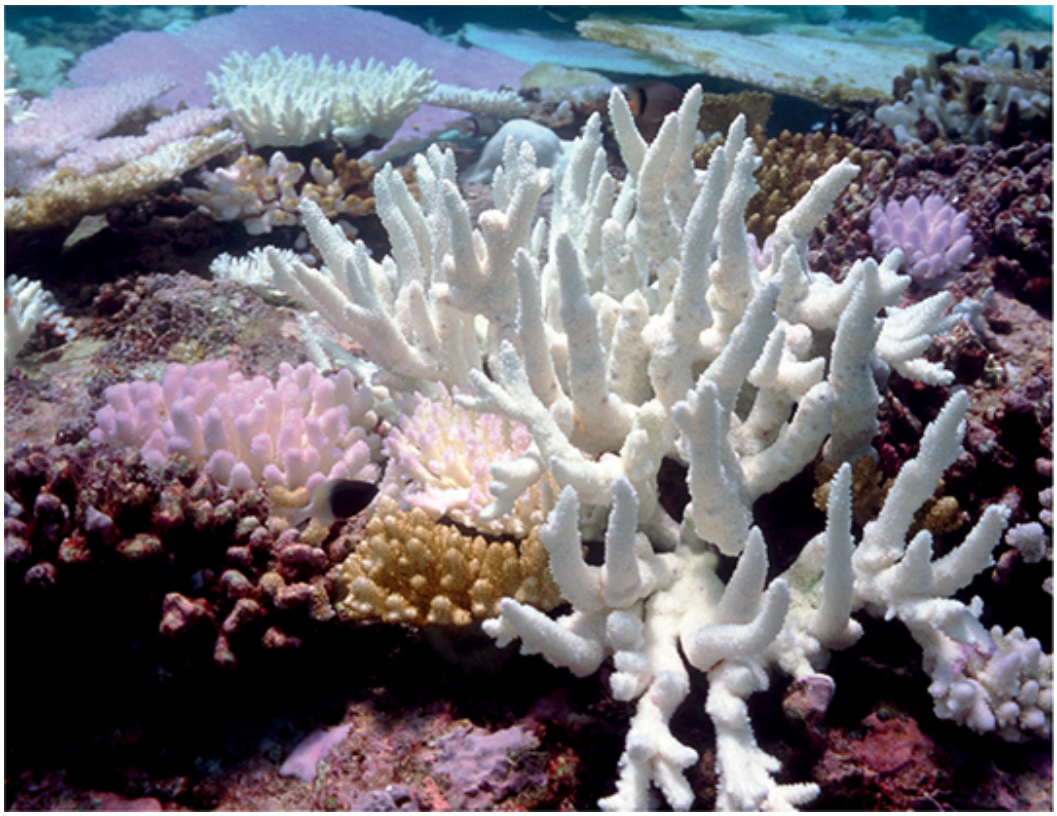


Figure 12.6.3: Coral bleaching.

- Range shift in temperate species – As the climate change alters the conditions in environmental niches, species are shifting their habitat ranges towards areas and conditions to which they are adapted and can survive. Individual species' ability to tolerate change decides which ones need to move and how far. In Europe a study of birds and butterflies showed that butterflies had moved further north to reach cooler conditions by 114 km while birds moved an average of 37 km north. Movements like these enable a species to survive but, as they arrive, incoming species can disrupt the structure of the community that lived there. For example, as the tundra warms and is being taken over by boreal forest, many species such as the arctic fox and snowy owl are losing their habitats. And as rivers and streams warm up,

fish that survive in warmer waters are expanding their ranges into areas that were occupied by cold water species and cold water fish such as trout are losing their habitats.

NATURE OF SCIENCE

Forest regeneration and restoration of peatlands

In Section 12.5 the possibilities of restoration and regeneration of ecosystems were discussed. Tree planting and the establishment of new forests is taking place all over the world.

Native woodlands and peatlands are two of our largest natural climate regulating ecosystems. They both have high biodiversity and are a priority for conservation and restoration. It is generally agreed that sustainable management and enhancement of both peatlands and woodlands are necessary to support biodiversity and reduce climate change. Many trees have been planted on peatlands to provide timber and to absorb carbon, but is new forest planting on peatland the best thing to do? Some of the plantations that have been established on peatlands are now being removed so that peatland can be restored.

Growing trees on peatland is not the most sustainable or cost-effective option for tackling climate change. The amount of carbon absorbed by trees on peatlands is not as high as that of trees grown on non-peat soils and planting trees on peat costs more to manage than planting in a forest.

The environmental importance of lower carbon absorption from trees on peatland must be considered and compared with the alternative of simply restoring the peatland. The optimum solution for both carbon absorption and biodiversity seem to

be to maintain peatlands which have not been planted with trees, restore forested peatland to open habitat and plant new trees on non-peat soils or other areas that can benefit the peatlands.

To consider:

1 Why is tree planting such an important aspect of restoration projects?

2 What other methods might be successful in restoration of a habitat?

12.6.2 Timing of biological events and global warming

KEY POINTS

phenology is the study of cyclic and seasonal events of plants and animals in relation to climate

photoperiod is the length of the day – how much light species receive over 24 hours

Phenology, the timing of annual cycles of plants and animals, is very sensitive to changes in climate. Organisms adjust the timing of certain events such as flowering or migration depending on changes in weather. The timing of these events provides information about climate change to investigators and others who monitor the growth of crops and forests. Changes in the phenology of species are one of the most obvious consequences of climate change as temperature increases and is the major abiotic factor that affects the seasonal timing of life histories and synchrony between different species.

The Arctic mouse ear chickweed (*Cerastium arcticum*) is a small plant that grows in southern Greenland, Iceland, Scotland and Norway. It has white flowers and grows in tufts in loose gravel, it forms part of the vegetation of the ecosystem and provides food for reindeer in summer. The chickweed has been affected by warming of the soil in winter. Climate change affects all seasons, but warming is more pronounced in winter than summer at high latitudes. Snow cover is important in ecosystem processes, but winter warming is decreasing the amount of cover and snow depth. Taller more productive species benefit from a

longer growing season without snow and become dominant, out-competing smaller, species such as mouse ear chickweed by shading them out. Changes like this cause changes in community composition which lead to changes not only in the above ground species but also to nutrient cycles and decomposition in the soil.

Figure 12.6.4: Alpine mouse-ear or alpine chickweed (Cerastium alpinum) growing on a hillside at the Ilulissat Icefjord in Greenland during unseasonably warm weather.

Migrating reindeer (*Rangifer tarandus*) live in boreal and montane ecosystems in the Arctic, but many populations are declining due to changes in their habitat. In the summer months reindeer eat mosses, herbs, such as mouse ear chickweed, ferns, grasses, and the shoots and leaves of shrubs. This diet is important for growth of their young, pregnancy and lactation in females. In winter, they scrape away snow with their hooves to feed on lichen (also called reindeer moss) and fungi. As the Arctic warms, vegetation patterns are changing. Climate change has changed the plants that grow in the Arctic and more rain and warmer winters mean that plants are covered with ice instead of snow. Ice prevents reindeer reaching the lichen beneath it.

Reindeer will need to adapt their range to the changed availability of food if they are to survive.

Another example of the effect of changing temperature on animal behaviour and feeding has been studied in an English oak woodland which is home to the bird, the great tit (*Parus major*). Over the last 60 years records show that temperatures have increase by 2.6°C. In response to this warming the birds have changed the timing of their mating and egg laying so it is now 16 days earlier in the spring than 60 years ago. Great tits feed their young on caterpillars, which they find living on oak tree leaves. But as spring has come earlier in the year oak leaves and caterpillars have been appearing earlier too. But the leaves and caterpillars are appearing even earlier than the young in the nests of the Great tits and this could mean that they could soon be out of synchrony with them.

Figure 12.6.5: Bark beetle on a dead spruce in the Harz mountain region in Germany. In April and for the previous two summers, there was very low rainfall so forests were stressed and more susceptible to pests.

Increases in number of insect life cycles

The spruce bark beetle (*Ips typographus*) is the most important bark beetle pest species in conifer forests of Europe, primarily attacking Norway spruce (*Picea abies).* Bark beetles infest and reproduce in the trees and can cause trees to die over a large area. In the United States species in the genera *Dendroctonus* and *Ips* are the primary pest species and in the years between 1997 and 2010 more than 5 million hectares of trees were affected by bark beetles. Bark beetle outbreaks have intensified in forests all over the world recently and are expected to increase further because of climate change.

Bark beetles have adapted to changing local conditions. Species that infest and reproduce in trees in warmer habitats have evolved mechanisms that allow them to produce several generations in a single year. Species with host trees in colder climates such as the spruce beetle have evolved to survive during cold winters and emerge as adults to attack trees during warm summer months. The effect of warming temperatures differs depending on the species and the season of warming but as changes in climate continue, many trees will be exposed to less suitable growing conditions and may become more susceptible to bark beetle attacks.

Evolution as a consequence of climate change

As climate change and warming temperatures have caused a reduction in snow cover there is one clear example of a species evolving to adapt to the changes.

Tawny Owls in southern Finland are evolving by changing the colour of their feathers. A study over about 30 years has recorded an increase in dark brown owls in the population that until recently has been dominated by light grey birds (Figure 12.6.6) which blend in with the pale snow coloured land. It may

be that selection pressure is no longer favouring the pale owls but the advantages that the dark owls have is still not clear as both colour birds seem to survive equally well. Natural selection acted in a similar way on the Peppered moth in the1800s in the UK when darker coloured individuals had an advantage in soot-covered trees (Section 11.2.2). We may well see natural selection acting on more species as climate change alters the conditions in the environment.

Figure 12.6.6: A grey tawny owl is camouflaged in a snow-covered terrain.

TEST YOUR UNDERSTANDING

36 Outline the effect of climate change on ocean currents

37 How are coral reefs affected by climate change?

38 Define phenology

REFLECTION

Could I summarise the effect that humans have had on ecological relationships and the survival of organisms? Which factors would I describe first to a classmate?

Links

- How do food chains and webs transfer matter and energy in ecosystems? (Chapter 12.2)
- How do human activities affect the evolution of species? (Chapter 11.3)

SELF-ASSESSMENT CHECKLIST

Think about the topics covered in this chapter. Which parts are you most confident with? Which topics require some extra practice?

I can...	Subsection	Needs more work	Nearly there	Confident to move on
distinguish between autotrophs and heterotrophs and chemosynthetic autotrophs	12.1.1			
identify *Euglena* as an autotroph and a heterotroph	12.1.1			
define consumers and saprotrophs by their methods of feeding	12.1.1			
outline the concept of trophic levels	12.1.1			
classify consumers as primary,	12.1.1			

secondary or tertiary consumers based on their diets				
describe the complexity of some trophic feeding relationships	12.1.2			
explain the adaptaions of carnivores, herbivores and humans to their method of feeding	12.1.2			
identify categories of heterotrophs that do not fit easily into food webs	12.1.2			
recognise other feeding relationships including mutualism, parasitism and commensalism	12.1.2			
state that most ecosystems rely on light energy from the Sun and	12.2.1			

that this is converted to chemical energy by photosynthesis				
autotrophs use carbon compounds for their energy needs	12.2.1			
describe how chemical energy in carbon compounds flows through food chains	12.2.1			
summarise how energy is lost as heat from ecosystems and from trophic levels as dead organic material	12.2.1			
state that energy losses limit the length of food chains and biomass accumulation	12.2.1			
describe how food chains and webs are represented in	12.2.1			

pyramids of energy				
distinguish between energy flow and nutrient cycling in an ecosystem	12.2.2			
summarise the events of the carbon cycle				
explain the terms population and community	12.3.1			
describe how to estimate populations using random sampling and the Lincoln index	12.3.2			
use the chi-squared test to assess whether there is association between two species	12.3.2			
outline the effects of intraspecific and interspecific competition	12.3.4			

draw and interpret population growth curves	12.3.3			
define the terms carrying capacity, density-dependent factors and density-independent factors, and explain their effects on population growth	12.3.3			
explain the difference between intra- and interspecific relationships	12.3.4			
summarise the effect of cooperative interspecific interactions (mutualistic) on plants and fungi	12.3.4			
describe the advantages and disadvantages of cooperative intraspecific relationships	12.3.4			

define and give examples of allelopathic relationships in plants, fungi and bacteria	12.3.5			
describe the effects of predator–prey interactions and how these can be modelled	12.3.6			
outline defence mechanisms used by prey animals and plants to avoid being eaten	12.3.6			
recognise that ecosystems may remain stable for long periods of time	12.4.1			
describe why some ecosystems may be reaching a tipping point	12.4.1			
describe the features of a mesocosm and define a	12.4.1			

sustainable ecosystem				
describe conditions that are required for sustainable harvesting from natural ecosystems	12.4.2			
describe factors affecting sustainability of agriculture	12.4.2			
outline human impacts on biogeochemical cycles	12.4.3			
define the terms ecological succession and climax community	12.4.4			
distinguish between primary, secondary and cyclical succession	12.4.4			
describe how human activities can prevent a	12.4.4			

climax community developing				
outline the factors needed for a stable ecosystem	12.5.1			
distinguish between non-renewable and renewable resources	12.5.1			
recall how tipping points can lead to irreversibe change	12.5.1			
outline how mesocosms can be used as models to investigate ecosystem stability	12.5.1			
recall the reasons for ecosystem loss	12.5.1			
summarise the causes and consequences of the 6th mass extinction	12.5.1			
list some of the ways humans	12.5.1			

interfere with stable ecosystems				
summarise the dangers of microplastics in the environment	12.5.2			
define an alien species and outline the problems they cause	12.5.2			
explain the difference between in situ and ex situ conservation, rewilding and regeneration	12.5.3			
explain how human activities disrupt the nitrogen cycle and cause eutrophication	12.5.4			
explain the importance of a keystone species	12.5.5			
define biomagnification and its effect on	12.5.5			

higher level consumers				
recall the reasons that are leading to climate change	12.6.1			
list the effects of climate change on ice melts, ocean currents, species ranges and coral reefs	12.6.1			
explain what is meant by a phenology and give examples of phenological events	12.6.2			
explain how climate change can lead to evolution.	12.6.2			

EXAM-STYLE QUESTIONS

You can find questions in the style of IB exams in the digital coursebook.

Acknowledgements

The authors and publishers acknowledge the following sources of copyright material and are grateful for the permissions granted. While every effort has been made, it has not always been possible to identify the sources of all the material used, or to trace all copyright holders. If any omissions are brought to our notice, we will be happy to include the appropriate acknowledgements on reprinting.

Thanks to the following for permission to reproduce images:

Cover Martin Harvey/GI; *Inside* **Unit 1:** Fernan Federici/GI; Arterra/Universal Images Group via GI; Paul Souders/GI; Anastasiia Usenko/Shutterstock; A. Barrington Brown/SPL; Universal History Archive/GI; **Unit 2:** Choksawatdikorn/Science Photo Library/GI; CNRI/SPL; David Parker/SPL; Nigel Cattlin/Alamy Stock Photo; DR Jeremy Burgess/SPL; **Unit 3:** KTSDESIGN/Science Photo Library/GI; J.C. Revy/ISM/SPL; DR Elena Kiseleva/SPL; Eye of Science/SPL; Reproduced with permission of Randy L' Jirtle at the University of Bedfordshire and the University of Wisconsin; **Unit 4:** SCIPRO/GI; Power & Syred/SPL; DenKuvaiev/GI; Sovereign,ISM/SPL; Masahiro Makino/GI; Despite Straight Lines (Paul Williams)/GI; **Unit 5:** Nanoclustering/Science Photo Library/GI; New Zealand American Submarine Ring of fire 2007 Exploration, NOAA Vents program, The Institute of Geological & Nuclear Sciences and NOAA-OE/SPL; Anadolu Agency/GI; M.I. Walker/SPL; Sinclair Stammers/SPL; BSIP, Sercomi/SPL; Ed Reschke/GI; DR Kari Lounatmaa/Science Photo Library; Astrid & Hansfreider Michler/SPL; National Institutes of Health/Stocktrek Images/GI; Fig 5.4.14 reproduced with permission of Visualhistology.com; Fig 5.2.16 reproduced with permission of Anton Page,

Biomedical Imaging Unit, University of Southampton; Don W. Fawcett/SPL; **Unit 6:** Raycat/GI; Steve Gschmeissner/SPL; Thinkstock Images/GI; Michael Abbey/SPL(X5); Lennart Nilsson, TT/SPL; **Unit 7:** Rost-9D/GI; BSIP/GI; vgajic/GI; **Unit 8:** Ed Reschke/GI; Tek Image/SPL; Science Photo Library/SPL; Science Photo Library-Ian Hooton./GI; DR Keith Wheeler/SPL; Garry Delong/SPL; DR Jeremy Burgess/SPL; Frank Fox/SPL; Gregory Dimijian/SPL; Astrid & Hanns-Frieder Michler/SPL; Garry Delong/SPL; damircudic/GI; Georg Ivan/GI; Sandra Standbridge/GI; **Unit 9:** koto_feja/GI; Kevin Schafer/GI; Mark Newman/GI; Digital Vision/GI; Dr Rosalind King/SPL; Martin Harvey/GI; George Karbus Photography/GI; **Unit 10:** dem10/GI; CNRI/SPL; Andrew Holt/GI; Sean Gladwell/GI; **Unit 11:** Natalia Crespo/GI; Sheila Terry/SPL; Ian_Redding/GI; John Durham/SPL; Mint Images/GI; La Su/GI; Chris Howarth/Galapagos/Alamy Stock Photo; Dr P. Marazzi/SPL; blickwinkel/Alamy Stock Photo; Fat Camera/GI; Chris Hellier/SPL; DEA/C.Galasso/GI; Ger Bosma/GI; sjharmon/GI; Ricardo Lima/GI; Martyn F. Chillmaid/SPL; Flowerphotos/GI; mtreasure/GI; robas/GI; TCYuen/GI; Annie Haycock/SPL; **Unit 12:** Abstract Aerial Art/GI; Nora Carol Photography/GI; Bob Gibbons/SPL; manonvanos/GI; George D.Lepp/GI; Avalon/GI; robertharding/Alamy Stock Photo; Martyn F.Chillmaid/SPL; ullstein bild/GI; Pat & Tom Leeson/SPL; lindamariecaldwell/GI; zmeel/GI; Sylvain Cordier/GI; Vittoriano Rastelli/GI; Art Wolfe/SPL; ©Nature Production/naturepl.com; Noel Hendrickson/GI; Arctic-Images/GI; Universal History Archive/GI; Education Images/GI; Nora Carol Photography/GI; Vaara/GI; Rosanne Tackaberry/Alamy Stock Photo; Peter Yates/SPL; Roslan Rahman/GI; Cal Holman/GI; Luis Robayo/GI; Georgette Douwma/SPL; Sean Gallup/GI; Jens Schlueter/GI; Sherry Epley/500px/GI

Key GI= Getty Images, SPL= Science Photo Library

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press & Assessment is a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781009039680

First published 2011
Second edition 2014
Third edition 2023

20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

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

ISBN 978-1-009-03968-0 Coursebook with digital access (2 years)
ISBN 978-1-009-01840-1 Digital coursebook (2 years)
ISBN 978-1-009-01843-2 Coursebook (eBook)

Additional resources for this publication at www.cambridge.org/go