Option B Biotechnology and bioinformatics

Introduction

Microorganisms, or microbes – so called because of their small size – have a vital role to play in all ecosystems. They recycle waste, fix nitrogen and can be used by humans to make foodstuffs such as bread, cheese, yoghurt, tofu and beer. Many microorganisms are used in industry and genetically modified bacteria also produce human proteins such as insulin and growth hormone. It has been estimated that microbes account for almost half the biomass on Earth.

Although the majority of microbes are harmless, a few species do cause disease or serious illness. Influenza, polio and HIV are viral diseases and bacteria cause tuberculosis, cholera and leprosy.

B1 Microbiology: organisms in industry

Classifying microorganisms

Microbes are difficult to classify because of their small size and in the last 30 years our understanding of their similarities and differences has changed as new techniques have helped in our study of their structure and biochemistry.

In Subtopic **5.3**, you learned that all living organisms are classified into a number of kingdoms. In the past, it was commonly agreed by scientists that the organisms in these kingdoms could be separated into two groups, based on their structures:

- **prokaryotes**, with little cellular organisation and no organelles such as a nucleus
- **eukaryotes**, which contain organelles including a nucleus

New data from the sequencing of RNA has shown that the group known as prokaryotes should be split into two distinct groups – the Archaea and the Eubacteria. The eukaryotes still form a separate group, called the Eukarya. These groups are called **domains** (Figure **B.1**). The Archaea and Eubacteria are the groups containing most microorganisms and, as they have been studied further, details about other structures and molecules have been discovered that support this three-domain model.

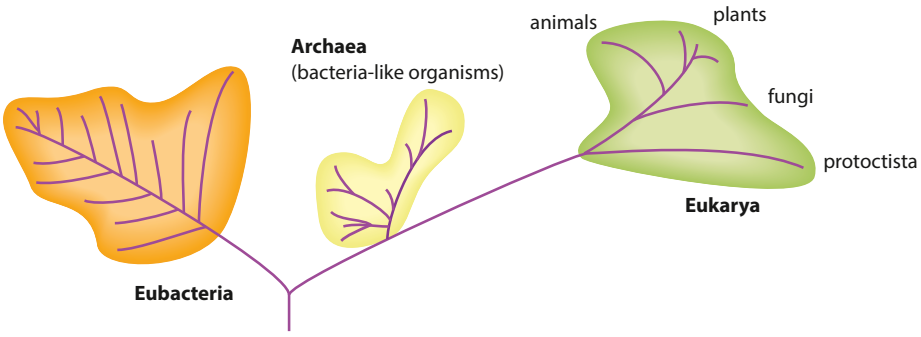


Figure B.1 A phylogenetic tree showing the grouping of organisms into three domains.

Learning objectives

You should understand that:

- There are many different types of microorganism, which have different metabolic reactions.
- Microorganisms are useful in industry because they are small and they grow and reproduce at a rapid rate.
- Pathway engineering is a way of optimising the regulatory and genetic processes that occur in microorganisms.
- In industrial processes, pathway engineering can be used to produce metabolites that are of interest.
- Useful metabolites can be produced on a large scale using fermenters.
- Growth of microorganisms in fermenters is carried out by batch or continuous culture.
- The growth of microorganisms in fermenters is limited by the accumulation of waste products.
- Conditions within fermenters are monitored by probes.
- Microorganisms are cultured by maintaining conditions at optimal levels for the growth of the organisms.

Diversity of metabolism in microorganisms

Microbes have a varied range of metabolic processes that enable them to use different sources of energy and carbon. As Table **B.1** shows, microbes are divided into four groups based on their methods of metabolism: **photoheterotrophs**, **chemoheterotrophs**, **photoautotrophs** and **chemoautotrophs**.

Method of metabolism	Energy source to generate ATP	Carbon source used to obtain organic compounds	Example
photoheterotrophic	light	organic compounds	*Heliobacter* is a photoheterotrophic bacterium found in waterlogged soils and paddy fields. It is able to fix nitrogen so it is probably important in soil fertility.
chemoheterotrophic	chemical reactions	organic compounds	Fungi such as the yeast *Saccharomyces* are chemoheterotrophic microbes. These organisms cannot photosynthesise, so they must use organic material as an energy source. They respire or ferment sugars to make ATP. Most bacteria are chemoheterotrophs.
photoautotrophic	light	inorganic carbon dioxide	An example of a photoautotroph is *Anabaena*, a cyanobacterium (filamentous blue–green bacterium) found among freshwater plankton and on grass. It fixes nitrogen and forms symbiotic relationships with some plants.
chemoautotrophic	chemical reactions	inorganic carbon dioxide	*Nitrobacter*, a nitrifying bacterium found in the soil, is an example of a chemoautotroph. Others include the sulfur-oxidising Archaea that live in hostile environments such as deep sea vents.

Table B.1 Different types of microbe metabolism.

Photoheterotroph microbe that uses light energy to generate ATP but which gets the organic compounds it needs from other organisms

Chemoheterotroph microbe that uses the chemical energy released from chemical reactions to generate ATP and obtains organic compounds from other organisms

Photoautotroph microbe that uses light energy to create ATP and to produce organic material from simple inorganic materials

Chemoautotroph microbe that uses energy released from chemical reactions to generate ATP and which makes its own organic material from simple inorganic materials

Diversity of Archaea and Eubacteria

The **Archaea** (Archaebacteria or ancient bacteria) inhabit some of the most extreme environmental conditions on the planet and those that reflect the conditions present in the early part of the Earth's existence. Three different groups of Archaea are identified from their metabolism and habitats.

- **Thermophilic bacteria** have evolved to survive at temperatures in excess of 70 °C and up to 100 °C in some cases. They inhabit hot, sulfurous springs in volcanic regions and hydrothermal vents on the ocean floor. One species, *Thermus aquaticus*, provides the enzyme DNA polymerase, vital for use in the polymerase chain reaction (PCR) for amplifying copies of DNA.
- **Halophilic bacteria** live in very salty environments like tidal mud flats and inland lakes (such as the Dead Sea) where the Sun has evaporated much of the water. They are also found in salt mines.
- **Methanogenic bacteria** are anaerobes found in the guts of ruminants (cows and sheep) and termites as well as in waste landfills, sewage works, paddy fields and marshland. They produce methane as a waste product of respiration. Methane is a major greenhouse gas.

Eubacteria evolved at the same time as the **Eukarya**, yet still possess primitive features. They have no internal organelle structure and often reproduce by simple binary fission. They are divided into groups based on their shapes, as shown in Figure **B.2**.

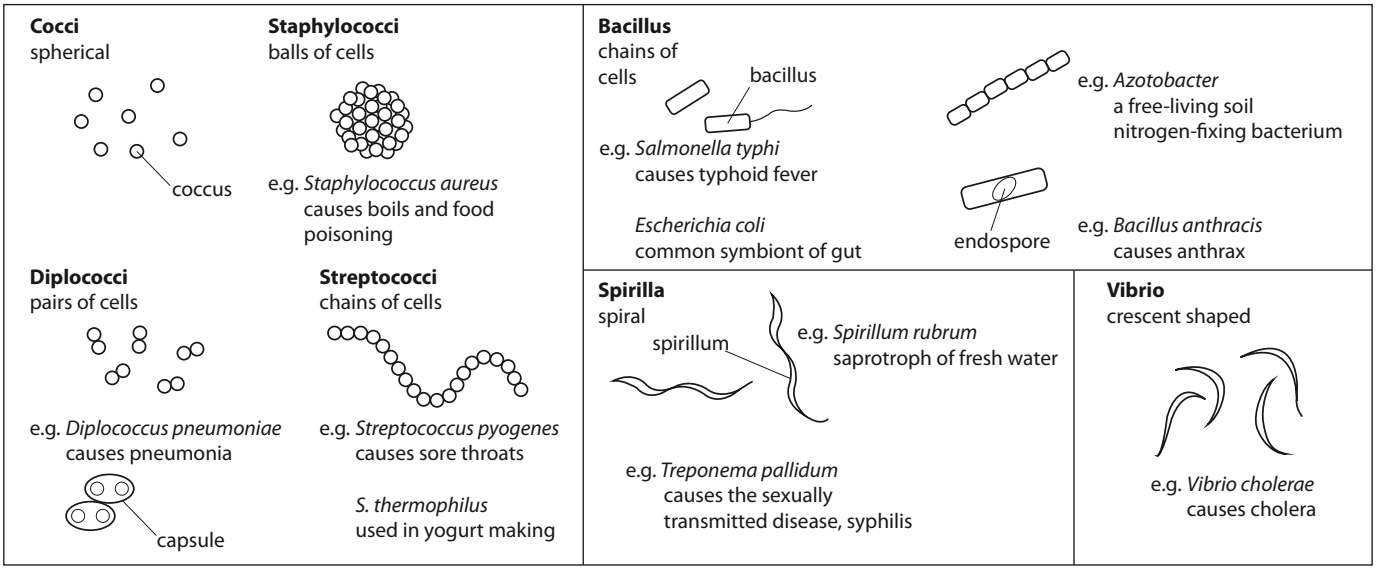


Figure B.2 Shapes of Eubacteria.

Exam tip
Remind yourself of the relationship between the three domains and the different groups of Archaeans by drawing a family tree or cladogram.

Gram-positive and Gram-negative Eubacteria

As well as being recognised by their shape (Figure **B.2**), bacteria can be separated into two main types by the structure of their cell walls. The Gram staining method is a useful way of differentiating these two types:

- Gram-negative bacteria do not retain the colour when the dye crystal violet is added.
- Gram-positive bacteria retain the dye and appear purple, even when washed in a decolourising solution.

These results are explained by the difference in structure of the cell walls – Gram-positive bacteria have large amounts of peptidoglycan, which retains the dye, in their cell walls whereas Gram-negative bacteria do not (Figure **B.3**). The main differences between Gram-positive and Gram-negative bacteria are outlined in Table **B.2**.

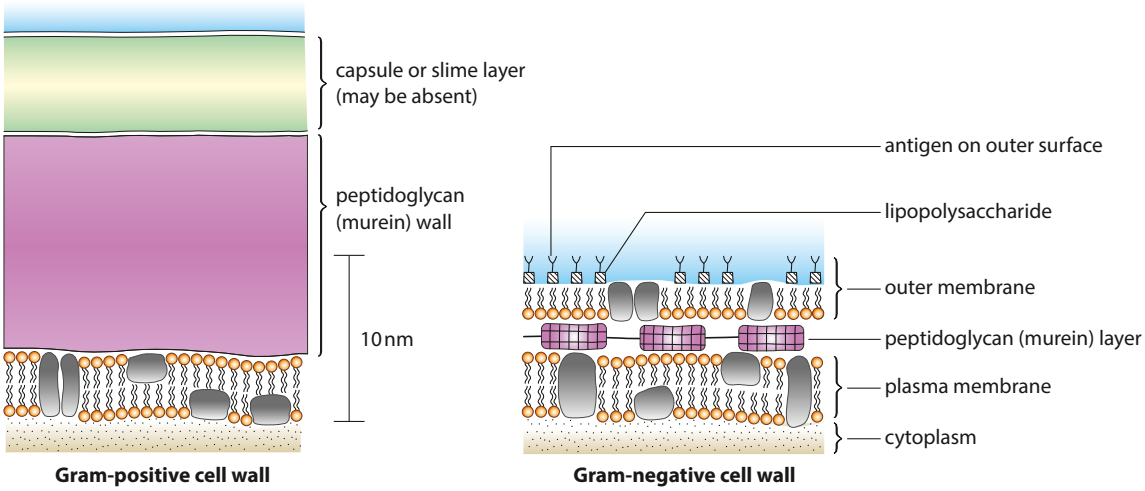


Figure B.3 Peptidoglycan consists of sugar molecules joined to polypeptides, which surround and protect the cell.

Gram-positive	Gram-negative
thick cell wall	thin cell wall
several layers of peptidoglycan connected by peptide bridges	layer of peptidoglycan sandwiched between inner and outer layer
no outer layer	outer layer contains lipopolysaccharide (LPS) and protein

Table B.2 Differences between Gram-positive and Gram-negative bacteria.

Diversity of microscopic eukaryotes

The microscopic eukaryotes are a diverse group of organisms. Using a microscope, it is possible to see the variety in their structure. They also vary in their methods of nutrition and of movement. Table **B.3** summarises the differences found in the group. The structures of the organisms listed in the table are shown in Figure **B.4**.

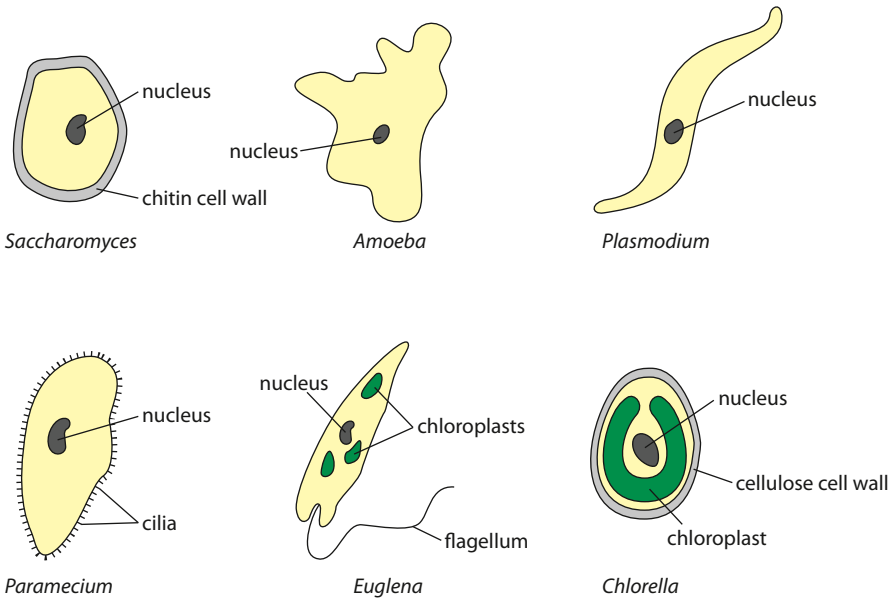


Figure B.4 Basic structures of the microscopic eukaryotic organisms listed in Table **B.3**.

Organism	Cell structure	Nutrition
Saccharomyces sp. (yeast)	cell wall made of chitin	heterotrophic, absorbs small molecules saprotrophically and feeds on sugars
Amoeba	no cell wall	heterotrophic, feeds on other organisms
Plasmodium (malarial parasite)	no cell wall	parasitic, some of its life cycle occurs in human cells
Paramecium	no cell wall	heterotrophic, takes in food by endocytosis
Euglena	no cell wall	autotrophic and heterotrophic, contains chlorophyll
Chlorella (single-celled green alga)	cell wall made of cellulose	autotrophic, photosynthesises using a chloroplast

Table B.3 Differences between types of eukaryotic organisms.

Microorganisms in industry

Microorganisms are used in commercial processes that produce foodstuffs, medicines and industrial products such as enzymes. Their fast growth rate and small size means that suitable amounts of product can be produced relatively quickly and efficiently.

Strict safety precautions are needed to ensure that products are pure and uncontaminated. This is very important where genetically engineered strains are used. Microorganisms may be cultured either on a solid substrate, which is a more traditional method used in cheese and sauerkraut production, or in an aqueous culture. Aqueous cultures may be produced by the batch culture method or the continuous culture method (see below).

Pathway engineering

Pathway engineering is a way of manipulating metabolic pathways in microorganisms so that particular metabolites of interest are produced in useful quantities. It is a developing new technology that has the potential to assist in many production methods based on biological organisms. It has been used in the production of both penicillin and citric acid to improve productivity, efficient use of carbon sources and product purity. Study of **genomics**, **proteomics**, **fluxomics** and **physiomics** are all used in pathway engineering, and can significantly increase the amount of product produced per cell.

Metabolic pathway engineers use four key strategies to achieve their results:

1. **optimisation of the primary metabolic pathway** by removal of factors that limit the rate of the reaction or transcription, as well as removing allosteric regulation, so that more product is produced
2. **genetic blocking** of any competing pathways
3. **maximising the amount of carbon substrate** that is directed to the metabolic pathway of interest and away from other pathways
4. **modification of secondary metabolic pathways** so that energy metabolism and availability of required enzymes and cofactors for the pathway of interest are maximised.

Two key examples of successful pathway engineering have been in the production of penicillin from *Penicillium* sp. bacteria and of citric acid from the fungus *Aspergillus niger*.

Knowledge of the **gene clusters** involved in penicillin production has enabled pathway engineers to increase output by the amplification of certain genes. This development first led to a 40% increase in output in the 1990s and later the introduction of structural genes and a whole gene cluster increased production still further.

Genetic improvement of *Aspergillus niger* used in industry was achieved by the use of mutagens, particularly gamma ray irradiation, so that citric acid yields increased more than three-fold. Further improvements were achieved by genetic engineering of the primary metabolism of *Aspergillus*. This increased the 'metabolic flux' or flow through the pathway forming citric acid, a product that would not normally accumulate. One method used is to decrease the flows through branches off the main pathway so that fewer by-products are formed. Another way is to engineer organisms so that they overproduce the enzymes needed for citric acid production so that the flux through the main pathway increases.

Genomics the study of the genome of an organism
Proteomics the study of the structure and functions of proteins
Fluxomics the study of the flow of fluids and molecules within cells
Physiomics the study of an organism's physiome, the interconnections of aspects of physiology that result from genes and proteins in the organism

Gene clusters sets of two or more genes that code for the same or similar products

Microorganisms naturally produce new genotypes as a result of mutations. Most mutations are harmful to the organism and are eliminated by natural selection but a few are beneficial. Others are not beneficial to the microorganism but are useful to humans. Microbiologists purposely apply **mutagens** (such as nitrosoguanidine, hydroxylamine, ultraviolet light and gamma rays) to different microorganisms, including *Aspergillus*, to modify their genomes – any useful mutations (for example, those that lead to an increase in the production of substances such as antibiotics) can then be detected by selection and screening processes and conserved in the genome. This has led to huge improvements in production in industrial microbiology since the 1960s.

Figure **B.5** shows some useful products that are made from glucose metabolism by fungi such as *Aspergillus* sp. and *Penicillium* sp.

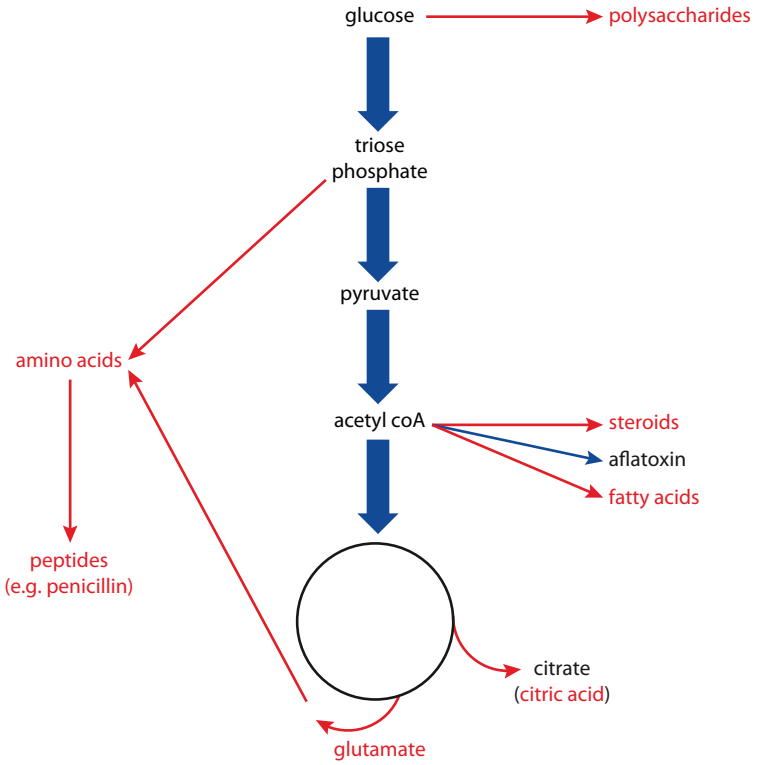


Figure B.5 Diagram to show some useful products of the metabolism of glucose by fungi. Useful substances are shown in red. Pathway engineers optimise genetic or regulatory processes in microorganisms to maximise production of these substances.

Large-scale production of metabolites of microorganisms

In order to produce and harvest the useful metabolites of microorganisms, a large-scale industrial process is needed. Most organisms are cultured in aqueous culture, a method in which the substrate has high water content. Two methods are commonly used: the batch culture method and the continuous culture method. In both cases, cells are grown in **fermenters** (Figure **B.6**), which are large tanks that can hold up to 200 000 litres. Cells are provided with nutrients and the exact environmental conditions they need to grow at their maximum rate. If the microorganisms used are aerobic, air is pumped in, and the culture is also stirred to ensure even mixing of nutrients throughout the culture. Growth in a fermenter can be limited by accumulation of waste products but in a continuous culture these can be removed.

Nutrients enter the fermenter through valve-operated pipes so that exact quantities can be controlled to maximise growth. All fermenters must be sterile before the process begins and this is usually achieved by passing steam through the pipes and tanks. People who work in the biotechnology industry must also take precautions and wear protective clothing to avoid contamination of the process or of themselves.

Although the culture vessels are known as fermenters, the metabolic process that occurs in them does not have to be fermentation (that is, anaerobic respiration).

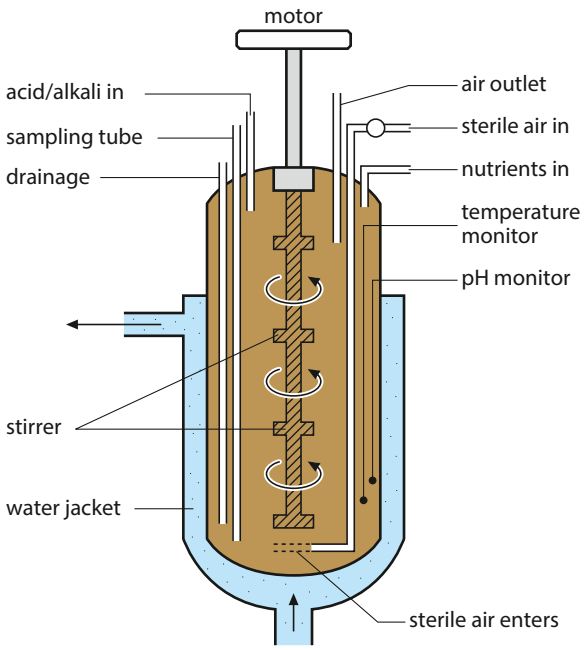


Figure B.6 The inputs and outputs of a typical fermenter.

Batch culture

Batch culture is used in the mass production of penicillin, and is carried out in a closed fermenter. The microorganism (*Penicillium* sp.), from a culture whose genome is known and stable, is added to the fermenter with the necessary nutrients, which have been pre-sterilised. The fermenter is then left for the process to take place. As the microorganisms grow and reproduce, nothing is added to or removed from the fermenter except waste gases, which are allowed to escape. As the curve in Figure **B.7** shows, the microorganisms multiply rapidly at first, but this exponential phase of growth is usually quite short.

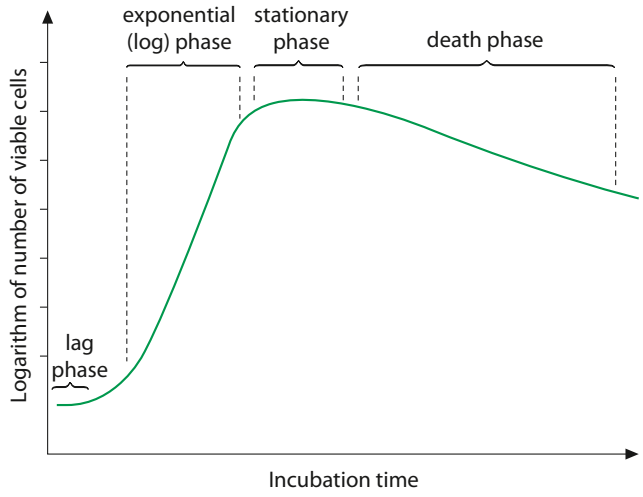


Figure B.7 Growth curve of microorganism in batch culture.

Secondary metabolites substances produced by microorganisms after the main growth phase has been completed, and which are not essential for normal cellular functions; penicillin produced by *Penicillium* sp. is an example of a secondary metabolite

As a batch culture proceeds, the temperature can rise by 1 °C per hour, and because penicillin is a protein and can easily be denatured by high temperatures, the fermenter must be cooled. Cooling coils or a water jacket may be used in some fermenters. Although the temperature is kept constant, other conditions in the fermenter change – up to 0.5 tonnes of glucose per day may be used and waste products begin to accumulate. Penicillin is a **secondary metabolite** so eventually the amount of product made will start to decline as the reproduction of the microorganisms slows down. At this point the fermenter will be shut down for cleaning and restocking.

Continuous culture

Continuous culture is used in the manufacture of citric acid by the fungus *Aspergillus niger*. Citric acid is difficult to produce from inorganic chemicals but is quickly and easily made by microorganisms. In continuous culture, nutrients are supplied continuously to the fermenter at a steady rate (usually the nutrients are molasses or hydrolysed corn starch, which are cheap sources of sugar). Continuous culture matches supply and demand so that the organisms can be kept in an exponential phase of growth. The conditions in the fermenter are carefully monitored by probes that check pH, nutrient concentration, oxygen levels and product levels. Used medium is drained off, together with the citric acid product. The fungi are filtered out of the liquid that leaves the fermenter and citric acid is extracted by precipitation with calcium hydroxide.

A continuous culture method means that an uninterrupted supply of product is made and the equipment can be used without breaks for cleaning and recharging the fermenter. Continuous culture can therefore be more economical than a batch culture method. Continuous and batch culture methods are compared in Table **B.4**.

Batch culture	Continuous culture
product is separated from the mixture at the end	product is harvested continuously
conditions inside the fermenter change (although temperature is monitored)	all environmental factors are monitored and kept as constant as possible (sometimes this is difficult and production can be disrupted)
cells have a relatively short time in the exponential growth phase	cells are kept in the exponential growth phase
larger 'deep tank' fermenters are needed	smaller fermenters can be used – process is more productive
usually less cost effective, but if a batch is contaminated only one batch is lost	continuous process is more economical, but if a fermenter is contaminated losses are greater

Table B.4 Comparing batch and continuous culture methods.

Uses of citric acid

Large volumes of citric acid are produced by fermentation every year – estimates put production at around a million tonnes per year. Citric acid is a natural preservative and is also used to add an acid taste to foods and soft drinks (in Europe it is classified as food additive number E330).

Citric acid is also used:

- as an emulsifying agent to prevent fats from separating in ice cream and other products
- with sodium bicarbonate in effervescent indigestion tablets
- as a grease remover in many cleaning products
- as a pH adjuster in creams and gels.

Biogas production

Biomass in the form of wood and agricultural waste, such as straw and animal manure, already provides a useful source of fuel. Now many countries are looking at the use of **biofuels** to reduce their dependence on fossil fuels. Biofuels are produced by converting biomass into ethanol or methane. This is done in **bioreactors**, either on a large industrial scale or in small fermenters on farms or in villages. Methane produced from animal manure and agricultural waste is known as **biogas**. A simple, small-scale fermenter that can be used to produce biogas is illustrated in Figure **B.8**. Manure and straw are fed into the bioreactor, where they decompose anaerobically as different groups of bacteria present in the manure break down the organic material. The slurry that remains is a useful fertiliser.

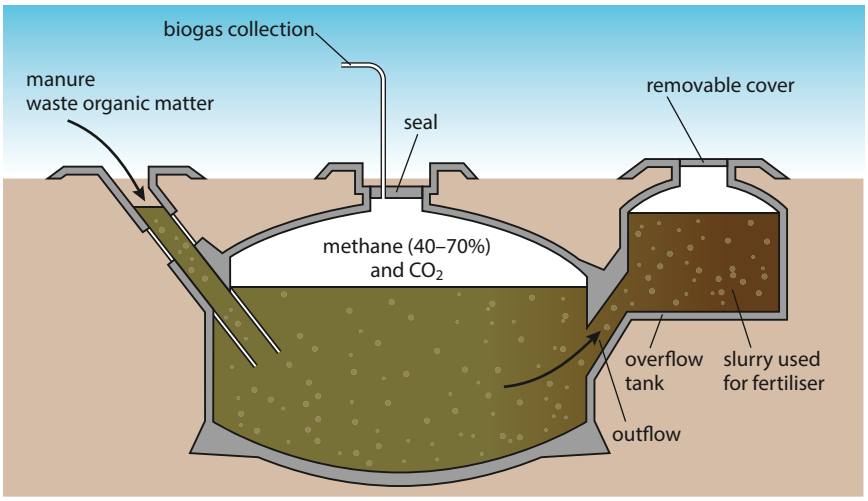


Figure B.8 Cross-section of a biogas reactor.

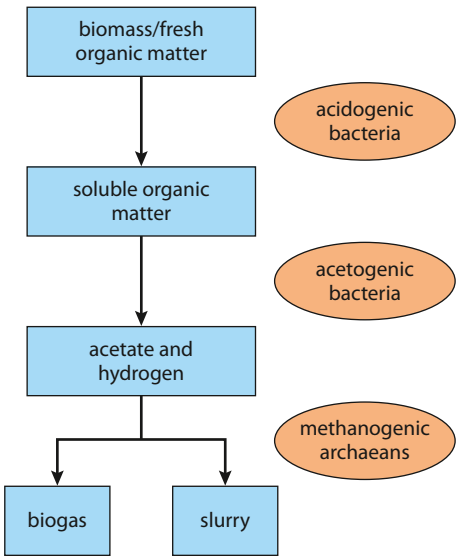


Figure B.9 Processes in biogas production. Three groups of microorganisms digest organic waste to produce methane

Three groups of microorganism produce the enzymes that digest organic material, and each stage breaks down the complex carbohydrates, fats and proteins into simpler compounds.

- Organic material is first converted to organic acids and ethanol by anaerobic, **acidogenic bacteria**, which occur naturally in manure.
- **Acetogenic bacteria** then use the organic acids and ethanol to produce acetate (ethanoate), carbon dioxide and hydrogen.
- Finally, **methanogenic archaeans** produce methane either from carbon dioxide and hydrogen or by breaking down acetate:

carbon dioxide + hydrogen → methane + water

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$$

acetate (ethanoate) → methane + carbon dioxide

$$CH_3COOH \rightarrow CH_4 + CO_2$$

The sequence of these processes is shown in Figure **B.9**.

The biogas that is produced contains up to 70% methane and about 30% carbon dioxide. Biogas can be used to produce electricity or burned directly as a renewable fuel, and the by-products of the reactions can be used as fertiliser.

The production of biofuels has the advantage of being **sustainable** because plants regrow each season. In addition, it makes use of the methane gas that is naturally produced by the anaerobic digestion of organic matter. Methane is a potent greenhouse gas, partly responsible for global warming. On a small scale, biofuels produced at a local level have many advantages but some people have voiced concerns that problems may arise if large areas of land are used to grow crops for biofuels rather than food.

Worldwide, a range of fuels is now produced from biomass on a large scale. Bioethanol can be used as a fuel for vehicles in its pure form, but it is usually used as a petrol additive to increase a car's performance and improve vehicle emissions. Bioethanol is widely used in the USA and in Brazil. Biodiesel made from oils and fats is most often used as a diesel additive to reduce levels of emissions, and is the most common biofuel used in Europe. In 2008, biofuels accounted for almost 2% of the world's transport fuel.

Nature of science

Serendipity in science – the discovery of penicillin

The discovery of antibiotics began by accident. On 3 September 1928, Professor Alexander Fleming (1881–1955) 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 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. This chance event led Fleming to consider the possibility that the bacteria around the circular mould had been killed off by a substance produced by the mould. He discovered that the mould could kill other bacteria and that it could be given to small animals without any harmful effects. However, he moved onto other research and it was not until 10 years later that Howard Florey (1898–1968) and Ernst Chain (1906–1979), 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. In 1945 Fleming, Chain and Florey shared the Nobel Prize for Medicine.

Luck or judgment?

Alexander Fleming noticed that one agar plate he was about to discard was contaminated by a mould. He was surprised to see that the mould had not only stopped the bacterial growth but had killed the bacteria.

Questions to consider

- To what extent was Fleming's discovery a lucky observation?
- Are we more likely to perceive those events that we are open to?

Test yourself

1 Outline the differences between batch culture and continuous culture of microorganisms.
2 State the name of the organism used in the commercial production of citric acid.
3 Outline the modes of nutrition of *Saccharomyces*, *Euglena* and *Chlorella*

B2 Biotechnology in agriculture

Learning objectives

You should understand that:

- Transgenic organisms have been created to produce proteins that were not part of their species proteome in the past.
- Genetic modification can overcome environmental resistance and therefore increase crop yields.
- Genetically modified crops can be used to produce new products.
- Bioinformatics has helped in identifying target genes.
- A target gene is attached to other sequences, which control its expression.
- An 'open reading frame' is defined as a significant length of DNA beginning at a start codon and ending with a stop codon.
- To create genetically modified plants, recombinant DNA must be inserted into cells and be taken up by chromosomes or chloroplast DNA.
- Successful uptake of introduced DNA is indicated by the use of marker genes.
- Recombinant DNA can be introduced into protoplasts, leaf discs or whole plants.
- Recombinant DNA can be introduced either indirectly by vectors or directly by physical or chemical methods.

Genetic modification in agriculture

Transgenic describes an organism that contains genes from another organism

Gene technology enables scientists to transfer genes from one species to another completely different species and create a **transgenic** organism in just one generation. For example, bacterial genes have been transferred to plants, human genes transferred to bacteria and spider genes transferred to a goat. Gene transfer is possible because the genetic code is universal – no matter what the species, the genetic code spells out the same information and produces an amino acid sequence in one species that is exactly the same in any other species.

Through gene transfer, transgenic organisms can be created that are able to produce proteins that were not previously part of their species' proteome. Animals have been farmed for the products of such inserted genes – for example, Factor XI, a clotting factor used in the treatment of hemophilia, can be produced in the milk of genetically modified sheep and alpha 1-antitrypsin, a protein whose deficiency leads to the breakdown of lung tissue, has also been produced in the same way. Research continues on transgenic goats, bred to produce a growth hormone that may bolster the immune systems of transplant recipients and reduce the side effects of chemotherapy.

Genetic modification in crop plants

By the first decade of the twenty-first century, almost 100 plant species had been genetically modified and many trials have taken place to assess their usefulness. In comparison, there are very few examples of genetically modified animal species. Most genetic engineering has involved commercial crops such as soybeans, maize, potatoes, tomatoes and cotton. Plants have been modified to make them resistant to pests and disease, tolerant to herbicides and able to produce novel products. Some have been modified to extend their ranges so that they are able to tolerate drier conditions and increase crop yields in these areas. Three examples of important genetically modified crops are described below.

Glyphosate-resistant soybeans

Herbicides are used to kill weeds in crop fields but they are expensive and can affect local ecosystems as well as cultivated areas. One commonly sprayed and very powerful herbicide is glyphosate, which is rapidly broken down by soil bacteria. For maximum crop protection, farmers needed to spray several times a year. But now, the genes from soil bacteria have been successfully transferred into soybean plants making them resistant to the herbicide.

Farmers can plant the modified seeds, which germinate along with the competing weeds. Spraying once with glyphosate kills the weeds and leaves the soybean plants unaffected. The crops then grow and out-compete any weeds that grow later when the glyphosate has broken down in the soil. Yields are improved and less herbicide has to be used.

Glyphosate-resistance genes are introduced into soybeans using a vector, the bacterium *Agrobacterium tumefaciens*. This bacterium causes tumours known as galls when it infects plants – it has a large tumour-inducing plasmid known as Ti, which it incorporates into the DNA of infected plants. This Ti plasmid has been used to introduce glyphosate-resistance genes into the genomes of both soybeans and maize (Figure **B.10**), both of which are vital agricultural crop plants. The genes conferring glyphosate resistance are transferred from plant cells of resistant species into the Ti plasmid using **restriction endonucleases** to open the plasmid. These recombinant plasmids are reintroduced into *A. tumefaciens* cells, which then infect plant cells susceptible to glyphosate. Plant cells with included resistance are grown in tissue culture and whole crop plants have been produced from them.

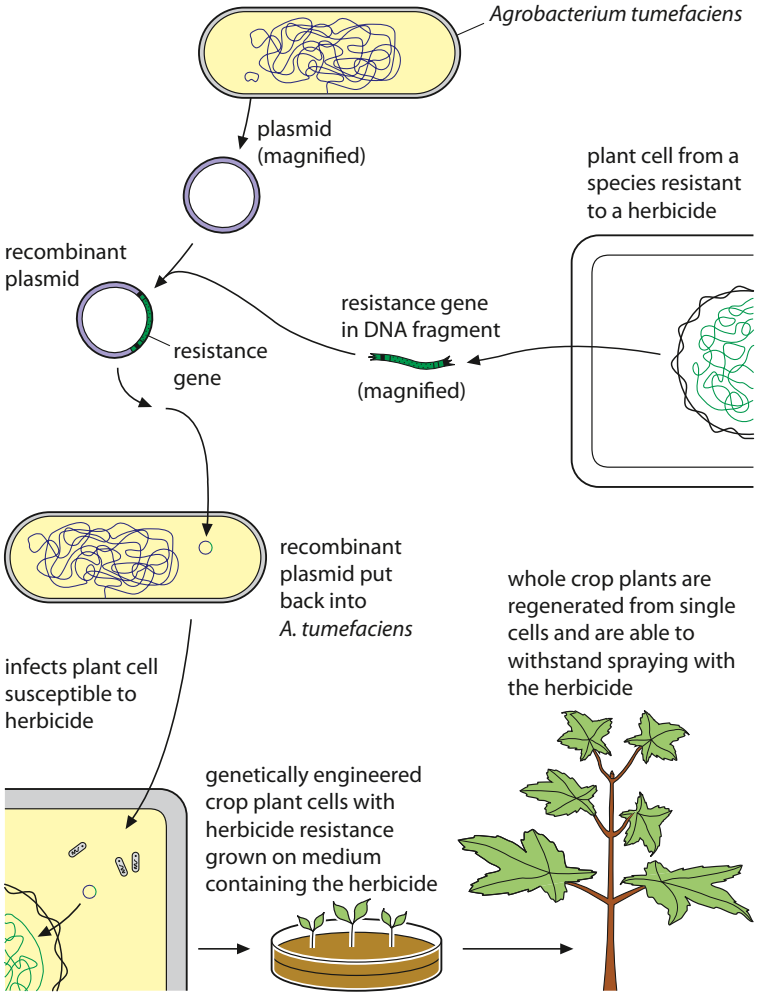


Figure B.10 Transfer of glyphosate-resistance genes by Ti plasmid.

Environmental impact of glyphosate-tolerant soybeans

Transgenic glyphosate-resistant soybeans are grown in many countries, where scientists and environmentalists closely monitor their effects on the environment. The number of transgenic soybean plants grown in the USA increased from approximately one third of all soybean plants in 1998 to more than a half, 2 years later. Studies showed that in the USA, in some areas, there was a 40% decrease in the use of herbicides on the glyphosphate-resistant beans compared with non-transgenic beans, between 1995 when the plants were introduced and 1998. The use of glyphosate-resistant soybeans has also meant that farmers have to dig and turn the soil less frequently, so environmental degradation due to agriculture has been reduced. Some groups have raised concerns that glyphosate-resistance genes in soybeans could be transferred to wild species, but this risk has been found to be low because soybean plants are almost completely self-pollinated. Furthermore, in many countries there are no wild relatives of soybean plants so transmission to other species is very unlikely. One of the possible negative impacts of the soybeans is weed resistance. Weeds that can resist glyphosate will become dominant, as will those that spring up at times when the glyphosate spray is not applied, and thus avoid its application.

Exam tip
If you are asked to discuss a subject, make sure you consider more than one point of view in your answer

Figure B.11 Tobacco leaves infected with TMV.

Producing vaccines in tobacco plants

Tobacco mosaic virus (TMV) is an RNA virus that infects plants. It has recently been adapted for vaccine development, because its simple structure – a strand of RNA encapsulated by just one type of coat protein – means that its genome can be easily manipulated. The virus has been modified to carry genes for several proteins, including those which induce antibody and T-cell responses in animals and so can be used to provide protection against a number of viral diseases such as flu and hepatitis B. This method has already produced a vaccine against Newcastle disease, a viral disease in chickens, and has been extensively investigated for use in the rapid production of flu vaccine in case of a pandemic among humans. Producing vaccines using plants infected with manipulated viruses is known as 'molecular farming'.

As its name suggests, TMV infects tobacco plants, which are fast-growing and have large leaves (Figure **B.11**) – they produce large quantities of biomass in a short period of time so are ideal candidates for vaccine production.

Vaccine production

Vaccines are made by first introducing multiple copies of the genetic information needed to produce the target proteins into the TMV viruses. This genetic information must be introduced into tobacco plants and the TMV is the ideal 'launch vector' to do this. Viruses can only reproduce inside living cells because they do not possess the molecular machinery to do so independently. So as the virus infects a tobacco plant, its genetic material is incorporated into that of the plant cells (Figure **B.12**), which then produce both the normal viral proteins (to make new viral units), and the proteins coded for by the introduced genetic material.

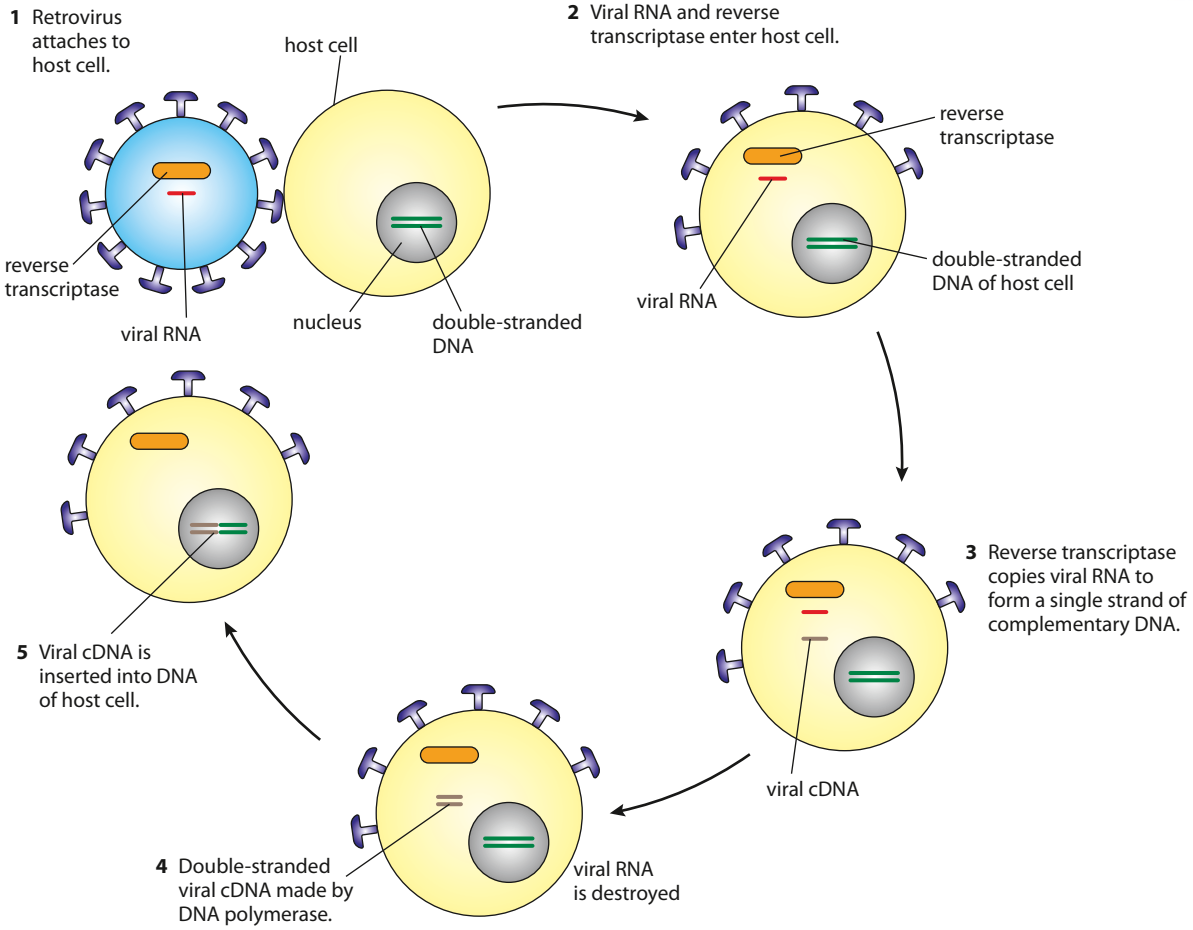


Figure B.12 How the genetic information from a retrovirus becomes part of the DNA of a host cell.

The proteins can then be harvested from the plants to make vaccines. Plants have the potential to make vaccines against any virus in this way, producing high yields of the precise protein that is needed.

The plants can be grown in **hydroponic culture** with carefully controlled conditions to maximise their growth. Although the viruses can infect tobacco plants naturally (through damaged leaves, for example), virus **vectors** are often introduced into mature plants artificially in order to speed up the production process. For example, in some production processes, tobacco plants are grown for 4 weeks before the virus vector is introduced by means of a new technique known as vacuum infiltration. To do this, the plants are turned upside down, and submerged in water containing TMV vectors. A vacuum is then created, drawing air away from the water and plants. As the vacuum is switched off, the water containing the virus vectors is 'sucked' into the plants. The plants are left to grow and in about 7 days they will have produced the target proteins in their leaves and stalks. The plants are harvested, the leaves are cut into small pieces and liquefied, and the proteins are extracted from the liquid.

In the past, vaccines against viral diseases such as influenza and hepatitis have been grown in birds' eggs. But tobacco-based plant vaccines have several advantages:

- vaccines can be produced more quickly
- there are no known allergies to plant-produced vaccines, unlike those from birds' eggs
- large tobacco leaves make harvesting easy
- many doses of vaccine can be made from a relatively low biomass.

Reverse transcriptase and its use in molecular biology

The enzyme reverse transcriptase was discovered in 1970 in a group of viruses known as retroviruses, which includes HIV, TMV and feline leukaemia virus. These viruses contain RNA as their genetic material. Reverse transcriptase enters along with the viral RNA when a retrovirus invades a host cell. It is used to transcribe the virus RNA into a single strand of DNA, using nucleotides from the host cell. The new complementary DNA (cDNA) is then converted to double-stranded DNA using the enzyme DNA polymerase. The original RNA is degraded and the double-stranded DNA is inserted into the host's chromosomes (Figure **B.12**).

Reverse transcriptase is widely used in genetic engineering. Molecular biologists are able to produce therapeutic proteins such as insulin and growth hormone by inserting the genes that code for them into the genetic material of bacteria or other organisms. These organisms then produce large amounts of the required protein as they grow and multiply.

Paper production and potatoes

Potato tubers store starch which, as well as being a good source of food, has a wide variety of uses in the manufacture of textiles, paper and adhesives. Potatoes (*Solanum tuberosum*) produce large amounts of starch and give higher yields per hectare than either corn or wheat. Potato starch consists of approximately 80% amylopectin and 20% amylose (Subtopic **2.3**) but most of the useful properties for industry, such as adhesion, come from amylopectin. In conventional manufacturing processes amylose is not needed and causes production problems because it forms a gel, which makes potato starch unstable.

Recently, researchers at BASF Plant Science succeeded in deactivating the genes for amylose in the potato genome and developed a new variety of genetically modified potato known as Amflora. The Amflora potato produces only amylopectin and so it is ideal for use in the paper-making and the adhesive industries. It enables concrete to stick to walls more effectively, keeps glue liquid for longer and improves the 'glossiness' of paper (Figure **B.13**).

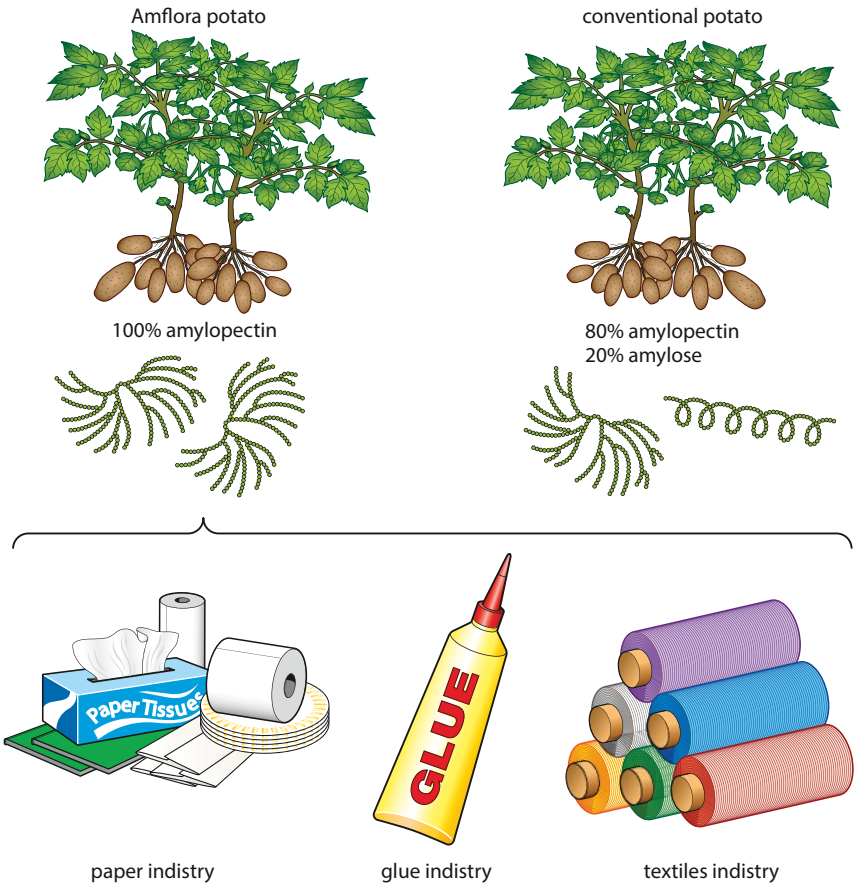


Figure B.13 Amflora potatoes and their uses.

Politics and GM

The modified potato received approval from the European Union in 2010 after an extensive period of review and testing. Amflora potatoes for planting cannot be sold through normal channels but are supplied directly to farmers who have contracts with BASF and who plant the Amflora potatoes at a fixed distance from other potato crops. When harvested, Amflora are taken to separate processing factories. Despite these precautions, and the fact that the European Food Safety Authority (EFSA) has stated that there are no risks to humans or animals from Amflora, the potatoes struggled to gain acceptance – indeed, by 2011, they were only being grown on one small site in Germany. This lead to the company, BASF, deciding to move its biotechnology headquarters to the USA – it has stopped production of genetically modified (GM) products in Europe, citing lack of political support for its work as the main reason. A strain of GM maize, developed by another biotechnology company, Monsanto, is now the only crop approved for use in the EU, apart from the Amflora potato. Other GM crops including maize, rice and tomatoes are widely grown in South America, the USA and Asia.

Support for science

Following the withdrawal of BASF from biotechnology production in Europe, a BBC report said '*No one from the political side supported it. There were no signals from the European Commission that any change was likely*.' Some reports indicated that a new EU health commissioner might freeze approval on new GM crops, while others said that he was trying to clarify the regulations on growing GM crops in Europe.

Question to consider

- To what extent should political and public pressure influence the advance and development of new scientific discoveries?

How recombinant DNA is transferred into plants

In gene technology, manipulated DNA containing genes for the protein or proteins of interest is called **recombinant DNA** (Subtopic **3.5**). The first stage in any gene technology process, such as that used for 'molecular farming' in tobacco plants described above, is to find and isolate the **target genes** for the protein to be produced. These genes are also linked to other genes, which may control the way they are expressed. **Bioinformatics**, which is the use of databases and accessible stores of information, has been crucial in the easy identification of DNA sequences coding for these desired proteins. The most useful sequences are known as 'open reading frames' – these, combined with **marker genes**, can be inserted into target organisms. The newly created transgenic organisms can then produce the novel products.

Open reading frames

An **open reading frame** (ORF) is a sequence of nucleotides in a DNA molecule that has no codons that terminate transcription within it. The site that terminates transcription is found after the ORF and beyond the translation stop codon. This means that, once transcribed and translated, the ORF has the potential to produce a complete polypeptide chain.

ORFs are very useful in helping to predict which sections of a DNA molecule are likely to be genes. Long ORFs can be used to identify regions of DNA likely to code for proteins. Researchers search for a start codon followed by an ORF that is long enough to code for a typical protein in the particular organism they are investigating. But even if ORFs are found, this does not necessarily mean that these regions of DNA are genes that are transcribed and translated. To check for an ORF's potential to be transcribed and translated, an Open Reading Frame Finder (ORF Finder) can be used. This is a graphical analysis tool used by researchers, which finds all open reading frames of a certain, chosen size in the DNA sample being studied. The ORFs may either be in a new sequence that is being investigated or in a sequence that has already been stored in a database. The Finder tool can deduce the amino acid sequences coded by lengths of DNA and these can be compared with known amino acid sequences using the BLAST server. BLASTn is a database that can identify similar nucleotide sequences in different organisms while BLASTp can match amino acid sequences in proteins. This is described in more detail in Subtopic **B.5**.

Introducing genes into plants

In order to modify plants, new DNA sequences must be introduced into their cells and taken up either by the plant's chromosomes or by the chloroplast DNA. Recombinant DNA can be introduced into whole plants, but is often introduced into leaf discs kept in tissue culture or into protoplasts (cells that have had their cell walls removed). When DNA has been taken up by plants, leaf discs or protoplasts they can be grown up to make new transgenic plants. As well as viral vectors, described above, several other methods are used to introduce recombinant DNA into plant cells.

Chemical methods

Sequences of DNA or bacterial plasmids can be encapsulated into tiny, neutral structures known as **liposomes**. Liposomes are vesicles, which have a phospholipid bilayer around them. Encapsulation is achieved by adding a mixture of ethanol and calcium chloride to a mixture of vesicles and plasmids. Liposomes provide an efficient method of transferring DNA to new cells. Cells are 'transfected' by the liposomes, which can actually fuse with the cell membrane of a target cell and release the DNA or plasmid directly into it (Figure **B.14**). In some of the transfected cells the DNA will be taken up by the cell's genome.

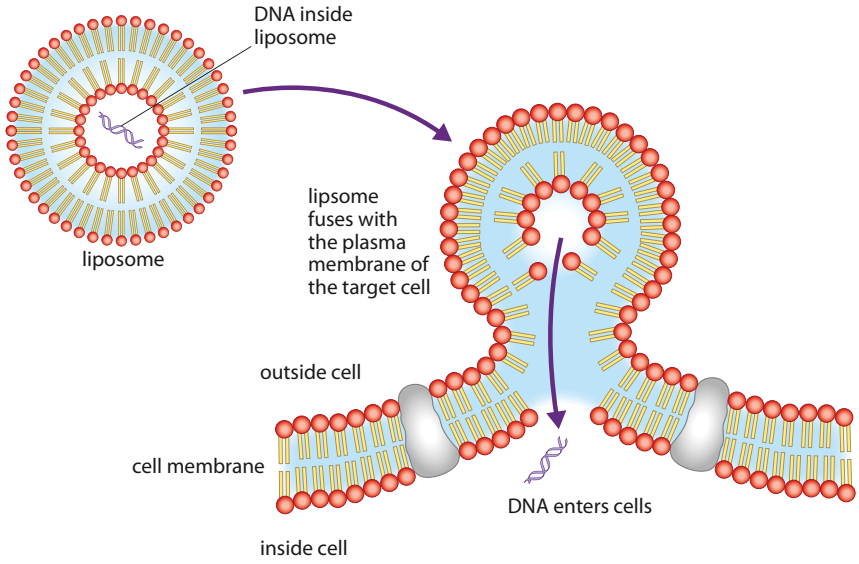


Figure B.14 Liposome structure and transinfection.

An alternative, but less efficient, method uses calcium phosphate and calcium chloride solutions. A solution containing phosphate ions is combined with calcium chloride solution containing the DNA to be transferred. As the two solutions are mixed, a precipitate of positively charged calcium and negatively charged phosphate forms. The DNA that is to be transferred binds to its surface. The transfer is completed by adding a suspension of the precipitate to plant cells grown in a single layer in tissue culture. The cells take some of the precipitate and DNA into their cells. The exact mechanisms of this process are not fully understood.

Physical methods

Three important physical methods of introducing genes into plant cells are electroporation, microinjection and biolistics (the so-called 'gunshot' method).

- **Electroporation** exposes cells to a strong electric field for a very short period of time. This has the effect of changing the permeability of cell membranes for sufficient time for new DNA to enter the cells.
- **Microinjection** involves using a micropipette to inject DNA into a living cell. This is usually carried out under a microscope, and is the same technique as that used to transfer material into animal eggs during *in vitro* fertilisation. New variations of this technique have been developed to bind DNA to large number of nanofibres, which can be inserted into cells or tissues in an automated process.

- **Biolistic** methods involve a 'gene gun', which shoots DNA attached to a nanoparticle of an unreactive material, often gold, directly into the nucleus of the target cell (Figure **B.15**).

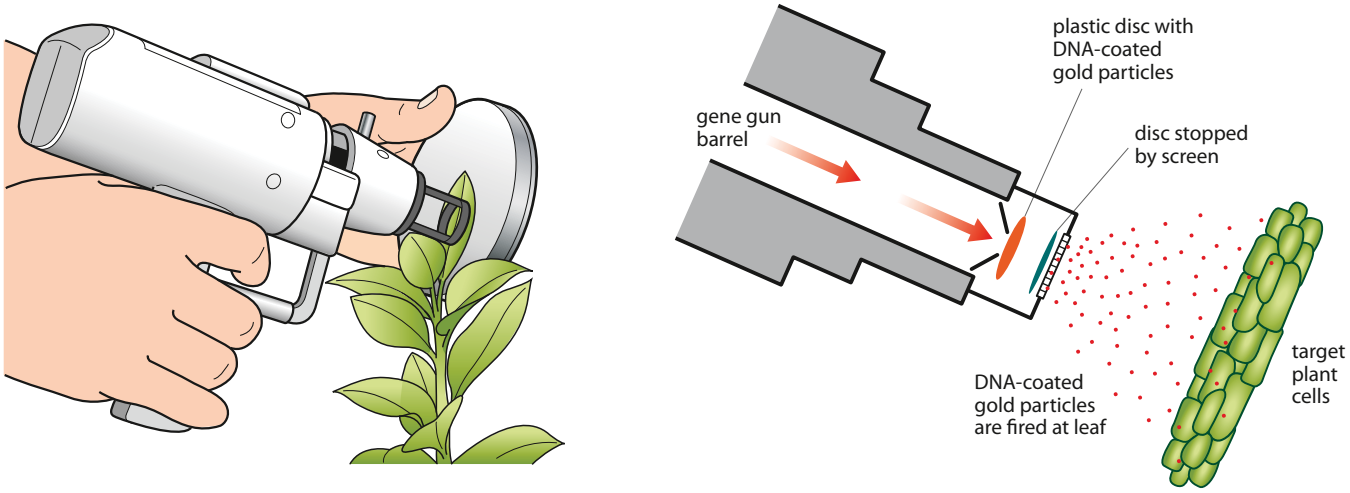


Figure B.15 A gene gun and its method of working.

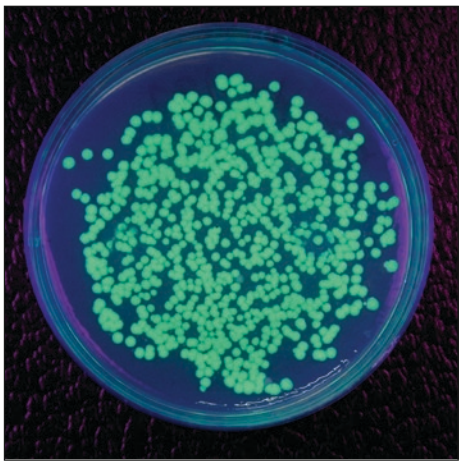


Figure B.16 Transformed bacteria colonies containing a jellyfish gene for GFP protein causing green bioluminescence. This gene can be used as a marker to indicate the successful uptake of a target gene.

Exam tip
Draw a flow diagram to summarise the ways in which DNA is introduced into one organism from another.

Marker genes

Marker genes are linked to sequences of DNA that are being transferred to new organisms so that researchers can check that the insertion has been successful. There are two types of marker genes: screening markers and selectable markers. **Screening markers** identify cells that have taken up new genes by their appearance. The three most common screening markers are:

- a green fluorescent protein (GFP), which makes cells glow under UV light (Figure **B.16**)
- the so-called GUS assay, which stains cells blue and is regularly used in experimental plant science but has limited use in applied work because it kills cells
- the blue–white method, used as both a plant and bacterial marker, which involves adding a bacterial gene coding for beta galactosidase enzyme so that if galactosides are added to a culture medium, cells containing the gene convert it to a blue substance, which can be seen easily.

Selectable markers (usually antibiotic resistance markers or ARMs) protect the cells of organisms that have taken up new DNA from substances that would otherwise kill them. Because very few cells take up new DNA, in most cases it is easier to kill those that have not and use the selectable marker to protect the remainder. Antibiotic markers are often used to kill chloroplasts during the genetic modification of plant cells. Any plant cells that have not taken up the new DNA after they have been treated, and therefore do not carry the resistance marker, can be targeted with suitable antibiotics, which will destroy them.

Nature of science

Assessing risk in science – concerns about genetic modification

Genetic modification of plants and animals has the potential to be very helpful to the human race, but risks and benefits must be evaluated. One of the important risks is the potential of herbicide resistance genes, such as the genes for glyphosate resistance in GM soybeans, escaping into wild populations of plants. If wild species should become resistant to herbicides it might be less easy to clear fields of weeds with the herbicides that we have today. But there are many sides to the discussion – some of the key points to consider are as follows.

- Modifying crop plants and animals to increase yields will provide more food for the growing human population. Plants can also be made tolerant to drought or salt water so that food can be grown in difficult areas.
- Crop plants that are disease resistant not only increase yields but also reduce the need to apply harmful pesticides, which have impacts on the wider ecosystem.
- Many substances – such as human growth hormone, a blood-clotting factor, antibodies, and vitamins – are already being made by genetically modified organisms to improve human health.

On the other hand:

- There is concern that there may be dangers in consuming products from genetically modified plants and animals.
- The long-term effects of genetically modified crops in the environment are not known. Plants or animals could 'escape' into the environment and their genes might become incorporated into wild populations, with unforeseeable effects.
- Human food crops could become controlled by a small number of biotechnology companies, which could make seeds and plants more expensive.
- More genetically modified organisms might lead to a reduction in natural biodiversity.

You can read how pollen from GM maize was alleged to have affected monarch butterflies in Subtopic **3.5**.

Test yourself

4 Define the term 'transgenic organism'.

5 List **three** methods of introducing new DNA into plant cells.

6 Outline the meaning of the term 'open reading frame'.

7 State the key difference between the genetically modified Amflora potato and a normal potato.

B3 Environmental protection

Learning objectives

You should understand that:

- Bioremediation together with physical and chemical procedures can be used in response to pollution incidents.
- Microorganisms are used in bioremediation.
- Microorganisms can metabolise some pollutants.
- Biofilms are cooperative aggregates of microorganisms.
- Biofilms show emergent properties.
- Microorganisms growing in biofilms are highly resistant to antimicrobial agents.
- In biofilms, microorganisms are able to cooperate through quorum sensing.
- Bacteriophages are used to disinfect water systems.

Exam tip

If you are asked to write an essay on bioremediation of an oil spill, make a checklist of all the different things that would need to be done. Don't forget physical and chemical cleaning.

Bioremediation

Bioremediation is the process that uses microbes to treat areas of land or sea that have been contaminated by pesticides, oil or solvents. Bioremediation is usually carried out in conjunction with physical and chemical procedures, such as clearing away contaminated soil or scraping away oil, and treatment with detergents sprays or remedial chemicals to neutralise contamination.

Cleaning up pollution at sea

Crude oil spills from tankers at sea contain many chemicals that harm the marine environment and seashore. Many different microbes are able to oxidise harmful hydrocarbons and break down the oil. To increase the numbers of bacteria and speed up the bioremediation process, nitrate and phosphate fertilisers are sometimes added, to encourage decomposition of the crude oil by the bacteria. Bioremediation like this can halve the time it takes to clean up an oil spill. A number of strains and species of bacteria in the genus *Marinobacter* have been found to be important degraders of hydrocarbons in the sea. These bacteria live in habitats that sometimes have extremes of pH or salt concentrations, and because they are **halophilic** (salt loving) they tolerate marine conditions well. *Marinobacter hydrocarbonoclastus*, for example, is able to degrade petroleum hydrocarbons, such as benzene, which are released during accidental spillage of oil at sea. In 2011, it was discovered that *M. hydrocarbonoclastus* is inhibited by the presence of certain chemicals used as dispersants in cleaning up oil spills, suggesting that dispersants should not be used at the same time as this bacterium in bioremediation treatments (Figure **B.17**).

Figure B.17 Workers using hoses and dispersants to clean up an oil spill. Bacteria may be able to do the job more efficiently in future. Experts from different parts of the world are brought together to deal with serious oil spills. Each one may have experience in a different aspect of the clean up.

Cleaning up pollution on land

On land, bioremediation has been used to remove pesticide residues, crude oil spills, heavy metals and solvents from contaminated soil. There are approximately 20 species of bacteria with the ability to metabolise petroleum hydrocarbons found in oil, which include members of the genera *Pseudomonas*, *Aeromonas*, *Bacillus*, *Flavobacterium* and *Micrococcus*. One of the most efficient hydrocarbon users on land is *Pseudomonas aeruginosa*, which can use crude oil as its sole source of carbon. It can also tolerate and thrive in high concentrations of oil, and if surfactants are used in combination with the bacteria they can degrade the oil more quickly. Bioremediation often happens very slowly when bacteria in the environment break down other toxic substances. To speed up the process, inorganic fertilisers added to the area increase the supply of nutrients and enhance the work of the bacteria. Spreading out polluted soils can also stimulate faster growth of bacteria.

Several species of bacteria have been found to contain the enzyme nitro-reductase, which gives them the ability to break down explosives such as TNT (2,4,6-trinitrotoluene). These species can be very helpful in cleaning up land contaminated with such substances. Over the last century, large quantities of explosives have been manufactured for military and industrial use. Many are highly resistant to biodegradation and large areas of land are contaminated with residues from their manufacture and storage. Incineration of affected soil is very expensive, so bioremediation is an attractive alternative.

In Australia, two species of bacteria, *Pseudomonas* sp. and *Azospirillum* sp., have been isolated from contaminated soils around disused sheep dips. These bacteria have been shown to be capable of breaking down organophosphate pesticides, which remain in the soil and pose a significant threat to the environment and public health. The bacteria are being used to develop a more general bioremediation strategy for the removal of the pesticides.

Removing methyl mercury

Mercury is a very toxic substance and, although all forms of it are harmful, major public health concerns are centred on methyl mercury, a neurotoxin that causes problems ranging from mild numbness to blindness, loss of balance, and in severe cases, death. Methyl mercury can enter the human body through the consumption of contaminated fish. Mercury compounds in sea water accumulate in the bodies of marine organisms, becoming more concentrated as they pass through food chains by a process called **biomagnification**, so that levels in larger fish taken for human food can be highly toxic. Methyl mercury was the compound responsible for the harm caused to residents of Minamata Bay in Japan in the 1950s following industrial pollution of the sea and the shellfish they used as food.

Bioremediation can be used to clean up mercury-contaminated waste water from processes such as gold mining, manufacture of plastics and the processing of fossil fuels, thus preventing methyl mercury from entering the food chain. Methyl mercury can be demethylated by bacteria such as *Pseudomonas balearica*, which carry mercury-resistance genes, and can decompose methyl mercury to produce inorganic mercury. Mercury is a liquid at room temperature and will vaporise on contact with air, but it is almost insoluble in water. If a community of mercury-resistant bacteria such as *Pseudomonas balearica* is established and held under a layer of water

in a bioreactor packed with inert material such as rubber fragments, mercury will precipitate and can be collected and removed.

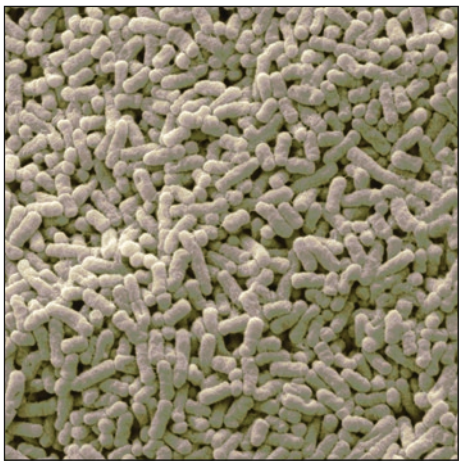


Figure B.18 Biofilm of bacteria forming on a tooth.

Emergent properties

An **aggregate** is a group of bacteria living together that has characteristics not displayed by the individual bacteria. For example, the marine bacterium *Vibrio fischeri*, which lives in the epidermis of sea anemones, does not emit light on its own, but in an aggregate the bacteria become bioluminescent. Other species produce bioluminescence in the light organs of the squid, when they are assembled in large enough groups. These aggregates, and other biofilms that develop different properties from the individual organisms, are said to have **emergent properties** – properties arising as a result of the aggregation.

Bacterial aggregates and biofilms

Most bacteria live as single cells but some (such as *Streptococcus mutans*, which occurs widely in the mouth) form **aggregates** or groups of cells that are connected together. *S. mutans* forms a layer known as a **biofilm** on teeth, at the junction with the gums (Figure **B.18**). The bacteria convert sucrose to a glue-like 'extracellular polysaccharide substance' (EPS), which allows the bacteria to stick to each other and form plaque. Together with the acid produced by bacterial metabolism, plaque leads to tooth decay.

Dental plaque is one example of a biofilm but there are many others, including the persistent slime that forms in a bathroom drain and the coating on the surface of submerged rock, which makes it slippery and dangerous. Some biofilms are useful (one example is their use in bioremediation in sewage treatment) but biofilms can cause major problems. If they form on food-production surfaces they can be a health hazard, and in other circumstances biofilms can cause clogging and corrosion in pipes. In each case, the aggregations are groups of microorganisms living in a matrix of adhesive EPS, which is used to incorporate new cells created by reproduction as well as other cells from the outside.

Biofilms start to form when individual microorganisms attach to a damp surface, which may be either living or non living. The reasons why a biofilm starts to form and attach to particular surface are not fully understood but the rate of water flow across the surface, nutrient availability and debris already present may be key factors. Others include the temperature and pH of the environment, and metabolic interactions between the cells. Keeping a surface perfectly clean is one way to prevent a biofilm forming, so hygiene and cleanliness in hospitals and kitchens are vitally important.

Biofilms are very flexible and are described as being **viscoelastic**, which means they can stretch and change their shape as a flow of liquid pulls or pushes them. Because of this, biofilms cannot be removed by rinsing them away – a flow of liquid over a biofilm may at best cause clumps to disconnect and fall away to settle elsewhere. This causes serious problems in hospitals and industry if biofilms attach to medical equipment or pipes. Some biofilms can be scraped off with gentle scrubbing but others are resistant even to pressure hoses. Biofilms that grow in areas where water flow is strongest tend to modify their growth to become thinner but more firmly attached.

Serious difficulties occur when biofilms become established in areas that are too difficult to reach or too delicate to treat. In hospitals they may form in catheters, nasal tubes or on heart valves. Some medical equipment can be replaced but biofilms that grow on implants can cause life-threatening infections. Biofilms protect the organisms within them from the human immune system and make them highly resistant to treatment with antibiotics and other antimicrobial agents.

Research efforts are being focussed on how to break up different biofilm colonies and make them detach from their surfaces, and on methods to weaken the EPS matrix so that antimicrobial agents can be used to kill the cells. Keeping surfaces clean is the best method of preventing a biofilm from forming in the first place.

Quorum sensing

Microorganisms living close to one another within a biofilm communicate using **quorum sensing**, by releasing signalling molecules into the environment. The accumulation of signalling molecules enables a single cell to detect the density of other cells in the area, and allows cells to coordinate their behaviour. If an environmental factor, such as the availability of nutrients changes, they can respond quickly in order to survive. They are able to defend themselves against competitors that share the same food source and avoid toxic compounds such as antibacterial agents. Quorum sensing systems exist in biofilms, as well as between marine and many pathogenic bacteria, which are known to coordinate their virulence in order to escape the immune system of their host. It may be that quorum sensing also establishes boundaries between different biofilms, and enables cells to reproduce or transfer genetic material between individuals.

About 60% of all hospital infections may be due to biofilms that have formed on medical devices. Catheters, lines and drips are all prone to biofilm infection and on implanted devices, such as valves and replacement joints, biofilms are very resistant to the human immune system and to antimicrobial agents.

Using biofilms in sewage treatment

Biofilms are very useful in sewage treatment. Their tenacity and resilience help them survive in the difficult conditions of a sewage treatment plant. The first stage in the treatment of raw sewage is to pass the material through coarse filters, which remove grit and other items in the sewage (Figure **B.19**). After this, the waste is left to settle in sedimentation tanks so that sludge can be taken away and the remaining liquid effluent given a secondary treatment.

Older methods of secondary treatment use trickle filter beds (Figure **B.20**), which contain biofilms of bacteria and fungi growing on the surface of a porous material known as clinker or on plastic piping, which can be up to 2 m thick. As water trickles through the biofilm, organic matter in it is broken down by the microorganisms in the biofilm, which use it as a source of nutrients. Oxygen levels are kept high by the movement of the trickling water and small invertebrates such as worms and protozoans grow over the biofilm, feeding on it so that it does not grow too thick and impede the filtering process.

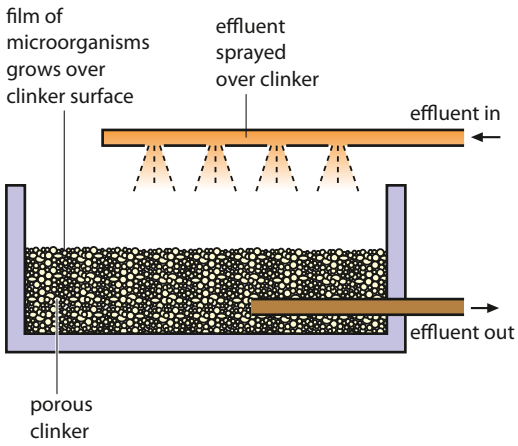


Figure B.20 Trickle bed filter.

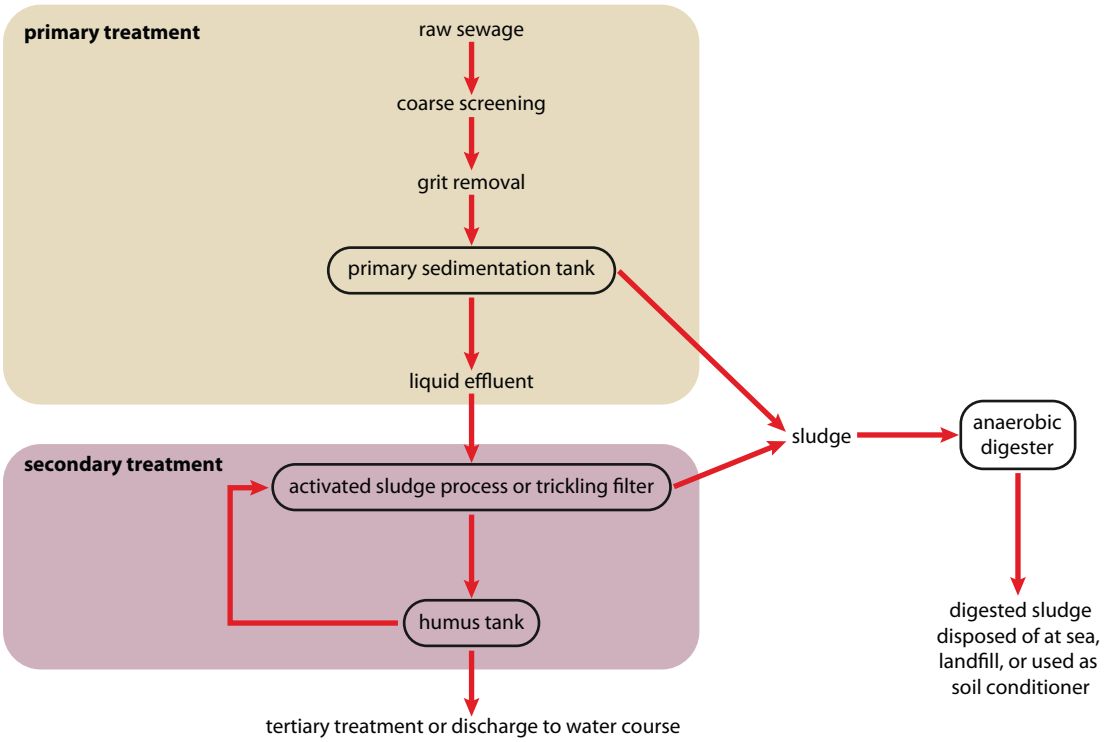


Figure B. 19 Sewage treatment.

Using bacteriophages in disinfection of water systems

Bacteriophages are viruses that infect and kill bacteria (Figure **B.21**) but which have no effect on plant or animal cells. Today, as antibiotic resistance is on the increase (Subtopic **5.2**) and it is becoming more and more difficult to kill bacteria, scientists are examining possible roles for bacteriophages as antibacterial agents.

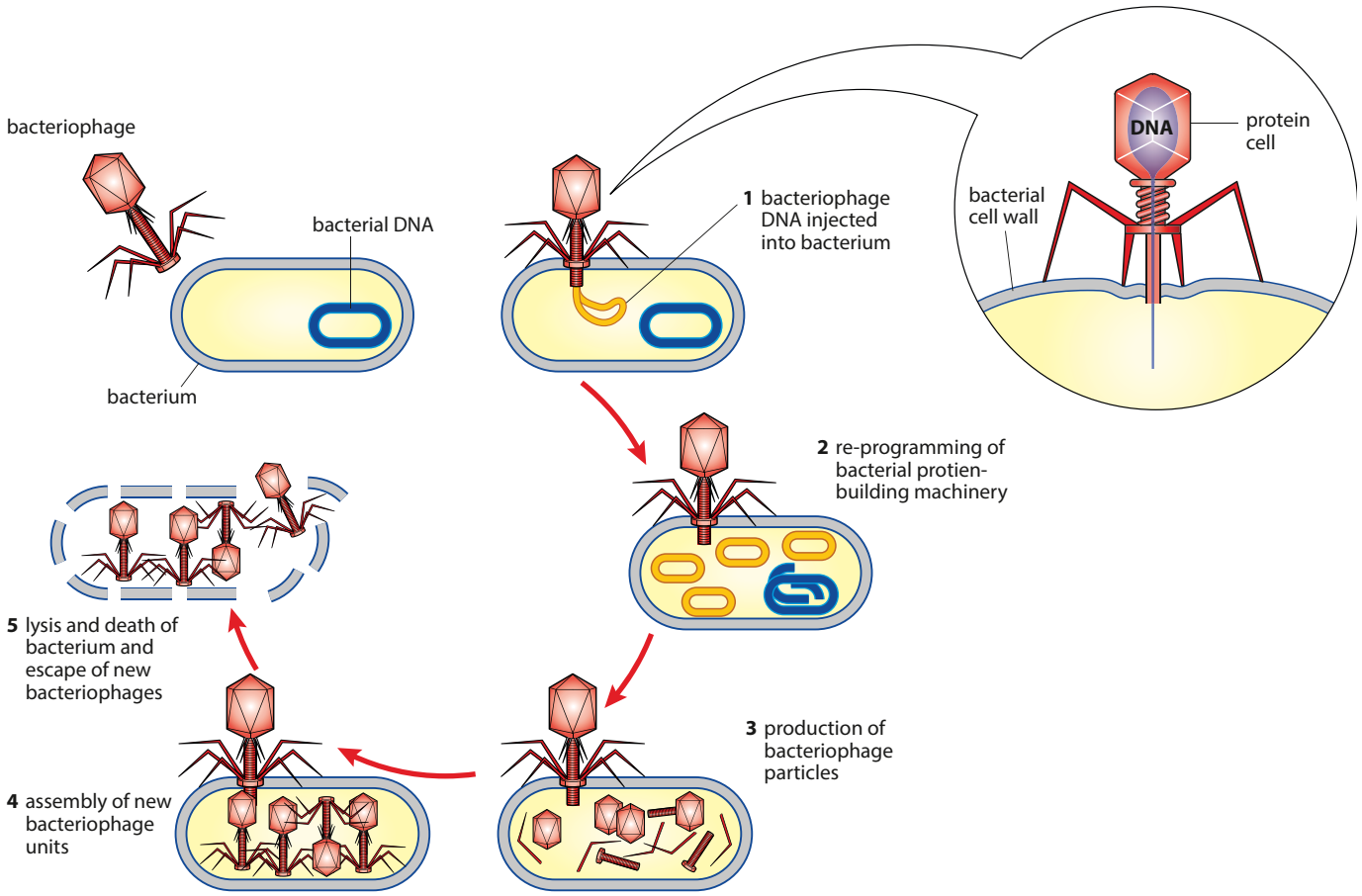


Figure B.21 How bacteriophages kill bacteria.

Water treatment is one area where exciting developments are taking place. Bacteriophages have been successfully used to kill the filamentous bacteria (*Sphaerotilus natans*) that grow in long threads during the settlement stage of sewage treatment. The bacteria interfere with sewage treatment by preventing sludge settling in the settlement tanks so that the remaining liquid cannot be led off for secondary treatment. Disinfectants such as chlorine can be used to kill the bacteria but if bacteriophages are employed, chlorine is no longer needed. After just a day the bacteriophages can remove enough bacteria to improve settlement rates and they have been found to remain active for many months.

Bacteriophages have also been used to remove biofilms of *Pseudomonas aeruginosa* bacteria, which can clog filters at water purification plants. Usually these bacteria must be cleaned away with chlorine and expensive flushing treatments. Studies show that bacteriophages can kill almost 90% of these biofilms and treatment is almost 100% effective if the bacteriophages are followed by a single treatment with chlorine.

Nature of science

Scientific advance follows technical innovation – the laser-scanning microscope

Advances in scanning microscopes have enabled scientists not only to examine the outer structure of bacteria and biofilms but also to build up 3D reconstructions of the complex interactions between individual cells in an aggregate. The **confocal laser-scanning microscope** is able to form images from selected depths within a biofilm in a process known as optical sectioning. The images obtained are built up into a 3D picture using computers. This has enabled researchers to develop a much clearer understanding not only of the interactions between individual bacteria but also the structure of their surrounding matrix. Without such new technology this would not have been possible. This type of microscopy also has uses in medicine and has helped in the diagnosis of keratomycosis, a fungal infection of the cornea. With laser images from the microscope, early diagnosis and treatment have been improved.

Test yourself

8 Outline one physical and one biological strategy used in a response to an oil spill.

9 List **three** properties of biofilms.

10 Outline what is meant by the term 'quorum sensing'.

11 State how bacteriophages can be used in water treatment.

B4 Medicine (HL)

Learning objectives

You should understand that:

- Genetic material or antigens from a pathogen can be used to detect infection by the pathogen.
- Markers can be used to detect predisposition to a genetic disease.
- DNA microarrays can be used to diagnose a disease or test for genetic predisposition to it.
- Metabolites indicating disease can be detected in urine or blood.
- Tracking experiments are used to gather information about the localisation and interaction of a desired protein.
- Genetically modified animals and plants are used in biopharming to produce proteins for therapeutic use.
- Viruses can be used as vectors in gene therapy.

Detecting infection using antigens or genetic material

A doctor may use a blood or urine test in an attempt to diagnose an infection. Antigens or genetic material from pathogens that are present in clinical specimens can be identified and used to pinpoint the cause of infection. Early diagnosis can then lead to faster treatment.

The three most important techniques used are:

- detection of pathogen-specific antibodies in the blood
- detection of the pathogen's antigens
- detection of genetic material from the pathogen.

Detection of antibodies

During an infection, **antibodies** bind to pathogens to disable them and prevent further infection of cells (Subtopic **11.1**). Two antibodies, known as immunoglobulins (IgM and IgG), are always produced as the immune system responds to an infection, and will be present in the blood. IgM is expressed on the surfaces of B cells and kills pathogens in the early stages of infection. These molecules are only produced for a few weeks, so detection of IgM in the blood tells a doctor that an infection is present at the time a blood test is taken. IgG is produced indefinitely by B cells and is used by the immune system to identify and neutralise bacteria and viruses. Presence of IgG provides evidence of a past infection.

Detection of antigens

Antigens (proteins derived from a pathogen) can be detected in blood samples using the ELISA test (Enzyme-Linked ImmunoSorbent Assay), which is one the most commonly used diagnostic tests (Figure **B.22**). It was developed for the detection of HIV but is now used to detect many other pathogens. The ELISA test is highly sensitive and very specific. It has the advantage that it can be used to screen large numbers of specimens at a time, making it invaluable for public health screening and in the screening of blood for transfusions.

ELISA is now a well-established tool for the rapid diagnosis of a wide variety of infectious diseases, helping doctors to choose the appropriate method to treat the infection at an early stage when antibiotics are most effective. Early diagnosis and targeted use of antibiotics may also reduce the development of antibiotic resistance.

The test uses an antibody that is specific to a particular antigen from a known pathogen, and a colour change shows whether or not antigens are present in a sample. The antigens chosen for use in the test are those that appear in blood or urine during an infection – for example, polysaccharides from viral capsules or proteins produced by bacteria.

The procedure involves:

1. immobilising a sample containing an unknown amount of the suspected antigen, usually by attaching it to a well in a plastic microtitre plate
2. attaching the antibodies being used to detect the pathogen to an enzyme and then adding this enzyme–antibody molecule to the plate so that it can form a complex with any antigens present

3 rinsing away any unbound material so that only enzyme–antibody molecules that have formed complexes with antigens remain
4 finally, adding a substrate for the enzyme to the surface of the plate – so that as the substrate binds with enzyme molecules attached to antibody–antigen complexes, a colour change occurs (Figure **B.22**). The degree of change in colour indicates the amount of antigen in the sample. Quantitative results can be obtained by measuring the optical density – the greater the colour change, the greater the quantity of antigen present.

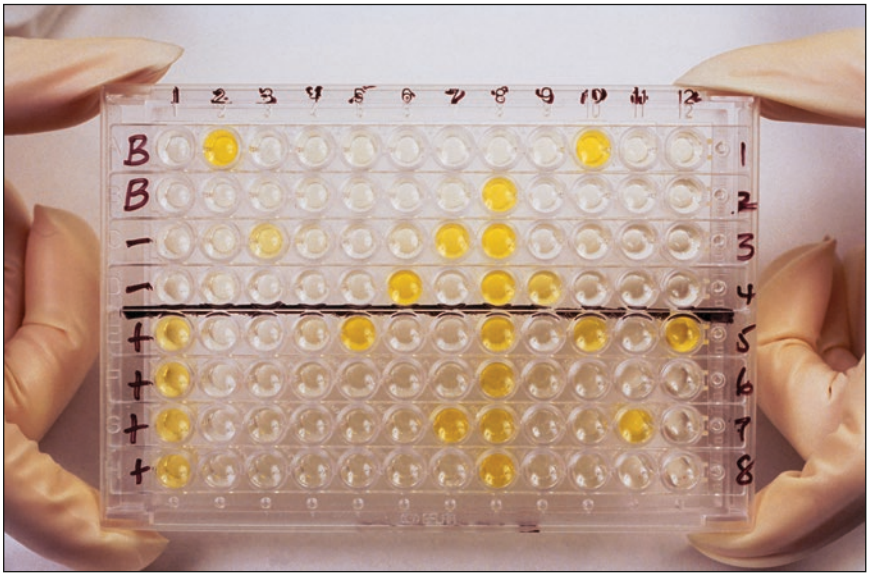


Figure B.22 The ELISA test can be carried out on many specimens at once. Each well in the plate is used for a separate patient's specimen. The yellow colour on this plate indicates a positive result – antigen from the pathogen is present, which shows that the patient is infected with this pathogen.

Detection of genetic material

Genetic material from pathogens such as influenza viruses can be detected using a rapid, sensitive diagnostic technique called the **reverse transcriptase polymerase chain reaction** (RT-PCR). This method uses genetic material from retroviruses such as influenza (which contain only RNA) as a template for reverse transcription. Complementary DNA is produced from the virus RNA. Alternatively, for other pathogens, DNA samples can be used directly. The DNA is amplified using the **polymerase chain reaction** (PCR) (Subtopic **3.5**) and detected either by gel electrophoresis or by using hybridisation with known sequences of DNA to assess the degree of matching and so identify the pathogen.

The sections of genetic material from the pathogen that are selected for amplification in the PCR are carefully chosen. The sections are removed from the genome using restriction endonucleases and by using different endonucleases different regions can be targeted. When trying to identify of bird flu viruses, for example, hemagglutinin (HA) and neuraminidase (NA) genes are targeted because these are unique to strains of bird flu and cannot be confused with material from human flu viruses. By using two different genes in the test a clinician can be more certain that positive results are reliable.

Following the amplification of genetic material, the DNA fragments are separated on agarose gels using electrophoresis. Gel electrophoresis separates the fragments by their size. All DNA fragments are negatively charged but when the electric current is switched on, smaller fragments will move further through the gel than larger ones, which cannot pass easily through it. DNA samples are mixed with a 'loading' dye, which makes them visible and the power supply is turned on. The samples and their dye markers move through the gel and the different molecules separate, within their different lanes. Then the current is switched off and the gel removed. DNA molecules are made visible by staining the gel with ethidium bromide, which binds to them. Ethidium bromide fluoresces under ultraviolet light so the series of bands spread through the gel can be seen if it is illuminated by a UV lamp (Figure **B.23**). Different strains of influenza viruses can be identified from profiles like these, because each genome produces a different characteristic range of DNA fragments when treated with endonucleases.

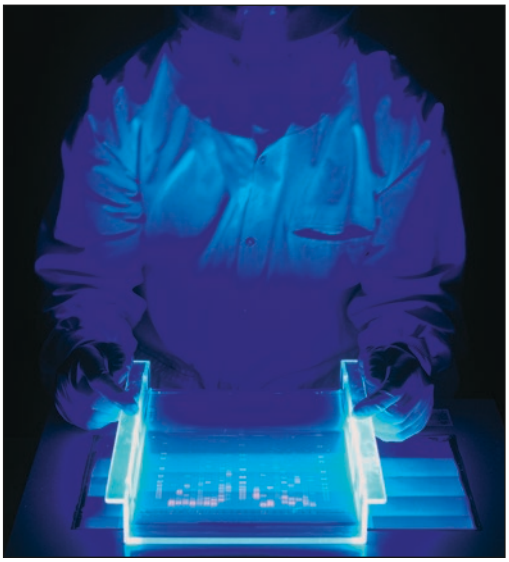


Figure B.23 Agarose gel stained with ethidium bromide. Samples of genetic material from different pathogens produce different banding patterns following electrophoresis, which can be used to identify them.

Genetic predisposition to disease, and diagnosis

Diseases can involve complex interactions between many genes and many are also closely linked to environmental factors. An individual may not be born with a disease but may be at high risk of acquiring it. This is known as **genetic predisposition** or susceptibility. A genetic susceptibility to a particular disease can be identified, in some cases, by **markers** present in a person's body. Recent research has moved the detection of these markers forward so that by understanding their genetic predisposition to disease individuals can make informed choices about lifestyle or medical treatments that could reduce their probability of developing a disease. This might include avoiding certain activities such as smoking, or taking more exercise.

Testing to assess a person's predisposition, or to diagnose the disease itself, is described below for three important diseases – breast cancer, Alzheimer's disease and prostate cancer.

In medical science, a **marker**, or **biomarker**, is any measurable characteristic that indicates the presence or severity of a disease, or a person's susceptibility to it. Biomarkers are often detected in blood or tissue, and may be proteins – such as specific gene products, enzymes or hormones – or other molecules, genes, or cells. High body temperature is a biomarker for fever, while blood pressure is a well-known indicator of the risk of stroke.

Breast cancer

Inherited mutations of two genes known as the *BRCA* genes on chromosomes 13 and 17, account for a small proportion of all breast cancers, but women who have these mutations have a substantial risk, greater than 70%, of developing breast cancer or ovarian cancer. Identification of people who carry the mutations, by genetic testing, allows for preventive measures (including mastectomy), clinical treatment and counselling. *BRCA* analysis is a genetic test that requires only a blood sample to determine whether a patient carries a faulty *BRCA1* or *BRCA2* gene. Genes may be detected using DNA microarrays, which assess the match between genes from the person being tested to known *BRCA* genes.

A **DNA microarray** is a 'biochip' that has thousands of microscopic DNA samples attached to its surface (Figure **B.24**). Each spot contains a specific DNA sequence, known as a **probe**, that is used to hybridise samples of **target DNA**. Hybridisation occurs when complementary nucleotide sequences in probe and target DNA molecules pair up by forming hydrogen bonds. The more complementary sequences present, the greater and stronger the bonding between the two strands. The microarray is washed after hybridisation and weakly or non-bonded strands are flushed away. Hybridisation is usually detected and quantified using luminescent labels attached to target DNA (Figure **B.25**).

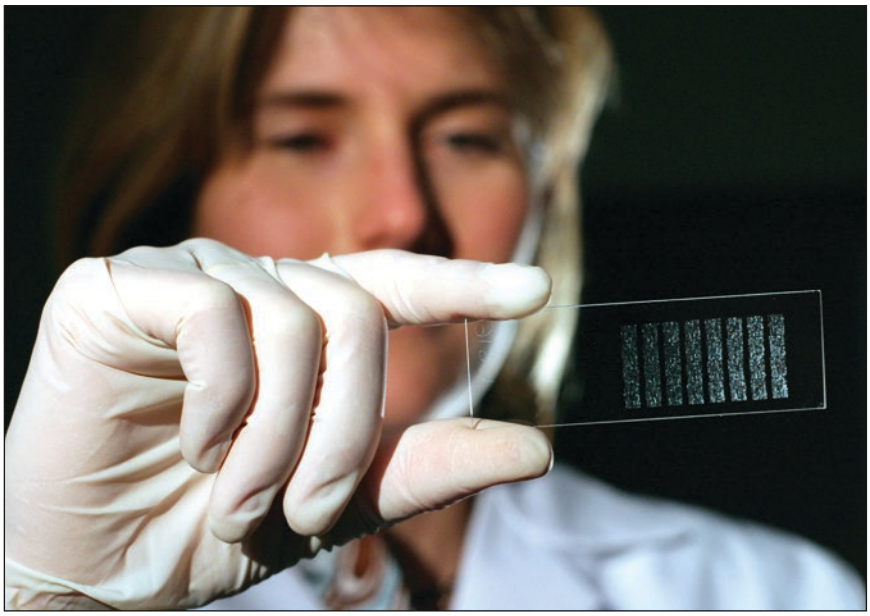


Figure B.24 A microarray contains tens of thousands of microscopic DNA probes.

Microarrays enable scientists to carry out thousands of genetic tests at the same time. They can be used to check for the presence of genes in a genome, to check for genetic predisposition to a disease, or to diagnose a condition. DNA microarrays are used to seek out single nucleotide polymorphisms (SNP) among alleles making them important tools in evaluating faulty genes or locating genetic mutations in cancers.

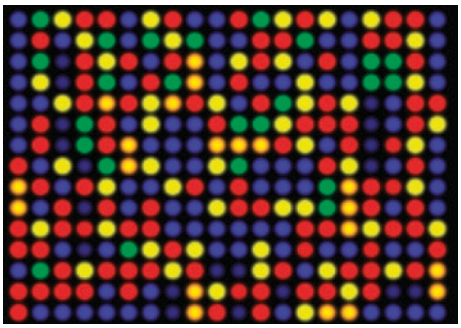


Figure B.25 Results from a microarray. An array of DNA sequences for a particular set of genes is created on the biochip. Each gene's position in the array is known. Samples of unbound DNA or mRNA (messenger RNA) are then labelled with differently coloured fluorescent markers and added to the chip. The genetic material in the samples binds to sites on the array that have a matching (complementary) sequence. The pattern of colours is then analysed.

Tau proteins are found in neurons where they are concerned with the formation of microtubules, tiny structures responsible for internal communication and transport. If they are folded they can become tangled and form insoluble aggregates, which interfere with cell function and are characteristic of Alzheimer's and other neurological diseases.

Alzheimer's disease

Research has revealed that testing either blood or cerebrospinal fluid may allow doctors to detect a person's predisposition to developing Alzheimer's disease, which affects memory and a person's ability carry out everyday tasks. Individuals whose blood contains higher than average levels of the marker peptide beta-amyloid 42 (Aβ42) are at increased risk of developing Alzheimer's disease. Researchers have found that plasma levels of Aβ42 increase before the onset of Alzheimer's disease and decline shortly afterwards. This may indicate that as the disease progresses the decline of Aβ42 in the bloodstream is linked to accumulation of Aβ42 in the brain, which occurs in people who have developed dementia.

More recently, genetic markers that identify mutations affecting the build-up of certain proteins in the brain have also been used to help diagnose Alzheimer's disease. High levels of **tau proteins** and the presence of folded molecules of phosphorylated tau are signs of the disease.

Prostate and other cancers

A **tumour marker** is a substance found in blood, urine or tissue samples that may become elevated if a person has developed cancer. There are many different tumour markers and each one is specific to a particular disease. Tumour markers can be produced either directly by the tumour or by non-tumour cells that are responding to the presence of a tumour. Most markers are tumour antigens.

Today, the most widely used tumour marker in cancer diagnosis is the prostate-specific antigen (PSA), which is detected in a blood test. The test is used to check men for prostate cancer. Men with prostate cancer usually have high PSA levels. But the test is not perfect because high PSA levels can occur in men without cancer, and a normal PSA level does not necessarily mean that no cancer is present. Other tumour markers have been identified which can indicate many other cancers including colon, stomach and breast cancers.

Metabolites in blood and urine

The serotonin production pathway is activated when a person consumes alcoholic drinks so this can be a metabolic marker indicating recent alcohol consumption.

As we have seen, a **biomarker** is a substance that can be measured and which can indicate the presence or severity of a disease. General biomarkers used in medicine include LDL (low density lipoprotein), which can be measured to indicate blood cholesterol levels, and genetic biomarkers for specific cancers or predisposition to them.

Metabolomics is a name given to the analysis of metabolites in a blood or urine sample. In many cases diseases cause disruption of metabolic pathways and the accumulation of, or lack of, metabolites can be biological indicators of certain diseases. Some examples of diseases that can be detected in this way are described below.

Diagnosis of diabetes – ketones in urine

If the body is deprived of carbohydrates, it must rely on the metabolism of fats to supply energy. This can occur if an individual is starving or following a high-protein diet, but it is also a symptom of diabetes. Fat metabolism occurs in several stages and the intermediate products of the process known as ketones can accumulate in the body if they are not metabolised completely. Ketones accumulate in the blood and are excreted in urine. Their presence in urine samples can be an indicator of diabetes.

Diagnosis of PKU – phenylalanine in blood samples

Phenylketonuria (PKU) is a genetic disorder caused by a mutation on chromosome 12. People who suffer from PKU lack the enzyme tyrosine hydroxylase, needed to process the amino acid phenylalanine. If phenylalanine levels in the blood are high, serious mental and physical health problems can arise. Babies are routinely tested for the amino acid with a simple blood test and if it is present suitable treatment can be given.

Glycogen storage disorders – creatine kinase in blood samples

Glycogen storage disorders are a range of conditions that can lead to low blood sugar levels and other symptoms such as fatigue and muscle problems. Diagnosis includes the use of blood tests to check for lipids and urate levels, and for the enzyme creatine kinase, which can indicate whether the glycogen storage disorder is affecting an individual's muscles.

Galactosemia – enzyme absent from blood or urine

Galactosemia is a rare genetic disorder affecting a person's ability to metabolise the sugar galactose (found in milk). It is due to a deficiency in an enzyme responsible for breaking down the galactose. A galactosemia test involves testing blood or urine for three enzymes. A person with galactosemia is lacking one of these enzymes so that high levels of galactose are present in their blood or urine.

Porphyria – accumulation of precursors in the blood or urine

Porphyria is caused by a deficiency in any of the eight enzymes involved in the synthesis of porphyrins and heme. Normally, the body uses porphyrins to produce heme, but if enzymes are missing, porphyrins and precursors in the reaction accumulate and insufficient heme is produced. (Heme is essential for the formation of hemoglobin and healthy red blood cells.) Sometimes the precursors are deposited in the skin and lead to photosensitivity. Some porphyrins may pass out of the body in feces or with urine, colouring it black. In other cases of porphyria, porphyrins such as porphobilinogen (PBG), one of the first substances in the synthesis pathway, can build up. Initial diagnosis of porphyria involves testing urine for the presence of PBG, followed by tests on blood and feces for other porphyrins.

Tracking tumour cells – transferrin and luminescent probes

Transferrins are glycoproteins found in the blood, which bind to iron and control the level of free iron in the plasma. **Transferrin receptor** (TfR) is a glycoprotein found in cell membranes that is involved in the uptake of iron from transferrin and also with the regulation of cell growth. Transferrins carrying iron bind to TfRs and iron can then be transported into the cell by endocytosis.

Recent studies have shown that there are higher levels of TfR on tumour cells than on normal cells. The TfR on normal cells is expressed at low levels but is expressed at greater levels on cells with a high proliferation rate, including cells of the epidermis and intestinal epithelium as well as tumour cells. The high levels of this protein receptor, in an accessible extracellular position, can be used as a target for the

location and treatment of cancerous cells. Tumour cells can be tracked by linking luminescent probes to transferrin molecules, which then attach to the TfRs. In this way, luminescence has been used both to measure levels of TfRs and to identify and track tumour cells in the body.

One new cancer treatment uses this transferrin technology to target and destroy cancerous cells. The technique involves a 'hybrid peptide' composed of two parts: one peptide binds to a target TfR, while a second, lytic peptide then destroys the cell membrane and kills the cancerous cell.

Biopharming

Biopharming refers to the use of genetic engineering to add genes to animals or plants so that they produce useful pharmaceuticals. The GM animals or plants express the added genes and so produce protein from them. As we have seen, proteins produced from recombinant DNA such as human insulin can be produced in bacteria or yeast in bioreactors (Subtopic **B.1**), but one of the main reasons for the increasing interest in animal biopharming is its potential to produce substances that cannot be made using these methods. Transgenic animals have an advantage over both plants and microorganisms because they are able to carry out complex post-translational modifications to the proteins they produce. Animals can therefore make high-quality, biologically active proteins for use as pharmaceuticals.

Harvesting the desired proteins from milk is the most popular way of accessing recombinant proteins from transgenic organisms. Milk production is plentiful and purification of the proteins from milk is straightforward. Goats are one species that is well suited to biopharming because they have a faster breeding cycle than cattle and produce more milk than sheep. Females mature quickly and have a gestation period of just 6 months.

Transgenic goats can be produced in one of two ways. Either by the fusion of cloned fragments of DNA coding for the required proteins with goat genes, which are then injected into eggs or by nuclear transfer (Figure **B.26**). The gene that codes for the chosen protein is attached to a promoter gene, usually the gene for casein, a milk protein, and then microinjected into the nucleus of a newly fertilised egg in the laboratory. Alternatively the DNA is injected along with goat genes into an enucleated egg. In both cases the gene may become incorporated with the genetic material of the embryo as the cell divides. If this happens, the new gene (or transgene) will be incorporated into every cell of the developing goat embryo. The embryo is cultured and then transferred to a goat who will become a surrogate mother. If a female kid is produced, she will produce milk with the new protein, which can be extracted from her milk. When the female kid becomes an adult and breeds, half of her offspring will have the new gene.

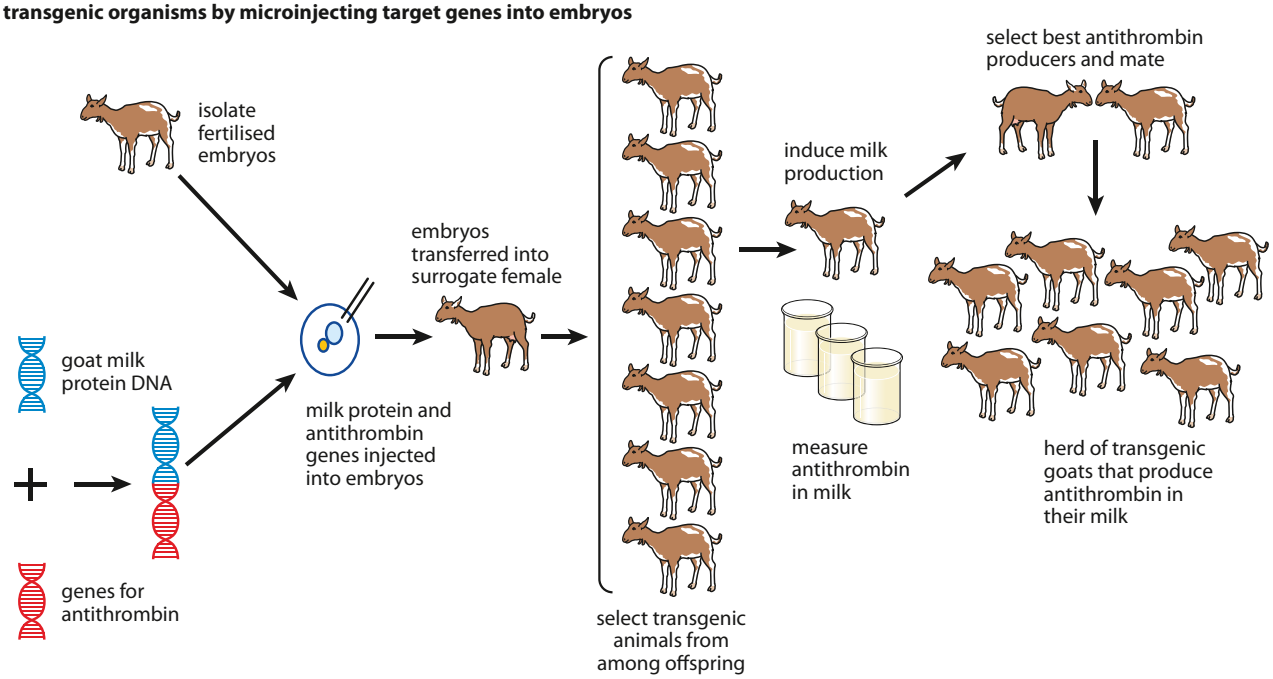
A herd of transgenic goats can be produced using microinjection of genes into enucleated eggs and using the transgenic offspring for reproduction. Cloned early-stage embryos can be used to create additional offspring that carry the modified genome. However, the low survival rate of embryos and animals generated by biopharming is one of the major problems of this technique.

Biotechnologists anticipate that cloning will become far more important as the biopharming of animals develops. Cloning embryos can increase the efficiency rate significantly. The rate of success for microinjection is low. If 1000 fertilised eggs are injected, only 100 embryos are likely to be produced and from these only one transgenic offspring is likely to be born – and there's only a 50% chance of its being female. Cloning embryos produced using adult genetic material and enucleated eggs has a much higher success rate.

In 2009, the US Government gave permission for the sale of the first drug to be produced by biopharming – the drug, called ATryn, is recombinant human antithrombin protein purified from the milk of genetically modified goats. Marketing permission for ATryn was granted in Europe in 2006. (Antithrombin activates enzymes needed for clotting blood so people who inherit antithrombin deficiency have blood that does not clot easily.)

As well as blood clotting factors many other therapeutic substances have been produced by biopharming, including antibodies, growth hormones, fibrinogen, albumin and recombinant enzymes. Treatments for cystic fibrosis using recombinant gastric lipase are also being developed.

Creating transgenic organisms by microinjecting target genes into embryos



Creating transgenic organisms by nuclear transfer

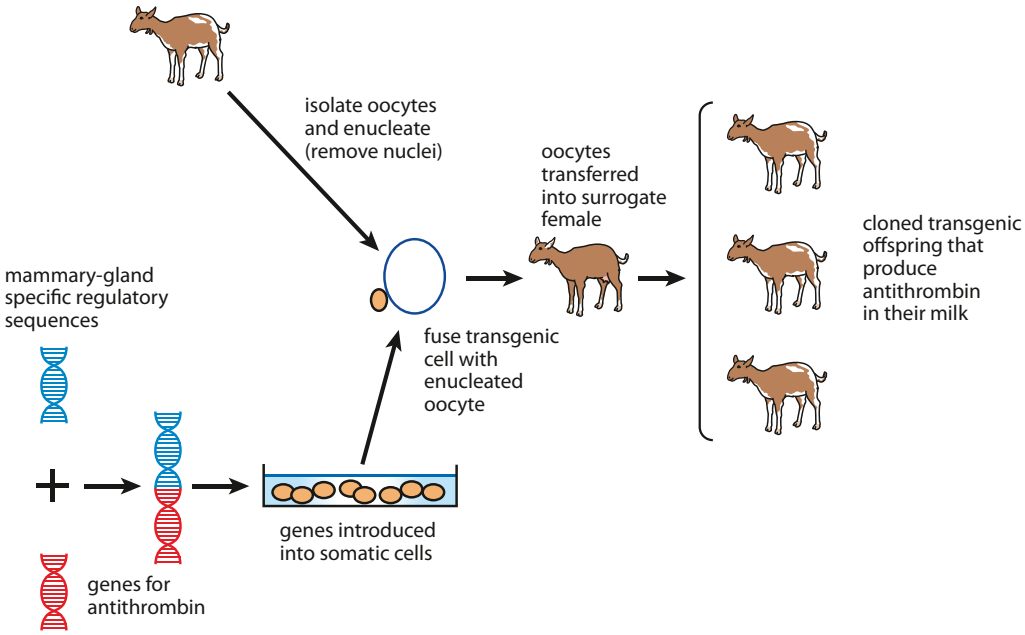


Figure B.26 Creating transgenic goats.

Therapeutic proteins from transgenic plants are also being developed. The proteins can be harvested from the plants' leaves or seeds and used to produce pharmaceuticals. (In Subtopic **B2**, you can read about how the tobacco plant is used to produce vaccine proteins, for example.) New developments in this area include the use of transgenic non-crop plants such as duckweed (*Lemna minor*) and moss. These organisms can be grown in bioreactors, where they secrete proteins produced from introduced recombinant DNA into their growth medium. This makes the purification of proteins for medical use much simpler than for a species grown in a field and then harvested.

Viral vectors in gene therapy

Viruses are very efficient at entering the cells of organisms, and can be used as **vectors** providing a powerful means for the delivery of therapeutic genes into cells. Some viruses are even able to incorporate the genes they carry into the cells they enter.

Before a virus can be used, it must be modified so that it will enter but not replicate inside a target cell, as this would lead to the cell's destruction. Viral genes that are involved in replication are removed or inactivated. Deletion of these genes also allows non-viral genetic material to be inserted and these viruses are then known as vectors.

Retroviruses are the most frequently used vectors. These single-stranded RNA viruses enter target cells via specific receptors and their RNA is converted into DNA and integrated into the genetic material of the cell (Subtopic **B2**), where it remains for the life of the cell. Integrated genes are also passed on when the cell divides.

There has been some success in treating a condition called severe combined immune deficiency (SCID) using retroviruses. Children who suffer from SCID have no immune system because a gene mutation prevents their cells producing the enzyme adenosine deaminase (ADA). Substrates for ADA build up in cells and are very toxic to developing lymphocytes. These cells fail to mature and the patient is left without a working immune system. Stem cells from bone marrow or umbilical cord blood can be taken out of the body and treated with viral vectors that transfer a normal copy of the ADA gene to them. If the treated cells are returned to the bone marrow, the replacement genes can begin to produce ADA. Bone marrow and stem cell transplants now save up to 80% of SCID patients.

Gene therapy

Gene therapy involves modification of genetic material in the cells of a patient in order to bring about a therapeutic effect. Modification is usually achieved by introducing DNA, using viral vectors or other means. Although gene therapy is still in its infancy in medical treatments, discussion of the ethical issues involves principles that apply to all clinical medicine. How are subjects selected for gene therapy trials? How can the safety of individuals who take part in the trials be safeguarded, and do scientists involved in the trials have a conflict of interest between wanting to conduct their research and the best interests of their patients?

In 2007, a 36-year-old woman with rheumatoid arthritis died while participating in a clinical trial for gene therapy. Some experts say she shouldn't have received such an unpredictable, potentially dangerous treatment at all. She was able to lead a full and active life with existing drugs keeping her disease under control. Soon after the experimental treatment a sudden infection caused her organs to fail and there is a suspicion that her death was linked to the therapy. Is it ethical to test unknown therapies on patients whose ailments are not life threatening? Priorities in medical research can raise troubling issues of social ethics.

Questions to consider

- Can 'good' and 'bad' uses of gene therapy be distinguished?
- Who decides which traits are normal and which constitute a disability or disorder worthy of gene therapy treatment?
- Can the safety of patients ever be guaranteed in treatments that are new or experimental?
- Are there conflicts of interest between patient safety and the need to conduct research?

Nature of science

Scientific advance follows technical innovation – technology in medicine

The use of computers to store and collate data, the use of microtechnology to produce DNA chips and the use of robotic plating in ELISA tests are all examples of the way that technology has enabled scientists to collect more, better and more precise information. The use of such technology in medicine is an area that has expanded rapidly in recent decades, allowing huge improvements in the understanding, diagnosis and treatment of disease.

Test yourself

12 List **two** diseases that can be detected by the presence of markers in a blood sample.

13 Describe how a microarray can be used to search for a sequence of nucleotides in a DNA sample.

14 Define the term 'biopharming'.

B5 Bioinformatics (HL)

Learning objectives

You should understand that:

- Databases allow scientists to access information easily.
- The quantity of data in databases is increasing exponentially.
- Similar sequences can be identified in the genomes or proteins of different organisms using BLAST searches.
- Gene function can be investigated by studying model organisms that have similar sequences.
- Sequences from different organisms can be compared using sequence alignment software.
- BLASTp software allows scientists to find alignment between amino acid sequences in proteins and known sequences stored in a database, while BLASTn software allows nucleotide sequence alignment.
- Searches of databases can allow newly identified sequences to be compared with sequences of known function in other organisms.
- Multiple sequence alignment is used in phylogenetic studies.
- Expressed sequence tags (ESTs) can be used to identify potential genes.

Databases

Biological databases are enormous electronic records of information from experiments, scientific papers and computer analyses of data. Rather like a library of books, a database is a collection of structured, searchable and up-to-date data maintained electronically. Information is stored from research in genomics, proteomics, metabolomics, microarray results on genes and phylogenetic studies of evolution, and can be accessed quickly and easily. In genetics, the information provides scientists with easy access to information about gene function, structure and location as well as storing information on the effects of mutations. Similarities between sequences and structures of DNA and proteins in different species can be obtained and compared with new information that is collected. Databases can help our understanding of the interaction of biomolecules or observations of metabolism.

The exponential increase in the quantity of data stored in databases has helped in the development of drugs, our ability to combat diseases and our understanding of evolution. Genome sequencing, in particular, along with many other large-scale research projects, has generated an explosive growth in biological data and there are many specialised databases. Species-specific databases are kept for species such as *Escherichia coli*, *Drosophila* and nematodes, which are model organisms used in many research projects. There are hundreds of public, freely accessible databases available in a matter of minutes to anyone worldwide. But recently some private genome-sequencing companies have started to charge for data that might be of commercial interest – for example, the public genome database for *Saccharomyces cerevisiae* (yeast), a model organism used by many scientists, changed from a free to a chargeable database in 2002.

Data are usually entered in databases and assigned an identifying number for quotation in scientific publications. Most of the data are not directly from the original source – the contents are extracted from other databases by a process of filtering, transformation and manual correction. Because the information is distributed among many general and specialised databases it is sometimes difficult to make sure that the information is consistent.

Model organism non-human species used for study and experimentation in order to understand particular pathways or processes. *Escherichia coli*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and the plant *Arabidopsis thaliana* are all model organisms. Experiments are done with the expectation that any discoveries made will increase understanding of the same processes in other organisms

Freedom of information

The information in the formerly public database containing information about yeast was taken over by a commercial company in 2000. They now charge US $2000 per lab per year for scientists to use the database. Much of the information in the database has come from scientific publications and personal communication between researchers who have used it.

Question to consider

- Should commercial interests ever be put ahead of free access to scientific information?

BLAST searches

BLAST is an **algorithm** used to compare sequence information, such as the amino acid sequences in different proteins or the nucleotide sequences in DNA from different organisms. The BLAST program was first published in 1990 and now researchers can use a BLAST search to compare a sequence of DNA or amino acids with a database containing known sequences. The search will reveal all the known sequences that have more than a certain proportion of resemblances to the sequence being investigated.

Algorithm a set of rules that defines a sequence of operations for solving a problem in a finite number of steps

The **BLASTp** program allows amino acid sequence alignments to be made and matches to be found between newly discovered sequences and known sequences. Questions it can help to answer include:

- Do analogous proteins from different species share similar amino acid sequences, and if so are the species likely to be related in evolutionary terms?
- Where in a protein is a particular sequence of amino acids located?

The **BLASTn** program allows nucleotide sequence alignments to be made. BLASTn can be used for several purposes such as the identification of species, establishment of evolutionary relationships, and DNA mapping and comparison:

- BLASTn can help to determine the origin of a certain sequence of DNA. It can help to identify a new or known species from a sample of its DNA and search for related species with similar sequences.
- Using the results from BLASTn, a phylogenetic tree can be built up – although phylogenies based on BLASTn alone are not 100% reliable and are always used in conjunction with other phylogenetic analyses.
- In DNA mapping, BLASTn can compare the chromosomal position of a gene sequence of a known species with the positions of other sequences in the database, or locate genes that are common in related species to map differences and similarities between one organism and another.

Studying gene function in model organisms

Knockout mice

Gene function can be studied using model organisms such as the mouse, which is the laboratory animal mostly closely related to humans. The mouse has many genetic sequences that are very similar to those of humans.

A **knockout mouse** is a genetically engineered animal that has had certain genes inactivated or 'knocked out'. The first knockout mouse was created in 1989 and since then thousands of different strains have been made. A gene in a knockout mouse may either be replaced or disrupted by the addition of a section of artificial DNA. Knocking out genes changes the phenotype of the mouse so that it looks or behaves in a way that researchers can see, or it has different biochemical characteristics that can be monitored (Figure **B.27**).

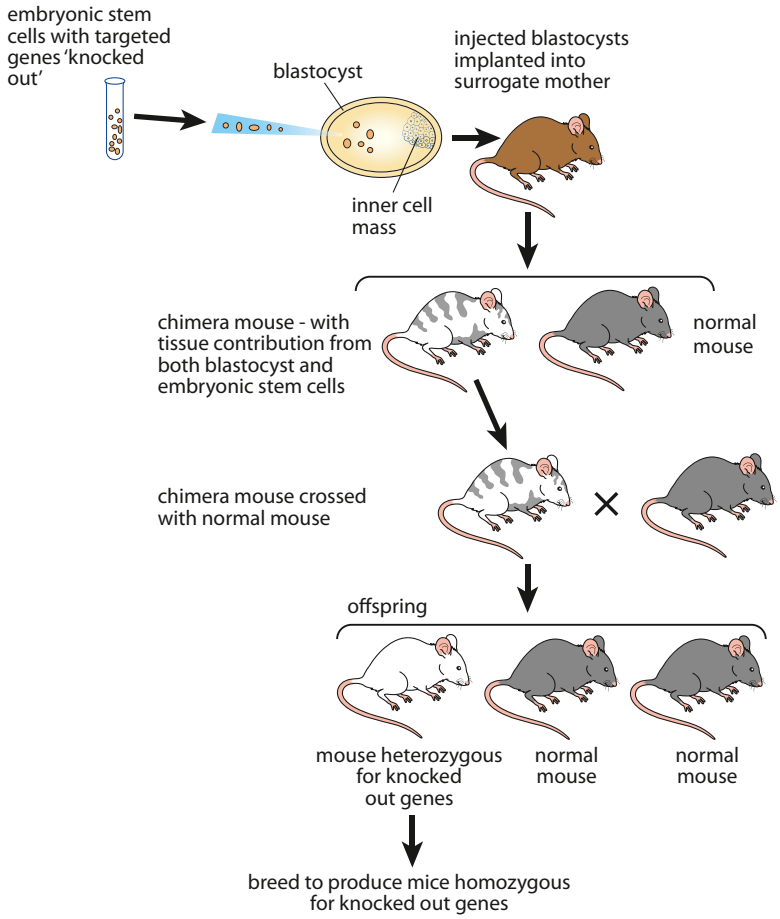


Figure B.27 Genes are replaced or disrupted in a blastocyst, which is implanted into a surrogate mother.

Knockout mice are important animal models for studying the genes that have been sequenced but whose functions are not known. By inactivating a specific mouse gene and observing any differences in behaviour or physiology, researchers can deduce its possible function. Different strains of model mice are named after the gene that has been deactivated. The 'p53 knockout mouse' is named after the gene that codes for a tumour-suppressing protein. The protein stops cell division and induces programmed cell death (apoptosis). In humans, the genetic disease Li–Fraumeni syndrome occurs in people who have a mutation that deactivates their p53 gene. These individuals are at a high risk of developing several types of bone, breast and blood cancers. However, in p53 knockout mice cancers develop in quite different tissues. So, as data is extrapolated from one species to another, scientists must be aware that a change in a gene in a knockout mouse may produce different effects in humans.

Discovery of genes using EST data mining

Improved understanding of normal gene expression can help our understanding of what happens when genes are altered – for example, in disease. Most studies of this type involve investigations of the proteins that are coded for by the genes involved. But finding exactly which gene codes for which protein is not easy and is very time consuming. The process has been improved considerably since the development of new technology to generate **expressed sequence tags** (ESTs). An EST is the nucleotide

sequence of a tiny portion of a known gene, which computers can use to scan databases to help make a match with unknown genes and to map their positions within a genome, assisting in the construction of genome maps. ESTs are short sequences (usually between 300 and 500 nucleotides long) that occur at one or both ends of a gene. ESTs effectively 'represent' the genes expressed in certain cells, and researchers use them as 'tags' to locate a gene in a portion of chromosomal DNA from a different organism, by matching up the base pairs. ESTs provide a route for finding new genes and for obtaining data on gene expression and regulation.

Gene sequences coding for analogous proteins are different in different organisms, varying with the size of the genome and the introns the genes contain. Humans have large numbers of introns and relatively few genes, which makes an individual gene almost impossible to locate. mRNA is used to help find those genes that are expressed and produce protein. But mRNA is an unstable molecule so for research purposes it is converted to complementary DNA (cDNA), which is more stable and contains none of the introns that were present in the original chromosomal DNA sequence. cDNA can be prepared in the laboratory using reverse transcriptase (Subtopic **B2**), which uses mRNA as a template to produce cDNA containing only exons (expressed DNA sequences) (Subtopic **7.1**). ESTs can be sequenced from either end of the cDNA but sequencing at the start of the molecule produces a **5' EST**, which is more useful, because this region tends to be conserved across species and is not significantly different in groups of closely related genes that produce similar proteins.

ESTs are crucial in the search for known genes because they greatly reduce the time needed to find them. For example, suppose a doctor suspects that a person has a disease or condition associated with a missing or defective protein, and wants to check the person's DNA to see if the gene for that protein is defective. ESTs for that known gene could be used to seek out the gene in a DNA sample from the person, and that gene analysed to see if it differs significantly from normal. In humans, genes involved in Alzheimer's disease, colon cancer and other diseases have already been found in this way.

In the 1990s, a new database called dbEST was set up to record data from ESTs. All ESTs that are submitted to the genetic database GenBank are checked and annotated and then lodged in dbEST, which stores a wide variety of genome data.

Exam tip
Make a list of key terms that are important in this and other topics in your course. Some might be **cDNA**, **bioassay**, **cladogram**, **PCR**. Check that you can define all of them.

Nature of science

Cooperation and collaboration – databases allow free access to information

Scientists all over the world use databases to access information on their research organisms, and to contribute to the body of information stored there. The internet has provided the opportunity for interaction in a way that has never been possible before and computers have enabled huge amounts of quantitative data to be stored and accessed. All this information would have taken years to collate and catalogue before the advent of computers.

Reliability of databases

Scientists use databases to provide data, which they use in their research. But how reliable can claims based on such data be? Is it important that the data on which a scientist bases his or her research may have come from researchers using different techniques and producing the data for different reasons?

Test yourself

15 State the type of data that can be accessed in a BLASTn search.

16 Outline what is meant by model organisms and why they are important in research.

17 Describe how ESTs are used to identify potential genes in DNA sequences.

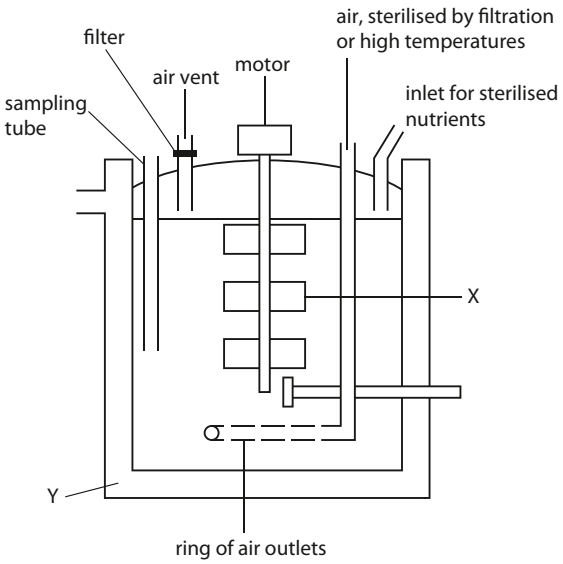
Exam-style questions

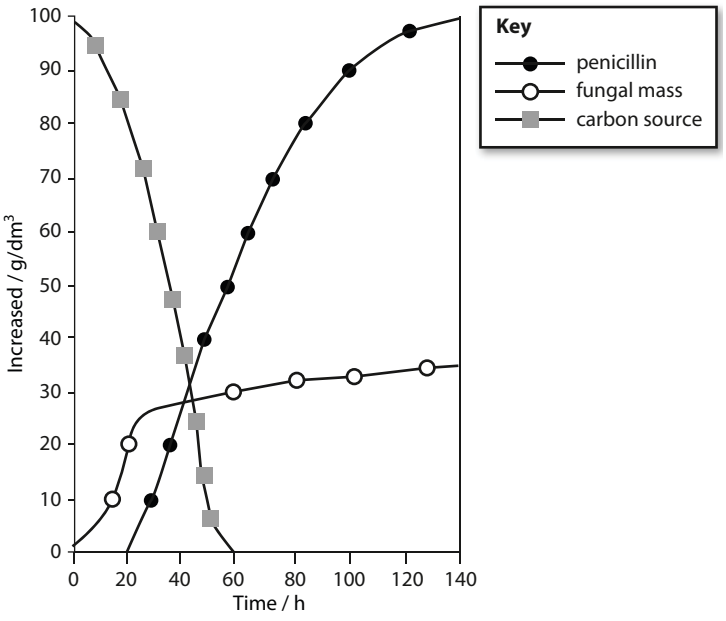
1 Penicillin is manufactured using a process called batch fermentation.

a Outline what is meant by 'batch fermentation'. **[2]**

b State how batch fermentation differs from continuous culture. **[1]**

A fermenter similar to the one shown in the diagram below is used to produce penicillin. The graph shows the rates of penicillin production and use of carbon by the *Penicillium* fungi.





c What is the function of the structure labelled X? **[1]**

d State the function of structure Y. **[1]**

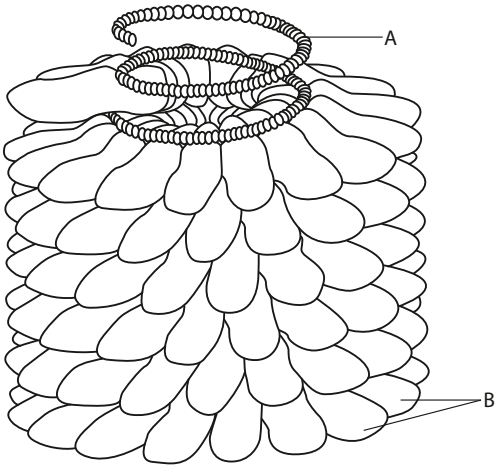
e Define the term 'secondary metabolite'. **[1]**

Using information shown in the graph:

f State the time when the fungal population reaches the plateau stage. **[1]**

g Explain why penicillin production is not increased if further nutrients are added at 30 hours. **[1]**

2 The diagram shows the structure of Tobacco Mosaic Virus (TMV), which infects the leaves of the tobacco plant.



a Name the molecule A. **[1]**

b Name the molecules that make up structure B. **[1]**

c Explain how TMV has been used to allow the bulk production of vaccines from tobacco plants. **[5]**

3 Outline **three** advantages of the Amflora potato in industry. **[3]**

4 Which of the following statements about transgenic plants is true:

A they contain genes from another species
B they are used to produce human antibodies
C they are from different geographical areas
D statements a and b only **[1]**

5 Which of the following microorganism is used for the production of citric acid?

A *Lactobacillus bulgaricus*
B *Saccharomyces cerevisiae*
C *Aspergillus niger*
D *Streptococcus lacis* **[1]**

6 Which of the following statements correctly describes transfection?

A The synthesis of mRNA from a DNA template
B The process by which a cell become malignant
C The introduction of genes from a different species in to a cell
D The synthesis of protein based on mRNA sequence **[1]**

7 Describe what is meant by the term 'open reading frame'. **[2]**

8 Complete the following table inserting a tick (✓) if the structure is present and a cross (✗) if it is not.

Microorganism	Nucleus	Mitochondria	Ribosomes
fungus			
bacterium			
virus			

[3]

9 Biofilms are said to possess emergent properties.

a Define the term biofilm. [1]

b State one emergent property that a biofilm might possess. [1]

c Biofilms co-operate using quorum sensing. Describe what advantages this may confer to the microorganisms. [2]

10 Animal pharming can be described as which of the following:

A Growing animals for farming
B Genetically modifying animals to produce novel products
C Producing transgenic animals for farming
D Creating clones of useful animals **[1]**

11 A micro array is an ordered array of microscopic elements on a flat substrate that allows the specific binding of which of the following:

A A gene or gene products
B A whole genome
C A virus
D Coloured markers **[1]**

12 Which of the following statements about nucleic acids contained in liposomes used for gene transfer in plants are true?

i They are protected from nuclease digestion
ii Nuclei acids are stable in liposomes
iii Nuclei acids are unstable in liposomes
iv They are not protected from nuclease digestion

A i only
B i and ii only
C i and iii only
D iii and iv only **[1]**