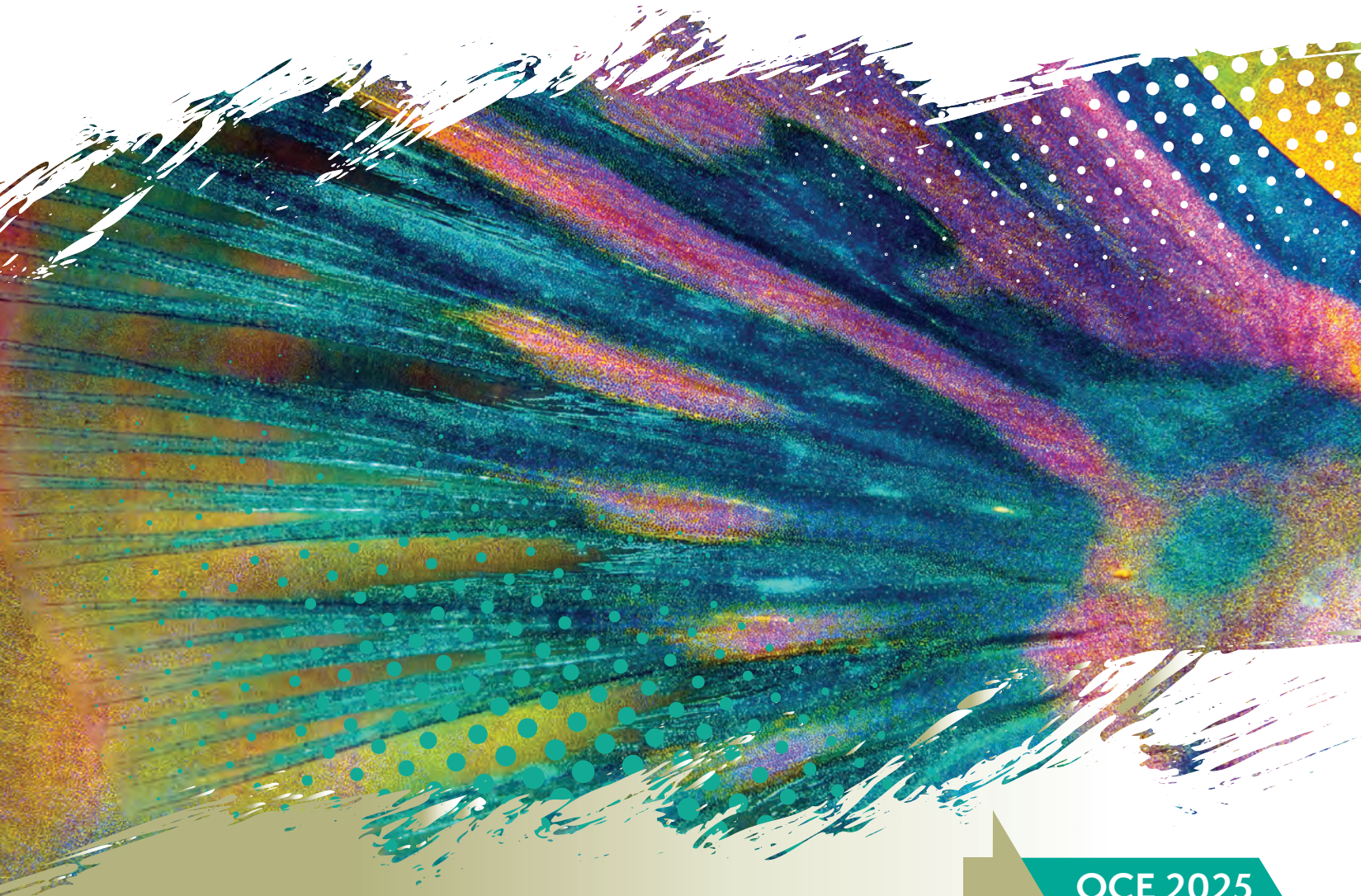
PEARSON

BIOLOGY

QUEENSLAND

UNITS 3 & 4

2nd EDITION



Student Book

QCE 2025
Biology
SYLLABUS

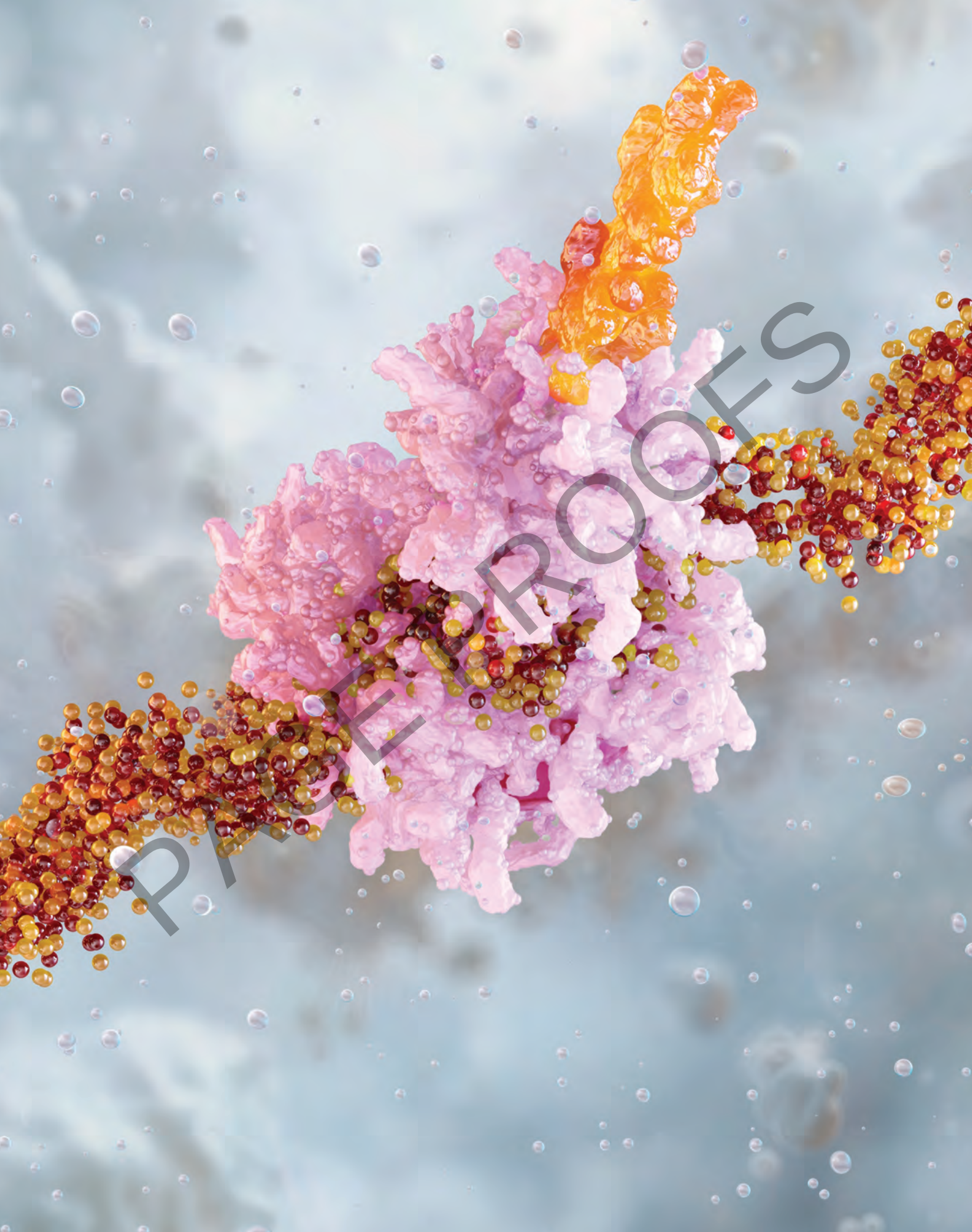
UNIT 4 Heredity and continuity of life

TOPIC 1 Genetics and heredity

TOPIC 2 Continuity of life on Earth

Unit 4 objectives

- Describe ideas and findings about genetics and heredity, and the continuity of life on Earth
- Apply understanding of genetics and heredity, and the continuity of life on Earth
- Analyse data about genetics and heredity, and the continuity of life on Earth
- Interpret evidence about genetics and heredity, and the continuity of life on Earth
- Evaluate processes, claims and conclusions about genetics and heredity, and the continuity of life on Earth
- Investigate phenomena associated with genetics and heredity, and the continuity of life on Earth



CHAPTER

05 DNA replication and protein synthesis

The molecular basis for inheritance is DNA. Its structure was determined by scientists in the 1950s in what was one of the most significant discoveries of the 20th century. Nucleic acids have the unique ability to direct their own replication and pass on the chemical code from which proteins are synthesised, which ultimately determines the characteristics of their offspring.

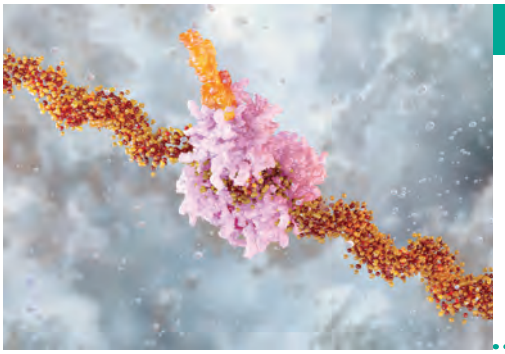
By the end of this chapter, you will be able to describe the structure and function of DNA and RNA and the role of these nucleic acids in protein synthesis. You will understand how DNA is replicated, how proteins are synthesised and how gene expression is regulated by chemical tags and transcription factors. You will also be able to identify different types of mutations and understand the effects of point and frameshift mutations.

Syllabus subject matter

Topic 1 • Genetics and heredity

- Describe the structure and function of DNA, genes and chromosomes in prokaryotes and eukaryotes, including
  - helical structure, nucleotide composition (nitrogenous base + sugar + phosphate), complementary base pairing, hydrogen bonds
  - introns and exons, promoter region
  - homologous chromosomes (i.e. sister chromatids, centromeres, telomeres, gene loci, alleles), role of histones
  - circular chromosomes (i.e. prokaryotes, mitochondria, chloroplasts) and plasmids **5.1**
- Describe the process of DNA replication with reference to helicase, DNA polymerase and the joining of Okazaki fragments **5.1**
- Explain how errors in DNA replication and damage by physical/chemical factors in the environment can lead to point and frameshift mutations **5.1**
- Explain the process of protein synthesis in terms of
  - transcription of a gene into messenger RNA in the nucleus
  - RNA processing (5' cap, RNA splicing, poly-A tail)
  - translation of mRNA into an amino acid sequence at the ribosome, referring to transfer RNA, codons and anticodons **5.2**
- Determine the effect of point and frameshift mutations on polypeptides using the genetic code **5.2**
- Explain how gene expression is regulated in response to environmental signals and to allow for cell differentiation, including
  - chemical tags that affect chromatin structure (heterochromatin versus euchromatin)
  - proteins that bind to the promoter region of a gene (transcription factors) **5.2**
- Explain how genes from the HOX transcription factor family regulate morphology **5.2**

5.1 DNA structure and replication



BY THE END OF THIS MODULE, YOU SHOULD BE ABLE TO:

- describe the structure and function of deoxyribonucleic acid (DNA), genes and chromosomes
- understand condensation polymerisation and complementary base pairing
- understand how DNA is packaged in chromosomes and the different types of chromosomes (linear, circular, homologous)
- describe the process of DNA replication and explain how DNA damage and errors in DNA replication can result in different types of mutations.

Nucleic acids are organic biomolecules that store and transmit inherited characteristics of organisms by encoding instructions for the synthesis of **proteins**. The two types of nucleic acids are **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. You will learn about DNA in this module and RNA in Module 5.2.

THE COMPOSITION OF NUCLEIC ACIDS

Nucleic acids (DNA and RNA) are made up of nucleotides.

Nucleotides

Nucleotides are the building blocks of DNA and RNA. A single nucleotide consists of three units:

- a phosphate group—the same in all nucleotides
- a five-carbon (pentose) sugar, either:
  - **deoxyribose** in DNA nucleotides
  - **ribose** in RNA nucleotides
- a nitrogenous (nitrogen-containing) **base**.

The five carbon atoms in a pentose sugar molecule are labelled 1′ to 5′ (pronounced ‘one prime’ and ‘five prime’). In a single nucleotide, the phosphate is always attached to the 5′ carbon and the base is always attached to the 1′ carbon (Figure 5.1.1).

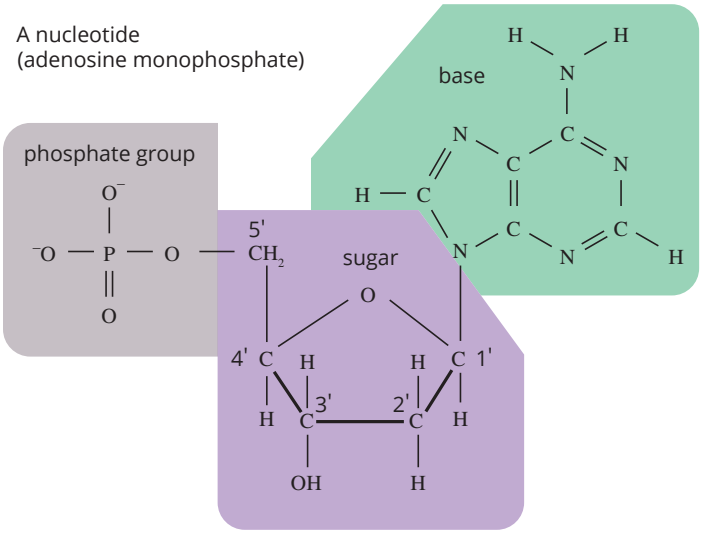


FIGURE 5.1.1 The structure of a DNA nucleotide, showing the phosphate group, the five-carbon sugar and the nitrogenous base (adenine, A, in this example).

Polynucleotide chains

Polymers consist of a long chain of repeating molecules (or **monomers**). Nucleic acids are polymers that consist of a long chain of repeating nucleotide monomers, also known as a **polynucleotide chain**, or strand (Figure 5.1.2). The ends of the polynucleotide chain are referred to as the 5′ and 3′ ends (pronounced 'five prime' and 'three prime'). At the 5′ end of the polynucleotide strand the fifth carbon atom is free. At the 3′ end of the polynucleotide strand the third carbon atom is free.

During DNA synthesis, the 3′ hydroxyl group of the deoxyribose sugar of one nucleotide reacts with the phosphate group of another incoming nucleotide. This reaction is called **condensation polymerisation** and it releases a water molecule and forms a type of covalent bond called a **phosphodiester bond** between the sugar and phosphate group. This process is repeated, and the sugar–phosphate backbone of DNA is formed as a result, with each nucleotide's sugar molecule connected to two phosphates: one at its 5′ carbon and the other at its 3′ carbon. The nitrogenous bases are attached to the sugars but are not part of the sugar–phosphate backbone. DNA and RNA are synthesised in the 5′ to 3′ direction.

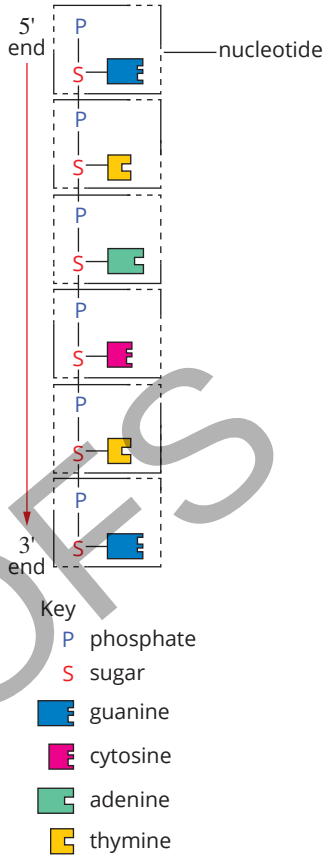


FIGURE 5.1.2 A single polynucleotide chain. Individual nucleotides (shown in boxes) are joined by phosphodiester bonds.

Nitrogenous bases

Attached to the pentose sugar within every nucleotide is a **nitrogenous base**. There are five different nitrogenous bases:

- **adenine (A)**
- **guanine (G)**
- **cytosine (C)**
- **thymine (T)**—in DNA only
- **uracil (U)**—in RNA only.

These five nitrogenous bases can be categorised into one of two groups based on their structure (Figure 5.1.3). **Pyrimidines** (C, T and U) have one ring in their structure. **Purines** (A and G) have two rings in their structure.

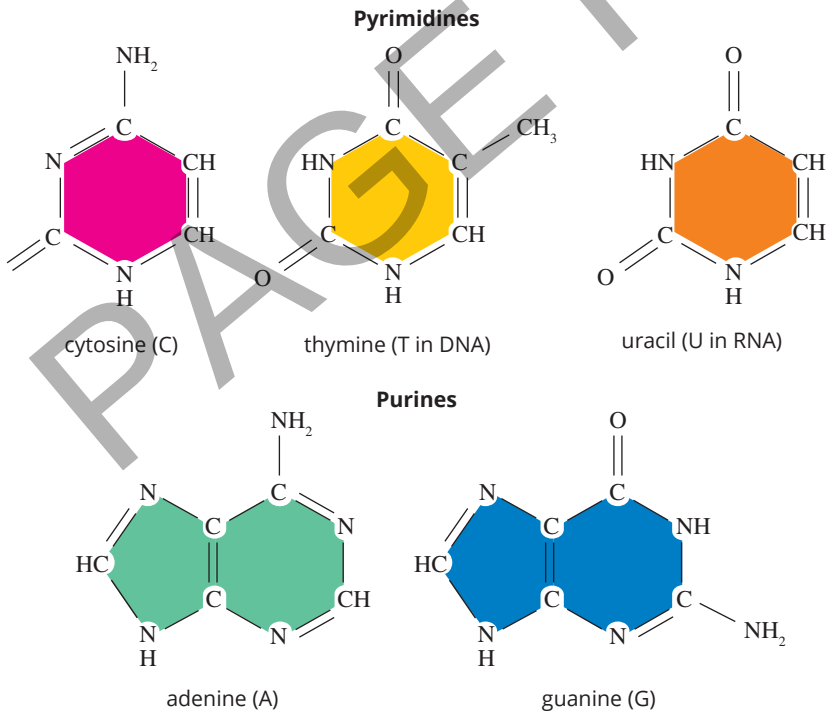


FIGURE 5.1.3 The structure of pyrimidines (cytosine, thymine and uracil) and purines (adenine and guanine).

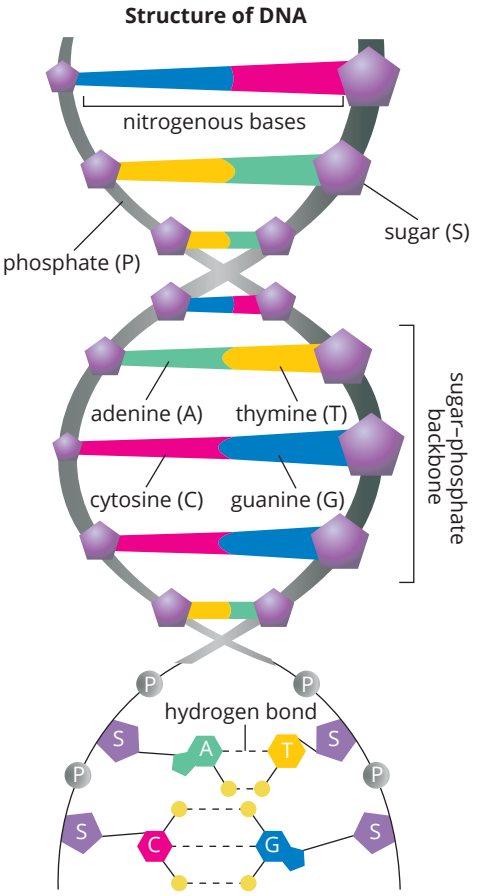


FIGURE 5.1.4 The structure of the DNA double helix biomolecule. The helical (spiral) structure of DNA is formed by two strands of complementary nitrogenous bases that are joined by hydrogen bonds. Each side of the helix is comprised of deoxyribose sugar and phosphate molecules, known as the sugar–phosphate backbone.

THE STRUCTURE AND FUNCTION OF DNA

The function of DNA is to carry the inheritable genetic code for the control of cell activities and protein synthesis. You will learn how the genetic code is expressed to produce proteins in Module 5.2.

The primary structure of DNA is a single strand of polynucleotides, which consists of a specific sequence of nitrogenous bases (A, T, C and G). The opposing nitrogenous bases from two strands of DNA are joined by **hydrogen bonds** (Figure 5.1.4). These hydrogen bonds stabilise the secondary structure of the DNA and form the double-stranded helix (spiral) structure often referred to simply as a **double helix** (Figure 5.1.5).

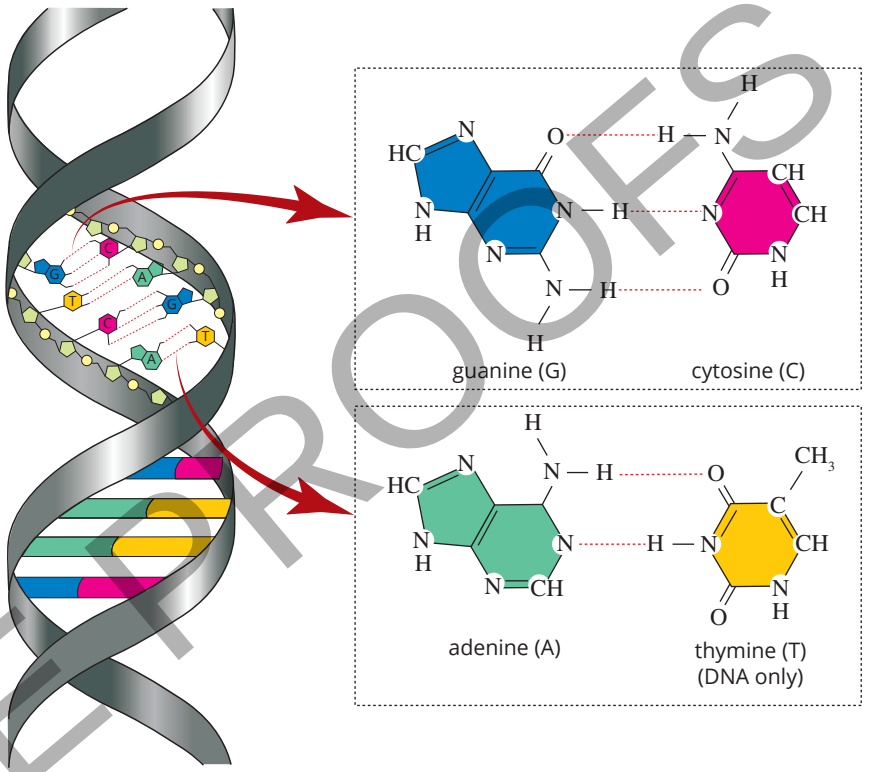


FIGURE 5.1.5 The helical structure of DNA. Two complementary strands form a double helix joined by base pairs guanine (G) and cytosine (C), and adenine (A) and thymine (T).

When DNA is uncoiled, the two strands can be represented as a ladder. The sides of the ladder consist of the sugar–phosphate backbone. The two strands of a DNA molecule run in opposite directions so they are **antiparallel**. The 3′ end of one strand matches up with the 5′ end of the other strand. The rungs of the DNA ladder are the nitrogenous bases of each nucleotide (Figure 5.1.6). The nitrogenous bases can occur in any order within the strand.

Chemical studies show the concentration of adenine in a DNA molecule is always equal to that of thymine, and that the concentration of guanine is always equal to that of cytosine. These observations are explained by a direct pairing between A and T and between G and C in the DNA molecule. This pairing of nitrogenous bases in DNA molecules is known as **complementary base pairing**.

In complementary base pairing:

- the purine adenine (A) always pairs with the pyrimidine thymine (T), held together with two weak hydrogen bonds
- the purine guanine (G) always pairs with the pyrimidine cytosine (C), held together with three weak hydrogen bonds.

The diameter of the DNA double helix is approximately 2.0 nanometres and there are approximately 10 pairs of nucleotide bases in each twist of the helix.

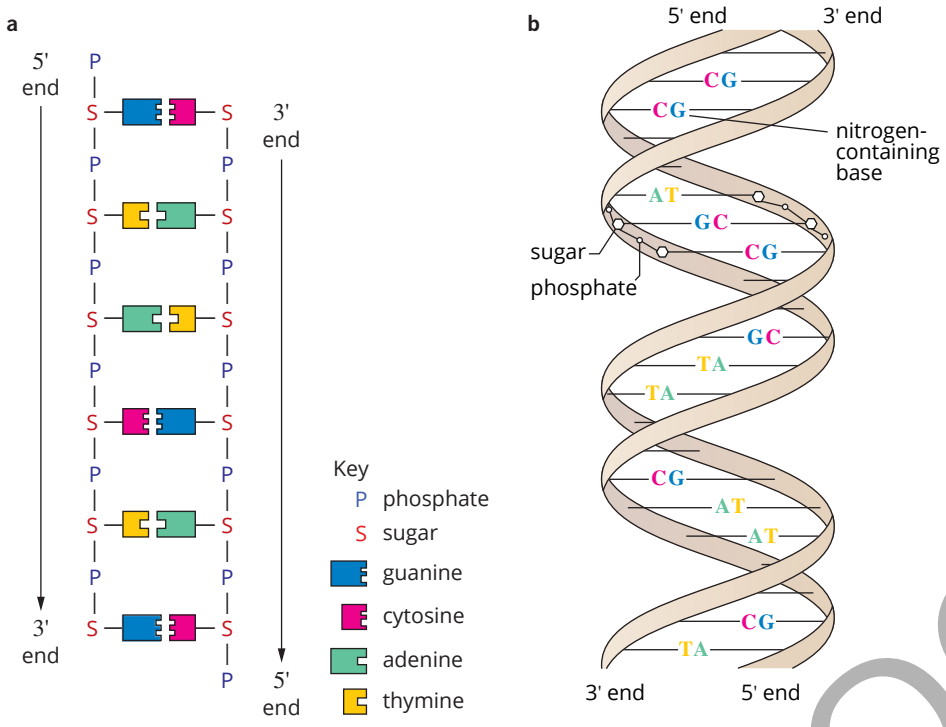


FIGURE 5.1.6 Two-dimensional representations of the DNA molecule. (a) Hydrogen bonds between the complementary base pairs hold the two antiparallel polynucleotide strands together. (b) The structure of DNA, showing the two complementary strands forming the double helix.

Eukaryotic and prokaryotic DNA

Eukaryotic cells have an average of 25 times more DNA and are more complex than prokaryotic cells (Figures 5.1.7, 5.1.8 and 5.1.9). The processes involved in DNA replication are much slower in eukaryotic cells because the cells are more complex—some bacterial cells take just 40 minutes to replicate their DNA, while in some animals this can take up to 10 hours.

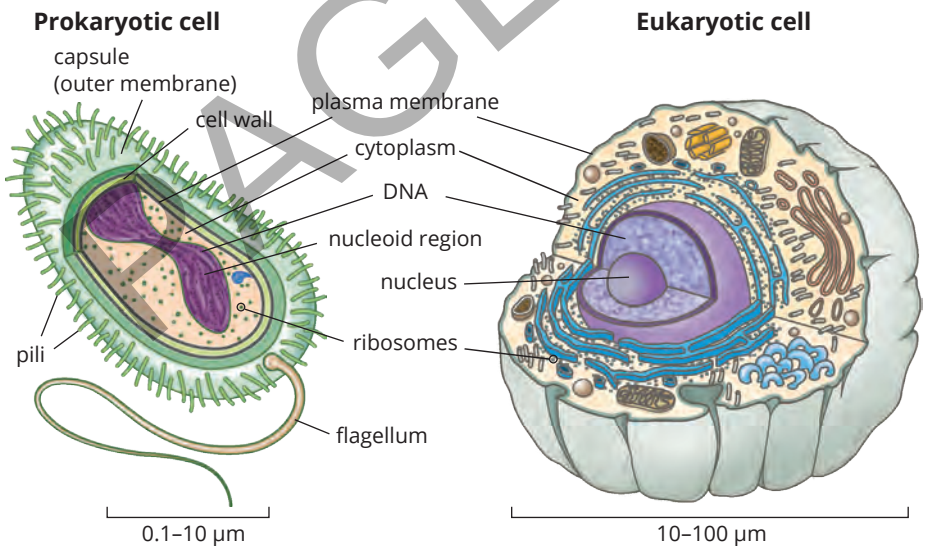


FIGURE 5.1.7 Prokaryotic cell compared to a eukaryotic cell. Prokaryotic cells and the processes involved in their DNA replication and gene expression are generally much simpler than in eukaryotic cells.

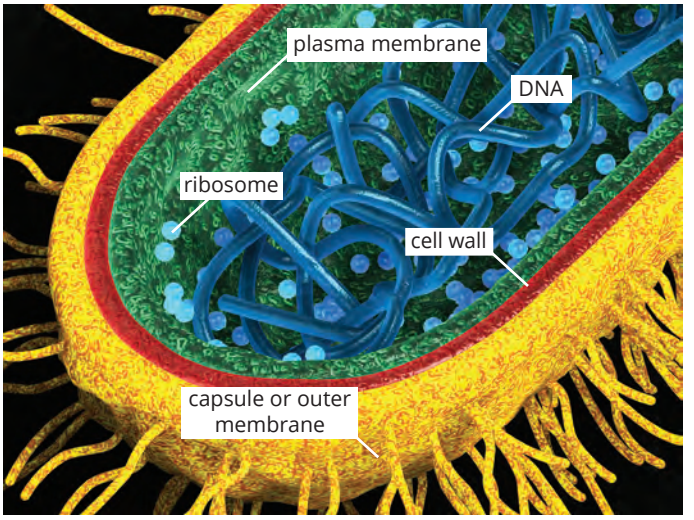


FIGURE 5.1.8 The inner structures of a prokaryotic cell. The pili (hair-like structures) and capsule are shown in yellow, the cell wall in red, the plasma membrane in green, the ribosomes in light blue and the DNA in dark blue.

FIGURE 5.1.9 The inner structures of a eukaryotic cell include the nucleus (centre), which has a membrane with nuclear pores (purple). Found inside the nucleus is the DNA (genetic material, white) and the nucleolus (blue).

DNA is found in the nucleoid region of prokaryotic cells and the nucleus of eukaryotic cells, where it is tightly coiled up to form structures called **chromosomes**. There is also mitochondrial DNA (mtDNA) in the mitochondria of all eukaryotic cells and chloroplast DNA (cpDNA) in the chloroplasts of photosynthetic eukaryotes.

CHROMOSOMES

Chromosomes are thread-like structures that serve as the storage and transmission units of the genetic information necessary for the growth, development, reproduction and functioning of an organism.

FIGURE 5.1.10 Coloured transmission electron micrograph of plasmids from *E. coli* bacteria. Plasmids are small, circular DNA molecules that are commonly found in bacteria.

Circular chromosomes

Circular chromosomes are composed of a DNA double helix where the ends of the DNA strand are covalently bonded to form a closed loop. Circular chromosomes store the genetic information in prokaryotes (e.g. bacteria and archaea), while eukaryotes have linear chromosomes. Prokaryotes have a single circular DNA chromosome that may be anchored to the cell membrane at a point called the origin. Some organelles of eukaryotic cells including mitochondria and chloroplasts also contain circular chromosomes, which reflects their evolutionary origins from ancient prokaryotes. Circular chromosomes are typically supercoiled, meaning that the DNA is twisted upon itself. Supercoiling means that the DNA is compacted, which is important due to the limited space in prokaryotes and organelles like mitochondria. In addition to the primary circular chromosome, some organisms have smaller, circular DNA molecules known as **plasmids** that carry extra genetic information (Figure 5.1.10).

Linear chromosomes

Linear chromosomes are structures that store genetic information in eukaryotic cells (Figure 5.1.11). Unlike circular chromosomes, linear chromosomes are generally larger and contain more non-coding DNA, including repetitive elements, and have a distinct linear shape with two ends, known as **telomeres**. In most somatic cells, chromosomes exist as single, continuous DNA molecules, not as duplicated structures. However, during the S phase (synthesis phase) of the cell cycle, chromosomes undergo DNA replication and form two identical halves called **chromatids**. These chromatids are joined at a region called the centromere, creating the duplicated chromosome structure often observed during cell division.

Table 5.1.1 summarises key similarities and differences between circular and linear chromosomes.

Supercoiling is the ability of DNA to twist upon itself in ways that impact its compaction and accessibility. Supercoiling is a key feature of circular chromosomes but also occurs in some parts of linear chromosomes (localised).

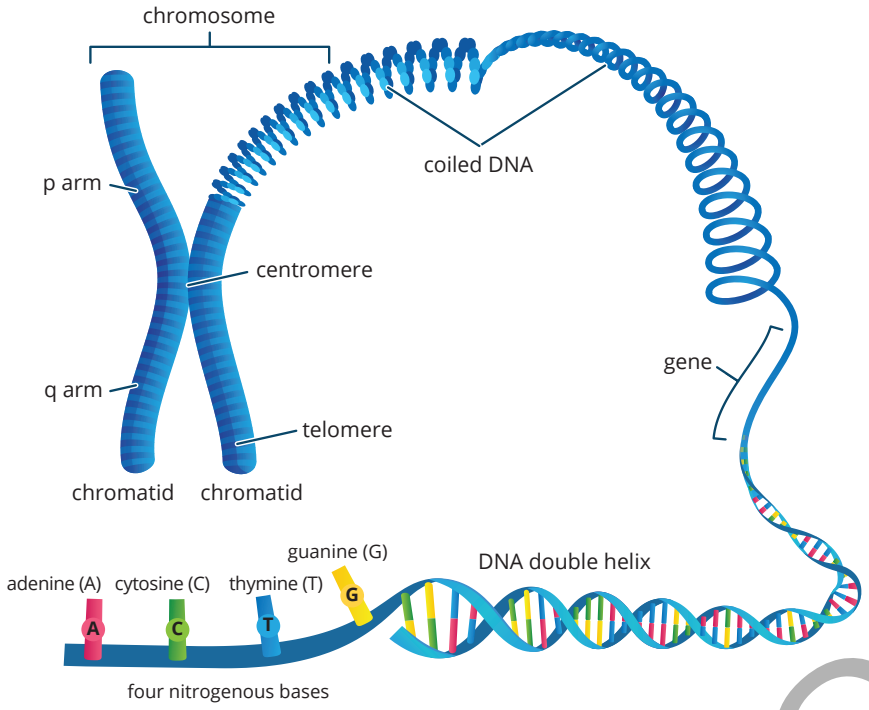


FIGURE 5.1.11 Linear chromosomes consist of two chromatids joined at the centromere. Each end of a chromatid is known as a telomere.

TABLE 5.1.1 Comparison of circular and linear chromosomes

Feature	Circular chromosomes	Linear chromosomes
shape	circular, closed loop	linear, with distinct ends called telomeres
location	prokaryotes / mitochondria and chloroplasts of eukaryotes	nuclei of eukaryotic cells
origins of replication	typically a single origin of replication	multiple origins of replication
size	generally smaller	generally larger
replication mechanism	bidirectional or rolling circle replication	bidirectional replication
presence of telomeres	absent	present
genome organisation	compact with less non-coding DNA	contains more non-coding and repetitive DNA
stability	more stable due to the absence of free ends	less stable; telomeres prevent degradation
supercoiling	common, to aid in compaction	occurs but is localised and less significant
examples	bacterial genomes, mitochondrial DNA, plasmids	human chromosomes, plant nuclear chromosomes

Subtypes of linear chromosomes

Each eukaryotic chromosome has a constriction point called the **centromere**, which divides the chromosomes into two sections. The regions above and below the centromere are referred to as the chromosome arms (Figure 5.1.12). The shorter arm is called the 'p arm' and the longer arm is called the 'q arm'. Photographs or diagrams of chromosomes are always arranged so that the p arm is at the top.

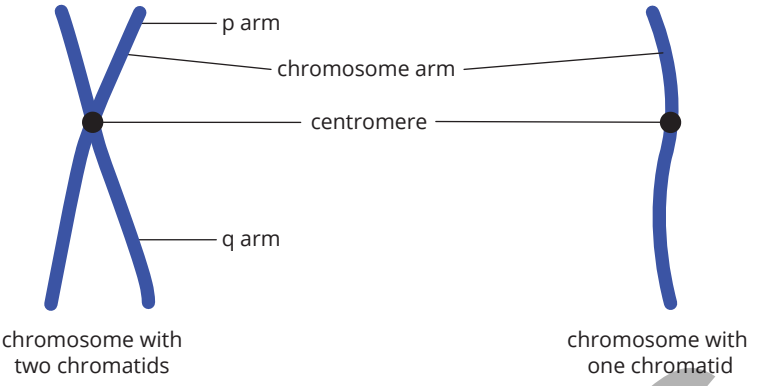


FIGURE 5.1.12 Parts of a chromosome.

There are four major types of chromosomes, based on the position of the centromere on the chromosomes as they appear during metaphase (Figure 5.1.13).

- Metacentric chromosomes have the centromere centrally positioned, giving arms of equal length.
- Submetacentric chromosomes have the centromere towards one end, resulting in arms of unequal length, with a long arm approximately twice the length of the short arm.
- Acrocentric chromosomes have the centromere very close to one end.
- Telocentric chromosomes have the centromere at the tip of the arms.

The names of the types of centromeres are derived from Greek:

- *meta* means in the middle
- *submeta* means below the middle
- *acro* means topmost, at the extremity
- *telo* means end, terminus.

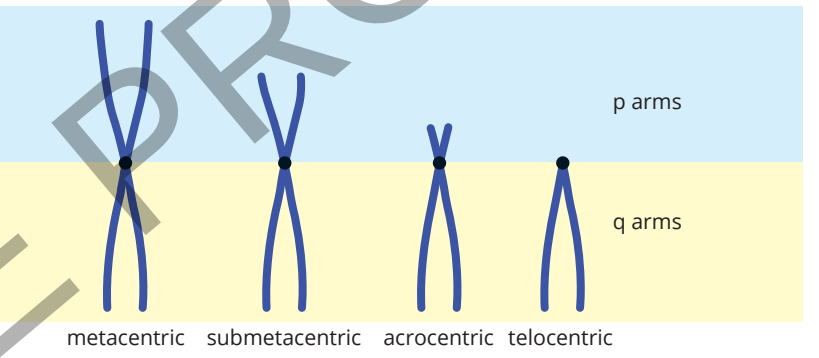


FIGURE 5.1.13 The four major types of chromosomes, named for the position of their centromere: metacentric, submetacentric, acrocentric and telocentric.

Structure of linear chromosomes

Chromosomes package DNA efficiently and protect it from enzymatic degradation. Each eukaryotic, linear chromosome consists of DNA wrapped around small proteins called **histones**, which contribute to localised supercoiling and help organise DNA into a compact structure ~10 nm in diameter called a **nucleosome**. Nucleosomes link together to form chromatin, which has the appearance of a string of beads, and higher order folding of chromatin forms ~30 nm diameter **chromatin** fibres (Figure 5.1.14).

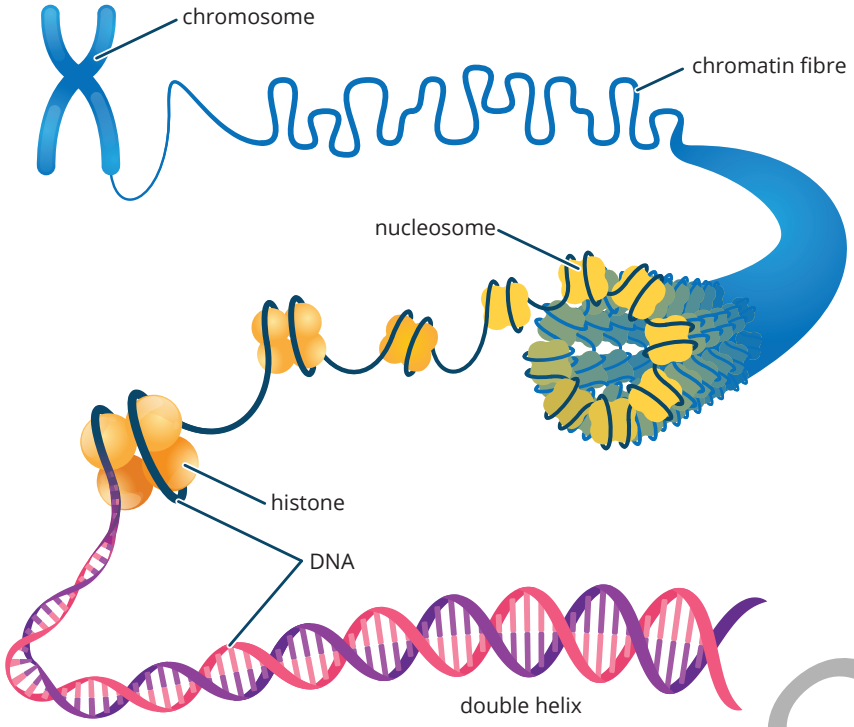


FIGURE 5.1.14 The unwound chromatid reveals the complex substructure that compacts DNA.

Chromatin fibres form chromatids through progressive folding and condensation steps that further compact the DNA to prepare for cell division (Figure 5.1.15). The 30 nm chromatin fibres are organised into **looped domains**, which are typically 50 000 to 200 000 base pairs long and are anchored by a protein scaffold that allows for more efficient packing. During the prophase stage of mitosis or meiosis, the looped domains are condensed (coiled and folded) into even tighter structures in a process mediated by proteins like condensin and cohesin, which stabilise the loops and hold **sister chromatids** together.

Sister chromatids are identical copies of a single chromosome produced during DNA replication.

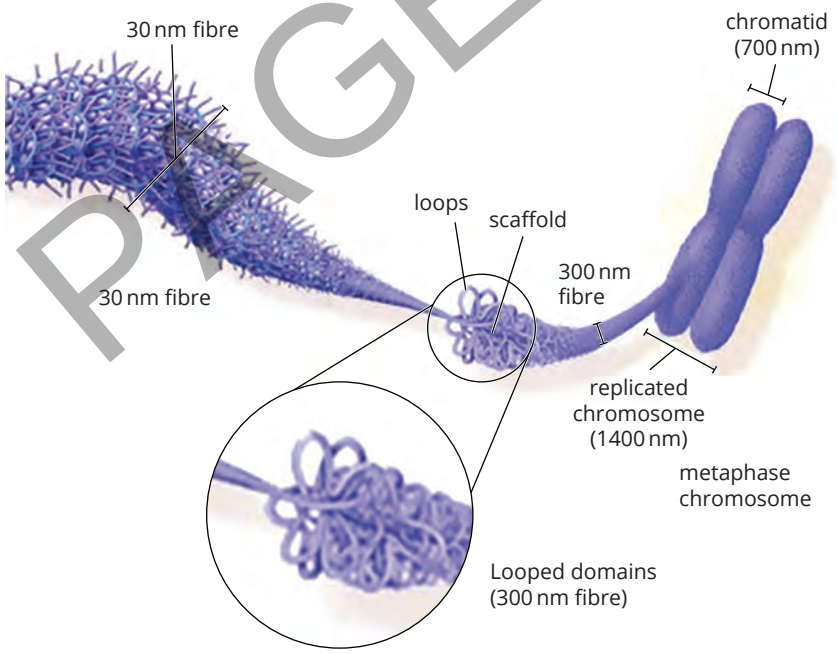


FIGURE 5.1.15 Chromatin fibres form looped domains held together by a protein scaffold.

Looped domains are held together by structural maintenance of chromosomes (SMC) proteins.

A gene is a DNA sequence that contains instructions for producing a functional product, located at a specific locus on a chromosome.

Alleles are all the variant forms of a particular gene.

Homologous chromosomes

Homologous chromosomes (or homologues) are a pair of chromosomes, one inherited from the biological mother and one inherited from the biological father (e.g. chromosome 1 from the mother and chromosome 1 from the father). Homologous chromosomes carry the same genes at the same locations (loci) but possibly have different alleles.

Since homologous chromosomes originate from different parents (maternal and paternal), their chromatids are inherently **non-sister chromatids** and may carry different alleles for the same genes. By comparison, sister chromatids are genetically identical and come from the same chromosome. Sister chromatids separate during mitosis and meiosis II, whereas non-sister chromatids are involved in crossing-over during meiosis I to exchange genetic material. Figure 5.1.16 shows chromosomes at metaphase. Each pair of chromosomes are said to be homologues because they contain the same gene sets.

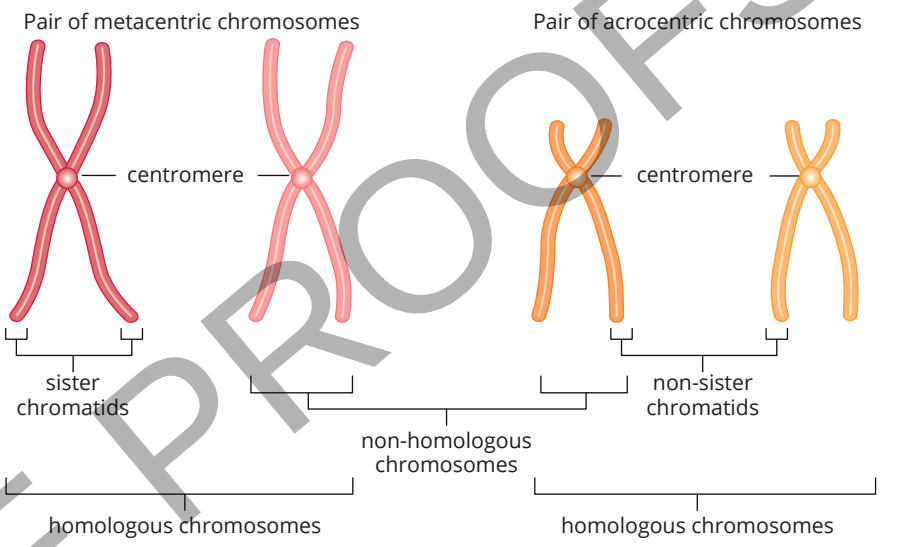


FIGURE 5.1.16 Homologous and non-homologous chromosomes.

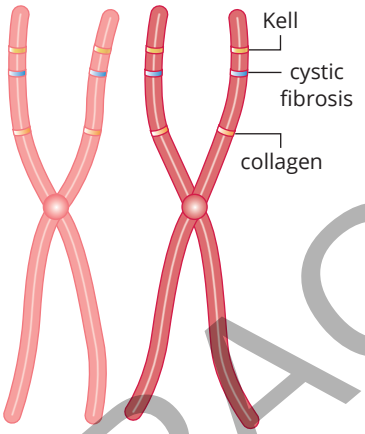


FIGURE 5.1.17 The two human chromosome 7 homologues during metaphase. The locus for three genes is shown: a collagen gene, the cystic fibrosis gene and the Kell gene.

The Kell gene is located on chromosome 7 and encodes the Kell glycoprotein, which is a key antigen found on red blood cells. The Kell blood group system includes various antigens, the most significant being K (K1) and k (K2), which play an important role in blood transfusions and immune responses.

Humans have 46 chromosomes: 23 inherited from the biological mother and 23 inherited from the biological father. Forty-four of these chromosomes form 22 matching pairs. The other two chromosomes are the sex chromosomes. The sex chromosomes for all females are homologues because females have a matching pair of X chromosomes. The sex chromosomes for males are non-homologous because males have an X and a Y chromosome, which contain different gene sets. Nevertheless, in most mammals, the X and Y chromosomes behave as a homologous pair during meiosis because some small regions of these chromosomes are homologous.

In Figure 5.1.17, two human chromosome 7 homologues are shown during metaphase. The loci for three genes are shown: a collagen gene, the cystic fibrosis gene and the Kell gene, which produces a protein involved in determining the Kell blood groups. The three genes are located in this same position on chromosome 7 in all cells in most individuals.

In actively dividing cells, there is a period during the cell cycle, after chromosome replication and before cytokinesis, when there are four copies of each gene in the diploid nucleus, but after cell division, cells have only two copies of each gene (i.e. two alleles for each gene, which may be identical or different, one of which has been inherited from each parent. Cell replication is explained in more detail in Chapter 6. Cells are not always actively dividing, and so most of the time non-dividing cells contain chromosomes that consist of s single strand (chromatid) (Figure 5.1.18).

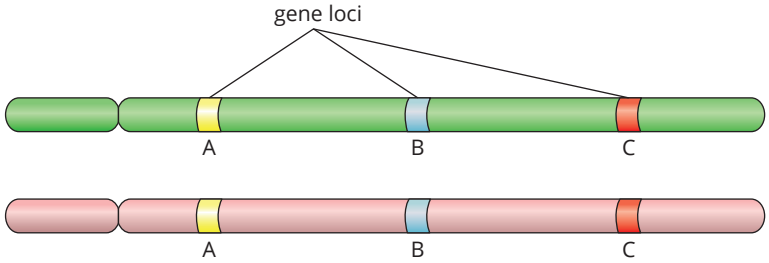


FIGURE 5.1.18 Chromosomes in non-dividing cells consist of one strand. Non-dividing cells have two copies of each chromosome and therefore two copies of each gene.

Sex chromosomes

Sex chromosomes (also called **allosomes**) are chromosomes involved in biological sex determination. Chromosomes that are not involved in sex determination are called **autosomes**. In humans and all other mammals, a pair of chromosomes known as the X and Y chromosomes determine the biological sex of an individual. Other types of organisms may have different types of sex-determining chromosomes (Table 5.1.2).

TABLE 5.1.2 Examples of sex determination in different organisms

Examples of organisms	Female	Male
humans, other mammals, fruit flies	XX	XY
birds, butterflies, strawberries	ZW	ZZ
grasshoppers, moths	XX	XO
plants	XX	XY

Individuals with two similar sex chromosomes are the **homogametic** sex. Individuals with different sex chromosomes are the **heterogametic** sex. Some organisms (such as fungi and algae) do not have sex-determining chromosomes and therefore do not have sexes; instead, they have 'mating types'.

GENOMES, GENES AND ALLELES

As chromosomes are composed of DNA, an organisms entire genome is contained within these chromosomes, including individual genes and their allele variants.

Genomes

The somatic cells (all cells in the body except gametes) of most sexually reproducing organisms are **diploid** ($2n$). This is because they contain two sets of chromosomes (one set from each parent). The DNA that makes up these chromosomes ranges in size from 50 million to 300 million base pairs.

The **genome** of an organism is the total of an organism's DNA measured in the number of base pairs contained in a **haploid** (n) set of chromosomes. The genome is measured in haploid cells because not all organisms have a diploid state. Some eukaryotic organisms, such as fungi and algae, are haploid organisms but will produce specialised diploid cells during sexual phases of their lifecycles. The genome of most prokaryotes is contained within one single circular chromosome, although some bacterial genomes consist of several different chromosomes. A typical human cell has two similar sets of chromosomes, and each set has DNA totalling 3234 million base pairs.

The human genome has approximately 3.234×10^9 DNA base pairs.

Genes

Recall that DNA is the molecule of life that encodes the information by which organisms are built. This information is stored as a sequence of nucleotides that encodes biological information.

A **gene** is a specific sequence of DNA that encodes for a functional product, namely a protein or functional RNA molecule such as tRNA (transfer RNA). When coding for a polypeptide, it is the sequence of nucleotides within a gene that contains the information for the protein to be synthesised. Genes can be millions of nucleotides in length. The longest known gene to date, which codes for the protein dystrophin, is 2.5 megabase pairs long (2.5×10^6 base pairs). Some of the shortest genes are fewer than 300 base pairs long.

The fixed position of the gene on the chromosome is called the **locus**. A DNA molecule consists of many genes, and these genes determine the characteristics of an organism.

Each gene carries a particular instruction, for example, how to make silk for a spider web (Figure 5.1.19). The web is made of silk fibroin, which is a structural protein. The genetic information on how to make this protein is in one gene called the silk fibroin gene. The process by which the information in a gene is decoded (in this case, to assemble silk fibroin) is called **gene expression**. You will learn about gene expression and regulation in Module 5.2.

Gene expression is the process by which the information stored in a gene is used to synthesise a functional gene product (protein or RNA).

FIGURE 5.1.19 The silk fibroin gene carries instructions for making the silk to create a spider web.

Nearly all genes specify the production of a polypeptide chain or protein that performs essential functions in the body's cells. For example, the gene that codes for salivary amylase (which produces sugars from starch) is called the *AMY1* gene and is located on chromosome 1. The protein **haemoglobin** in red blood cells carries oxygen (Figure 5.1.20). It is formed by linking four sub-units of polypeptides, two β-globin molecules and two α-globin molecules. The genes that code to produce functional haemoglobin sub-units are called *HB-B*, located on chromosome 11, and *HB-A1*, located on chromosome 16.

Proteins control cellular functions and genes control the production of proteins; therefore, inherited genes ultimately govern the functions of organisms. The length of the sequence of DNA and the precise order of the base pairs in a gene are the critical factors that determine what the gene product will be like and what it will do in a cell.

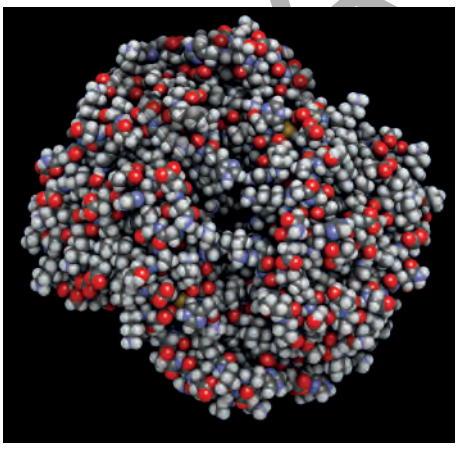


FIGURE 5.1.20 The information found in the genes *HB-B* and *HB-A1* is used to produce haemoglobin, a protein that transports oxygen around the body in red blood cells.

Alleles

Alleles are all the variant forms of a particular gene. For example, a population might have a gene for eye colour with different alleles for brown, blue or green eyes. At the molecular level, an allele is a variant of a gene distinguished by its different nucleotide sequence at the same gene locus. Somatic cells of a diploid organism contain two alleles for every gene; one allele for each gene in an organism is inherited from each biological parent.

You might wonder how there can be identical alleles when alleles are defined as variants of the same gene. Alleles are referred to as variants because across a population there are often multiple versions of a gene. However, within an individual, if both inherited copies are the same version (i.e. the same nucleotide sequence) then the alleles are identical, and the individual is **homozygous** for that gene. By comparison, if an individual has two different versions (i.e. different nucleotide sequences) of an allele at a gene locus, then that individual is **heterozygous** for that gene.

Alleles give rise to variation in physical characteristics (such as eye colour, height and the ability to roll your tongue) within populations and between individuals. This variation occurs because an individual gene that is responsible for a specific trait will have different alleles. For example, the *TAS2R38* gene found on chromosome 7 has two common alleles that result in variation in the ability to taste phenylthiocarbamide (PTC), a chemical detected by the bitter-taste receptors on the tongue and found in many common foods and drinks like coffee and bitter-tasting vegetables (Figure 5.1.21).

The shape of the tongue's bitter-taste receptor determines how strongly PTC will bind to it, and the stronger it binds the more a person is able to taste the chemical. One allele for *TAS2R38* gene is known as a tasting allele because it codes for a receptor protein that strongly binds PTC. The other is known as a non-tasting allele because it codes for a receptor protein with a different shape that results in binding to PTC that is weak by comparison. Both alleles code for a taste receptor protein on the tongue and both are found in the locus on chromosome 7, but differences in the DNA sequence of bases for these two alleles impacts the structure of the taste receptor protein and has an impact on taste as a result.

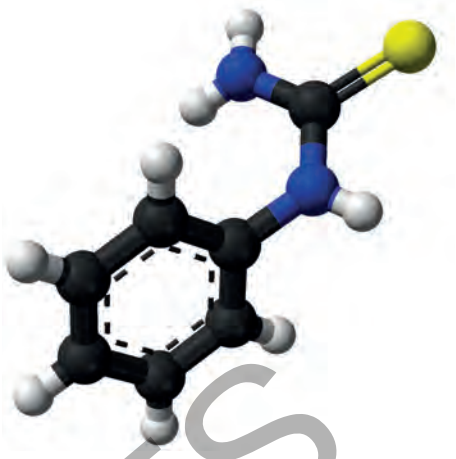


FIGURE 5.1.21 The chemical structure of phenylthiocarbamide (PTC) is similar to other toxins found in poisonous plants.

The genetic code

The **genetic code** is a universal set of rules that defines how the information in nucleic acids (DNA and RNA) is translated into proteins and functional RNA molecules for all organisms on Earth. The information in DNA and RNA is stored as a three-letter code of nucleotides. In DNA, this three-letter code is called a **triplet**. When a DNA triplet is transcribed into mature mRNA (messenger RNA), the triplet is called a **codon**. Each triplet or codon codes for one amino acid (Figure 5.1.22), and may also provide specific instructions, such as 'start translation' and 'stop translation'. These amino acids form polypeptide chains and the polypeptide chains form proteins.

The set of alleles present in the DNA of an individual organism is known as the organism's genotype.

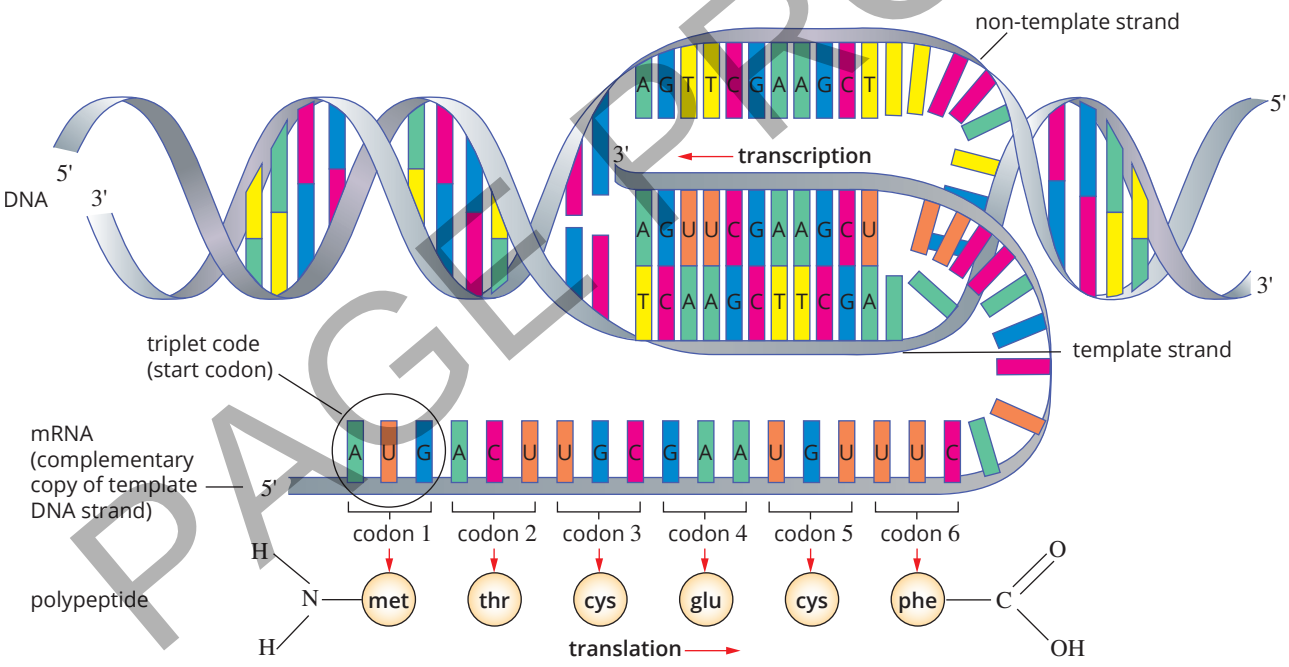


FIGURE 5.1.22 Each codon codes for a particular amino acid.

Messenger RNA is produced during transcription and then translated to produce an amino acid chain (polypeptide).

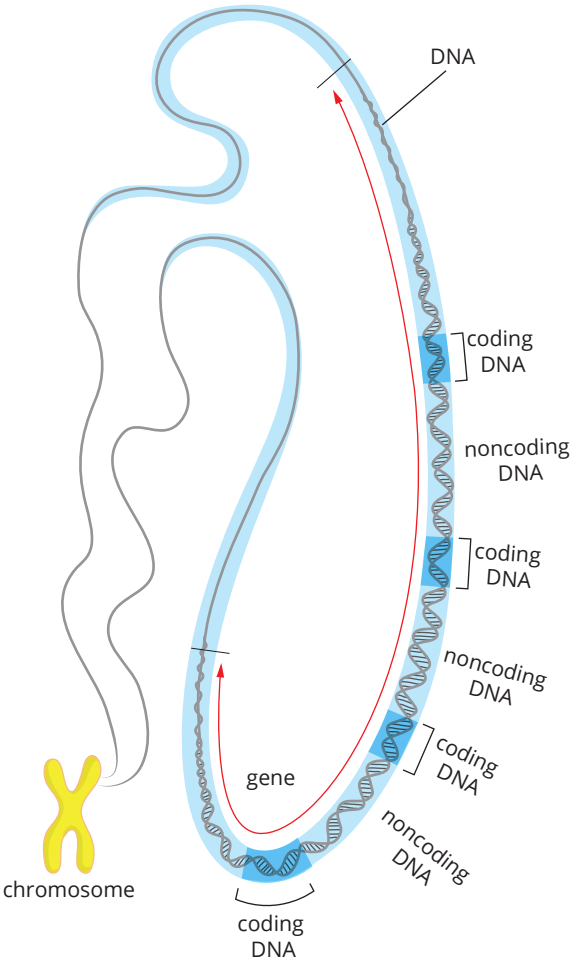


FIGURE 5.1.23 The location of non-coding DNA sequences.

Non-coding DNA

Genes consist of regions of DNA that can either code for amino acids or that are **non-coding DNA** sequences that do not code for proteins. Non-coding DNA sequences are also found between genes (Figure 5.1.23). Non-coding DNA, once misleadingly called 'junk' DNA, is now known to be important for **gene regulation**.

> Gene regulation refers to the mechanisms and processes that control the timing, location and amount of gene expression.

One type of non-coding DNA, called structural non-coding DNA, regulates the separation of chromosomes during mitosis by ensuring chromosome alignment, segregation and stability. (Mitosis is part of eukaryotic cell replication and will be explored in Chapter 6.) Structural non-coding DNA is found in the centromere and telomere regions of the chromosome. In telomeres, structural non-coding DNA consists of short repeating nucleotide sequences (in humans, it is TTAGGG) that help protect the DNA molecule from being eroded during cell replication. Without this telomeric DNA, the DNA strands become shorter and shorter with each mitotic division and, if repeated often enough, entire genes could be deleted. Structural non-coding DNA also interacts with DNA-binding proteins to form a protein structure at the centromere called the kinetochore complex, which serves as the attachment point for spindle microtubules that pull sister chromatids apart during anaphase.

Degeneracy

The genetic code is said to be **degenerate** because more than one codon can code for the same amino acid. Differences in codons encoding the same amino acid usually occur at the second or third base. As the genetic code has four different nucleotides and three sequential nucleotides are read at the same time to code for an amino acid, there are 64 possible codons ($4^3 = 64$) available to code for the total 20 amino acids (Figure 5.1.24).

Second base of codon

First base of codon	U		C		A		G		Third base of codon
U	UUU	phenylalanine (Phe)	UCU	serine (Ser)	UAU	tyrosine (Tyr)	UGU	cysteine (Cys)	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine (Leu)	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	tryptophan (Trp)	G
C	CUU	leucine (Leu)	CCU	proline (Pro)	CAU	histidine (His)	CGU	arginine (Arg)	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	glutamine (Gln)	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	isoleucine (Ile)	ACU	threonine (Thr)	AAU	asparagine (Asn)	AGU	serine (Ser)	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lysine (Lys)	AGA	arginine (Arg)	A
	AUG	methionine (Met) START	ACG		AAG		AGG		G
G	GUU	valine (Val)	GCU	alanine (Ala)	GAU	aspartic acid (Asp)	GGU	glycine (Gly)	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	glutamic acid (Glu)	GGA		A
	GUG		GCG		GAG		GGG		G

FIGURE 5.1.24 The genetic code for the 20 amino acids and stop codons. Codon charts use uracil (U) instead of thymine (T) because they represent the codons found in mRNA (which uses uracil), not DNA (which uses thymine).

The structure of a gene

All eukaryotic genes have several structural features in common (Figure 5.1.25):

- start and stop triplets
- promoter regions
- exons
- introns.

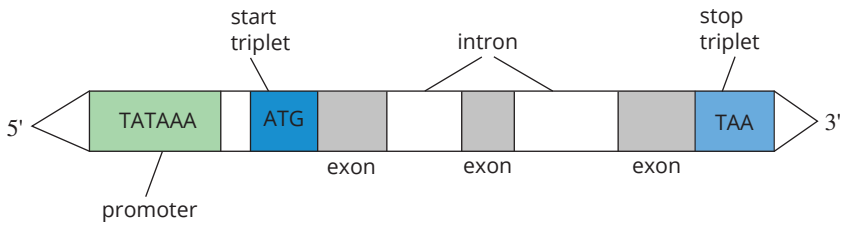


FIGURE 5.1.25 Eukaryotic genes have start and stop triplets, promoter regions, coding exons and non-coding introns.

Start and stop triplets

A start triplet indicates where the first stage of gene expression will begin. When transcribed into mRNA, the DNA triplet ATG becomes the start codon AUG. AUG is the most common start codon in mRNA. The start codon initiates **translation** and codes for the amino acid methionine. Most functional proteins start with methionine but there are rare exceptions. For example, a protein in the fungus *Candida albicans* uses GUG as a start codon.

A stop triplet indicates where **transcription** will end. The stop triplet does not code for an amino acid. When stop triplets are transcribed into mRNA, they become the codons UAA, UAG and UGA.

Promoter regions

Promoter regions are sections of a gene that are found before the start triplet, at the 5′ end of the site where transcription will begin. A promoter region:

- is the location where the **RNA polymerase** (the enzyme that initiates transcription) attaches to the gene
- identifies which DNA strand will be transcribed
- identifies where transcription of the gene will start
- identifies in which direction transcription will occur.

In many eukaryotic genes, the promoter region is coded for by the sequence of bases TATAAA, which is sometimes called the TATA box.

Exons and introns

Exons are regions of a gene that are usually expressed as proteins or RNA. Exons come together to make up mRNA, which is then translated into proteins. Together all of the exons form the **exome** of the cell or organism.

In eukaryotes, not all sections of a gene are translated. **Introns** (or spacer DNA) are a type of non-coding DNA that are transcribed into pre-mRNA but spliced out (removed) during mRNA processing. The functions of introns are not fully understood but they may contain regulatory elements or contribute to alternative splicing (see Module 5.2).

There are no rules governing the number of exons and introns in a gene. For example, 99% of the length of the dystrophin gene is made up of introns, but the gene for the protein insulin consists of three exons and two introns.

DNA REPLICATION

DNA is passed on from the parent cell to the two daughter cells that are produced when the parent cell divides. To transmit an exact copy of the DNA, without losing any instructions, cells need a mechanism for accurately copying DNA. This mechanism occurs during the S phase (synthesis phase) of interphase. It is called **DNA replication**.

Most of the DNA in eukaryotic cells is arranged in chromosomes in the nucleus. During replication, the double-stranded DNA separates into two **template strands** and a new complementary strand is synthesised onto each template strand (Figure 5.1.26).

DNA replication is the process used by a parent cell to make an exact copy of its DNA. DNA replication is essential for cell division, as both daughter cells need a complete set of genetic information to function.

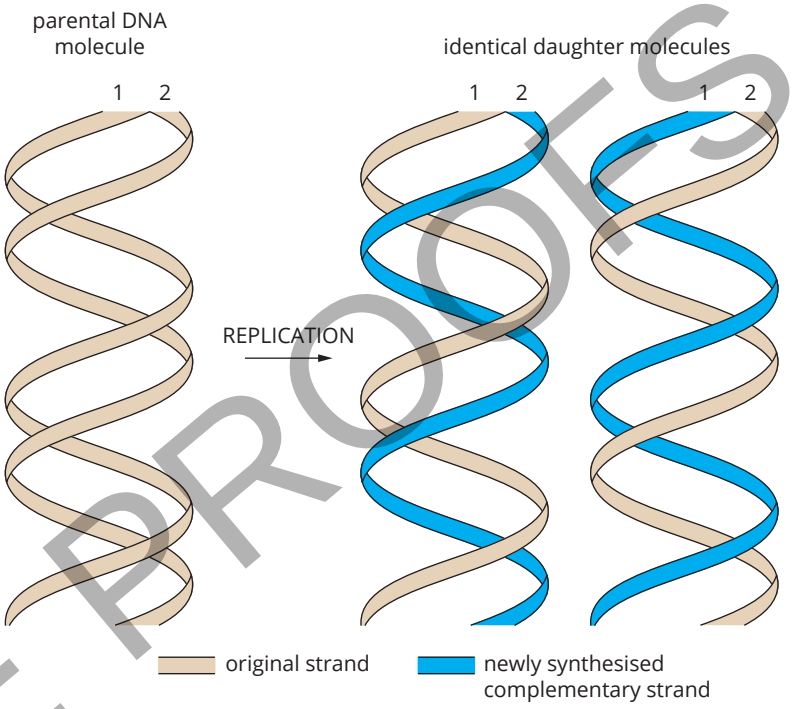


FIGURE 5.1.26 Each strand of the parent DNA molecule acts as a template for the synthesis of a new strand.

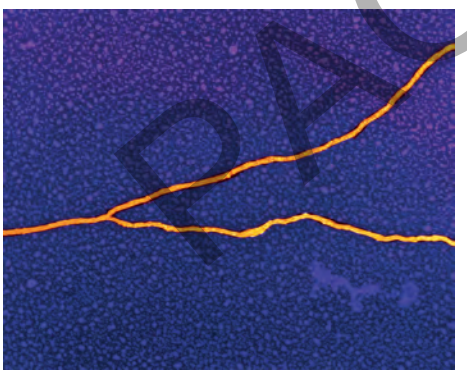


FIGURE 5.1.27 A coloured transmission electron micrograph of human DNA showing the replication fork.

Separation of the parent double strand of DNA into two template strands is achieved by an enzyme called **DNA helicase**. DNA helicase unwinds and then 'unzips' the parent DNA by breaking the hydrogen bonds between the complementary base pairs. As DNA helicase unzips the DNA, it forms a **replication fork** (Figure 5.1.27).

At the replication fork, a new polynucleotide chain is synthesised by adding complementary nucleotide units according to the sequence of the template strand. To make the new strand, new nucleotides pair up with the template strand, using the rules of complementary base pairing. A complementary strand is formed because adenine (A) always pairs with thymine (T), and cytosine (C) always pairs with guanine (G). Each daughter DNA molecule produced in this process consists of one old strand from the parent DNA and one newly-synthesised strand, so DNA replication is described as semi-conservative.

The enzyme responsible for adding the nucleotides in the correct sequence is called **DNA polymerase**. DNA polymerases rarely add the wrong nucleotides, but if this does happen there are proofreading and repair enzymes to correct the error. Mutations only occur when these backup systems fail.

DNA replication is an extremely accurate process with three distinct phases (Figure 5.1.28).

1 The parent DNA molecule unwinds, 'unzips' and a replication fork forms, achieved by the action of the enzyme DNA helicase.
2 The DNA is copied, with complementary nucleotide bases (A–T and G–C) attaching to the parent strands. Nucleotides are added by the enzyme DNA polymerase.
3 Replication results in two identical strands. One was the template strand and the other is a newly synthesised complementary strand.

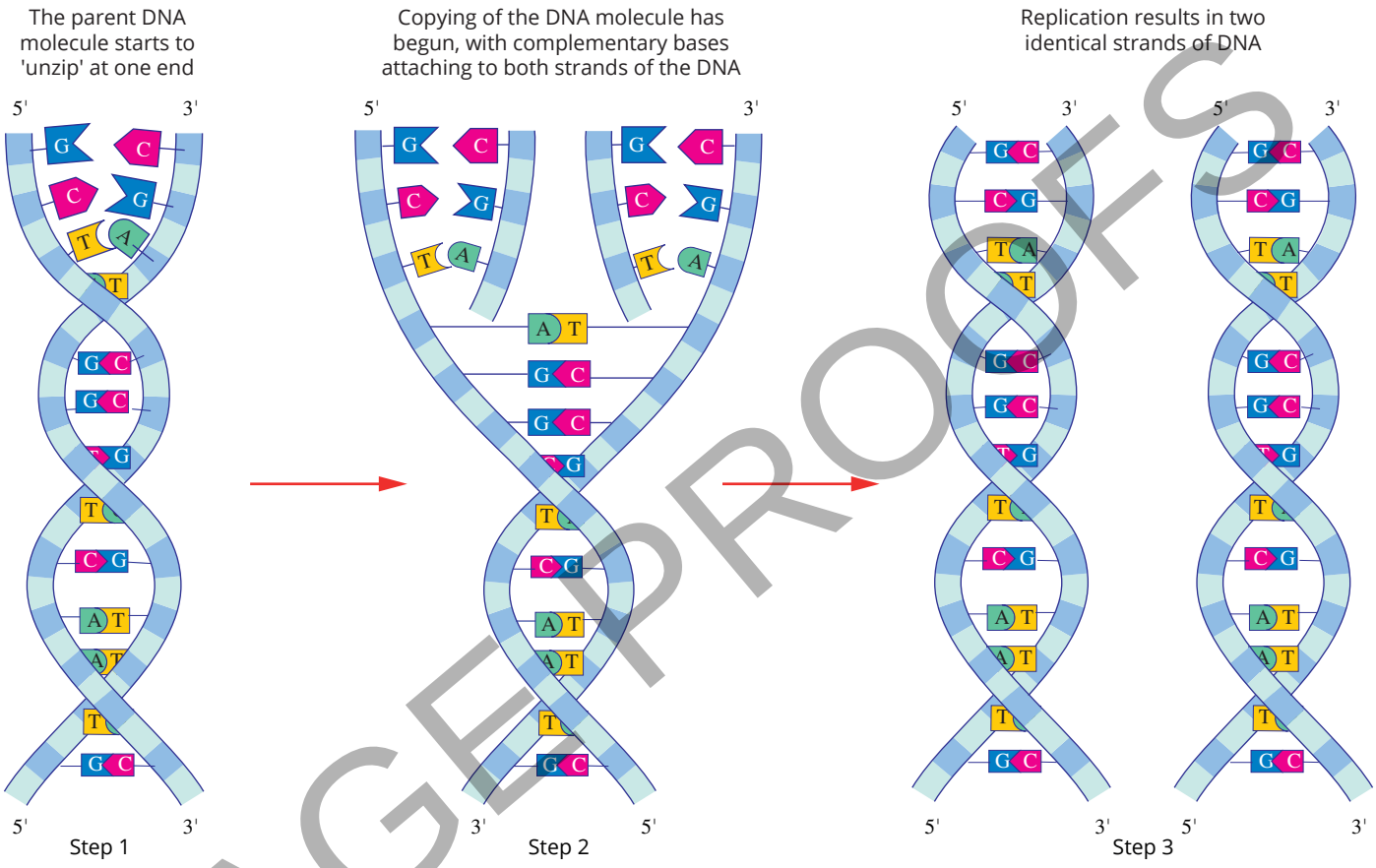


FIGURE 5.1.28 DNA replication involves three distinct phases.

Direction of DNA replication

Recall that DNA is made up of two antiparallel polynucleotide chains (strands). DNA polymerase can only add nucleotides to the 3′ end of the growing strand. This means that one strand is being replicated in the same direction as the replication fork is moving. This strand is called the leading strand because its synthesis is less complex and occurs more quickly than its complement. The other strand is called the lagging strand. The lagging strand (with its orientation being antiparallel and the nucleotides being added to the 3′ end) is synthesised in the opposite direction to the movement of the replication fork.

Once synthesis begins at the replication fork, the fork moves on. This leaves a gap on the lagging strand that requires another DNA polymerase to attach and begin replication again (Figure 5.1.29). Replication on the lagging strand is discontinuous and results in short sections of newly synthesised DNA called Okazaki fragments, which were discovered by Dr Reiji Okazaki. The discontinuous fragments are eventually joined to produce a continuous strand of DNA.

DNA replication is directional. The leading strand is synthesised in the same direction as the replication fork. The lagging strand is synthesised in the opposite direction, which results in small segments of DNA known as Okazaki fragments.

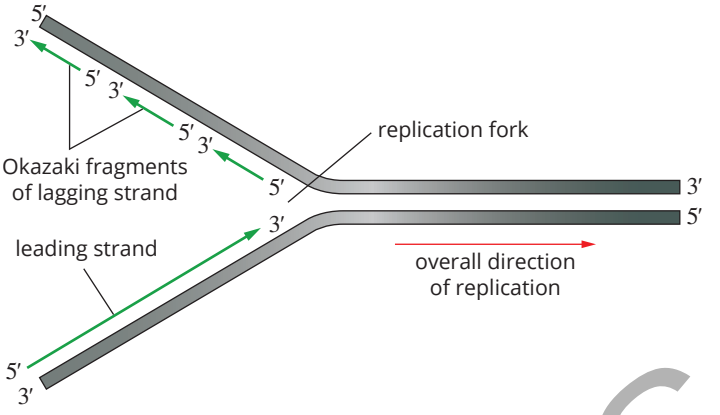


FIGURE 5.1.29 DNA replication showing the replication fork, the direction of replication and the leading and lagging strands. The lagging strand is made up of Okazaki fragments.

MUTATIONS AND MUTAGENS

Mutations are changes in the sequence of DNA nucleotides that may affect a single gene, multiple genes or entire chromosomes. Once formed, a mutation is established in the DNA unless changed by another mutation. Cytosine-to-thymine mutations occur more often than any other. Mutations can arise from random errors in DNA replication or damage from environmental factors.

Spontaneous mutations

Spontaneous mutations arise naturally as random changes in the base sequence of DNA. They may occur because of rare, undetected and unrepaired errors of DNA synthesis. Spontaneous mutations occur at an average rate of approximately one in a million; that is, an error occurs in a gene about once in every million replicated base pairs.

Induced mutations

The rate of mutation can be increased above spontaneous (background) levels when cells are exposed to biological, chemical or physical **mutagens** (Figure 5.1.30). Mutations that occur following deliberate or accidental exposure to mutagens are called **induced mutations**.

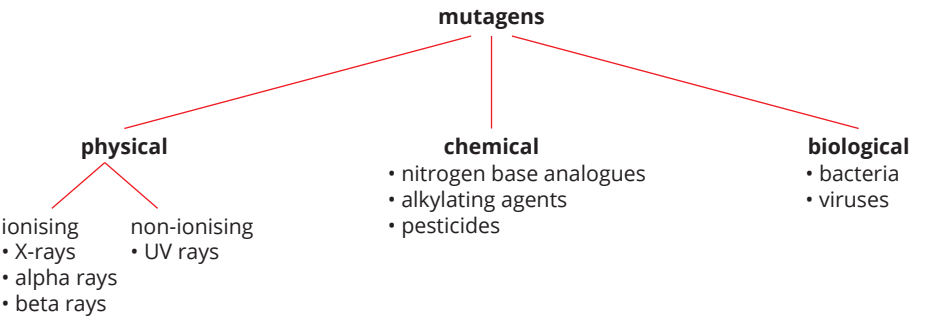


FIGURE 5.1.30 Types of mutagens.

Biological mutagens, such as viruses and bacteria, change the genetic composition of the cells by integrating their DNA into the human genome during cell division. This can lead to changes in the normal functioning of the genes.

Chemical mutagens can interfere with correct DNA replication by inserting themselves into the DNA and distorting the double helix. Chemicals that are chemically similar to the nitrogen bases can substitute into the DNA strand and result in faulty base pairing, misreading of the DNA sequence and disruption of the replication machinery. Other chemical mutagens may modify the nitrogen

bases by deamination or alkylation processes, which remove or transfer parts of these molecules. These processes significantly disrupt the regular base pairing and introduce frameshift mutations, which you will learn about in Module 5.2. Some pesticides are chemical mutagens, and their resulting base pair changes have been linked to neurodegenerative diseases such as Parkinson's disease.

Since the industrial revolution, and more dramatically in recent decades, the number of chemicals added to the environment because of agricultural and industrial practices and warfare has increased. Some of these chemicals are mutagens (which cause changes in DNA), some are **carcinogens** (which promote cancer) and some are both. Carcinogens often lead to increased cell division and uncontrolled growth of affected tissues, although they do not always act by directly altering genetic material. If a carcinogen does cause genetic changes, then it is also a mutagen. Mutations in certain genes called proto-oncogenes can turn these genes into oncogenes, which promote cancer. A mutagen that induces mutations in these genes may also act as a carcinogen by driving cancer development.

Radiation as a mutagen

Physical mutagens include forms of ionising and non-ionising radiation. Radioactive particles emit sufficient energy to disrupt the atoms or molecules in the DNA structure.

Forms of radiation that are encountered regularly are UV radiation (via sunlight), ionising radiation (such as via X-rays used in medical diagnosis) and high-energy radiation from radioactive elements. Normally the mutagenic effects of radiation are relatively minor. However, overexposing your body to UV radiation—for example, by sunbaking at the beach—can result in mutagenesis and skin cancer.

Public awareness of the risk of overexposure to UV radiation has increased dramatically in Australia in recent times. The Sun Smart campaign has had a significant influence. UV radiation may damage DNA, causing bases of the same strand to join, or it may cause the loss of a base. These changes cause distortions to the DNA strand and when DNA replication takes place, the distortions interrupt it, leaving a gap in the strand that is synthesised, which can cause cancers like melanoma (Figure 5.1.31).

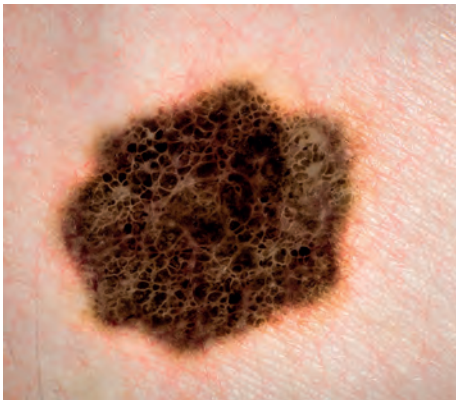


FIGURE 5.1.31 Malignant (life-threatening) melanoma skin cancer.

Radiologists wear lead-lined protective clothing to reduce their long-term exposure to X-rays when treating patients. The protective equipment covers the genital regions, to reduce the chance of causing mutation to the DNA in the germline cells (eggs or sperm). X-rays, which have a higher energy than UV radiation, can break DNA and even cause chromosome breakage, resulting in cell death or mutational change. Radiation from radioactive elements can have similar effects. Both spontaneous and induced mutations may result in similar genetic changes.

Heat as a mutagen

Recall that DNA is a long-chain polymer of nucleotides linked by phosphodiester bonds between the deoxyribose sugar molecules and the phosphate groups, and that in addition to the deoxyribose sugar and phosphate group, each nucleotide also contains a nitrogenous base. The chemical bond that links the nitrogenous base to the sugar molecule is called a β-glycosidic bond.

When DNA is exposed to heat, the bond between the deoxyribose sugar and the nitrogenous base can be broken, which results in the nitrogenous base being removed from the nucleotide, leaving a 'baseless' site in the DNA strand. This resulting sugar–phosphate only (SP) site is unstable and rapidly degrades leaving a gap in the DNA template. It is thought that as many as 10 000 SP sites are produced in each human cell each day. Despite this, heat mutagenesis is not a common form of mutation because cells have highly effective repair mechanisms in place. However, heat mutagenesis may be responsible for the decline in male fertility, as a result of testes being exposed to increased temperatures, affecting spermatogenesis and the production of viable sperm.

DNA repair

All cells have DNA repair systems, which are important to cell survival and frequently called into action. Cells have enzymes that can repair radiation damage to DNA. If damage to the template strand of a DNA molecule is not repaired, the cell dies. In most cases, DNA repair occurs during replication by randomly incorporating bases into a gap (Figure 5.1.32). If the base incorporated is not the same as the missing base, this results in a mutation in the synthesised strand.

There are also back-up repair systems that are used when other repair mechanisms fail. For example, bacteria have a repair system that acts as an effective last resort when there has been large-scale damage to DNA.

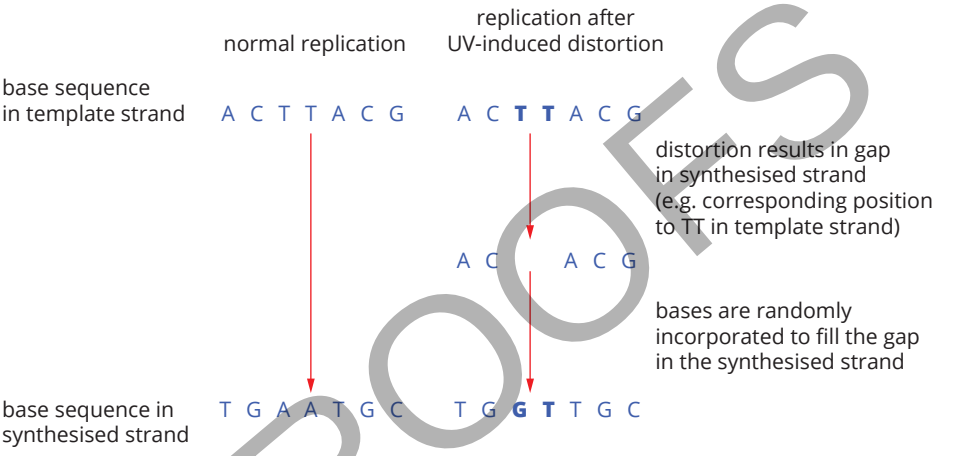


FIGURE 5.1.32 UV damage to DNA may result in a gap in the DNA strand that is filled in by randomly incorporated bases. In this case, a mutation occurs in the synthesised strand. The base sequence during normal replication is shown for comparison.

5.1 Review

SUMMARY

- Nucleotides (which consist of a five-carbon sugar, a phosphate and a nitrogenous base) are joined in a condensation polymerisation reaction to form polynucleotides and the two nucleic acids (DNA and RNA).
- DNA is a long, coiled, double-stranded nucleic acid that forms a double helix.
  - Each strand of DNA contains nucleotides that are made up of deoxyribose sugar, a phosphate and one of four nitrogenous bases (adenine, cytosine, guanine and thymine).
  - The two strands of DNA are joined by complementary base pairing between the nitrogenous bases. Adenine joins with thymine by two weak hydrogen bonds, while cytosine joins with guanine by three weak hydrogen bonds.
  - The two strands of DNA are antiparallel. One runs in the 5′ to 3′ direction, while the other runs in the opposite direction, from 3′ to 5′.
- In DNA, three-letter codes of nucleotides known as triplets are transcribed into mature mRNA (messenger RNA). Triplets are known as codons in mRNA.
- Chromosomes store, organise and transmit genetic information. Circular chromosomes are found in prokaryotes and some eukaryotic organelles, whereas linear chromosomes are found in the nucleus of eukaryotic cells.
- Genes on homologous chromosomes occur at the same loci but may have different alleles.
- Features of eukaryotic genes are start and stop triplets (three-letter nucleotide codes that indicate where transcription starts and stops), promoters (where RNA polymerase attaches to begin transcription), exons (coding DNA) and introns (non-coding DNA).
- Mutations are changes in the DNA sequence that result from errors in DNA replication or DNA damage caused by biological, chemical or physical mutagens.

KEY QUESTIONS

Describe

1 Nucleic acids are polymers that consist of repeating monomer units of chemical building blocks.
- **a** Recall the name given to this group of chemical building blocks.
- **b** Identify the three basic components of their structure.
- **c** Illustrate the structure with a suitable schematic diagram.

2 Recall:
- **a** the primary and secondary structure of DNA
- **b** the function of DNA
- **c** the structure of a nucleosome.

3 Define:
- **a** DNA
- **b** genome
- **c** gene
- **d** allele
- **e** DNA replication
- **f** mutation.

Apply

4 Why is a ladder often used as a metaphor for the structure of DNA? Refer to the specific components of the nucleotide in your response.

5 Explain the meaning of the term ‘complementary’ with reference to the strands of the DNA molecule. Include a diagram to illustrate your explanation.

6 One strand of DNA has the sequence ATTCCGTA. Write the sequence of the complementary strand.

continued over page

5.1 Review *continued*

7 Copy this table and complete it by assigning 'purine' or 'pyrimidine' to each nitrogenous base and filling in its complementary base in DNA.

Base	Purine or pyrimidine structure	Complementary base	Complementary base purine or pyrimidine structure
adenine			
guanine			
cytosine			
thymine			
uracil			

8 Explain how the antiparallel DNA strands ensure the stability of the secondary structure of the molecule.

9 Distinguish between the following terms:
- **a** autosome and allosome
- **b** homologous and non-homologous chromosomes.

10 Why is the genetic code described as universal?

Analyse

11 The following table shows the composition of bases in the cells in a variety of species.

Species	Thymine (%)	Adenine (%)	Guanine (%)	Cytosine (%)
humans	30.1	30.4	19.6	19.9
cattle	28.7	29.0	21.2	21.2
salmon	29.1	29.7	20.8	20.4
wheatgerm	27.4	28.1	21.8	22.7
E. coli	23.6	24.7	26.0	25.7
sea urchin	32.1	32.8	17.7	17.3

Argue whether or not this evidence supports the existence of complementary base pairs.

12 The mutation rate induced by different mutagens depends on several factors including the type and dose of the mutagen, the DNA repair mechanisms of the organism, and the specific genomic region affected. The following table compares the effects of different mutagens on mutation rates in DNA.

Mutagen	Mutations per million base pairs
control (no mutagen)	0.3
UV radiation	6.5
benzene	3.2
ethyl methanesulfonate	7.5

Use your understanding of mutations and the data available to answer the following questions.
- **a** Rank the mutagens from greatest to least impact on mutation rate.
- **b** Why does the control group have a mutation rate?
- **c** Benzene and ethyl methanesulfonate are both chemical mutagens. Why are their mutation rates so different?

13 The following table shows the frequency of thymine dimers detected in skin cells after varying durations of UV exposure. Thymine dimers are bonds between adjacent thymine nucleotides on the same DNA strand.

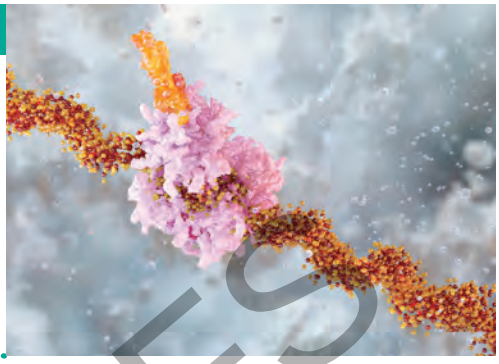
Duration of UV exposure (min)	Average thymine dimers per 1000 base pairs
0	0
5	3
10	8
20	15
30	25

Use your understanding of mutations and the data available to answer the following questions.
- **a** Describe the trend in thymine dimer formation as UV exposure increases.
- **b** What potential consequences could these thymine dimers have on DNA replication?

5.2 Protein synthesis and gene regulation

BY THE END OF THIS MODULE, YOU SHOULD BE ABLE TO:

- understand the structure and function of RNA as compared to DNA
- recognise that the purpose of gene expression is to synthesise a functional product such as a protein or RNA
- explain the process of protein synthesis in eukaryotes
- identify factors that regulate the expression of genes
- distinguish between structural and regulatory genes and recall examples of regulatory transcription factors.



In the previous module, you learned that the **genome** is the complete set of DNA present in an organism and that the genetic information stored in DNA nucleotides and read as triplet codes is used in **protein synthesis** to synthesise the amino acid sequences that form proteins. This process is called gene expression.

In this module, you will learn about the different steps of gene expression and how RNA plays a major role in protein synthesis. You will also learn about different types of mutations and the effects they can have.

THE STRUCTURE AND FUNCTION OF RNA

Ribonucleic acid (RNA) is typically single-stranded but can adopt double-stranded regions or even be fully double stranded in certain circumstances. For example, some viruses (e.g. rotaviruses) have double-stranded RNA (dsRNA) genomes. RNA molecules are usually much shorter than DNA molecules. The nucleotides of RNA have the same basic structure as those of DNA, with a few differences (Figure 5.2.1). DNA contains deoxyribose sugar, while RNA contains ribose sugar. The nitrogenous base thymine in DNA is replaced by uracil in RNA. Both of these bases pair with adenine. Uracil is more stable in single-stranded polynucleotides. Table 5.2.1 compares DNA and RNA.

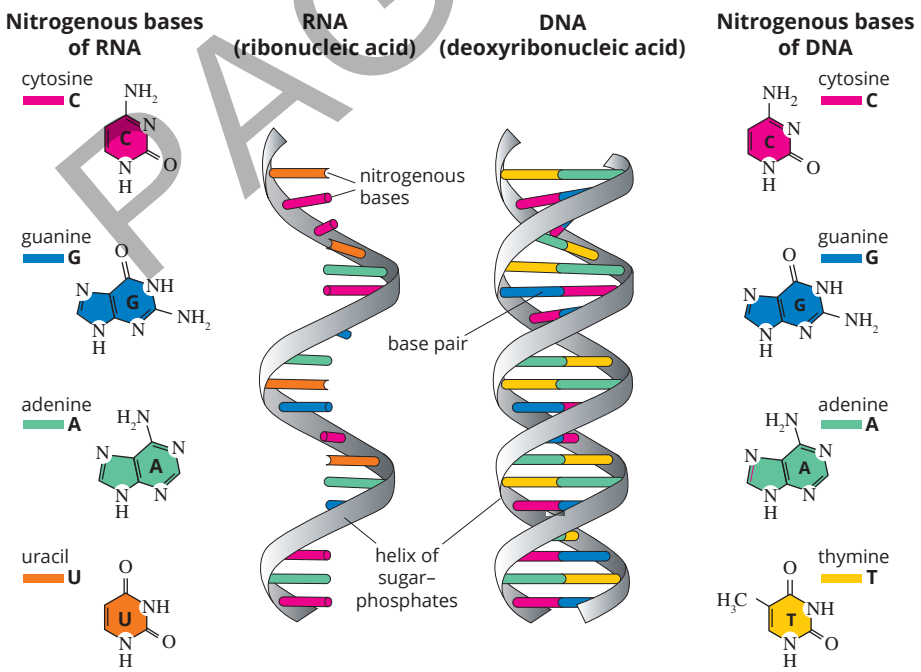
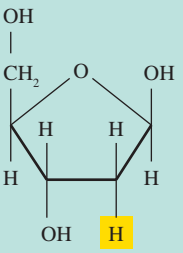
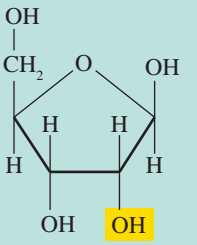
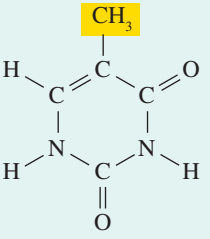
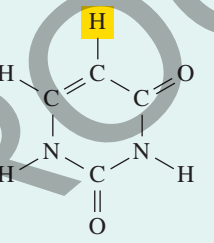
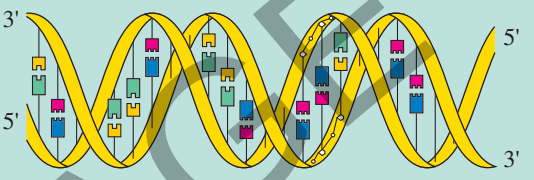


FIGURE 5.2.1 Comparison of the structures of RNA and DNA. RNA is typically single-stranded and DNA is double-stranded. Both molecules are made up of nitrogenous bases and a sugar–phosphate backbone. In RNA ribose sugar replaces deoxyribose sugar and uracil replaces thymine.

TABLE 5.2.1 A summary of the differences between DNA and RNA

	DNA	RNA
relative length	long	short
sugar	deoxyribose  deoxyribose in DNA	ribose  ribose in RNA
bases	adenine cytosine guanine thymine  thymine in DNA	adenine cytosine guanine uracil  uracil in RNA
strands	double 	usually single 
base pairing	adenine–thymine cytosine–guanine	adenine–uracil cytosine–guanine

In eukaryotic cells, RNA molecules are formed in the nucleus and pass into the cytoplasm where protein synthesis occurs. There are several types of RNA that are categorised as either RNA involved in protein synthesis, regulatory RNA or specialised RNA. Our focus is the three types of RNA involved in protein synthesis.

- **Messenger RNA (mRNA)** nucleotide sequences can fold back on themselves to form secondary structures. mRNA carries a copy of the DNA's nucleotide sequence to serve as a template for ribosomes to translate into proteins (Figure 5.2.2).
- **Ribosomal RNA (rRNA)** molecules fold into complex structures with double-stranded regions essential for ribosome function. rRNA forms ribosomes, the site of translation of the mRNA into proteins (Figure 5.2.3). It also helps catalyse peptide bond formation.

- **Transfer RNA (tRNA)** molecules have double-stranded regions formed by internal base pairing, giving them a cloverleaf structure (Figure 5.2.4). tRNA molecules can 'read' and translate the DNA information because they have an anticodon region that pairs with the complementary codon on mRNA. This enables tRNA to deliver specific amino acids to ribosomes and build the polypeptide chain during translation.

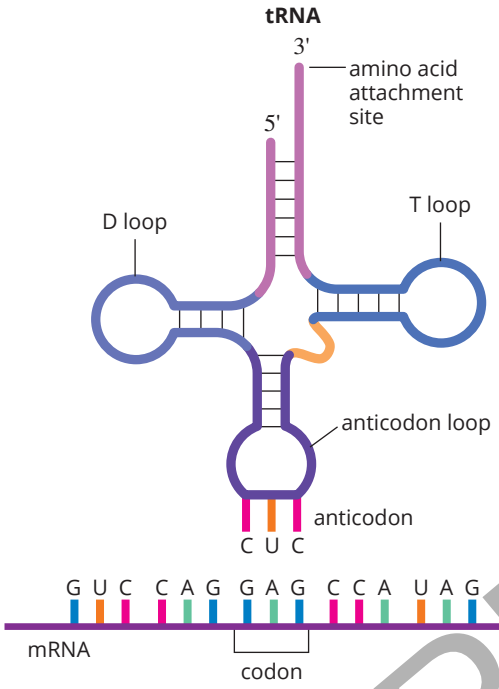


FIGURE 5.2.4 Transfer RNA (tRNA).

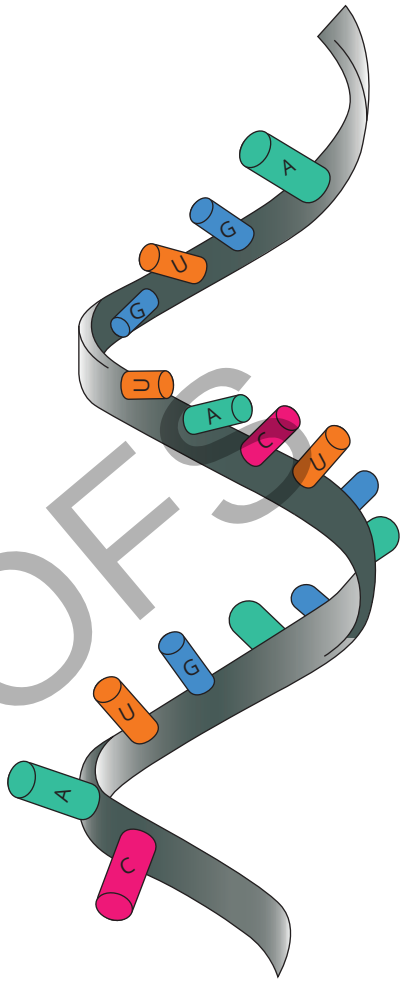


FIGURE 5.2.2 Messenger RNA (mRNA).

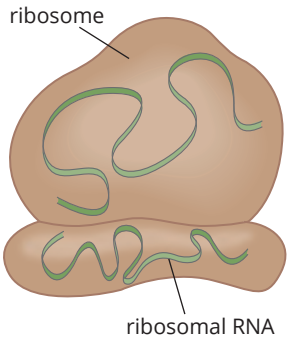


FIGURE 5.2.3 Ribosomal RNA (rRNA).

GENE EXPRESSION AND PROTEIN SYNTHESIS

Gene expression is the process by which the information stored in a gene is used to synthesise a functional gene product, namely protein or RNA (Figure 5.2.5).

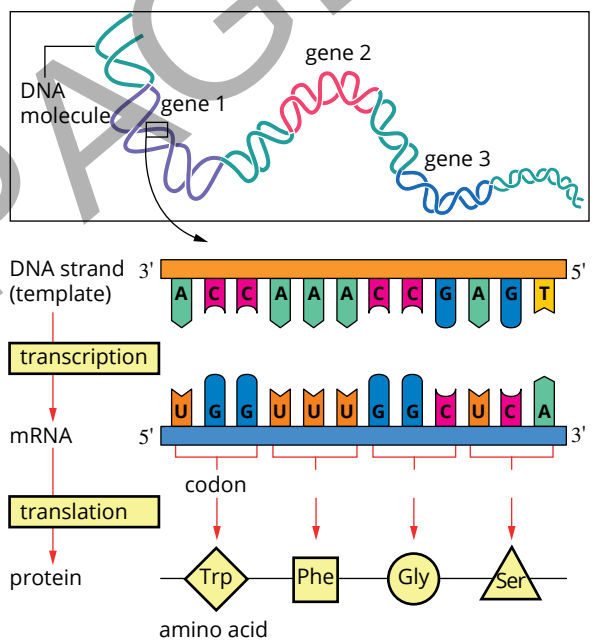


FIGURE 5.2.5 Overview of gene expression and resulting protein synthesis.

DNA and RNA both play vital roles in protein synthesis (Figure 5.2.6). DNA provides the instructions that are translated by RNA into proteins that carry out all the functions essential to life. RNA plays an important role in expressing the information contained in the sequence of a gene to synthesise proteins. Each different type of RNA has a specific role in the process of protein synthesis.

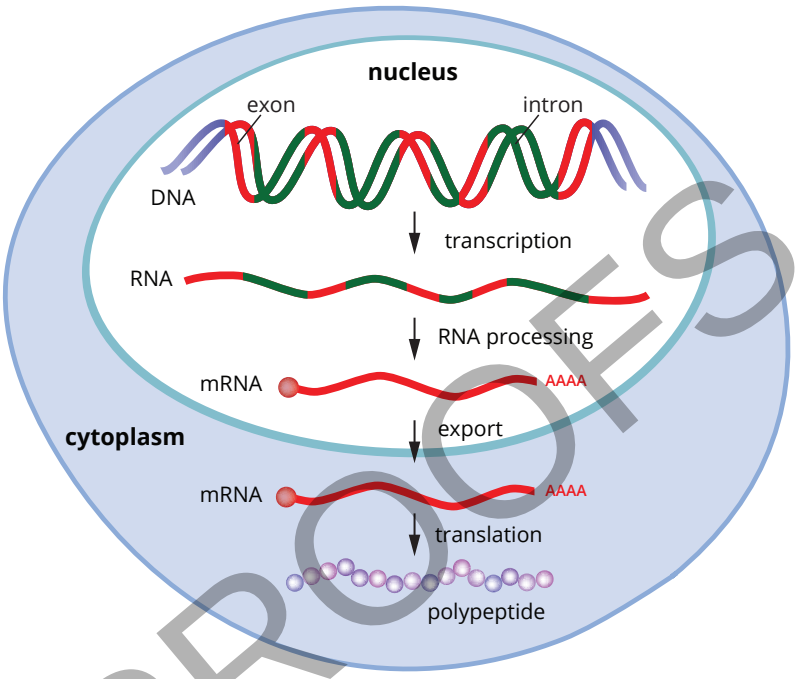


FIGURE 5.2.6 The information stored in a gene is used to synthesise a polypeptide, which will become a protein.

Messenger RNA is formed in the nucleus by the process of transcription. It carries a copy of the nucleotide sequence of DNA that specifies the amino acid sequence for a particular protein. During transcription, pre-mRNA is first formed by the enzyme RNA polymerase. Pre-mRNA is then processed (post-transcriptional modification) to form mature mRNA, which is a single-stranded copy of the **coding DNA** (or gene). The mature mRNA travels from the nucleus to the cytosol where it binds to ribosomes ready for translation (Figure 5.2.7).

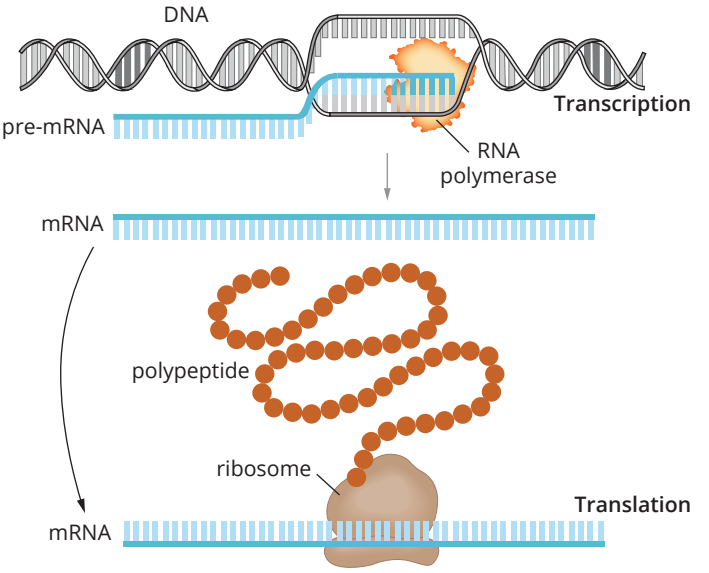


FIGURE 5.2.7 Messenger RNA is formed by transcription inside the nucleus. Mature mRNA then travels outside the nucleus to a ribosome, where translation occurs.

Ribosomal RNA is synthesised in the nucleolus of the cell nucleus and is based on the nucleotide sequence of the DNA. Together with proteins, rRNA forms a small organelle called a ribosome. Ribosomes are the sites where the information in the mRNA is translated into a chain of amino acids (Figure 5.2.8).

Transfer RNA molecules transfer amino acids from the cytoplasm to the ribosomes, where they are joined to form a polypeptide chain based on the arrangement of nucleotides in the mRNA. There are 61 different tRNA molecules, each of which combines with only one specific amino acid at one end of its molecule. (There are no tRNA molecules that recognise the three stop codons, which is how translation is terminated.)

There are three places for tRNA to bind to the ribosome: the exit (E), peptidyl (P) and aminoacyl (A) sites (Figure 5.2.9). At the other end of the tRNA molecule, there is a sequence of nucleotides known as the **anticodon**. The anticodon recognises a particular sequence of nucleotides in the mRNA. This enables an amino acid to be positioned in the correct place on a protein.

Gene expression leading to protein synthesis in eukaryotic cells occurs in three stages:

- transcription
- RNA processing
- translation.

Transcription—copying the genetic code

The production of single-stranded mRNA from DNA is called **transcription** and occurs within the nucleus of eukaryotic cells. Transcription occurs in three steps: initiation, elongation and termination. The DNA segment that undergoes transcription is known as the transcription unit.

Initiation

Transcription factors are proteins that bind to DNA sequences close to the promoter region of a gene. In eukaryotic cells, transcription factors are required for RNA polymerase to attach to the DNA. RNA polymerase then attaches to the promoter, unwinding and unzipping the DNA molecule by breaking the weak hydrogen bonds between the two strands to expose the bases (Figure 5.2.10a).

Elongation

During transcription, the RNA polymerase molecule covers a region of approximately 30 base pairs. Within this region, a segment of about 15 base pairs is uncoiled. This forms a transcription bubble. As the RNA polymerase moves along the gene, DNA strands behind the transcription bubble are coiled again. The RNA polymerase moves along the DNA molecule, producing a strand of mRNA. It uses a strand of DNA as a template, attaching nucleotides (A, U, G, C) by complementary base pairing.

Messenger RNA is always synthesised in the 5′ to 3′ direction, with new nucleotides added to the 3′ end. The initial mRNA molecule transcribed is called a primary RNA transcript (Figure 5.2.10b). The primary RNA transcript is then processed into mature mRNA.

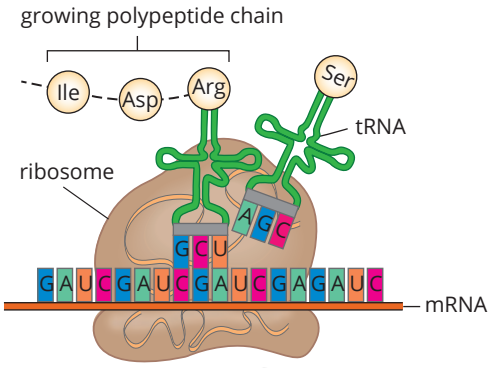


FIGURE 5.2.8 Ribosomal RNA and proteins form a small organelle called a ribosome.

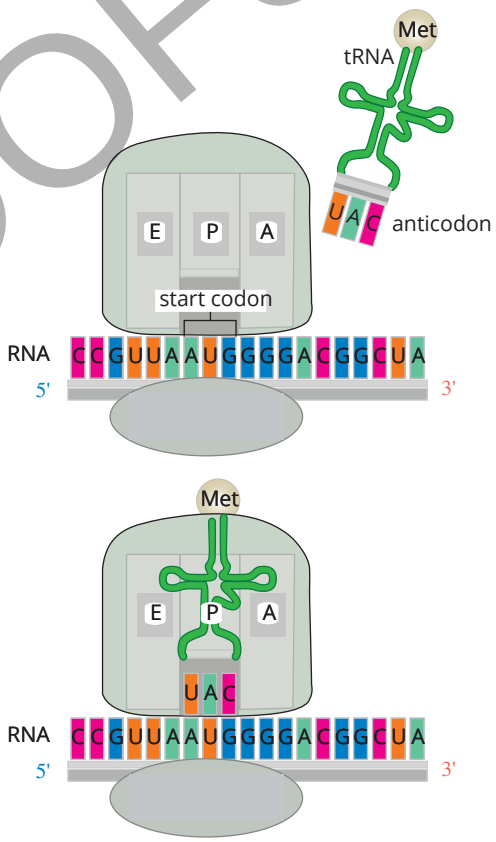


FIGURE 5.2.9 Transfer RNA molecules carry amino acids from the cytoplasm and bind them to a specific spot on the ribosome: the exit (E), peptidyl (P) or aminoacyl (A) sites.

Transcription is the synthesis of messenger RNA from a DNA template.

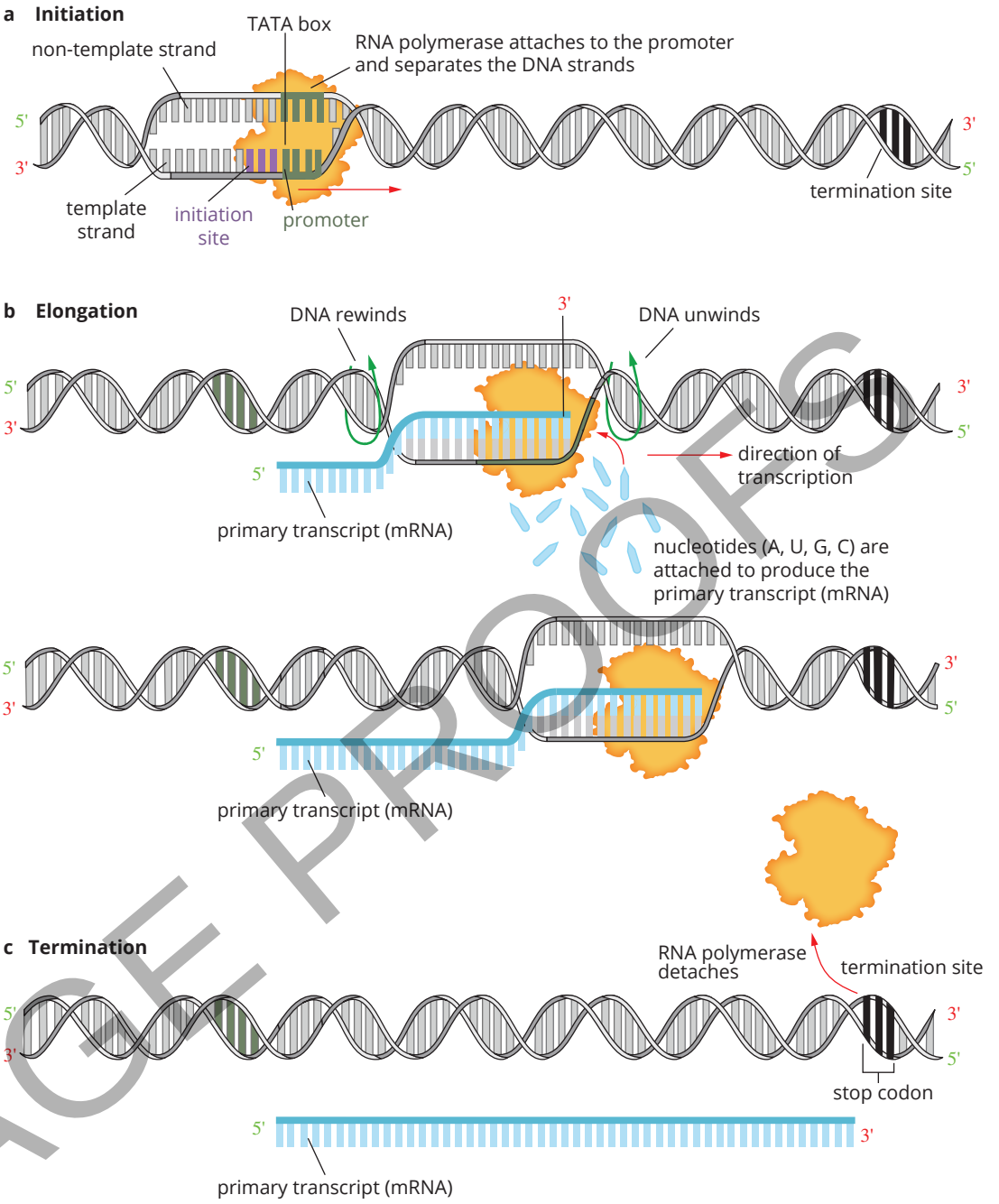


FIGURE 5.2.10 Transcription occurs in three stages: (a) initiation, (b) elongation and (c) termination.

Termination

Transcription ends when RNA polymerase reaches the termination site of the gene. This region contains a stop triplet code, which binds release factors that signal termination. The RNA polymerase detaches, releasing the mRNA and allowing the DNA molecule to re-form (Figure 5.2.10c).

Many RNA polymerase molecules may attach to the gene being transcribed, producing many of the same mRNA molecules. The strand of DNA that is transcribed to the mRNA is known as the template strand, and the other complementary strand is known as the **coding strand**. The mRNA carries the same base sequence as the coding strand, except it contains uracil in place of thymine.

RNA processing

After transcription, the primary RNA transcript undergoes processing before it is translated. This stage of gene expression is called **RNA processing** and includes:

- the addition of a 5′ cap
- the addition of a poly-A tail
- splicing (removal) of the introns (mRNA maturation).

5′ cap and poly-A tail

A cap consisting of a methylguanosine triphosphate molecule, called a **5′ cap**, is added to the 5′ end of the primary RNA transcript while it is being synthesised during transcription. Once transcription has finished, a chain of up to 250 adenine nucleotides is added to the 3′ end of the primary RNA transcript. This chain is called a **poly-A tail**.

These modifications to either end of the primary RNA transcript increase its stability and prevent it from degrading. The 5′ cap also helps bind the ribosome to the mRNA at the beginning of translation.

Splicing

Recall that eukaryotic genes have protein-coding regions called exons and non-protein-coding regions called introns. In eukaryotes, before a protein can be produced, the introns must be cut out of the primary RNA transcript to form the mature mRNA molecule. This process is known as **splicing**. During splicing, a complex molecule composed of protein and RNA molecules, called a **spliceosome**, removes the introns from the primary RNA transcript and joins the exon sections together to make mature mRNA (Figure 5.2.11). Note that not all exons will necessarily be included in the final mRNA, as alternative splicing allows for the production of different proteins from the same gene. The single-stranded mature mRNA molecule then exits the nucleus via a nuclear pore.

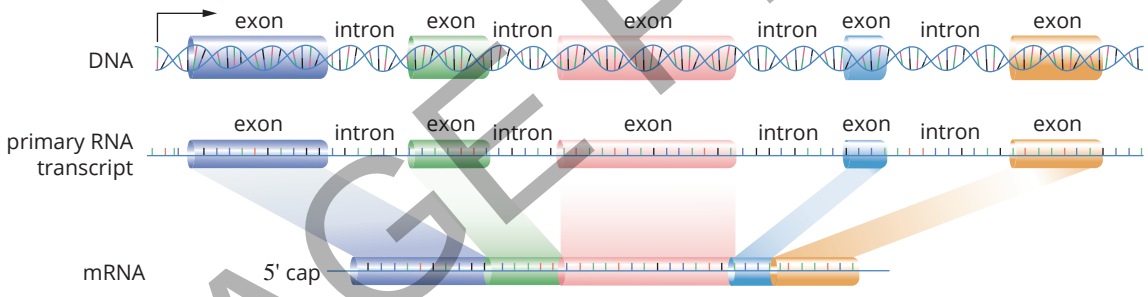


FIGURE 5.2.11 During RNA processing, the introns are spliced from the primary RNA transcript, resulting in mature messenger RNA.

Most prokaryotes contain only exons and therefore splicing does not occur in prokaryotes. In fact, little or no RNA processing occurs in prokaryotes.

Alternative splicing

Enzymes in eukaryotic nuclei modify pre-mRNA in specific ways before translation. During this RNA processing, both ends of the primary transcripts are altered and, in most cases, interior sections of the RNA molecule (introns and exons) are cut out and the remaining parts are spliced together. These modifications produce an mRNA molecule ready for translation.

A pre-mRNA transcript can be spliced in many ways, resulting in alternative mature mRNA strands from a single gene and therefore different proteins. This is the result of some exons being removed along with the introns during RNA processing. In the example shown in Figure 5.2.12, a particular gene may result in a mature mRNA that contains all exons 1–5, but the same gene may result in another mature mRNA that contains only exons 1, 2, 4 and 5. Alternative splicing is one reason that the 21 000 genes of humans can produce many more than 21 000 proteins.

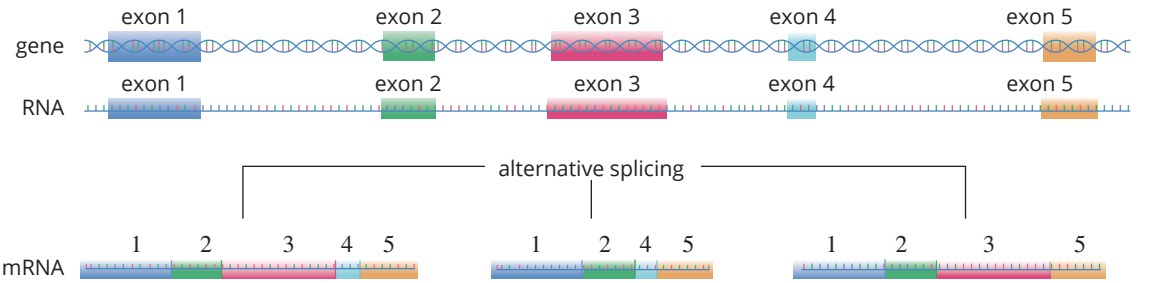


FIGURE 5.2.12 Alternative splicing of a single gene gives rise to alternative mRNA molecules, resulting in many different proteins.

In early research on gene structure, introns were called 'junk DNA' because it was believed they had no role in protein production. It is now known that protein expression is much more complex than first thought and that, by splicing mRNA during RNA processing, different proteins may be made from the same gene.

Translation—assembling the protein

Translation is another three-step process (initiation, elongation and termination) in which the codons on mRNA are translated into a sequence of amino acids, resulting in a polypeptide. This process occurs when ribosomes bind to an mRNA molecule and act as docking stations for the tRNA molecules to deposit their specific amino acids. A part of the tRNA, called an anticodon, recognises and binds to the codon on the mRNA by complementary base pairing. Each tRNA molecule carries a specific amino acid related to the codon to which it binds.

Initiation

To begin protein synthesis, a small ribosomal sub-unit attaches to the 5′ end of an mRNA strand. It moves along the mRNA until it reaches a start codon (AUG). The sequence AUG, which codes for the amino acid methionine, is the most common initiation triplet of mRNA (there are some rare exceptions).

A tRNA molecule with the anticodon UAC then brings the methionine to the mRNA. The tRNA molecule joins to the mRNA start codon, attaching by complementary base pairing between the codon and anticodon. A large ribosomal sub-unit also attaches to the tRNA and the small ribosomal sub-unit.

The binding of both ribosomal sub-units causes the formation of three sites for tRNA to bind: the exit (E), peptidyl (P) and aminoacyl (A) sites. The attachment of amino acids to their corresponding tRNA molecules occurs in the cytosol—a process that is catalysed by enzymes.

Elongation

Following the attachment of methionine, another tRNA molecule with a complementary anticodon to the next codon on the mRNA attaches and adds its specific amino acid to the growing polypeptide chain (Figure 5.2.13). The deposited amino acid joins by a peptide bond to the first amino acid through a condensation polymerisation reaction. The ribosome then releases the tRNA and moves further along the mRNA strand. At each codon, a new tRNA binds and adds another amino acid. The tRNA molecules can be reused, allowing them to pick up more of their specific amino acids and return to the mRNA molecule.

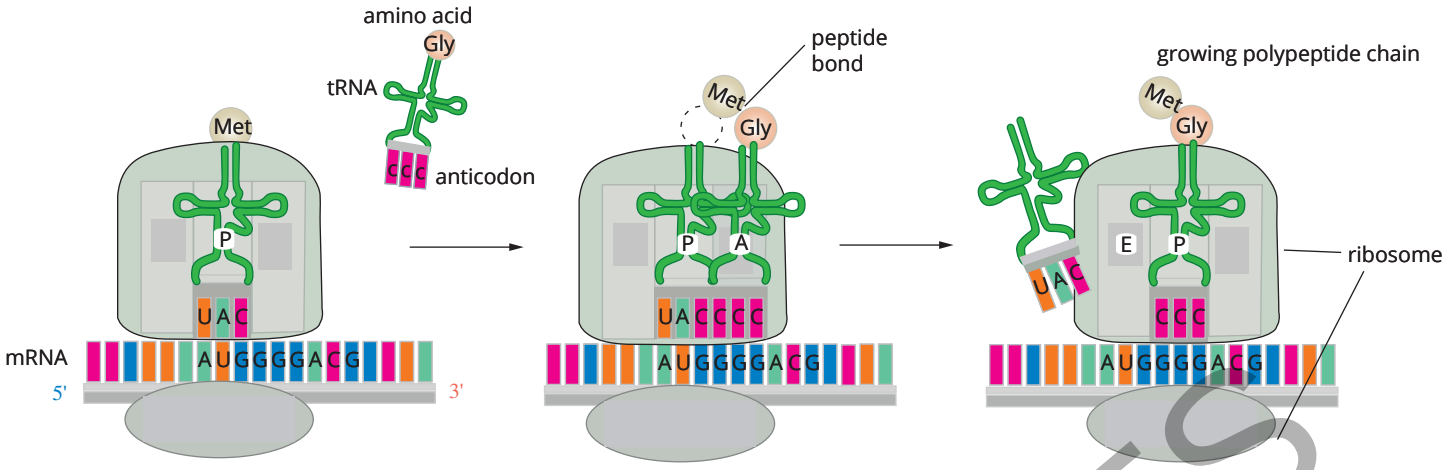


FIGURE 5.2.13 The initiation and elongation steps in the translation process.

Termination

Attachment of amino acids continues until a stop codon is reached. The polypeptide chain is then released from the ribosome into the cytoplasm or the endoplasmic reticulum. Some proteins consist of more than one polypeptide. The polypeptides of these proteins associate in the cytoplasm or the Golgi apparatus to form the fully functional protein.

Many ribosomes can translate the same, single strand of mRNA, enabling many polypeptide chains to be produced at the same time. Once the polypeptides are fully functional, they either remain in the cell for use or are exported from the cell via exocytosis for use elsewhere in the organism.

Gene expression and protein synthesis in eukaryotes and prokaryotes

Gene expression is tightly controlled in both eukaryotes and prokaryotes. However, as the process of gene expression in eukaryotes is more complex, gene regulation occurs in a greater number of stages in eukaryotes than in prokaryotes.

Eukaryotic cells have compartmentalised organelles, such as the nucleus, which separates transcription and translation into different parts of the cell. In eukaryotic cells, transcription and RNA processing occur within the nucleus and translation occurs in the cytoplasm (Figure 5.2.14). Gene expression in eukaryotes can occur during the processes of transcription, RNA processing or translation. It is highly controlled and can be regulated at any of these stages.

Gene expression in prokaryotic cells consists only of transcription and translation and occurs in the cytoplasm of cells. This is because prokaryotic cells do not have a nucleus or any other membrane-bound organelles. In the absence of a nucleus, the prokaryote can simultaneously transcribe and translate the same gene, and the newly made protein can quickly diffuse to its site of function. Ribosomes can attach to the mRNA while it is being transcribed, so translation can occur at the same time. Most prokaryotes also do not contain introns, so splicing is not necessary prior to translation, and the translation of mRNA can begin as soon as the leading 5′ end of the mRNA molecule peels away from the DNA template.

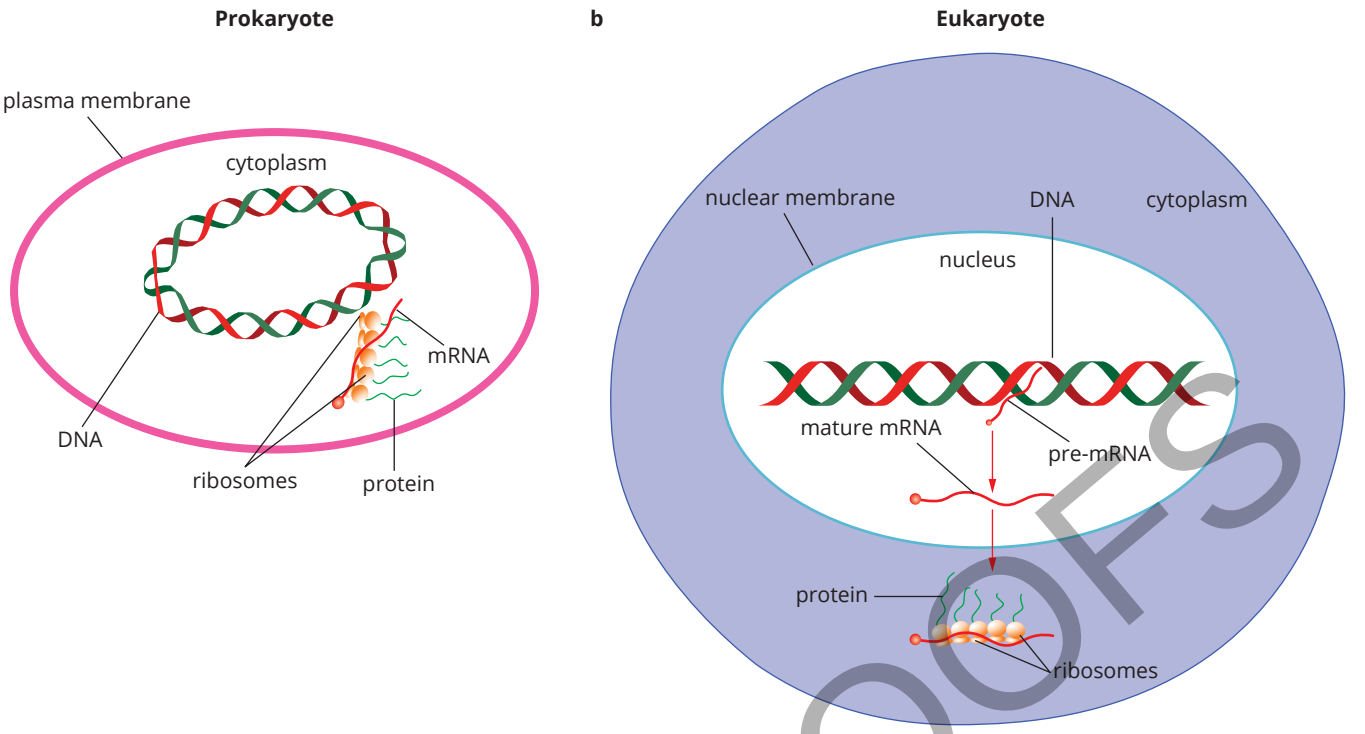


FIGURE 5.2.14 (a) In prokaryotic cells, transcription and translation occur simultaneously in the cytoplasm. (b) In eukaryotic cells, transcription and RNA processing occur in the nucleus and translation occurs in the cytoplasm.

In prokaryotes, gene expression consists only of transcription and translation. In eukaryotes, it involves transcription, RNA processing and translation.

Figure 5.2.15 illustrates the many differences between protein synthesis in prokaryotic and eukaryotic cells. These differences have been used to develop drugs that target protein synthesis in prokaryotes only.

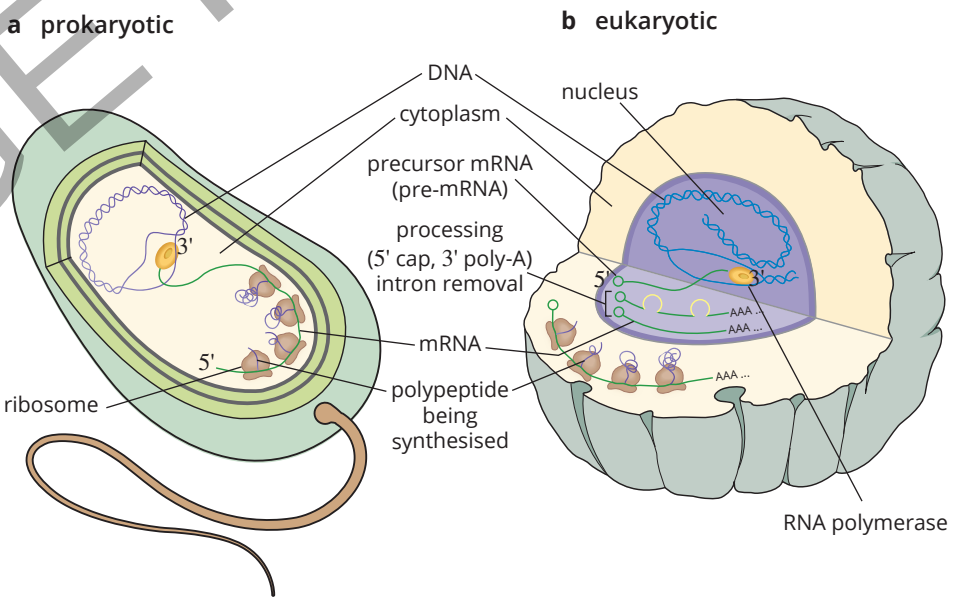


FIGURE 5.2.15 Comparison of protein synthesis in (a) prokaryotic and (b) eukaryotic cells.

GENE REGULATION

Recall gene *expression* is the process by which the information stored in a gene is used to synthesise a functional gene product (protein or RNA) and that gene *regulation* refers to the mechanisms and processes that control the timing, location and amount of gene expression. Gene regulation ensures cellular:

- efficiency – the ability to regulate gene expression conserves energy and materials (nucleotides and amino acids) in the cell as proteins or RNA molecules are only produced when they are required
- adaptability – for example, exposure to UV radiation results in increased expression of the genes in cells that produce melanin, with the resulting increased melanin helping to protect against UV damage to DNA
- differentiation – multicellular organisms can have specialised cells that require a specific set of proteins. For example, in humans, the cells in connective tissue and bone require the protein fibrillin to form elastic fibres, and skin cells require the enzyme tyrosinase to produce melanin and other pigments.

Structural and regulatory genes

Constitutive genes are always switched on, so they are expressed constitutively (continually) in cells. For other genes, transcription may be **induced** or **repressed** by transcription factors as needed, depending on the cell type, stage or the specific environmental conditions.

Structural genes code for proteins and RNAs that are not involved in gene regulation. For example, they can code for enzymes, protein channels, protein components for the cytoplasmic skeleton, or tRNA, among others.

Regulatory genes code for transcription factors. Transcription factors are proteins that control gene expression at the transcription stage. They bind to DNA sequences close to the promoter region of a gene or to the RNA polymerase to induce or repress the expression of specific genes (Figure 5.2.16).

Genes may be switched on all the time (constitutively expressed) or may be induced or repressed as needed by transcription factors.

Structural genes code for RNA and proteins that are not involved in gene regulation. Regulatory genes control the expression of structural genes via the production of transcription factors.

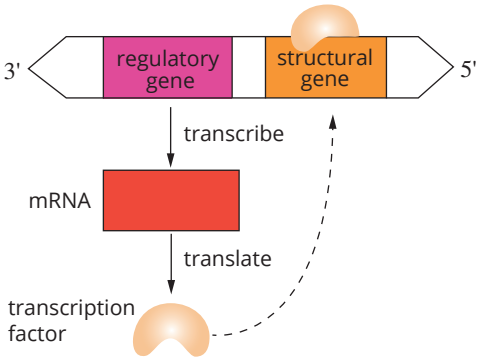


FIGURE 5.2.16 Regulatory genes code for transcription factors that induce or repress structural genes.

The *lac* operon

As gene regulation in eukaryotes is complex, a simple prokaryotic model, the ***lac* operon,** is commonly used to illustrate how transcription factors regulate gene transcription.

The *lac* operon is found in *E. coli* and some other bacteria. In prokaryotes, an **operon** is a unit of DNA under the regulation of a single promoter that codes for several proteins. The *lac* operon is an example of an inducible operon, meaning that it can be switched on or off. It expresses three structural genes that code for three enzymes, but only when the sugar **lactose** is available. The enzymes break down lactose into the usable forms, glucose and galactose.

Producing the enzymes that break down lactose constitutively would be a misuse of energy for the bacteria, because lactose is not the preferred energy source for *E. coli*. It is important that the *lac* operon genes are inducible and are only expressed when lactose is present in the environment.

The *lac* operon consists of:

- a promoter—the binding site of the RNA polymerase
- an **operator**—the binding site of the transcription factor, which, in this case, is a repressor
- three structural genes—*lacZ* (β-galactosidase), *lacY* (β-galactoside permease) and *lacA* (β-galactoside transacetylase)—that code for three different enzymes (Figure 5.2.17).

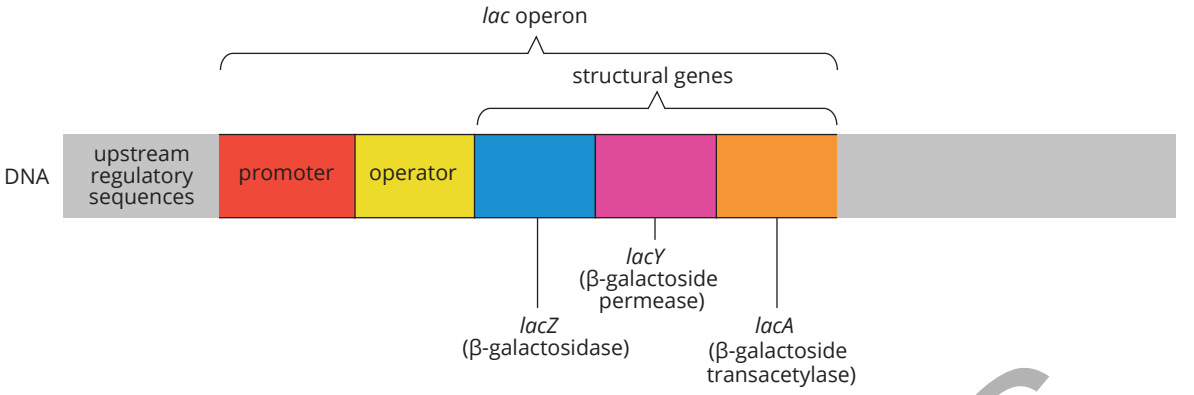


FIGURE 5.2.17 The *lac* operon.

Adjacent to the *lac* operon is the regulatory gene ***lacI***. *LacI* codes for a transcription factor called the ***lac* repressor**. The *lacI* gene is expressed constitutively (or continually) so the *lac* repressor is always present. This transcription factor binds to the operator in the *lac* operon, blocking the RNA polymerase from binding to and transcribing the structural genes in the *lac* operon, preventing the synthesis of the three enzymes involved in lactose metabolism. However, when lactose is present, the lactose binds to the *lac* repressor, inhibiting the transcription factor from binding to the operator. This enables the RNA polymerase to attach to the promoter in the *lac* operon and transcribe the three structural genes, resulting in the production of the enzymes involved in lactose metabolism (Figure 5.2.18).

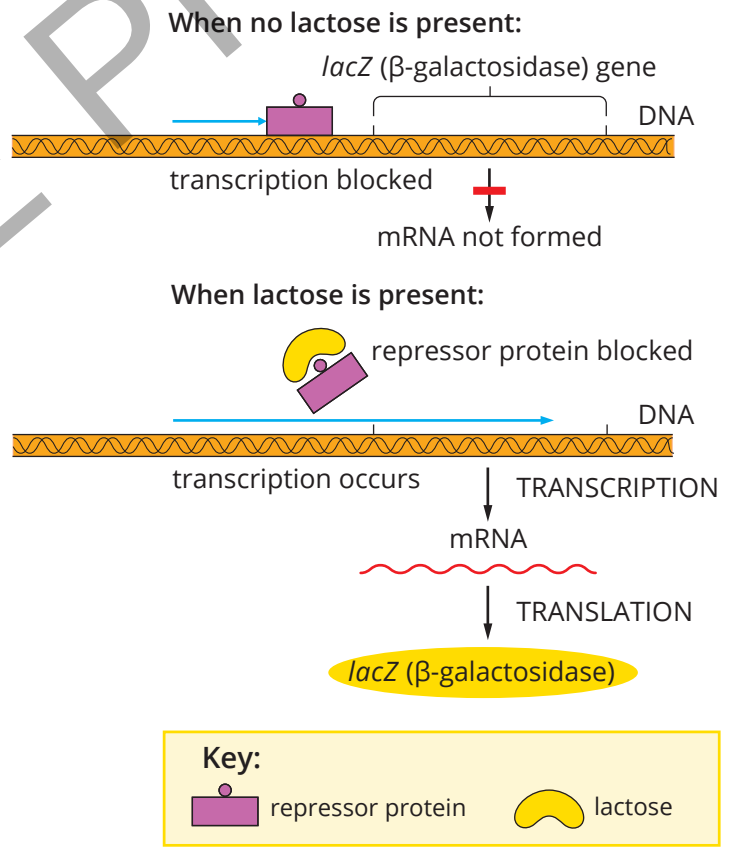


FIGURE 5.2.18 The genes of the *lac* operon are only expressed in the presence of lactose.

Tumour-suppressor genes

Cell differentiation is a complex process controlled by regulatory genes. Knowledge of regulatory genes is constantly developing, and much of what we know about cell differentiation and gene regulation comes from research into cancer.

Cancer is a set of diseases in which cells escape from the control mechanisms that normally limit their growth. Genes that normally regulate cell growth and division during the cell cycle include genes for growth factors, their receptors and molecules for signalling pathways. Mutations that alter any of these genes in cells can lead to cancer.

In addition to genes whose products normally promote cell division, cells contain genes whose normal products inhibit cell division. These genes are called **tumour-suppressor genes** because the tumour-suppressor proteins they encode for can prevent uncontrolled cell growth. Tumour-suppressor proteins are produced in response to exposure to mutagens that cause damage to the structure of DNA. Some tumour-suppressor proteins repair damaged DNA, preventing the cell from accumulating cancer-causing mutations.

An important tumour-suppressor protein is p53 (Figure 5.2.19). If DNA damage has occurred, the p53 protein binds to specific sites on the DNA to repress genes that play a role in the continuation of the cell cycle. This inhibits cell division and prevents damaged DNA from replicating. If the damage is minor, p53 activates genes that repair DNA. In cases where the damage cannot be repaired, p53 will initiate apoptosis (cell death). The p53 protein plays a major role in the prevention of cancers. If the gene coding for the p53 protein is deleted or mutated, the risk of cancer is greatly increased. In half of all cancers, p53 has been found to be inactive.

FIGURE 5.2.19 A molecular model of the tumour-suppressor protein p53 (left and right) bound to a DNA molecule (centre).

Epigenetic inheritance

The structure of DNA was described in Module 5.1. Recall that the DNA of eukaryotic cells is packaged with proteins called histones, which together form nucleosomes (Figure 5.2.20). Chromatin is the entire complex of nucleosomes found in the nucleus of eukaryotic cells. Chromatin compacts and organises the genetic material within the nucleus of eukaryotic cells and plays a crucial role in regulating gene expression, DNA replication and cell division by controlling DNA accessibility.

Epigenetic inheritance is the transmission of changes in gene expression or cellular phenotype from one generation to the next that do not involve alterations to the DNA nucleotide sequence but are instead the result of **chemical tags**. These tags are molecules that attach to parts of the DNA or proteins and can activate or silence genes. Epigenetic inheritance is an example of gene regulation because it involves chemical changes to DNA, histones or other regulatory-associated proteins that affect gene expression, without altering the DNA sequence itself. These modifications can be inherited across cell divisions and sometimes even between generations, affecting gene activity in a heritable yet reversible manner.

Chemical tags and chromatin remodelers

Chemical tags and chromatin remodeling complexes influence how accessible or compact chromatin is, which determines whether chromatin adopts a loose, transcriptionally active euchromatin state or a tightly packed, transcriptionally inactive heterochromatin state (Figure 5.2.21).

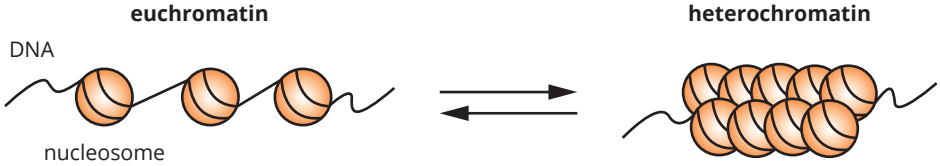


FIGURE 5.2.21 Chromatin transitions between a transcriptionally active, loosened euchromatin state and a transcriptionally silent, compacted heterochromatin state.

Histones are proteins found in eukaryotic cells that tightly package DNA into structures called nucleosomes.

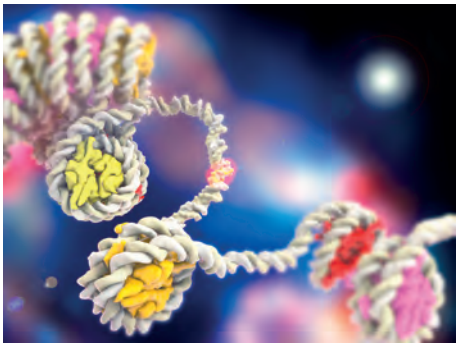


FIGURE 5.2.20 DNA strands coil around histones to form nucleosomes.

Euchromatin is a form of chromatin that is loosely or lightly packed. Heterochromatin is a form of chromatin that is densely or tightly packed. The presence of euchromatin usually indicates that a cell is actively transcribing DNA to mRNA.

Chemical tags are covalent modifications added to DNA or histones, namely DNA methylation and histone modifications.

DNA methylation is the addition of a methyl group ($-CH_3$) to cytosine bases in DNA, catalysed by enzymes called DNA methyltransferases. Low DNA methylation levels are associated with open chromatin and active gene transcription (euchromatin). High DNA methylation levels compact chromatin and silence gene expression (heterochromatin) by recruiting proteins such as methyl-CpG-binding domain proteins that attract histone deacetylases and chromatin remodelers. Methylation patterns can be passed on during DNA replication.

The eight histone proteins that form a nucleosome are two copies each of the core histone proteins H2A, H2B, H3 and H4.

Histone modifications include (among others that are less well understood):

- Methylation – the addition of methyl groups to lysine or arginine residues. Histone methylation can either activate or repress transcription depending on the specific modification.
- Acetylation – the addition of an acetyl group ($-COCH_3$) to lysine residues, catalysed by enzymes called histone acetyltransferases. In euchromatin, histone acetylation neutralises the positive charge of histones, which reduces their affinity for negatively charged DNA and loosens chromatin for transcription. Conversely, the heterochromatin state is associated with histone deacetylation, catalysed by histone deacetylases that remove the acetyl group.
- Phosphorylation – the addition of phosphate groups to serine or threonine residues. In euchromatin, histone phosphorylation introduces negative charges on histones that weaken their interaction with negatively charged DNA and makes DNA more accessible. In heterochromatin, histone phosphorylation compacts chromatin by enhancing the binding of structural proteins (e.g. heterochromatin protein 1, HP1) and prepares chromatin for DNA damage repair or mitosis.
- Ubiquitination – the addition of ubiquitin proteins to histones. Histone ubiquitination plays an important role in responding to DNA damage and can either activate or repress transcription depending on the molecular context.

Chromatin remodelers are protein complexes that use energy from adenosine triphosphate (ATP) hydrolysis to slide, eject or restructure nucleosomes, acting together with chemical tags to maintain chromatin in either euchromatin or heterochromatin states. For example, the switch/sucrose non-fermentable (SWI/SNF) and imitation switch (ISWI) families of chromatin remodelers have different protein complexes that are associated with gene activation in euchromatin or gene repression in heterochromatin. Mutations in chromatin remodeling genes are common in human cancers, and similar mutations have also been observed in non-human species.

Table 5.2.2 provides a summary of the chemical tags and chromatin remodeling complexes that allow cells to switch between euchromatin and heterochromatin states depending on environmental signals, developmental cues or other regulatory needs.

TABLE 5.2.2 Summary of chemical tags and chromatin remodeling complexes and their effects

Modification	Euchromatin (active genes)	Heterochromatin (compacted chromatin)
DNA methylation	low levels	high levels
histone acetylation	high levels	low levels
histone methylation	activation marks (e.g. H3K4me3, H3K36me3)	repression marks (e.g. H3K9me3, H3K27me3)
histone phosphorylation	loosens chromatin for transcription (e.g. H3S10ph)	role in chromatin condensation in mitosis (e.g. H3S10ph) and enhances binding with structural proteins like HP1
histone ubiquitination	typically associated with H2B ubiquitination	typically associated with H2A ubiquitination
chromatin remodelers	open chromatin (e.g. SWI/SNF, ISWI)	closed chromatin (e.g. SWI/SNF, ISWI)

Twin studies

The study of **epigenetics** and molecular markers that alter the expression of genes can be used to identify biological mechanisms that cause diseases. Monozygotic (identical) twin studies contribute to our understanding of the genetic component of age-related traits and diseases. Because identical twins have developed from the same fertilised egg, they have the same genome. Twins who are raised together also have shared environments, therefore any difference in their phenotype (observable traits) must be due to epigenetic modification. Epigenetic variations might explain, for example, why one identical twin acquires a genetically based disease such as schizophrenia while the other does not, despite them having identical genomes.

As twins age, they acquire different characteristics because of their environment. It has been proposed that these phenotypic differences are the result of epigenetic differences accumulating in the chromatin of the twins. The older the twins become, the greater the number of differences in the histones surrounding the DNA that regulates gene expression. Research indicates that not all age-related diseases are heritable. For every heritable disease, there is also an important non-genetic component that may influence the onset and severity of the disease.

Different responses to diseases in identical twins may be due to differences in their embryonic development or the ways their bodies respond to environmental factors. These responses could be random **somatic mutations** in non-gamete cells at any of the stages of DNA replication or protein synthesis processes. Studying identical twins allows researchers to control for contributing factors such as age, gender, maternal effects and most in utero and environmental influences.

While there is a significant body of evidence available to suggest that the less time twins spend together, the greater the epigenetic variability, further research is required to determine how environmental changes directly cause epigenetic changes in the chromatin. Smoking is one factor that has been reliably detected as a significant factor in changing the epigenetic pattern.

New technology, such as CRISPR (explored in Chapter 8), combined with an improved understanding of the epigenetic markers that contribute to disease, can potentially allow for targeted epigenome editing and reprogramming of genes. It may soon be possible to manufacture more effective treatments for cancer based upon our understanding of epigenetics, as we may be able to repair or replace mutations in genes such as the tumour-suppressor gene and other oncogenes.

Somatic mutations are changes in DNA that occur after conception in any cell of the body, except for germ cells (sperm and egg cells).

Master regulatory genes

The development of a complex, trillion-celled adult organism from a single, fertilised cell occurs in a series of steps. Master regulatory genes code for transcription factors that turn genes on and off in different cells in the developing embryo. They can also start a sequence of events by turning other regulatory genes on and off, leading to the production of transcription factors that will, in turn, regulate other genes, and so on. A single master regulatory gene can, in this way, control the development of a complex structure such as an eye or nervous system, or a whole organism.

Some of the most important master regulatory genes are the **homeotic genes**. Homeotic genes control the structure and organisation of body segments during embryonic development. The proteins of homeotic genes bind to regulatory regions of target genes, which then activate or repress cellular activity, directing the development of the organism.

Hox genes

Hox genes are master regulatory genes that affect the spatial pattern of expression of other genes. Hox genes control genes that form specific tissues of an organism at specific times during embryonic development. This group of genes determines the body plan along the head-to-tail axis during embryonic development. The expression of these master regulatory genes during the development of an organism determines the structures that will be formed at a given position.

In the fruit fly, *Drosophila*, Hox genes determine the position of the abdomen, thorax and head, and the location of the antenna and legs. As an embryo develops, the Hox genes direct messages to groups of embryonic cells to, for example, 'grow into a head' or 'develop into an eye'.

A mutation in a single Hox gene can lead to a fully functional leg growing where the antenna should be (Figure 5.2.22). Because the production of a leg requires the coordinated action of many genes, the base change must have occurred in a gene that switches the leg genes 'on' and other genes 'off'. This means a Hox gene has control over the genes that are involved in the development of all the cell types in both leg and antenna.

This important gene family has been highly conserved throughout animal evolution. Because of the high degree of genetic similarity in these genes across animal species, researchers can use model organisms, such as *Drosophila*, to investigate human birth defects and diseases (Figure 5.2.23).

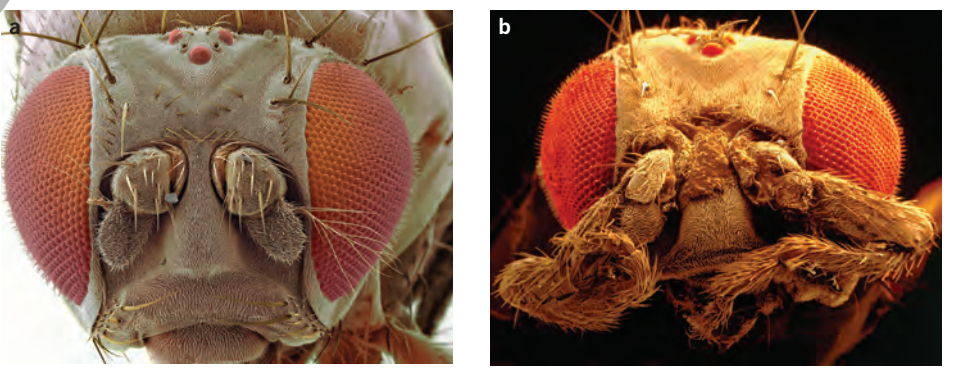


FIGURE 5.2.22 (a) A normal fly. (b) A mutation in a Hox gene results in a fly with legs growing where antennae should be.

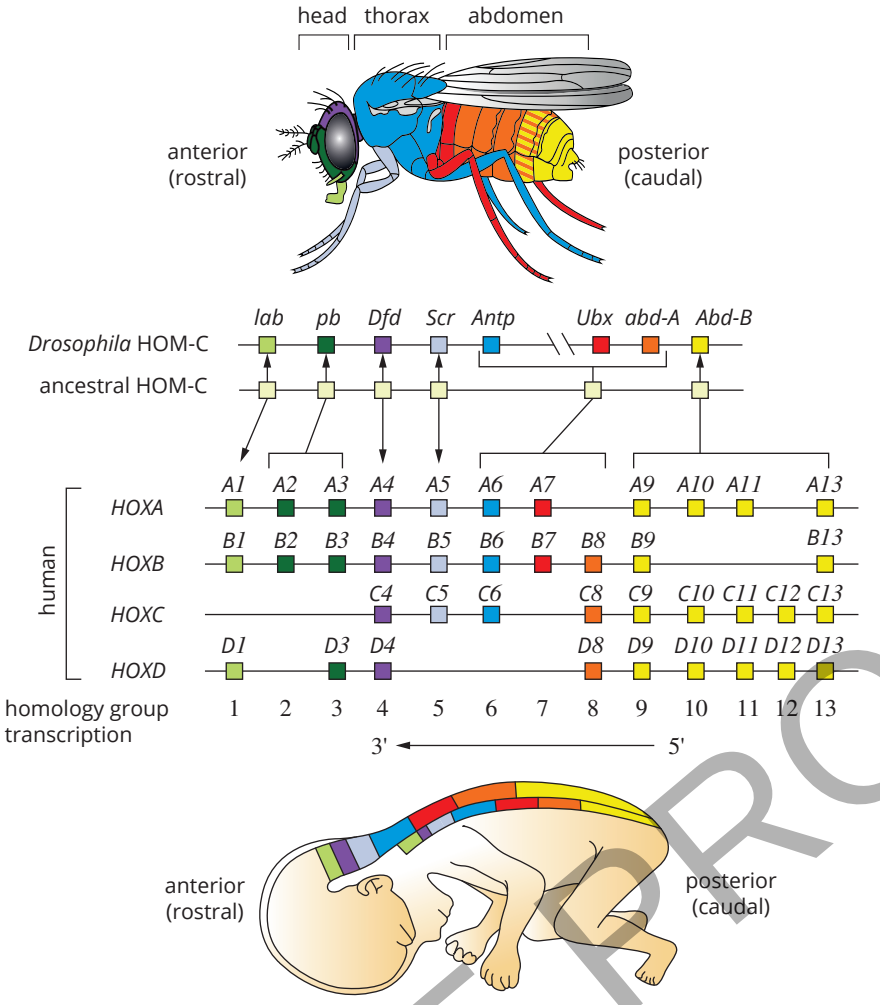


FIGURE 5.2.23 The genomic organisation and expression patterns of the Hox gene family in *Drosophila* and humans.

Sex determination

In humans, sex is determined by the presence of a pair of sex chromosomes (Figure 5.2.24). Females have two X chromosomes, and males have one copy of the X chromosome and one copy of the Y chromosome (Figure 5.2.25).

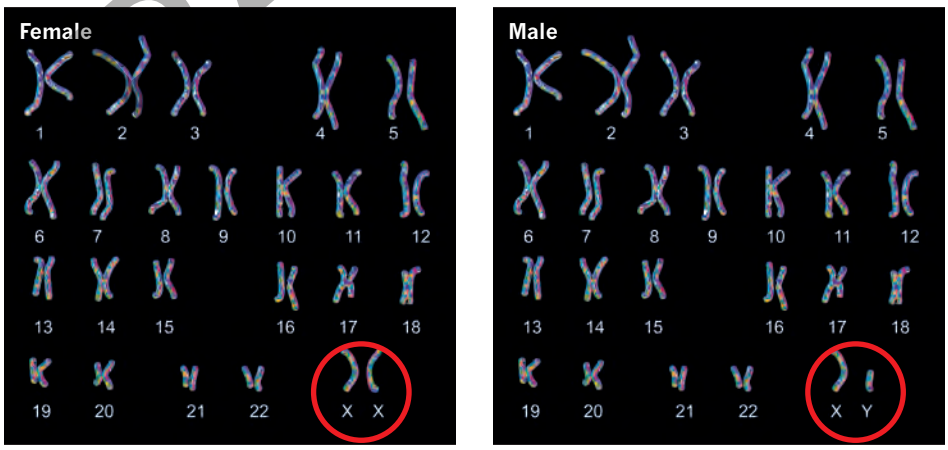


FIGURE 5.2.25 Human karyotype diagrams highlighting the difference between female and males. Females have two X chromosomes (a), and males have an X and Y chromosome (b).

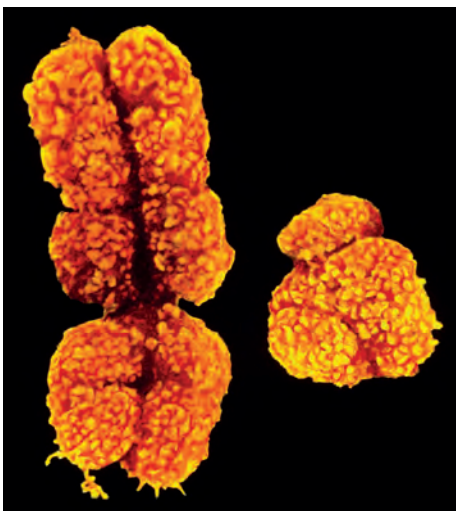


FIGURE 5.2.24 Coloured scanning electron microscope image of human sex chromosomes: X (left) and Y (right).

FIGURE 5.2.26 An SRY protein (red, curled) bound to a DNA helix.

One of the major genes for the regulation of sexual characteristics in mammals is found on the Y chromosome. The **sex-determining region Y (SRY)** gene codes for a protein that directs the development of the gonads into testes. Other autosomal genes regulate the embryonic development of the gonads initially.

The SRY gene is less than 1000 base pairs long and forms part of an SRY-containing region that is 900 base pairs long and that includes regulatory sequences. It is located on the short arm of the Y chromosome and codes for a 223-amino acid peptide and transcription factor that activates male-determining genes (Figure 5.2.26). SRY gene expression must occur within a narrow window of embryonic development (around 6 to 7 weeks of gestation in humans) to initiate male sex differentiation. If this does not occur, then the default female ovarian differentiation pathway is activated. The ovarian differentiation pathway is considered the default pathway because, in the absence of a Y chromosome and the SRY gene, the embryo will naturally develop ovaries and follow the female developmental trajectory.

MUTATIONS AND THEIR EFFECTS

Mutations result in different forms of the same gene, known as alleles. When different alleles of the same gene are examined using DNA sequence analysis, differences in the DNA base pair sequence are detected. Some alleles have changes in the DNA sequence that do not affect the structure of the polypeptide product of the gene but only affect the rate of gene expression. At the other extreme, some mutations create alleles that produce non-functional polypeptides.

Most mutations are detected and repaired by enzymes. Those that are not repaired fall into one of three categories.

- Neutral mutations have no effect on survival.
- Harmful mutations decrease the likelihood of survival.
- Beneficial mutations increase the likelihood of survival.

One mutation database has over 645 000 harmful human mutations identified from research. It also has over 5 227 000 DNA regions with approximately 34 870 900 human mutations, many of which are of unknown phenotypic function. These are classified as neutral mutations until they are better understood.

Beneficial mutations are dependent on population genetics and natural selection (Chapter 9). They are currently modelled, as there is no human database available (although many are being investigated).

Types of mutations

Somatic mutations occur in body cells. They only affect the individuals in which they arise and cannot be inherited by future generations. **Germline mutations** are heritable because they affect gametes (meaning they occur in the DNA of egg or sperm cells) and can be passed on to offspring. A germline mutation may bring a new allele into a gene pool, potentially influencing the allele frequencies in a population.

The alteration of a single base in DNA is referred to as a **point mutation** (Figure 5.2.27) and every amino acid change results from a change in the base sequence of DNA (Figure 5.2.28). For reference, recall Figure 5.1.24, which shows the genetic code for all 20 amino acids and their stop codons. However, not every change in the DNA base sequence results in an amino acid change. For example, the DNA sequence ATA is transcribed to the mRNA sequence as UAU and specifies the amino acid tyrosine. A mutation producing the DNA sequence ATG results in the mRNA sequence UAC, which still specifies tyrosine. Mutations like this, which have no effect on the polypeptide, are called **synonymous mutations** (Figure 5.2.28b). The DNA strand will have an optimal sequence of nucleotides to produce functional proteins.

FIGURE 5.2.27 A point mutation involves just one nucleotide in RNA (or one nucleotide pair in DNA).

a Normal sequence

DNA sequence transcribed	TAC	ATA	CAT	ATA	CAT	ATA	CAT
mRNA sequence	AUG	UAU	GUA	UAU	GUA	UAU	GUA
Amino acid sequence	Met	Tyr	Val	Tyr	Val	Tyr	Val

b Synonymous mutation (base substitution → no amino acid replacement)

Normal DNA sequence	TAC	ATA	CAT	ATA	CAT	ATA	CAT
Mutant DNA sequence	TAC	AT**G**	CAT	ATA	CAT	ATA	CAT
mRNA sequence	AUG	UA**C**	GUA	UAU	GUA	UAU	GUA
Amino acid sequence	Met	**Tyr**	Val	Tyr	Val	Tyr	Val

c Non-synonymous substitution mutation (base substitution → amino acid replacement)

Normal DNA sequence	TAC	ATA	CAT	ATA	CAT	ATA	CAT
Mutant DNA sequence	TAC	**G**TA	CAT	ATA	CAT	ATA	CAT
mRNA sequence	AUG	**C**AU	GUA	UAU	GUA	UAU	GUA
Amino acid sequence	Met	**His**	Val	Tyr	Val	Tyr	Val

d Nonsense mutation (base substitution → new 'stop' codon shortens the polypeptide)

Normal DNA sequence	TAC	ATA	CAT	ATA	CAT	ATA	CAT
Mutant DNA sequence	TAC	ATA	CAT	AT**T**	CAT	ATA	CAT
mRNA sequence	AUG	UAU	GUA	UA**A**	GUA	UAU	GUA
Amino acid sequence	Met	Tyr	Val	**STOP**			

e Single base insertion (frameshift—changes all amino acids after the point of insertion)

Normal DNA sequence	TAC	ATA	CAT	ATA	CAT	ATA	CAT	
Mutant DNA sequence	TAC	ATA	**G**CA	TAT	ACA	TAT	ACA	T
mRNA sequence	AUG	UAU	**C**GU	AUA	UGU	AUA	UGU	A
Amino acid sequence	Met	Tyr	**Arg**	**Ile**	**Cys**	**Ile**	**Cys**	

f Single base deletion (frameshift—changes all amino acids after the point of deletion)

Normal DNA sequence	TAC	ATA	**C**AT	ATA	CAT	ATA	CAT
Mutant DNA sequence	TAC	ATA	_ATA	TAC	ATA	TAC	AT
mRNA sequence	AUG	UAU	UAU	AUG	UAU	AUG	UA
Amino acid sequence	Met	Tyr	**Tyr**	**Met**	**Tyr**	**Met**	

FIGURE 5.2.28 A portion of the DNA template strand of a hypothetical gene showing the effect of different types of point mutations on the amino acid chain.

When there is a change in the DNA sequence that results in a different codon that still codes for the same amino acid (a synonymous mutation), the structure of the mRNA may differ. This post-transcriptional difference between the optimal codon sequence and the mutation codon is thought to affect the stability of the mRNA and the way that it is spliced during processing. There is now sufficient evidence to suggest that synonymous mutations can significantly influence the efficiency of gene expression and protein synthesis. Some of these changes have been linked to diseases, including cancer, by altering gene regulation and protein production.

Point mutations that result in an amino acid replacement are called **non-synonymous substitution mutations** (Figure 5.2.28c). Sickle-cell anaemia is an example of a disease caused by a non-synonymous substitution mutation.

Mutations that result in the generation of a 'stop' codon—for example UAU (tyrosine) to UAA (stop)—are termed **nonsense mutations** because no further amino acids are added after the site of mutation (Figure 5.2.28d). Because nonsense mutations stop translation, they may have severe effects, particularly if the mutation occurs early in a gene. Nearly all nonsense mutations lead to non-functional proteins.

If a single base pair is added to or deleted from the DNA sequence, the reading frame (codon sequence) of the mRNA is altered, and the wrong amino acids will be incorporated for the remainder of the sequence (Figure 5.2.28e, f). This type of change is called a **frameshift mutation** if the number of nucleotides inserted or deleted is not a multiple of three (which would instead delete a single codon and therefore remove a single amino acid from the polypeptide chain). In Figure 5.2.28e, the base guanine (G) is added to the sixth position from the left. In Figure 5.2.28f, the base cytosine (C) is deleted from the sixth position from the left.

Synonymous mutations do not result in an amino acid change. However, they do affect the messenger RNA structure and therefore may change the rate of protein expression, conformation (spatial arrangement of the component parts of the protein) and protein function.

Frameshift mutations can have very significant effects on the polypeptide formed. All bases after the point of a frameshift mutation will be improperly grouped into codons, and the result will be extensive missense (different amino acids being added to the polypeptide chain), usually ending sooner or later in nonsense and premature termination. Unless the frameshift is very near the end of the gene, the protein is almost certain to be non-functional.

Sickle-cell anaemia

Most people have normal haemoglobin in their red blood cells. However, some people have a type of haemoglobin that is distorted, resulting in 'sickle-shaped' red blood cells (Figure 5.2.29). The genetic basis for sickle-cell anaemia is a mutation of a single nucleotide pair in the gene that encodes the beta (β)-globin polypeptide of haemoglobin.

Haemoglobin consists of four polypeptide chains: two alpha (α) and two β-chains produced from α-globin and β-globin genes respectively. The difference between normal and sickle-cell haemoglobin is a single amino acid alteration in the β-chain, which changes the shape of the molecule. This chain consists of 146 amino acids. If the amino acid sequences of the polypeptide of β-chains of normal individuals and those with sickle-cell anaemia are compared, only one difference in amino acid sequence is observed.

The replacement of glutamic acid (in normal haemoglobin) with valine (in sickle-cell haemoglobin) dramatically alters the structure of the haemoglobin protein, changing the shape of red blood cells. This has serious consequences because it reduces the capacity of the haemoglobin molecule to carry oxygen to tissues. All the other amino acids in the polypeptides are identical.

This single point mutation has a major effect on people who have the sickle-cell allele. People with sickle-cell anaemia will die if not treated quickly when they show symptoms of the disease. Even with treatment, they seldom live beyond puberty.

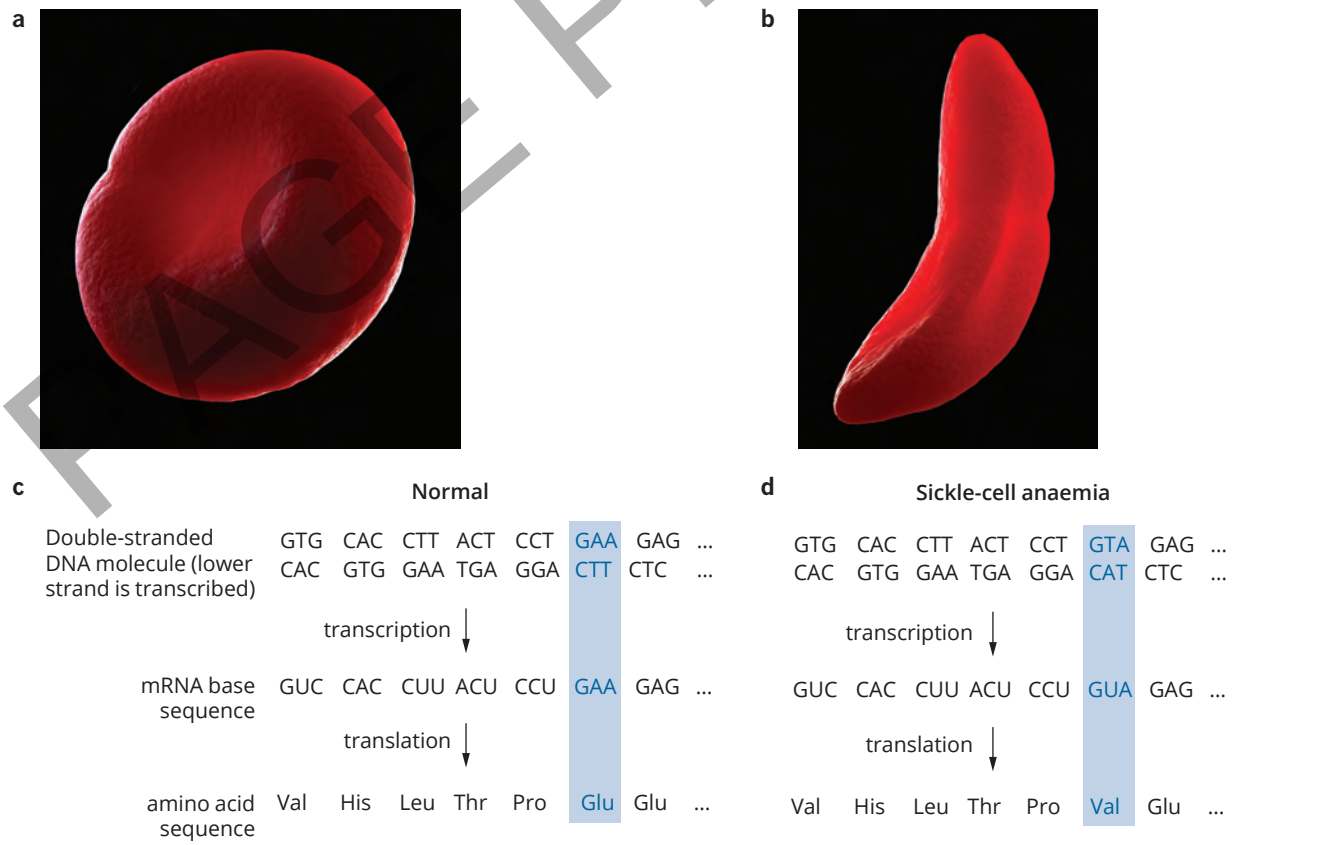


FIGURE 5.2.29 Red blood cells of (a) individuals without and (b) individuals with sickle-cell anaemia. DNA, mRNA and the resulting amino acid sequences of the β-chain of haemoglobin proteins of (c) individuals without and (d) individuals with sickle-cell anaemia.

People who carry an allele from sickle-cell anaemia are very rare in Australia, but its prevalence has been increasing due to immigration of people from other countries where the genetic disorder is more common.

Cystic fibrosis

The addition or deletion of bases from DNA results in mutations that can have very significant effects. In the disease **cystic fibrosis**, an affected individual carries two alleles with mutations of the *CF* gene on chromosome 7. The normal copy of the *CF* gene produces a protein named CFTR (an abbreviation of cystic fibrosis transmembrane conductance regulator), which forms a channel in the outer cell membrane. The CFTR channel controls the flow of chloride out of the cell. Mutations in the gene can lead to defective CFTR channel proteins being produced or, for some mutations, no protein being produced.

Cystic fibrosis symptoms vary from person to person. Affected individuals have high salt levels in their sweat and frequently produce sticky mucus in the lungs, which blocks the airways and increases the chance of infection. The pancreas does not work efficiently in most sufferers and some individuals may also have liver problems.

The CF polypeptide consists of 1480 amino acids. Molecular analysis of the DNA of individuals with cystic fibrosis has shown that one particular mutation, called Δ*F508*, is common. It is present in up to 90% of cases. The Δ*F508* allele has a deletion of three base pairs found in the normal copy of the *CF* gene (Figure 5.2.30). The deletion of the three base pairs results in the deletion of the amino acid phenylalanine from the polypeptide produced by the Δ*F508* allele.

The loss of one amino acid out of 1480 may not seem significant but in this case, it is. In fact, it is so significant that the cell recognises the polypeptide produced by the Δ*F508* allele as being defective and destroys it before it gets to the cell membrane.

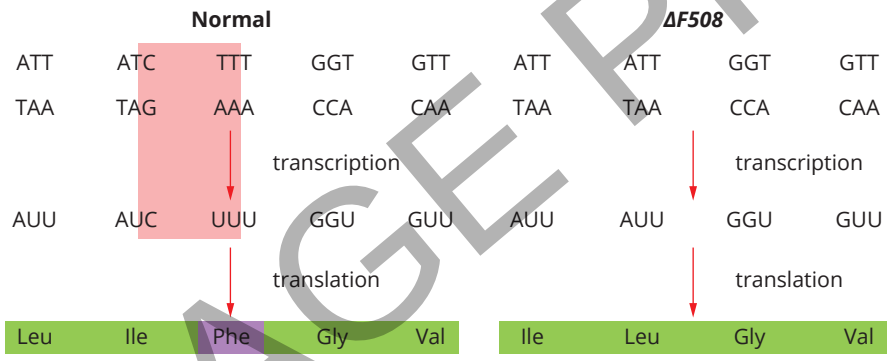


FIGURE 5.2.30 Comparison of the base mRNA and DNA sequences and the resulting amino acid sequences of the CFTR protein from normal individuals and those with cystic fibrosis.

APOA1 gene variant

A possible example of a beneficial substitution point mutation is one that involves the *APOA1* gene. This gene codes for a protein (apolipoprotein A1) that is normally involved in the transport of cholesterol and phospholipids to the liver, where they are then redistributed or broken down and excreted. One of the mutated forms of the protein, ApoA-I Milano, involves a substitution of the amino acid arginine for cysteine. This mutation enhances the protein's antioxidant properties, reducing cholesterol deposition in arteries and significantly decreasing the risk of cardiovascular disease.

Research is being done on *APOA1* mutations to understand links with familial visceral amyloidosis, a condition that involves protein misfolding and deposition in internal organs such as the liver, kidneys and heart.

The mutant form of the ApoA1 protein (Figure 5.2.31) was first identified in Milan, which is why the mutated gene was named after that city. Further investigation, including blood tests of an entire Italian village, traced the origin of the germline mutation to a single man. The 3.5% frequency of the gene in that village population can be attributed to the descendants of this one man.

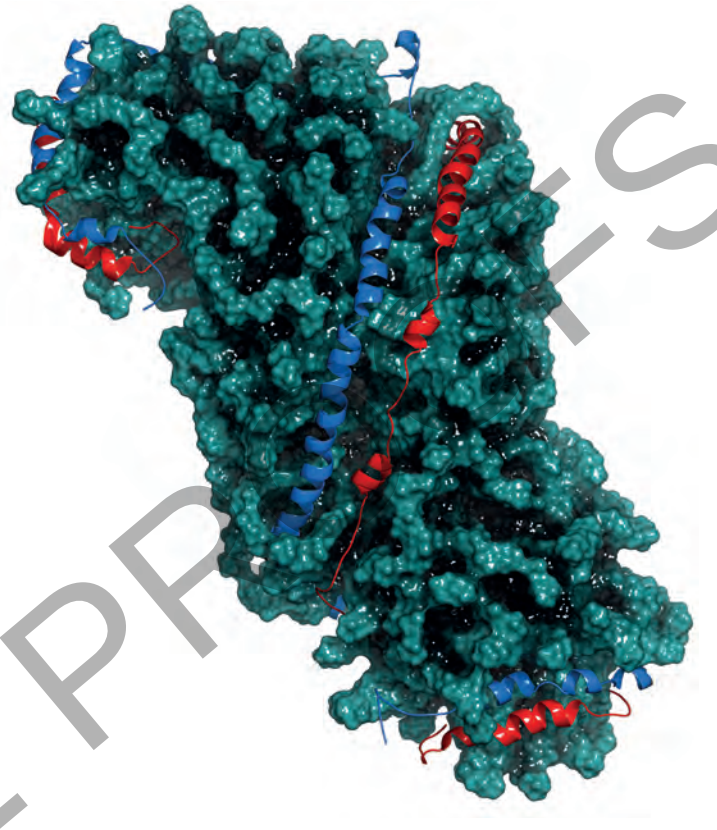


FIGURE 5.2.31 ApoA-I Milano is a mutated form of a protein that can reduce cholesterol levels in the human blood stream. ApoA-I Milano is caused by a beneficial point mutation.

To be considered a beneficial mutation in evolutionary terms (Chapter 9), the ApoA-I Milano allele would need to confer an adaptive advantage that increases its frequency in the population through natural selection. While it has shown potential cardiovascular benefits, its role in broader evolutionary fitness remains a topic of ongoing investigation.

5.2 Review

SUMMARY

- RNA is a short, usually single-stranded nucleic acid that contains nucleotides made up of ribose sugar, a phosphate and one of four nitrogenous bases (adenine, cytosine, guanine and uracil). Types of RNA involved in protein synthesis are: mRNA, rRNA and tRNA.
- The genetic code is the set of rules that governs how the instructions carried in nucleic acids are translated to synthesise gene products (proteins and RNA).
- The stages of protein synthesis are transcription, RNA processing (eukaryotes only) and translation. Gene regulation occurs at any of these stages in eukaryotes and only during transcription in prokaryotes.
- Constitutive genes are expressed continually. Structural genes code for proteins and RNA not involved in gene regulation. Regulatory genes code to produce transcription factors, which are proteins that control gene expression at the transcription stage.
- Master regulatory genes encode transcription factors that control the expression of other genes during embryonic development. Hox genes are a family of master regulatory genes that determine the spatial arrangement of body parts along the anterior–posterior axis.
- Chemical tags are covalent modifications added to DNA or histones that work together with chromatin remodelling proteins to transition chromatin between transcriptionally active euchromatin and transcriptionally inactive heterochromatin states.
- There are many types of DNA mutations. A point mutation is a genetic alteration that occurs when a single base pair in a DNA sequence (or single base in an RNA sequence) is changed, deleted or inserted. A frameshift mutation is a genetic mutation that occurs when a DNA sequence is altered by the insertion or deletion of one or more nucleotides that is not a multiple of three, causing a change in all bases after the point of the mutation.

KEY QUESTIONS

Describe

1 Describe the following structural features of eukaryotic genes and their functions:
- **a** stop and start codons
- **b** promoter regions
- **c** exons
- **d** introns

2 Define:
- **a** transcription factor.
- **b** regulatory gene.
- **c** constitutive gene.

3 Describe the function of the following transcription factor genes:
- **a** Hox genes
- **b** SRY genes

Apply

4 What is the difference between gene expression and gene regulation?

5 What are the different roles of tRNA and rRNA in protein synthesis?

6 Identify the complementary nucleotides and indicate the 5′ and 3′ ends of each nucleotide sequence. Assume no RNA processing occurs.
- **a** coding strand DNA: 5′-A T G T A T G C C A A T C G A 3′
- **b** non-coding strand DNA: _′-T _____ ___-__′
- **c** mRNA: _′-A_____ __-′
- **d** anticodons on tRNA: _′-__ __/__ __/__ __/__ __/__ __/__-__′

7 Construct a table that includes the stages for the following processes and the events that occur at each stage.
- **a** transcription
- **b** translation

8 Explain:
- **a** why single nucleotide substitution mutations are usually much less damaging than frameshift mutations.
- **b** how the p53 protein, encoded by a tumour suppressor gene, prevents cancer.

9 Explain the structural differences between prokaryotic and eukaryotic cells and the effect they have on the rate of protein synthesis.

continued over page

5.2 Review *continued*

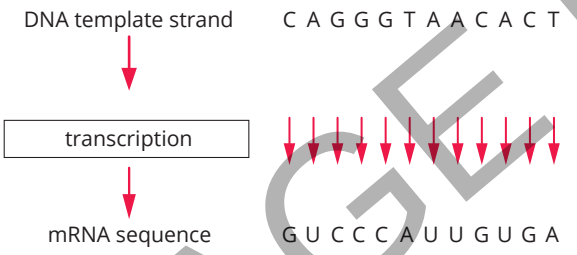
Analyse

To answer questions **10–12**, refer to the genetic code in the table below. With a few rare exceptions, the genetic code is accepted as being universal.

Genetic code translation table

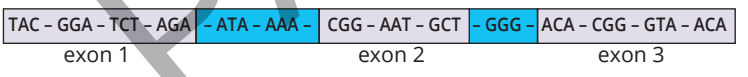
First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

10 The following flow chart represents transcription.



Outline the significance of the fourth codon.

11 The DNA sequence of a particular gene is shown below.



Determine the functional amino acid sequence coded for by this segment of DNA, showing the mature mRNA strand.

12 The DNA sequence of the promoter and first exon of a gene are shown below.

GGGCTCTATAAAGGGTACCACTTCAATGCT

a Determine the amino acid sequence coded for by this DNA template and provide reasoning.

b Suggest the likely consequence for the organism if a mutation changes the sequence of the TATA box.

13 A sample of double-stranded DNA is obtained and the mRNA from this DNA is transcribed. The base composition of each of the two DNA strands and the mRNA strand is analysed and the results are provided in the following table. The numbers indicate the percentage of each base in the strand.

Base composition of DNA and mRNA from sample 1

	A	G	C	T	U
strand 1	40.1	28.9	9.9	0.0	21.1
strand 2	21.5	9.5	29.9	39.1	0.0
strand 3	40.0	29.0	9.7	21.3	0.0

a Identify which of these strands is mRNA.

b Justify which strand is the template strand for the mRNA.

Identify the strands and justify your answer.

14 Construct a table or diagram to compare the structures of RNA and DNA. Include at least three distinct characteristics.

15 The following table shows mutation rates in the p53 gene across various cancer types. Explain the relationship between p53 mutations and the prevalence of cancer.

Cancer type	p53 mutation rate (%)	Prevalence in population (%)
lung cancer	50	12
breast cancer	45	10
colorectal cancer	30	8

Chapter review

05

KEY TERMS

5′ cap
adenine (A)
allele
allosome
anticodon
antiparallel
autosome
base
carcinogen
centromere
chemical tag
chromatid
chromatin
chromosome
circular chromosome
coding DNA
coding strand
codon
complementary base pairing
condensation polymerisation
cystic fibrosis
cytosine (C)
degenerate
deoxyribonucleic acid (DNA)
deoxyribose
diploid
DNA helicase
DNA polymerase
DNA replication
double helix
epigenetic inheritance
epigenetics
exome
exon
frameshift mutation
gene
gene expression
gene regulation
genetic code
genome
germline mutation
guanine (G)
haemoglobin
haploid
heterogametic
heterozygous
histone
homeotic gene
homogametic
homologous chromosome
homozygous
Hox gene
hydrogen bond
induce
induced mutation
intron
lac operon
lac repressor
lacI
lactose
linear chromosome
locus
looped domain
messenger RNA (mRNA)
monomer
mutagen
mutation
nitrogenous base
non-coding DNA
nonsense mutation
non-sister chromatid
non-synonymous substitution mutation
nucleic acid
nucleosome
nucleotide
operator
operon
phosphodiester bond
plasmid
point mutation
poly-A tail
polymer
polynucleotide chain
promoter region
protein
protein synthesis
purine
pyrimidine
regulatory gene
replication fork
repressed
ribonucleic acid (RNA)
ribose
ribosomal RNA (rRNA)
RNA polymerase
RNA processing
sex-determining region Y (SRY)
sister chromatid
somatic mutation
spliceosome
splicing
spontaneous mutation
structural gene
synonymous mutation
telomere
template strand
thymine (T)
trait
transcription
transcription factor
transfer RNA (tRNA)
translation
triplet
tumour-suppressor gene
uracil (U)

KEY QUESTIONS

Describe

1 The three parts of a nucleotide are:
- **A** sugar, phosphate, base
- **B** phosphate, base, protein
- **C** thymine, sugar, protein
- **D** protein, sugar, phosphate

2 Identify the correct statement regarding the structure of DNA.
- **A** Deoxyribose is a six-carbon sugar.
- **B** The base C always pairs with the base G.
- **C** The building blocks are called nucleosomes.
- **D** The DNA molecule is a single-stranded helix.

3 In polypeptide synthesis, the function of the ribosome is to:
- **A** synthesise amino acids.
- **B** ensure that DNA transcription is complete.
- **C** provide the energy needed for the synthesis.
- **D** enable the information in the mRNA to be translated into a chain of amino acids.

4 Identify which of the following statements is true about the DNA code for the synthesis of polypeptides.
- **A** The code is read as sets of three bases called triplets.
- **B** Every codon codes for its own exclusive amino acid.
- **C** Each triplet codes for at least two different amino acids.
- **D** There are 20 different amino acids, therefore there are 20 different codons.

5 Identify the name given to the nucleotide sequences that are included in the final mRNA product.
- **A** exons
- **B** introns
- **C** termons
- **D** spliced codons

6 A promoter is:
- **A** a specific sequence of DNA to which DNA polymerase may bind.
- **B** a specific sequence of DNA to which RNA polymerase may bind.
- **C** a specific sequence of DNA to which a repressor may bind.
- **D** a specific sequence of RNA in mRNA.

7 Describe the difference between the DNA of eukaryotic cells and prokaryotic cells.

8 Describe the function of the codons UAA and AUG.

9 Explain the effect of a single base pair deletion in the DNA of a gene.

10 Sketch a labelled diagram to symbolise a:
- **a** nucleotide with adenine as its base.
- **b** strand of DNA, including paired bases.

Apply

11 Identify the three main types of RNA involved in transcription and translation. Summarise the differences between the types of RNA by completing the table.

Type of RNA	Where it is produced	Where and how it functions in cells

12 Identify the genetic structure represented by the nucleotide sequence
A G U G A C C A A.
- **A** a sequence of mRNA
- **B** a section of double helix
- **C** the amino acid chain of a polypeptide
- **D** part of the DNA template of a particular gene

13 In analysing the number of different bases in a DNA sample, determine which formula would be consistent with the base-pairing rules.
- **A** A = G
- **B** A = C
- **C** G = T
- **D** A + T = G + T
- **E** A + G = T + C

14 3′-T T C A G T C G T-5′
Write the mRNA sequence and the coding DNA sequence, indicating the 5′ and 3′ ends, then compare the two sequences.

15 The template strand of a gene includes this sequence:
3′-T A C T T G T C C G A T A T C-5′
It is mutated to:
3′-T A C T T G T C C A A T A T C-5′
Using the following codon chart, justify the effect of the mutation on the amino acid sequence.

First position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

16 Contrast splicing and alternative splicing in eukaryotic cells to explain how human cells can make 75 000–100 000 different proteins, given that there are about 20 000 genes.

17 Explain how somatic mutations differ in their effect on generations compared with germline mutations.

18 What type of mutation leads to sickle-cell anaemia and how does it cause this disease?

19 Complete the table below to show the differences between synonymous mutations, non-synonymous mutations, nonsense mutations and frameshift mutations.

Mutation type	Characteristic	Implications of mutation
synonymous		
non-synonymous		
nonsense		
frameshift		

20 Compare a spontaneous mutation to an induced mutation.

21 Explain the main function of homeotic genes and describe a consequence if a mutation occurs in one of these genes.

22 Explain the roles of the following enzymes in DNA replication.

a DNA helicase

b DNA polymerase

23 How does semi-conservative replication ensure the accuracy of DNA replication?

24 What is the function of a stop codon?

25 How do transcription factors regulate gene expression?

26 Explain why replication of the lagging strand is different to that of the leading strand.

Analyse

27 Compare the gene expression processes of prokaryotic and eukaryotic organisms.

28 Using the codon chart from Question **15**, determine the 5′–3′ sequence of nucleotides in the DNA template strand for an mRNA coding for the polypeptide sequence Phe–Pro–Lys.

A 5′-GAACCCCTT-3′

B 5′-AAAGGCTTT-3′

C 5′-CTTCGGGAA-3′

D 5′-AAACCCUUU-3′

E 5′-UUUGGGAAA-3′

29 A strand of nucleic acid is shown below.

AUG AAU CCU UAU GGU GGC UUU UAA

The protein produced as a result of the information encoded in this strand is shown below.

Met–Asn–Pro–Phe

a Justify whether the strand given is DNA, pre-mRNA or mRNA.

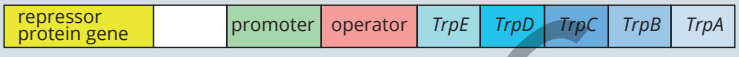
b Determine the amino acid being transported by tRNA for translation of the nucleic acid sequence above at the second codon.

30 Outline why, in humans, a mature adult intestinal cell will produce digestive enzymes while endocrine gland cells produce hormones, even though both differentiated cells contain the same human genome and secrete proteins.

31 Predict the effect on protein synthesis of a mutation to the TATA box due to X-rays.

32 A mutation in *E. coli* changes the *lac* operator so that the active repressor cannot bind. Predict the effect on the cell's production of β-galactosidase. Refer to the type of gene expression involved.

33 Tryptophan is an amino acid that prokaryotic cells are able to synthesise when it is in short supply in the environment. Tryptophan production is controlled by a series of five enzymes that are coded for by five genes (*trp A*, *B*, *C*, *D* and *E*), which form a single unit called the tryptophan operon. This operon has been extensively studied in the bacterium *E. coli*. The *trp* operon and some of its upstream region are shown below.



The gene for the repressor protein is constitutively expressed and is produced in an inactive form. It becomes activated when it binds to tryptophan.

a Deduce the expression of genes when *E. coli* has an abundance of tryptophan in its diet.

b Compare the means of repression of the *lac* operon and the *trp* operon.

34 Single-celled organisms such as amoeba often live in environments that change quickly. Consider how the control of gene expression ensures that an amoeba is able to respond effectively to frequent short-term environmental change.

35 The table below shows the effects of point mutations on codons and their corresponding amino acids. Interpret the data to explain how these mutations could impact protein synthesis.

Mutation type	Original codon	Mutated codon	Result
none	AUG	AUG	methionine
substitution	AUG	GUG	valine
nonsense mutation	AUG	UAA	stop

36 With reference to chromatin states and cell specialisation, explain how the expression of genes involved in the synthesis of contractile proteins (like myosin and actin) and neurotransmitters would differ between muscle and brain cells.

Interpret

37 Consider which of the following mutations be would likely to cause the most harmful effect on an organism. Provide an evaluation of each.

a a nucleotide-pair substitution

b a deletion of three nucleotides near the middle of a gene

c a single nucleotide deletion in the middle of an intron

d a single nucleotide deletion near the end of the coding sequence

e a single nucleotide insertion downstream of and close to the start of the coding sequence

Data analysis

DATA SET 1

The information below applies to Questions **1–5**.

An experiment was conducted to investigate the effect of nucleotide availability and temperature on the rate of DNA replication in a prokaryotic organism. The experiment measured the time taken to replicate a circular chromosome under different conditions. A second part of the experiment measured the rate of protein synthesis by analysing the production of a fluorescent protein encoded by the organism's genome. The data is presented in the table below.

Effect of nucleotide availability and temperature on DNA replication

Temperature (°C)	Nucleotide concentration (%)	Replication time (min)	Protein production (units/min)
25	50	15	200
25	100	10	350
37	50	8	400
37	100	5	600
45	50	20	150
45	100	12	250

Question 1 (2 marks)
Identify the independent and dependent variables in this experiment.

Question 2 (3 marks)
Compare the impact of temperature on DNA replication at 25°C, 37°C, and 45°C at a nucleotide concentration of 100%.

Question 3 (3 marks)
Analyse the relationship between nucleotide concentration and fluorescent protein production at 37°C.

Question 4 (4 marks)
Predict and justify how replication time and protein production might change at 50°C based on the data provided.

Question 5 (3 marks)
Explain why DNA replication and protein production are interdependent processes.