

Molecules (Continuity and change) revision notes

How DNA is copied, how its base sequence is used to make polypeptides, and how changes to that sequence arise and can be made on purpose. The group covers DNA replication with PCR and gel electrophoresis, protein synthesis, and mutation and gene editing.

For IB Diploma Biology, Standard Level and Higher Level (first assessment 2025). **HL only** is additional higher level content, for HL students only.

1 DNA replication **Spec D1.1**

Exact copies by semi-conservative replication **Spec D1.1.1, D1.1.2**

DNA replication is the production of exact copies of DNA with identical base sequences. It happens before a cell divides, so that each daughter cell receives a full copy of the genome.

- **Reproduction:** DNA must be copied so that it can be passed to offspring, in asexual reproduction and in the production of gametes.
- **Growth:** a multicellular organism grows by mitosis, and every new cell needs a complete copy of the DNA.
- **Tissue replacement:** cells lost through damage or wear (for example skin cells or red blood cell precursors) are replaced by dividing cells, which again need copied DNA.

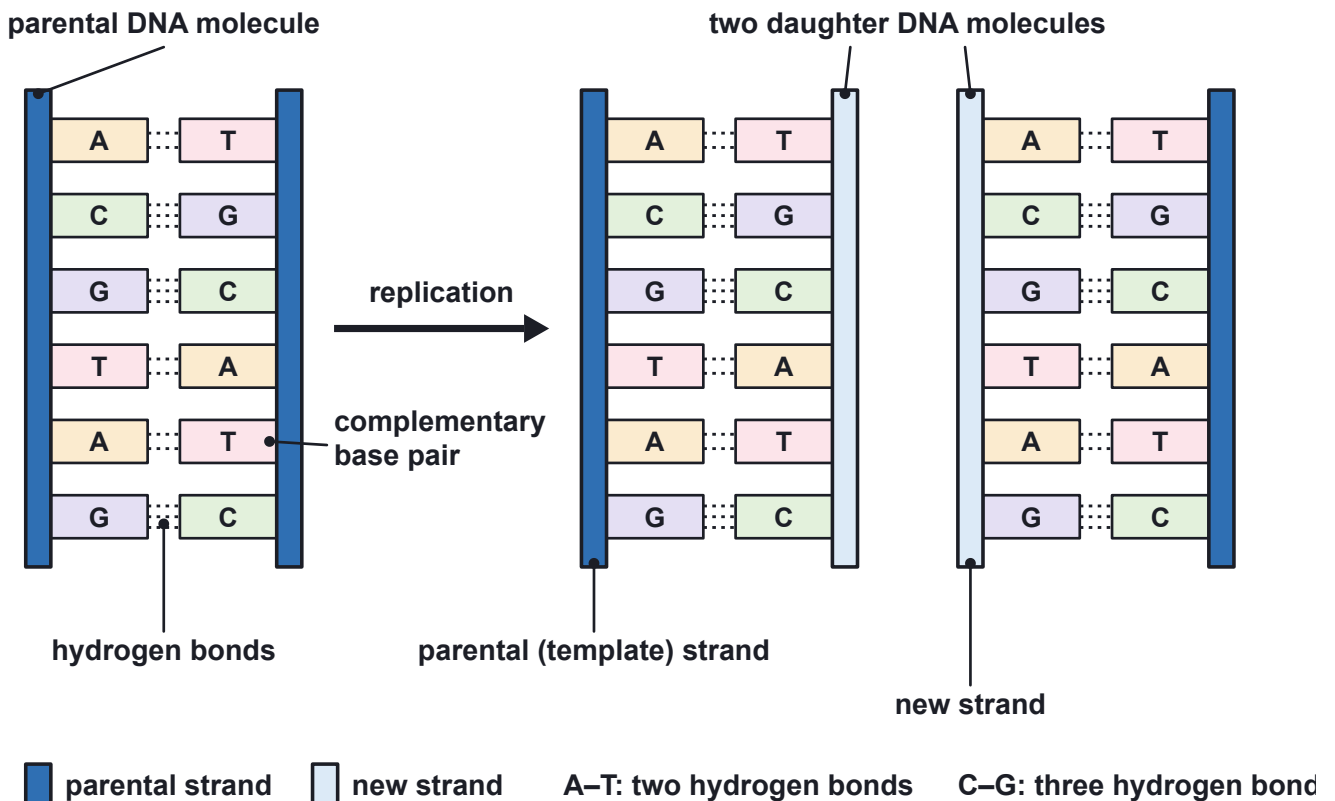
Semi-conservative replication

The two strands of the double helix separate and each one is used as a **template** for a new strand. Each of the two DNA molecules produced therefore has one strand from the original molecule (conserved) and one new strand. This is called **semi-conservative replication**.

Complementary base pairing and accuracy

Free DNA nucleotides are added to the new strand according to **complementary base pairing**: adenine pairs only with thymine (two hydrogen bonds) and cytosine pairs only with guanine (three hydrogen bonds). Because the base on the template strand decides which base is added, the new strand has exactly the sequence complementary to the template, and the new double helix is identical to the original.

A base that does not pair correctly does not fit or form the right hydrogen bonds, so mismatches are rare. This is why replication copies base sequences with a very high degree of accuracy.



[\(opens full size in a new tab\)](#)

Semi-conservative replication: each daughter molecule keeps one parental strand and gains one new strand.

Common mistake: Semi-conservative does not mean half of each strand is new. Each daughter molecule has one complete old strand and one complete new strand.

Roles of helicase and DNA polymerase Spec D1.1.3

Two enzymes are central to replication.

Helicase	Unwinds the double helix and separates the two strands by breaking the hydrogen bonds between complementary bases. Each separated strand can then act as a template.
DNA polymerase	Assembles a new strand: it brings free DNA nucleotides into position opposite their complementary bases on the template strand and links them together with covalent (phosphodiester) bonds to form the sugar-phosphate backbone.

1. Helicase moves along the DNA, unwinding it and breaking hydrogen bonds between base pairs, so the strands separate.
2. Free DNA nucleotides pair with exposed bases on each template strand (A-T, C-G).
3. DNA polymerase links the paired nucleotides into a continuous new strand.
4. Two DNA molecules result, each with one old and one new strand.

Exam tip: Say helicase breaks hydrogen bonds between bases, not covalent bonds. DNA polymerase forms covalent bonds between adjacent nucleotides of the new strand.

PCR and gel electrophoresis Spec D11.4

Polymerase chain reaction (PCR)

The **polymerase chain reaction (PCR)** makes millions of copies of a chosen DNA sequence from a tiny sample. The reaction mixture contains the sample DNA, free DNA nucleotides, two **primers** and **Taq polymerase**, in a buffer. A thermal cycler changes the temperature in repeated cycles.

One cycle of PCR

Stage	Typical temperature	What happens
Denaturation	about 95 °C	Hydrogen bonds between the strands break, so the double helix separates into single strands.
Annealing	about 50–65 °C (often about 55 °C)	Primers bind by complementary base pairing to the ends of the target sequence on each single strand.
Extension	about 72 °C	Taq polymerase adds nucleotides to the primers, building a new complementary strand.

- **Primers** are short single-stranded sequences of DNA. Two different primers are used, one for each strand, and they mark the two ends of the sequence to be copied. DNA polymerase can only add nucleotides to an existing strand, so the primer gives it a starting point.
- **Taq polymerase** is a DNA polymerase from *Thermus aquaticus*, a bacterium of hot springs. It is not denatured at 95 °C and works best at about 72 °C, so it survives every cycle and does not need to be replaced.
- **Temperature changes** drive the cycle: high temperature separates strands, a lower temperature allows primers to bind, and an intermediate temperature suits Taq polymerase.

Each cycle doubles the number of copies of the target sequence, so the amount of DNA increases exponentially. Thirty cycles typically take one to two hours.

$$\text{number of copies} = \text{starting number of copies} \times 2^n \text{ (n = number of cycles)}$$

Maths skill: copies after PCR

A sample contains 50 copies of a target DNA sequence. How many copies are present after 20 cycles of PCR, assuming each cycle doubles the DNA?

1. $2^{20} = 1\,048\,576$

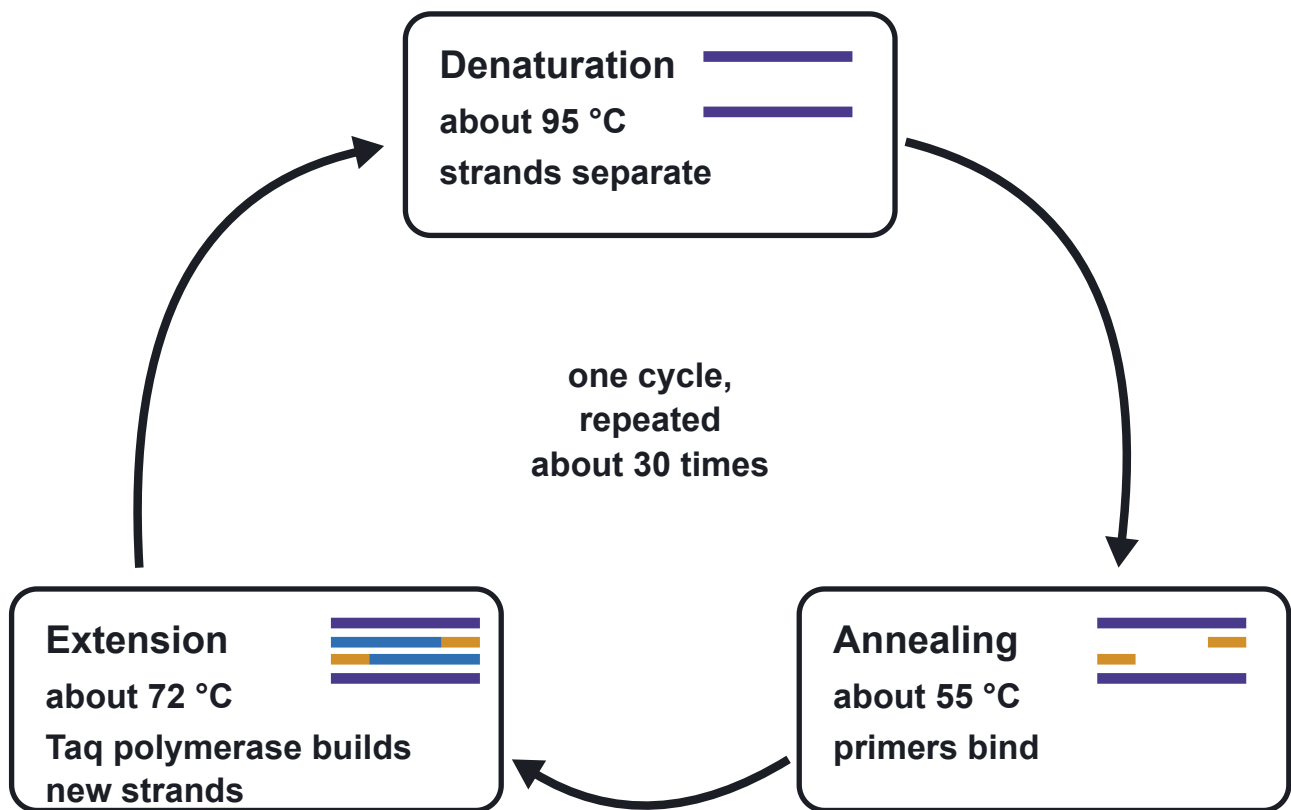
2. $50 \times 1\,048\,576 = 52\,428\,800$

Answer: About 5.2×10^7 copies.

Gel electrophoresis

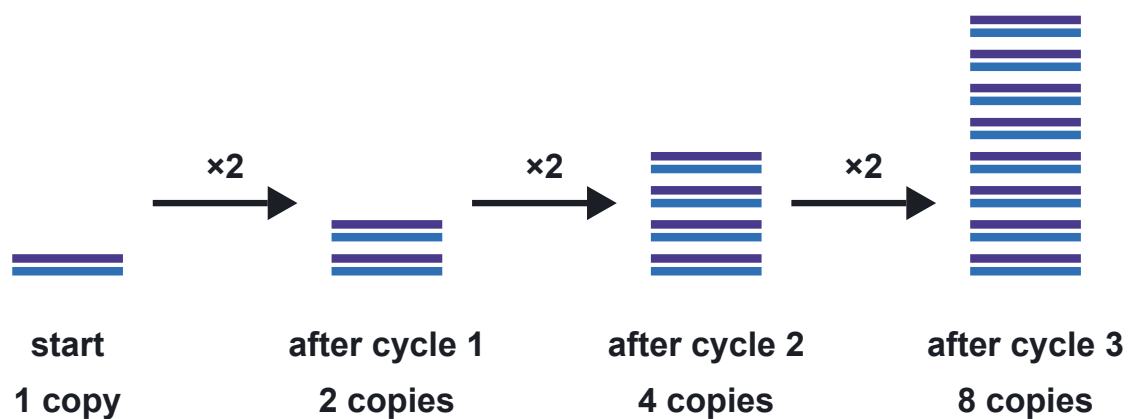
Gel electrophoresis separates DNA fragments according to their length. Samples are placed in wells at one end of an agarose gel and a voltage is applied. DNA has a **negative charge** (from its phosphate groups), so fragments move through the gel towards the positive electrode. The gel acts like a mesh: **smaller fragments move faster and travel further** than larger ones in the same time.

A DNA ladder (a mixture of fragments of known lengths) is run in one lane so that fragment sizes can be estimated. The DNA is stained so that the bands can be seen.



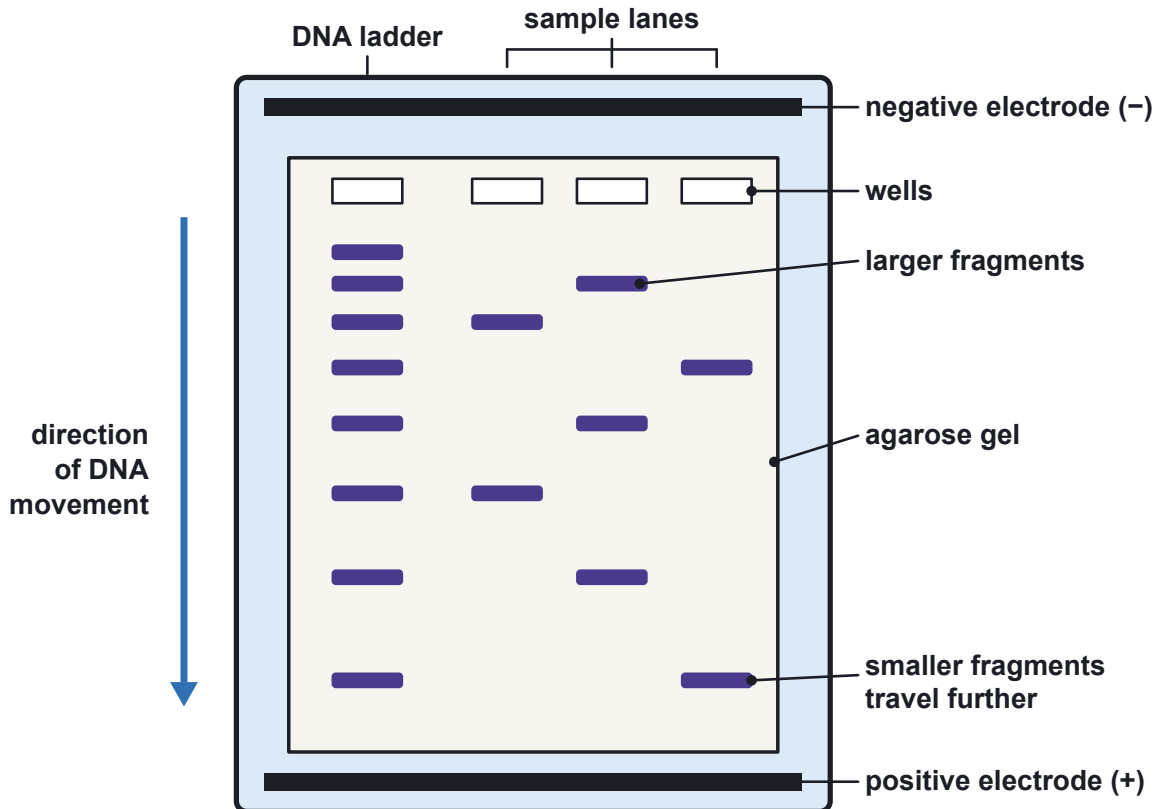
— template DNA
 — primer
 — new strand

copies of the target sequence double every cycle



[\(opens full size in a new tab\)](#)

One cycle of **PCR**. Each cycle doubles the number of copies of the target sequence.



[\(opens full size in a new tab\)](#)

DNA is negatively charged, so it moves towards the **positive electrode**; **smaller fragments travel further**.

Practical skill: When reading a gel, compare each band with the ladder lane. Bands at the same distance from the wells in different lanes are fragments of the same length.

Applications of PCR and gel electrophoresis Spec D1.1.5

Because PCR can amplify a chosen sequence from a tiny sample and gel electrophoresis can separate fragments by length, the two techniques have a broad range of applications.

- **DNA profiling** for paternity and forensic investigations.
- **Detecting pathogens:** PCR detects DNA (or RNA copied into DNA) of a virus or bacterium in a patient sample, as in tests for viral infections.
- **Genetic screening:** testing for alleles that cause genetic conditions.
- **Research:** amplifying genes for sequencing or for comparing species.

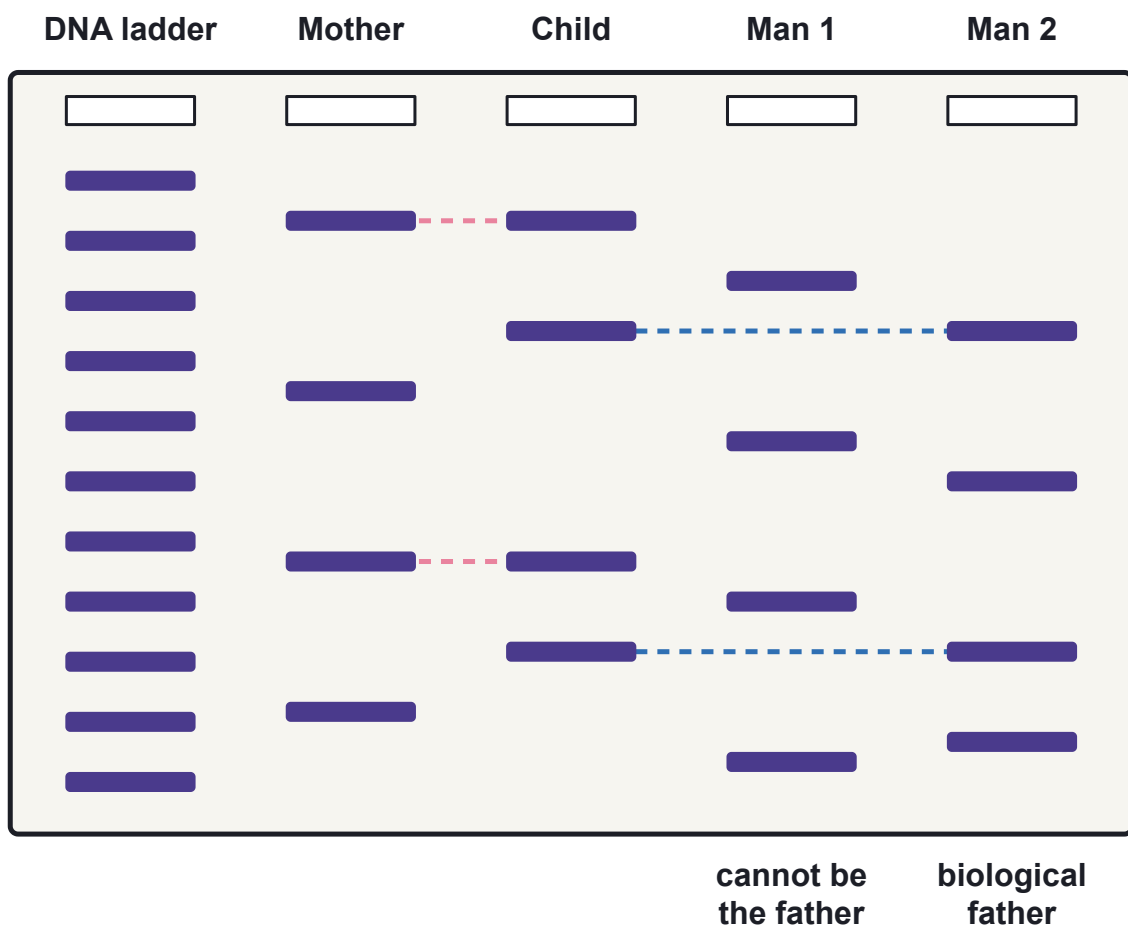
DNA profiling

Some regions of the genome contain **short tandem repeats (STRs)**: a short base sequence repeated many times. The number of repeats at each region varies between individuals, so the length of the region varies. In **DNA profiling**, several of these regions are amplified by PCR and the fragments are separated by gel electrophoresis, giving a pattern of bands (a DNA profile) that is almost unique to each person.

Using DNA profiles

Use	How profiles are compared
Paternity	A child inherits one allele of each region from each parent. Every band in the child's profile must match a band in the mother's or in the father's profile. A man whose profile lacks the non-maternal bands cannot be the father.
Forensic investigation	DNA from blood, hair or skin cells at a crime scene is amplified and profiled, then compared with profiles from suspects. A match links a person to the scene; a mismatch excludes them.

Exam tip: Nature of science: reliability is enhanced by increasing the number of measurements. In DNA profiling, using more markers (more STR regions) reduces the probability of a false match between two different people.



--- child's band matches a band from the mother

--- child's band matches a band from Man 2

[\(opens full size in a new tab\)](#)

Paternity test: each band in the child comes from one parent, so Man 2 is the father and Man 1 is excluded.

Directionality: leading and lagging strands Spec D1.1.6, D1.1.7 HL only

5' and 3' ends

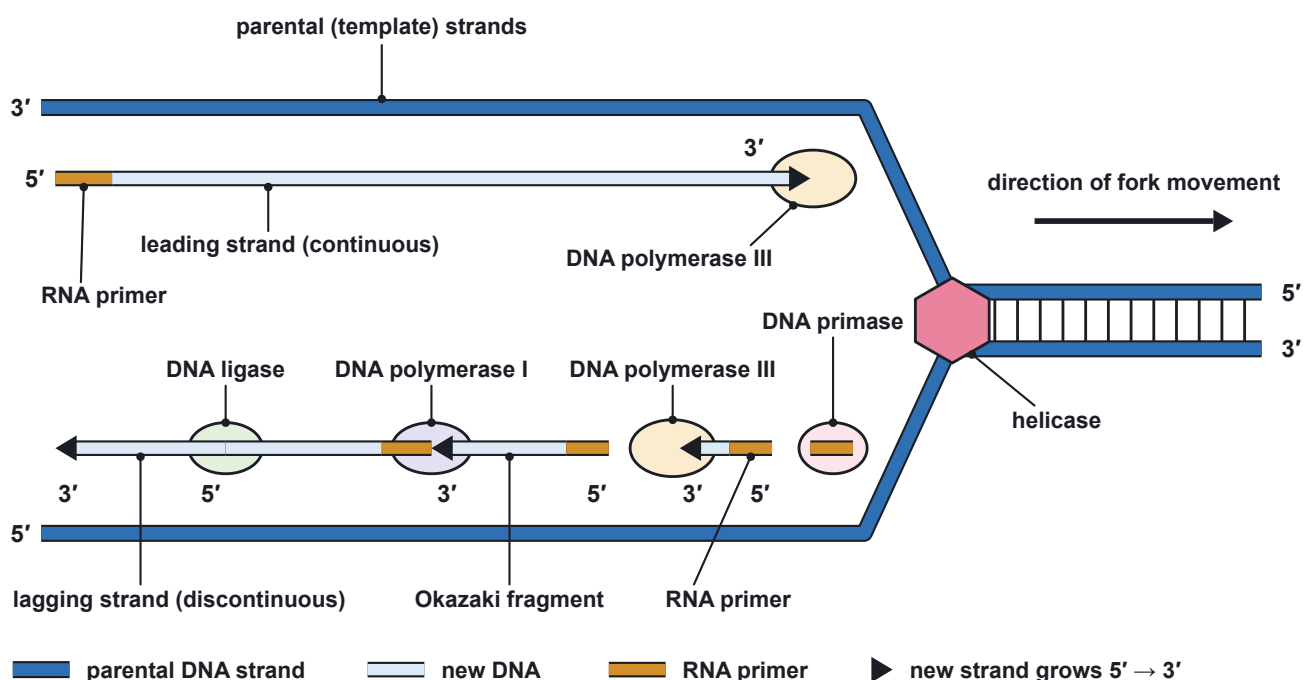
Each nucleotide in a strand is linked to the next through its deoxyribose. The **5' terminal** of a strand has a phosphate group attached to carbon 5' of the sugar; the **3' terminal** has a free -OH group on carbon 3'. The two strands of the double helix are antiparallel: they run in opposite directions.

DNA polymerases add the 5' of a DNA nucleotide to the 3' end of a strand of nucleotides. The phosphate on carbon 5' of the incoming nucleotide is bonded to the 3' -OH of the strand. Free nucleotides arrive as deoxynucleoside triphosphates, and the release of two phosphates provides the energy for the bond. New DNA therefore always grows in the 5' → 3' direction.

Leading and lagging strands

Replication at a replication fork		
Feature	Leading strand	Lagging strand
Direction of synthesis relative to the fork	Same direction as the fork moves	Opposite direction to the fork
Type of replication	Continuous	Discontinuous, in short sections called Okazaki fragments
RNA primers	Needed only once, at the start	Needed repeatedly, one for each Okazaki fragment
Joining	Little joining: the single primer is replaced by DNA and the one nick sealed by DNA ligase	Fragments joined by DNA ligase

Because the template strands are antiparallel and synthesis can only go 5' → 3', one new strand can follow the helicase as the fork opens, while the other must be made in short pieces backwards from the fork, each started afresh as more template is exposed.



[\(opens full size in a new tab\)](#)

At a replication fork, new DNA always grows 5' → 3': the **leading strand** is continuous and the **lagging strand** is made in Okazaki fragments.

Exam tip: Linking question: what biological mechanisms rely on directionality? Replication, transcription and translation all proceed 5' to 3', which follows from the structure of the sugar-phosphate backbone.

Replication enzymes and DNA proofreading Spec D1.1.8, D1.1.9 HL only

The functions below are for the **prokaryotic system**, as the IB requires.

Enzymes of prokaryotic DNA replication

Enzyme	Function
DNA primase	Synthesises a short RNA primer complementary to the template. This provides the 3' end that DNA polymerase III needs to start adding nucleotides. Once on the leading strand; at the start of every Okazaki fragment on the lagging strand.
DNA polymerase III	Adds DNA nucleotides to the 3' end of the primer and then of the growing strand, 5' → 3', following complementary base pairing. Makes most of the new DNA. Also proofreads.
DNA polymerase I	Removes the RNA primers nucleotide by nucleotide and replaces them with DNA nucleotides.
DNA ligase	Seals the nicks between adjacent sections of DNA (for example between Okazaki fragments) by forming a covalent bond in the sugar-phosphate backbone.

DNA proofreading

Occasionally DNA polymerase III adds a nucleotide with a base that does not pair with the template. A mismatched base pair distorts the double helix at the 3' end of the new strand. DNA polymerase III detects this, **removes the nucleotide with the mismatched base from the 3' terminal** and then adds a correctly matched nucleotide in its place before continuing. **Proofreading** greatly reduces the error rate of replication.

Common mistake: In IB answers, proofreading of the new strand is credited to DNA polymerase III. The role to give for DNA polymerase I is removing RNA primers and replacing them with DNA (it can also correct errors, but this is not required).

Quick check

1. Why is DNA replication described as semi-conservative?
2. What is the role of helicase?
3. Why is Taq polymerase used in PCR?
4. Why do smaller DNA fragments travel further in gel electrophoresis?
5. **HL only** Why are many RNA primers needed on the lagging strand?

2 Protein synthesis Spec D1.2

Transcription and gene expression Spec D1.2.1, D1.2.2, D1.2.3, D1.2.4

Transcription is the synthesis of RNA using a DNA template. Only one of the two DNA strands, the **template strand**, is transcribed. The other strand is the coding strand: the RNA has the same base sequence as the coding strand, except that uracil replaces thymine.

Roles of RNA polymerase

1. RNA polymerase binds to DNA at the start of a gene.
2. It unwinds the double helix and separates the two strands by breaking hydrogen bonds between bases, over a short stretch of DNA.
3. Free RNA nucleotides pair with exposed bases on the template strand by complementary base pairing.
4. RNA polymerase links the RNA nucleotides with covalent bonds to form a single strand of RNA.
5. Behind the polymerase, the RNA separates from the template and the DNA strands re-form their hydrogen bonds and rewind.

Hydrogen bonding and complementary base pairing

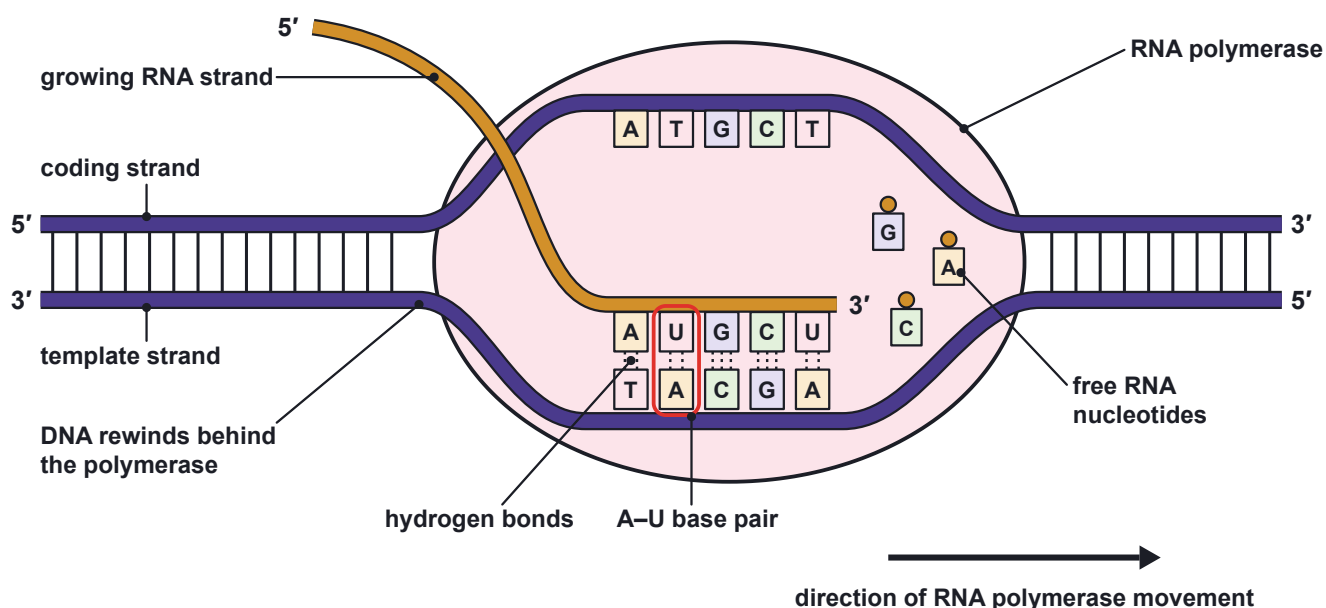
Each RNA nucleotide is held in place by **hydrogen bonds** to the complementary base on the template strand. The pairs are: **adenine (A) on the DNA template with uracil (U) on the RNA**, thymine (T) on the template with adenine (A), cytosine (C) with guanine (G) and guanine (G) with cytosine (C). The hydrogen bonds are weak, so the RNA can separate from the template once its nucleotides are linked.

Stability of DNA templates

A single DNA strand can be used as a template for transcribing a base sequence without the DNA base sequence changing: only hydrogen bonds are broken and then re-formed. In **somatic cells that do not divide**, such as many neurones, the same DNA must be transcribed again and again, so its sequences must be conserved throughout the life of the cell.

Transcription and gene expression

Not all genes in a cell are expressed at any given time. Transcription is the **first stage of gene expression**, so it is a key stage at which the expression of a gene can be switched on or off. A gene that is not transcribed produces no mRNA and so no polypeptide.



[\(opens full size in a new tab\)](#)

Transcription: RNA polymerase builds RNA that is **complementary to the template strand**, held in place by hydrogen bonds.

Exam tip: Linking question: what biological processes depend on hydrogen bonding? Replication, transcription and codon-anticodon pairing all rely on hydrogen bonds being strong enough to hold bases in place but weak enough to break easily.

Translation at the ribosome Spec D1.2.5, D1.2.6, D1.2.7, D1.2.10

Translation is the synthesis of polypeptides from mRNA. The base sequence of the mRNA is translated into the amino acid sequence of a polypeptide. It takes place at ribosomes in the cytoplasm.

Roles of the molecules in translation

Molecule	Role
mRNA	Carries the base sequence copied from a gene. It binds to the small subunit of the ribosome . Its bases are read in groups of three (codons).
Ribosome	Made of a small and a large subunit (rRNA and protein). Two tRNAs can bind simultaneously to the large subunit , so their amino acids are held next to each other. The ribosome catalyses peptide bond formation.
tRNA	Each tRNA carries a specific amino acid and has an anticodon of three bases that pairs with a matching codon on the mRNA.

Codons and anticodons

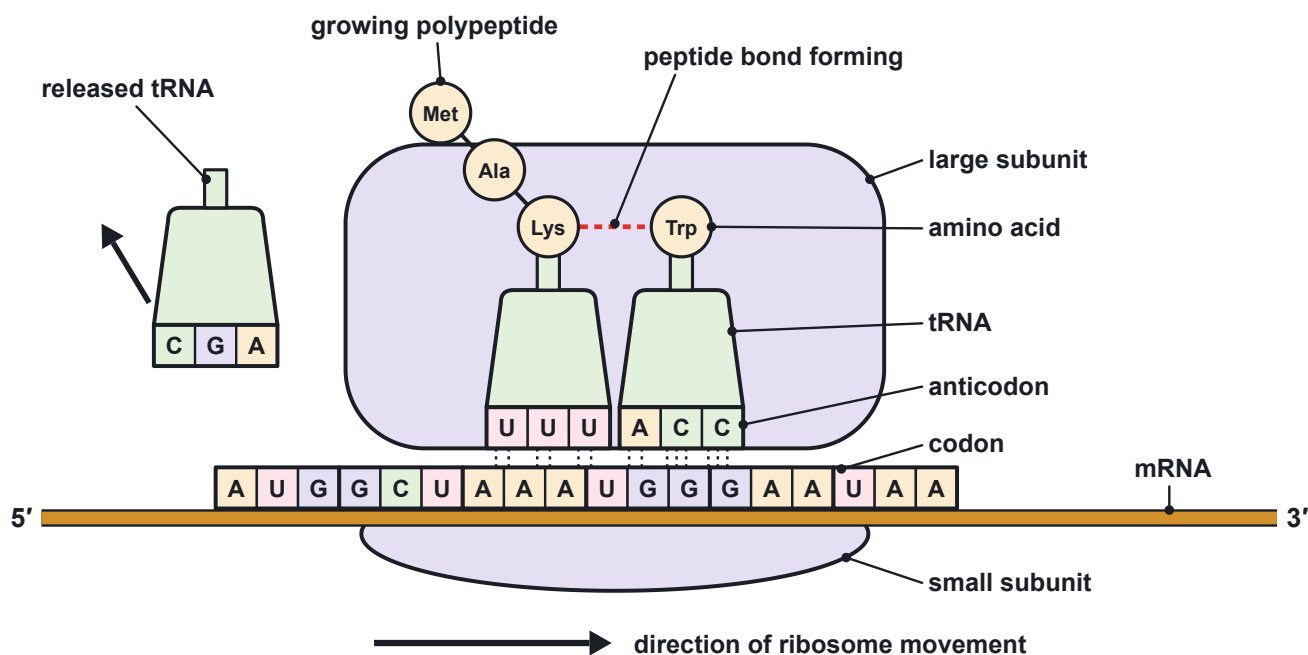
A **codon** is a sequence of three bases on mRNA that codes for one amino acid. An **anticodon** is the complementary sequence of three bases on a tRNA. The anticodon pairs with the codon by complementary base pairing and hydrogen bonding (A with U, C with G), which ensures the correct amino acid is added. For example, the codon AUG pairs with the anticodon UAC.

Elongation of the polypeptide

1. A tRNA carrying the growing polypeptide is bound to the ribosome, with its anticodon paired to a codon on the mRNA.

2. A second tRNA, whose anticodon is complementary to the next codon, binds alongside it, bringing its amino acid.
3. A **peptide bond** forms between the amino acid on the second tRNA and the polypeptide, so the polypeptide is transferred to the second tRNA and becomes one amino acid longer.
4. The ribosome moves along the mRNA by one codon (three bases). The first tRNA, now without an amino acid, is released and can collect another amino acid.
5. The next codon is exposed and the cycle repeats, adding one amino acid at a time until a stop codon is reached.

This stepwise movement of the ribosome along the mRNA, one codon at a time, means the amino acids are linked in exactly the order set by the codons.



[\(opens full size in a new tab\)](#)

Elongation: **two tRNAs bind to the large subunit at once**, so a peptide bond can join the polypeptide to the next amino acid.

Common mistake: Codons are on mRNA and anticodons are on tRNA. Do not describe the anticodon as part of the mRNA or the codon as part of the tRNA.

The genetic code Spec D1.2.8, D1.2.9

Why a triplet code

There are four bases in mRNA and 20 amino acids in proteins. A code of single bases would give 4 combinations, and pairs of bases would give $4^2 = 16$, which is not enough. Triplets give $4^3 = 64$ codons, which is more than enough for 20 amino acids plus start and stop signals.

Degeneracy	More than one codon can code for the same amino acid. For example, GAA and GAG both code for glutamic acid. Of the 64 codons, 61 code for amino acids and 3 are stop codons.
Universality	The same codons code for the same amino acids in almost all living organisms, with only minor variations. This is evidence for common ancestry and allows genes to be transferred between species.
Start codon	AUG, which codes for methionine and marks where translation begins.
Stop codons	UAA, UAG and UGA, which do not code for an amino acid and signal the end of translation.

Using a table of mRNA codons

Codon tables give the amino acid for each mRNA codon. Read the mRNA from the start codon in groups of three, without overlapping, and look up each codon. If you are given DNA, first work out the mRNA: it is complementary to the template strand, and the same as the coding strand with U in place of T.

Part of the mRNA codon table (practice extract)			
Codon	Amino acid	Codon	Amino acid
AUG	Met (start)	GAA	Glu
GCU	Ala	GAG	Glu
AAA	Lys	GUG	Val
UGG	Trp	UUU	Phe
CAU	His	UAA	stop

Maths skill: deducing an amino acid sequence

The template strand of a short DNA sequence reads TAC CGA TTT ACC CTT ATT. Deduce the amino acid sequence.

1. Write the complementary mRNA (A→U, T→A, C→G, G→C): AUG GCU AAA UGG GAA UAA.
2. Look up each codon: AUG = Met, GCU = Ala, AAA = Lys, UGG = Trp, GAA = Glu, UAA = stop.

Answer: Met - Ala - Lys - Trp - Glu (then stop).

Exam tip: Check whether a sequence is DNA (contains T) or mRNA (contains U) before using a codon table. Codon tables are always written for mRNA.

Mutations that change protein structure Spec D1.2.11

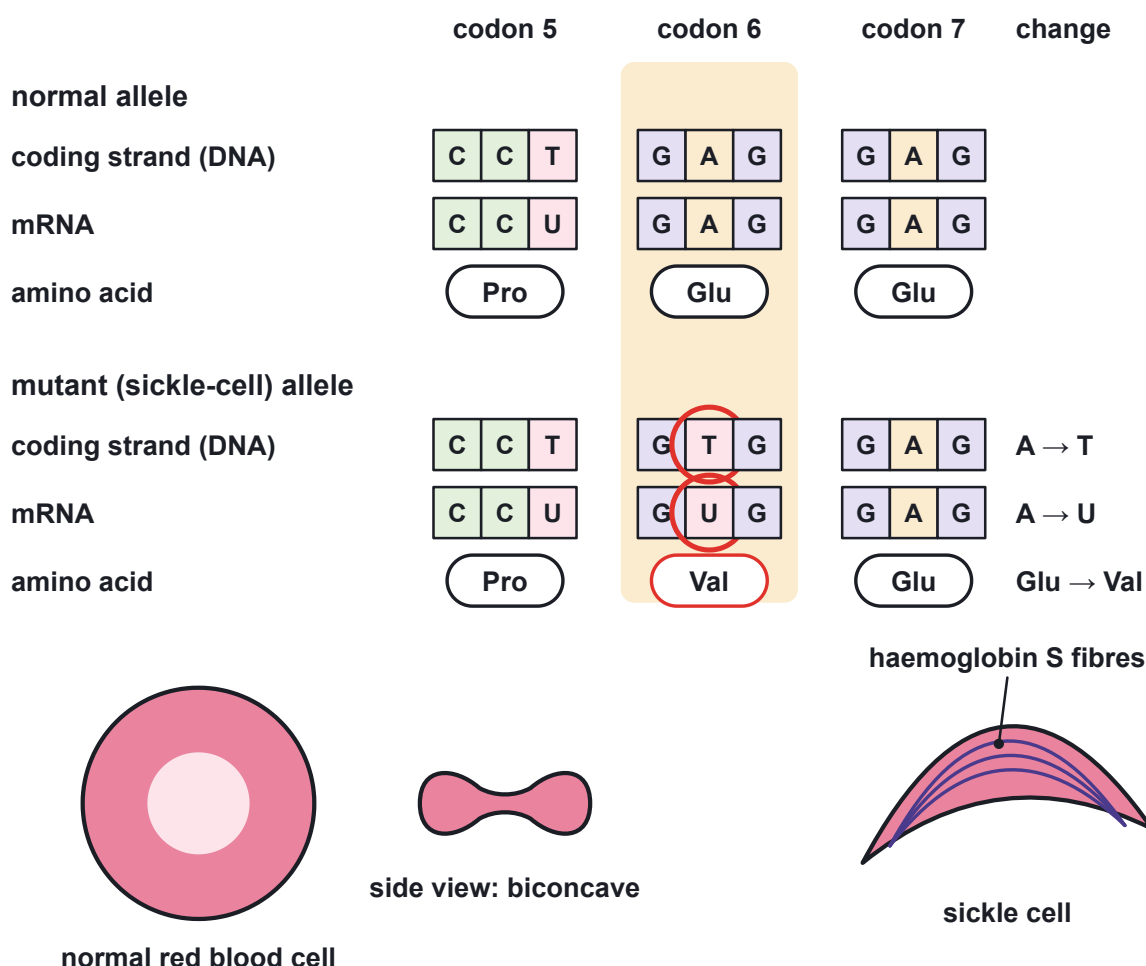
The amino acid sequence of a polypeptide decides how it folds and therefore its three-dimensional structure. A change in the base sequence of a gene can change the codons in the mRNA, and so the amino acid sequence, the folding and the function of the protein.

Example: sickle-cell anaemia

1. The gene for the beta chain of haemoglobin has a **base substitution**: on the coding strand, A is replaced by T (GAG becomes GTG).
2. The mRNA codon becomes **GUG instead of GAG**.

- At the sixth amino acid of the beta chain, **valine replaces glutamic acid**. The resulting form is haemoglobin S.
- Glutamic acid is charged and hydrophilic; valine is non-polar and hydrophobic. At low oxygen concentrations, haemoglobin S molecules stick together into long fibres.
- The fibres distort red blood cells into a rigid sickle shape. Sick cells can block capillaries and carry less oxygen, and they are broken down faster, causing anaemia.

A change to just one base of the gene is enough to alter the protein. This shows how closely protein structure depends on base sequence.



[\(opens full size in a new tab\)](#)

Sickle-cell anaemia: one base substitution (GAG → GTG) changes **glutamic acid to valine** in haemoglobin.

Exam tip: In a sickle-cell answer, give the base change, the codon change (GAG → GUG), the amino acid change (Glu → Val) and the effect on haemoglobin and red blood cells. Each is a separate marking point.

Directionality and initiation at the promoter Spec D1.2.12, D1.2.13 HL only

5' to 3' transcription

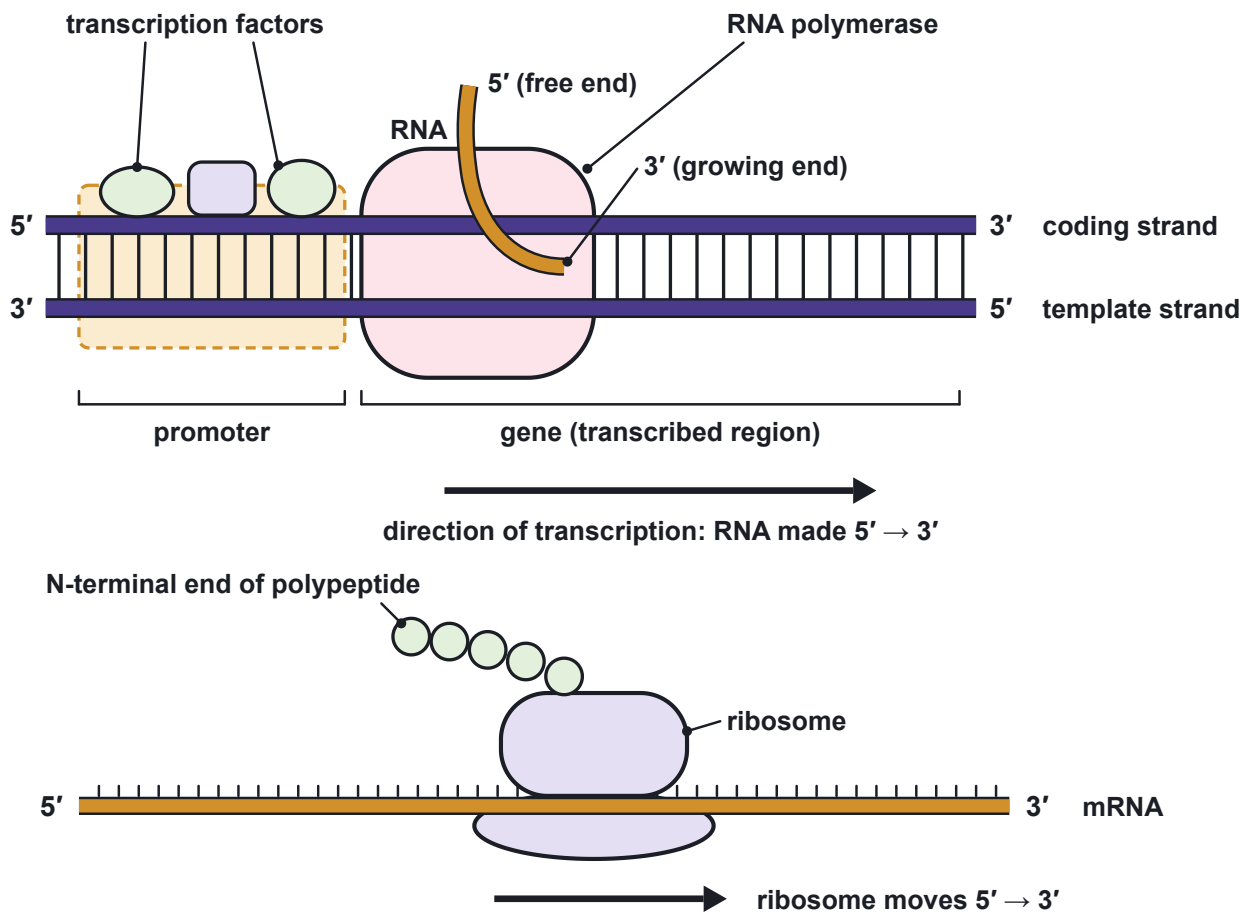
RNA polymerase, like DNA polymerase, adds the 5' end of each new RNA nucleotide to the 3' end of the growing RNA strand. RNA is therefore synthesised in the **5' → 3' direction**. Because the template strand is antiparallel to the RNA, RNA polymerase moves along the template strand in its 3' → 5' direction.

5' to 3' translation

Ribosomes bind near the 5' end of mRNA and move towards the 3' end, reading codons in order. Translation is therefore **5' → 3'** along the mRNA. The first amino acid added (from the start codon) forms the amine (N) terminal of the polypeptide, and the last forms the carboxyl (C) terminal.

The promoter

The **promoter** is a base sequence in DNA just before the start of a gene. It is not itself transcribed. Transcription is initiated when proteins called **transcription factors** bind to the promoter. They allow RNA polymerase to bind at the correct site and start transcribing in the right direction along the right strand. Whether transcription factors bind is one way in which the expression of a gene is switched on or off.



[\(opens full size in a new tab\)](#)

Transcription factors at the **promoter** let RNA polymerase start; RNA is made **5' → 3'** and translated 5' → 3'.

Non-coding DNA, mRNA processing and alternative splicing Spec D1.2.14, D1.2.15, D1.2.16

HL only

Non-coding sequences

In eukaryotes most of the DNA does not code for polypeptides. **Non-coding sequences** include:

- **Regulators of gene expression**, such as promoters, where proteins bind to control whether a gene is transcribed.
- **Introns**: sequences within a gene that are transcribed but removed from the mRNA before translation.
- **Telomeres**: repetitive sequences at the ends of chromosomes that protect the coding DNA from being lost during replication.

- **Genes for rRNA and tRNA:** these are transcribed, but the RNA is used directly in ribosomes and in carrying amino acids, so it is never translated into a polypeptide.

Post-transcriptional modification

In eukaryotic cells the RNA made by transcription is **pre-mRNA**. It is modified in the nucleus to form mature mRNA:

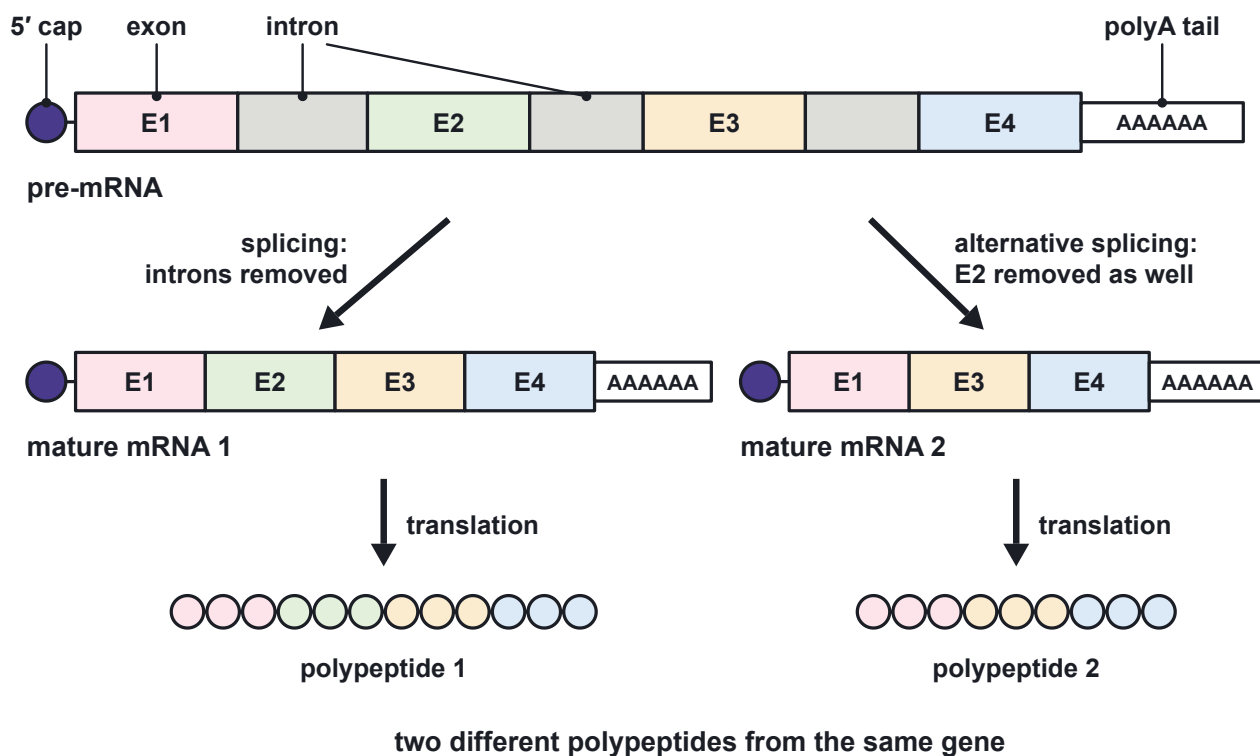
1. A **5' cap** (a modified guanine nucleotide) is added to the 5' end.
2. A **3' polyA tail** (a long chain of adenine nucleotides) is added to the 3' end.
3. **Introns are removed** and the **exons are spliced together** to give a continuous coding sequence.

The cap and tail stabilise the mRNA transcript, protecting it from breakdown in the cytoplasm.

Prokaryotes do not do this: their genes generally have no introns and mRNA can be translated while it is still being transcribed.

Alternative splicing

Alternative splicing is the splicing together of different combinations of exons from the same pre-mRNA. Some exons may be left out in one cell type but included in another. This allows one gene to code for different polypeptides, so a genome can produce many more proteins than it has genes.



[\(opens full size in a new tab\)](#)

Alternative splicing joins different combinations of exons, so one gene can code for different polypeptides.

Exam tip: Linking question: how does the diversity of proteins produced contribute to the functioning of a cell? Alternative splicing lets different cells make different variants of a protein from the same gene.

Initiation of translation and the A, P and E sites Spec D1.2.17 HL only

Initiation

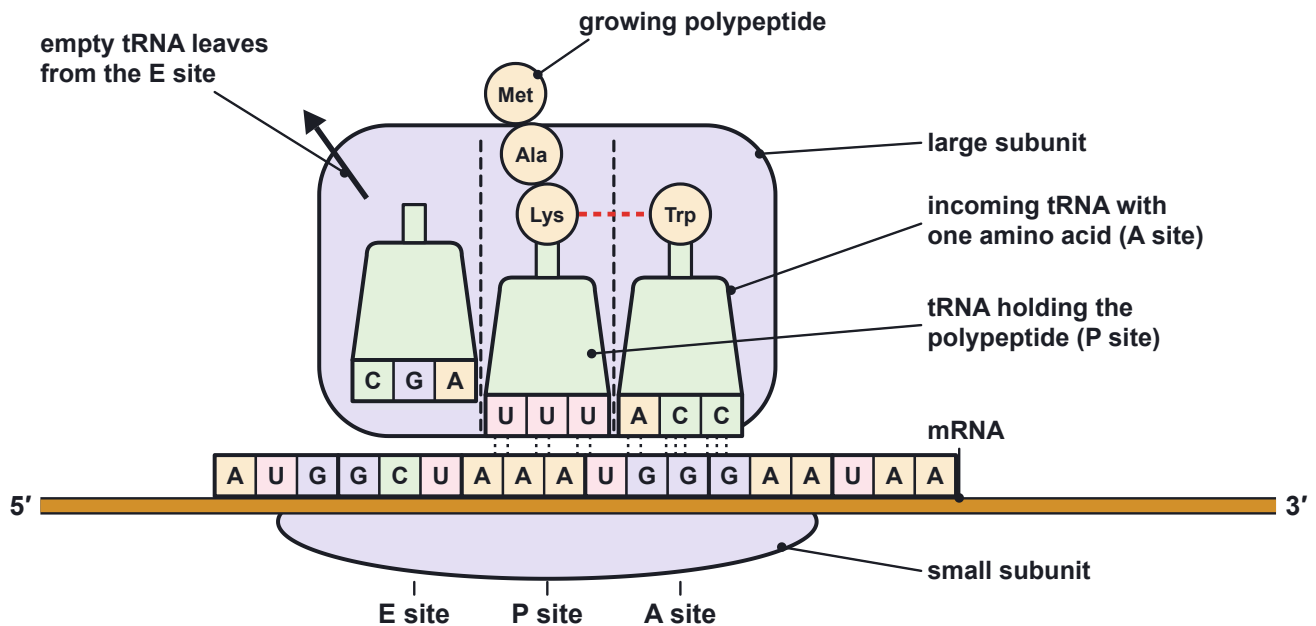
1. The **small ribosomal subunit** attaches to the 5' terminal of the mRNA.
2. It moves along the mRNA towards the 3' end until it reaches the **start codon (AUG)**.
3. The **initiator tRNA**, with anticodon UAC and carrying methionine, pairs with the start codon.

4. The **large subunit attaches**, so the initiator tRNA occupies the P site.
5. Another tRNA, with the anticodon complementary to the next codon, binds at the A site, and elongation begins.

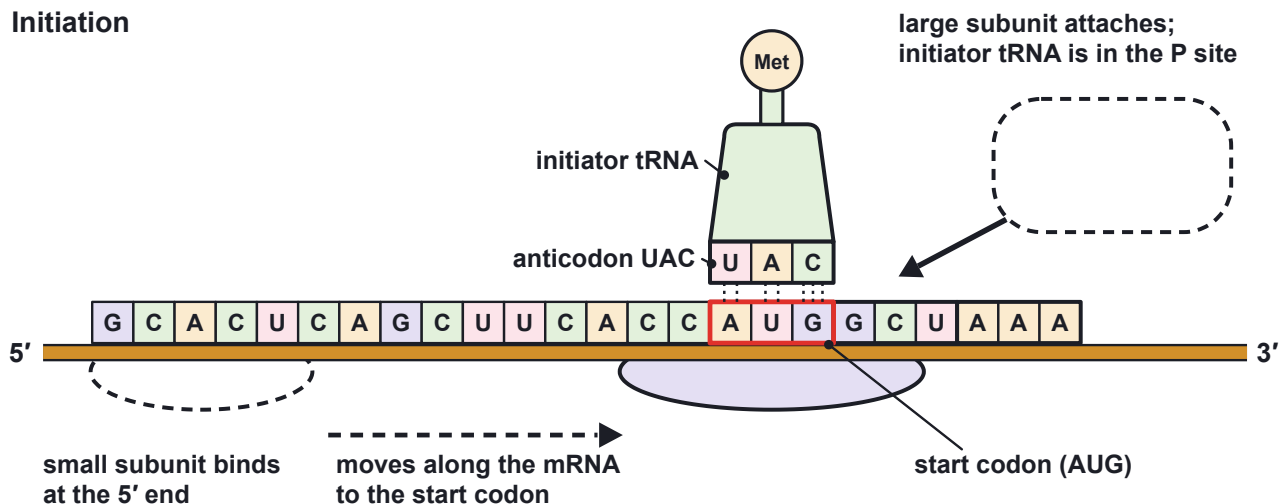
The three tRNA binding sites

Roles of the tRNA binding sites during elongation	
Site	Role
A site (aminoacyl)	A tRNA carrying a single amino acid arrives and its anticodon pairs with the codon exposed there.
P site (peptidyl)	Holds the tRNA attached to the growing polypeptide. The polypeptide is transferred to the amino acid on the tRNA in the A site by formation of a peptide bond.
E site (exit)	After the ribosome moves one codon along, the tRNA that has given up its polypeptide moves to the E site and is released.

After each peptide bond forms, the ribosome moves along the mRNA by one codon. The tRNA in the A site, now carrying the polypeptide, moves to the P site; the empty tRNA moves from the P site to the E site and leaves; the A site is free for the next tRNA.



Initiation



[\(opens full size in a new tab\)](#)

The A, P and E sites, and initiation at the **start codon AUG**.

Modification of polypeptides and proteasomes Spec D1.2.18, D1.2.19 HL only

The polypeptide released from the ribosome is often not yet functional. It may need to be folded, cut, joined to other polypeptides or have other groups (such as carbohydrates) added.

Pre-proinsulin to insulin

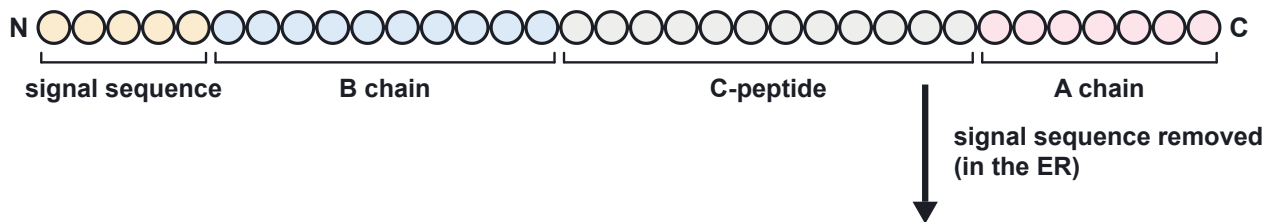
1. Translation on ribosomes of the rough endoplasmic reticulum produces **pre-proinsulin**. It has a signal sequence at its amine end that directs it into the endoplasmic reticulum.
2. **Stage 1:** the signal sequence is removed in the endoplasmic reticulum, giving **proinsulin**. Proinsulin folds and three disulfide bonds form.
3. **Stage 2:** in the Golgi apparatus and secretory vesicles, enzymes cut out the middle section, the **C-peptide**.
4. The remaining two chains, the A chain (21 amino acids) and B chain (30 amino acids), stay linked by disulfide bonds. This is active insulin, which is secreted.

Recycling of amino acids by proteasomes

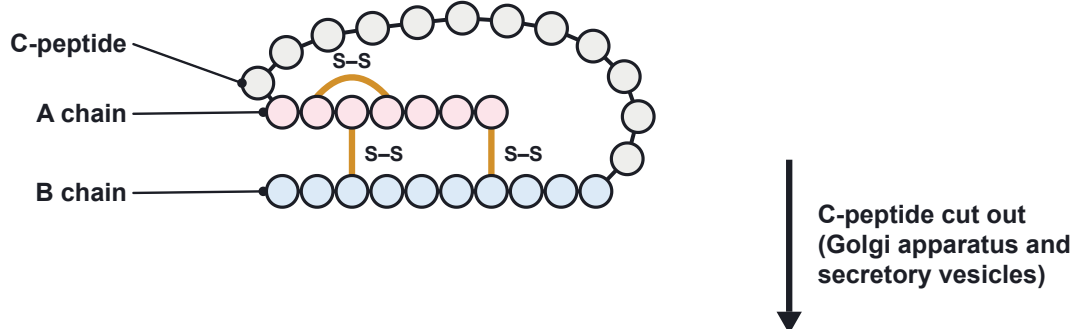
The **proteome** is the set of all proteins produced by a cell. Proteins are constantly being damaged or misfolded, or are no longer needed. Such proteins are first tagged for destruction (with a small protein called ubiquitin). **Proteasomes**, large barrel-shaped protein complexes in the cytoplasm and nucleus, recognise the tag and hydrolyse the tagged protein into short peptides. Other enzymes (peptidases) break the peptides into amino acids, which are reused in protein synthesis.

Sustaining a functional proteome requires constant protein breakdown and synthesis. Breakdown removes faulty proteins and allows the cell to change which proteins it contains as conditions change.

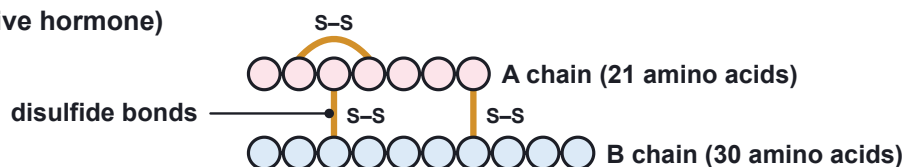
pre-proinsulin



proinsulin



insulin (active hormone)



[\(opens full size in a new tab\)](#)

Insulin is made from one polypeptide, **pre-proinsulin**, by removing the signal sequence and then the C-peptide.

Common mistake: Insulin is two polypeptide chains, but they come from a single gene and a single polypeptide (pre-proinsulin). The chains are separated by cutting, not translated separately.

Quick check

1. Which base on RNA pairs with adenine on the DNA template strand?
2. Why must DNA in non-dividing somatic cells stay unchanged by transcription?
3. What is meant by the degeneracy of the genetic code?
4. How many tRNAs can bind to the large ribosomal subunit at once?
5. **HL only** What is removed from proinsulin to produce insulin?

3 Mutation and gene editing Spec D1.3

Substitutions, insertions and deletions Spec D1.3.1, D1.3.2, D1.3.3

A **gene mutation** is a structural change to a gene at the molecular level: a change to its base sequence. There are three types.

Types of gene mutation		
Type	Change	Example (coding strand)
Substitution	One base is replaced by a different base.	GAG → GTG
Insertion	One or more extra nucleotides are added.	GAG CTA → GAA GCT A...
Deletion	One or more nucleotides are lost.	GAG CTA → GGC TA...

Consequences of base substitutions

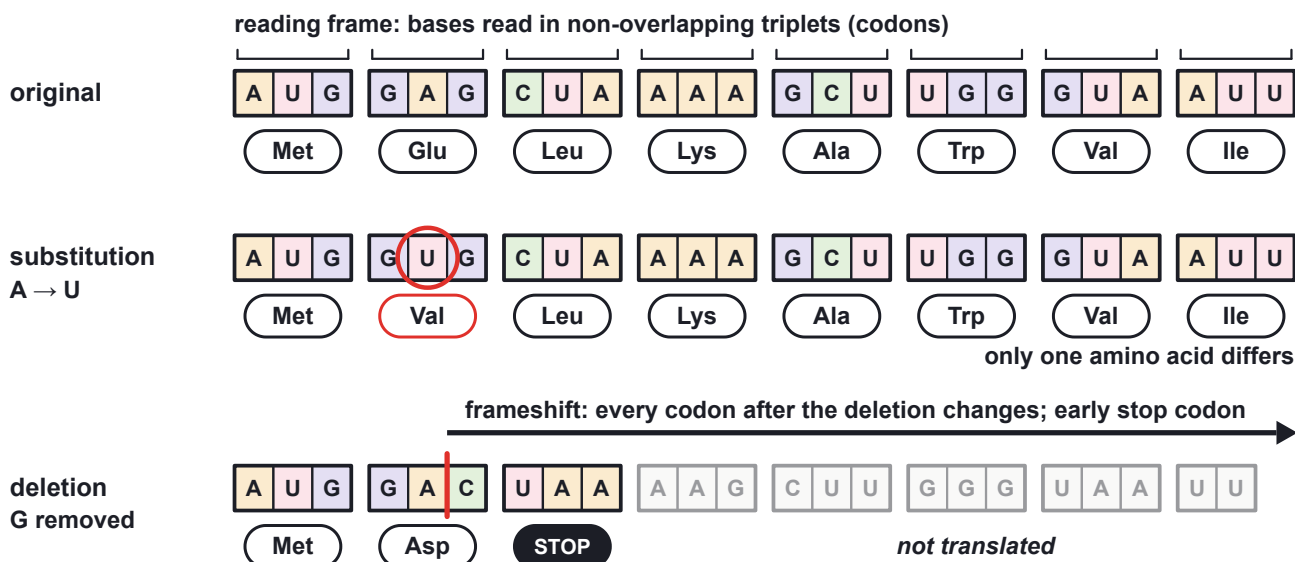
Single-nucleotide polymorphisms (SNPs) are positions in the genome where a single base differs between individuals. They are the result of base substitution mutations. A substitution changes only one codon, and because of the **degeneracy of the genetic code** it may or may not change a single amino acid:

- **Same amino acid:** the new codon codes for the same amino acid (for example GAA → GAG, both glutamic acid). The polypeptide is unchanged.
- **Different amino acid:** one amino acid in the polypeptide is changed (for example GAG → GUG, glutamic acid → valine). The effect may be small or, as in sickle-cell anaemia, large.
- **Stop codon:** the codon becomes a stop codon (for example UAC → UAA), so the polypeptide is cut short and is usually non-functional.

Consequences of insertions and deletions

Because mRNA is read in non-overlapping triplets, inserting or deleting one or two nucleotides shifts the reading frame. This **frameshift** changes every codon after the mutation, so the amino acid sequence from that point is altered and a premature stop codon often appears. The polypeptide is very likely to cease to function.

Inserting or deleting a multiple of three nucleotides does not cause a frameshift, but adds or removes amino acids. Major insertions or deletions of many nucleotides also make it likely that the polypeptide stops functioning.



[\(opens full size in a new tab\)](#)

A substitution changes one codon at most; a deletion causes a **frameshift** that changes every codon after it.

Exam tip: Linking question: how does variation in subunit composition of polymers contribute to function? A single changed nucleotide in DNA can change one amino acid in a protein and alter its function.

Causes of mutation and randomness Spec D1.3.4, D1.3.5

Causes of gene mutation

A **mutagen** is an agent that increases the rate of mutation.

Causes of gene mutation		
Cause	Examples	How it causes mutation
Chemical mutagens	Benzo[a]pyrene and nitrosamines in tobacco smoke; mustard gas	React with bases or bind to DNA, causing bases to pair incorrectly during replication.
Mutagenic radiation	Ultraviolet (UV) light, X-rays, gamma rays, alpha particles from radioactive isotopes such as radon	UV causes adjacent pyrimidine bases (such as two thymines) to bond to each other; high-energy radiation ionises molecules and can break DNA strands.
Errors in DNA replication	A mismatched nucleotide not corrected by proofreading	A wrong base is copied into the new strand and then into every later copy.
Errors in DNA repair	Damaged DNA repaired incorrectly	The repaired region has a changed base sequence.

Randomness in mutation

Mutations are **random**. They can occur anywhere in the base sequences of a genome, although some bases have a higher probability of mutating than others (for example, a cytosine with a methyl group added, often where C is followed by G, can lose an amino group and become thymine, so C → T substitutions are especially common at these sites). There is **no known natural mechanism for making a deliberate change to a particular base** with the purpose of changing a trait. An organism cannot produce a mutation because it would be useful.

Common mistake: Do not write that organisms mutate in order to adapt. Mutations happen at random; natural selection then acts on the variation they produce.

Germ and somatic cells, and mutation as a source of variation Spec D1.3.6, D1.3.7

Germ cells and somatic cells

Consequences of mutation in different cells		
	Germ cells	Somatic cells
What they are	Cells that produce gametes (and the gametes themselves)	All other body cells
Inherited?	Yes: a mutated gene can be passed to offspring	No: not passed to offspring
Cells affected	Every cell of an offspring that inherits it	Only the mutated cell and the cells produced from it by mitosis
Possible consequence	Inherited genetic disease, or new variation in the population	Cancer , if mutations affect genes that control cell division

Cancer develops when mutations in a somatic cell cause it to divide repeatedly without control, forming a tumour. Usually several mutations in the same cell line are needed, which is one reason cancer becomes more common with age and with exposure to mutagens.

Mutation as a source of genetic variation

Gene mutation is the **original source of all genetic variation**: it is the only process that creates new alleles. Meiosis and fertilisation shuffle existing alleles into new combinations, but cannot create new ones. Most mutations are either harmful or neutral for an individual organism, but in a species over the long term they are **essential for evolution by natural selection**, because selection can only act on variation that already exists.

Exam tip: Nature of science: commercial genetic tests can give information about potential future health and disease risk. Without expert interpretation, this information could be misunderstood and cause needless worry or false reassurance.

Exam tip: Linking question: how can natural selection lead to both a reduction in variation and an increase in biological diversity? Mutation supplies the variation; selection removes some alleles within a population but can drive populations apart into new species.

Gene knockout Spec D1.3.8 HL only

Gene knockout is a technique for investigating the function of a gene by changing it to make it inoperative. The gene's base sequence is altered, or part of it removed, so that no functional product is made.

1. Choose a gene whose function is unknown.
2. Make the gene inoperative in an organism (a knockout organism).
3. Compare the phenotype of the knockout organism with that of a normal (wild-type) organism.
4. Any difference, such as in development, behaviour or metabolism, suggests what the gene normally does.

Details of knockout techniques are not required. Students should know that **libraries of knockout organisms** are available for some species used as models in research. For example, collections of yeast strains exist with nearly every gene knocked out one at a time, and international projects are building a

knockout mouse for each mouse gene. Researchers can obtain an organism with a particular gene knocked out instead of making it themselves.

Practical skill: A knockout experiment needs a control: wild-type organisms of the same strain, raised in the same conditions, so that any difference can be attributed to the missing gene.

CRISPR-Cas9 gene editing Spec D1.3.9 HL only

How CRISPR-Cas9 editing works

1. A **guide RNA** is made with a base sequence complementary to the target sequence in the gene to be edited.
2. The guide RNA is combined with **Cas9**, an enzyme that cuts DNA (an endonuclease).
3. The guide RNA binds by complementary base pairing to the target sequence, bringing Cas9 to exactly that site.
4. Cas9 cuts both strands of the DNA at the target site.
5. The cell repairs the break. Repair often adds or removes a few nucleotides, which disables (knocks out) the gene. If a DNA template with a desired sequence is supplied, the cell can use it during repair, inserting or correcting a sequence.

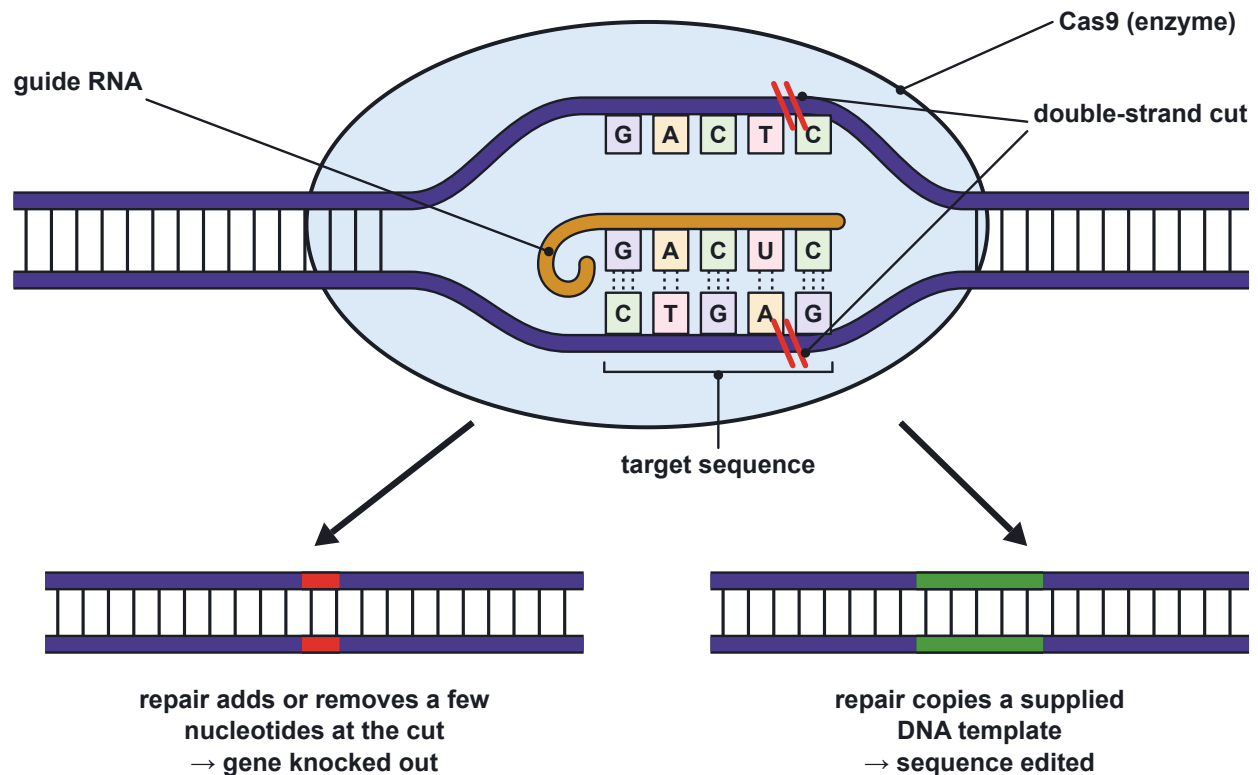
The role of the CRISPR-Cas system in prokaryotes is not required.

Example: treating sickle-cell disease

The first approved CRISPR-Cas9 therapy, exagamglogene autotemcel (exa-cel, sold as Casgevy), treats sickle-cell disease. It was authorised in the UK by the MHRA on 16 November 2023 and in the USA by the FDA on 8 December 2023, for patients aged 12 and over. Blood stem cells are collected from the patient's blood and edited outside the body. Cas9 cuts an enhancer (a regulatory sequence) of the gene *BCL11A*, so red blood cell precursors make much less of the BCL11A protein. BCL11A normally switches off the genes for **fetal haemoglobin** after birth. The patient receives chemotherapy to clear the old bone marrow, then the edited cells are infused. They produce red blood cells containing fetal haemoglobin, which does not form fibres, and most treated patients in the trials had no severe pain crises for at least a year.

Ethics and regulation

Certain potential uses of CRISPR raise ethical issues that must be addressed before implementation, especially editing embryos or germ cells, because changes would be inherited by future generations without their consent, and editing for enhancement rather than treating disease. Scientists across the world are subject to **different regulatory systems**, so there is an **international effort to harmonise regulation** of genome editing technologies such as CRISPR.



[\(opens full size in a new tab\)](#)

CRISPR-Cas9: the guide RNA takes Cas9 to the target sequence, Cas9 makes a **double-strand cut** and the cell repairs it.

Conserved and highly conserved sequences Spec D1.3.10 HL only

Conserved sequence	A base sequence (or amino acid sequence) that is identical or similar across a species or a group of species.
Highly conserved sequence	A sequence that has remained identical or similar over long periods of evolution, so it is shared by distantly related groups.

Examples of highly conserved genes include those for histone proteins, which package DNA, and for ribosomal RNA, which are similar in organisms as different as yeast and humans.

Hypotheses to account for conservation

Two hypotheses	
Hypothesis	Explanation
Functional requirements for the gene products	The protein or RNA must have a very precise structure to work, for example because it binds many other molecules. Almost any mutation makes it less functional, so individuals carrying it are selected against and the mutation is removed from the population. The sequence stays the same because changes are eliminated, not because they do not happen.
Slower rates of mutation	Some sequences may mutate less often than others, for example because of their position in the genome or more effective repair, so fewer changes arise in the first place.

For protein-coding genes the two hypotheses make different predictions. If functional requirements are the cause, base substitutions that do not change the amino acid (possible because the code is

degenerate) should still build up at a normal rate, while substitutions that change the amino acid are rare. If the mutation rate is slower, both kinds of change should be rare. In mammals, histone genes show almost no amino acid changes but a normal rate of the silent changes, which supports the functional-requirements hypothesis for these genes.

Exam tip: Name both hypotheses when asked to account for conserved sequences: functional requirements of the gene product (with selection removing changes) and a slower rate of mutation.

Quick check

1. What is a single-nucleotide polymorphism (SNP)?
2. Why does deleting one base usually have a larger effect than substituting one base?
3. Give one chemical mutagen and one form of mutagenic radiation.
4. Why are somatic mutations not inherited?
5. **HL only** What does the guide RNA do in CRISPR-Cas9 editing?

Quick check answers

DNA replication

1. Each new DNA molecule contains one strand from the original molecule and one newly synthesised strand.
2. It unwinds the double helix and breaks hydrogen bonds between bases so the strands separate.
3. It comes from a hot-spring bacterium and is not denatured at 95 °C, so it survives the denaturation step of every cycle.
4. They pass more easily through the mesh of the gel, so they move faster towards the positive electrode.
5. It is made discontinuously as Okazaki fragments, and each fragment must start from its own primer.

Protein synthesis

1. Uracil.
2. The same sequences must be transcribed throughout the life of the cell, so they must be conserved.
3. More than one codon can code for the same amino acid.
4. Two.
5. The C-peptide.

Mutation and gene editing

1. A position in the genome where a single base differs between individuals, resulting from a base substitution mutation.
2. It shifts the reading frame, changing every codon after the mutation.
3. Benzo[a]pyrene (in tobacco smoke); ultraviolet light (or X-rays, gamma rays).
4. They are not in germ cells, so they are not passed into gametes.
5. It binds by complementary base pairing to the target DNA sequence and brings Cas9 to it.