



Protein synthesis

Revision notes • IB Biology D1.2: transcription, translation, the genetic code, mutations that change proteins, and HL mRNA splicing, initiation and protein modification.

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

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

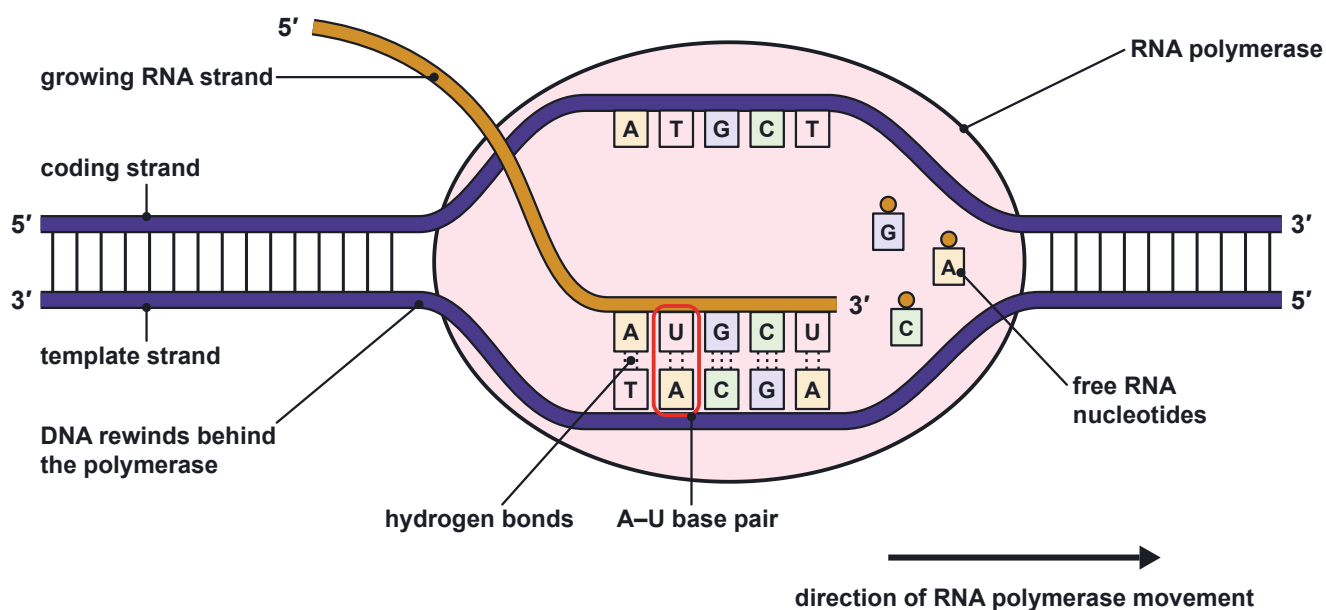
Each RNA nucleotide is held in place by **hydrogen bonds** to the complementary base on the template strand. The pairs are: **adenine (A) on the DNA template with uracil (U) on the RNA**, thymine (T) on the template with adenine (A), cytosine (C) with guanine (G) and guanine (G) with cytosine (C). The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

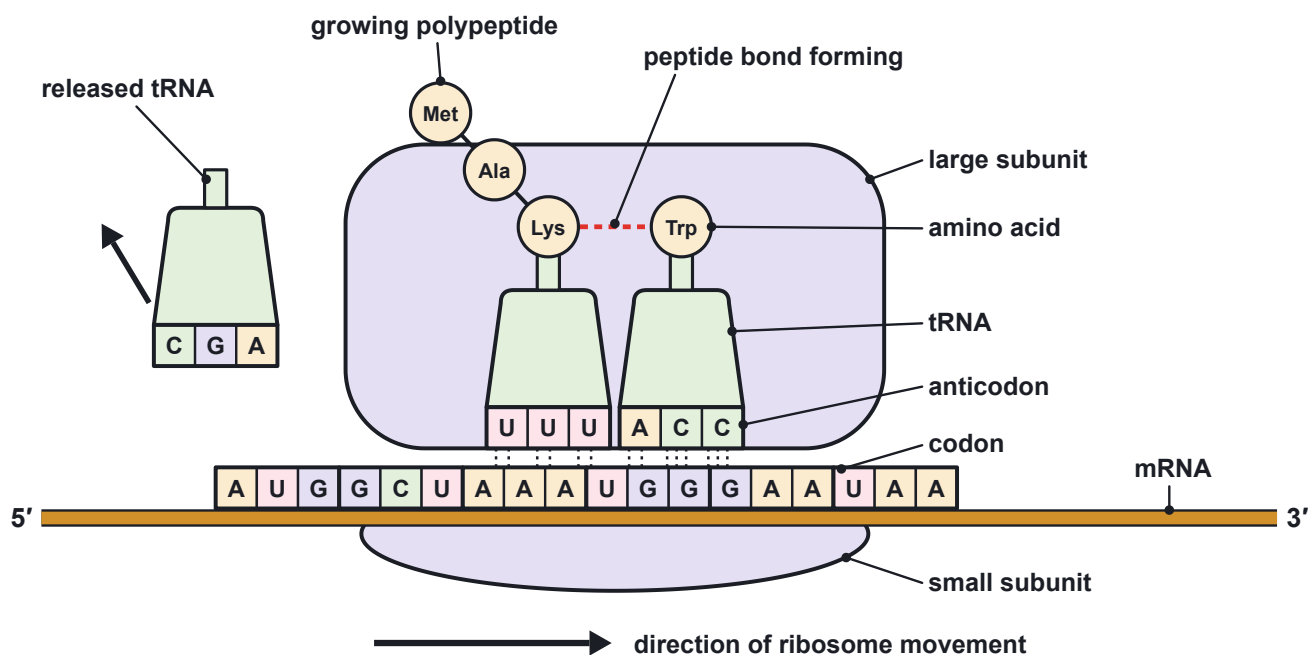
A **codon** is a sequence of three bases on mRNA that codes for one amino acid. An **anticodon** is the complementary sequence of three bases on a tRNA. The anticodon pairs with the codon by complementary base pairing and hydrogen bonding (A with U, C with G), which ensures the correct amino acid is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

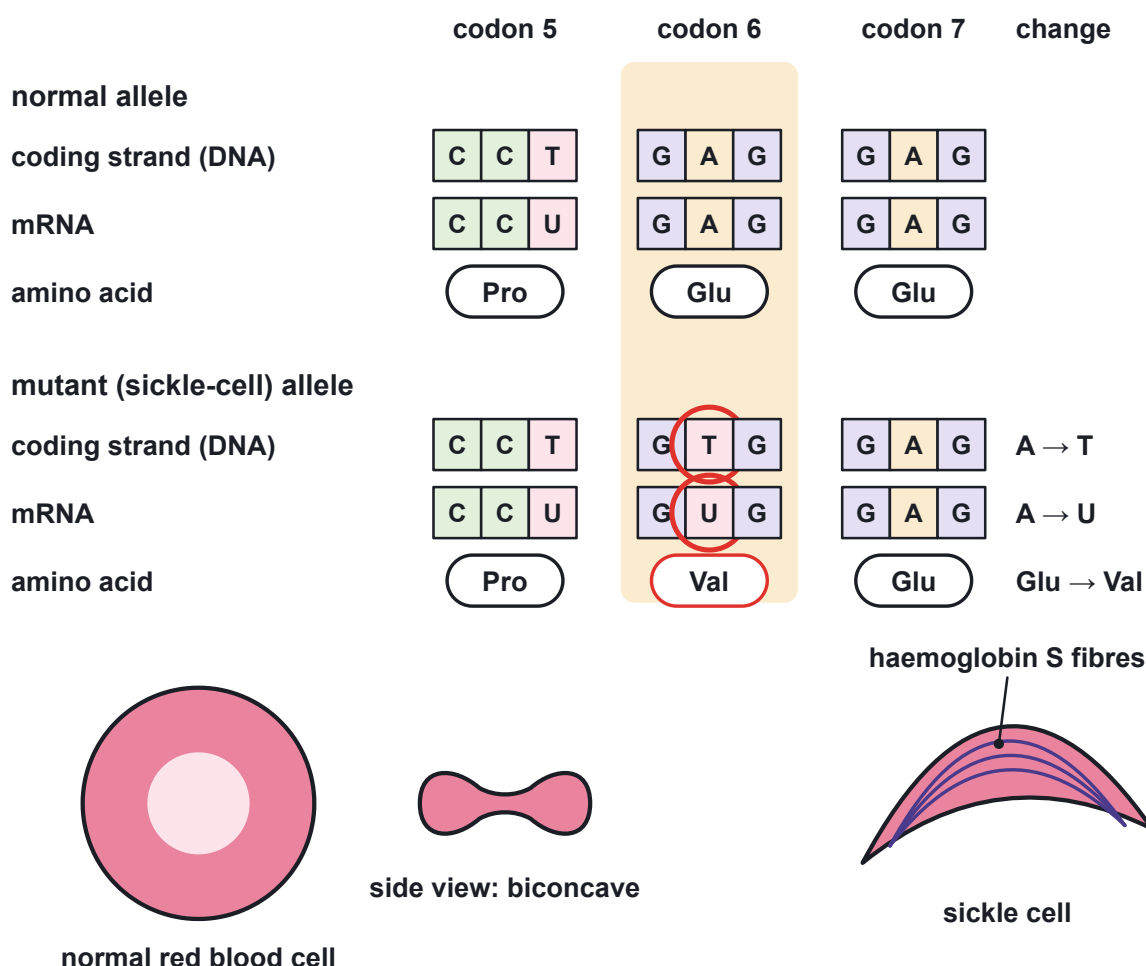
The amino acid sequence of a polypeptide decides how it folds and therefore its three-dimensional structure. A change in the base sequence of a gene can change the codons in the mRNA, and so the amino acid sequence, the folding and the function of the protein.

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

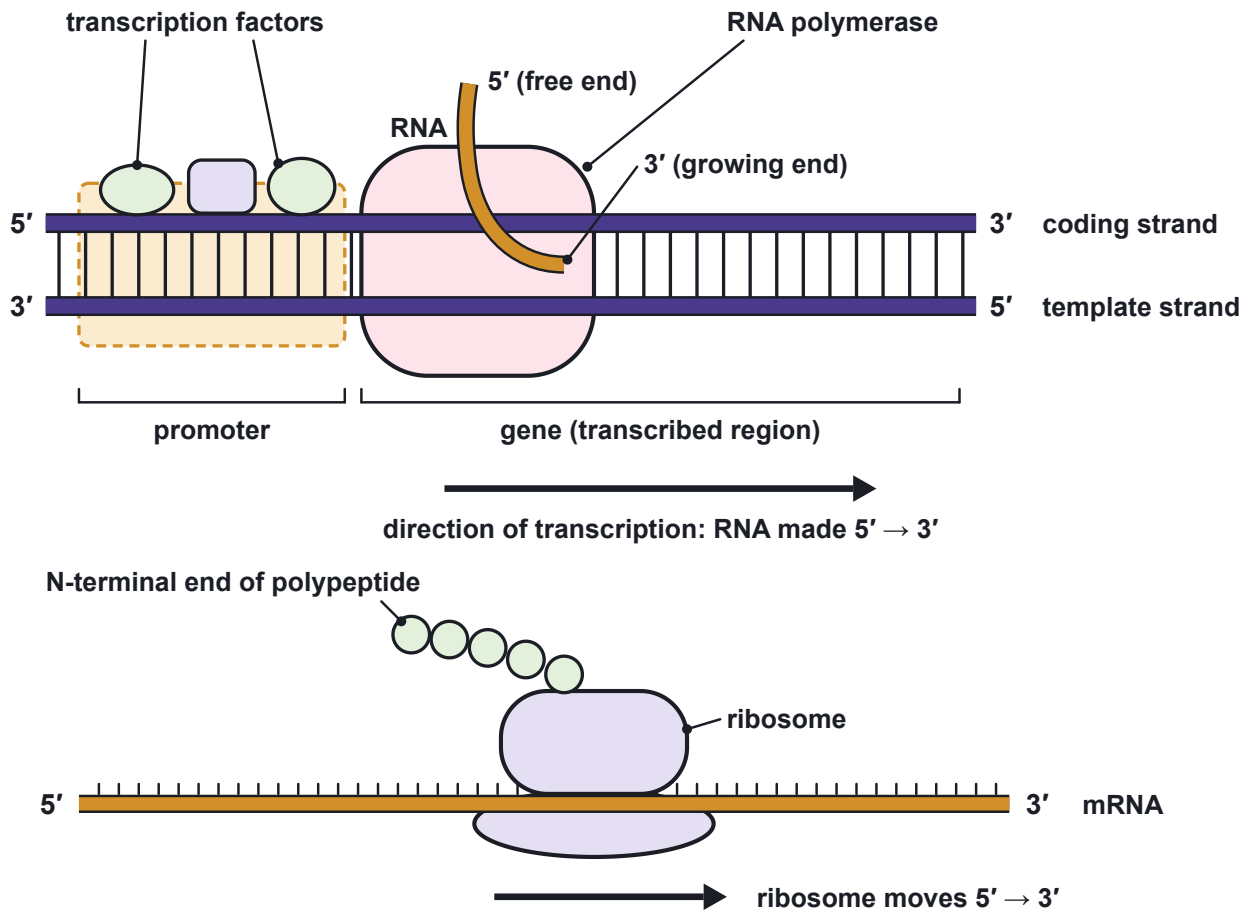
RNA polymerase, like DNA polymerase, adds the 5' end of each new RNA nucleotide to the 3' end of the growing RNA strand. RNA is therefore synthesised in the **5' → 3' direction**. Because the template strand is antiparallel to the RNA, RNA polymerase moves along the template strand in its 3' → 5' direction.

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:

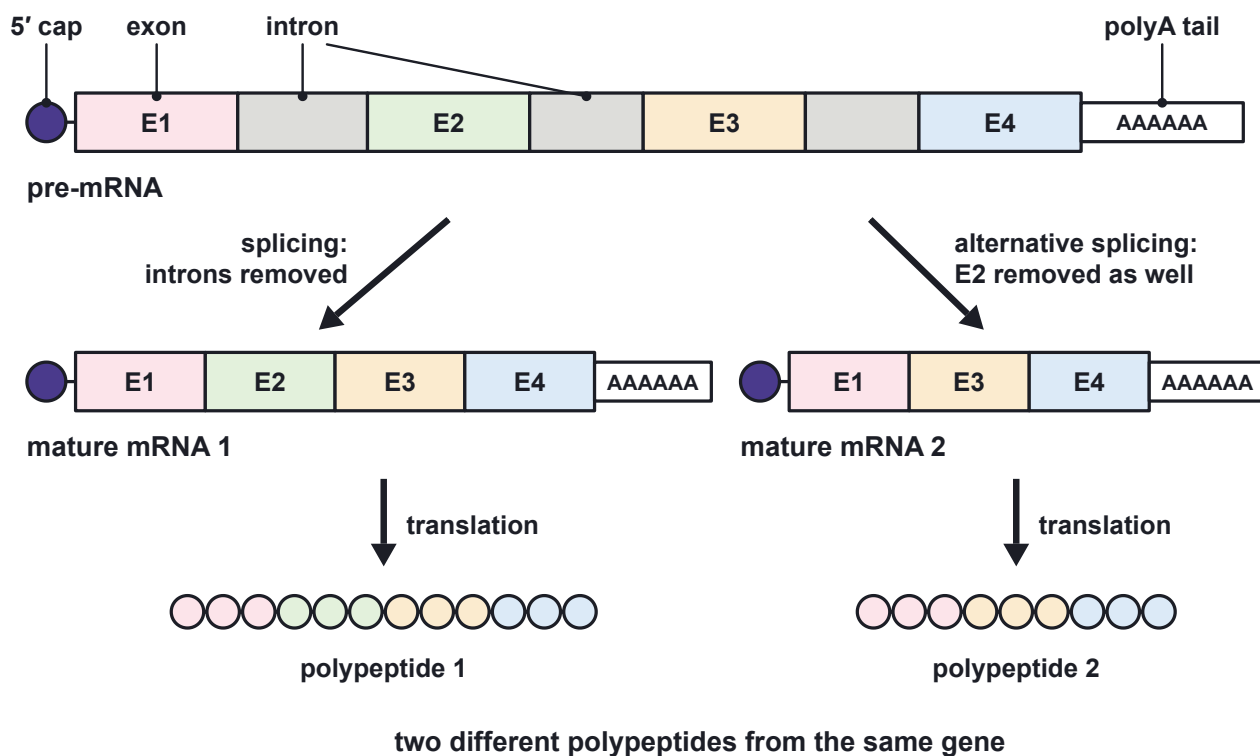
1. A **5' cap** (a modified guanine nucleotide) is added to the 5' end.
2. A **3' polyA tail** (a long chain of adenine nucleotides) is added to the 3' end.
3. **Introns are removed** and the **exons are spliced together** to give a continuous coding sequence.

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

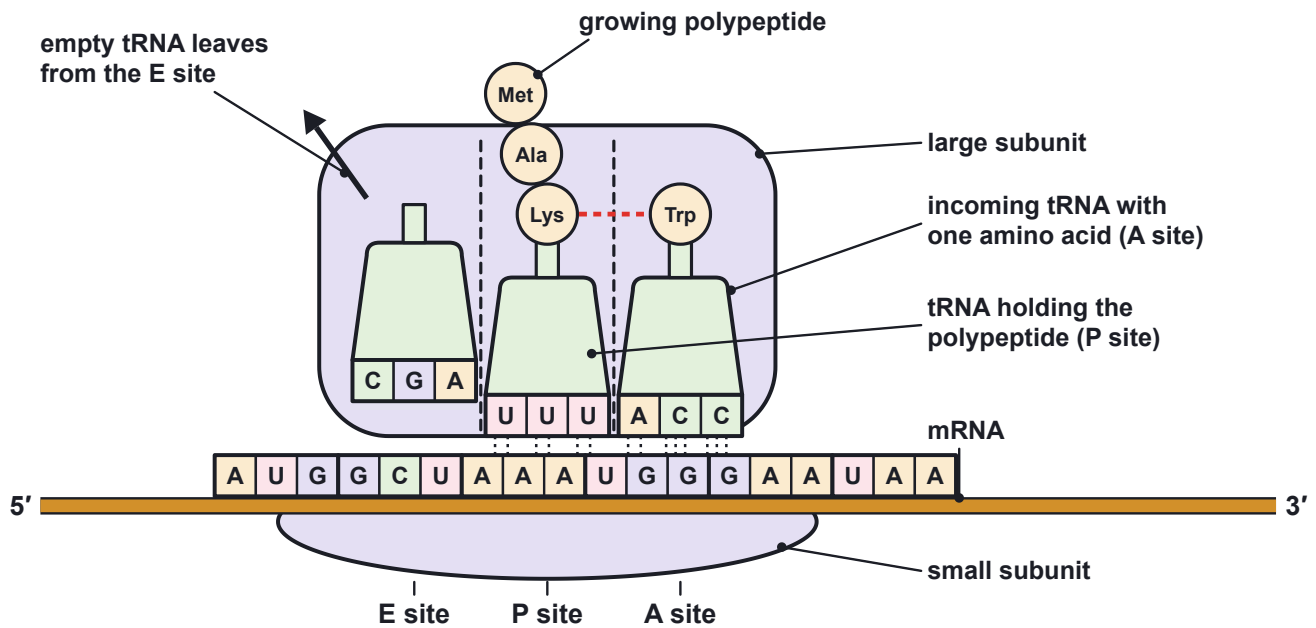
1. The **small ribosomal subunit** attaches to the 5' terminal of the mRNA.
2. It moves along the mRNA towards the 3' end until it reaches the **start codon (AUG)**.
3. The **initiator tRNA**, with anticodon UAC and carrying methionine, pairs with the start codon.

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.

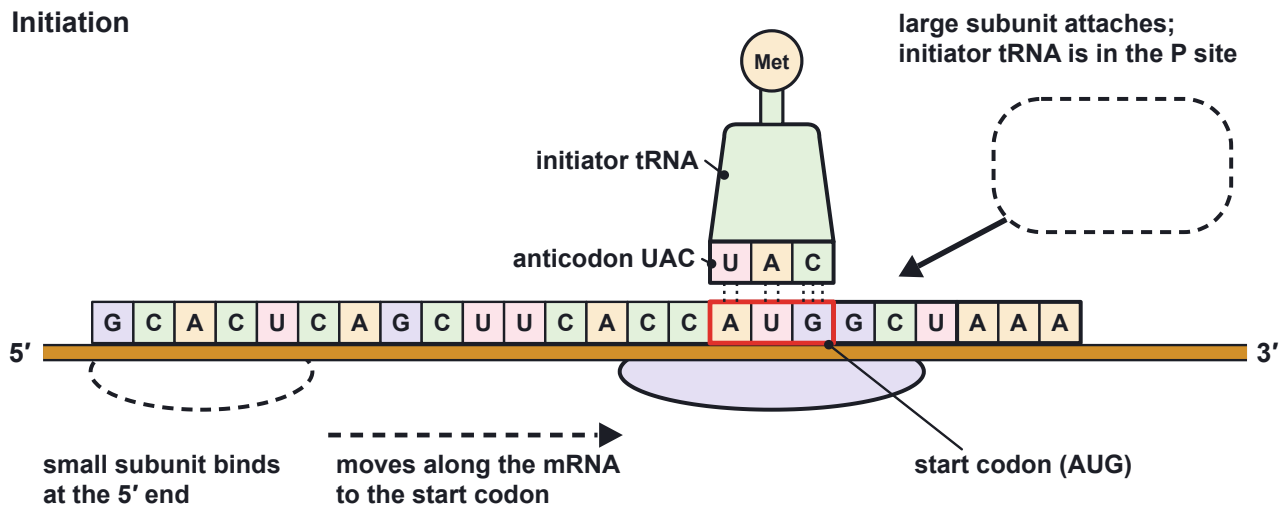
The three tRNA binding sites

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

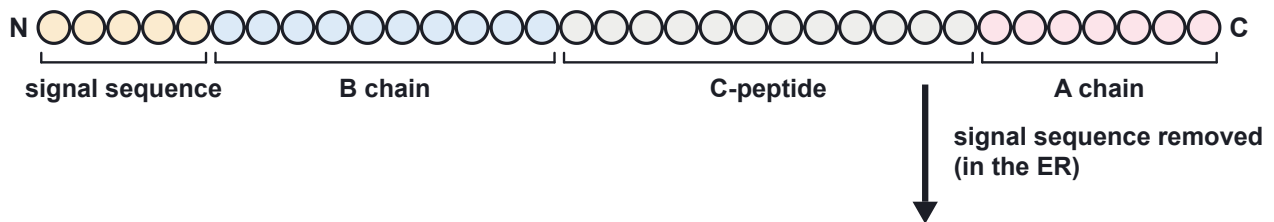
1. Translation on ribosomes of the rough endoplasmic reticulum produces **pre-proinsulin**. It has a signal sequence at its amine end that directs it into the endoplasmic reticulum.
2. **Stage 1:** the signal sequence is removed in the endoplasmic reticulum, giving **proinsulin**. Proinsulin folds and 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

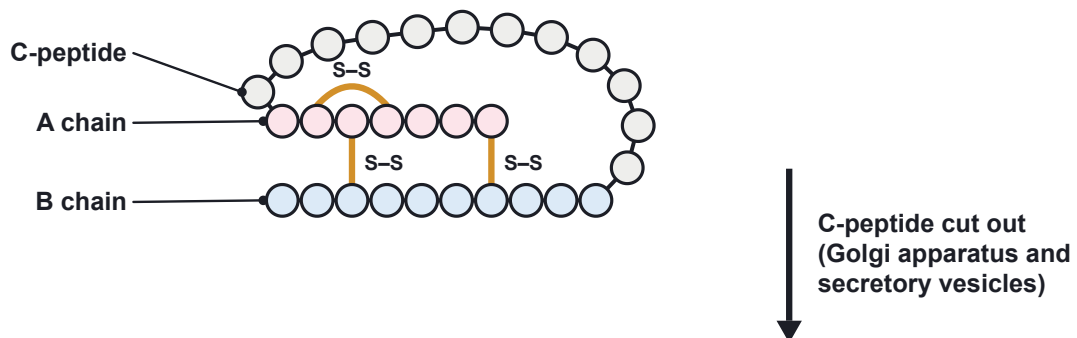
The **proteome** is the set of all proteins produced by a cell. Proteins are constantly being damaged or misfolded, or are no longer needed. Such proteins are first tagged for destruction (with a small protein called ubiquitin). **Proteasomes**, large barrel-shaped protein complexes in the cytoplasm 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.

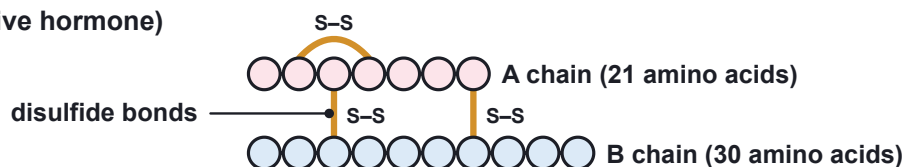
pre-proinsulin



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?

Quick check answers

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.