

# DNA, the genetic code and gene expression

Revision notes · The structure of DNA and RNA, semi-conservative replication, the genetic code, transcription, translation and gene mutations.

For Pearson Edexcel International AS Biology XBI11 and International A Level Biology YBI11. **A Level only** is IA2 content (Units 4 to 6), not needed for the International AS.

## DNA, the genetic code and gene expression Spec 2.9-2.14

### The structure of mononucleotides, DNA and RNA Spec 2.9

A **mononucleotide** has three parts: a pentose sugar, a phosphate group and a nitrogen-containing base. In DNA the sugar is **deoxyribose**; in RNA it is **ribose**. DNA has four bases: adenine (A), thymine (T), cytosine (C) and guanine (G). In RNA, thymine is replaced by **uracil (U)**.

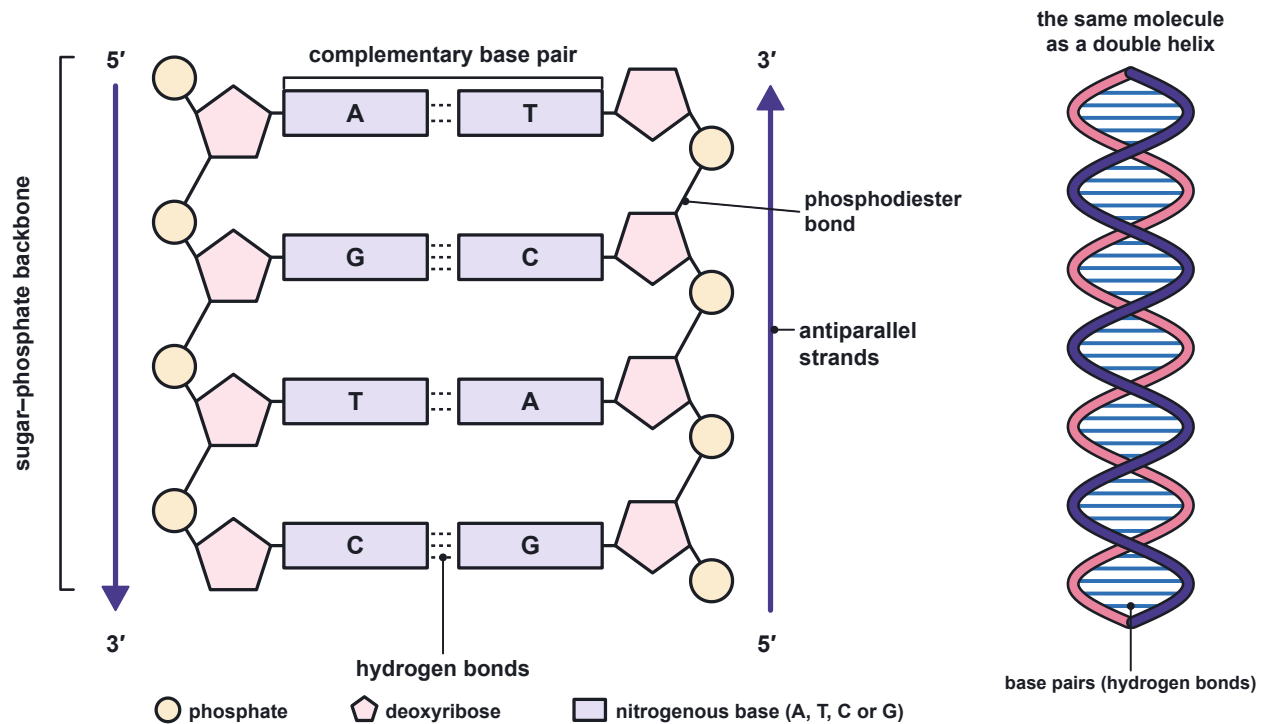
Mononucleotides join in condensation reactions. A **phosphodiester bond** forms between the phosphate group of one nucleotide and the sugar of the next, releasing water. Many nucleotides in a chain make a **polynucleotide**, with a sugar-phosphate backbone and bases sticking out to the side.

### The DNA double helix

- A DNA molecule is made of two polynucleotide strands wound into a double helix.
- The strands are **antiparallel**: they run in opposite directions (one runs 5' to 3', the other 3' to 5').
- **Complementary base pairing**: adenine pairs with thymine (two hydrogen bonds) and cytosine pairs with guanine (three hydrogen bonds). The base pairs lie across the middle of the helix.
- Many **hydrogen bonds** hold the two strands together. Each bond is weak, so the strands can be separated for replication and transcription, but together they make DNA stable.
- The sugar-phosphate backbones protect the bases on the inside.

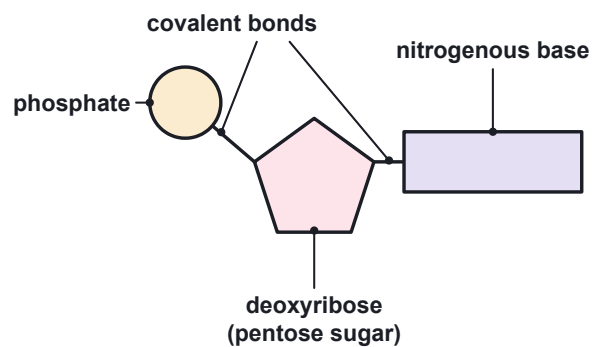
| Comparing DNA and RNA |                                          |                                                                 |
|-----------------------|------------------------------------------|-----------------------------------------------------------------|
| Feature               | DNA                                      | RNA                                                             |
| <b>Sugar</b>          | Deoxyribose                              | Ribose                                                          |
| <b>Bases</b>          | A, T, C, G                               | A, U, C, G                                                      |
| <b>Strands</b>        | Two polynucleotide strands, double helix | One polynucleotide strand                                       |
| <b>Size</b>           | Very long                                | Relatively short                                                |
| <b>Types</b>          | —                                        | Messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA) |

**Common mistake: Phosphodiester bonds join nucleotides along one strand; hydrogen bonds join bases between two strands. Do not mix them up, and do not say hydrogen bonds link A with G.**

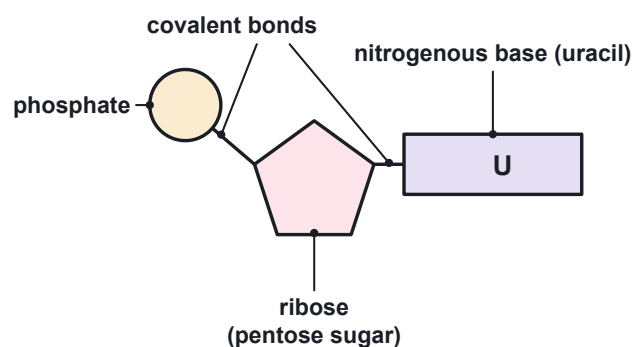


The **double helix**: antiparallel strands joined by **hydrogen bonds** between complementary bases.

#### DNA mononucleotide



#### RNA mononucleotide



A DNA nucleotide contains **deoxyribose**; an RNA nucleotide contains **ribose** and the base **uracil**.

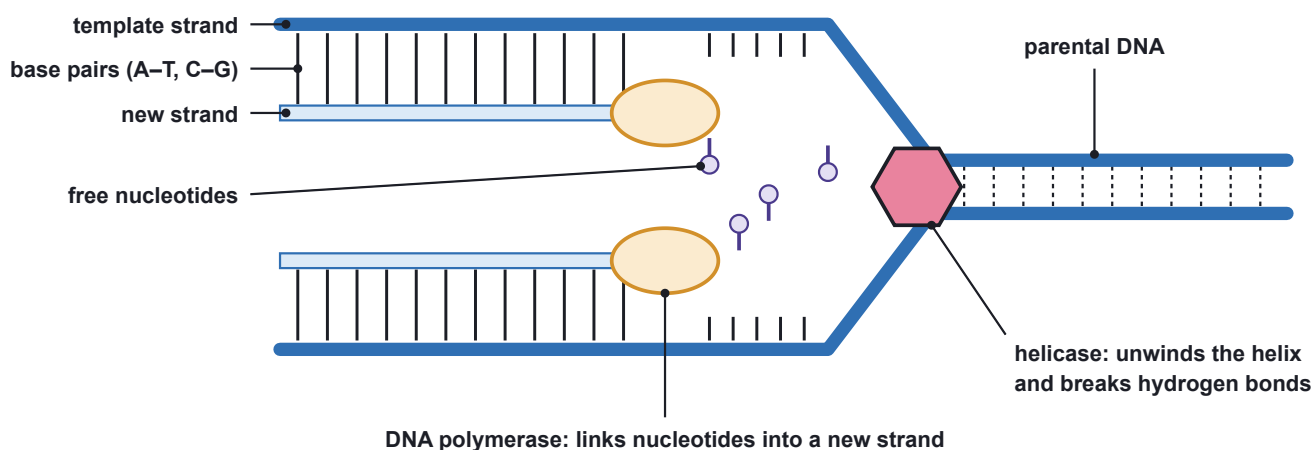
## DNA replication and the role of DNA polymerase Spec 2.10

Before a cell divides, its DNA must be copied. In **semi-conservative replication**, each of the two new DNA molecules is made of one original (old) strand and one newly synthesised strand.

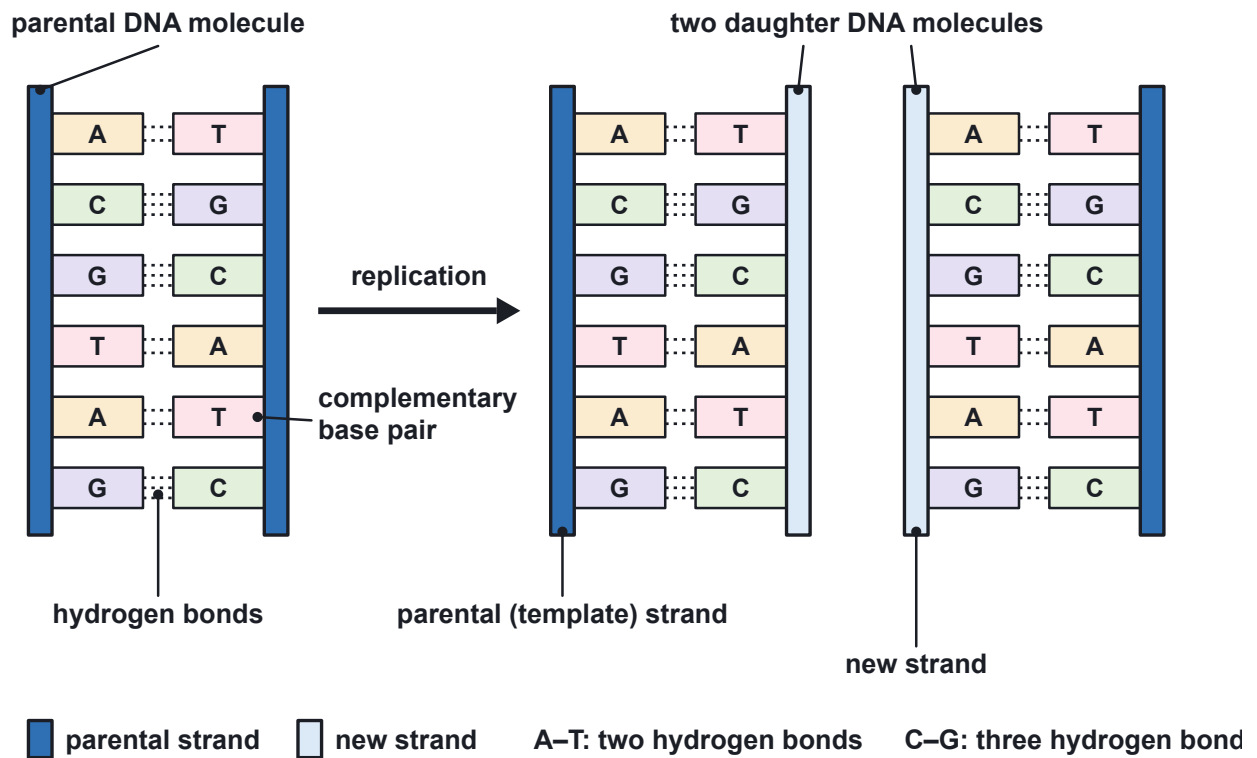
1. The enzyme DNA helicase unwinds the double helix and breaks the hydrogen bonds between the complementary bases, exposing the bases of the two template strands.
2. Free DNA nucleotides in the nucleus pair up with their complementary bases on each template strand: A with T, C with G.
3. **DNA polymerase** joins the nucleotides next to each other, forming phosphodiester bonds between them, which makes a new sugar-phosphate backbone. It adds nucleotides to the growing strand in one direction only.
4. Hydrogen bonds form between the bases of the template and new strands, and each pair twists into a double helix.
5. Two identical DNA molecules result, each with one old strand and one new strand.

The base sequence of the new strand is determined by the template strand, so the two new molecules have identical base sequences to the parent molecule (unless a mistake occurs, see gene mutations). This ensures that daughter cells receive the same genetic information.

*Exam tip: A full-mark replication answer names DNA helicase (unwinding, breaking hydrogen bonds), complementary base pairing, DNA polymerase (joining nucleotides, forming phosphodiester bonds) and says 'semi-conservative'.*



At the replication fork, **DNA helicase** unwinds the helix and **DNA polymerase** joins nucleotides to the new strand.



**Semi-conservative:** each daughter molecule keeps one old strand and one new strand.

## The Meselson-Stahl experiment Spec 2.10

Three models were proposed for how DNA could be copied. In the **conservative** model the original molecule stays intact and a completely new molecule is made. In the **semi-conservative** model each new molecule has one old and one new strand. In the **dispersive** model, each strand is a mixture of old and new DNA.

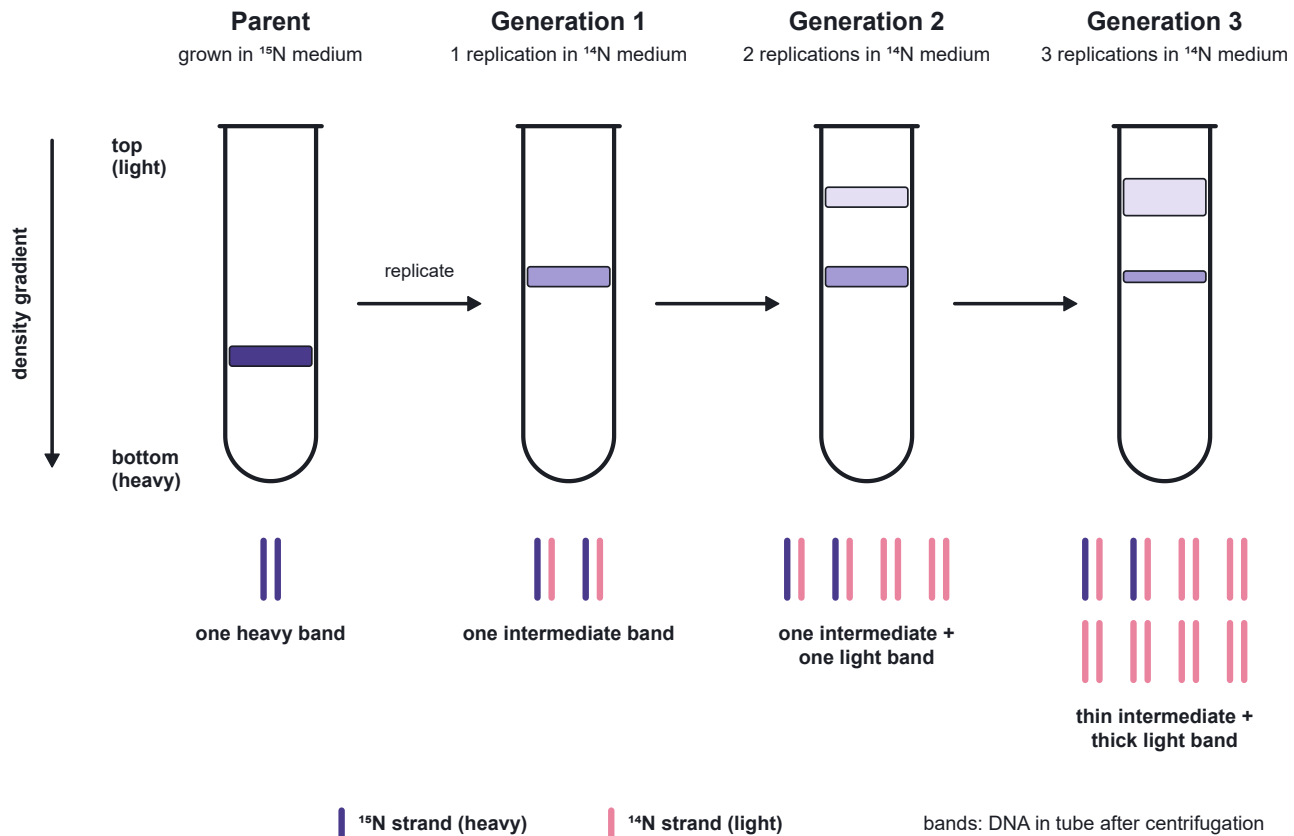
### The experiment

1. Bacteria (*E. coli*) were grown for many generations in a medium with heavy nitrogen,  $^{15}\text{N}$ , so all their DNA contained  $^{15}\text{N}$  (heavy DNA).
2. The bacteria were transferred to a medium containing only normal, lighter nitrogen,  $^{14}\text{N}$ , and allowed to divide. New DNA could only be made using  $^{14}\text{N}$ .
3. Samples were taken after each generation. The DNA was extracted and spun in a centrifuge in a caesium chloride solution, which separates DNA by density into bands.

| Results and what each model predicted         |                                   |                                           |                                        |                                                    |
|-----------------------------------------------|-----------------------------------|-------------------------------------------|----------------------------------------|----------------------------------------------------|
| Sample                                        | Result observed                   | Conservative model                        | Semi-conservative model                | Dispersive model                                   |
| <b>Start (all <math>^{15}\text{N}</math>)</b> | One band, heavy                   | One heavy band                            | One heavy band                         | One heavy band                                     |
| <b>After 1 generation</b>                     | One band, intermediate density    | Two bands: heavy and light (not matching) | One intermediate band (matches)        | One intermediate band (matches)                    |
| <b>After 2 generations</b>                    | Two bands: intermediate and light | Heavy and light bands (not matching)      | Intermediate and light bands (matches) | One band, lighter than intermediate (not matching) |

After one generation, a single band of intermediate density rules out the conservative model, which would give separate heavy and light bands. After two generations, the presence of both an intermediate and a light band rules out the dispersive model, which would give a single band of lower density. Only the semi-conservative model fits both results. The new data therefore **supported** the accepted theory and **refuted** the competing theories.

*Exam tip: In a data question, state what the band positions mean in terms of nitrogen (heavy = both strands  $^{15}\text{N}$ , intermediate = one  $^{15}\text{N}$  strand and one  $^{14}\text{N}$  strand, light = both strands  $^{14}\text{N}$ ), then compare with each model's prediction.*



The **Meselson-Stahl** results support **semi-conservative replication**.

## The genetic code and the gene Spec 2.11, 2.12

A **gene** is a sequence of bases on a DNA molecule that codes for a sequence of amino acids in a polypeptide chain. The order of bases in the gene therefore determines the order of amino acids in the polypeptide, and so the protein's structure and function.

### Features of the genetic code

- Triplet code:** a sequence of three bases (a triplet in DNA; a **codon** in mRNA) codes for one amino acid. There are four bases, so there are  $4^3 = 64$  possible triplets, and these are enough to code for the 20 amino acids. (A doublet code would give only  $4^2 = 16$  combinations, too few.)
- Non-overlapping:** each base is part of only one triplet, so the sequence is read in separate groups of three, one after another.
- Degenerate:** most amino acids are coded for by more than one triplet. For example, several codons code for glycine. This means that some changes in the base sequence do not change the amino acid.
- Some triplets are not amino acids: the codon AUG is a start codon (it also codes for methionine) and three codons (UAA, UAG, UGA) are stop codons, which end translation.

### Reading a sequence in triplets

A section of the template strand of DNA reads TAC GGG TTT. State the number of amino acids coded for by these nine bases, and write the sequence of the mRNA transcribed from it.

1. The code is a non-overlapping triplet code, so divide the bases into groups of three:  $9 \div 3 = 3$  triplets.
2. Each triplet codes for one amino acid: 3 amino acids.
3. mRNA bases are complementary to the template: A-U, T-A, C-G, G-C.
4. TAC → AUG; GGG → CCC; TTT → AAA.

**Answer:** 3 amino acids. mRNA: AUG CCC AAA.

*Common mistake: 'Degenerate' does not mean the code is damaged. It means there is more than one codon for an amino acid. 'Non-overlapping' does not mean there are gaps between triplets.*

### Transcription Spec 2.13

Protein synthesis has two stages: **transcription** (making mRNA from DNA in the nucleus) and **translation** (making a polypeptide from mRNA at a ribosome). DNA stays in the nucleus, so a copy of the gene, **messenger RNA (mRNA)**, carries the instructions to the cytoplasm.

1. RNA polymerase attaches to the DNA at the start of the gene. The two strands separate as the enzyme breaks the hydrogen bonds between bases.
2. One strand, the **template (antisense) strand**, is used to make the mRNA. The other strand, the sense (coding) strand, has the same base sequence as the mRNA (but with T instead of U).
3. Free RNA nucleotides pair with the complementary bases on the template strand: A-U, T-A, C-G, G-C.
4. RNA polymerase joins the RNA nucleotides together by phosphodiester bonds, to make a strand of mRNA.
5. The DNA strands re-join behind the enzyme. When RNA polymerase reaches a stop signal at the end of the gene, mRNA is released and leaves the nucleus through a nuclear pore.

#### Template and mRNA: an example

| Strand                             | Base sequence     |
|------------------------------------|-------------------|
| Template (antisense) strand of DNA | T A C G C A T T T |
| mRNA transcribed                   | A U G C G U A A A |
| Sense (coding) strand of DNA       | A T G C G T A A A |

*Exam tip: Questions on the 'role of the template strand' want: it is the strand copied by RNA polymerase, and the mRNA base sequence is complementary to it. A common error is to transcribe the sense strand.*



1. The ribosome attaches to the mRNA and moves along it to the **start codon** (AUG). The tRNA with the complementary anticodon (UAC) and its amino acid (methionine) binds to the start codon.
2. A second tRNA, with the anticodon complementary to the next codon, binds. The ribosome holds two tRNAs at a time.
3. A peptide bond forms between the two amino acids, by a condensation reaction, and the first tRNA is released to pick up another amino acid.
4. The ribosome moves along the mRNA by one codon (three bases) and the next tRNA binds. The polypeptide chain gets longer one amino acid at a time.
5. This continues until the ribosome reaches a **stop codon** (UAA, UAG or UGA). No tRNA matches a stop codon, so translation ends and the polypeptide is released.

7 / 11

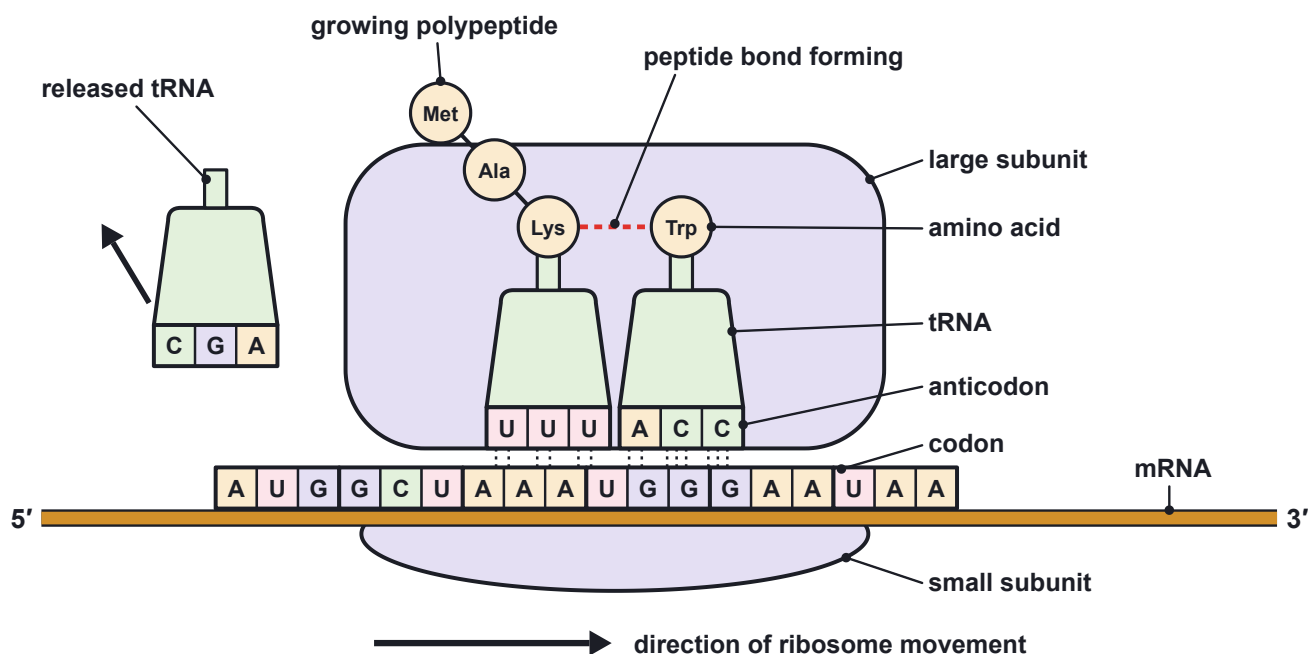
## From DNA to polypeptide

A gene has the template strand sequence TAC TTT CCC ATT. Using the table of codons (AUG = methionine, AAA = lysine, GGG = glycine, UAA = stop), state the mRNA, the tRNA anticodons and the amino acid sequence.

1. mRNA is complementary to the template: TAC → AUG; TTT → AAA; CCC → GGG; ATT → UAA.
2. tRNA anticodons are complementary to the mRNA codons: AUG → UAC; AAA → UUU; GGG → CCC (no tRNA for UAA).
3. Translate each codon: AUG = methionine; AAA = lysine; GGG = glycine; UAA = stop.

**Answer:** mRNA: AUG AAA GGG UAA. Anticodons: UAC, UUU, CCC. Polypeptide: methionine - lysine - glycine.

*Common mistake: The anticodon is on the tRNA; the codon is on the mRNA. The anticodon is complementary to the mRNA codon, which means it has the same base sequence as the template strand but with U instead of T.*



**Translation** at the ribosome: tRNA anticodons pair with mRNA codons and a **peptide bond** forms between amino acids.

## Gene mutations Spec 2.14

A **mutation** is a change in the sequence of bases in DNA. Mutations can arise from errors in DNA replication, when the wrong nucleotide is inserted, or when a base is added or missed out.

### Types of gene mutation

| Type                | What happens                     | Possible effect                                                                                                                                                |
|---------------------|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Substitution</b> | One base is replaced by another. | Only one codon changes. The amino acid may stay the same (code is degenerate), change to a different amino acid, or become a stop codon (shorter polypeptide). |
| <b>Insertion</b>    | An extra base is added.          | Reading frame shifts (frameshift): every codon after the insertion changes, so a completely different sequence of amino acids is likely.                       |
| <b>Deletion</b>     | A base is lost.                  | Frameshift: every codon after the deletion changes. (Adding or deleting three bases would add or remove one amino acid without shifting the rest.)             |

### Effect of a deletion

The mRNA sequence AUG GAA CCU UUA is read in codons. The first G after AUG is deleted, giving AUG AAC CUU UA. State the effect on the codons.

1. Original codons: AUG GAA CCU UUA (four codons).
2. After the deletion the bases are AUGAACCUUUA, read in threes: AUG AAC CUU UA.
3. Every codon after the deletion has changed (GAA → AAC, CCU → CUU), so the amino acid sequence after the mutation is changed.

**Answer:** A frameshift: all the codons after the deletion are changed, giving a different amino acid sequence.

### Consequences of mutations

- Some mutations cause **cancer**, if they occur in genes that control cell division, leading to uncontrolled cell division and a tumour.
- Some cause **genetic disorders**, if the protein made is faulty, for example the mutation that causes cystic fibrosis.
- Many mutations have **no observable effect**: the code is degenerate so the amino acid may not change; the change may be in a part of DNA that does not code for a protein; or the new amino acid may not alter the shape of the protein.

*Exam tip: In a question about the effect of a mutation, work through the chain: base sequence → mRNA codon → amino acid sequence → tertiary structure → function. Say whether it is a substitution or a frameshift.*

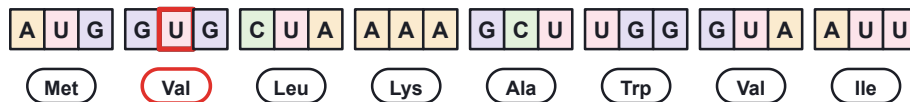
mRNA codons, and the amino acids they code for

original



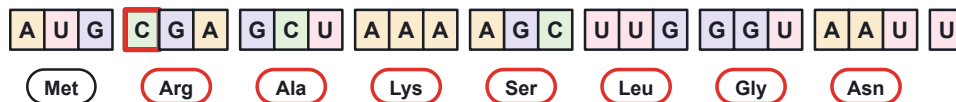
reading frame: triplets (codons)

substitution



A → U: one amino acid changes (Glu → Val)

insertion



extra C: frameshift, every codon after it changes

deletion



first G of GAG lost: frameshift, early STOP codon

A **substitution** changes at most one amino acid; an **insertion** or **deletion** causes a **frameshift**.

**Quick check**

- 1 . Name the bond that joins nucleotides in one DNA strand.
- 2 . Which bases pair together in DNA?
- 3 . What does semi-conservative mean?
- 4 . What is the role of RNA polymerase in transcription?
- 5 . What is an anticodon?

## Quick check answers

### DNA, the genetic code and gene expression

1. A phosphodiester bond, formed by a condensation reaction.
2. Adenine with thymine (two hydrogen bonds), cytosine with guanine (three hydrogen bonds).
3. Each new DNA molecule has one original strand and one newly made strand.
4. It breaks the hydrogen bonds, and joins RNA nucleotides complementary to the template strand by phosphodiester bonds.
5. Three bases on a tRNA molecule that are complementary to a codon on mRNA.