

Cambridge AS & A Level · Biology 9700 · Notes

Nucleic acids and protein synthesis AS

Read online at pastopic.com/cambridge/biology-9700/notes/nucleic-acids-and-protein-synthesis

Last checked 30 September 2026

Read first: [Biological molecules](#)

© 2026 Pastopic. Free for personal study. Please share the link, not the file.

What you need to know

By the end of this topic you should be able to:

Nucleic acids and DNA replication

1. **Describe a nucleotide:** a pentose sugar, a phosphate group and a nitrogenous base. Describe **ATP** as a nucleotide with three phosphate groups. You don't need structural formulae.
2. **Name the two kinds of base:** adenine and guanine are **purines** (two rings); cytosine, thymine and uracil are **pyrimidines** (one ring). You don't need their formulae.
3. **Describe DNA as a double helix:** two **antiparallel** strands (one runs 5' to 3', the other 3' to 5') held together by **complementary base pairing**, with **two hydrogen bonds** in each A–T pair and **three** in each C–G pair, and nucleotides in each strand joined by **phosphodiester bonds**.
4. **Describe semi-conservative replication** in the **S phase** of the cell cycle, including what **DNA polymerase** and **DNA ligase** do, and why one new strand (the **leading strand**) is made in one piece while the other (the **lagging strand**) is made in fragments: DNA polymerase can only add nucleotides in the 5' to 3' direction. Other replication enzymes are not needed.
5. **Describe an RNA molecule**, using messenger RNA (mRNA) as the example.

Protein synthesis

6. **State that a gene** is a sequence of nucleotides, part of a DNA molecule, that codes for a polypeptide.
7. **Describe the universal genetic code:** triplets of bases code for particular amino acids, or act as start and stop signals, and the same triplets mean the same thing in every organism.
8. **Describe transcription and translation**, including the roles of RNA polymerase, mRNA, codons, transfer RNA (tRNA), anticodons and ribosomes.
9. **Name the two DNA strands:** the one copied in transcription is the **transcribed** (or **template**) strand; the other is the **non-transcribed** strand.
10. **Explain splicing:** in eukaryotes the first RNA made (the **primary transcript**) has its non-coding sequences (**introns**) removed and its coding sequences (**exons**) joined to make mRNA.
11. **State that a gene mutation** is a change in the sequence of base pairs in DNA that may give an altered polypeptide.

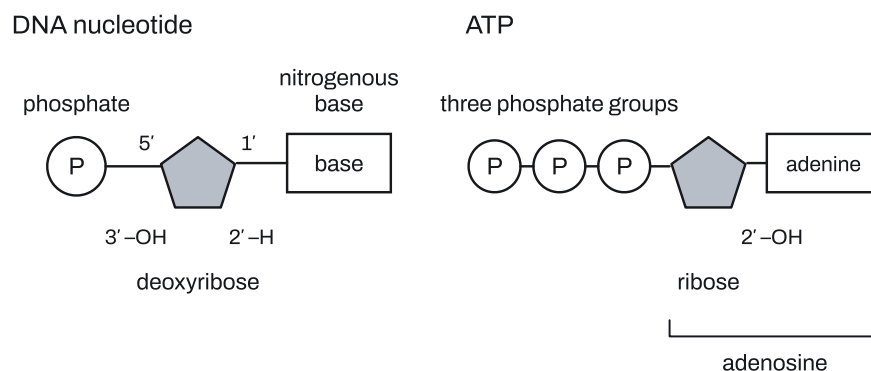
12. **Explain the three kinds of gene mutation**, substitution, deletion and insertion of nucleotides, and outline how each may change the polypeptide.

Understand it

Nucleotides: the building blocks

A **nucleotide** has three parts joined by condensation reactions:

- a **pentose** (five-carbon) sugar: **deoxyribose** in DNA, **ribose** in RNA. The carbons are numbered 1' to 5'. The base joins carbon 1', and the phosphate joins carbon 5'. Deoxyribose has only H on carbon 2' where ribose has –OH (that is what "deoxy" means).
- a **phosphate group**;
- a **nitrogenous base**.



ATP (adenosine triphosphate) is a nucleotide too: **adenine + ribose + three phosphate groups**. It is the adenine RNA nucleotide with two extra phosphates. Adenine joined to ribose is called **adenosine**. ATP is made in respiration and is the cell's energy currency. It differs from the activated DNA nucleotide that carries adenine (which also has three phosphates) only in its sugar: **ribose, not deoxyribose**.

Bases: purines and pyrimidines

There are five bases. Two are large, with a **double ring**: the **purines**, **adenine (A)** and **guanine (G)**. Three are smaller, with a **single ring**: the **pyrimidines**, **cytosine (C)**, **thymine (T)** and **uracil (U)**. A memory aid: "pure As Gold" for the purines.

- DNA contains A, T, C and G. RNA contains A, **U**, C and G. So no nucleic acid contains all five: T is only in DNA and U is only in RNA.
- In a base pair, a purine always pairs with a pyrimidine. Every pair is therefore three rings wide, and the two strands stay the **same distance apart** all along the molecule.

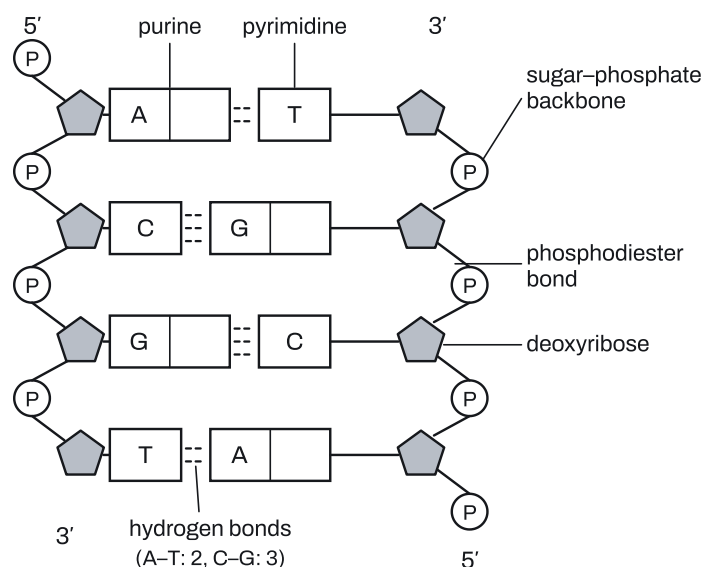
DNA: a double helix

A **polynucleotide** forms when nucleotides join by **condensation**: the phosphate on carbon 5' of one nucleotide bonds to the –OH on carbon 3' of the next. This covalent link, sugar–O–P–O–sugar, is a **phosphodiester bond**. The alternating sugars and phosphates make a strong **sugar–phosphate backbone**, with the bases sticking out to the side.

One end of a strand has a free phosphate on carbon 5' (the **5' end**); the other has a free –OH on carbon 3' (the **3' end**).

DNA has **two** polynucleotide strands, twisted into a **double helix**:

- The backbones are on the outside and the bases point inwards, protected.
- The strands are **antiparallel**: they run in opposite directions, one 5' → 3' and the other 3' → 5'.
- Bases on the two strands pair up by **complementary base pairing**: **A with T** (two hydrogen bonds) and **C with G** (three hydrogen bonds). Only these pairs fit, because of their sizes and the positions of the atoms that form hydrogen bonds.
- Each hydrogen bond is weak, but there are millions of them, so the molecule is stable. They can also be broken easily enough to separate the strands for copying. DNA with more C–G pairs has more hydrogen bonds, so it needs more energy to separate.



Because A always pairs with T and C with G, in double-stranded DNA the **amount of A equals the amount of T**, and the **amount of C equals the amount of G**. So purines make up 50% of the bases. This is true for the whole molecule, **not** for each strand on its own.

RNA and mRNA

RNA is also a polynucleotide, but:

	DNA	mRNA
Strands	two, a double helix	one
Sugar	deoxyribose	ribose
Bases	A, T, C, G	A, U , C, G
Base pairing	between the two strands, by hydrogen bonds	none within the molecule (only briefly, with DNA or tRNA)
Length	very long (a whole chromosome)	short (a copy of one gene)
Where	nucleus (also mitochondria and chloroplasts)	made in the nucleus, used in the cytoplasm

Uracil pairs with adenine, using two hydrogen bonds, just as thymine does.

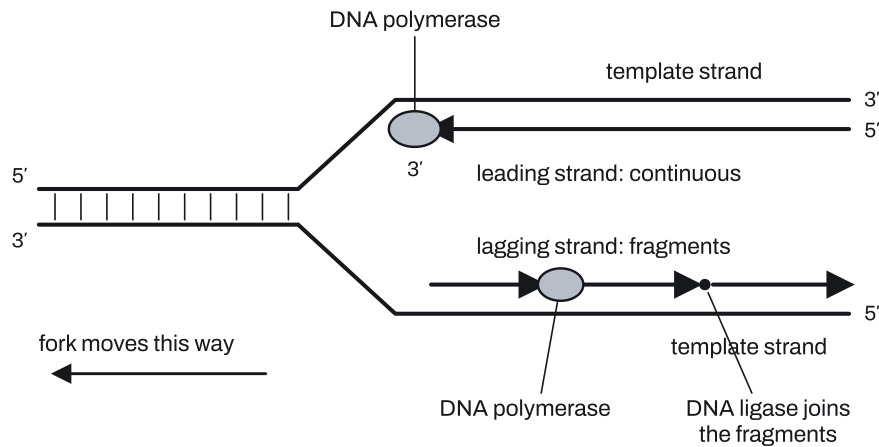
DNA replication

Before a cell divides, its DNA is copied during the **S phase of interphase**, in the nucleus. The process is **semi-conservative**: each new DNA molecule has **one strand from the original molecule and one newly made strand**.

1. The double helix **unwinds** and the **hydrogen bonds between the bases break**, so the two strands separate. Both strands act as **templates**.
2. Free **activated nucleotides** (each carrying three phosphate groups) in the nucleus pair with the exposed bases by **complementary base pairing**: A with T, C with G. Hydrogen bonds form between the paired bases.
3. **DNA polymerase** joins each new nucleotide to the end of the growing strand, forming a **phosphodiester bond** between the sugar of one nucleotide and the phosphate of the next (a condensation reaction). The two spare phosphates are released, which supplies the energy. DNA polymerase also **proofreads**: it checks each new base pair and removes a wrongly paired nucleotide.
4. The result is two DNA molecules, each with one old and one new strand, each identical to the original.

DNA polymerase can only add a nucleotide to the free **3' end** of a strand, so a new strand always grows in the **5' → 3' direction**. As the new strand is antiparallel to its template, the polymerase moves along the template from its 3' end towards its 5' end. The two templates run in opposite directions, and this creates a difference:

- **Leading strand**: its template lets DNA polymerase move **towards** the point where the helix is opening (the **replication fork**), so this strand is made **continuously**, in one piece.
- **Lagging strand**: on the other template, 5' → 3' means moving **away** from the fork. The polymerase makes a short piece, then starts a new one nearer the fork as more DNA unwinds. This strand is made in **short fragments** (Okazaki fragments).
- **DNA ligase** joins the fragments of the lagging strand by forming **phosphodiester bonds** between them, making one continuous strand.



Replication also happens in mitochondria and chloroplasts, which have their own DNA.

Genes and the genetic code

A **gene** is a sequence of nucleotides, part of a DNA molecule, that codes for **one polypeptide**. The order of bases in the gene decides the order of amino acids in the polypeptide (its primary structure). A protein made of two different polypeptides needs two genes; one made of several copies of the same polypeptide needs only one.

The **genetic code** is read in **triplets**: three bases code for one amino acid. With four bases there are $4^3 = 64$ triplets (pairs would give only $4^2 = 16$, too few for 20 amino acids).

- **Degenerate**: most amino acids have more than one triplet (64 triplets for 20 amino acids). Leucine has six; tryptophan has only one.
- **Start and stop**: the mRNA codon **AUG** starts translation (it also codes for methionine). **UAA, UAG and UGA** are **stop** codons: they code for no amino acid and end the polypeptide.
- **Non-overlapping**: each base belongs to one triplet only, read one after another from the start codon.
- **Universal**: the same triplet codes for the same amino acid in (almost) all organisms. This is why a human gene put into a bacterium makes the human protein.

Because of degeneracy, you cannot work out the exact DNA sequence from a polypeptide's amino acid sequence.

Transcription: DNA → primary transcript (nucleus)

1. The part of the DNA holding the gene **unwinds** and the hydrogen bonds between its bases break.
2. **RNA polymerase** binds to the DNA at the start of the gene. Only one strand, the **transcribed (template)** strand, is copied. The other is the **non-transcribed** strand.
3. Free **RNA nucleotides** pair with the exposed bases of the template: **A-U, T-A, C-G, G-C**.
4. RNA polymerase joins the RNA nucleotides with **phosphodiester bonds**, making an RNA strand $5' \rightarrow 3'$ as it moves along the template.

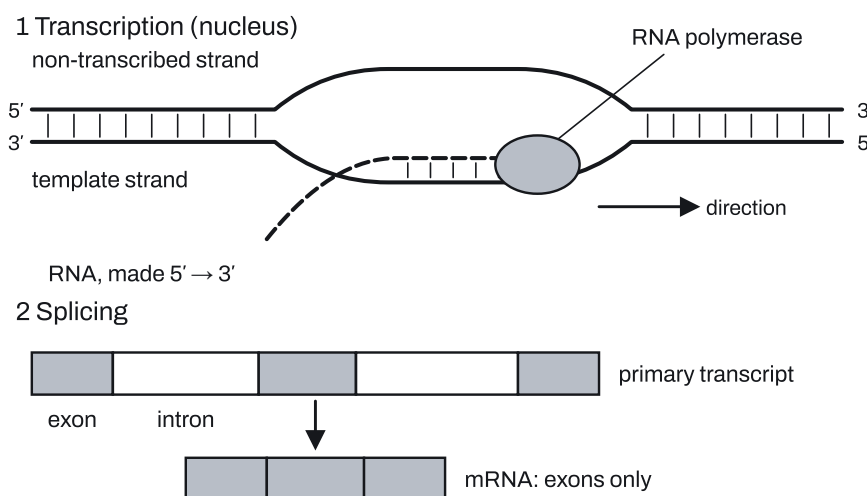
5. At the end of the gene the RNA is released, and the two DNA strands re-form their hydrogen bonds.

The RNA is **complementary** to the template strand, so its sequence is the **same as the non-transcribed strand, with U in place of T**.

Only one strand of a gene is transcribed. The other strand has a different base sequence, so RNA copied from it would code for a different sequence of amino acids (or none that works), and the polypeptide would not function.

Splicing: primary transcript → mRNA (eukaryotes)

In eukaryotes a gene contains **exons** (sequences that code for amino acids) separated by **introns** (non-coding sequences). Transcription copies both, making the **primary transcript**. Before it leaves the nucleus, the primary transcript is modified: the **introns are cut out and the exons are joined together**. This gives the **mRNA**, which is shorter than the primary transcript. The mRNA leaves the nucleus through a nuclear pore.



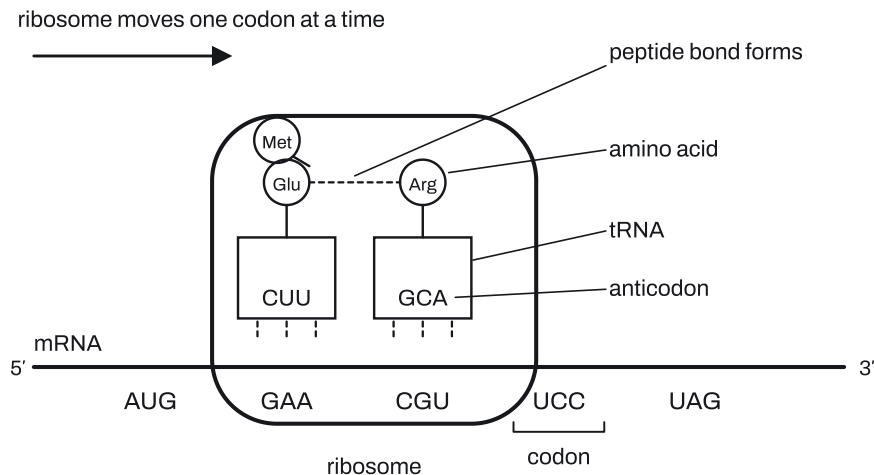
Prokaryotes have no introns and no nucleus, so their mRNA needs no splicing, and it can be translated while it is still being made.

Translation: mRNA → polypeptide (ribosomes in the cytoplasm)

Each group of three bases on mRNA is a **codon**. A **tRNA** molecule carries one specific amino acid at one end and has three bases, its **anticodon**, at the other. The anticodon is complementary to a codon.

1. The mRNA binds to a **ribosome**. The ribosome holds two codons at a time.
2. A tRNA with the anticodon complementary to the start codon (AUG), carrying methionine, pairs with it by **hydrogen bonds** between complementary bases.
3. A second tRNA, carrying its amino acid, pairs with the next codon.
4. A **peptide bond** forms between the two amino acids (a condensation reaction, catalysed by the ribosome). The first tRNA, now without its amino acid, leaves and can collect another.
5. The ribosome **moves along the mRNA by one codon**, and another tRNA pairs with the new codon. The chain grows one amino acid at a time.

6. At a **stop codon** no tRNA pairs, and the finished polypeptide is released. It then folds into its secondary and tertiary structure.



Several ribosomes can translate one mRNA at the same time. Ribosomes in the eukaryotic cytoplasm are **80S**; those in prokaryotes (and in mitochondria and chloroplasts) are smaller, **70S**. Translation needs mRNA, ribosomes, tRNA, amino acids and ATP; it does **not** need DNA or RNA polymerase.

Following a sequence through

Molecule	Example
DNA template (transcribed) strand	TAC CTT GCA
DNA non-transcribed strand	ATG GAA CGT
mRNA codons	AUG GAA CGU
tRNA anticodons	UAC CUU GCA
Amino acids	Met – Glu – Arg

The mRNA matches the **non-transcribed** strand (T → U). The anticodon matches the **template** strand (T → U).

Gene mutations

A **gene mutation** is a change in the sequence of base pairs in DNA. It is copied into the mRNA, so it may change the polypeptide.

- **Substitution**: one base pair replaced by another. Only **one codon** changes. The effect can be:
 - **none**, if the new codon codes for the same amino acid (the code is degenerate);
 - **one different amino acid**. This may change the R group, the bonds that hold the tertiary structure, and so the shape (for example, an enzyme's active site). Sickle cell anaemia is caused by a substitution: the codon GAG becomes GUG, so glutamic acid is replaced by valine in the β -globin polypeptide.
 - a **stop codon**, so the polypeptide is **shorter** and usually does not work.

- **Deletion** or **insertion** of one nucleotide: every triplet after it is read differently (a **frameshift**). All the amino acids after the mutation may change, and a stop codon may appear early or be lost. The protein is usually non-functional.
- Deleting or inserting **three** nucleotides (one whole triplet) removes or adds one amino acid without shifting the frame.
- A mutation in an **intron** usually has no effect on the polypeptide, since introns are removed.

Key facts and formulas

This topic has no formulas in the list printed in the syllabus, so everything below is **Learn it**.

Formula	Status
Base pairing in DNA: A–T (2 hydrogen bonds), C–G (3 hydrogen bonds)	Learn it
Base pairing with RNA: A–U (2), T–A (2), C–G (3)	Learn it
Double-stranded DNA: $\%A = \%T$, $\%C = \%G$, $\%A + \%G = 50\%$, so $\%C = \frac{100 - 2 \times \%A}{2}$	Learn it
Hydrogen bonds = $2 \times (\text{A–T pairs}) + 3 \times (\text{C–G pairs})$	Learn it
Phosphodiester bonds: one strand of n nucleotides has $n - 1$; linear double-stranded DNA of n base pairs has $2(n - 1)$; circular DNA of n base pairs has $2n$	Learn it
Triplet code: $4^3 = 64$ triplets for 20 amino acids; start codon AUG (methionine); 3 stop codons (UAA, UAG, UGA); 61 codons for amino acids	Learn it
Number of amino acids = $\frac{\text{coding nucleotides in mRNA}}{3} - 1$ (the last codon is a stop); nucleotides for the amino acids alone = $3 \times \text{amino acids}$	Learn it
Replication forks = $\frac{\text{base pairs}}{\text{bp per second per fork} \times \text{time in seconds}}$	Learn it
Number of amino acids $\approx \frac{\text{mass of polypeptide}}{\text{mean mass of an amino acid}}$	Learn it
Meselson–Stahl: after g generations of ^{15}N DNA in ^{14}N , fraction hybrid = $\frac{2}{2^g}$, the rest light, none heavy	Learn it

How to do it

In the 165 questions we have tagged for this topic (129 from Paper 1, 36 from Paper 2, 2021–2025), the types came up this often. 33 questions mix more than one type, so the counts add up to 215. Paper 1 questions are single multiple-choice marks; Paper 2 parts often sit inside a question on another topic.

Question type	Questions
Classify bases and apply base-pairing rules	31
Describe transcription and translation; name the parts	31
Nucleotides, ATP and polynucleotides: parts and bonds	23
Describe DNA replication	22
Convert sequences and read codon tables	21
Identify a gene mutation and predict its effect	21
Genes, one gene one polypeptide, the genetic code	18
Explain splicing of the primary transcript	14
Calculate base numbers, lengths, bond counts or amino acids	14
Compare DNA and RNA	8
Explain how an inhibitor, analogue or interfering RNA blocks a process	8
Predict heavy, hybrid and light DNA (Meselson–Stahl)	4

1. Classify bases and apply base-pairing rules (31 questions)

- Purine or pyrimidine?** Two rings (A, G) → purine; one ring (C, T, U) → pyrimidine. For an unfamiliar or synthetic base in a diagram, **count the rings** and say so: "it has a double ring, so it is a purine".
- Drawing hydrogen bonds:** dotted lines, 2 for A–T and 3 for C–G, between atoms that face each other (an H on one base, an O or N on the other).
- Identify a base from a diagram:** ring count gives purine or pyrimidine; hydrogen bonds to its partner give A/T (2) or C/G (3).

Variations you will meet:

- Tables** ("name of base", "purine or pyrimidine", "in RNA, DNA or both"): A, G and C are in **both**; T is **DNA** only; U is **RNA** only. Spell thymine and uracil in full.
- Why the strands are a constant distance apart:** a purine always pairs with a pyrimidine, so each pair is the same width.
- RNA pairing with RNA** (a hairpin, or double-stranded RNA): the same rules, with U in place of T: A–U and G–C, held by hydrogen bonds.
- Judge a synthetic base pair:** the DNA pair most like it is the one with the same number of hydrogen bonds (3 → C–G, 2 → A–T) and the same purine–pyrimidine arrangement.

- **Statements to judge:** base pairs are all the same length; purines and pyrimidines are **different** sizes; C–G has three bonds, not two; pairing happens in replication, transcription **and** translation.

2. Describe transcription and translation; name the parts (31 questions)

1. **One-word answers:** the enzyme is **RNA polymerase**; the strand used is the **transcribed** or **template** strand (not "coding" or "leading"); the first RNA made is the **primary transcript**; translation happens at **ribosomes**; the process that makes RNA is **transcription**.
2. **Describe** (4–6 marks): give the steps from "Understand it" in order. The mark-earning words for translation are ribosome, **specific** amino acid, **anticodon complementary to codon**, hydrogen bonds, **peptide bond**, ribosome moves one codon, **stop codon**; for transcription, hydrogen bonds break, RNA polymerase, free RNA nucleotides, template strand, phosphodiester bonds, primary transcript.

Variations you will meet:

- **Order of events** from a list: in transcription, hydrogen bonds break (DNA separates), form (RNA nucleotides pair), break (RNA leaves the DNA) and form (DNA strands rejoin).
- **What translation needs:** mRNA, ribosomes, tRNA and amino acids (and ATP). Not DNA, not RNA polymerase, not free nucleotides.
- **Label a diagram:** codon on mRNA, anticodon on tRNA, polypeptide attached to the tRNA, ribosome. In a prokaryote the ribosome is **70S**.
- **Which strand matches the mRNA?** The non-transcribed strand, with T for U.
- **Role of tRNA** (not of translation in general): carries a specific amino acid to the ribosome; its anticodon pairs with the complementary codon, so amino acids are placed in the order the mRNA codes for. After the peptide bond forms it leaves and is reused.
- **Statements about anticodons:** an anticodon is complementary to the codon, so it has the same bases as the template-strand triplet **except U where the DNA has T**; "always the same as the template triplet" is false. There are **more than 20** different anticodons, because most amino acids have several codons. Codon–anticodon hydrogen bonds must form **before** the peptide bond can.
- **Why only one strand is transcribed:** the other strand has a different base sequence, so its RNA would code for a different amino acid sequence and a non-functional polypeptide.

3. Nucleotides, ATP and polynucleotides: parts and bonds (23 questions)

1. Name the part asked for (pentose, phosphate, base; adenosine; ribose) and the bond: **phosphodiester** along a strand (covalent, condensation), **hydrogen** between strands.

Variations you will meet:

- **"State the meaning of monomers of DNA":** the repeating units, nucleotides, that are joined to make the polymer.
- **How ATP differs from a DNA nucleotide with three phosphates:** its sugar is ribose, which has –OH (not –H) on carbon 2'.
- **Circle one nucleotide** on a strand: include one phosphate, one sugar and one base, all joined together.

- **Why a single base pair on its own has no phosphodiester bond:** those bonds join nucleotides **along** a strand, and each strand here has only one nucleotide.
- **Elements:** nucleic acids contain C, H, O, N and P (proteins have no P; most contain S).
- **Formulae of the sugars:** ribose is $C_5H_{10}O_5$ and deoxyribose $C_5H_{10}O_4$ (one oxygen fewer), so ribose has carbon and oxygen 1 : 1 and deoxyribose 5 : 4.
- **Unfamiliar nucleotides** (a changed sugar, base or phosphate): name which of the three parts has been replaced.

4. Describe DNA replication (22 questions)

1. **When and where:** the **S phase** of interphase (write "S phase", not just "S"), in the nucleus.
2. **Describe or explain:** follow "Understand it" and Example 4. Name **DNA polymerase** and **DNA ligase** in full, and give each a role. DNA polymerase: adds activated nucleotides complementary to the template, forms phosphodiester bonds between adjacent nucleotides, works $5' \rightarrow 3'$ only, and **proofreads** for mismatched bases. DNA ligase: joins the lagging-strand fragments with phosphodiester bonds, completing the sugar–phosphate backbone. A question on both enzymes wants something about each.

Variations you will meet:

- **Diagram of the direction** of polymerase on the two templates: pick the one where each arrow points towards the template's **5' end**, so the two arrows point opposite ways.
- **Rows about the leading and lagging strands:** here "leading strand" and "lagging strand" mean the **new** strands. DNA polymerase builds **both** $5' \rightarrow 3'$ (so "moves along the leading and lagging strands $5' \rightarrow 3'$ " is right). Each lagging-strand fragment grows $5' \rightarrow 3'$, but because every new fragment starts nearer the fork, the lagging strand **as a whole** grows $3' \rightarrow 5'$, and ligase joins the fragments; the leading strand grows continuously $5' \rightarrow 3'$.
- **The sequence of events at one nucleotide** from a structural diagram: an activated nucleotide (three phosphates) pairs with its complementary base by hydrogen bonds **first**; then DNA polymerase forms the phosphodiester bond between its phosphate and the $3' -OH$ of the growing strand, releasing two phosphates.
- **Which enzyme at which stage** of a flow chart: ligase at the last stage (joining fragments). Polymerase adds nucleotides.
- **Products:** two DNA molecules, each with one original and one new strand, identical to each other.

5. Convert sequences and read codon tables (21 questions)

1. **Write each molecule under the last**, three bases at a time:
 - template strand $\rightarrow$ mRNA: complementary, with **U opposite A**;
 - mRNA $\rightarrow$ anticodon: complementary again, so the **anticodon has the same bases as the template triplet**, with U for T;
 - non-transcribed strand $\rightarrow$ mRNA: same bases, T $\rightarrow$ U.

2. **Check which strand the table uses** before reading it. A table of "DNA triplets" may be the template strand or the non-transcribed strand, and a table of codons is mRNA. Test with methionine: its template triplet is TAC, non-transcribed ATG, codon AUG. In 9700 papers a table headed just "DNA triplets" with no Met to test is taken to be the **template** strand.
3. **Check the direction.** mRNA is read 5' → 3', and it runs antiparallel to the template. If a template strand is printed 5' → 3' from left to right, reverse it (read it 3' → 5') before writing the complement.
4. Read the amino acids in order, stopping at a stop triplet. The stop triplets are UAA, UAG, UGA on mRNA, TAA, TAG, TGA on the non-transcribed strand and **ATT, ATC, ACT** on the template strand.

Variations you will meet:

- **Anticodon for an amino acid** from a table of template-strand triplets: swap T for U in the triplet. From a table of mRNA codons: write the complement.
- **Many possible DNA sequences** for one short polypeptide: multiply the numbers of triplets for each amino acid (degeneracy).
- **A codon wheel** usually shows **mRNA codons**, so convert any DNA to mRNA first. If the wheel has T instead of U, check its label: a wheel of **non-transcribed-strand** DNA triplets is read with the mRNA codon written with T for U (mRNA UUA → TTA on the wheel). Find the first base in the centre, the second in the middle ring and the third in the outer ring; the amino acid is printed at the edge next to it. For example, a template triplet printed 5' CAT 3' is read 3' TAC 5', so the codon is AUG (Met). Complementing it without reversing gives GUA (Val), which is the trap option.
- **mRNA → non-transcribed strand**: same bases with T for U (GAU → GAT). The template strand is the complement.
- **A stop triplet given a new meaning** (a changed genetic code): the new tRNA's anticodon is still the template triplet with U for T (template ATC → anticodon AUC).
- **RNA from a telomere or other non-coding DNA**: still complementary to the strand it was copied from.

6. Identify a gene mutation and predict its effect (21 questions)

1. **Name the type** by lining up the old and new sequences: same length, one base different → **substitution**; one fewer → **deletion**; one more → **insertion**.
2. **Translate both** with the table, codon by codon.
3. **Explain the effect**: changed codon(s) → changed amino acid(s) (or a stop codon) → changed **primary structure** → changed bonds between R groups → changed **tertiary structure / shape** → protein (for example, an enzyme's active site) may not work.

Variations you will meet:

- **Substitution with no effect**: say the new triplet codes for the **same amino acid**, because the code is **degenerate**.
- **Substitution giving a stop codon**: the polypeptide is **shorter**.
- **Deletion or insertion**: say "**frameshift**: every triplet after the mutation changes", then give the new amino acids from the table. Deleting a whole triplet is **not** a frameshift: one amino acid is lost.

- **"Which mutation gives the shorter chain / two different amino acids?":** work through each option; don't guess from the mutation type.
- **Where can the change be seen?** In the DNA, the mRNA and the amino acid sequence.
- **Mutation in an intron** (or in a regulatory sequence): no change in the polypeptide, though it can change how much is made.
- **An insertion and a deletion close together:** only the triplets between them change, and the reading frame is restored after the second one. The amino acid sequence changes and the protein may not work. Check each changed triplet for a stop (ATT, ATC, ACT on the template strand; TAA, TAG, TGA on the non-transcribed strand): if none is a stop, reject any statement that stop codons appear, and the mRNA can still be translated.
- **Which change makes a stop?** On the template strand a stop triplet is ATT, ATC or ACT, so look for a change that creates one of these.

7. Genes, one gene one polypeptide, the genetic code (18 questions)

1. **Gene:** "a sequence of nucleotides, part of a DNA molecule, that codes for a polypeptide". Say both parts.
2. **Universal:** the same triplets code for the same amino acids in all organisms. **Degenerate:** most amino acids have more than one triplet. (Not "degenerative".)

Variations you will meet:

- **How many genes for a protein?** Count the **different** polypeptides: two different chains → two genes; identical chains → one gene; three chains that could be the same or different → 1, 2 or 3.
- **Why 20 amino acids but 64 triplets:** several triplets per amino acid, plus 3 stop codons.
- **Differences between two genes** on one chromosome: different base sequences, different lengths, different numbers of introns and exons, coding for different polypeptides.
- **What DNA codes for directly:** polypeptides and so proteins (such as enzymes). Carbohydrates and lipids are made by enzymes, not coded directly.
- **Why you can't work back from a polypeptide to one DNA sequence:** degeneracy (and introns are not in the polypeptide).

8. Explain splicing of the primary transcript (14 questions)

1. Say both halves: **introns are removed** from the primary transcript and **exons are joined**, giving a shorter mRNA.

Variations you will meet:

- **Relationships** ("template strand and primary transcript", "primary transcript and mRNA"): the primary transcript is a complementary RNA copy of the template strand; the mRNA is the primary transcript with its introns removed and exons joined.
- **Introns and exons compared:** both are sequences in the gene and in the primary transcript; only exons code for amino acids and stay in mRNA.

- **A gene with no introns** (or a prokaryote): no splicing is needed; the primary transcript already has the same sequence as the mRNA.
- **Extra ends added in the nucleus** (a cap and a tail, if a question shows them): suggest that they protect the mRNA from enzymes that break it down, and help it leave through a nuclear pore and attach to a ribosome.

9. Calculate base numbers, lengths, bond counts or amino acids (14 questions)

1. **Percentages:** $A = T$ and $C = G$ in double-stranded DNA. If guanine is 23%, cytosine is 23%, leaving 54% for A and T, so each is 27%. For a **number** of bases, remember that N base pairs hold $2N$ bases: a 5000 bp molecule with 23% guanine has $0.23 \times 10\,000 = 2300$ cytosine bases.
2. **Strand tables:** A on one strand = T on the other; C on one = G on the other. Fill in each strand, then add up. If only totals are needed, use the whole molecule: total T = total A, and the rest splits equally, so total G = $\frac{2N - 2 \times \text{total A}}{2}$. For 60 bp with 20 A on one strand and 14 on the other: total A = 34, so $G = (120 - 68) \div 2 = 26$.
3. **Lengths:** 3 nucleotides of mRNA per amino acid, plus 3 for the stop codon. A 300-amino-acid polypeptide needs $300 \times 3 = 900$ nucleotides for its amino acids, or 903 counting the stop codon. Work out **both**: when the options show the "× 3" number but not the "+ 3" one, the question is counting only the codons for amino acids. If the first amino acid (methionine) was removed after translation, add it back before multiplying.
4. **From a mass:** amino acids $\approx \text{mass} \div \text{mean amino acid mass}$ ($44\,000 \text{ Da} \div 110 \text{ Da} = 400$). The mRNA is about three times that many bases; gene length minus mRNA length is roughly the introns.
5. **Bonds:** hydrogen bonds = $2(A-T) + 3(C-G)$. Phosphodiester bonds: $n - 1$ per linear strand, so a 10 bp linear double strand has 18; a circle has as many as it has nucleotides (a 10 bp circle has 20).
6. **Replication forks:** forks = base pairs $\div$ (base pairs per second for one fork $\times$ time in seconds). A 120 million bp chromosome copied in 50 minutes at 40 bp per second per fork needs $120\,000\,000 \div (40 \times 3000) = 1000$ forks. Convert the time to seconds first.

Variations you will meet:

- **Maximum amino acids from double-stranded DNA of N nucleotides:** only one strand is read, so $\frac{N}{2}$ bases, and $\frac{N}{6}$ amino acids at most (for a "maximum", ignore the stop codon and introns). Check whether N counts nucleotides or base pairs.
- **Why a long gene gives few amino acids** (Paper 2): three bases code for one amino acid; introns are non-coding and are removed from the primary transcript, and only exons are joined into mRNA; the stop codon codes for no amino acid; the first methionine may be removed after translation; the gene also includes non-coding regulatory sequences, such as the region where RNA polymerase binds.
- **Percentage of a base in RNA** made from a strand: it equals the complementary base's percentage **on the template strand**, not on the whole DNA. The RNA has the **same** base make-up as the **non-transcribed** strand ($T \rightarrow U$): if 35% of the non-transcribed strand's bases are purines, 35% of the transcript is A + G, so C + U is $100 - 35 = 65\%$.

- **Approximate lengths from a mass:** when the numbers are rounded (a gene in kb, a protein in kDa), mRNA $\approx 3 \times$ amino acids and the stop codon is too small to matter.

10. Compare DNA and RNA (8 questions)

1. Choose three structural differences: **number of strands** (two / one), **sugar** (deoxyribose / ribose), **base** (thymine / uracil), **base pairing** (hydrogen bonds between strands / none), **length** (longer / shorter).
2. Write **both** halves of each point: "DNA has deoxyribose; mRNA has ribose". A one-sided statement gets nothing when the question says "include information about both".

Variations you will meet:

- **mRNA compared with a double-stranded RNA** (such as siRNA): siRNA is double-stranded, with paired and unpaired nucleotides held by hydrogen bonds; mRNA is single-stranded, longer, and codes for amino acids.
- **Tracking RNA made in cells** with a label: use a label on **uracil** or **ribose**, which are only in RNA.

11. Explain how an inhibitor, analogue or interfering RNA blocks a process (8 questions)

These are "suggest" questions in unfamiliar contexts. Work out **which process** is blocked, then **how**, then **the result**:

1. **Name the step:** replication (DNA polymerase), transcription (RNA polymerase) or translation (ribosome, tRNA, mRNA).
2. **Mechanism**, using what you know:
 - binds the enzyme's **active site** (competes with nucleotides) or another site that **changes the active site's shape**;
 - binds the **DNA template** so the polymerase can't bind or move along;
 - a **nucleotide analogue** is built into the new strand but has **no –OH on carbon 3'**, so the next **phosphodiester bond cannot form** and the strand stops growing;
 - a short RNA or DNA **complementary to the mRNA** binds it by base pairing (a double-stranded RNA must first release one strand to pair), so ribosomes can't read it, or the mRNA is cut;
 - binds **70S** (bacterial) ribosomes but not **80S** (human) ones, so only bacteria are affected;
 - different steps need different enzymes, because each active site fits one specific site, and the enzymes may work in different places (nucleus or cytoplasm).
3. **Result:** no mRNA → no polypeptide; cut mRNA may be broken down before it reaches a ribosome, may be unable to attach to a ribosome, may lose its start codon (not translated) or its stop codon (the chain is not released); a shorter or wrong polypeptide may be broken down in the cytoplasm, or doesn't fold into its tertiary structure (R groups form different bonds) → no functional protein.

For example, cordycepin (3'-deoxyadenosine) is adenosine without the –OH on carbon 3'. As a triphosphate, RNA polymerase can add it to a growing RNA strand opposite a T in the template, because it pairs like adenine (a purine). Then no nucleotide can join after it, because there is no 3' –OH to form the next phosphodiester bond, so the RNA stops short.

12. Predict heavy, hybrid and light DNA (Meselson–Stahl) (4 questions)

Bacteria grown on heavy nitrogen (^{15}N) have heavy DNA. They are moved to light nitrogen (^{14}N), and their DNA is spun in a centrifuge.

1. The two original heavy strands survive, each paired with a light strand, so every original molecule leaves **exactly 2 hybrid molecules**; the rest are light.
2. After g generations there are 2^g molecules per original, so the fraction hybrid is $\frac{2}{2^g}$.
3. If the question counts **strands**, not molecules: there are 2×2^g strands and 2 of them are heavy, so the fraction of heavy strands is $\frac{1}{2^g}$ (25% after 2 generations).
4. In the centrifuge tube, heavy DNA forms the **lowest** band, hybrid the middle band and light DNA the **highest** (upper) band.

Generation	Heavy	Hybrid	Light
0	100%	0	0
1	0	100%	0
2	0	50%	50%
3	0	25%	75%

If the question gives a division time, divide the time by it to get the number of generations first.

Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- **M1** method mark: for a correct method, even if the numbers slip.
- **A1** accuracy mark: for a correct answer; it needs the M mark before it.
- **B1** mark for a correct result on its own, with no method needed.

Example 1

A section of double-stranded DNA is 40 base pairs long. Strand 1 contains 8 adenine and 11 cytosine bases. Strand 2 contains 7 adenine bases.

(a) Calculate the number of guanine bases in strand 1 and the number of cytosine bases in strand 2. **[2]**

(b) Calculate the total number of hydrogen bonds between the two strands. **[2]**

(c) State the number of phosphodiester bonds in this section. **[1]**

(d) A second section of DNA of the same length has only 10 C–G pairs. Explain which section needs more energy to separate its strands. **[2]**

(a) Adenine in strand 2 pairs with thymine in strand 1, so strand 1 has 7 T. Strand 1: $40 - 8 - 11 - 7 = 14$ guanine. **B1**

Guanine in strand 1 pairs with cytosine in strand 2, so strand 2 has **14** cytosine. **B1**

(b) A–T pairs = $8 + 7 = 15$; C–G pairs = $11 + 14 = 25$. **M1**

Hydrogen bonds = $15 \times 2 + 25 \times 3 = 105$. **A1**

(c) Each strand of 40 nucleotides has 39 phosphodiester bonds, so 78. **B1**

(d) The first section, which has 25 C–G pairs. **B1** C–G pairs have three hydrogen bonds and A–T pairs two, so it has more hydrogen bonds to break. **B1**

Example 2

The table shows the template-strand DNA triplets for some amino acids.

Amino acid	Template-strand DNA triplets
Met	TAC
Glu	CTT, CTC
Arg	GCA, GCG, GCT, GCC, TCT, TCC
Ser	AGG, AGA, AGT, AGC, TCA, TCG
Asp	CTA, CTG
Val	CAA, CAG, CAT, CAC
Pro	GGA, GGG, GGT, GGC
stop	ATT, ATC, ACT

The start of the template strand of a gene is:

TAC CTT GCA AGG ATC

(a) Write the mRNA sequence transcribed from this DNA. **[1]**

(b) State the sequence of amino acids coded for. **[1]**

(c) State the anticodon of the tRNA that carries the third amino acid. **[1]**

(d) Explain the effect of each of these mutations on the polypeptide. Number the bases 1 to 15 from the left.

(i) The T at position 6 is replaced by C. **[2]**

(ii) The C at position 4 is replaced by A. **[2]**

(iii) The T at position 5 is deleted. **[3]**

(a) AUG GAA CGU UCC UAG **B1** (complementary to the template, with U opposite A)

(b) Met – Glu – Arg – Ser, then stop. **B1**

(c) The third codon is CGU, so the anticodon is **GCA** (the same bases as the template triplet). **B1**

(d)(i) CTT becomes CTC, which still codes for Glu. **B1** The polypeptide is unchanged, because the genetic code is **degenerate**. **B1**

(ii) CTT becomes ATT, a stop triplet. **B1** Translation stops after Met, so no functional polypeptide is made (only a single amino acid). **B1**

(iii) The new triplets are TAC CTG CAA GGA ... **M1** This is a **frameshift**: every triplet after the deletion changes, **B1** giving Met – Asp – Val – Pro ..., a different primary structure, so the polypeptide will not fold into its normal tertiary structure. **A1**

Example 3

A eukaryotic gene contains three exons of 120, 210 and 165 nucleotides, separated by two introns of 900 and 1500 nucleotides. Assume the exons contain only the sequence from the start codon to the stop codon.

(a) Calculate the length of the primary transcript and of the mRNA. **[2]**

(b) Calculate the number of amino acids in the polypeptide. **[2]**

(c) Calculate the percentage of the primary transcript removed during splicing. **[1]**

(d) Describe how the primary transcript is changed into mRNA. **[2]**

(e) After translation, the first amino acid (methionine) is removed from this polypeptide. Calculate the number of amino acids in the finished protein and the number of mRNA nucleotides that code for amino acids. **[2]**

(a) Primary transcript = $120 + 900 + 210 + 1500 + 165 = 2895$ nucleotides. **B1** mRNA = $120 + 210 + 165 = 495$ nucleotides. **B1**

(b) $495 \div 3 = 165$ codons. **M1** The last codon is a stop codon, which codes for no amino acid, so 164 amino acids. **A1**

(c) $\frac{900 + 1500}{2895} \times 100 = 82.9\%$ **B1**

(d) The introns (non-coding sequences) are removed **B1** and the exons (coding sequences) are joined together. **B1**

(e) $164 - 1 = 163$ amino acids in the finished protein. **B1** The mRNA still codes for all 164 amino acids:
 $164 \times 3 = 492$ nucleotides (495 including the stop codon). **B1**

Example 4

A cell is in the S phase of its cell cycle.

- (a) Describe how a new strand of DNA is made against a template strand. **[4]**
- (b) Explain why one of the two new strands at a replication fork is made in short fragments. **[3]**
- (c) State the role of DNA ligase. **[1]**
- (d) Explain why the replication is described as semi-conservative. **[1]**
- (e) Part of a template strand reads 3' AGCTTA 5'. Write the new strand made against it, labelling its ends. **[1]**
- (a) Any four of:
- Hydrogen bonds between bases break and the strands **separate**. **B1**
 - Free (activated) **DNA nucleotides** pair with the exposed bases by **complementary base pairing**, A with T and C with G. **B1**
 - **Hydrogen bonds** form between the paired bases. **B1**
 - **DNA polymerase** joins adjacent nucleotides by forming **phosphodiester bonds** (condensation). **B1**
 - Nucleotides are added to the **3' end**, so the strand grows **5' → 3'**. **B1**
 - DNA polymerase **proofreads**, removing mismatched nucleotides. **B1**
- (b) The two strands of DNA are **antiparallel**. **B1** DNA polymerase can only add nucleotides in the **5' → 3'** direction, **B1** so on one template it must work **away from** the replication fork, making a short piece each time more DNA unwinds (the lagging strand). **B1**
- (c) It joins the fragments of the lagging strand by forming phosphodiester bonds. **B1**
- (d) Each new DNA molecule contains one strand from the original molecule and one newly made strand. **B1**
- (e) 5' TCGAAT 3': complementary bases, and antiparallel to the template. **B1**

Common mistakes

- **Mixing up introns and exons, and primary transcript and mRNA.** Introns are removed from the **primary transcript**; exons are kept and joined. Call the RNA "mRNA" only once it has been spliced.
- **Giving DNA polymerase or ligase the wrong job.** Neither unzips DNA or forms hydrogen bonds, and phosphodiester bonds join adjacent nucleotides in one strand, never complementary bases. Ligase is not helicase and does not proofread; it only joins the lagging-strand fragments.
- **"S" instead of "S phase", or "G1".** Replication is in the S phase of **interphase**, not in mitosis.

- **Writing T in RNA.** Uracil replaces thymine in every RNA sequence: examiners have seen GTG written for GUG.
- **"Code" for "codon".** A codon is a triplet on mRNA; the code is the whole system.
- **Using "base" when you mean "nucleotide".** Nucleotides are added to a strand; bases pair.
- **Calling a chain-stopping drug a "stop codon".** A nucleotide analogue stops the chain because no phosphodiester bond can form after it; a stop codon is a triplet on mRNA and only acts in translation.
- **Base pairs versus bases.** A percentage of bases in double-stranded DNA applies to $2N$ bases for N base pairs.
- **Reading a template strand the wrong way round.** Check which end is 5' before converting it to mRNA or using a codon wheel.

How to get the marks

- **Use the exact terms:** transcribed/template strand, non-transcribed strand, primary transcript, RNA polymerase, DNA polymerase, DNA ligase, phosphodiester bond, codon, anticodon, complementary. Examiners want the named enzyme ("DNA polymerase", not "polymerase").
- **Describe** (a process) means the steps in the right order, each with its key word: bonds break → pairing → enzyme joins → product. Don't wander into the other process: a translation answer earns nothing for transcription steps.
- **Explain** needs a reason with each point: "the strands are antiparallel **and** DNA polymerase only works 5' → 3', **so** ...".
- **Suggest** means use the syllabus ideas in a new context: name the process affected, the mechanism and the result.
- **Sequences:** write in triplets and use U in RNA. For multiple choice, work the full sequence out yourself before looking at the options.
- **Calculations:** whole numbers of bases and amino acids; subtract one codon for the stop codon when you count amino acids, and check whether the options count the stop codon at all.
- **Both enzymes:** if a question names DNA polymerase and DNA ligase, give a role for each; proofreading is an extra creditable point for DNA polymerase.

Quick check

1. One strand of a short piece of DNA reads 5' GATTACGC 3'. Write the complementary strand, labelling its 5' and 3' ends, and calculate the number of hydrogen bonds holding the two strands together.
2. A circular bacterial chromosome has 4 600 000 base pairs, and 30% of its bases are adenine. (a) How many cytosine bases does it contain? (b) How many phosphodiester bonds? (c) It is copied from 2 replication forks, each adding 800 base pairs per second. How many minutes does replication take, to 3 significant figures?
3. A tRNA molecule has the anticodon CGA. Give the mRNA codon it pairs with and the matching triplet on each DNA strand.

4. In a gene, three adjacent nucleotides that form one whole triplet in an exon are deleted. In a second copy, one nucleotide in an intron is replaced by another. Explain the likely effect of each on the polypeptide.
5. Bacteria with only ^{15}N in their DNA divide every 25 minutes. They are moved to a medium with only ^{14}N . What percentage of their DNA molecules are hybrid, and what percentage light, after 100 minutes?

Answers

1. The complementary strand is 3' CTAATGCG 5' (written 5' → 3': GCGTAATC). There are 4 A–T and 4 C–G pairs, so $4 \times 2 + 4 \times 3 = 20$ hydrogen bonds.
2. (a) The molecule has $2 \times 4\,600\,000 = 9\,200\,000$ bases. Adenine 30%, so thymine 30%, leaving 40% for C and G: cytosine is 20%, $0.2 \times 9\,200\,000 = 1\,840\,000$. (b) A circle has one phosphodiester bond per nucleotide: 9 200 000. (c) $4\,600\,000 \div (2 \times 800) = 2875$ s, and $2875 \div 60 = 47.9$ minutes.
3. The codon is **GCU** (complementary to the anticodon). The template triplet has the same bases as the anticodon with T for U: **CGA**. The non-transcribed triplet matches the codon: **GCT**. (GCU codes for alanine.)
4. Deleting a whole triplet removes **one amino acid** and does not shift the reading frame, so the rest of the polypeptide is normal; it may still fold differently if that amino acid mattered to its shape. The intron substitution has **no effect** on the polypeptide, because introns are removed from the primary transcript before translation.
5. $100 \div 25 = 4$ generations, so $2^4 = 16$ molecules per original molecule, 2 of them hybrid: $\frac{2}{16} = 12.5\%$ hybrid and 87.5% light (none heavy).

Past questions to try

- **9700/22/F/M/24 Q4**: template strand to primary transcript, a codon table, a substitution and a deletion, then replication.
- **9700/23/O/N/24 Q3**: purines and pyrimidines, then extending a new strand and what antiparallel means for replication.
- **9700/22/M/J/23 Q5**: a base table, then a nucleotide analogue that stops DNA polymerase.
- **9700/23/M/J/24 Q6**: three differences between mRNA and DNA, then judging synthetic base pairs.