Biology Units 3 & 4

Chapter 2 — From DNA to proteins

Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Contents

Theory

Every living cell must store a set of instructions for building the proteins it needs, and pass a copy of those instructions on when it divides. The molecules that store and carry these instructions are the **nucleic acids**. There are two kinds: **DNA** (deoxyribonucleic acid) and **RNA** (ribonucleic acid). Because the instructions are held as a sequence that can be read and copied, nucleic acids are described as **information molecules** — the information is the *order* of their building blocks along the chain.

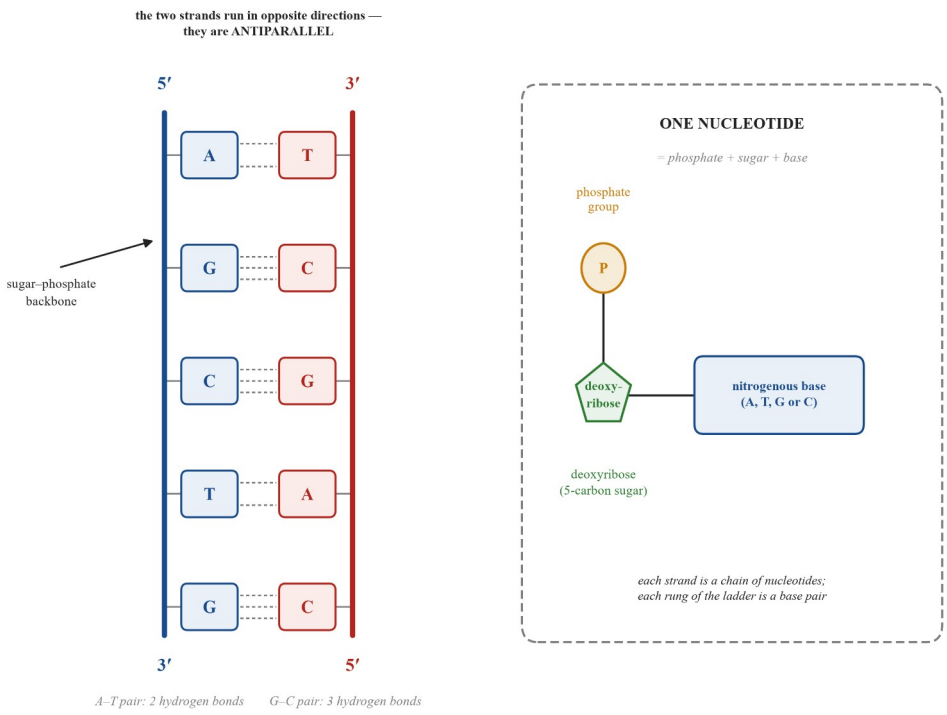
Both DNA and RNA are long chains (polymers) built from repeating units called **nucleotides**. The instruction is written in the order of the nucleotides, in much the same way that the order of letters carries the meaning of a word. A **gene** is a length of DNA whose nucleotide sequence carries the instructions to build one polypeptide (protein chain). How that sequence is actually read to make a protein is the subject of the next subtopic — here we look only at the *structure* of the molecules that hold and carry the message.

The nucleotide. A single nucleotide has three parts joined together:

- a **phosphate group**;
- a **five-carbon (pentose) sugar** — in DNA this sugar is **deoxyribose**;
- a **nitrogenous base**.

In DNA there are four possible bases: **adenine (A)**, **thymine (T)**, **guanine (G)** and **cytosine (C)**. Two nucleotides differ only in which base they carry, so the "letters" of the genetic message are really the four bases.

The structure of DNA — the double helix. A DNA molecule is made of **two** nucleotide strands wound around each other into a **double helix** (a twisted ladder). Within each strand, the phosphate of one nucleotide is joined to the sugar of the next, over and over, forming a continuous **sugar–phosphate backbone**. The bases point inwards from the two backbones, towards each other.



Complementary base pairing. The two strands are held together by **hydrogen bonds** between the bases that face each other. The pairing is highly specific:

- **A always pairs with T**, joined by **2 hydrogen bonds**;

- **G always pairs with C**, joined by **3 hydrogen bonds**.

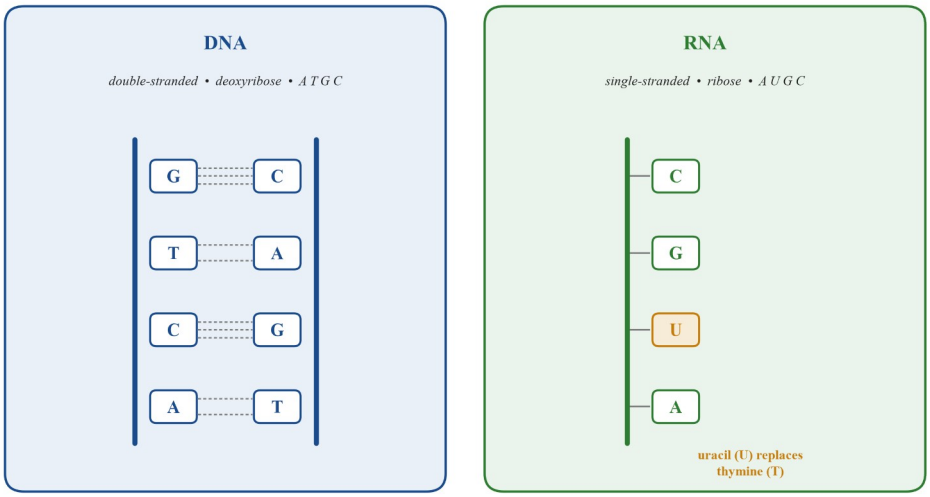
Because each base can only pair with its partner, the two strands are **complementary** — the base sequence of one strand exactly determines the base sequence of the other. A useful consequence is that a G–C-rich region (three hydrogen bonds per pair) is held together a little more strongly than an A–T-rich region (two hydrogen bonds per pair).

Antiparallel strands and the 5′ and 3′ ends. Each strand has a direction, set by the way the sugars are joined: one end is called the **5′ (five-prime) end** and the other the **3′ (three-prime) end**. In the double helix the two strands run in **opposite** directions — where one strand runs 5′→3′, its partner runs 3′→5′. Strands arranged this way are said to be **antiparallel**. Running antiparallel is what allows the complementary bases along the whole length to line up and pair correctly.

Chargaff's rule. Because A pairs only with T and G only with C, in any sample of double-stranded DNA the amount of adenine equals the amount of thymine, and the amount of guanine equals the amount of cytosine. As percentages: %A = %T and %G = %C, and the four bases together make up 100%. This lets you work out all four base percentages of a double-stranded DNA sample from just one of them.

The structure of RNA. RNA is also a nucleic acid built from nucleotides, but it differs from DNA in three important ways:

- RNA is normally **single-stranded** (one chain), not a double helix;
- the sugar in an RNA nucleotide is **ribose**, not deoxyribose;
- RNA uses the base **uracil (U)** in place of thymine, so its four bases are A, U, G and C. Where a base pairs in RNA, U pairs with A.



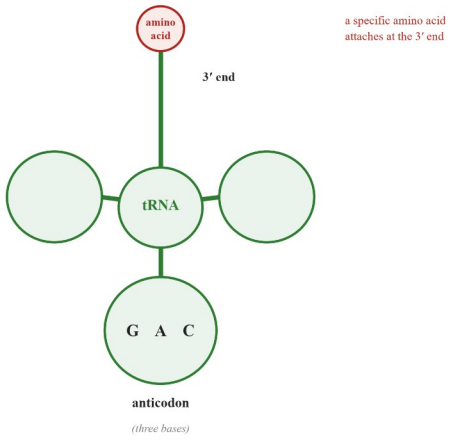
The differences between a DNA nucleotide and an RNA nucleotide can be summarised:

Feature	DNA nucleotide	RNA nucleotide
Sugar	deoxyribose	ribose
Bases used	A, T, G, C	A, U, G, C
Strandedness of the molecule	double-stranded (double helix)	single-stranded
Phosphate group	present	present

The three forms of RNA. Cells make three main forms of RNA, each with its own job in protein synthesis:

- **Messenger RNA (mRNA)** — a **long, unfolded single strand** that is a copy of the coded message of a gene, so its length depends on the length of the gene. Its role is to **carry the coded instructions from the DNA to a ribosome**, where the message will later be read to build a protein. (mRNA is the "messenger" only; how the message is read is covered in the next subtopic.)

- **Ribosomal RNA (rRNA)** — combines with proteins to form the **ribosomes**. rRNA is both a **structural** part of the ribosome and has a **catalytic** role, helping to join amino acids together as the protein is built.
- **Transfer RNA (tRNA)** — a small RNA folded into a **clover-leaf** shape. Each tRNA **carries one specific amino acid** and has a region of three bases called an **anticodon**, which lets it match its amino acid to the correct place on the mRNA.



A single RNA strand folded back on itself into a clover-leaf shape.

In short: DNA is the stable, double-stranded master copy that stores the instructions; the three forms of RNA are single-stranded working molecules that carry the message (mRNA), build the machine that reads it (rRNA) and deliver the amino acids (tRNA).

Worked examples

Example 1 — scaffolded

One strand of a section of DNA reads **5′-A G G C T A-3′**. Write the base sequence of the complementary strand, showing its 5′ and 3′ ends, and calculate the total number of hydrogen bonds holding the two strands together.

Working	Comment
Apply the base-pairing rules to each base: A pairs with T, G with C, C with G, T with A. Pairing 5′-A G G C T A-3′ gives the bases **T C C G A T**.	A–T and G–C are the only allowed pairs; write the partner of each base in turn.
The complementary strand is **antiparallel**, so its 5′ end lies opposite the 3′ end of the given strand. Aligned under the original it reads **3′-T C C G A T-5′**.	The partner strand runs in the opposite direction; label the ends the opposite way around.
Written in its own 5′→3′ direction the complementary strand is **5′-T A G C C T-3′**.	Read the antiparallel partner from its own 5′ end to give the strand on its own.
Count hydrogen bonds: the pairs are A–T (2), G–C (3), G–C (3), C–G (3), T–A (2), A–T (2). Total = 2 + 3 + 3 + 3 + 2 + 2 = **15 hydrogen bonds**.	A–T pairs contribute 2 bonds each and G–C pairs contribute 3 each.

Example 2 — scaffolded

A sample of double-stranded DNA is found to contain **28% adenine**. Determine the percentage of each of the other three bases, and explain the reasoning.

Working	Comment
Adenine pairs only with thymine, so in double-stranded DNA %A = %T. Therefore **thymine = 28%**.	State Chargaff's rule (A pairs with T) before calculating.

Working	Comment
Adenine and thymine together make up 28% + 28% = 56% of the bases. The remaining 100% − 56% = 44% must be guanine and cytosine.	The four bases add to 100%.
Guanine pairs only with cytosine, so %G = %C. The 44% is shared equally: **guanine = 22%** and **cytosine = 22%**.	G–C are equal because each G is paired with a C.
Final: A 28%, T 28%, G 22%, C 22% (check: 28 + 28 + 22 + 22 = 100%).	Always check the four percentages sum to 100.

Example 3 — unscaffolded

Compare a DNA nucleotide with an RNA nucleotide.

A DNA nucleotide and an RNA nucleotide are alike in that **each is made of a phosphate group, a five-carbon sugar and a nitrogenous base**. They differ in two ways. First, the **sugar** is **deoxyribose in a DNA nucleotide** but **ribose in an RNA nucleotide**. Second, one of the **bases** differs: a DNA nucleotide may carry adenine, thymine, guanine or cytosine, whereas an RNA nucleotide carries adenine, **uracil**, guanine or cytosine — that is, **uracil replaces thymine** in RNA. (At the level of the whole molecule, DNA is double-stranded while RNA is single-stranded.) ∴ the two nucleotides share the same three-part plan but differ in their sugar and in one of their four bases.

Example 4 — unscaffolded

In a sample of double-stranded DNA the percentage of adenine always equals the percentage of thymine, yet in a single-stranded RNA molecule the percentage of adenine need not equal the percentage of uracil. Explain why.

In **double-stranded DNA**, every base on one strand is **paired with its complementary base** on the other strand: each adenine is hydrogen-bonded to a thymine. Counting across both strands, every A is matched by a T, so the number of adenines must equal the number of thymines and therefore **%A = %T**. An **RNA molecule is single-stranded**, so its bases are **not each paired** with a complementary base within the same molecule. There is no rule forcing every A to sit opposite a U, so the amount of adenine can differ from the amount of uracil. ∴ the equal-base rule follows from complementary base pairing between two strands, which single-stranded RNA does not have.

Practice questions

Scaffolded

Q1. Answer the following about the building blocks of nucleic acids. (a) Name the two types of nucleic acid found in cells. (b) Name the three components that make up a single nucleotide. (c) Name the four nitrogenous bases found in DNA.

Q2. A DNA molecule has a double-helix structure. (a) State what is meant by the term *double helix*. (b) Name the two components that make up the sugar–phosphate backbone. (c) State which base pairs with which, and give the number of hydrogen bonds in each pair.

Q3. One strand of a section of DNA reads **5′-T A C G G A T-3′**. (a) State the base-pairing rules for DNA. (b) Write the base sequence of the complementary strand, showing its 5′ and 3′ ends. (c) Explain what is meant by describing the two strands as *antiparallel*.

Q4. RNA differs from DNA in several ways. (a) Name the sugar found in an RNA nucleotide. (b) Name the base found in RNA that takes the place of thymine. (c) State how the number of strands in an RNA molecule differs from the number in a DNA molecule.

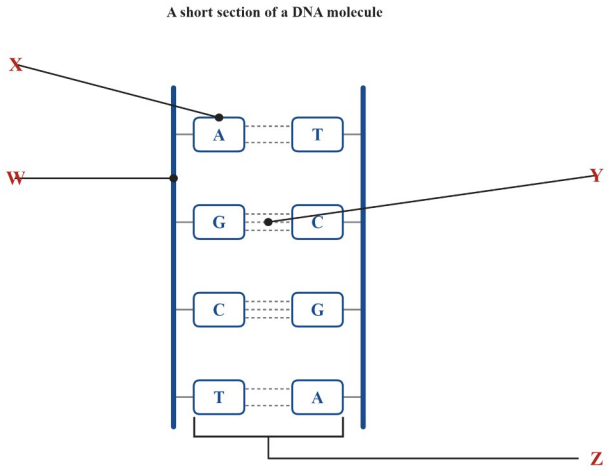
Q5. Cells make three main forms of RNA. (a) Name the three forms of RNA. (b) State the role of mRNA. (c) State the role of tRNA.

Q6. A sample of double-stranded DNA is found to contain **32% cytosine**. (a) State the percentage of guanine in the sample, and explain your answer. (b) Calculate the combined percentage of adenine and thymine. (c) State the percentage of adenine.

Unscaffolded

Q7. Describe how the nucleotides within one strand of a DNA molecule are joined to one another, and state where the bases of that strand lie in relation to the backbone. *(2 marks)*

Q8. *(Stimulus.)* The diagram shows a short section of a DNA molecule with four features labelled **W**, **X**, **Y** and **Z**.



Identify the feature indicated by each of **W**, **X**, **Y** and **Z**. *(4 marks)*

Q9. One strand of a section of DNA reads **5′-C A T T A G C-3′**. A student writes the complementary strand as **5′-G T A A T C G-3′**. Identify the mistake the student has made, write the complementary strand correctly, and calculate the total number of hydrogen bonds between the two strands. *(3 marks)*

Q10. *(Data interpretation.)* The table shows the base composition (as a percentage of all bases) of the nucleic acid extracted from three organisms.

Organism	% A	% T	% G	% C
Species P	30	30	20	20
Species Q	21	?	29	?
Virus R	24	32	18	26

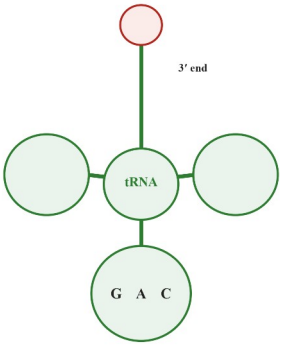
(a) Assuming the nucleic acid of Species Q is double-stranded DNA, determine the two missing values.

(b) The nucleic acid of one organism is unlikely to be double-stranded. Identify this organism and justify your choice using the data. *(4 marks)*

Q11. Explain what is meant by describing the two strands of DNA as *antiparallel*, and explain why complementary base pairing means the base sequence of one strand determines the base sequence of the other. *(3 marks)*

Q12. Complete a comparison of the three forms of RNA by giving, for each, its full name and its role in the cell. *(3 marks)*

Q13. *(Stimulus.)* The diagram shows a molecule of transfer RNA (tRNA).



(a) Name the three-base region on the tRNA that allows it to recognise the correct mRNA codon, and state what type of molecule attaches at the 3′ end.
(b) Using the diagram, state one piece of evidence that tRNA is a nucleic acid and not a protein. *(3 marks)*

Q14. Explain why a region of DNA rich in G–C base pairs is held together more strongly than a region of the same length that is rich in A–T base pairs. *(2 marks)*

Q15. Nucleic acids are described as *information molecules*. Explain what is meant by this, and identify the feature of a DNA molecule that stores the information. *(2 marks)*

Q16. A sample of double-stranded DNA contains a total of **4000 bases**, of which **1200 are guanine**. Calculate the number of adenine, thymine and cytosine bases in the sample, showing your reasoning. *(3 marks)*

Q17. Distinguish between the structure of an mRNA molecule and the structure of a tRNA molecule. *(2 marks)*

Q18. DNA is described as the cell's stable master copy of its instructions, while the three forms of RNA are described as working molecules. Using the structure of each, explain why DNA is suited to storing the instructions permanently and RNA is not. *(3 marks)*

Q19. *(Synthesis.)* A researcher isolates a single gene from a plant cell. The gene is a section of double-stranded DNA that is **900 base pairs** long, and **270** of those base pairs are G–C pairs. (a) State what is meant by the term *gene*, and calculate the total number of nucleotides in this section of DNA. (b) Calculate the total number of hydrogen bonds holding the two strands of this gene together, showing your reasoning. (c) A copy of this gene's message is carried to a ribosome. Name the form of RNA that does this, and state one way in which a nucleotide of that RNA differs from a nucleotide of this gene. *(7 marks)*

Answers

Q1. (a) DNA and RNA. (b) A phosphate group, a (five-carbon/pentose) sugar and a nitrogenous base. (c) Adenine, thymine, guanine and cytosine. **Q2.** (a) Two nucleotide strands wound around each other into a twisted-ladder shape. (b) The sugar (deoxyribose) and the phosphate group. (c) A pairs with T (2 hydrogen bonds); G pairs with C (3 hydrogen bonds). **Q3.** (a) A pairs with T and G pairs with C. (b) 3′-A T G C C T A-5′ (that is, 5′-A T C C G T A-3′). (c) The two strands run in opposite directions — one 5′→3′, the other 3′→5′. **Q4.** (a) Ribose. (b) Uracil (U). (c) An RNA molecule is normally single-stranded (one chain), whereas a DNA molecule is double-stranded (two chains wound into a double helix). **Q5.** (a) mRNA, rRNA and tRNA. (b) Carries the coded message from the DNA to a ribosome. (c) Carries a specific amino acid (and has an anticodon that matches it to the mRNA). **Q6.** (a) 32% (G = C, because guanine pairs only with cytosine). (b) 36%. (c) 18%. **Q7.** The phosphate of one nucleotide is joined to the sugar of the next, repeatedly, forming a continuous sugar–phosphate backbone; the bases are attached to the sugars and point inwards from the backbone, towards the bases of the other strand. **Q8.** W = sugar–phosphate backbone; X = a nitrogenous base; Y = hydrogen bonds; Z = a (complementary) base pair. **Q9.** The bases are paired correctly, but the student has written them in the same direction as the given strand, so the 5′ and 3′ labels are the wrong way around — those bases are 3′-G T A A T C G-5′. Written in its own 5′→3′ direction the complementary strand is 5′-G C T A A T G-3′; 17 hydrogen bonds. **Q10.** (a) T = 21%, C = 29%. (b) Virus R, because %A ≠ %T and %G ≠ %C, so its bases are not complementarily paired — it is single-stranded. **Q11.** Antiparallel = the strands run in opposite directions (5′→3′ and 3′→5′); because each base pairs only with its complement (A–T, G–C), fixing one strand's sequence fixes the other's. **Q12.** mRNA (messenger RNA) — carries the coded message from DNA to the ribosome; rRNA (ribosomal RNA) — a structural and catalytic component of ribosomes; tRNA (transfer RNA) — carries a specific amino acid and has an anticodon. **Q13.** (a) The anticodon is a group of three bases; an amino acid attaches at the 3′ end. (b) The diagram shows nitrogenous bases on the molecule (G, A and C) — or, equally, labels one end as the 3′ end; bases and 5′/3′ ends belong to nucleotides, so the molecule is a nucleic acid, whereas a protein would be a chain of amino acids. **Q14.** Each G–C pair has 3 hydrogen bonds but each A–T pair has only 2, so a G–C-rich region has more hydrogen bonds and is held together more strongly. **Q15.** The information is the sequence (order) of bases along the DNA; different base orders spell different instructions for building proteins. **Q16.** C = 1200; A = 800; T = 800. **Q17.** mRNA is a long single strand that stays extended (its length set by the gene it copies); tRNA is a much shorter single strand folded back on itself into a fixed clover-leaf shape. **Q18.** DNA is double-stranded, with every base hydrogen-bonded to its complementary partner and the bases turned inwards, so the molecule is stable and suited to permanent storage; RNA is a single strand with unpaired, exposed bases, so it is not held in this way and serves as a short-lived working molecule. **Q19.** (a) A gene is a length of DNA whose base sequence carries the instructions to build one polypeptide; 900 × 2 = 1800 nucleotides. (b) 630 A–T pairs; (270 × 3) + (630 × 2) = 810 + 1260 = 2070 hydrogen bonds. (c) mRNA; an RNA nucleotide has ribose instead of deoxyribose (or uracil instead of thymine).

Fully-worked solutions

Q1. (a) The two nucleic acids are **DNA** (deoxyribonucleic acid) and **RNA** (ribonucleic acid). (b) Every nucleotide is made of a **phosphate group**, a **five-carbon (pentose) sugar** and a **nitrogenous base**. (c) The four bases in DNA are **adenine (A)**, **thymine (T)**, **guanine (G)** and **cytosine (C)**.

Q2. (a) A **double helix** is two nucleotide strands wound around each other, like a ladder twisted along its length. (b) The backbone of each strand is made of alternating **sugar (deoxyribose)** and **phosphate** groups joined together. (c) **Adenine pairs with thymine**, held by **2 hydrogen bonds**, and **guanine pairs with cytosine**, held by **3 hydrogen bonds**.

Q3. (a) The base-pairing rules are **A pairs with T** and **G pairs with C**. (b) Pairing each base of 5′-T A C G G A T-3′ gives the partner bases A T G C C T A. Because the strands are antiparallel, the complementary strand aligns as **3′-A T G C C T A-5′**, which written in its own 5′→3′ direction is **5′-A T C C G T A-3′**. (c) **Antiparallel** means the two strands run in **opposite directions** — where one strand runs 5′→3′, its partner runs 3′→5′.

Q4. (a) The sugar in an RNA nucleotide is **ribose** (a DNA nucleotide contains deoxyribose). (b) The base that takes the place of thymine in RNA is **uracil (U)**, so RNA's four bases are A, U, G and C. (c) An RNA molecule is normally **single-stranded** — it is a single chain of nucleotides — whereas a DNA molecule is **double-stranded**, two chains wound around each other into a double helix.

Q5. (a) The three forms are **mRNA (messenger RNA)**, **rRNA (ribosomal RNA)** and **tRNA (transfer RNA)**. (b) **mRNA carries the coded message from the DNA to a ribosome**, where it will later be read to build a protein. (c) A **tRNA carries a specific amino acid** and has an **anticodon** (three bases) that matches its amino acid to the correct place on the mRNA.

Q6. (a) Guanine pairs only with cytosine, so in double-stranded DNA %G = %C; therefore **guanine = 32%**. (b) Cytosine and guanine together make up 32% + 32% = 64%, so adenine and thymine together make up 100% − 64% = **36%**. (c) Adenine equals thymine (%A = %T), so the 36% is shared equally: **adenine = 18%**.

Q7. Within a strand, the **phosphate group of one nucleotide is joined to the sugar of the next nucleotide**, and this join is repeated along the whole chain, so the strand carries a continuous **sugar–phosphate backbone** of alternating sugar and phosphate groups. The **nitrogenous bases are attached to the sugars and point inwards**, away from the backbone and towards the bases of the partner strand, where they can pair. *(1 mark: the phosphate of one nucleotide is joined to the sugar of the next, giving an alternating sugar–phosphate backbone; 1 mark: the bases point inwards from the backbone, towards the other strand.)*

Q8. W points to the **sugar–phosphate backbone** (the outer strand of alternating sugar and phosphate groups). **X** points to a **nitrogenous base**. **Y** points to the **hydrogen bonds** that hold the two bases of a pair together. **Z** points to a **base pair** — two complementary bases (here an A–T pair) joined across the middle of the molecule. *(1 mark each.)*

Q9. Pairing each base of 5′-C A T T A G C-3′ (C→G, A→T, T→A, T→A, A→T, G→C, C→G) does give the partner bases **G T A A T C G**, so the student's base pairing is correct. The mistake is one of **direction**: the two strands are **antiparallel**, so those partner bases lie under the given strand as **3′-G T A A T C G-5′**. By labelling them "5′-G T A A T C G-3′" the student has put the 5′ and 3′ ends the wrong way around; the strand must be **reversed** when it is written out in its own 5′→3′ direction, giving **5′-G C T A A T G-3′**. The pairs are C–G (3), A–T (2), T–A (2), T–A (2), A–T (2), G–C (3), C–G (3), giving a total of 3 + 2 + 2 + 2 + 2 + 3 + 3 = **17 hydrogen bonds**. *(1 mark: identifies that the complement has been written in the wrong direction / the 5′ and 3′ ends are reversed; 1 mark: correct strand 5′-G C T A A T G-3′; 1 mark: 17 hydrogen bonds.)*

Q10. (a) If Species Q is double-stranded DNA then %A = %T and %G = %C. Given A = 21%, thymine must be **21%**; given G = 29%, cytosine must be **29%** (check: 21 + 21 + 29 + 29 = 100%). (b) The organism whose nucleic acid is unlikely to be double-stranded is **Virus R**. In its data **%A (24) does not equal %T (32)** and **%G (18) does not equal %C (26)**, so its bases cannot all be joined in complementary A–T and G–C pairs. This means the nucleic acid is **single-stranded** (there is no complementary partner strand). *(1 mark: T = 21%; 1 mark: C = 29%; 1 mark: Virus R; 1 mark: justified by %A ≠ %T and %G ≠ %C.)*

Q11. Antiparallel means the two strands of the double helix run in **opposite directions** — one strand runs 5′→3′ while its partner runs 3′→5′ — which allows the bases along the strands to line up and pair. Because base pairing is **complementary and specific** (A pairs only with T, and G only with C), each base on one strand fixes the identity of the base opposite it. So once the sequence of one strand is known, the sequence of the other strand is **completely determined**: for example, an A on one strand must have a T opposite it, and a G must have a C. *(1 mark: antiparallel = opposite 5′→3′ directions; 1 mark: pairing is specific A–T, G–C; 1 mark: one strand therefore determines the other.)*

Q12. mRNA (messenger RNA) carries the **coded message copied from the DNA to a ribosome**, where the message will be read. **rRNA (ribosomal RNA)** is a **structural and catalytic component of the ribosome** — it combines with proteins to form the ribosome and helps join amino acids. **tRNA (transfer RNA)** carries a **specific amino acid** and has an **anticodon** that matches its amino acid to the correct place on the mRNA. *(1 mark for each RNA's name-and-role.)*

Q13. (a) The **anticodon** is a group of **three bases** on the tRNA; a **specific amino acid** attaches at the **3′ end** of the molecule. (b) The diagram shows the molecule carrying **nitrogenous bases** — the bases **G**, **A** and **C** are drawn in the lower loop — and it labels one end of the molecule as the **3′ end**. Bases, and the 5′/3′ way of naming the ends, are features of **nucleotides**, the units from which a **nucleic acid** is built. A protein contains no bases and has no 3′ end: it is a chain of **amino acids** joined by peptide bonds. *(1 mark: cites a nucleic-acid feature actually shown on the diagram — the bases G, A and C, or the labelled 3′ end.)*

Q14. Each **G–C pair is held by 3 hydrogen bonds**, whereas each **A–T pair is held by only 2**. A region rich in G–C therefore has **more hydrogen bonds along its length** than an A–T-rich region of the same length, so more energy is

needed to separate the strands and the G–C-rich region is held together **more strongly**. *(1 mark: G–C has 3 bonds vs A–T has 2; 1 mark: more bonds means held more strongly.)*

Q15. Describing nucleic acids as **information molecules** means they **store instructions** — in particular the instructions for building the cell's proteins. The information is stored as the **sequence (order) of the nitrogenous bases** along a DNA strand: just as changing the order of letters changes a word, a different order of bases spells a different instruction, and so specifies a different protein. *(1 mark: they store the instructions for making proteins; 1 mark: the information is the base sequence/order.)*

Q16. Guanine pairs only with cytosine, so **cytosine = guanine = 1200 bases**. Guanine and cytosine together account for 1200 + 1200 = 2400 bases, leaving 4000 − 2400 = 1600 bases for adenine and thymine. Adenine pairs only with thymine, so these are shared equally: **adenine = 800** and **thymine = 800**. *(1 mark: C = 1200 with reason; 1 mark: A + T = 1600; 1 mark: A = 800 and T = 800.)*

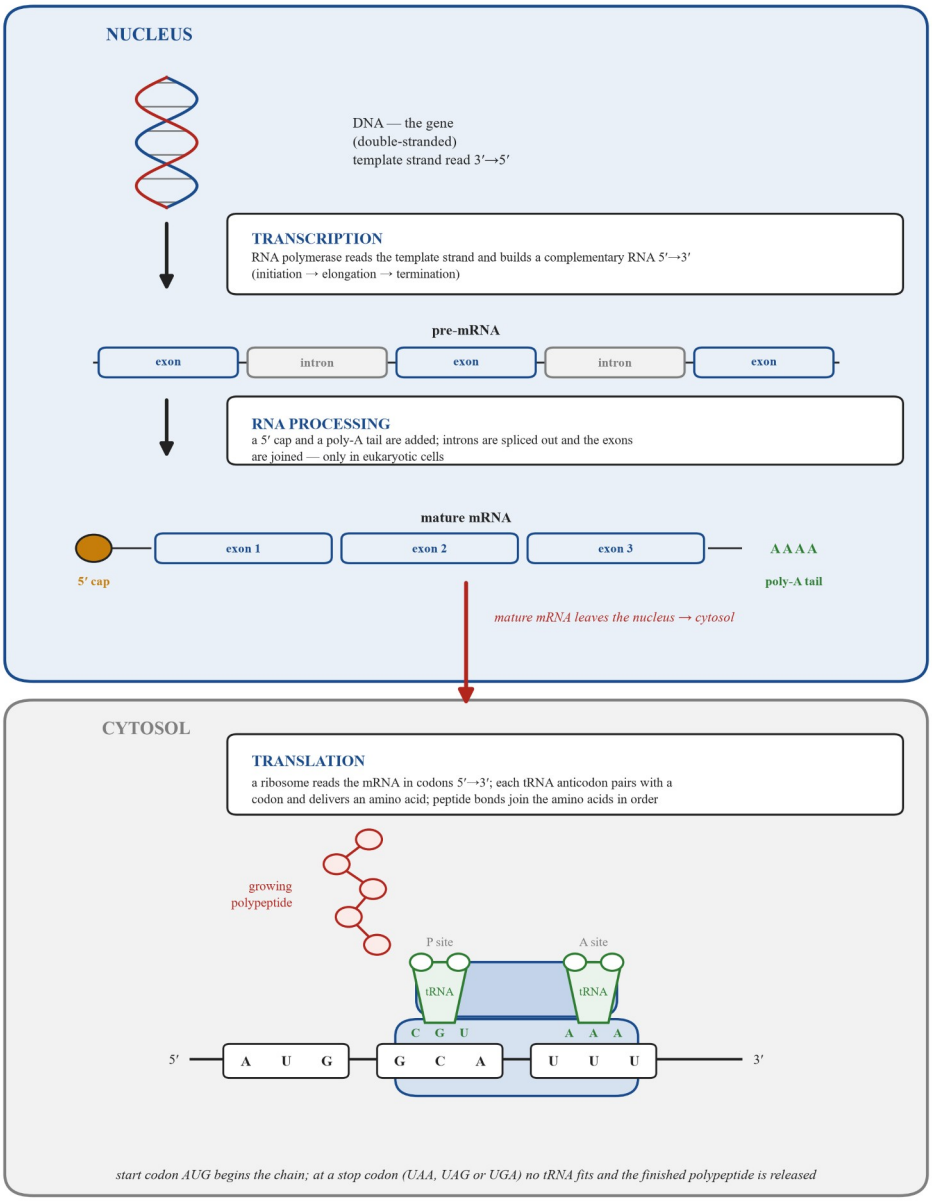
Q17. An **mRNA** molecule is a **long single strand** of RNA nucleotides that stays **extended**: it is a straight copy of the coded message of a gene, so how long it is depends on how long that gene is. A **tRNA** molecule is **much shorter**, and although it too is a single strand it is **folded back on itself into a fixed clover-leaf shape** (the fold is what creates its anticodon loop and its amino-acid attachment site). So both are single strands of RNA nucleotides, but they differ in length and in whether the strand remains extended (mRNA) or folds into a set three-dimensional shape (tRNA). *(1 mark: mRNA — a long, extended/unfolded single strand copying a gene; 1 mark: tRNA — a much shorter single strand folded into a clover-leaf shape. Congruent, both sides present.)*

Q18. DNA is double-stranded. Every base along one strand is **hydrogen-bonded to its complementary base** on the other strand (A–T, G–C), and the two strands wind together as a double helix, so the bases that carry the information are turned **inwards**, held between the two sugar–phosphate backbones. This makes the molecule **stable**, which is what a permanent store of instructions needs. **RNA is single-stranded**: its bases are **not paired** with a partner strand and are **exposed**, so the molecule is not held together in the same way and does not last in the cell. ∴ the double-stranded structure of DNA suits it to storing the instructions permanently, while the single strands of mRNA, rRNA and tRNA suit them to being short-lived working molecules that carry the message, build the ribosome and deliver the amino acids. *(1 mark: DNA double-stranded, bases paired and hydrogen-bonded between the strands and turned inwards, so it is stable; 1 mark: RNA single-stranded, bases unpaired and exposed, so it is not held together in the same way; 1 mark: links each structure to its function — permanent store vs short-lived working molecule.)*

Q19. (a) A **gene** is a **length of DNA whose sequence of bases carries the instructions to build one polypeptide (protein chain)**. Each base pair is made of **two nucleotides**, one in each of the two strands, so a section 900 base pairs long contains 900 × 2 = **1800 nucleotides**. (b) Of the 900 base pairs, **270 are G–C pairs**, so the remaining 900 − 270 = **630 are A–T pairs**. Each **G–C pair is held by 3 hydrogen bonds**, giving 270 × 3 = **810** bonds, and each **A–T pair by 2 hydrogen bonds**, giving 630 × 2 = **1260** bonds. The total is 810 + 1260 = **2070 hydrogen bonds**. (c) The form of RNA that carries the coded message of a gene to a ribosome is **mRNA (messenger RNA)**. A nucleotide of that mRNA differs from a nucleotide of this gene in containing the sugar **ribose** in place of deoxyribose (or in carrying the base **uracil** in place of thymine). ∴ the structure of the gene both stores the instruction and, once copied into mRNA, allows it to be carried to the ribosome. *(a: 2 marks — the meaning of gene; 1800 nucleotides. b: 3 marks — 630 A–T pairs; 810 and 1260 hydrogen bonds; total 2070. c: 2 marks — mRNA named; one nucleotide difference.)*

Theory

A gene is a length of DNA that carries the instructions for building a polypeptide (a chain of amino acids). **Gene expression** is the two-stage process by which that instruction is used to make the polypeptide: **transcription** copies the gene into a messenger RNA (mRNA) molecule, and **translation** reads the mRNA to assemble the amino acids in the correct order. This subtopic follows the information from DNA through to a finished polypeptide.



The genetic code. The order of bases in a gene is a coded message. The code is read in groups of three bases. In mRNA a group of three bases is called a **codon**, and each codon specifies one amino acid (or a signal to start or stop). Because the code is read three bases at a time, it is called a **triplet code**. There are four RNA bases (A, U, G, C), so there are $4 \times 4 \times 4 = 64$ possible codons. These 64 codons must specify only about 20 different amino acids, plus start and stop signals.

The genetic code has three features you must be able to describe:

- **It is a triplet code.** Three bases (one codon) code for one amino acid. The codon **AUG** is the **start codon** (it also codes for the amino acid methionine) and sets the reading frame; three codons (**UAA, UAG, UGA**) are **stop codons** that signal the end of the polypeptide and do not code for any amino acid. Because the bases are counted off in non-overlapping threes from the start codon, the **reading frame** decides how every later base is

grouped. Replacing one base alters only the codon it sits in, but inserting or deleting one base regroups **every codon after it** (a **frameshift**), so most of the amino acids from that point on are changed.

- **It is degenerate (redundant).** Because there are 64 codons but only about 20 amino acids, most amino acids are specified by **more than one codon**. For example, GGU, GGC, GGA and GGG all code for glycine. "Degenerate" means many codons can specify the same amino acid — it does **not** mean the code is ambiguous: each individual codon still codes for only **one** amino acid. Degeneracy also means that a base change which swaps one codon for another codon of the **same** amino acid leaves the polypeptide unaltered; such a change is described as **silent**.
- **It is universal (nearly).** The same codons specify the same amino acids in almost every living organism, from bacteria to plants to humans. This shared code is evidence of common ancestry, and it is what makes genetic engineering possible: a bacterium supplied with an **intron-free copy** of a human coding sequence reads those codons exactly as a human cell would and builds the same protein. The copy must be intron-free because a bacterium cannot splice (see RNA processing below) — a universal code guarantees the codons are read the same way, not that a whole eukaryotic gene will work in a bacterium.

The mRNA codon table. The table below shows which amino acid each codon specifies. Notice the degeneracy: several amino acids have four or six codons. Codons are always read in the **5′→3′** direction.

Amino acid	mRNA codons (read 5′→3′)
Alanine (Ala)	GCU, GCC, GCA, GCG
Arginine (Arg)	CGU, CGC, CGA, CGG, AGA, AGG
Asparagine (Asn)	AAU, AAC
Aspartate (Asp)	GAU, GAC
Cysteine (Cys)	UGU, UGC
Glutamate (Glu)	GAA, GAG
Glutamine (Gln)	CAA, CAG
Glycine (Gly)	GGU, GGC, GGA, GGG
Histidine (His)	CAU, CAC
Isoleucine (Ile)	AUU, AUC, AUA
Leucine (Leu)	UUA, UUG, CUU, CUC, CUA, CUG
Lysine (Lys)	AAA, AAG
Methionine (Met) — START	AUG
Phenylalanine (Phe)	UUU, UUC
Proline (Pro)	CCU, CCC, CCA, CCG
Serine (Ser)	UCU, UCC, UCA, UCG, AGU, AGC
Threonine (Thr)	ACU, ACC, ACA, ACG
Tryptophan (Trp)	UGG
Tyrosine (Tyr)	UAU, UAC
Valine (Val)	GUU, GUC, GUA, GUG
STOP (no amino acid)	UAA, UAG, UGA

Transcription — copying the gene into mRNA. Transcription makes an RNA copy of one gene. In a eukaryotic cell it takes place in the **nucleus**, where the DNA is. The steps are:

- **Initiation.** The enzyme **RNA polymerase** binds to a control region (the promoter) at the start of the gene and unwinds the double helix, exposing the bases.
- **Elongation.** Only **one** strand of the DNA, the **template strand**, is read. RNA polymerase moves along the template strand in the **3′→5′** direction and builds a complementary RNA strand in the **5′→3′** direction. It adds RNA nucleotides using complementary base pairing, but in RNA the base **uracil (U)** pairs with adenine instead of thymine (so a DNA A on the template gives a U in the RNA).

- **Termination.** When RNA polymerase reaches a termination signal at the end of the gene, it stops and the new RNA strand is released.

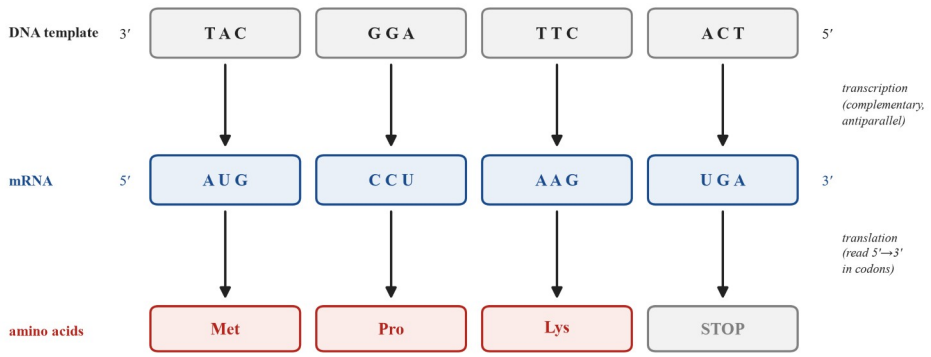
The RNA made directly from the gene is called **pre-mRNA** (or primary transcript). It is complementary and antiparallel to the template strand — and therefore has the same base sequence as the non-template (coding) strand, with U in place of T.

RNA processing — making a mature mRNA (eukaryotes only). Before it can be used, the pre-mRNA of a eukaryote is modified inside the nucleus in three ways:

- A **5′ cap** (a modified guanine nucleotide) is added to the 5′ end. It helps the mRNA bind to the ribosome and protects the molecule.
- A **poly-A tail** (a string of many adenine nucleotides) is added to the 3′ end. It protects the mRNA from being broken down and helps it leave the nucleus.
- **Splicing** removes the non-coding sections. A eukaryotic gene contains coding sections called **exons** and non-coding sections called **introns**. During splicing the **introns are cut out** and the **exons are joined together** in order. Only the exons remain in the final message.

The result is a shorter **mature mRNA** that then leaves the nucleus through a nuclear pore and enters the cytosol. **Prokaryotic cells** (such as bacteria) have no nucleus and their genes generally lack introns, so they do **not** carry out this RNA processing; their transcription and translation happen together in the cytosol.

Translation — reading the mRNA to build a polypeptide. Translation takes place in the **cytosol**, at a **ribosome** (ribosomes may be free in the cytosol or attached to the rough endoplasmic reticulum). The players are the mature mRNA (the message), **transfer RNA (tRNA)** molecules (each carrying a specific amino acid and displaying a three-base **anticodon**), and the ribosome (which has two subunits and three tRNA binding sites: the **A site**, **P site** and **E site**).



the codon AUG starts the chain; UGA is a stop codon — no amino acid is added

The steps are:

- **Initiation.** The small ribosomal subunit binds the mRNA near the 5′ cap and moves to the **start codon, AUG**. A tRNA carrying methionine, with the anticodon that pairs with AUG, binds; the large subunit then joins to complete the ribosome. AUG fixes the **reading frame** — the way the following bases are grouped into codons.
- **Elongation.** The ribosome reads the mRNA **one codon at a time in the 5′→3′ direction**. For each codon, a tRNA whose **anticodon is complementary and antiparallel** to that codon enters the **A site**, delivering the matching amino acid. A **peptide bond** forms between the new amino acid and the growing chain held in the **P site**. The ribosome then moves along by one codon; the empty tRNA leaves from the **E site** to be recharged with another amino acid. In this way amino acids are joined in the exact order dictated by the codons.
- **Termination.** When the ribosome reaches a **stop codon (UAA, UAG or UGA)**, no tRNA has a matching anticodon. Translation stops, the completed **polypeptide is released**, and the ribosome separates from the mRNA.

A codon and the anticodon that reads it pair by complementary, antiparallel base pairing. For example the codon **5′-AUG-3′** is read by a tRNA with the anticodon **3′-UAC-5′**.

Where the two stages happen. In a eukaryotic cell, **transcription and RNA processing occur in the nucleus** and **translation occurs in the cytosol at a ribosome**. Separating the two stages is why the mRNA must be finished (capped, tailed and spliced) before it leaves the nucleus to be translated.

Worked examples

Example 1 — scaffolded

The template strand of a short gene reads **3′-TAC-CCG-AAA-ATT-5′**. Write the mRNA transcribed from it (showing its direction), then use the codon table to give the sequence of amino acids in the polypeptide.

Working	Comment
RNA polymerase reads the template **3′→5′** and builds mRNA **5′→3′** by complementary base pairing, using **U instead of T**. Pair each template base: T→A, A→U, C→G, G→C.	State the direction and the base-pairing rule first; a DNA A on the template gives a U in the mRNA.
Template 3′-T A C-C C G-A A A-A T T-5′ gives mRNA **5′-AUG-GGC-UUU-UAA-3′**.	Keep the codons grouped so they are easy to read; check the direction is labelled 5′→3′.
Read the codons 5′→3′ using the table: AUG = Met (start), GGC = Gly, UUU = Phe, UAA = stop.	AUG is the start codon; the stop codon adds no amino acid.
Polypeptide: **Met – Gly – Phe**.	Do not include the stop codon in the amino-acid sequence.

Example 2 — scaffolded

An mRNA has the sequence **5′-AUG-GCA-UUU-GGU-UAA-3′**. Using the codon table, give the polypeptide it codes for, and state the anticodon of the tRNA that reads the first codon.

Working	Comment
Read the mRNA in codons, 5′→3′, starting at AUG.	Translation always begins at the start codon AUG.
AUG = Met, GCA = Ala, UUU = Phe, GGU = Gly, UAA = stop.	Look each codon up in the table; UAA is a stop codon.
Polypeptide: **Met – Ala – Phe – Gly**.	The stop codon ends the chain but is not translated into an amino acid.
The first codon is 5′-AUG-3′; the anticodon pairs with it in an antiparallel, complementary way: **3′-UAC-5′**.	Write the anticodon 3′→5′ so it lines up base-for-base with the codon.

Example 3 — unscaffolded

A student says: “In a eukaryote, the mRNA made by RNA polymerase is used straight away by a ribosome.” Explain why this statement is incorrect, referring to what happens to the pre-mRNA before translation.

The statement is incorrect because the RNA made directly from the gene is **pre-mRNA**, which must first be processed in the **nucleus**. Three modifications occur: a **5′ cap** is added to the 5′ end, a **poly-A tail** is added to the 3′ end, and **splicing** removes the **introns** and joins the **exons** together. Only after this processing does the shorter **mature mRNA** leave the nucleus and enter the cytosol, where a ribosome translates it. ∴ the pre-mRNA cannot be used straight away; it must be capped, tailed and spliced into mature mRNA first.

Example 4 — unscaffolded

Part of an mRNA reads **5′-AUG-ACU-CAU-GGA-UAA-3′**. A single base substitution changes the third codon from CAU to **CAC**. Using the codon table, determine the effect of this change on the polypeptide, and explain what it shows about the genetic code.

The original codons give AUG = Met, ACU = Thr, CAU = His, GGA = Gly, UAA = stop, so the polypeptide is **Met – Thr – His – Gly**. After the substitution the third codon is **CAC**, which also codes for **histidine (His)**. The amino acid sequence is therefore **unchanged** — this is a **silent mutation**. This happens because the genetic code is **degenerate**: CAU and CAC are two different codons that specify the same amino acid, so changing the third base makes no difference to the protein. (Had the substitution instead produced UAU, the amino acid would have changed to tyrosine, altering the protein.) ∴ because the code is degenerate, this substitution is silent and the polypeptide is unaffected.

Practice questions

Scaffolded

Q1. Define each of the following terms: (a) codon; (b) degenerate (as applied to the genetic code); (c) universal (as applied to the genetic code).

Q2. The template strand of a gene reads **3′-TAC-AAG-GGT-ACT-5′**. (a) Name the enzyme that carries out transcription. (b) Write the mRNA sequence produced, showing its 5′ and 3′ ends. (c) Using the codon table, give the amino-acid sequence of the polypeptide.

Q3. A eukaryotic pre-mRNA is converted to a mature mRNA before translation. (a) State where in the cell this processing occurs. (b) Name the two structures added, one to each end of the molecule. (c) Explain what happens to the introns and exons during splicing.

Q4. Translation occurs at a ribosome. (a) Name the codon at which translation begins. (b) Describe the role of a tRNA molecule in translation. (c) State what happens when the ribosome reaches a stop codon.

Q5. An mRNA has the sequence **5′-AUG-UUA-CGU-GGA-UAA-3′**. (a) Identify the start codon and state the amino acid it codes for. (b) Using the codon table, give the full amino-acid sequence of the polypeptide. (c) State the anticodon of the tRNA that reads the codon UUA.

Q6. Use the mRNA codon table to answer the following. (a) State the amino acid specified by the codon 5′-CAG-3′. (b) State how many different codons specify leucine. (c) Describe the pattern in the codons that specify the same amino acid, using one amino acid from the table as your example.

Unscaffolded

Q7. Distinguish between the terms *universal* and *degenerate* as used to describe the genetic code. *(2 marks)*

Q8. Describe the process of transcription, including the role of RNA polymerase, the template strand and the direction of synthesis. *(3 marks)*

Q9. A tRNA molecule has the anticodon **3′-GAC-5′**. State the mRNA codon it pairs with, name the amino acid that codon specifies (use the codon table), and explain why the anticodon must be written in the opposite direction to the codon. *(3 marks)*

Q10. Describe the three modifications a eukaryotic pre-mRNA undergoes during RNA processing, and state the purpose of each. *(3 marks)*

Q11. Explain how a ribosome builds a polypeptide during translation, from the start codon through to the release of the finished polypeptide. *(4 marks)*

Q12. State the location of transcription and of translation in a eukaryotic cell, and explain one reason the mRNA must be modified before it leaves the nucleus. *(3 marks)*

Q13. Explain **two** differences in gene expression between a bacterium (a prokaryote) and a eukaryotic cell, with reference to RNA processing and the location of the processes. *(3 marks)*

Q14. *(Stimulus.)* A section of the template strand of a gene from the soil bacterium-derived study organism *Terravella nova* reads:

3′-TAC-ACC-GGA-CTT-ATC-5′

Using the codon table: (a) write the mRNA transcribed from this template, showing its 5′ and 3′ ends; (b) give the amino-acid sequence of the polypeptide produced. *(4 marks)*

Q15. Part of an mRNA reads **5′-…-ACU-UCA-GGU-…-3′**. A single base substitution changes the middle codon from **UCA** to **UCG**. Using the codon table, state the effect of this change on the amino-acid sequence and name the type of mutation this is. *(3 marks)*

Q16. *(Stimulus.)* An mRNA reads **5′-AUG-GCA-CAU-UGG-UAA-3′**, coding for Met–Ala–His–Trp. A mutation **deletes the fourth nucleotide** (the G at the start of the second codon). (a) Write the new sequence of codons that would be read from AUG, and use the table to give the new amino-acid sequence. (b) Explain why a single deletion like this usually has a much greater effect on the polypeptide than a single base substitution. *(4 marks)*

Q17. During elongation each tRNA passes through the ribosome and eventually leaves it without an amino acid attached. Describe how a tRNA moves through the three binding sites, and state what happens to it after it leaves the ribosome. *(2 marks)*

Q18. Describe the roles of the start codon and the stop codons in translation. *(3 marks)*

Q19. *(Data interpretation.)* The table shows the length of a eukaryotic gene's transcript at two stages.

Stage	Length (nucleotides)
pre-mRNA (introns + exons)	3600
combined exons after splicing	1200

(a) Calculate the total length of the introns removed during splicing.

(b) Explain why the spliced (coding) sequence is shorter than the pre-mRNA.

(c) Assuming the 1200 exon nucleotides are all read as codons, calculate the number of codons and state the maximum number of amino acids in the polypeptide. *(4 marks)*

Q20. *(Synthesis.)* A researcher sequences a short polypeptide made in a human liver cell and finds the amino-acid sequence **Met – Cys – Val**. Translation of its mRNA ended when the ribosome reached the stop codon **5′-UAG-3′**. She now wants to work back from the polypeptide to the gene. (a) Using the codon table, write **one** mRNA sequence, shown 5′→3′ and ending in that stop codon, that could code for this polypeptide. (b) Write the DNA template strand from which your mRNA would be transcribed, showing its 5′ and 3′ ends, and state how the base sequence of the non-template (coding) strand compares with the mRNA. (c) Explain why part (a) has several correct answers, while working the other way — from a known mRNA to the polypeptide — always gives one answer only. *(7 marks)*

Answers

Q1. (a) A group of three mRNA bases that codes for one amino acid (or a start/stop signal). (b) More than one codon can code for the same amino acid. (c) The same codons specify the same amino acids in (almost) all organisms. **Q2.** (a) RNA polymerase. (b) 5′-AUG-UUC-CCA-UGA-3′. (c) Met – Phe – Pro. **Q3.** (a) In the nucleus. (b) A 5′ cap (5′ end) and a poly-A tail (3′ end). (c) Introns are removed and the exons are joined together in order. **Q4.** (a) AUG (the start codon). (b) A tRNA carries a specific amino acid and its anticodon pairs with the matching mRNA codon, delivering that amino acid to the ribosome. (c) No tRNA matches the stop codon; translation ends and the polypeptide is released. **Q5.** (a) AUG = methionine (Met). (b) Met – Leu – Arg – Gly. (c) 3′-AAU-5′. **Q6.** (a) Glutamine (Gln). (b) Six. (c) They usually share their first two bases and differ only in the third (e.g. the four proline codons CCU, CCC, CCA, CCG). **Q7.** Universal = the same codons code for the same amino acids in nearly all organisms; degenerate = most amino acids are coded for by more than one codon. **Q8.** RNA polymerase binds the promoter, unwinds the DNA, reads the template strand 3′→5′ and builds a complementary mRNA 5′→3′ (with U instead of T) until it reaches a termination signal. **Q9.** Codon = 5′-CUG-3′; amino acid = leucine (Leu); codon and anticodon pair antiparallel, so the anticodon's 3′ end lies opposite the codon's 5′ end and the bases line up base-for-base. **Q10.** 5′ cap (helps ribosome binding / protection); poly-A tail (protection / export from nucleus); splicing removes introns and joins exons (leaves only the coding message). **Q11.** A tRNA carrying Met binds the start codon AUG; each following codon is read 5′→3′; a matching tRNA delivers an amino acid to the A site; peptide bonds join the amino acids; at a stop codon the polypeptide is released. **Q12.** Transcription = nucleus; translation = cytosol/ribosome; the mRNA must be capped, tailed and spliced (processed) before it leaves the nucleus, because processing only happens in the nucleus. **Q13.** Prokaryotes carry out no RNA processing (no introns to splice, no cap/tail) whereas eukaryotes do; in prokaryotes transcription and translation both occur in the cytosol (and can be coupled), whereas in eukaryotes transcription is in the nucleus and translation in the cytosol. **Q14.** (a) 5′-AUG-UGG-CCU-GAA-UAG-3′. (b) Met – Trp – Pro – Glu. **Q15.** No change to the amino acid (UCA and UCG both code for serine); it is a silent mutation. **Q16.** (a) AUG-CAC-AUU-GGU-… = Met – His – Ile – Gly – … (b) A deletion causes a frameshift, changing every codon after it, whereas a substitution changes at most one codon. **Q17.** It enters the A site with its amino acid, shifts to the P site as the ribosome moves along, then leaves from the E site with no amino acid attached; in the cytosol it is recharged with another molecule of its specific amino acid and used again. **Q18.** Start codon AUG begins translation, codes for methionine and sets the reading frame; a stop codon (UAA, UAG, UGA) ends translation and causes the polypeptide to be released; stop codons code for no amino acid. **Q19.** (a) 2400 nucleotides. (b) The introns were removed during splicing, leaving only the exons. (c) 1200 ÷ 3 = 400 codons; up to 399 amino acids (one codon is a stop codon). **Q20.** (a) e.g. 5′-AUG-UGU-GUU-UAG-3′. (b) Template 3′-TAC-ACA-CAA-ATC-5′; the coding strand has the same base sequence as the mRNA with T in place of U (5′-ATG-TGT-GTT-TAG-3′). (c) The code is degenerate, so Cys and Val each have more than one codon (1 × 2 × 4 = 8 possible mRNAs); but each codon specifies only one amino acid, so one mRNA gives one polypeptide.

Fully-worked solutions

Q1. (a) A **codon** is a group of **three consecutive bases in mRNA** that codes for one amino acid (or acts as a start or stop signal). (b) **Degenerate** means that the code is redundant: **most amino acids are specified by more than one codon** (for example, four different codons all code for glycine). It does not mean the code is ambiguous — each codon still specifies only one amino acid. (c) **Universal** means that **the same codons specify the same amino acids in almost all living organisms**, from bacteria to humans.

Q2. (a) The enzyme is **RNA polymerase**. (b) RNA polymerase reads the template 3′→5′ and builds a complementary mRNA 5′→3′ using U in place of T. Pairing each base of 3′-TAC-AAG-GGT-ACT-5′ gives **5′-AUG-UUC-CCA-UGA-3′**. (c) Reading the codons 5′→3′: AUG = Met, UUC = Phe, CCA = Pro, UGA = stop. The polypeptide is **Met – Phe – Pro** (the stop codon adds no amino acid).

Q3. (a) Processing occurs in the **nucleus**. (b) A **5′ cap** is added to the 5′ end and a **poly-A tail** is added to the 3′ end. (c) During **splicing**, the **introns** (non-coding sections) are **cut out** of the pre-mRNA and the **exons** (coding sections) are **joined together in order**, so that the mature mRNA contains only the coding message.

Q4. (a) Translation begins at the **start codon, AUG**. (b) Each **tRNA** carries one **specific amino acid** and displays a three-base **anticodon**. Its anticodon pairs, in a complementary and antiparallel way, with the matching **codon** on the mRNA, thereby **delivering the correct amino acid** to the ribosome in the order set by the mRNA. (c) When the

ribosome reaches a **stop codon**, no tRNA has a matching anticodon, so no amino acid is added; translation **stops** and the completed **polypeptide is released** from the ribosome.

Q5. (a) The start codon is **AUG**, which codes for **methionine (Met)**. (b) Reading 5′→3′: AUG = Met, UUA = Leu, CGU = Arg, GGA = Gly, UAA = stop. The polypeptide is **Met – Leu – Arg – Gly**. (c) The codon is 5′-UUA-3′, so the anticodon pairs with it antiparallel and complementary: **3′-AAU-5′**.

Q6. (a) Reading the table, the codon **5′-CAG-3′** specifies **glutamine (Gln)**. (b) **Leucine** has **six** codons (UUA, UUG, CUU, CUC, CUA, CUG). (c) In most cases the codons for one amino acid **share their first two bases and differ only in the third**: all four **proline** codons begin with CC (CCU, CCC, CCA, CCG), and all four valine codons begin with GU. The three amino acids with six codons are the exceptions — leucine, serine and arginine each have **two separate groups** of codons rather than one. This third-base pattern is why a change to the last base of a codon so often leaves the amino acid unchanged.

Q7. *Universal* and *degenerate* describe different features. **Universal** means the **same codons code for the same amino acids in nearly all organisms**, which is why a coding sequence from one species is translated into the same amino acids in another. **Degenerate** means **most amino acids are coded for by more than one codon** (there are 64 codons but only about 20 amino acids). One feature is about the code being shared across species; the other is about there being spare codons for each amino acid.

Q8. In **transcription**, the enzyme **RNA polymerase** binds to the promoter at the start of the gene and unwinds the DNA. It reads only the **template strand**, moving along it in the **3′→5′** direction, and builds a **complementary mRNA strand in the 5′→3′ direction** by adding RNA nucleotides (using **U instead of T** opposite adenine). When RNA polymerase reaches a termination signal, it stops and the pre-mRNA is released. *(1 mark: RNA polymerase reads the template strand; 1 mark: template read 3′→5′ / mRNA built 5′→3′; 1 mark: complementary base pairing with U replacing T.)*

Q9. An anticodon of **3′-GAC-5′** pairs, complementary and antiparallel, with the mRNA codon **5′-CUG-3′**. From the codon table, CUG codes for **leucine (Leu)**. The anticodon is written in the **opposite direction** to the codon because the two pair **antiparallel**: the anticodon's **3′ end lies opposite the codon's 5′ end**. Setting the codon 5′-CUG-3′ above the anticodon 3′-GAC-5′ puts each base directly opposite its complementary partner (C–G, U–A, G–C). Writing the same anticodon 5′→3′ would reverse it to 5′-CAG-3′, and the bases would no longer line up with the codon.

Q10. During RNA processing a eukaryotic pre-mRNA is modified in three ways. **(1)** A **5′ cap** is added to the 5′ end, which helps the mRNA bind to the ribosome and protects it from breakdown. **(2)** A **poly-A tail** (many adenine nucleotides) is added to the 3′ end, which protects the mRNA and helps it leave the nucleus. **(3) Splicing** removes the **introns** and joins the **exons** together, so the mature mRNA carries only the coding message.

Q11. Translation begins when the ribosome binds the mRNA and reaches the **start codon AUG**; a tRNA carrying **methionine** pairs with it. The ribosome then reads the mRNA **one codon at a time in the 5′→3′ direction**. For each codon, a tRNA whose **anticodon is complementary** to that codon enters the ribosome and delivers its **amino acid**; a **peptide bond** forms, joining that amino acid to the growing chain. The ribosome moves along codon by codon, and the empty tRNAs leave to collect more amino acids. When a **stop codon (UAA, UAG or UGA)** is reached, no tRNA matches it, so translation ends and the **completed polypeptide is released**. *(1 mark each: start at AUG; codons read 5′→3′ with matching tRNA anticodons; peptide bonds join amino acids in order; stop codon ends translation and releases the polypeptide.)*

Q12. Transcription (and RNA processing) occur in the **nucleus**; **translation** occurs in the **cytosol**, at a **ribosome**. The mRNA must be **capped, tailed and spliced (processed) before it leaves the nucleus** because RNA processing only takes place inside the nucleus — once the mRNA passes through a nuclear pore into the cytosol it can no longer be processed, so it must already be a mature mRNA ready for translation.

Q13. First difference — RNA processing: a **eukaryote** processes its pre-mRNA (adds a 5′ cap and poly-A tail and splices out introns), whereas a **prokaryote** does **not** — its genes generally have no introns, so there is no splicing and no cap or tail added. **Second difference — location and timing:** in a **eukaryote**, transcription happens in the **nucleus** and translation later in the **cytosol**, so the two are separated; in a **prokaryote** there is **no nucleus**, so both transcription and translation occur in the **cytosol** and translation can begin on an mRNA while it is still being transcribed.

Q14. (a) RNA polymerase reads the template 3′→5′ and builds mRNA 5′→3′ with U replacing T. Pairing each base of 3′-TAC-ACC-GGA-CTT-ATC-5′ gives **5′-AUG-UGG-CCU-GAA-UAG-3′**. (b) Reading the codons 5′→3′ from the table: AUG = Met, UGG = Trp, CCU = Pro, GAA = Glu, UAG = stop. The polypeptide is **Met – Trp – Pro – Glu** (the stop codon is not translated). *(2 marks: correct mRNA with directions; 2 marks: correct amino-acid sequence.)*

Q15. The middle codon changes from **UCA to UCG**. From the codon table, **both UCA and UCG code for serine (Ser)**, so the amino acid at that position is **unchanged** and the whole polypeptide sequence is the same. This is a **silent mutation**, made possible because the genetic code is **degenerate** (serine has several codons that differ only in the third base).

Q16. (a) Deleting the fourth nucleotide (the first G of GCA) shifts all the following bases along by one, so the bases are regrouped into new codons. Reading from AUG: **AUG-CAC-AUU-GGU-…**, which gives **Met – His – Ile – Gly – …** — completely different from the original Met–Ala–His–Trp, and the original stop codon is lost. (b) A single deletion causes a **frameshift**: because bases are read in threes from a fixed starting point, removing one base **changes the grouping of every codon after the deletion**, so most of the amino acids downstream are wrong (and a stop codon may be gained or lost). A **substitution**, by contrast, changes only the **one codon** it falls in, so at most one amino acid is affected — and sometimes none, if the code's degeneracy makes it silent. *(1 mark: new codons; 1 mark: new amino-acid sequence; 2 marks: frameshift alters all downstream codons versus one codon for a substitution.)*

Q17. A tRNA first enters the **A site**, where its **anticodon** pairs with the codon being read and the **amino acid** it carries is joined to the growing chain by a **peptide bond**. As the ribosome moves along by one codon the tRNA shifts to the **P site**, which holds the growing polypeptide, and then to the **E site**, from which it is released **with no amino acid attached**. Back in the cytosol that empty tRNA is **recharged** with another molecule of its own specific amino acid, so the same tRNA can be used over and over while the polypeptide is built.

Q18. The **start codon AUG** marks where translation begins: it binds the first tRNA (carrying methionine) and **sets the reading frame**, fixing how the following bases are grouped into codons. The **stop codons (UAA, UAG and UGA)** mark where translation ends: **no tRNA has a matching anticodon**, so no amino acid is added, translation stops and the **finished polypeptide is released**. Stop codons therefore code for no amino acid. *(1 mark: AUG starts translation / codes for Met / sets reading frame; 1 mark: stop codon ends translation; 1 mark: no tRNA matches, polypeptide released.)*

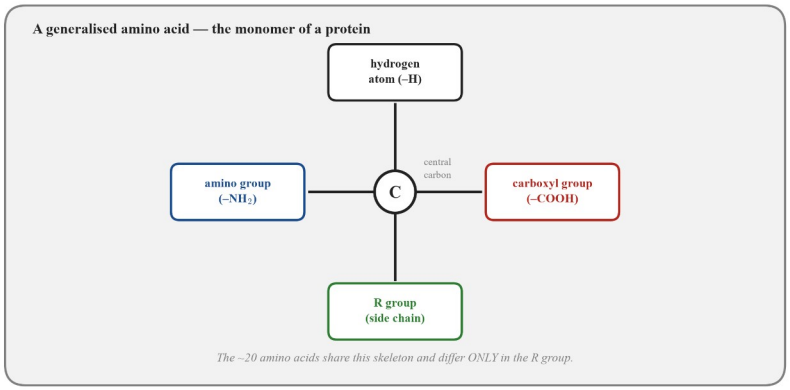
Q19. (a) The introns removed = pre-mRNA length − exon length = 3600 − 1200 = **2400 nucleotides**. (b) The spliced sequence is shorter because **splicing removes the introns** from the pre-mRNA and joins only the **exons**, so the coding sequence contains fewer nucleotides than the original pre-mRNA. (c) The number of codons = 1200 ÷ 3 = **400 codons**. Because one codon must be a **stop codon** (which adds no amino acid), the polypeptide contains a maximum of **399 amino acids**. *(1 mark: 2400; 1 mark: introns removed; 1 mark: 400 codons; 1 mark: 399 amino acids.)*

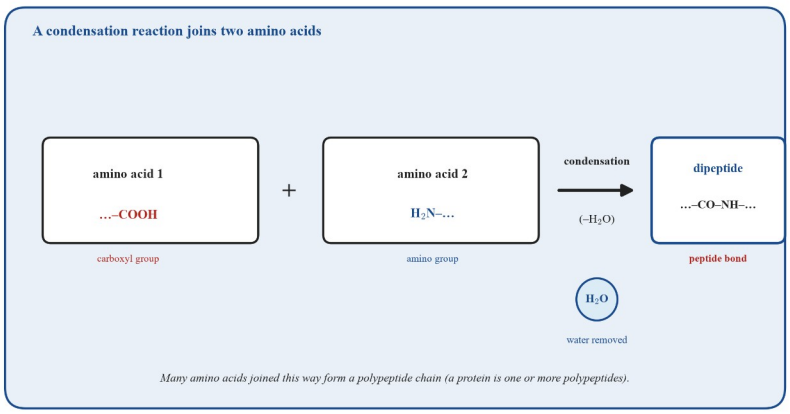
Q20. (a) From the codon table, **methionine has only one codon, AUG**, while **cysteine** is specified by **UGU or UGC** and **valine** by **GUU, GUC, GUA or GUG**. Choosing one codon for each amino acid and ending with the stop codon given, one correct answer is **5′-AUG-UGU-GUU-UAG-3′**. (b) The mRNA was built on the template strand by **complementary, antiparallel** base pairing, so the template is recovered by pairing each mRNA base with its DNA partner (A opposite U, G opposite C, and **T opposite A**): the template strand is **3′-TAC-ACA-CAA-ATC-5′**. The **non-template (coding) strand** is the complement of that template, so it has the **same base sequence as the mRNA with T in place of U**: 5′-ATG-TGT-GTT-TAG-3′. (c) The genetic code is **degenerate**, so most amino acids are specified by more than one codon — cysteine by two and valine by four. There are therefore 1 × 2 × 4 = **8 different mRNA sequences** that would give Met – Cys – Val with this stop codon, and the polypeptide alone cannot say which was used: working backwards is **one-to-many**. Reading forwards is **one-to-one**, because degenerate does **not** mean ambiguous — each individual codon specifies **only one** amino acid, so a known mRNA always yields exactly one polypeptide. ∴ the researcher can narrow the gene to eight possible sequences but cannot identify it exactly from the polypeptide. *(a: 2 marks — codons matching Met–Cys–Val; correct 5′→3′ direction including the stop codon; b: 3 marks — template complementary and antiparallel to the mRNA; 5′ and 3′ ends labelled; coding strand = the mRNA sequence with T for U; c: 2 marks — more than one codon per amino acid so several mRNAs are possible; each codon specifies only one amino acid so a given mRNA gives one polypeptide.)*

Theory

Proteins are a large and diverse group of molecules that carry out most of the work of a cell. A protein is a **polymer** built from smaller subunits called **amino acids**. What lets one protein speed up a reaction, another carry oxygen and another form a strong fibre is the same thing in every case: the **three-dimensional (3-D) shape** the chain folds into. This subtopic follows a protein from its monomers, through the levels of structure that give it its shape, to the pathway by which a cell exports proteins it makes for secretion. (Proteins are assembled at ribosomes by **translation**, which is covered in 2B; here we begin with the finished chain of amino acids.)

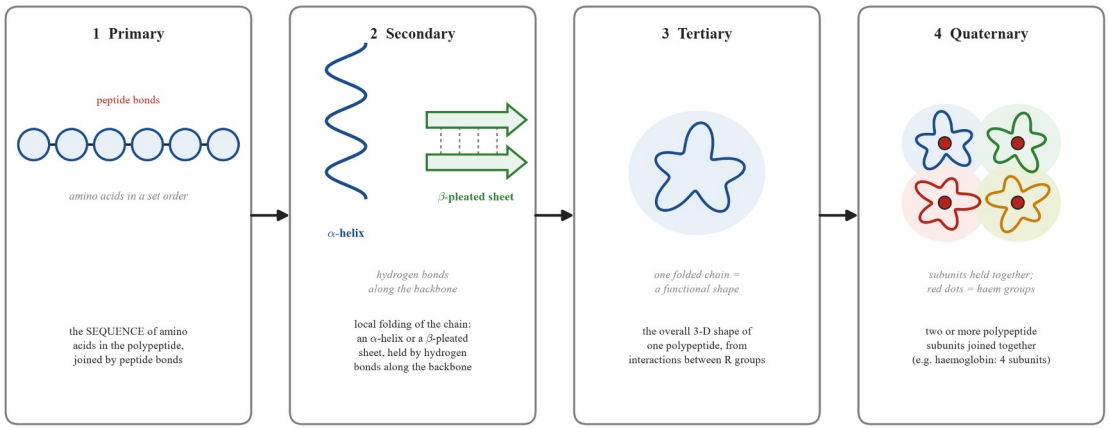
Amino acids — the monomers of a protein. Every amino acid is built on the same skeleton: a **central carbon** atom bonded to four things — an **amino group (–NH₂)**, a **carboxyl group (–COOH)**, a single **hydrogen atom (–H)**, and a **variable R group** (also called the side chain). There are about **20** different amino acids, and they all share this skeleton — they differ **only in their R group**. The R group can be small or large, charged, polar or non-polar, and it is these differences that give each amino acid its chemical character. Because the order of amino acids can be varied without limit, a huge number of different proteins can be built from just 20 building blocks.





The peptide bond — joining amino acids. Two amino acids are joined by a **condensation reaction**: the **carboxyl group** of one amino acid reacts with the **amino group** of the next, a molecule of **water is removed**, and a **peptide bond** forms between them. Two amino acids joined this way make a **dipeptide**; many amino acids joined in a chain make a **polypeptide**. A protein is one or more polypeptide chains folded into a working shape. (In a cell this joining is carried out by a ribosome during translation — see 2B.)

The four levels of protein structure. A functional protein is organised at four hierarchical levels. Each level builds on the one before it.



- **Primary structure** — the **sequence (order) of amino acids** in the polypeptide chain, held together by peptide bonds. The primary structure is set by the gene, and it determines every level above it.
- **Secondary structure** — **local folding of the chain** into regular, repeating shapes. The two common forms are the **α-helix** (a coil) and the **β-pleated sheet**. Secondary structure is held in place by **hydrogen bonds between atoms of the polypeptide backbone** (not between R groups).
- **Tertiary structure** — the **overall 3-D shape of a single polypeptide**, produced when the chain folds back on itself. It is stabilised by **interactions between the R groups** — such as hydrogen bonds, ionic bonds, disulfide bridges (between the R groups of two cysteines) and hydrophobic interactions. For a protein made of one chain, the tertiary structure is its functional shape.
- **Quaternary structure** — present only in proteins made of **two or more polypeptide subunits**. The separate folded subunits join together to form the working protein. A classic example is **haemoglobin**, which has **four** polypeptide subunits (each holding a haem group that binds oxygen). Not all proteins have a quaternary level.

Shape determines function. Each protein folds into one particular 3-D shape, and that shape is what lets it do its job — for example, an enzyme's shape creates an **active site** that fits its substrate, and a receptor's shape creates a **binding site** that fits a specific signalling molecule. Because the primary structure determines the folding, a **change in the amino-acid sequence** (for example, from a mutation) can change the shape and therefore **change or destroy the protein's function**. The 3-D shape can also be lost by **denaturation**: heat or an extreme pH breaks the hydrogen bonds and other interactions that hold the secondary and tertiary structure together, so the protein **unfolds** and no longer works. (The peptide bonds of the primary structure are usually not broken, but the shape — and the function — is lost, and this is often permanent.)

The proteome. The **proteome** is the **complete set of proteins expressed by a cell or organism**. It is different from the genome (the organism's complete set of genes): the proteome is **larger and more dynamic** than the genome. A cell can express many more proteins than it has genes, and its proteome **changes over time** — with the type of cell, its stage of development and the conditions it is in — whereas essentially the same genome is present in every cell of the organism.

A diversity of functions. Proteins are the most functionally varied group of molecules in a cell. The table gives the main categories in the study of biology, with an example of each.

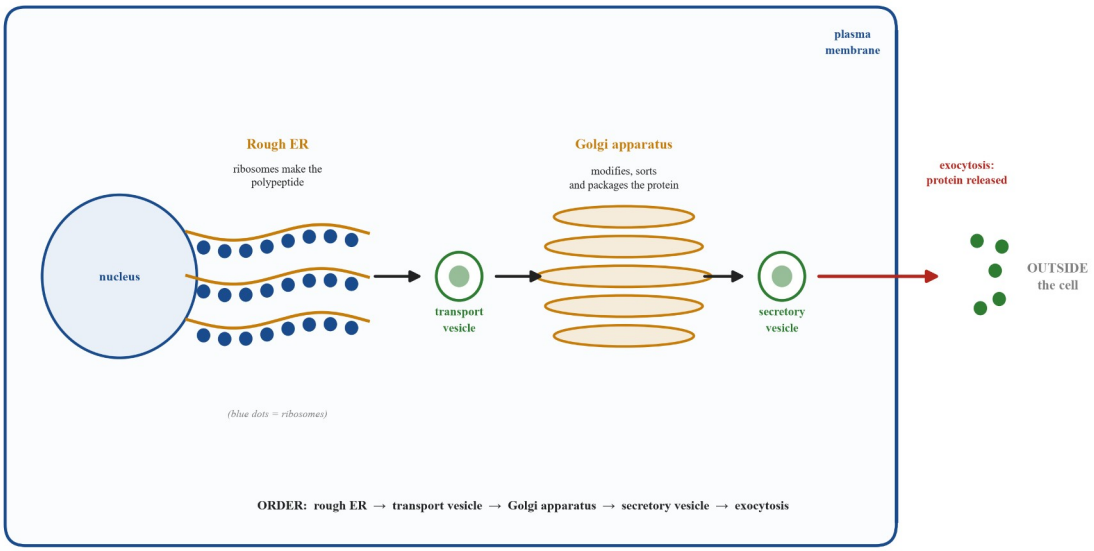
Functional category	What the protein does	Example
Enzyme (catalyst)	speeds up a specific biochemical reaction	amylase (breaks starch down to sugars)
Transport	carries a substance around the body or across a membrane	haemoglobin (carries oxygen)
Structural	forms strong or supporting structures	keratin (hair, nails); collagen (connective tissue)
Hormone (signalling)	carries a chemical message between cells	insulin (regulates blood glucose)
Antibody (defence)	binds to a specific antigen in an	IgG (binds a bacterial antigen)

Functional category	What the protein does	Example
	immune response	
Receptor	binds a signalling molecule and triggers a response	the insulin receptor (on a liver cell)

Notice that each Example names an **actual protein**, not the category again. If a question asks you to *name a protein* that catalyses a reaction, "an enzyme" scores nothing — you must name one, such as **amylase** or **catalase**.

Enzymes are the one category named directly in this AoS: they are protein **catalysts** that speed up reactions in the biochemical pathways of a cell. A **biochemical pathway** is a sequence of reactions in which the product of one step becomes the starting material of the next, and **each step is catalysed by its own specific enzyme** — an enzyme's 3-D shape gives it an **active site** that fits only one substrate, so it speeds up only one reaction of the pathway. **Amylase** (starch → sugars) and **catalase** (hydrogen peroxide → water + oxygen) are named examples. (How enzymes work, and the factors that affect their activity, are covered in Chapter 4.)

The protein secretory pathway. Proteins that a cell **exports** (secretes) — such as hormones, antibodies and digestive enzymes — follow a set route through the cell's membrane-bound organelles. The order is important. Proteins that stay and work **inside** the cell (haemoglobin, for example) are made on **free ribosomes in the cytosol** and never enter this pathway, which is why only cells that export a lot of protein carry a lot of rough ER and Golgi apparatus.



1. **Rough endoplasmic reticulum (rough ER).** Ribosomes attached to the rough ER **synthesise the polypeptide**, which enters the rough ER and begins to fold. The rough ER packages the protein into a small **transport vesicle** that buds off.
2. **Transport vesicle → Golgi apparatus.** The transport vesicle carries the protein to the **Golgi apparatus**, where it **modifies, sorts and packages** the protein (for example, by adding carbohydrate groups and labelling it for its destination).
3. **Golgi → secretory vesicle → exocytosis.** The finished protein is packaged into a **secretory vesicle**, which moves to and fuses with the **plasma membrane**, releasing the protein outside the cell by **exocytosis**.

The order to remember is: **rough ER → transport vesicle → Golgi apparatus → secretory vesicle → exocytosis at the plasma membrane.**

Worked examples

Example 1 — scaffolded

The diagram of a generalised amino acid shows a central carbon bonded to four groups. (a) Name the four groups or atoms bonded to the central carbon, and state which one makes amino acids differ from one another. (b) Name the

reaction that joins two amino acids and the bond it forms. (c) State the small molecule removed when they join, and name the product formed from just two amino acids.

Working	Comment
The central carbon is bonded to: an **amino group (–NH_2)**, a **carboxyl group (–COOH)**, a **hydrogen atom (–H)** and an **R group (side chain)**.	List all four. The amino and carboxyl groups are the reactive ends used to join amino acids.
The **R group** is the part that varies — amino acids differ only in their R group.	This is what gives each of the ~20 amino acids its own chemical character.
Two amino acids are joined by a **condensation reaction**, which forms a **peptide bond**.	Name both the reaction (condensation) and the bond (peptide).
A molecule of **water (H_2O)** is removed; the product of joining two amino acids is a **dipeptide**.	Condensation always removes water. Two = dipeptide; many = polypeptide.

Example 2 — scaffolded

A protein is described at four points. Match each description to the correct level of protein structure (primary, secondary, tertiary or quaternary). (i) The exact order of amino acids along the chain. (ii) A stretch of the chain coiled into an α-helix, held by hydrogen bonds along the backbone. (iii) The whole chain folded into a compact 3-D shape by interactions between its R groups. (iv) Four separate folded polypeptides joined together to make the working protein.

Working	Comment
(i) The **sequence of amino acids** is the **primary** structure.	Primary = the order of the monomers, held by peptide bonds.
(ii) An **α-helix held by backbone hydrogen bonds** is **secondary** structure.	Secondary = local α-helix / β-pleated sheet from backbone H-bonds.
(iii) The **overall 3-D fold from R-group interactions** is **tertiary** structure.	Tertiary = the shape of one whole polypeptide, set by R-group interactions.
(iv) **Several polypeptide subunits joined together** is **quaternary** structure.	Quaternary = two or more subunits (e.g. haemoglobin's four).

Example 3 — unscaffolded

A single amino acid buried in the interior of an enzyme is replaced, by mutation, with an amino acid that has a very different R group. Explain how this one change could stop the enzyme working.

The identity of that amino acid is part of the enzyme's **primary structure** (its amino-acid sequence). Because the primary structure determines how the chain folds, changing one R group can change the **R-group interactions** that hold the **tertiary structure** together. If those interactions are altered, the polypeptide folds into a **different 3-D shape**. An enzyme's shape forms its **active site**, so a change in shape can distort the active site; if the active site no longer fits the substrate, the enzyme **can no longer catalyse its reaction**. ∴ a single change to the primary structure can alter the higher levels of structure and, through them, destroy the enzyme's function.

Example 4 — unscaffolded

A goblet cell in the lining of the gut makes and secretes mucin, a large protein, to the outside of the cell. Describe the path mucin takes from where it is made to where it is released, naming the organelles and vesicles in order.

Mucin is a secreted protein, so it follows the **protein secretory pathway**. It is first synthesised by **ribosomes on the rough endoplasmic reticulum**, entering the rough ER as it is made. The rough ER packages it into a **transport vesicle**, which carries it to the **Golgi apparatus**. The Golgi **modifies, sorts and packages** the mucin and buds it off in a **secretory vesicle**. The secretory vesicle moves to the **plasma membrane** and fuses with it, releasing the mucin outside the cell by **exocytosis**. ∴ the order is rough ER → transport vesicle → Golgi apparatus → secretory vesicle → exocytosis.

Practice questions

Scaffolded

Q1. Define each of the following terms: (a) amino acid; (b) polypeptide; (c) denaturation.

Q2. A protein is a polymer built from amino acid monomers. (a) Distinguish between a dipeptide and a polypeptide. (b) State what must happen to a newly made polypeptide before it can work as a functional protein. (c) Explain how about 20 different amino acids can give rise to an enormous number of different proteins.

Q3. A functional protein is organised at four levels of structure. (a) State what is meant by the primary structure of a protein. (b) Name the two common forms of secondary structure and the type of bond that holds them in place. (c) State what type of interaction gives rise to the tertiary structure.

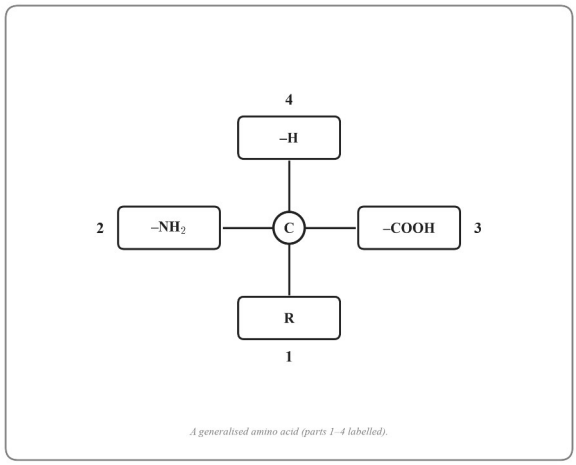
Q4. The function of a protein depends on its 3-D shape, and that shape can be lost. (a) Explain why the 3-D shape of a protein is important for its function. (b) Name two conditions that can cause a protein to denature. (c) State what happens to the primary structure of a protein when it denatures.

Q5. Proteins are a diverse group of molecules that collectively make up an organism's proteome. (a) Define the proteome. (b) Name a protein that carries out each of these functions: transport; defence; catalysis. (c) State the role of enzymes in the biochemical pathways of a cell.

Q6. The Golgi apparatus and its associated vesicles complete the export of a protein. (a) State two changes the Golgi apparatus makes to a protein before it is exported. (b) Distinguish between a transport vesicle and a secretory vesicle. (c) Describe what happens when a secretory vesicle reaches the plasma membrane.

Unscaffolded

Q7. *(Stimulus.)* The diagram shows a generalised amino acid, with four parts labelled 1–4.



A generalised amino acid (parts 1–4 labelled).

(a) Name the part labelled 1, 2, 3 and 4.
(b) State which numbered part varies between different amino acids. *(5 marks)*

Q8. A polypeptide is a single chain of 150 amino acids. State how many peptide bonds hold the chain together and how many molecules of water were released while it was being built, and explain your answer. *(4 marks)*

Q9. A polypeptide folds back on itself, so that amino acids that are far apart in the chain end up side by side. Name two types of interaction between their R groups that hold the resulting tertiary structure in place. *(2 marks)*

Q10. Distinguish between the secondary and tertiary levels of protein structure. *(3 marks)*

Q11. In a cell, substance P is converted into substance S in three steps, each catalysed by a different enzyme. Explain why three different enzymes are needed rather than one, and predict what happens in the cell if the enzyme that catalyses step 2 is missing. *(4 marks)*

Q12. Explain why the proteome of an organism is larger and more dynamic (changeable) than its genome. *(3 marks)*

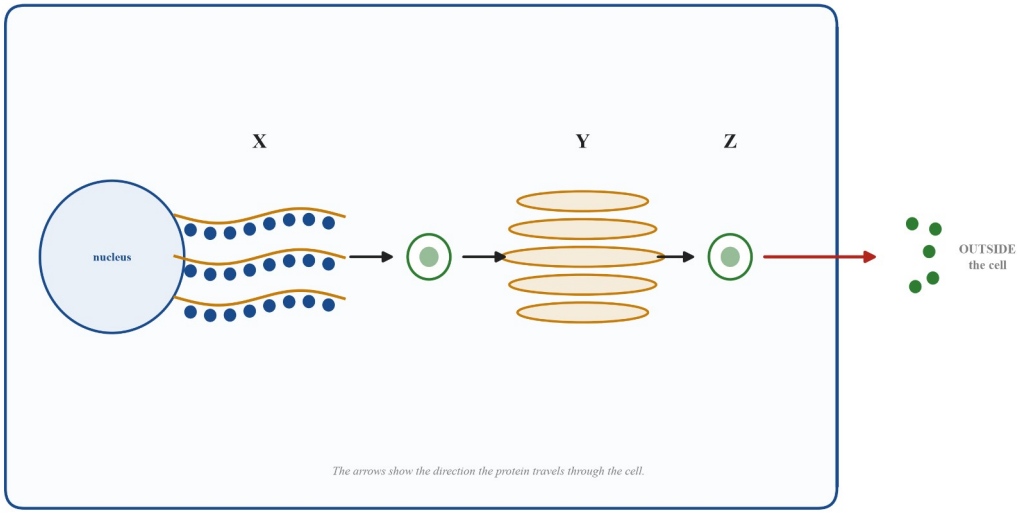
Q13. *(Stimulus.)* A researcher studying a marine invertebrate records four of its proteins and what each one does:

Protein	Role in the organism
ferrizoin	carries oxygen in the blood
pepsase	catalyses the breakdown of food proteins in the gut
spinelin	forms the tough fibres of the body wall
glucostat	released into the blood to regulate sugar levels

State the functional category (for example: enzyme, transport, structural, hormone) of each protein. *(4 marks)*

Q14. A plasma cell exports antibodies. The same cell also makes an enzyme that works in its own cytosol and is never released. State where in the cell each of these two proteins is synthesised, and explain why only the antibody passes through the Golgi apparatus. *(4 marks)*

Q15. *(Stimulus.)* The diagram shows a cell that manufactures and secretes a protein. Three structures are labelled X, Y and Z.



(a) Name structures X, Y and Z.
(b) Name the structure (shown, between X and Y, but not labelled) that carries the protein from X to Y.
(c) Name the process occurring at the plasma membrane on the right, by which the protein is released. *(5 marks)*

Q16. A membrane receptor protein depends on a specific 3-D shape to bind its signalling molecule. The cell is exposed to a very high temperature. (a) Explain what happens to the receptor protein's structure. (b) Explain the effect this has on its ability to bind the signalling molecule. *(4 marks)*

Q17. Haemoglobin is a protein made of four polypeptide subunits joined together. (a) Name the level of protein structure this describes. (b) Distinguish between the tertiary and quaternary levels of protein structure. *(3 marks)*

Q18. *(Stimulus.)* A pancreatic beta cell manufactures and secretes the protein hormone insulin. A researcher observes that the cell contains an unusually large amount of rough endoplasmic reticulum and Golgi apparatus. (a) Explain why a cell that secretes a large amount of protein would contain a large amount of rough ER. (b) The researcher then treats the cells with a drug that stops secretory vesicles fusing with the plasma membrane. Predict the effect on the amount of insulin released and on the contents of the cytoplasm, and explain your prediction. *(4 marks)*

Q19. *(Synthesis.)* A gland cell manufactures and secretes a protein hormone, "somatrin", which is built from a chain of 84 amino acids. A researcher attaches a label to newly made somatrin and records where the label is found: after 3 minutes it is inside the rough endoplasmic reticulum; after 20 minutes it is inside a stack of flattened membrane sacs near the nucleus; after 45 minutes it is outside the cell. (a) Name the structure holding the label at 20 minutes, and name the structure that carried somatrin to it from the rough ER. (b) Explain how the primary structure of somatrin

gives rise to its final functional 3-D shape, referring to the levels of protein structure. (c) State one change the structure named in (a) makes to somatrin, then name the structure that carries somatrin over the last stage of its journey and the process by which it is released. *(8 marks)*

Answers

Q1. (a) The monomer of a protein: a central carbon bonded to an amino group, a carboxyl group, a hydrogen atom and a variable R group. (b) A chain of many amino acids joined by peptide bonds. (c) The loss of a protein's 3-D shape, and therefore its function, when the bonds holding that shape are broken. **Q2.** (a) A dipeptide is two amino acids joined by one peptide bond; a polypeptide is many amino acids joined in a chain by peptide bonds. (b) It must fold into its specific 3-D shape (and in some proteins join with other subunits). (c) The 20 amino acids can be joined in any order and in chains of any length, and each different sequence folds into a different shape, so a huge number of proteins is possible. **Q3.** (a) The sequence (order) of amino acids in the polypeptide chain. (b) The α-helix and the β-pleated sheet, held by hydrogen bonds (along the polypeptide backbone). (c) Interactions between the R groups (side chains). **Q4.** (a) The specific 3-D shape creates a binding site / active site that fits a specific molecule, so the shape determines what the protein can do. (b) Heat (a high temperature) and an extreme pH. (c) It is unchanged — the peptide bonds are not broken, so the amino acids stay joined in the same order. **Q5.** (a) The complete set of proteins expressed by a cell/organism. (b) Transport = haemoglobin; defence = immunoglobulin G (IgG); catalysis = amylase (or catalase). (c) Each step of a biochemical pathway is catalysed by its own specific enzyme, which speeds that step up. **Q6.** (a) Any two of: it modifies the protein (e.g. adds carbohydrate groups); it sorts and labels it for its destination; it packages it into a vesicle. (b) A transport vesicle buds from the rough ER and carries the protein to the Golgi; a secretory vesicle buds from the Golgi and carries the finished protein to the plasma membrane. (c) It fuses with the plasma membrane and releases the protein outside the cell — exocytosis. **Q7.** (a) 1 = R group (side chain); 2 = amino group ($–NH_2$); 3 = carboxyl group (–COOH); 4 = hydrogen atom (–H). (b) Part 1, the R group. **Q8.** 149 peptide bonds and 149 molecules of water: joining n amino acids needs n − 1 bonds, and each bond forms in a condensation reaction that removes one molecule of water. **Q9.** Any two of: hydrogen bonds, ionic bonds, disulfide bridges, hydrophobic interactions. **Q10.** Secondary = local folding of the backbone into an α-helix or β-pleated sheet, held by hydrogen bonds along the backbone; tertiary = the overall 3-D shape of the whole polypeptide, held by interactions between R groups. **Q11.** Each enzyme's active site fits only one substrate, so one enzyme cannot catalyse all three different reactions; without enzyme 2 the pathway stops after step 1, so the step-1 intermediate builds up and substance S is never made. **Q12.** One genome can be expressed as many more proteins (one gene → more than one protein; proteins can also be modified), and the proteome changes with cell type/conditions/time, whereas the genome is the same in every cell. **Q13.** ferrizoin = transport; pepsase = enzyme (catalyst); spinelin = structural; glucostat = hormone. **Q14.** The antibody is made by ribosomes attached to the rough ER; the cytosolic enzyme is made by free ribosomes in the cytosol. Only proteins made on the rough ER enter the secretory pathway, so only the antibody is carried to the Golgi to be modified, sorted and packaged for export. **Q15.** (a) X = rough ER; Y = Golgi apparatus; Z = secretory vesicle. (b) A transport vesicle. (c) Exocytosis. **Q16.** (a) The heat breaks the hydrogen bonds/interactions holding its shape, so it denatures (unfolds) and loses its 3-D shape. (b) Its binding site changes shape, so it can no longer bind the signalling molecule. **Q17.** (a) Quaternary structure. (b) Tertiary = the 3-D shape of a single polypeptide; quaternary = two or more polypeptide subunits joined together. **Q18.** (a) Rough ER carries the ribosomes that synthesise secreted proteins, so a cell making a lot of secreted protein needs a lot of rough ER. (b) Almost no insulin would be released, and insulin-filled secretory vesicles would build up in the cytoplasm, because exocytosis requires the vesicle to fuse with the plasma membrane. **Q19.** (a) The Golgi apparatus; a transport vesicle. (b) The sequence (primary) folds into α-helices/ β-pleated sheets (secondary) via backbone H-bonds, then into an overall 3-D shape (tertiary) via R-group interactions, giving the functional shape. (c) The Golgi modifies it (e.g. adds carbohydrate groups) and sorts/labels it; a secretory vesicle carries it to the plasma membrane; it is released by exocytosis.

Fully-worked solutions

Q1. (a) An **amino acid** is the monomer (building block) of a protein: a **central carbon** atom bonded to an **amino group**, a **carboxyl group**, a **hydrogen atom** and a **variable R group**. (b) A **polypeptide** is a chain of many amino acids joined together by **peptide bonds**. (c) **Denaturation** is the **loss of a protein's 3-D shape** — and therefore of its function — when the **bonds and interactions holding its secondary and tertiary structure** together are broken, for example by heat or an extreme pH.

Q2. (a) A **dipeptide** is **two** amino acids joined by a **single peptide bond**; a **polypeptide** is a chain of **many** amino acids joined by peptide bonds. Both are built by the same condensation reaction — they differ in the number of monomers. (b) The new chain must **fold into its specific 3-D shape**; some proteins must also **join with other polypeptide subunits** (a quaternary structure) before they will work. (c) The 20 amino acids can be joined in **any**

order and in chains of any length, so the number of possible sequences is enormous. Because the **primary structure determines the folding**, each different sequence gives a **different 3-D shape** — and therefore a different protein.

Q3. (a) The **primary structure** is the **sequence (order) of amino acids** in the polypeptide chain, joined by peptide bonds. (b) The two common forms of **secondary structure** are the **α-helix** and the **β-pleated sheet**; both are held in place by **hydrogen bonds between atoms of the polypeptide backbone**. (c) The **tertiary structure** arises from **interactions between the R groups** (side chains) of the amino acids — for example hydrogen bonds, ionic bonds, disulfide bridges and hydrophobic interactions.

Q4. (a) A protein folds into one specific **3-D shape**, and that shape forms a **binding site or active site** with a particular shape and chemistry. The protein can only carry out its job (e.g. bind a substrate or a signalling molecule) if that site has the correct shape, so **shape determines function**. (b) **Heat (a high temperature)** and an **extreme pH** (strongly acidic or strongly alkaline conditions). (c) The **primary structure is unchanged**. Denaturation breaks the hydrogen bonds, ionic bonds and other interactions holding the secondary, tertiary and quaternary structure together, but the **peptide bonds of the backbone are not broken** — so the amino acids remain joined in the same order even though the shape, and with it the function, are lost.

Q5. (a) The **proteome** is the **complete set of proteins expressed by a cell or organism** (at a given time). (b) **Transport** = **haemoglobin** (carries oxygen in red blood cells); **defence** = **immunoglobulin G (IgG)**, an antibody that binds a specific antigen; **catalysis** = **amylase** (breaks starch down to sugars). Catalase, pepsin or any other named enzyme is equally acceptable — but "an enzyme" is **not** an answer here: the question asks you to *name a protein*, so it must be a named example rather than the functional category it was given. (c) A **biochemical pathway** is a sequence of reactions in which the product of one step is the starting material of the next. **Each step is catalysed by its own specific enzyme**, which **speeds that reaction up** without being used up, so enzymes control the rate at which the pathway runs.

Q6. (a) Any **two** of: it **modifies** the protein (for example by **adding carbohydrate groups**); it **sorts** the protein and **labels** it for its destination; it **packages** the finished protein into a vesicle. (b) A **transport vesicle** buds from the **rough ER** and carries the protein **to the Golgi apparatus**; a **secretory vesicle** buds from the **Golgi apparatus** and carries the finished protein **to the plasma membrane**. Both are small membrane-bound sacs — they differ in where they bud from and where they deliver. (c) The secretory vesicle **fuses with the plasma membrane**, so its membrane becomes part of the plasma membrane and its contents are **released outside the cell**. This process is **exocytosis**.

Q7. (a) Reading the formula printed in each box: **1 = the R group (side chain)**; **2 = the amino group (–NH_2)**; **3 = the carboxyl group (–COOH)**; **4 = the hydrogen atom (–H)**. (b) The part that varies between different amino acids is **part 1, the R group** — all amino acids share the same skeleton and differ only in their R group. *(1 mark for each of the four parts correctly named; 1 mark for identifying the R group as the variable part.)*

Q8. A peptide bond is formed between **each neighbouring pair** of amino acids, so joining **n** amino acids into one chain needs **n − 1** peptide bonds. For 150 amino acids that is **149 peptide bonds**. Every one of those bonds is formed by a **condensation reaction**, and each condensation reaction removes **one molecule of water** — so **149 molecules of water** were released while the chain was being built. *(1 mark: 149 peptide bonds; 1 mark: 149 molecules of water; 2 marks for the reasoning — n amino acids need n − 1 bonds, and one water is removed per condensation reaction / per bond formed. Total 4.)*

Q9. The **tertiary structure** is held in place by **interactions between the R groups** of amino acids that the folding has brought close together. Any **two** of: **hydrogen bonds**, **ionic bonds**, **disulfide bridges** (a covalent link between the R groups of two cysteines) and **hydrophobic interactions**. *(1 mark each, for any two.)*

Q10. Secondary structure is the **local folding of the polypeptide backbone** into a regular shape — an **α-helix** or a **β-pleated sheet** — held together by **hydrogen bonds between atoms of the backbone**. **Tertiary structure** is the **overall 3-D shape of the whole polypeptide**, produced when the chain folds back on itself, and it is held together by **interactions between the R groups** (side chains). So secondary is local and backbone-based, whereas tertiary is the shape of the entire chain and R-group-based. *(1 mark: secondary = α-helix/β-sheet from backbone H-bonds; 1 mark: tertiary = overall 3-D shape of the polypeptide; 1 mark: tertiary held by R-group interactions.)*

Q11. An enzyme's 3-D shape gives it an **active site** that fits **only one substrate**, so each enzyme can catalyse **only one reaction**. The three steps that convert P into S each act on a **different substrate**, so **three different enzymes** are

needed — one specific to each step. If the enzyme that catalyses **step 2** is missing, **step 2 cannot proceed**, so the pathway **stops after step 1**: the intermediate made in step 1 **accumulates** in the cell, and **substance S is never produced**. *(2 marks: the active site is specific to one substrate, so a single enzyme cannot catalyse all three different reactions; 2 marks: the pathway halts at step 2, so the step-1 intermediate builds up and S is not made. Total 4.)*

Q12. The **genome** (the full set of genes) is essentially the **same in every cell** of an organism, but the **proteome** — the proteins actually present — is larger and more changeable for two reasons. First, **one gene can be expressed as more than one protein**, and proteins can be **modified after they are made**, so there are more proteins than genes. Second, **which proteins a cell makes changes** with the **type of cell, its stage of development and its environment/conditions**, so the proteome differs from cell to cell and from moment to moment, whereas the genome does not. *(1 mark: more proteins than genes; 1 mark: proteome changes with cell type/conditions/time; 1 mark: genome is the same in all cells / over time.)*

Q13. Each protein is placed in the category that matches its role: **ferrizoin — transport** (it carries oxygen in the blood); **pepsase — enzyme / catalyst** (it catalyses the breakdown of food proteins); **spinelin — structural** (it forms the tough fibres of the body wall); **glucostat — hormone** (it is released into the blood to regulate sugar as a chemical message). *(1 mark each.)*

Q14. The **antibody is exported**, so it is synthesised by **ribosomes attached to the rough endoplasmic reticulum** and enters the rough ER as it is made. The **cytosolic enzyme is not exported**, so it is synthesised by **free ribosomes in the cytosol**. Only proteins made on the rough ER enter the **protein secretory pathway**, so only the antibody is carried by a transport vesicle to the **Golgi apparatus** to be modified, sorted and packaged for release. The cytosolic enzyme never enters a membrane-bound organelle — it stays and works in the cytosol. *(1 mark: antibody made on ribosomes of the rough ER; 1 mark: enzyme made on free ribosomes in the cytosol; 2 marks: only proteins destined for export enter the secretory pathway, so only the antibody reaches the Golgi. Total 4.)*

Q15. (a) **X = the rough endoplasmic reticulum (rough ER)** (the membrane system studded with ribosomes, next to the nucleus); **Y = the Golgi apparatus** (the stack of flattened cisternae); **Z = a secretory vesicle** (budding from the Golgi towards the plasma membrane). (b) The unlabelled structure carrying the protein from X to Y is a **transport vesicle**. (c) The protein is released at the plasma membrane by **exocytosis**. *(a: 1 mark each for X, Y, Z; b: 1 mark; c: 1 mark.)*

Q16. (a) The high temperature **breaks the hydrogen bonds and other interactions** that hold the receptor's secondary and tertiary structure together, so the protein **denatures — it unfolds and loses its specific 3-D shape**. (b) The receptor binds its signalling molecule at a **binding site with a specific shape**. Once the protein has unfolded, that **binding site changes shape**, so the signalling molecule **no longer fits and can no longer bind** — the receptor cannot carry out its function. *(a: 2 marks — the bonds/interactions holding the shape are broken; the protein denatures and loses its specific 3-D shape. b: 2 marks — the binding site changes shape; the signalling molecule no longer fits, so it cannot bind. Total 4.)*

Q17. (a) A protein made of **two or more polypeptide subunits joined together** has a **quaternary structure**. (b) **Tertiary structure** is the **overall 3-D shape of a single polypeptide chain** (from R-group interactions), whereas **quaternary structure** is the arrangement of **two or more separate polypeptide subunits joined together** to form the working protein. Haemoglobin's four subunits are an example of quaternary structure, and each subunit's own fold is its tertiary structure. *(1 mark: quaternary named; 1 mark: tertiary = one polypeptide's 3-D shape; 1 mark: quaternary = two or more subunits joined.)*

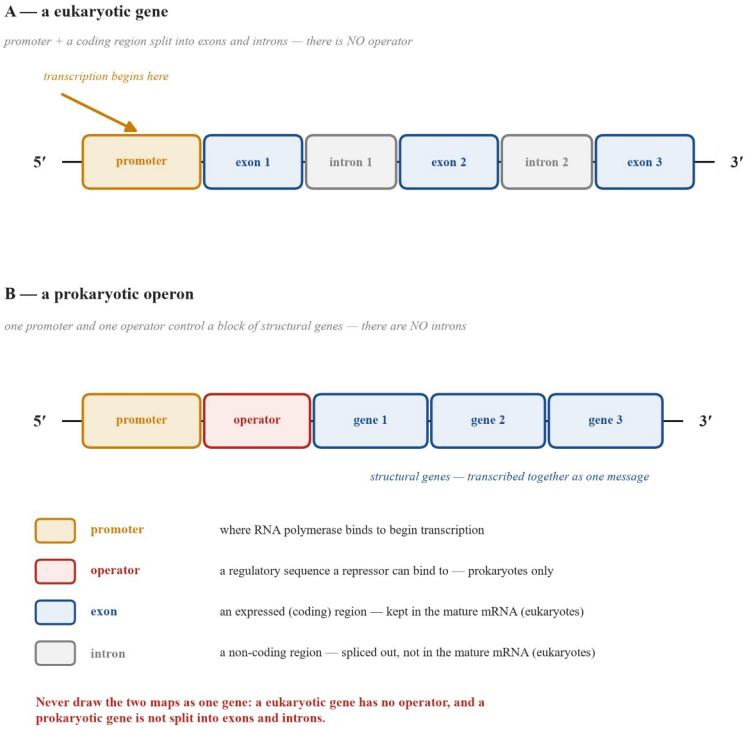
Q18. (a) The **rough ER carries the ribosomes** that synthesise proteins destined for **secretion**, and it is where those proteins begin to fold before being packaged into transport vesicles. A cell such as a beta cell that makes and exports a **large amount of protein** therefore needs a **large amount of rough ER** to synthesise and process all of that protein. (b) Insulin would still be made and processed normally, but **almost none would be released** from the cell. Exocytosis occurs only when a secretory vesicle **fuses with the plasma membrane**; if that fusion is blocked, the vesicles cannot empty their contents outside, so **insulin-filled secretory vesicles accumulate in the cytoplasm** just inside the plasma membrane. *(a: 2 marks — the rough ER bears the ribosomes that make secreted proteins; more secretion therefore demands more rough ER. b: 2 marks — insulin release falls to almost nothing; secretory vesicles build up, because exocytosis requires vesicle–membrane fusion. Total 4.)*

Q19. (a) A **stack of flattened membrane sacs near the nucleus** is the **Golgi apparatus**, so at 20 minutes the label is in the Golgi apparatus. Somatrin was carried there from the rough ER inside a **transport vesicle**, which buds off the rough ER and delivers its contents to the Golgi. (b) The **order of the 84 amino acids is the primary structure**, and it determines all the folding that follows. Parts of the chain fold into **secondary structures — α-helices and β-pleated sheets — held by hydrogen bonds along the backbone**. The whole chain then folds into its **tertiary structure — its overall 3-D shape — through interactions between the R groups** (such as hydrogen bonds, ionic bonds and disulfide bridges). This final 3-D shape is what makes somatrin a **functional hormone**. (If somatrin were made of more than one subunit, joining the subunits would give a quaternary structure — but the functional shape here is set by the tertiary fold.) (c) In the Golgi apparatus somatrin is **modified** — for example by having **carbohydrate groups added** — and is **sorted and labelled for its destination** (any one of these changes earns the mark). It is then packaged into a **secretory vesicle**, which carries it over the last stage of its journey to the **plasma membrane**, fuses with that membrane and releases somatrin outside the cell by **exocytosis**. ∴ between 20 and 45 minutes the label travels Golgi apparatus → secretory vesicle → outside the cell. *(a: 2 marks — Golgi apparatus; transport vesicle. b: 3 marks — primary sequence, secondary α-helix/β-sheet from backbone H-bonds, tertiary 3-D shape from R-group interactions. c: 3 marks — one named Golgi modification; secretory vesicle; exocytosis. Total 8.)*

Theory

A gene is a length of DNA that carries the instructions for making a polypeptide. Not every part of a gene, though, is a coded message for the protein — a gene is organised into regions that do different jobs. Some regions **carry the code** for the polypeptide, and other regions are **control regions** that decide **whether, and when, the gene is switched on**. A cell does not read every gene at once; building a protein costs energy and raw materials, so cells switch genes on only when the protein is needed. This subtopic looks at the parts of a gene and at one worked example of how a gene is switched on and off — the prokaryotic **trp operon**.

The parts of a gene. The course names four regions: the promoter, the operator, exons and introns. These four do **not** all occur in the same gene. The **operator** is a **prokaryotic** regulatory sequence, part of an operon; **exons and introns** are a feature of **eukaryotic** genes. So the two cases are drawn below as **two separate maps**, each read in the **5′→3′** direction: (A) a eukaryotic gene, and (B) a prokaryotic operon.



- **Promoter.** A control region at the **start** of the gene, present in **both** eukaryotes and prokaryotes. It is the sequence that **RNA polymerase binds to in order to begin transcription** (the copying of the gene into RNA — see 2B). The promoter decides **where** transcription starts; if RNA polymerase cannot bind the promoter, the gene cannot be transcribed.
- **Operator.** A short **regulatory sequence** found in **prokaryotes**, lying between the promoter and the structural genes of an operon. A **repressor protein can bind to the operator**. A repressor sitting on the operator **covers the promoter site, so RNA polymerase cannot bind there and transcription never begins**. The operator is therefore an **on/off switch**, not a coding region.
- **Exons.** The **expressed** (coding) regions of a **eukaryotic** gene. Their name is a reminder: an **ex**on is **ex**pressed. Exons are kept in the finished (mature) mRNA, so their code ends up in the polypeptide.
- **Introns.** The **non-coding** regions that lie in between the exons of a **eukaryotic** gene. During RNA processing the **introns are spliced out** and the exons are joined together (the mechanism of splicing is covered in 2B), so introns are **not** present in the mature mRNA and are not translated into protein.

Never draw the two maps as one gene. A eukaryotic gene has **no operator**, and a prokaryotic gene is **not split into exons and introns**. A diagram showing a promoter, an operator, an exon and an intron in a single gene describes no real gene, and is a standard distractor in exam questions on gene structure.

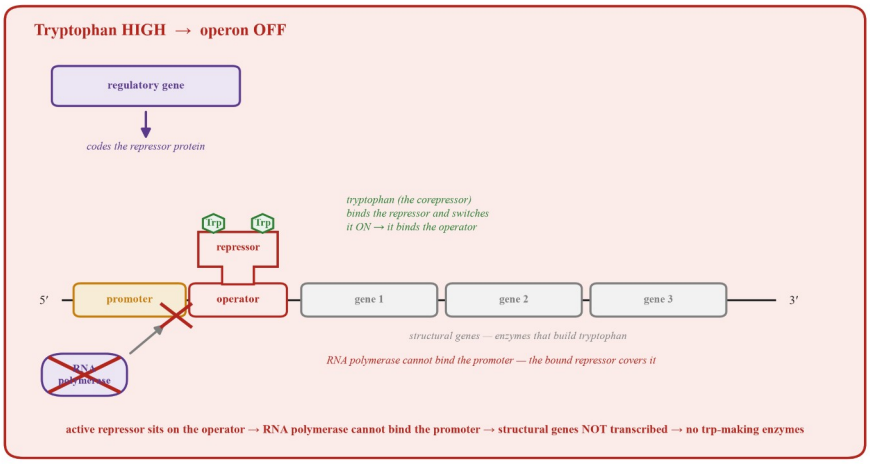
Structural genes and regulatory genes. Genes can be grouped by the **job of the protein they code for**.

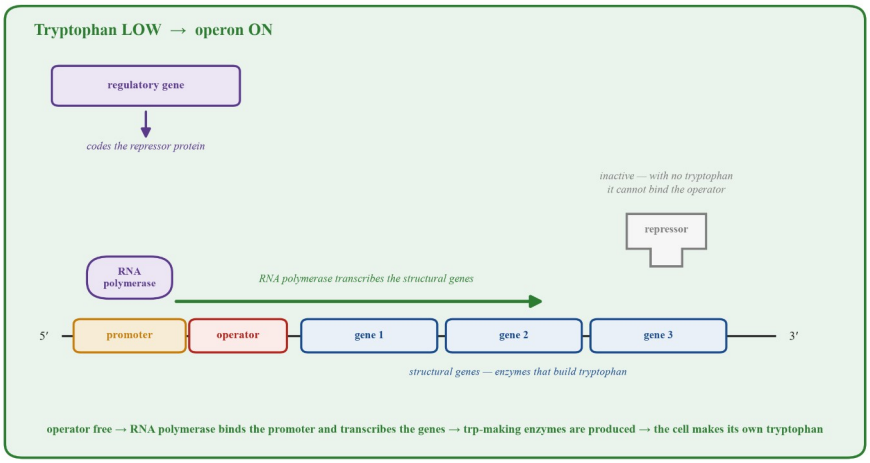
- A **structural gene** codes for a protein that **carries out a function in the cell** — for example an enzyme, or a structural protein such as collagen. In the trp system, the structural genes code for the **enzymes that build the amino acid tryptophan**.
- A **regulatory gene** codes for a protein that **controls the expression of other genes** — for example a **repressor**. Its product does not do a job of metabolism itself; it is a controller that decides whether other genes are switched on or off.

Both are lengths of DNA that are transcribed and translated into a protein. The difference is what that protein does: a structural gene's product **does the cell's work**, while a regulatory gene's product **controls whether other genes are expressed**.

Gene regulation and the operon. In prokaryotes, several structural genes with a **related job** are often grouped together and controlled as a single unit called an **operon**: one promoter and one operator control a block of structural genes that are transcribed **together** as one message. This lets the cell switch a whole set of related genes on or off at once.

The trp operon — a worked example of regulation. The trp operon controls the making of the amino acid **tryptophan**. It contains a **promoter**, an **operator**, and the **structural genes** that code for the enzymes that build tryptophan. A separate **regulatory gene** (with its own promoter) codes for the **trp repressor**, which is produced in an **inactive** form that cannot, on its own, bind the operator. Whether the operon is on or off depends on **how much tryptophan is present**.





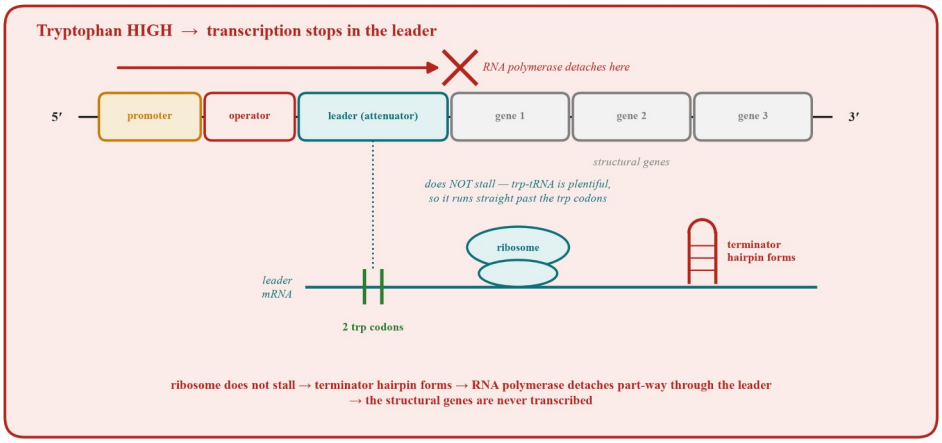
- **When tryptophan is HIGH (operon OFF).** If tryptophan is already plentiful, tryptophan itself acts as a **corepressor**: it **binds to the inactive repressor and changes its shape**, switching it to the **active** form. The

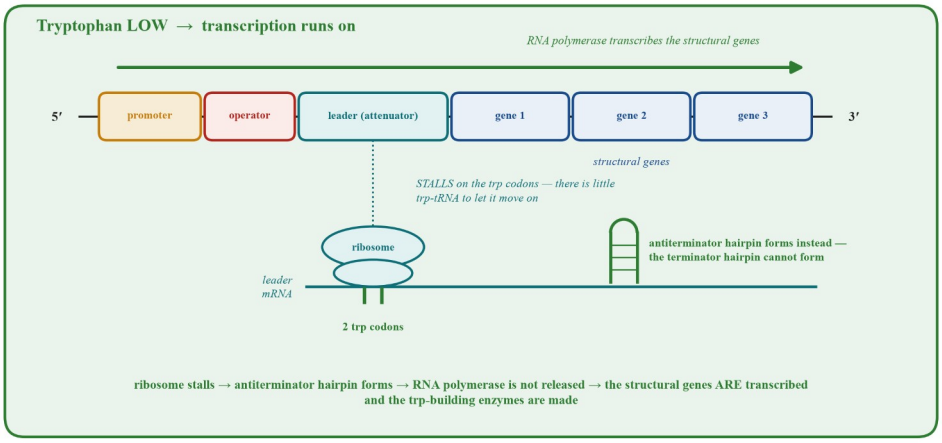
active repressor binds the operator, and in doing so it covers the promoter site, so **RNA polymerase cannot bind the promoter** and transcription never begins. The structural genes are **not transcribed** and the trp-building enzymes are **not made**. This is sensible — there is already plenty of tryptophan, so the cell does not need to make any more.

- **When tryptophan is LOW (operon ON).** If tryptophan is scarce, there is little or no tryptophan to act as a corepressor, so the **repressor stays inactive** and **cannot bind the operator**. The operator is free, so **RNA polymerase binds the promoter and transcribes the structural genes**. The trp-building enzymes are made, and the cell **synthesises its own tryptophan**. Once enough tryptophan builds up, it switches the repressor back on and the operon is turned off again.

Condition	Tryptophan	Repressor	Operator	RNA polymerase	Operon	trp-synthesis enzymes
Trp plentiful	HIGH	active (trp bound)	bound by repressor	cannot bind the promoter	**OFF**	not made
Trp scarce	LOW	inactive	free	binds the promoter and transcribes	**ON**	made

A second control — attenuation. Binding at the operator is not the only way the operon responds to tryptophan. Between the operator and the first structural gene lies a short stretch called the **leader** (or **attenuator**) sequence, and what happens there gives a second, finer control called **attenuation**. A prokaryote has no nucleus, so a **ribosome begins translating the leader mRNA while RNA polymerase is still transcribing it**. The leader contains **two tryptophan (trp) codons**, and the fate of the operon turns on whether the ribosome can get past them.





- **When tryptophan is HIGH.** Charged trp-tRNA is plentiful, so the ribosome **does not stall** at the two trp codons and runs straight past them. This allows a **terminator hairpin** to form in the leader mRNA, and RNA polymerase **detaches part-way through the leader**. Transcription stops **before it reaches the structural genes**, so the trp-building enzymes are **not made**.

- **When tryptophan is LOW.** There is little charged trp-tRNA, so the ribosome **stalls on the two trp codons**. An **antiterminator hairpin** forms instead, the terminator hairpin cannot form, RNA polymerase is **not released**, and transcription **runs on into the structural genes**, so the enzymes **are made**.

Feature	Repression	Attenuation
Where it acts	at the **operator**	in the **leader (attenuator)** sequence
When it acts	**before** transcription begins	**after** transcription has begun
What senses tryptophan	free tryptophan, acting as a **corepressor**	the supply of **charged trp-tRNA**, sensed by the ribosome
Effect when trp is HIGH	repressor binds the operator → RNA polymerase cannot bind the promoter	ribosome does not stall → terminator hairpin → RNA polymerase detaches in the leader
Effect when trp is LOW	repressor inactive → operator free → transcription begins	ribosome stalls → antiterminator hairpin → transcription runs on

Both mechanisms are triggered by **high tryptophan** and both cut the supply of trp-synthesis enzymes, so they work **together**: repression is the coarse on/off switch at the operator, and attenuation catches the transcription that repression lets through, ending it early in the leader.

Why this is efficient. The cell makes the tryptophan-building enzymes **only when tryptophan is scarce**, and stops making them as soon as tryptophan is plentiful. This **saves energy and raw materials** — the cell is not wasting resources making enzymes for a product it already has enough of.

Because a **rise in the end product (tryptophan) switches the operon OFF**, the trp operon is described as a **repressible operon**: it is normally on, and is switched off when its product accumulates. (This is the opposite logic to a system that is normally off and switched on by a signal.)

Worked examples

Example 1 — scaffolded

A **eukaryotic** gene is shown running 5′→3′ with the following regions, in order: **promoter – exon 1 – intron 1 – exon 2**. Identify the region RNA polymerase binds to, name the regions whose code reaches the polypeptide, state what happens to intron 1, and explain why no operator appears on this map.

Working	Comment
RNA polymerase binds the **promoter** — the control region at the start of the gene where transcription begins.	The promoter is a binding site that sets where transcription starts, not a coding region.
The **exons (exon 1 and exon 2)** are the expressed, coding regions; they are kept in the mature mRNA, so their code reaches the polypeptide.	"Exon = expressed." Only the exons carry through to the protein.
Intron 1 is non-coding; it is **spliced out** and is not in the mature mRNA, so it is not translated.	Introns are removed before translation (the splicing mechanism is covered in 2B).
There is no operator because an operator is a **prokaryotic** regulatory sequence, found in an operon between the promoter and the structural genes. A eukaryotic gene does not have one.	Examiners mark a eukaryotic gene drawn with an operator as wrong. Keep the two maps apart: no operator in eukaryotes, no introns in prokaryotes.

Example 2 — scaffolded

A culture of bacteria is grown in a medium containing a **high** concentration of tryptophan. Using the trp operon, predict the state of the repressor, whether it binds the operator, whether RNA polymerase can transcribe the structural genes, and whether the trp-synthesis enzymes are made. Explain your reasoning.

Working	Comment
Tryptophan is HIGH, so tryptophan acts as a	The key point examiners look for — tryptophan

Working	Comment
corepressor: it binds the inactive repressor and changes its shape to the **active** form.	**activates** the repressor (it is a corepressor), it does not inactivate it.
The active repressor now **binds the operator**.	Only the active shape fits the operator. Note it binds the **operator**, not the promoter.
With the repressor sitting on the operator, **RNA polymerase cannot bind the promoter**, so transcription of the structural genes never begins — the operon is **OFF**.	Write "cannot **bind** the promoter", not "cannot get past". The credited mechanism is that the bound repressor blocks RNA polymerase from attaching in the first place.
Therefore the **trp-synthesis enzymes are not made**.	Efficient: tryptophan is already plentiful, so none needs to be built.

Example 3 — unscaffolded

Explain how **attenuation** switches off the production of the trp-synthesis enzymes when tryptophan is **high**.

Between the operator and the first structural gene lies a short **leader (attenuator)** sequence that contains **two trp codons**. Because a prokaryote has no nucleus, a **ribosome begins translating the leader mRNA while RNA polymerase is still transcribing it**. When tryptophan is high there is plenty of charged **trp-tRNA**, so the ribosome **does not stall** at the two trp codons and moves straight past them. This allows a **terminator hairpin** to form in the leader mRNA, and the hairpin causes **RNA polymerase to detach part-way through the leader**. Transcription therefore stops **before it reaches the structural genes**, so the enzymes that build tryptophan are **not made** and the cell does not spend energy or raw materials on a product it already has. ∴ attenuation switches trp synthesis off by ending transcription early in the leader — a control that acts **after** transcription has begun, and so works alongside repression at the operator rather than replacing it.

Example 4 — unscaffolded

In a mutant strain of bacteria, a change in the **operator** sequence means the repressor can **no longer bind** to it. Predict how this mutant expresses the trp-synthesis genes when tryptophan is **high**, and explain the effect on the cell.

Normally, when tryptophan is high the active repressor binds the operator and switches the operon **off**. In this mutant the repressor **cannot bind the altered operator**, so even when tryptophan is high the operator stays **free**. **Repression is therefore lost**: RNA polymerase can bind the promoter and **begin transcribing regardless of the tryptophan level**, so the trp-synthesis enzymes go on being made when the cell already has plenty of tryptophan. This is **wasteful** — the cell spends energy and raw materials building tryptophan-making enzymes (and tryptophan) it does not need. Note that **attenuation is unaffected**, because the leader sequence is intact: some of the transcripts that now start will still be terminated early in the leader, so the mutant makes **more** enzyme than the wild type at high tryptophan, but not the full amount it makes when tryptophan is scarce. ∴ a non-functional operator removes the cell's main switch for trp synthesis, leaving only attenuation to limit it.

Practice questions

Scaffolded

Q1. Define each of the following parts of a gene: (a) exon; (b) intron; (c) promoter.

Q2. A student draws a **human** gene running 5′→3′ as: **promoter – operator – exon 1 – intron 1 – exon 2**. (a) Identify the one region that should not appear in this gene, and explain why. (b) State the type of organism whose genes contain that region, and name what lies immediately on either side of it. (c) After the student's error is corrected, state which of the remaining regions are **not** present in the mature mRNA.

Q3. Genes can be described as structural or regulatory. (a) State what a structural gene codes for. (b) State what a regulatory gene codes for. (c) In the trp operon, state which type of gene codes for the repressor.

Q4. In the trp operon, a short **leader (attenuator)** sequence lies between the operator and the first structural gene. (a) State how many tryptophan codons the leader contains. (b) State whether the ribosome stalls at those codons when tryptophan is **high**, and explain your answer. (c) State what happens to RNA polymerase as a result.

Q5. A bacterium is grown in a medium with a **very low** concentration of tryptophan. (a) State the state of the repressor. (b) State whether the repressor can bind the operator. (c) State what RNA polymerase does and what the cell is then able to make.

Q6. Cells regulate which of their genes are expressed. (a) Explain one reason a cell does not express all of its genes at once. (b) State what an operon is. (c) Explain one advantage to a bacterium of controlling several related genes from a **single** promoter and operator, rather than each of those genes having its own.

Unscaffolded

Q7. Distinguish between an **exon** and an **intron**. *(2 marks)*

Q8. Describe the function of the **promoter** and the function of the **operator**, and state how their roles differ. *(3 marks)*

Q9. Distinguish between a **structural gene** and a **regulatory gene**. *(2 marks)*

Q10. When tryptophan levels are **high**, the trp operon is controlled by both **repression** and **attenuation**. Compare these two mechanisms. *(4 marks)*

Q11. Explain what happens to the trp operon when tryptophan levels in the cell are **low**, and state what the cell is then able to produce. *(4 marks)*

Q12. Explain why regulating tryptophan production in this way is efficient for a bacterial cell. *(3 marks)*

Q13. The trp operon is described as a **repressible** operon. Explain what this means, with reference to the role of tryptophan. *(2 marks)*

Q14. A mutation in the **regulatory gene** produces a repressor protein that binds the operator **even when no tryptophan is present**. Predict how this mutant strain expresses the trp-synthesis genes when tryptophan is **low**, and explain the consequence for the cell. *(3 marks)*

Q15. A mutation in the **promoter** means that RNA polymerase can no longer bind to it. Predict the effect of this mutation on the production of the trp-synthesis enzymes when tryptophan is **low**, and explain why. *(3 marks)*

Q16. *(Stimulus.)* A bacterium makes an amino acid (call it **amino acid X**) using a repressible operon that works in the same way as the trp operon. Reading the operon 5′→3′, four regions are labelled:

Region (5′→3′)	Description
1	the sequence RNA polymerase binds to
2	a short sequence that a repressor protein can bind to
3 and 4	two genes coding the enzymes that build amino acid X

(a) Name region **1** and region **2**.
(b) State whether the genes at regions **3 and 4** are structural or regulatory genes, and justify your answer.
(c) Amino acid X is a corepressor for this operon. Predict whether the operon is ON or OFF when amino acid X is **abundant**, and explain. *(4 marks)*

Q17. *(Data interpretation.)* Bacteria were grown at different tryptophan concentrations and the relative amount of trp-synthesis enzyme in the cells was measured.

Tryptophan in medium (relative units)	Trp-synthesis enzyme in cells (relative units)
100 (high)	5
50	32

Tryptophan in medium (relative units)	Trp-synthesis enzyme in cells (relative units)
10	84
0 (none)	100

(a) Describe the relationship between the tryptophan concentration and the amount of trp-synthesis enzyme.

(b) Explain this relationship using the trp operon.

(c) Predict the relative amount of enzyme you would expect if the tryptophan concentration were raised to 200 relative units, and justify your prediction. *(4 marks)*

Q18. A student writes: "Tryptophan switches the trp operon **on**, because when tryptophan is present the genes are transcribed." Identify the error in this statement and rewrite it correctly. *(2 marks)*

Q19. *(Synthesis.)* The soil bacterium *Bacillus scriptus* makes tryptophan using a trp operon organised in the same way as the one studied above. Two strains were grown with and without tryptophan, and the relative amount of trp-synthesis enzyme in the cells was measured.

Strain	Enzyme with NO tryptophan	Enzyme with HIGH tryptophan
wild type	100	4
strain L — leader (attenuator) sequence deleted; promoter, operator and repressor all normal	100	35

(a) State what the **wild-type** results show about the effect of tryptophan on this operon, and name the control mechanism that acts at the operator.

(b) Strain L still lowers its enzyme level when tryptophan is high. Explain why.

(c) Explain why strain L's enzyme level in high tryptophan is much higher than the wild type's, and state what this shows about how the two control mechanisms normally work together. *(7 marks)*

Answers

Q1. (a) An exon is an expressed (coding) region of a gene, kept in the mature mRNA. (b) An intron is a non-coding region, spliced out and not in the mature mRNA. (c) A promoter is the region RNA polymerase binds to in order to begin transcription. **Q2.** (a) The operator — an operator is a prokaryotic sequence, and a human gene is eukaryotic. (b) Prokaryotes (bacteria), in an operon; the promoter lies on its 5′ side and the structural genes on its 3′ side. (c) The promoter (never transcribed) and intron 1 (spliced out). **Q3.** (a) A protein that carries out a function in the cell (e.g. an enzyme). (b) A protein that controls the expression of other genes (e.g. a repressor). (c) A regulatory gene. **Q4.** (a) Two. (b) No — trp-tRNA is plentiful, so the ribosome runs straight past them. (c) A terminator hairpin forms and RNA polymerase detaches part-way through the leader, so the structural genes are not transcribed. **Q5.** (a) Inactive. (b) No. (c) RNA polymerase binds the promoter and transcribes the structural genes; the cell can then make the enzymes that build tryptophan (and so make tryptophan). **Q6.** (a) Making proteins costs energy and materials, so a cell makes a protein only when it is needed. (b) A group of structural genes with a related job controlled together by one promoter and one operator. (c) All the enzymes of the pathway are switched on and off together from one switch, so the cell makes the complete set only when the pathway is needed. **Q7.** An exon is an expressed, coding region kept in the mature mRNA; an intron is a non-coding region spliced out and not in the mature mRNA. **Q8.** The promoter is where RNA polymerase binds to begin transcription; the operator is a regulatory sequence a repressor can bind to, stopping RNA polymerase binding the promoter. The promoter starts transcription; the operator controls (switches off) transcription. **Q9.** A structural gene codes for a protein that does a job in the cell (e.g. an enzyme); a regulatory gene codes for a protein that controls the expression of other genes (e.g. a repressor). **Q10.** Both act when tryptophan is high and both stop the trp-synthesis enzymes being made, saving resources. Repression: tryptophan acts as a corepressor, activating the repressor, which binds the operator so RNA polymerase cannot bind the promoter — transcription never starts. Attenuation: transcription does start, but the ribosome does not stall at the leader's two trp codons, a terminator hairpin forms and RNA polymerase detaches part-way through the leader — transcription stops after it has begun. **Q11.** With little tryptophan there is no corepressor, so the repressor stays inactive and cannot bind the operator; RNA polymerase binds the promoter and transcribes the structural genes (operon ON); the enzymes are made and the cell synthesises its own tryptophan. **Q12.** The trp-synthesis enzymes are made only when tryptophan is scarce and not when it is plentiful, so the cell does not waste energy or raw materials making a product it already has. **Q13.** Repressible = normally ON, switched OFF when the end product accumulates. A rise in tryptophan activates the repressor, switching the operon off. **Q14.** The genes are not transcribed even when tryptophan is low — the operon stays OFF, because the mutant repressor binds the operator without needing a corepressor. The cell cannot make the trp-synthesis enzymes and so cannot make its own tryptophan; it can only grow if tryptophan is supplied from outside. **Q15.** No trp-synthesis enzymes are made, even when tryptophan is low, because RNA polymerase cannot bind the mutated promoter and so cannot transcribe the genes; the cell cannot make its own tryptophan. **Q16.** (a) Region 1 = promoter; region 2 = operator. (b) Structural genes — they code for enzymes that carry out a function (building amino acid X). (c) OFF; abundant amino acid X acts as a corepressor, activating the repressor, which binds the operator so RNA polymerase cannot bind the promoter. **Q17.** (a) As tryptophan concentration increases, the amount of trp-synthesis enzyme decreases (an inverse relationship). (b) High tryptophan activates the repressor, which binds the operator and switches the operon off, so little enzyme is made; low tryptophan leaves the operon on, so a lot of enzyme is made. (c) About 5 units or slightly lower (very low) — the operon is already switched off at 100 units, so raising tryptophan further keeps it off and enzyme stays minimal. **Q18.** Error: tryptophan switches the operon OFF, not on. Correct: tryptophan switches the trp operon off, because when tryptophan is present it activates the repressor, which binds the operator so RNA polymerase cannot bind the promoter. **Q19.** (a) Tryptophan switches the operon off — enzyme falls from 100 to 4 relative units when tryptophan is high (an inverse relationship); the mechanism acting at the operator is repression. (b) Strain L has a normal operator and repressor, so repression still works: high tryptophan activates the repressor, which binds the operator so RNA polymerase cannot bind the promoter. (c) Strain L has no leader, so attenuation cannot occur; transcription that escapes repression runs straight through into the structural genes instead of stopping in the leader, so more enzyme is made (35 vs 4). The two mechanisms normally work together, attenuation catching the transcription repression lets through.

Fully-worked solutions

Q1. (a) An **exon** is an **expressed (coding) region** of a gene; it is kept in the mature mRNA, so its code reaches the polypeptide. (b) An **intron** is a **non-coding region** lying between exons; it is spliced out during RNA processing and

is **not** present in the mature mRNA. (c) A **promoter** is the control region at the start of a gene that **RNA polymerase binds to in order to begin transcription**.

Q2. (a) The **operator** should not appear. An operator is a **prokaryotic** regulatory sequence that forms part of an operon; a human gene is **eukaryotic**, so it has no operator. (b) Operators are found in the genes of **prokaryotes** (bacteria), within an **operon**: the **promoter** lies immediately on its 5′ side and the **structural genes** immediately on its 3′ side. (c) The corrected gene reads promoter – exon 1 – intron 1 – exon 2. The **promoter** is not present in the mature mRNA because it is never transcribed, and **intron 1** is not present because it is **spliced out** during RNA processing. Only the exons remain.

Q3. (a) A **structural gene** codes for a protein that **carries out a function in the cell**, such as an enzyme or a structural protein. (b) A **regulatory gene** codes for a protein that **controls the expression of other genes**, such as a repressor. (c) In the trp operon, the repressor is coded for by a **regulatory gene**.

Q4. (a) The leader contains **two** tryptophan codons. (b) **No** — when tryptophan is high there is plenty of charged **trp-tRNA**, so the ribosome has the tryptophan it needs at those codons and runs **straight past them** without stalling. (c) Because the ribosome does not stall, a **terminator hairpin** forms in the leader mRNA and **RNA polymerase detaches part-way through the leader**, so transcription stops before it reaches the structural genes and the trp-synthesis enzymes are not made.

Q5. (a) The repressor is **inactive**. (b) **No** — without tryptophan as a corepressor the repressor cannot bind the operator. (c) RNA polymerase **binds the promoter and transcribes the structural genes**, so the cell makes the **enzymes that build tryptophan**, and can therefore **make its own tryptophan**.

Q6. (a) Making a protein **costs energy and raw materials**, so a cell makes a given protein **only when it is needed** rather than all the time. (b) An **operon** is a group of structural genes with a related job, controlled together by a **single promoter and operator** and transcribed as one unit. (c) The enzymes of one pathway are needed **at the same time and in the same circumstances**, so putting them under a **single promoter and operator** lets the cell transcribe them **together as one message** and switch the whole pathway on or off with **one switch**. The cell therefore makes the complete set of enzymes only when the pathway is needed, and none of them when it is not, instead of having to regulate each gene separately.

Q7. An **exon** is an **expressed (coding) region** of a gene that is **kept in the mature mRNA**, so its code appears in the polypeptide. An **intron** is a **non-coding region** that is **spliced out** and is **not** present in the mature mRNA. *(1 mark each side, congruent: coding + kept vs non-coding + removed.)*

Q8. The **promoter** is the sequence that **RNA polymerase binds to in order to begin transcription** — it marks where transcription starts. The **operator** is a **regulatory sequence** that a **repressor protein can bind to**; when the repressor is bound it **stops RNA polymerase binding the promoter**, so transcription does not begin. The roles differ in that the **promoter starts** transcription, whereas the **operator controls (can switch off)** transcription. *(1 mark: promoter = RNA polymerase binding / start; 1 mark: operator = repressor binding site; 1 mark: the difference — start vs control/off.)*

Q9. A **structural gene** codes for a protein that **carries out a function in the cell** (for example an enzyme such as one of the tryptophan-building enzymes). A **regulatory gene** codes for a protein that **controls the expression of other genes** (for example the trp repressor). The distinction is the **job of the protein**: doing the cell's work versus controlling other genes.

Q10. In **repression**, high tryptophan acts as a **corepressor**: it binds the inactive repressor and changes its shape to the **active** form, which then **binds the operator**. With the repressor there, **RNA polymerase cannot bind the promoter**, so transcription of the structural genes **never begins**. In **attenuation**, RNA polymerase **does** begin transcribing, but the **leader (attenuator)** sequence carries **two trp codons**; with tryptophan plentiful there is ample charged **trp-tRNA**, so the ribosome translating the leader **does not stall**, a **terminator hairpin** forms and RNA polymerase **detaches part-way through the leader**, before reaching the structural genes. The two are **similar** in that both are triggered by **high tryptophan** and both stop the trp-synthesis enzymes being made, saving the cell energy and raw materials. They **differ** in when and where they act: repression prevents transcription from **starting**, at the operator, whereas attenuation stops transcription that has **already started**, in the leader. *(1 mark: repression — corepressor activates repressor → binds operator → RNA polymerase cannot bind the promoter; 1 mark: attenuation — no ribosome stall*

→ terminator hairpin → RNA polymerase detaches in the leader; 1 mark: similarity — both act when trp is high and both reduce enzyme production; 1 mark: difference — before transcription begins vs after it has begun.)

Q11. When tryptophan is **low**, there is little or no tryptophan to act as a corepressor, so the **repressor stays inactive** and **cannot bind the operator**. With the operator **free**, **RNA polymerase binds the promoter and transcribes the structural genes** (operon **ON**). The **trp-synthesis enzymes are made**, and the cell is able to **synthesise its own tryptophan**. *(1 mark: no corepressor, repressor inactive; 1 mark: repressor cannot bind the operator; 1 mark: RNA polymerase transcribes the genes; 1 mark: enzymes made / cell makes tryptophan.)*

Q12. The regulation means the trp-synthesis enzymes are produced **only when tryptophan is scarce** (operon on) and **not when tryptophan is plentiful** (operon off). Because making enzymes and tryptophan **costs energy and raw materials**, the cell **avoids wasting resources** on a product it already has enough of, only spending them when tryptophan is actually needed. *(1 mark: made only when trp scarce; 1 mark: not made when trp plentiful; 1 mark: saves energy/materials — efficiency.)*

Q13. A **repressible** operon is one that is **normally switched on** but can be **switched off** when its end product builds up. In the trp operon, when **tryptophan (the end product) accumulates** it acts as a corepressor, activating the repressor, which binds the operator and **switches the operon off**. So tryptophan represses its own production once there is enough of it.

Q14. The trp-synthesis genes are **not transcribed, even when tryptophan is low** — the operon is stuck **OFF**. Normally, low tryptophan means there is no corepressor, so the repressor stays inactive, the operator is **free**, and RNA polymerase can bind the promoter and transcribe. This **mutant repressor binds the operator without needing tryptophan**, so the operator is occupied at all tryptophan levels and **RNA polymerase cannot bind the promoter**. The consequence is that the cell **cannot make the trp-synthesis enzymes and so cannot synthesise its own tryptophan**; it can grow only if tryptophan is supplied from its surroundings. *(1 mark: genes not transcribed / operon stuck off even at low trp; 1 mark: mutant repressor binds the operator without a corepressor, so RNA polymerase cannot bind the promoter; 1 mark: cell cannot make its own tryptophan / depends on an external supply.)*

Q15. No trp-synthesis enzymes are made, even when tryptophan is low. Transcription can only begin if **RNA polymerase binds the promoter**; if the mutated promoter can no longer be bound by RNA polymerase, the structural genes **cannot be transcribed** at all, regardless of the operator or the repressor. So even in low tryptophan — when the operon would normally be on — the enzymes are not made and the cell **cannot synthesise its own tryptophan**. *(1 mark: no enzymes made / genes not transcribed; 1 mark: RNA polymerase cannot bind the promoter; 1 mark: linked to low-tryptophan condition — operon would normally be on.)*

Q16. (a) Region **1** is the **promoter**; region **2** is the **operator**. (b) The genes at regions **3 and 4** are **structural genes**, because they **code for enzymes that carry out a function** (they build amino acid X) rather than controlling other genes. (c) The operon is **OFF**. When amino acid X is **abundant**, it acts as a **corepressor**: it binds the repressor and activates it, so the repressor binds the operator (region 2), and **RNA polymerase can no longer bind the promoter (region 1)**, so the structural genes are not transcribed. *(1 mark: promoter + operator; 1 mark: structural, with justification; 1 mark: OFF; 1 mark: corepressor activates repressor → binds operator → RNA polymerase cannot bind the promoter.)*

Q17. (a) There is an **inverse relationship**: as the tryptophan concentration **increases**, the amount of trp-synthesis enzyme **decreases** (from 100 units of enzyme at zero tryptophan down to 5 units at 100 units of tryptophan). (b) When tryptophan is **high**, it activates the repressor, which binds the operator and **switches the operon off**, so **little enzyme** is made (attenuation reinforces this by ending any transcription that does start early in the leader). When tryptophan is **low**, the repressor is inactive and the operon is **on**, so **a lot of enzyme** is made. (c) The enzyme level would stay **very low — about 5 units or a little lower**. At 100 units of tryptophan the operon is already switched off, so raising tryptophan to 200 units keeps the repressor active and the operon off, and the enzyme level cannot fall much further than it already has. *(1 mark: inverse / directional trend; 1 mark: high trp → operon off → little enzyme; 1 mark: low trp → operon on → much enzyme; 1 mark: prediction ~5 units with justification that the operon is already off.)*

Q18. The **error** is that tryptophan switches the operon **on** — in fact tryptophan switches the trp operon **OFF**. A correct version is: "**Tryptophan switches the trp operon off**, because when tryptophan is present it acts as a corepressor that **activates the repressor**; the active repressor **binds the operator, so RNA polymerase cannot bind**

the promoter, and the genes are **not** transcribed." *(1 mark: identifies that tryptophan switches the operon off, not on; 1 mark: correct mechanism — activates repressor → binds operator → RNA polymerase cannot bind the promoter.)*

Q19. (a) In the wild type, tryptophan **switches the operon off**: enzyme falls from **100 relative units with no tryptophan to 4 with high tryptophan**, an **inverse** relationship. The mechanism acting at the operator is **repression** — tryptophan acts as a corepressor, activating the repressor, which binds the operator so RNA polymerase cannot bind the promoter. (b) Strain L has lost only the **leader**; its **promoter, operator and repressor are all normal**, so **repression still operates**. When tryptophan is high the activated repressor still binds the operator and still prevents RNA polymerase binding the promoter, so most transcription is still shut down and the enzyme level falls (100 → 35). (c) Strain L has **no leader (attenuator) sequence**, so **attenuation cannot occur**: there are no trp codons for a ribosome to run past and no terminator hairpin can form. Any RNA polymerase that **escapes repression** — for example when the repressor momentarily detaches from the operator — therefore transcribes **straight through into the structural genes** instead of detaching in the leader, so far more enzyme is made (**35 vs 4**). This shows the two mechanisms normally **work together**: repression is the coarse switch at the operator, and **attenuation catches the transcription that repression lets through**, so the two in combination shut the operon down far more tightly than repression alone. *(a: 2 marks — inverse relationship / operon off when trp high, quoting the data; repression named. b: 2 marks — operator and repressor still normal so repression still occurs; activated repressor binds operator → RNA polymerase cannot bind promoter. c: 3 marks — no leader so no attenuation / no terminator hairpin; transcription escaping repression runs on into the structural genes, so more enzyme (35 vs 4); the two mechanisms normally act together, attenuation backing up repression.)*