DP IB Biology: SL

Your notes

2.6 Transcription & Translation

Contents

2.6.1 Transcription

Your notes

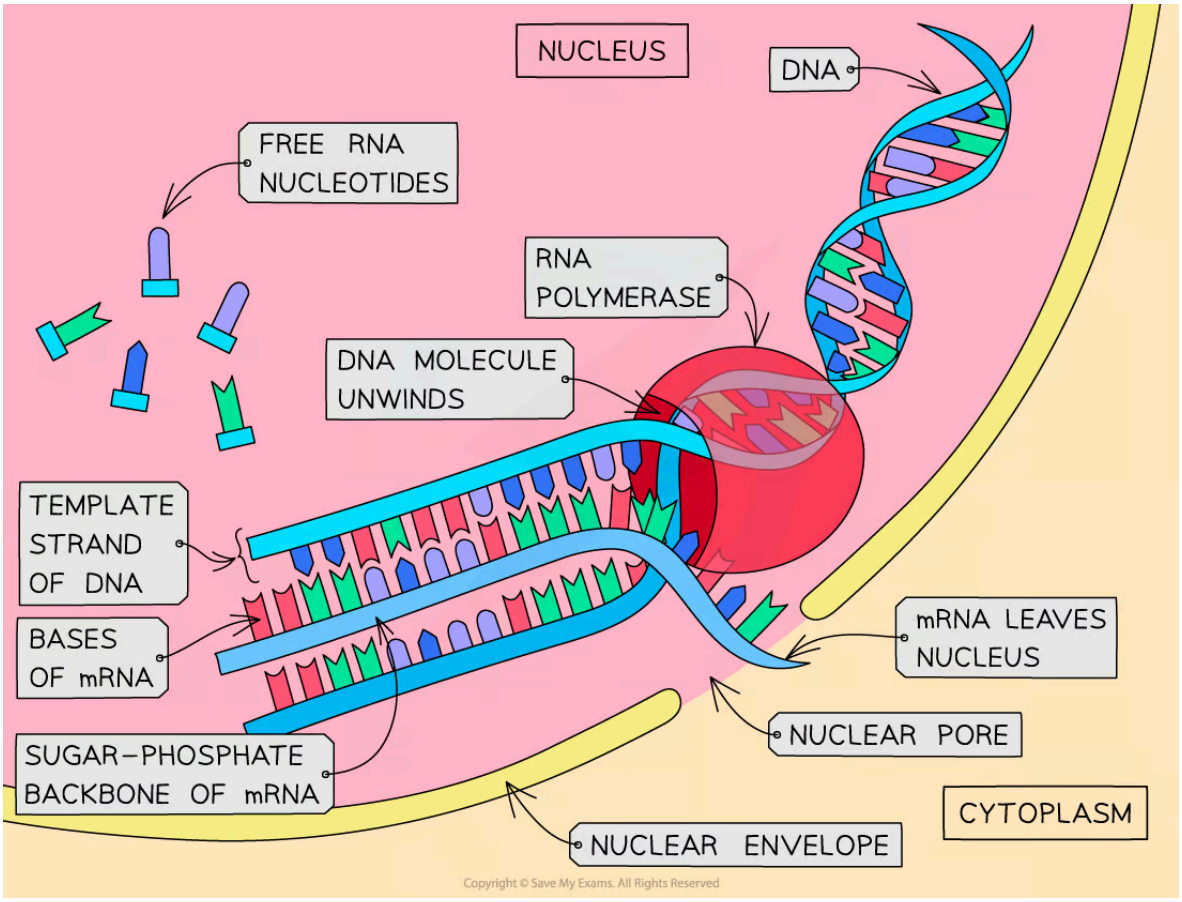
Transcription

- This process of protein synthesis occurs in **two stages**:
  - **Transcription** – **DNA** is transcribed and an **mRNA** molecule is produced
    - mRNA is a single stranded RNA molecule that transfers the information in DNA from the nucleus into the cytoplasm
    - mRNA production requires the enzyme RNA polymerase
  - **Translation** – **mRNA** (messenger RNA) is translated and an **amino acid sequence** is produced

The process of transcription

- This stage of protein synthesis occurs **in the nucleus** of the cell
- Part of a DNA molecule **unwinds** (the **hydrogen bonds** between the complementary base pairs **break**)
- This exposes the **gene** to be transcribed (the gene from which a particular polypeptide will be produced)
- A complementary copy of the code from the gene is made by building a **single-stranded nucleic acid molecule known as mRNA** (messenger RNA)
- **Free RNA nucleotides** pair up (via hydrogen bonds) with their complementary (now exposed) bases on one strand (the template strand) of the 'unzipped' DNA molecule
- The sugar-phosphate groups of these RNA nucleotides are then **bonded** together by the enzyme **RNA polymerase** to form the sugar-phosphate backbone of the mRNA molecule
- When the gene has been transcribed (when the mRNA molecule is complete), the hydrogen bonds between the mRNA and DNA strands break and the **double-stranded DNA molecule re-forms**
- The mRNA molecule then **leaves the nucleus** via a pore in the nuclear envelope
  - This is where the term *messenger* comes from - the mRNA is despatched, **carrying a message**, to another part of the cell
  - DNA can't make this journey; **it's too big to fit** through the pores in the nuclear envelope

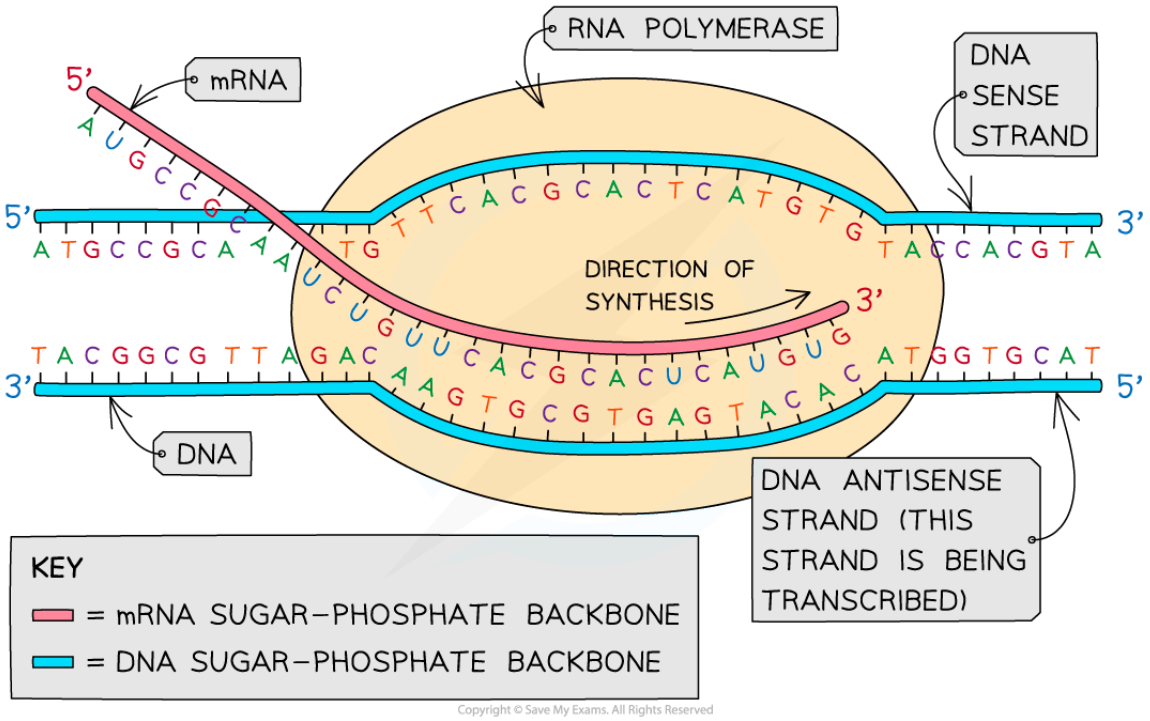
Your notes



DNA is transcribed and an mRNA molecule is produced

Sense and anti-sense strands

- In the **transcription** stage of protein synthesis, free RNA nucleotides pair up with the exposed bases on the DNA molecule but **only with those bases on one strand of the DNA molecule**
- The RNA will have a complementary base sequence to the DNA strand (with the substitution of Thymine with Uracil)
- The strand of the DNA molecule that carries the genetic code is called the **sense strand**
- The opposite DNA strand is called the **antisense** strand
- To get **an RNA transcript of the sense strand**, the **antisense strand is the one that is transcribed** to form the mRNA molecule
  - This mRNA molecule will later be translated into an amino acid chain



The antisense strand of the DNA molecule is the one that is transcribed

Analogy: Think of transcription and translation as being like converting between languages

- Each language has its **alphabet**, just as nucleic acids and proteins have their **monomers**
- **Transcription** is like converting text from **English** to **French**
  - The same characters are used, but there are slight differences
  - French uses the same alphabet as English but employs occasional accented characters like â, é, or ç
  - DNA and RNA employ largely the same monomers, but with the slight differences of the two pentose sugars and of U replacing T.
- **Translation** is like converting text from a western language to a language that uses a different alphabet, like **Japanese**
  - A completely **different set of characters** is used
  - The sequence of characters is **unrecognisable** from the original
  - If we could see them, a chain of amino acids would look nothing like a chain of nucleotides

Transcription and Translation Can be Likened to Conversion Between Languages Table

Your notes

Transcription	DNA → RNA	Similarities	English → French		Similarities
DNA → RNA	TTACAGCTC → AAUGUCGAG	Both use a similar set of monomers (with a slight difference; U replaces T)	"I received biology lessons at my school"	"J'ai reçu des cours de biologie à mon école"	Both use a similar alphabet (with slight differences: ç,à,é,Ô etc)

Translation	RNA → Protein	Differences	French → Japanese		Differences
RNA → protein	AAUGUCGAG → Asn–Val–Glu	Both use different monomers (nucleotides & amino acids)	"J'ai reçu des cours de biologie à mon école"	学校で生物学の授業を受けました	Both use different alphabets

Exam Tip

Be careful – DNA polymerase is the enzyme involved in DNA replication; RNA polymerase is the enzyme involved in transcription – don't get these confused.

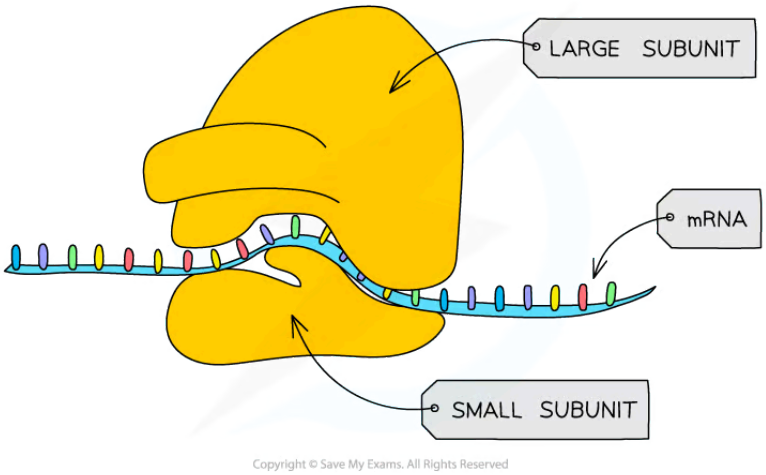
Your notes

2.6.2 Translation

Your notes

Translation

- Translation is the synthesis of polypeptides on ribosomes
- This stage of protein synthesis occurs **in the cytoplasm** of the cell
- After leaving the nucleus, the **mRNA molecule attaches to a ribosome**
- A ribosome is a complex structure that is made of a large and small subunit
  - Ribosomes are themselves made of **proteins** and **RNA** (called ribosomal RNA or **rRNA**)
- There are **binding sites on the subunits** for the various other molecules involved in translation



A ribosome is built of large and small subunits, ribosomal RNA and an area on the surface that catalyses the formation of peptide bonds in a newly-synthesised protein

Exam Tip

Make sure you learn both stages of protein synthesis fully. Don't forget WHERE these reactions take place – transcription occurs in the nucleus but translation occurs in the cytoplasm!

Genetic Code & mRNA

- The amino acid sequence of polypeptides is determined by mRNA according to the genetic code
- mRNA **varies in length**, depending on the length of the gene, but is around 2,000 nucleotides long, on average (in mammals)
- Only certain genes are transcribed in a particular cell, depending on the function of that cell
  - The gene for **rhodopsin** (a light-sensitive protein in the eye) is transcribed to mRNA in **retina cells**, but not transcribed in other body cells where rhodopsin is not required; that would be **a waste of cellular energy**

Exam Tip

Most RNA exists as mRNA but don't forget the other types; transfer RNA (tRNA) and ribosomal RNA (rRNA).

Your notes

Codons

Your notes

- Codons of **three bases** on mRNA correspond to **one amino acid** in a polypeptide
- The four nucleotide bases in mRNA are not enough to code for **20 separate amino acids**
- Pairs of nucleotides would only give 16 combinations ($4^2 = 16$), which is **still not enough**
- Triplets of nucleotides would yield **64 combinations** ($4^3 = 64$), which is **more than enough**
- Different triplets code for **the same amino acid**, giving **some protection against mutation**
  - A **triplet** is a sequence of three DNA bases that codes for a specific amino acid
  - A **codon** is a sequence of three **mRNA** bases that codes for a specific amino acid
  - A codon is transcribed from the triplet and is complementary to it
- An **anticodon** is a sequence of three **transfer RNA** (tRNA) bases that are complementary to a codon
  - The transfer RNA **carries the appropriate amino acid** to the ribosome
  - The amino acid can then be condensed **onto the growing polypeptide chain**
- Certain codons carry the command to **stop translation** when the polypeptide chain is complete ('**Stop codons**')

mRNA Codons and Amino Acids Table

FIRST LETTER	SECOND LETTER: U	SECOND LETTER: C	SECOND LETTER: A	SECOND LETTER: G	THIRD LETTER
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G
A	AUU, AUC, AUA: Ile AUG: Met	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn AAA, AAG: Lys	AGU, AGC: Ser AGA, AGG: Arg	U C A G
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp GAA, GAG: Glu	GGU, GGC, GGA, GGG: Gly	U C A G

Worked example

Use the **rules of base-pairing** and the **mRNA Codons and Amino Acids Table** (above) to deduce the amino acid sequence coded for by the following DNA **sense strand** sequenceTTC GAG CAT TAC GCC

Step 1: Work out the antisense sequence using A-T and C-G base pairing rules

AAG CTC GTA ATG CGG

Step 2: Work out the mRNA codons, complementary to the antisense strand

UUC GAG CAU UAC GCC

Step 3: Use the mRNA Codons and Amino Acids Table (above) to work out the first amino acid

First base in codon = U, second base = U, third base = C

So we're looking in the top-left box of the table; this amino acid is **Phe**

Step 4: Repeat for the remaining 4 codons

GAG = Glu

CAU = His

UAC = Tyr

GCC = Ala

Answer: The final sequence of amino acids is Phe-Glu-His-Tyr-Ala

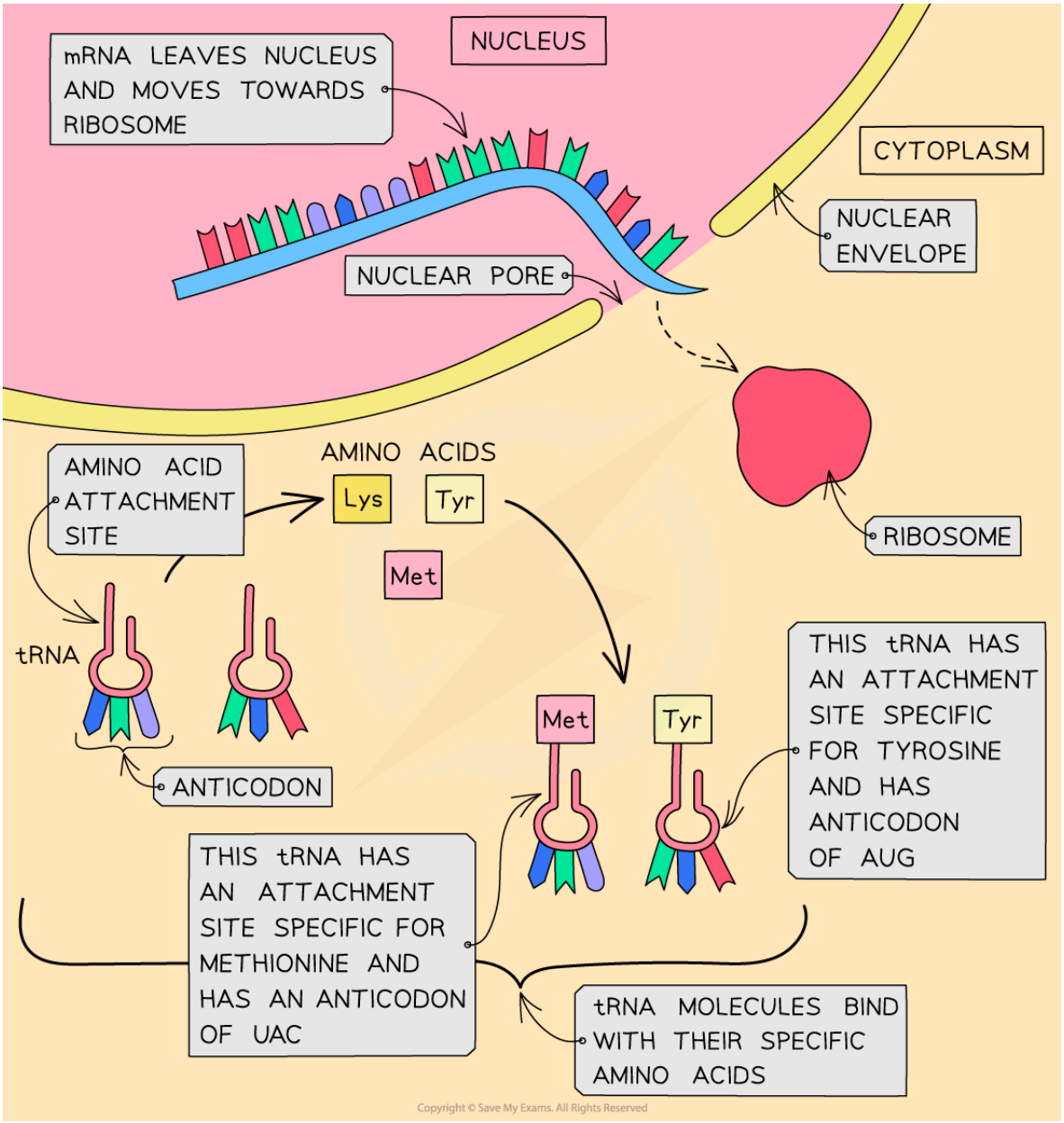
Your notes

Codons & Anticodons


Your notes

- Translation depends on complementary base pairing between codons on mRNA and anticodons on tRNA
- In the cytoplasm, there are **free molecules of tRNA** (transfer RNA)
- The **tRNA molecules bind with their specific amino acids** (also in the cytoplasm) and bring them to the mRNA molecule on the **ribosome**
- The triplet of bases (anticodon) on each tRNA molecule pairs with a complementary triplet (codon) on the mRNA molecule
- **Two tRNA molecules fit onto the ribosome at any one time**, bringing the amino acid they are each carrying side by side
- A **peptide bond** is then formed (by condensation) between the two amino acids
  - The formation of a peptide bond between amino acids is an **anabolic** reaction
  - It **requires energy,** in the form of **ATP**
  - The ATP needed for translation is provided by the **mitochondria** within the cell
- This process continues until a **'stop' codon** on the mRNA molecule is reached – this acts as a signal for translation to stop and at this point the amino acid chain coded for by the mRNA molecule is complete
- This amino acid chain then **diffuses away** from the ribosome and forms the final polypeptide

Your notes



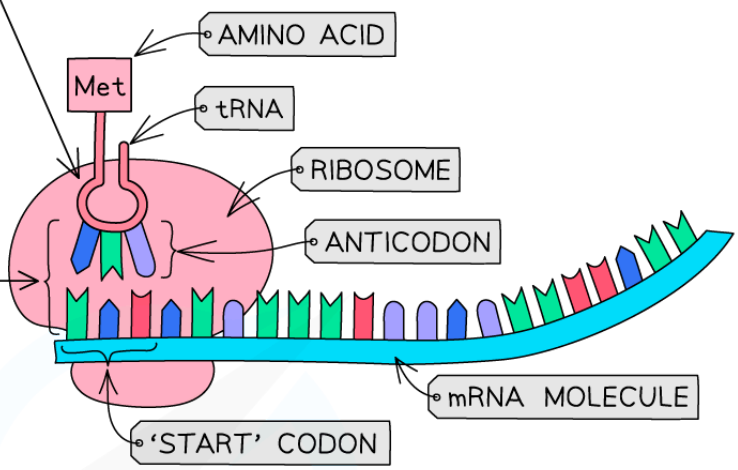
The translation stage of protein synthesis – tRNA molecules bind with their specific amino acids

Your notes

1 IN THE CYTOPLASM THE mRNA ATTACHES TO A RIBOSOME

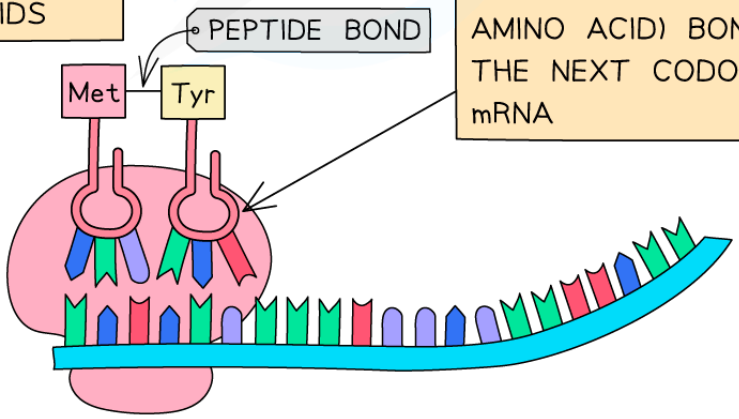
2 EACH tRNA HAS THE COMPLEMENTARY **ANTICODON** TO THE CODON ON THE mRNA

3 THE FIRST tRNA (WHICH ALWAYS CARRIES THE METHIONINE AMINO ACID) FORMS HYDROGEN BONDS WITH THE FIRST OR **'START' CODON** (AUG) ON THE mRNA.



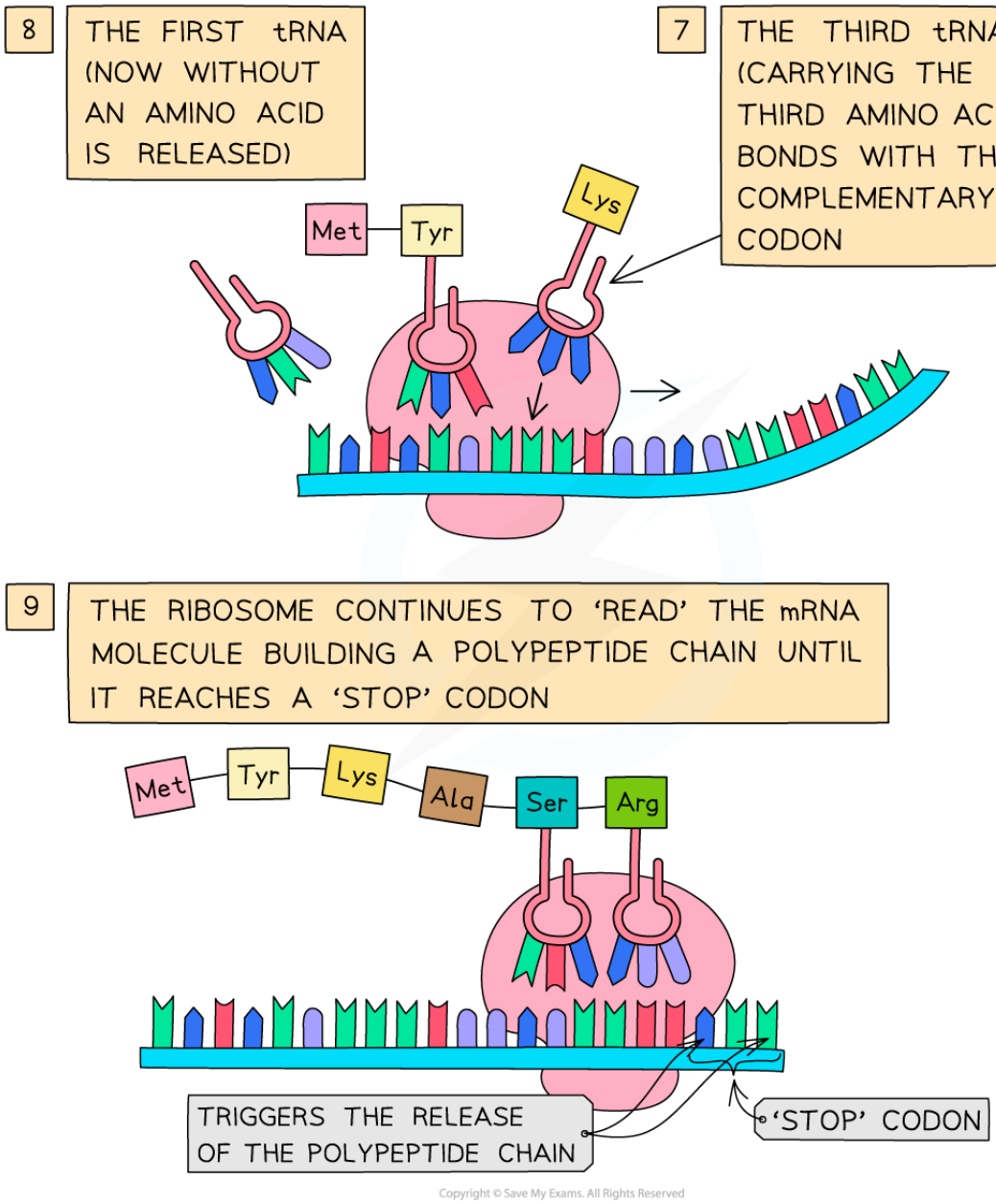
5 A PEPTIDE BOND FORMS BETWEEN THE AMINO ACIDS

4 THE SECOND tRNA (BRINGING THE SECOND AMINO ACID) BONDS WITH THE NEXT CODON ON THE mRNA



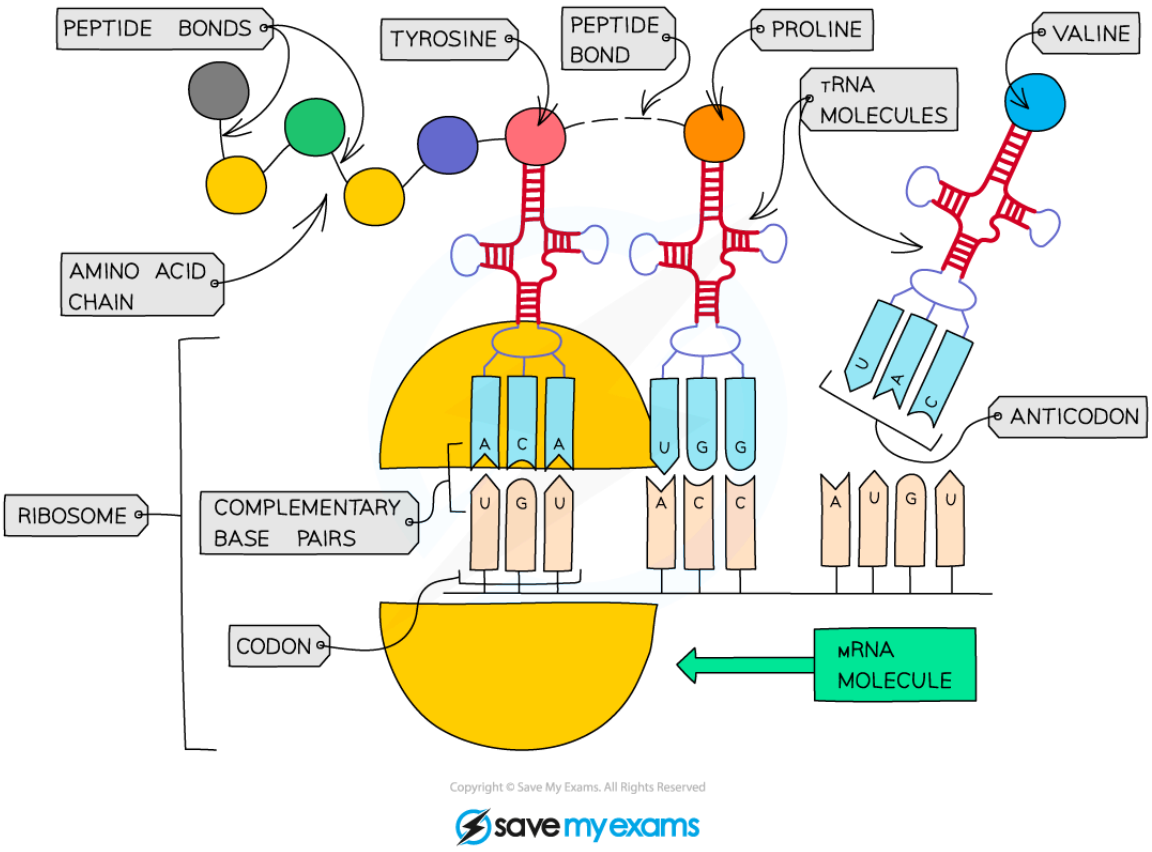
6 THE RIBOSOME MOVES ALONG THE mRNA (IN A 5' TO 3' DIRECTION) 'READING' THE NEXT CODON

Your notes



The translation stage of protein synthesis – an amino acid chain is formed

Your notes



An polypeptide forms as peptide bonds are added in sequence

2.6.3 Biotechnology

Your notes

Polymerase Chain Reaction

Use of Taq DNA polymerase to produce multiple copies of DNA rapidly by the polymerase chain reaction (PCR)

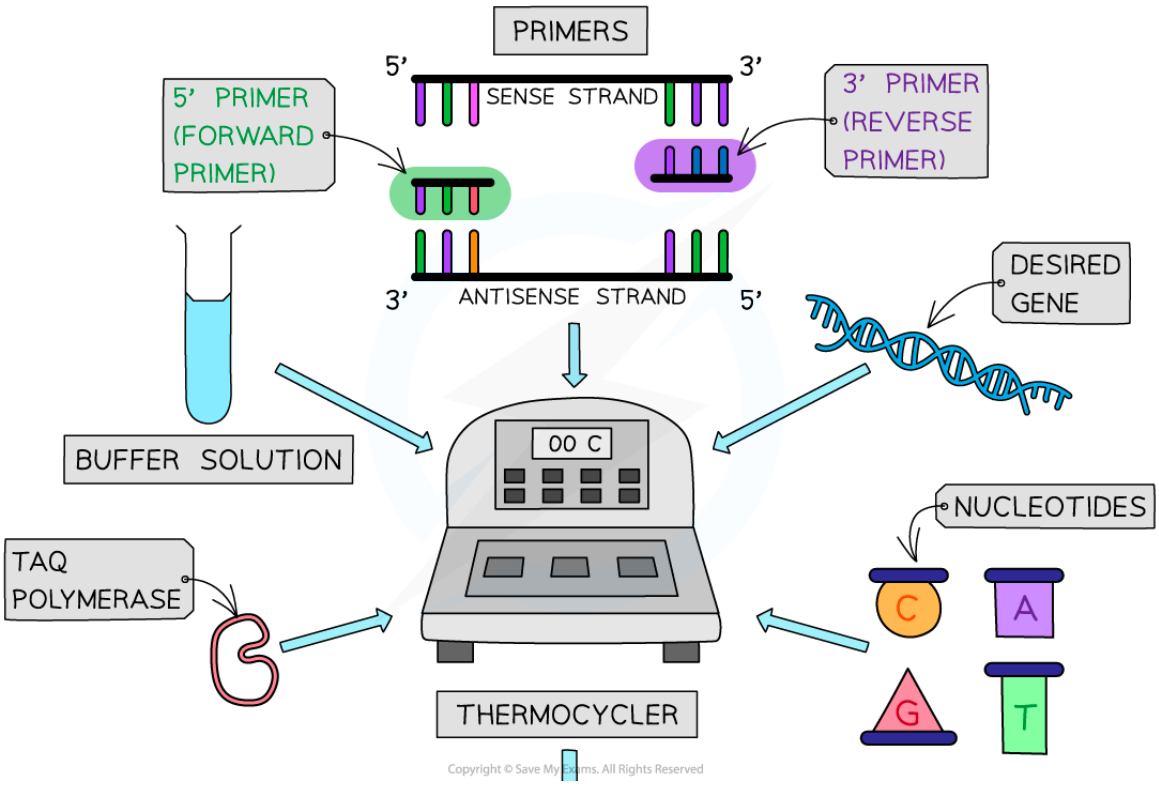
- **Polymerase chain reaction** (PCR) is a common molecular biology technique used in most applications of gene technology, for example, DNA profiling (eg. identification of criminals and determining paternity) or genetic engineering
  - PCR is also used in routine **COVID-19 testing** to detect and amplify small amounts of viral RNA
- It can be described as the ***in vitro* method of DNA amplification**
- It is used to produce **large quantities** of specific fragments of DNA or RNA from very small quantities (even just one molecule of DNA or RNA)
- By using PCR scientists can have **billions of identical copies** of the DNA or RNA sample within a few hours
- The PCR process involves **three key stages** per cycle
- In each cycle, DNA is doubled so, in a standard run of 20 cycles, 1 million DNA molecules are produced. The three stages are undertaken in a PCR instrument (or **thermal cycler**) which automatically provides the **optimal temperature** for each stage and controls the **length of time** spent at each stage

The process of PCR

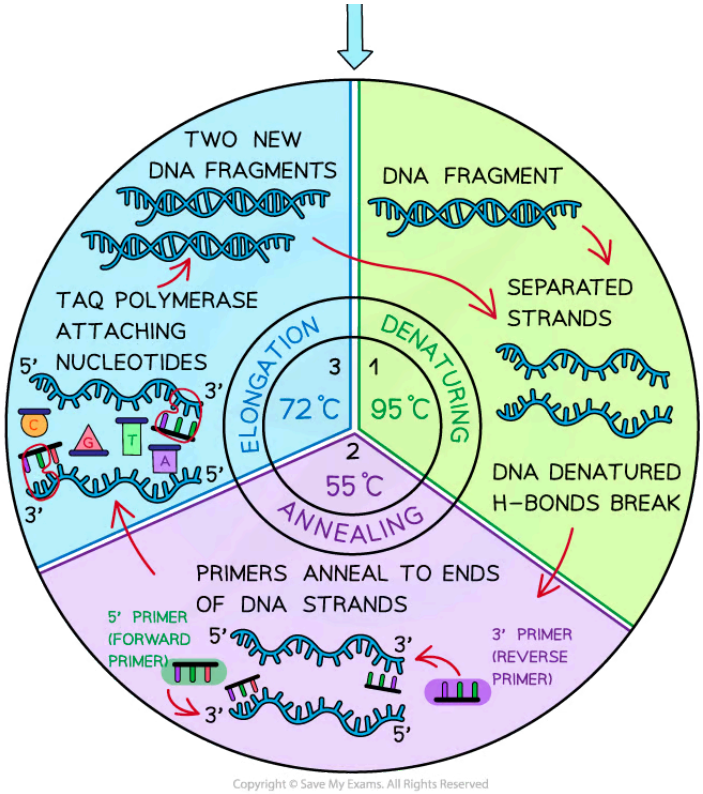
- Each PCR reaction requires:
  - **Target DNA** or **RNA** that is being amplified
  - **Primers** (forward and reverse) – these are short sequences of single-stranded DNA that define the region that is to be amplified by showing the DNA polymerase **where to begin building** the new strands
  - **DNA polymerase** – is the enzyme used to build the new DNA or RNA strand.
    - The most commonly used polymerase is ***Taq* polymerase** as it comes from a thermophilic bacterium *Thermus aquaticus*
    - This bacterium lives in **hot springs** in geothermal areas
    - Taq polymerase **does not denature** at the high temperature involved during the first stage of the PCR reaction
    - The enzyme's optimum temperature is **high enough to prevent annealing** of the DNA strands that have not been copied yet
  - **Free nucleotides** – used in the construction of the DNA or RNA strands
  - **Buffer solution** – to provide the optimum pH for the reactions to occur in
- The three stages are:
  - **Denaturation** – the double-stranded DNA is heated to 95°C for 15 seconds, which breaks the hydrogen bonds that hold the two DNA strands together
  - **Annealing** – the temperature is decreased to between 50 - 60°C so that primers (forward and reverse ones) can attach to the ends of the single strands of DNA by hydrogen bonding

- **Elongation / Extension –** the temperature is increased to 72°C for at least a minute
  - This is the optimum temperature for *Taq* polymerase to build the complementary strands of DNA
  - To produce the new identical double-stranded DNA molecules
- The three stages of a cycle take 2–3 minutes, so many cycles can be completed in a short space of time

Your notes



Your notes



The substances required for the Polymerase Chain Reaction to occur and the three key stages of the reaction

Exam Tip

It is important to know the three stages and the temperatures the reactions occur at during the different stages. You must also know why the *Taq* polymerase is used in PCR.

Production of Human Insulin


Your notes

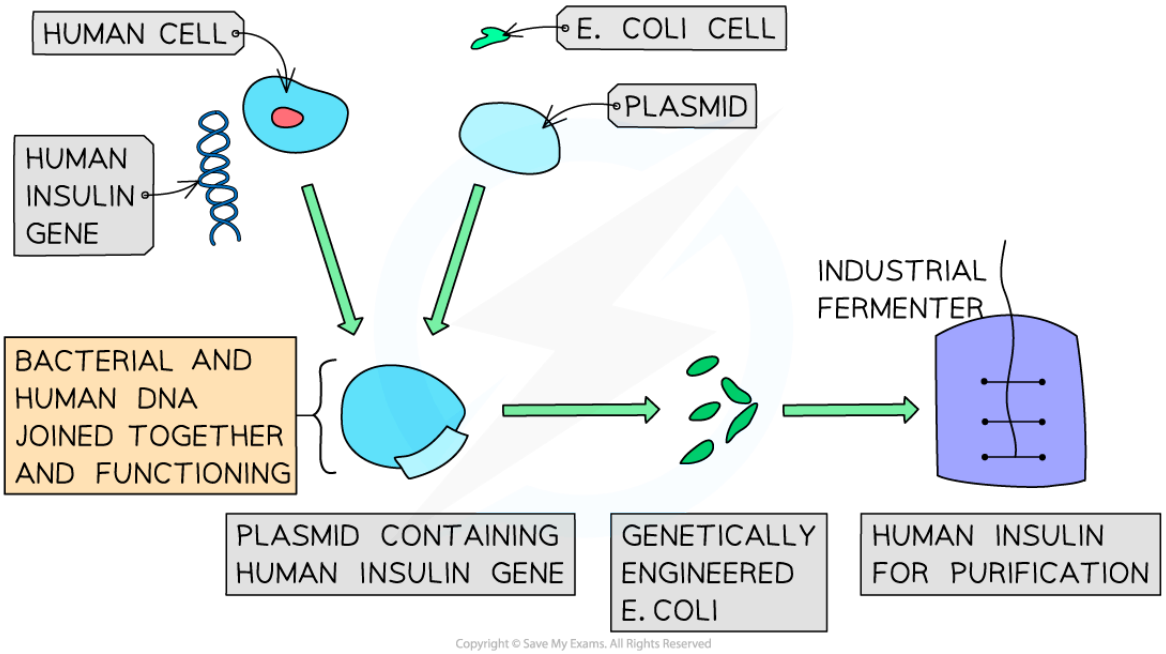
Production of human insulin in bacteria as an example of the universality of the genetic code allowing gene transfer between species

- Prior to the mid-1980s, some insulin-dependent diabetics would need to inject **pig or cattle insulin** as a substitute for human insulin to control their blood sugars
  - Some diabetics developed an **allergy** so could not use it
- Human insulin can be produced by other organisms **by transferring the human insulin gene to them** for **large-scale expression** of that gene
- DNA can be transferred from a human to a prokaryote (eg. *E. coli*), a eukaryote single-celled organism (eg. yeast) or a plant (eg. safflower species)
  - All these organisms express the human insulin gene
  - The insulin produced can be **harvested for medical use**
- The fact that DNA can be transferred from one organism to another **across kingdoms** (and can still do the same job) demonstrates the **universality of the genetic code**
  - The presence of nucleotides is a marker between living and non-living entities
- This was an early successful example of **genetic modification**

Human and Bacterial DNA working together

- In 1982, **insulin** was the first genetically engineered human protein to be approved for use in **diabetes** treatment
- Bacterial **plasmids** are modified to incorporate the human insulin gene
- These genetically modified plasmids are then **inserted into *Escherichia coli***
- The newly-adjusted bacteria are isolated, purified and placed into large scale **fermenters** that provide **optimal conditions**
- The genetically engineered bacteria multiply by binary fission, and express the human protein - insulin, which is eventually extracted and purified
- The advantages for scientists to use genetically engineered insulin are:
  - It is **identical to human insulin**, unless modified to have different properties (eg. act faster, which is useful for taking immediately after eating or to act more slowly)
  - There is a **reliable supply available** to meet demand (no need to depend on the availability of meat stock)
  - **Fewer ethical**, **moral** or **religious** concerns (proteins are not extracted from cows or pigs)
  - **Fewer rejection problems** or **side effects** or **allergic reactions**
  - **Cheaper** to produce in large volumes
  - That it is useful for people who have **animal insulin intolerance**

Your notes



The production of human insulin; the combination of DNA from two widely different organisms demonstrates the universality of the genetic code

Exam Tip

The details of the steps of production of human insulin are not required here. The main learning is that DNA and RNA are a universal code that applies to **all life forms**, as demonstrated by our ability to transfer genes successfully across species and kingdoms.

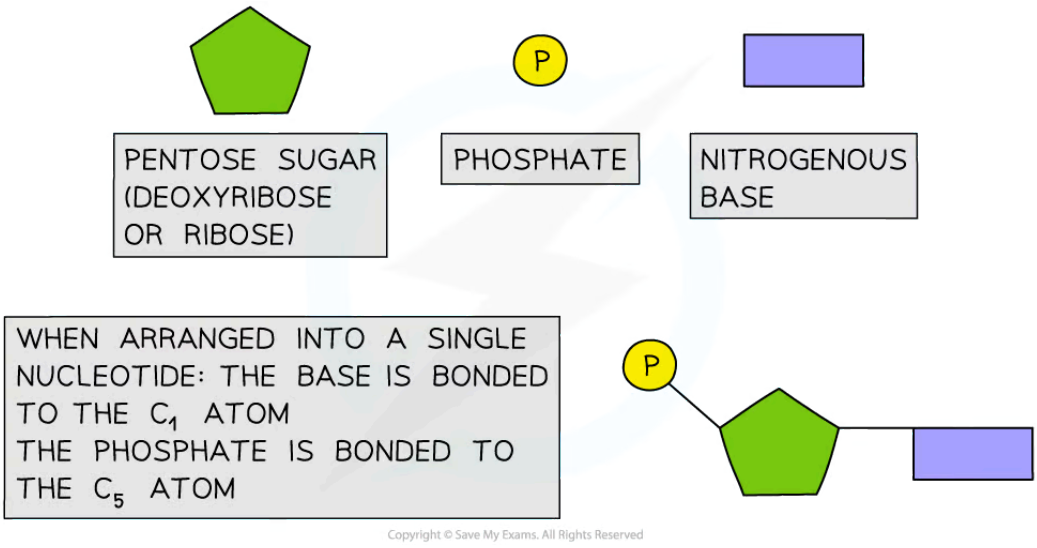
2.6.4 Skills: DNA, RNA & Protein Synthesis

Your notes

Drawing DNA & RNA Nucleotides & DNA Double Helix

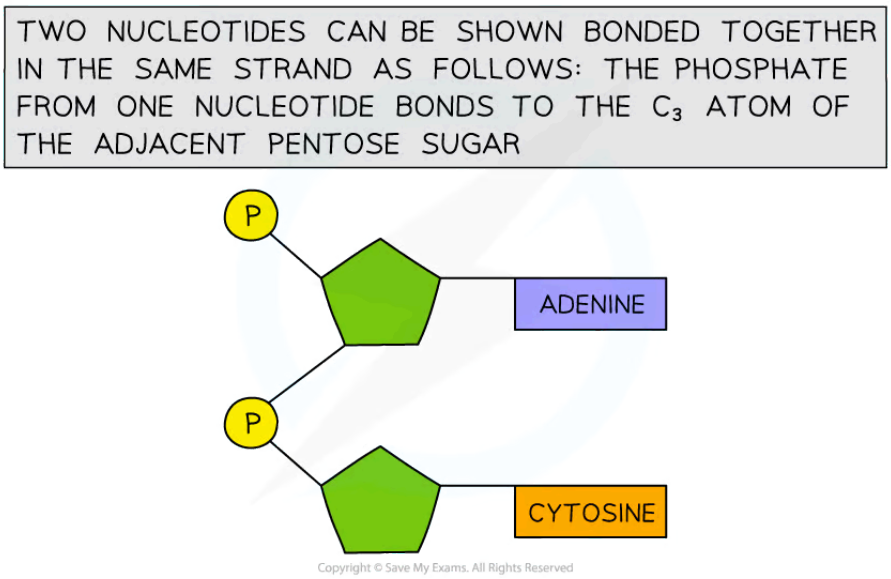
Drawing simple diagrams of the structure of single nucleotides of DNA and RNA

- **Simple shapes** can be used to draw the **main building blocks of nucleotides** and the DNA double helix
  - Advanced drawing skills are not required!
- **Pentagons** can represent **pentose sugars**
- **Circles** can represent **phosphates**
  - Often shown as a circle with the letter P inside: Ⓟ
- **Rectangles** can represent **bases**
- **Covalent bonds** can be shown with **solid lines**
- **Hydrogen bonds** can be shown with **dashed lines**
  - Or with complementary shapes that fit together (see diagrams)

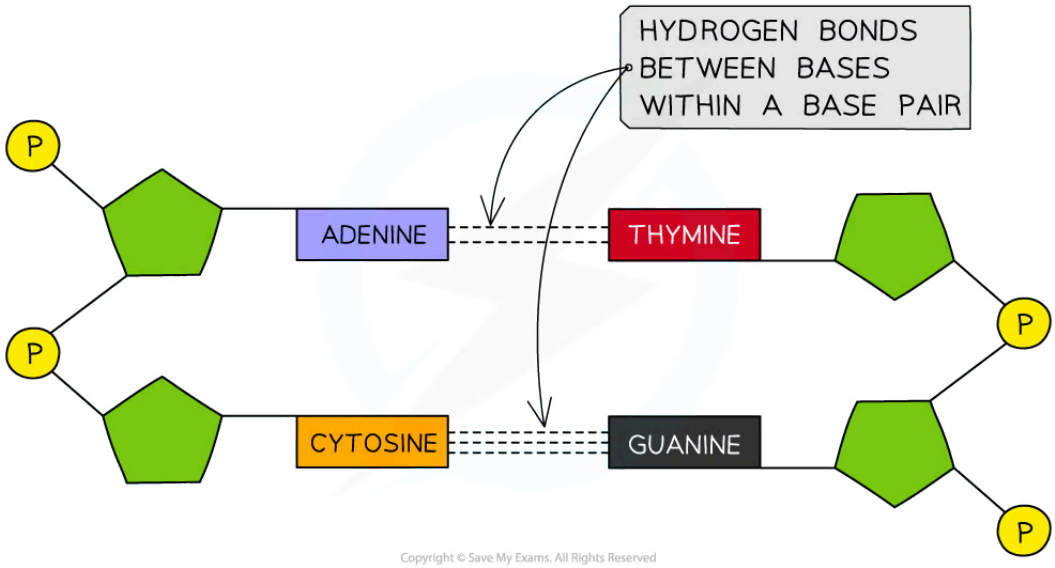


Simple shapes can be used to represent parts of nucleotide molecules

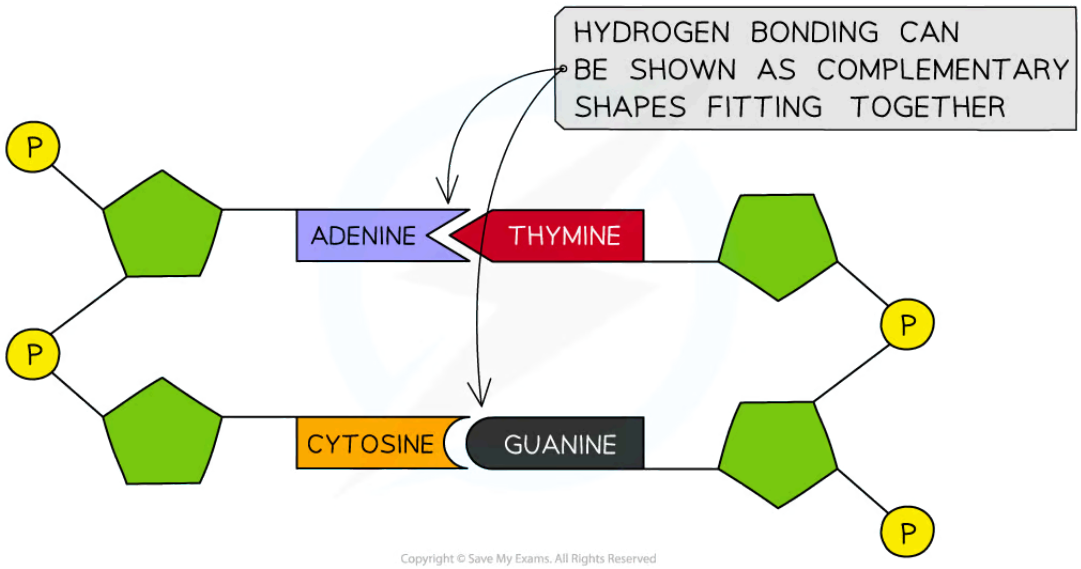
Your notes



Two nucleotides shown bonded together covalently within a strand



When drawing the base pairing, the opposite strand should be antiparallel to the first. The presence of hydrogen bonding is shown, but the numbers/lengths of bonds is not required



An alternative way to draw a DNA strand is to use complementary shapes for the bases

Exam Tip

Simple, hand-drawn shapes will suffice in an exam. Expert tip - a **large** drawing is always easier for an examiner to read (and award marks for) than a small one!Read the question carefully; examiners often want a **whole nucleotide** to be identified in your diagram and to ensure your diagram includes **all 4 complementary bases.** You don't have to remember the number of hydrogen bonds between the bases.Also, remember to draw DNA strands as **antiparallel** (one upside-down versus the other) but you don't have to be able to draw a helix shape!

Your notes

2.6.5 Skills: Interpreting Sequences

Your notes

Determination of mRNA Base Sequence

- The **rules of base pairing** and the **table of mRNA codons** are needed to be able to convert between DNA sequences, mRNA base sequences and amino acid sequences
- A **triplet** is a sequence of three DNA bases that codes for a specific amino acid
- A **codon** is a sequence of three **mRNA** bases that codes for a specific amino acid
- A codon is transcribed from the triplet and is complementary to it
- When comparing the genetic code to amino acid sequences, **mRNA codons are often used**
- The four bases found in RNA molecules (adenine, uracil, cytosine and guanine) have the ability to form **64** different codons
- Multiple mRNA codons can encode the same amino acid
  - This means that a **change in the genetic code doesn't necessarily result in a change in the amino acid sequence**
    - For example, UGU and UGC both code for the amino acid, cysteine
- Some send important signals to the transcription machinery
  - The **START codon** initiates the process of transcription and ensures it starts in the right location (this is always the amino acid **methionine** in eukaryotic cells, coded for by the codon AUG)
  - **STOP codons** cause transcription to terminate and do not code for an amino acid e.g. UAA
- The genetic code is **non-overlapping**
  - Each base is only read once in the codon it is part of

The rules of base pairing

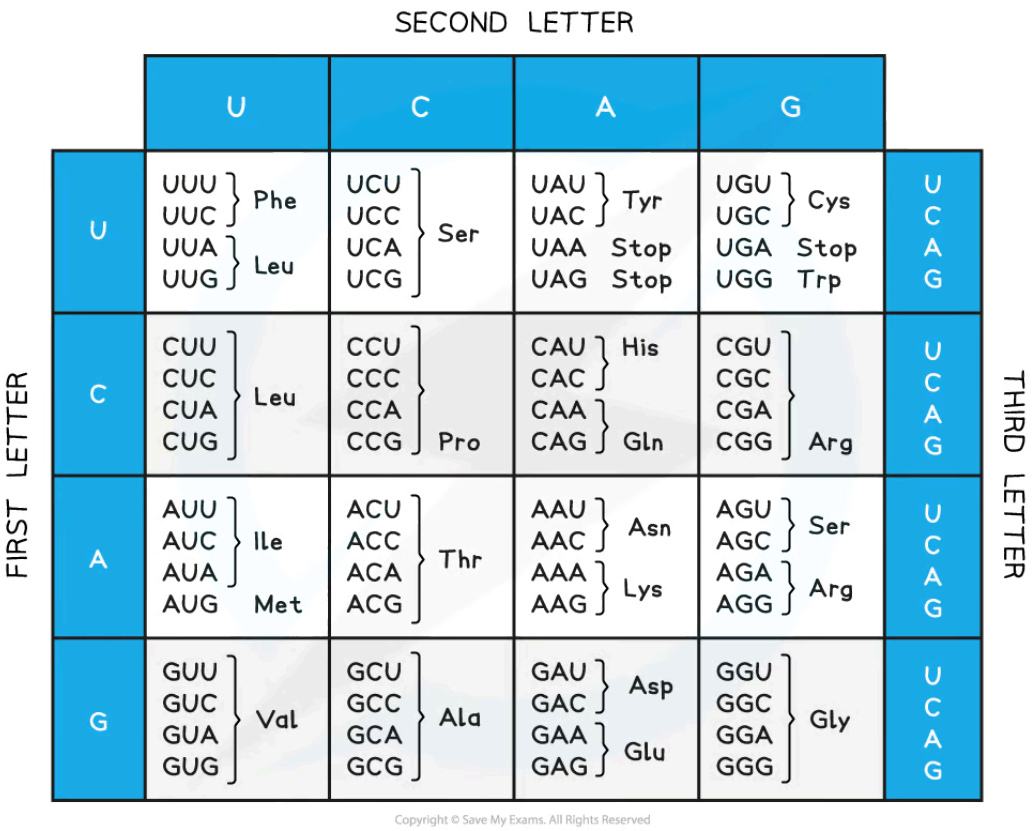
- In **DNA**,
  - **Adenine** (A) always pairs with **Thymine** (T)
  - **Cytosine** (C) always pairs with **Guanine** (G)
- In **RNA**, Thymine (T) is replaced by **Uracil** (U)
  - This means that the base Adenine (A) in DNA is transcribed to Uracil (U) in the mRNA strand

The mRNA Codons and Amino Acids table

- The first three bases of an mRNA strand form the first codon
- The **first base** of the codon is read from the **first column** of the table
- The **second base** of the codon is read from the **top row** of the table
- The **third base** of the codon is read from the **final column** of the table

mRNA Codons and Amino Acids Table

Your notes



Intrepreting the Genetic Code


Your notes

- Use a table of the genetic code to deduce which codon(s) corresponds to which amino acid
- The **20 amino acids** are all coded for in the Table of mRNA Codons
- Some amino acids are **only coded for by one codon** eg. methionine (Met), coded by AUG
- Other amino acids **have several codons that code for them** eg. arginine (Arg), coded by CGU, CGC, CGA, CGG, AGA and AGG

Worked example

Use the table of mRNA Codons and Amino Acids to identify the mRNA codons that code for the following amino acids:

1. Histidine (His)
2. Tryptophan (Trp)
3. Glycine (Gly)
4. Leucine (Leu)

Step 1: Look up His in the table

Codon CAU (no others)

Step 2: Look up Trp in the table

Codon UGG (no others)

Step 3: Look up Gly in the table

Codons GGU, GGC, GGA and GGG

Step 4: Look up Leu in the table

Codons CUU, CUC, CUA, CUG, UUA, UUG

Deducing the Sequences

- Use a table of mRNA codons and their corresponding amino acids to deduce the sequence of amino acids coded by a short mRNA strand of a known base sequence

Your notes

Worked example

Deduce the amino acid sequence coded for by the mRNA sequenceAUGACUGGGCCUCCCCAAUAUUAG

Step 1: split the mRNA sequence into triplets

AUG ACU GGG CCU CCC CAA UAU UAG

Step 2: look up the first triplet in the mRNA Codons and Amino Acids Table:

AUG = Met

Step 3: repeat for the remaining triplets

ACU = Thr

GGG = Gly

CCU = Pro

CCC = Pro

CAA = Gln

UAU = Tyr

UAG = Stop

Step 4 : link the amino acids together

Met-Thr-Gly-Pro-Pro-Gln-Tyr

Exam Tip

In an exam, you may be asked to predict the effect of specific mutations in the genetic code. Remember that the genetic code allows more than one amino acid to be coded for by triplets and is non-overlapping!You will not be required to memorise specific codons and the amino acids that they code for.