DNA AND GENOMICS

Learning Outcomes

Core Topic 2 – DNA and Genomics

Candidates should be able to

(a) Describe the structure and roles of DNA and RNA (tRNA, rRNA and mRNA). (Mitochondrial DNA is not required.)

(b) Describe the process of DNA replication and the experimental evidence for semi-conservative replication.

(c) Describe how the information on DNA is used to synthesize polypeptides in prokaryotes and eukaryotes. (Description of the processes of transcription, formation of mRNA from pre-mRNA and translation is required.)

(d) Explain how a change in the sequence of DNA nucleotide (gene mutation) may affect the amino acid sequence in a protein, and hence the phenotype of the organism e.g. sickle cell anaemia and cystic fibrosis. (Knowledge of substitution, addition, deletion and frameshift mutation is required.)

Content Outline

1. Introduction
2. Nucleic Acids
   (a) Structure of nucleotides
   (b) Structure of polynucleotides
   (c) Deoxyribonucleic acid (DNA)
   (d) Ribonucleic acid (RNA)
   (e) Similarities and differences between DNA and RNA
3. DNA Replication
   (a) Models of DNA replication
   (b) Evidence for semi-conservative DNA replication
   (c) Semi-conservative replication
   (d) Importance of base-pairing and hydrogen bonding in DNA
4. Gene Expression
   (a) The Central Dogma of molecular biology
   (b) Transcription
   (c) Post-transcriptional modification
   (d) The Genetic Code
   (e) Translation
   (f) Differences between prokaryotic and eukaryotic gene expression
5. Gene Mutation and its Effects
   (a) Types of gene mutation
   (b) Effects of gene mutation
   (c) Sickle Cell Anaemia
   (d) Cystic fibrosis

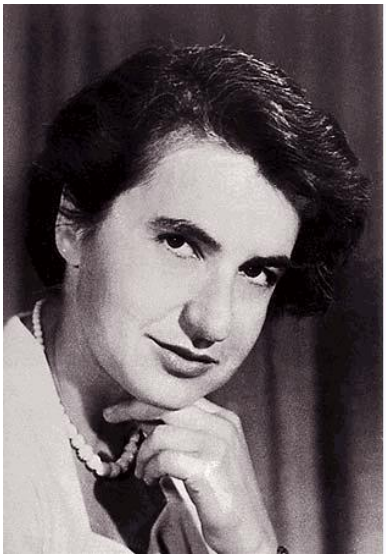
References

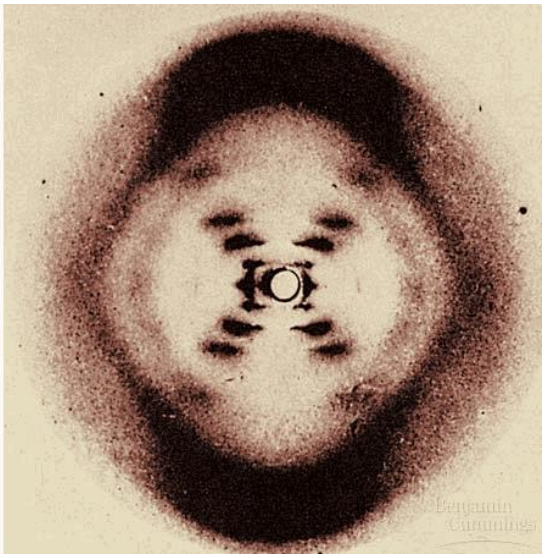
1. Alberts B., Johnson A., Lewis J., Raff M., Roberts K. and Watts P. (2002) Molecular Biology of the Cell. 4th Edition. Chapter 6: How cells read the genome: From DNA to Protein. Garland Science, Taylor and Francis Group.
2. Campbell N. A. and Reece J. B. (2008). Biology. Chapter 5: The structure and function of biological molecules. 8th Edition. Benjamin Cummings Publishing, Inc.
3. Campbell N. A. and Reece J. B. (2008). Biology. Chapter 16: The molecular basis of inheritance. 8th Edition. Benjamin Cummings Publishing, Inc.
4. Campbell N. A. and Reece J. B. (2008). Biology. Chapter 17: From gene to protein. 8th Edition. Benjamin Cummings Publishing, Inc.
5. Clegg C. J. and MacKean D. J. (2000). Advanced Biology – Principles and Applications. Chapter 9: The nucleus in division and interphase. 2nd Edition. John Murray Publishers Ltd.
6. Jones M. and Jones G. (2004). Advance Biology. Chapter 5: DNA and protein synthesis. 1st Edition. Cambridge University Press.

1. Introduction

The American James Watson and Englishman Francis Crick solved the puzzle of deoxyribonucleic acid (DNA) structure.

Watson saw an X-ray diffraction image of DNA (below, right) revealed by a technique called X-ray crystallography produced by Rosalind Franklin (below, left) and Maurice Wilkins.

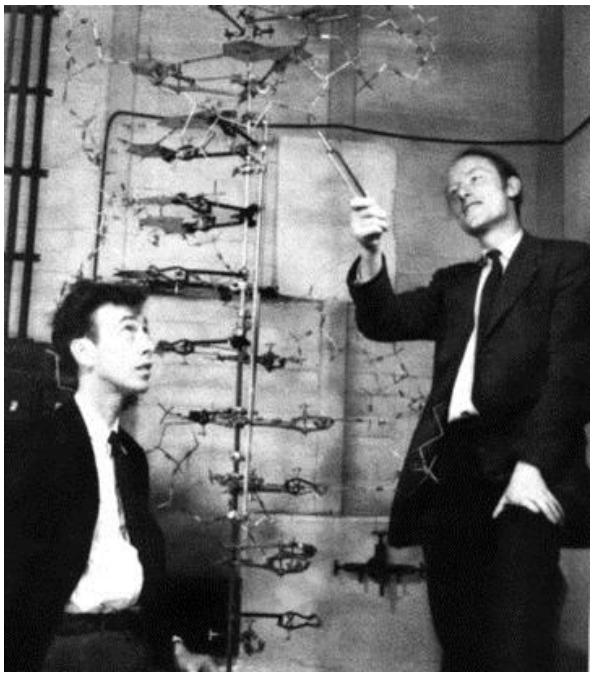




From the analysis of X-ray diffraction photo of DNA, the structure of DNA molecule:

- is a helix made up of two strands.
- has a uniform width of 2nm.
- is a helix that makes one full turn every 3.4 nm.
- has a 0.34nm distance between stacked nitrogenous bases.

In April 1953, Watson and Crick (below, left) surprised the scientific world with a succinct, one-page paper in the British Journal, *Nature*. The paper reported their molecular model for DNA: the double helix which has since become the symbol of molecular biology.



No. 4356 April 25, 1953 NATURE 737

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxy-ribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configuration) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

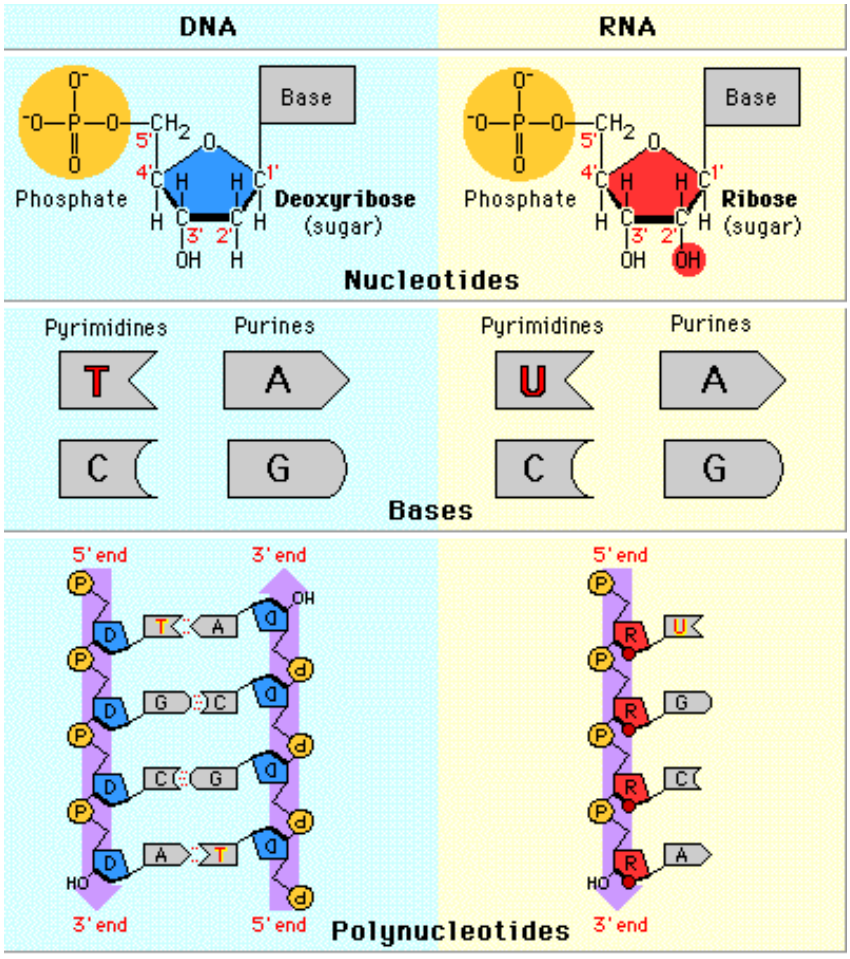
It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at



Structure of DNA and RNA

2. Nucleic Acids

There are 2 types of nucleic acids:

- **deoxyribonucleic acid (DNA)**
- **ribonucleic acid (RNA)**.

Nucleotides are the monomers of **nucleic acids** (**polynucleotides**).

- Deoxyribonucleotides are the monomers of **deoxyribonucleic acid (DNA)**
- Ribonucleotides are the monomers of **ribonucleic acid (RNA)**.

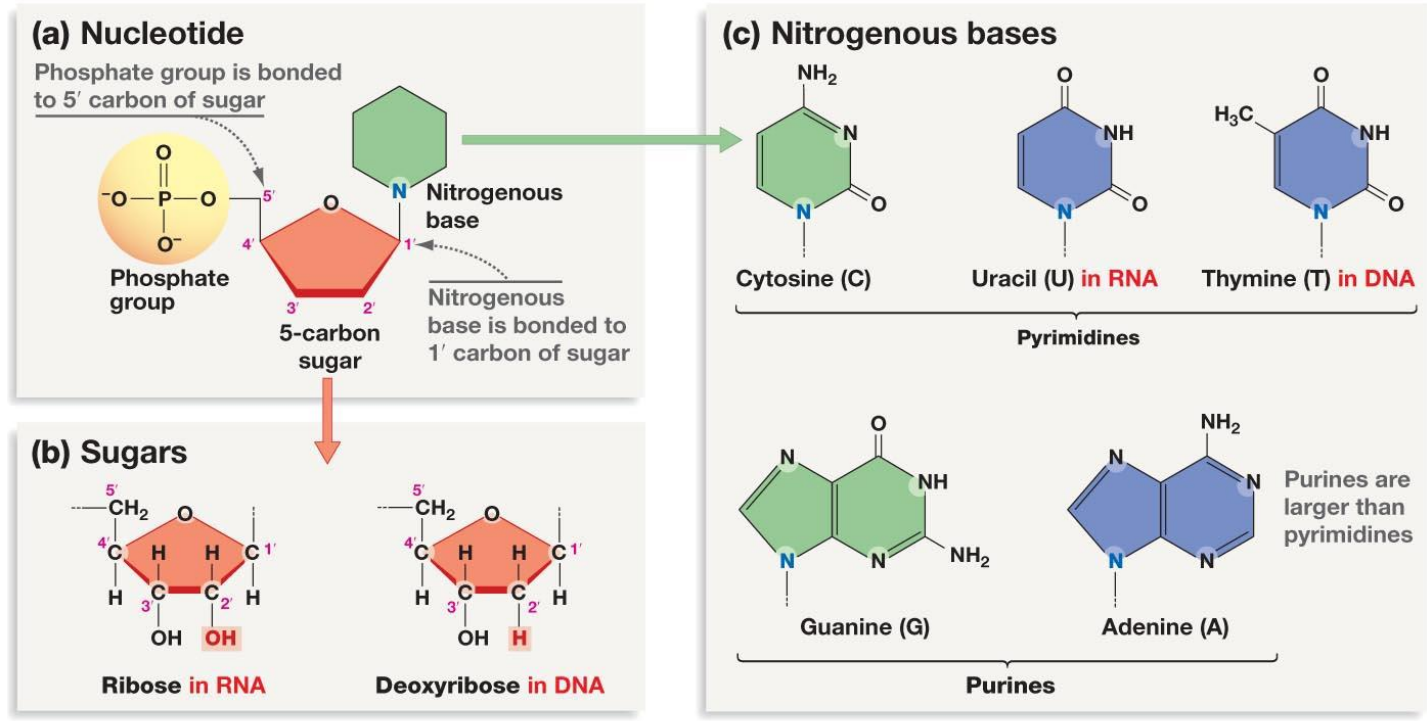
(a) Structure of nucleotides

A **nucleotide** consists of 3 components:

(i) a **pentose sugar**,
(ii) a **nitrogenous base** &
(iii) one or more **phosphate groups** (derived from phosphoric acid).

The three components are joined by **condensation reaction**.

- The **nitrogenous base is bonded to 1' carbon** of the pentose sugar molecule.
- The **phosphate group is bonded to 5' carbon** of the pentose sugar molecule.
- Two molecules of water are formed.



Structure of nucleotide and their components

A **nucleoside** is a molecule consisting of:

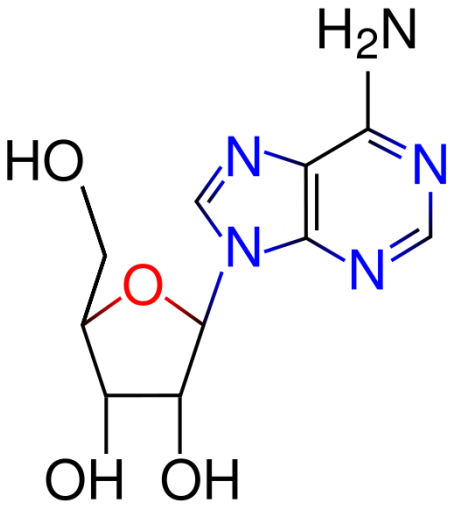
- a **pentose sugar** and
- a **nitrogenous base**.

E.g. Adenosine is a nucleoside with a pentose sugar (ribose) and the nitrogenous base (adenine).

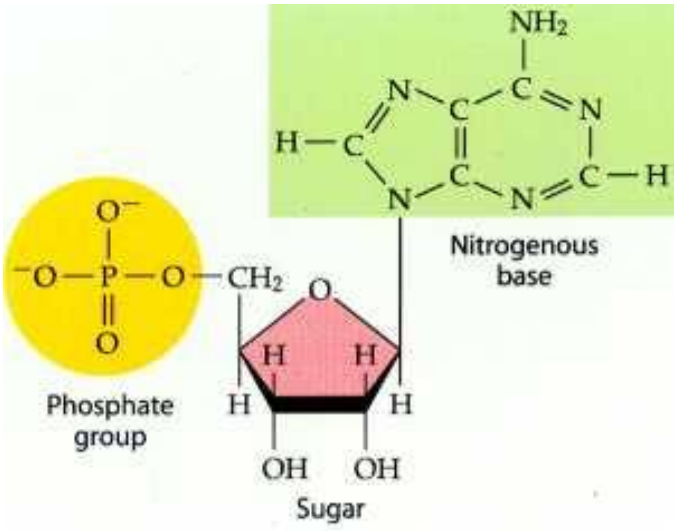
A **nucleotide** is a molecule consisting of:

- **nucleoside** (pentose sugar + nitrogenous base) and
- **phosphate** group(s)

E.g. Adenosine monophosphate is a nucleotide.



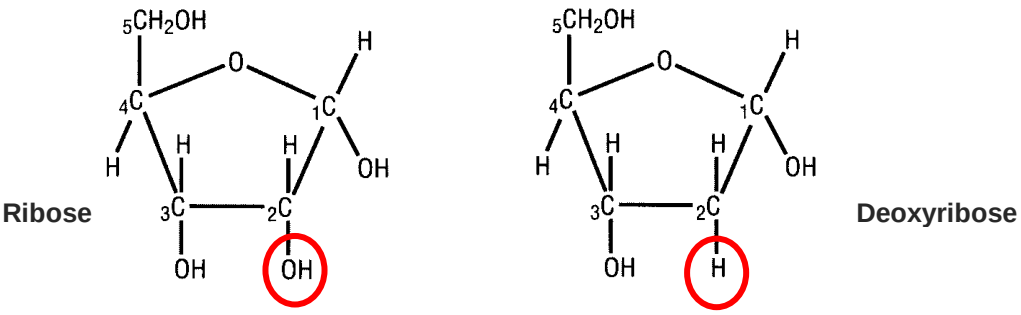
Structure of adenosine



Structure of nucleotide (adenosine monophosphate)

(i) **Pentose Sugar**

- **5-carbon** sugar.
- 2 different types of pentose sugars: **ribose** and **deoxyribose**.
- In **ribonucleic acid** (RNA), the pentose sugar is **ribose**.
  - Ribose has a hydroxyl group at 2' carbon.
- In **deoxyribonucleic acid** (DNA), the pentose sugar is **deoxyribose**.
  - Deoxyribose has a hydrogen atom at 2' carbon.



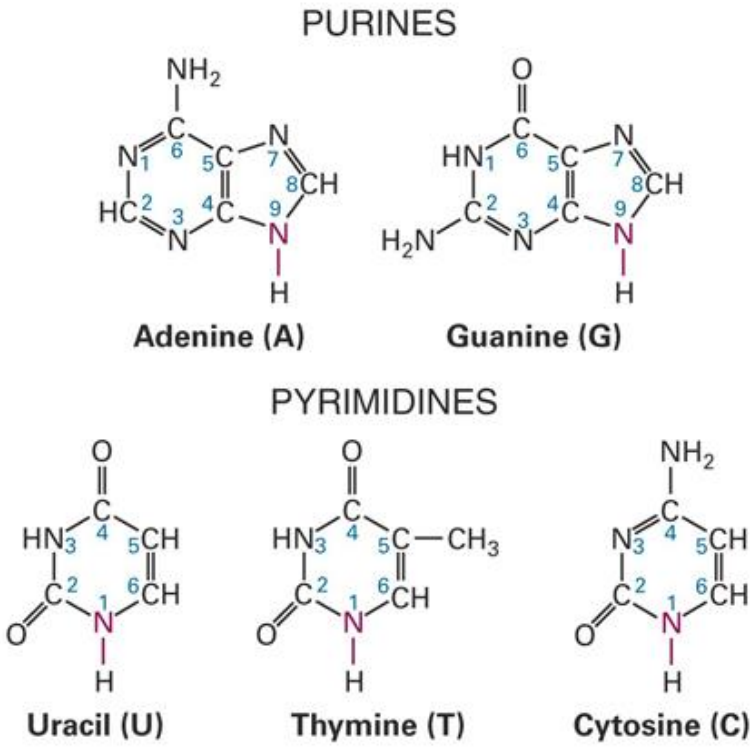
Structure of ribose and deoxyribose

(ii) **Nitrogenous bases**

- Has one or two ring structures containing nitrogen atoms.
- Are known as a "base" due to the presence of lone pair of electrons on nitrogen atoms which tend to accept H^+ from solution.
- 2 categories of nitrogenous bases: **purines** and **pyrimidines**.
- 5 different types of nitrogenous bases:
  adenine, **guanine**, **cytosine**, **thymine** & **uracil**.

- Purines
  - 2 rings: one 6-membered ring fused to one 5-membered ring
  - **adenine** (A) & **guanine** (G)

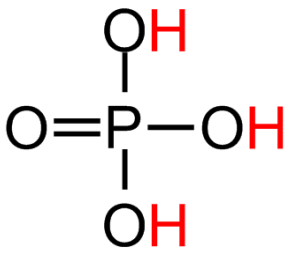
- Pyrimidines
  - 1 ring: one 6-membered ring
  - **cytosine** (C), **thymine** (T) & **uracil** (U)
    - T is only found in DNA
    - U is only found in RNA



Structure of nitrogenous bases

(iii) **Phosphate group**

- Derived from phosphoric acid
- **attached to 5' carbon of pentose sugar**



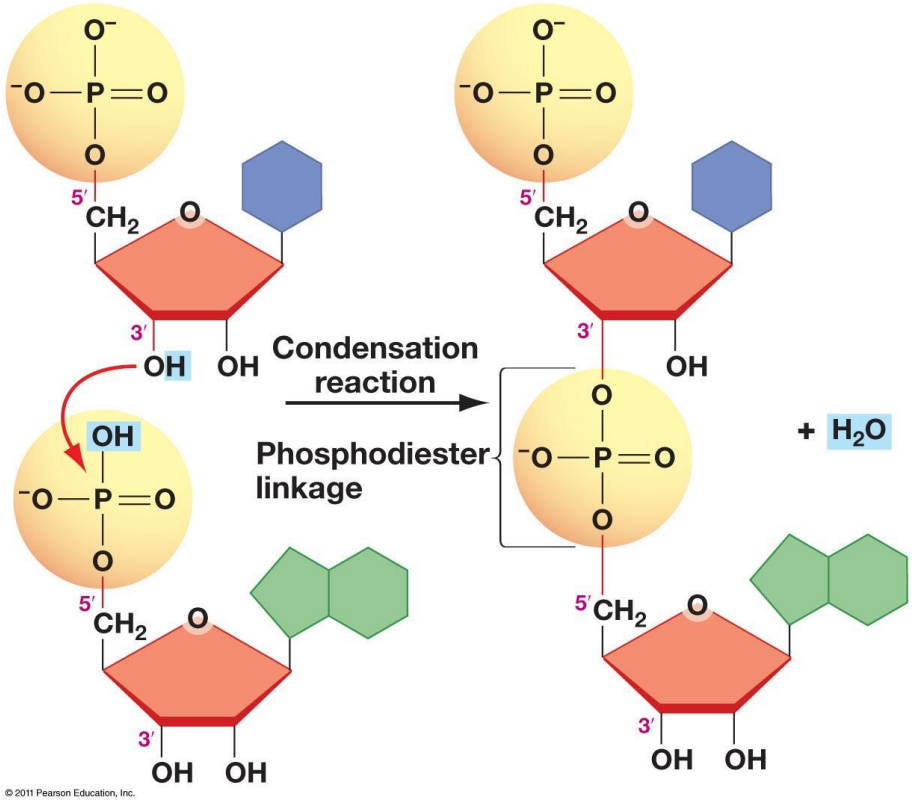
H_3PO_4

Structure of phosphoric acid

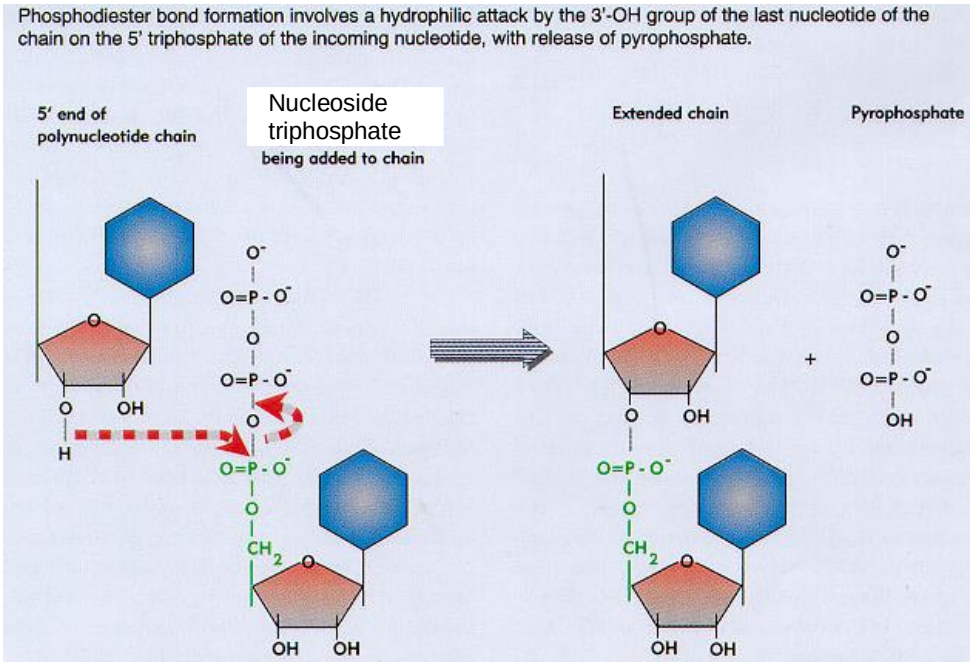
(b) Structure of polynucleotides

Many **nucleotides** joined together to form **polynucleotides**.
They are joined by strong covalent bonds called **phosphodiester bonds / linkages**.
Phosphodiester bonds are formed by **condensation reaction** between:

- **–OH group on 3' carbon** of pentose sugar of one nucleotide and
- **phosphate group on 5' carbon** of pentose sugar of the next nucleotide.



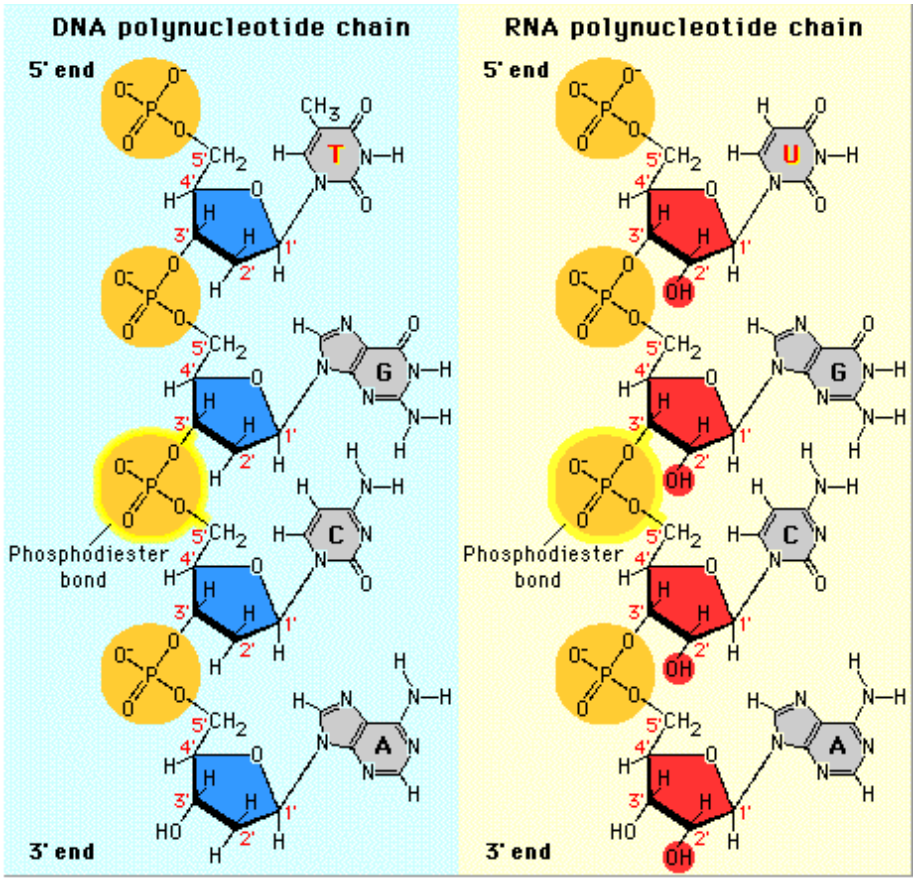
Formation of phosphodiester bonds



The resulting structure is a backbone with a repeating pattern of sugar-phosphate called the **sugar-phosphate backbone**, with nitrogenous bases projecting out.

The two ends of the polynucleotide are different from each other.

- **5' end** of a polynucleotide has **a phosphate group** attached to 5' carbon of sugar.
- **3' end** of a polynucleotide has a **hydroxyl** / **–OH group** on 3' carbon of sugar.

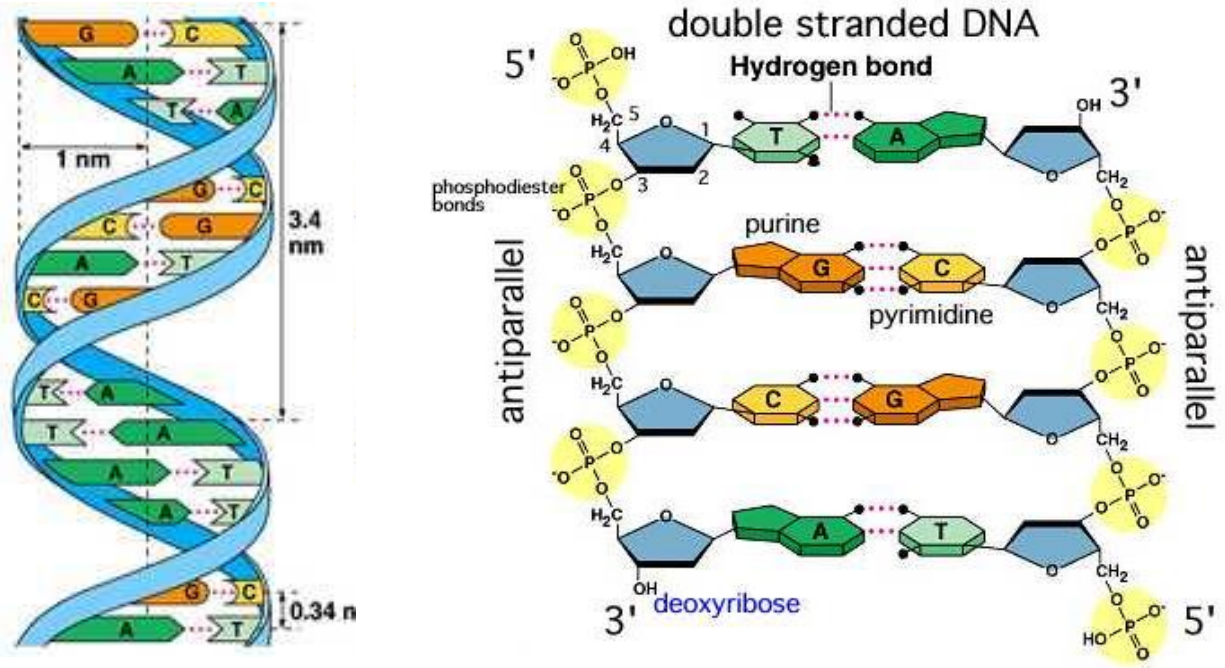


Structure of polynucleotide

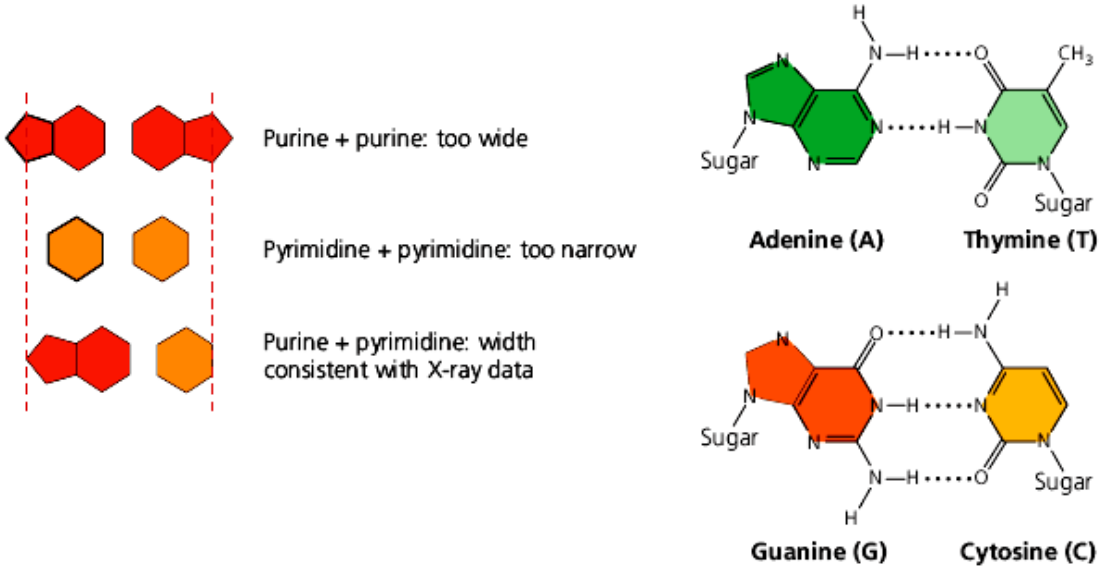
(c) Deoxyribonucleic Acids (DNA)

(i) Structure

- DNA molecule consists of **2 polynucleotide chains** spiralled around an imaginary axis to form a **double helix**.
- The 2 polynucleotide chains are **antiparallel** (run in opposite directions)
  - One strand runs in 5' to 3' direction.
  - The other strand runs in 3' to 5' direction.
- DNA molecule has a **uniform width of 2 nm**.
- The nitrogenous bases are stacked **0.34 nm** apart and the helix **makes one full turn every 3.4 nm**, as such there are **10 base pairs in each turn** of the helix.
- The two (hydrophilic) **sugar-phosphate backbones are on the outside of helix**
- The (hydrophobic) **nitrogenous bases are paired in the interior of the helix**.
- The **stacking of base pairs** results in **hydrophobic interactions**.
- The two polynucleotide chains or strands are held together by **hydrogen bonds between the paired nitrogenous bases**.
- The sequence of bases along one strand is **complementary** to the sequence of bases along the other strand.
  - **adenine** (A) always pairs with **thymine** (T) with **two** hydrogen bonds via **complementary base pairing**.
  - **guanine** (G) always pairs **cytosine** (C) with **three** hydrogen bonds via **complementary base pairing**.
- The two grooves between the backbones are called the **major groove** and **minor groove**.



Structure of DNA



Complementary base pairing in DNA

- Chargaff's rules:
  - The base composition varies between species.
  - Within a species, the number of A and T bases are equal and the number of G and C bases are equal.
  - The total amount of pyrimidine nucleotides (T + C) always equals the total amount of purine nucleotides (A + G).

	Mol% of bases			
Source	A	G	C	T
ϕX174	24.0	23.3	21.5	31.2
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Chicken	28.0	22.0	21.6	28.4
Rat	28.6	21.4	20.5	28.4
Human	29.3	20.7	20.0	30.0

Chargaff's base composition data

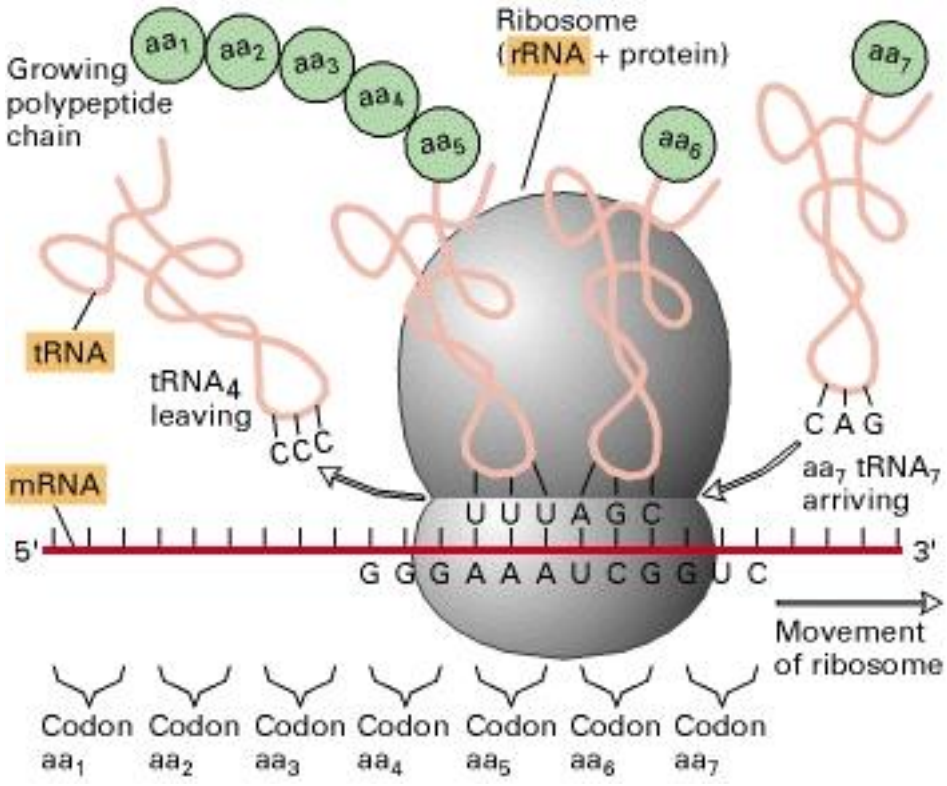
(ii) **Function**

- DNA is the **genetic material** that organisms **inherit** from their parents.
- Most DNA molecules are very long and each DNA molecule contains numerous **genes**. Each gene occupies a specific region along the DNA molecule.
- The gene is a **unit of inheritance**, which store coded instructions for the **synthesis** of specific molecules like **protein** and **RNA**.
- During cell division, **replication of DNA** occurs. The structure of DNA (i.e. two polynucleotide chains) makes it possible to accurately copy the entire DNA.
- During cell division, the two strands of DNA separate. Each strand serves as a **template** from which a new complementary strand is made (i.e. **semi-conservative replication**).

(d) Ribonucleic Acids (RNA)

(i) Structure

- The RNA molecule is a **single polynucleotide chain**.
- RNA has much fewer nucleotides compared to DNA.
- RNA contains ribose as its pentose sugar.
- RNA is **less stable than DNA** because it is more prone to hydrolysis by intracellular enzymes.
- RNA molecule can contain 4 different nitrogenous bases similar to those in DNA except that in RNA, **thymine is replaced by uracil**.
- RNA can develop **secondary structures** which are formed by **complementary base pairing within the RNA molecule**. The secondary structures function to promote stability of the molecules.
  - **adenine** (A) always pairs with **uracil** (U) with **two** hydrogen bonds.
  - **guanine** (G) always pairs **cytosine** (C) with **three** hydrogen bonds.
- Cells synthesize several types of RNA. The different forms of RNA include:
  - **messenger RNA (mRNA)**,
  - **ribosomal RNA (rRNA)** and
  - **transfer RNA (tRNA)** that are involved in the synthesis of polypeptides (translation).



Role of mRNA, rRNA and tRNA in protein synthesis

(ii) **Messenger RNA**

- A mRNA molecule is a single-stranded RNA.
- mRNA is synthesised by the **transcription** of DNA in the **nucleus**.
- mRNA is transported to cytoplasm for **translation** by **ribosomes**.
- mRNA is used as an **template** to **synthesis proteins**.
- Eukaryotic mRNA requires processing after being transcribed before it is transported out of the nucleus, while prokaryotic mRNA does not.

(iii) **Ribosomal RNA**

- rRNA molecule is single-stranded RNA.
- rRNA is **synthesized in the nucleolus**.
- rRNA and ribosomal proteins are then **assembled into the large and the small subunits within the nucleolus**.
- The two subunits are then transported out of the nucleus into the cytoplasm where they associate to form **ribosomes**.
- Ribosomes are the **sites of protein synthesis**.

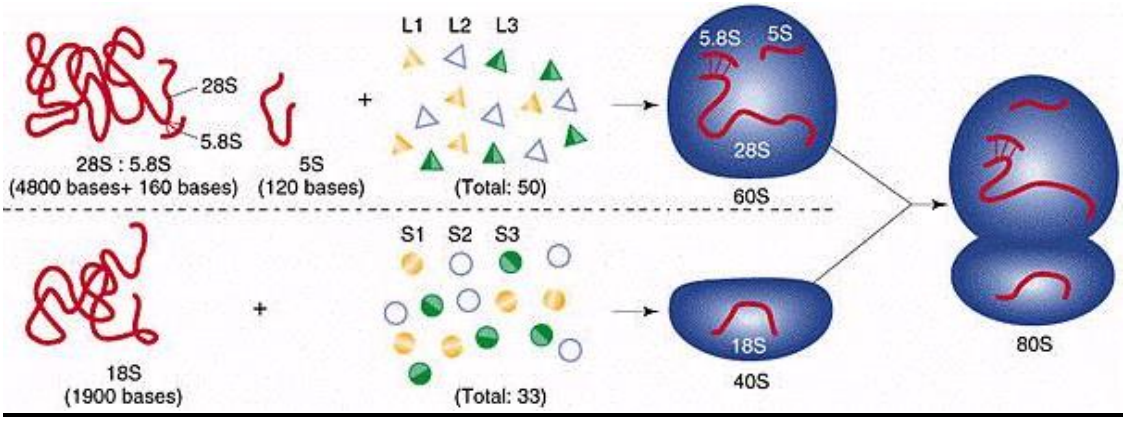
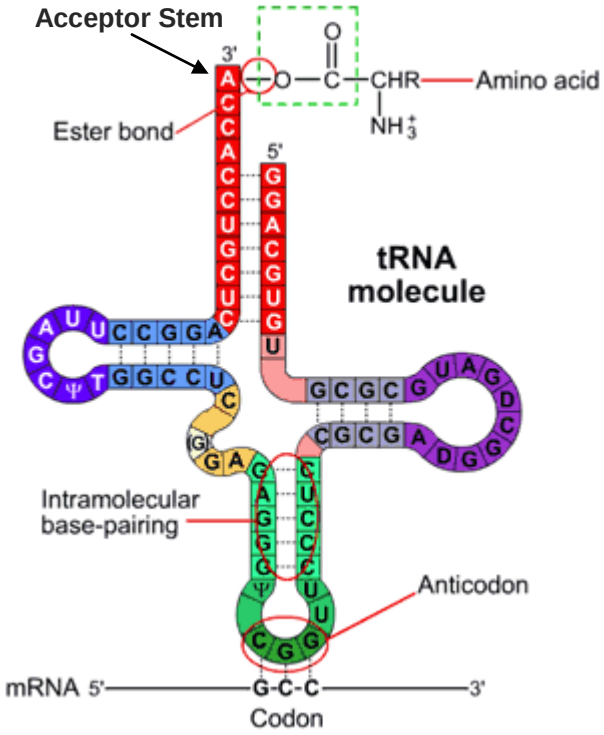


Diagram showing assembly of eukaryotic ribosomal subunits from rRNA and ribosomal proteins

(iv) **Transfer RNA**

- A tRNA molecule is a single-stranded RNA molecule.
- A triplet base sequence known as the **anticodon** is present on the tRNA.
  - **Anticodon** on tRNA forms hydrogen bonds with a codon (triplet sequence) on mRNA via **complementary base pairing**.
- The **3' end** is known as the **acceptor stem**.
  - the site for **amino acid attachment**.
- The **clover leaf structure** of the tRNA further folds into a **3-dimensional structure**.
  - tRNA is able to **fit into the ribosome** during translation (refer to Section 4e).
- tRNA transfers the **correct amino acid** to the ribosome during translation.



2D structure of tRNA

3D structure of tRNA

(e) Similarities and Differences between DNA and RNA

(i) Similarities

- Both are polynucleotides made up of **nucleotides**, which consist of a **pentose sugar**, a **nitrogenous base** and a **phosphate group**.
- Polynucleotides have a sugar-phosphate backbone joined by **phosphodiester bonds** formed by **condensation reaction** between nucleotides.
- Both make use of 3 common nitrogenous bases: purines (adenine & guanine) & pyrimidine (cytosine)

(ii) Differences

Features	DNA	RNA
Number of polynucleotide chains	Double stranded DNA, in the form of double helix	Single stranded RNA
Monomer	Deoxyribonucleotide consist of: Deoxyribose sugar 4 nitrogenous bases (A, T, G, C) Phosphate group	Ribonucleotide consist of: Ribose sugar 4 nitrogenous bases (A, U, G, C) Phosphate group
Type of pentose sugar	Deoxyribose has a **hydrogen atom** at 2' carbon.	Ribose has a **hydroxyl group** at 2' carbon.
Type of pyrimidine	**thymine** (T) & cytosine (C)	**uracil** (U) & cytosine (C)
Complementary base-pairing	**adenine** base pairs with **thymine** **guanine** base pairs with **cytosine**	**adenine** base pairs with **uracil** **guanine** base pairs with **cytosine**
Molecular mass	Relatively large with long chain of nucleotides	Relatively small with short chain of nucleotides
Function	DNA is the genetic material that organisms **inherit** from their parents	mRNA, rRNA & tRNA Each serves a different function in protein synthesis
Enzyme involved in polymerisation	DNA polymerase	RNA polymerase
Stability	Very stable	Less stable (degraded after it serves its function)
Location	Found almost entirely in nucleus as linear DNA (circular DNA found in chloroplasts & mitochondria & bacteria)	mRNA, tRNA are transcribed in nucleus and transported to cytoplasm rRNA is transcribed in the nucleolus and is part of the structure of ribosomes.
Quantity present in each cell	Same for all cells of the same species (except in gametes where number is halved)	Varies from cell to cell within the same organism (dependent on the cell's needs at a particular time)

3. DNA Replication

(a) Models of DNA replication

(i) **Semi-Conservative Model** (Watson & Crick's DNA model)

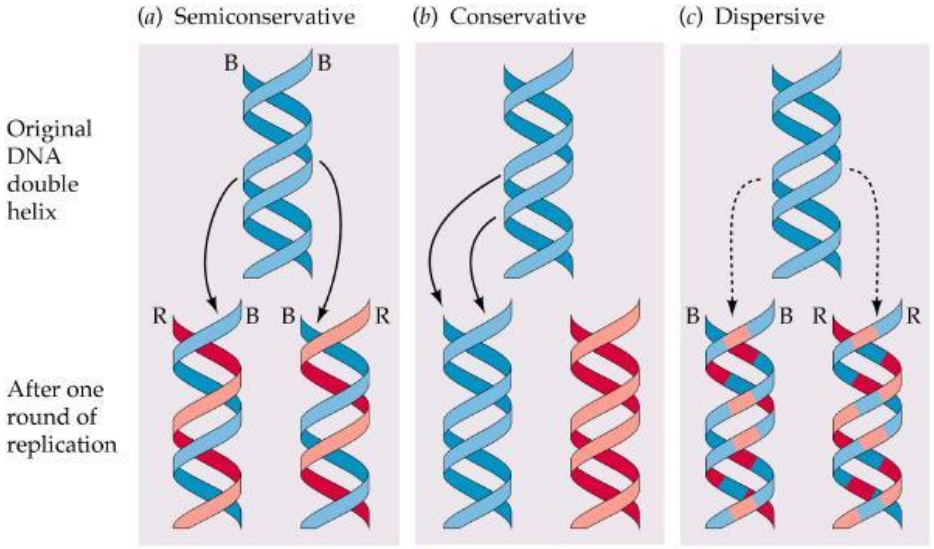
- This model proposes that that the double helix **parental molecule** replicates and each of the daughter molecules has a **parental strand** and a **newly synthesised strand**.
- The two strands of the parental molecule **separate** and each strand acts as **template** for synthesis of newly synthesised strand.
- Each daughter **molecule** has **one strand conserved from the parental molecule** and the other **newly synthesised strand**.

(ii) **Conservative Model**

- This model proposes that the parental molecule remains intact and the resulting daughter molecule is formed from two newly synthesised DNA strands.
- Both strands of the DNA molecule act as templates for the synthesis of an entirely new DNA molecule.
- The two parental strands reassociate to restore the parental molecule.

(iii) **Dispersive Model**

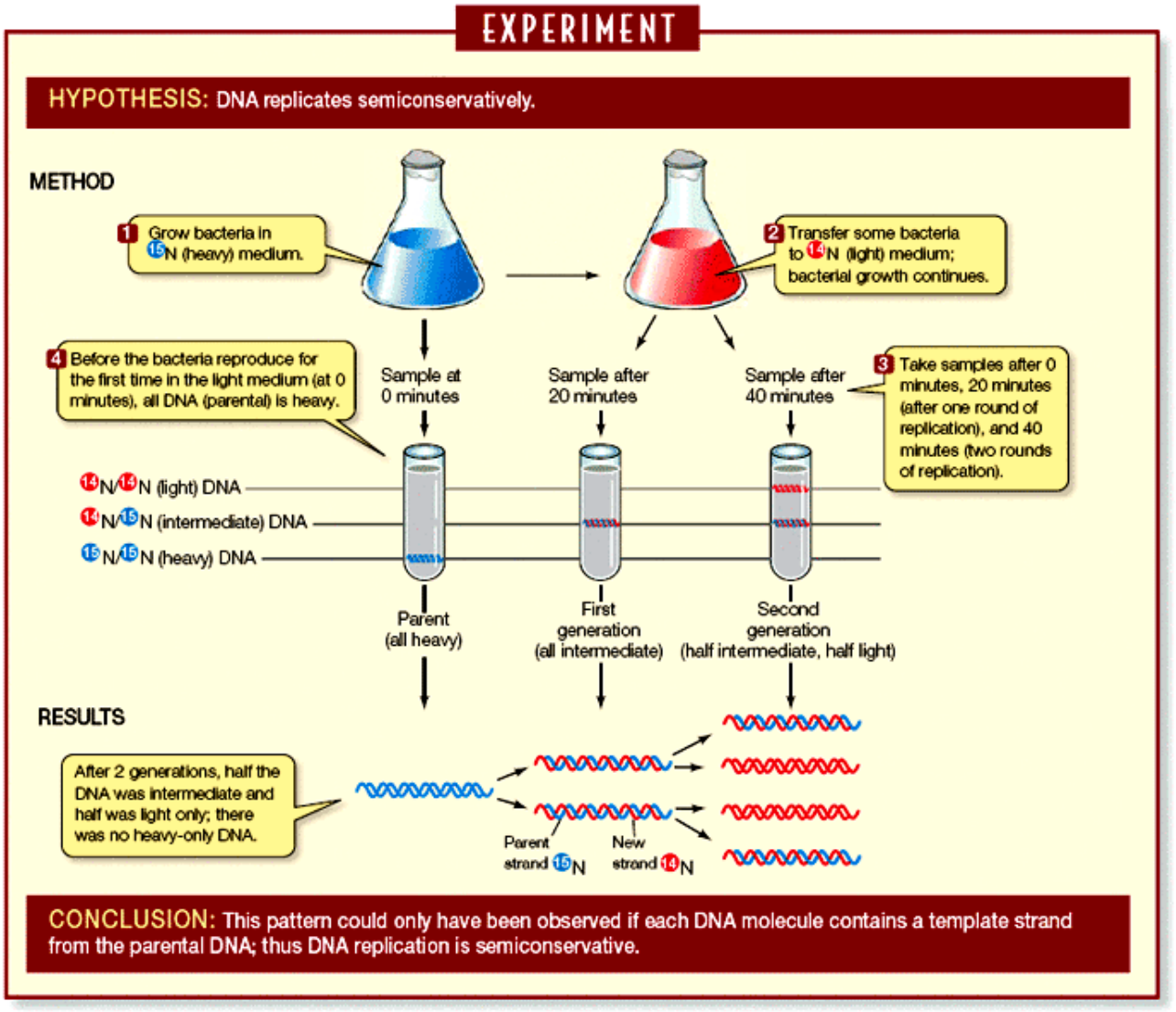
- This model proposes that each strand of the both daughter molecules contains a mixture of old and newly synthesised DNA.
- The parental DNA molecule breaks up into short segments, which act as templates for the synthesis of DNA.
- The segments are then joined together, resulting in both old and new DNA interspersed along each strand in both daughter DNA molecules.



Three Models of DNA replication

(b) Evidence for semi-conservative DNA replication

In 1958, American scientists Matthew Meselson & Franklin Stahl carried out an experiment in bacterium *Escherichia coli*, to see if DNA is copied semi-conservatively. *Escherichia coli* is a common and usually harmless rod-shaped bacterium found in the human gut.



The Meselson and Stahl experiment

Bacteria were used rather than eukaryotic cells because:

- Bacterial cells reproduce quickly such that many generations can be studied in a short period of time.
- They are easy to culture in large quantities.

1 ***E. coli*** were grown in a medium (food source) containing **ammonium chloride (NH_4Cl) with a heavy nitrogen isotope, ^{15}N** for many generations.
   - Every time a cell divides, its DNA replicates and ^{15}N is incorporated into nucleotides which are used to synthesize new DNA.
   - All the bases in the DNA molecules contain ^{15}N and their DNA will be 'heavy'.

2 The bacteria were **transferred** into a medium containing the **lighter isotope of nitrogen, ^{14}N** and allowed to grow. At various times of one, two or more generations after the transfer, samples of bacteria were collected.

3 The DNA from each group of bacteria was then extracted and put into a **caesium chloride solution** and spun at 40000 g in a **centrifuge**.

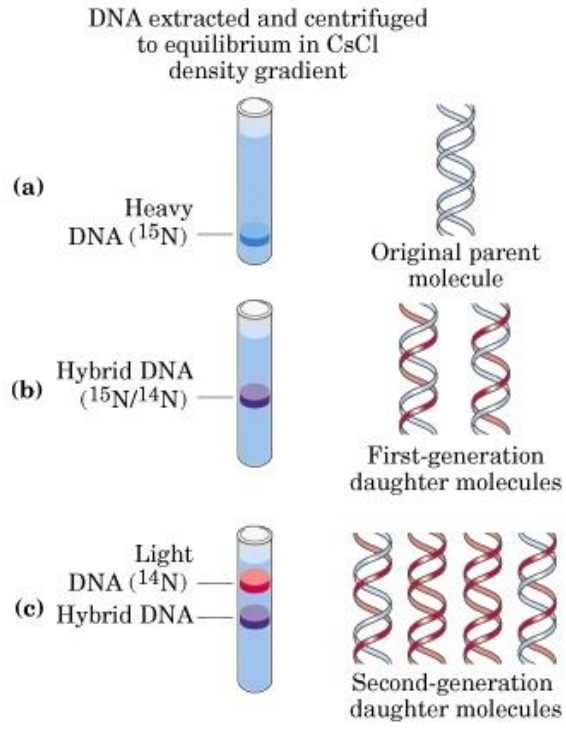
4 The caesium chloride molecules sink to the bottom of the test tubes creating a **density gradient**. The DNA molecules will position at their corresponding level of density where its density equals to caesium chloride solution
   - DNA molecules containing ^{15}N are heavier than those containing ^{14}N, so they ended up nearer the base of the tubes.

5 These centrifuge tubes were observed under UV rays. DNA appeared as fine bands in the tubes at different heights according to their density.

Results

6 By **semi-conservative replication**, DNA of **first generation** would be of **intermediate density** as all DNA molecules comprise one ^{15}N strand and one ^{14}N strand.

7 Half of DNA molecules from **second generation** would be of **intermediate density** and half of DNA molecules would be of **light density** as 50% of DNA molecules comprise one ^{15}N strand and one ^{14}N strand and 50% of DNA molecules comprise two ^{14}N strands.



Results proving DNA replication is semi-conservative

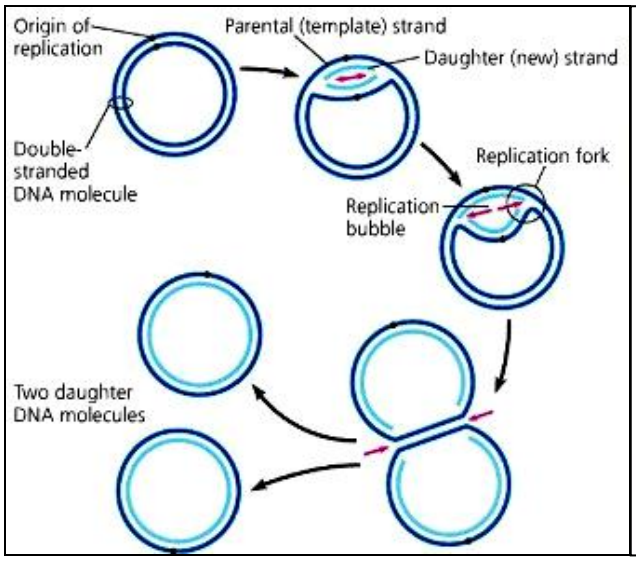
(c) Semi-conservative replication

During cell division, DNA replication must occur so that each daughter cell has a complete set of genes. **DNA replication** occurs in **synthesis phase of interphase** during the mitotic and meiotic cell cycle.

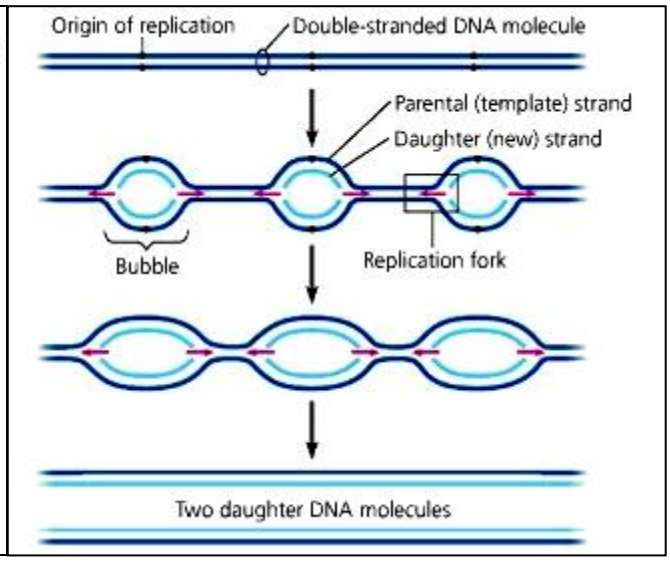
(i) Stage 1: **Unwinding**

- A portion of double helix is **unwound** and **unzipped** at the **origin of replication** by **DNA helicase**. As helicase moves along the double helix just in front of the DNA polymerase, the two parental strands are **separated** by breaking **hydrogen bonds** between nitrogenous bases. Helicase uses energy from ATP to unwind and unzip the DNA helix.

  - Replication begins at a specific sequence of nucleotides along the DNA molecule called **origins of replication**.
  - As the two DNA strands separate, a **replication bubble** is formed with two **replication forks**, a Y-shaped region at the two ends of the bubble where the parental molecule is being unzipped.
  - DNA replication proceeds in both directions from each origin of replication.
  - For a prokaryotic chromosome, only a **single origin of replication** is present. However, a eukaryotic chromosome may have **multiple origins of replication**. Multiple replication bubbles form and eventually fuse, thus speeding up the copying of the very long DNA molecules.

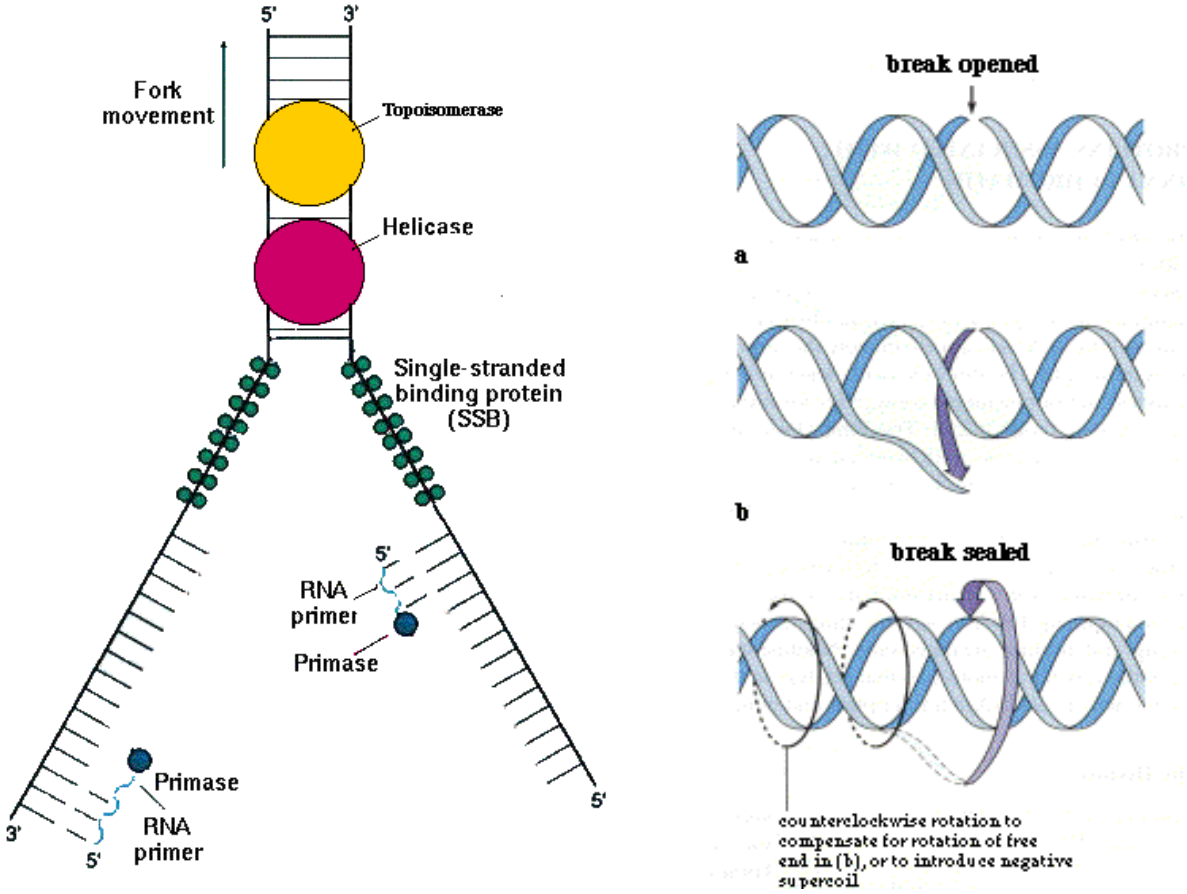


One origin of replication in prokaryotes



Multiple origins of replication in eukaryotes

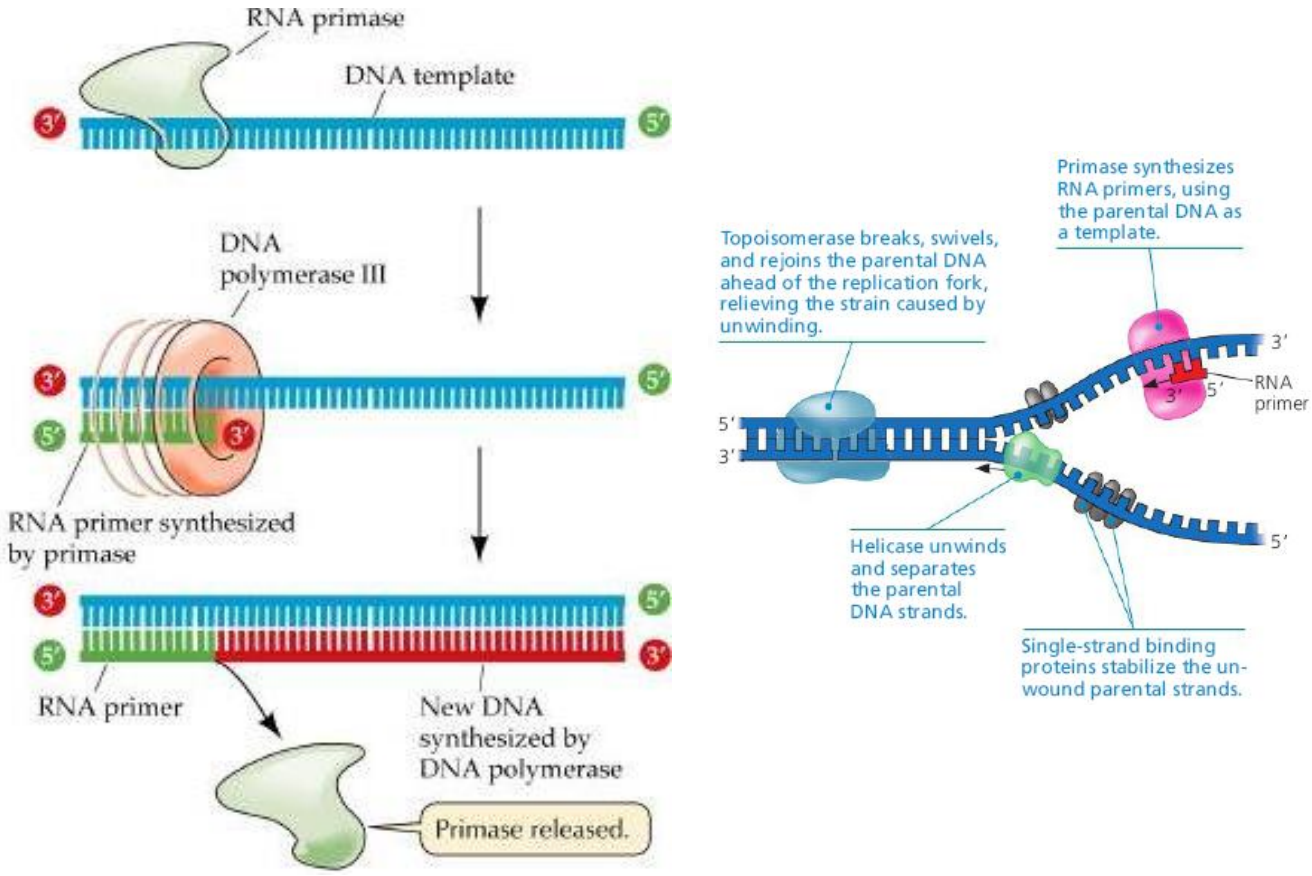
- Each strand is bound and stabilised by **single-stranded binding proteins**, preventing them from rewinding behind the replication fork.
- The unwinding of double helix causes tighter twisting and strain in front of the replication fork resulting in a positive supercoil. **DNA topoisomerase** introduces a **break in a single strand**, thus allowing the strand to **rotate around the break**, and **reseals the strand**, eliminating the positive supercoil in front of the replication fork.
- Each parental strand acts as a **template** for the synthesis of daughter strand.
- However, DNA polymerase cannot initiate the synthesis of a polynucleotide; they can only add nucleotides to the end of an already existing chain.



Role of DNA helicase, single-stranded DNA binding proteins and DNA topoisomerase

(ii) Stage 2: **Priming**

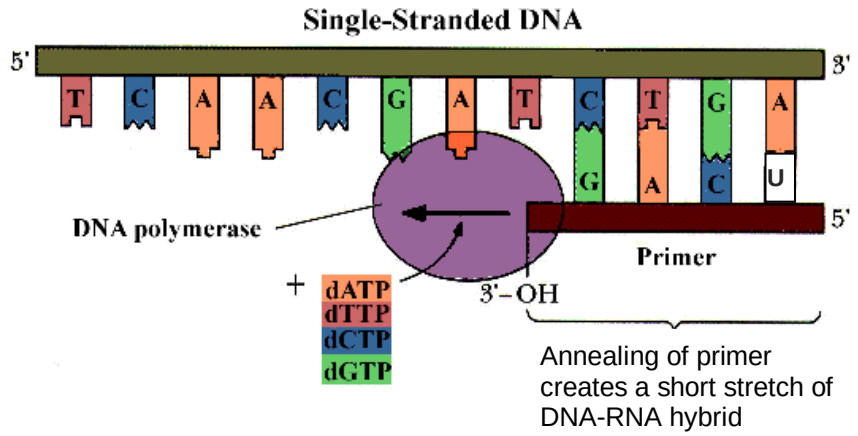
- RNA primers are short segments of RNA of about 5 to 10 nucleotides required for DNA polymerase to initiate elongation as the RNA primers provide the 3'OH end.
- **Primase** (a specialized RNA polymerase) binds to the single-stranded DNA template and synthesises **RNA primers** in the **5' to 3' direction**.
- Ribonucleotides are added one at a time via **complementary base pairing**, using the parental DNA strand as a template on both sides of a replication fork.
- The single-stranded binding proteins are displaced where the RNA primers are.

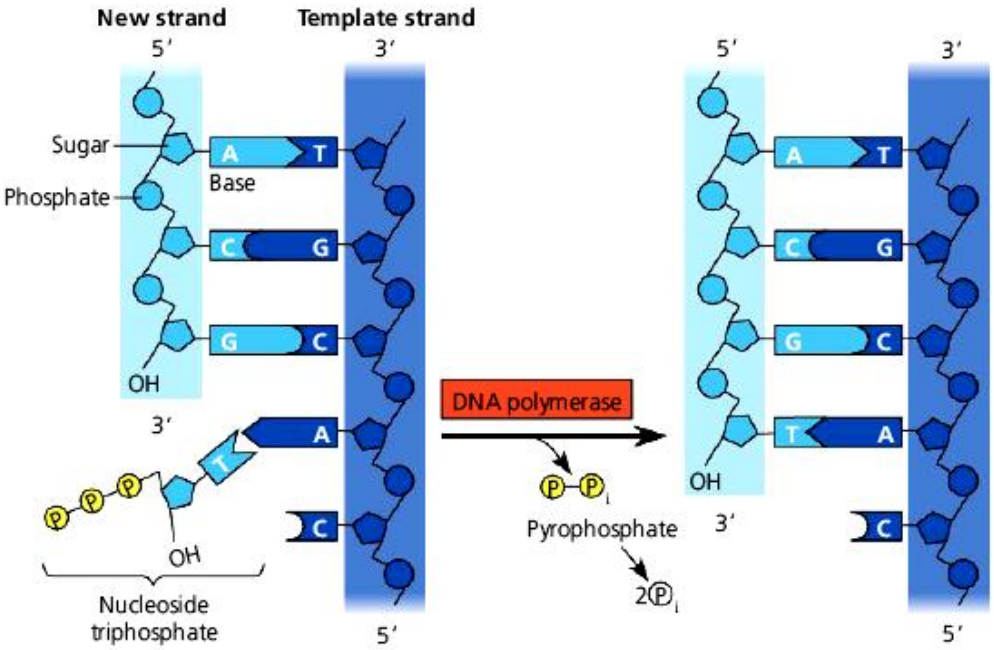


RNA primers synthesised by primase

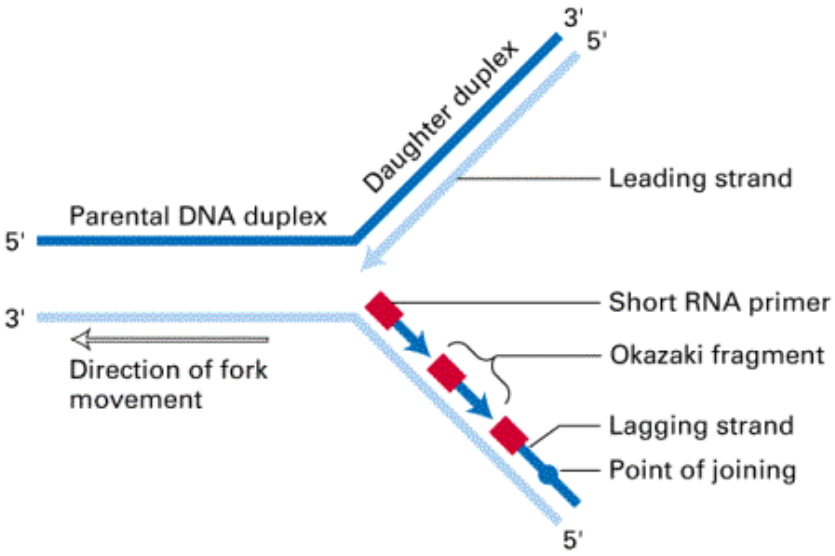
(iii) Stage 3: **Elongation**

- Before the start of DNA replication, free deoxyribonucleotides are synthesised in the cytoplasm and transported into the nucleoplasm via nuclear pores.
- DNA polymerase adds **deoxyribonucleotides** (more specifically deoxyribonucleoside triphosphates) to the free 3'OH end of the RNA primer as its active site is specific for the –OH group on the nucleotide.
- **DNA polymerase** catalyses the synthesis of a new strand of DNA in the **5' to 3' direction** via **complementary base pairing**, **A=T and G≡C** with the parental strand.
- DNA polymerase catalyses the formation of a **phosphodiester bond** between the 3'OH end of the primer and 5' phosphate group of the dNTP added.
- The energy for this process comes from the two phosphate groups which are removed as the nucleotide joins the growing end.



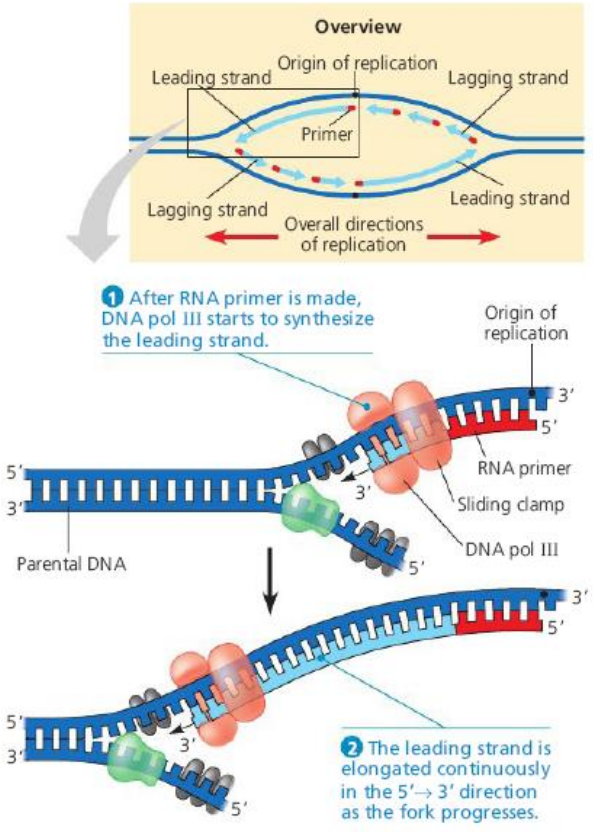


- As the two parental DNA strands are antiparallel and polymerization proceeds in the **5' to 3' direction** on both sides of the replication fork,
  - One daughter strand is synthesised towards the replication fork (**leading strand**).
  - Another daughter strand is synthesised away from the replication fork (**lagging strand**).



Leading strand

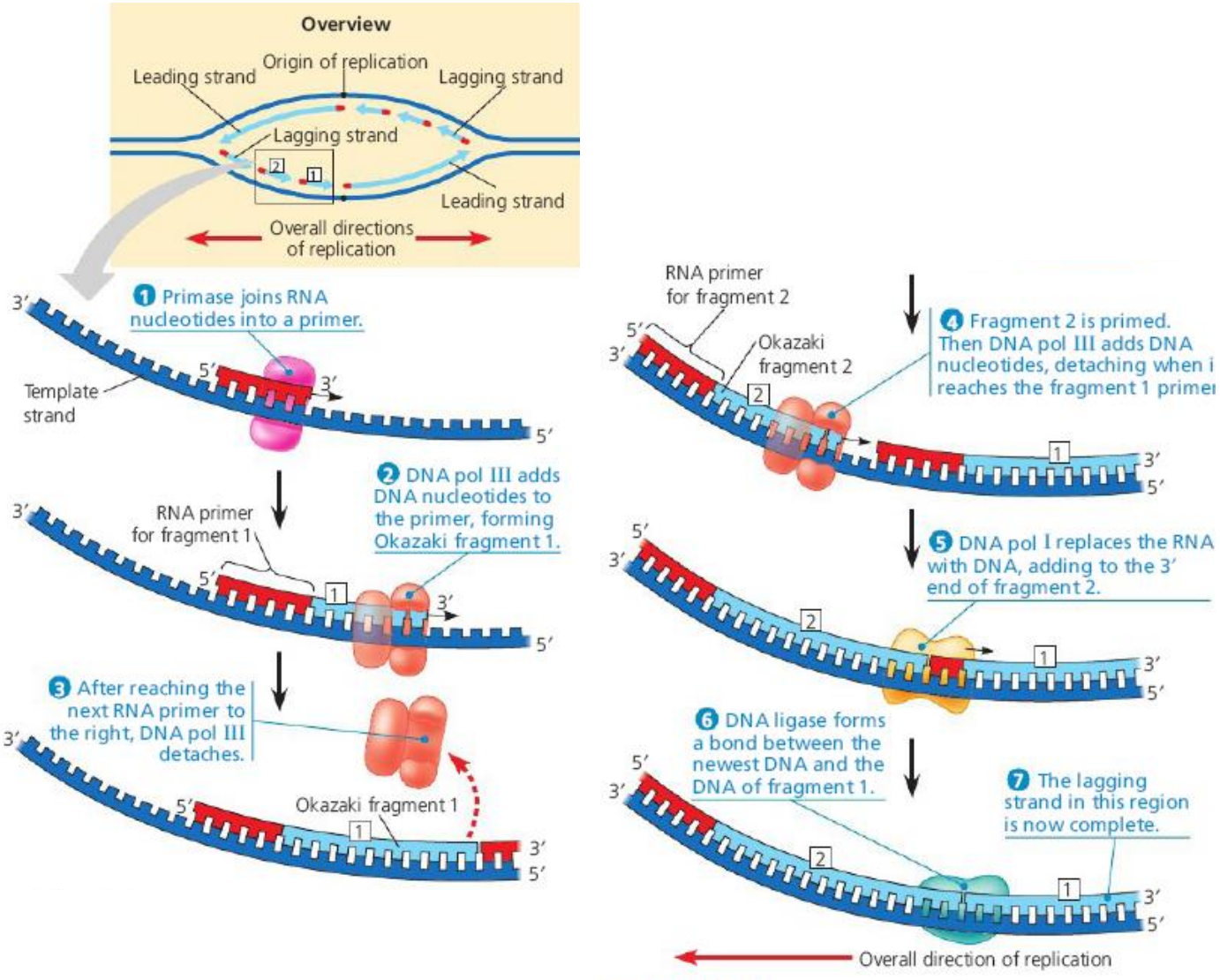
- The leading strand is synthesised **continuously** (in the 5' to 3' direction) towards the replication fork.
- Only one primer is required for DNA polymerase to synthesise the leading strand per replication fork.



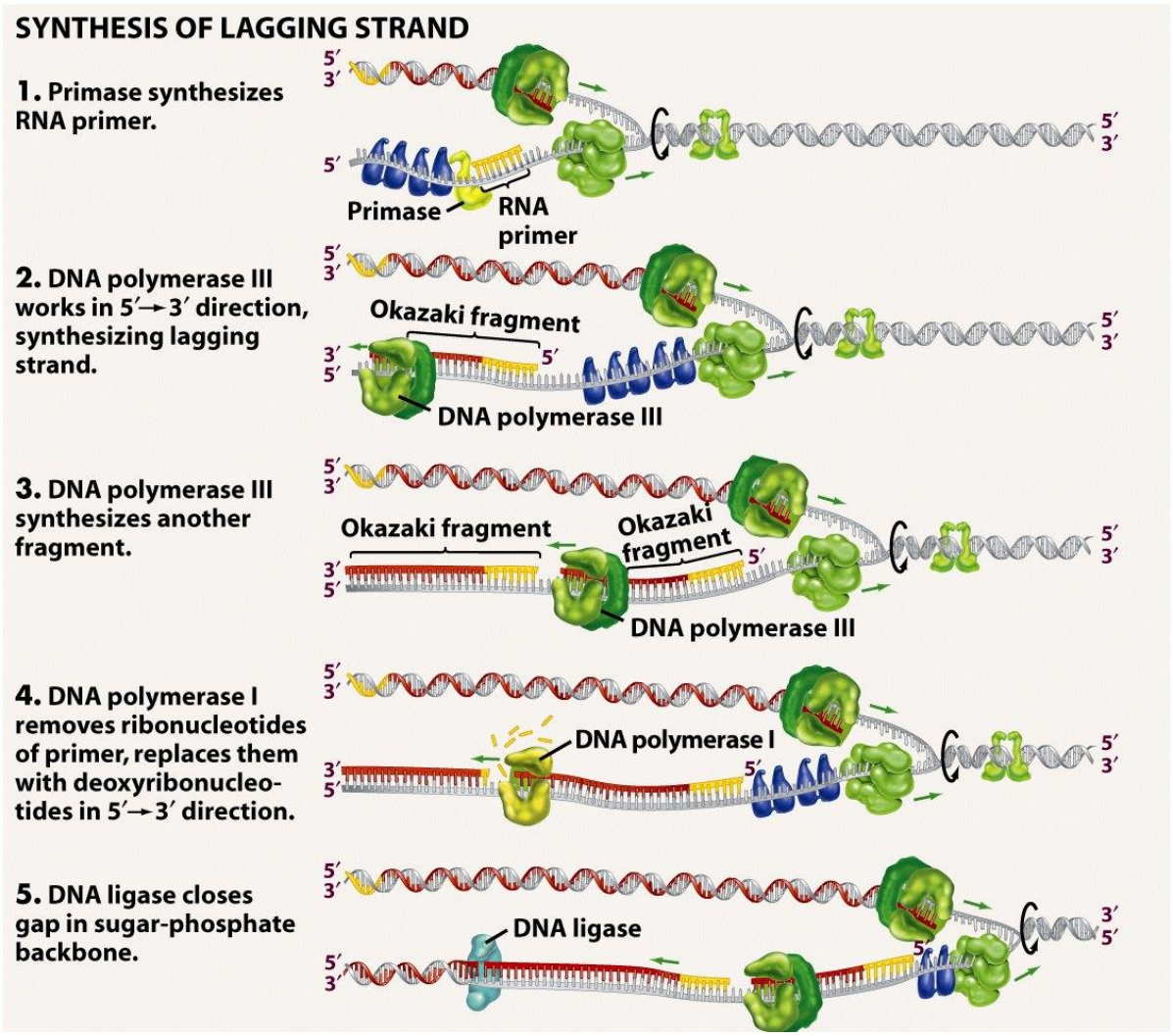
Synthesis of leading strand

Lagging strand

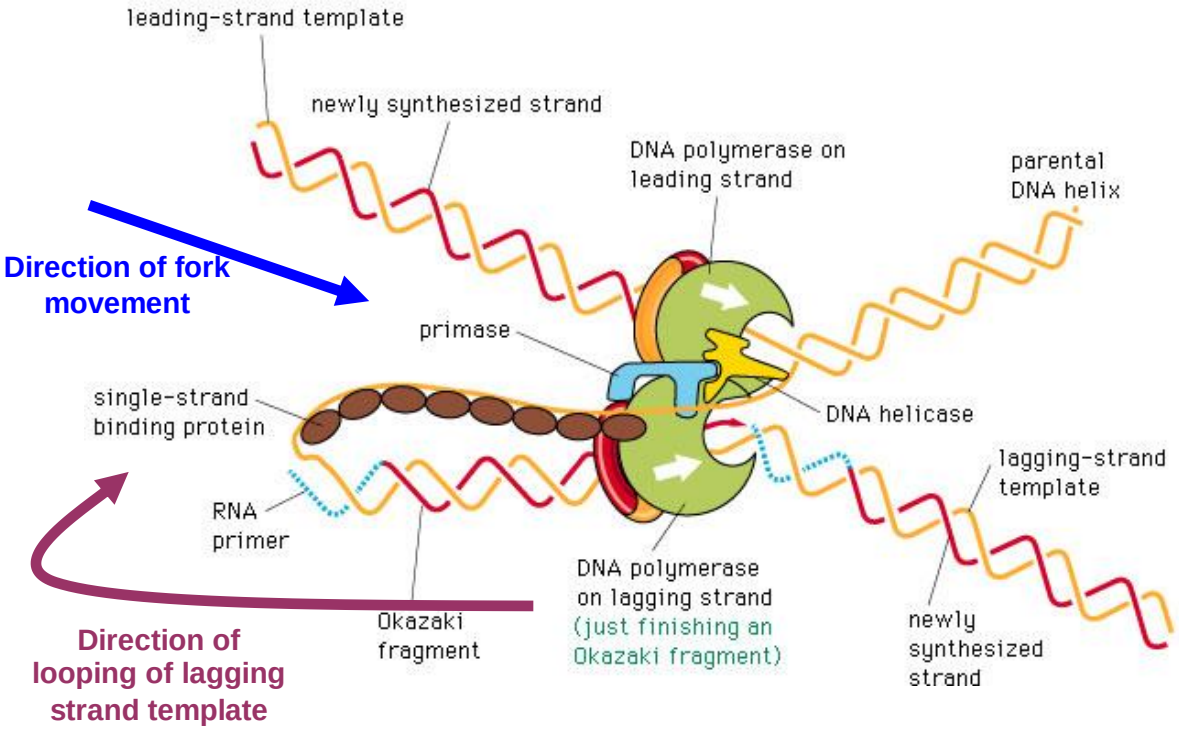
- The lagging strand is synthesised **discontinuously**, via a series of **Okazaki fragments** (in the 5' to 3' direction) away from the replication fork.
- Each Okazaki fragment needs to be primed separately.
- The **RNA primers** are **excised** and **replaced with deoxyribonucleotides** by another **DNA polymerase**.
- The adjacent Okazaki fragments must be linked together. The 3'OH end of one fragment is adjacent to the 5' phosphate end of the previous fragment. **DNA ligase** catalyses the formation of **phosphodiester bond** between the two Okazaki fragments.



Synthesis of lagging strand



Synthesis of lagging strand



Current model of DNA replication complex

(iv) Step 4: **Termination**

- The product of replication is thus **two daughter DNA molecules** formed from one original parental DNA molecule. Each of the daughter DNA molecules is identical to the original parental molecule.
- Each daughter **molecule** contains **one strand conserved from the parental molecule** and **one newly synthesised strand**.

Summary of DNA replication

1. A portion of double helix is **unwound** and **unzipped** at the **origin of replication** by **DNA helicase** and **hydrogen bonds** between parental strands are broken.
2. Each strand is bound and stabilised by **single-stranded binding proteins**, preventing them from rewinding behind the replication fork.
3. **DNA topoisomerase** introduces a **break in a single strand**, thus allowing the strand to **rotate around the break**, and **reseals the strand**, thus eliminating positive supercoil in front of the replication fork.
4. Each parental strand acts as a **template** for the synthesis of new daughter strand.
5. **Primase** synthesises **RNA primers** in the **5' to 3' direction** by adding ribonucleotides via complementary base pairing using parental DNA strand as template.
6. **DNA polymerase** adds **deoxyribonucleotides** to 3'OH end of RNA primer, via **complementary base-pairing** with the parental strand and catalysing the formation of **phosphodiester bonds** between nucleotides.
7. **Adenine** (A) always pairs with **thymine** (T) with **two** hydrogen bonds and **guanine** (G) always pairs **cytosine** (C) with **three** hydrogen bonds.
8. The leading strand is synthesised **continuously** in the **5' to 3' direction** and the lagging strand is synthesised **discontinuously**, via a series of **Okazaki fragments** in the **5' to 3' direction**.
9. The **RNA primers** are **excised** and **replaced with deoxyribonucleotides** by another **DNA polymerase**.
10. **DNA ligase** catalyses the formation of **phosphodiester bond** between the two Okazaki fragments.
11. The product of **semi-conservative replication** is two **DNA daughter molecules** formed from one original parental DNA molecule and each daughter **molecule** contains **one strand conserved from the parental molecule** and **one newly synthesised strand**.

DNA polymerases

Three main DNA polymerase enzymes have been characterised in *E.coli.* DNA polymerase I is involved in the DNA repair and excising RNA primers and replacing with deoxyribonucleotides. DNA polymerase II is probably involved in DNA repair. DNA polymerase III, a multi-subunit protein, responsible for synthesis of new strands of DNA.

At least 11 eukaryotic DNA polymerase enzymes have been identified. DNA polymerase α, β δ and ε are involved in nuclear DNA replication and repair and with DNA polymerase γ is involved in mitochondrial DNA replication.

A common feature of all DNA polymerases is that they cannot initiate synthesis of a DNA chain from free deoxyribonucleotides. They require a primer to provide a free 3'OH end before they can add deoxyribonucleotides to initiate polynucleotide synthesis.

(d) Importance of base-pairing and hydrogen bonding in DNA

(i) Stability of DNA molecule

- In the DNA double helix, the nitrogenous bases are held together by **hydrogen bonds**; two hydrogen bonds between adenine and thymine, and three hydrogen bonds between cytosine and guanine.
- **Hydrogen bonds**, together with **hydrophobic interactions** between the stacked bases, stabilise the structure of the double helix.
- The adjacent nucleotides within each polynucleotide strand are held together by **strong covalent phosphodiester bonds**, which are not easily broken. In this way, the integrity of the DNA base sequence is maintained.

(ii) DNA replication

- DNA is the hereditary material and thus the structure of DNA has to allow for its own replication. It is essential that DNA replicates itself accurately prior to mitosis so that the daughter nuclei have **identical copies of DNA** as the parent nucleus.
- The double helical structure of DNA enables **semi-conservative replication** to occur. During replication, the two parental strands separate and each strand acts as a template for synthesis of new strand via **complementary base pairing**. Each original DNA molecule will give rise to two daughter DNA molecules with **identical base sequence**.
- DNA polymerase might improve the specificity of complementary base pairing at two stages:
  - scrutinise the incoming nucleotide for the proper complementarity with the template (pre-synthetic error control).
  - scrutinise the nucleotide against the template as soon as it is added to the growing strand and DNA polymerase removes the incorrectly paired nucleotide and resumes synthesis (proofreading).

(iii) DNA repair

- DNA is subjected to environmental factors that can cause changes in base sequence, i.e. mutations. The structure of DNA allows for a repair mechanism to operate in the event of such a mutation.
- In the event of a mutation, the intact complementary strand can be used as a template to guide the repair by **DNA repair enzymes**. Such repair mechanisms ensure that the integrity of the base sequence of the DNA molecule is maintained.

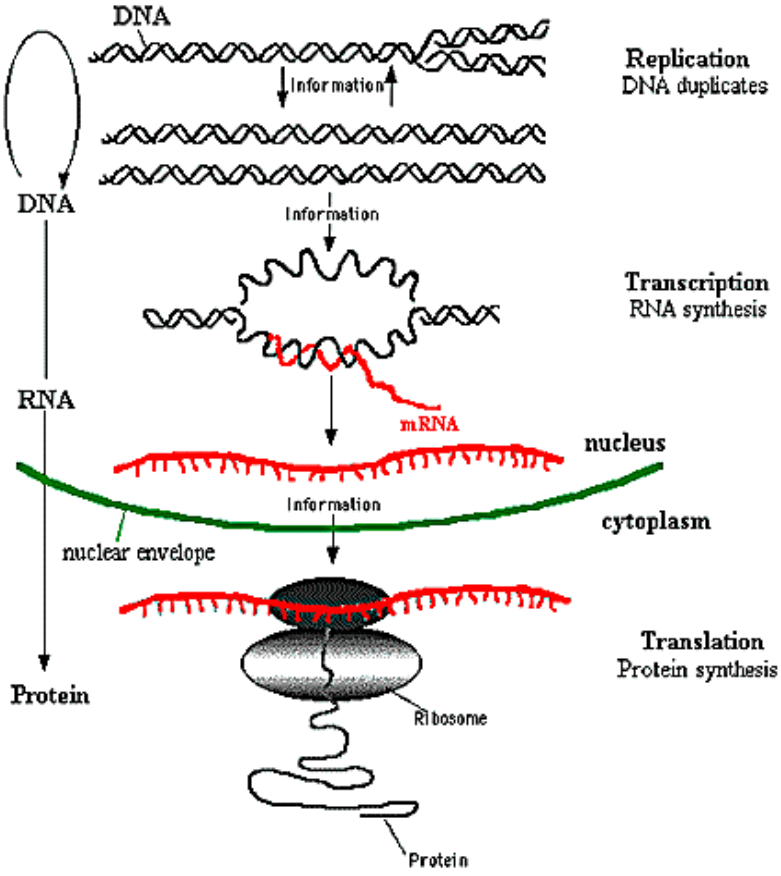
4. Gene Expression in Eukaryotes

(a) The Central Dogma Of Molecular Biology

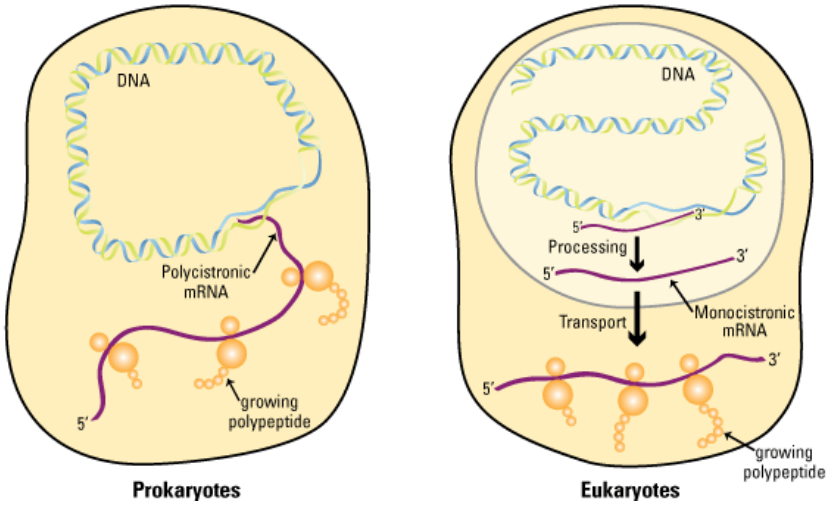
Gene expression is the process by which DNA directs the synthesis of proteins (in some cases, RNA).

The central dogma of molecular biology is a framework for understanding the sequence of transfer of information from DNA, to RNA and finally to protein.

The three stages involved in the central dogma are **replication**, **transcription** and **translation**.



The Central Dogma of Molecular Biology



Transcription and translation in prokaryotes and eukaryotes

(b) Transcription

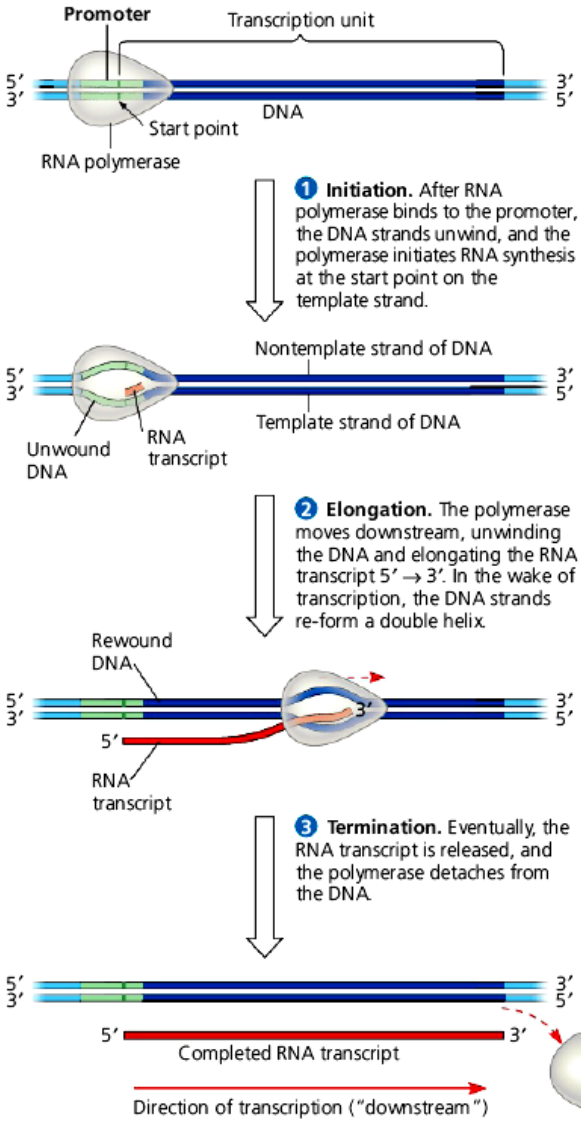
Transcription is the synthesis of a RNA molecule using one of the DNA strands as a **template**.

There are three main stages in transcription:

(i) **Initiation**

(ii) **Elongation**

(iii) **Termination**



Overview of Transcription

Sequences are conventionally written so that transcription proceeds from left (upstream) to right (downstream), which corresponds to the direction of transcription.

The synthesised RNA is

- similar to the **non-template DNA strand**.
- complementary to the **template DNA strand**.

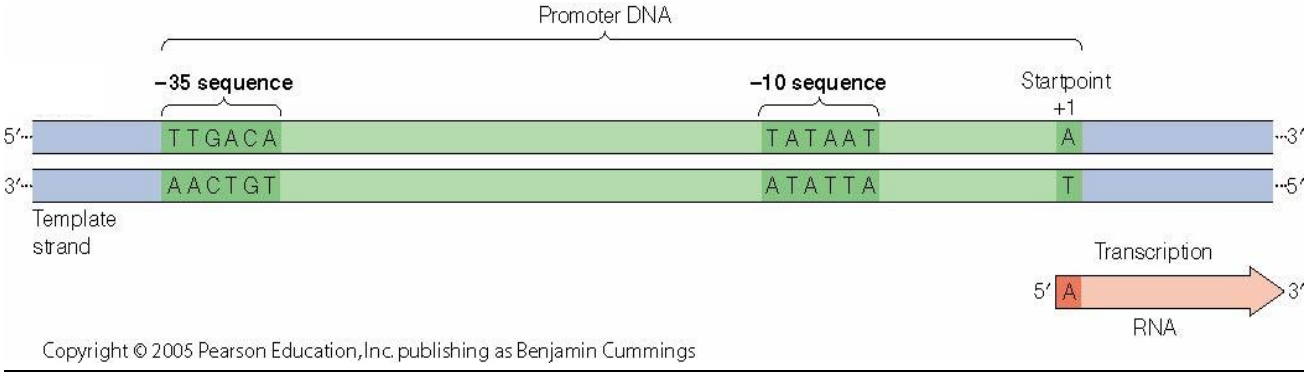
Transcription is the general term for the synthesis of any kind of RNA using DNA as a template. There are three main types of RNA molecules synthesised by transcription:

- messenger RNA (mRNA)
- transfer RNA (tRNA)
- ribosomal RNA (rRNA)

(i) **Initiation**

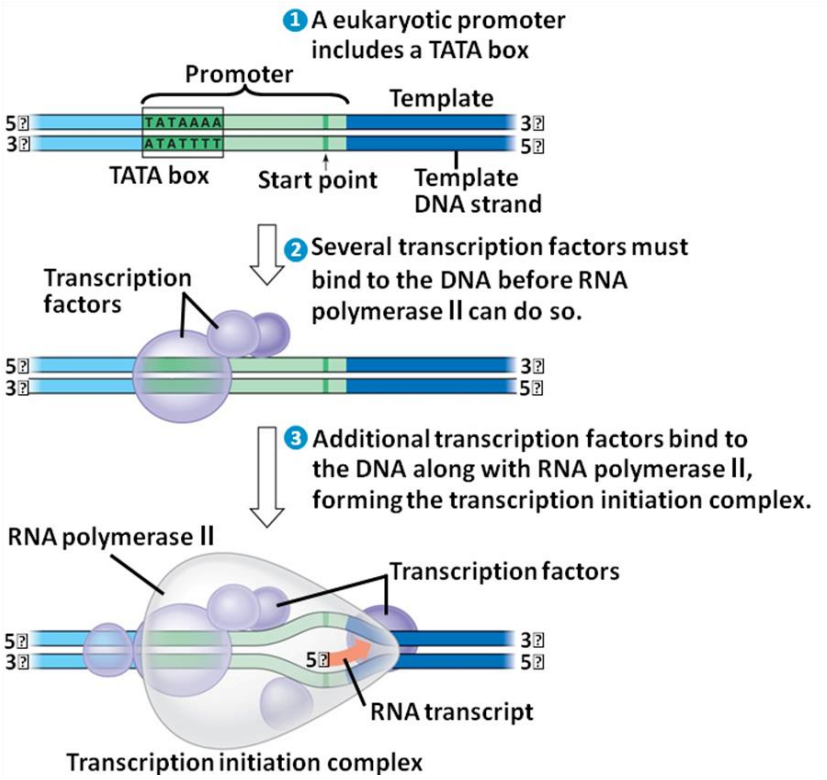
In prokaryotes,

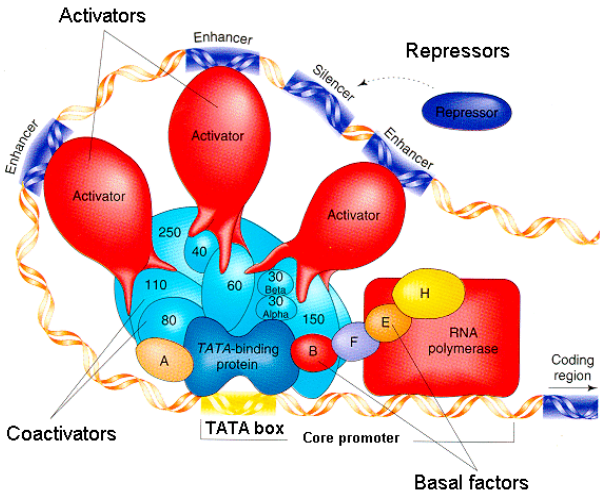
- Sigma factor of RNA polymerase **recognises** and **binds** to the double stranded DNA at the both the **-35 and -10 sequences** of **promoter**.
  - The prokaryotic promoter has two consensus sequences located at 35 and 10 base pairs upstream of the transcription start site.
  - The -10 sequence is a consensus sequence of 5'–TATAAT–3' known as a **Pribnow box**.
- Sigma factor is then released from the core enzyme.
- RNA polymerase unwinds and **separates** the two strands of DNA by breaking **hydrogen bonds** between bases.
- The template strand is available for complementary base pairing with **ribonucleotides**.



In eukaryotes,

- **TATA binding protein** (TBP) **recognises** and **binds** to the **TATA box** of **promoter**.
  - The eukaryotic promoter has a sequence called the **TATA box**, located at 25 base pairs upstream of the transcription start site (i.e. **-25 position**).
  - The promoter can be dispersed over more than 200 base pairs, bound by transcription factors.
- **General transcription factors and RNA polymerase** are recruited to form the **transcription initiation complex**.
- RNA polymerase unwinds and **separates** the two strands of DNA by breaking **hydrogen bonds** between bases.
- The template strand is available for complementary base pairing with **ribonucleotides**.



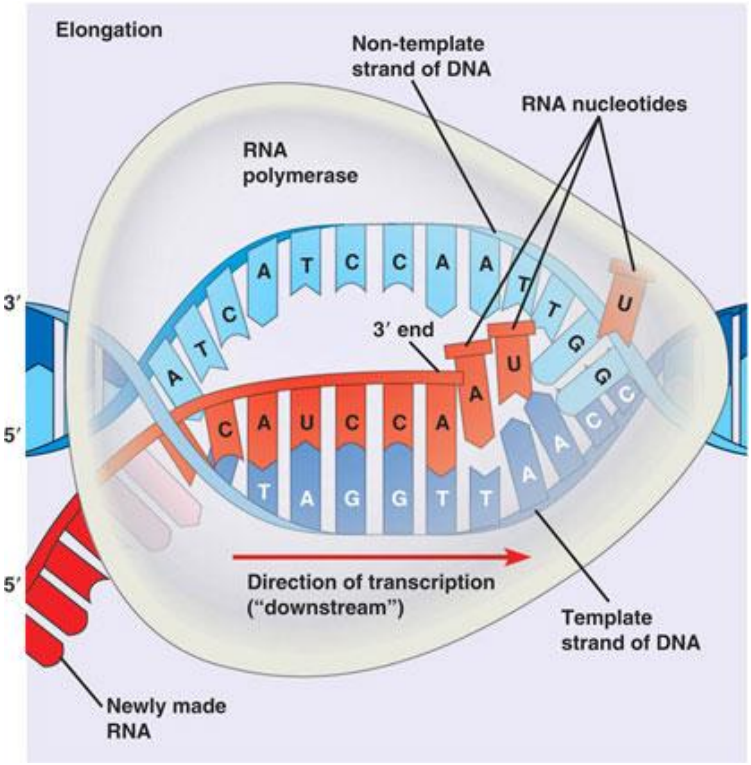


Transcription initiation complex

(ii) **Elongation**

In prokaryotes and eukaryotes,

- **RNA polymerase** adds **ribonucleotides** (more specifically ribonucleoside triphosphates, NTP), to the free **3'OH** end of the growing RNA chain
- **RNA polymerase** catalyses the synthesis of a new strand of RNA in the **5' to 3' direction** via **complementary base pairing** with the **template strand**.
- RNA polymerase catalyses the formation of a **phosphodiester bond** between the 3'OH end of the RNA and 5' phosphate group of the NTP added.
- The energy for this process comes from the two phosphate groups which are removed as the nucleotide joins the growing end.
- As the enzyme moves, it separates the DNA helix to expose a new segment of the template strand.
- A short RNA-DNA hybrid is formed in the unwound region. Behind the unwound region, the DNA template strand pairs with its non-template strand to reform the double helix. The RNA emerges as a free single strand.

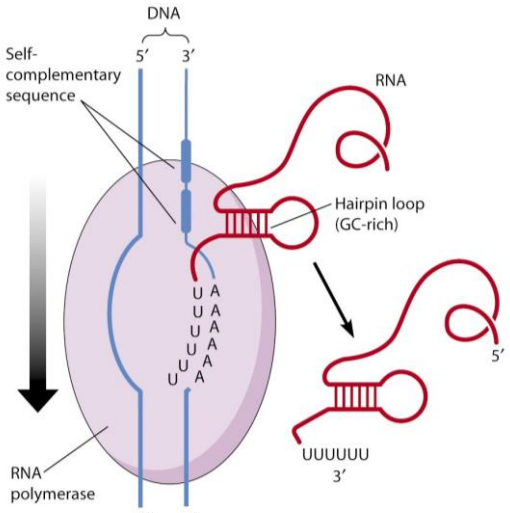


Transcription elongation

In prokaryote *E. coli*, two termination mechanisms have been identified: **rho (ρ)-dependent** and **ρ-independent termination**.

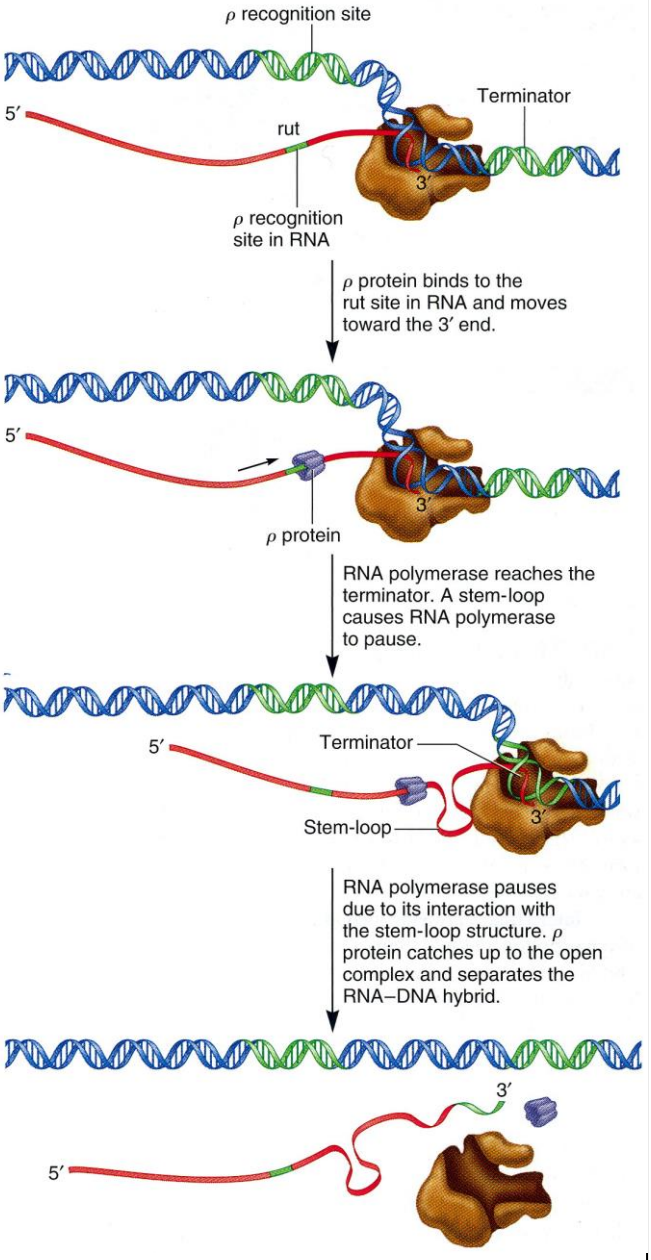
ρ-independent termination

- DNA sequence slightly upstream from terminator sequence encodes RNA sequence that is GC-rich.
- This spontaneously forms the stem-loop structure, which causes the RNA polymerase to pause in synthesis of the RNA transcript.
- The pause occurs after the sequence of uracils is transcribed and the UA base pairs spontaneously denature, releasing the RNA transcript and RNA polymerase from DNA.
- This is possible as UA base pairs have only two hydrogen bonds and are less stable than GC base pairs.



ρ-dependent termination

- ρ-dependent termination requires the ρ-protein which functions as helicase to separate RNA-DNA hybrid regions.
- A ρ-recognition site for the binding of ρ-protein can be found near the 3' end of newly synthesised RNA. The ρ-protein binds to RNA and moves in the direction of RNA polymerase.
- DNA sequence slightly upstream from terminator sequence encodes RNA sequence that is GC-rich. This spontaneously forms the stem-loop structure which causes RNA polymerase to pause in RNA synthesis.
- The pause allows for ρ-protein to catch up with the RNA polymerase, passing through and thus breaking the hydrogen bonds between DNA and RNA. The completed RNA strand is thus separated from DNA along with RNA polymerase.



(iii) **Termination**

In prokaryotes,

- Termination occurs after a **terminator sequence** found on the DNA template strand is transcribed.
- The short RNA-DNA hybrid is separated, releasing the newly synthesised RNA transcript and RNA polymerase.
- No further modification is required before translation.

In eukaryotes,

- Termination occurs after the terminator sequence, **polyadenylation signal sequence** found on the DNA template strand, is transcribed.
- It codes for a **polyadenylation signal sequence (AAUAAA)** in the **pre-mRNA**.
- At a point about 10-35 nucleotides downstream from the AAUAAA sequence, pre-mRNA transcript is cleaved, releasing the pre-mRNA and RNA polymerase.
- After the pre-mRNA is released, post-transcriptional modification occurs before the **mature mRNA** is transported to the ribosomes in cytoplasm via the nuclear pore.

About RNA Polymerases

In prokaryotes, the RNA polymerase is a holoenzyme which can be separated into 2 components.

- **Core enzyme**: made up of 5 subunits: α^{I}, α^{II}, β, β' and ω subunits
- **Sigma factor**: ensures RNA polymerase binds in a stable manner to DNA at the promoter. Sigma factor is released when the RNA chain reaches 8-9 bases, leaving the core enzyme to undertake elongation.

In eukaryotes, there are 3 main classes of RNA polymerases.

- **RNA polymerase I** transcribes genes that code for **rRNA**.
- **RNA polymerase II** transcribes genes for **mRNA** and
- **RNA polymerase III** transcribes genes that code for **rRNA, tRNA and other small RNAs**.

These three classes of enzymes have similar structures, with 2 large subunits and many smaller subunits.

None of these RNA polymerases recognise their promoters directly and require transcription factors for them to bind to their promoters.

(c) **Post-Transcriptional Modification**

In eukaryotes, a newly synthesised mRNA,

- is known as the **primary mRNA / pre-mRNA**.
- must undergo **post-transcriptional modification** before it can be translated.

The post-transcription modifications are:

(i) **5' Cap**

(ii) **3' Poly-A Tail**

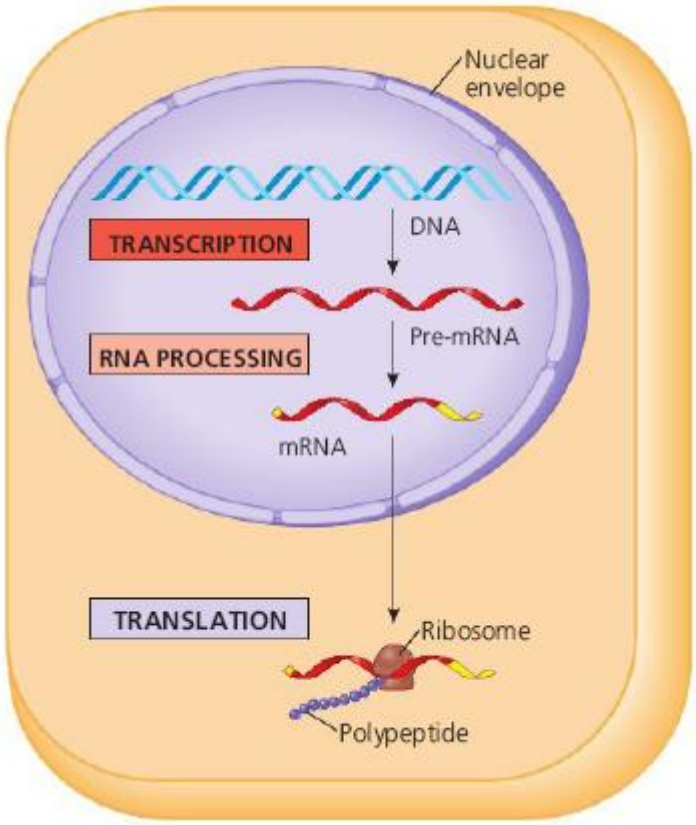
(iii) **RNA Splicing**

These modifications **occur in the nucleus**.

After post-transcriptional modifications, a **mature RNA** is produced that is ready to be transported to ribosomes for translation.

Prokaryotes do not carry out post-transcriptional modification as

- prokaryotes lack a nucleus
- translation occur simultaneously while transcription is still taking place.

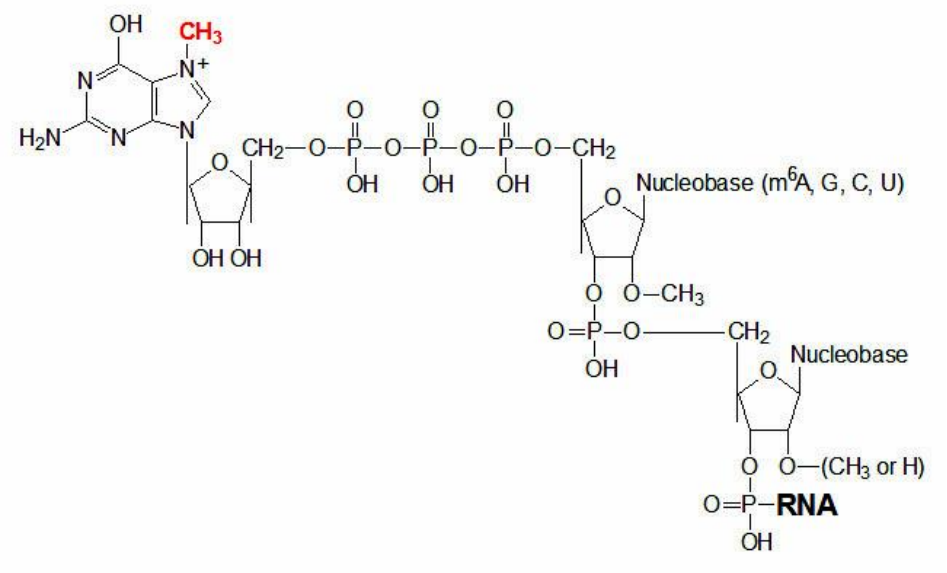


Post transcriptional modification

(i) **5' Cap**

Structure

- **Methyl guanosine nucleoside triphosphate** is added to the first nucleotide by a **5' – 5' triphosphate linkage**. This structure is called a **5' methylguanosine cap**.
- Addition of the 5' methylguanosine cap is catalysed by a nuclear enzyme, **guanylyl transferase**.
- 5' cap is added after transcription of the first 20-40 nucleotides.



5' methylguanosine cap

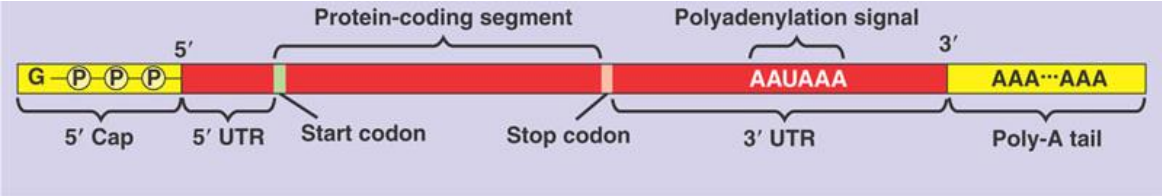
Function

- **Facilitate the export of the mature mRNA from the nucleus into the cytoplasm**.
- **Protects the mRNA from 5' exonucleases** (enzymes that hydrolyses phosphodiester bonds between nucleotides from the end of polynucleotide chain), hence confers **stability** to the mRNA.
- **Facilitates in binding of ribosomes** to mRNA.

(ii) **3' Poly-A Tail**

Structure

- The polyadenylation signal sequence, AAUAAA is a signal for transcription termination and polyadenylation of the 3'end of mRNA
- About 200 **adenine** residues are added to the **3' end** of the pre-mRNA called the **3' poly-A tail**.
- Addition of **poly-A tail** is catalysed by the enzyme **poly-A polymerase**.
- The poly-A sequence is not coded in the DNA but is added to the RNA in the nucleus after transcription.



5' methylguanosine cap and 3' Poly-A tail

Function

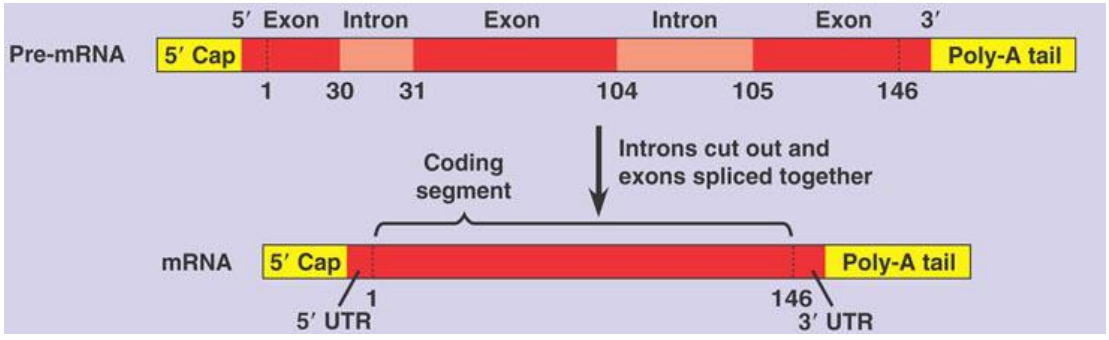
- **Facilitate the export of the mature mRNA from the nucleus into the cytoplasm**.
- Slow down degradation by **3' exonucleases**. The longer the poly-A tail, the longer is the half-life of mRNA.

(iii) **RNA Splicing**

The average length of a transcription unit along a human DNA molecule is about 27000 nucleotide pairs, so the primary RNA transcript is also that long.

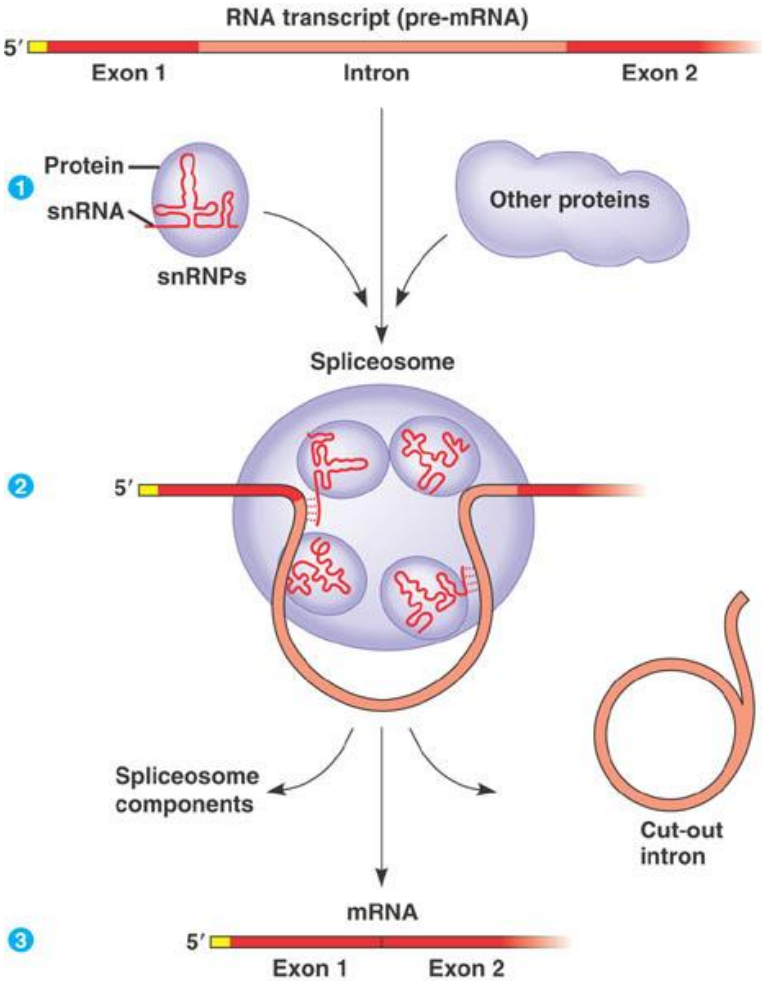
But it takes only about 1200 nucleotides in RNA to code for an average-sized protein of 400 amino acids. This means that most eukaryotic genes and their RNA transcripts have long non-coding stretches of nucleotides (i.e. regions that are not translated).

- **RNA splicing** is the **removal of introns** and **joining of the exons** together.
- Non-coding sequences on mRNA are called **introns**. They are interspersed between coding sequences called **exons**.
- Exons are eventually translated into amino acid sequences (except UTRs of the exons at the ends of the RNA.)
- The introns are **removed** from the pre-mRNA and the exons are **ligated** together, forming an mRNA molecule with a continuous coding sequence.



RNA splicing

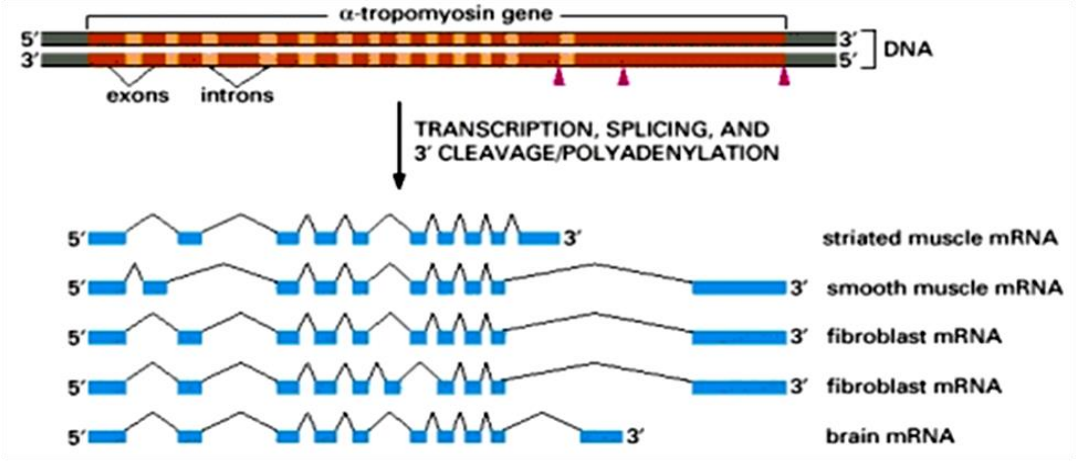
- Small nuclear ribonucleoproteins (snRNPs) are located in the nucleus and comprise RNA and proteins.
- Specific snRNPs recognises and bind to 5' splice site and the 3' splice site.
- Additional proteins interact with snRNPs to form **spliceosome**.
- As the exons are brought closer together, intron will loop.
- Spliceosome **excise the introns** and **join the exons** that flanked the intron, releasing introns in a lariat structure (which is subsequently degraded).



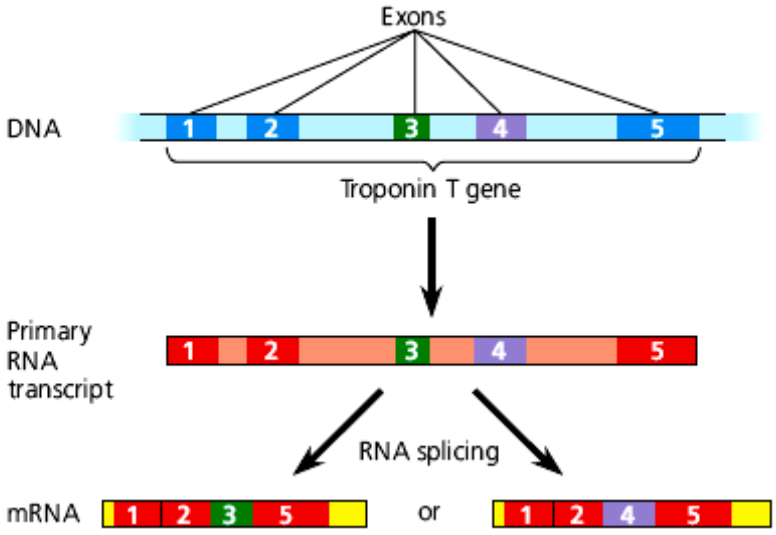
The role of spliceosome in RNA Splicing

Alternative Splicing

- When pre-mRNAs are processed by joining exons in different combinations, **different mature mRNAs** are produced from the same pre-mRNA.
- Thus, a single gene can code for more than one kind of polypeptide.
- It increases number and variety of proteins in the cell without increasing the genome size.
- The number of proteins produced by an organism can be greater than the number of genes.
- e.g. α-tropomyosin gene or Troponin T gene undergo alternative splicing in different cell types e.



Alternative splicing of the α-tropomyosin gene from rat



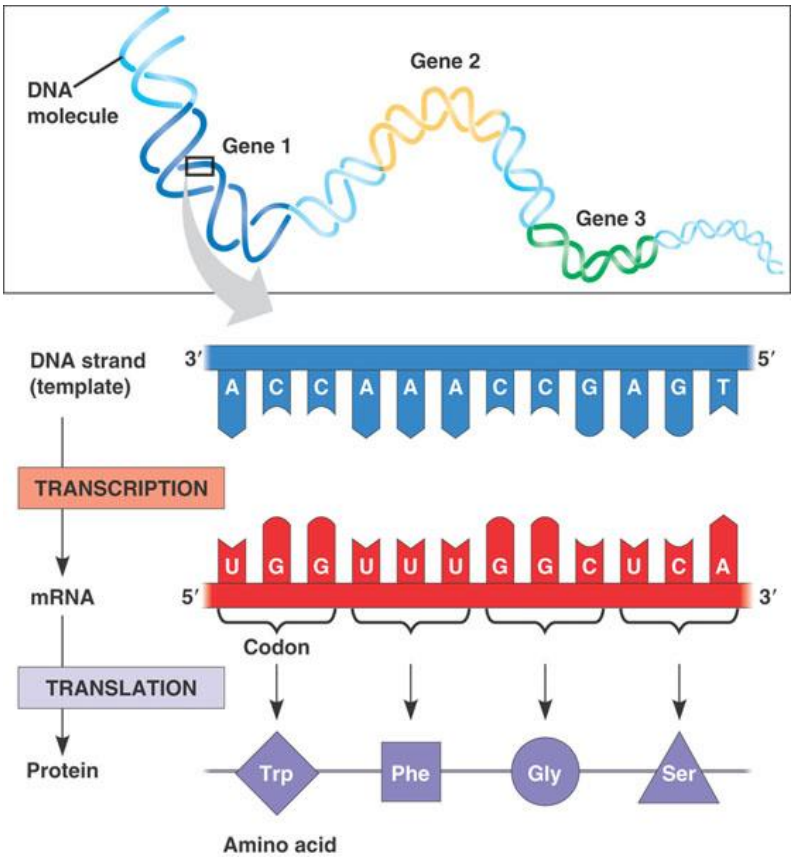
Alternative splicing

(d) The Genetic Code

The genetic code defines a mapping between triplets of bases in genetic material (usually RNA) and amino acids in a polypeptide

- **Gene:** a specific nucleotide sequence of DNA which codes for **a functional product (RNA or protein)**

- **Codon: three consecutive nucleotide bases (a triplet of bases)** in mRNA that codes for a specific amino acid in a protein. e.g. base triplet GGC codes for glycine.



Characteristics of Genetic Code:

(i) **Triplet** : Three consecutive nucleotide bases specify for one amino acid

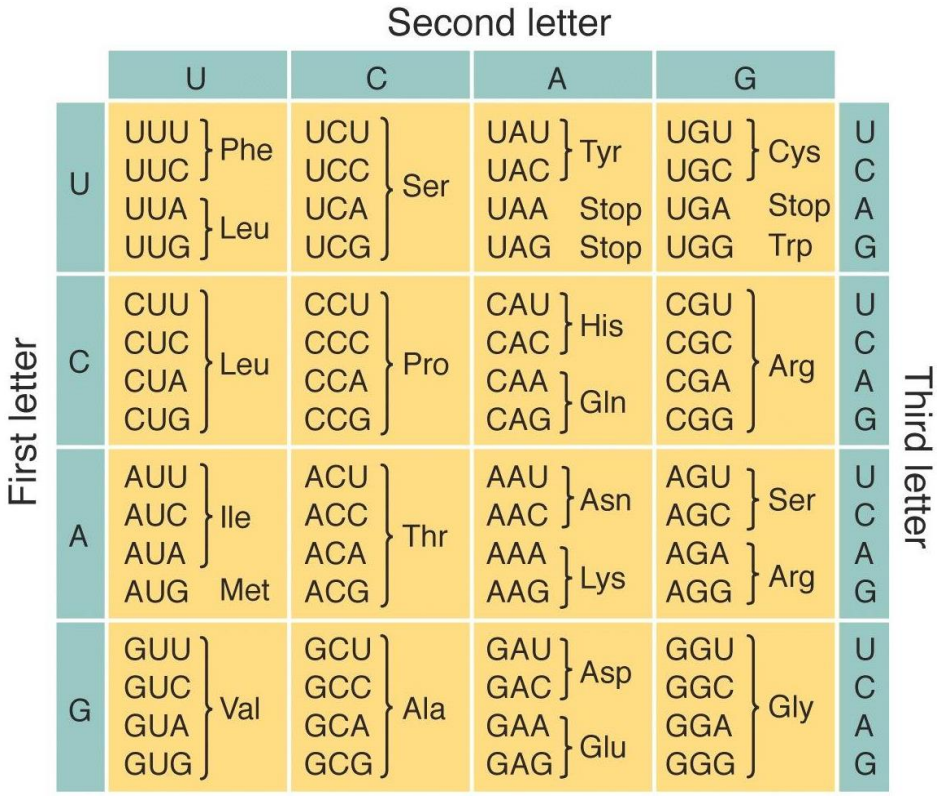
(ii) **Degenerate** / **redundant** : More than one codon can code for the same amino acid. The codons that code for the same amino acid usually vary in the 3rd base

(iii) **Unambiguous** : No codon specifies for more than one amino acid

(iv) **Non-Overlapping** : Successive triplets are read in order with no overlapping of nucleotide bases

(v) **Unpunctuated** : There are no gaps between adjacent codons

(vi) **Universal** : The genetic code is common to almost all organisms, although there are a number of exceptions to this rule, particularly in unicellular eukaryotes and organelles genes (e.g. mitochondrion) of some species.



The codon table for mRNA

- Some codons signal the start or termination of synthesis of a polypeptide chain.
  - **start codon**: **AUG** (codes for methionine)
  - **stop codons**: **UGA / UAA / UAG**
- 61 of the 64 triplets code for amino acids. The three codons that do not code for amino acids are the stop codons.

(e) **Translation**

Translation is the synthesis of protein using mRNA as a **template**.

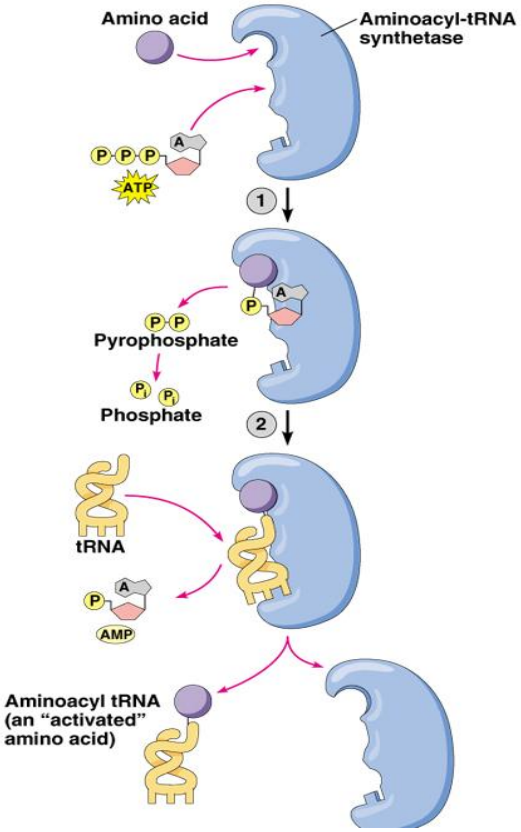
The accurate translation of a genetic message requires two recognition steps:

- a correct match between a tRNA and an amino acid during **amino acid activation**,
- a correct match between a tRNA anticodon and the mRNA codon (i.e. **complementary base pairing** of codon and anticodon).

Amino Acid Activation

Before translation can take place, amino acid activation must occur. A tRNA that binds to an mRNA codon coding for a specific amino acid must carry that amino acid to the ribosome.

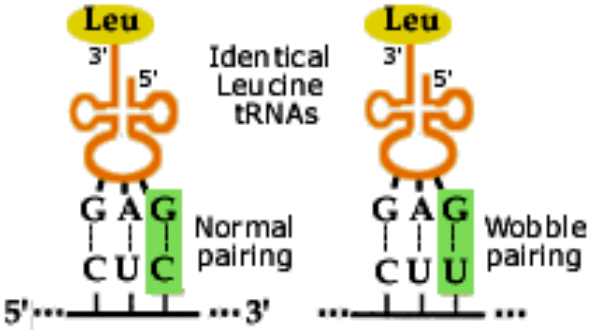
- Each amino acid is joined to the correct tRNA by a specific enzyme called an **aminoacyl-tRNA synthetase**.
- The active site of **aminoacyl-tRNA synthetase** recognises and binds to **a specific pair of amino acid and tRNA**.
- There are at least 20 different synthetases, one for each amino acid.
- The synthetase catalyses the **covalent attachment of an amino acid** to the **3' acceptor stem of the corresponding tRNA**.
- Each tRNA has a specific **anticodon** which binds to the **codon** on mRNA by **complementary base-pairing**.
- This process is driven by energy from the hydrolysis of ATP.
- The resulting **aminoacyl tRNA**, also called an **activated amino acid**, is released from the enzyme and delivers its amino acid to a growing polypeptide chain on a ribosome.
- This process is genetically controlled and takes place continuously in the cytoplasm.



Amino acid activation

Pairing of Codon and Anticodon

- The second recognition step involves correct complementary base pairing between the **tRNA anticodon** and an **mRNA codon**.
- If one tRNA existed for each mRNA codon specifying an amino acid, there would be 61 tRNAs. However, there are only about 45, meaning some tRNAs must be able to bind to more than one codon.
- The rules for base pairing between the third nucleotide base of a codon and the corresponding base of a tRNA anticodon are not as strict.
- This relaxation of the base-pairing rules is called **wobble effect**. The wobble effect explains why the synonymous codons for a given amino acid can **differ in their third base**, but not usually in their other bases.
- For example, the base G at the 5' end of a tRNA anticodon can pair with either C or U in the third position of an mRNA codon.



Wobble effect

There are three main stages in translation:

(i) **Initiation**

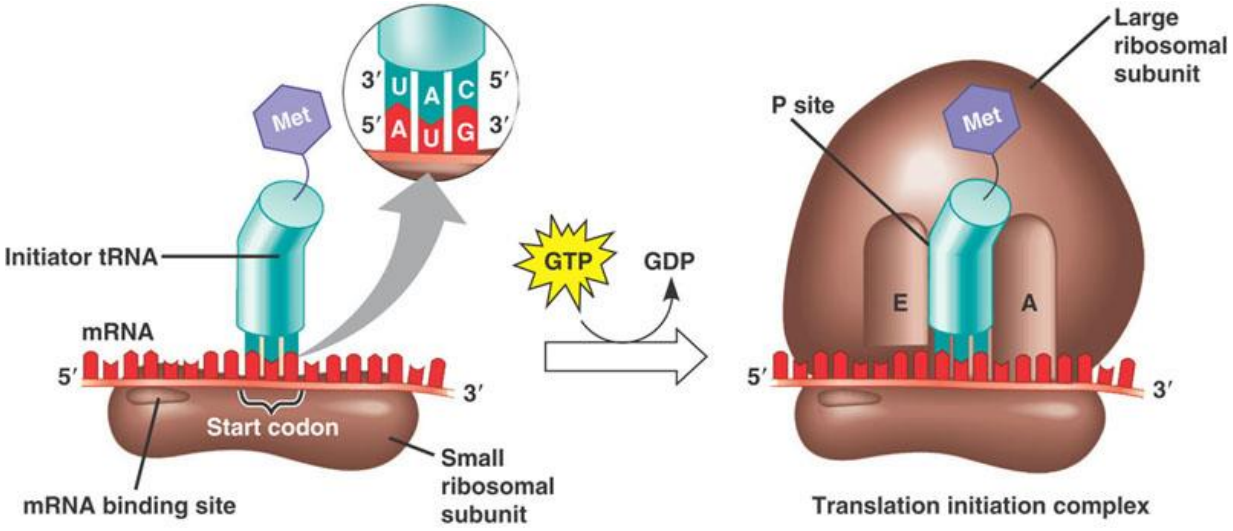
(ii) **Elongation**

(iii) **Termination**

(i) **Initiation**

- A **small ribosomal subunit** binds to **both 5' end of mRNA and a specific initiator tRNA**, which carries the amino acid **methionine**, with the aid of proteins known as **initiation factors**.
  - Initiator tRNA with anticodon, 3'-UAC-5' carries N-formyl methionine in prokaryotes
  - Initiator tRNA with anticodon, 3'-UAC-5' carries methionine in eukaryotes

- The small subunit scans, downstream along the mRNA until it reaches the **start codon**, **AUG**, which signals the start of translation; this establishes the start of a **reading frame** for the mRNA.
- Hydrogen bonds are formed between the **complementary bases** of the methionine-tRNA anticodon and the start codon on mRNA.
- A **large ribosomal subunit** attaches and binds to mRNA, forming the **translation initiation complex**.
- At the completion of the initiation process, the **initiator tRNA sits in the P (peptidyl) site of the ribosome**, and the **vacant A (aminoacyl) site is ready for the next aminoacyl tRNA**.
- A polypeptide is always synthesized in one direction, from the initial methionine at the amino end (i.e. **N-terminus**), toward the final amino acid at the carboxyl end (i.e. **C-terminus**).



Translation initiation

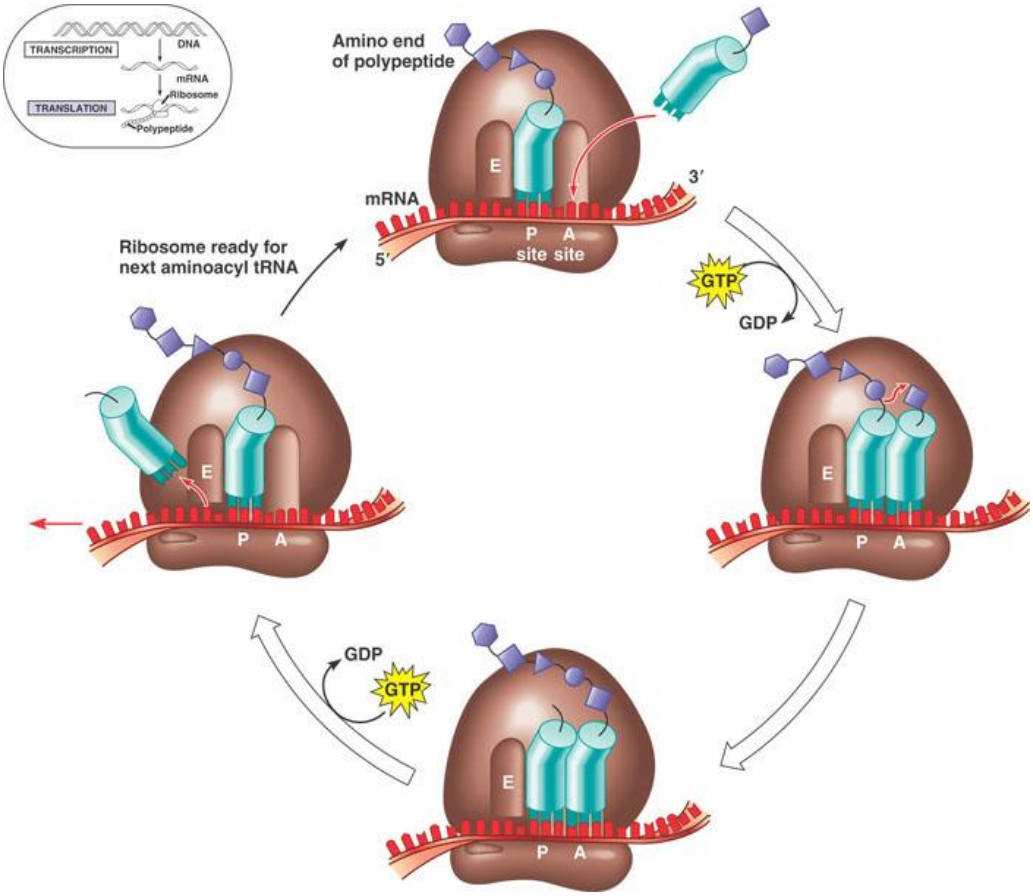
(ii) **Elongation**

Amino acids are added one by one to the preceding amino acid via formation of peptide bonds.

- **Codon recognition**:
  - The anticodon of an incoming aminoacyl tRNA **complementary base pairs** with the mRNA codon in the A site.
  - Energy released by hydrolysis of GTP is required for codon recognition.

- **Peptide bond formation**:
  - **Peptidyl transferase**, found in a section of the rRNA molecule of the large subunit **catalyses the formation of a peptide bond**.
  - Peptide bond is formed between the **new amino acid in the A site** and the **carboxyl end of the growing polypeptide in the P site**.
  - This step transfers the polypeptide to the tRNA in the A site.

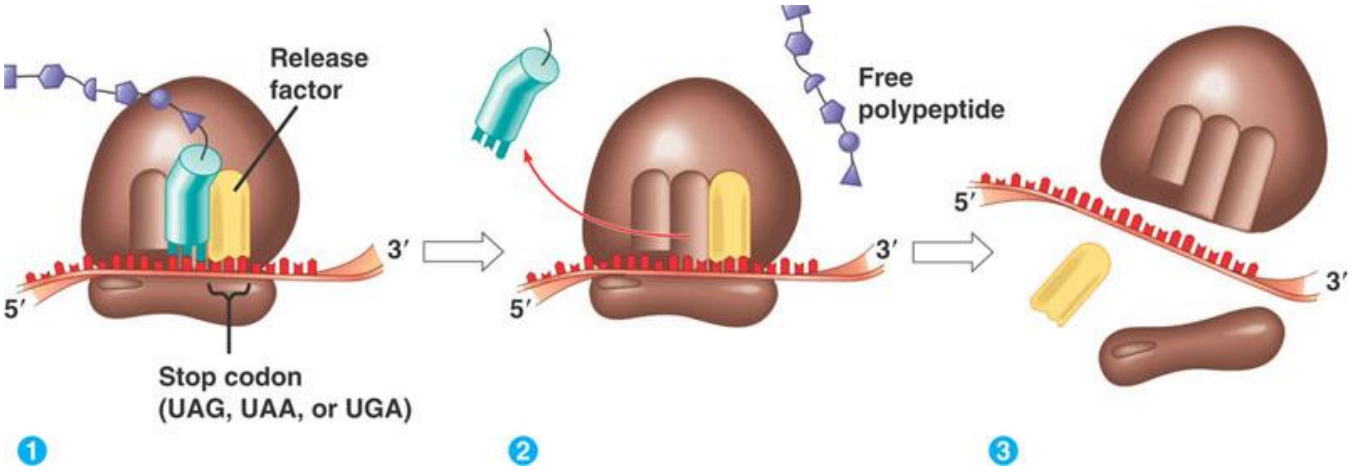
- **Translocation:**
  - The ribosome **translocates** by advancing **3 nucleotides** / **a codon** along the mRNA.
  - The tRNA carrying the polypeptide in the A site is now found in the P site.
  - The empty tRNA in the P site is now in the **E (exit) site, where it is released**.
  - The ribosome then has an empty A site ready for entry of the amino-acyl tRNA corresponding to the next codon on the mRNA.
  - Energy released by hydrolysis of GTP is required for translocation.



Translation elongation

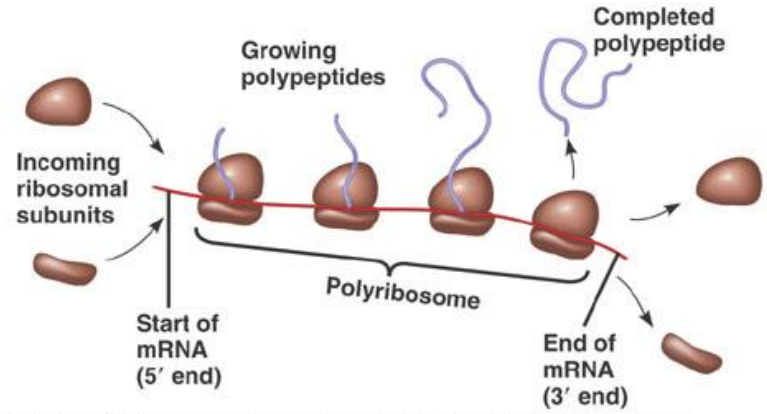
(iii) **Termination**

- Elongation continues until a stop codon in the mRNA reaches the A site of the ribosome.
- The base triplets **UAG**, **UAA** and **UGA** do not code for amino acids but instead signal to stop translation.
- A protein called a **release factor** binds directly to the stop codon in the A site.
- The release factor causes the **addition of a water molecule** instead of an amino acid to the polypeptide chain.
- This reaction **hydrolyses the completed polypeptide from the tRNA in the P site**, releasing the polypeptide through the exit tunnel of the ribosome's large subunit.
- The translation assembly then comes apart.
- Energy released by hydrolysis of GTP is required.

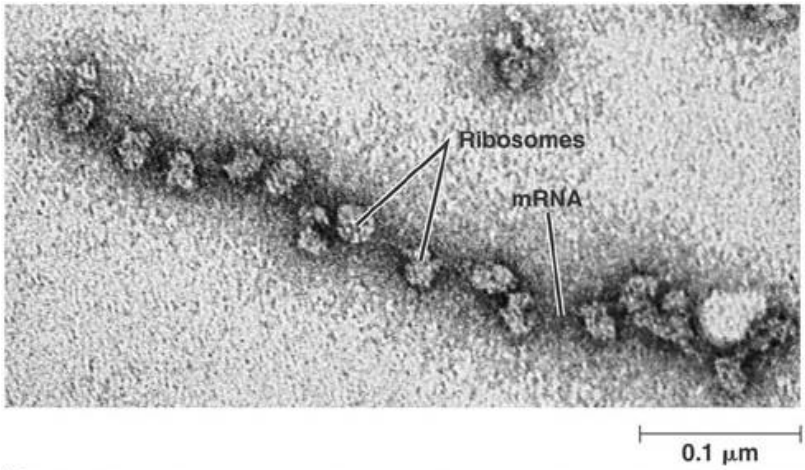


Polyribosomes

- A single mRNA can be used to make many copies of a polypeptide simultaneously because several ribosomes can translate the message from one mRNA at the same time.
- Once a ribosome moves past the start codon, a second ribosome can attach to the mRNA; thus, a number of ribosomes can translate one mRNA at the same time. Such strings of ribosomes are called **polyribosomes / polysomes**.
- They are found in both prokaryotic and eukaryotic cells, enabling a cell to **make many copies of a polypeptide rapidly**.



(a) An mRNA molecule is generally translated simultaneously by several ribosomes in clusters called polyribosomes.



(b) This micrograph shows a large polyribosome in a prokaryotic cell (TEM).

5. Gene Mutation and its Effects

Gene mutation describes **any change in the nucleotide sequence of DNA**.

(a) Types of gene mutations

There are three types of mutation:

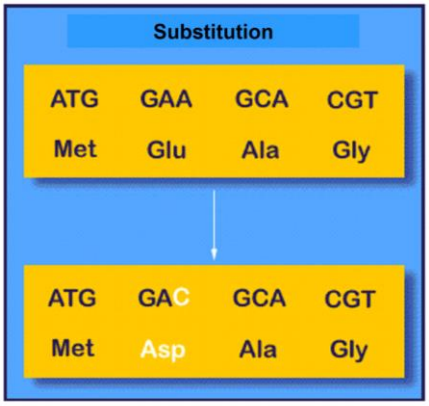
(i) **Substitution**
(ii) **Addition**
(iii) **Deletion**

Point mutations refer to changes in **one nucleotide pair of a gene**. If a point mutation occurs in a gamete or cell that gives rise to gametes, it may be inherited by its offspring and to succession of future generations.

If the mutation has an adverse effect on the phenotype of an organism, the mutant condition is referred to as a **genetic disorder**, or **hereditary disease**.

(i) Substitution

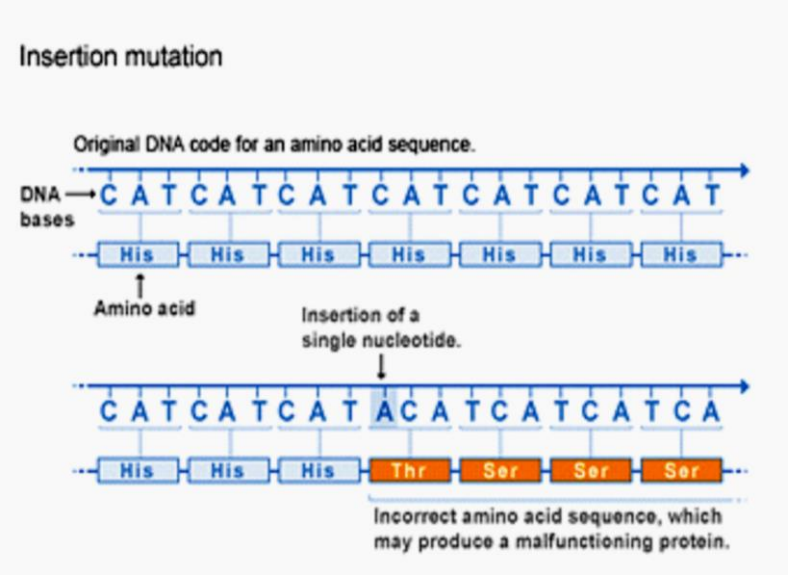
- **Substitution** is the **replacement of one nucleotide pair with another pair of nucleotides**.
- Sometimes, more than one nucleotide pair can be substituted at one time.



Substitution

(ii) **Addition**

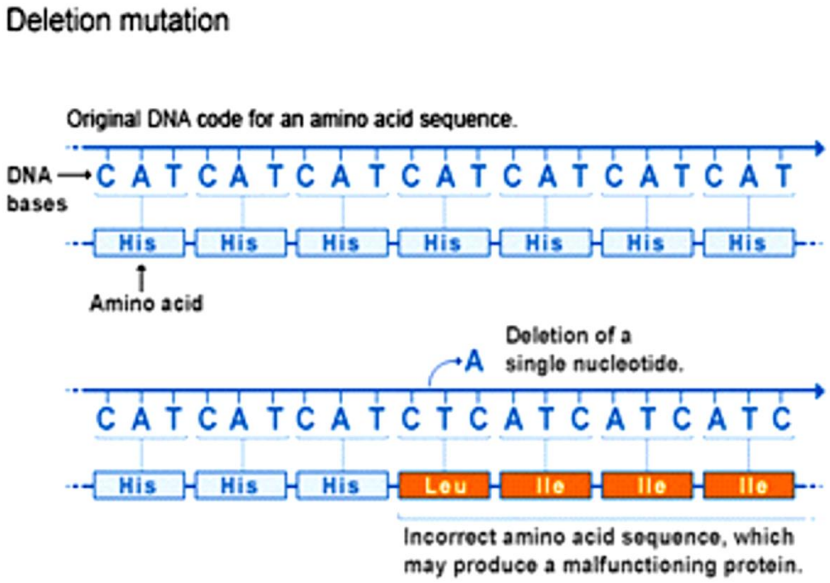
- An **addition** is the **insertion of one or more nucleotide pairs into a DNA sequence**.



Addition

(iii) **Deletion**

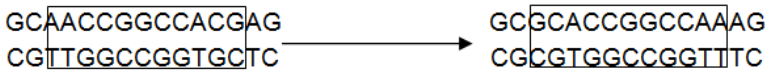
- A **deletion** is a mutation in which **one or more nucleotide pairs are removed from a DNA sequence**.



Other than the three types of gene mutations, there can be two other forms of gene mutations, namely, inversion and duplication.

(iv) Inversion

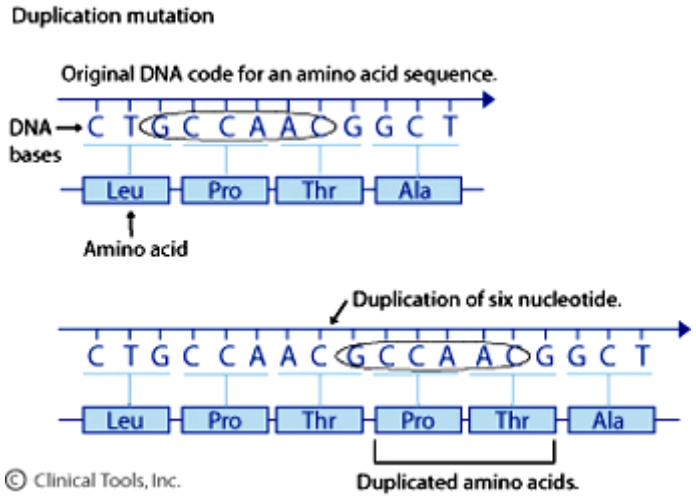
- An inversion is a type of gene mutation in which a section of the gene (more than one nucleotide at a time) is cut, rotated 180^{o} and spliced back to the gene.



Inversion

(v) Duplication

- Duplication is a gene mutation in which a short sequence of DNA is copied next to an identical section, increasing the length of the gene.



Duplication

(b) Effects of Gene Mutations

Base pair substitution may result in **silent**, **missense** or **nonsense** mutations.

(i) Silent Mutations

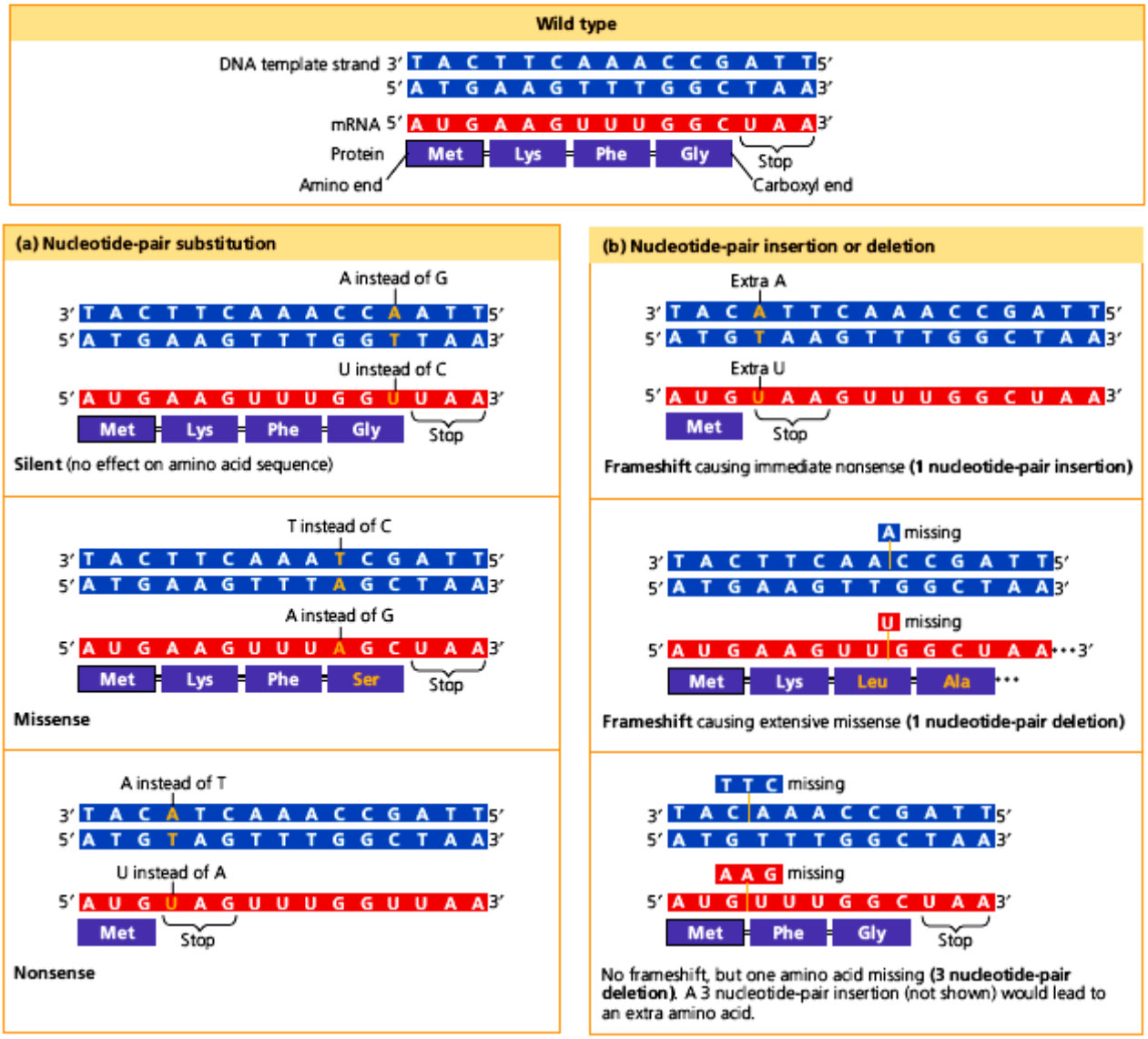
- Silent mutations are mutations that have no effect on amino acid sequence
- No observable effect on phenotype of organism.
- Some substitution mutations are silent due to the **redundancy / degeneracy of the genetic code**.
- E.g. if 3'–CCG–5' on the template strand mutated to 3'-CCA-5', the mRNA codon that used to be GGC would become GGU, and a glycine would still be inserted at the proper location in the protein.

(ii) Missense Mutations

- Other substitutions may change an amino acid but have little effect on the protein because new amino acid may have properties similar to those of the amino acid it replaces.
- The replacement amino acid may also be in a region of the protein where the exact sequence of amino acids is not essential to the protein's function.
- However, if the base-pair substitutions occur in a crucial area in a protein, e.g. in the active site of an enzyme, it may result in an alteration of a single amino acid that will significantly alter protein activity.
- Such mutations occasionally lead to an improved protein or one with novel capabilities, but are usually detrimental, leading to a useless or less active protein that impairs cellular function.
- Substitutions usually results in **missense mutations**. The altered codon still codes for amino acid and thus makes sense, although not necessarily the right sense. This means that although a protein is still produced, the protein may not be as effective.

(iii) Nonsense Mutations

- A point mutation can also change a codon for an amino acid into a stop codon. It **causes translation to be terminated prematurely**.
- This **nonsense mutation** results in the polypeptide formed being shorter than the polypeptide encoded by the normal gene. Nearly all nonsense mutations lead to **non-functional proteins**.



Effects of mutation

Addition and deletion may result in **frameshift mutation**.

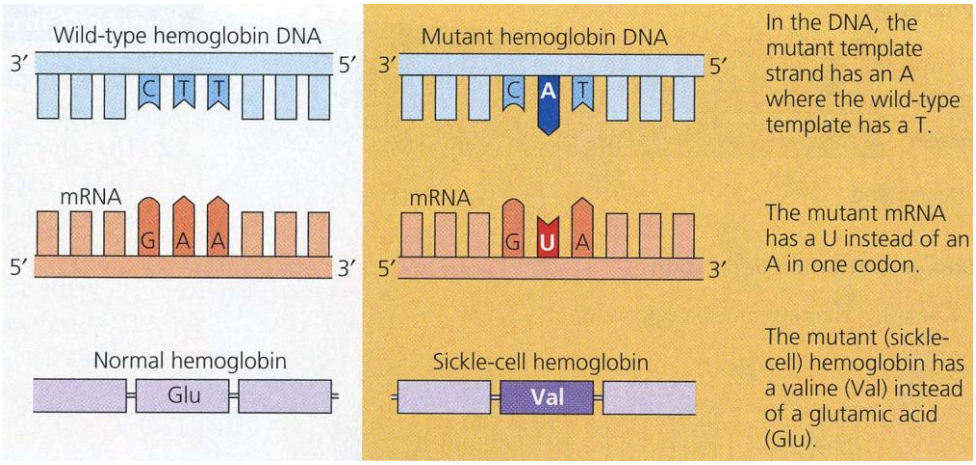
(iv) Frameshift Mutations

- **Additions** or **deletions** often have a disastrous effect on the resulting protein.
- mRNA is read as a series of triplet nucleotide during translation, addition or deletion of nucleotide pair may **alter the reading frame** (triplet grouping) of the genetic message.
- **Frameshift mutations** occur whenever the number of nucleotides inserted or deleted **is not a multiple of three**.
- All the nucleotides that are downstream of the addition or deletion will be **improperly grouped into codons**. It results in extensive **change in the sequence of amino acids**.
- The change in codons may also result in **premature termination**.
- Unless the frameshift is very near the end of the gene, it will produce a protein that is almost certain to be **non-functional**.
- Insertions and deletions of three nucleotide pairs result in polypeptides with additional or missing amino acids.

(c) Diseases associated with gene mutation

(i) Sickle cell anaemia

- **Autosomal recessive** disorder
- The genetic basis of **sickle-cell anaemia** is due to mutation of a single base pair in the gene that codes for one of the polypeptides of haemoglobin.
- The **substitution** of a single nucleotide, from CTT to CAT in the DNA's template strand of **chromosome 11** leads to a change in the codon of mRNA.



Substitution mutation in haemoglobin

- The original amino acid coded for is **glutamate is changed to valine** at the **sixth position**.
- Glutamate and valine are amino acids with very different properties. **Glutamate is hydrophilic** while **valine is hydrophobic**.
- The mutated haemoglobin tends to **polymerise into long rigid chains when not bound to oxygen**.
- The **long fibres distort the membrane** of the red blood cell giving it its **distinct sickle shape**.
- This results in the **decreased oxygen-carrying ability** of the red blood cells.
- Sickled red blood cells may clump and clog small blood vessels, leading to organ damage and even paralysis.
- In individuals who are **homozygous** for the mutant allele, altered haemoglobin results in the sickling of red blood cells and produces the multiple symptoms associated with sickle-cell disease, such as shortness of breath and dizziness.

(ii) Cystic fibrosis

- **Autosomal recessive** disorder
- The genetic basis of **cystic fibrosis** is due to **mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene**, which codes for **CFTR protein**.
- Most cystic fibrosis patients have a mutated CFTR gene on **chromosome 7,** consisting of **deletion of three consecutive nucleotides**.
- Deletion results in **deletion of phenylalanine** (an amino acid) from CFTR protein.
- In some other cases, the whole CFTR is missing from the membranes of these cells.
- CFTR protein function as **chloride ion channels** embedded in **membranes** of cells that line the respiratory tract, pancreas, sweat glands, intestine and other organs. It regulates the **transport of chloride ions out of these cells.**
- Mutated CFTR gene encodes a **non-functional CFTR protein**, resulting in **accumulation of chloride ions in cells**, causing **build up of thick sticky mucus on the surfaces of these cells**.
- Sufferers often have **reduced lifespan** due to the symptoms associated with this disease:
  - **Excessively salty sweat**, a largely benign effect of the mutant gene due to chloride ions being trapped in the cells. This draws in positive charged ions, like sodium and form the salt, sodium chloride which is lost in the sweat.
  - Respiratory problems as lungs are clogged with thick mucus, which results in **interference with gaseous exchange**. Patients are often breathless due to inefficient gaseous exchange.
  - Mucus in lungs also tend to **trap bacteria leading to chronic lung infections** e.g. pneumonia.
  - Thick mucus often results in blocked pancreatic ducts and **reduced secretion of pancreatic enzymes needed for proper digestion**.
  - Building up of mucus in the digestive tract **interferes with absorption of digested nutrients** causing individuals to be malnourished no matter how much they eat.
  - **Males may be sterile and females have reduced fertility**. Men make normal sperms but are missing the vas deferens whereas females have fertility difficulties due to thickened cervical mucus (sperms cannot penetrate) and malnutrition disrupting ovulation and causing amenorrhea, the absence of a menstrual period in a woman of reproductive age.
  - Patients often **develop diabetes** although the cause is still relatively unknown.

Mutagens

Errors during DNA replication, repair, or recombination can lead to base-pair substitutions, insertions, or deletions. Such mutations are called **spontaneous mutations**.

There are also a number of factors which can cause DNA mutations. Physical and chemical agents, called **mutagens**, interact with DNA in ways that cause mutations. Most **carcinogens** (cancer-causing chemicals) are mutagenic and, conversely, most mutagens are carcinogenic.

- **Physical mutagen** – X-rays, UV light and other forms of ionising radiation, have been discovered to cause genetic changes.
- **Chemical mutagen** – Base analogs are chemicals (e.g. ethidium bromide) that are similar to normal DNA bases but pair incorrectly during DNA replication. Some other chemical mutagens interfere with correct DNA replication by inserting themselves into the DNA and distorting the double helix. Still other mutagens cause chemical changes in bases that change their pairing properties.