DNA & Genomics

1. Overview of Topic

Genetic information is stored in an organism's DNA; expression of genes results in the synthesis of functional products, such as rRNA, tRNA and proteins. These products play a role in intra- and extra-cellular biochemical pathways and influence the physiological processes in organisms. Genomes contain heritable information necessary for continuity of life at all levels: cell, organism and system. This information is stored and passed on to subsequent generations via DNA. Reproduction can occur at the cellular or organismal level; each progeny needs to receive heritable genetic information from its parent/s.

2. Learning Outcomes

(a) describe the structure and roles of DNA and RNA (tRNA, rRNA and mRNA) (knowledge of mitochondrial DNA is not required)

(b) describe the process of DNA replication and how the end replication problem arises

(c) describe how the information on DNA is used to synthesise polypeptides in prokaryotes and eukaryotes (description of the processes of transcription, formation of mRNA from pre-mRNA and translation is required)

(d) describe the structure and organisation of viral, prokaryotic and eukaryotic genomes (including DNA/RNA, single-/double-stranded, number of nucleotides, packing of DNA, linearity/circularity and presence/absence of introns)

3. References

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V. and Jackson, R. B. (2011) Campbell Biology (Ninth Edition) (Pearson Higher Education) ISBN 0321739752

Contents

4. Nucleotides

There are two types of nucleic acids: Deoxyribonucleic acid (DNA) and Rribonucleic acid (RNA).

DNA can be found in eukaryotes, prokaryotes and viruses. For eukaryotes, DNA is found in the nucleus, mitochondria and chloroplasts

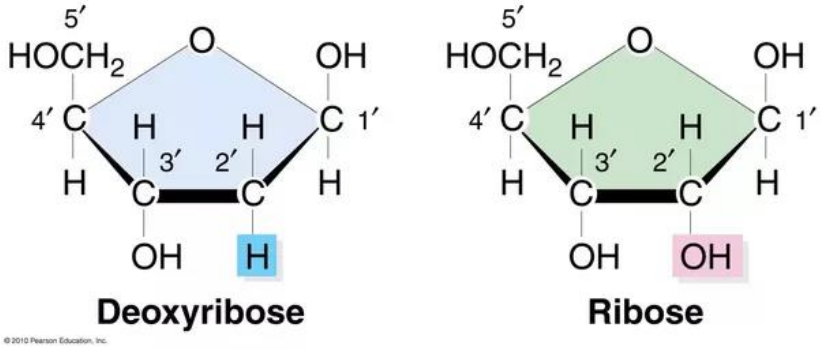
There are three different types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). mRNA is transcribed from DNA. rRNA which is synthesised in the nucleolus will combine with ribosomal proteins to form ribosomes in the cytoplasm. tRNA plays an active role in the process of translation.

Nucleic acids are polymers of nucleotides, thus it is known as polynucleotides. Monomers of DNA and RNA are deoxyribonucleotides and ribonucleotides respectively.

Structure of nucleotides

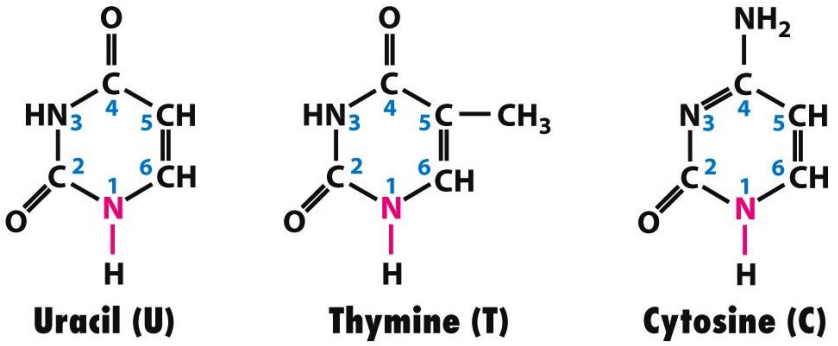
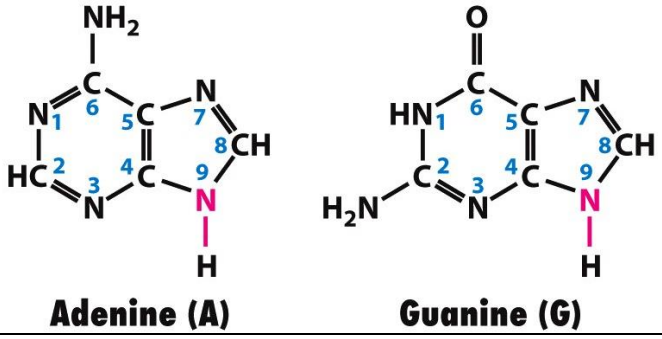
Individual nucleotide comprises of a pentose (5 carbon sugar), a phosphate and a nitrogenous base.

There two types of pentose: deoxyribose and ribose sugar, the former has the oxygen removed from the -OH group at its carbon 2 while the latter retains the -OH group's oxygen. Deoxyribonucleotide would contain a deoxyribose sugar while Ribonucleotide would contain ribose sugar.

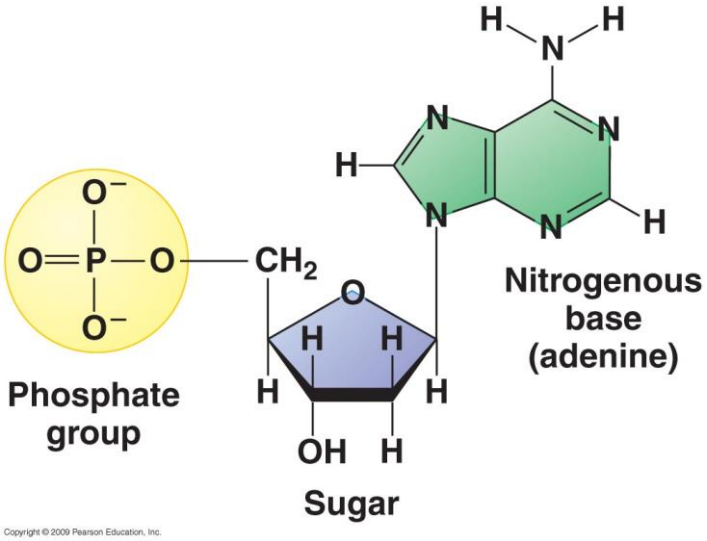


The phosphate group is covalently bonded to the pentose as carbon 5 while the Nitrogenous base is covalently bonded to the pentose as carbon 1. In short, formation of a deoxyribonucleotide / ribonucleotide involves 2 condensation reactions.

There are five different types of nitrogenous bases: adenine (A), guanine (G), cytosine (C), thymine (T) & uracil (U). These five bases are classified into two families, they are either a Pyrimidine base or a Purine base.

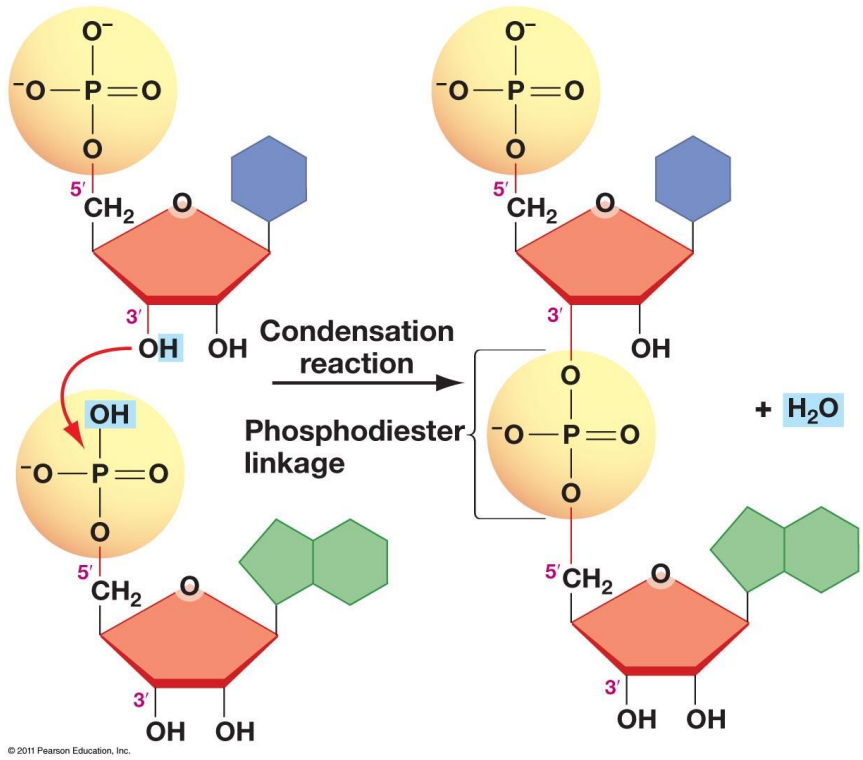
Base Type	Structure and Representative Bases
Pyrimidine Single ring, each with six sides	
Purine Two rings – six-sided and five-sided	

DNA contains A, G, C, T while RNA contains A, G, C, U bases.

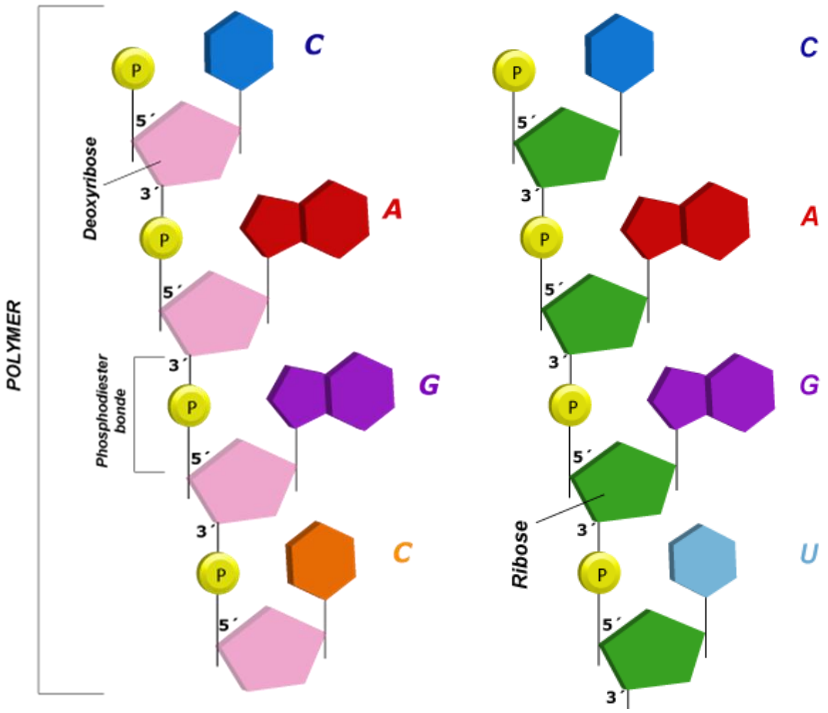


Polymerisation of Nucleotides

DNA and RNA are formed from the polymerisation of deoxyribonucleotides and ribonucleotides respectively. Nucleotides polymerise through consecutive condensation reactions between the phosphate group at carbon 5 of the pentose sugar of one nucleotide and the hydroxyl group at carbon 3 of the pentose sugar of the adjacent nucleotide to form a phosphodiester bond. During condensation reaction, a water molecule is removed. Phosphodiester linkages are strong covalent bonds which confers strength and stability to the nucleic acid.



DNA Polynucleotide Chain **RNA Polynucleotide Chain**



Some nucleotides serve important biological functions other than as a genetic material.

Table showing common nucleotides and their abbreviation and functions:

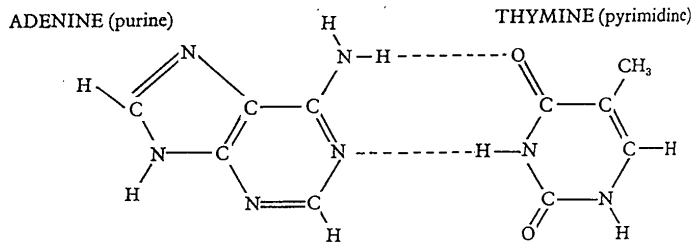
Molecule	Abbreviation	Function
• Deoxyribonucleic Acid	• DNA	Contains the genetic information of cells or viral particle
• Ribonucleic Acid	• RNA	All three types play a vital role in protein synthesis
• Adenosine monophosphate • Adenosine diphosphate • Adenosine triphosphate	• AMP • ADP • ATP	Coenzymes important in making energy available for cells for metabolic activities.
• Nicotinamide adenine dinucleotide • Flavine adenine dinucleotide	• NAD • FAD	Electron carrier important in respiration in transferring high energy electrons to the Electron Transport Chain and various reactions in the respiratory chain.
• Nicotinamide adenine dinucleotide phosphate	• NADP	Electron carrier important in photosynthesis for accepting electrons from the chlorophyll molecule and making them available for the photolysis of water
• Coenzyme A	• CoA	Coenzyme important in respiration in combing with pyruvic acid to form acetyl coenzyme A and transferring the acetyl group into the Krebs' cycle

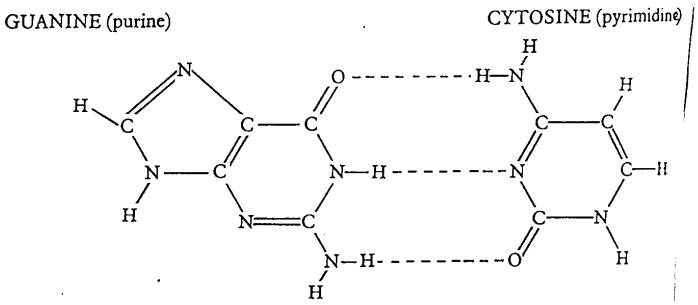
5. DNA structure

The DNA Consists of 2 polynucleotide chains. Each chain is a right – handed spiral; 2 chains coiled around each other to form a **double helix**.

Each chain is made up of a **sugar-phosphate backbone**, found on the outside. Chains run **anti-parallel** to each other: 3' end of one chain lies opposite the 5' end of the other. Nitrogenous bases project at **right angle** inward. **Hydrogen bonds** formed between the base of one chain and the corresponding base of the opposite chain. Each base pair consist of **one purine + one pyrimidine**

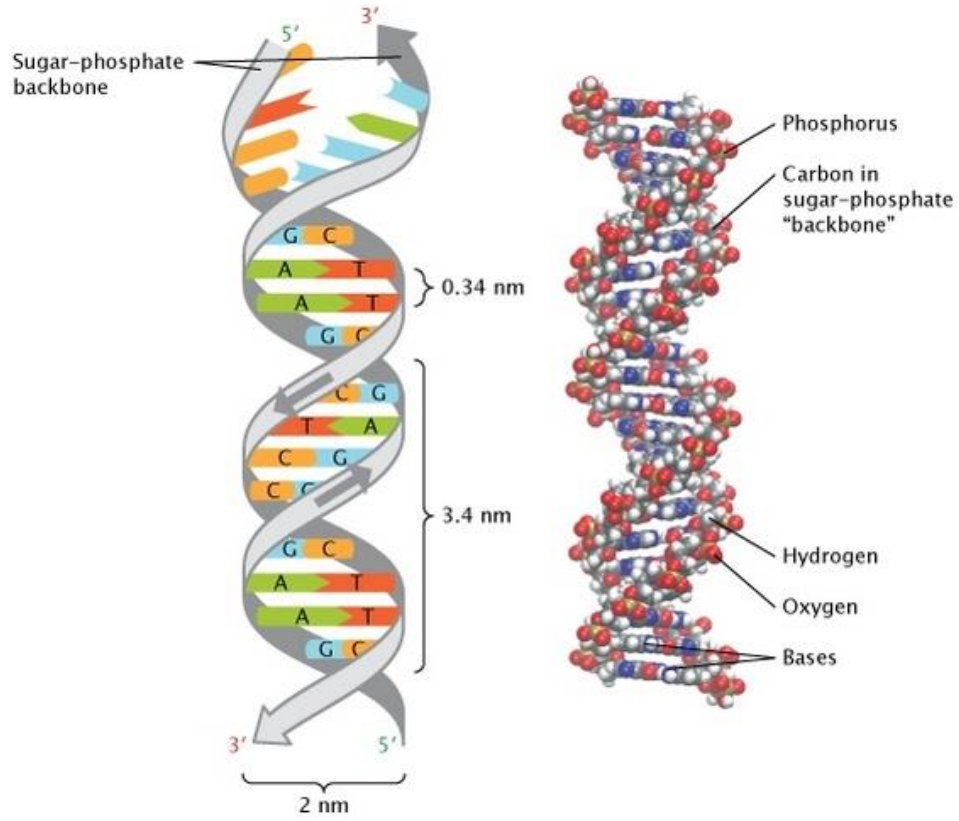
The positions of the hydrogen atoms in relation to the shape of the molecule ensures that **A** can only link with **T**, and **C** with **G**, like fitting together of complementary pieces in a jig-saw puzzle. 2 hydrogen bonds formed: A = T (for DNA only) or A = U (for RNA only) while 3 hydrogen bonds formed: C ≡ G





Width between 2 chains is constant, i.e. the width of one base pair (2nm). As the distance between adjacent base pairs is 0.34nm, a complete turn of the double helix (made up of Ten nucleotides, 10 base pairs) would be 3.4nm.

Base sequence in one chain is not restricted, but because of rules of base pairing, sequence of one chain determines that in the other, and the 2 chains are complementary to each other



Base ratio of D*NA*

The number of purine bases (A + G) present equals the number of pyrimidine bases (T + C). The number of adenine bases = the number of thymine bases while the number of guanine bases = the number of cytosine bases. This is because Adenine is always base paired with thymine (A = T) while guanine is base paired with cytosine (G ≡ C).

Allowing accurate replication because of this rule of base pairing.

Relative amounts of bases in DNA from various organisms

Source of DNA	*Adenine*	*Guanine*	*Thymine*	*Cytosine*
Human	30.9	19.9	29.4	19.8
Sheep	29.3	21.4	28.3	21.0
Hen	28.8	20.5	29.2	21.5
Turtle	29.7	22.0	27.9	21.3
Salmon	29.7	20.8	29.1	20.4
Sea urchin	32.8	17.7	32.1	17.3
Locust	29.3	20.5	29.3	20.7
Wheat	27.3	22.7	27.1	22.8
Yeast	31.3	18.7	32.9	17.1
Escherichia coli (a bacterium)	24.7	26.0	23.6	25.7
ϕX174 bacteriophage (a virus)	24.6	24.1	32.7	18.5

Amounts are in molar proportions on a percentage basis.

6. RNA Structure

RNA are essential in living cells and viruses for the process of protein synthesis. They are normally single-stranded and are polynucleotides made up of many ribonucleotides joined by 3' → 5' phosphodiester bonds

There are 3 types of RNA, all involved in protein synthesis:

- mRNA : messenger RNA
- tRNA : transfer RNA
- rRNA : ribosomal RNA

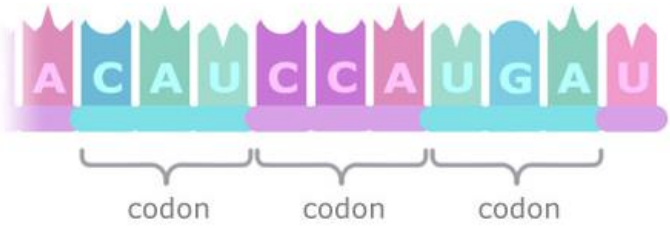
The process of making RNA from DNA is known as transcription. For Eukaryotes, transcription takes place in the nucleus. These RNA strands will then get transported to the cytoplasm through the nuclear pore.

Messenger RNA

This form of RNA is found to make up 3 – 5% of the total cellular RNA. mRNAs are long **single-stranded** polynucleotide. Base sequence of mRNA is **complementary** to the template DNA strand.

Carries the DNA message from the nucleus to the cytoplasm in the form of a series of **codons** (sequence of 3 nucleotides), acting as a template for protein synthesis.

These codons provide specific binding sites for a series of complementary aminoacyl-tRNA complexes and hence coding for specific amino acid sequence in polypeptide chain during translation.

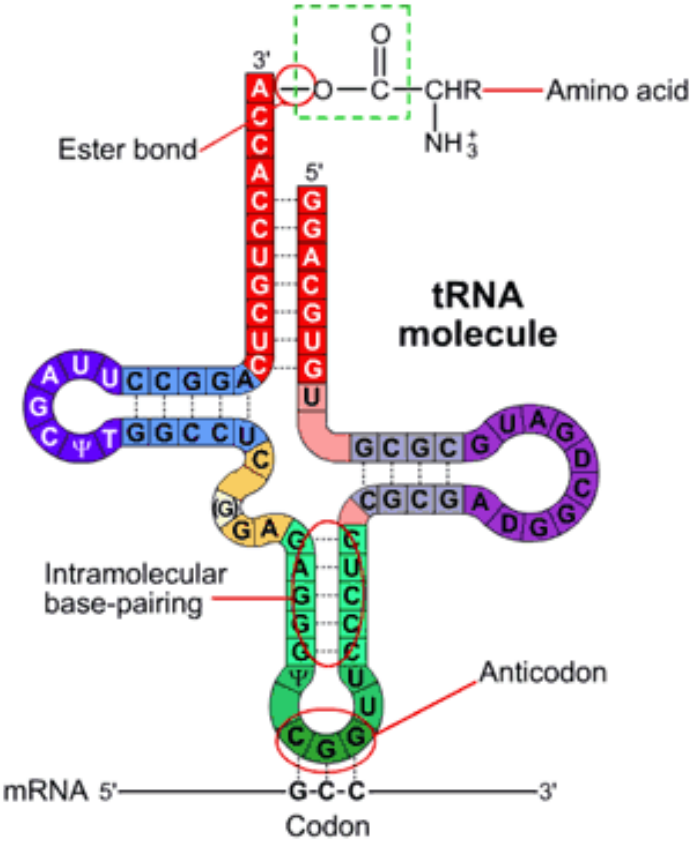


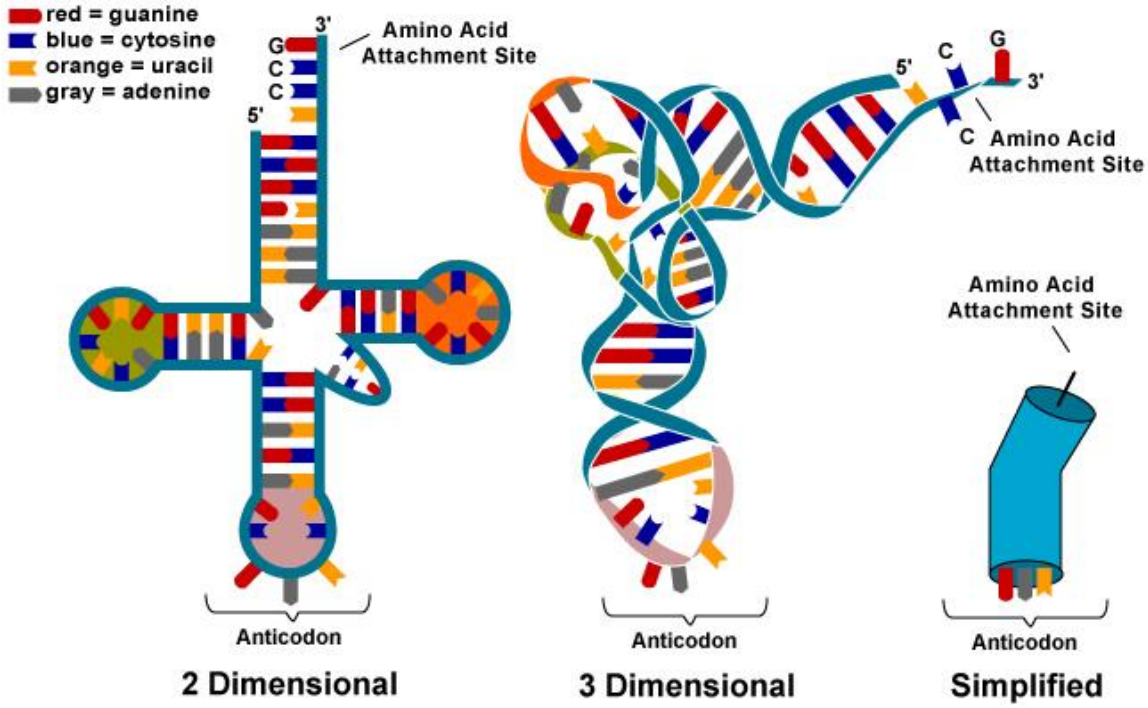
Transfer RNA

This form of RNA is found to make up of 15% of the total cellular RNA. It averages to be about 80 nucleotides per molecule, comprising a single strand. Hydrogen bonding between complementary bases along chain cause strand to fold up, forming a clover-shaped secondary structure. The 5' end of tRNA always ends with a base, guanine, whilst the 3' end always ends with the base sequence of three nucleotides - C C A

The tRNA contains **anticodons** which are sequences of 3 nucleotides that are complementary to the codons on mRNA. Hence, using mRNA as a template for the coding of specific amino acid sequence.

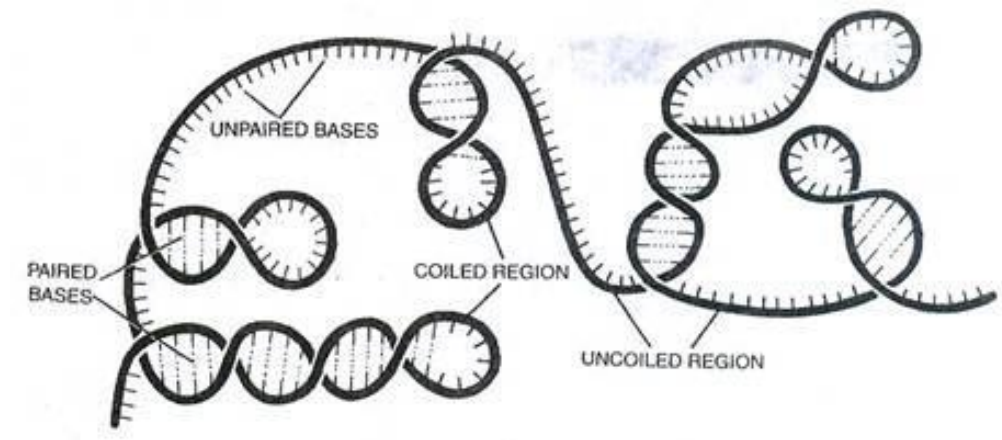
There are at least 20 different kinds of tRNA, each one carrying a different amino acid. The amino acid binds to the 3' end of the tRNA via a condensation reaction to form an ester linkage. This is catalysed by an enzyme called amino-acyl tRNA synthetase (further elaborated under Translation)


Ester bond
Amino acid
tRNA molecule
Intramolecular base-pairing
Anticodon
mRNA 5'
G-C-C
3'
Codon

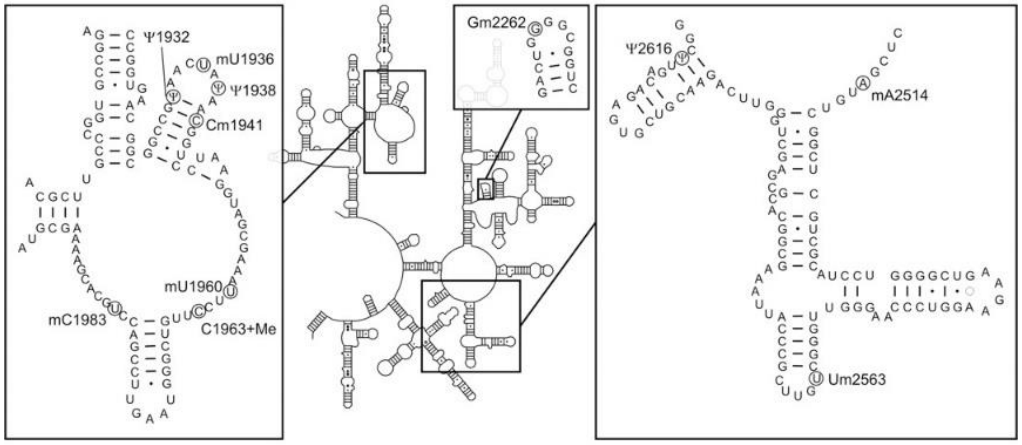

red = guanine
blue = cytosine
orange = uracil
gray = adenine
Amino Acid Attachment Site
Amino Acid Attachment Site
Amino Acid Attachment Site
Anticodon
Anticodon
Anticodon
2 Dimensional
3 Dimensional
Simplified

Ribosomal RNA

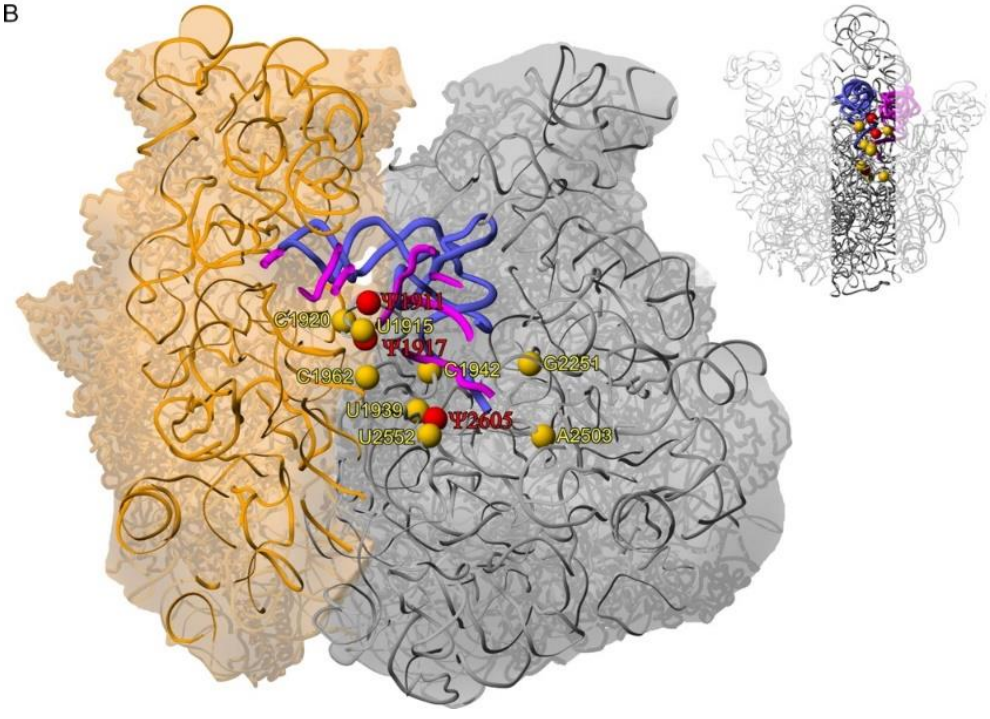
This form of RNA is found to make up of 80% of the total cellular RNA. It is a complex molecule made up of double or single helices.



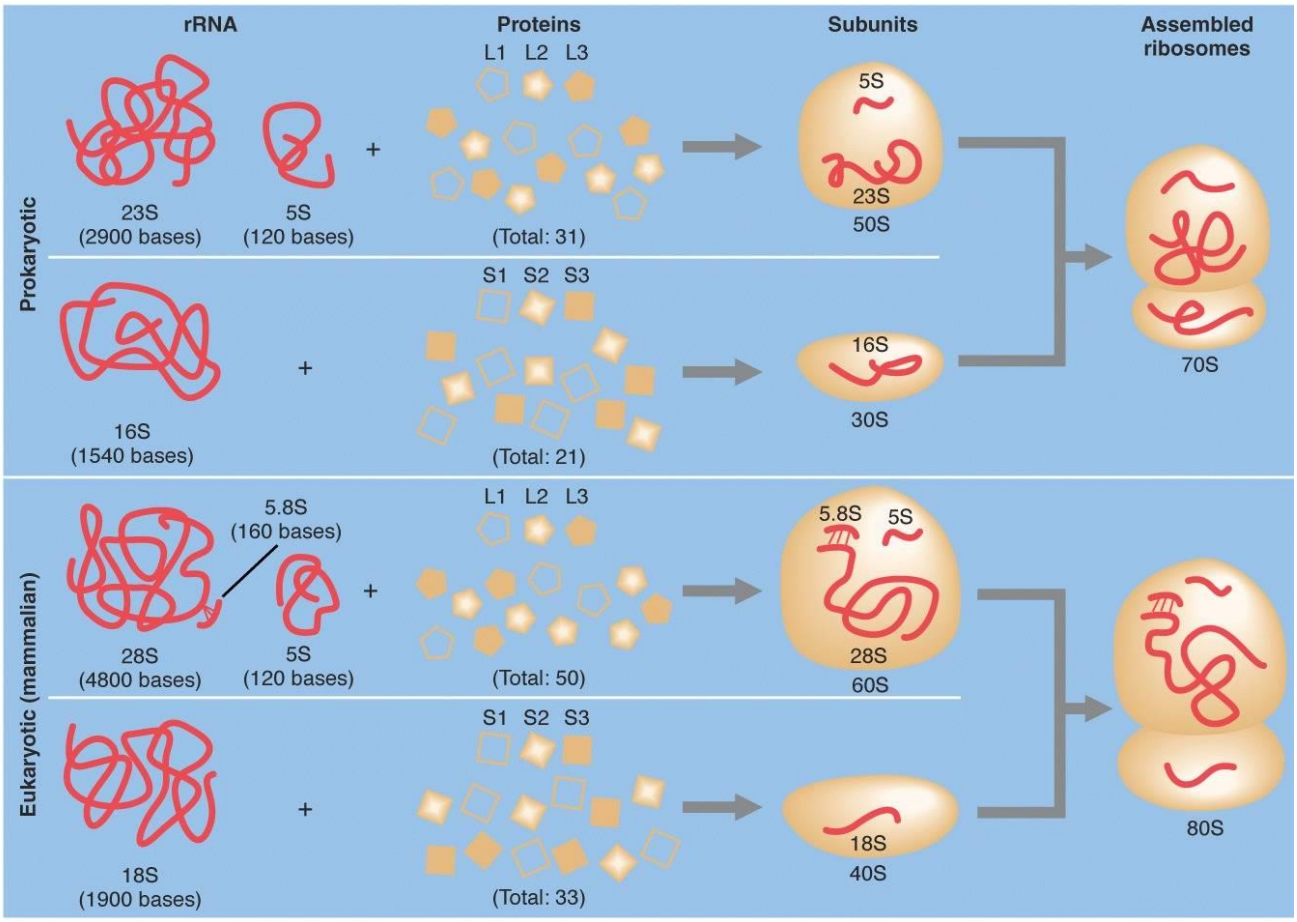
A



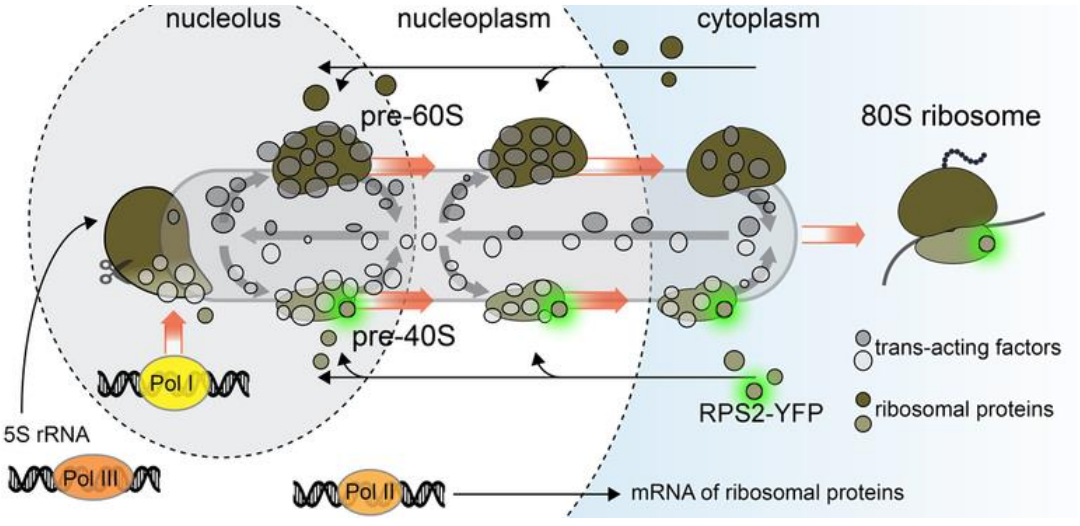
B



Prokaryotic and eukaryotic ribosomes are slightly different in terms of the compositional makeup of the ribosome. However, they are similar in terms of their structure. Both ribosome types are composed of a large subunit and a small subunit.

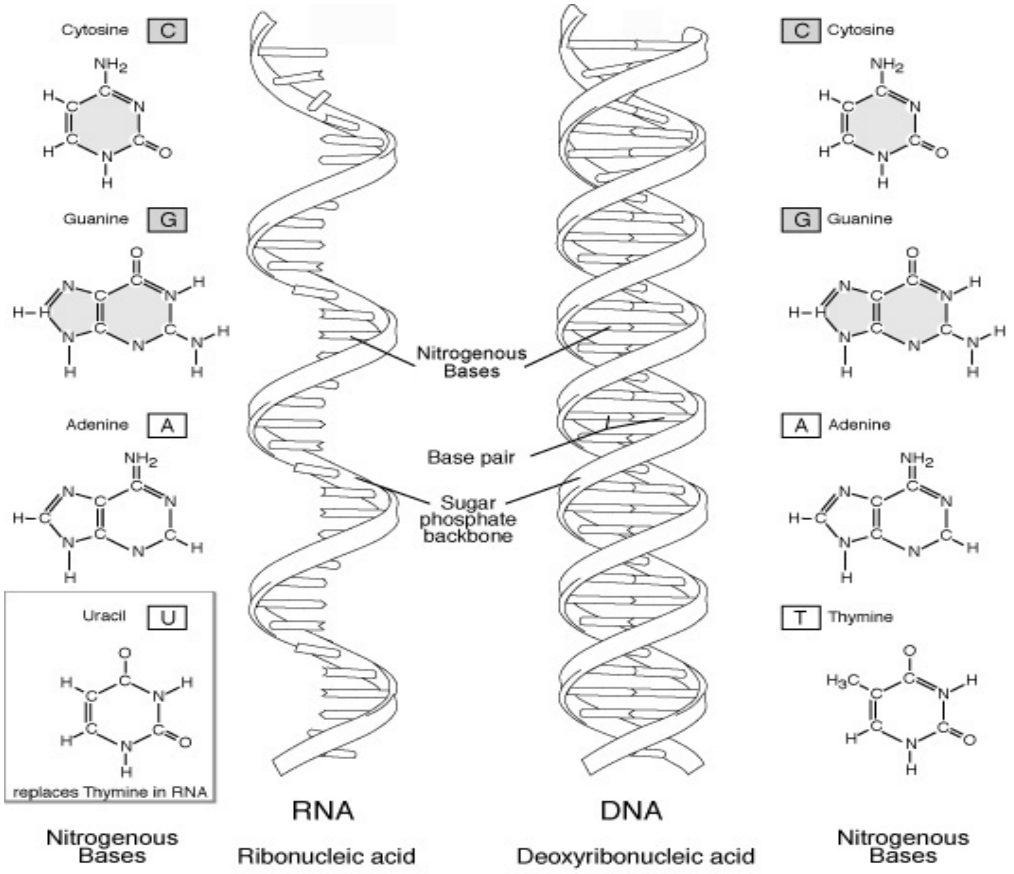


For eukaryotes, the rRNA is synthesised in the nucleolus within the nucleus. The rRNA synthesised will enter will be associated with ribosomal proteins to form ribosomes within the nucleus. Ribosome provides the structural framework for holding mRNA and tRNA; helps to stabilize the temporary union between mRNA and tRNA. In prokaryotes, these are occurring in the cytoplasm due to the lack of membrane-bound organelles within these cells.



rRNA keeps the ribosome functional by orientating a large variety of enzymes (aminoacyl-tRNA synthetases) involved in protein synthesis in the correct way/ spatial arrangement. rRNA helps to align mRNA in ribosome in such a way that incoming aminoacyl-tRNA complexes can base-pair conveniently.

Try This !



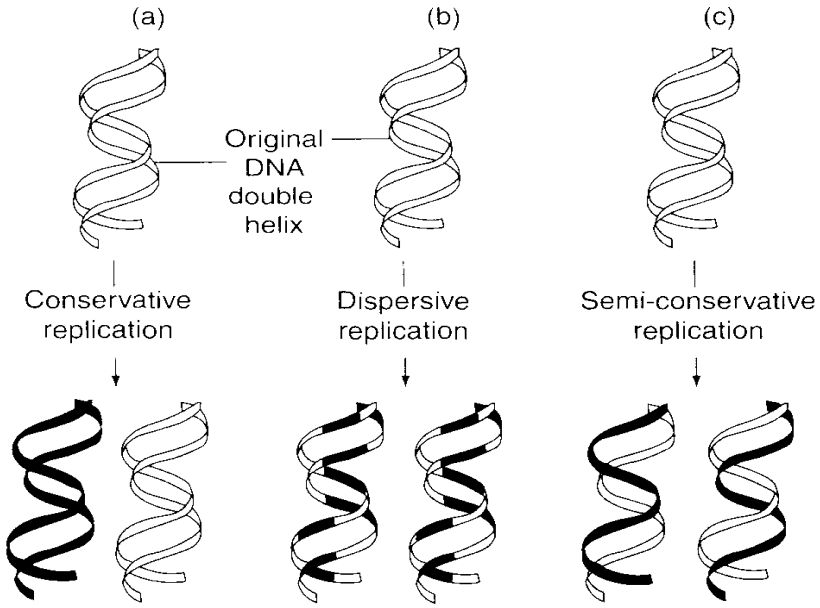
QNS: Based on the above diagram, list down the differences between a DNA and an RNA.

7. DNA replication

DNA Replication is the process of making an exact copy of DNA to increase amount of DNA prior to cell division.

DNA is replicated **Semi-conservatively** where both strands of DNA molecule separate and act as templates for the synthesis of two daughter DNA strands via complementary base pairing. Hydrogen bonds form between the bases of one original parent strand and one newly synthesised strand to form a complete DNA molecule. Each daughter cell, after cell and nuclear division, will inherit DNA molecules that are hybrids, each consisting of one original parent strand and one newly synthesised daughter strand.

There are actually 2 other possible models of DNA replication:



The two original strands of DNA are shown in white while the newly synthesized DNA is shown in black.

Conservative replication

Both strands of the DNA molecule act as **templates** for the synthesis of an entirely new DNA molecule. The parent DNA molecule is intact and goes into one daughter cell, and the newly synthesised DNA molecule goes into the other daughter cell.

Dispersive replication

Parent DNA molecule breaks up into short segments, which act as templates for the synthesis of DNA. The segments are then joined together. This results in two DNA molecules with sections of both original and newly synthesised DNA molecules interspersed along each strand.

The experimental proof that DNA replication is semi-conservative is found in the BROYO section on page 13.

BROYO

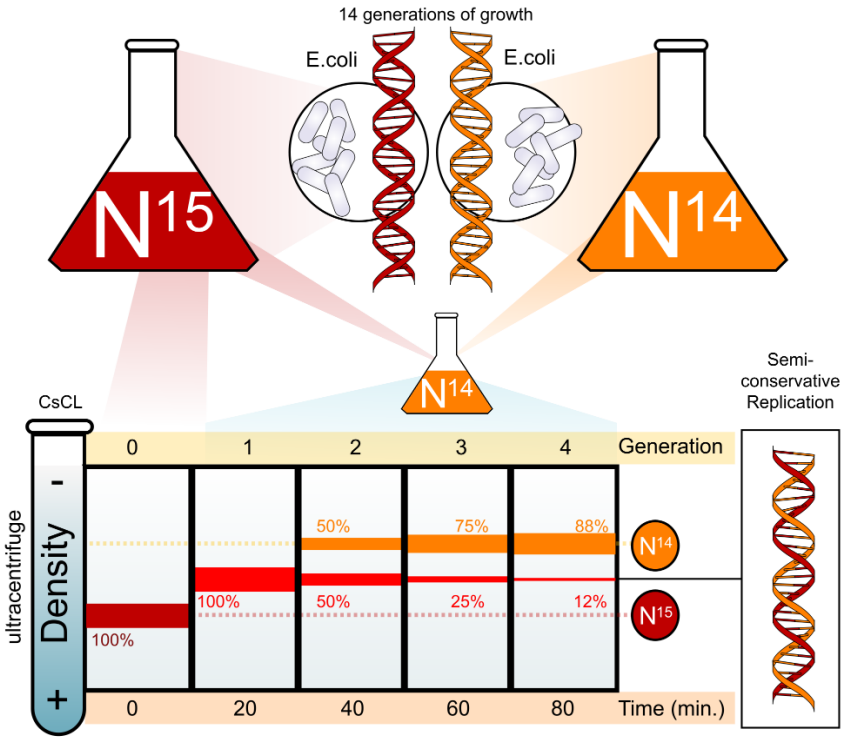
Evidence for semi-conservative replication

Matthew Meselson and Franklin Stahl (1958) used two different isotopes of nitrogen for their experiments – ^{14}N is the ordinary isotope of nitrogen; ^{15}N is the heavy isotope of nitrogen.

Culture bacteria *Escherichia coli* (*E. coli*) for many generations (e.g. 14 generations) in a medium that contained $^{15}NH_4Cl$ as the sole source of nitrogen, until all ^{14}N in parental DNA are replaced by ^{15}N (for all the 14 generations). These cells were transferred into a culture containing $^{14}NH_4Cl$ (normal isotope of nitrogen) as the sole source of nitrogen and allowed to grow.

At a fixed time interval – corresponding to generation time for *E. coli*, samples were removed. One sample of bacterial cells was allowed to divide once to provide the "first generation" cells. Another sample of cells was allowed to divide twice to obtain the "second generation" cells. DNA was extracted and centrifuged at 40 000 times gravity in a solution of caesium chloride (CsCl).

DNA containing only ^{15}N (^{15}N ^{15}N strands) are denser than DNA containing only ^{14}N (^{14}N ^{14}N strands). The former would form a band below the latter in a caesium chloride solution. Hybrid DNA, consisting of one ^{14}N strand and one ^{15}N strand (^{14}N ^{15}N), would form a band halfway between the $^{14}N^{14}N$ strands DNA band and $^{15}N^{15}N$ DNA band.



First generation:
Density of the DNA was intermediate between DNA of bacterium with only $^{15}N^{15}N$ strands and those DNA with only $^{14}N^{14}N$ strands, i.e. it contained equal amounts of each

Second generation:
Two bands observed: Half of the DNA molecules were of the hybrid type $^{14}N^{15}N$ while the other half were pure $^{14}N^{14}N$ DNA.

Mechanism of DNA replication

STEP 1: Unwinding of the DNA double helix

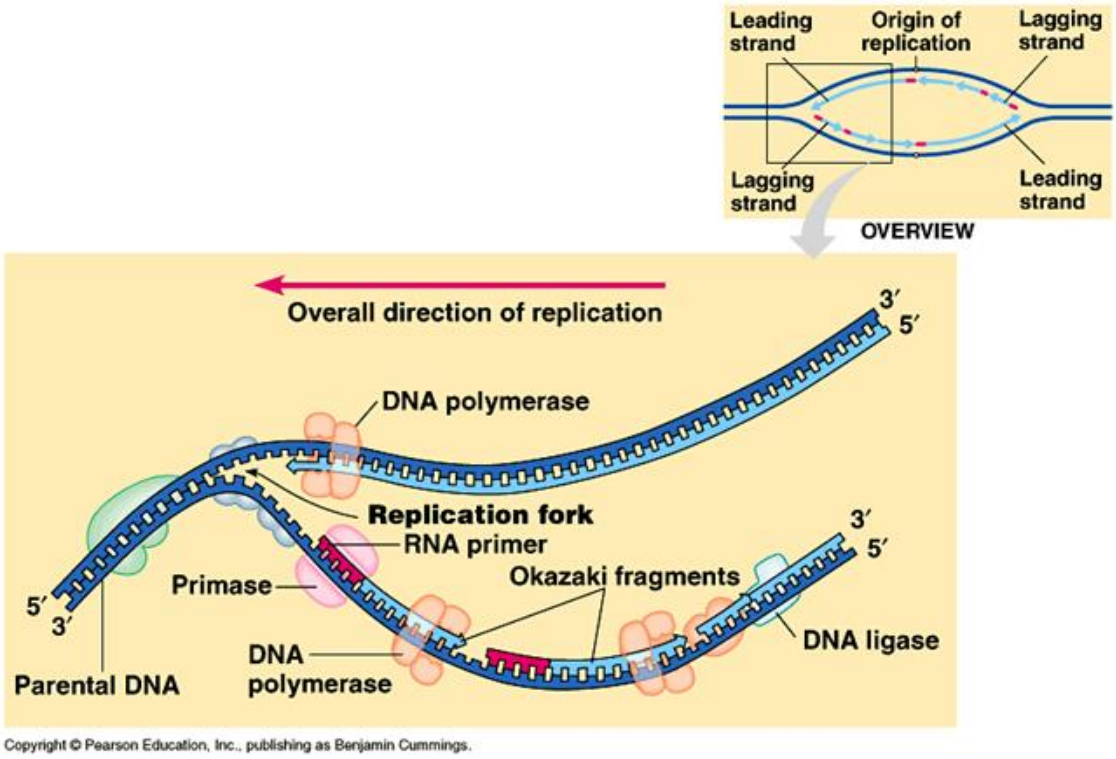
Before the start of DNA replication, free **deoxyribonucleotides** are synthesised in the cytoplasm.

Replication of DNA begins at specific sites called the **origins of replication**. The bacterial chromosome, which is circular, has a single origin of replication. Eukaryotic chromosome may have hundreds or even thousands of origins of replication. Multiple replication bubbles form and eventually fuse, thus speeding up the copying of the very long DNA molecule.

The ATP-dependent enzyme **helicase** causes the DNA molecule to **unwind at the origin of replication** by breaking **hydrogen bonds** between the complementary bases**,** causing the DNA strands to separate. This form a **replication fork** that will progressively move away from the origin of replication.

As the template unwinds, there is a tendency for the remaining double strands to create super coils and this can be alleviated by the action of the enzyme DNA topoisomerase. This enzyme introduces nicks on the DNA before the fork and allow the supercoiled strand to reduce in tension.

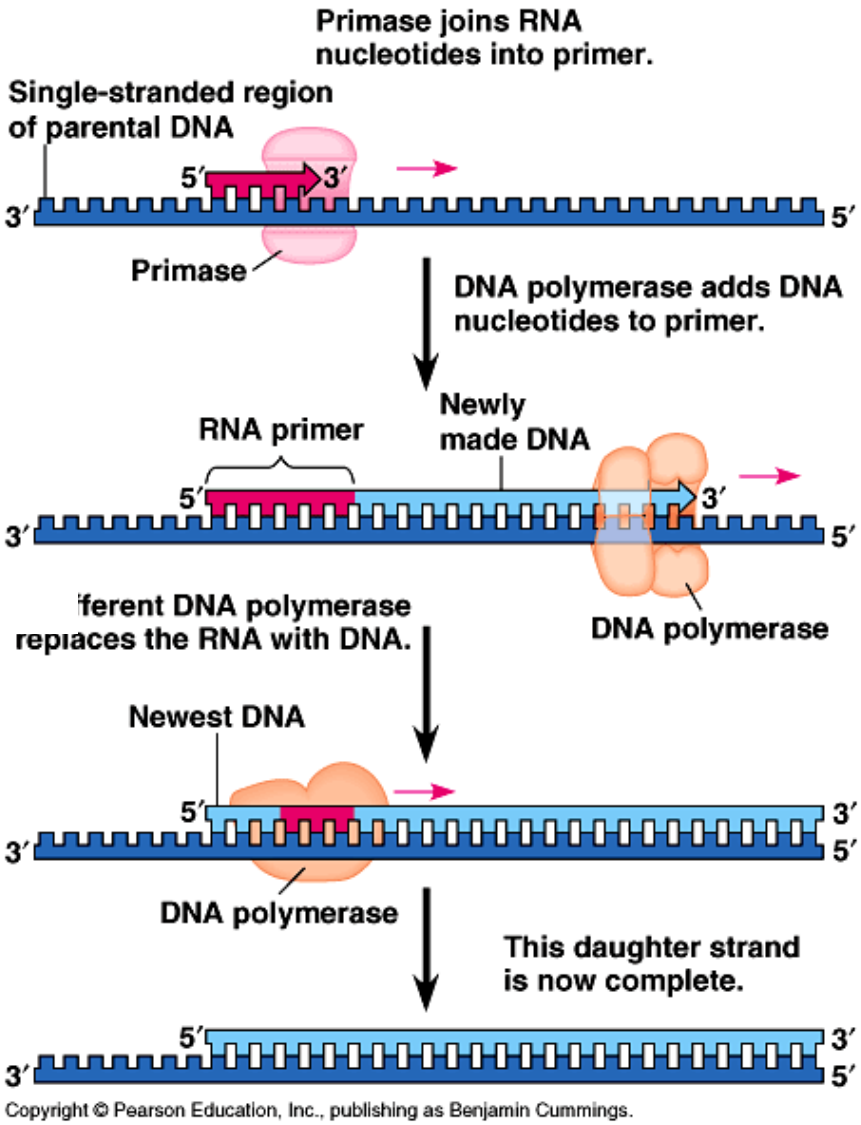
The separated strands of parental DNA interact with the **single-stranded DNA binding protein**. The role of the single-stranded DNA binding protein is to **stabilise the single-stranded DNA** formed by the action of helicase so that the unwound region can serve as a template.



STEP 2: Formation of the RNA primer

The enzyme **primase** (also known as **DNA-dependent RNA polymerase**) attaches to the unwound chains behind the replication fork and catalyses the formation of a short RNA strand that is complementary to one of the DNA template strands.

This short RNA strand is known as the **primer** which consists of about **10 ribonucleotides.** Primers are the beginnings of all new DNA chains (both leading & lagging strands). **DNA polymerase**, responsible for the synthesis of DNA can only add deoxyribonucleotides to the **3' end** of an already existing chain that is base-paired with the template strand.



STEP 3: Synthesis of the new DNA strand

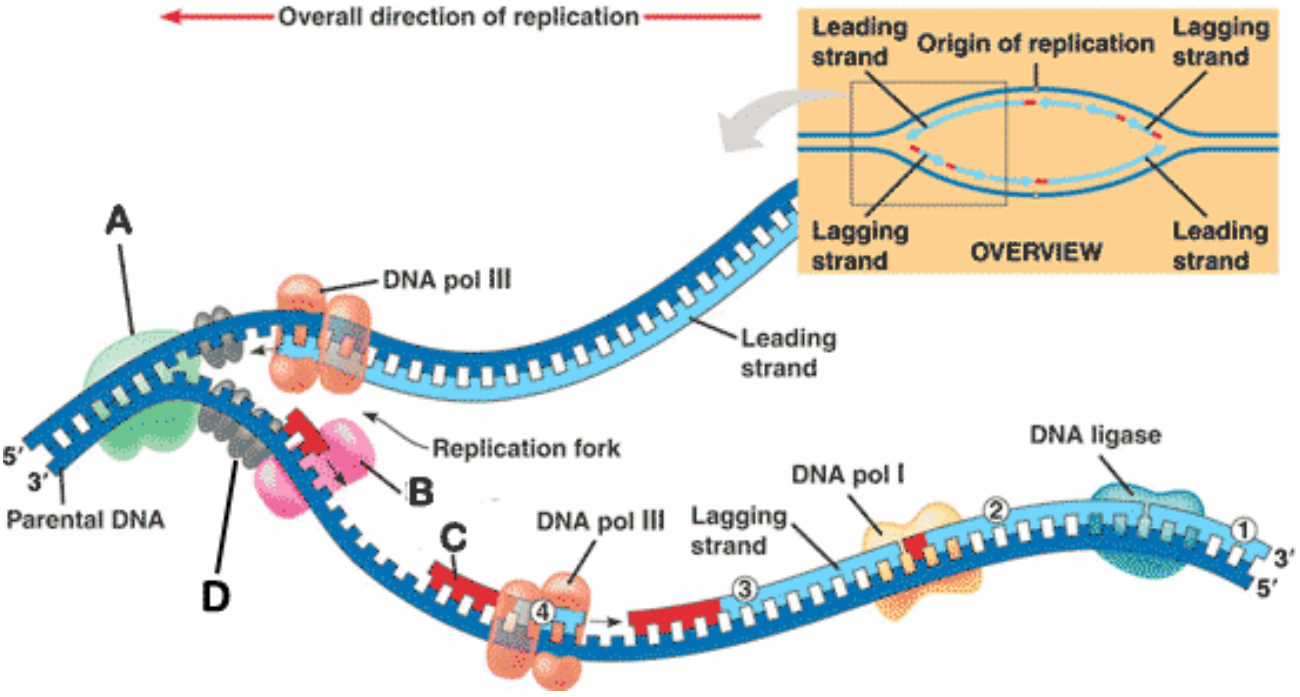
DNA polymerase synthesizes new DNA, continuous from the RNA primer strand; using the parent strand as the template, recognises the bases and fit in free deoxyribonucleotides that are complementary to those on the parental strand. Adenine pairs with thymine; cytosine pairs with guanine.

The phosphodiester bond between the incoming deoxyribonucleotide and the preceding deoxyribonucleotide can only be formed if the base on the incoming deoxyribonucleotide is complementary to the base on the template.

The enzyme DNA polymerase can remove the mismatched one before proceeding with polymerisation, i.e. DNA polymerase III can perform a proofreading or an editing function.

DNA polymerase works only in **5' to 3' direction** with respect to the growing daughter strands. Therefore, only one new chain that is complementary to the **parent strand** that runs from **3' to 5'** direction would be formed as a continuous chain. It is also known as the **leading strand**.

Okazaki fragments will be formed for the other parent strand that runs from 5' to 3' direction. It will be known as the **lagging strand**.



STEP 4: Removal of the primer

The RNA portion (RNA primer) of the RNA-DNA hybrid is then hydrolysed.

Removal of the RNA primer leaves subsequent gaps between the DNA fragments. The filling of these gaps by adding complementary deoxyribonucleotides, is also catalysed by a DNA polymerase.

After the removal of the RNA segments (the primer) and the filling of the gaps with DNA, there remain points in the DNA backbone where the phosphodiester bond is broken. These breaks, called nicks, will be sealed by DNA ligase.

STEP 5: Joining Okazaki Fragments and Sealing Nicks

DNA ligase will join the short fragments together; joins the free hydroxyl group of the 3' end of one fragment with the phosphate group at the 5' end of another fragment, forming a phosphodiester bond.

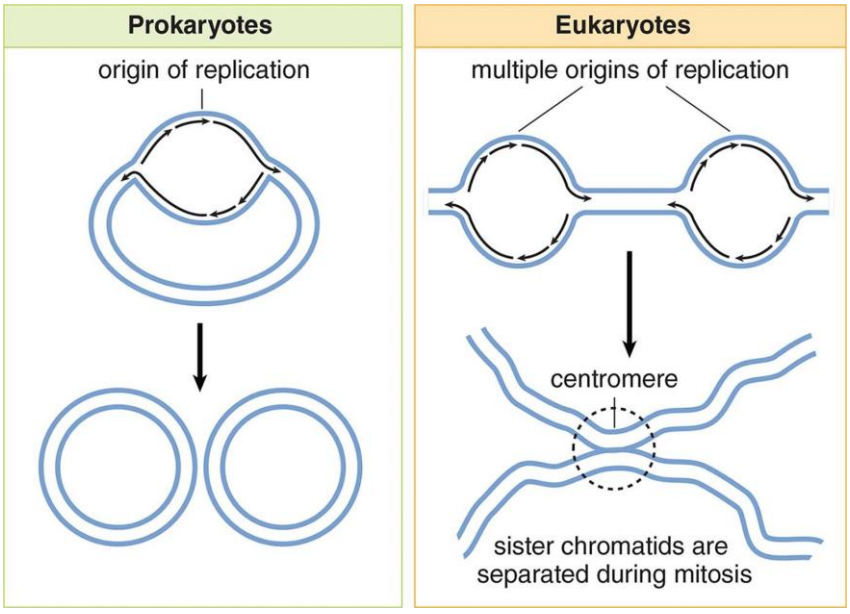
Two DNA molecules are now formed. Each containing a newly synthesised strand and a parent strand, rewind into a double helix. The process is semi-conservative since each resultant double helix consists of one original parental strand and one newly synthesised daughter strand.

Difference between Prokaryotic and Eukaryotic DNA Replication

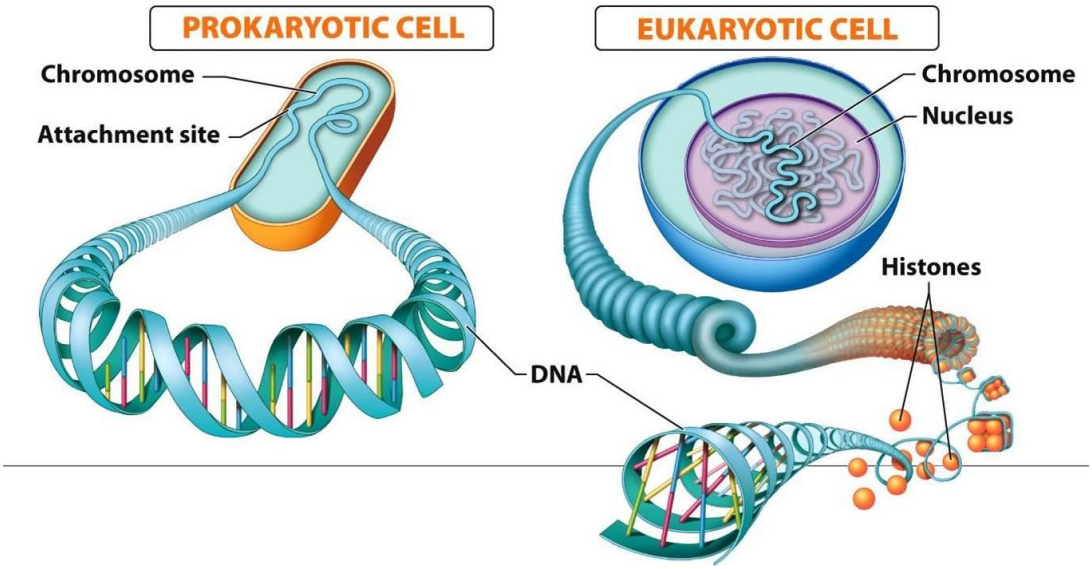
Although the process of DNA replication in all organism is based upon the need of the enzyme DNA polymerase, key chaperone proteins and largely similar essential steps to copy the DNA semi-conservatively, there are still key differences of this process in prokaryotes and eukaryotes.

Table showing some key differences between prokaryotic and eukaryotic DNA replication

Features	Prokaryotes	Eukaryotes
RNA primer length	50 nucleotides	9 nucleotides
DNA polymerase	I, II, III	α, β, γ, δ, ε
Number of Origins of Replication	Single	Multiple
Nucleotide length of Okazaki fragments	1000 – 2000 nucleotides	200 nucleotides
Rate of replication	500 nucleotides/ second	50 nucleotides/ second

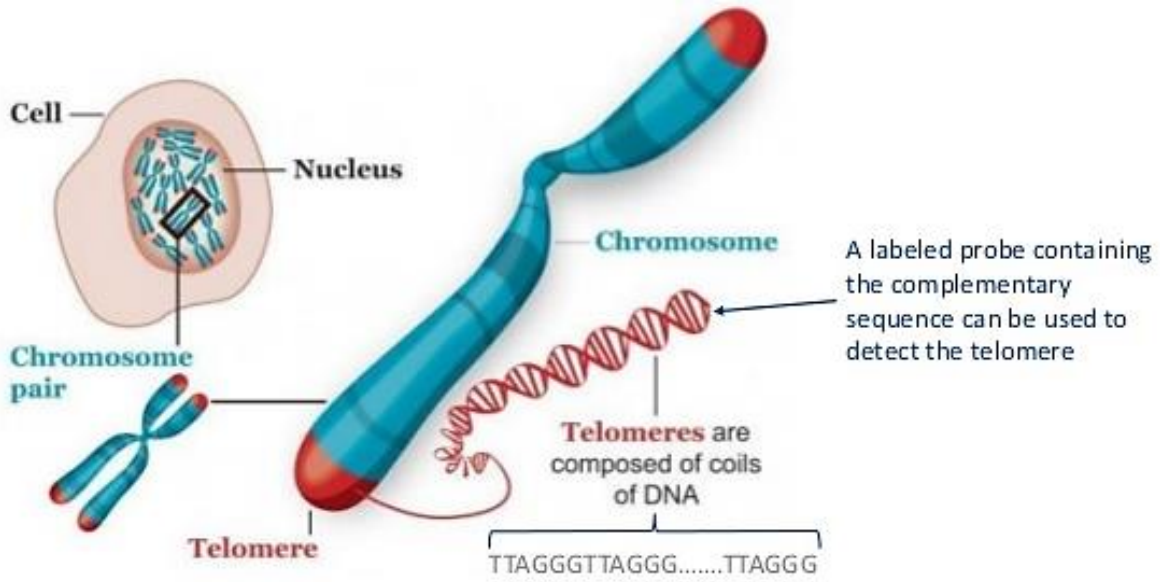


Another key difference would be the problem that is experienced during the process of DNA replication due to the type of DNA. In prokaryotes, DNA is circular while in eukaryotes, DNA is linear.



End replication problem during DNA replication in eukaryotes

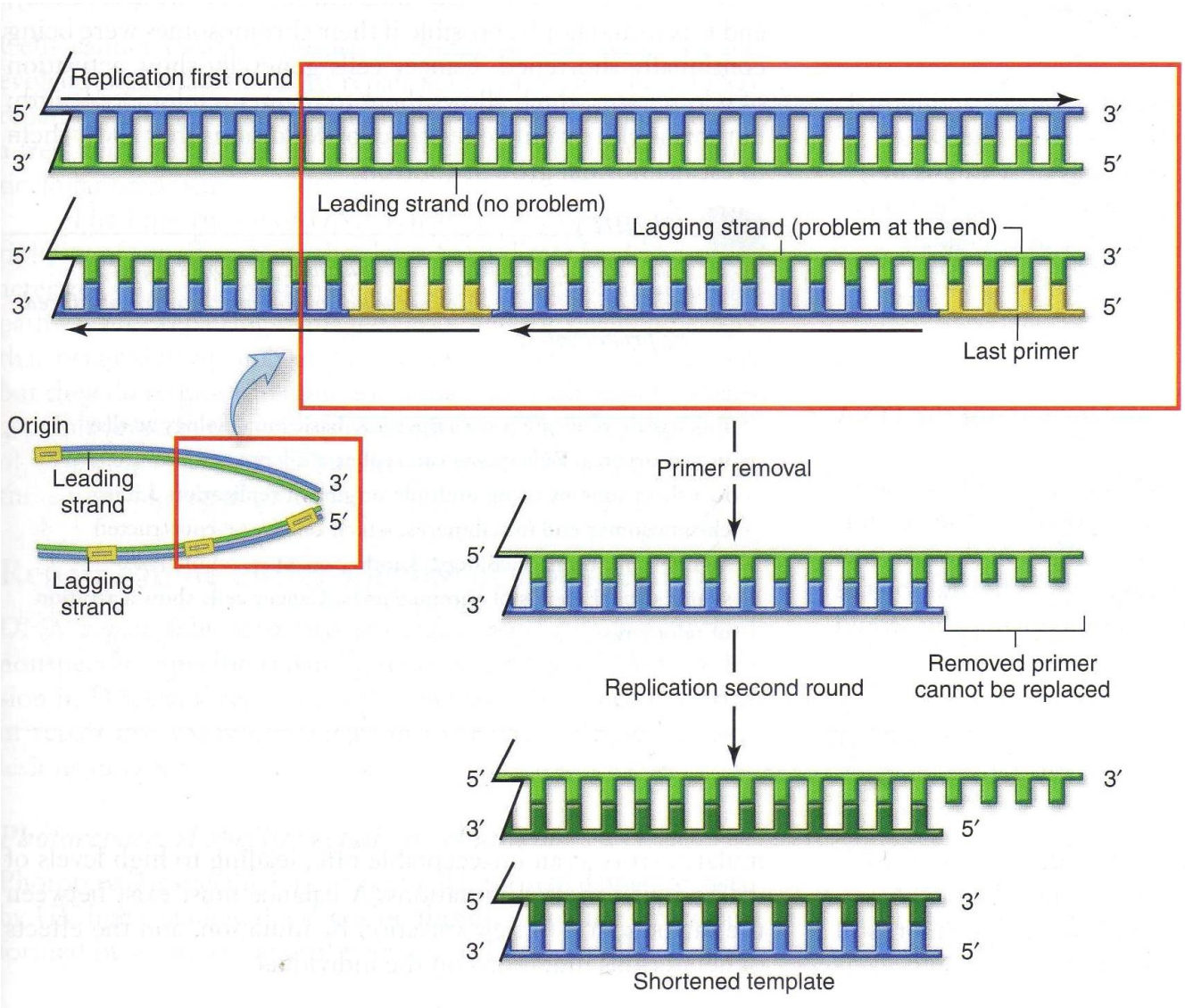
The linear chromosomes of eukaryotic cells terminate at both ends in protective caps called telomeres, which is composed of DNA associated with proteins, these caps contain no protein-encoding genes, but are crucial in **preserving the integrity of each chromosome**. Telomeres are composed of the base sequence 5'-TTAGGG-3' repeated 250 – 1000 times, and is highly conserved in all mammals. (*The role of telomere will be discussed in greater depth in future lectures*)



During replication, at the position of the last primer, the DNA polymerase involved in removing the RNA primer and is unable to replacing it with the DNA nucleotides due to the lack of a 3'OH end preceding it to attach new nucleotides.

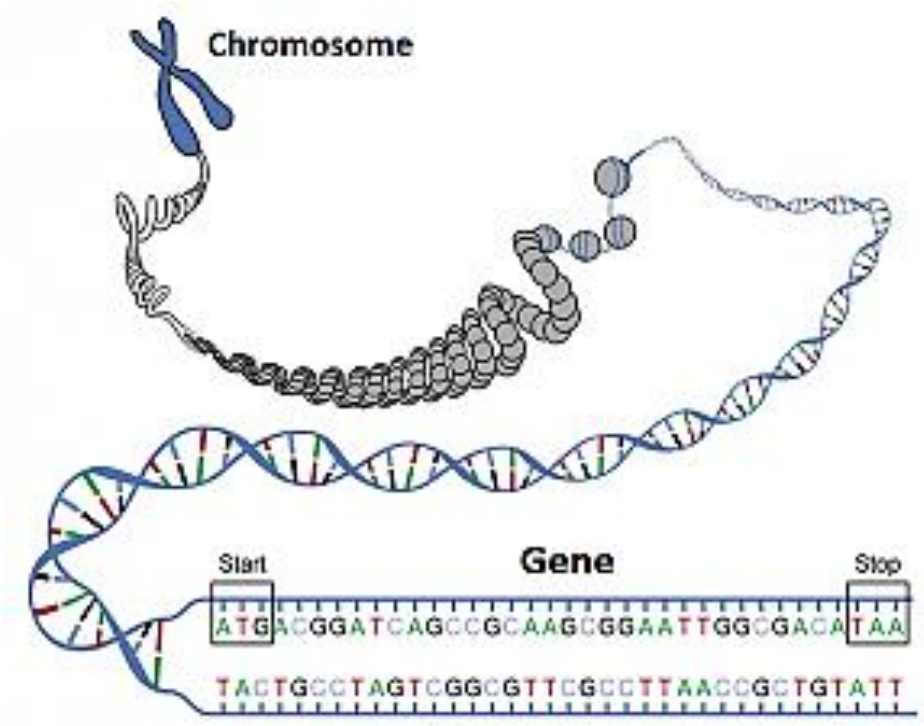
The result would be a gradual shortening of chromosomes with each round of cell division (loss of 50-200bp per round of replication). Thus, resulting in "end-replication problem" in eukaryotes. Prokaryotes do not undergo end-replication problem as it has a circular DNA.

Cellular senescence is triggered when telomeres reach a critical length (on average 4-6kb). The cell can no longer replicate, and dies by a process known as apoptosis.



8. The Genetic Code

The role of DNA is to **instruct the cell what specific proteins to make.** Genetic information is stored in the form of a specific sequence of nucleotides in a DNA molecule. It then determines the specific sequence of amino acids of the polypeptide chain (i.e. primary structure of protein).



Main features of the Genetic Code

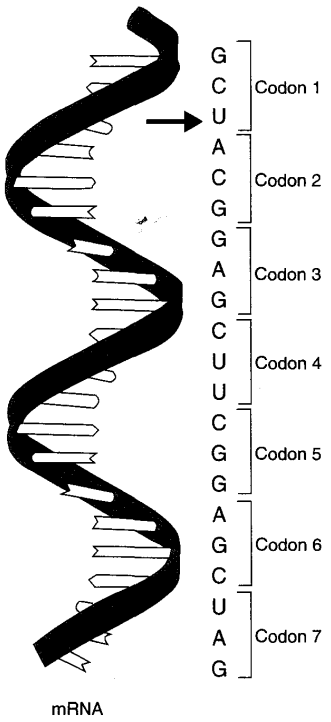
Triplet Codon: Three nucleotides will make up one codon, coding for one amino acid.

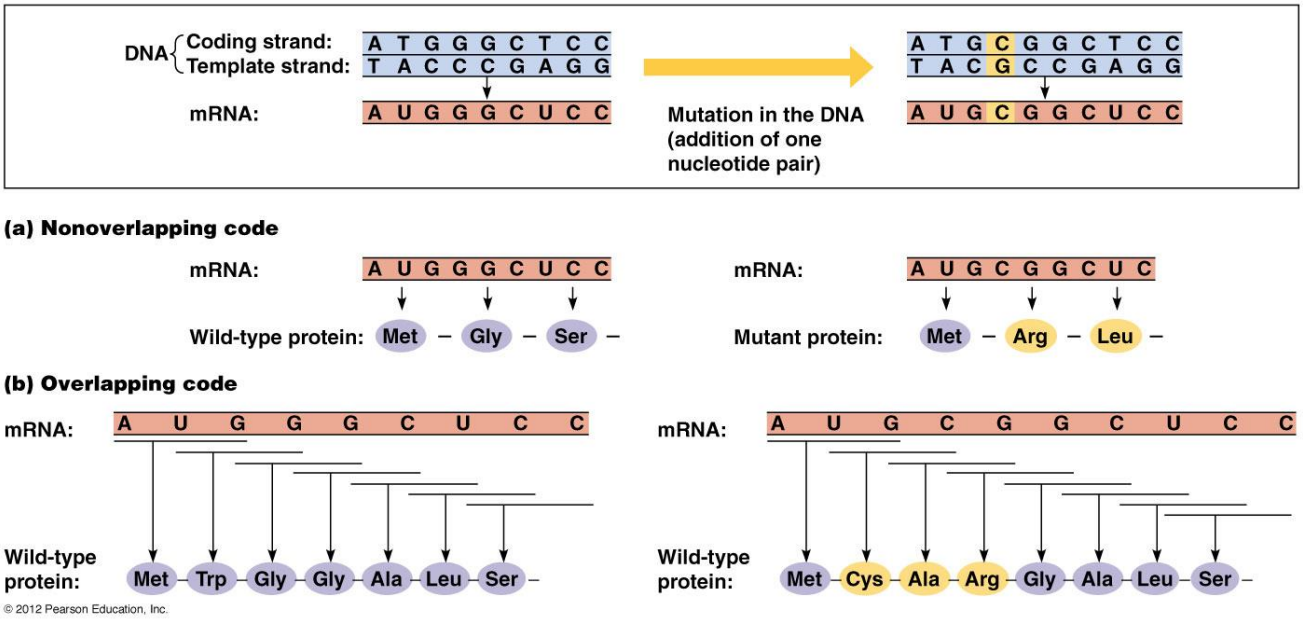
If it is a singlet code, 1 nucleotide will code for 1 amino acid and there will be only a maximum of 4 codons, and therefore, 4 amino acids only. If it is a duplet code, 2 nucleotides will code for 1 amino acid and there will be a maximum of $4^2 = 16$ codons, and therefore, 16 amino acids only. If it is a triplet code, 3 nucleotides will code for 1 amino acid and there will be a maximum of $4^3 = 64$ codons, and therefore would be able to code for more than 20 amino acids.

Universal Code: The same triplet codon codes for the same amino acid in all organisms.

Degenerate Code: A given amino acid may be coded by more than 1 codon. e.g. Valine is coded by GUU, GUC & GUA

Non-Overlapping Triplet Code: Each nucleotide in a triplet can be used only once and is read once as a triplet by the ribosome





Continuous: The codon sequence is read continuously eg. A-U-G-U-U-C as two codons AUG and UUC

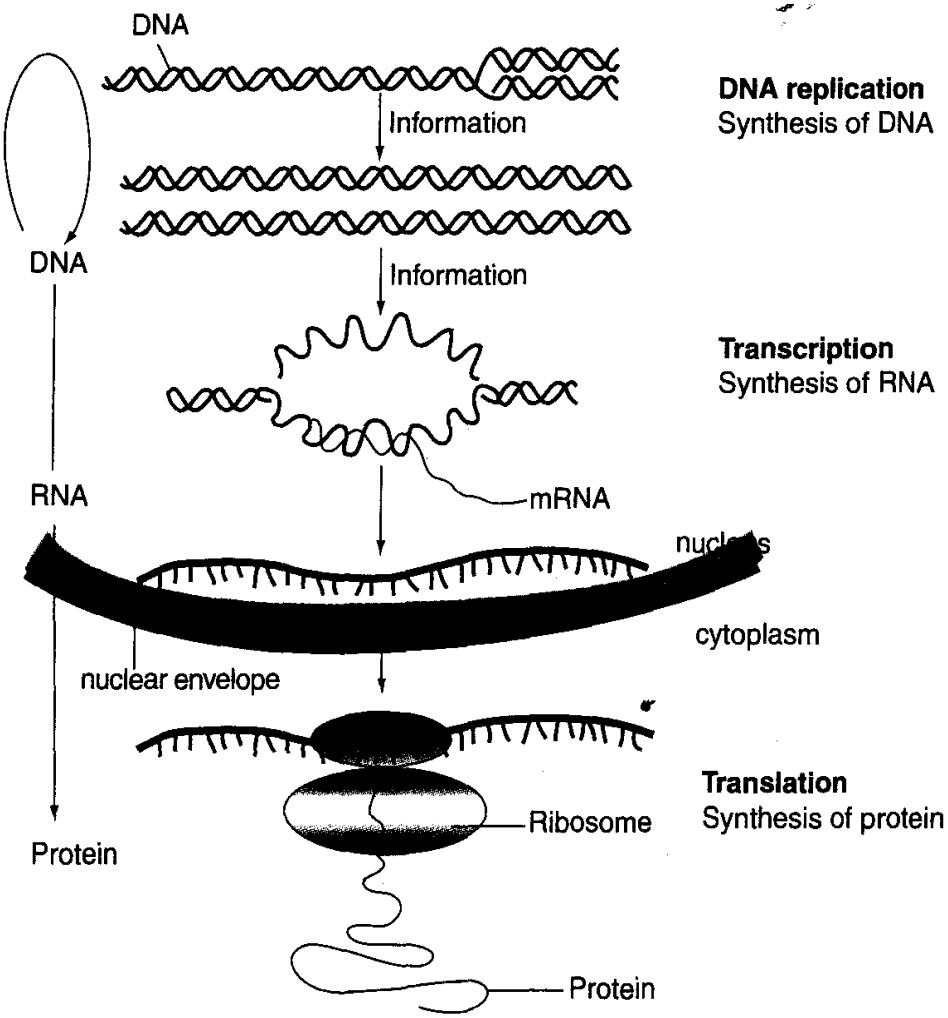
Punctuated: There are codons for initiation and termination of translation. The codon AUG is a start codon and it signals the initiation of translation of an mRNA into a polypeptide. The three codons, UGA, UAG and UAA, are stop codons and they signal the termination of translation.

	Amino acid
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartic acid
Cys	cysteine
Gln	glutamine
Glu	glutamic acid
Gly	glycine
His	histidine
Ile	isoleucine
Leu	leucine
Lys	lysine
Met	methionine
Phe	phenylalanine
Pro	proline
Ser	serine
Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine

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

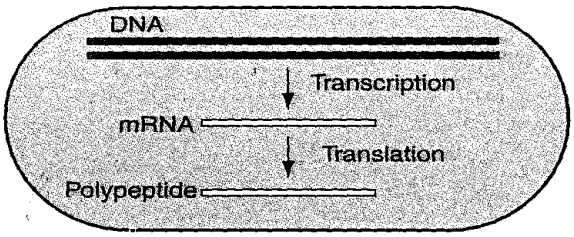
9. Protein synthesis

Protein synthesis is the transfer of the coded information from DNA to mRNA (TRANSCRIPTION) and mRNA to polypeptides (TRANSLATION). Both these stages involve ribonucleic acid (RNA). The three different RNA are messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). Protein synthesis allows genes to be expressed into proteins to allow them to carry out physiological functions and impacts.

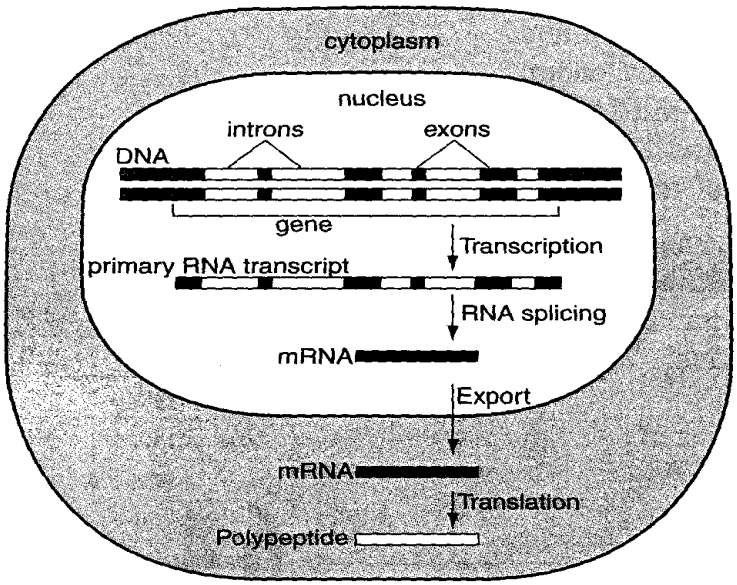


Proteins synthesis in prokaryotes and eukaryotes are different due to their differences in cellular and genomic structures. In terms of location, prokaryotic transcript and translation occur in the cytoplasm while for eukaryotes, transcription occurs in the nucleus and translation occurs in the cytoplasm. In terms of presence of post-transcriptional modification, in prokaryotes, there are no such modification and mRNA will be directly translated after transcription. However, for eukaryotes, the mRNA undergoes post-transcriptional modifications to form the mature mRNA before being translated.

(a) PROKARYOTES



(b) EUKARYOTES



Transcription

The enzyme **RNA polymerase** binds to a region of the DNA template strand near the beginning of the gene to be transcribed. This region is called the **promoter**. This region serves as a recognition site, it will not be transcribed.

In prokaryotes, the promoter contains the **Pribnow box** (TATAAT sequence) while the eukaryotic promoter contains the **TATA box**.

The TATA box is named for its conserved DNA sequence, which is most commonly TATAAA. Many eukaryotic genes have a conserved TATA box located 25-35 base pairs before the transcription start site of a gene. Proteins called transcription factors can bind to the TATA box and recruit an enzyme called RNA polymerase, which synthesizes RNA from DNA.

The Pribnow box has a function similar to the TATA box, it is recognized and bound by a subunit of RNA polymerase during initiation of transcription. This region of the DNA is also the first place where base pairs separate during prokaryotic transcription to allow access to the template strand.

These promoter regions the eukaryotes and prokaryotes are recognised and bound to by proteins which will recruit and facilitate the binding of RNA polymerase to the promoter. In prokaryotes, the Pribnow box is recognised by sigma factor while the TATA box in eukaryotes is recognised by the TATA binding Protein (TFIID, which stands for Transcription Factor II D). There other proteins which are involved in facilitating transcription.

Diagram showing Prokaryotic Transcription:

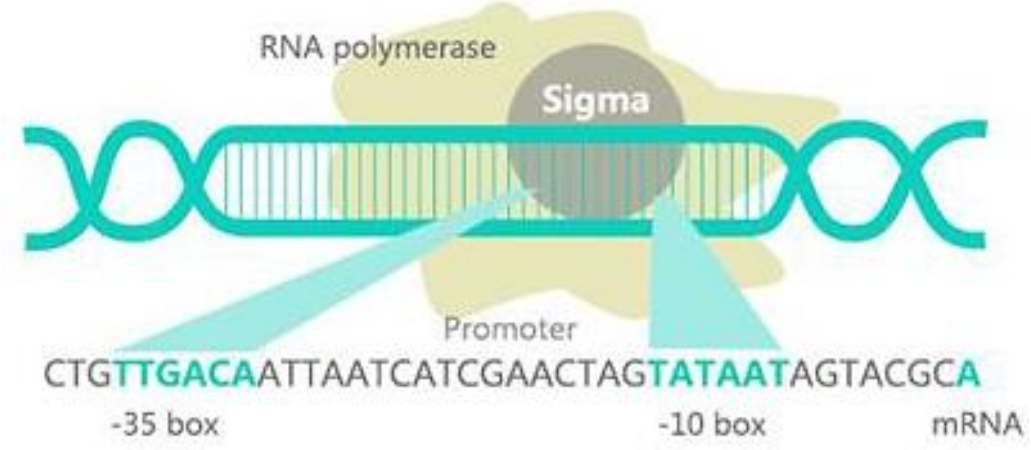
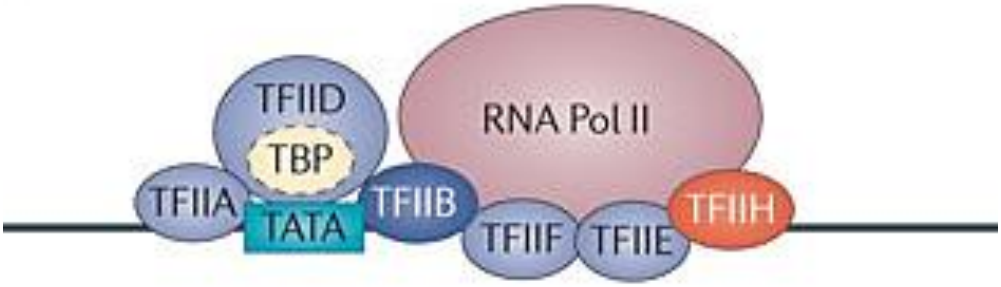


Diagram showing Eukaryotic Transcription:



After binding of the RNA polymerase to the promoter, helicase activity (within RNA polymerase in the case of prokaryotic RNA polymerase and transcription factors in the case for eukaryotes) causes the unwinding of DNA by catalysing the breaking of the relatively weak hydrogen bonds between the bases of the two strands, thus exposing single strands of DNA.

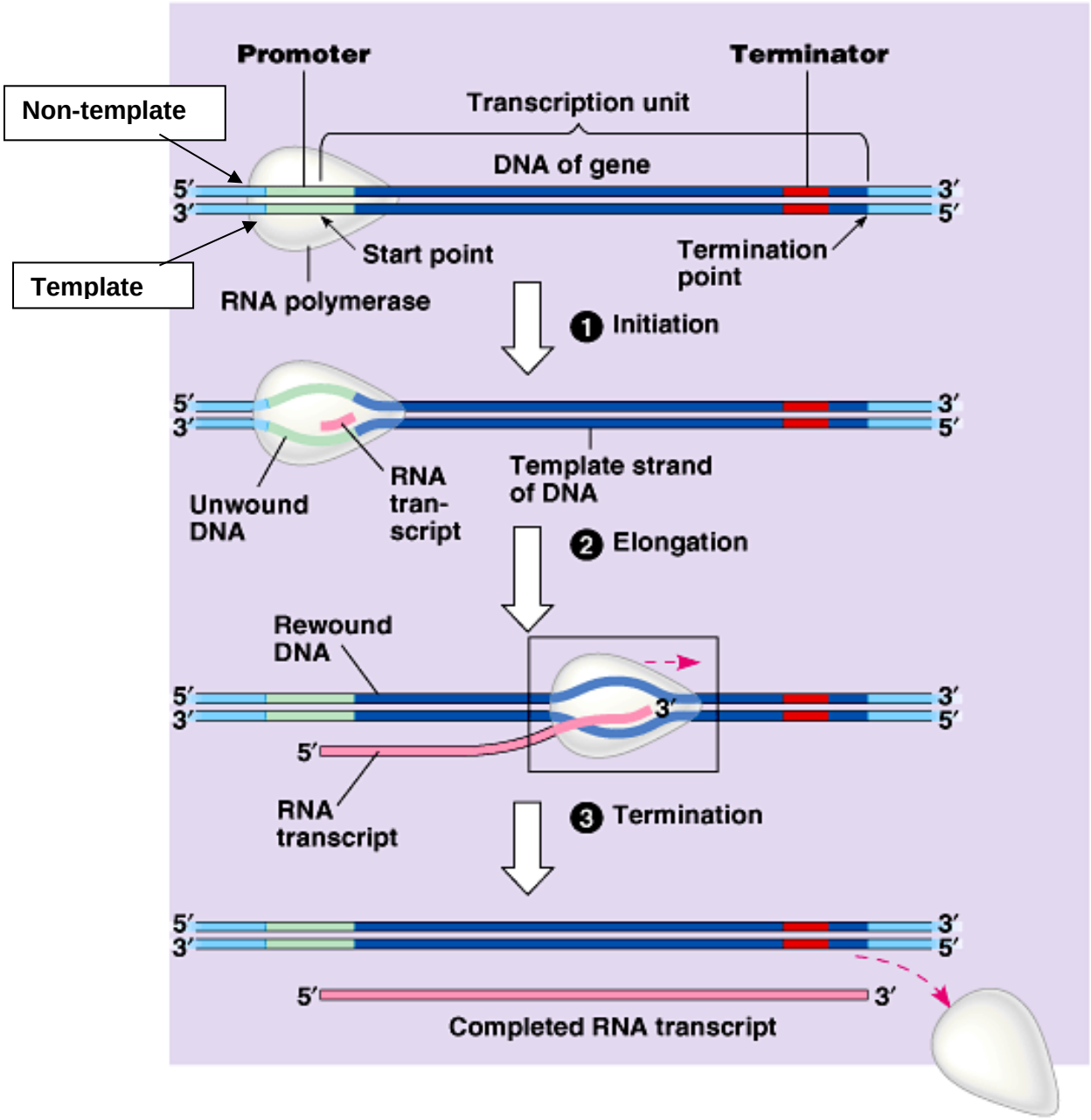
One of the exposed strands of DNA in the **3' → 5' direction**, called the **template strand**, is used as a **template** for transcription, to synthesise RNA which can be translated into functional proteins.

The other complementary DNA strand, in the direction of **5' → 3'**, is usually **not transcribed**. It is also known as the **non-template strand**, because it is a direct message of the RNA. If transcription of the DNA strand running in the **5' → 3'** takes place, then the RNA that is being synthesised would be not be able to be translated into functional proteins.

Actual transcription starts about 6 bases downstream from the promoter. As RNA polymerase moves along the anti-sense strand (also known as the coding strand) from the 3' to 5' end, free ribonucleotides, which are complementary to the DNA deoxyribonucleotides, are taken up and matched up with the DNA template by **complementary base pairing and the new mRNA strand** is synthesised in the **5' to 3' direction**.

Adenine in the DNA pairs with uracil in the RNA, thymine in the DNA pairs with adenine in the RNA, guanine in the DNA pairs with cytosine in the RNA and cytosine in the DNA pairs with guanine in the RNA.

Adjacent ribonucleotides are joined together by phosphodiester bonds to form the sugar-phosphate backbone via condensation. When the RNA polymerase reaches the end of the gene, transcription ends after reading the terminator sequence. Do note that the stop codons are also transcribed into the primary mRNA transcript.



Post-Transcriptional Modification

The primary RNA transcript is transcribed from a eukaryotic gene that has both exons and introns which contain the coding and the non-coding sequences respectively.

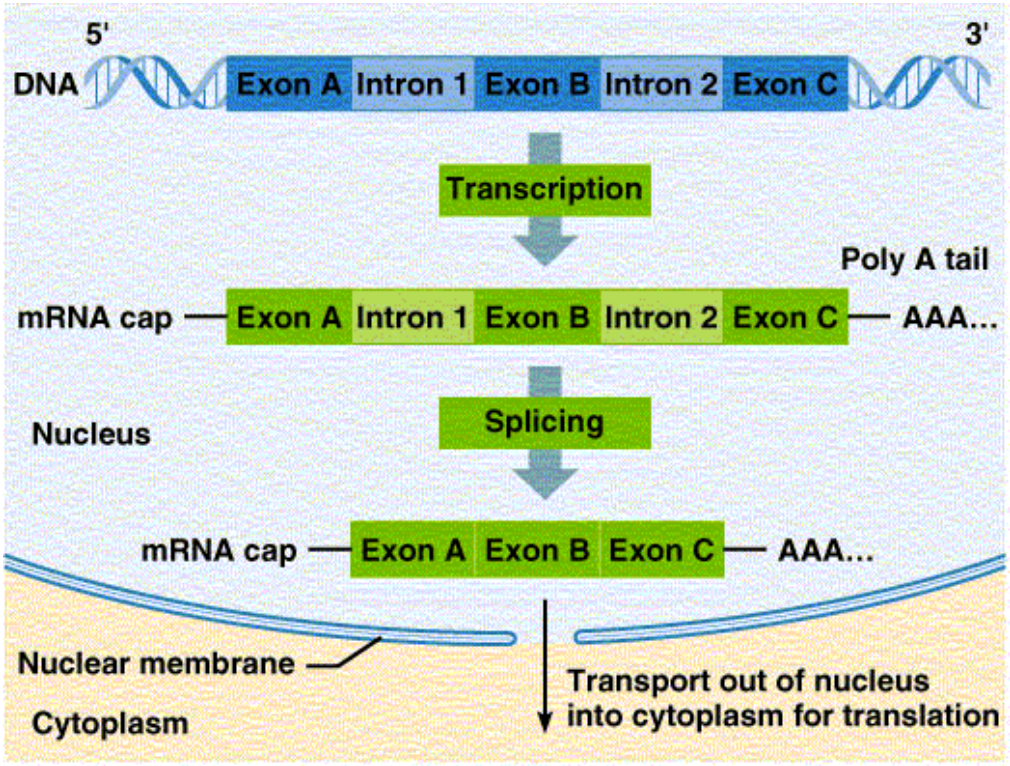
In eukaryotes, before the mRNA is translated, it undergoes mRNA processing consisting 3 steps:

1. **Addition of 5' cap [5'cap=modified form of guanine]**
Which acts as a signal promoting translation, once the mRNA reaches a ribosome

2. **Addition of 3' poly-A tail [A : Adenine]**
Which acts as a signal for the export of the mRNA from the nucleus and it is also considered to offer some protection from nucleases present in the cytoplasm that degrade the 3' end of the mRNA.

3. **Splicing**
The introns (non-coding sequences) are cut out from the primary RNA transcript while the coding sequences are simultaneously joined together to form the functional mature mRNA through a process called Splicing facilitated by the Spliceosome.

The mature mRNA then leaves the nucleus through the pores in the nuclear envelope.



Translation

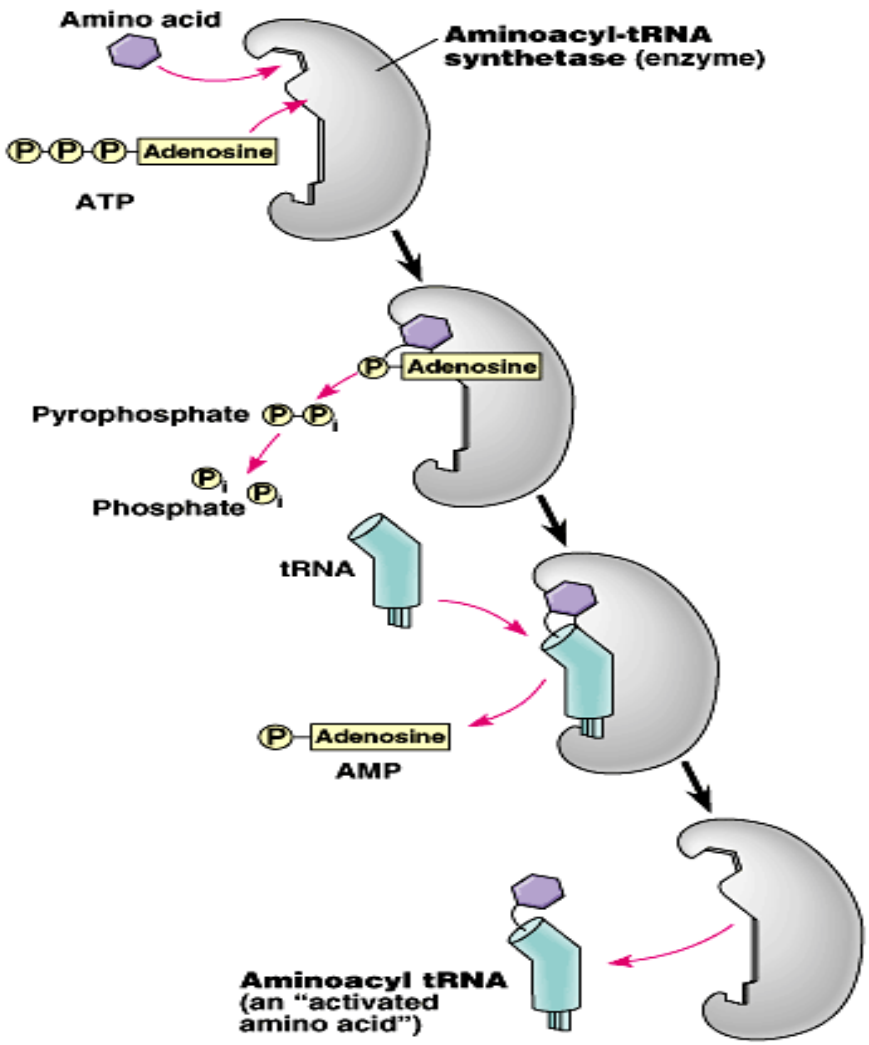
Translation is the process by which a sequence of nucleotides in an mRNA molecule is converted into a sequence of amino acids in a polypeptide. It involves the following steps:

Amino acid activation

The covalent attachment of an amino acid to its specific tRNA is catalysed by a group of enzymes known as **aminoacyl tRNA synthetases**, energy in the form of ATP is required.

An aminoacyl tRNA synthetase has an active site, which specifically binds one type of amino acid to a specific tRNA molecule. There are at least 20 aminoacyl tRNA synthetases, each one specific in the amino acid it binds to.

Amino acid is attached to a specific tRNA molecule to form an aminoacyl-tRNA complex. The 3' end of tRNA ends with the base sequence CCA. The specific amino acid will be attached to the hydroxyl group at carbon atom 3 of ribose of the last nucleotide that contains the base adenine.



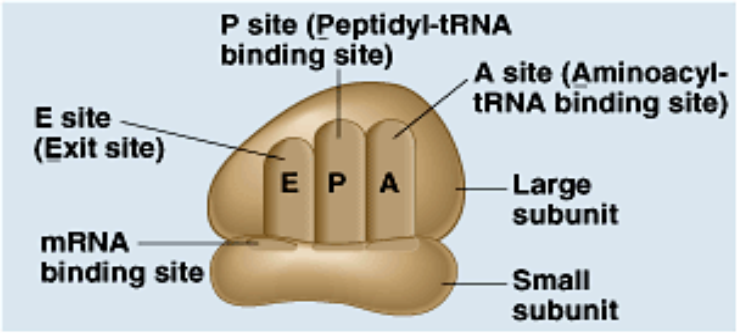
Try This !

Qn: How does the tRNA recognize their corresponding amino acid?

STEP 1: Translation Initiation

The structure of the bacterial and eukaryotic ribosome is roughly similar. A ribosomal subunit is an aggregate of ribosomal RNA molecules and proteins. A ribosome is composed of a larger subunit and a small subunit.

A ribosome has an mRNA binding site and three tRNA binding sites, known as the Exit (E) site, Peptidyl-tRNA binding (P) site and Aminoacyl tRNA binding (A) site



The information processing region of the **small subunit** of a ribosome binds to the mRNA at its 5' end at the **ribosome binding site**. The information processing region of the ribosome can only cover two codons on the mRNA at any one time and this scans the mRNA's 5' **untranslated region (UTR)** to locate the **translation initiation codon** (AUG – start codon).

In prokaryotes, the length of the 5′ UTR tends to be 3-10 nucleotides long, while in eukaryotes it tends to be anywhere from 100 to several thousand nucleotides long. The elements of a eukaryotic and prokaryotic 5′ UTR differ greatly. The prokaryotic 5′ UTR contains a ribosome binding site, also known as the **Shine Dalgarno sequence** (AGGAGGU), which is usually 3-10 base pairs upstream from the initiation codon. In contrast, the eukaryotic 5′ UTR contains the **Kozak consensus sequence** (ACCAUGG), which contains the initiation codon.

A tRNA fits into a binding site when its anticodon base-pairs with an mRNA codon. The P site holds the tRNA attached to the growing polypeptide. The A site holds the tRNA carrying the next amino acid to be added to the polypeptide chain. Discharged tRNA without an amino acid leaves via the E site.

Each tRNA has a specific code made up of 3 nucleotides, known as the **anticodon.** The anticodon is complementary to the codon on the mRNA. Hence, a specific aminoacyl-tRNA complex will bind to a specific codon on the mRNA strand due to complementary codon-anticodon base pairing.

An aminoacyl-tRNA complex, with the anticodon UAC, will therefore bind to the AUG (Start Codon) on the mRNA.

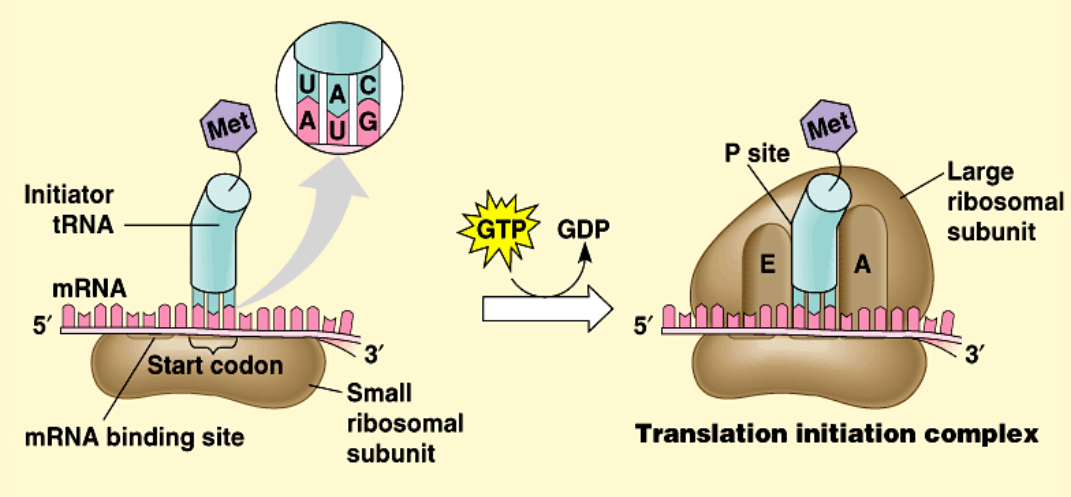
tRNA, with anticodon UAC, carries the amino acid **methionine**, which is always the first amino acid in a polypeptide.

So the **first aminoacyl-tRNA complex** is always **methionine-tRNA(Met-tRNA) complex**.

Formylation of the first amino acid methionine is to ensure that the addition of the subsequent amino acid is at the carboxyl end of methionine, such that the polypeptide is synthesised from the amino end to the carboxyl end. This is because the formyl group blocks the amino end of methionine, thus oxidation of peptide bond at the amino end of the methionine cannot take place.

Met-tRNA binds to the start codon on the mRNA. The **large ribosomal subunit** now binds to form the **translation initiation complex** and protein synthesis now begins.

The aminoacyl-tRNA complex is positioned at the P (peptidyl-tRNA binding) site of the **large subunit of the ribosome**.



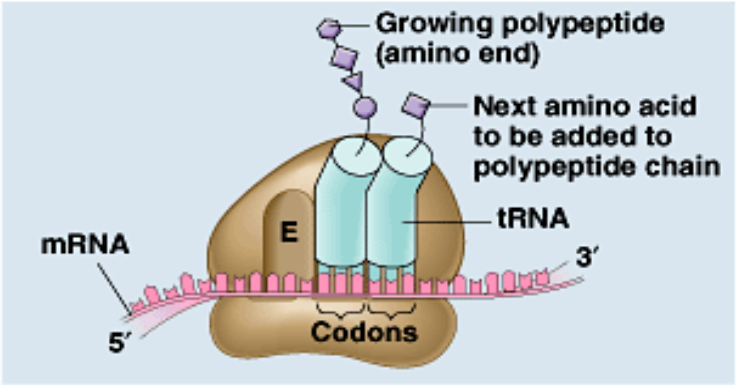
STEP 2: Translation Elongation

The amino acids are added on, one at a time, to a growing polypeptide chain, the order is specified by the base sequence on the mRNA codon.

The second aminoacyl-tRNA complex, with anticodon complementary to the second codon of the mRNA, binds with the second codon.

The second aminoacyl-tRNA complex is held at the A (aminoacyl-tRNA binding) site of the large subunit of the ribosome.
The first amino acid (methionine) is transferred from its tRNA to the aminoacyl tRNA complex at the A site, forming a **peptidyl-tRNA**.

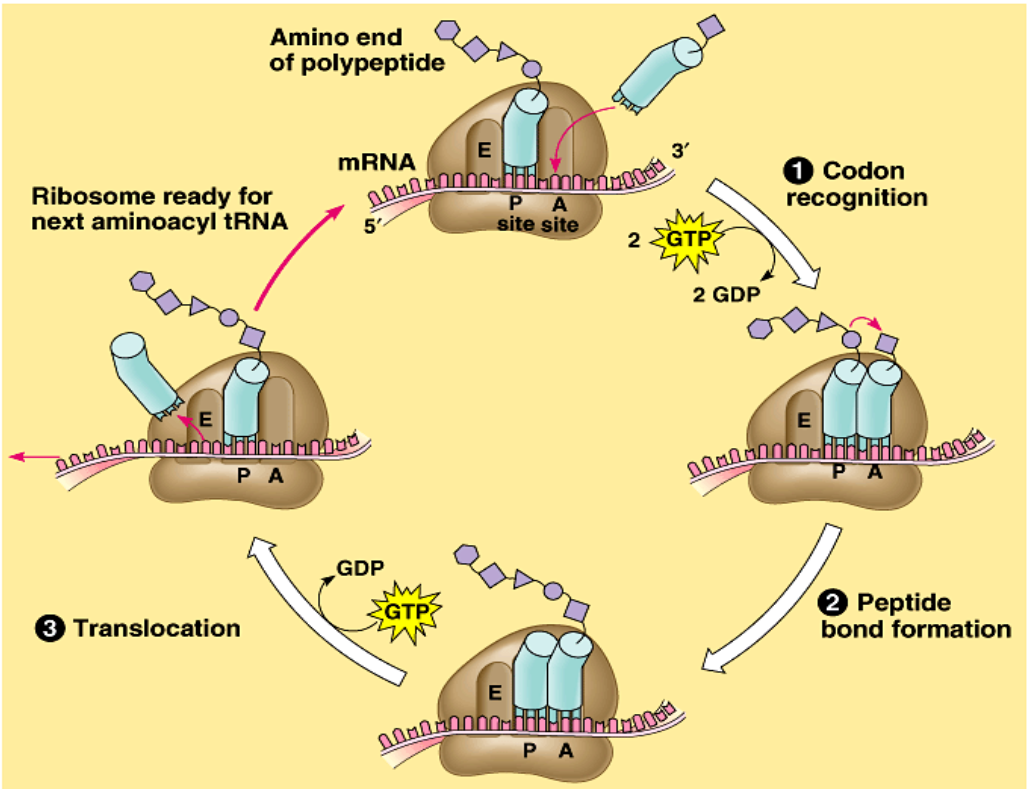
The two amino acids are joined by a peptide bond, and the reaction is catalysed by the enzyme **peptidyl transferase.**



The **tRNA at the P** site is now "**uncharged**", without an amino acid, and it is ejected from the ribosome into the cytoplasm, where it can be **recycled**. This means the tRNA molecule can reattach to a new amino acid.

The peptidyl-tRNA now enters Peptidyl-tRNA binding (P) site from A site as the ribosome is translocated one codon down the mRNA.

The A site now has the third codon of the mRNA, allowing it to receive another aminoacyl-tRNA complex, which has an anticodon complementary to the third codon on the mRNA.



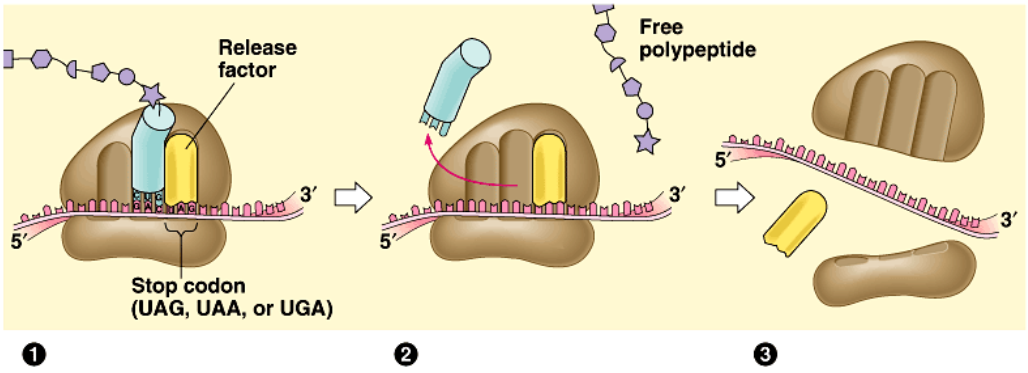
The process is repeated until the ribosome reaches the stop codon on the mRNA, i.e. UAA, UAG or UGA.

STEP 5: Chain termination

Once the stop codon (UAA, UAG or UGA) is read, a release factor occupies the A site and the bond between the final amino acid and tRNA is hydrolysed.

The polypeptide is released from the ribosome and it assumes its secondary and/or tertiary structures automatically in the lumen of the rER. The polypeptide may also undergo chemical modification at the Golgi body. For cytoplasmic ribosomes, folding would be done within the cytoplasm.

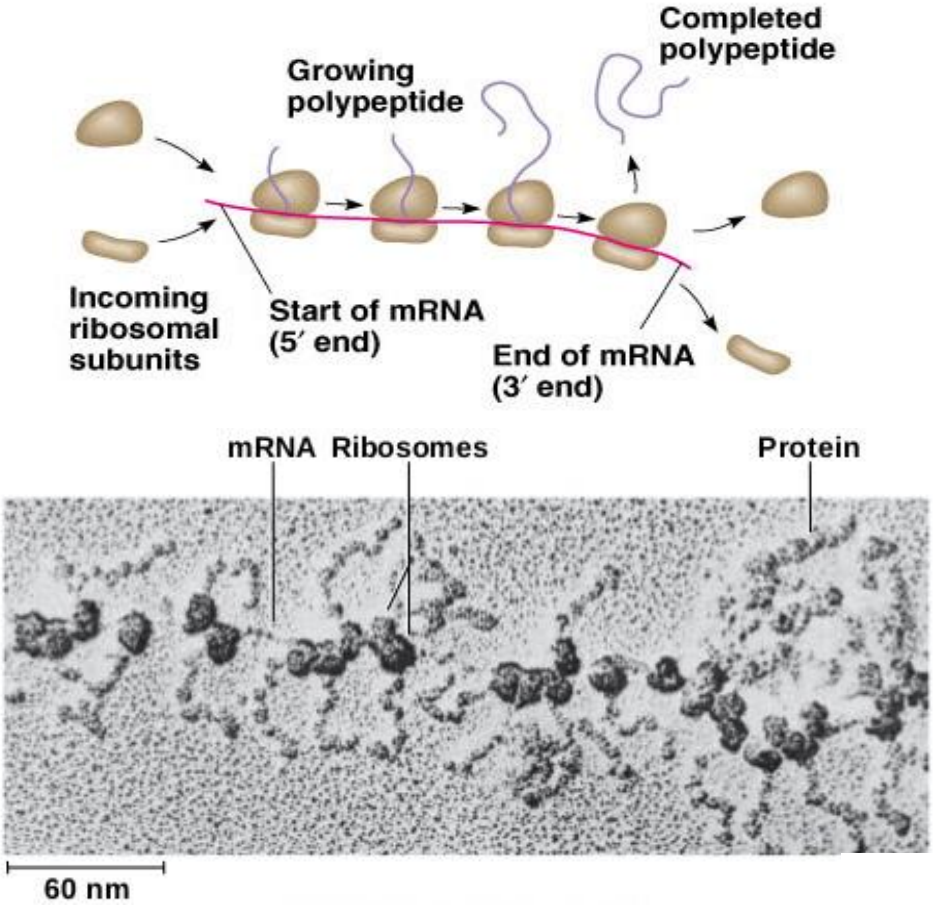
If the ribosome is not in use, it will then disassemble into its subunits. It may be reassembled for future use.



BROYO

Polyribosome

The second and subsequent ribosomes may pass along the mRNA immediately behind the first ribosome. The whole structure of mRNA with many ribosomes attached to it is called **polyribosome or polysome**. This would mean that a large number of similar polypeptides can be assembled on a single mRNA in a relatively short period.



Try This !

Compare the differences between DNA Replication Transcription and Translation:

Feature	DNA Replication	Transcription	Translation
Location of Process			
Start Site			
Product(s)			
Template			
Enzymes involved			
Other components/ factors involved			
Building blocks/ raw materials			
Reading of genetic message			

Note that this least of points of comparison is non-exhaustive. There are more features that can be compared.