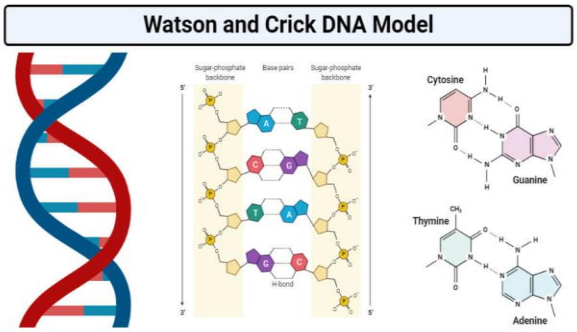
Paper V – Molecular Biology
UNIT 1 – Genome Structure

Watson and Crick DNA Model

- **DNA** stands for Deoxyribonucleic acid which is a molecule that contains the instructions an organism needs to develop, live and reproduce.
- It is a type of nucleic acid and is one of the four major types of macromolecules that are known to be essential for all forms of life.



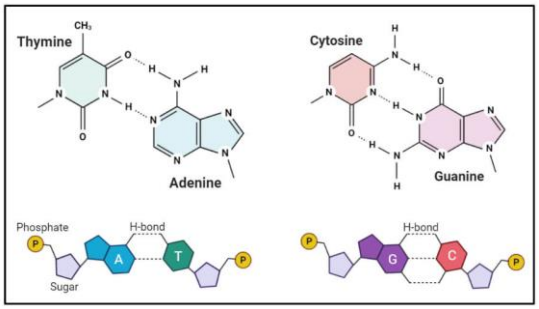
DNA Model

- The three-dimensional structure of DNA, first proposed by James D. Watson and Francis H. C. Crick in 1953, consists of two long helical strands that are coiled around a common axis to form a double helix.
- Each DNA molecule is comprised of two biopolymer strands coiling around each other.
- Each strand has a 5′end (with a phosphate group) and a 3′end (with a hydroxyl group).
- The strands are antiparallel, meaning that one strand runs in a 5′to 3′direction, while the other strand runs in a 3′to 5′direction.
- The diameter of the double helix is 2nm and the double-helical structure repeats at an interval of 3.4nm which corresponds to ten base pairs.
- The two strands are held together by hydrogen bonds and are complementary to each other.
- The two DNA strands are called polynucleotides, as they are made of simpler monomer units called nucleotides. Basically, the DNA is composed of deoxyribonucleotides.
- The deoxyribonucleotides are linked together by 3′- 5′phosphodiester bonds.
- The nitrogenous bases that compose the deoxyribonucleotides include adenine, cytosine, thymine, and guanine.
- The structure of DNA -DNA is a double helix structure because it looks like a twisted ladder.
- The sides of the ladder are made of alternating sugar (deoxyribose) and phosphate molecules while the steps of the ladder are made up of a pair of nitrogen bases.
- As a result of the double-helical nature of DNA, the molecule has two asymmetric grooves. One groove is smaller than the other.
- This asymmetry is a result of the geometrical configuration of the bonds between the phosphate, sugar, and base groups that forces the base groups to attach at 120-degree angles instead of 180 degrees.
- The larger groove is called the major groove, occurs when the backbones are far apart; while the smaller one is called the minor groove, and occurs when they are close together.

- Since the major and minor grooves expose the edges of the bases, the grooves can be used to tell the base sequence of a specific DNA molecule.
- The possibility for such recognition is critical since proteins must be able to recognize specific DNA sequences on which to bind in order for the proper functions of the body and cell to be carried out.

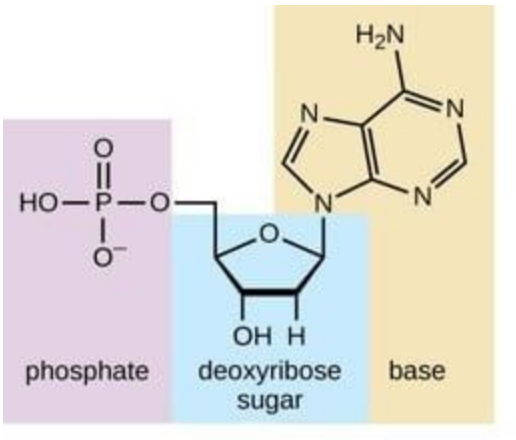
Components of DNA Double Helix Structure

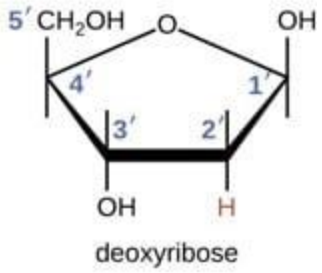
The Nitrogen Bases or Nucleotides



- DNA strands are composed of monomers called nucleotides.
- These monomers are often referred to as bases because they contain cyclic organic bases.
- Four different nucleotides, abbreviated A, T, C, and G, (adenine, thymine, cytosine, and guanine) are joined to form a DNA strand, with the base parts projecting inward from the backbone of the strand.
- Two strands bind together via the bases and twist to form a double helix.
- The nitrogen bases have a specific pairing pattern. This pairing pattern occurs because the amount of adenine equals the amount of thymine; the amount of guanine equals the amount of cytosine. The pairs are held together by hydrogen bonds.
- Each DNA double helix thus has a simple construction: wherever one strand has an A, the other strand has a T, and each C is matched with a G.
- The complementary strands are due to the nature of the nitrogenous bases. The base adenine always interacts with thymine (A-T) on the opposite strand via two hydrogen bonds and cytosine always interacts with guanine (C-G) via three hydrogen bonds on the opposite strand.
- The shape of the helix is stabilized by hydrogen bonding and hydrophobic interactions between bases.

Deoxyribose Sugar





- Deoxyribose, also known as D-Deoxyribose and 2-deoxyribose, is a pentose sugar (monosaccharide containing five carbon atoms) that is a key component of the nucleic acid deoxyribonucleic acid (DNA).
- It is derived from the pentose sugar ribose. Deoxyribose has the chemical formula $C_5H_{10}O_4$.
- Deoxyribose is the sugar component of DNA, just as ribose serves that role in RNA (ribonucleic acid).
- Alternating with phosphate bases, deoxyribose forms the backbone of the DNA, binding to the nitrogenous bases adenine, thymine, guanine, and cytosine.
- As a component of DNA, which represents the genetic information in all living cells, deoxyribose is critical to life. This ubiquitous sugar reflects a commonality among all living organisms.

The Phosphate Group (Phosphate Backbone)

Created with BioRender.com

- The sugar-phosphate backbone forms the structural framework of nucleic acids, including DNA.
- This backbone is composed of alternating sugar and phosphate groups and defines the directionality of the molecule.
- DNA are composed of nucleotides that are linked to one another in a chain by chemical bonds, called ester bonds, between the sugar base of one nucleotide and the phosphate group of the adjacent nucleotide.
- The sugar is the 3′ end, and the phosphate is the 5′ end of each nucleotide.
- The phosphate group attached to the 5′ carbon of the sugar on one nucleotide forms an ester bond with the free hydroxyl on the 3′ carbon of the next nucleotide.
- These bonds are called phosphodiester bonds, and the sugar-phosphate backbone is described as extending, or growing, in the 5′ to 3′ direction when the molecule is synthesized.
- In double-stranded DNA, the molecular double-helix shape is formed by two linear sugar-phosphate backbones that run opposite each other and twist together in a helical shape.
- The sugar-phosphate backbone is negatively charged and hydrophilic, which allows the DNA backbone to form bonds with water.

Functions of DNA

DNA has a crucial role as genetic material in most living organisms. It carries genetic information from cell to cell and from generation to generation.

Thus its major functions include:

- Storing genetic information
- Directing protein synthesis
- Determining genetic coding
- Directly responsible for metabolic activities, evolution, heredity, and differentiation.

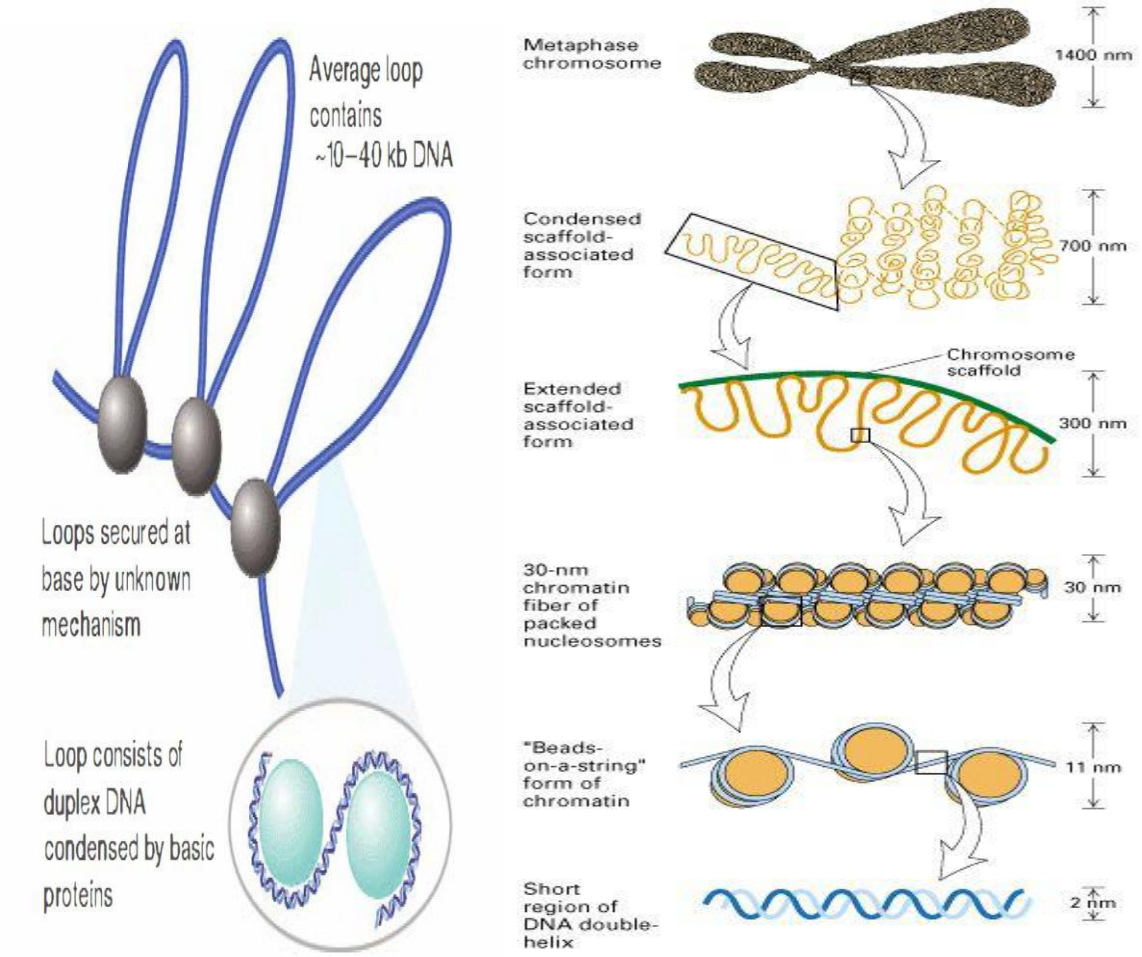
It is a stable molecule and holds more complex information for longer periods of time.

Genome organization:

- The genome of an organism encompasses all of the genes of that organism.
- **Gene** is a distinct sequence of nucleotides forming part of a chromosome, the order of which determines the order of monomers in a polypeptide or nucleic acid molecule.
- Thus a protein-coding gene is defined as a region of DNA that encodes a single polypeptide or a set of closely related polypeptides.
- Genes are contained in chromosomes.
- Chromosomes are thus structures within cells that contain hundreds to thousands of genes.

Prokaryotic Chromosomes

- The DNA of a bacterial cell, such as *Escherichia coli*, is a circular double-stranded molecule often referred to as the bacterial chromosome.
- The circular DNA is packaged into a region of the cell called the nucleoid where it is organized into 50 or so loops or domains that are bound to a central protein scaffold, attached to the cell membrane.
- The DNA is negatively supercoiled, that is, it is twisted upon itself.
- It is complexed with several DNA-binding proteins, the most common of which are proteins HU, HLP-1 and H-NS. These are histone-like proteins.



Eukaryotic Chromosomes

- The large amount of genomic DNA in a eukaryotic cell is tightly packaged in chromosomes contained within a specialized organelle, the nucleus.
- With the exception of the sex chromosomes, diploid eukaryotic organisms such as humans have two copies of each chromosome, one inherited from the father and one from the mother.
- Chromosomes contain both DNA and protein.
- Most of the protein on a weight basis is histones, but there are also many thousands of other proteins found in far less abundance and these are collectively called **non- histone proteins (NHP).**
- This nuclear DNA–protein complex is called chromatin.
- In the nucleus, each chromosome contains a single linear double-stranded DNA molecule.
- The length of the packaged DNA molecule varies. In humans, the shortest DNA molecule in a chromosome is about 1.6 cm and the longest is about 8.4 cm.
- The extensive packaging of DNA in chromosomes results from three levels of folding involving nucleosomes, 30 nm filaments and radial loops.

1. Nucleosomes

- The first level of packaging involves the binding of the chromosomal DNA to histones.
- Overall, in chromosomes, the ratio of DNA to histones on a weight basis is approximately 1:1.
- There are five main types of histones called H1, H2A, H2B, H3 and H4.
- Histones are very basic proteins; about 25% of their amino acids are lysine or arginine so histones have a large number of positively charged amino acid side-chains.
- These positively charged groups therefore bind to the negatively charged phosphate groups of DNA

2. 30 nm fiber

- If nuclei are lysed very gently, the chromatin is seen to exist as a 30 nm diameter fiber.
- The fiber is formed by a histone H1 molecule binding to the linker DNA of each nucleosome at the point where it enters and leaves the nucleosome.
- The histone H1 molecules interact with each other, pulling the nucleosomes together.

3. Radial loops

- When chromosomes are depleted of histones, they are seen to have a central fibrous 'protein scaffold' (or nuclear matrix) to which the DNA is attached in loops.
- Therefore, in vivo it seems likely that the next order of packaging involves the attachment of the 30 nm fiber to multiple locations on this central protein scaffold in a series of radial loops.
- The mitochondria and chloroplasts of eukaryotic cells also contain DNA but, unlike the nuclear DNA, this consists of double-stranded circular molecules resembling bacterial chromosomes.

DNA as a genetic material

Transforming Principle:

- Frederick Griffith in 1928, carried out a series of experiments with Streptococcus pneumoniae (a bacterium that cause pneumonia). He observed that when these bacteria (Streptococcus pneumonia) are grown on a culture plate, some of them produce smooth, shiny colonies (S-type), whereas, the others produce rough colonies (R-type). This difference in character (smooth/rough) is due to a mucous (polysaccharide) coat present in the S-strain bacteria, which is not present in the R-strain.
- In his experiments, he first infected two separate groups of mice. The mice that were infected with the S-strain die from pneumonia.
- 'S' strains are the virulent strains causing pneumonia.
- The mice that were infected with the R-strain do not develop pneumonia and they live.
- S-strain (virulent strain) → Inject into mice → Mice die
- R-strain (non-virulent strain) → Inject into mice → Mice live
- In the next set of experiments, Griffith killed the bacteria by heating them. The mice that were injected heat-killed S-strain bacteria did not die and lived, whereas the mice that were injected a mixture of heat-killed S-strain and live R-strain bacteria, died due to unexpected symptoms of pneumonia.
- S-strain (heat killed) → Inject into mice → Mice live
- S-strain (heat killed) + R-strain (live) → Inject into mice → Mice die
- Griffith concluded that the live R-strain bacteria, were transformed by the heat-killed S-strain bacteria.

- He proved that there was some ‘transforming principle’ that was transferred from the heat-killed S-strain, which helped the R-strain bacteria to synthesise a smooth polysaccharide coat and thus, become virulent. That was due to the transfer of the genetic material.
- However, he was not able to define the biochemical nature of genetic material from his experiments.

Biochemical Characterisation of Transforming Principle:

- Oswald Avery, Colin MacLeod and Maclyn McCarty (1933-44) worked to determine the biochemical nature of ‘transforming principle’ in Griffith’s experiment in an in vitro system.
- From the heat-killed S-cells, they purified biochemicals (proteins, DNA, RNA, etc.) to observe, that which biochemicals could transform live R-cells into S-cells.
- Therefore, they discovered that DNA alone from heat-killed S-type bacteria caused the transformation of non-virulent R-type bacteria into S-type virulent bacteria.
- Protein-digesting enzymes (proteases) and RNA-digesting enzymes (RNases) did not cause this transformation. This proved that the ‘transforming substance’ was neither the protein no RNA.
- DNA-digesting enzyme (DNase) caused inhibition of transformation, which suggests that the DNA caused the transformation. Thus, these scientists came to the conclusion that DNA is the hereditary material.

Hershey and Chase Experiment:

- The proof for DNA as a genetic material came from the experiment. Alfred Hershey and Martha Chase (1952) carried out some experiments with the viruses that infect bacteria. These viruses are called bacteriophages.
- The genetic material of bacteriophage enters the bacterial cell after the bacteriophage gets attached to the bacteria. The bacterial cell treats the genetic material of the virus (bacteriophage) like its own genetic material and then produces more virus particles. Hershey and Chase experimented to find out whether it was protein or DNA from the virus that had entered into the bacteria.
- **For this, they took two separate media for growing these bacteriophages:**
- (i) Out of two, one medium contained radioactive phosphorus and the other medium contained radioactive sulphur. Viruses (bacteriophage) were then grown on each medium.
- (a) The viruses grown in the presence of radioactive phosphorus (^{32}P) contained radioactive DNA (but not radioactive protein). This is because DNA contains phosphorus not protein.

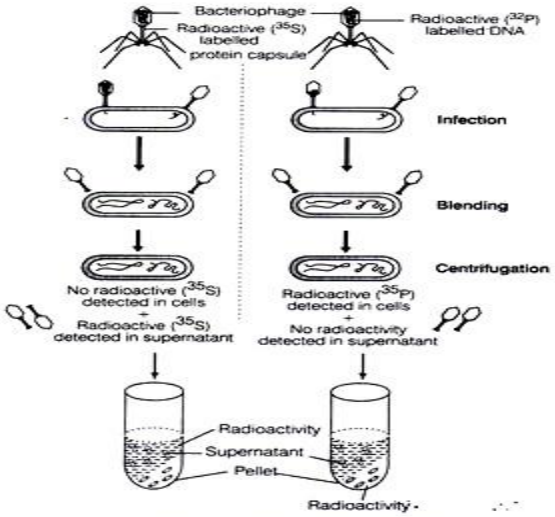
- (b) In the same way, the viruses grown in the medium containing radioactive sulphur (^{35}S) now contained radioactive protein (not radioactive DNA). This is because DNA does not contain sulphur.
- (ii) These radioactive viruses (bacteriophages) were then allowed to attach to bacteria (E. colt). As the process of infection with virus continued, the bacteria were agitated in a blender and the viral coats of the bacteria were removed.
- (iii) When they were spinned in a centrifuge, the virus particles were separated from the bacteria.
- (iv)They observed that the bacteria that were infected with virus containing radioactive DNA were radioactive, whereas the bacteria that were infected with radioactive proteins were not radioactive.
- (v) This indicates that only DNA not the protein coat entered the bacterial cell.
- (vi) Thus, the genetic material that is passed from virus to bacteria is DNA.
- 

Fig. 6.5 Hershey and Chase experiment

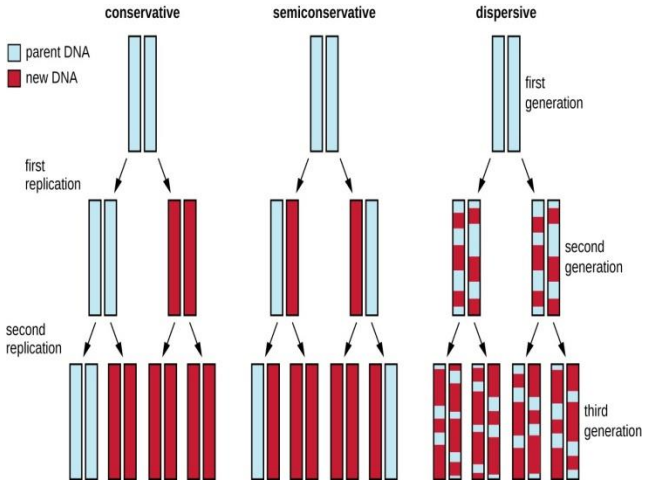
Properties of Genetic Material:
From the Hershey and Chase experiment, the fact was established that DNA acts as a genetic material. But later, studies revealed that in some viruses (e.g., Tobacco Mosaic Viruses, QB bacteriophage, etc.) RNA is the genetic material.

UNIT 2 – DNA Replication

Modes of DNA Replication

- **Semi-conservative replication.** In this model, the two strands of DNA unwind from each other, and each acts as a template for synthesis of a new, complementary strand. This results in two DNA molecules with one original strand and one new strand.
- **Conservative replication.** In this model, DNA replication results in one molecule that consists of both original DNA strands (identical to the original DNA molecule) and another molecule that consists of two new strands (with exactly the same sequences as the original molecule).

- **Dispersive replication.** In the dispersive model, DNA replication results in two DNA molecules that are mixtures, or "hybrids," of parental and daughter DNA. In this model, each individual strand is a patchwork of original and new DNA.



There were three models suggested for DNA replication. In the conservative model, parental DNA strands (blue) remained associated in one DNA molecule while new daughter strands (red) remained associated in newly formed DNA molecules. In the semiconservative model, parental strands separated and directed the synthesis of a daughter strand, with each resulting DNA molecule being a hybrid of a parental strand and a daughter strand. In the dispersive model, all resulting DNA strands have regions of double-stranded parental DNA and regions of double-

The Meselson-Stahl experiment

Meselson and Stahl conducted their famous experiments on DNA replication using *E. coli* bacteria as a model system.

They began by growing *E. coli* in medium, or nutrient broth, containing a "heavy" isotope of nitrogen, N^{15}. (An **isotope** is just a version of an element that differs from other versions by the number of neutrons in its nucleus.) When grown on medium containing heavy N^{15}, the bacteria took up the nitrogen and used it to synthesize new biological molecules, including DNA.

After many generations growing in the N^{15} medium, the nitrogenous bases of the bacteria's DNA were all labeled with heavy N^{15}. Then, the bacteria were switched to medium containing a "light" N^{14} isotope and allowed to grow for several generations. DNA made after the switch would have to be made up of N^{14}, as this would have been the only nitrogen available for DNA synthesis.

Meselson and Stahl knew how often *E. coli* cells divided, so they were able to collect small samples in each generation and extract and purify the DNA. They then measured the density of the DNA (and, indirectly, its N^{15} and N^{14} content) using **density gradient centrifugation**.

This method separates molecules such as DNA into bands by spinning them at high speeds in the presence of another molecule, such as cesium chloride, that forms a density gradient from the top to the bottom of the spinning tube. Density gradient centrifugation allows very small differences—like those between N^{15} and N^{14} -labeled DNA—to be detected.

The experiment done by Meselson and Stahl demonstrated that DNA replicated semi-conservatively, meaning that each strand in a DNA molecule serves as a template for synthesis of a new, complementary strand.

Although Meselson and Stahl did their experiments in the bacterium *E. coli*, we know today that semi-conservative DNA replication is a universal mechanism shared by all organisms on planet Earth. Some of your cells are replicating their DNA semi-conservatively right now!

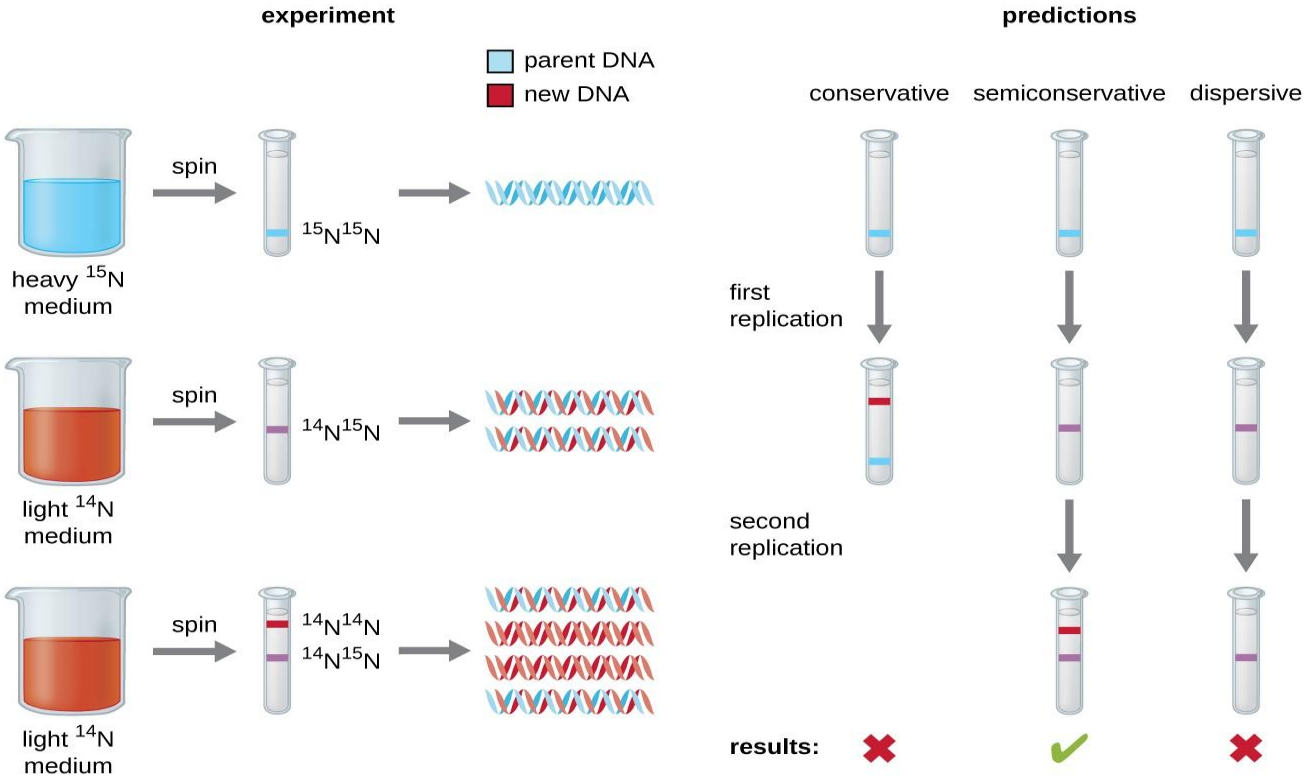


Figure 2. Meselson and Stahl experimented with *E. coli* grown first in heavy nitrogen (^{15}N) then in ^{14}N. DNA grown in ^{15}N (blue band) was heavier than DNA grown in ^{14}N (red band), and sedimented to a lower level on ultracentrifugation. After one round of replication, the DNA sedimented halfway between the ^{15}N and ^{14}N levels (purple band), ruling out the conservative model of replication. After a second round of replication, the dispersive model of replication was ruled out. These data supported the semiconservative replication model.

ROLLING CIRCULAR DNA REPLICATION

• Rolling circle replication describes a process of **unidirectional nucleic acid replication** that can rapidly synthesize **multiple copies of circular molecules** of DNA or RNA, such as **plasmids**, the genomes of bacteriophages, and the circular RNA genome of viroids.

• Some **eukaryotic viruses** also replicate their DNA via the rolling circle mechanism. Usually under the name **rolling circle amplification**, the mechanism is also widely used in the laboratory in molecular biology research and in nanotechnology.

Three basic steps:

i) Initiation

ii) Elongation

iii) Termination

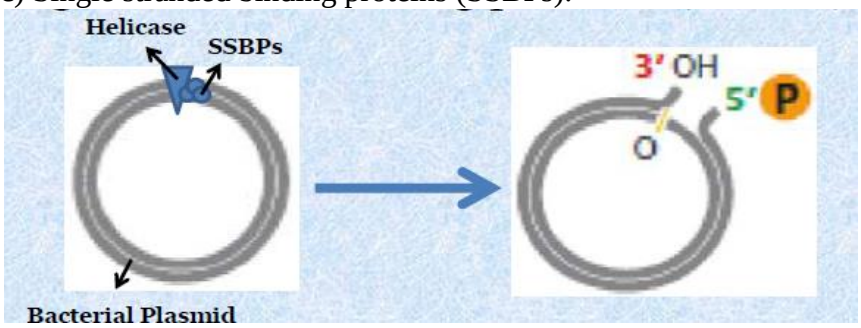
Initiation

Initiates by the production of nick on one of the two strands producing free 3́-OH and 5́phosphate ends, by the action of:

a) Helicase

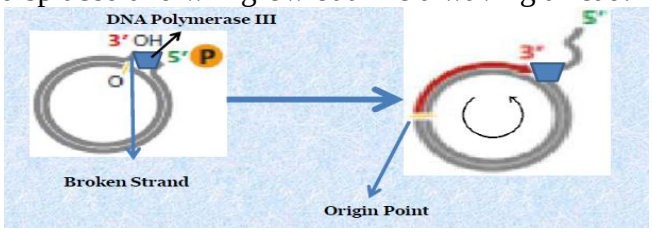
b) Topoisomerases

c) Single stranded binding proteins (SSBPs).



Elongation

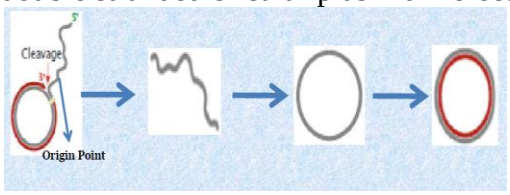
For **Elongation,** the **DNA polymerase III** binds to the 3́-OH group of broken strand, using the unbroken strand as a template. The polymerase will start to move in a circle for elongation, due to which it is named as **Rolling circle model**. As the elongation proceeds, the 5́end will be displaced and will grow out like a waving thread.

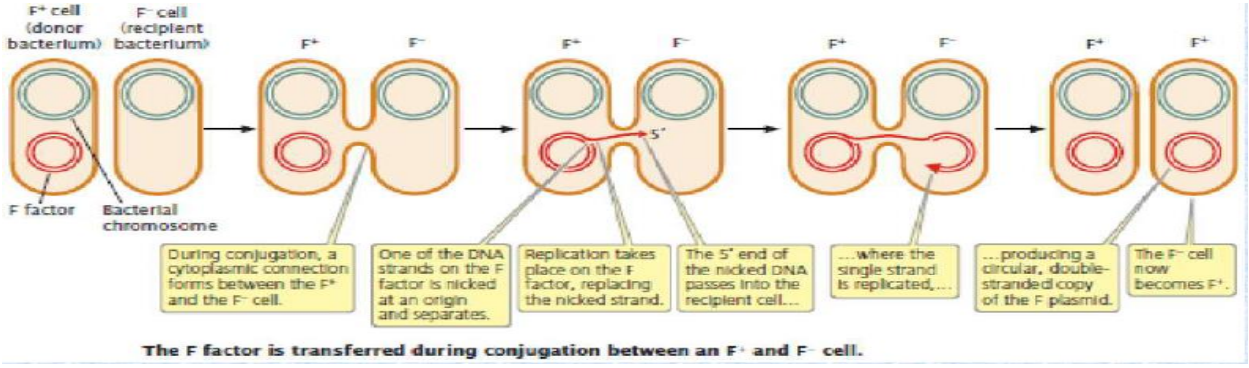


Termination

At the point of termination, the linear **DNA molecule is cleaved from the circle**, resulting in a **double stranded circular DNA molecule** and **a single-stranded linear DNA molecule.**

The linear single stranded molecule is circularized by the action of **ligase** and then replication to double stranded circular plasmid molecule.





DNA REPLICATION

• DNA replication is the process by which DNA makes a copy of itself during cell division.

• In molecular biology, DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule. This process occurs in all living organisms and is the basis for biological inheritance.

• DNA is made up of a double helix of two complementary strands.

STEPS FOR DNA REPLICATION:

INITIATION -

STEP 1 :. In DNA replication, the first step is to '**unzip**' the double helix structure of the DNA molecule.

STEP 2: DNA **gyrase**, often referred to simply as gyrase, is an enzyme that relieves strain while double-strand DNA is being unwound by helicase.

STEP 3: This is carried out by an enzyme called **helicase** which breaks the hydrogen bonds holding the complementary bases of DNA together (A with T, C with G).

STEP 4: The separation of the two single strands of DNA creates a 'Y' shape called a replication 'fork'. The two separated strands will act as templates for making the new strands of DNA.

STEP 5: One of the strands is oriented in the 3' to 5' direction (towards the replication fork), this is the leading strand. The other strand is oriented in the 5' to 3' direction (away from the replication fork), this is the lagging strand.

• As a result of their different orientations, the two strands are replicated differently.

ELONGATION -

Leading Strand:

• A short piece of RNA called **a primer** (produced by an enzyme called primase) comes along and binds to the end of the leading strand. The primer acts as the starting point for **DNA synthesis**.

• **DNA polymerase III** binds to the leading strand and then 'walks' along it, **adding** new complementary Nucleotide bases (A, C, G and T) to the strand of DNA in the 5' to 3' direction.

• This sort of replication is called **continuous.**

Lagging strand:

• **Numerous RNA primers** are made by the primase enzyme and bind at various points along the lagging strand.

• Chunks of DNA, called **Okazaki fragments**, are then added to the lagging strand also in the 5'to 3' direction.

• This type of replication is called **discontinuous** as the Okazaki fragments will need to be joined up later.

Step 9: Once all of the bases are matched up (A with T, C with G), an enzyme called **exonuclease (DNA Polymerase I)** strips away the primer(s). The gaps where the primer(s) were are then filled by yet more complementary nucleotides.

TERMINATION -

Step 10: While removing the primers, another type of exonuclease proofread the new stands, checking, removing, and replacing any errors formed during synthesis.

Step 11: Finally, an enzyme called **DNA ligase** seals up the sequence of DNA into two continuous double strands. The ends of the parent strand consist of a repetition of DNA sequences known as telomeres which act as protective caps at the ends of chromosomes preventing the fusion of nearby chromosomes.

- The telomeres are synthesized by a special type of DNA polymerase enzyme known as telomerase.
- It catalyzes the telomere sequences at the end of the DNA.

Step 12: The result of DNA replication is two DNA molecules consisting of one new and one old chain of nucleotides. This is why DNA replication is described as **semi-conservative,** half of the chain is part of the original DNA molecule, half is brand new.

Step 13: Following replication the new DNA **automatically** winds up into a **double helix.**

DNA replication enzymes and Proteins

DNA polymerase

- DNA polymerases are enzymes used for the synthesis of DNA by adding nucleotide one by one to the growing DNA chain. The enzyme incorporates complementary amino acids to the template strand.
- DNA polymerase is found in both prokaryotic and eukaryotic cells. They both contain several different DNA polymerases responsible for different functions in DNA replication and DNA repair mechanisms.

DNA Helicase enzyme

- This is the enzyme that is involved in unwinding the double-helical structure of DNA allowing DNA replication to commence.
- It uses energy that is released during ATP hydrolysis, to break the hydrogen bond between the DNA bases and separate the strands.
- This forms two replication forks on each separated strand opening up in opposite directions.
- At each replication fork, the parental DNA strand must unwind exposing new sections of single-stranded templates.
- The helicase enzyme accurately unwinds the strands while maintaining the topography on the DNA molecule.

DNA primase enzyme

- This is a type of RNA polymerase enzyme that is used to synthesize or generate RNA primers, which are short RNA molecules that act as templates for the initiation of DNA replication.

DNA ligase enzyme

- This is the enzyme that joins DNA fragments together by forming phosphodiester bonds between nucleotides.

Exonuclease

- These are a group of enzymes that remove nucleotide bases from the end of a DNA chain.

Topoisomerase

- This is the enzyme that solves the problem of the topological stress caused during unwinding.
- They cut one or both strands of the DNA allowing the strand to move around each other to release tension before it rejoins the ends.

- And therefore, the enzyme catalysts the reversible breakage it causes by joining the broken strands.
- Topoisomerase is also known as DNA gyrase in *E. coli.*

Telomerase

- This is an enzyme found in eukaryotic cells that adds a specific sequence of DNA to the telomeres of chromosomes after they divide, stabilizing the chromosomes over time.

E. coli Gene	Enzyme/Protein Function	Description
dnaA	Initiator Protein	Melts DNA at oriC, exposing two template ssDNA strands
dnaB	Helicase	Unwinds the DNA helix at the front end of each replication fork during replication
dnaC	Helicase Loader	Loads the DnaB Helicase onto the ssDNA template strands
dnaG	Primase	Synthesizes RNA primers used to initiate DNA synthesis
dnaE	α-Catalytic Subunit of DNA Polymerase III	Catalytic subunit of the main replicative polymerase during DNA replication
dnaQ	ε-Editing Subunit of DNA Polymerase III	Editing subunit of the main replicative polymerase during DNA replication
dnaN	β-clamp subunit of DNA Polymerase III	Clamping subunit of the main replicative polymerase during DNA replication
polA	DNA Polymerase I	Processes Okazaki fragments and also fills in gaps during DNA repair processes
polB	DNA Polymerase II	Proofreading and editing, especially on lagging strand synthesis and some involvement in DNA repair
ssb	Single Stranded Binding Proteins (SSB)	Bind with single-stranded regions of DNA in the replication fork and prevent the strands from rejoining
A dimer encoded by *gyrA* and *gyrB*	DNA Gyrase	Type II Topoisomerase involved in releiving positive supercoiling tension caused by the action of Helicase
A dimer encoded by *parC* and *parE*	Topoisomerase IV	Type II Topoisomerase involved in decatenation of daughter chromosomes during DNA replication
ligA	DNA Ligase	Fixes nicks in the DNA backbone during DNA replication, DNA damage, and DNA repair processes
Note: Only the genes involved in the formation of the catalytic domain of DNA polymerase III are listed		

UNIT 3 – Transcription

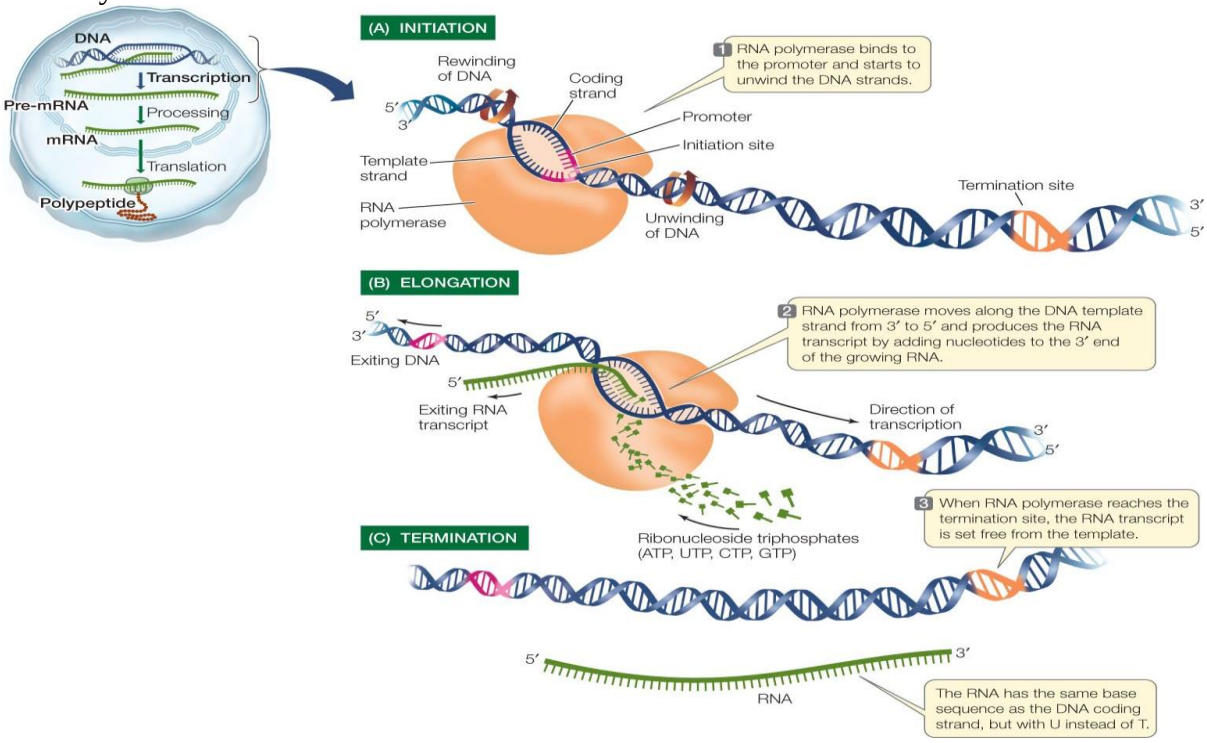
Prokaryotic Transcription

- Transcription is the process by which the information in a strand of **DNA** is copied into a new molecule of messenger **RNA** (mRNA).
- In prokaryotic organisms transcription occurs in three phases known as initiation, elongation and termination.

Enzyme(s) Involved

- RNA is synthesized by a single RNA polymerase enzyme which contains multiple polypeptide subunits.
- In *E. coli*, the RNA polymerase has subunits: two α, one β, one β' and one ω and σ subunit (α2ββ'ωσ). This complete enzyme is called as the holoenzyme.

- The σ subunit may dissociate from the other subunits to leave a form known as the core enzyme.



Initiation Phase

During initiation, RNA polymerase recognizes a specific site on the DNA, upstream from the gene that will be transcribed, called a **promoter site** and then unwinds the DNA locally.

Promoters and Initiation

- The holoenzyme binds to a promoter region about 40–60 bp in size and then initiates transcription a short distance downstream (i.e. 3 to the promoter).
- Within the promoter lie two 6 base pair sequences that are particularly important for promoter function.
- They are highly conserved between species.
- Using the convention of calling the first nucleotide of a transcribed sequence as +1, these two promoter elements lie at positions –10 and –35, that is about 10 and 35 bp, respectively, upstream of where transcription will begin.
- The –10 sequence has the consensus because this element was discovered by Pribnow, it is also known as the Pribnow box. It is an important recognition site that interacts with the σ factor of RNA polymerase.
- The –35 sequence has the consensus **TTGACA** and is important in DNA unwinding during transcriptional initiation.

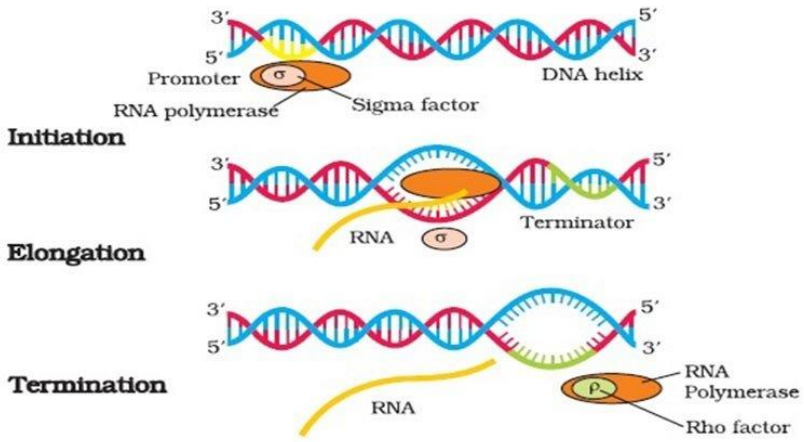
- RNA polymerase does not need a primer to begin transcription; having bound to the promoter site, the RNA polymerase begins transcription directly.

Elongation Phase

- After transcription initiation, the σ factor is released from the transcriptional complex to leave the core enzyme (α2 ββω) which continues elongation of the RNA transcript.
- The core enzyme contains the catalytic site for polymerization, probably within the β subunit.
- The first nucleotide in the RNA transcript is usually pppG or pppA.
- The RNA polymerase then synthesizes RNA in the 5' →3' direction, using the four ribonucleoside 5-triphosphates (ATP, CTP, GTP, UTP) as precursors.
- The 3-OH at the end of the growing RNA chain attacks the α phosphate group of the incoming ribonucleoside 5-triphosphate to form a 3'5′ phosphodiester bond.
- The complex of RNA polymerase, DNA template and new RNA transcript is called a **ternary complex** (i.e. three components) and the region of unwound DNA that is undergoing transcription is called the transcription bubble.
- The RNA transcript forms a transient RNA–DNA hybrid helix with its template strand but then peels away from the DNA as transcription proceeds.
- The DNA is unwound ahead of the transcription bubble and after the transcription complex has passed, the DNA rewinds.
- Thus, during the elongation, the RNA polymerase uses the antisense (-) strand of DNA as template and synthesizes a complementary RNA molecule.
- The RNA produced has the same sequence as the non-template strand, called the sense (+) strand (or coding strand) except that the RNA contains U instead of T.
- At different locations on the bacterial chromosome, sometimes one strand is used as template, sometimes the other, depending on which strand is the coding strand for the gene in question.
- The correct strand to be used as template is identified for the RNA polymerase by the presence of the promoter site.

Termination Phase

- Transcription continues until a termination sequence is reached.
- The most common termination signal is a GC-rich region that is a palindrome, followed by an AT-rich sequence.
- The RNA made from the DNA palindrome is self- complementary and so base pairs internally to form a hairpin structure rich in GC base pairs followed by four or more U residues.
- However, not all termination sites have this hairpin structure. Those that lack such a structure require an additional protein, called **rho**, to help recognize the termination site and stop transcription.
- Thus the RNA polymerase encounters a termination signal and ceases transcription, releasing the RNA transcript and dissociating from the DNA.



RNA processing

- In prokaryotes, RNA transcribed from protein-coding genes (messenger RNA, mRNA), requires little or no modification prior to translation.
- Many mRNA molecules begin to be translated even before RNA synthesis has finished.
- However, since ribosomal RNA (rRNA) and transfer RNA (tRNA) are synthesized as precursor molecules, they require post-transcriptional processing.

Significance

- Transcription of DNA is the method for regulating gene expression.
- It occurs in preparation for and is necessary for protein translation.

RNA polymerase definition

Ribonucleic Acid (RNA) Polymerase (RNAP) enzyme is a multi-subunit enzyme that applies its activity in the catalyzation of the transcription process of RNA synthesized from a DNA template.

- And therefore, RNA polymerase enzyme is responsible for the copying of DNA sequences into RNA sequences during transcription.
- The function of RNA polymerase is to control the process of transcription, through which copying of information stored in DNA into a new molecule of messenger RNA (mRNA.)
- During transcription, the RNA polymer is contemporary to the template DNA that is synthesized in the direction of 5′ to 3′.
- The enzyme RNA polymerase interacts with proteins to enable it to function in catalyzation of the synthesis of RNA.
- The collaborator proteins assist in enabling the specific binding of RNA polymerase, assist in the unwinding of the double chemical structure of DNA, moderate the enzymatic activities of RNA polymerase and to control the speed of transcription.
- The RNA polymerase enzyme has an interrupted mechanism whereby it continuously synthesizes RNA polymers of over four thousand bases per minute but they pause or stop occasionally to maintain fidelity.
- RNA polymerase is an enzyme that is responsible for copying a DNA sequence into an RNA sequence, during the process of transcription. As a complex molecule composed of protein

subunits, RNA polymerase controls the process of transcription, during which the information stored in a molecule of DNA is copied into a new molecule of messenger RNA.

- RNA polymerases have been found in all species, but the number and composition of these proteins vary across taxa.
- For instance, bacteria contain a single type of RNA polymerase, while eukaryotes (multicellular organisms and yeasts) contain three distinct types.
- In spite of these differences, there are striking similarities among transcriptional mechanisms.
- For example, all species require a mechanism by which transcription can be regulated in order to achieve spatial and temporal changes in gene expression.

Types of RNA polymerase

Prokaryotic (Bacteria, viruses, archaea) organisms have a single type of RNA polymerase that synthesizes all the subtypes of RNA, while eukaryotes (multicellular organisms) have 5 different types of RNA polymerases which perform different functions in the synthesis of different RNA molecules.

Prokaryotic RNA polymerase

- The prokaryotes have a single type of RNA polymerase (RNAP) which synthesizes all the classes of RNA, i.e mRNA, tRNA, rRNA, sRNA.
- The RNA Polymerase molecule is made up of 2 domains and 5 subunits:

i. Core and holoenzyme
ii. Subunits (β, β', α (αI and αII), ω,)

- The promoter is the sequence of DNA that is required for accurate and specific initiation of transcription, and also, it is the sequence of DNA to which RNA polymerase binds accurately to initiate transcription.
- The 'a' subunit is made up of two distinct domains. The N-terminal domain (a-NTD) and the C-terminal.
- The N-terminal is involved in dimerization forming a2 and further assembly of the RNA polymerase.
- The C-terminal domain functions such as binding to the Upstream Promoter (UP) DNA sequence at promoters for rRNA and tRNA genes and in communication with several transcriptional activators.

Sigma Factor: Sigma factors are **subunits of all bacterial RNA polymerases**. They are responsible for determining the specificity of promoter DNA binding and control how efficiently RNA synthesis (transcription) is initiated. It was found that RNA polymerase activity was associated with two protein species.

Role of the sigma subunit of RNA polymerase:

The additional sigma (σ) subunit (factor), which binds to the RNAP core to form the RNAP holoenzyme, is **responsible for promoter recognition and subsequent transcription initiation**. Thus, σ factors are the regulators that are necessary for the transcription initiation of each bacterial gene.

When the sigma subunit, σ^{70}, is added, it can bind to **core forming a holoenzyme (α2ββ′σ)** that is capable of specific engagement with duplex DNA at the beginning of genes (promoters) as well as efficient initiation of transcription.

- Each of the subunit structure is as follows:

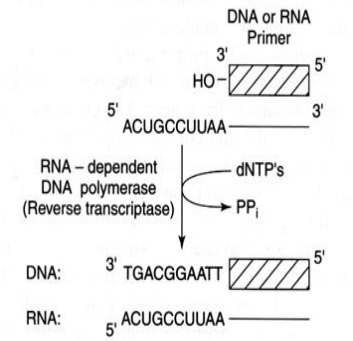
Prokaryotic RNA polymerase Subunits

Subunit	Size	Function
β	150.4 kDa	The β' + β form the catalytic center, responsible for RNA synthesis.
β'	155.0 kDa	The β' + β form the catalytic center, responsible for RNA synthesis.
α (αI and αII)	36.5 kDa	It is made up of the enzyme assembly, and it also binds the UP sequence in the promoter.
ω	155.0 kDa	It confers specificity for promoter; and binds to -10 and -35 sites in the promoter.

Functions of RNA Polymerase

- Generally, the RNA molecule is a messenger molecule that is used to export information that is coded in DNA out of the cell nucleus, to synthesize proteins in the cell cytoplasm.
- RNA polymerase is used in the production of molecules that play a wide range of roles, of which one of its functions is to regulate the number and type of RNA transcript that is formed in response to the requirements of the cell.
- The RNA polymerase enzyme interacts with different molecular proteins, transcription factors, and signaling molecules on the carboxyl-terminal, which regulates its mechanisms, which play a major role in gene expression and gene specialization in multicellular (eukaryotic) organisms.
- The RNA enzyme also ensures irregularities and errors during the conversion of DNA to RNA (transcription). Such as ensuring that the right nucleotide is added to the newly synthesized RNA strand, inserting the right amino acid-base which is complementary to the template of the DNA strand.
- When the right nucleotides have been inserted, the RNA polymerase can then catalyze and elongate the RNA strand, at the same time, proofread the new strand and remove incorrect bases.
- RNA polymerase is also involved in the post-transcription modification of RNAs, converting them into functional molecules that facilitate the transportation of molecules from the nucleus to their site of action.
- Besides its role in the synthesis of proteins, RNA performs other functions such as
- Protein coding
- Regulation of gene expression
- Act as enzymes
- Formation of gametes by the non-coding RNA (ncRNA)
- Production of regulatory molecules.

Reverse Transcription: **Reverse transcription** (which occurs in both prokaryotes and eukaryotes) is the synthesis of DNA from an RNA template. A class of RNA viruses, called **retroviruses**, are characterized by the presence of an RNA-dependent DNA polymerase (reverse transcriptase). The virus that causes AIDS, Human Immunodeficiency Virus (HIV), is a retrovirus. Because nuclear cell division doesn't use reverse transcriptase, the most effective anti-HIV drugs target reverse transcriptase, either its synthesis or its activity.



Like other DNA polymerases, reverse transcriptases are primer- and template-dependent. They also possess an RNase H activity (H stands for hybrid) that can degrade the RNA template after it is used for synthesis of the first DNA strand. The enzyme then can copy the first strand of DNA to make a double-stranded molecule.

Reverse transcription is error prone relative to DNA replication because reverse transcriptases don't have an editing (3′-5′) exonucleolytic activity. This has one important consequence for HIV treatment and epidemiology. HIV mutates very rapidly. In advanced AIDS patients, the virus that is isolated from the bloodstream often bears very little resemblance to the original infecting strain. This rapid variation means that drug-resistant mutant strains of the virus arise frequently, and drug treatment doesn't work well. Secondly, the rapid mutation rate complicates vaccine development—new strains that are not neutralized by the vaccine can appear and infect individuals that were vaccinated against the original strain.

UNIT 4 – Gene Expression and Regulation

Gene Regulation – Lac Operon

- E.coli can use lactose as a source of carbon.
- The enzymes required for the use of lactose as a source of carbon are synthesised only when the lactose is available as carbon source.
- The lac operon is an example of negatively controlled inducible operon.

Structure

The lac operon consists of 5 structural units.

1) Promoter

2) Operator
3) Structural genes
4) CAP binding sites 5) Regulatory gene

1) promoter

 It is the RNA polymerase binding site.
 As a result transcription starts.

2) operator

 It is the regulator binding site.
 The regulator proteins activator or repressor attaches the operator.

3) structural genes

 There are 3 structural genes which are placed adjacent to one another. They are

i) Lac z- which codes for beta galactosidase, an enzyme responsible for the hydrolysis of lactose to galactose and glucose.

ii) Lac y- which codes for galactoside permease which is responsible for the transport of lactose across bacterial cell wall.

iii) Lac a- which codes for thiogalactoside transacetylase whose physiological role is unclear.

 The structural genes z,y,a transcribe to form a single large mRNA with three independent translation units for the synthesis of three distinct enzymes.

 Such a mRNA coding for more than one protein is called "polycistronic mRNA" which is a characteristics of prokaryotes.

4) CAP binding site

 CAP binding site is a positive regulatory site that is bound by catabolic activator protein(CAP).

 CAP act as a glucose sensor. When glucose levels are low, it activates transcription of the operon by the binding of RNA polymerase to the promoter.

 CAP is not always active. So it is regulated by a small molecule called **cyclic AMP**(cAMP).

 cAMP is a " hunger signal" made by E.coli when glucose levels are low.

5) Regulatory gene

 The regulatory gene is the igene that code for the repressor protein.

 The igene is expressed in all the time, hence it is also known as *constitutive gene*.

Gene expression of a Lac operon

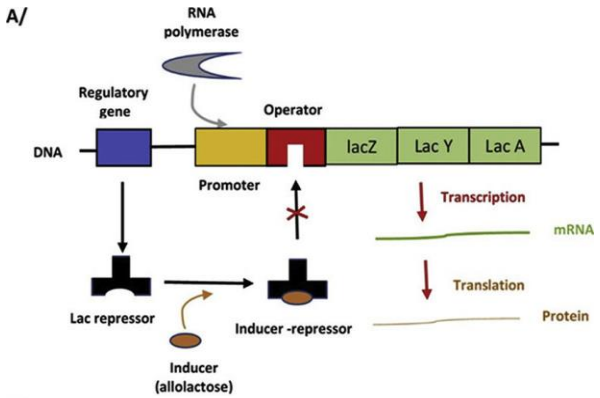
1) Negative control of lac operon

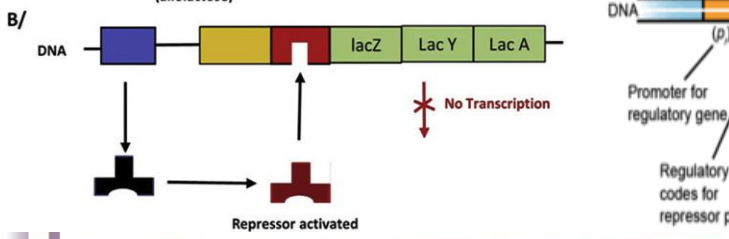
• The regulatory gene (igene) secrets an active repressor called lac repressor.
• It attaches to the promoter site. So the RNA polymerase does not attach the operator.
• Ie. the reaction is turn off(it will not produce any kind of protein or enzymes)
• Now the inducer (**lactose**) attaches to the active repressor. As a result active repressor become inactive repressor.
• So the inactive repressor is not capable of attaching promoter site.
• As a result RNA polymerase attaches the operator site and transcription begins.

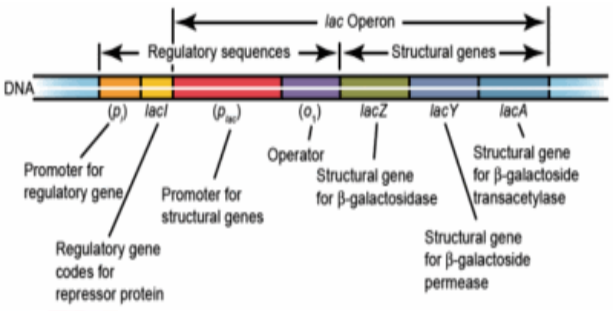
2) Positive control of lac operon

 If glucose and laactose are both present in the medium, cells will use up glucose first.

 The presence of glucose in the cell switch off the lac operon by a mechanism called catabolic repression, which involves a protein called catabolite activator protein(CAP).



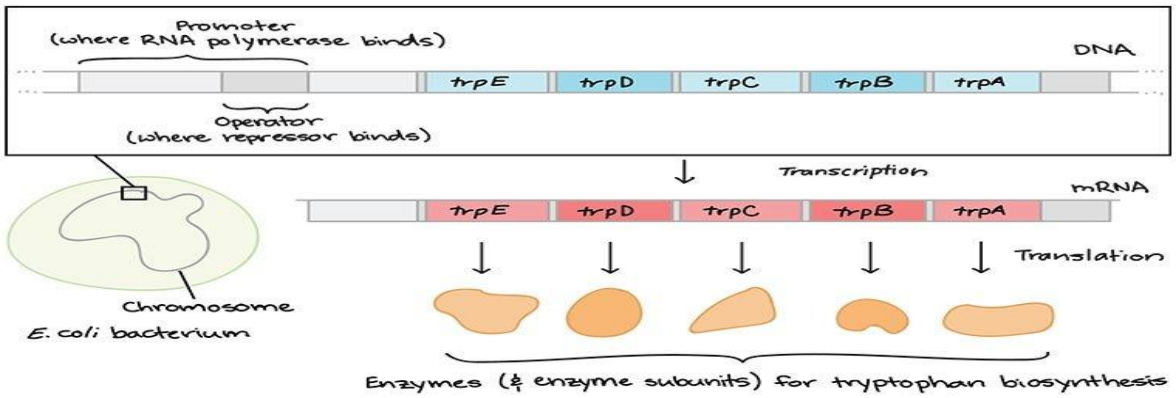




Summary

Carbohydrates	Activator protein	Repressor protein	RNA polymerase	*lac* Operon
+ GLUCOSE + LACTOSE	Not bound to DNA	**Lifted off operator site**	Keeps falling off promoter site	**No transcription**
+ GLUCOSE - LACTOSE	Not bound to DNA	Bound to operator site	Blocked by the repressor	**No transcription**
- GLUCOSE - LACTOSE	**Bound to DNA**	Bound to operator site	Blocked by the repressor	**No transcription**
- GLUCOSE + LACTOSE	**Bound to DNA**	**Lifted off operator site**	**Sits on the promoter site**	**Transcription**

Trp Operon

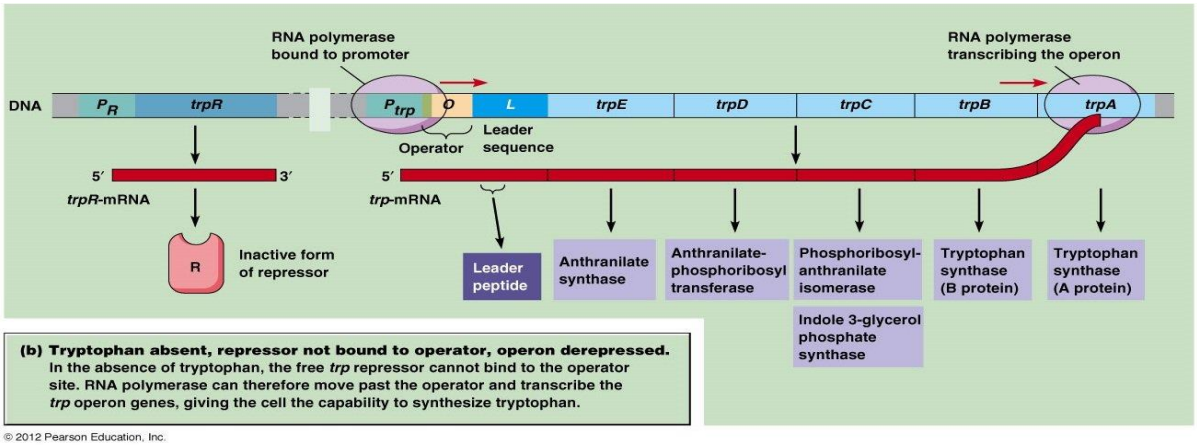


- The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan.
- The tryptophan (trp) operon contains five structural genes encoding enzymes for tryptophan biosynthesis with an upstream trp promoter (Ptrp) and trp operator sequence (Otrp).
- Structural genes are TrpE, TrpD, TrpC, TrpB and TrpA

1. trpE: It enodes the enzyme Anthranilate synthase I
2. trpD: It encodes the enzyme Anthranilate synthase II
3. trpC: It encodes the enzyme N-5'-Phosphoribosyl anthranilate isomerase and Indole-3-glycerolphosphate synthase
4. trpB: It encodes the enzyme tryptophan synthase-B sub unit
5. trpA: It encode the enzyme tryptophan synthase-A sub unit

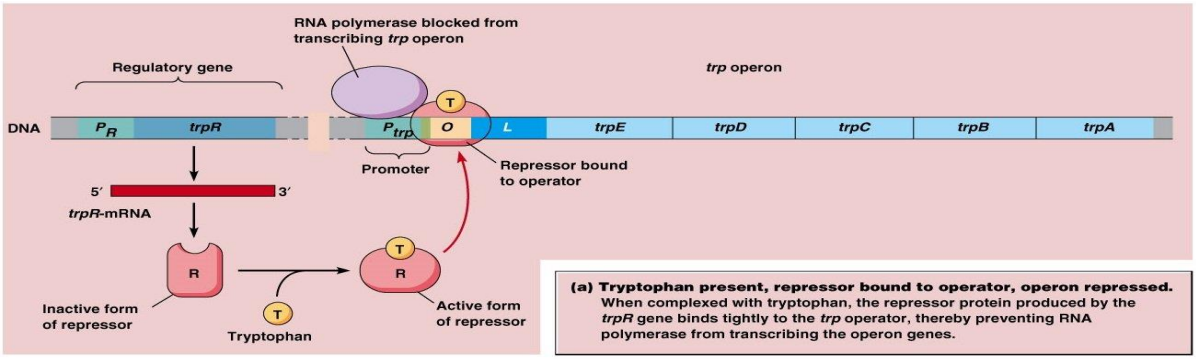
- The trp operator region partly overlaps the trp promoter.
- The operon is regulated such that transcription occurs when tryptophan in the cell is in short supply.

In the Absence of Tryptophan



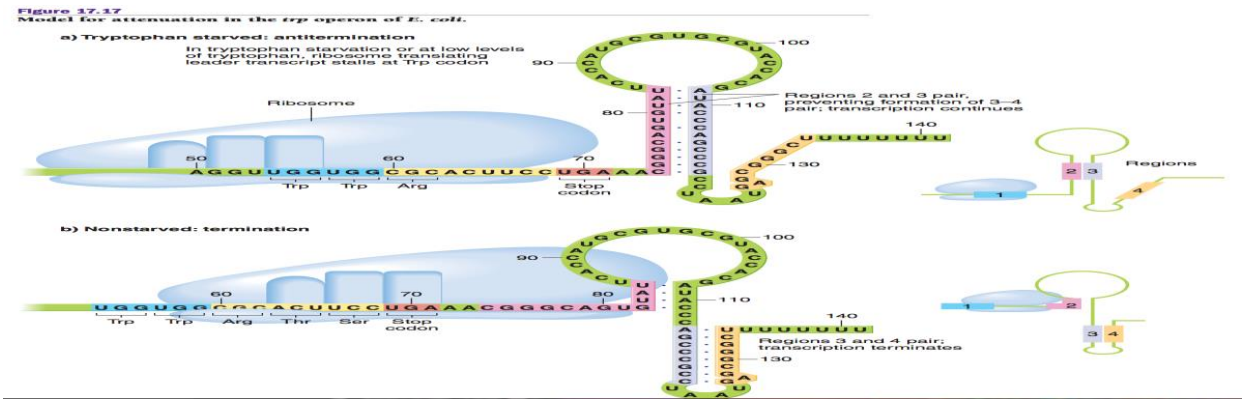
- In the absence of tryptophan, a trp repressor protein encoded by a separate operon, trpR, is synthesized and forms a dimer.
- However, this is inactive and so is unable to bind to the trp operator and the structural genes of the trp operon are transcribed.

In the Presence of Tryptophan



- When tryptophan is present, the enzymes for tryptophan biosynthesis are not needed and so expression of these genes is turned off.
- This is achieved by tryptophan binding to the repressor to activate it so that it now binds to the operator and stops transcription of the structural genes.
- Binding of repressor protein to operator overlaps the promoter, so RNA polymerase cannot bind to the prometer. Hence transcription is halted.
- In this role, tryptophan is said to be a co-repressor. This is negative control, because the bound repressor prevents transcription.

Trp Operon Attenuation



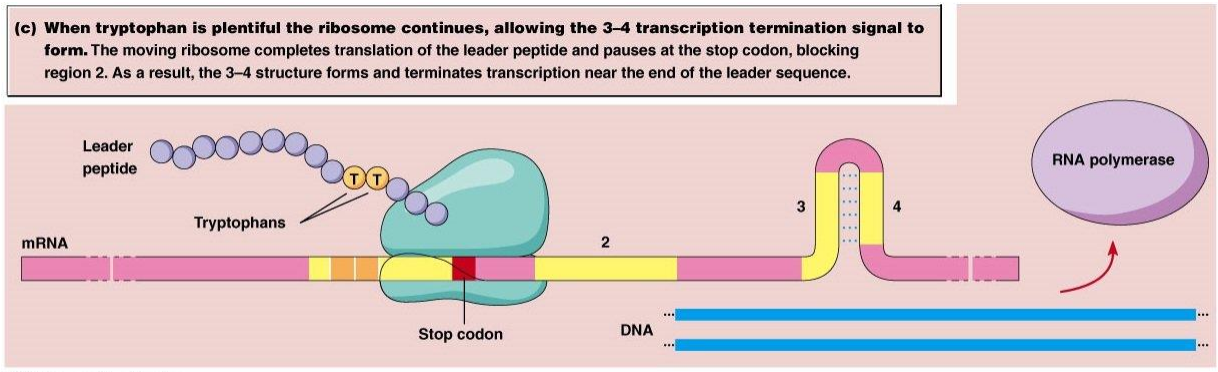
- A second mechanism, called attenuation, is also used to control expression of the trp operon.
- The 5′ end of the polycistronic mRNA transcribed from the trp operon has a leader sequence upstream of the coding region of the trpE structural gene.
- This **leader sequence** encodes a 14 amino acid leader peptide containing two tryptophan residues.
- The function of the leader sequence is to fine tune expression of the trp operon based on the availability of tryptophan inside the cell.

The leader sequence contains four regions (numbered 1–4) that can form a variety of base paired stem-loop (‘hairpin’) secondary structures.

- The regions are: Region 1, region 2, region 3 and Region 4. Region 3 is complementary to both region 2 and region 4.
- If region 3 and region 4 base pair with each other, they form a loop like structure called attenuator and it function as transcriptional termination. If pairing occur between region 3 and region 2, then no such attenuator form so that transcription continues.

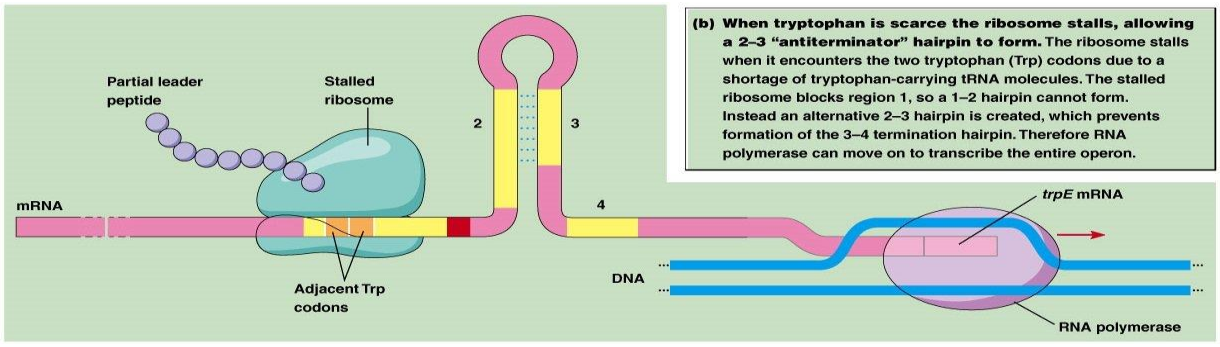
Attenuation depends on the fact that, in bacteria, ribosomes attach to mRNA as it is being synthesized and so translation starts even before transcription of the whole mRNA is complete.

When Trypophan is abundant



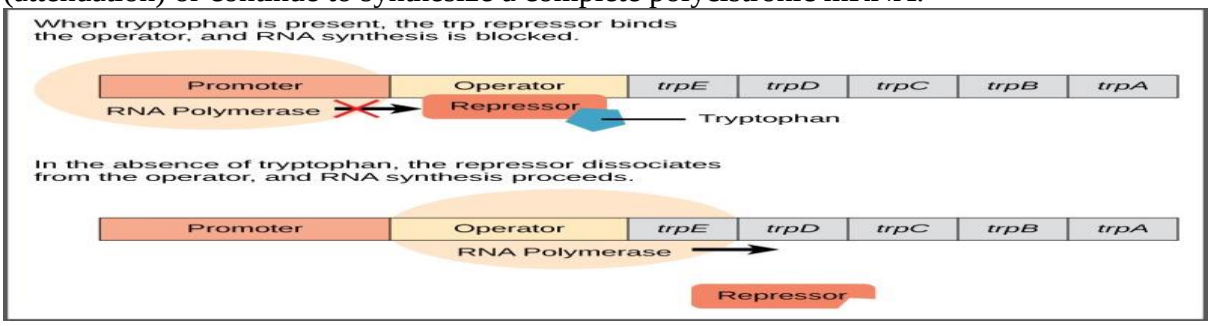
- When tryptophan is abundant, ribosomes bind to the trp polycistronic mRNA that is being transcribed and begin to translate the leader sequence.
- Now, the two trp codons for the leader peptide lie within sequence 1, and the translational Stop codon lies between sequence 1 and 2.
- During translation, the ribosomes follow very closely behind the RNA polymerase and synthesize the leader peptide, with translation stopping eventually between sequences 1 and 2.
- At this point, the position of the ribosome prevents sequence 2 from interacting with sequence 3.
- Instead sequence 3 base pairs with sequence 4 to form a 3:4 stem loop which acts as a transcription terminator.
- Therefore, when tryptophan is present, further transcription of the trp operon is prevented.

When Trypophan is scarce



- If, however, tryptophan is in short supply, the ribosome will pause at the two trp codons contained within sequence 1.
- This leaves sequence 2 free to base pair with sequence 3 to form a 2:3 structure (also called the anti-terminator), so the 3:4 structure cannot form and transcription continues to the end of the trp operon.

Hence the availability of tryptophan controls whether transcription of this operon will stop early (attenuation) or continue to synthesize a complete polycistronic mRNA.



Regulation of Trp Operon

Overall, for the trp operon, repression via the trp repressor determines whether transcription will occur or not and attenuation then fine tunes transcription.

Monocistronic and Polycistronic mRNA

The **main difference** between monocistronic and polycistronic mRNA is that the **monocistronic mRNA produces a single protein while polycistronic mRNA produces several proteins that are functionally-related.** Furthermore, eukaryotes have monocistronic mRNA while prokaryotes have polycistronic mRNA.

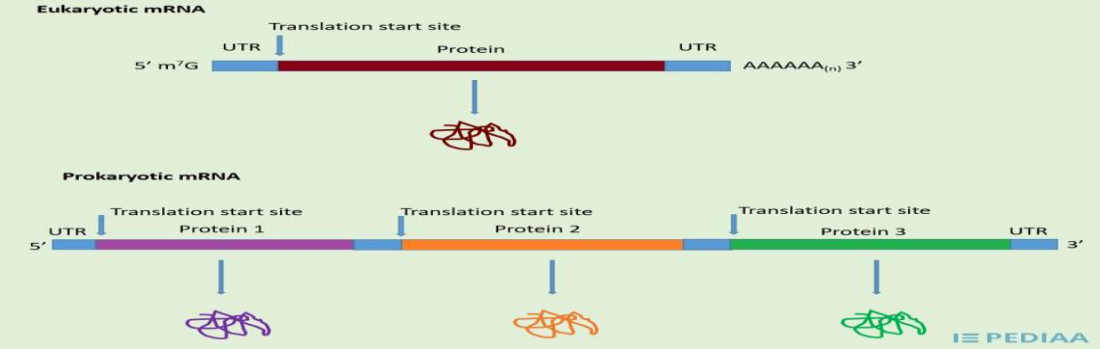
Monocistronic and polycistronic mRNA are two types of mRNA molecules, which can be decoded into polypeptide sequences. Most mRNA are monocistronic while fewer are polycistronic.

Similarities Between Monocistronic and Polycistronic mRNA

- Monocistronic and polycistronic mRNA are two types of mRNA that can be translated into proteins.
- Untranslated regions flank the protein coding region in both mRNA.
- Both are transcribed under a single promoter.

What is Monocistronic mRNA

Monocistronic mRNA refers to eukaryotic mRNA that consists of a single cistron. Therefore, it can produce a single protein. The nascent transcripts of genes are called the pre-mRNA. Pre-mRNA and other nuclear RNA are collectively called heterogeneous nuclear RNA (hnRNA). They associate with proteins to form heterogeneous nuclear ribonucleoprotein (hnRNP) inside the nucleus. Some hnRNA undergoes post-transcriptional modifications in order to become mature RNA, which travels to the cytoplasm for translation.



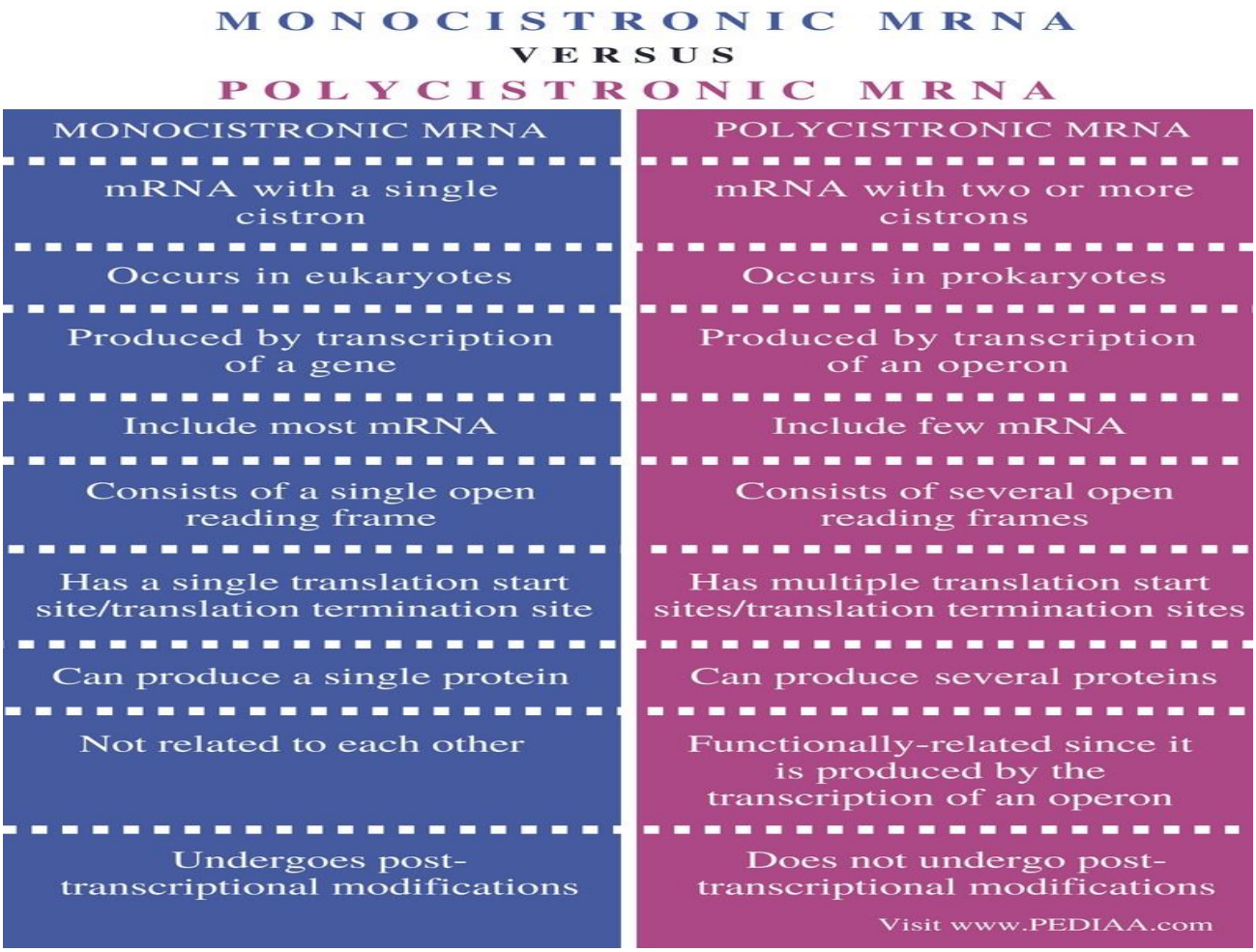
The **addition of the 5' cap** is the initial stage of post-transcriptional modifications. It is catalyzed by a dimeric capping enzyme associated with the CTD (phosphorylated carboxyl-terminal tail domain) of RNA polymerase II. Another post-transcriptional modification is **3' polyadenylation**, which covalently adds a polyadenylyl moiety at the 3' end. Poly-A tail promotes the export of the mRNA molecule into the cytoplasm while protecting it from degradation. **RNA splicing** is the other event in the post-transcriptional modifications. During this, the introns are removed and exons are spliced together.

What is Polycistronic mRNA

Polycistronic mRNA refers to prokaryotic mRNA consisting of two or more cistrons. The mRNA produced by chloroplasts and mitochondria are also polycistronic. In prokaryotes, functionally-related genes assemble in groups in such a way that all proteins can be transcribed at once when needed. *Lac* operon is one such famous operon. *lacZ*, *lacY*, and *lacA* are the three genes in the *Lac* operon, which encode the beta-galactosidase, beta-galactoside permease, and beta-galactoside transacetylase enzymes respectively. All enzymes are involved in the lactose metabolism.

Lac Operon

An operon that produces polycistronic mRNA consists of a leader and a trailer sequence. The leader sequence is the first gene followed by an intercistronic region and the sequence of the second gene. The trailer sequence is the last gene. Transcription of the operon is regulated by a single promoter. However, each cistron is flanked by transcription initiation site and transcription termination site.



UNIT 5 – Genetic code and protein synthesis

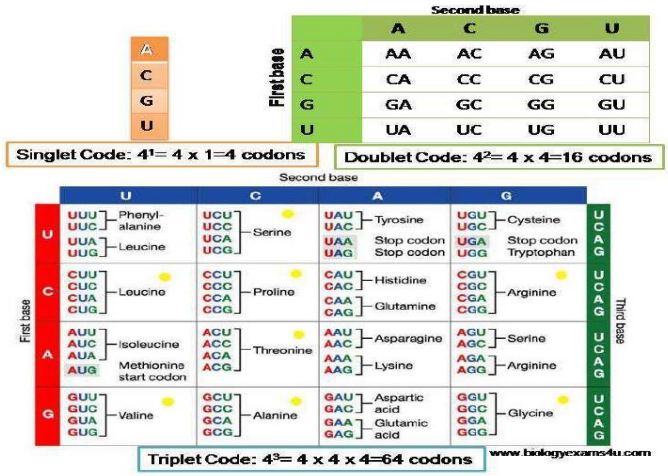
Genetic Code

- The sequence of nitrogenous bases in mRNA molecule which encloses information for the synthesis of protein molecules is known as Genetic code.
- The nucleotide or nucleotide sequence in mRNA which codes for a particular amino acid is known as Codon.
- The main problem of genetic code was to determine the exact number of nucleotides in a codon which codes for one amino acid.

A single code consisting of only one nucleotide provides for just four codons A, C, G and U. These are insufficient to code for 20 amino acids.

Similarly a combination of two nitrogenous bases (double code) provides $4^{x}4 = 16$ codons still insufficient for 20 amino acids.

George Gamow (1954) pointed out the possibility of three letter code, i.e., each codon consisting of three nitrogenous bases. This will give $4^{x}4^{x}4 = 64$ codons, which are more than enough to code for 20 amino acids.



The triple code is the most acceptable one as it has found support from experimental evidences of both genetic and biochemical nature.

Properties of genetic code

The genetic code possesses the following important properties which have now been proved by definite experimental evidences.

1) The code is triplet
2) The code is degenerate
3) The code is non-overlapping
4) The code is commaless
5) The code is non-ambiguous
6) The code is universal
7) The code is polarity
8) The codes as act as start code 9) The codes as act as stop code

1) The code is triplet:

- A triplet or three letter code was first suggested by a physicist Gamow in 1954.
- A codon of the present day genetic code that specifies one amino acid in a polypeptide chain comprises of a sequences of three nitrogenous bases on mRNA in a specific sequence.
- The first experimental evidence supporting the concept of a triplet code was provided by Crick and Co-workers (1961) in T4 bacteriophage.

2) The code is degenerate:

- When a particular amino acid is coded by more than one codon, it is called degenerate. e.g., except for *tryptophan* and *methionine*, which have a single codon each, all other 18 amino acids have more than one codon.
- The code degeneracy is basically of two types:
    - Partial degeneracy occurs when first two nucleotides are identical but the third nucleotide of the degenerate codons differs, e.g., CUU and CUC code for leucine.
    - Complete degeneracy occurs when any of the four bases can take third position and still code for same amino acid, e.g., UCU, UCC, UCA and UCG code for serine.

3) The code is non-overlapping:

- The code is non-overlapping which means that the same latter is not used for different codons.

- The evidencies are available to show that translation of genetic code in an mRNA begins at the correct point and there is no overlapping of codons. i.e., genetic code is non overlapping.
- Although the code is non overlapping but in the bacteriophage ϕ x 174 there is a possibility of overlapping genes and codons.

4) The code is comma less:

- There are no punctuations or comma etc. between two codons, In the genetic code.
- In other words no codon is reserved for punctuations.

5) The code is non ambiguous:

- The genetic code inside the cell medium (*in vivo*) is said to be non abmiguos because a particular codon always codes for the same amino acid may be coded by more than one codons (degeneracy), but one codon never codes for two different amino acids.

6) The code is universal:

- Same genetic code is found valid for all organisms ranging from bacteria to man.
- Such universality of the code was demonstrated by Marshall, Caskey and Nirenberg (1967) who found that E.coli (bacterium) and guinea pig (mammal) amino acyl – tRNA use almost the same code.

7) The code has polarity:

- The code is always read in a fixed direction, i.e., in the 5' 3' direction. In other words, the Codon has a polarity.
- Reading from left to right and right to left will specify for different amino acids.

8) Some codes act as start codons:

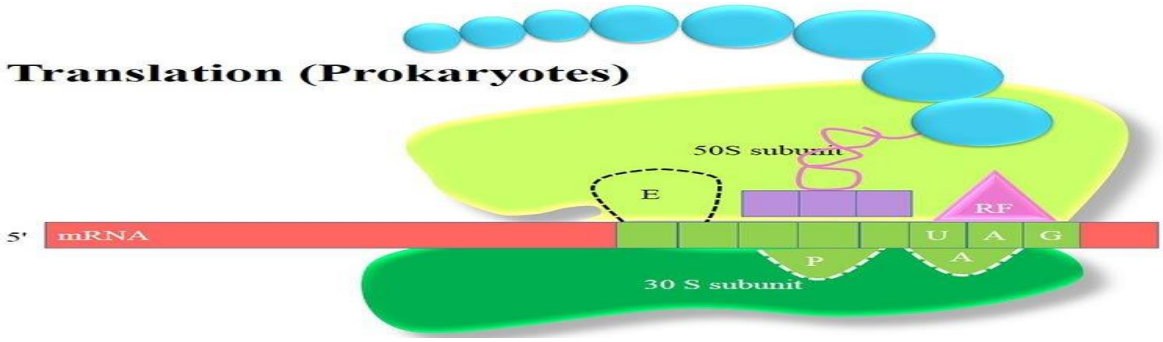
- In most organisms, AUG codon is the start or initiation codons the polypeptide chain starts either with methionine or formaylmethionine.
- Normally GUG codes for valine but when normal AUG codon is lost by deletion only than GUG is use as initiation codon.

9) Some codon act as stop codons:

- Three codons UAG, UAA, and UGA are the chain stop or termination codon.
- They do not code for any of the amino acids.
- The UAA is known as ochre, UAG is known as amber and UGA is known as opal.
- Signals of stop or termination codons are read by proteins called release factors.

Protein Synthesis – Translation in Prokaryotes

- Translation involves translating the sequence of a messenger **RNA** (mRNA) molecule to a sequence of amino acids during protein synthesis.
- It is the process in which ribosomes in the cytoplasm or ER synthesize proteins after the process of transcription of **DNA** to RNA.



The Ribosomes

- Ribosomes exist normally as separate subunits that are composed of protein and rRNA.
- The subunits come together to form a ribosome when they bind to an mRNA, near its 5' end.
- On binding to an mRNA, the ribosome reads the nucleotide sequence from the 5' to 3' direction, synthesizing the corresponding protein from amino acids in an N-terminal (amino-terminal) to C-terminal (carboxyl terminal) direction.
- Ribosomes are located in the cytosol, either freely floating or associated with the endoplasmic reticulum.
- They serve to synthesize proteins.

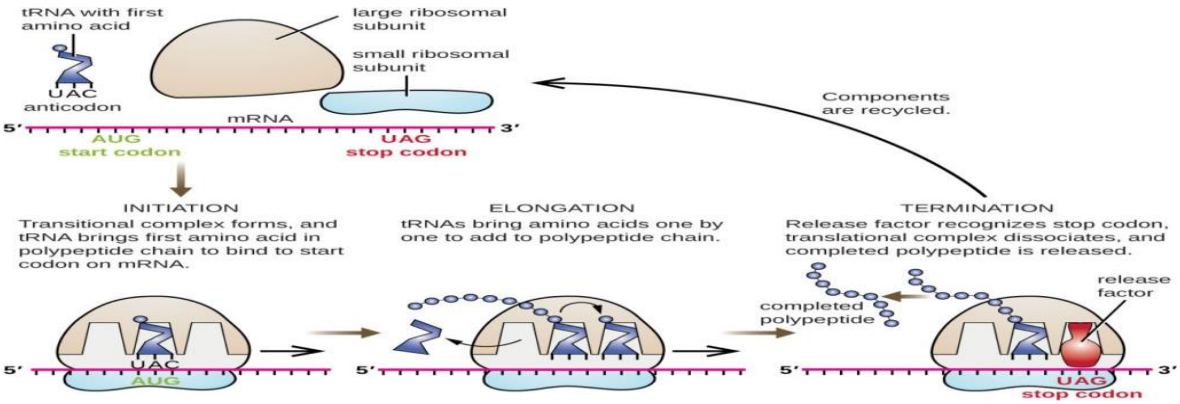
Ribosomal Sites for Protein Translation

Each prokaryotic ribosome, shown schematically, has three binding sites for tRNAs.

1. **The aminoacyl-tRNA binding site** (or A site) is where, during elongation, the incoming aminoacyl-tRNA binds.
2. **The peptidyl-tRNA binding site** (or P site) is where the tRNA linked to the growing polypeptide chain is bound.
3. **The exit site** (or E site) is a binding site for tRNA following its role in translation and prior to its release from the ribosome.

All three sites (A, P and E) are formed by the rRNA molecules in the ribosome.

THE PROCESS OF TRANSLATION



Protein synthesis (or translation) takes place in three stages:

1. Initiation
2. Elongation and
3. Termination.

- During initiation, the mRNA–ribosome complex is formed and the first codon (always AUG) binds the first aminoacyltRNA (called initiator tRNA).
- During the elongation phase, the other codons are read sequentially and the polypeptide grows by addition of amino acids to its C-terminal end.
- This process continues until a termination codon (Stop codon), which does not have a corresponding aminoacyl-tRNA with which to base pair, is reached.
- At this point, protein synthesis ceases (termination phase) and the finished polypeptide is released from the ribosome.

Synthesis of aminoacyl-tRNA

- Synthesis of aminoacyl-tRNAs is crucially important for two reasons:

1. Each amino acid must be covalently linked to a tRNA molecule in order to take part in protein synthesis, which depends upon the 'adaptor' function of tRNA to ensure that the correct amino acids are incorporated.
2. The covalent bond that is formed between the amino acid and the tRNA is a high energy bond that enables the amino acid to react with the end of the growing polypeptide chain to form a new peptide bond.

For this reason, the synthesis of aminoacyl-tRNA is also referred to as **amino acid activation.**

- Each tRNA molecule has a cloverleaf secondary structure with the anticodon accessible at the end of the anticodon stem loop.
- During synthesis of the aminoacyl-tRNA, the amino acid is covalently bound to the A residue of the CCA sequence at the 3' end.
- Each tRNA molecule carries only a single amino acid.
- The attachment of an amino acid to a tRNA is catalyzed by an enzyme called **aminoacyl-tRNA synthetase.**
- A separate aminoacyl-tRNA synthetase exists for every amino acid, making 20 synthetases in total.

The synthesis reaction occurs in two steps.

1. The first step is the reaction of an amino acid and ATP to form an aminoacyl-adenylate (also known as aminoacyl-AMP).
2. In the second step, without leaving the enzyme, the aminoacyl group of aminoacyl-AMP is transferred to the 3' end of the tRNA molecule to form aminoacyl-tRNA

The overall reaction is:

Amino acid + ATP + tRNA → aminoacyl-tRNA + AMP + PP_i

Initiation of Protein Synthesis

- The first codon translated in all mRNAs is the start codon or initiation codon, AUG which codes for methionine.

- Two different tRNAs are used for the two types of AUG codon; $tRNA_f^{Met}$ is used for the initiation codon and is called the initiator tRNA whereas $tRNA_m^{Met}$ is used for internal AUG codons.
- In prokaryotes the first amino acid of a new protein is N-formylmethionine (abbreviated fMet). Hence the aminoacyl-tRNA used in initiation is fMet-$tRNA_f^{Met}$.
- A short sequence rich in purines (5'-AGGAGGU-3'), called the **Shine–Dalgarno sequence**, lies 5' to the AUG initiation codon and is complementary to part of the 16S rRNA in the small ribosomal subunit.
- Therefore this is the binding site for the 30S ribosomal subunit which then migrates in a 3' direction along the mRNA until it encounters the AUG initiation codon.
- Initiation of protein synthesis requires proteins called initiation factors (IFs).
- In prokaryotes, three initiation factors (IF-1, IF-2 and IF-3) are essential.
- Because of the complexity of the process, the exact order of binding of IF-1, IF-2, IF-3, fMet-tRNAf is controversial.

Steps Involved

1. Initiation begins with the binding of IF-1 and IF-3 to the small (30S) ribosomal subunit.
- Their role is to stop the 30S subunit binding to the 50S subunit in the absence of mRNA and fMet-$tRNA_f^{Met}$ which would result in a nonfunctional ribosome.
2. The small subunit then binds to the mRNA via the Shine–Dalgarno sequence and moves 3' along the mRNA until it locates the AUG initiation codon.
3. The initiator tRNA charged with N-formylmethionine and in a complex with IF-2 and GTP (fMet-tRNAfMet/IF-2/GTP) now binds.
4. IF-3 is released.
5. The complex of mRNA, fMet-tRNAf Met, IF-1, IF-2 and the 30S ribosomal subunit is called the 30S initiation complex.
6. The large (50S) ribosomal subunit now binds, with the release of IF-1 and IF-2 and hydrolysis of GTP, to form a 70S initiation complex.

Elongation of Protein Synthesis

- At the start of the first round of elongation, the initiation codon (AUG) is positioned in the P site with fMet-$tRNA_f^{Met}$ bound to it via codon–anticodon base pairing.
- The next codon in the mRNA is positioned in the A site.
- Elongation of the polypeptide chain occurs in three steps called the elongation cycle, namely aminoacyl-tRNA binding, peptide bond formation and translocation:

Aminoacyl-tRNA binding

- The corresponding aminoacyl-tRNA for the second codon binds to the A site via codon–anticodon interaction.
- Binding of the aminoacyl-tRNA requires elongation factor EF-Tu and GTP which bind as an aminoacyl-tRNA/EF-Tu/GTP complex.
- Following binding, the GTP is hydrolyzed and the EF-Tu is released, now bound to GDP.
- Before the EF-Tu molecule can catalyze the binding of another charged tRNA to the ribosome, it must be regenerated by a process involving another elongation factor, EF-Ts.

This regeneration is called the EF-Tu–EF-Ts exchange cycle.

- First, EF-Ts binds to EF-Tu and displaces the GDP. Then GTP binds to the EF-Tu and displaces EF-Ts. The EF-Tu-GTP is now ready to take part in another round of elongation.

Peptide bond formation

- The second step, peptide bond formation, is catalyzed by peptidyl transferase.
- In this reaction the carboxyl end of the amino acid bound to the tRNA in the P site is uncoupled from the tRNA and becomes joined by a peptide bond to the amino group of the amino acid linked to the tRNA in the A site.

Translocation

- In the third step, a complex of elongation factor EF-G (also called translocase) and GTP (i.e. EF-G/GTP) binds to the ribosome.
- Three concerted movements now occur, collectively called translocation:

1. the deacylated tRNA moves from the P site to the E site
2. the dipeptidyl-tRNA in the A site moves to the P site, and
3. the ribosome moves along the mRNA (5' to 3') by three nucleotides to place the next codon in the A site.

- During the translocation events, GTP is hydrolyzed to GDP and inorganic phosphate, and EF-G is released ready to bind more GTP for another round of elongation.
- After translocation, the A site is empty and ready to receive the next aminoacyltRNA.
- The A site and the E site cannot be occupied simultaneously. Thus the deacylated tRNA is released from the E site before the next aminoacyl-tRNA binds to the A site to start a new round of elongation.
- Elongation continues, adding one amino acid to the C-terminal end of the growing polypeptide for each codon that is read, with the peptidyl-tRNA moving back and forth from the P site to the A site as it grows.

Termination

- Eventually, one of three termination codons (also called Stop codons) becomes positioned in the A site. These are UAG, UAA and UGA.
- Unlike other codons, prokaryotic cells do not contain aminoacyl-tRNAs complementary to
- Stop codons. Instead, one of two release factors (RF-1 and RF-2) binds instead.
- RF-1 recognizes UAA and UAG whereas RF-2 recognizes UAA and UGA. A third release factor, RF-3, is also needed to assist RF-1 or RF-2 interaction with the ribosome. Thus either RF-1 + RF-3 or RF-2 + RF-3 bind depending on the exact termination codon in the A site.
- RF-1 (or RF-2) binds at or near the A site whereas RF-3/GTP binds elsewhere on the ribosome.
- The release factors cause the peptidyl transferase activity to transfer the polypeptide to a water molecule instead of to aminoacyl-tRNA, effectively cleaving the bond between the polypeptide and tRNA in the P site.

The free polypeptide now leaves the ribosome, followed by the mRNA and free tRNA, and the ribosome dissociates into 30S and 50S subunits ready to start translation again.

RNA Type # 1. Transfer RNA (tRNA):

It delivers amino acids to ribosome and decodes the information of mRNA. Each nucleotide triplet codon on mRNA represents an amino acid. The tRNA plays the role of an adaptor and matches each codon to its particular amino acid in the cytopolasmic pool.

The tRNA has two properties:

(a) It represents a single amino acid to which it binds covalently.

(b) It has two sites. One is a trinucleotide sequence called anticodon, which is complementary to the codon of mRNA. The codon and anticodon form base pairs with each other. The other is amino acid binding site.

There are many different kinds of tRNA molecules in a cell. Each tRNA is named after the amino acid it carries. For example if tRNA carries amino acid tyrosine it is written as tRNATyr. Sometimes there are more than one tRNA for an amino acid, then it is denoted as tRNA1Try and tRNA2Try. A minimum of 32 tRNAs are required to translate all 61 codons.

The tRNA charged with an amino acid is called amino acyl tRNA.

Clover Leaf Structure of tRNA:

The primary structure of all tRNA molecules is small, linear, single stranded nucleic acid ranging in size from 73 to 93 nucleotides. The tRNA due to its property of having stretches of complementary base pairs forms secondary structure, which is in the form of a cloverleaf.

Several regions of the single stranded molecule form double stranded stems or arms and single stranded loops due to folding of various regions of the molecule. These double stranded stems have complementary base pairs. A typical tRNA has bases numbering from 1-76, using the standard numbering convention where position 1 is the 5′ end and 76 is the 3′ end.

The various regions of the clover leaf model of tRNA are as follows:

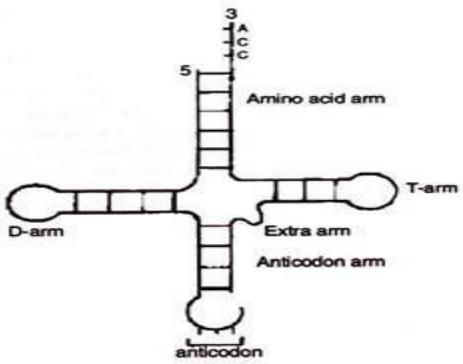


Fig. 7.3. Clover leaf model of tRNA.

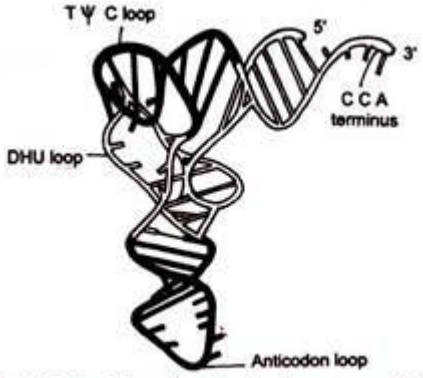


Fig. 7.4. L-shaped three dimensional tertiary structure of tRNA.

1. Amino acid arm:
It has a seven base pairs stem formed by base pairing between 5′ and 3′ ends of tRNA. At 3′ end a sequence of 5′-CCA-3′ is added. This is called CCA arm or amino acid acceptor arm. Amino acid binds to this arm during protein synthesis.

2. D-arm:
Going from 5′ to 3′ direction or anticlockwise direction, next arm is D-arm. It has a 3 to 4 base pair stem and a loop called D-loop or DHU-loop. It contains a modified base dihydrouracil.

3. Anticodon arm:
Next is the arm which lies opposite to the acceptor arm. It has a five base pair stem and a loop in which there are three adjacent nucleotides called anticodon which are complementary to the codon of mRNA.

4. An extra arm:
Next lies an extra arm which consists of 3-21 bases. Depending upon the length, extra arms are of two types, small extra arm with 3-5 bases and other a large arm having 13-21 bases.

5. T-arm or TψC arm:
It has a modified base pseudouridine ψ. It has a five base pair stem with a loop.

There are about 50 different types of modified bases in different tRNAs, but four bases are more common. One is ribothymidine which contains thymine which is not found in RNA. Other modified bases are pseudouridine ψ, dihyrouridine and inosine.

Three Dimensional Structure of tRNA:
X-ray crystallographic analysis of tRNA shows three dimensional structure called tertiary structure. The molecule is folded and has two helical double stranded branches. One branch consists of acceptor arm and T ψ C arm. The other arm consists of DHU loop and anticodon arm with loop.

The tRNA molecule is L- shaped. The tertiary structure creates two double helices at right angle to each other. The amino acid binding site is opposite to the anticodon arm. This facilitates protein synthesis.

The tRNA constitutes about 10% of the total cellular RNA.

RNA Type # 2. Messenger RNA (mRNA):

Messenger RNA is a linear molecule transcribed from one strand of DNA. It carries the base sequence complementary to DNA template strand. The base sequence of mRNA is in the form of consecutive triplet codons. Ribosomes translate these triplet codons into amino acid sequence of polypeptide chain.

Length of mRNA:

Length of mRNA depends upon the length of polypeptide chain it Codes for. Polypeptide length varies from a chain of a few amino acids to thousands of amino acids. For example, a sequence of 600 nucleotides will code for a polypeptide having a chain of 200 amino acids. The message is read in the groups of three consecutive bases from a fixed starting point.

Life Span of mRNA:

In bacteria, mRNA is transcribed and translated in a single cellular compartment and the two processes are so closely linked that they occur simultaneously. Transcription begins when the enzyme RNA polymerase binds to DNA and then moves along making a copy of one strand. As soon as the transcription begins, the ribosomes attach to the 5′ end (free end) of the mRNA and start translation while the other end of mRNA is still under synthesis.

This is known as coupled transcription and translation in prokaryotes. After the translation of whole of mRNA is completed, the mRNA is then degraded in 5′ → 3′ direction. The mRAN is synthesized, translated and degraded all in rapid succession and all in 5′ → 3′ direction. An individual mRNA molecule survives only for a minute or less.

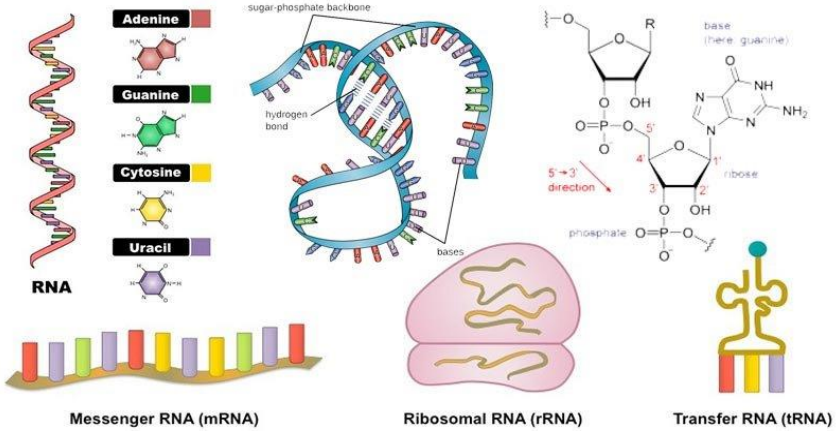
In eukaryotes, transcription occurs in the nucleus while translation takes place in cytoplasm. Eukaryotic mRNA is quite stable and survives from a few minutes to more than a day. In mammalian RBC, through the nucleus is lost, mRNA continues to produce haemoglobin for many days.

Eukaryotic mRNA constitutes only a small proportion of the total cellular RNA. It is only about 3% of the total RNA.

Coding and Non-coding Regions:

All mRNAs have two types of regions. The coding region consists of series of codons. But the mRNA is longer than the coding regions. Length of newly synthesized mRNA is much larger than the length of mRNA used for translation. The coding regions are called exons. Between the

coding regions lie various non-coding regions called introns. Genes with these intervening sequences are called Split genes or Interrupted genes.



Adaptor hypothesis: The adaptor hypothesis is part of a scheme to explain how information encoded in DNA is used to specify the amino acid sequence of proteins. It was formulated by Francis Crick in the mid-1950s, together with the central dogma of molecular biology and the sequence hypothesis.

Explanation : The adaptor hypothesis was framed to explain how information could be extracted from a nucleic acid and used to put together a string of amino acids in a specific sequence, that sequence being determined by the nucleotide sequence of the nucleic acid (DNA or RNA) template. Crick proposed that each amino acid is first attached to its own specific "adaptor" z ▪ piece of nucleic acid (in an enzyme-catalysed reaction). The order of assembly of the amino acids is then determined by a specific recognition between the adaptor and the nucleic acid which is serving as the informational template. In this way the amino acids could be lined up by the template in a specific order. Coupling between adjacent amino acids would then lead to the synthesis of a polypeptide whose sequence is determined by the template nucleic acid.

Basis: Crick's thinking behind this proposal was based on a general consideration of the chemical properties of the two classes of molecule — nucleic acids and proteins. The amino acids are characterised by having a variety of side chains which vary from being hydrophilic to hydrophobic: their individual characters reside in the very different properties these side chains have. By contrast, a nucleic acid is composed of a string of nucleotides whose sequence presents a geometrically defined surface for hydrogen bonding. z ▪ This makes nucleic acids good at recognising each other, but poor at distinguishing the varied side chains of amino acids. It was this apparent lack of any possibility of specific recognition of amino acid side chains by a nucleotide sequence which led Crick to conclude that amino acids would first become attached to a small nucleic acid — the adaptor — and that this, by base-pairing with the template (presumably as occurs between DNA strands in the double helix), would carry the amino acids to be lined up on the template.

Proof: ▪ That such adaptors do exist was discovered by Mahlon Hoagland and Paul Zamecnik in 1958. These "soluble RNAs" are now called transfer RNAs and mediate the translation of messenger RNAs on ribosomes according to the rules contained in the genetic code. Crick imagined that his adaptors would be small, perhaps 5-10 nucleotides long. In fact, they are much

larger, having a more complex role to play in protein synthesis, and are closer to 100 nucleotides in length.

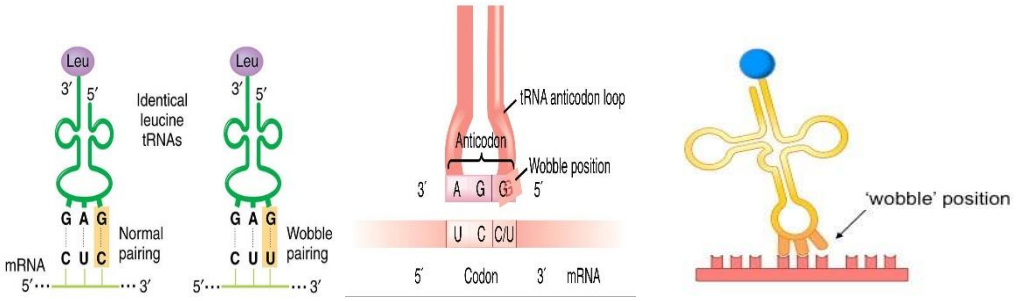
The Wobble Hypothesis

- There are more than one codon for one **amino acid**. This is called degeneracy of genetic code.
- To explain the possible cause of degeneracy of codons, in 1966, Francis Crick proposed **"the Wobble hypothesis".**
- According to this hypothesis, only the first two bases of the codon have a precise pairing with the bases of the anticodon of tRNA, while the pairing between the third bases of codon and anticodon may Wobble (wobble means to sway or move unsteadily).
- The phenomenon permits a single tRNA to recognize more than one codon. Therefore, although there are 61 codons for amino acids, the number of tRNA is far less (around 40) which is due to wobbling.

The Wobble Hypothesis

The wobble hypothesis states that the base at 5′ end of the anticodon is not spatially confined as the other two bases allowing it to form hydrogen bonds with any of several bases located at the 3′ end of a codon. This leads to the following conclusions:

- The first two bases of the codon make normal (canonical) H-bond pairs with the 2nd and 3rd bases of the anticodon.
- At the remaining position, less stringent rules apply and non-canonical pairing may occur. The wobble hypothesis thus proposes a more flexible set of base-pairing rules at the third position of the codon.
- The relaxed base-pairing requirement, or "wobble," allows the anticodon of a single form of tRNA to pair with more than one triplet in mRNA.
- The rules: first base U can recognize A or G, first base G can recognize U or C, and first base I can recognize U, C or A.



Crick's hypothesis hence predicts that the initial two ribonucleotides of triplet codes are often more critical than the third member in attracting the correct tRNA.

Wobble base pairs

- A wobble base pair is a pairing between two nucleotides in RNA molecules that does not follow Watson-Crick base pair rules.
- The four main wobble base pairs are guanine-uracil (G-U), hypoxanthine-uracil (I-U), hypoxanthine-adenine (I-A), and hypoxanthine-cytosine (I-C).

- In order to maintain consistency of nucleic acid nomenclature, “I” is used for hypoxanthine because hypoxanthine is the nucleobase of inosine.
- Inosine displays the true qualities of wobble, in that if that is the first nucleotide in the anticodon then any of three bases in the original codon can be matched with the tRNA.

Significance of the Wobble Hypothesis

- Our bodies have a limited amount of tRNAs and wobble allows for broad specificity.
- Wobble base pairs have been shown to facilitate many biological functions, most clearly proven in the bacterium ***Escherichia coli***.
- The thermodynamic stability of a wobble base pair is comparable to that of a Watson-Crick base pair.
- Wobble base pairs are fundamental in RNA secondary structure and are critical for the proper translation of the genetic code.
- Wobbling allows faster dissociation of tRNA from mRNA and also protein synthesis.
- The existence of wobble minimizes the damage that can be caused by a misreading of the code; for example, if the Leu codon CUU were misread CUC or CUA or CUG during transcription of mRNA, the codon would still be translated as Leu during protein synthesis.