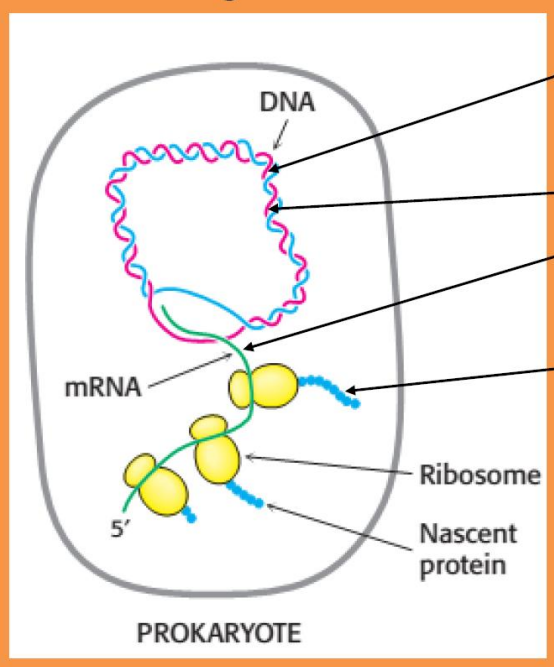
6. MOLECULAR BASIS OF INHERITANCE

Chapter @ a Glance

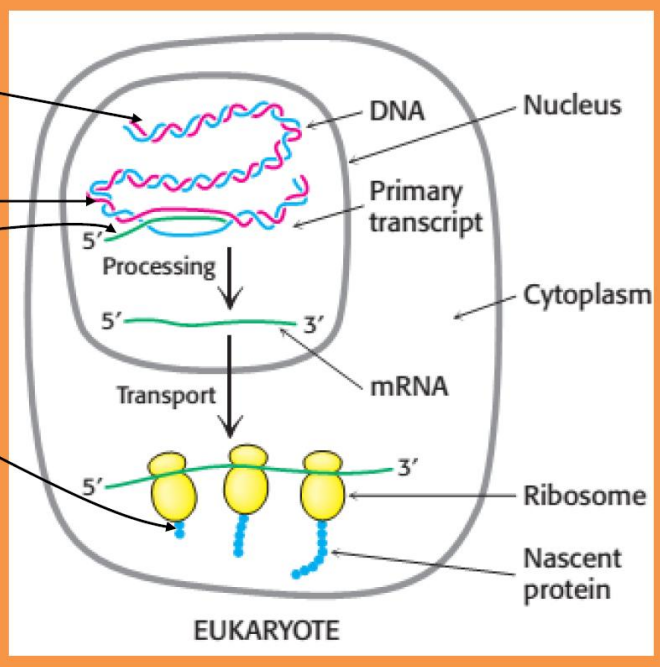
I. Search for genetic Material



II. Structure of DNA

III. Central Dogma
1. DNA replication
2. Transcription
- Types of RNA
- Genetic Code
3. Translation

4. Regulation of Gene Expression
e.g.: Lac Operon



IV. Human Genome Project
V. DNA finger printing

Genetic material is the substance that carries the biological information regarding the structural, functional, developmental and behavioural properties of organisms. It also serves as the agent of **inheritance**, i.e., transmits coded biological information from parents to progeny.
DNA & RNA (***Nucleic acids***) are the 2 types of *genetic material* found in living organisms.

Properties of Genetic Material

A genetic material must
(i) Be able to **generate its replica** (Replication).
(ii) Chemically and structurally be **stable**.
(iii) Potentiality to **generate variation** (through **mutation**s or recombination) that are required for evolution.
(iv) Be able to express itself as **'Mendelian Characters'** (characters specified by it).

THE SEARCH FOR GENETIC MATERIAL

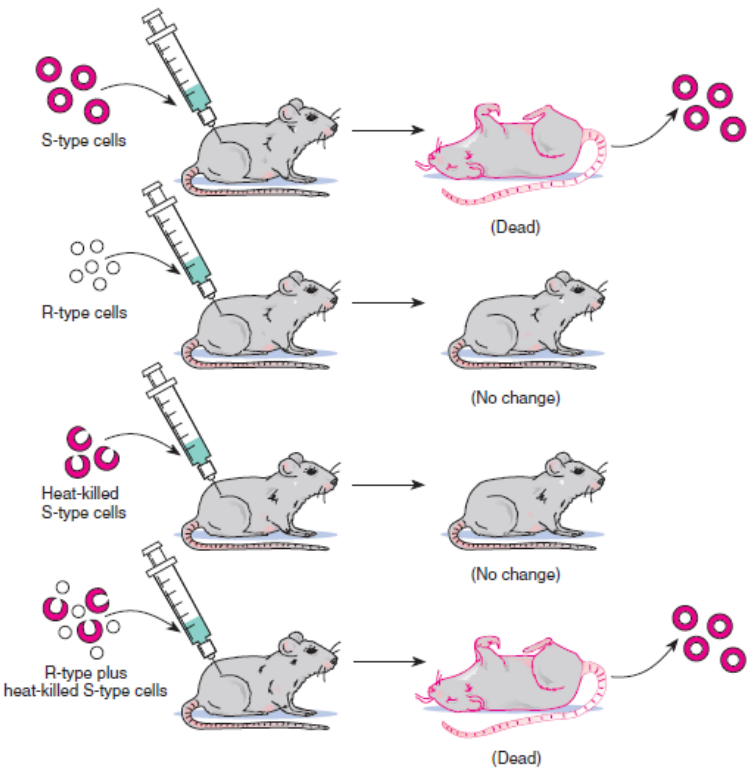
1. **Griffith's Experiment** (Transforming principle)

- **Frederick Griffith** (1928) performed experiments on **bacterial transformation** (transfer of genetic material from one bacterium to another) in *Streptococcus pneumoniae* (causing pneumonia).

This bacterium has 2 strains-

***S** (smooth) strain*	***R** (Rough) strain*
▪ Has a polysaccharide coat(hence they produce smooth colonies on culture plate)	▪ No coat (hence they produce rough colonies)
▪ Virulent (causes pneumonia)	▪ Non virulent

Experiment:
- S-strain (live) → Inject into mice → Mice die (caused pneumonia)
- R-strain (live) → Inject into mice → Mice live (no ill-effect)
- S-strain (Heat killed) → Inject into mice → Mice live
- S-strain (Hk) + R-strain (live) → Inject into mice → Mice die



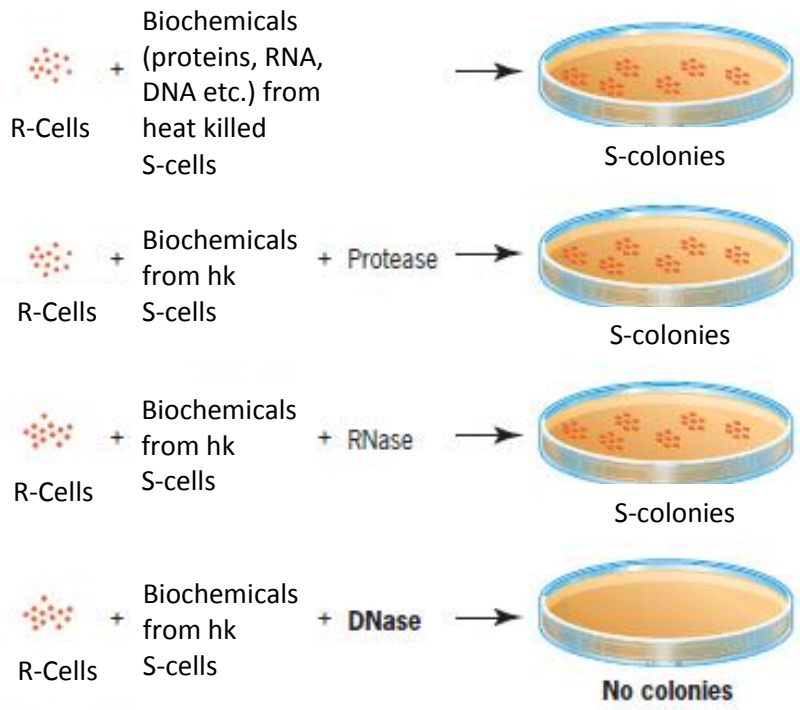
Conclusion:-
Some chemical substances present in the killed S-strain **transform** R-strain to synthesis polysaccharide coat & become virulent S-strain. (This must be due to the transfer of genetic material).

2. Biochemical explanation for Griffith's observation

Avery, **MacLeod** and **McCarty** (1944) perfected the biochemical nature of substance responsible for transformation in Griffith's experiment.

- They purified biochemicals (proteins, RNA, DNA etc.) from the heat-killed S cells to see which ones could transform live R cells into S cells. They had following observations:
  - Digestion of protein and RNA (using Proteases and RNases) did not affect transformation.
  - Digestion of DNA (with DNase) inhibited transformation.

It means that DNA from S strain bacteria causes R strain to be transformed (to virulent).



Conclusion: - Because only DNase destroyed the transforming substance, the transforming substance is DNA and it is the genetic material.

3. Experiment to confirm DNA as the genetic material (Harshey-Chase Experiment)

Experiments of **Hershey** and **Chase** (1952) with bacteriophages again proved that DNA is the genetic material.

Fact:-Phages are viruses that infect bacteria. A phage has 2 parts- **protein coat** (contain S, but no P) and **DNA** (contain P, but no S). When phage are mixed with bacteria, they first attaches to the bacteria and then some material from them enters the bacterial cell. The bacterial cell treats the viral (genetic) material as if it was its own and subsequently manufactures more virus particles.

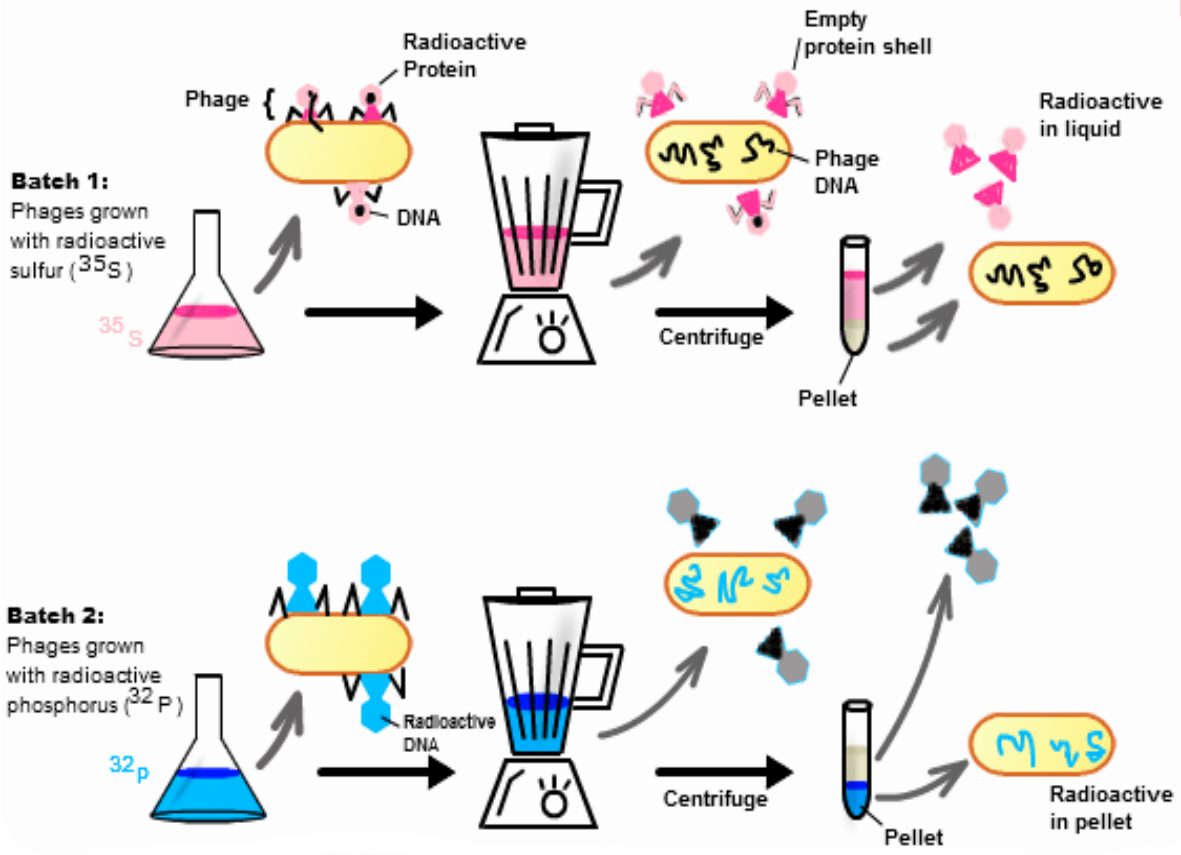
This experiment was to determine whether it is the phage DNA or protein that is injected to the bacterium during infection.

Experimental steps:-

i. Prepared two cultures of bacteriophage, T_2 –
   In one, **protein coat** were radioactively labelled with ^{35}**S**
   In the second, **DNA** is radioactively labelled with ^{32}P.

ii. **Radioactive phages** were used separately to infect *E. coli*.

iii. After infection, the *E. coli* was **blended** and **centrifuged**.
   As bacterial cells were heavier, they settled at the bottom. The supernatant contain lighter viral particle.



Observation:-

- Bacteria which was infected with viruses that had radioactive DNA were radioactive
- Bacteria that were infected with viruses that had radioactive proteins were not radioactive.

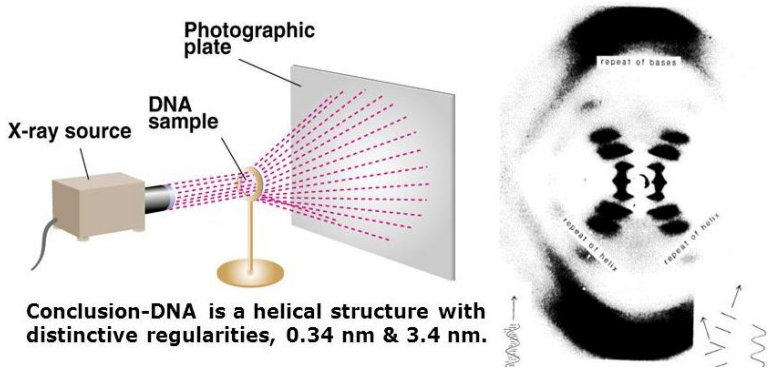
Inference:-

When a virus infects a bacterium, only viral DNA gets into bacterium and the viral protein remains outside. So it is the viral DNA-not protein- contain genetic information to make new viral material.

Structure of DNA

- **Friedrich Meischer (1869):** Identified DNA and named it as **'Nuclein' (**as a constituent of nucleus**).**
- **James Watson** & **Francis Crick (1953):** Proposed **double helix model of DNA.** It was based on-

I. The X-ray diffraction data produced by **Maurice Wilkins** & **Rosalind Franklin** (1953).



Conclusion-DNA is a helical structure with distinctive regularities, 0.34 nm & 3.4 nm.

II. **Chargaff's rules** of base pairing.

The analytical study by Erwin Chargaff (1950) reveals that-

(i) The amount of A is always equal to the amount of T and the amount of G is always equal to the amount of C.

(ii) A is joined to T with 2 H-bonds while G to C by 3 H-bond

(iii)The ratio of A to T and that of G to C is always equal to one.

i.e., $\frac{A}{T} = \frac{G}{C} = 1$

STRUCTURE OF POLYNUCLEOTIDE CHAIN

Polynucleotides are the polymer of **nucleotides.** DNA & RNA are polynucleotides. A nucleotide has 3 components:

1. A nitrogenous base, 2 types -

- **Purines:** It includes **Adenine (A)** and **Guanine (G).**
- **Pyrimidines:** It includes **Cytosine (C), Thymine (T)** & **Uracil (U).**

Thymine (5-methyl Uracil) present only in DNA and Uracil only in RNA.

b. A pentose sugar (ribose in **RNA & deoxyribose** in **DNA)**

c. A phosphate group – provides acidity to the nucleic acids

- A nitrogenous base is linked to the pentose sugar through an **N-glycosidic linkage** to form **nucleoside.**

Nitrogen base + sugar + phosphate group = Nucleotide (deoxyribonucleotide).

Nucleoside

- In RNA, every nucleotide residue has an additional –OH group present at 2′-position in the ribose.
- 2 nucleotides are linked through **3'-5' phosphodiester bond** to form **dinucleotide.** When more nucleotides are linked it forms **polynucleotide.**
- Each polynucleotide chain has 2 free ends- **3 prime (3')** end and the opposite **5 prime (5')** end**.**
  In 3' end, the 3rd C-atom of the sugar is free, i.e., it is not linked to any nucleotide.
  Similarly in the 5' end, the 5th C of sugar is free.

Salient features of double helix model of DNA:

(i) DNA is made of 2 polynucleotide chains coiled in a right handed fashion like a spiral staircase. Its rail is formed of sugar & phosphates and the H-bonded base pairs as the steps.

(ii) The 2 chains have **anti-parallel polarity**, i.e. one chain has the polarity **5'→3'** and the other has **3'→5'.**

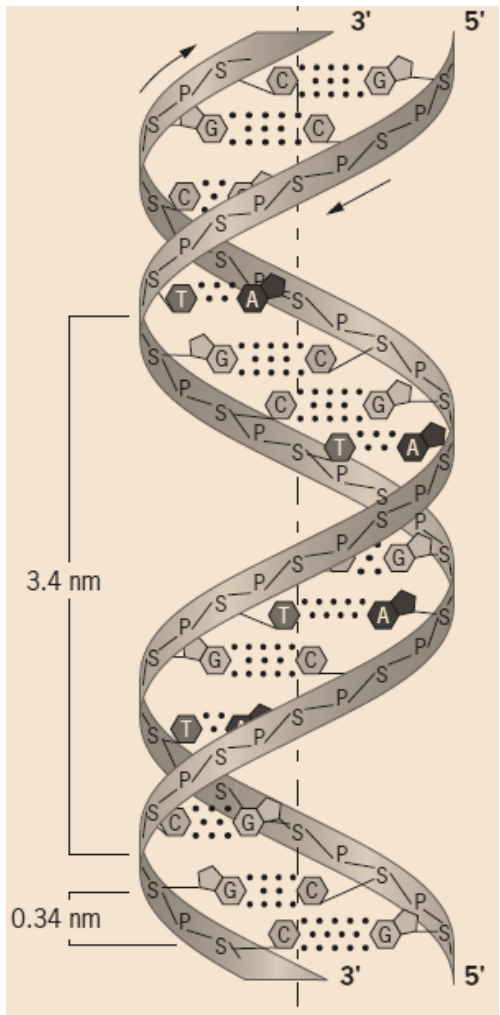
(iii) The bases in 2 stands are paired through **H-bonds** forming **base pairs (bp).**
A=T (2 H bonds) C≡G (3 H bonds). As a result, purine comes opposite to a pyrimidine, this generates uniform distance between the 2 strands.

➔ They are said to be complementary to each other, and therefore if the sequence of bases in one strand is known then the sequence in other strand can be predicted.

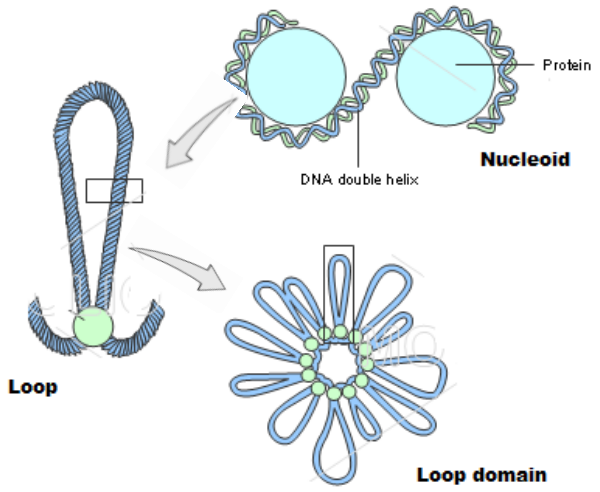
(iv) **Length of DNA** = number of base pairs **X** distance between two adjacent base pairs (**0.34nm**).
This also is the characteristic of an organism.
For ex: Φ 174 (a bacteriophage) - 5386 nucleotides (single stranded)
Lambda (a bacteriophage) - 48502 bp
E. coli - 4.6×10^6 bp
Human (haploid) - 3.3×10^9 bp

Number of base pairs in human = 6.6×10^9
Hence, the length of DNA = $6.6 \times 10^9 \times 0.34 \times 10^{-9} = 2.2$ m
In *E. coli*, length of DNA = 1.36 mm (1.36×10^{-3} m)
∴ The number of base pairs = $\frac{1.36 \times 10^{-3}}{0.34 \times 10^{-9}} = 4 \times 10^6$ bp

(v) The plane of one base pair stacks over the other in a double helix. This provides stability of the helical structure.



Packaging of DNA Helix

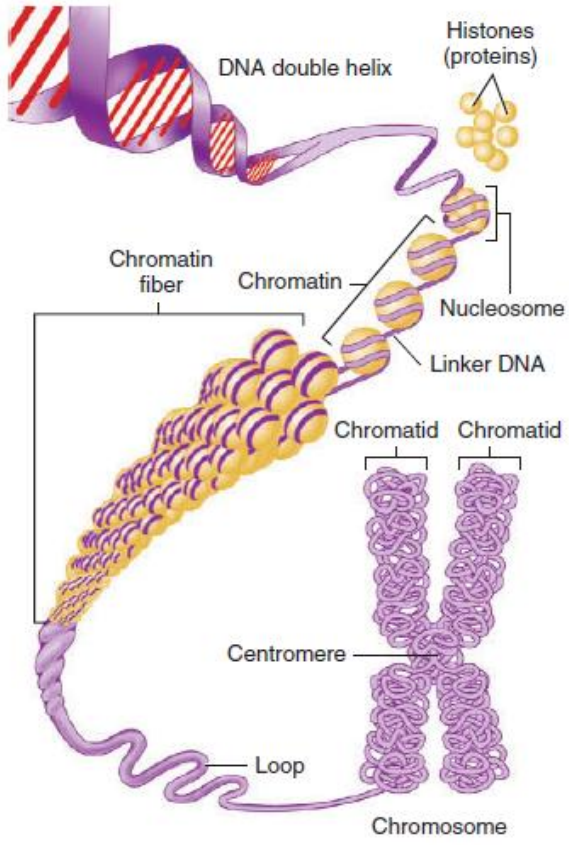


In prokaryotes

- ♥ The prokaryotes lack defined nucleus. So the DNA is found in a compact structure called **nucleoid**.
- ♥ DNA (negatively charged) is held with some proteins (that have positive charges) and form '**nucleoid'**. The DNA in nucleoid is organised in about 50 large **loops** held by proteins.

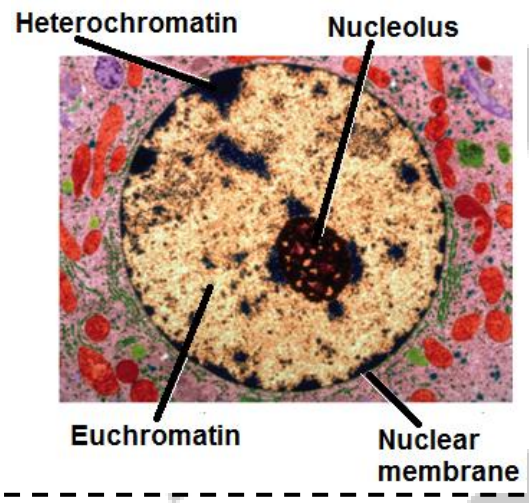
In eukaryotes

- ♥ In eukaryotes, DNA is found in the nucleus.
- ♥ DNA is wrapped around a unit of 8 molecules of positively charged protein **Histone** (histone octamer) to form a structure called **nucleosome.**
  It contains 200 bp of DNA helix.
- ♥ Nucleosomes constitute the repeating unit like 'beads-on-string' structure in nucleus called **chromatin**,.



❖ Chromatins include -

Euchromatin	Heterochromatin
1. Loosely-packed region of chromatin	1. Densely-packed region
2. Lightly-stained regions	2. Darly-stained regions
3. Transcriptionally active	3. Transcriptionally inactive

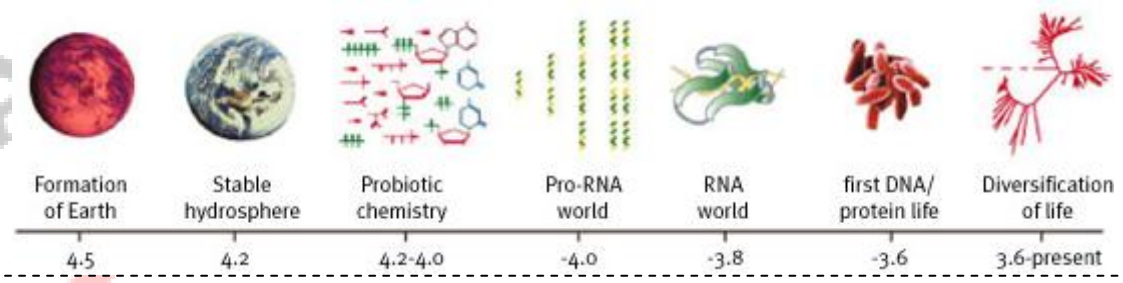


- ♥ Chromatin is packaged to form **chromatin fibers**.
- ♥ Chromatin fibers that are further coiled and condensed at metaphase stage of cell division to form **chromosomes**. (Chromosome= 40% DNA+ 50% histones + 8.5% protein + 1.5% RNA)

➔ Higher level packaging of chromatin requires **non-histone chromosomal (NHC) proteins.**

RNA WORLD

- ➢ RNA was the first genetic material.
- ➢ It acts as genetic material as well as catalyst.
- ➢ Essential life processes (metabolism, translation, splicing etc) evolved around RNA.
- ➢ DNA evolved from RNA for stability.



DNA V/S RNA

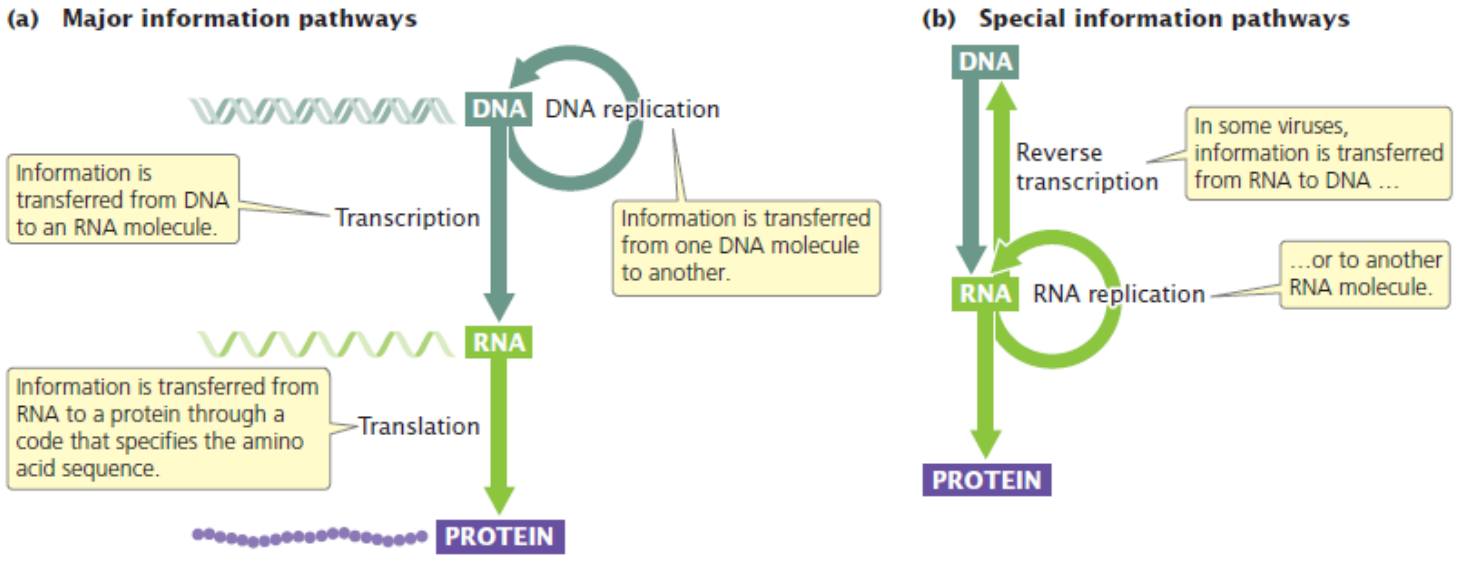
DNA	RNA
Acts as the genetic material in most of the organisms	- Mostly functions as messenger. - Acts as the genetic material in some viruses (***Retro viruses***- Tobacco Mosaic viruses, QB bacteriophage, HIV etc.) - Also functions as adapter and structural molecules. - Act as enzyme (***ribozyme***).
DNA is a better genetic material ∵ Chemically less reactive and structurally more stable due to • Being double stranded • Presence of thymine (5-methyl Uracil). • Absence of 2′-OH in deoxyribose sugar	**RNA is better for the transmission of genetic information** ∵ Labile and easily degradable due to • Being single stranded • Presence of uracil (less stable compared to thymine). • Presence of 2′-OH in ribose sugar (a reactive group)
Resist mutation by repair mechanism	Mutate at a faster rate
Dependent on RNA for synthesis of proteins	Can directly code for the synthesis of proteins, hence can easily express the characters.

CENTRAL DOGMA OF MOLECULAR BIOLOGY

It is proposed by **Francis Crick.** It states that the genetic information flows from DNA → RNA → Protein.

Central Dogma Reverse-

* In some viruses, the flow of information is in reverse direction. i.e., RNA→DNA -**Reverse transcription.** Virus capable of reverse transcription are called **retroviruses** (eg: HIV).



DNA Replication

- *It is the copying of DNA from parental DNA.*
- **Watson** & **Crick** (1953) proposed **Semi-conservative model** of replication.
  It suggests that the 2 parental DNA strands separate and each act as **template** for the synthesis of new complementary strands. After the completion of replication, each DNA molecule would have one parental and one new strand.

Experimental proofs for semi conservative DNA replication

1. **Matthew Messelson & Franklin Stahl** (1958) experiment on *E. coli*

Steps:-

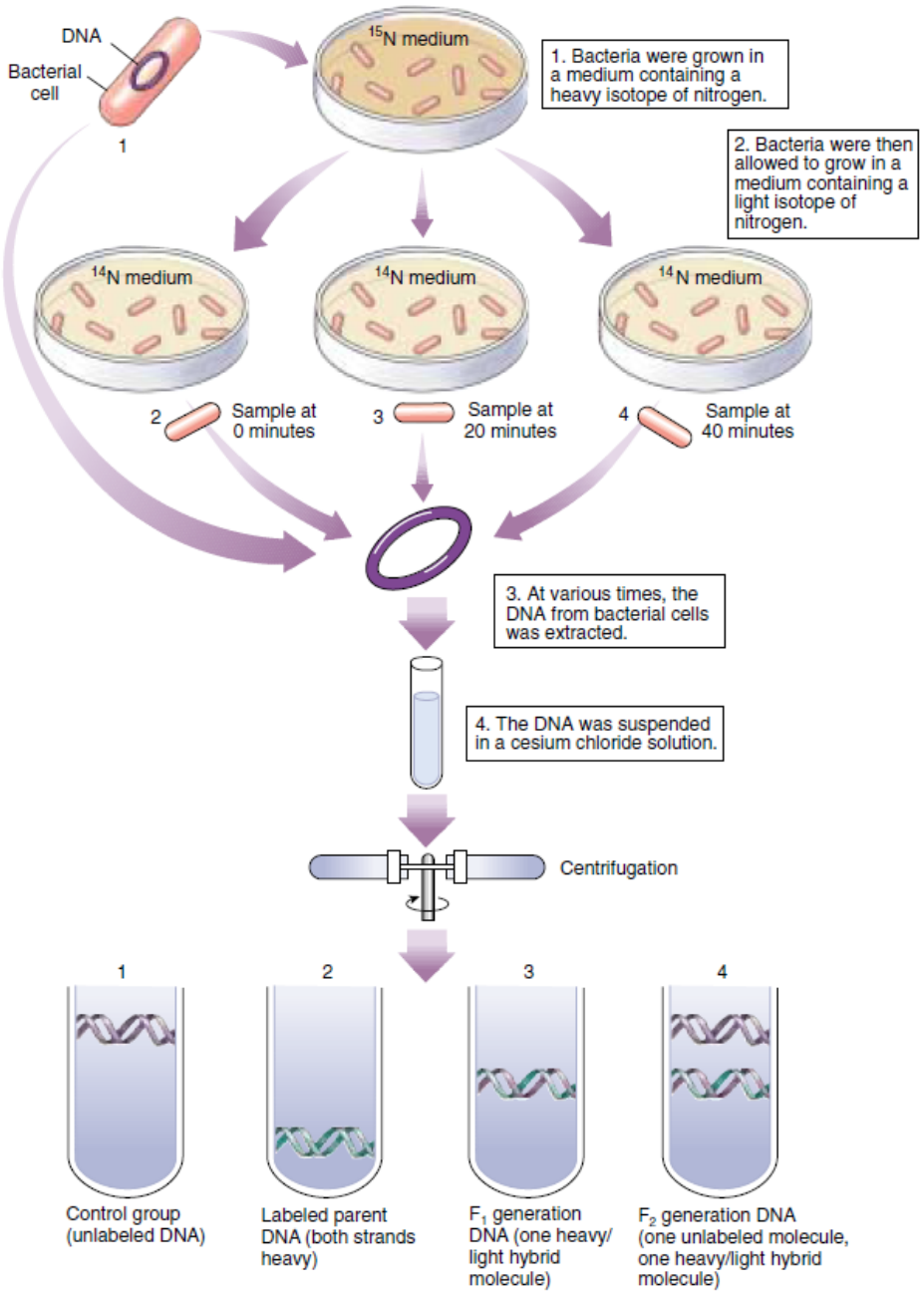
(i) Cultured *E. coli* in a medium containing ^{15}**N** (heavy isotope of N) as the only nitrogen source for many generations. (Thus, ^{15}**N** will be incorporated into DNA and become **heavier)**.
(ii) The cells were transferred into a medium with normal ^{14}**N**
(iii) Took samples at various definite time intervals to extract DNA.
(iv) Then it is undergone for centrifugation in ***CsCl*** and measured to get their densities. (To distinguish heavy DNA molecule from the normal DNA).

Observation:-

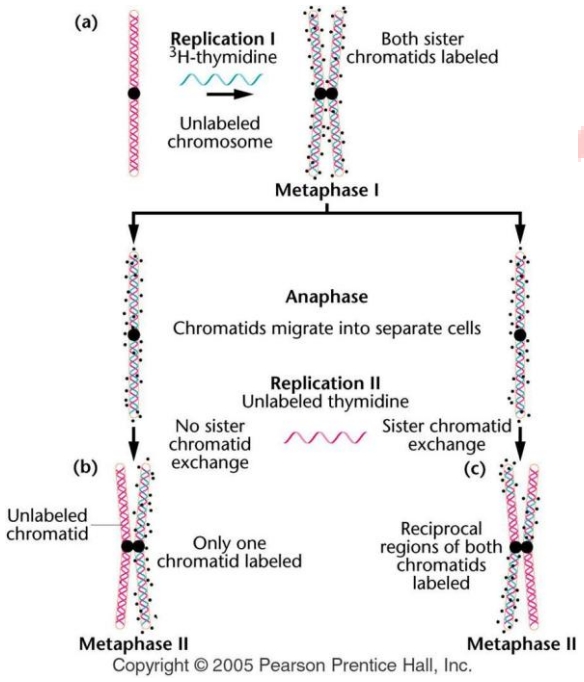
- The DNA that was extracted from the culture 1st generation after (20 min) the transfer from ^{15}N to ^{14}N medium intermediate density.
- DNA extracted from the culture after 2nd generation [i.e., after 40min.] was composed of equal amounts of this hybrid DNA and of 'light' DNA.

Inference:-

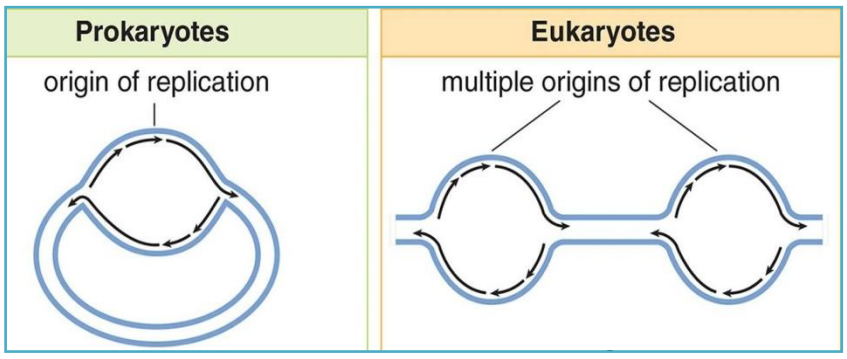
The newly synthesised DNA obtains one of its strands from the parent. i.e., replication is semi-conservative.



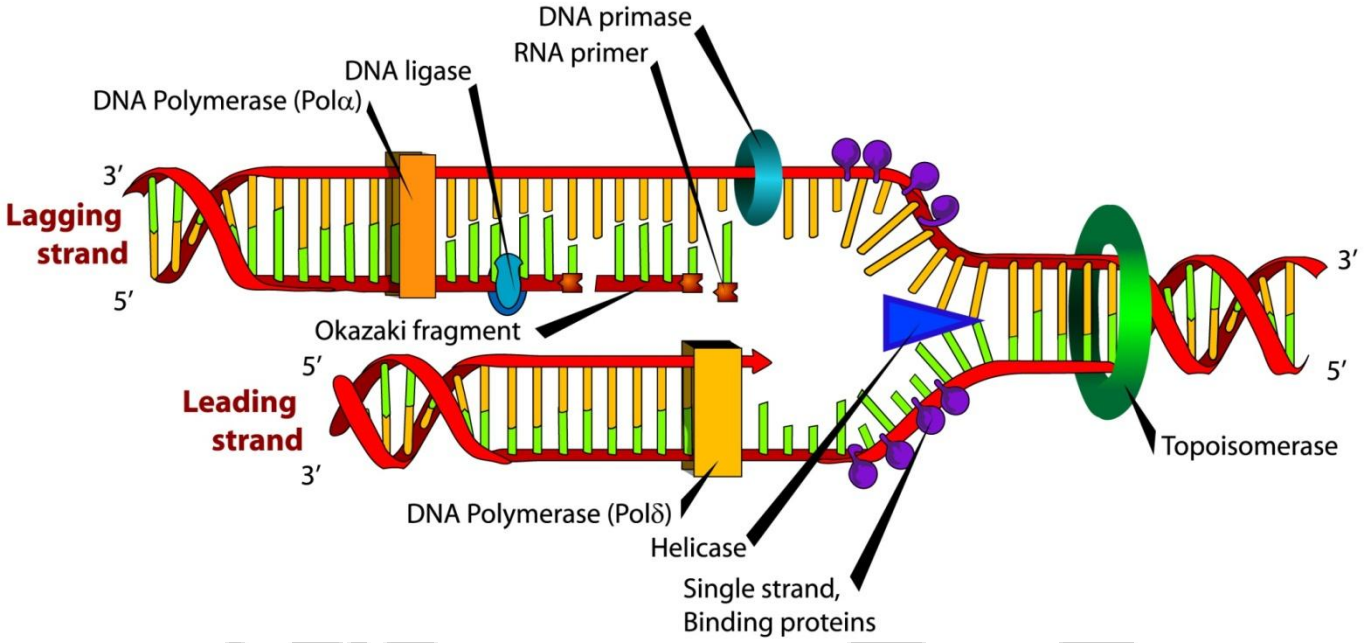
2. **Taylor** and colleagues (1958) on *Vicia faba*.
   Using **radioactive thymidine** to detect distribution of newly synthesized DNA in the chromosomes.



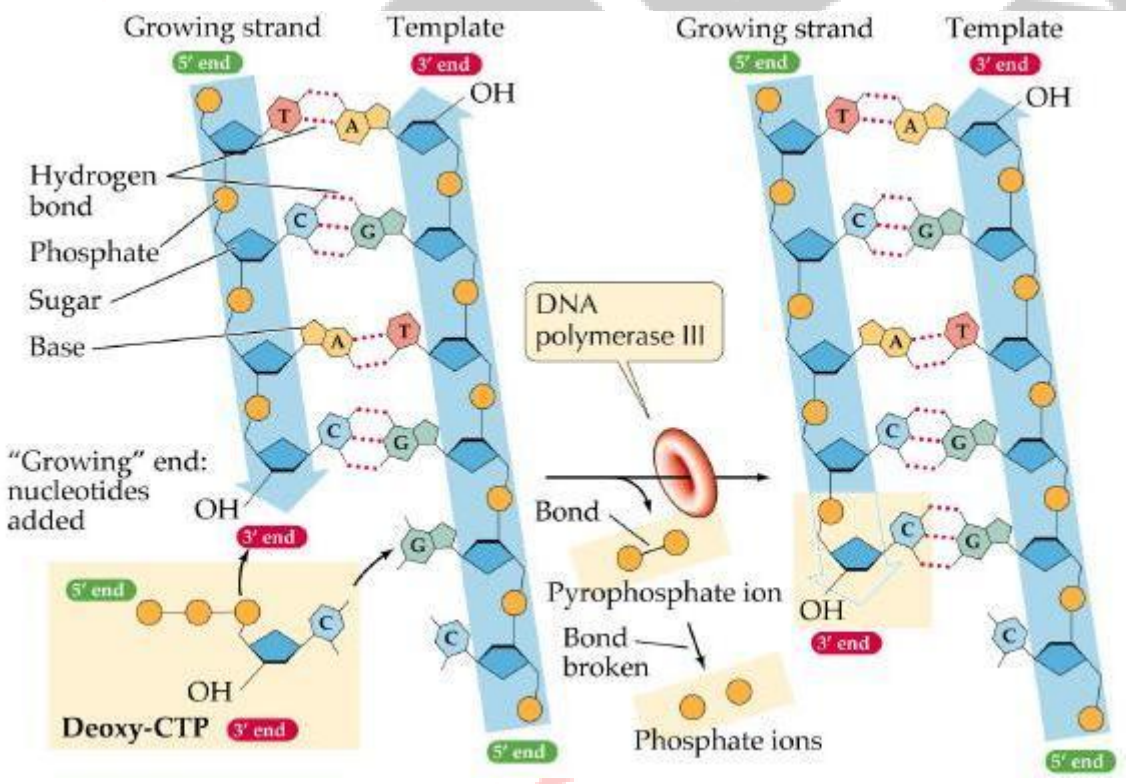
MECHANISM OF DNA REPLICATION



1. The replication of DNA takes place at S-phase of the cell-cycle.
2. DNA replicates in 5′➝3′
3. DNA replication starts at a point called **origin *(ori)***. Bacterial DNA has single *ori* while eukaryotes have many, which finally merge each other.
4. ***Helicase*** unwind the DNA strand. For long DNA molecules, since the 2 strands of DNA cannot be separated in its entire length (due to very high energy requirement), the replication occur within a small opening of the DNA helix, referred to as **replication fork**. A non-enzymatic **single-stranded DNA binding (SSB) protein,** stabilises the chain in single-stranded form to reduce the energy needed to unwind the DNA helix.
   Unwinding creates a coiling tension ahead of the moving replication fork. Enzyme ***Topoisomerases*** cut and rejoin one strand of DNA to facilitate uncoiling.
5. A short chain of RNA (**RNA primer**) formed from DNA template at the 5’ end. The RNA primer is formed because the ***DNA polymerase*** cannot initiate the synthesis of a new strand, although it can catalyse the growth of a DNA chain. (At last, RNA primers are removed from 5’ end by ***exonuclease*** and gap filled by dNTPs)
6. **DNA polymerase**, uses DNA template to catalyse the polymerisation of deoxynucleotides.



7. **Deoxyribonucleoside triphosphates** (dATP, dGTP, dCTP & dTTP) serve dual purposes-
   - They act as **substrates** for DNA synthesis.
   - They **provide energy** for polymerisation reaction (the two terminal phosphates in a dNTPs are high-energy phosphates, same as in case of ATP).



8. 2 new strands are synthesised along the replication fork- **Leading strand** and **Lagging strand**.
   On one strand (the template with polarity 3′➝ 5’), the replication is <u>continuous</u> (Leading strand), while on the other (the template with polarity 5′➝3’), it is <u>discontinuous-</u> **Okazaki fragments** (Lagging strand).
   DNA ligase joins the Okazaki fragments which are formed on lagging strand.

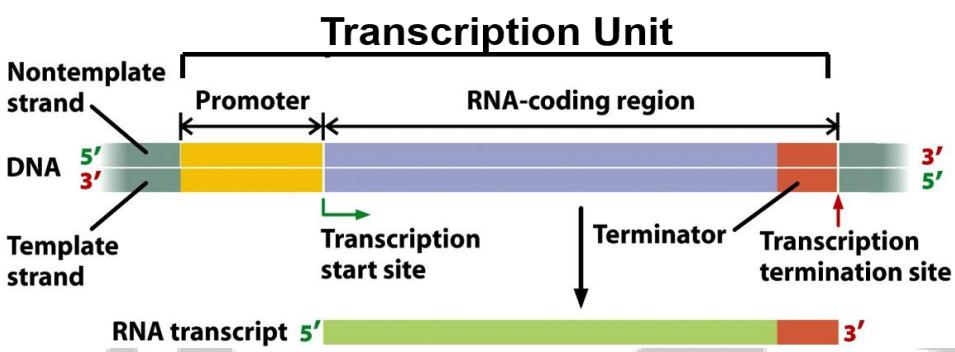
   ➔ **Caution**: The replication of DNA and cell division cycle should be highly coordinated. A failure in cell division after DNA replication results into polyploidy (a chromosomal anomaly).

Transcription (RNA Synthesis)

- *It is the process of copying genetic information from one strand of the DNA into RNA.* The genes (DNA) are 'master document from which 'working copies' are prepared in the form of mRNA.
- Here, adenine pairs with uracil instead of thymine.
- Both strands are not copied during transcription, because
  - ∵ The code for proteins is different in both strands. This complicates the translation.
  - ∵ If 2 RNA molecules are produced simultaneously this would be complimentary to each other, hence form a double stranded RNA. This prevents translation.

Transcription Unit

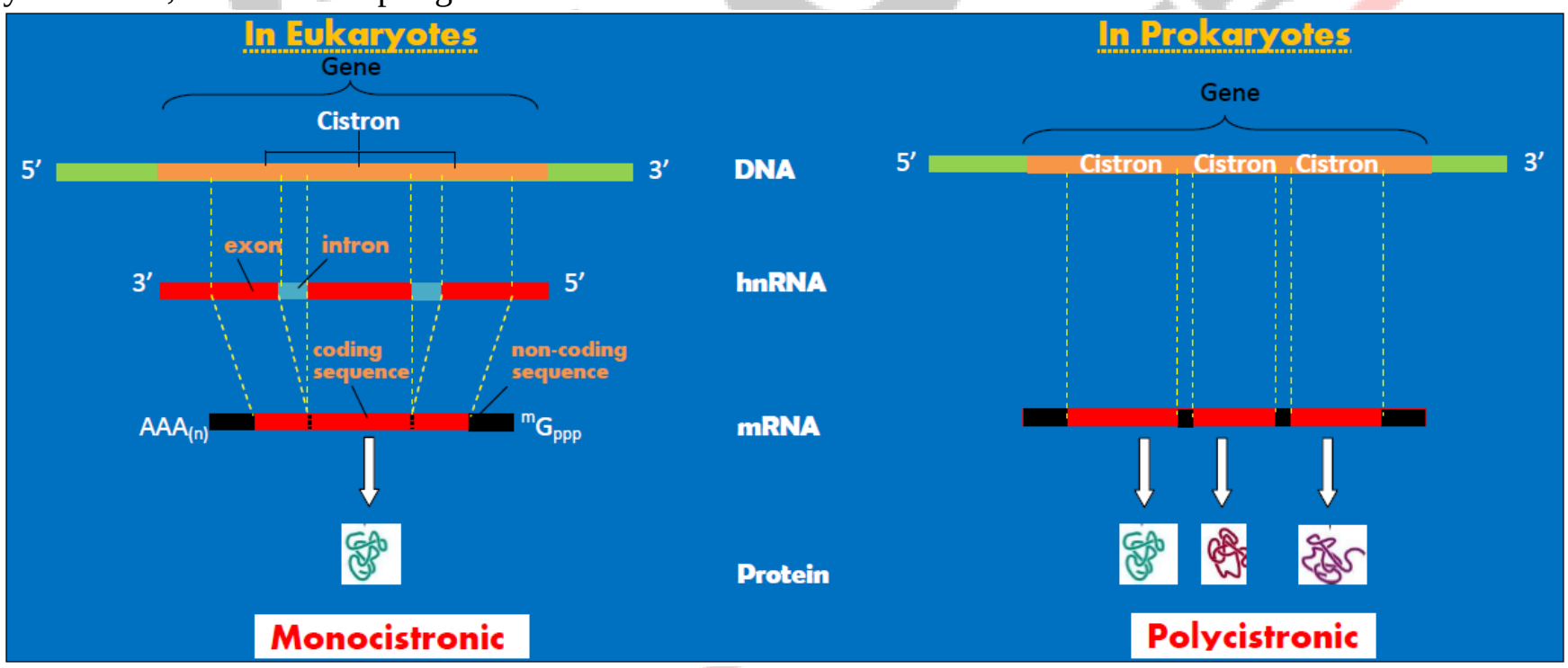
- *It is the segment of DNA which takes part in transcription*. It consists of 3 regions:
  - **A promoter (Transcription start site):** Binding site for RNA polymerase. Located at 5'end (upstream) of coding strand.
  - **Structural gene:** The region of template strand where transcription takes place. It is **monocistronic** in eukaryotes and **polycistronic** in prokayotes.
  - **A terminator:** The site where transcription stops. Located at 3' end (downstream) of coding gene.
- The **DNA- dependent *RNA polymerase*** catalyzes the polymerization only in 5'→3'direction.
- 3'→5' acts as **template strand**. 5'→3' acts as **coding strand/ non-template strand.**



The Gene and Cistron

- **Gene:** Functional unit of inheritance. It is the DNA sequence coding for **RNA molecule**.
- **Cistron:** A segment of DNA coding for a **polypeptide.**

- **Monocistronic structural genes (split genes)**: Cistron codes for a mRNA that specifies a single polypeptide. It is seen in eukaryotes. Here, the **coding sequences (expressed sequences** or **exons)** are interrupted by **introns** (intervening sequences).
- **Polycistronic structural genes**: Cistron codes for a long mRNA that specifies more than one polypeptide. It is seen in prokaryotes. Here, there are no split genes.

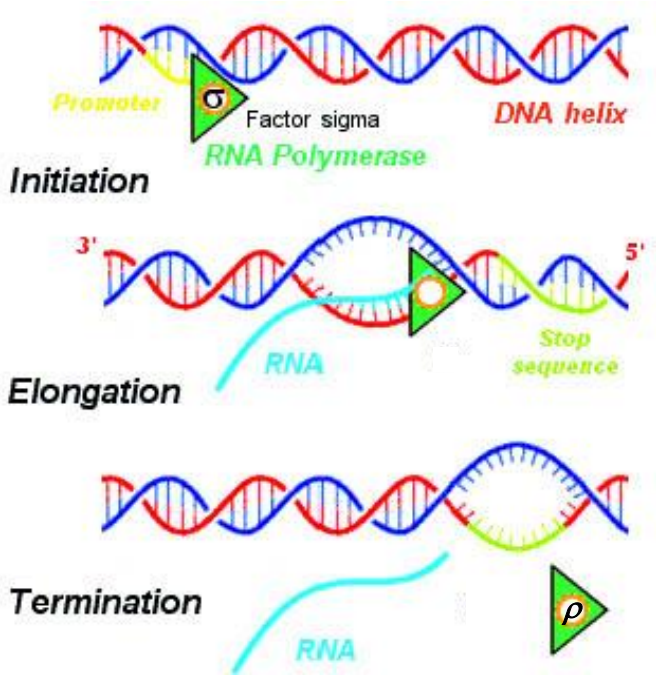


Steps of transcription

In prokaryotes

- **Initiation:** The promoter site of DNA is recognised by ***initiation factor (σ factor)*** of the ***RNA polymerase***. This causes the local unwinding of the DNA double helix.
- **Elongation:** The *RNA polymerase* after initiation of transcription loses the ***σ factor.*** It synthesize RNA chain in the 5'→3' direction by polymerising activated **ribonucleoside triphosphates** (ATP, GTP, UTP & CTP).
- **Termination:** Once the *RNA polymerase* reaches at the terminator region of DNA, a ***termination factor (ρ factor)*** binds to the it and terminates the transcription.

➔ In bacteria (Prokaryotes) transcription and translation can be coupled (Translation can begin before mRNA is fully transcribed) because
- ∵ mRNA requires no processing to become active.
- ∵ Transcription and translation take place in the same compartment (no separation of cytosol and nucleus).



In eukaryotes, there are 2 additional complexities:

1. There are 3 RNA polymerases:

- **RNA polymerase I:** Transcribes rRNAs (28S, 18S & 5.8S).
- **RNA polymerase II:** Transcribes the heterogeneous nuclear RNA (hnRNA). It is the precursor of mRNA.
- **RNA polymerase III:** Transcribes tRNA, 5S rRNA and snRNAs (small nuclear RNAs- helps in processing of rRNA and mRNA in the nucleus).

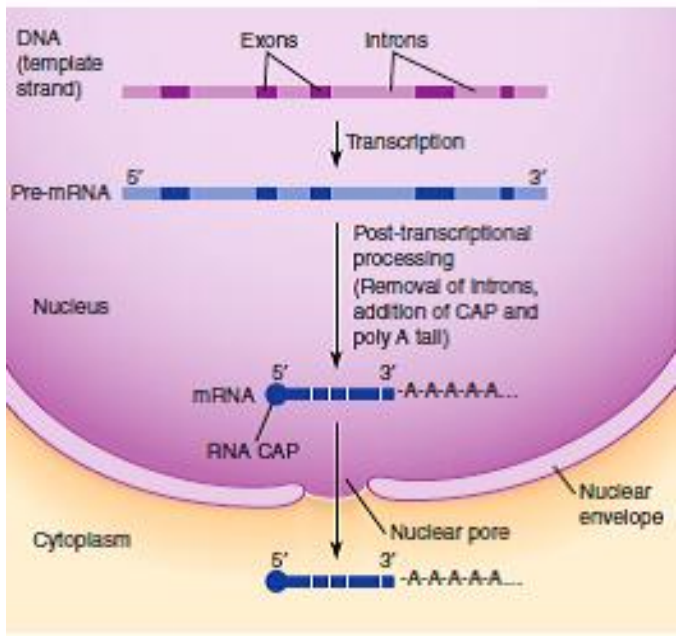
2. The primary transcripts (hnRNA) contain both the exons and introns and are non-functional. Hence introns have to be removed. For this, it undergoes the following processes:

➔ **Capping:** Here, a nucleotide **methyl guanosine triphosphate (cap)** is added to the 5' end of hnRNA.

➔ **Splicing:** From hnRNA introns are removed and exons are spliced (joined) together.

➔ **Tailing (Polyadenylation):** Here, **adenylate residues** (200-300) are added at 3'-end. It is the fully processed hnRNA, now called ***mRNA*** which moves to the ribosome to be translated.

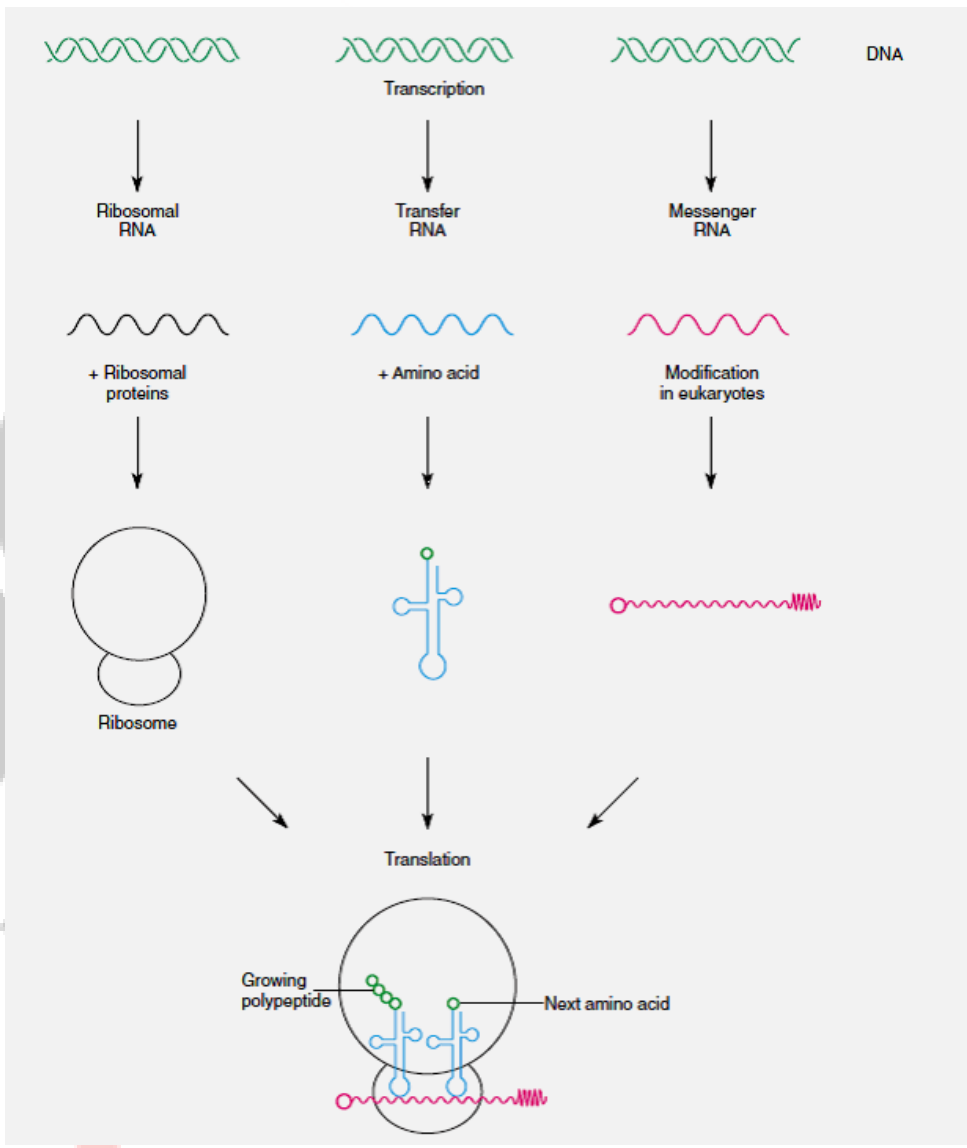
☯ Capping and tailing checks degradation by hydrolytic enzymes in cytoplasm.



TYPES OF RNA

- **mRNA (messenger RNA):** Provide template for translation (protein synthesis).
- **rRNA (ribosomal RNA):** Structural & catalytic role during translation. E.g. 23S rRNA in bacteria acts as ribozyme. (Eukaryotic **Ribosome** consists of larger 60Ssubunit made up of 28S, 5.8S and 5S rRNA's & over 45 different proteins. Smaller 40Ssubunit made up of 18S rRNA and 33 proteins)
- **tRNA (transfer RNA or sRNA or soluble RNA,** being smallest**):** Carries specific amino acids for protein synthesis and reads the genetic code.

Relationship among the 3 types of RNA

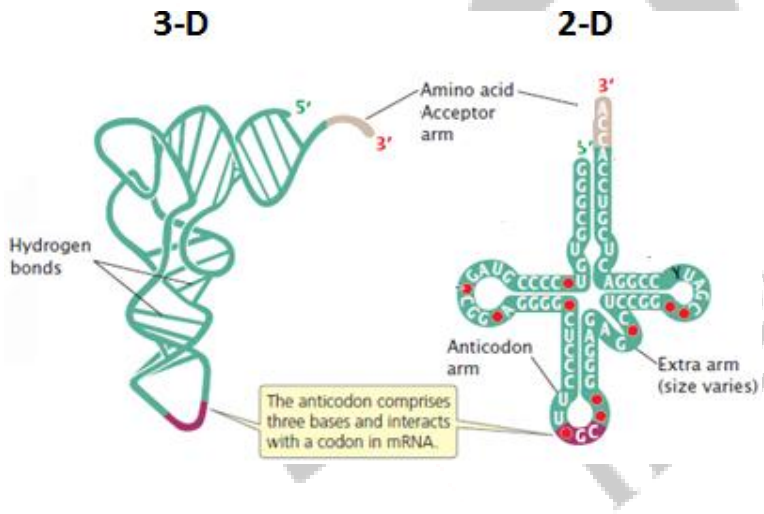


tRNA- the adapter molecule

tRNA has

- **An Anticodon loop** that has bases complementary to the code.
- **An amino acid acceptor end** to which amino acid binds.

- For initiation, there is another tRNA called **initiator tRNA.**
- There are no tRNAs for stop codons.
- **Secondary (2-D)** structure of tRNA looks like a **clover-leaf. 3-D structure** looks like **inverted 'L'.**



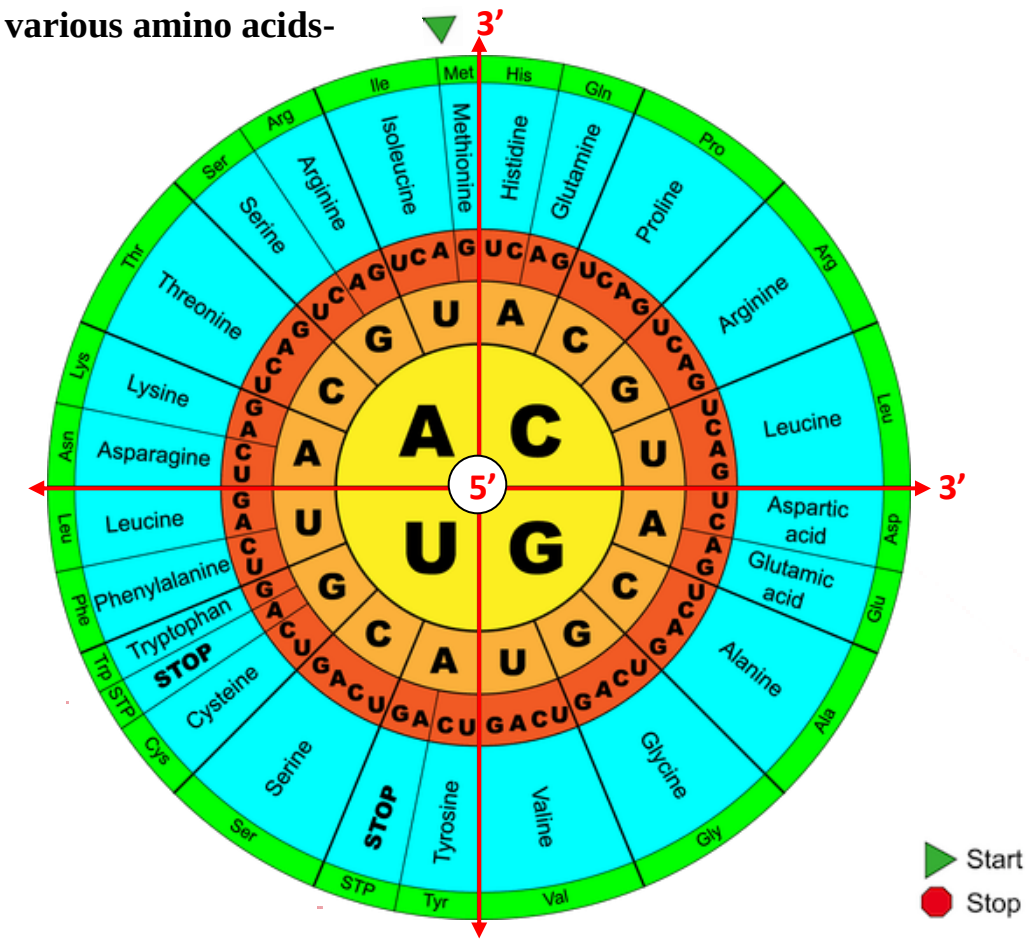
Genetic Code

- *It is the nucleotide sequence in mRNA that specifies the amino acid sequence of protein.*
  20 amino acids involved in translation. In other words, *the 4 letter genetic language of nucleic acids is translated into 20 letter language of the proteins viz. the* ***genetic code****.*

Scientists involved in revealing Genetic Code

- **George Gamow** (1954)**:** Suggested that for coding 20 amino acids, the code should be made up of 3 nucleotides.
- **Har Gobind Khorana:** Developed the chemical method in synthesizing RNA molecules with defined combinations of bases (homopolymers & copolymers).
- **Marshall Nirenberg** (1961)**:** Developed cell-free system for protein synthesis.
- **Severo Ochoa:** ***Polynucleotide phosphorylase*** is used to polymerize RNA with defined sequences in a template independent manner.
- **Frederick Sanger:** Developed method for determination of amino acid sequences in proteins.

The codons for the various amino acids-



Salient features of Genetic Code

- ***Triplet code*:** A codon is a set of 3 nucleotides and it codes for a specific amino acids.
  61 codons code for amino acids. 3 codons (**UAA, UAG** & **UGA)** do not code for any amino acids. They function as **stop codons.**
- ***Nearly Universal***: The same code applies to all animals with exceptions.
  Some exceptions are found in mitochondrial codons (for 6 codons out of 64), and in some protozoans.
- ***Comma less:*** No punctuations b/w adjacent codons. The codon is read in mRNA in a contiguous fashion.
- ***Degeneracy:*** A single amino acid is represented by many codons (except AUG for methionine & UGG for tryptophan).
- ***Unambiguous:*** Genetic code is **specific.** i.e. one codon specifies only one amino acid.
  AUG has dual functions. It codes for Methionine (met), and also acts as **initiator codon.** In eukaryotes, ***methionine*** is the first amino acid and ***formyl methionine*** in prokaryotes.

Translation (Protein Synthesis)

- *It is the process of formation of proteins from mRNA.*
- It takes place in ribosomes. Includes 4 steps –

1. Charging of tRNA (aminoacylation of tRNA)

- Activated amino acids (+ **ATP**-provide energy for the formation of peptide bond) linked to their specific tRNA in the presence of ***aminoacyl tRNA synthetase.*** So the tRNA becomes charged.

2. Initiation

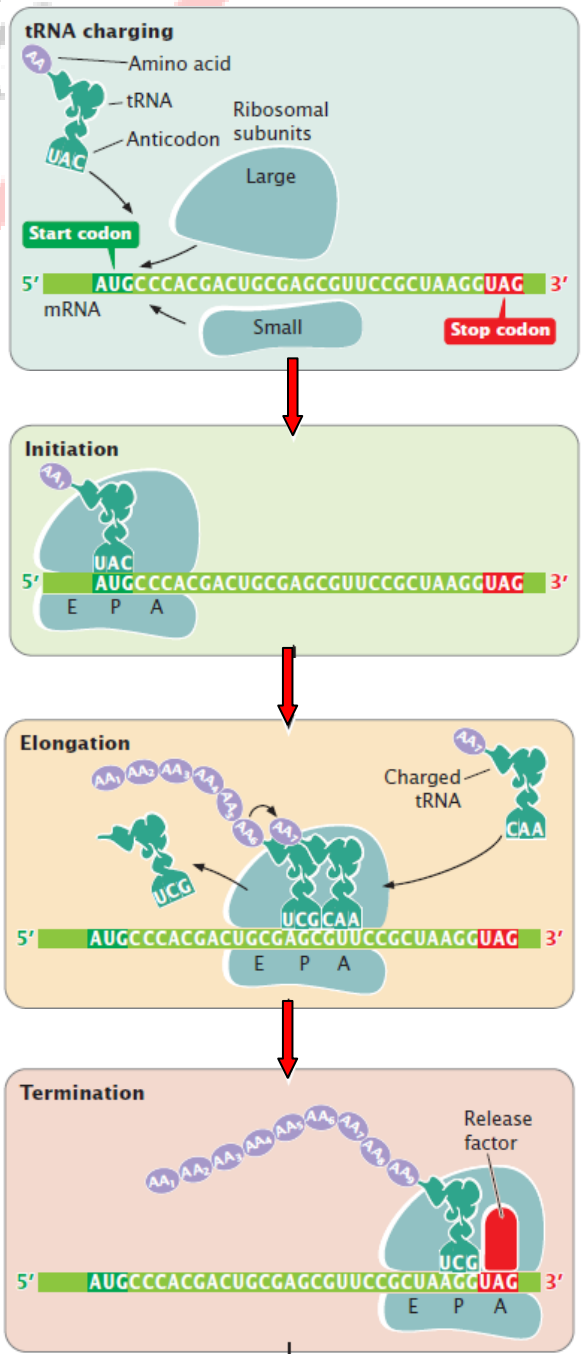
- The**smaller subunit** of ribosome binds at start codon (AUG) of mRNA @ 5'-end**.**
- The charged tRNA with anticodon UAC (carrying **methionine**) pair with AUG of mRNA.
- Then the **larger subunit** binds to smaller subunit to form the **initiation complex**.
  Large subunit has 2 binding sites for tRNA- **aminoacyl tRNA binding site (A-site)** and **peptidyl site (P-site).** Initiator tRNA directly attaches to P-site.

3. Elongation

- A second charged tRNA complex with an appropriate amino acid enters to A-site of the ribosome. Its anticodon binds to the 2nd codon on the mRNA
- A peptide bond is formed between 1st and 2nd amino acids (by ***peptidyl transferase*).**
- 1st amino acid and its tRNA are broken. This tRNA is removed from P-site and second tRNA at the A-site is pulled to P-site along with mRNA. This is called **translocation.**
- Then 3rd codon comes into A-site and a suitable tRNA with 3rd amino acid binds at the A-site. This process is repeated.

4. Termination

- When aminoacyl tRNA reaches the termination codon like**UAA, UAG** & **UGA,** the termination of translation occurs.
  The polypeptide and tRNA are released from the ribosomes.
- The ribosome dissociates into large and small subunits at the end of protein synthesis.



➔ An mRNA has additional sequences that are not translated **(untranslated regions or UTR).** UTRs are present at both 5'-end (before start codon) and 3'-end (after stop codon). They are required for efficient translation process.

Regulation of Gene Expression

- Gene express through the formation of polypeptides.
- Expression of gene should be at right time, at correct amount. Otherwise energy deprivation and metabolic chaos would occur in the cell.
- The metabolic, physiological and environmental conditions regulate expression of genes.

In prokaryotes:-

- *A set of genes regulating a metabolic reaction constitute an* ***Operon.***
  E.g. **lac operon, trp operon, ara operon, his operon, val operon etc.**
- When a substrate is added to growth medium of bacteria, a set of genes is switched on to metabolize it. This is called **induction.**
- When a metabolite (product) is added, the genes to produce it are turned off. This is called **repression.**

Lac operon in *E. coli*:

The operon controlling lactose metabolism. It was first elucidated by **Francois Jacob** and **Jacque Monod** in 1961.
It consists of –

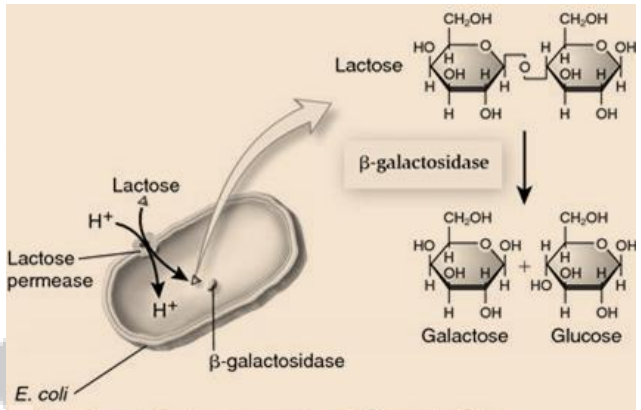
a) **3 structural genes:** *The part of DNA which transcribe mRNA for polypeptide synthesis.*
   - **z gene:** Codes for ***β- galactosidase*** (lactose $\xrightarrow{\beta\text{- gal}}$ galactose + glucose).
   - **y gene:** Codes for ***permease*** (increase permeability of the cell to lactose by active transport. Small amount of permease is present in the plasma membrane every time).
   - **a gene:** Codes for a ***transacetylase.***

b) Promoter gene (P): *It is the site of attachment of RNA polymerase and initiates transcription.*

c) Operator gene (O): *It is the DNA segment which controls the structural gene when the repressor binds. It lies very close to the structural gene.*

d) **A regulatory or inhibitor (*i*) gene:** *Codes for the* ***repressor protein***. It binds to operator and blocks the movement of the enzyme RNA polymerase. As a result, the transcription will be switched off.

e) **Inducer** (Here, Lactose): It is a chemical which can which can bind with repressor and thus prevents the repressor from binding to the operator gene. Inducer keep switch on and allow the structural gene to transcribe mRNA to synthesize the enzymes.



Functions of lactose permease and β-galactosidase

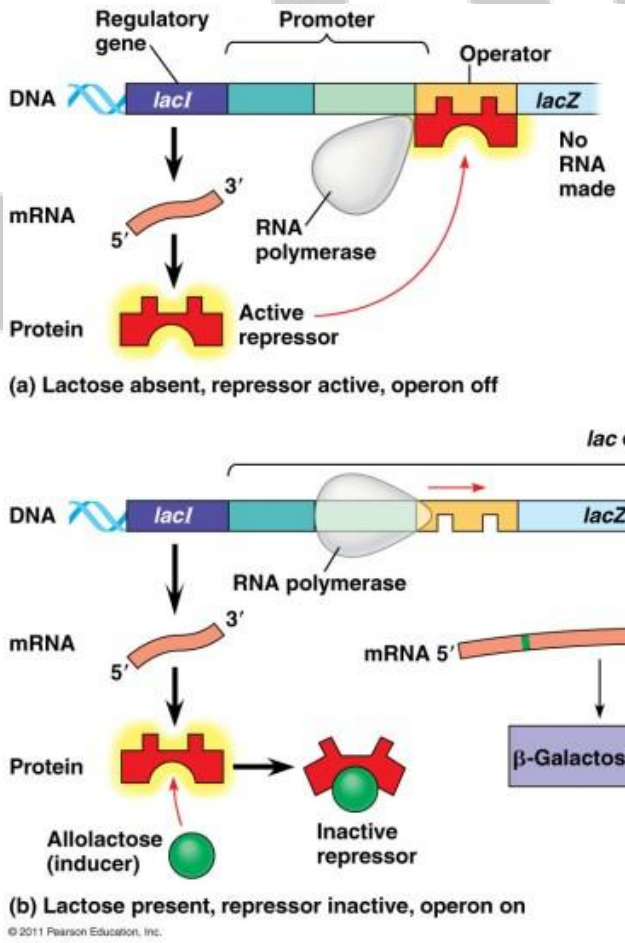
- Functioning of Lac operon:-

When lactose (inducer) is absent:

Step 1. The regulator gene synthesizes mRNA produces **repressor protein**
Step 2. Repressor binds to the **operator region** (blocks RNA polymerase movement)
Step 3. Prevents the transcription of mRNA from structural gene (remains switched off).

When lactose (inducer) is present:

Step 1. Lactose binds to the repressor protein making it inactive.
Step 2. Repressor fails to binds to the operator region
Step 3. The RNA polymerase binds with promoter gene (*lac* operon "switched on") and transcribe structural genes



In eukaryotes:-

The regulation includes the following levels:-

1. Transcriptional level (formation of primary transcript)
2. Processing level (regulation of splicing)
3. Transport of mRNA from nucleus to the cytoplasm
4. Translational level.

HUMAN GENOME PROJECT

- *The entire DNA in the haploid set of chromosome of an organism is called a **Genome**.* In Human genome, DNA is packed in 23 chromosomes.
- **Human Genome Project** (1990-2003) is the first effort in identifying the sequence of nucleotides and mapping of all the genes in human genome.
  Genetic and physical maps on the genome were generated using information on polymorphism of restriction endonuclease recognition sites, and **some repetitive DNA sequences (microsatellites).**
- Coordinated by the **U.S. Department of Energy** and the **National Institute of Health**.

GOALS of HGP

(i) Identify all the approximately 20,000-25,000 genes in human DNA.
(ii) Determine the sequences of the 3 billion chemical base pairs that make up human DNA
(iii) Store this information in databases
(iv) Improve tools for data analysis.

> **Bioinformatics:** Application of computer science and information technology to the field of biology & medicine. Usually applies in analyzing DNA sequence data.

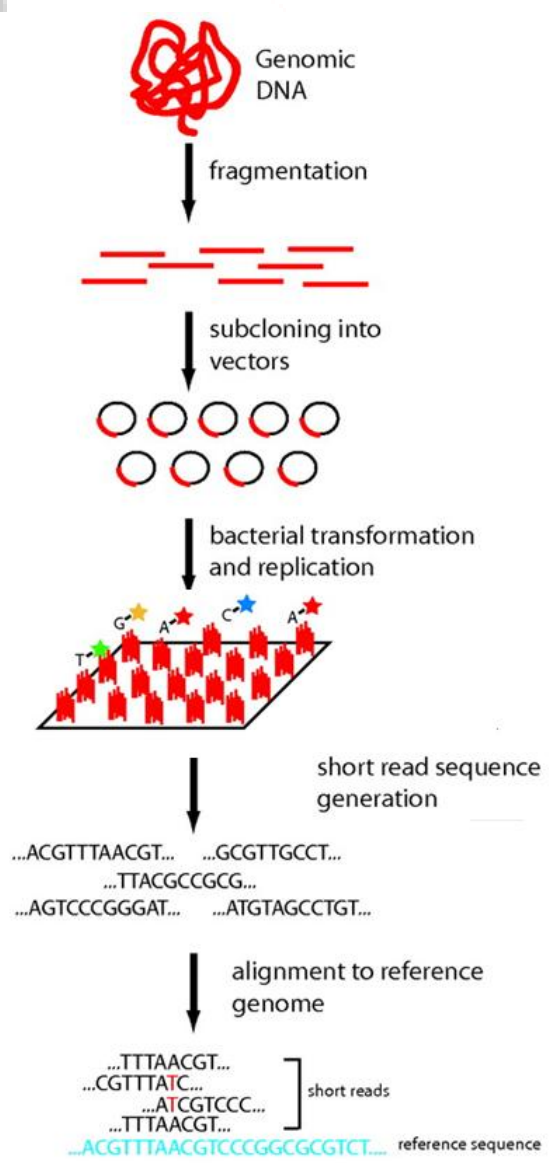
(v) Transfer related technologies to other sectors, such as industries
(vi) Address the **ethical, legal, and social issues (ELSI)** that may arise from the project.

➔ Many non-human model organisms, such as bacteria, yeast, *Caenorhabditis elegans* (a free living non-pathogenic nematode), *Drosophila* (the fruit fly), plants (rice and *Arabidopsis*), etc., have also been sequenced.

Methodologies :

♥ 2 major approaches.

1. **Expressed Sequence Tags** (ESTs)-Identifying all the genes that are expressed as RNA.
2. **Sequence Annotation** -Sequencing the whole set of genome containing all the coding & non-coding sequence, and later assigning different regions in the sequence with functions.
   Procedure: -
   i. The total DNA from a cell is isolated
   ii. Converted into random fragments of relatively smaller sizes
   iii. Cloned in suitable host (e.g.: bacteria and yeast) using specialised vectors (BAC and YAC) for amplification (now by PCR).
   iv. Fragments are sequenced using **automated DNA sequencers** (using Frederick-Sanger method).
   v. These sequences were then arranged based on some overlapping regions present in them.
   vi. Alignment of these sequences using computer programs.
   vii. These sequences were subsequently annotated and were assigned to each chromosome.



Salient Features of Human Genome

i. Human genome contains **3164.7 million** nucleotide bases.
ii. The total number of genes = 30,000
iii. Average gene consists of **3000 bases**, but sizes vary.
The **largest known human gene** (**dystrophin**-the gene responsible for Duchenne Muscular dystrophy present on X-chromosome) contains 2.4 million bases.
The **smallest gene** is the gene responsible for **TDF** (Testis Determining Factor) present on Y-chromosome. It is only 14 bp long.
iv. **99.9 %** nucleotide bases are exactly the same in all people. **0.1%** makes each of us unique.
v. Functions of over **50 %** of the discovered genes are unknown.
vi. Less than **2%** of the genome codes for proteins.
vii. **Repeated sequences** make up very large portion of the human genome.
They have no direct coding functions, but they **shed light** on chromosome structure, dynamics and evolution.
viii. **Chromosome 1** has most genes (2968), and the **Y** has the fewest (231).
ix. About **1.4 million** locations where single-base DNA differences (**SNPs – single nucleotide polymorphism**) occur in humans.

Applications

- To study the effects of DNA variations among individuals helps to diagnose, treat and prevent the disorders in human beings. **SNP's** helps in finding chromosomal locations for disease-associated sequences and tracing human history.
- Provide clues to understand human biology.
- Can learn natural capabilities of non-human organisms and can be applied toward solving challenges in health care, agriculture, energy production, environmental remediation.

DNA Fingerprinting

- *It is the technique to identify the variation in individuals at DNA level.*
- Developed by **Alec Jeffreys (**British geneticist **1985).**

Basis of DNA fingerprinting

- **Repetitive DNA:** DNA carries some non-coding repeated sequences**.**
  Number of repeats is specific from person to person (except in the case of monozygotic twins).
- **Satellite DNA:** These are highly-repeated short sequences in the repetitive DNA.
  Satellite DNA can be separated from bulk genomic DNA as different peaks during density gradient centrifugation. During centrifugation, the **bulk DNA** forms a **major peaks** and the **satellite DNA** forms the small peaks**.**

Satellite DNA can be classified on the basis of following:-

Base composition (A:T rich or G:C rich), Length of segment and Number of repetitive units.

Important types of Satellite DNA

A. Micro-satellites/ SSR(Simple Sequence Repeat)- 5-8 bp long

B. Mini-satellites- 11-60 bp long, e.g.: **VNTR (Variable Number of Tandem Repeats)**

VNTR shows high degree of polymorphism. So, they are used as basis of DNA fingerprinting.

DNA from every tissue such as blood stains, semen stains, skin, bone, saliva, hair roots etc. from an individual show the same degree of polymorphism.

- Mutations accumulate generation after generation and cause polymorphism.

Steps of DNA fingerprinting (Southern Blotting Technique)

a. Isolate DNA from the cell.

b. If DNA is in poor condition, make copies **(amplification)** by **polymerase chain reaction (PCR).**

c. Digest DNA by **restriction endonucleases.**

d. Separate DNA fragments by **gel electrophoresis.**

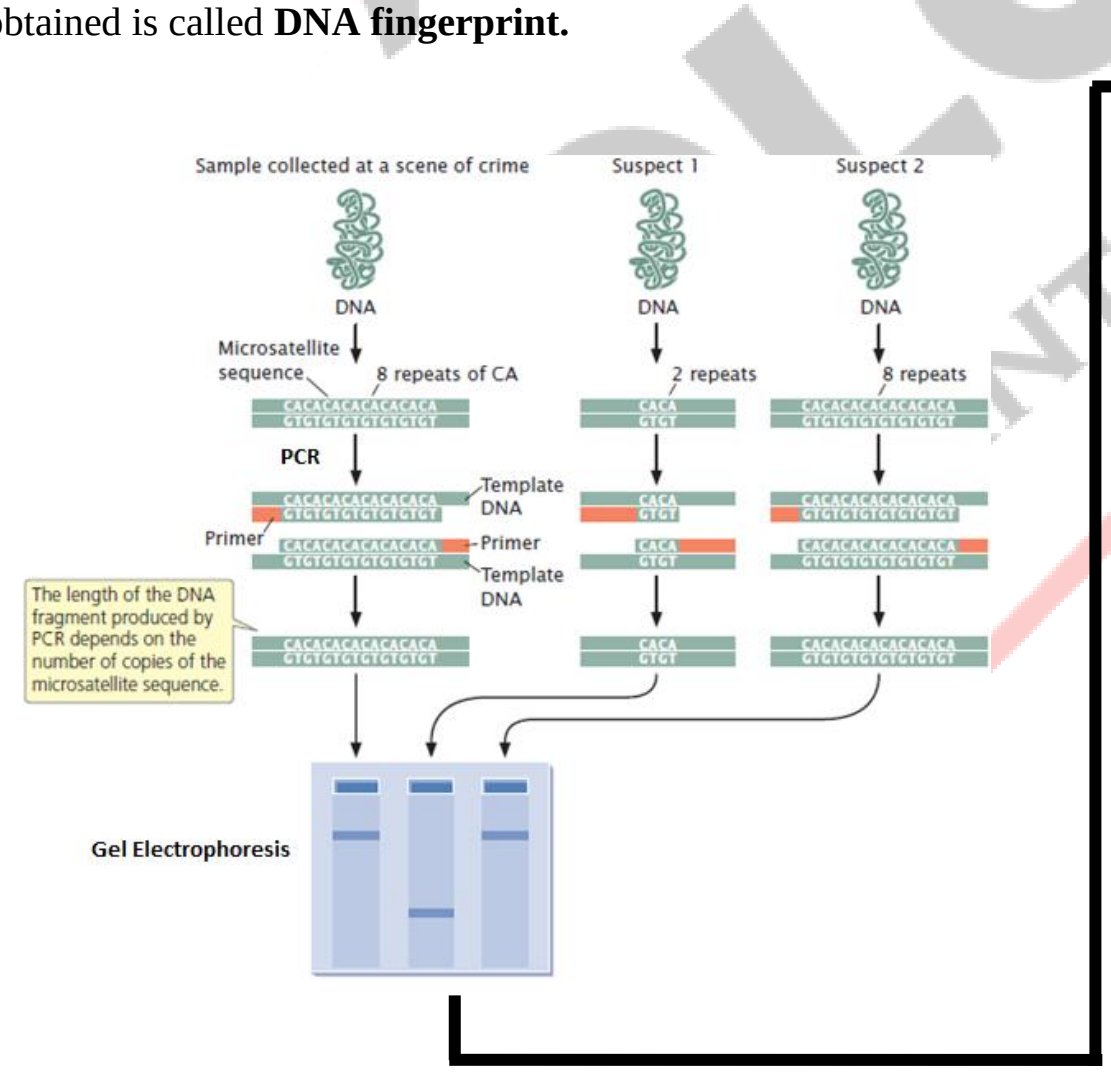
e. Treat with **alkali solution (NaOH)** to denature DNA bonds in the gel into single strands.

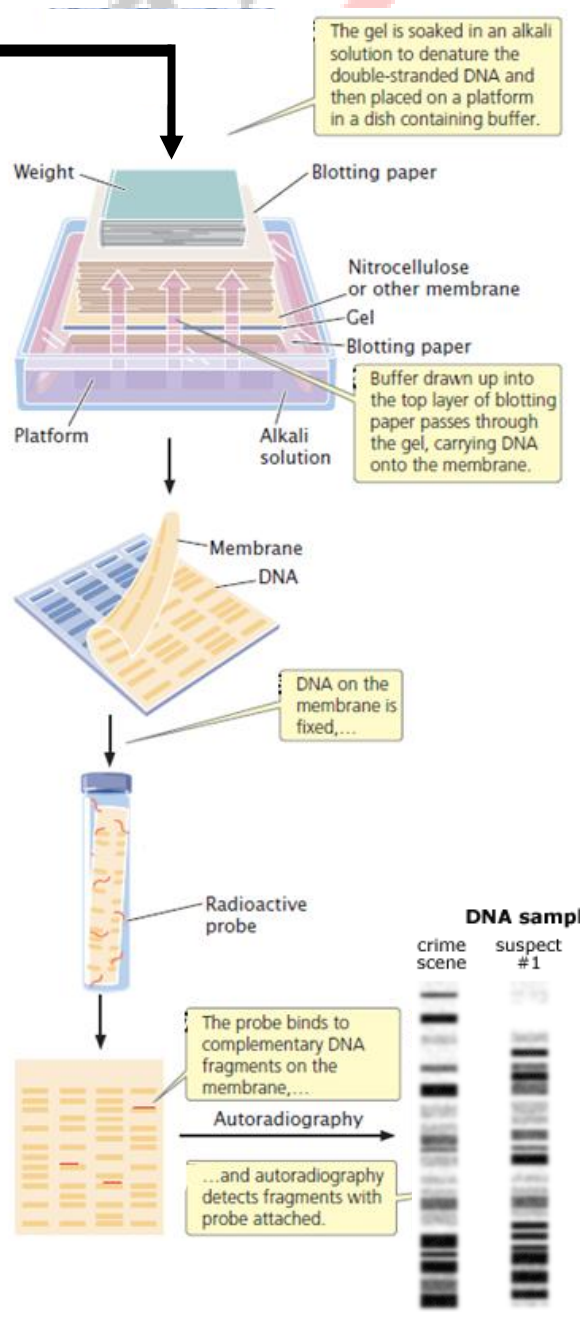
f. Transfer **(blotting)** single stranded DNA fragments to synthetic membranes such as **nitrocellulose** or **nylon**, and then baked in a vacuum oven at 80^{o}C for 3-5 hours (to fix the DNA fragment on the membrane).

g. Nitrocellulose filter paper is placed in a solution containing radioactive labelled single stranded **DNA probe** (contain repeated sequences of nucleotides complementary to those on VNTRs). The DNA probe binds with the complimentary sequences of the DNA fragment on the membrane to form a **hybridized DNA.**

h. The filter paper is washed to remove unbound probe.

i. The hybridized DNA is photographed on to an X-ray film by **autoradiography.** The image (in the form of dark & light bands) obtained is called **DNA fingerprint.**





Applications of DNA fingerprinting

- ✓ **Forensic tool** to solve paternity, rape, murder etc.
- ✓ For the diagnosis of **genetic diseases.**
- ✓ To determine **phylogenetic status** of animals. (For evolution & speciation, polymorphisms play important role).