36. MOLECULAR BASIS OF INHERITANCE

- Nucleic acids (DNA & RNA) are the building blocks of genetic material.
- DNA is the genetic material in most of the organisms.
- RNA is the genetic material in some viruses. RNA mostly functions as **messengers.**

THE DNA

STRUCTURE OF POLYNUCLEOTIDE CHAIN

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

1. **A nitrogenous base.**
2. **A pentose sugar (ribose** in **RNA & deoxyribose** in **DNA).**
3. **A phosphate group.**

Nitrogen bases are 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.

A nitrogenous base is linked to the OH of 1' C pentose sugar through an **N-glycosidic linkage** to form **nucleoside.**

Nucleosides in RNA	Nucleosides in DNA
Adenosine	Deoxyadenosine
Guanosine	Deoxyguanosine
Cytidine	Deoxycytidine
Uridine	Deoxythymidine

A phosphate group is linked to OH of 5' C of a nucleoside through **phosphoester linkage** to form **nucleotide** (or deoxynucleotide).

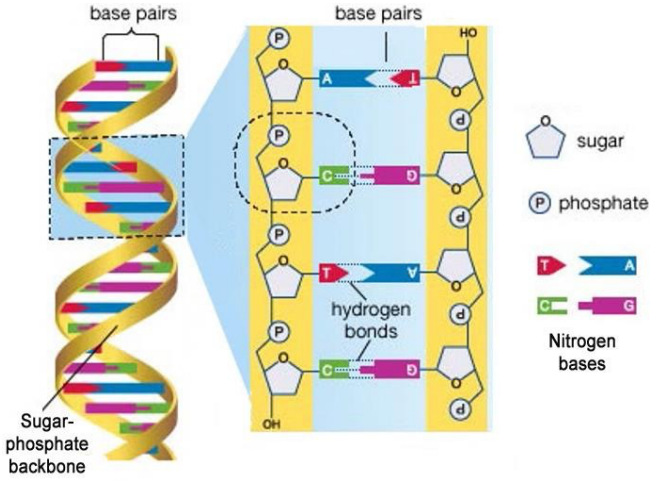
In RNA, each nucleotide has an additional –OH group at 2′ C of the ribose (2'- OH).

2 nucleotides are linked through **3'-5' phosphodiester bond** to form **dinucleotide.**

When more nucleotides are linked, it forms **polynucleotide.**

STRUCTURE OF THE DNA

- **Friedrich Meischer (1869):** Identified DNA and named it as **'Nuclein'.**
- **James Watson** & **Francis Crick (1953)** proposed **double helix model of DNA.** It was based on X-ray diffraction data produced by **Maurice Wilkins** & **Rosalind Franklin.**



Pitch of helix= 3.4 nm (34 Å)
Number of base pairs in each turn= 10
Distance between adjacent base pairs= 0.34 nm (3.4 Å)

- DNA is made of 2 polynucleotide chains coiled in a right-handed fashion. Its backbone is formed of sugar & phosphates. The bases project inside.
- The 2 chains have **anti-parallel polarity**, i.e. one chain has the polarity **5'→3'** and the other has **3'→5'.**
- The bases in 2 strands are paired through **H-bonds** forming **base pairs (bp).**

  A=T (2 hydrogen bonds) C≡G (3 hydrogen bonds)
- Purine comes opposite to a pyrimidine. This generates uniform distance between the 2 strands.
- **Erwin Chargaff's rule:** In DNA, the proportion of A is equal to T and the proportion of G is equal to C.

  ∴ **[A] + [G] = [T] + [C]**

  or [A] + [G] / [T] + [C] =1

- **Φ 174 (a bacteriophage)** has 5386 nucleotides.
- **Bacteriophage lambda** has **48502 base pairs (bp).**
- ***E. coli*** has **4.6x10^6 bp.**
- Haploid content of human DNA is **3.3x10^9 bp.**

Length of DNA = number of base pairs X distance between two adjacent base pairs.

Number of base pairs in human = 6.6 x 10^9

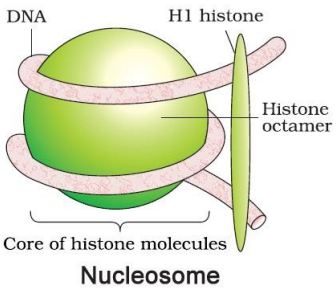
Hence, the length of DNA = 6.6 x10^9 x 0.34x 10^{-9} = 2.2 m

In *E. coli*, length of DNA =1.36 mm (1.36 x 10^{-3} m)

∴ The number of base pairs $= \frac{1.36 \times 10^{-3}}{0.34 \times 10^{-9}}$ = 4 x 10^6 bp

PACKAGING OF DNA HELIX

- In prokaryotes (E.g. *E. coli*), the DNA is not scattered throughout the cell. DNA is negatively charged. So it is held with some positively charged proteins to form **nucleoid.**
- In eukaryotes, there is a set of positively charged, basic proteins called **histones.**
- Histones are rich in positively charged basic amino acid residues ***lysines*** and ***arginines.***
- **8 histones** form **histone octamer.**
- Negatively charged DNA is wrapped around histone octamer to give **nucleosome.**



Nucleosome

- A typical nucleosome contains **200 bp.**

  Therefore, total number of nucleosomes in human = $\frac{6.6 \times 10^9 \text{bp}}{200} = 3.3 \times 10^7$
- Nucleosomes constitute the repeating unit to form **chromatin.** Chromatin is the thread-like stained bodies.

- Nucleosomes in chromatin = ***'beads-on-string'.***
- Chromatin is packaged → chromatin fibres → coiled and condensed at metaphase stage → chromosomes.
- Higher level packaging of chromatin requires **non-histone chromosomal (NHC) proteins.**
- Chromatin has 2 forms:
  - **Euchromatin:** Loosely packed and transcriptionally active region of chromatin. It stains light.
  - **Heterochromatin:** Densely packed and inactive region of chromatin. It stains dark.

THE SEARCH FOR GENETIC MATERIAL

Griffith's Transforming Principle experiment (1928)

Frederick Griffith used **mice** & ***Streptococcus pneumoniae.*** ***Streptococcus pneumoniae*** has 2 strains**:**

- **Smooth (S) strain (Virulent):** Has polysaccharide mucus coat. Cause pneumonia.
- **Rough (R) strain (Non-virulent):** No mucus coat. Do not cause Pneumonia.

Experiment:

- S-strain → Inject into mice → Mice die
- R-strain → Inject into mice → Mice live
- S-strain (Heat killed) → Inject into mice → Mice live
- S-strain (Hk) + R-strain (live) → Inject into mice → Mice die

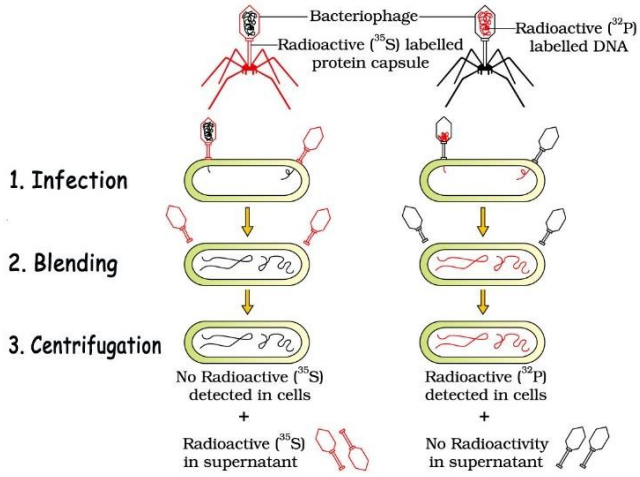
He concluded that some **'transforming principle'** transferred from heat-killed S-strain to R-strain. It enabled R-strain to synthesize smooth polysaccharide coat and become virulent. This must be due to the transfer of genetic material.

2. Biochemical characterization of transforming principle (1933-44)

- **Oswald Avery, Colin MacLeod & Maclyn McCarty** worked to determine the biochemical nature of **'transforming principle'** in Griffith's experiment.
- They purified biochemicals (proteins, DNA, RNA etc.) from heat killed S cells using suitable enzymes.
- They discovered that
  - Digestion of protein and RNA (using *Proteases* and *RNases*) did not affect transformation. It means that the transforming substance was not a protein or RNA.
  - Digestion of DNA with *DNase* inhibited transformation. It means that DNA caused transformation of R cells to S cells. It proves that DNA was the transforming principle.

3. Hershey-Chase Experiment (Blender Experiment)-1952

- **Hershey & Chase** grew some bacteriophage viruses on a medium containing radioactive phosphorus (**P^{32}**) and some others on medium containing radioactive sulphur (**S^{35}**).
- Viruses grown in P^{32} got **radioactive DNA** because only DNA contains phosphorus. Viruses grown in S^{35} got **radioactive protein** because protein contains sulphur.
- These preparations were used separately to infect ***E. coli.***
- After infection, the *E. coli* cells were gently agitated in a blender to remove the virus particles from the bacteria.
- Then the culture was centrifuged to separate lighter virus particles from heavier bacterial cells.
- Bacteria infected with viruses having radioactive DNA were radioactive. i.e., DNA had passed from the virus to bacteria. Bacteria infected with viruses having radioactive proteins were not radioactive. i.e., proteins did not enter the bacteria from the viruses. This proves that DNA is the genetic material.



PROPERTIES OF GENETIC MATERIAL (DNA v/s RNA)

A genetic material must have the following properties:

- Ability to generate its replica (Replication).
- Chemical and structural stability.
- Provide the mutations that are required for evolution.
- Ability to express as Mendelian Characters.

Reasons for stability (less reactivity) of DNA	Reasons for mutability (high reactivity) of RNA
Double stranded	Single stranded
Presence of thymine	Presence of Uracil
Absence of 2'-OH in sugar	Presence of 2'-OH in sugar

- RNA is unstable. So, RNA viruses (E.g. *Q.B bacteriophage, Tobacco Mosaic Virus* etc.) mutate and evolve faster.
- DNA strands are complementary. On heating, they separate. In appropriate conditions, they come together. In Griffith's experiment, some properties of DNA of the heat killed bacteria did not destroy. It indicates the stability of DNA.
- For the storage of genetic information, DNA is better due to its stability. But for the transmission of genetic information, RNA is better.
- RNA can directly code for the protein synthesis, hence can easily express the characters. DNA is dependent on RNA for protein synthesis.

RNA WORLD

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

CENTRAL DOGMA OF MOLECULAR BIOLOGY

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

Central dogma

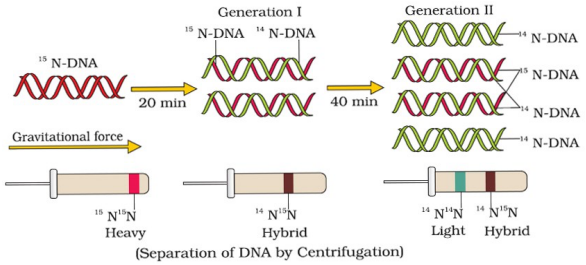
- In some viruses, flow of information is in reverse direction (from RNA to DNA). It is called **reverse transcription.**

DNA REPLICATION

- Replication is the copying of DNA from parental DNA.
- **Watson & Crick** proposed **Semi-conservative model** of replication. It suggests that the parental DNA strands act as **template** for the synthesis of new complementary strands. After replication, each DNA molecule would have one parental and one new strand.
- **Matthew Messelson & Franklin Stahl** (1958) experimentally proved Semi-conservative model.

Messelson & Stahl's Experiment

- They grew ***E. coli*** in **$^{15}NH_4Cl$ medium** (^{15}N = heavy isotope of nitrogen) as the only nitrogen source. As a result, ^{15}N was incorporated into newly synthesised DNA **(heavy DNA or ^{15}N DNA).**
- Heavy DNA can be distinguished from normal DNA (light DNA or ^{14}N DNA) by centrifugation in a **cesium chloride (CsCl) density gradient.**
- *E. coli* cells from ^{15}N medium were transferred to $^{14}NH_4Cl$ medium. After one generation (i.e. after 20 minutes), they isolated and centrifuged the DNA. Its **density** was **intermediate (hybrid)** between ^{15}N DNA and ^{14}N DNA. This shows that in newly formed DNA, one strand is old **(^{15}N type)** and one strand is new **(^{14}N type).** This confirms semi-conservative replication.
- After II generation (i.e. after 40 minutes), there was equal amounts of hybrid DNA and light DNA.

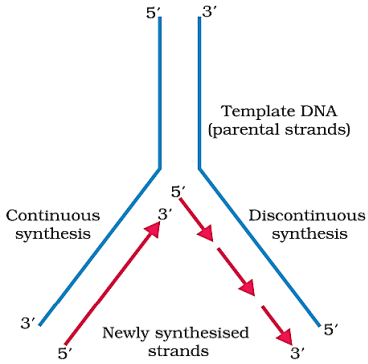


(Separation of DNA by Centrifugation)

Taylor & colleagues (1958) performed similar experiments on ***Vicia faba*** (faba beans) using **radioactive thymidine** to detect distribution of newly synthesized DNA in the chromosomes. It proved that the DNA in chromosomes also replicate semi-conservatively.

The Machinery and Enzymes for Replication

- DNA replication starts at a point called ***origin* (*ori*).**
- A unit of replication with one origin is called a ***replicon.***
- During replication, the 2 strands unwind and separate by breaking H-bonds in presence of an enzyme, ***Helicase.***
- Unwinding of the DNA molecule at a point forms a 'Y'-shaped structure called **replication fork.**
- The separated strands act as **templates** for the synthesis of new strands.
- DNA replicates in the **5'→3' direction.**



- ***Deoxyribonucleoside triphosphates*** (dATP, dGTP, dCTP & dTTP) act as substrate and provide energy for polymerization.
- Firstly, a small **RNA primer** is synthesized in presence of an enzyme, ***primase***.
- In presence of an enzyme, DNA dependent ***DNA polymerase*,** many nucleotides join with one another to primer strand and form a polynucleotide chain (new strand).
- During replication, one strand is formed as a continuous stretch in 5'→ 3' direction **(Continuous synthesis).** This strand is called **leading strand.**
- The other strand is formed in small stretches **(Okazaki fragments)** in 5'→ 3' direction **(Discontinuous synthesis).**
- The Okazaki fragments are then joined together to form a new strand by an enzyme, ***DNA ligase***. This new strand is called **lagging strand.**
- If a wrong base is introduced in the new strand, DNA polymerase can do **proof reading.**
- *E. coli* completes replication within 18 minutes. i.e. 2000 bp per second.
- In eukaryotes, the replication of DNA takes place at **S-phase** of the cell cycle. Failure in cell division after DNA replication results in **polyploidy.**

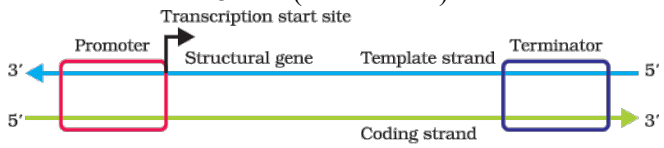
TRANSCRIPTION

- It is the process of copying genetic information from one strand of the **DNA into RNA.**
- Here, adenine pairs with uracil instead of thymine.
- The **DNA- dependent *RNA polymerase*** catalyzes the polymerization only in 5'→3'direction.
- 3'→5' acts as **template strand**. RNA is built from this.
- 5'→3' acts as **coding strand.** This is copied to RNA.
  3'-ATGCATGCATGCATGCATGCATGC-5' template strand.
  5'-TACGTACGTACGTACGTACGTACG-3' coding strand.
- During transcription, both strands are not copied 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. It forms a **double stranded RNA** and prevents translation.

Transcription Unit

- It is the segment of DNA between the sites of initiation and termination of transcription. It consists of 3 regions:

- **A promoter:** Binding site for *RNA polymerase.* Located towards 5'-end (upstream).
- **Structural gene:** The region between promoter and terminator where transcription takes place.
- **A terminator:** The site where transcription stops. Located towards 3'-end (downstream).



Transcription unit and gene

Gene is a functional unit of inheritance. It is the DNA sequence coding for an RNA (mRNA, rRNA or tRNA).

Cistron is a segment of DNA coding for a **polypeptide** during protein synthesis. It is the largest element of a gene.

Structural gene in a transcription unit is 2 types:

- **Monocistronic structural genes (split genes)**: It is seen in eukaryotes. Here, coding sequences ***(exons or expressed sequences)*** are interrupted by ***introns*** (intervening sequences). Exons appear in processed mRNA. Introns do not appear in processed mRNA.
- **Polycistronic structural genes**: It is seen in prokaryotes. Here, there are no split genes.

Transcription in prokaryotes

In bacteria (Prokaryotes), synthesis of all types of RNA are catalysed by a single *RNA polymerase.* It has 3 steps:

- **Initiation:** Here, the enzyme ***RNA polymerase*** binds at the promoter site of DNA. This causes the local unwinding of the DNA double helix. An ***initiation factor (σ factor)*** present in *RNA polymerase* initiates the RNA synthesis.
- **Elongation:** RNA chain is synthesized in 5'-3' direction. In this process, activated **ribonucleoside triphosphates** (ATP, GTP, UTP & CTP) are added. This is complementary to the base sequence in the DNA template.
- **Termination:** A ***termination factor (ρ factor)*** binds to the *RNA polymerase* and terminates the transcription.

In bacteria, transcription and translation can be coupled (translation begins before mRNA is fully transcribed) because

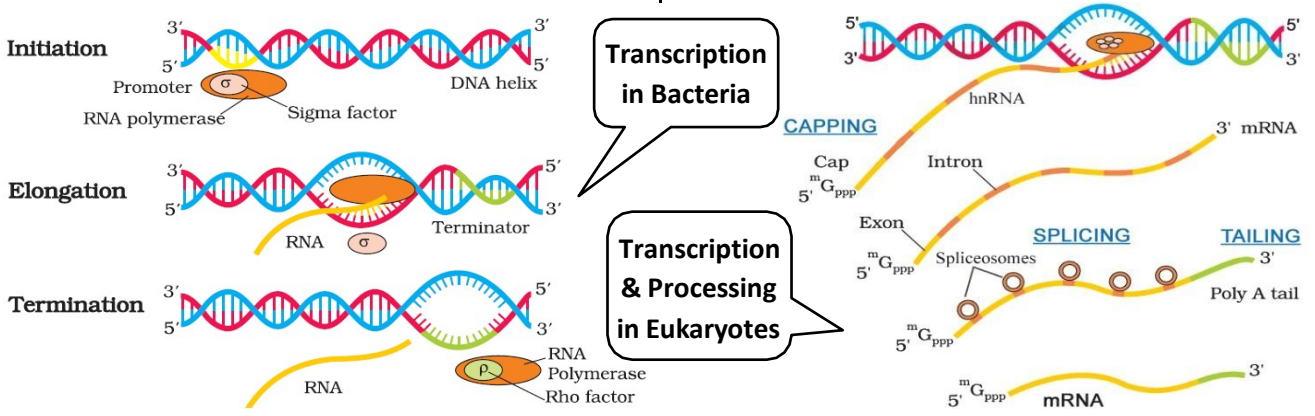
- mRNA requires no processing to become active.
- Transcription and translation take place in the same compartment (no separation of cytosol and nucleus).

Transcription in eukaryotes

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).
2. **The primary transcripts (hnRNA)** contain exons and introns and are non-functional. Hence introns must be removed. For this, it undergoes the following processes:
   - **Splicing:** From hnRNA, introns are removed (by the spliceosome) and exons are spliced (joined) together.
   - **Capping:** Here, a nucleotide **methyl guanosine triphosphate (cap)** is added to the 5' end of hnRNA.
   - **Tailing (Polyadenylation):** Here, **adenylate residues** (200-300) are added at 3'-end.

   Now, it is the fully processed hnRNA, called **mRNA.**



GENETIC CODE

- It is the sequence of nucleotides (nitrogen bases) in mRNA that contains information for protein synthesis (translation).
- The sequence of 3 bases determining a single amino acid is called **codon.**
- **George Gamow** suggested that for coding 20 amino acids, the code should be made up of 3 nucleotides. Thus, there are 64 codons ($4^3 = 4 \times 4 \times 4$).
- **Har Gobind Khorana** developed the chemical method in synthesizing RNA molecules with defined combinations of bases (homopolymers & copolymers).
- **Marshall Nirenberg** developed cell-free system for protein synthesis.
- **Severo Ochoa (*polynucleotide phosphorylase*) enzyme** is used to polymerize RNA with defined sequences in a template independent manner.

20 types of amino acids involved in translation

1. Alanine (Ala)
2. Arginine (Arg)
3. Asparagine (Asn)
4. Aspartic acid (Asp)
5. Cystein (Cys)
6. Glutamine (Gln)
7. Glutamic acid (Glu)
8. Glycine (Gly)
9. Histidine (His)
10. Isoleucine (Ile)
11. Leucine (Leu)
12. Lysine (Lys)
13. Methionine (Met)
14. Phenyl alanine (Phe)
15. Proline (Pro)
16. Serine (Ser)
17. Threonine (Thr)
18. Tryptophan (Trp)
19. Tyrosine (Tyr)
20. Valine (Val)

The codons for various amino acids

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

Salient features of genetic code

- **Codon is triplet** (three-letter code).
- 61 codons code for amino acids. 3 codons (**UAA, UAG & UGA)** do not code for any amino acids. They act as **stop codons (Termination codons or non-sense codons).**
- Genetic code is **universal.** E.g. From bacteria to human UUU codes for Phenylalanine. Some exceptions are found in mitochondrial codons, and in some protozoans.
- **No punctuations** b/w adjacent codons (comma less code). The codon is read in mRNA in a contiguous fashion.
- Genetic code is **non-overlapping.**
- An amino acid is coded by more than one codon (except AUG for methionine & UGG for tryptophan). Such codons are called **degenerate codons.**
- Genetic code is **unambiguous** and **specific.** i.e. one codon specifies only one amino acid.
- AUG has dual functions. It codes for Methionine and acts as **initiator codon.** In eukaryotes, ***methionine*** is the first amino acid and ***formyl methionine*** in prokaryotes.

Mutations and Genetic Code

- Relationship between genes & DNA are best understood by mutation studies. Deletions & rearrangements in a DNA may cause loss or gain of a gene and so a function.
- **Insertion or deletion** of **one or two bases** changes the reading frame from the point of insertion or deletion. It is called **frame-shift insertion or deletion mutations.**
- Insertion/ deletion of **three or its multiple bases** insert or delete one or multiple codon. The reading frame remains unaltered from that point onwards. Hence one or multiple amino acids are inserted /deleted.
- It proves that codon is a triplet and is read contiguously.

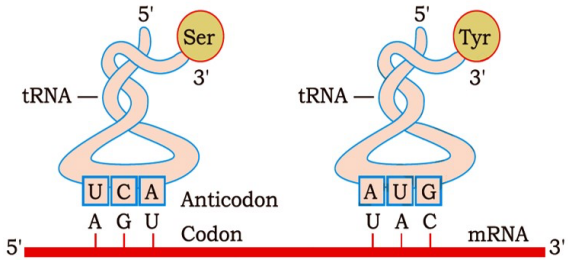
TYPES OF RNA

- **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.
- **tRNA (transfer RNA or sRNA or soluble RNA):** Brings amino acids for protein synthesis and reads the genetic code.

Francis Crick postulated presence of an adapter molecule that can read the code and to link with amino acids.

tRNA is called **adapter molecule** because it has

- **An Anticodon (NODOC) loop** that has bases complementary to the codon.
- **An amino acid acceptor end** to which amino acid binds.
- **Ribosome binding loop.**
- **Enzyme binding loop.**



- 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'.**

TRANSLATION (PROTEIN SYNTHESIS)

- It is the process of **polymerisation of amino acids** to form a polypeptide based on the sequence of codons in mRNA.
- It takes place in **ribosomes.** Ribosome consists of structural RNAs and about 80 types of proteins.
- Ribosome also acts as a **catalyst** (23S rRNA in bacteria is the enzyme- ribozyme) for the formation of peptide bond (***peptidyl transferase*** enzyme in large subunit of ribosome).
- Translation includes 4 steps:

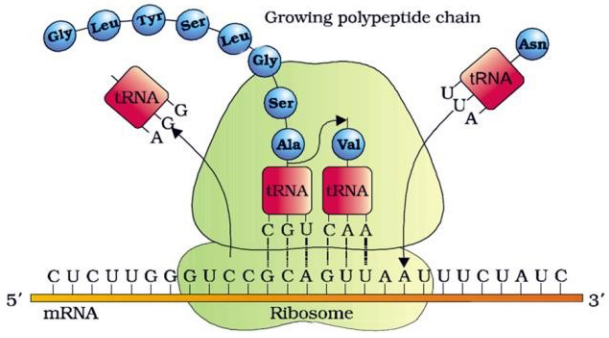
1. **Charging of tRNA**
2. **Initiation**
3. **Elongation**
4. **Termination**

1. Charging (aminoacylation) of tRNA

- Formation of peptide bond needs energy obtained from ATP.
- For this, amino acids are activated (amino acid + ATP) and linked to their cognate tRNA in presence of ***aminoacyl tRNA synthetase.*** Thus, the tRNA becomes charged.

2. Initiation

- In this, small subunit of ribosome binds to mRNA at the **start codon (AUG).**
- Now large subunit binds to small subunit to form **initiation complex.**
- Large subunit consists of **aminoacyl tRNA binding site (A site)** and **peptidyl site (P site).**
- The **initiator tRNA** (which carries methionine) binds on P site. Its **anticodon (UAC)** recognises start codon **AUG.**



3. Elongation

- Second aminoacyl tRNA binds to the A site of ribosome. Its anticodon binds to the second codon on the mRNA and

a **peptide bond** is formed between first and second amino acids in presence of ***peptidyl transferase.***

- First amino acid and its tRNA are broken. This tRNA is removed from P site and second tRNA from A site is pulled to P site along with mRNA. This is called **translocation.**
- These processes are repeated for other codons in mRNA.
- During translation, ribosome moves from codon to codon.

4. Termination

- When a **release factor** binds to stop codon, the translation terminates.
- The polypeptide and tRNA are released from the ribosomes.
- The ribosome dissociates into large and small subunits.

A group of ribosomes associated with a single mRNA for translation is called a **polyribosome (polysomes).**

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

In eukaryotes, gene expression occurs by following levels:

1. **Transcriptional** level (formation of primary transcript).
2. **Processing** level (splicing, capping etc.).
3. **Transport** of mRNA from nucleus to the cytoplasm.
4. **Translational** level (formation of a polypeptide).

The metabolic, physiological and environmental conditions regulate gene expression. E.g.

- In *E. coli,* the ***beta-galactosidase*** enzyme hydrolyses lactose into galactose & glucose. In the absence of lactose, the synthesis of *beta-galactosidase* stops.
- The development and differentiation of embryo into adult are a result of the expression of several set of genes.

If a substrate is added to growth medium of bacteria, a set of genes is switched on to metabolize it. It is called **induction.**

When a metabolite (product) is added, the genes to produce it are turned off. This is called **repression.**

OPERON CONCEPT

- ***"Each metabolic reaction is controlled by a set of genes"***
- All the genes regulating a metabolic reaction constitute an ***Operon.*** E.g. **lac operon, trp operon, ara operon, his operon, val operon etc.**

Lac Operon in *E. coli*

- The operon controlling lactose metabolism.
- It is proposed by **Francois Jacob & Jacque Monod.**

It consists of

a) A regulatory or inhibitor (i) gene: Codes for repressor protein.

b) 3 structural genes:

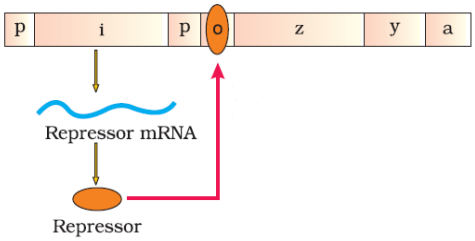
i. z gene: Codes for β ***galactosidase.*** It hydrolyses lactose to galactose and glucose.

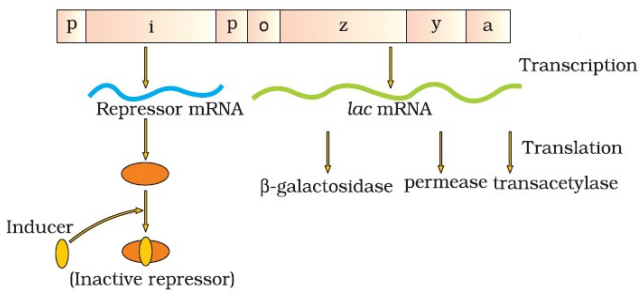
ii. y gene: Codes for ***permease.*** It increases permeability of the cell to β-galactosides (lactose).

iii. **a gene:** Codes for a ***transacetylase.***

- Genes in the operon function together in the same or related metabolic pathway.
- If there is no **lactose (inducer),** lac operon remains switched off. The regulator gene synthesizes mRNA to produce **repressor protein.** This protein binds to the **operator** region and blocks *RNA polymerase* movement. So the structural genes are not expressed.
- If lactose or allolactose is provided in the growth medium, it is transported into ***E. coli*** cells by the action of ***permease.*** Lactose (inducer) binds with repressor protein. So repressor protein cannot bind to operator region. The operator region becomes free and induces the *RNA polymerase* to bind with **promoter.** Then transcription starts.
- Regulation of lac operon by repressor is called **negative regulation.**

In the absence of inducer





In the presence of inducer

HUMAN GENOME PROJECT (HGP)

- The entire DNA in the haploid set of chromosomes of an organism is called a **Genome**.
- In Human genome, DNA is packed in 23 chromosomes.
- Human genome contains about **$3x10^9$ bp.**
- **Human Genome Project** (1990-2003) was the first mega project for the sequencing of nucleotides and mapping of all the genes in human genome.
- HGP was coordinated by **U.S. Department of Energy and the National Institute of Health.**

Goals of HGP

a. Identify all the estimated genes in human DNA.
b. Sequencing of 3 billion chemical base pairs of human DNA.
c. Store this information in **databases.**
d. Improve tools for data analysis.
e. Transfer related technologies to other sectors.
f. Address the **ethical, legal and social issues (ELSI)** that may arise from the project.

Methodologies of HGP: 2 major approaches.

- **Expressed Sequence Tags (ESTs):** Focused on identifying all the genes that are expressed as RNA.
- **Sequence annotation:** Sequencing whole set of genome containing all the coding & non-coding sequence and later assigning different regions in the sequence with functions.

Procedure of sequencing:

Isolate DNA from a cell → Convert into random fragments → Clone in a host (bacteria & yeast) using vectors (e.g. BAC & YAC) for amplification → Sequencing of fragments using Automated DNA sequencers (Frederick Sanger method) → Arrange the sequences based on overlapping regions→ Alignment of sequences using computer programs.

BAC= Bacterial Artificial Chromosomes
YAC= Yeast Artificial Chromosomes

- **Sanger** has also developed method for **sequencing of amino acids** in proteins.
- DNA is converted to fragments as there are technical limitations in sequencing very long pieces of DNA.
- HGP was closely associated with **Bioinformatics. Bioinformatics:** Application of computer science and information technology to the field of biology & medicine.
- Of the 24 chromosomes (22 autosomes and X & Y), the last sequenced one is **chromosome 1** (May 2006).
- DNA sequencing also have been done in bacteria, yeast, *Caenorhabditis elegans* (a free living non-pathogenic nematode), *Drosophila,* plants (rice & *Arabidopsis*), etc.

Salient features of Human Genome

a. Human genome contains **3164.7 million** nucleotide bases.
b. Total number of genes= about **30,000.**
c. Average gene consists of **3000 bases,** but sizes vary. Largest known human gene (**dystrophin** on X-chromosome) contains 2.4 million bases.
d. 99.9% nucleotide bases are same in all people. Only **0.1%** ($3x10^6$ bp) difference makes every individual unique.
e. Functions of over 50% of discovered genes are unknown.
f. Chromosome I has most genes **(2968)** and Y has the fewest **(231).**
g. Less than 2% of the genome codes for proteins.
h. Very large portion of human genome is made of **Repeated (repetitive) sequences.** These are stretches of DNA sequences that are repeated many times. They have no direct coding functions. They shed light on chromosome structure, dynamics and evolution.
i. About **1.4 million** locations have single-base DNA differences. They are called **SNPs (Single nucleotide polymorphism or 'snips').** This helps to find chromosomal locations for disease-associated sequences and tracing human history.

DNA FINGERPRINTING (DNA PROFILING)

- It is the technique to identify the similarities and differences of the DNA fragments of 2 individuals.
- It is developed by **Alec Jeffreys (1985).**

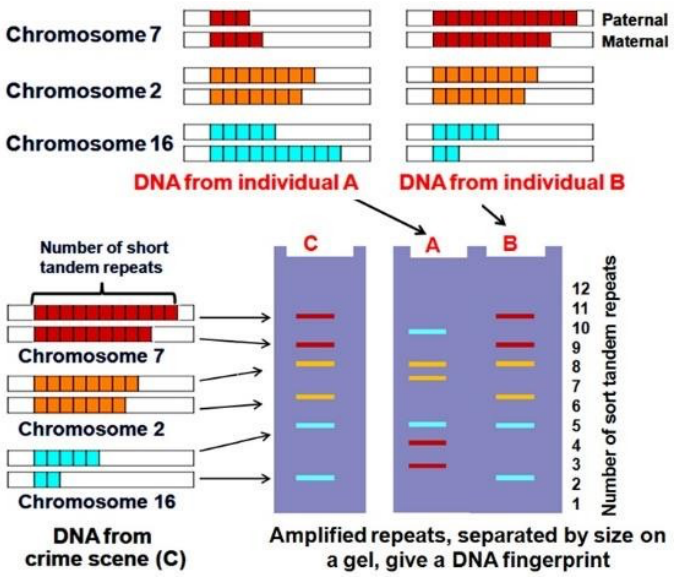
Basis of DNA fingerprinting

- DNA carries some non-coding **repetitive sequences.**
- Repetitive DNA can be separated from bulk genomic DNA as different peaks during density gradient centrifugation.
- The **bulk DNA** forms a **major peak** and the small peaks are called **satellite DNA.**
- **Satellite DNA** is classified as **micro-satellites, mini-satellites** etc. based on base composition (A:T rich or G:C rich), length of segment and number of repetitive units.
- A DNA sequence which is tandemly repeated in many copy numbers is called **variable number tandem repeats (VNTR).** It belongs to mini-satellite DNA.
- In a person, copy number varies in each chromosome.
- The two alleles (paternal and maternal) of a chromosome also contain different copy numbers of VNTR.
- VNTR is specific from person to person.
- The size of VNTR varies from **0.1 to 20 kb.**
- Any difference in the nucleotide sequence (inheritable mutation) observed in a population is called **DNA polymorphism** (variation at genetic level).
- Polymorphism is higher in non-coding DNA sequence because mutations in these sequences may not affect an individual's reproductive ability. These mutations accumulate generation to generation causing polymorphism.
- Polymorphisms have great role in evolution & speciation.

Steps of DNA fingerprinting (Southern Blotting Technique)

a. **Isolation** of DNA (from any cells or blood stains, semen stains, saliva, hair roots, bone, skin etc.).
b. Digestion of DNA by **restriction endonucleases.**
c. Separation of DNA fragments by **gel electrophoresis.**
d. **Transferring (blotting)** DNA fragments to synthetic membranes such as **nitrocellulose** or **nylon**.
e. Hybridization using radioactive labelled **VNTR probe.**
f. Detection of hybridized DNA by **autoradiography.**

The autoradiogram gives an image in the form of dark & light bands. It is called **DNA fingerprint.**



DNA fingerprint differs in everyone except in **monozygotic (identical) twins.**

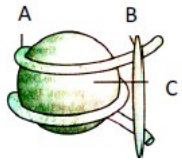
The sensitivity of the technique can be increased by use of **polymerase chain reaction (PCR).** Therefore, DNA from a single cell is enough for DNA fingerprinting.

Application of DNA fingerprinting

- **Forensic tool** to solve paternity, rape, murder etc.
- For the diagnosis of **genetic diseases.**
- To determine **phylogenetic status** of animals.
- To determine **population and genetic diversities.**

MODEL QUESTIONS

1. Note the relationship between first two words and fill up the fourth place.
   a. DNA: Thymine & cytosine RNA: ..... b. UGG: Tryptophan AUG: .....
2. The percentage of adenosine phosphate in DNA isolated from human liver is observed to be 30.7%. What is the expected percentage of four nitrogen bases?
3. Schematically represent Griffith's experiment.
4. Analyze the following diagram



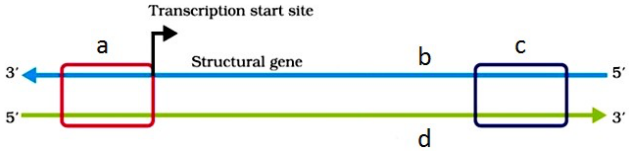
   a. Identify and copy the diagram.
   b. Label A, B & C.
   c. Mention the advantage of this arrangement.

5. Analyse the below flowchart which represent the central dogma of molecular biology and answer the following.

DNA ⟶(A) RNA ⟶(B) Proteins

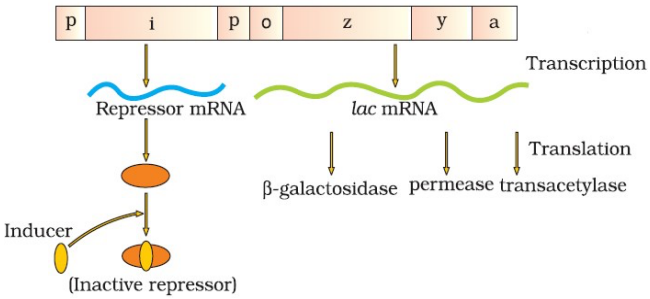
   a. Mention the processes represented by the letters A and B
   b. How the central dogma is modified with discovery of reverse transcriptase?

6. Observe the diagram given below:



   a. Label a, b, c, d in the figure.
   b. During transcription, both DNA strands are not copied. Explain the reason.

7. Find the odd one and give reason: UAA, AUG, UAG, UGA
8. The sequence of template strand of a gene is given below. Construct the base sequence of mRNA transcribed from this.
   3' CACGTGGACTGAGGACTCCTC 5'
9. Given below is the DNA sequence, representing a part of the gene. Analyse this and answer the following questions.
   5' ATGGGGGTGCTCAATATATGCCCCCGTAGTTAA 3'
   3' TACCC CCACGAGTTATATACGGGGGCATCAATT 5'
   a. Construct the mRNA molecule which will be transcribed from this DNA sequence.
   b. Make a processed mRNA (assume that all the codons containing a C represent the intron DNA).
   c. How many amino acid residues will make up the polypeptide corresponding to this processed mRNA?
10. "Protein synthesis will fail in the absence of tRNA molecule". Justify this statement.
11. The coding region of a gene is estimated to consist of 450 nucleotide base pairs. Avoiding stop codons & introns, how many amino acids would the corresponding polypeptide chain contain? Justify your answer.
12. Observe the diagrammatic representation of the lac operon given below and answer the questions.



   a. What is the inducer in lac operon?
   b. How does it ensure the "switching-on" of genes?
   c. Mention the structural genes in lac operon.
   d. What is the role of β-galactosidase & permease?

13. Expand these abbreviations:
   a. NHC b. UTRs c. HGP d. ESTs e. SNPs f. VNTR g. BAC h. YAC
14. Draw a flowchart of steps involved in DNA fingerprinting.