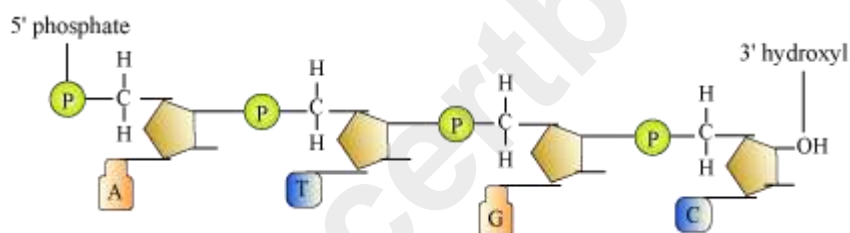


# Molecular Basis of Inheritance

## DNA : Structure of Polynucleotide Chain

### DNA

- DNA – Polymer of deoxyribonucleotides
- Nucleoside = Nitrogenous base + Pentose sugar (linked through *N* – **glycosidic bond**)  
Example – adenosine, deoxyadenosine, cytidine, etc.
- Nucleotide = Nucleoside + Phosphate group (linked through **phosphodiester bond**)
- Many nucleotides link together through 3' – 5' **phosphodiester bond** to form polynucleotide chain (as in DNA and RNA).
- In course of formation of polynucleotide chain, a phosphate moiety remains free at 5' end of ribose sugar (5' end of polymer chain) and one -OH group remains free at 3' end of ribose (3' end of polymer chain).

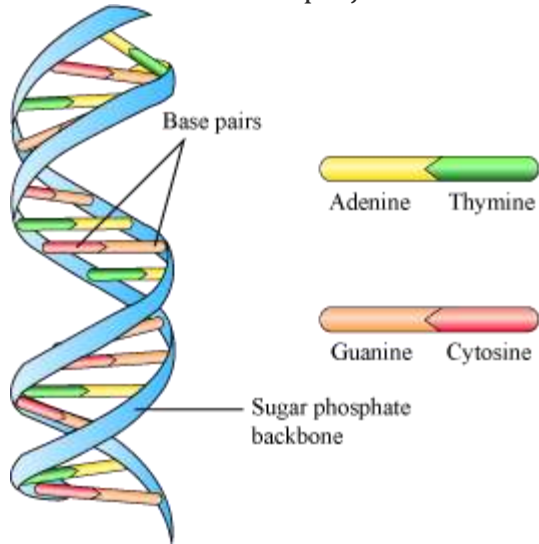


### Double Helix Model for the Structure of DNA

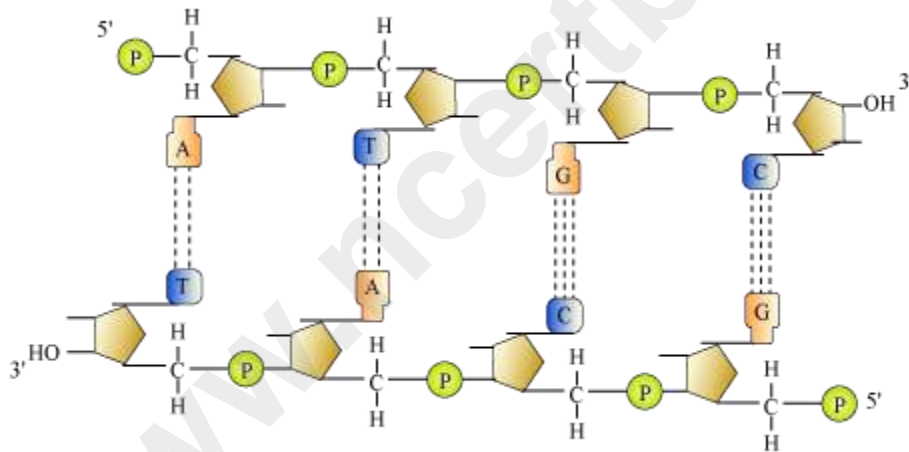
- Scientists involved
- **Friedrich Meischer** – First identified DNA as an acidic substance present in nucleus and named it as 'Nuclein'
- **Wilkins and Franklin** – Produced X-ray diffraction data for DNA structure
- **Watson and Crick** – Proposed double helix structure model for DNA based on X-ray diffraction data
- **Erwin Chargaff** – Proposed that in ds DNA, ratios A:T and C:G remain same and are equal to one

### Features of double helix structure of DNA:

- In a DNA, two polynucleotide chains are coiled to form a helix. Sugar-phosphate forms backbone of this helix while bases project inwards to each other.



- Complementary bases pair with each other through hydrogen bond. Purines always pair with their corresponding pyrimidines. Adenine pairs with thymine through two hydrogen bonds while guanine pairs with cytosine through three hydrogen bonds.

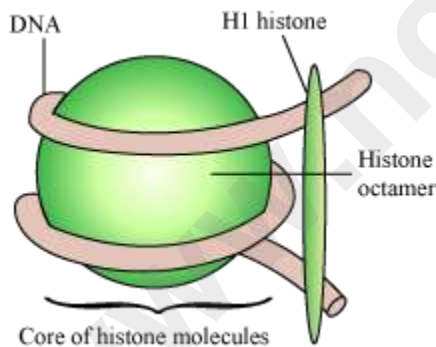


- The helix is right-handed.  
Pitch - 3.4 nm  
10 bp in each turn
- The plane of one base pair stacks over the other in a double helix. This provides stability to the helix along with hydrogen bonding.

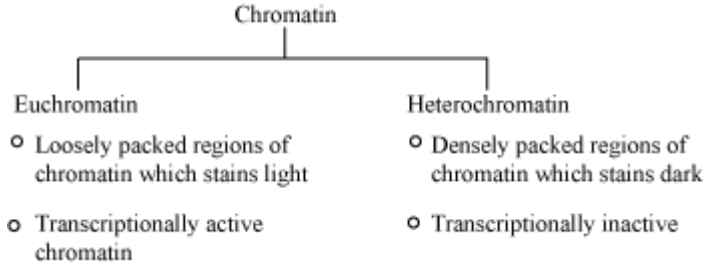
Packaging of DNA Helix

**Packaging of DNA Helix**

- Distance between two consecutive base pairs in a DNA = 0.34 nm =  $0.34 \times 10^{-9}$  m
- Total number of base pairs in a human DNA =  $6.6 \times 10^9$  bp
- Total length of human DNA =  $0.34 \times 10^{-9} \times 6.6 \times 10^9$   
=  $\sim 2.2$  m
- 2.2 m is too large to be accommodated in the nucleus ( $10^{-6}$  m).
- Organisation of DNA in prokaryotes:
- They do not have nucleus. DNA is not scattered throughout the cell.
- In certain regions called nucleoids, DNA (negatively charged) is organised in large loops and is held by some proteins (positively charged).
- Organisation of DNA in eukaryotes:
- They have positively charged basic proteins called histones (positive and basic due to presence of positive and basic amino acid residues, lysine and arginine).
- Histone octamer – Unit of eight molecules of histone
- DNA (negatively charged) winds around histone octamer (positively charged) to form nucleosome.



- 1 nucleosome has approx. 200 bp of DNA.
- Nucleosomes in a chromatin resemble beads present on strings.
- Beads on string structure in chromatin are further packaged to form chromatin fibres, which further coil and condense to form chromosomes during metaphase.
- Non-histone chromosomal proteins – Additional set of proteins required for packaging of chromatin at higher level



Transforming Principle

### Discovery of DNA as a Genetic Material

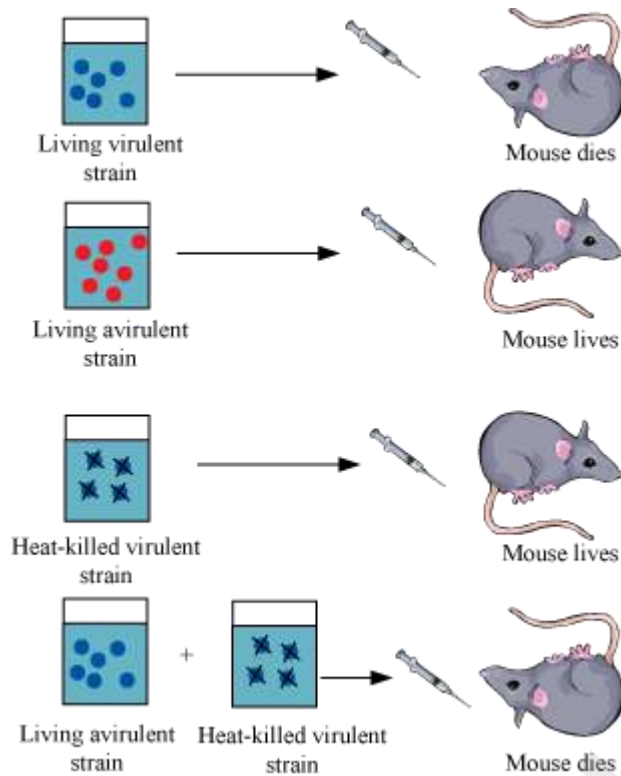
- Though principles of inheritance and discovery of chromosomes in nucleus were achieved long time back, there was confusion about which molecule acted as genetic material.

### Transforming Principle

- Griffith performed experiments with the bacteria *Streptococcus pneumoniae*. This bacterium has two strains – S strain and R strain.

S strain Bacteria	R strain Bacteria
Produce smooth colonies on culture plate	Produce rough colonies on culture plate
Have a polysaccharide coat	Do not have a polysaccharide coat
Virulent (causes pneumonia)	Non-virulent (does not cause pneumonia)

- Griffith's experiment



- Live R strain in the presence of heat-killed S strain produce virulence because somehow R strain bacteria is transformed by heat-killed S strain bacteria. Hence, it was concluded that there must be transfer of genetic material.

### Biochemical Nature of Transforming Material

- Avery, McLeod, and McCarthy worked to determine the biochemical nature of genetic material responsible for transformation.

Heat-killed S strain + Live R strain  $\xrightarrow[\text{Digest RNA and proteins}]{\text{RNase, Proteases}}$  Transformation occurs  
and mouse dies

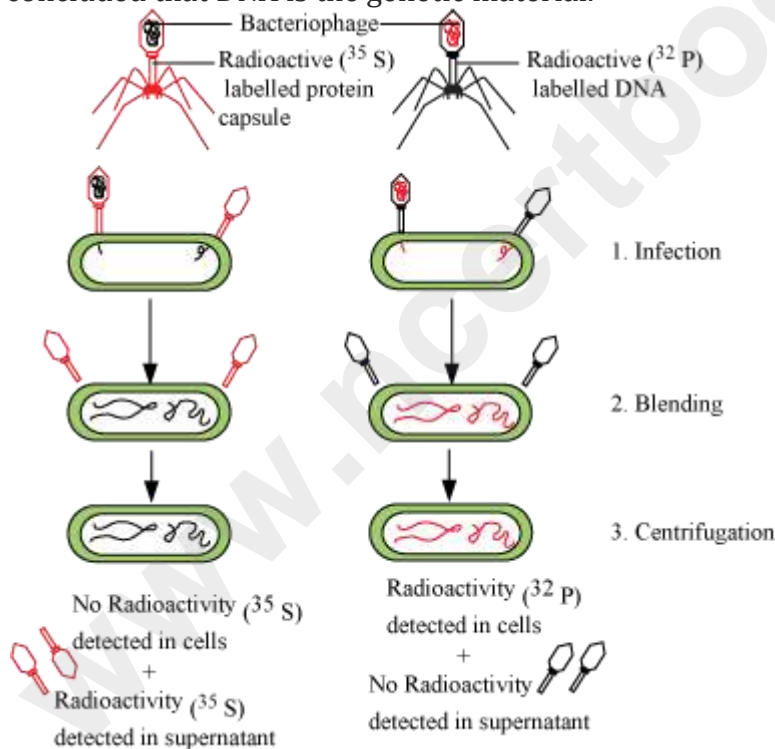
Heat-killed S strain + Live R strain  $\xrightarrow[\text{Digest DNA}]{\text{DNase}}$  Transformation does not  
occur and mouse survives

- This suggests that DNA has to be the genetic material.

### Hershey and Chase Experiment to Confirm DNA as the Genetic Material

- Hershey and Chase worked on bacteriophages (viruses that infect bacteria).

- When a bacteriophage infects a bacterium, the viral genetic material gets attached with the bacterial genetic material and bacteria then treats the viral genetic material as its own to synthesise more viral particles.
- Hershey and Chase worked to discover whether it was a protein or DNA that entered the bacteria from virus.
- They labelled some phages with radioactive sulphur and the others with radioactive phosphorus.
- These radioactive phages were used to infect *E. coli*.
- *E. coli* was then blended and centrifuged to remove viral particles.
- It was observed that bacteria with radioactive DNA were radioactive while those with radioactive proteins lost their radioactivity.
- This showed that it is the DNA that enters the bacteria from viruses and not proteins. Hence, it was concluded that DNA is the genetic material.



### Properties of the Genetic Material

- It should be able to replicate (duplicate to produce its identical copy).
- It should be chemically and structurally stable.
- It should have scope for changes that are essential for evolution.

- It should follow the Mendelian principles of inheritance.
- Difference between DNA and RNA:

DNA	RNA
Has deoxyribose sugar	Has ribose sugar
5-methyl uracil (thymine) is present.	Uracil is present in place of thymine.
Mostly DNA acts as the genetic material.	RNA acts as a messenger and adaptor. It acts as a genetic material in some viruses.
DNA is stable.	Presence of 2' OH group at every nucleotide makes RNA labile and easily biodegradable.
Chemically less reactive, mutates slowly	Mutation in RNA is faster.
DNA requires RNA for protein synthesis. DNA → RNA → Protein	RNA directly codes for proteins.

#### Why DNA is more stable than RNA?

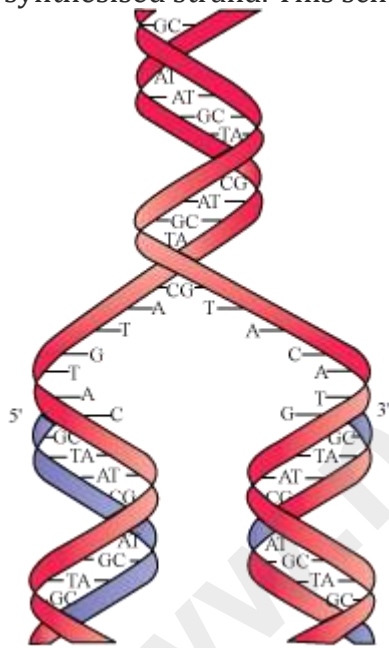
- In RNA, a 2' OH group is present at every nucleotide. This makes RNA unstable and degradable.
- Presence of thymine in place of uracil confers additional stability to DNA.
- RNA being a biocatalyst is more reactive.

- DNA is double-stranded having complementary strand, which resists the changes by repair mechanism.

## DNA Replication

### What is DNA Replication?

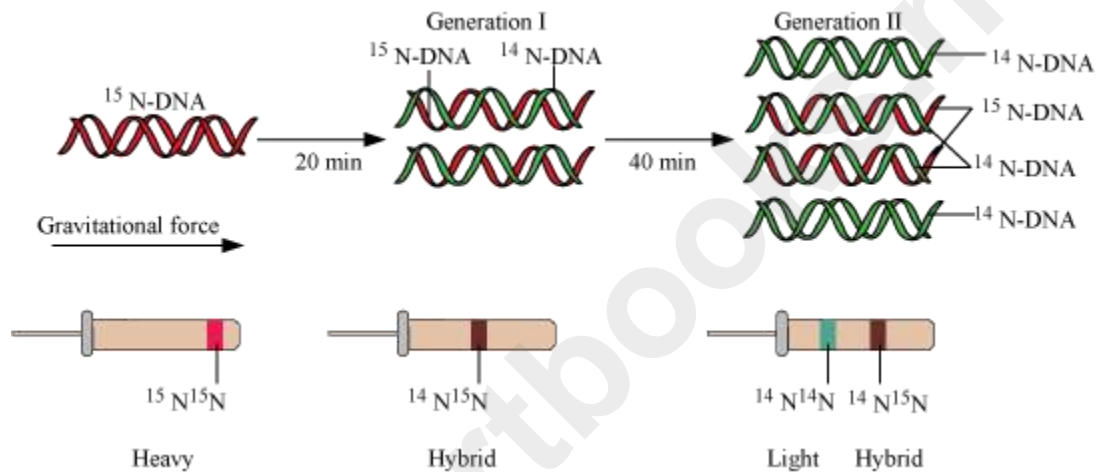
- DNA replication is the phenomenon in which a duplicate copy of DNA is synthesised.
- In replication, two strands of the DNA helix separate and each strand acts as a template for synthesising new complementary strands.
- After completion of replication, the two copies so produced will have one parental and one newly synthesised strand. This scheme of replication is called semi-conservative replication.



### Experiment to Prove That DNA Replicates Semi-Conservatively

- Performed by – Messelson and Stahl
- *E.coli* was grown in a medium containing heavy isotope  $^{15}\text{N}$  as the nitrogen source.
- $^{15}\text{N}$  was incorporated into newly synthesised DNA as well and the DNA became heavy DNA.
- Heavy DNA molecule can be differentiated from normal DNA by density gradient centrifugation using cesium chloride as the gradient.

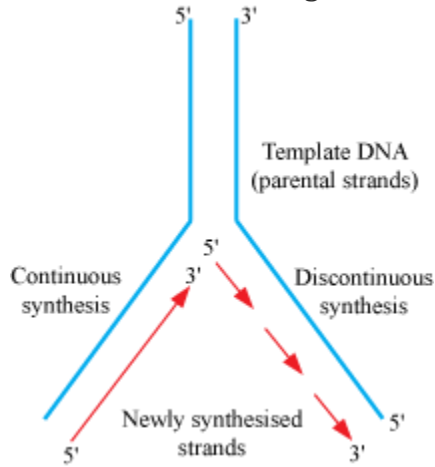
- Then, cells were again transferred into a medium with  $^{14}\text{N}$  as nitrogen source. Samples were taken from this media and their DNA was extracted.
- *E. coli* divides every 20 minutes. Therefore, the DNA extracted after 20 minutes had a hybrid density.
- DNA extracted after 40 minutes had equal amount of hybrid and light intensities.
- This implies that the newly synthesised DNA obtained one of its strands from the parent. Thus, replication is semi-conservative.



### Mechanism of DNA Replication

- Replication occurs in S phase of cell cycle.
- Enzyme involved - DNA polymerase (DNA dependent DNA polymerase)
- Replication requires energy.  
Source of energy – Deoxyribonucleoside triphosphates (DNTPs)
- DNTPs have dual purpose – Act as substrates and provide energy also
- Replication initiates at specific regions in DNA called origin of replication.
- DNA polymerase polymerises a large number of nucleotides in a very short time.
- During the course of replication, two parent strands do not completely open, but a small opening forms in which replication occurs. This small opening forms a replication fork.
- DNA polymerase can polymerise only in one direction that is  $5' \rightarrow 3'$ .

- Therefore, replication occurs smoothly at 3' to 5' end of DNA. (continuous replication, but occurs discontinuously at 5' to 3' end)
- The discontinuous fragments so formed are joined by DNA ligase.



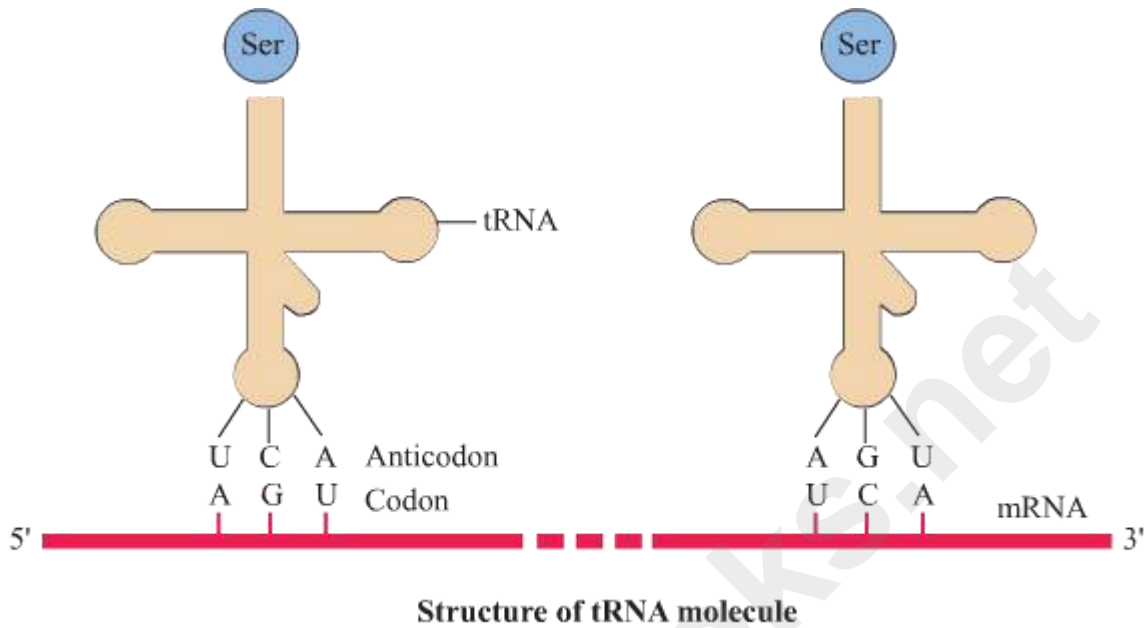
Types of RNA

### Types of RNA

- mRNA (messenger RNA) – It serves as a template for protein synthesis. DNA is transcribed to form an mRNA, which in turn is translated to form protein. [Central dogma of molecular biology]
- tRNA (transfer RNA) – It brings amino acids during translation and reads the genetic code.
- rRNA (ribosomal RNA) – These are the work benches of translation. They play a structural and catalytic role during translation.

### tRNA

- tRNA is an adapter molecule. On one hand, it reads the genetic code and on the other hand, it binds to specific amino acids.
- tRNA has an **anticodon loop** that has bases complementary to the mRNA code and an **amino acid acceptor end** where it binds to the corresponding amino acid.
- Initiation tRNA – This tRNA is essential for initiation of translation and has UAC in anticodon loop and Met in amino acid acceptor end.
- There are no tRNAs for stop codons.



Transcription Unit - Structure and its Relationship with a Gene

### Transcription

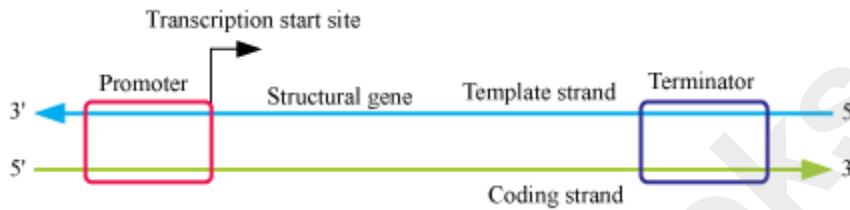
- Transcription is the process of formation of RNA molecules from the DNA.
- During transcription, only a segment of DNA from only one of the strands participates.
- Both strands are not copied during transcription because:
- If both strands get transcribed at the same time since the sequences of amino acid would be different in both (due to complementarity), then two RNA molecules with different sequences will be formed, which in turn give rise to two different proteins. Therefore, one DNA would end up giving rise to two different proteins.
- Two RNA molecules so formed will be complementary to each other, hence would end up forming a double-stranded RNA leaving the entire process of transcription futile.

### Transcriptional Unit

- A transcriptional unit has primarily three regions:
- Promoter – Marks the beginning of transcription; RNA polymerase binds here
- Structural gene – Part of the DNA that is actually transcribed
- Terminator – Marks the end of transcription

## Template Strand and Coding Strand

- Enzyme involved in transcription, RNA polymerase (DNA dependent RNA polymerase), catalyses in only one direction i.e., 5' to 3'.
- Therefore, the strand with polarity 3' → 5' acts as a template (Template Strand).
- The strand with polarity 5' → 3' acts as coding strand (which is a misnomer since it does not code for anything). Coding strand has sequence similar to RNA formed after transcription except for the change that thymine is present instead of uracil.



## Gene

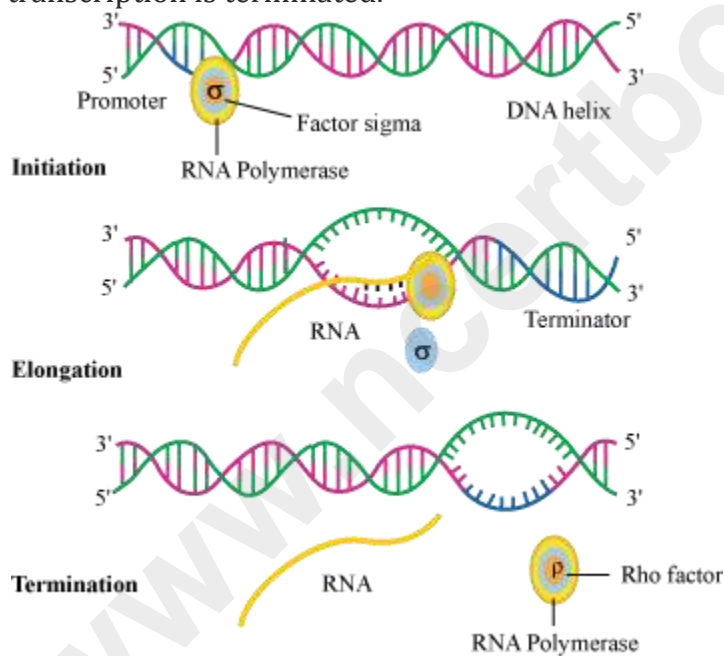
- The DNA sequence which codes for tRNA or rRNA molecule defines a gene.
- Cistron – Segment of DNA that contains the genetic code for a single polypeptide
- The structural genes could be of two types:
  - Monocistronic (mostly in eukaryotes)
  - Polycistronic (mostly in prokaryotes)
- Monocistronic genes have two parts:
  - Exon – Sequences that code for a particular character and is expressed in a matured and processed mRNA
  - Intron – Interrupting sequences that do not appear in a mature and processed mRNA
- Regulatory genes – Sequences that do not code for anything, but have regulatory functions

## Process of Transcription

### Process of Transcription

- Transcription has three steps – initiation, elongation, and termination.

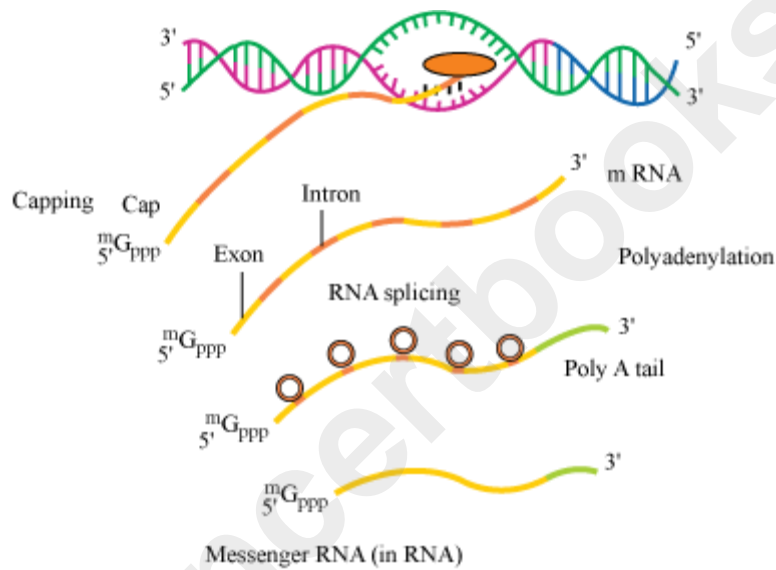
- Initiation:
  - RNA polymerase binds with the promoter to initiate the process of transcription.
  - Association with initiation factor ( $\sigma$ ) alters the specificity of RNA polymerase to initiate the transcription.
- Elongation:
  - RNA polymerase uses nucleoside triphosphate as substrate, and polymerisation occurs according to complementarity.
- Termination:
  - Termination occurs when termination factor (P) alters the specificity of RNA polymerase to terminate the transcription.
  - As the RNA polymerase proceeds to perform elongation, a short stretch of RNA remains bound to the enzyme. As the enzyme reaches the termination region, this nascent RNA falls off and transcription is terminated.



### Complexities Associated with Transcription

- In prokaryotes:
  - There is no clear demarcation between cytosol and nucleus. Therefore, translation can begin even before transcription is completed. Thus, in prokaryotes, transcription and translation are coupled.

- In eukaryotes:
- Three different kinds of RNA polymerases are present.  
RNA polymerase I transcribes rRNA.  
RNA polymerase II transcribes hnRNA (mRNA precursor).  
RNA polymerase III transcribes tRNA, snRNA, and srRNA.
- The precursor of mRNA, i.e. hnRNA, contains both introns and exons. Introns are removed and exons are joined by a process called splicing.
- Capping – In this, methyl guanosine triphosphate is added to the 5' end of hnRNA.
- Tailing – In this, adenylate residues are added to the 3' end of hnRNA.
- When hnRNA is fully processed, it is known as mRNA, which is transported out of the nucleus to get translated.



## Genetic Code

### Genetic Code

- Genetic code directs the sequence of amino acids during the synthesis of proteins.
- George Gamow proposed that if 20 amino acids are to be coded by 4 bases, then the code should be made up of three nucleotides.  $4^3 = 64$  ( $4^2 = 16$ ), which is less than 20; so, the codon was proposed to be triplet.
- Har Gobind Khorana developed a chemical method to synthesise RNA molecules with defined combination of bases.
- Nirenberg developed cell-free systems for protein synthesis, which helped the code to be deciphered.

- The enzyme known as Severo Ochoa enzyme (polynucleotide phosphorylase) helped to polymerise RNA with defined sequences in a template independent manner.
- It finally gave rise to the checker-board for genetic code.

**The Codons for the Various Amino Acids**

	Second position								Third position
	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.  $4^3 = 64$  (61 codons code for amino acids while 3 are stop codons)
  - One codon codes for a single specific amino acid. Codons are unambiguous.
  - Codons are degenerate since some amino acids are coded by more than one codon.
  - Genetic code is universal. 1 codon codes for same amino acid in all species.
  - Codons are read continuous. They lack punctuations.
  - AUG has dual functions – Codes for Methionine and acts as a start codon

## Mutation

### Mutation

- Alteration of DNA sequence resulting in changes in genotype and phenotype of organisms
- DNA helix runs in a chromatid, hence any change (insertion or deletion) in the DNA sequence affects the chromosome.
- Point Mutation – Mutation arising due to change in single base pair of DNA as in sickle cell anaemia
- Frameshift Mutation – Mutations arising due to deletion or insertion in DNA sequence
- Mutagens – Chemical or physical agents that lead to mutations  
Example – UV radiations

## Effects of Mutations on Genetic Code

- Mutations include insertions, deletions, and rearrangements.
- Mutation results in changed phenotype and diseases such as sickle cell anaemia. (Change Glu → Val in gene coding for beta globin chain of haemoglobin) Such mutations are called **point mutations**.
- Insertion or deletion of a single base pair disturbs the entire reading frame in mRNA. Such mutations are called **frameshift mutations**.
- Frameshift mutations hold the proof of the fact that codon is triplet because if we insert three or multiple of three bases followed by the deletion of same number of bases, then the reading frame will remain unaltered.

## Translation

### Translation

- The mRNA contains the genetic information, which is translated into the amino acid sequence with help of tRNA. Amino acids are polymerised to form a polypeptide.
- Amino acids are joined by peptide bond.
- First of all, charging of tRNA (amino-acylation of tRNA) takes place. In this, amino acids are activated in the presence of ATP and are linked to their corresponding tRNA.
- Ribosomes are the workbenches for translation. Ribosomes have 2 subunits: a large subunit and a small subunit.
- Smaller subunit comes in contact with mRNA to initiate the process of translation.
- Translational unit in an mRNA is the region flanked by start codon and stop codon.
- Untranslated regions (UTR) are the regions on mRNA that are not themselves translated, but are required for efficient translation process. They may be present before start codon (5' UTR) or after stop codon (3' UTR).
- Initiator tRNA recognises the start codon. (Initiation)
- Then t-RNA-amino acid complexes bind to their corresponding codon on the mRNA and base pairing occurs between codon on mRNA and tRNA anticodon.
- tRNA moves from codon to codon on the mRNA and amino acids are added one by one. (Elongation)
- Release factor binds to stop codon to terminate the translation. (Termination)

## Gene Regulation

### Gene Regulation

- Regulation of gene expression could be exerted at following levels.

- Transcriptional level (formation of primary transcripts)
- Processing level (splicing)
- Transport of mRNA from nucleus to cytoplasm
- Translational level
- In addition, metabolic, physiological, or environmental conditions regulate the expression of genes.
- Expression of genes coding for enzymes is required only when substrate for that enzyme is available.  
For example:  
Lactose  $\xrightarrow{\text{beta-galactosidase}}$  Glucose + Galactose  
*E.coli* synthesises beta-galactosidase, only when lactose is available.

- Regulation in prokaryotes
- Gene expression is regulated by controlling the rate of transcriptional initiation.
- The activity of RNA polymerase at a given promoter is regulated by accessory proteins. The accessory proteins affect the ability of a promoter to recognise start sites.
- A regulatory protein could be activator or repressor.
- Accessibility of promoter is also affected by operators. Operator is the region located adjacent to promoter.
- Each operon has a specific operator and a specific repressor.
- Usually operator binds to a repressor protein.

### **Lac Operon**

- Operon – An arrangement where a polycistronic gene is regulated by a common promoter and regulatory genes
- *Lac* operon, *trp* operon, *his* operon, *val* operon are the examples of such systems.
- The elucidation of *lac* operon as a transcriptionally active system was first done by geneticist Jacob and biochemist Monod.
- Genes constituting *lac* operon:

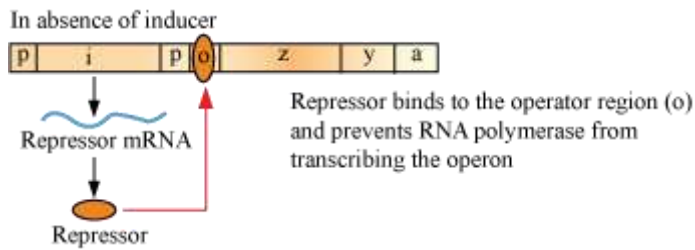
--

Gene	Nature	Function
<i>i</i> gene	Inhibitor	It codes for repressor of <i>lac</i> operon.
<i>z</i> gene	Structural	It codes for $\beta$ -galactosidase. Lactose $\xrightarrow{\beta\text{-galactosidase}}$ Galactose + Glucose
<i>y</i> gene	Structural	It codes for permease, which increases permeability of cell to $\beta$ -galactosidase.
<i>a</i> gene	Structural	It codes for transacetylase.

- All genes involved in *lac* operon are required for metabolism of lactose.
- **Inducer** – Lactose acts as an inducer for *lac* operon since it regulates the switching on and off of the operon.
- If lactose is provided to the growth media of bacteria in absence of any other carbon source, then it is transported inside the cells by permease.
- For permease to be present and lactose to enter inside the cells, low level of expression of *lac* operon must be present all the time.

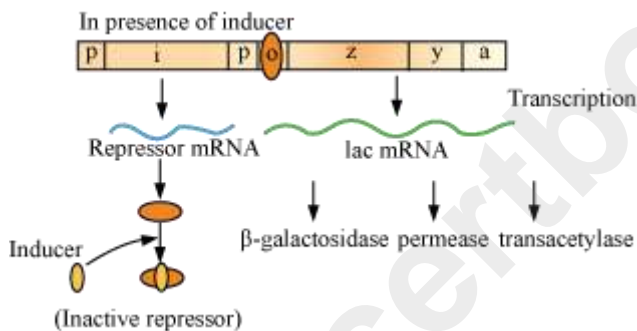
### Regulation in Absence of Inducer

- In absence of inducer, *i* gene transcribes to synthesise repressor mRNA, which translates to form repressor.
- This repressor binds with the operator region of operon and prevents RNA polymerase to transcribe genes – *z*, *y*, and *a* (negative regulation).
- Therefore, in absence of the products of these genes, metabolism of lactose ceases.



### Regulation in Presence of Inducer

- Inducer binds with the protein product of gene *i* (repressor) and inactivates it.
- This inactivated repressor is unable to inactivate RNA polymerase enzyme and *z*, *y*, and *a* genes synthesise their respective mRNA, which in turn gets translated to form  $\beta$ -galactosidase, permease, and transacetylase.
- In presence of all these enzymes, the metabolism of lactose proceeds in a normal manner.



### Genome Project

#### Human Genome Project

- Joint venture of US Department of Energy and National Institute of Health (NIH); later joined by Wellcome Trust (UK)
- Launched in 1990, completed in 2003
- This project worked towards the determination of complete DNA sequence of humans.
- DNA is the storehouse of genetic information and determining its sequence of base pairs can solve many medical, agricultural, environmental, and evolutionary mysteries.

#### Relationship of HGP with Bioinformatics

- Human genome (genome refers to the totality of genes that are present in a human being) contains  $3 \times 10^9$  base pairs.

- Cost of sequencing 1 bp = US \$ 3  
Cost of sequencing  $3 \times 10^9$  bp = US \$ 9 billion
- Enormous sequence data so generated would have required 3300 books containing 1000 pages each just for a human genome.
- Hence, for storing, retrieving, and analysing this enormous data, a new branch of biology has been developed known as bioinformatics.
- Genomes of many non-human models such as bacteria, yeast, *Caenorhabditis elegans*, *Drosophila*, plants (rice and *Arabidopsis*) have also been sequenced.

### **Methods to Identify Genes**

- Two methods – identifying ESTs (Expressed sequence Tags) and sequence annotation
- ESTs – As the name suggests, this refers to the part of DNA that is expressed, i.e. transcribed, as mRNA and translated into proteins thereafter. It basically focuses on sequencing the part denoting a gene.
- Annotation – In this approach, entire genome (coding + non-coding) is sequenced and later on function is assigned to each region in the genome.

### **Genome Sequencing**

- DNA from the cells is isolated and is randomly broken into fragments of smaller sizes.
- These fragments are cloned into suitable host using vectors.
- Cloned fragments amplify in the host. Amplification facilitates an easy sequencing.
- Common vectors used – BAC (Bacterial artificial chromosomes) and YAC (Yeast artificial chromosomes)
- Common hosts – Bacteria and yeasts
- Automated sequencers are used to sequence these smaller fragments (Sanger sequencing).
- The sequences so obtained are arranged based on overlapping regions within them (alignment).
- Alignment of the sequences is also done automatically by computer programs.
- Then these sequences are annotated and assigned to each chromosome.

### **Preparation of Genetic and physical maps on Genome**

- 2 methods are used – restriction polymorphism and microsatellites
- Restriction polymorphism – Specialized enzymes called restriction endonucleases are used to cut the genome at specialized sites called restriction endonuclease recognition site and maps are prepared based on it.
- Microsatellites – These are repetitive DNA sequences.

### **Observations from HGP**

- Human genome contains  $3 \times 10^9$  (3164.7 million) nucleotide bases.
- An average gene consists of 3000 bases. However, the size of genes varies. Largest gene is dystrophin (2.4 m bases).
- Total number of genes in human genome – 30,000
- Over 50% of the discovered genes have unknown functions.
- Less than 2% of genome is coding.
- Large portion of genome consists of repeating sequences.
- Repetitive sequences have no coding function. They are repeated over hundred to thousand times. They may have a role in evolution, chromosome structure, and dynamics.
- Chromosome with most genes – Chromosome 1 (2968)
- Chromosome with fewest genes – Chromosomes Y (231)
- SNPs (single nucleotide polymorphism) occur at about 1.4 million locations in human DNA. They are believed to have significance in explaining diseases and evolutionary history of human beings.

### **Rice Genome Project**

- Rice (*Oryza sativa*) is one of the most important food crops in the world.
- Rice, wheat, and maize account for about half of the world's food production, and rice is the principal food of half of the world's population.
- The rice genome is well mapped and well characterised.
- It is the smallest of the major cereal crop genomes at an estimated 400 to 430 Mb.
- The International Rice Genome Sequencing Project (IRGSP) was started in September 1997.

### **Genome Sequencing and Physical Mapping**

- Sequencing of the rice genome is being performed mainly from genomic BAC or PAC libraries created from the Nipponbare variety.

- Large genomic libraries, used as the primary sequencing templates, are constructed in bacterial artificial chromosomes (BACs) or P1-derived artificial chromosomes (PACs).
- Fingerprinting and physical mapping are used to create minimal tiling paths and to anchor BAC clones to physical positions along the length of a chromosome.
- 98 PACs/BACs corresponding to ~12 Mb of sequence have been completed and released to the DNA Data Bank of Japan.
- 53 of the PACs/BACs have been annotated.
- RGP is also working on an expressed sequence tag (EST) map and has generated ~5000 EST markers, which will be added to the rice database.
- FISH (Fluorescence *in situ* hybridisation) and optical mapping are two other methods that are used to aid genome sequencing projects.

### **Observations of RGP**

- ~3.5% of the genome has been completed (15 of 430 Mb) and another 3 to 5% is in production.
- A major part of the nuclear genomes of most plants and many eukaryotes, is composed of repetitive DNA elements.
- Repetitive DNA is estimated to constitute at least 50% of the rice genome.
- Complete sequencing of the rice genome will provide valuable information on the effect of repetitive elements on genome organisation and evolution in plants.

### DNA Fingerprinting

#### **Introduction**

- DNA fingerprinting is a method for comparing the DNA sequences of any two individuals.
- 99.9% of the base sequences in all human beings are identical. It is the remaining 0.1% that makes every individual unique.
- It is a really difficult and time-consuming task to sequence and compare all  $3 \times 10^9$  bases in two individuals. So, instead of considering the entire genome, certain specific regions called repetitive DNA sequences are used for comparative study.

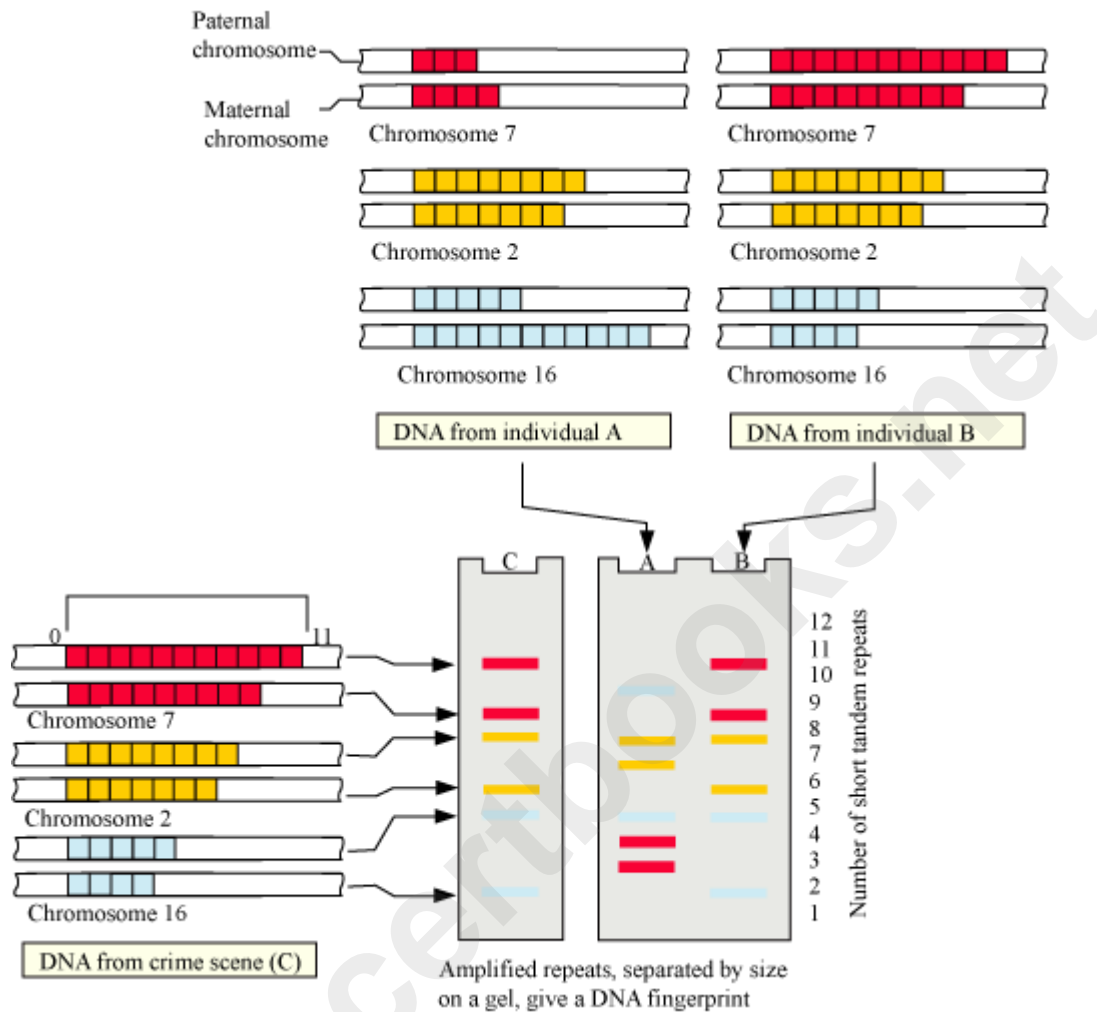
#### **Basis of DNA Fingerprinting**

- Repetitive DNA is separated from bulk genomic DNA since it appears as a distinct peak during density gradient centrifugation.
- Major peak: Formed by bulk DNA  
Smaller peak: Satellite DNA
- Satellites are of two types—micro-satellites and mini satellites, depending upon the base composition, length of segment and the number of repetitive units.

- Satellites do not code for proteins, but have a major role to play in DNA fingerprinting.
- Polymorphism is actually a result of mutation. A germ cell mutation (which can pass on to the next generation through sexual reproduction) gives rise to polymorphism in populations.
- In other words, an inheritable mutation if observed in higher frequencies in a population is known as polymorphism.
- Polymorphisms arise normally in non-coding sequences because mutations in non-coding sequences do not affect an individual's reproductive ability.

### **Methodology of DNA fingerprinting**

- VNTR (variable number of tandem repeats) are satellite DNAs that show high degree of polymorphism.
- VNTRs are used as probes in DNA fingerprinting.
- First of all, DNA from an individual is isolated and cut with restriction endonucleases.
- Fragments are separated according to their size and molecular weight on gel electrophoresis.
- Fragments separated on electrophoresis gel are blotted (immobilised) on a synthetic membrane such as nylon or nitrocellulose.
- Immobilised fragments are hybridised with a VNTR probe.
- Hybridised DNA fragments can be detected by autoradiography.
- VNTRs vary in size from 0.1 to 20 kb.
- Hence, in the autoradiogram, band of different sizes will be obtained.
- These bands are characteristic for an individual. They are different in each individual, except identical twins.



### Applications of DNA Fingerprinting

- DNA fingerprinting is widely used in forensics since every DNA of every tissue from an individual has the same degree of polymorphism.
- DNA fingerprinting forms the basis of paternity testing since a child inherits polymorphism from both its parents.
- It can be used for studying genetic diversity in a population and evolution.