Molecular Basis of Inheritance–

Structure of Nucleic Acids: Nucleic acids are the biomolecules which play a very important role in the process of Inheritance.

Two types of nucleic acids exist: DNA (Deoxyribo Nucleic Acid) and RNA (Ribo Nucleic Acid).

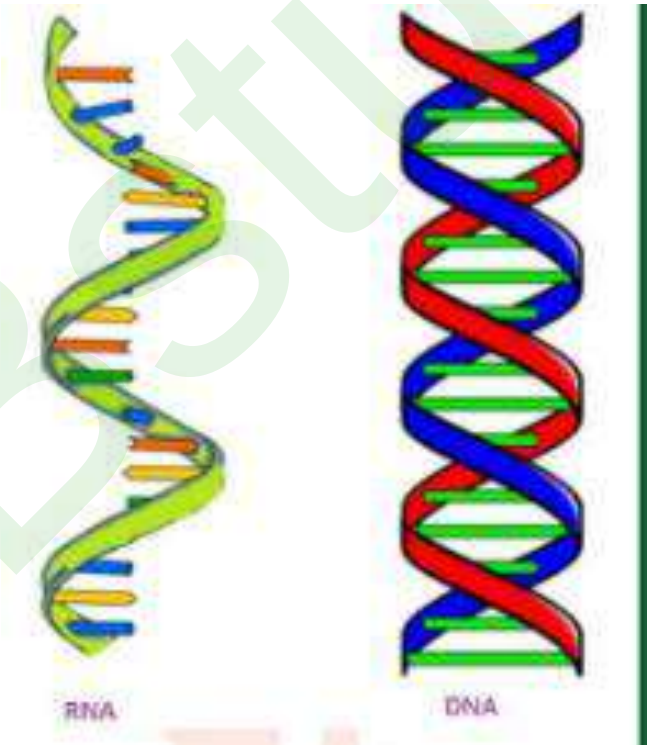
a) DNA: It has a double– stranded structure.

· It is a polynucleotide whose monomer units are deoxyribonucleotide.

· Length of DNA is determined by number of nucleotides in it.

b) RNA: It has a single– stranded structure.

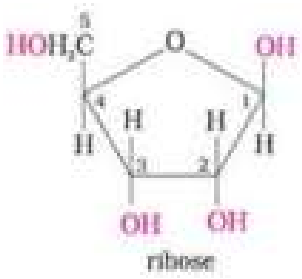
· It is also a polymer whose monomer units are ribonucleotide.

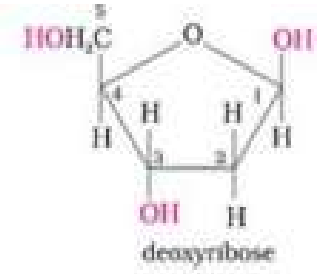


➢ A nucleotide has 3 components:

1) Pentose sugar:

Monosaccharide with 5 Carbon atoms.



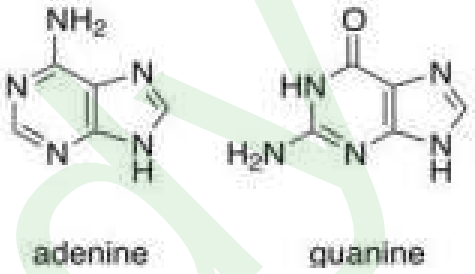


- Ribose sugar in RNA
- Deoxyribose sugar in DNA

2) Nitrogenous base: Nitrogen containing compound with properties of a base.

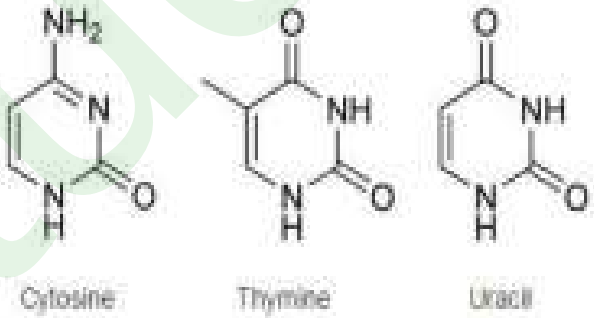
- It is of 2 types: Purines, Pyrimidines.

a) Purine: Heterocyclic aromatic organic compound.



- 9- membered ring.
- Examples: Adenine, Guanine.

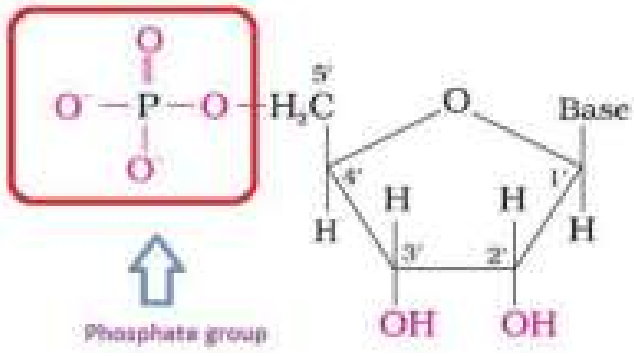
b) Pyrimidine: Heterocyclic aromatic organic compound.



- 6- membered ring.
- Examples: Cytosine, Uracil, Thymine.

3) Phosphate group: Inorganic salt of phosphorus.

- Forms backbone of polynucleotide chain along with the sugar.

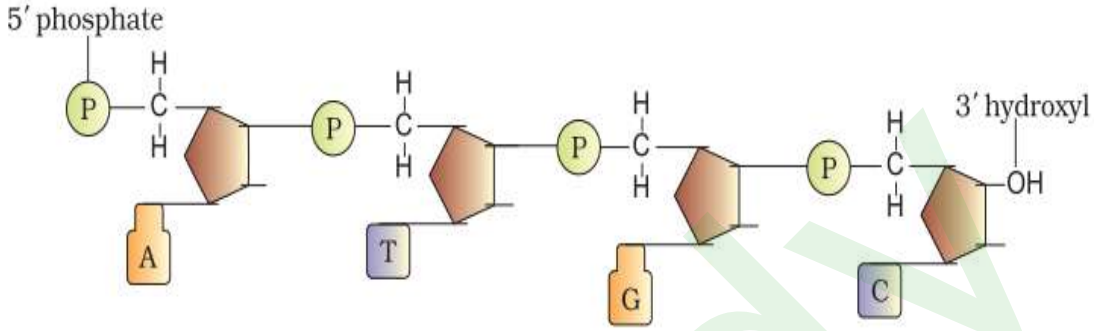


➢ **Formation of a polynucleotide takes place using the following linkages:**

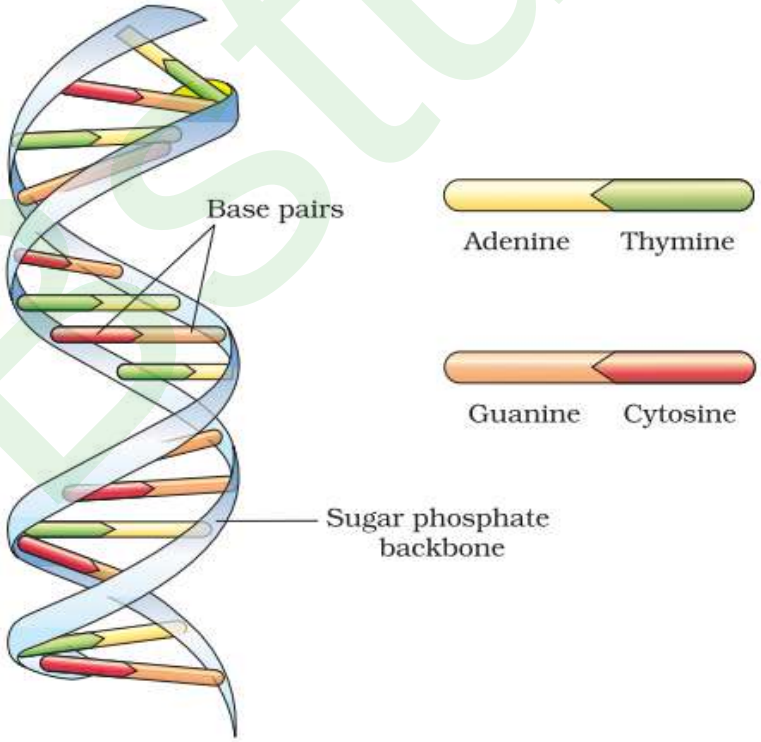
· Nitrogenous base is linked to the pentose sugar through a N– glycosidic bond to form a nucleoside.

· A phosphate group is linked to 5'– OH of a nucleoside through phosphoester bond to form a nucleotide.

· Multiple nucleotides are joined together through 3'– 5' phosphodiester bond to form a polynucleotide.



DNA Double– helix model: DNA is composed of two polynucleotide chains.



i. Sugar– phosphate forms the backbone.

ii. Nitrogenous bases form the interior, paired through H– bonds.

· Complementary base pairing is an important feature of DNA structure.

· The two polynucleotide chains have anti– parallel polarity.

· Two chains are coiled in a right handed fashion forming a right– handed helix.

· Uniform distance is maintained between the two strands of helix.

Packaging of DNA helix: Length of DNA is found to be far greater than dimension of a typical nucleus.

· Total number of base pairs in a typical mammalian cell= $6.6 * 10^9$.

· Distance between two base pairs= 0.34nm.

· Therefore, Length of DNA= $0.34 * 10^{-9} * 6.6 * 10^9$= 2.2m.

· Size of a nucleus is of the order of $10^{-6}m$.

· So, Length of DNA is greater than the size of nucleus.

· DNA is packaged very strategically to fit inside the nucleus.

· In Prokaryotes, DNA is organized into loops held by proteins. The region where DNA is present is termed as ' Nucleoid' .

· In Eukaryotes, there exist positively charged basic proteins called Histones.

· DNA wrap around the histone octamer (group of 8 histone proteins) to form a Nucleosome. Each nucleosome contains 200 base pairs of DNA helix. Nucleosomes in chromatin are seen as ' beads– on– string' under Electron microscope.

➢ Based on different types of DNA packaging, there are two forms of Chromatin.

1) Euchromatin: Less condensed structure with looser DNA packaging.

· Lightly stained when observed under microscope.

· Contains less DNA.

· Transcriptionally active.

· Found in eukaryotes & prokaryotes.

2) Heterochromatin: Highly condensed structure with tighter DNA packaging.

· Dark stained when observed under microscope.

· Contains more DNA.

· Transcriptionally inactive, as those regions of the genes which need to interact with proteins for transcription is inaccessible.

· Found in eukaryotes.

Search for the Genetic Material:

1) Griffith experiment: Experiment with Streptococcus pneumoniae bacteria. This bacteria causes Pneumonia. Two strains of this bacteria were used R– strain & S– strain.

S– strain R– strain

a) S strain: Smooth mucous polysaccharide coat.

· Resistant to immune system.

· Virulent.

b) R strain: Lacks the coat.

· Non- virulent.

· Destroyed by immune system of the host.

➢ **The experiment was performed in multiple steps.**

i. S- strain (virulent) was injected into mouse. It was found that mouse died of pneumonia.

ii. R- strain (non- virulent) was injected into mouse. It was found that mouse remained alive.

iii. Heat killed S- strain (S- strain bacteria were killed by heating) was injected into mouse. It was found that the mouse remained alive.

iv. Heat killed S- strain & live R- strain were injected into mouse. It was found that the mouse died of pneumonia.

Griffith thus arrived at the following conclusion:-

· Something caused bacteria to change from one type (R) to another type (S).

· Some ' Transforming principle' , transferred from heat killed S strain to R strain and transformed it virulent.

2) Bacteriologists experiment: He performed a series of experiments to identify the Transforming Principle.

· Transforming principle precipitated with alcohol. This showed it was not Carbohydrate.

· Transforming principle could not be destroyed with Proteases. Thus, it was not Protein.

· Transforming principle could not be destroyed with Lipases.This proved it was neither Lipids.

· Transforming principle could not be inactivated with Ribonuclease, hence not RNA.

· Transforming principle could be inactivated with Deoxyribonuclease.

Transforming principle was DNA. Therefore, DNA was the genetic material.

3) Hershey– Chase Experiment: Hershey– Chase experiment was performed to confirm that DNA was the genetic material. They experimented with Bacteriophages. Bacteriophages are the viruses that infect & replicate within bacteria.

· Some bacteriophages were grown in radioactive phosphorus medium. It was found that these Bacteriophages came up with radioactive DNA.

· Some bacteriophages were grown in radioactive sulfur medium. It was found that these Bacteriophages with radioactive protein.

It was therefore concluded that it was not the proteins, rather DNA which entered into the bacteria. Therefore, DNA causes the replication of viruses inside the bacteria. DNA was thus proved to be the genetic material.

Criteria for Genetic Material: DNA was found to be the prominent genetic material in most organisms. Exceptions were some viruses where RNA was the genetic material.

Important criteria to be fulfilled to be a genetic material are:

· Capable of replicating itself.

· Chemically & structurally stable.

· Provide scope for mutation which can lead to evolution

· Capable of expressing itself in the form of 'Mendelian Characters' .

· Most of the other molecules like proteins, carbohydrates, lipids failed to fulfill the above mentioned criteria.

· However, RNA could also fulfill the criteria; still DNA was a preferred genetic material over RNA because of the following reasons:

· DNA is structurally more stable than RNA.

· DNA is chemically more stable than RNA.

· DNA has double– stranded structure which provides better ability to rectify errors during replication.

· DNA can' t code directly for protein synthesis & thus depends on RNA.

Therefore, DNA was used for storage of genetic information due to its structural & chemical stability. RNA, on the other hand was used for expression of genetic information as it could directly code for proteins.

RNA World: RNA world was a kind of hypothetical world where RNA performed all the activities which are today performed by DNA & proteins. DNA later evolved from RNA with chemical modifications which made it more stable.

DNA Replication: Replication is the process of reproducing or creating a copy of something. Various hypotheses were proposed by various scientists regarding the replication model of DNA.

· **Semi- conservative DNA replication model:** According to this model, the two strands of DNA separate.

· **Conservative DNA replication:** According to this model, Complete DNA molecule (and not just one strand) acts as a template for new synthesis.

· **Dispersive DNA replication:** According to the Dispersive model, the new DNA is synthesized in short pieces.

Meselson- Stahl Experiment: This experiment was performed to prove the semi conservative nature of DNA replication experimented with bacteria E.coli.

· Presence of intermediate density excluded Conservative hypothesis. Had it been Conservative hypothesis, Generation I would have been either Blue(N-15) or Green(N-14); and not an Intermediate one.

· Presence of N-14 DNA in Generation II excluded Dispersive hypothesis. If it was Dispersive, each DNA should have had the same density. But, in Generation II, we could see 50% of the DNA have intermediate density, whereas remaining 50% have N-14 density.

· Semi- conservative hypothesis could explain the entire experimental result. Separation of strands concept could explain the outcomes of Generation I & II.

Thus, it was proved that DNA replication is Semi-conservative in nature.

Machinery & Enzymes for Replication: Enzymes play an important role acting as catalysts during the process of DNA replication.

· Energy source is needed to provide energy during the replication process. Deoxyribonucleoside triphosphates act as substrates & provide energy for polymerization reaction.

1) DNA polymerase: DNA polymerase creates DNA from nucleotides. It reads the existing DNA strands to create two new strands that match the existing ones.

It is a highly efficient enzyme, as it can replicate a large number of base pairs in a very short time. Rate of replication or Rate of polymerization is approx 2000 bp per second. A total of 4.6 * 106 base pairs are replicated within 18 minutes.

2) Helicase: Enzyme Helicase unwinds DNA from tightly double stranded structure. Only after the strands are separated, DNA polymerase can do its job of creating the new strands. This enzyme separates the strands by breaking the hydrogen bonds between the bases of the two strands.

3) Primase: This enzyme creates a short fragment of RNA (primer) paired with the template DNA strand. This enzyme initiates the process of creation of new strands. DNA polymerase cannot initiate the process on its own. Therefore, primase initiates the same.

Process of DNA replication: Replication cannot be initiated in any random part of DNA. Region in a DNA where replication initiates is termed as ' Origin of Replication' .

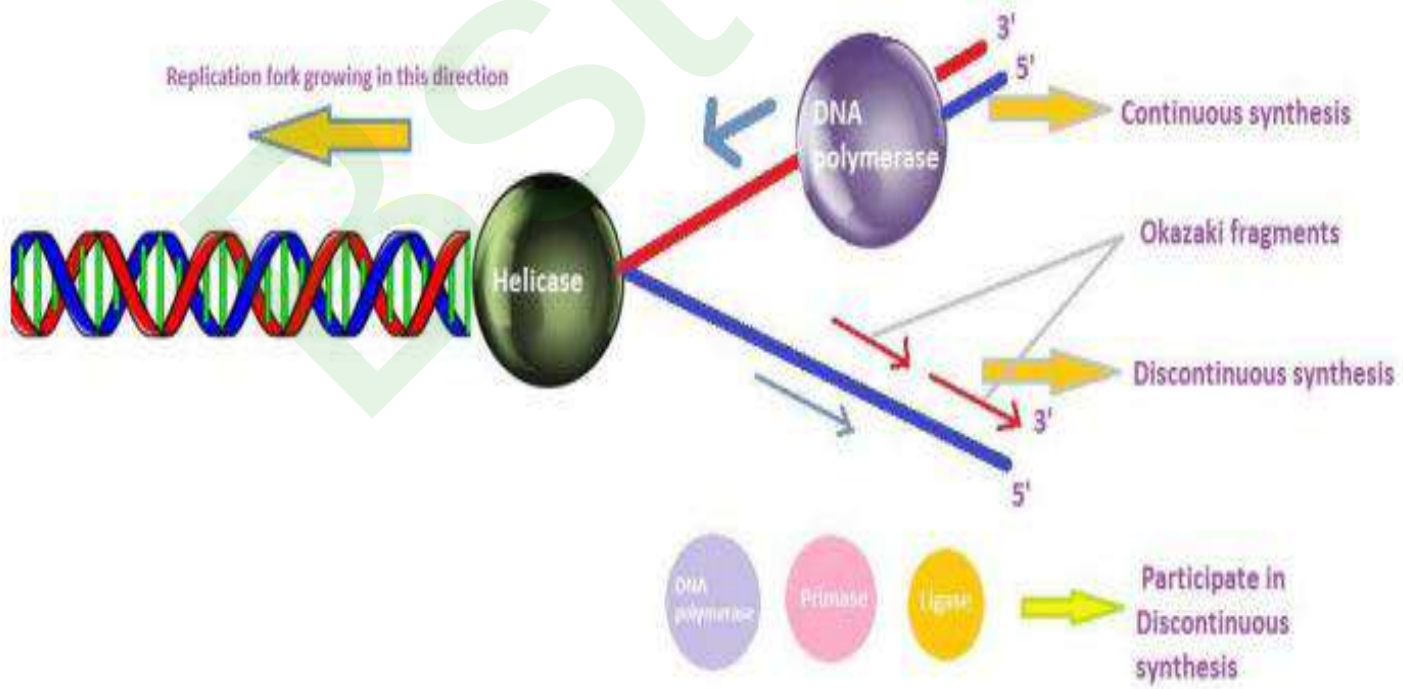
➢ **Step 1-** Enzyme Helicase breaks hydrogen bonds, thus separating the two strands of DNA. Replication fork structure is formed.

➢ **Step 2-**

· Continuous synthesis takes place in the Leading strand. In this strand, DNA is synthesized in the same direction as the growing replication fork. Observe the direction of movement of Helicase & DNA polymerase.

· Discontinuous synthesis takes place in the Lagging strand. Synthesis in this strand is more complicated than the Leading strand. DNA polymerase can add new free nucleotides to the 3' end of the new strand.

These DNA fragments are termed as ' Okazaki fragments' after the name of the scientist who first described the process of Discontinuous synthesis on Lagging strand.



This entire process of DNA replication occurs during S-phase of cell cycle in eukaryotes. Research is still going on for more detail on the replication process.

Transcription: The process of copying genetic information from one strand of the DNA into RNA is termed as transcription.

· In transcription only a segment of DNA and only one of the strands is copied into RNA because if both strands act a template, they would code for RNA molecule with different sequences and the sequences of amino acids in the coded protein would be different.

· The two RNA molecules would be complementary to each other and would form a double stranded RNA which would prevent translation.

Transcription unit: It consists of a Promoter, the Structural gene & a Terminator.

· The two strands of the DNA in the structural gene of a transcription unit is termed as template strand and coding strand.

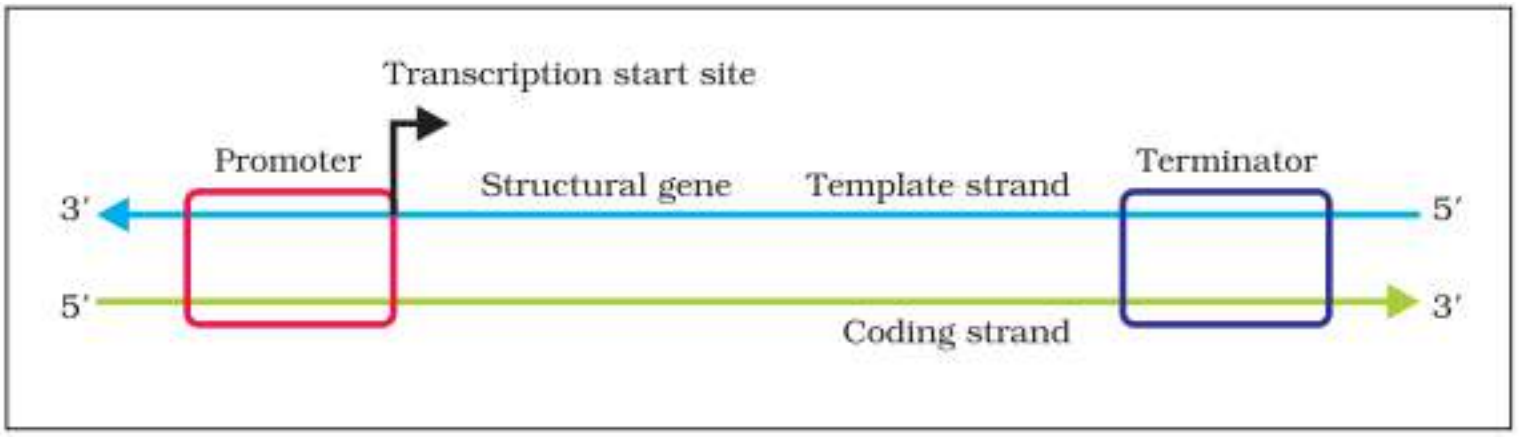
· The strand that has the polarity 3'→5' acts as a template, is referred as template strand.

· The other strand which has the polarity (5'→3') is referred as coding strand.

· The promoter and terminator flank the structural gene in a transcription unit.

· The promoter is located towards 5'– end (upstream) of the structural gene which provides binding site for RNA polymerase.

· The terminator is located towards 3'– end (downstream) of the coding strand which defines the end of the process of transcription.



Structure of a gene: A gene is defined as the functional unit of inheritance.

· A gene also referred as a cistron defined as a segment of DNA coding for a polypeptide.

· The structural gene in a transcription unit could be said as monocistronic mostly in eukaryotes or polycistronic mostly in bacteria or prokaryotes.

· Exons are the coding sequences or expressed sequences that appear in mature or processed RNA.

· Introns are the intervening sequences which interrupt exons and do not appear in mature or processed RNA.

Types of RNA: There are three major types of RNAs:–

1) mRNA (messenger RNA),

2) tRNA (transfer RNA),

3) rRNA (ribosomal RNA).

· All three RNAs are needed to synthesize a protein in a cell.

· The mRNA provides the template, tRNA brings amino acids and reads the genetic code, and rRNAs play structural and catalytic role during translation.

Process of transcription

1) In prokaryotes: Transcription takes place in three steps–

i. Initiation: RNA polymerase binds to promoter and initiates transcription.

· Initiation factor or sigma (σ) recognizes the promoter of the DNA.

ii. Elongation: RNA polymerase facilitates opening of the helix and continues elongation.

· RNA polymerase uses nucleoside triphosphates as substrate and polymerizes in a template depended fashion following the rule of complementarity.

· Only a short stretch of RNA remains bound to the enzyme.

iii. Termination: Once the polymerases reaches the terminator region RNA polymerase binds with the termination– factor (ρ) to terminate transcription.

· The nascent RNA falls off with the RNA polymerase which results in termination of transcription.

· The transcription and translation can be coupled in bacteria as the mRNA does not require any processing to become active, and also transcription and translation take place in the same compartment.

2) In eukaryotes: There are two additional complexities in eukaryotes.

i. The first complexity is that there are at least three RNA polymerases in the nucleus.

· The RNA polymerase I transcribes rRNAs (28S, 18S, and 5.8S).

· The RNA polymerase II transcribes precursor of mRNA, the heterogeneous nuclear RNA (hnRNA).

· The RNA polymerase III is responsible for transcription of tRNA, 5srRNA, and snRNAs (small nuclear RNAs).

ii. The second complexity is that the primary transcripts contain both the exons and the introns and are non-functional.

· Primary transcripts are subjected to a process called splicing where the introns are removed and exons are joined in a defined order.

· hnRNA undergo two additional processing called as capping and tailing.

· In capping an unusual nucleotide (methyl guanosine triphosphate) is added to the 5'- end of hnRNA.

· In tailing, adenylate residues (200-300) are added at 3'- end in a template independent manner and the fully processed hnRNA is called mRNA.

· mRNA is transported out of the nucleus for translation.

Significance of complexities: The split-gene arrangements represent probably an ancient feature of the genome.

· The presence of introns is reminiscent of antiquity, and the process of splicing represents the dominance of RNA-world.

Genetic code: The sequence of nucleotides on DNA which determines the sequence of amino acids in a polypeptide chain is termed as Genetic code.

· The process of translation requires transfer of genetic information from a polymer of nucleotides to a polymer of amino acids but there is no complementarity between nucleotides and amino acids which led to the proposition of a genetic code that could direct the sequence of amino acids during synthesis of proteins.

· The salient features of genetic code are as follows:-

· The codon is triplet, 61 codons code for amino acids and 3 codons do not code for any amino acids, hence they function as stop codons.

· One codon codes for only one amino acid thus it is unambiguous and specific.

· Some amino acids are coded by more than one codon, hence the code is degenerate.

· The codon is read in mRNA in a contiguous fashion, there are no punctuations.

· The code is nearly universal. For example, from bacteria to human UUU would code for Phenylalanine (phe).

· AUG has dual functions, it codes for Methionine (met), and it also act as initiator codon.

Mutation and genetic code: The relationships between genes and DNA are best understood by mutation studies.

· Point mutation is the insertion or deletion of a single gene in the structural gene.

· Frame shift mutation is the insertion and deletion of three or its multiple bases which insert or delete one or multiple

codon hence one or multiple amino acids, and reading frame remains unaltered from that point onwards.

Example– cystic fibrosis.

Translation: Translation refers to the process of polymerization of amino acids to form a polypeptide.

· The order and sequence of amino acids are defined by the sequence of bases in the mRNA and the amino acids are joined by a bond which is known as a peptide bond.

· Formation of a peptide bond requires energy and thus amino acids are activated in the presence of ATP and linked to their cognate tRNA by the process of charging of tRNA.

· If charged tRNAs are brought close enough, a peptide bond forms which is enhanced by the presence of a catalyst such as ribosome.

· Ribosome in its inactive state exists as two subunits; a large subunit and a small subunit.

· There are two sites in the large subunit, for subsequent amino acids to bind to and thus become close enough to each other for the formation of a peptide bond.

· A translational unit in mRNA is the sequence of RNA that is flanked by the start codon (AUG) and the stop codon and codes for a polypeptide.

· An mRNA also has some additional sequences that are not translated and are referred as untranslated regions (UTR).

· After activation of amino acids, translation starts with its three steps.

i. Initiation: the ribosome binds to the mRNA at the start codon (AUG) that is recognised only by the initiator tRNA.

ii. Elongation: The ribosome proceeds to the elongation phase of protein synthesis.

· The ribosome moves from codon to codon along the mRNA.

· Amino acids are added one by one, translated into Polypeptide sequences dictated by DNA and represented by mRNA.

iii. Termination: At the end, a release factor binds to the stop codon, terminating translation and releasing the complete polypeptide from the ribosome.

Regulation of gene expression: Gene expression results in the formation of a polypeptide and it can be regulated at several levels such as–

· transcriptional level (formation of primary transcript),

· processing level (regulation of splicing),

· transport of mRNA from nucleus to the cytoplasm,

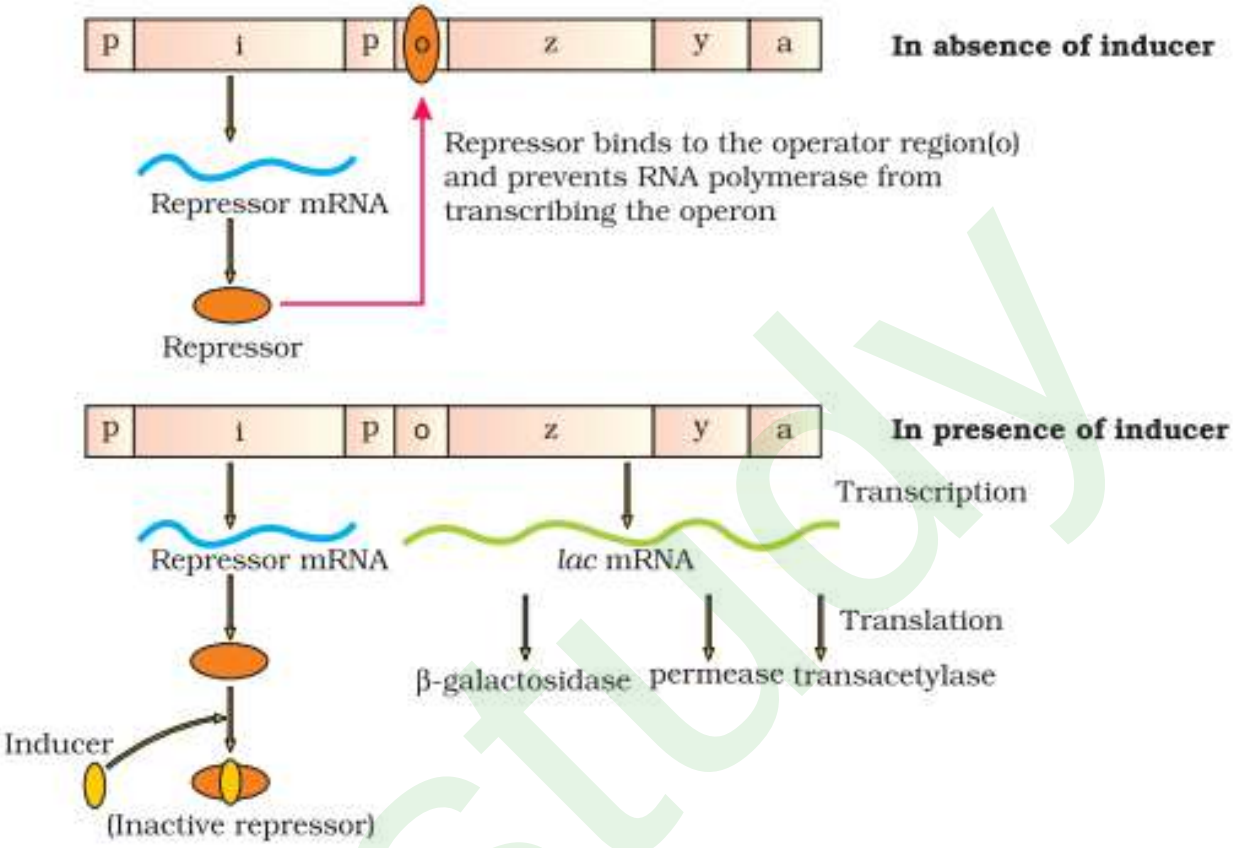
· translational level.

In prokaryotes, control of the rate of transcriptional initiation is the predominant site for control of gene expression.

In a transcription unit, the activity of RNA polymerase at a given promoter is in turn regulated by interaction with accessory proteins which can act both positively (activators) and negatively (repressors). Regulation of gene expression can be studied with the help of Lac operon.

Lac operon: The lac operon consists of one regulatory gene the i gene which codes for the repressor of the lac operon and three structural genes (z, y, and a).

· The z gene codes for beta– galactosidase (β– gal), which hydrolyses disaccharide, lactose into galactose and glucose.



· The y gene codes for permease, which increases permeability of the cell to β galactosides.

· The a gene encodes a transacetylase.

· Lactose is termed as inducer as lactose is the substrate for the enzyme beta– galactosidase and it regulates switching on and off of the operon.

· In the absence of inducer, The repressor of the operon is synthesized (all– the– time – constitutively) from the i gene.

· The repressor protein binds to the operator region of the operon and prevents RNA polymerase from transcribing the operon.

· In the presence of inducer, The repressor is inactivated by interaction with the inducer which allows RNA polymerase access to the promoter and transcription proceeds.

· Regulation of lac operon by repressor is referred to as negative regulation.

Human genome project: The scientific project which deal with the study of base sequences of DNA molecules of complete set of chromosomes is called human genome project.

· HGP was closely associated with the rapid development of a new area in biology called as bioinformatics.

➤ Goals of Human Genome Project

· Identify all the approximately 20,000–25,000 genes in human DNA.

· Determine the sequences of the 3 billion chemical base pairs that make up human DNA.

· Store this information in databases;

· Improve tools for data analysis;

· Transfer related technologies to other sectors, such as industries;

· Address the ethical, legal, and social issues (ELSI) that may arise from the project.

Methodologies

· To identifying all the genes that expressed as RNA referred to as ESTs.

· Sequencing the whole set of genome that contained all the coding and non– coding sequence, and later assigning different regions in the sequence with functions is called as Sequence Annotation.

· The commonly used vectors are BAC (bacterial artificial chromosomes), and YAC (yeast artificial chromosomes).

· The sequences were subsequently annotated and were assigned to each chromosome.

➢ Salient Features of Human Genome

· The human genome contains 3164.7 million nucleotide bases.

· The average gene consists of 3000 bases with the largest known human gene being dystrophin at 2.4 million bases.

· The total number of genes is estimated at 30,000.

· 9 per cent nucleotide bases are exactly the same in all people.

· The functions are unknown for over 50 per cent of discovered genes.

· Less than 2 per cent of the genome codes for proteins.

· Chromosome 1 has most genes (2968), and the Y has the fewest (231).

· Scientists have identified about 1.4 million locations where single base DNA differences (SNPs – single nucleotide polymorphism) occur in humans.

➤ Applications of HGP

· All the genes in a genome can be studied together.

· Helps to understand how tens of thousands of genes and proteins work together in interconnected networks.

· Helps to diagnose and treat genetic diseases.

DNA fingerprinting: The process of comparison of DNA from different sources to establish the identity is called DNA fingerprinting.

· DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as repetitive DNA.

· The bulk DNA forms a major peak and the other small peaks are referred to as satellite DNA.

· Satellite DNA is of two types based on base composition, length of segment, and number of repetitive units.

i. micro– satellites.

ii. mini– satellites.

Satellite DNA: Satellite DNA sequences normally do not code for any proteins, but they form a large portion of human genome.

· An inheritable mutation occurring in a population at high frequency, is referred to as DNA polymorphism.

· Repeated nucleotide sequences in the non–coding DNA of an individual is called Variable Number of Tandem Repeats (VNTR).

· The size of VNTR varies in size from 0.1 to 20 kb.

· DNA fingerprinting includes the following steps–

i. isolation of DNA.

ii. digestion of DNA by restriction endonucleases.

iii. separation of DNA fragments by electrophoresis.

iv. transferring (blotting) of separated DNA fragments to synthetic membranes, such as nitrocellulose or nylon.

v. hybridization using labeled VNTR probe.

vi. detection of hybridized DNA fragments by autoradiography.

Applications–

· In identification of criminals.

· In determining population and genetic diversities.

· In solving parental disputes.