CHAPTER 6

MOLECULAR BASIS OF INHERITANCE

There are two types of nucleic acids present in living organisms. They are- Ribonucleic acid (RNA) and Deoxyribonucleic acid (DNA).

The length of DNA in some organisms

Bacteriophage ϕX 174 has 5386 nucleotides
Bacteriophage lambda has 48,502 base pairs (bp)
Escherichia coli has 4.6×10^6 bp
Haploid cell of Human DNA contains 3.3×10^9 bp.

The structure of POLYNUCLEOTIDE CHAIN (DNA AND RNA)

A nucleotide is the basic unit of polynucleotide chain of DNA or RNA.
Each nucleotide is composed of three components:

i) A nitrogenous base
ii) Pentose sugar (ribose in RNA, and deoxyribose in DNA) and

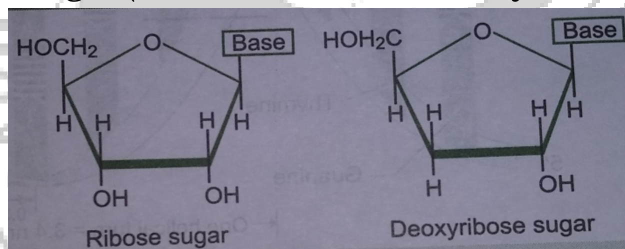


Figure: Types of sugar molecules.

iii) A phosphate group

Nitrogenous base: It is of two types, purines (Adenine and Guanine) and pyrimidines (Cytosine and Thymine in case of DNA and instead of Thymine, Uracil is present in RNA).

A nitrogenous base is attached to the pentose sugar by an N-glycosidic linkage to form a nucleoside.

When a phosphate group is attached to 5’-OH of a nucleoside through phosphodiester linkage, a nucleotide is formed.

(**nucleotide** = nucleoside +phosphate; **nucleoside** =pentose sugar + a nitrogenous base)

Two nucleotides are joined through 3’-5’ phosphodiester linkage and a dinucleotide is formed. Thus, when numerous nucleotides are joined, a polynucleotide chain id formed.

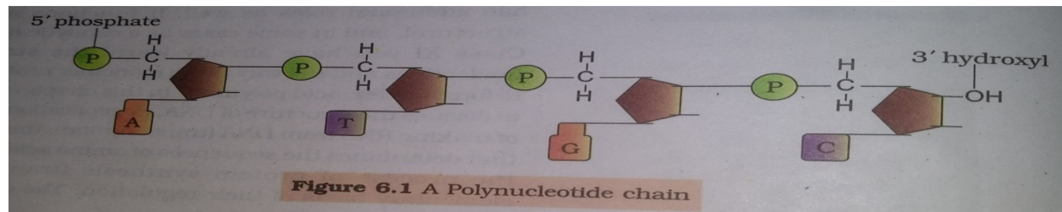


Figure 6.1 A Polynucleotide chain

Free OH is present on 5’ (five prime) end as well as on 3’ (three prime) end of the polynucleotide chain.
Sugar and phosphate constitute the backbone of polynucleotide chain and nitrogenous bases are linked to sugar moiety which projects from the backbone.

SALIENT FEATURES OF DOUBLE HELICAL DNA

James Watson and Francis Crick in 1953 proposed the double helix model of DNA based on the X-ray diffraction data produced by Maurice Wilkins and Rosalind Franklin and Erwin Chargaff’s rules of base pairing.
Chargaff’s rule: i) The amount of adenine is always equal to the amount of thymine, and the amount of guanine is always equal to the amount of cytosine i.e., [A=T], [G=C]
ii) Adenine always pairs with thymine with a double hydrogen bond, and guanine always pairs with cytosine with a triple hydrogen bond.
iii) The ratio of adenine and guanine to that of thymine and cytosine is always equal to one, i.e., [A=G]

[T+C]

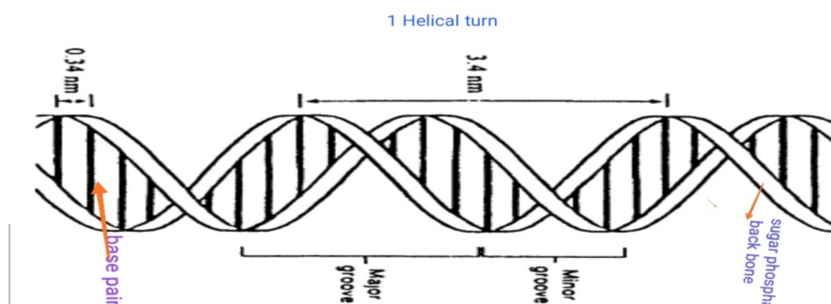


Fig: DNA double helix

Some features of DNA

i) DNA is made up of two polynucleotide chains, where the backbone is made up of sugar and phosphate groups and the nitrogenous bases project towards the centre.

ii) There is complementary base pairing between the two strands of DNA.

iii) The two strand are coiled in right handed fashion and are anti-parallel in orientation. One chain has 5'→3' polarity and the other has 3'→5'.

iv) The diameter of the DNA double helix is always constant due to pairing between purine and pyrimidine i.e., pairing between A and T, which are complementary and, between G and C which are complementary.

v) The distance between two consecutive base pairs in a helix is 0.34 nm and 1 helical turn contains approximately 10 base pairs, so length of 1 helical turn is 3.4 nm.

CENTRAL DOGMA

Francis Crick proposed the central dogma of molecular biology which states that genetic information flows from DNA to $_{m}$RNA (transcription) and then from $_{m}$RNA to polypeptide chain/protein (translation) always unidirectionally (except in some viruses it is bidirectional and the process is called reverse transcription /teminism).

Packaging of DNA Roger Kornberg (1974) reported that chromosome

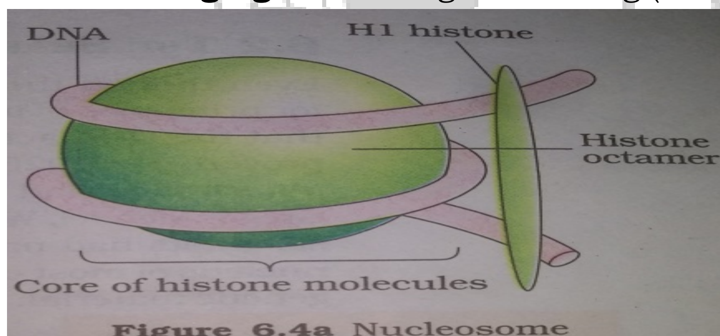


Figure 6.4a Nucleosome

is made up of DNA and protein.

Later Beadle and Tatum reported that chromatin fibres looks lie beads on a string, where beads re repeated units of proteins.

The proteins associated with DNA are of two types-basic proteins (histone and protamine) and acidic non-histone chromosomal (NHC) proteins.

The negatively charged DNA molecules wraps around the positively charged histone proteins to form a structure called **nucleosome.**

The nucleosome core is made up of four types of histone proteins (H_2A, H_2B, H_3, H_4 – occurring in pairs).

About 200bp DNA helix wraps around the nucleosome by $1^{3}/_{4}$ turns, plugged by H_1, histone protein. Repeating units of nucleosomes form the chromatin in nucleus, which is thread like structure. The chromatin is packed to form a solenoid structure of 30 nm in diameter. Solenoid further condenses to form a looped structure called the chromatin fibre of about 700 nm in diameter. These chromatin fibres further coil and condense at metaphase stage of cell division to form chromosome, about 1400 nm in diameter.

Difference between Euchromatin and Heterochromatin

Sl no	Euchromatin	Heterochromatin
i)	These are regions of chromatin which are loosely packed	These are regions of chromatin which are densely packed
ii)	These are stained lightly with Feulgen stain.	These are intensely/dark stained with Feulgen stain.
iii)	Euchromatin contains active genes.	Heterochromatin contains inactive genes.
iv)	They do not contain repetitive DNA sequences.	They are enriched with highly repetitive tandemly arranged DNA sequences.
v)	It is transcriptionally active.	It is transcriptionally inactive.

Transforming principle

Frederick Griffith (1928) conducted experiments with bacterium *Streptococcus pneumonia* which cause pneumonia. He observed two strains – one smooth (s) shiny colonies with capsule and the other rough (R) without capsule.

Griffith's experiment

1. When live S-type cells are injected into mice, they died due to pneumonia.
2. When live R-type cells are injected into mice, they survived.
3. When heat-killed S-type cells were injected, the mice survived. When heat killed S-type cells were mixed with R-type cells and injected to the mice, they died, and live S-type cells were obtained from dead mice.
4. He concluded that heat-killed S-type transformed the R-type into S-type but was unable to understand the cause of this bacterial transformation.

BIOCHEMICAL CHARACTERISATION OF TRANSFORMING PRINCIPLE

Oswald Avery, Colin Macleod and Maclyn McCarthy repeated Griffith's experiment in an *in vitro* system in order to determine biochemical nature of transforming principle.
They reported that DNA from heat-killed S-type bacteria caused the transformation of non-virulent R-type bacteria into virulent S-type bacteria.
They also discovered that proteases and RNAses did not affect transformation while DNAse inhibit the process.
They concluded that DNA is the hereditary material.
PROOF OF DNA AS THE GENETIC MATERIAL.
Hershey and Chase conducted experiments on bacteriophage to prove that DNA is the genetic material.

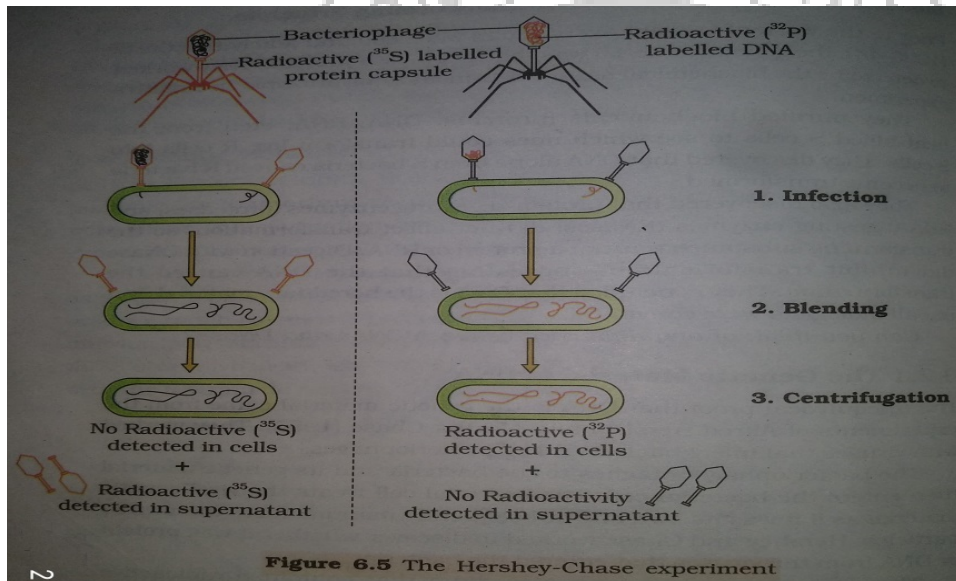


Figure 6.5 The Hershey-Chase experiment

PROCEDURE:
1. Some bacteriophage virus were grown on a medium that contained radioactive phosphorus (^{32}P) and some are grown in medium containing radioactive sulphur (^{35}S).
2. Viruses grown in ^{32}P containing medium contain radioactive DNA.
3. Viruses grown in ^{35}S containing medium contain radioactive protein coat.
4. Both types of bacteriophages are allowed to infect bacterial cells separately.

5. After infection, the bacterial cells were gently agitated in blender to remove bacteriophage cells from bacteria.
6. Then the culture were also centrifuged to separate viral particles from bacterial cells.

Observation and conclusion:

1. Only ^{32}P was found in bacterial cells and ^{35}S were found only in surrounding medium.
2. This indicates that only DNA and not protein coat entered the bacterial cell.
3. This proves that DNA is the genetic material which is passed from the virus to the bacterial cells and not the protein.

CHARACTERISTICS OF GENETIC MATERIAL

1. It should be able to generate its own replicas.
2. It should be chemically and structurally stable.
3. It should have room for mutation for bringing evolution.
4. It should be able to store genetic information which can be inherited.
5. It should be able to express itself in the form of 'Mendelian characters'.

RIBONUCLEIC ACID (RNA) world

1. RNA was the first genetic material and is single stranded.
2. RNA behaves as genetic material as well as a catalyst.
3. RNA can synthesize DNA (reverse transcription) and it is estimated that DNA has evolved from RNA from chemical modification.
4. The backbone of RNA is made up of pentose sugar (ribose) and a phosphate.
5. The 2'-OH group of ribose makes RNA labile and easily degradable.
6. RNA has two types of nitrogenous bases: i) purines-Adenine and Guanine. ii) pyrimidines-Cytosine and Uracil.

Ribonucleotide = Ribonucleoside + phosphate
Ribonucleoside = Ribose sugar + Nitrogenous base

DIFFERENCES BETWEEN DNA AND RNA

Sl no	DNA	RNA
1	Sugar present is deoxyribose.	Sugar present is ribose.
2	Nitrogenous bases are Adenine, Guanine, Thymine and Cytosine	Nitrogenous bases are Adenine, Guanine, Uracil and Cytosine
3	It is always double stranded	It is single stranded and or double

		stranded (tRNA folded structure)
4	It is the genetic material of almost all the living organisms	It is the genetic material of a few viruses.
5	It is chemically less reactive and structurally more stable.	It is chemically more reactive and structurally less stable.

DNA Replication

Watson and Crick in 1953 proposed that DNA replication is semi-conservative in nature.

Here, the two parental strands separate and each strand acts as a template for synthesising a complementary strand over it. After completion of replication, each DNA had one parental strand and one newly synthesised strand.

i) **Experimental proof of semi-conservative mode of DNA replication**: Mathew Meselson and Franklin Stahl in 1958 performed experiments on *E. coli* to prove the semiconservative mode of DNA replication.
They grew *E. coli* in a medium containing $^{15}NH_4Cl$ (^{15}N is heavy isotope) for many generations.
As a result ^{15}N got incorporated into newly synthesized DNA.
This heavy DNA can be differentiated from the normal DNA by centrifugation in Caesium Chloride (CsCl) Density gradient.
Then they are transferred into a normal $^{14}NH_4Cl$ containing medium and took the samples at various definite time intervals as the cells multiply.
The extracted DNAs were centrifuged and measure their densities.
The DNA extracted after one generation of transfer from ^{15}N medium to ^{14}N medium (i.e., after 20 minutes; *E.coli* divides after every 20 minutes) showed an intermediate hybrid density.
The DNA extracted after 40 minutes showed equal amount of light/normal DNA and heavy DNA.
Similar experiment on *Vicia faba* by Taylor and colleagues in 1958 prove DNA in chromosome also replicate semi-conservatively.

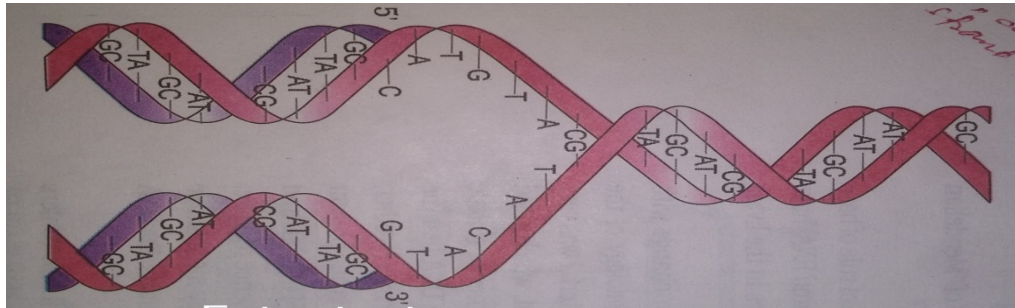


Fig: Watson and Crick's model for semiconservative DNA replication

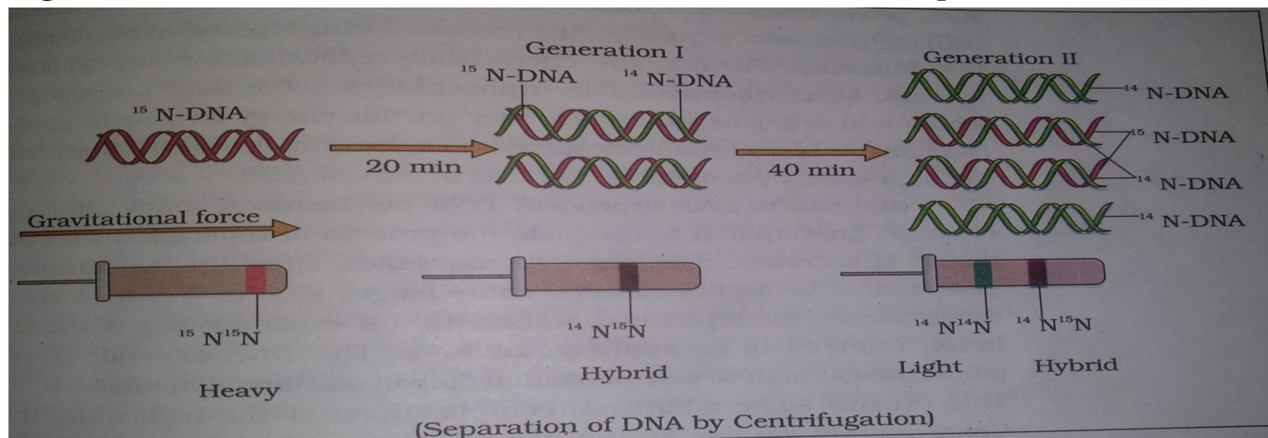


Fig: Meselson and Stahl's Experiment.

ii) **Enzymes for DNA replication**

DNA-dependant DNA polymerase catalyses the polymerisation of deoxynucleotides on DNA template.

Helicase unwinds the DNA to form the replication fork.

DNA ligase joins the okazaki fragments.

iii) **Process of replication**

DNA replication begins at a unique and fixed point called 'origin of replication' or 'ori'.

INITIATION

The complementary strands of DNA double helix are separated by DNA gyrase and DNA helicase. This is called unwinding of double stranded DNA.

The separated strands tend to rewind, therefore these are stabilised by single stranded binding proteins (ssBPs), which bind to the separated strands**. As** the unwinding continues more tension is built up, topoisomerase enzyme cuts and reseals the strands to release the built up tension on the strands.

Unwinding of the double stranded DNA forms a Y-shaped structure in the DNA-duplex, which is known as replication fork.

ELONGATION

An enzyme called RNA primase initiates replication on the template strand in 3’→5’ direction. This generates about 10-60 nucleotide long primer RNA in 5’→3’ direction.

The free 3’-OH of this RNA primer provides the initiation point for DNA polymerase for sequential addition of deoxyribonucleotides.

DNA polymerase progressively adds deoxyribonucleotides to the free 3’-end of the growing polynucleotide chain so that replication of 3’→5’ strand of template DNA molecule is continuous and the complementary newly synthesised DNA is called Continuous strand/ leading strand. While the replication of the second strand (5’→3’ template strand) of the DNA molecules is discontinuous and it is called lagging strand.

The lagging strand is made up of small polynucleotide fragments called the ‘Okazaki fragments’ (after R. Okazaki, who first identifies them) which are joined together by ligase enzyme.

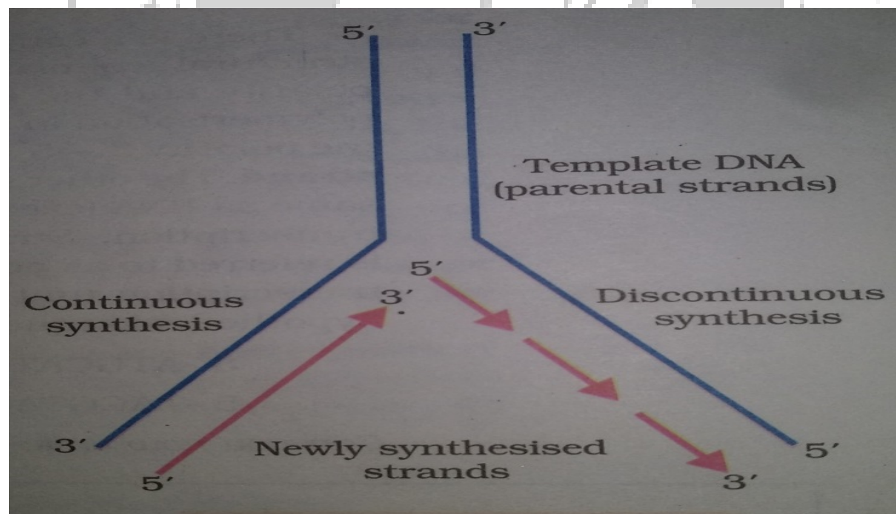


Fig: Replication fork

Transcription

The process of copying genetic information from one strand of DNA into RNA is called transcription.
The principle of complimentarity governs t the process, except that adenosine new base pairs with uracil instead of thymine, as in replication. Unlike replication, only a single stranded fragment of DNA gets copied into RNA.

TRANSCRIPTION UNIT

The transcription unit of DNA has three regions:

i) A promoter: It is the binding site for RNA polymerase for initiation of transcription and is located towards 5' end of coding strand, upstream of the structural gene.
ii) A structural gene: It codes for enzymes or protein for structural functions.
iii) A terminator: It is the region where transcription ends and is located towards 3'-end of coding strand, downstream of the structural gene.

The DNA-dependant RNA polymerase helps in RNA transcription by catalysing the polymerisation in 5'→3' direction.
The DNA strand that has the 3'→5' direction acts as template and is called the template strand.
The DNA strand that has 5'→3' direction is called coding strand and has the base sequence similar to the RNA thus formed.

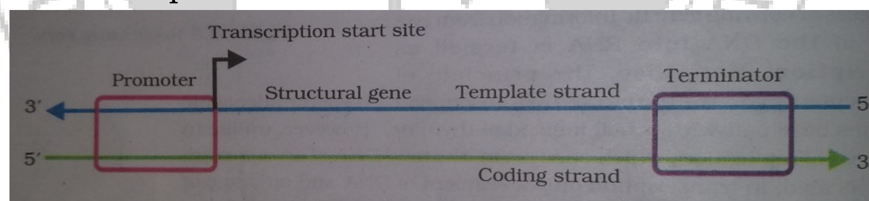


Fig: Schematic structure of a transcription unit

TRANSCRIPTION UNIT AND THE GENE

The segment of DNA coding for a polypeptide is called cistron.
In eukaryotes, the transcription unit possess a structural gene specific only for a single polypeptide, so it is called monocistronic.
In prokaryotes, the transcription unit possess structural genes coding for many polypeptides which are part of a single metabolic pathway so, it is called polycistronic.
The gene in eukaryote is split into coding regions known as **exons** and non-coding regions which are known as **introns**.

TRANSCRIPTION IN PROKARYOTES

In prokaryotes, structural gene is polycistronic and continuous.

In bacteria, the transcription of all the three types of RNA (*m*RNA, *t*RNA and *r*RNA) is catalysed by a single DNA-dependant enzyme, called the RNA polymerase which has co-factors β, β', α, α', ω and σ (sigma) to catalyse the process.

The transcription is completed in three steps.

Initiation : σ factor recognises the start signal and promoter region, and binds to DNA along with RNA polymerase to initiate transcription.

Elongation: The RNA polymerase loses σ factor after initiation, but continues the polymerisation of ribonucleotides to form RNA.

Termination : Once the RNA polymerase reaches the termination region of DNA, the RNA polymerase is separated from DNA-RNA hybrid, as a result nascent RNA separates. This process is called termination and is facilitated by rho (ρ) factor.

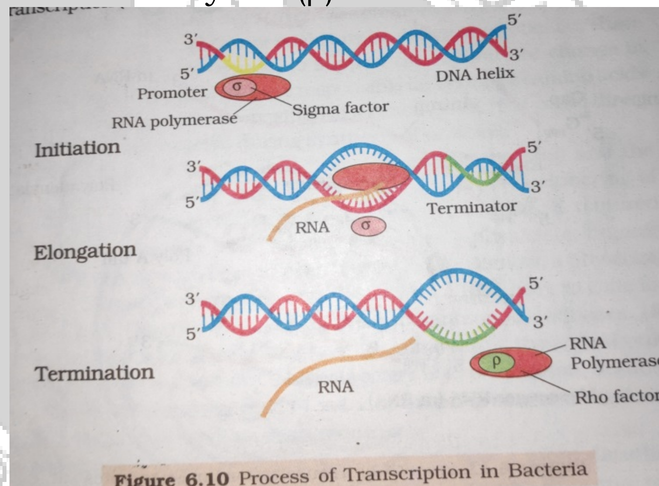


Figure 6.10 Process of Transcription in Bacteria

Transcription in Eukaryotes

The process of transcription is similar to that of prokaryotes and takes place in nucleus.

Coding genes called exons formed the part of *m*RNA and non-coding sequences called introns are removed during splicing.

In eukaryotes, three types of RNA polymerase are found in the nucleus: i) RNA polymerase I transcribes *r*RNA (28S, 18S and 5S). ii) RNA polymerase II transcribes the precursor *m*RNA (heterogenous nuclear RNA or *hn*RNA). iii) RNA polymerase III transcribes tRNA, 5S *r*RNA and *sn*RNAs (small nuclear RNAs).

POST TRANSCRIPTIONAL MODIFICATIONS

The primary transcripts are non-functional, containing both exon and introns in RNA and are called *hn*RNA. The *hn*RNA undergoes capping process where an unusual methyl guanosine triphosphate is added to the 5'-end of RNA and tailing processes where adenylate residues (about 200-300) are added at 3'-end of RNA in a template independent manner.
Now the *hn*RNA undergoes a process where introns are removed and exons are joined to form *m*RNA and the process is called splicing.

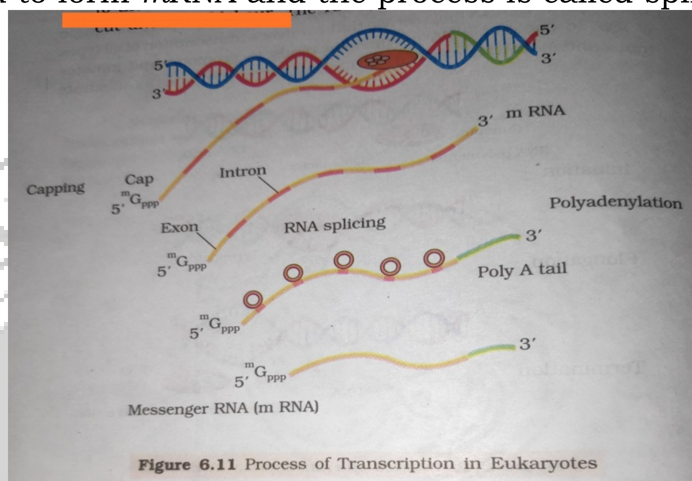


Figure 6.11 Process of Transcription in Eukaryotes

GENETIC CODE

The relationship between the sequence of nucleotides in the *m*RNA and the sequence of amino-acids in the polypeptide chain is called genetic code. It is made up of codons.
George Gamow suggested codon should be triplet (must contain only 3 nitrogenous bases as there are only 4 nitrogenous bases to code for 20 amino acids i.e., 4^3=4x4x4=64).
Har Govind Khorana developed chemical method for synthesizing RNA molecules with defined base combinations (homopolymers and copolymers) to develop the genetic code.
Marshal Nirenberg artificially synthesized proteins to understand the nature of codons.
Servo Ochoa demonstrated that polynucleotide phosphorylase enzyme helps in polymerising RNA with defined sequences in a template independent manner (enzymatic synthesis of RNA).

Table 6.1: The Codons for the Various Amino Acids

First position	Second position: U	C	A	G	Third position
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 codon

1. The codon is triplet. 61 codons code for amino acids. 3 codons (UAA, UAG and UGA) do not code for any amino acids hence function as terminating or and stop codons.
2. **One** codon codes only for a particular amino-acid, hence it is unambiguous or **specific**
3. Some amino acids are coded by more than one codon, hence the code is degenerate.
4. There are no punctuations in a reading frame and read continuously or in a contiguous fashion.
5. The code is universal i.e., a particular codon codes for the same amino acid in all the living organisms (except in mitochondria and some protozoans).
6. AUG has dual functions. It codes for methionine (met) and also acts as initiator codon.

MUTATIONS

Mutation is defined as the sudden inheritable change in the genetic material. It can be of the following types: i) Point mutation: It is the mutation in a single base pair , which is replaced by another base pair. For

example, in sickle-cell anaemia, point mutation in beta globin gene results in the change of amino acid **glutamate to valine.**

ii) Frameshift mutation: It is the change in the reading frame due to insertion or deletion'

Insertion: It is the addition of one or more nucleotides in the DNA segment. Insertion of three or its multiple bases do not change the reading frame but add new amino acid/s.

Deletion: It is the removal of one or more nucleotides from the DNA segment. Deletion of three or its multiple bases do not change the reading frame but removes amino acid/s.

NORMAL DNA reading frame: **ATC GAT CGA**
INSERTION: IF 'C' is added between C and G at the 3rd and 4th position- the reading frame becomes
ATC CGA TCG A
DELETION: IF 'G' at the 4th position in the reading frame is removed it becomes **ATC ATC GA**

tRNA-the adaptor molecule

Francis Crick proposed the presence of an adaptor molecule which could read the code on one hand and on the other end would bind to specific amino acids.

However *t*RNA was known before genetic code was postulated and was called *s*RNA (solute RNA). Its role as an adaptor was reported later.

The secondary structure of *t*RNA is a cloverleaf like, but the three-dimensional tertiary structure is a compact L-shaped molecule.

*t*RNA has five loops or arms. i) Anticodon loop has bases complementary to the code on reading frame.

ii) Amino acid acceptor end to which amino acid binds.

iii)T loop helps in binding to ribosome.

iv) D loop helps in binding amino acyl synthetase.

v) Variable loop.

Each *t*RNA is specific for a particular amino acid. A specific *t*RNA for initiation is called initiator *t*RNA. there are no *t*RNA for stop codons.

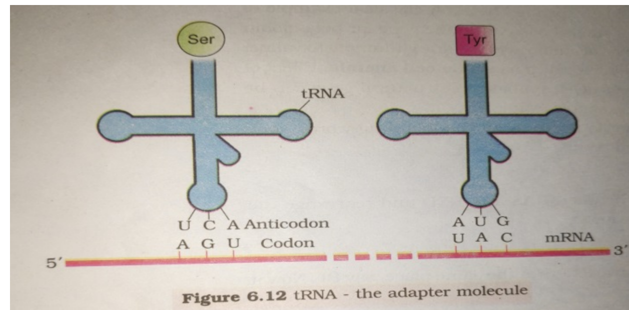


Figure 6.12 tRNA - the adapter molecule

Translation

Translation is the process of synthesis of protein from *m*RNA with the help of ribosome.

A translational unit in *m*RNA from 5’→3’ comprises of a start codon, region coding for a polypeptide, a stop codon and untranslated regions at both 5’-end and 3’-end for efficient process.

There are three stages of protein synthesis. i) Initiation: Assembly of ribosome on *m*RNA. Activation of amino acids and its delivery to *t*RNA.

ii) Elongation: Repeated cycle of amino acid delivery. Peptide bond formation and movement along *m*RNA called translocation.

iii)Termination: The release of polypeptide chain.

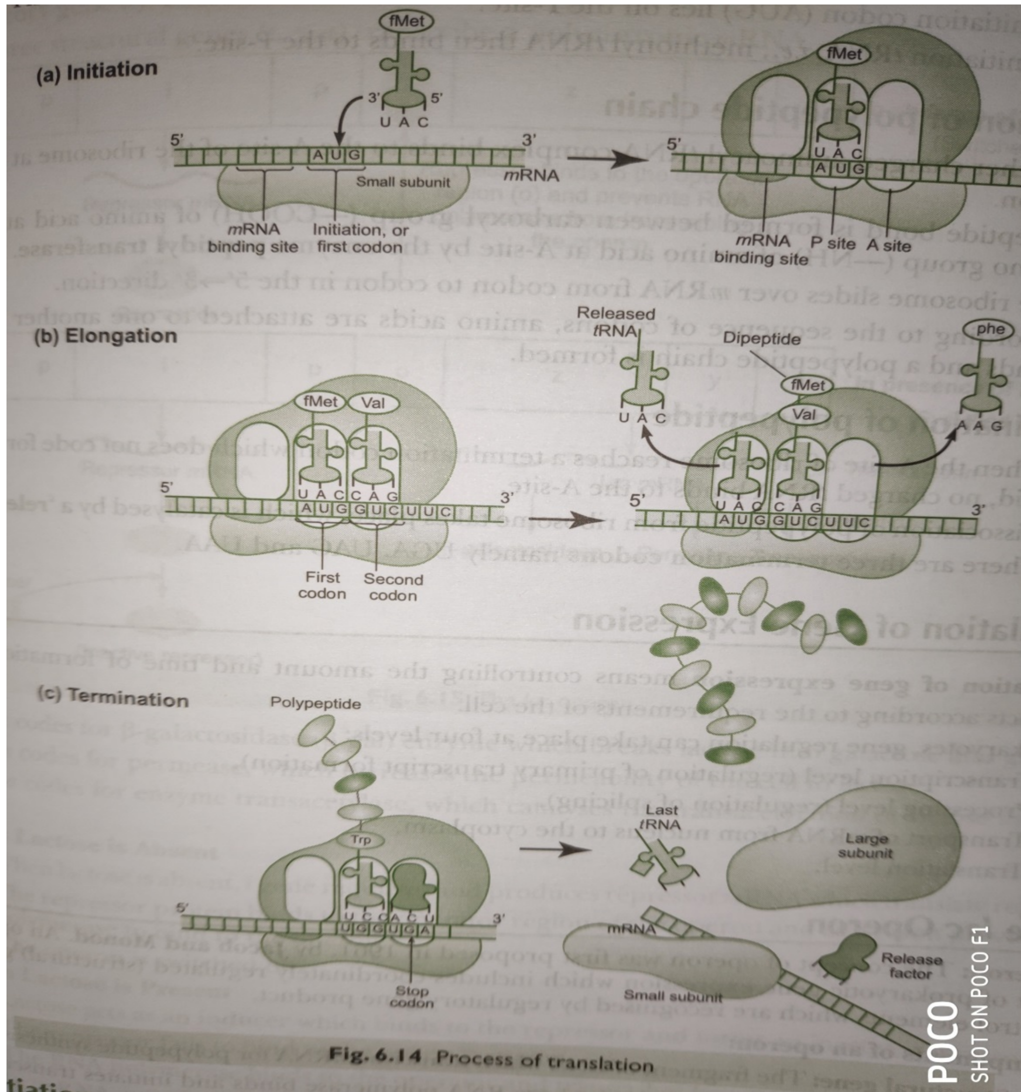


Fig. 6.14 Process of translation

POCO
SHOT ON POCO F1

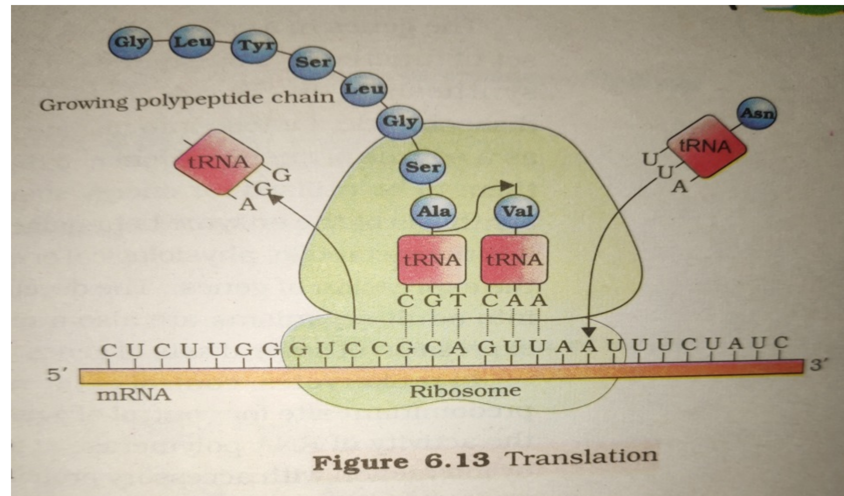
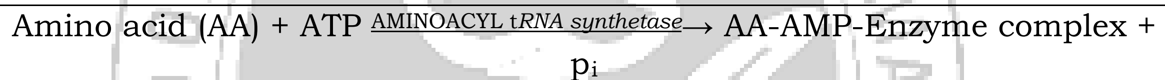


Figure 6.13 Translation

i) Initiation

In prokaryotes, initiation requires the *m*RNA, the small and large sub-units of ribosome, three initiation factors (IFS) and initiator *t*RNA.

Activation of amino acid: Amino acid becomes activated by binding with aminoacyl RNA synthetase enzyme in presence of ATP.



Amino acid (AA) + ATP $\xrightarrow{\text{AMINOACYL } t\text{RNA synthetase}}$ AA-AMP-Enzyme complex + p_i

Transfer of amino acid to *t*RNA: The AA-AMP complex formed reacts with specific *t*RNA to form aminoacyl-*t*RNA complex.

AA-AMP-Enzyme complex +*t*RNA → AA-*t*RNA + AMP + Enzyme

The cap region of *m*RNA binds to the subunit of ribosome.

The bigger subunit of ribosome has three sites E-site (exit), A-site (aminoacyl site) and P-site (peptidyl site).

The smaller subunit first binds to the initiator *m*RNA and then binds to larger subunit so that initiation codon (AUG) lies on P-site.

The initiator *t*RNA i.e., metionyl*t*RNA then binds to the P-site.

ii) Elongation of polypeptide chain

Another charged aminoacyl *t*RNA complex binds to the A-site of the ribosome at the second codon.

A peptide bond is formed between carboxyl group (-COOH) of amino acid at P-site and amino group (-NH) of amino acid at A-site by the enzyme peptidyl transferase.

The ribosome slides over *m*RNA from codon to codon in 5’→3’ direction, but actually the *m*RNA is pulled i.e., mRNA moves and not the ribosome.

According to sequence of codons, amino acids are attached to one another by peptide bonds and a polypeptide chain is formed.

iii) Termination of polypeptide chain

When A-site reaches a termination codon (either of the three-UAA, UAG, UGA) no charged *t*RNA binds to the A-site.

The dissociation of polypeptide from ribosome and also dissociation of large and small subunit also takes place due by addition of release factor.

Regulation of Gene Expression

Regulation/control of gene expression means the mechanism allowing the expression of certain genes but at the same time suppressing the expression of other genes as required in the cell.

In Eukaryotes gene expression can be regulated at four levels:

i) Transcription level (regulation of primary transcript formation).
ii) Processing level (regulation of splicing)
iii) Transport of *m*RNA from nucleus to the cytoplasm
iv) Translation level.

Example of gene regulation in prokaryotes – The lac Operon

<u>Operon</u>: The concept of operon was first proposed by Jacob and Monod in 1961.An operon is a segment of DNA that comprises of one or more adjacent structural genes, an operator gene and a regulator gene. Operon works in cooperation with a repressor(gene product) and or an inducer substance.

<u>Components of operon</u>:

<u>i)</u> Structural genes: The fragment of DNA which transcribes polypeptide synthesis.

<u>ii)</u> Promoter: The fragment of DNA where RNA polymerase binds and initiates transcription of structural genes is promoter.

<u>iii)</u> Operator: The fragment of DNA adjacent to promoter where specific repressor protein binds is called operator.

<u>iv)</u> Regulator gene: The gene that codes for the repressor protein that binds to the operator and suppress its activity as a result of which transcription will be switched off.

<u>v)</u> Inducer: The substrate that prevents the repressor from binding to operator, is called an inducer. As a result transcription is switched on. It a chemical of diverse nature like metabolite hormones etc.

The lac/lactose operon: The *lac z, lac y, lac a* genes are transcribed from a *lac* transcription unit under the control of a single promoter. They encode enzyme required for the use of lactose as a carbon

source. The *lac i* gene product, the *lac* repressor, is expressed from a separate transcription unit upstream from the operator.

Lac operon consists of three structural genes (*z, y, a)*, operator (*o*), promoter (*p)* and a separate regulator gene (*i*). Lactose is the inducer in *lac* operon. The three structural genes (*z,y,a*) transcribe a polycistronic *m*RNA.

Gene *z* codes for β-galactosidase enzyme which breaks lactose into galactose and glucose.

Gene *y* codes for permease, which increases the permeability of of the lactose to the cell.

Gene *a* codes for enzyme transacetylase, which catalyses the transacetylation of lactose in its active form.

Mechanism of *Lac* operon

When lactose is absent-i) *i* gene produces repressor protein, and binds to promoter region which suppresses the transcription of structural genes *lac (z, y, a)*. So, the operon is switched off.

When lactose is present it acts as an inducer by binding to the repressor which results in changing the conformity of the repressor. Thus repressor now cannot bind to the operator giving way to the RNA polymerase to bind and thus transcribes the structural genes *lac (z, y, a)* and these translates into their respective enzymes i.e., β galactosidase, permease and transacetylase. So, the operon is switched on.

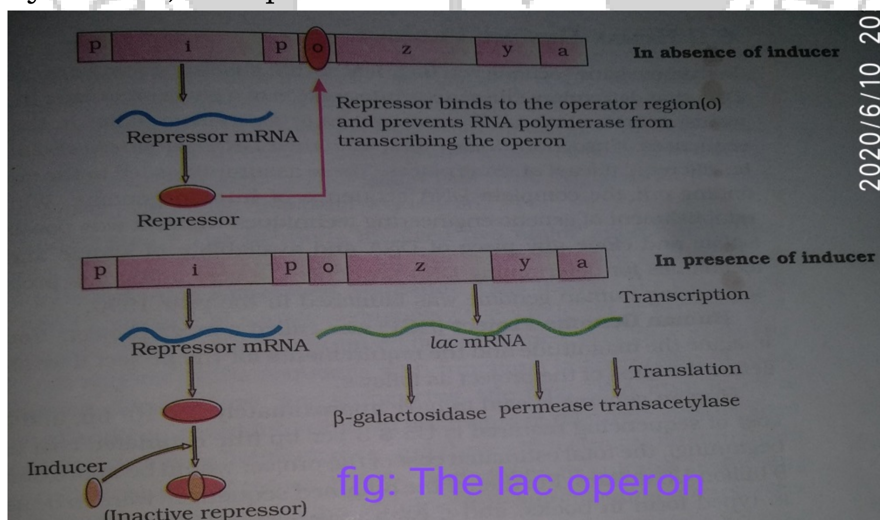


fig: The lac operon

Human Genome Project (HGP)

Human genome project is a hugh/mega project to sequence the genome of human being. The total cost of the project was 9 billion US dollar and an

enormous amount of data was generated, if written in typed form of book, it would require 3300 books of 1000 pages each and each page having 100 letters. The project was a 13-year project by the US Department of Energy and National Institute of Health. It was launched in 1990 and completed in 2003. During the early years of HGP, the Welcome Trust (UK) became a major partner and additional contribution came from Japan, France, Germany and China. HGP was closely associated with development of bioinformatics.

Goals of HGP

i) To identify 20000-25000 genes in human.

ii) To determine all the 3 billion base pair sequences of human genome DNA

iii) To store the above information in database.

iv) To improve tools for data analysis.

v) To transfer the technologies to other sectors such as industries.

vi) To address the ethical, legal and social issues (ELSI) that may arise from the project.

Advantages of HGP

i) Provides clues to understand human biology.

ii) The effect of DNA variation can be studied among individuals which can lead to new ways to diagnose and treat many disorders or diseases.

iii) More information can be obtained about other organisms like bacteria, yeast, nematode, fruit fly, rice and other plants etc.

Methodologies of HGP

Two major approaches are there for sequencing human genome:

i) EXPRESSED SEQUENCE TAGS (ESTs): This method focusses on identifying all the genes that are expressed .

ii) SEQUENCE ANNOTATION: this is the method of sequencing the whole set of genome both coding and non-coding regions and later assigning different regions in the sequence with functions. The second method is used in HGP.

For sequencing, the total DNA from the cell is isolated and broken into relatively small sizes as fragments.

These DNA fragments are cloned in suitable host using suitable vectors. When bacteria is used as vector, they are called bacterial artificial chromosome (BAC) and when yeast is used as vectors they are called yeast artificial chromosome (YAC).

Frederick Sanger developed a principle according to which the fragments of DNA were sequenced by DNA automated sequencers. On the basis of overlapping regions in DNA fragments, these sequences are arranged accordingly.
For alignment of these sequences, specialised computer-based programmes were developed. Finally, the genetic and physical maps of the genome were constructed by collecting information about certain repetitive DNA sequences and DNA polymorphism, based on endonuclease restriction sites.

Salient Features Of HGP

I) The human genome contains 3164.7 million base pairs.
II) The average consists 3000 bases, the largest known human gene is Dystrophin gene at 2.4 million bases.
III) The total number of gene is estimated to be about 30,000 and 99.9 percent nucleotide bases are exactly the same exactly the same in all the people.
IV) Of all the discovered genes less than 50% genes' function is known.
V) Less than 2% of the genome codes for proteins.
VI) The human genome contains large repeated sequences.
VII) The repeated sequences is thought to have no direct coding functions but they throw light on chromosome structures, dynamics and evolution.
VIII) Chromosome I has most genes (2968) and Y has the fewest (231).
IX) Scientists have identified about 1.4 million locations where single base DNA sequence differences occur and are called SNPs or single nucleotide polymorphism, occur in humans.

Applications and future challenges

When complete sequence of human genome is properly studied, new approach to biological research will develop.
All genes in a genome and transcripts in a particular tissue, organ or tumour can be studied.
It enables us to understand how enormous number of genes and proteins work together.

DNA Fingerprinting

Dr. Alec Jeffreys developed the technique of DNA fingerprinting in an attempt to identify DNA marker for inherited diseases.

DNA fingerprinting is a technique to identify an individual using unique sequences called variable number tandem repeats (VNTRs) in DNA which is unique to each individual except monozygotic twins. VNTRs are inheritable from one generation to the next.

Methodology and technique

1. DNA is extracted and isolated from the sample cell or tissue.
2. If the sample is very little the VNTRs are amplified by polymerase chain reaction (PCR).
3. DNA is cut into small fragments by treating with restriction endonucleases.
4. DNA fragments are separated by agarose gel electrophoresis.
5. The separated DNA fragments are split into single strands by treating with alkaline solution .
6. The isolated DNA on the agarose gel is transferred or copied to a nylon nitrocellulose paper by placing it on the gel. This technique is Southern blotting after its inventor EM Southern.
7. Radioactive probes (DNA fragments with known sequences) are then hybridised
8. The hybridised DNA fragments are detected by autoradiography.

   APPLICATION OF DNA FINGERPRINTING

   1. It is used as a forensic tool to identify criminals in criminal investigations.
   2. It is used to establish biological parent of a child in paternity/maternity disputes.
   3. It is used to determine population and genetic diversities to study evolution.

Paternal chromosome

Maternal chromosome

Chromosome 7

Chromosome 2

Chromosome 16

DNA from individual A

Chromosome 7

Chromosome 2

Chromosome 16

DNA from individual B

Number of short tandem repeats

0 11

Chromosome 7

Chromosome 2

Chromosome 16

DNA from crime scene (C)

C A B

12 11 10 9 8 7 6 5 4 3 2 1

Number of stort tandem repeats

Amplified repeats,separated by size on a gel, give a DNA fingerprint

Figure 6.16 Schematic representation of DNA fingerprinting: Few representative chromosomes have been shown to contain different copy number of VNTR. For the sake of understanding different colour schemes have been used to trace the origin of each band in the gel. The two alleles (paternal and maternal) of a chromosome also contain different copy numbers of VNTR. It is clear that the banding pattern of DNA from crime scene matches with individual B, and not with A.