Post-Graduate Degree Programme (CBCS)

in

ZOOLOGY

SEMESTER-IV

HARD CORE THEORY PAPER

MOLECULAR BIOLOGY & BIOTECHNOLOGY &

TOOLS &TECHNIQUE

ZCORT-412

SELF LEARNING MATERIAL

DIRECTORATE OFOPEN AND DISTANCE LEARNING

UNIVERSITY OF KALYANI

KALYANI, NADIA,

W.B. INDIA

Content Writer:

Dr. Subhabrata Ghosh, Assistant Professor, Postgraduate Department of Zoology, Asutosh College, Kolkata

May 2024

Directorate of Open and Distance Learning, University of Kalyani.

Published by the Directorate of Open and Distance Learning,
University of Kalyani, Kalyani-741235, West Bengal.

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani, a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2020 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor (Dr.) Amalendu Bhunia, Hon'ble Vice- Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks are also due to the Course Writers-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and coordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self-writing and collected from e-book, journals and websites.

Director
Directorate of Open and Distance Learning
University of Kalyani

HARD CORE THEORY PAPER (ZCORT -412)

MOLECULAR BIOLOGY AND BIOTECHNOLOGY AND TOOLS AND TECHNIQUE

Part A: MOLECULAR BIOLOGY AND BIOTECHNOLOGY

Module	Unit	Content	Credit	Page No.
ZHT - 412 (MOLECULAR BIOLOGY AND	I	Transcriptional control of gene expression- positive and negative regulations, RNA polymerases, promoters and regulatory sequences, activators and repressors of transcription, transcription initiation by RNA polymerases, regulation of transcription factor activity, elongation and termination of transcription.		
	II	Post-transcriptional gene control -types of introns and their splicing, evolution of introns, catalytic RNA, alternative splicing and proteome diversity, regulation of Pre-mRNA Processing, micro RNA and other non-coding RNAs, degradation of RNA.		
	III	Transport across the nuclear envelope and stability of RNA- structure of nuclear membrane and nuclear pore complexes,		

		processes of nuclear import and export and their regulation, degradation of RNA. Translational machinery and translational control - energetics of amino acid polymerization, tRNAs and their modifications, aminoacyl tRNA synthetases, accuracy during aminoacylation of tRNA, regulation of initiation of translation in eukaryotes, elongation and its control, inhibitors of translations.		
	IV	**Basic recombinant DNA techniques, cutting and joining DNA molecules, restriction modification systems, various enzymes used in recombinant DNA technology, restriction maps and mapping techniques; footprinting, methyl interference assay.**		
	V	**Basic biology of cloning vectors: plasmids, phages, single stranded DNA vectors, high capacity vectors, retroviral vectors, expression vectors and other advanced vectors in use. Gene cloning strategies: methods of transforming E. coli with rDNA; methods of**		

		selection and screening of transformed cells; construction of genomic and cDNA libraries; phage display.	
	VI	**Manipulating genes in animals: gene transfer to animal cells, genetic manipulation of animals, transgenic technology, application of recombinant DNA technology; genetically modified organisms: gene knockouts, mouse disease models, gene silencing, gene therapy, somatic and germ- line therapy**	
	VII	**Genome manipulation-CRISPR-Cas9 System.**	
	VIII	**Polymerase chain reaction– methods and applications.**	
	IX	**Basic Concepts of Microscopy: Magnification, Resolution, Limit of Resolution, Chromatic Aberrations. Types of microscopies: Bright Field Microscopy, Dark Field Microscopy, Phase Contrast Microscopy and Differential Interference Contrast Microscopy: Fluorescent Microscopes, Confocal microscopy.**	

	X	**Electronic Imaging Systems- Electron Microscopy, TEM Vs. SEM. Different fixation and staining techniques for EM, freeze-etch and freeze-fracture methods for EM, image processing methods in microscopy.**	
	XI	**Basic concept of flow cytometry**	
	XII	**Cell Fractionation Methods: i) Ultracentrifugation ii) Gradient Centrifugation.**	
	XIII	**Separation of Cell Constituents: Chromatography: ion exchange; gel filtration and HPLC.**	
	XIV	**Separation of Cell Constituents: Electrophoresis-PAGE, SDS-PAGE (One and Two dimentional).**	
	XV	**Spectroscopy: UV-spectroscopy, Circular Dichroism, surface plasma resonance methods.**	
	XVI	**Blotting Methods: Southern, Northern & Western blotting.**	

	XVII	**RFLP, RAPD and AFLP techniques.**	
	XVIII	**Pesticide formulation**	
	XIX	**Database search tool; Sequence alignment and database searching; Computational tools and biological databases, NCBI, EBL, Sequence similarity tools; Blast and FASTA**	
	XX	**Methods for analysis of gene expression at RNA and protein level, large scale expression, such as micro array based techniques.**	

Unit-I

Transcriptional control of gene expression- positive and negative regulations, RNA polymerases, promoters and regulatory sequences, activators and repressors of transcription, transcription initiation by RNA polymerases, regulation of transcription factor activity, elongation and termination of transcription

Objective: In this unit you will know about various aspects of transcription i.e. initiation, elongation and termination. You will also know about Gene regulation mechanisms in prokaryotes and eukaryotes. There will also be discussion about transcription regulatory events.

Introduction:

Transcription is a process in which ribonucleic acid (RNA) is synthesized from DNA. The word gene refers to the functional unit of the DNA that can be transcribed. Thus, the genetic information stored in DNA is expressed through RNA. For this purpose, one of the two strands of DNA serves as a template (non-coding strand or sense strand) and produces working copies of RNA molecules.

The other DNA strand which does not participate in transcription is referred to as coding strand or antisense strand (frequently referred to as coding strand since with the exception of T for U, primary mRNA contains codons with the same base sequence).

Transcription is Selective:

The entire molecule of DNA is not expressed in transcription. RNAs are synthesized only for some selected regions of DNA. For certain other regions of DNA, there may not be any transcription at all. The exact reason for the selective transcription is not known. This may be due to some inbuilt signals in the DNA molecule.

The product formed in transcription is referred to as primary transcript. Most often, the primary RNA transcripts are inactive. They undergo certain alterations (splicing, terminal additions, base modifications etc.) commonly known as post- transcriptional modifications, to produce functionally active RNA molecules. There exist certain differences in the transcription between prokaryotes and eukaryotes. The RNA synthesis in prokaryotes is given in some detail. This is followed by a brief discussion on eukaryotic transcription.

Transcription in Prokaryotes:

A single enzyme—DNA dependent RNA polymerase or simply RNA polymerase— synthesizes all the RNAs in prokaryotes. RNA polymerase of E. coli is a complex holoenzyme (mol wt. 465

kDa) with five polypeptide subunits— 2α, 1β and 1β' and one sigma (σ) factor (Fig. 4.2). The enzyme without sigma factor is referred to as core enzyme ($\alpha_2\beta\beta'$).

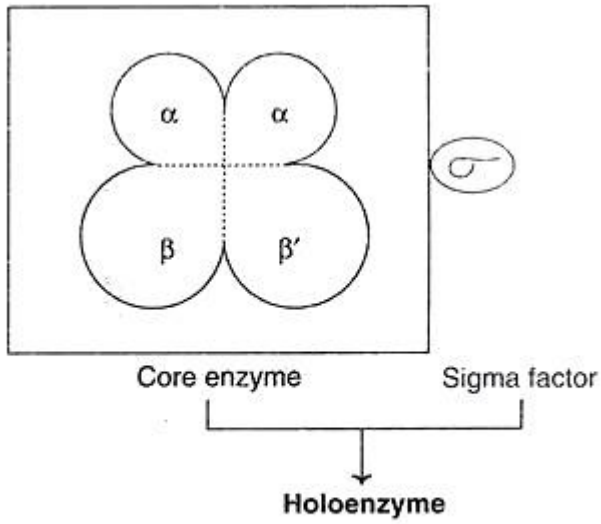


Fig. 4.2 : RNA polymerase of E. coli.

Initiation:

The binding of the enzyme RNA polymerase to DNA is the prerequisite for the transcription to start. The specific region on the DNA where the enzyme binds is known as promoter region. There are two base sequences on the coding DMA strand which the sigma factor of RNA polymerase can recognize for initiation of transcription (Fig. 4.5).

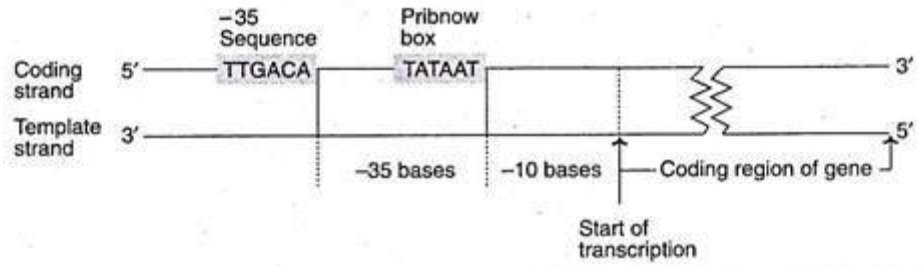


Fig. 4.5 : Promoter regions of DNA in prokaryotes.

1. Pribnow box (TATA box):

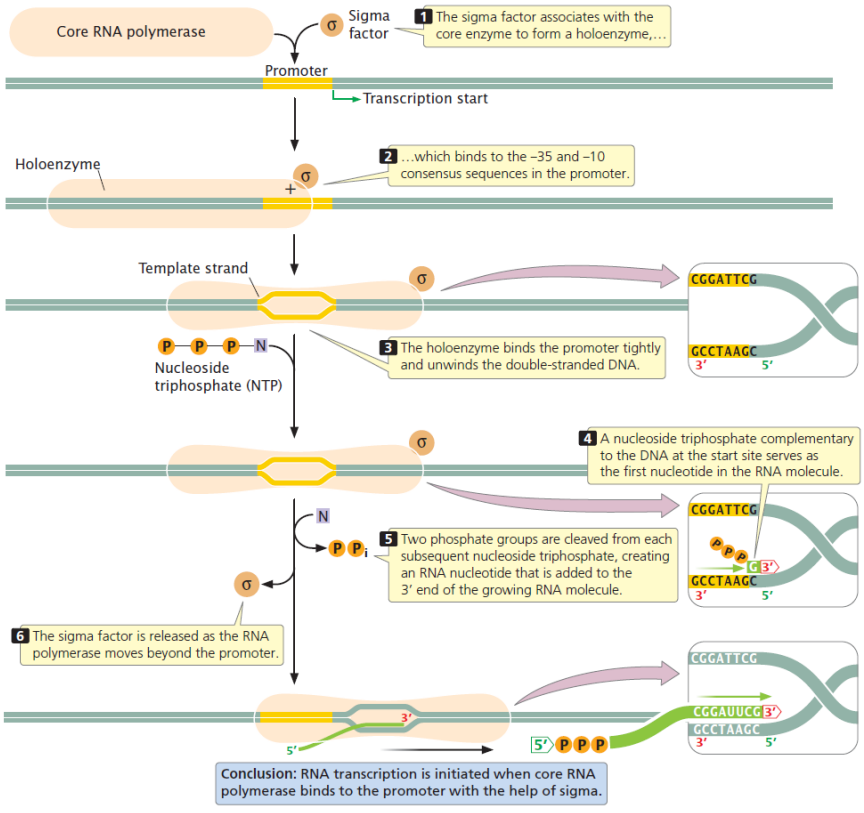
This consists of 6 nucleotide bases (TATAAT), located on the left side about 10 bases away (upstream) from the starting point of transcription.

2. The '-35' sequence:

This is the second recognition site in the promoter region of DNA. It contains a base sequence TTGACA, which is located about 35 bases (upstream, hence -35) away on the left side from the site of transcription start.

The initiation consists of the following steps:

(i) RNA polymerase (RNAP) binds to one of several specificity factors, to form a holoenzyme. In this form, it can recognize and bind to' specific promoter regions in the DNA. At this stage, the DNA is double-stranded ("closed"). This holoenzyme/wound-DNA structure is referred to as the closed complex.
(ii) The DNA is unwound and becomes single-stranded ("open") in the vicinity of the initiation site (defined as + 1). This holoenzyme/unwound-DNA structure is called the open complex.
(iii) The RNA polymerase transcribes the DNA, but produces about 10 abortive (short, non-productive) transcripts which are unable to leave the RNA polymerase because the exit channel is blocked by the cr-factor.
(iv) The a-factor eventually dissociates from the holoenzyme, and elongation proceeds. Most transcripts originate using adenosine-5′-triphosphate (ATP) and, to a lesser extent, guanosine-5′-triphosphate (GTP) (purine nucleoside triphosphates) at the +1 site. Uridine-5′-triphosphate (UTP) and cytidine-5′-triphosphate (CTP) (pyrimidine nucleoside triphosphates) are dis-favoured at the initiation site.



2. Elongation:

In transcription only one strand of DNA [called template strand or non-coding strand] takes part as a template. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy.

Although RNA polymerase traverses the template strand from 3′ —> 5′, the coding (non-template) strand is usually used as the reference point, so transcription is said to go from 5′ -> 3′.This produces an RNA molecule from 5′ 3′, an exact copy of the coding strand (except that thymine are replaced with uracil, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).

In the prokaryotes, the elongation starts with the **"abortive initiation cycle".** During this cycle RNA Polymerase will synthesize mRNA fragments 2-12 nucleotides long. This continues to occur until the σ factor rearranges, which results in the transcription elongation complex (which gives a 35 bp moving footprint). The a factor is released before 80 nucleotides of mRNA are synthesized.

3. Termination:

In prokaryotes, two different modes of transcription termination, viz:

(i) Rho-independent and
(ii) Rho-dependent are well known.
These are briefly discussed as follows:

(i) Rho-independent termination:

It is also known as intrinsic transcription termination. It involves terminator sequences within the RNA that signal the RNA polymerase to stop. The terminator sequence is usually A palindromic sequence that forms a stem-loop hairpin structure that leads to the dissociation of the RNAP from the DNA template.

In the Rho-independent transcription termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C rich hairpin loop, followed by a run of U's, which makes it detached the DNA template.

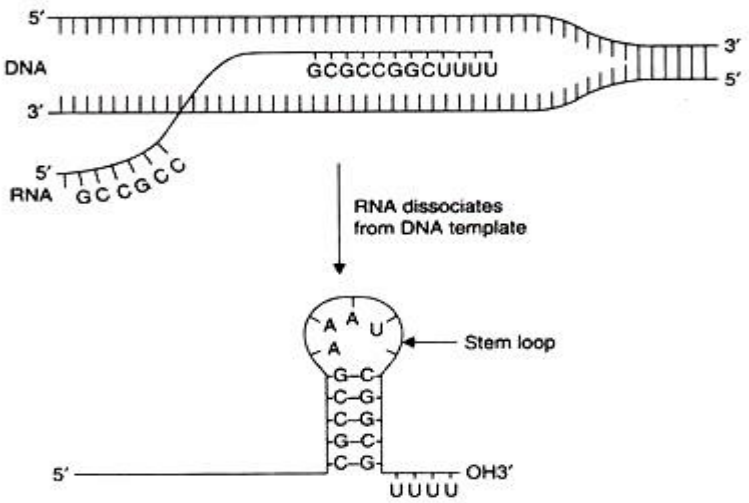


Fig. 15.8 Termination of transcription is signaled by a GC-rich inverted repeat followed by four A residues. The formation of a stem-loop structure by the inverted repeat in the RNA, causes RNA to dissociate from the DNA template.

(ii) Rho-dependent termination:

In the **"Rho-dependent"** type of termination, a protein factor called **"Rho"** is used to stop RNA synthesis at specific sites. This protein binds at a Rho utilisation site on the nascent RNA strand and runs along the mRNA towards the RNA polymerase.

When p-factor reaches the RNAP, it causes RNAP to dissociate from the DNA, terminating transcription. In other words, it destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

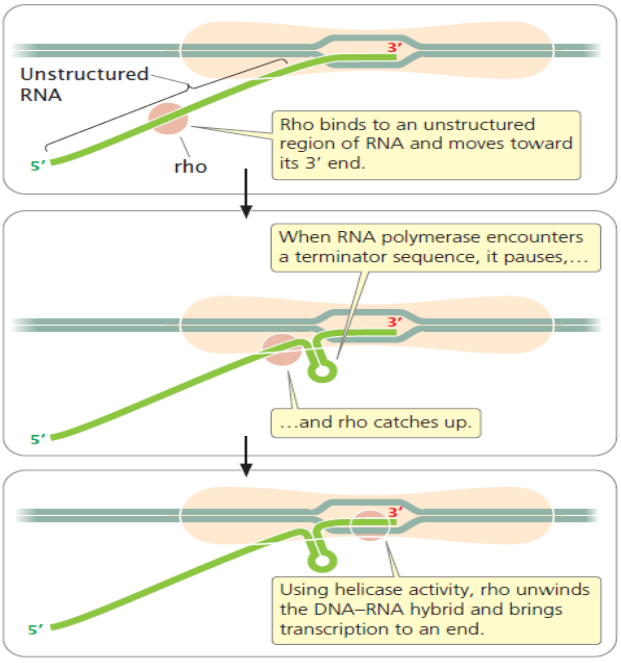


Figure: Rho-dependent terminator

Reverse Transcription in Prokaryotes:

Synthesis of DNA from RNA molecule in the presence of enzyme reverse transcriptase is referred to as reverse transcription. Reverse transcription was first reported by Temin and Baltimore in 1970 for which they were awarded Nobel Prize in 1975.

Reverse transcription is also known as Teminism. Some viruses (such as HIV, the cause of AIDS), have the ability to transcribe RNA into DNA. HIV has an RNA genome that is duplicated into DNA. The resulting DNA can be merged with the DNA genome of the host cell. The main enzyme responsible for synthesis of DNA from an RNA template is called reverse transcriptase. In the case of HIV, reverse transcriptase is responsible for synthesizing a complementary DNA strand (cDNA) to the viral RNA genome.

An associated enzyme, ribonuclease H, digests the RNA strand, and reverse transcriptase synthesizes a complementary strand of DNA to form a double helix DNA structure. This cDNA is integrated into the host cell's genome via another enzyme (integrase), causing the host cell to generate viral proteins which reassemble into new viral particles. Subsequently, the host cell undergoes programmed cell death (apoptosis).

TABLE 21.1. Differences between transcription and reverse transcription

S.No.	*Particulars*	*Transcription*	*Reverse Transcription*
1.	Occurs in	Both prokaryotes and eukaryotes	Reported in some viruses
2.	Molecule synthesized	RNA from DNA	DNA from RNA
3.	Enzyme involved	Transcriptase	Reverse transcriptase
4.	Template used	DNA	RNA

Detection of Transcription in Prokaryotes:

Transcription can be measured and detected in a variety of ways.

The commonly used methods of detecting transcription are given below:

1. Nuclear Run-on assay, measures the relative abundance of newly formed transcripts.
2. RNAse protection assay and ChlP-Chip of RNAP, detect active transcription sites.
3. RT-PCR, measures the absolute abundance of total or nuclear RNA levels, which may however-differ from transcription rates.
4. DNA microarrays measures the relative abundance of the global total or nuclear RNA levels, which may however differ from transcription rates.
5. In situ hybridization, detects the presence of a transcript.

Transcription in Eukaryotes:

There are two major differences between prokaryotic and eukaryotic transcription systems.

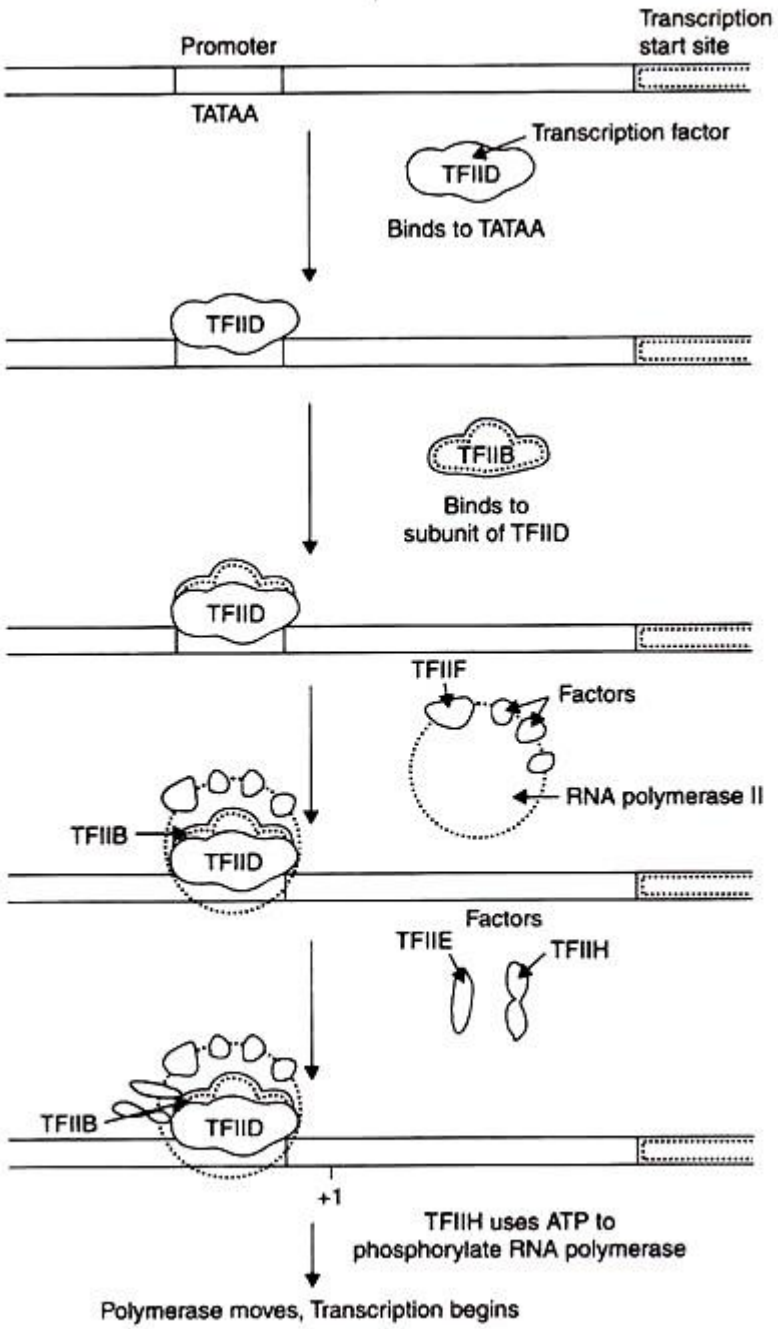
First, a single RNA polymerase is able to transcribe all genes in bacteria, whereas eukaryotic cells have multiple different RNA polymerases that transcribe distinct classes of genes. Second, eukaryotic RNA polymerases do not bind directly to promoter sequences, but interact with a number of proteins to specifically initiate transcription. The complexity of the transcription process in eukaryotes is presumed to be related with the regulation of gene expression

required to control activities of many different cell types in multicellular forms. Three distinct nuclear RNA polymerases transcribe different classes of genes in eukaryotic cells (Table).

Table. Classes of genes transcribed by different eukaryotic RNA polymerases

RNA Polymerase	*Type of RNA Synthesised*
II	*m*RNA, small nuclear snRNA
III	*t*RNA, 5S *r*RNA
I	*r*RNA 5.8*S*, 18*S*, 28*S*, small cytoplasmic scRNA

RNA polymerase II transcribes protein coding genes in nucleus to yield mRNAs; RNA polymerases I and III transcribe ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). The three largest species of tRNAs are transcribed by RNA polymerase I. The genes for transfer RNA and the smallest species of ribosomal RNA (5S rRNA), as well as some small nuclear (snRNAs) and cytoplasmic RNAs (scRNAs) involved in splicing and protein transport are transcribed by RNA polymerase III. In addition, mitochondria and chloroplasts contain separate RNA polymerases similar to bacterial RNA polymerases that specifically transcribe DNA in these organelles.



Stages in initiation of transcription

Transcription by RNA Polymerase II:

The different mode of action of transcription in eukaryotic cells was noted in 1979 when it was found that RNA polymerase II is able to initiate transcription only if additional proteins are added to the reaction. In contrast with the bacterial sigma factors, transcription in eukaryotic cells requires distinct initiation factors that were not associated with the polymerase.

Specific proteins acting as transcription factors have now been identified that are required by RNA polymerase II to initiate transcription. Two types of transcription factors have been defined: general transcription factors involved in transcription from all polymerase II promoters; additional transcription factors involved in control of expression of individual

genes. Experiments using in vitro systems have indicated that five general transcription factors are required for initiation of transcription by RNA polymerase II. The promoters of many genes transcribed by polymerase II contain a sequence similar to TATAA 25 to 30 nucleotides upstream of the transcription start site.

This sequence referred to as the TATA box is similar to the -10 sequence of bacterial promoters and is involved in initiation of transcription as follows: first, a general transcription factor called TFIID (TF indicates transcription factor, II denotes polymerase II) binds to the TATA box. TFIID has multiple subunits including the TATA-binding protein (TBP). The TBP binds specifically to the TATAA consensus sequence and 10-12 other polypeptides called TBP-associated factors (TAFs). Second, TBP binds to a second general transcription factor (TFIIB) forming a TBP-TFIIB complex at the promoter. Following recruitment of RNA polymerase II to the promoter, two additional factors (TFIIE and TFIIH) are required for initiation of transcription.

Two subunits of TFIIH are helicase that unwind DNA around initiation site, while another subunit is a protein kinase that phosphorylates repeated sequences in the largest subunit of RNA polymerase II. In spite of the development of in vitro systems, much remains to be elucidated about polymerase II transcription in eukaryotic cells.

Transcription by RNA Polymerases I and III:

Like RNA polymerase II, the other two polymerases I and III also require additional transcription factors to associate with appropriate promoter sequences. Although the three eukaryotic polymerases recognise distinct types of promoters, a common transcription factor, the TATA- binding protein (TBP) seems to be required for initiation of transcription by all 3 polymerases.

RNA polymerase I transcribes ribosomal RNA genes which are present in tandem repeats, to yield a large 45S pre-rRNA, which is then processed to derive the 28S, 18S and 5.8S rRNAs (Fig. 15.9). The promoter of rRNA genes consists of 150 base pairs just upstream of the transcription initiation site. These promoter sequences are recognised by two transcription factors, UBF (upstream binding factor) and SL1 (selectivity factor 1) which bind to the promoter and then recruit polymerase I to form an initiation complex.

One of the four protein subunits of the SL1 transcription factor is TBP. Thus, TBP is a common transcription factor required by all 3 types of eukaryotic RNA polymerases. The promoter of ribosomal RNA genes does not contain TATA box, therefore, TBP does not bind to specific promoter sequences. Thus TBP associates with ribosomal RNA genes through the binding of other proteins in the SL1 complex to the promoter.

The genes for tRNAs, 5S rRNA and some of the small RNAs involved in splicing and protein transport are transcribed by Polymerase III. These genes are characterised by promoters that lie within, and not upstream of the transcribed sequence.

The process of termination of transcription was found out from experiments in prokaryotes in which nucleus is not bound by a membrane. Therefore, synthesis of proteins on ribosomes occurs simultaneously with synthesis of mRNA on DNA.

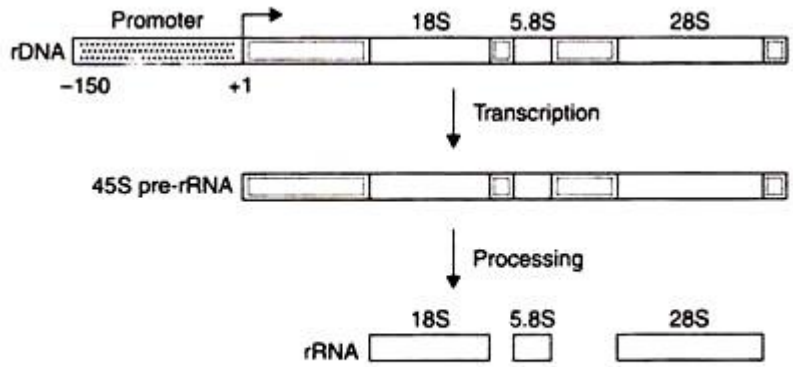


Fig. 15.9 The eukaryotic ribosomal RNA gene (rDNA) is transcribed into a large RNA molecule, the 45S pre-rRNA which is cleaved into the different rRNAs.

The main points related to transcription in eukaryotes are briefly discussed below:

1. Synthesis:

RNA is synthesized from a DNA template. The RNA is processed into messenger RNA [mRNA], which is then used for synthesis of a protein. The RNA thus synthesized is called messenger RNA (mRNA), because it carries a genetic message from the DNA to the protein- synthesizing machinery of the cell.

The main difference between RNA and DNA sequence is the presence of U, or uracil in RNA instead of the T, of thymine of DNA.

2. Template used:

The RNA is synthesized from a single strand or template of a DNA molecule. The stretch of DNA that is transcribed into an RNA molecule is called a transcription unit. A transcription unit codes the sequence that is translated into protein. It also directs and regulates protein synthesis.

The DNA strand which is used in RNA synthesis is called template strand; because it provides the template for ordering the sequence of nucleotides in an RNA transcript. The DNA strand which does not take part in DNA synthesis is called coding strand, because, its nucleotide sequence is the same as that of the newly created RNA transcript.

3. Enzyme Involved:

The process of transcription is catalysed by the specific enzyme called RNA polymerase. DNA sequence is enzymatically copied by RNA polymerase to produce a complementary nucleotide RNA strand. In eukaryotes, there are three classes of RNA polymerases: I, II and III which are involved in the transcription of all protein genes.

4. Genetic Information Copied:

In this process, the genetic information coded in DNA is copied into a molecule of RNA. The genetic information is transcribed or copied, from DNA to RNA. In other words, it results in the transfer of genetic information from DNA into RNA.

5. First Step:

The expression of a gene consists of two major steps, viz., transcription and translation. Thus transcription is the first step in the process of gene regulation or protein synthesis.

6. Direction of Synthesis:

As in DNA replication, RNA is synthesized in the 5′ —> 3′ direction. The DNA template strand is read 3′ –> 5′ by RNA polymerase and the new RNA strand is synthesized in the 5′ -> 3′ direction. RNA polymerase binds to the 3′ end of a gene (promoter) on the DNA template strand and travels toward the 5′ end.

The regulatory sequence that is before, or 5′, of the coding sequence is called 5′ un-translated region (5′ UTR), and sequence found following, or 3′, of the coding sequence is called 3′ un-translated region (3′ UTR). Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA; therefore, transcription has a lower copying fidelity than DNA replication.

Mechanism of Transcription in Eukaryotes:

The mechanism of transcription consists of five major steps, viz:

(1) Pre-initiation,
(2) Initiation,
(3) Promoter clearance,
(4) Elongation and
(5) Termination.

These are briefly discussed as follows:

1. Pre-Initiation:

The initiation of transcription does not require a primer to start. RNA polymerase simply binds to the DNA and, along with other cofactors, unwinds the DNA to create an initiation bubble so that the RNA polymerase has access to the single-stranded DNA template. However, RNA Polymerase does require a promoter like sequence.

Proximal (core) Promoters:

TATA promoters are found around -30 bp to the start site of transcription. Not all genes have TATA box promoters and there exists TATA-less promoters as well. The TATA promoter consensus sequence is TATA(A/T)A(A/T).

2. Initiation:

In eukaryotes and archaea, transcription initiation is far more complex. The main difference is that eukaryotic polymerases do not recognize directly their core promoter sequences. In eukaryotes, a collection of proteins called transcription factors mediate the binding of RNA polymerase and the initiation of transcription.

Only after attachment of certain transcription factors to the promoter, the RNA polymerase binds to it. The complete assembly of transcription factors and RNA polymerase bind-to the

promoter, called transcription initiation complex. Initiation starts as soon as the complex is opened and the first phosphodiester bond is formed. This is the end of Initiation.
RNA Pol II does not contain a subunit similar to the prokaryotic factor, which can recognize the promoter and unwind the DNA double helix. In eukaryotes, these two functions are carried out by a set of proteins called general transcription factors. The RNA Pol II is associated with six general transcription factors, designated as TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, where "TF" stands for "transcription factor" and "II" for the RNA Pol II.
TFIID consists of TBP (TATA-box binding protein) and TAFs (TBP associated factors). The role of TBP is to bind the core promoter. TAFs may assist TBP in this process. In human cells, TAFs are formed by 12 subunits. One of them, TAF250 (with molecular weight 250 kD), has the histone acetyltransferase activity, which can relieve the binding between DNA and histones in the nucleosome. The transcription factor which catalyses DNA melting is TFIIH. However, before TFIIH can unwind DNA, the RNA Pol II and at least five general transcription factors (TFIIA is not absolutely necessary) have to form a pre-initiation complex (PIC).

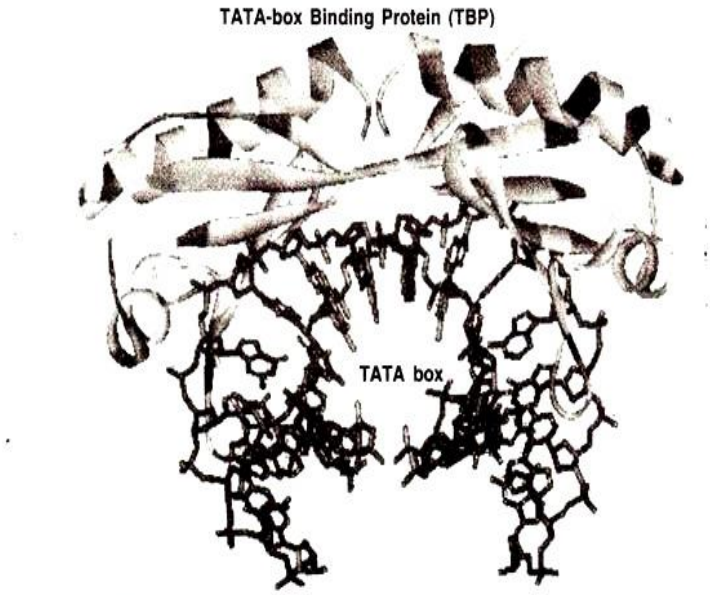


Fig. 22.1. Structure of the human TBP core domain complexed with DNA as determined by X-ray crystallography. The DNA includes the TATA element. PDB ID = 1CDW.

3. Promoter Clearance:

After the first bond is synthesized the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called abortive initiation and is common for both Eukaryotes and Prokaryotes.
Once the transcript reaches approximately 23 nucleotides it no longer slips and elongation can occur. This is an ATP dependent process. Promoter clearance also coincides with Phosphorylation of serine 5 on the carboxy terminal domain which is phosphorylated by TFIIH

4. Elongation:

For RNA synthesis, one strand of DNA known as the template strand or non-coding strand is used as a template. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy.

Although RNA polymerase traverses the template strand from 3′ —> 5′, the coding (non-template) strand is usually used as the reference point, so transcription is said to go from 5′ —> 3′.

This produces an RNA molecule from 5′ —> 3′, an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).

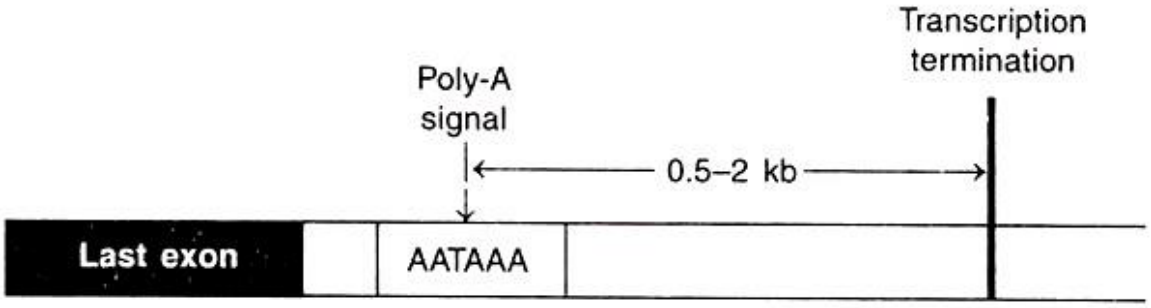
After pre-initiation complex [PIC] is assembled at the promoter, TFIIH can use its helicase activity to unwind DNA. This requires energy released from ATP hydrolysis. The DNA melting starts from about -10 bp. Then, RNA Pol II uses nucleoside triphosphates (NTPs) to synthesize a RNA transcript. During RNA elongation, TFIIF remains attached to the RNA polymerase, but all of the other transcription factors have dissociated from PIC.

The carboxyl-terminal domain (CTD) of the largest subunit of RNA Pol II is critical for elongation. In the initiation phase, CTD is un-phosphorylated, but during elongation it has to be phosphorylated. This domain contains many proline, serine and threonine residues.

5. Termination:

In eukaryotic transcription the mechanism of termination is not very clear. In other words, it is not well understood. It involves cleavage of the new transcript, followed by template-independent addition of As at its new 3′ end, in a process called polyadenylation.

Eukaryotic protein genes contain a poIy-A signal located downstream of the last exon. This signal is used to add a series of adenylate residues during RNA processing. Transcription often terminates at 0.5-2 kb downstream of the poly-A signal.



Transcription Factories in Eukaryotes:

Active transcription units that are clustered in the nucleus, in discrete sites are called ‘transcription factories’. Such sites could be visualized after allowing, engaged polymerases to extend their transcripts in tagged precursors (Br-UTP or Br-U), and immuno-labelling the tagged nascent RNA.

Transcription factories can also be localized using fluorescence in situ hybridization, or marked by antibodies directed against polymerases. There are ~10,000 factories in the nucleoplasm of a HeLa cell, among which are ~8,000 polymerase II factories and ~2,000 polymerase III factories. Each polymerase II factory contains ~8 polymerases.

As most active transcription units are associated with only one polymerase, each factory will be associated with ~8 different transcription units. These units might be associated through promoters and/or enhancers, with loops forming a ‘cloud’ around the factory.

Reverse Transcription in Eukaryotes:

Synthesis of DNA from RNA molecule in the presence of enzyme reverse transcriptase is referred to as reverse transcription. Reverse transcription was first reported by Temin and Baltimore in 1970 for which they were awarded Nobel prize in 1975. Reverse transcription is also known as Teminism. Some viruses (such as HIV, the cause of AIDS), have the ability to transcribe RNA into DNA.

In some eukaryotic cells, an enzyme is found with reverse transcription activity. It is called telomerase. Telomerase is a reverse transcriptase that lengthens the ends of linear chromosomes. Telomerase carries an RNA template from which it synthesizes DNA repeating sequence, or "junk" DNA. This repeated sequence of "junk" DNA is important because every time a linear chromosome is duplicated, it is shortened in length.

With "junk" DNA at the ends of chromosomes, the shortening eliminates some repeated, or junk sequence, rather than the protein-encoding DNA sequence that is further away from the chromosome ends. Telomerase is often activated in cancer cells to enable cancer cells to duplicate their genomes without losing important protein-coding DNA sequence. Activation of telomerase can be part of the process that allows cancer cells to become immortal.

Role of Transcription Factors in Eukaryotes*:*

In eukaryotes, the association between DNA and histones prevents access of the polymerase and general transcription factors to the promoter. Histone acetylation catalys ed by HATs can relieve the binding between DNA and histones. Although a subunit of TFIID (TAF250 in human) has the HAT activity, participation of other HATs can make transcription more efficient. The following rules apply to most (but not all).

(i) Binding of activators to the enhancer element recruits HATs to relieve association between histones and DNA, thereby enhancing transcription.

(ii) Binding of repressors to the silencer element recruits histone deacetylases (denoted by HDs or HDACs) to tighten association between histones and DNA.

Regulation of transcription factor activity:

The term **transcription factor** is broadly used to describe proteins that influence the ability of RNA polymerase to transcribe a given gene. Such transcription factors can regulate the binding of the transcriptional apparatus to the core promoter and/or control the switch from the initiation to the elongation stage of transcription. Two categories of transcription factors play a key role in these processes. General transcription factors (GTF) are required for the binding of RNA polymerase to the core promoter and its progression to the elongation stage.

General transcription factors are necessary for a basal level of transcription. In addition, eukaryotic cells possess a diverse array of regulatory transcription factors that serve to regulate the rate of transcription of target genes. The importance of transcription factors is underscored by the number of genes that encode this category of proteins. Regulatory transcription factors exert their effects by influencing the ability of RNA polymerase to begin transcription of a particular gene. They typically recognize *cis*-acting elements that are located in the vicinity of the core promoter. These DNA sequences are analogous to the operator sites found near bacterial promoters. In eukaryotes, these DNA sequences are generally known as

control elements, or regulatory elements**.** When a regulatory transcription factor binds to a regulatory element, it affects the transcription of an associated gene. For example, the binding of regulatory transcription factors may enhance the rate of transcription. Such a transcription factor is termed an activator**,** and the sequence it binds to is called an enhancer**.** Alternatively, regulatory transcription factors may act as repressors by binding to elements called silencers and preventing transcription from occurring.

By studying transcriptional regulation, researchers have discovered that most eukaryotic genes, particularly those found in multicellular species, are regulated by many factors. This phenomenon is called combinatorial control because the combination of many factors determines the expression of any given gene.

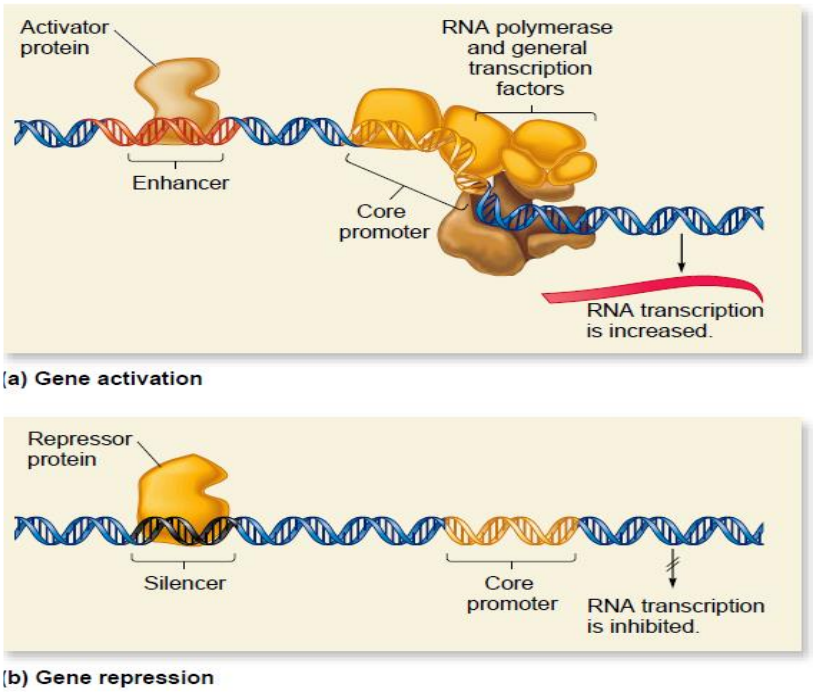


Fig 13: Overview of transcriptional regulation by regulatory transcription factors

At the level of transcription, the following are common factors that contribute to combinatorial control:

1. One or more activator proteins may stimulate the ability of RNA polymerase to initiate transcription.
2. One or more repressor proteins may inhibit the ability of RNA polymerase to initiate transcription.
3. The function of activators and repressors may be modulated in a variety of ways, including the binding of small effector molecules, protein–protein interactions, and covalent modifications.
4. Regulatory proteins may alter the composition or arrangements of nucleosomes in the vicinity of a promoter, thereby affecting transcription
5. DNA methylation may inhibit transcription, either by preventing the binding of an activator protein or by recruiting proteins that cause the chromatin to become more compact.

All five of these factors can contribute to the regulation of a single gene, or possibly only three or four will play a role. In most cases, transcriptional regulation is aimed at controlling the initiation of transcription at the promoter. The functions of the regulatory transcription factors themselves must also be modulated.

The genes they control must be turned on at the proper time, in the correct cell type, and under the appropriate environmental conditions. Therefore, eukaryotes have evolved different ways to modulate the functions of these proteins. The functions of regulatory transcription factor proteins are controlled in three common ways: through (1) the binding of a small effector molecule, (2) protein-protein interactions, and (3) covalent modifications.

Usually, one or more of these modulating effects are important in determining whether a transcription factor can bind to the DNA or influence transcription by RNA polymerase. For example, a small effector molecule may bind to a regulatory transcription factor and promote its binding to DNA.

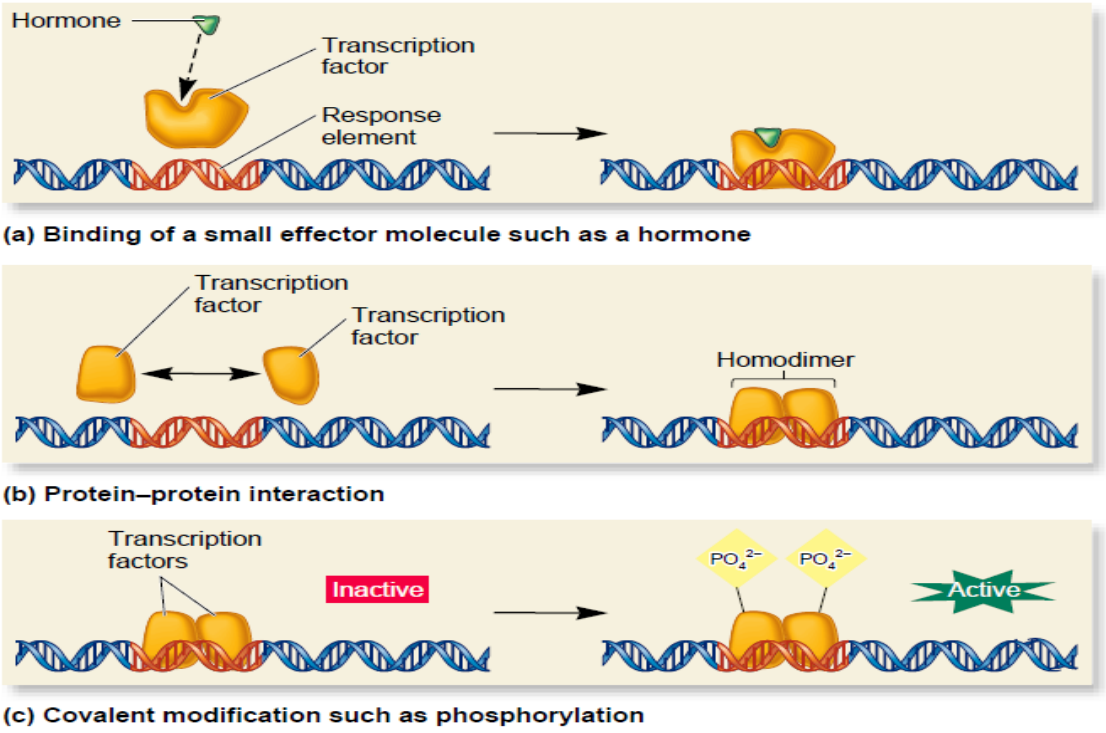


Fig 14: Mechanism of modulating regulatory transcription factor

Genes that encode general and regulatory transcription factor proteins have been identified and sequenced from a wide variety of eukaryotic species, including yeast, plants, and animals. Several different families of evolutionarily related transcription factors have been discovered. In recent years, the molecular structures of transcription factor proteins have become an area of intense research. Transcription factor proteins contain regions, called domains, that have specific functions. For example, one domain of a transcription factor may have a DNA-binding function, and another may provide a binding site for a small effector molecule. When a domain or portion of a domain has a very similar structure in many different proteins, such a structure is called a motif. The protein secondary structure known as an α helix is frequently found in transcription factors. The α helix is the proper width to bind into the major groove of the DNA double helix. In helix-turn-helix and helix-loop-helix motifs, an α helix called the recognition helix makes contact with and recognizes a base sequence along the major groove of the DNA

Major groove is a region of the DNA double helix where the bases contact the surrounding water in the cell. Hydrogen bonding between an α helix and nucleotide bases is one way that a transcription factor can bind to a specific DNA sequence. In addition, the recognition helix often contains many positively charged amino acids (e.g., arginine and lysine) that favourably interact with the DNA backbone, which is negatively charged. Such basic domains are a common feature of many DNA-binding proteins.

Transcriptional gene regulation in eukaryotes:

Different mechanisms have been discovered that explain how a regulatory transcription factor can bind to a regulatory element and thereby affect gene transcription. For most genes, more than one mechanism is involved. The net effect of a regulatory transcription factor is to influence the ability of RNA polymerase to transcribe a given gene. However, most regulatory transcription factors do not bind directly to RNA polymerase.

For structural genes in eukaryotes, regulatory transcription factors commonly influence the function of RNA polymerase II by interacting with other proteins that directly bind to RNA polymerase II. Two protein complexes that communicate the effects of regulatory transcription factors are TFIID and mediator.

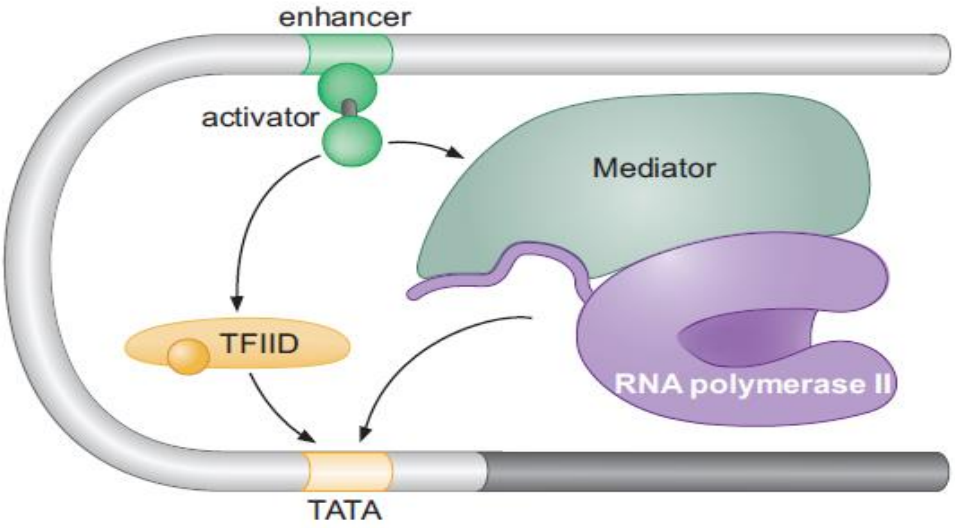
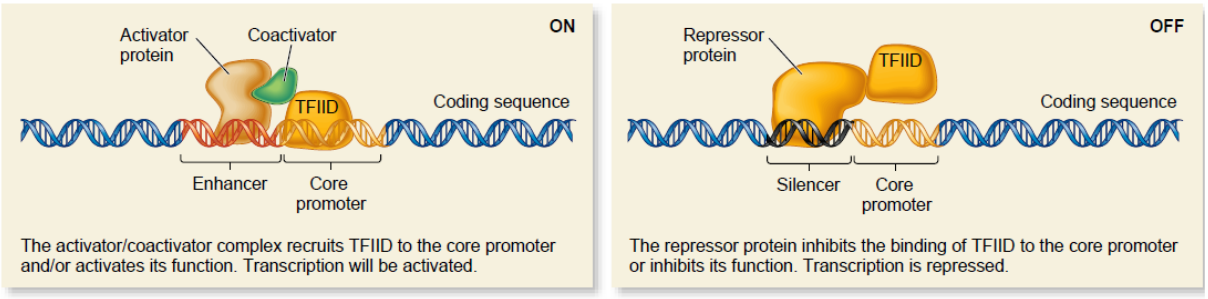


Fig 15. Activation transcription initiation by recruiting transcription machinery

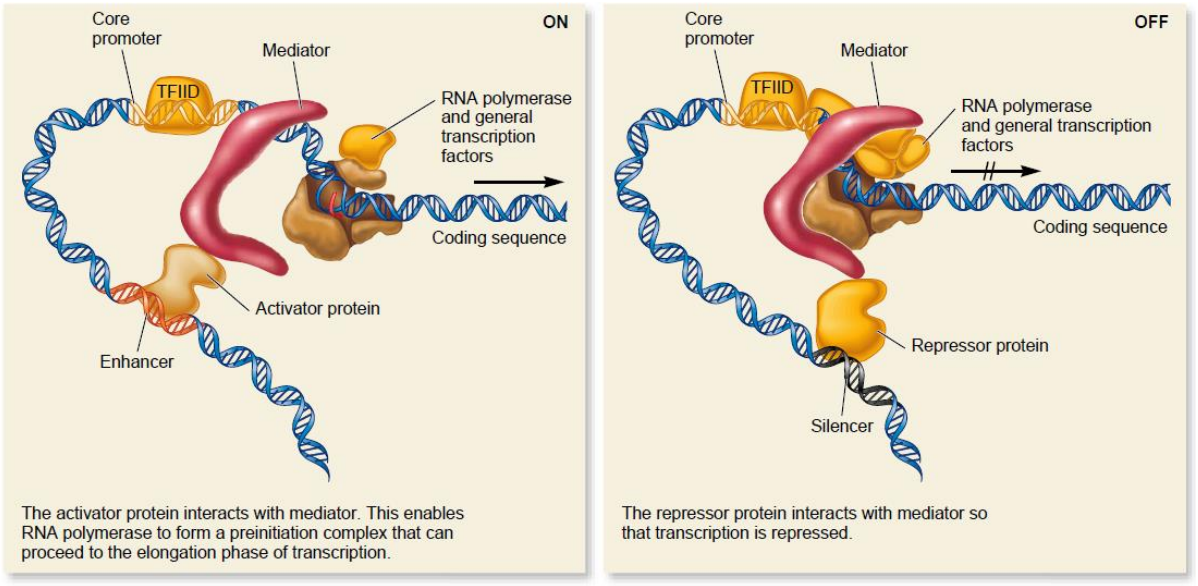
Some regulatory transcription factors bind to a regulatory element and then influence the function of TFIID. TFIID is a general transcription factor that binds to the TATA box and is needed to recruit RNA polymerase II to the core promoter. Activator proteins are expected to enhance the ability of TFIID to initiate transcription. One possibility is that activator proteins could help recruit TFIID to the TATA box. Or they could enhance the function of TFIID in a way that facilitates its ability to bind RNA polymerase II. In some cases, activator proteins exert their effects by interacting with coactivators-proteins that increase the rate of transcription but do not directly bind to the DNA itself. In contrast, repressors inhibit the function of TFIID. They could exert their effects by preventing the binding of TFIID to the TATA box or by inhibiting the ability of TFIID to recruit RNA polymerase II to the core promoter.



(a) Transcriptional activation via TFIID (b) Transcriptional repression via TFIID

Fig 16. Transcription regulation via TFIID

A second way that regulatory transcription factors control RNA polymerase II is via mediator-a protein complex discovered by Roger Kornberg and colleagues in 1990. The term mediator refers to the observation that it mediates the interaction between RNA polymerase II and regulatory transcription factors. Mediator controls the ability of RNA polymerase II to progress to the elongation stage of transcription. Transcriptional activators stimulate the ability of mediator to facilitate the switch between the initiation and elongation stages, whereas repressors have the opposite effect. An activator binds to a distant enhancer element. The activator protein and mediator are brought together by the formation of a loop within the intervening DNA.



(a) Transcriptional activation via mediator (b) Transcriptional repression via mediator

Fig 17: Effect of effector protein on mediator

A third way that regulatory transcription factors can influence transcription is by recruiting proteins to the promoter region that affect nucleosome positions and compositions. For example, certain transcriptional activators can recruit proteins to the promoter region that promote the conversion of chromatin from a closed to an open conformation.

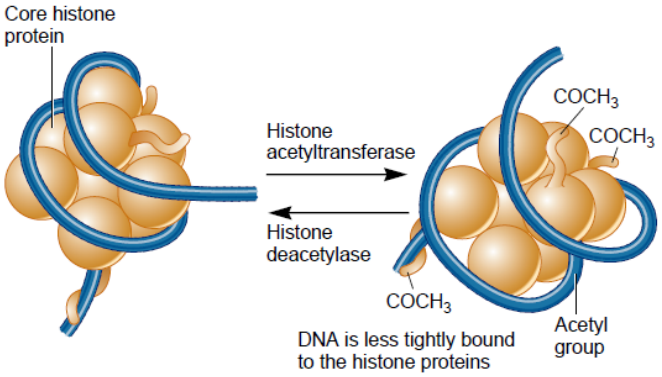


Fig 18: Effect of acetylation on nucleosome structure

Prokaryotic vs. Eukaryotic Transcription:

The process of transcription is the same in both eukaryotes and prokaryotes in several aspects. However, there are some differences in transcription of these two groups as highlighted below.

TABLE 21.2. Comparison of transcription in prokaryotes and eukaryotes

S.No.	*Particulars*	*Prokaryotes*	*Eukaryotes*
1.	Transcription and translation	Occur simultaneously	Occur seperately
2.	Process	Simple	More complicated
3.	mRNA	No need of modification	Needs modification
4.	Enzymes involved	RNA polymerase α, β, β′ and σ	RNA polymerases : I, II, and III
5.	Termination	Well known	Less clear
6.	Site	Cytoplasm	Nucleus
7.	DNA	Circular and free	Linear, packed with histones

1. The process is much more complicated in eukaryotes than prokaryotes.

2. In eukaryotes, transcription and translation take place separately in nucleus and cytoplasm respectively while in prokaryotes both processes take place simultaneously in the cytoplasm.

3. The eukaryotic mRNA contains introns and hence needs modification before taking part in protein synthesis. In prokaryote, the mRNA does not require modification.

4. Eukaryotes have DNA in the nucleus, whereas in prokaryotes DNA is in the cytoplasm.

Visualisation of Transcription:

The idea of visualizing gene transcription originally arose from light microscopic studies of lampbrush chromosomes in amphibian oocytes performed by Callan and Lloyd, and Gall in the 1960's; similar studies were done on the puffed polytene chromosomes of insects by Beermann and his colleagues. Oocyte-chromosomes are highly extended in the lampbrush state and contain thousands of chromosomal loci active in RNA synthesis.

Another favourable attribute of oocytes is that there is amplification (manifold increase) of rRNA genes during early oogenesis giving rise to hundreds of extra nucleoli in a nucleolus. In this system, transcription of ribosomal cistrons has been visualised.

During transcription on oocyte lampbrush chromosomes, the DNA in the condensed, beadlike chromomere unravels and is spun out into a loop and transcribed. The loop axis becomes covered by the transcribed RNA fibrils embedded in a protein matrix (Fig. 15.12). At the base of each RNP (ribonucleoprotein) fibril, an RNA polymerase molecule is attached. In male meiosis and somatic cells transcription produces fine hair-like outgrowths from the chromosomes (Fig. 15.13).

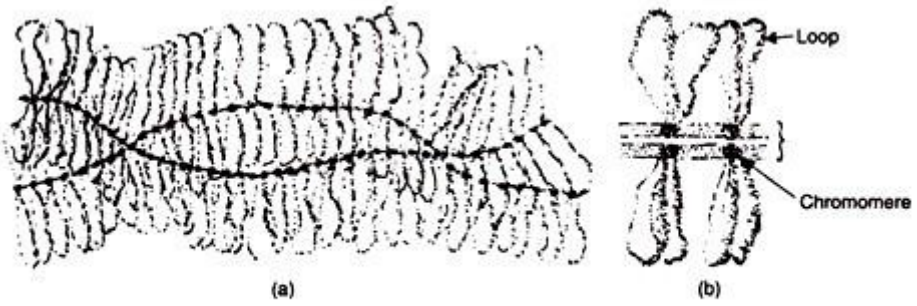


Fig. 15.12 (*a*) a portion from lampbrush chromosomes in amphibian oocytes and (*b*) schematic outline to show details of loop formation.

Fig. 15.13 The lampbrush state at meiotic prophase in the human male.

Puffing in the giant polytene chromosomes in the salivary glands of insects represents a direct way of correlating chromosome structure with gene transcription. Puff formation indicates genes that are actively transcribing RNA which would be translated into salivary proteins. Further work of Grossbach (1969) and others made it possible to relate the synthesis of a specific protein with a specific puff.

In 1969 Miller and Beatty developed a spreading technique for chromatin by which nascent RNP transcripts could be visualised in the electron microscope. The technique has since been applied to various materials. It allows us to visualize the spatial relationships between DNA, RNA polymerase and the RNA transcripts in situ.

EM studies have confirmed that most of the DNP (de-oxy-ribonucleoprotein) fibrils are entangled in the chromomeric mass and only a small proportion is extended into lateral loops. The initiation and termination sites for transcription are located at the two ends of the loop i.e., the thick and thin insertion sites of the loop. The lateral RNP fibrils which contain the nascent RNA transcripts are of increasing length. One loop is said to represent one transcriptional unit. Many times an active loop shows tandemly arranged transcription units separated by spacer

regions (Fig. 15.14). The spacers represent non-transcribed regions. Up to 5 transcriptional units may be present on a loop; the loop is therefore a multi-gene structure.

These authors also found initiation sites of two transcriptional units overlapping each other. The drug actinomycin-D which inhibits transcription removes RNP fibrils from the template and causes loops to collapse. The initial products of transcription observed in EM are longer than the average hnRNA molecules isolated by biochemical techniques. Similar studies on transcription have also been conducted on interphase nuclei of somatic cells in some organisms.

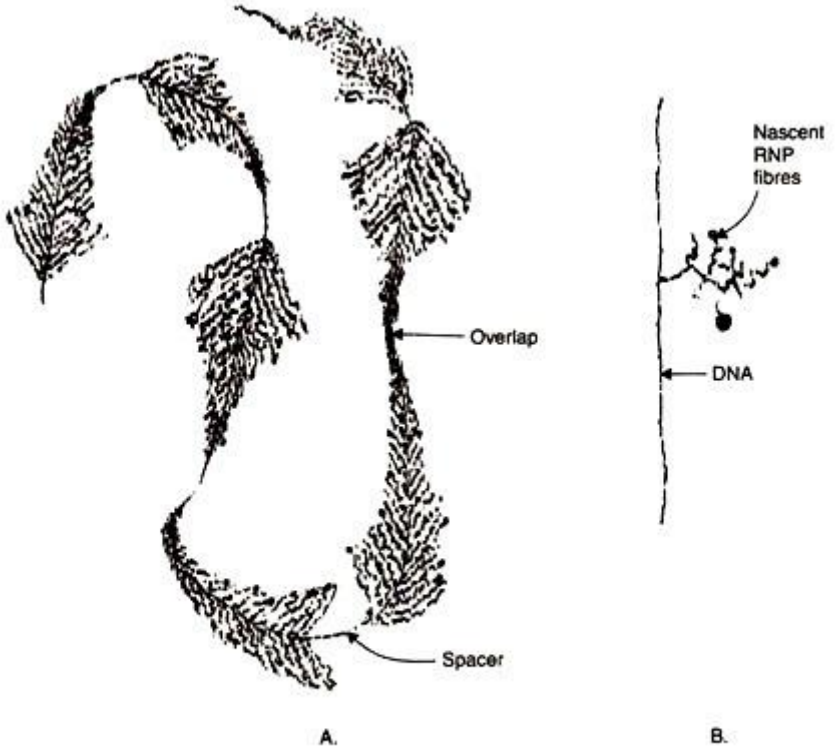


Fig. 15.14 Visualisation of transcribing genes. A—Transcription units are tandemly repeated on a DNA axis. Fig. A shows transcription of ribosomal cistrons and Fig. B exhibits non-ribosomal transcription.

Inhibitors of Transcription:

Several compounds can inhibit transcription of DNA by RNA polymerase. One group of compounds acts by binding non-covalently to the DNA template and modifying its structure; the other group binds to the RNA polymerase and inhibits its catalytic function. The most important inhibitor is actinomycin-D (AMD), an antibiotic produced by streptomyces.

Its phenoxazene ring intercalates between two GC pairs, while its two peptide side chains form H-bonds with guanine bases and project into minor groove of the double helix.

AMD does not interfere with the binding of RNA polymerase to DNA but inhibits chain elongation by preventing movement of core enzyme along the template. AMD does not interfere with replication of DNA. Aflatoxin, ethidium bromide and 2- acetyl-amino-fluorine also inhibit transcription by binding to DNA.

A group of bacterial antibiotics called rifamycins act by inhibiting bacterial RNA polymerases. One such compound known as rifampicin binds non-covalently to the β subunit of RNA polymerase so that chain initiation is inhibited, but does not affect chain elongation, α-

amanitin blocks one of the RNA polymerase enzymes present in eukaryotic cells; but bacterial mitochondrial or chloroplast RNA polymerases are not affected by it.

Gene Regulation in Prokaryotes:

Transcriptional Regulation of Gene Expression in Prokaryotes:

Gene transcription is regulated in bacteria through a complex of genes termed operon. These are transcriptional units in which several genes, with related functions, are regulated together. Other genes also occur in operons which encode regulatory proteins that control gene expression. Operons are classified as inducible or repressible.

Inducible and Repressible System:

The β galactosidase in E. coli is responsible for hydrolysis of lactose into glucose and galactose.

$$\text{Lactose} \xrightarrow{\beta\text{-galactosidase}} \text{glucose} + \text{galactose}$$

If lactose is not supplied to E. coli cells, the presence of β galactosidase is hardly detectable. But as soon as lactose is added, the production of β galactosidase enzyme increases. The enzyme falls as quickly as the substrate (lactose) is removed.

Such enzymes whose synthesis can be induced by adding the substrate are known as inducible enzymes and the genetic system responsible for the synthesis of such an enzyme is called inducible system. The substrate whose addition induces the synthesis of an enzyme is inducer.

In some other cases, the situation is reverse. For instance, when no amino acids are supplied from outside, the E. coli cells can synthesize all the enzymes needed for the synthesis of different amino acids. However, if a particular amino acid, for instance, histidine, is added, the production of histidine synthesizing enzyme falls. In such a system, the addition of the end product of biosynthesis checks the synthesis of the enzymes needed for the biosynthesis. Such enzymes whose synthesis can be checked by the addition of the end product are repressible enzymes and the genetic system is known as repressible system. The end product, the addition of which check the synthesis of the enzyme is co-repressor.

A class of molecules called repressors are found in cells and these repressors check the activity of genes. An active repressor can be made inactive by adding inducer, while an inactive repressor can be made active by adding a co-repressor.

Operon Model:

A hypothesis to explain the induction and repression of enzyme synthesis was first proposed by Jacob and Monod. The scheme proposed by them is called Operon Model.

This consists of the components:

(i) Structural genes
(ii) Promoter genes
(iii) Operator genes
(iv) Regulator genes
(v) Effector or inducer

a. Structural Gene:

These are directly concerned with the synthesis of cellular proteins. They produce the mRNAs through transcription and determine the sequence of amino acids in the synthesized proteins. All the structural genes under an operon may form one long poiycistronic or polygenic mRNA molecule.

b. Operator Gene:

This is located adjacent to the structural gene. It determines whether the structural genes are to be repressed by the repressor protein, a product of regulator gene. The operator gene is the site of binding of the repressor protein, the latter binds to the operator forming an operator-repressor complex. When the repressor binds to the operator, transcription of the structural genes cannot occur.

c. Regulator Gene:

These genes synthesize repressor. Repressor may be either an active repressor or an inactive repressor. Repressor protein has one active site for operator recognition and other active site for inducer. In absence of an inducer protein, the repressor binds to the operator gene and blocks the path of RNA polymerase. Thus the structural genes are unable to transcribe mRNA and consequently protein synthesis does not occur. n presence of an inducer, the repressor protein binds to the inducer to form an inducer-repressor complex. The repressor when binds with inducer undergoes a change and becomes ineffective and as a result it cannot bind to the operator gene and the protein synthesis is possible.

d. Promoter Gene:

The actual site of transcription initiation is known as promoter gene which lies to the left of the operator gene. It is believed that RNA polymerase binds to and moves from the promoter site.

e. Effector or Inducer:

Effector is a small molecule (sugar or amino acid) that can be linked to a regulator protein and will determine whether repressor will bind the operator or not. In the inducible operon, these

effector molecules are called inducer. In repressible operon, these effector molecules are called co-repressor.

I. Inducible Operon:

Lac Operon:

The best known operon is the lac operon. In 1961, F. Jacob and J. Monod proposed the operon model to explain the genetic basis of enzyme induction and repression in prokaryotes.

A few years later (1965), these two investigators were awarded the Nobel Prize for their most incisive work. Although Jacob and Monod's original operon model applied specifically to the regulation of the genes for lactose metabolism in E. coli, additional findings by these two scientists as well as the work of many others have revealed the mechanisms by which other operons function.

The lac operon exercises both positive and negative control. Negative control is in the sense that the operon is normally **"on"** but is kept **"off"** by the regulator gene, i.e., the genes are not allowed to express unless required.

The lac repressor exercises negative control. Positive control is that in which the regulator gene will stimulate the production of the enzyme. Catabolite activator protein (CAP) facilitates transcription, so it exercises positive control. Two unique proteins are thus involved in the regulation of the lac operon which are lac repressor and CAP. Lactose is a disaccharide molecule. In order to utilize lactose as a carbon and energy source, the lactose molecules must be transported from the extracellular environment into the ceil, and then undergo hydrolysis into glucose and galactose. These reactions are catalysed by three enzymes. The lac operon consists of three structural genes (lac Z, Y, A) which code for these three enzymes (Fig. 17.2).

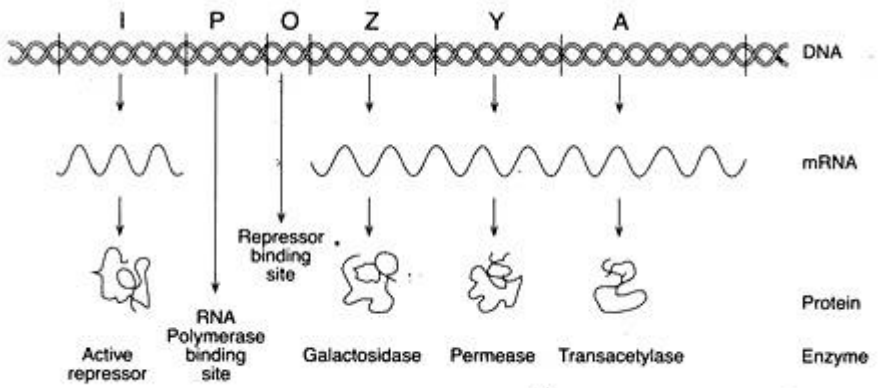


Fig. 17.2: The model of gene regulation of Jacob and Monod as applied to the lac operon in E. coli (from G. S. Stent)

The normal inducer of the lac operon is allolactose, which is produced from lactose by β-galactosidase (Fig. 11-13). A few copies of this enzyme are present in E. coli cells even in the uninduced state. When allolactose combines with the repressor, a steric change occurs that causes the repressor to be released from the operator. As a result, transcription of the structural genes by RNA polymerase begins and the three enzymes quickly appear in the cells.

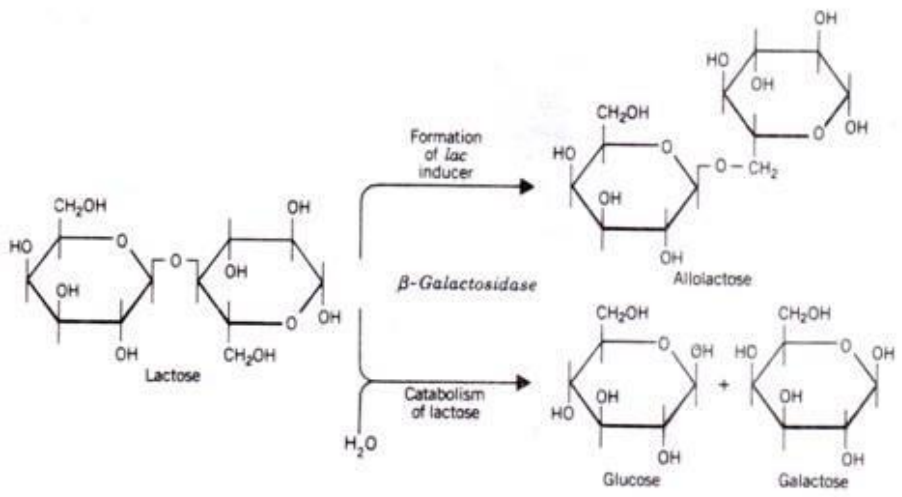


FIGURE 11–13 The two reactions that are catalyzed by the enzyme *β-galactosidase*. Allolactose produced in the upper reaction is an inducer of the *lac* operon.

lac Z gene — codes for enzyme β galactosidase which breaks lactose into galactose and glucose

lac Y gene — codes for permease which transports lactose into the cell

lac A gene — codes for transacetylase which transfer the acetyl group from acetyl CoA to galactose.

Negative Control of lac Operon:

lac repressor is synthesized through the activity of the lac I gene called the regulator gene. This repressor is an allosteric protein

(i) That can bind the lac DNA at the operator site, or

(ii) That can bind to inducer.

In the absence of inducer, DNA binding site of repressor is functional. The repressor protein binds to the DNA at the operator site of the lac locus and blocks the transcription of the lac genes by RNA polymerase. Thus lac enzyme synthesis is inhibited (Fig. 17.3A).

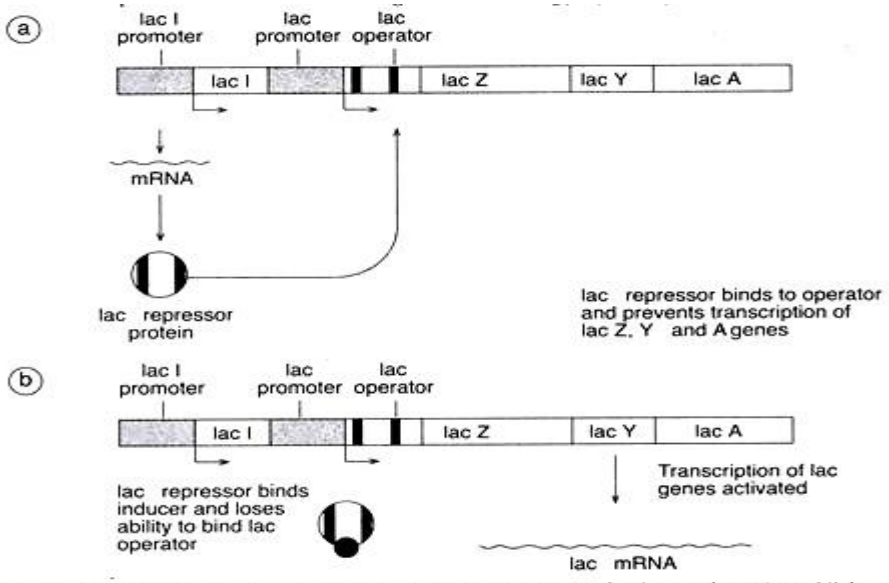


Fig. 17.3A: Regulation of lac operon (a) in the absence and (b) in the presence of inducer (after Winter, Hickey & Fletcher)

Lactose is not the real inducer of the lac operon. It binds to repressor to increase its affinity for operator. On the other hand, the bound protein of the inactive repressor is the allolactose. While β galactosidase breaks lactose into glucose and galactose, a side reaction changes galactose to allolactose and galactobiose.

This allolactose prevents the anti-inducing effect of lactose. When the allolactose (inducer) binds to the repressor, it changes the form of DNA binding site making the repressor inactive and release from- the operator site. Thus transcription of lac genes are possible.

Positive Control of lac Operon:

It is an additional regulatory mechanism which allows the lac operon to sense the presence of glucose, an alternative and preferred energy source to lactose. If glucose and lactose are both present, cells will use up the glucose first and will not utilize energy splitting lactose into its component sugars.

The presence of glucose in the cell switches off the lac operon by a mechanism called catabolite repression which involves a regulatory protein called the catabolite activator protein (CAP). CAP binds to a DNA sequence upstream of the lac promoter and enhances binding of the RNA polymerase and transcription of the operon is enhanced (Fig. 17.3B).

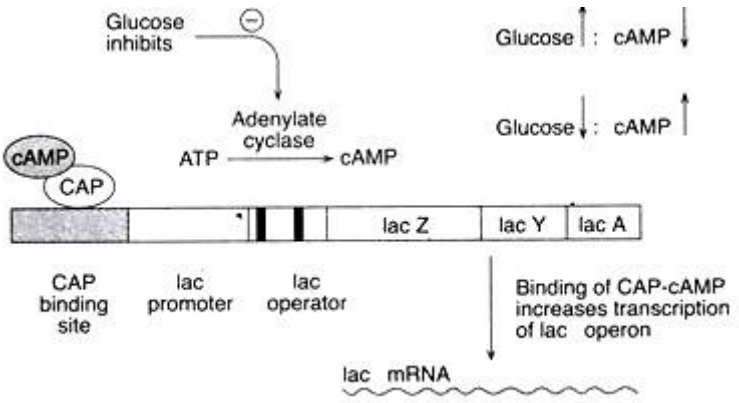


Fig. 17.3B: Catabolite repression of lac operon

CAP only binds in the presence of a derivative of ATP called cyclic adenosine monophosphate (cAMP) whose levels are influenced by glucose. The enzyme adenylate cyclase catalyses the formation of cAMP and is inhibited by glucose. When glucose is available to the cell, adenylate cyclase is inhibited and cAMP levels are low.

Under these conditions CAP does not bind upstream of the promoter and the lac operon is transcribed at a very low level. Conversely, when glucose is low, adenylate cyclase is not inhibited, cAMP is higher and CAP binds increasing the level of transcription from the operon.

If glucose and lactose are present together, the lac operon will only be transcribed at a low level. However when the glucose is used up, catabolite repression will end and transcription from the lac operon increases allowing the available lactose to be used up.

Catabolic Repression:

Catabolic repression is a specific type of repression of enzyme production in which a metabolite such as glucose acts to repress the formation of enzymes that would allow the catabolism of other, related metabolites. For example, when E. coli cells are cultured in a medium that is rich in glucose, the glucose represses the formation of β-galactosidase even if lactose (an inducer of this enzyme) is added to the medium. Glucose is even known to repress the production of enzymes that formerly were thought to be constitutive.

A common phenomenon in bacteria is the suppression of aerobic respiration and electron transport by high glucose concentrations, even in the presence of ample oxygen. Under this condition, the cells utilize the glycolytic and fermentative pathways. The manner in which the catabolite brings about the effect is only partially understood. In the case of catabolite repression by glucose in bacteria, it appears that glucose affects the amount of cyclic AMP (cAMP) present in the cells.

When the concentration of glucose is high, the concentration of cAMP is low; and low levels of glucose are accompanied by high concentrations of cAMP. It is possible that glucose affects the synthesis of cAMP, which is formed from ATP by adenyl cyclase (Fig. 11-14).

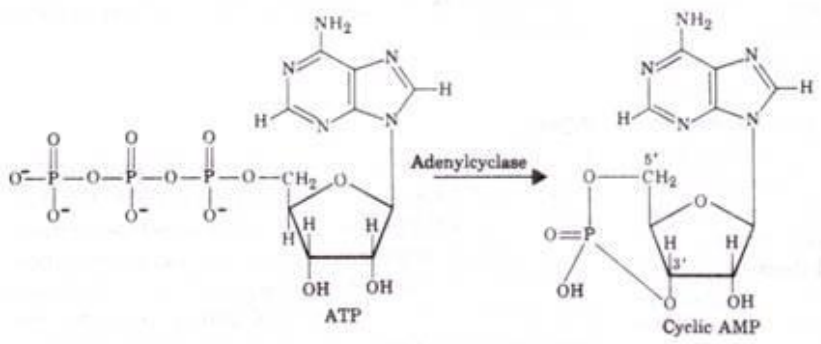


FIGURE 11–14 Formation of cyclic AMP from ATP by the enzyme, *adenylcyclase*.

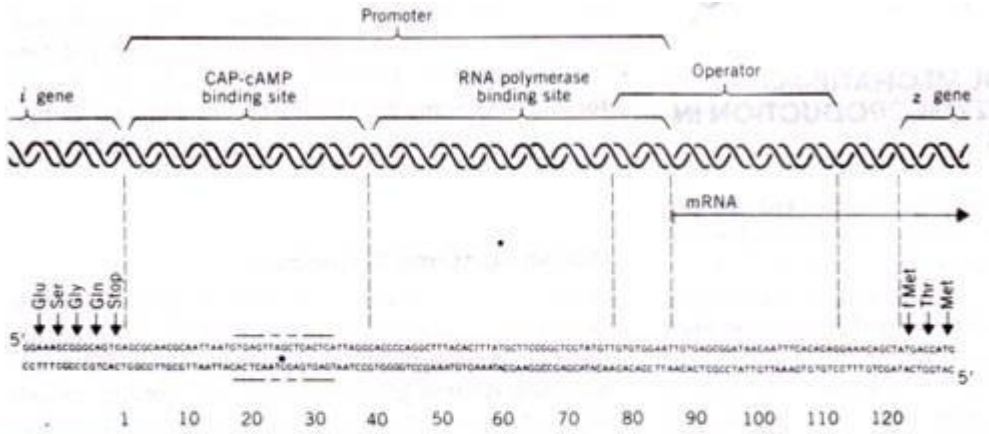


FIGURE 11–15 Organization and nucleotide sequences of the *lac* operon. The promoter consists of two components: the site that binds the cAMP–CAP complex and the RNA polymerase binding site. The adjoining segments of the repressor (*i*) and *β-galactosidase* (*z*) structural genes are also shown. The horizontal arrow (labeled mRNA) shows the position at which the transcription of the operon begins (i.e., what will be the 5′ end of the mRNA). The numbers at the bottom of the figure give the distances (in base pairs) from the end of the *i* gene. (From Gardner and Snustad, *Principles of Genetics*, 7th ed., p. 389 Copyright © 1984 by John Wiley & Sons, Inc.)

cAMP is necessary for the CAP to bind to the promoter site. cAMP binds to the CAP, and once the cAMP-CAP complex binds to the promoter (Fig. 11-15), the RNA polymerase can attach to the promoter and begin transcription. For example, in the case of the lac operon, when the glucose level is high (even if lactose is also present), the cAMP is low, and therefore the cAMP—CAP complex is not available to bind to the promoter and allow transcription to start.
However, in the absence of glucose and in the presence of lactose (which forms a complex with the repressor), cAMP is plentiful and is available to combine with the CAP so that transcription proceeds. The rate of lac operon transcription in the absence of glucose is 50 times as great as in the presence of glucose.

II. Repressible Operon:
Trp Operon:
The trp operon consists of the following components:

(i) Structural genes (trp E, D, C, B and A):
This operon contains five structural genes encoding enzymes involved in biosynthesis of the amino acid tryptophan. The genes are expressed as a single mRNA transcribed from an upstream promoter.

(ii) Promoter gene (trp P):
It is the promoter region which is the binding site for RNA polymerase.

(iii) Operator gene (trp O):
It is the operator region which binds with the repressor.

(iv) Leader gene (trp L):
It is the leader region which is made of 162 nucleotides prior to the first structural gene trp E. It has four regions, region 1 has the codon for tryptophan, region 2, 3 and 4 regulate the mRNA synthesis of the structural genes.

Expression of the operon is regulated by the level of tryptophan in the cell (Fig. 17.4). A regulatory gene upstream of the trp operon encodes a protein called the trp repressor. This protein binds a DNA sequence called the trp operator which lies just downstream of the trp promoter partly overlapping it.

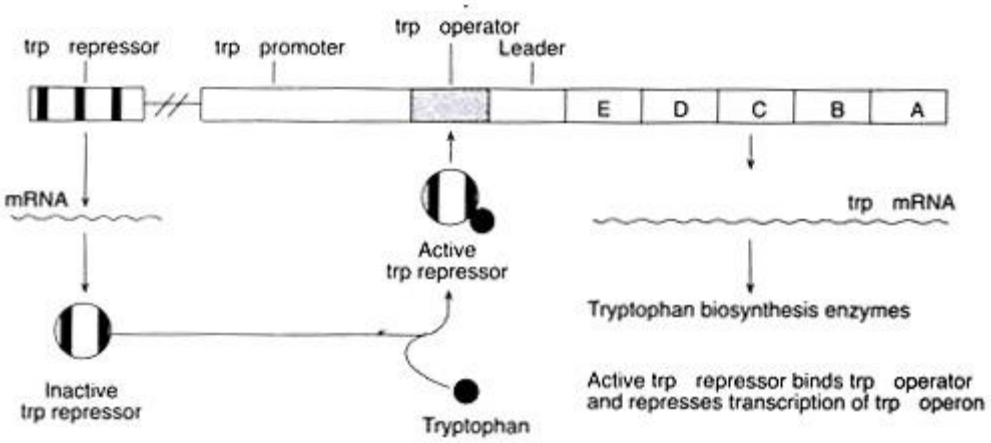


Fig. 17.4: The trp operon

When tryptophan is present in the cell it binds to the trp repressor protein enabling it to bind the trp operator sequence, obstructing binding of the RNA polymerase to the trp promoter and preventing transcription of the operon.

In the absence of tryptophan, the trp repressor is incapable of binding the trp operator and transcription of the operon proceeds. Tryptophan, the end product of the enzymes encoded by the trp operon, thus acts as a co-repressor with the trp repressor protein and inhibits its own synthesis by end product inhibition.

Attenuation:

Attenuation is an alternative regulatory mechanism that allows fine adjustment of expression of the trp operon and other operons (phe, his, leu, thr operon). The transcribed mRNA sequence between the trp promoter and the first trp gene are capable of forming either a large stem-loop structure that does not influence transcription or a smaller stem loop which acts as transcription terminator (Fig. 17.5).

The relative position of the sequences does not allow the formation of both stem-loops at a time. Attenuation depends on the fact that transcription and translation are linked, i.e., ribosomes attach to mRNAs as they are being transcribed and begin translating them into protein. Binding of ribosomes to the trp mRNA influences which of the two stem-loops can form and so determines whether termination occurs or not (Fig. 17.5).

A short coding region upstream of the stem-loop region contains tryptophan codons which is translated before the structural genes. When tryptophan levels are adequate, RNA polymerase transcribes the leader region closely followed by a ribosome which prevents formation of the larger stem-loop, allowing the terminator loop to form ending transcription. If tryptophan is lacking, transcription is initiated, but not subsequently terminated because the ribosome is stalled, the RNA polymerase moves ahead and the large stem-loop forms. Formation of the terminator loop is blocked and transcription of the operon proceeds. When tryptophan present at intermediate levels, some transcripts will terminate and others not.

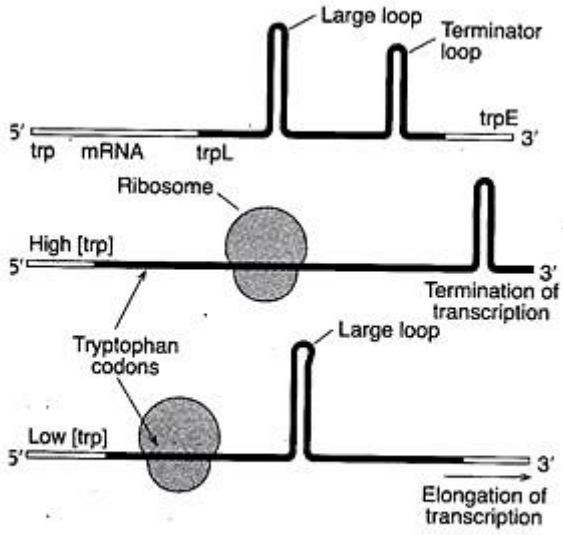


Fig. 17.5A: Transcriptional attenuation in the trp operon (from Winter, Hickey, Fletcher)

Attenuation thus allows the cell to synthesize tryptophan according to its exact requirements. Overall, the trp repressor determines whether the operon is switched on or off and attenuation determines how efficiently it is transcribed. The sequence of the mRNA suggests that ribosome stalling influences termination at the attenuator. The ability of the ribosome to proceed through the leader region may control transition between these structures. The structure determines whether the mRNA can provide the features needed for termination or not.

When tryptophan is present, ribosomes are able to synthesize the leader peptide. They will continue along the leader section of the mRNA to the UGA codon, which lies between regions 1 and 2. By progressing to this point, the ribosomes extend over region 2 and prevent it from base pairing. The result is that region 3 is available to base pair with region 4, generating the terminator hairpin. Under these conditions, therefore, RNA polymerase terminates at the attenuator.

However, when there is no tryptophan, ribosomes initiate translation of the leader peptide but stall at the trp codons which is at the region 1. Thus the region 1 cannot base pair with region 2. If this happens, even while the mRNA itself is being synthesized, region 2 and 3 will be base-paired before region 4 has been transcribed. This compels region 4 to remain in a single stranded form. In the absence of the terminator hairpin, RNA polymerase continues transcription past the attenuator.

The two operon models described above can be summarized as given below:

(i) Inducible System:

Active Repressor + Operator → System OFF

Active Repressor + Inducer = Inactive Repressor → System ON

(ii) Repressible System:

Apo-repressor and co-repressor complex = Active repressor → System OFF

Apo-repressor = Inactive Repressor → System ON

Importance of Gene Regulation:

1. There are two types of gene action – constitutive and regulated.
2. The constitutive gene action occurs in those systems which operate all the time and the cell cannot live without them, e.g., glycolysis. It does not require repression. Therefore, regulator and operator genes are not associated with it.
3. In regulated gene action all the genes required for a multistep reaction can be switched on or off simultaneously.
4. The genes are switched on or off in response to particular chemicals whether required for metabolism or are formed at the end of a metabolic pathway.
5. Gene regulation is required for growth, division and differentiation of cells. It brings about morphogenesis.

Probable Questions:

1. Write down the composition of RNA polymerase of E.coli. What is sigma factor.
2. Write down the role of consensus sequences in initiation of transcription.
3. Discuss rho dependent and rho independent termination of transcription.
4. What is reverse transcription ? describe different step of reverse transcription.
5. Write down the differences between transcription and reverse transcription.
6. How transcription is initiated in Eukaryotes?
7. What is promoter clearance?
8. How transcription factors regulate transcription in eukaryotes?
9. Differentiate prokaryotic and eukaryotic transcription.
10. Write short note on inhibition of transcription.
11. Define inducible and repressible system? Give examples.
12. Describe different components of an operon.
13. Write down the structural components of Lac operon.
14. How allolactose induce lac operon? Explain with suitable diagram.
15. Describe positive role of lac operon.
16. What is catabolic repression? Explain.
17. Describe structural components of trp operon.
18. How attenuation controls trp operon.
19. Describe the significance of gene regulation.

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.

Unit-II

Post-transcriptional gene control: Regulation of Pre-mRNA Processing; Splicing, Types of introns and their splicing, evolution of introns, catalytic RNA, alternative splicing and proteome diversity, micro RNA and other non-coding RNAs.

Objective: In this unit we will discuss about different types of post transcriptional mechanisms and their regulations. We will also discuss about catalytic RNA and different types of non coding RNA including micro RNA. Alternative splicing and their role in protein diversity generation will also be discussed.

Introduction:

Almost all types of RNA molecules undergo post synthesis transformation which is called RNA processing. Prokaryotic mRNA is generally not processed. In prokaryotes 5′-end of prokaryotic mRNA starts translation while the 3′-end is still under synthesis. Eukaryotic mRNA undergoes maximum processing. Both prokaryotic and eukaryotic tRNAs and rRNAs undergo processing. Except prokaryotic mRNA, all other kinds of RNA are processed immediately after synthesis.

Processing of Eukaryotic mRNA:

Newly synthesized mRNA is called primary transcription or precursor mRNA. It is quite different from the mRNA that takes part in protein synthesis. Large-scale changes take place in precursor mRNA. These changes are called processing of mRNA. Both 5′-end 3′-end of mRNA are modified. Non-coding regions are removed by splicing. The changes lead to the formation of mature mRNA which takes part in protein synthesis.

There are 4 main processes by which eukaryotic RNA are processed: These are described below:

A. 5’ Methyl Guanosine Capping:

Eukaryotic pre-mRNAs are altered at their 5′ end by a modification known as capping which involves addition of the modified nucleotide, 7-methylguanosine. The cap is added by the enzyme guanyl transferase which joins GTP by an unusual 5′—- > 5′ triphosphate linkage to the first nucleotide of the mRNA.

Methyl transferase enzymes then add a -CH_3 group to the 7-nitrogen of the guanine ring and, usually, to the 2′ hydroxyl group on the ribose sugar of the next two nucleotides. Capping protects the mRNA from being degraded from the 5′ end by exonucleases in the cytoplasm and is also a signal allowing the ribosome to recognize the start of a mRNA molecule (Fig. 16.9A).

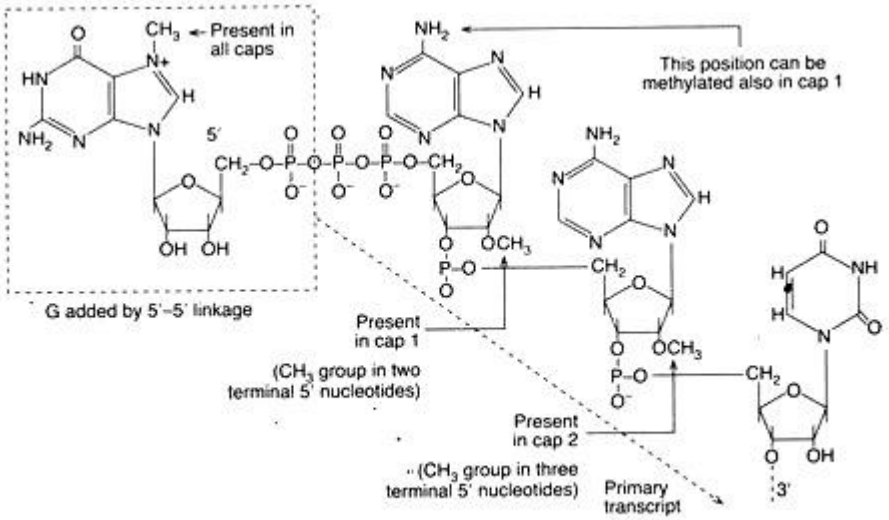


Fig. 16.9A: Capping of 5' end of mRNA by cap0, cap1 and cap2 due to methylation at several positions (from P. K. Gupta)

B. Polyadenylation:

Most eukaryotic pre- mRNA are modified at their 3′ ends by the addition of a sequence of up to 250 adenines, known as a poly A tail. This modification is called polyadenylation and requires the presence of signal sequences in the pre-mRNA.

These consist of the polyadenylation signal sequence, 5′ AAUAAA 3′, which occurs near the 3′ end of the pre-mRNA. The sequence YA (Y = pyrimidine) occurs in the next 11-20 bases and a GU rich sequence is often present further downstream. A number of specific proteins recognize and bind these signal sequences forming a complex which cleaves the mRNA about 20 nucleotides downstream of the 5′ AAUAAA 3′ sequence. The enzyme poly(A) polymerase then adds adenines to the 3′ end of the molecule. The purpose of the poly A tail may be to protect the mRNA from degradation of the coding sequence at the 3′ end by exonucleases (Fig. 16.9B).

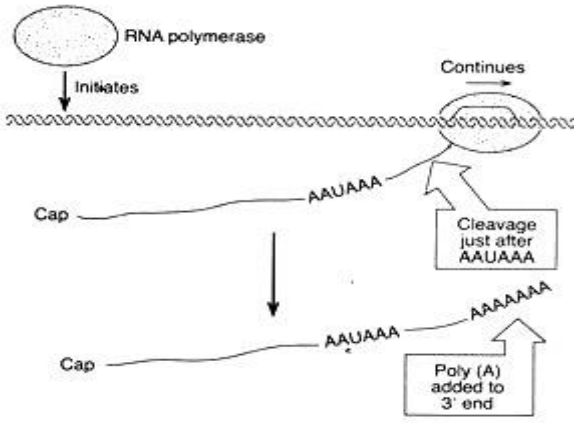


Fig. 16.9B: Cleavage at AAUAAA and polyadenylation at the cleaved 3' end

C. Intron Splicing :

Till recently it was believed that coding sequence of DNA and amino acids of polypeptide is collinear. The coding sequences are continuous and codon for one amino acid is adjacent to the codon for the next amino acid. The open reading frame is a single stretch of codons without any gap.

But now it has been discovered that coding sequence of most of eukaryotic genes is split into stretches of codons interrupted by stretches of non-coding sequences. Most human genes are discontinuous. The coding sequences of DNA of the gene are called exons. In between exons, there are intervening non-coding sequences called introns. This type of genes are called split genes or interrupted genes. They are most common in eukaryotes. They are also found in viruses but rarely in bacteria. The terms exons and introns were given by Gilbert in 1977. It was discovered in Amphibia, mammals and some other animals that genes are not represented by continuous sequence of nucleotides. Introns are removed by excision and discarded.

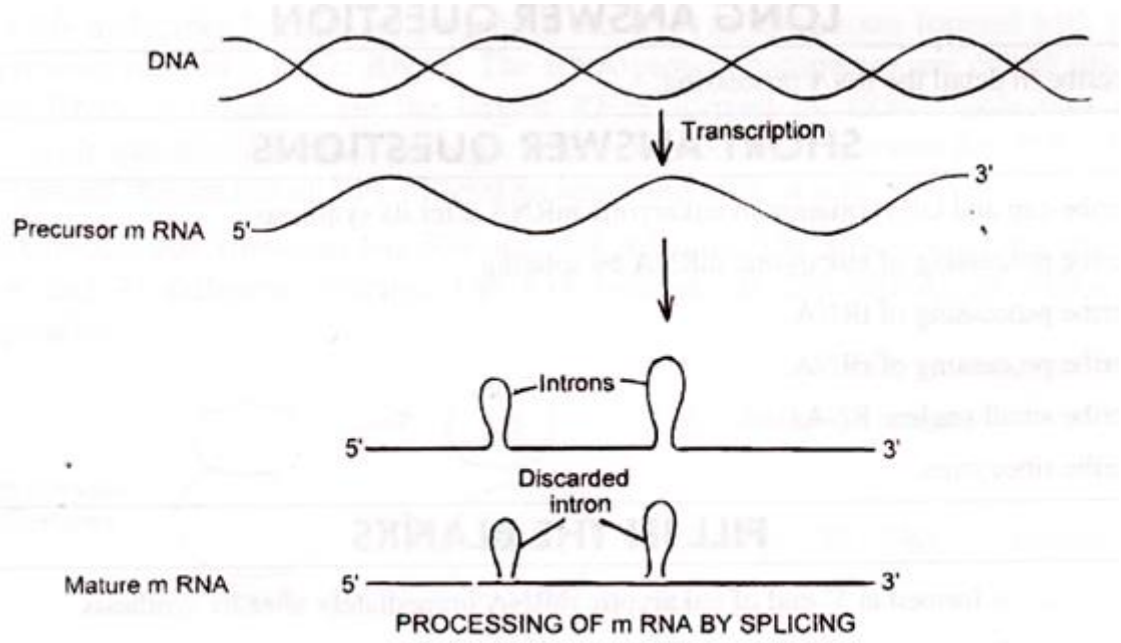


Fig. 9.1. Processing of mRNA by Splicing

The size of introns and exons varies Introns are usually much larger than the exons. A typical exon consists of small number of nucleotides whereas an intron may consist of thousands of nucleotides. Moreover, the introns constitute a large portion of the genome.

During the processing of mRNA in eukaryotes, the amount of discarded RNA ranges from 50-90 percent of the primary transcript. The remaining segments of mRNA or exons are joined together to form the finished mature mRNA. The split genes are transcribed into a single mRNA copy of the entire gene, called primary transcript. It consists of both exons and introns.

The removal of introns by excision is called RNA splicing. The 5′-segment or the cap and the 3′-segment of the trailer are never discarded. Some examples of number of introns are — a-globin has two introns. Yeast genes have one intron, ovalbumin has seven, a-collagen has 52 introns and Titin gene of human beings has 363 introns.

Translation takes place only after the splicing is completed:

Phillip Sharp and Richard Robert won Nobel Prize for medicine in 1993 for their work on split genes. They studied the hybrid of adenovirus mRNA and one (template) strand of DNA under

the electrons microscope. The mRNA was found to hybridize with discontinuous stretches of genomic DNA. The intervening stretches of DNA were in the form of loops. These loops represented introns.

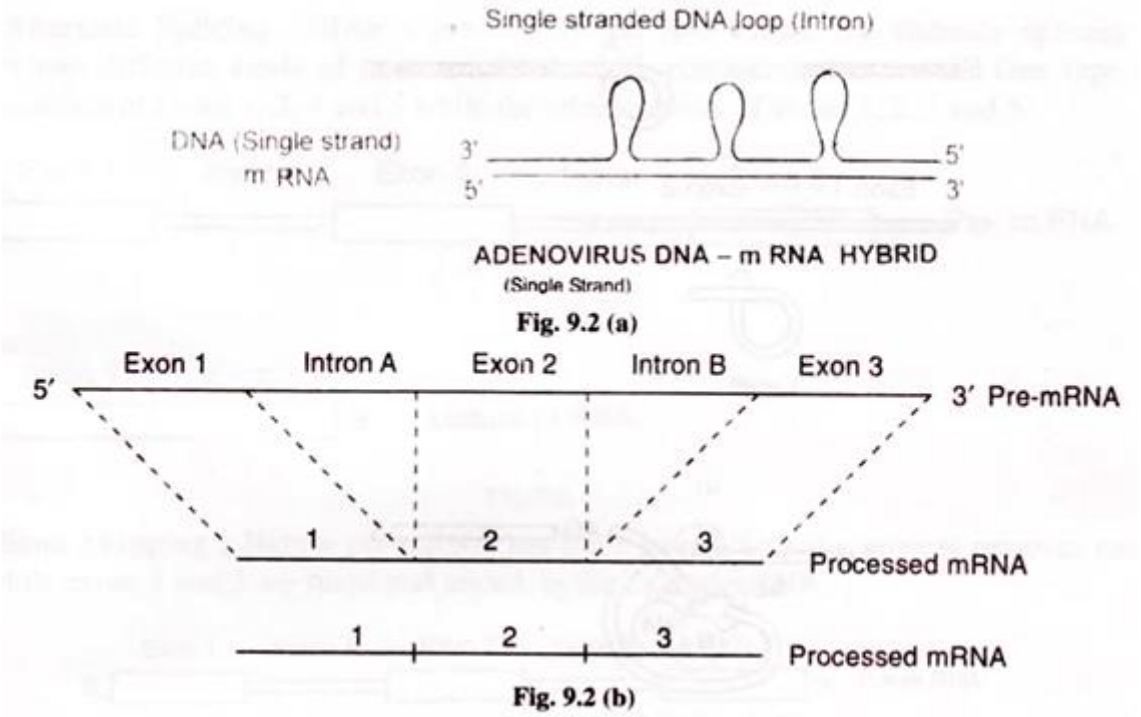


Fig. 9.2 (a)

Fig. 9.2 (b)

The splicing by excision is remarkably precise and accurate and cuts or nicks are made at highly precise position. Because even if an error of one base occurs, the correct reading frame would to disturbed and wrong amino acid will be coded. Chambon studied the boundaries of the introns.

The end of intron always have GU towards the 5′-end (5′-GU-3′) and AG towards the 3′-end (5′-AG-3′). This is known as GU-AG rule or chambon's rule according to which an intron begins with GU and ends with AG. The boundary between exon-intron at 5′-end is called 5′-splice site, while boundary between intron-exon at 3′-end is called 3′-splice site.

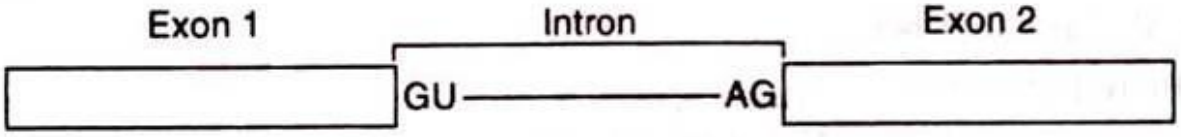


Fig. 9.3.

Introns are excised one by one. The AG at the 3′-end is preceded by a pyrimidine rich sequence called polypyrimidine tract (Py tract). About 10-40 bases upstream of the polypyrimidine tract is a sequence called branchpoint sequence (A), which is 5′-UACUAAC- 3′ in yeast. This sequence is required in lariat formation during splicing process.

RNA Splicing:

Splicing is generally performed by endonuclease enzymes cleaving the introns at both ends. Phosphodiester bond between sugar and phosphate at the junction between intron and exon is cleaved. The freed 5′-end of the intron joins the branch point sequence of form lariat.

Spliceosome:

Splicing is performed by a large complex called spliceosome. The spliceosome is made up of small nuclear ribonuclear proteins (sn RNP) called snurps. These consist of RNAs which are rich in uracil and are of several types U1, U2, U4, U5 and U6 which are collectively called small nuclear RNAs (sn RNA).

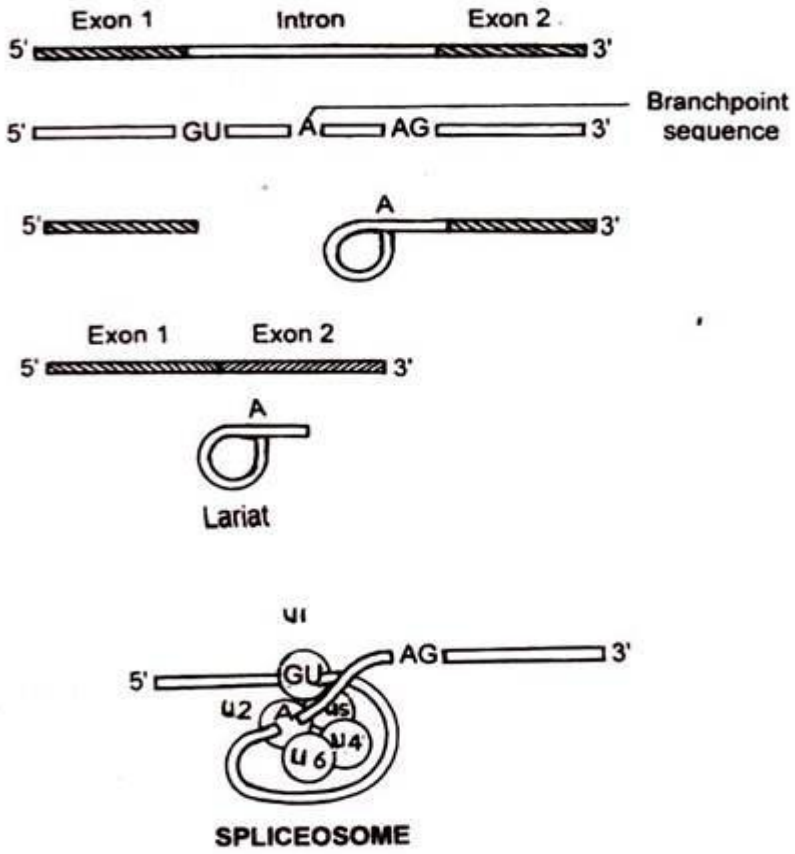


Fig. 9.4.

RNA Splicing:

Splicing is generally performed by endonuclease enzymes cleaving the introns at both ends. Phosphodiester bond between sugar and phosphate at the junction between intron and exon is cleaved. The freed 5′-end of the intron joins the branch point sequence of form lariat.

Spliceosome:

Splicing is performed by a large complex called spliceosome. The spliceosome is made up of small nuclear ribonuclear proteins (sn RNP) called snurps. These consist of RNAs which are

rich in uracil and are of several types U1, U2, U4, U5 and U6 which are collectively called small nuclear RNAs (sn RNA).

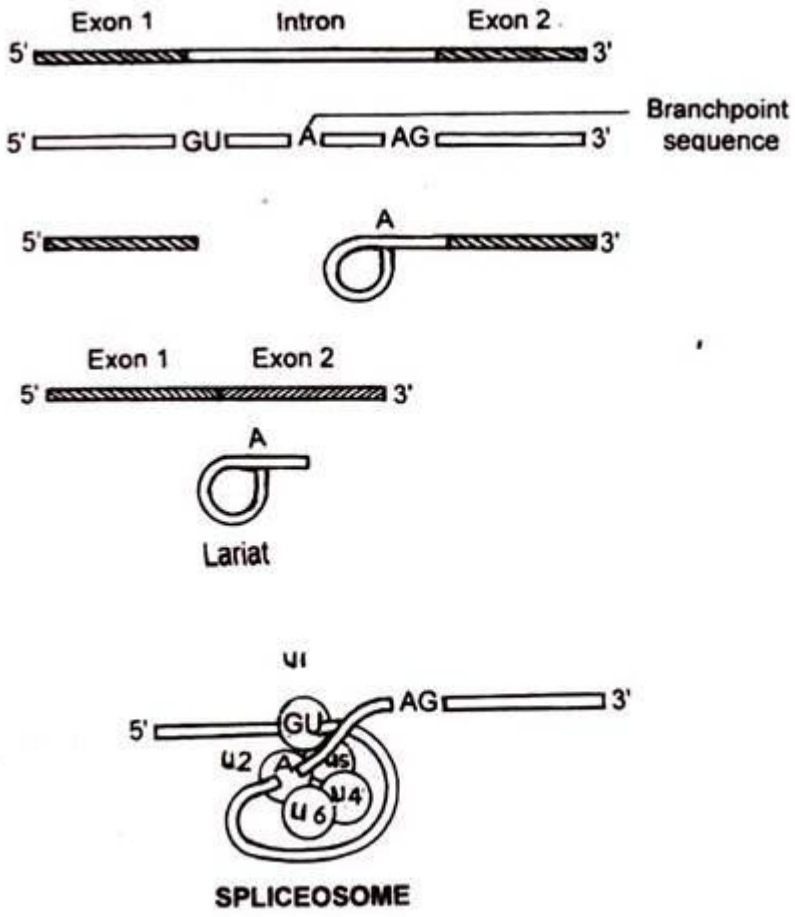


Fig. 9.4.

Mechanism of Splicing:

At first U1 snRNP recognize and-break the 5'splice site of the intron and bring it closer to branch site. Then the complex of sn RNP of U2, U4, U5 and U6 bind to the intron. The complex of snRNPs and precursor mRNA of the intron is called spliceosome. The spliceosome is looped out. This loop of intron is called lariat which is discarded and degraded. The exons on either side of the removed intron are brought closer and ligation seals them together.

Self Splicing:

In some cases specially group I introns, the intron itself folds and catalyses its own splicing. Here RNA of the intron functions as an enzyme and behaves like an endonuclease to splice out the intron. The RNA which acts as an enzyme is called ribozyme. There the splicing of the intron sequences requires no other enzyme. The intron is released in a linear form which is subsequently degraded.

Alternate Processing of Pre-mRNA:

Primary transcript (pre-mRNA) generally exhibits one pattern of RNA splicing to produce a single type of mature mRNA. But in higher eukaryotes, pre-mRNA exhibits alternate or differential processing patterns to yield different mRNAs containing different exons which encode different proteins. One pre-mRNA after alternate splicing produces more than one type of mature mRNAs.

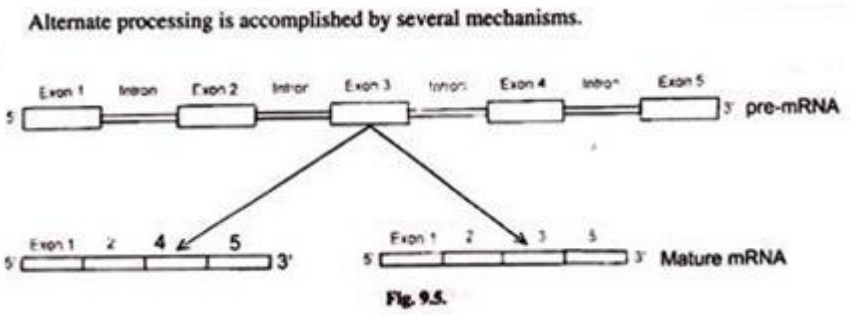


Fig. 9.5.

1. Alternate Splicing:

Here a pre-mRNA has five exons. By alternate splicing it produces two different kinds of mature mRNAs both containing four exons. One type of mRNA consists of exons 1, 2, 4 and 5 while the other consists of exons 1, 2, 3 and 5.

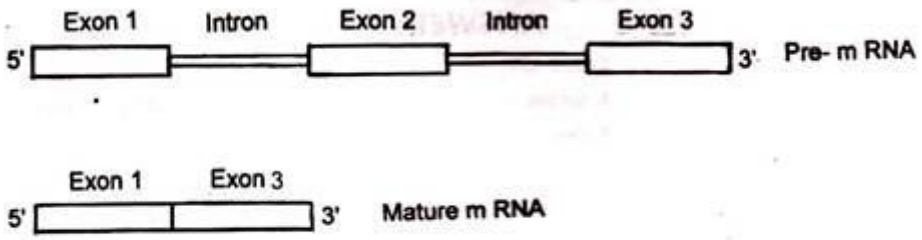


Fig. 9.6.

2. Exon Skipping:

Here a pre-mRNA has three exons. Splicing process removes exon no. 2, while exons 1 and 3 are fused and sealed, in the mature mRNA.

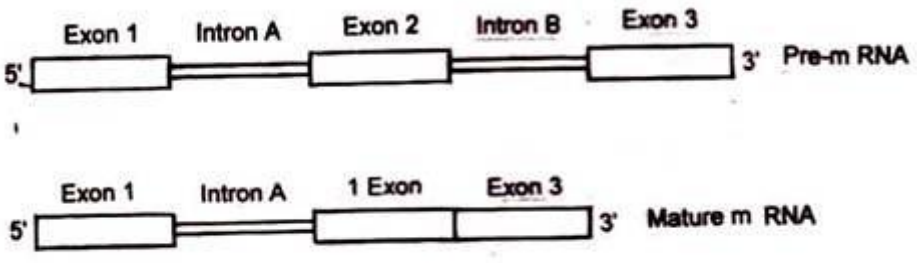


Fig. 9.7.

3. Intron Retaining:

Here a pre-mRNA has three exons and two introns. Splicing process removes the intron no. B while intron no. A is retained in the mature mRNA. It is possible to retain only a part of an intron.

D. RNA Editing:

Pre-mRNA may also undergo RNA editing in which the sequence of the pre- mRNA is altered by the insertion, deletion or substitution of bases. RNA editing was first identified in the mitochondrial gene in which the transcripts were found to be extensively modified by the insertion of uracil residues. RNA editing, may involve either the whole gene, i.e., pan-editing or just a few bases, i.e., minor editing.

Different types of RNA editing are:

(a) Base insertion/deletion type (Fig. 16.9C)

(i) U-insertion/deletion editing in the kinetoplastid, mitochondria is catalysed by editosome through cleavage, addition or removal and ligation utilizing guide RNAs (gRNAs).
(ii) C-insertion/dinucleotide (GC, GU, CU, AA, AU) editing in mitochondria of slime moulds is achieved by slippery transcription.
(iii) G or A-insertion editing found in negative strand RNA viruses and Ebola virus respectively.

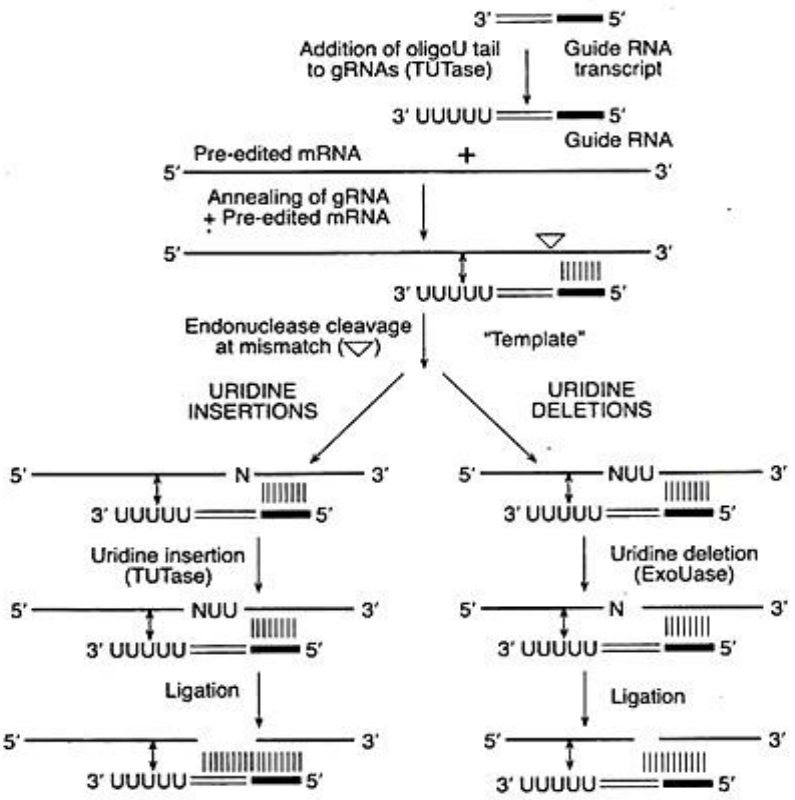


Fig. 16.9C: Mechanism of U insertion and deletion editing in trypanosome mitochondria. Enzymes that carry out each of the steps in U insertion and U deletion editing are shown. The site of pre-edited mRNA cleavage upstream of the anchor duplex is shown by a triangle. Guide RNA (gRNA) domains are indicated, with the 'anchor' as solid black segment, the 'guiding region' as double lined segment, and the 'oligo (U) tail' as UUUUU (from P. K. Gupta)

(b) Base substitution/modification type (Fig. 16.9D)

(i) C to U editing in plant mitochondria and chloroplasts, mammalian apo B, etc. involves transition due to deamination of cytosine by cytidine deaminase.
(ii) A to I (inosine) editing in glutamate receptor, hepatitis delta virus, etc. occurs through deamination of adenosine with the help of adenosine deaminases acting on RNA (ADARs) by specifically target- ting single nucleotides within partially double stranded pre-mRNAs.
(iii) A to G or U to A or U to G editing found in vertebrate mRNAs.

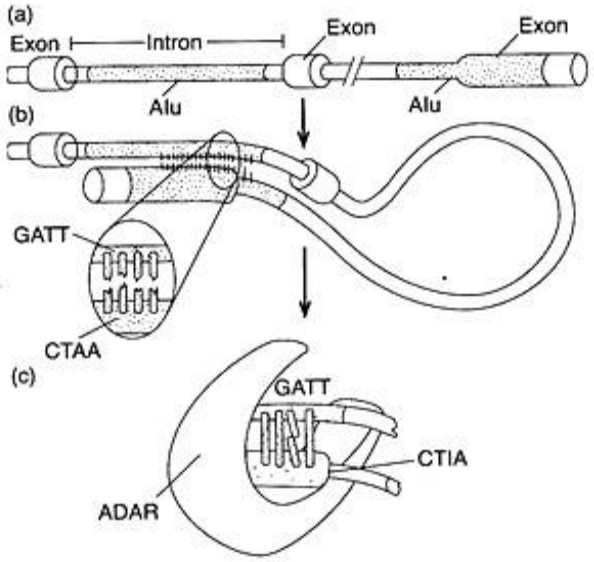


Fig. 16.9D: ADAR-mediated editing. (a) Pre-mRNA is transcribed from DNA. The gene contains two Alu repeats with opposite orientations, one of which overlaps with an exon. (b) The two oppositely oriented Alu sequnces form a dsRNA structure. (c) An enzyme of the ADAR family edits some of the adenosines (A) in the dsRNA stqucture into inosines (I) (from P. K. Gupta)

Processing of tRNA:

tRNA undergoes extensive processing. The mature tRNAs consists of 80-90 nucleotides. But the precursor tRNA is much longer. For example $tRNA^{Tyr}$ which is a tyrosine carrying tRNA contains 350 nucleotides. Processing discards useless sequences. This is done by enzymes RNase D, RNase E, RNase F and RNase P. Nucleotides are removed from both 5′ and 3′ ends. Endonucleases also remove many sequences. Cleaving is done after the primary transcript has folded and formed characteristic stems and loop structure by extensive complementary base pairing. RNase P is a ribozyme. The 5′-CCA-3′ sequence at 3′-end of mature tRNA is added by the enzyme tRNA nucleotidyl transferase. This generates the mature 3′-end the tRNA.

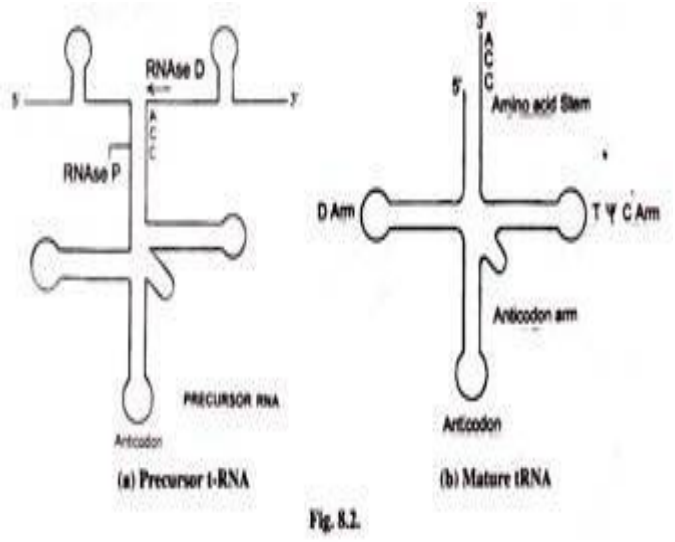


Fig. 8.2.

Several unusual bases are formed by the modification of normal existing bases A, G, C and U by the enzymatic action. These modified bases are pseudouridine (Ψ), 2- isopentenyladenosine (2 ip A), 2-O-methylguanosine (2m G), 4-thiouridine (4 t μ), Ribothymidine, dihyrouridine and inosine.

rRNA Processing:

In rRNA processing in both prokaryotes and eukaryotes, the primary transcript undergoes some changes. Some nucleotides are removed by exonucleases and endonucleases. New nucleotides are added both at 5′-end and 3′-end. Certain nucleotides are modified.

rRNA Processing in Prokaryotes:

In E. coli there are several different operons in rRNA. Each operon contains one copy of 5S, 16S and 23S rRNA sequences. In addition, several coding sequences of tRNA molecules are also present in rRNA operons. These primary transcriptions are processed to give both rRNA and tRNA molecules. All these molecules are cleaved from a continuous transcript of more than 5000 nucleotides.

rRNA Processing in Eukaryotes:

The precursor rRNA in eukaryotes contain one copy of 18S coding region, one copy of 5.8S and one copy of 28S coding region. This rRNA precursor is transcribed in the nucleolus by RNA polymerase I. The eukaryotic 5S rRNA is transcribed by RNA polymerase III from an unlinked gene. In both prokaryotes and eukaryotes rRNA molecules form secondary structures of numerous double stranded stems by complementary base pairing and single stranded loops.

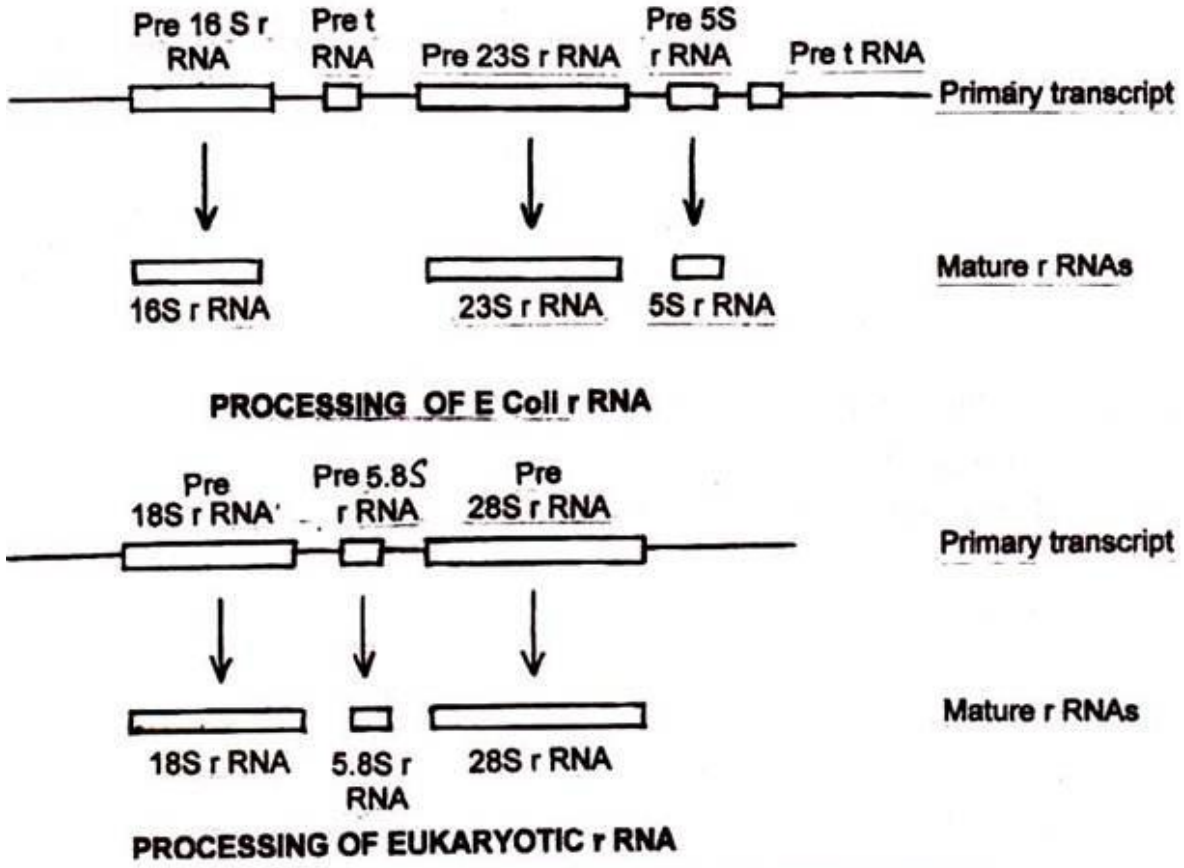


Fig. 8.3. Processing of rRNA primary transcript

Small nucleolar RNA (snoRNA) is required for the processing of rRNA molecules in eukaryotes. The snoRNAs are present in the nucleolus where rRNA is processed. Ribosomes are also assembled in nucleolus.

The RNA molecules in cells are present in the form of complexes formed with proteins. Specific proteins bind to specific RNAs. The RNA-protein complexes are called ribonuclear proteins or RNPs. Ribosomes are the largest RNPs formed by rRNA molecules forming complexes with specific ribosomal proteins. In *E. coli* ribosomes account for 25% of the dry weight of the cell. It consists of 10% of total proteins and 80% of total RNA. In prokaryotes 70S ribosome has 30S and 50S subunits. The 30S subunit has one copy of 16S rRNA and 21 different proteins. The 50S subunit has 23S rRNA, 5S rRNA and 31 different proteins.

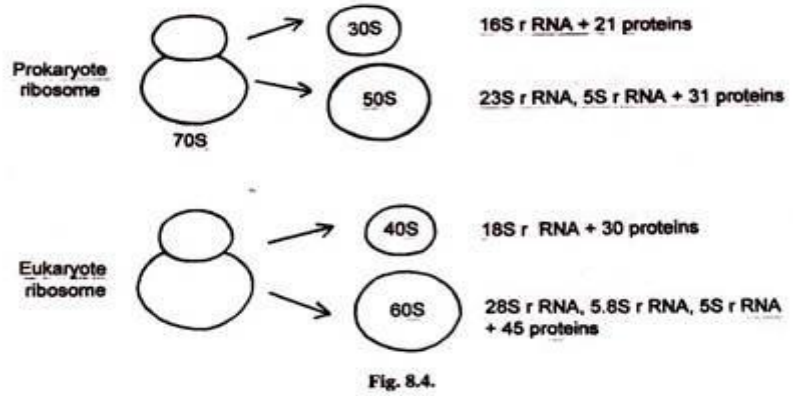


Fig. 8.4.

In eukaryotes 80S ribosome has 40S and 60S subunits. The 40S subunit has 18S rRNA and 30 different proteins. The 60S subunit has 28S rRNA, 5.8S rRNA, 5.8S rRNA and 45 different proteins.

Catalytic RNA: Like protein enzymes some RNA molecules function as enzymes. Like protein enzymes some RNA sequences form complex tertiary structures and play the role of biological catalysts. Such RNA enzymes are known as ribozymes. Like enzymes they have an active site, a binding site for a substrate and a binding site for a co-factor such as metal ion.

Ribozymes have been extensively studied in rRNA of a ciliate protozoan *Tetrahymena thermophila* and are also found in fungal mitochondria, in phage T_4 and E. coli bacteria.

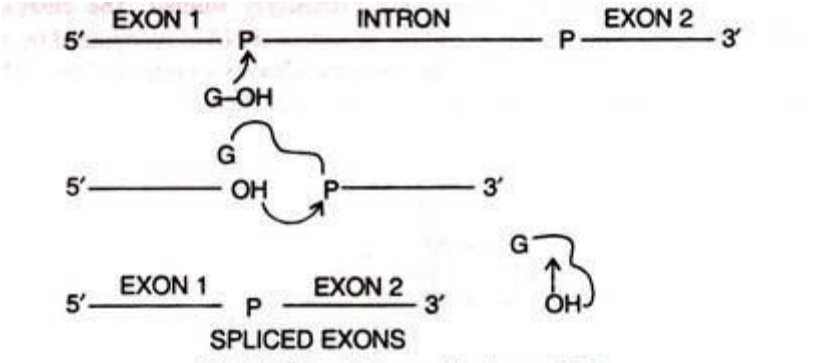


Fig. 10.1. Self splicing reaction in pre-rRNA.

Group 1 introns have an intrinsic ability to splice themselves which is called self splicing or autosplicing. Here no enzyme is required and RNA sequence functions as its own ribozyme and behaves like an endonuclease. In rRNA of Tetrahymena the splicing process involves binding of a single guanosine (G) to a site within the intron which is to be spliced out.

The splicing involves two transesterification reactions. First a free guanosine nucleotide attacks the phosphodiester bond at the 5′ splice site of the intron cleaving the 5′ end of intron. Second trans-esterification reaction splices the phosphodiester bond at the 3′-splice site of next exon causing cleavage. This causes the 3′ end of exon 1 to form a phosphodiester bond with 5′ end of exon 2, thus joining them. The intron is released in a linear form, which is subsequently degraded. It this way the RNA molecule functions as its own ribozyme as it splices itself without the help of any protein enzyme.

Another ribozyme is ribonuclease P (RNAse P) which is a ribonuclease involved in the processing of pre-tRNA. The RNAse P is an ribozyme composed of one RNA component and one protein component. Its RNA component is a ribozyme which catalyses the processing of precursors tRNA.

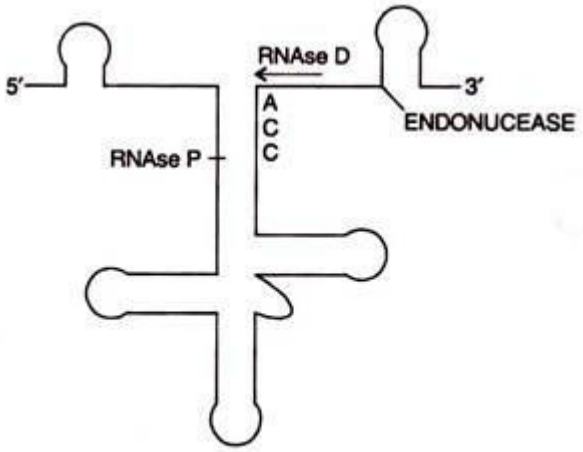


Fig. 2

Fig. 10.2. Role of ribozymes in processing of pre t-RNA.

Precursor tRNA is a very long molecule. During processing its large portion is removed and degraded. Precursor tRNA is cleaved at both 5′-end and 3′-end. In E. coli the tRNA which carries amino acid tyrosine ($tRNA^{lyr}$) has been extensively studied. The cleavage occurs differently at 5′-end and 3′-end. At the 5′-end the precursor tRNA is recognized by an enzyme known as RNAse P which is a ribozyme. This ribozyme cleaves a portion of pre-tRNA so that correct 5′-end of mature tRNA is produced. This proves that all enzymes are not proteins.

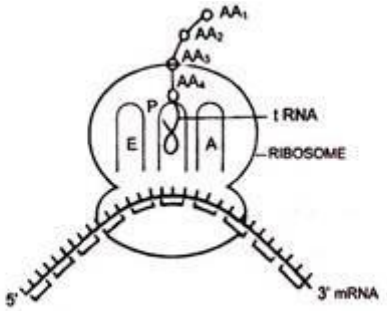


Fig. 10.3. 23S rRNA is a ribozyme responsible for catalyzing peptide bond formation between successive amino acids.

Role of Ribozyme in Protein Synthesis:

In protein synthesis, during elongation phase, the amino acids are added one-by-one. During elongation phase, a tRNA brings a new amino acid to the ribosome according to the codons of mRNA. Each new amino acid is then attached to the end of the growing polypeptide chain. During chain elongation, a peptide bond is formed between the amino acid at "A" site and growing polypeptide chain, which is transferred from "P" site to "A" site. The transfer is known as peptidyl transfer reaction.

This transfer involves formation of peptide bond between amino acid at the "A" site and polypeptide chain, thus lengthening the chain by one amino acid. It was discovered that the peptidyl transferase. Which catalyses the peptide bond formation between successive amino acids consists of several proteins and a $_{23}$S rRNA molecule present in the ribosome. This 23S rRNA is a ribozyme and is responsible for catalysing peptide bond formation between successive amino acid.

Hammerhead Ribozyme:

Hammerhead ribozyme is another small RNA which performs self cleavage reaction. It is found in viroids, which have RNA as genetic material and they infect the plants. When viroid replicates by rolling circle, it produces a continuous RNA chain, which consists of multiple copies of RNA. This continuous chain RNA undergoes cleavage to form single viroids or monomers. The self-cleaving RNA sequence at the junction of monomers is called hammerhead because of the shape of its secondary structure.

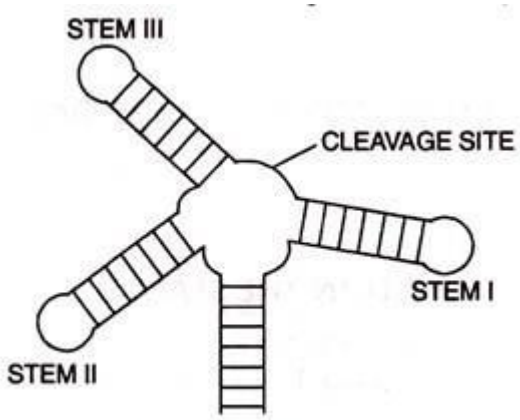


Fig. 10.4. Hammerhead Ribozyme

It consists of three complementary base paired stems. These are stem I, stem II and stem III surrounding a core of non-complementary nucleotides. This core lies at their junction and undergoes self-cleavage reaction. The catalytic reaction centre lies between stem 2/3 and stem I. This centre contains a magnesium ion that initiates hydrolysis reaction. The reaction breaks the RNA chain producing a 2′, 3′ cyclic phosphate and a free 5′-hydroxyl end. Like antisense RNA, which is used to silence the genes, rihozyme is also used to turn off specific genes.

Discovery of Split Genes:

During 1970, in some mammalian viruses (e.g. adenoviruses) it was found that the DNA sequences coding for a polypeptide were not present continuously but were split into several pieces. Therefore, these genes were variously named as split genes or introns, interrupted genes or intervening sequences, inserts, Junk DNA. For the discovery of split genes in adenoviruses and higher organisms, Richards J. Roberts and Phillip Sharp were awarded Nobel Prize in 1993.

As shown in Fig 6.4 a DNA sequence codes for mRNA but the complete corresponding sequence of DNA is not found in mRNA. Certain sequences of DNA are missing in mRNA. The sequences present in DNA but missing in mRNA are called intervening sequences or introns, and the sequences of DNA found in RNA are known as exons. The exons code for mRNA. For the first time W. Gilbert used the term introns and exons. After transcription a limited RNA transcript has the intron. Genes coding for rRNA and tRNA may also be intervened. The introns are also found in some eubacteria, cyanobacteria and archaeobacteria. For some time it was not certain how mRNA is synthesized from a DNA containing introns.

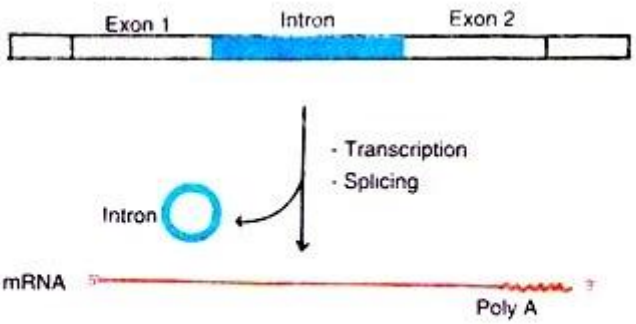


Fig. 6.4 : The split genes have exons separated by introns. Removal of introns through RNA splicing.

Evolution of Split Genes*:*

Before the discovery of split genes in 1977, all the genes analysed in detail were the bacterial genes. Bacteria were considered to resemble with the simpler cell from which eukaryotes must have been evolved. Now, it is supposed that split genes are the ancient condition and bacteria lost their introns only after evolution of most of their proteins.

Evidence for the ancient origin of introns has been obtained by the examination of the gene that encodes the ubiquitous enzyme, triose phosphate isomerase (TPI). The TPI is coded by a gene that contains six introns (in vertebrates), five of these are present at the same position as in maize. This shows that five introns were present in the gene before evolution of eukaryotes about 10^9 years ago.

The TPI plays a key role in cell metabolism that catalyses the inter-conversion of glyceraldehyde 3-phosphate and dihydroxy acetone phosphate- a central step in glycolysis and glycogenesis. By comparing this enzyme in various organisms it appears that the TPI evolved before the divergence of prokaryotes and eukaryotes from a common ancestor cell progenote.

The unicellular organisms under a strong selection pressure minimised the superfluous genome in their cell, whereas there was no such pressure on multicellular organisms. That is why Aspergillus has five introns and Saccharomyces has none. Precise loss of introns would have occurred by deletion in prokaryotes. The loss of introns requires the exact rejoining of DNA coding sequence.

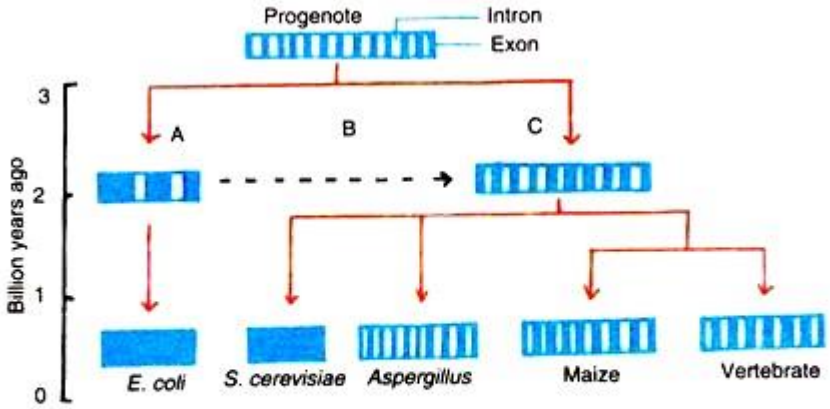


Fig. 6.5 : Outline of evolution of a particular gene. Introns were present in progenote before the evolution of archaeobacteria, eubacteria and eukaryotes. A, eubacteria that formed mitochondria and chloroplasts; B, approximate time of the endosymbiotic events that gave to mitochondria and chloroplast, and C, anaerobic eukaryotes.

The most likely source of the information is needed for such event in an mRNA transcript of the original gene from which introns are to be removed. The mRNA may be copied back into DNA by reverse transcriptase.

The recombination enzymes allow the DNA copies to become paired with the original sequence resulting in intron less form by a gene-conversion type of event. This pathway of intronless has been demonstrated in laboratory in S. cerevisiae.

Alternative splicing:

Alternative splicing, or differential splicing, is a regulated process during gene expression that results in a single gene coding for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final, processed messenger RNA (mRNA) produced from that gene. Consequently, the proteins translated from alternatively spliced mRNAs will contain differences in their amino acid sequence and, often, in their biological functions. Notably, alternative splicing allows the human genome to direct the synthesis of many more proteins than would be expected from its 20,000 protein-coding genes.

Alternative splicing occurs as a normal phenomenon in eukaryotes, where it greatly increases the biodiversity of proteins that can be encoded by the genome; in humans, ~95% of multi-exonic genes are alternatively spliced. There are numerous modes of alternative splicing observed, of which the most common is exon skipping. In this mode, a particular exon may be included in mRNAs under some conditions or in particular tissues, and omitted from the mRNA in others.

The production of alternatively spliced mRNAs is regulated by a system of trans-acting proteins that bind to cis-acting sites on the primary transcript itself. Such proteins include splicing activators that promote the usage of a particular splice site, and splicing repressors that reduce the usage of a particular site. Mechanisms of alternative splicing are highly variable, and new examples are constantly being found, particularly through the use of high-throughput techniques. Researchers hope to fully elucidate the regulatory systems involved in splicing, so that alternative splicing products from a given gene under particular conditions ("splicing variants") could be predicted by a "splicing code". Abnormal variations in

splicing are also implicated in disease; a large proportion of human genetic disorders result from splicing variants. Abnormal splicing variants are also thought to contribute to the development of cancer, and splicing factor genes are frequently mutated in different types of cancer.

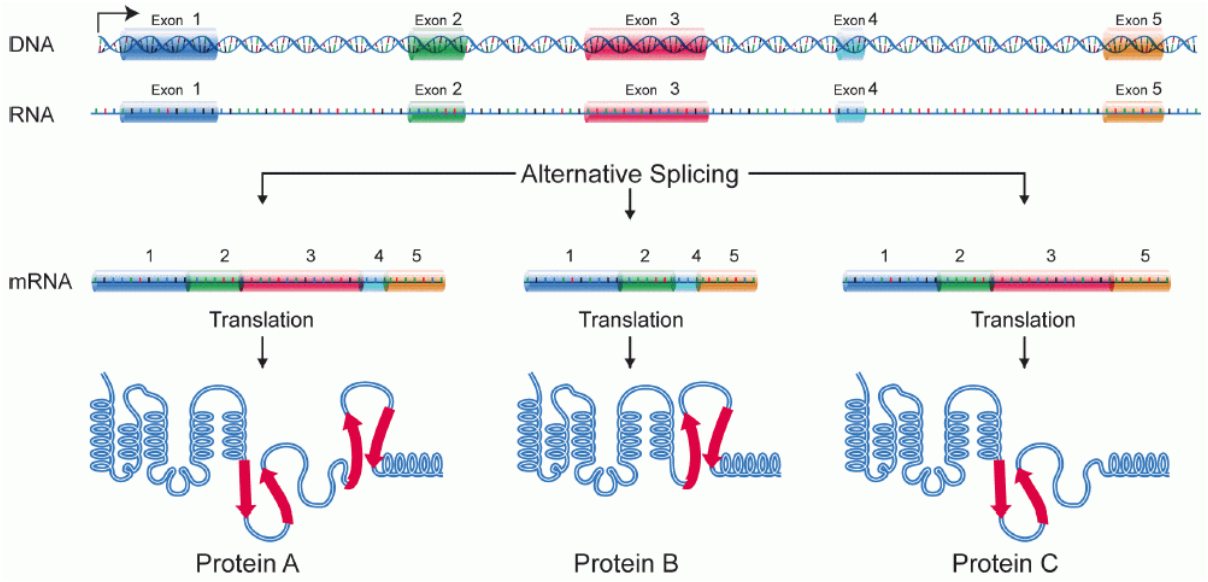


Figure: Alternative splicing produces three protein isoforms.

Five basic modes of alternative splicing are generally recognized.

- Exon skipping or cassette exon: in this case, an exon may be spliced out of the primary transcript or retained. This is the most common mode in mammalian pre-mRNAs.
- Mutually exclusive exons: One of two exons is retained in mRNAs after splicing, but not both.
- Alternative donor site: An alternative 5' splice junction (donor site) is used, changing the 3' boundary of the upstream exon.
- Alternative acceptor site: An alternative 3' splice junction (acceptor site) is used, changing the 5' boundary of the downstream exon.
- Intron retention: A sequence may be spliced out as an intron or simply retained. This is distinguished from exon skipping because the retained sequence is not flanked by introns. If the retained intron is in the coding region, the intron must encode amino acids in frame with the neighbouring exons, or a stop codon or a shift in the reading frame will cause the protein to be non-functional. This is the rarest mode in mammals.

In addition to these primary modes of alternative splicing, there are two other main mechanisms by which different mRNAs may be generated from the same gene; multiple promoters and multiple polyadenylation sites. Use of multiple promoters is properly described as a transcriptional regulation mechanism rather than alternative splicing; by starting transcription at different points, transcripts with different 5'-most exons can be generated. At the other end, multiple polyadenylation sites provide different 3' end points for

the transcript. Both of these mechanisms are found in combination with alternative splicing and provide additional variety in mRNAs derived from a gene.

Adaptive significance:

Alternative splicing is one of several exceptions to the original idea that one DNA sequence codes for one polypeptide (the One gene-one enzyme hypothesis). It might be more correct now to say "One gene – many polypeptides". External information is needed in order to decide which polypeptide is produced, given a DNA sequence and pre-mRNA. Since the methods of regulation are inherited, this provides novel ways for mutations to affect gene expression.

It has been proposed that for eukaryotes alternative splicing was a very important step towards higher efficiency, because information can be stored much more economically. Several proteins can be encoded by a single gene, rather than requiring a separate gene for each, and thus allowing a more varied proteome from a genome of limited size. It also provides evolutionary flexibility. A single point mutation may cause a given exon to be occasionally excluded or included from a transcript during splicing, allowing production of a new protein isoform without loss of the original protein. Studies have identified intrinsically disordered regions as enriched in the non-constitutive exons[35] suggesting that protein isoforms may display functional diversity due to the alteration of functional modules within these regions. Such functional diversity achieved by isoforms is reflected by their expression patterns and can be predicted by machine learning approaches. Comparative studies indicate that alternative splicing preceded multicellularity in evolution, and suggest that this mechanism might have been co-opted to assist in the development of multicellular organisms.

Research based on the Human Genome Project and other genome sequencing has shown that humans have only about 30% more genes than the roundworm *Caenorhabditis elegans*, and only about twice as many as the fly *Drosophila melanogaster*. This finding led to speculation that the perceived greater complexity of humans, or vertebrates generally, might be due to higher rates of alternative splicing in humans than are found in invertebrates. However, a study on samples of 100,000 ESTs each from human, mouse, rat, cow, fly (*D. melanogaster*), worm (*C. elegans*), and the plant *Arabidopsis thaliana* found no large differences in frequency of alternatively spliced genes among humans and any of the other animals tested.[41] Another study, however, proposed that these results were an artefact of the different numbers of ESTs available for the various organisms. When they compared alternative splicing frequencies in random subsets of genes from each organism, the authors concluded that vertebrates do have higher rates of alternative splicing than invertebrates.

Micro RNA and other non coding RNA:

Micro RNAs were first described for the worm C. elegans in 1993 by Lee and colleagues in the Victor Ambros lab. However, the term micro RNA was only introduced in 2001. By 2002, miRNAs have been confirmed in various plants and animals, including C. elegans, human and the plant *Arabidopsis thaliana*. Work at the University of Louisville has resulted in the

production of microarrays containing all known miRNAs for human, mouse, rat, dog, C. elegans and Drosophila.

In genetics, micro RNAs (miRNA) are single-stranded RNA molecules of about 21-23 nucleotides in length which regulate gene expression. In other words, a non-coding RNA molecule of approximately 21-23 nucleotides that inhibits mRNA expression is known as micro RNA.

The main points about micro RNA are given below:

1. Micro RNA is involved in regulation of gene expression.
2. In a cell, miRNA is transcribed from DNA but not translated into proteins.
3. Micro RNAs are non-coding molecules of approximately 21-23 nucleotides.
4. Micro RNAs inhibit the expression of mRNA molecule.
5. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules.
6. The main function of miRNAs is to down regulate gene expression.
7. It has been reported that a typical mammalian cell contains as many as 50,000 different miRNAs.
8. Micro RNAs were first described nor the worm *C. elegans* in 1993.
9. The term micro RNA was only introduced in 2001.
10. Only one strand of DNA can function as templates to give rise to miRNA.

Formation of Micro RNAs (miRNA):

The formation of micro RNAs consists of three important steps, viz:
(i) Formation of primary miRNA,
(ii) Formation of precursor miRNA, and
(iii) Formation of mature functional miRNA.

These are discussed below:

1. Formation of Primary miRNA:

The primary transcript is synthesized from DNA template. The miRNAs are first transcribed as primary transcripts or pri-miRNA with a cap and poly-A tail and then processed to pre- miRNA. Either the sense strand or antisense strand of DNA can function as templates to give rise to **miRNA.**

2. Formation of Precursor miRNA from pre-miRNA:

A short, 70-nucleotide stem-loop structure known as pre-miRNA is formed from primary miRNA in the cell nucleus. This processing is performed in animals by a protein complex known as the Microprocessor complex, consisting of the nuclease Drosha and the double-stranded RNA binding protein Pasha.

3. Formation of Mature miRNA from pre-miRNA:

These pre-miRNAs are then processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiates the formation of the RNA-induced silencing complex (RISC). This complex is responsible for the gene silencing observed due to miRNA expression and RNA interference. The genes encoding miRNAs are much longer than the processed mature miRNA molecule.

Thus microRNA (miRNA) is produced from precursor microRNA (pre-miRNA), which is formed from a microRNA primary transcript (pri-miRNA). The process of formation of miRNA can be represented as follows.

DNA ⟶ Primary transcript (pri-miRNA) ⟶ Precursor miRNA (pre-miRNA) ⟶ mature functional miRNA

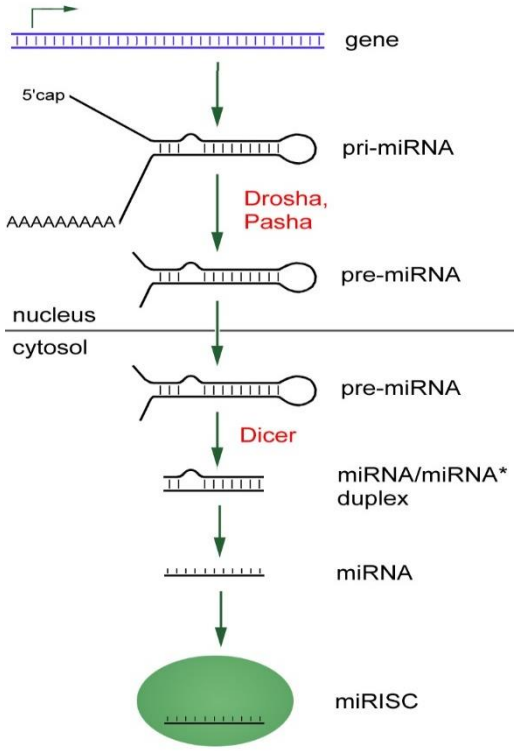


Figure: Different steps of Micro RNA formation

Processing of Micro RNAs (miRNA):

Efficient processing of pre-miRNA by Drosha requires the presence of extended single-stranded RNA on both 3′- and 5′-ends of hairpin molecule. These ssRNA motifs could be' of different composition but their length is important for processing. A bioinformatics analysis of human and fly pri-miRNAs revealed very similar structural regions, called 'basal segments', 'lower stems', 'upper stems' and 'terminal loops'.

Based on these conserved structures, thermodynamic profiles of primiRNA have been determined. The Drosha complex cleaves RNA molecule -2 helical turns away from the terminal loop and ~1 turn away from basal segments. In most analysed molecules this region

contains unpaired nucleotides and the free energy of the duplex is relatively high compared to lower and upper stem regions.
Most pre- miRNAs do not have a perfect double-stranded RNA (dsRNA) structure topped by a terminal loop. Clear similarities between pri-miRNAs encoded in respective (5′- or 3′-) strands have been demonstrated. The pre-miRNA stem-loop is cleaved by Dicer into two complementary short RNA molecules, but only one is integrated into the RISC complex. This strand is known as the guide strand and is selected by the argonaute protein. The remaining strand, known as the anti-guide or passenger strand, is degraded as a RISC complex substrate.
After integration into the active RISC complex, miRNAs base pair with their complementary mRNA molecules and induce mRNA degradation by argonaute proteins, which are catalytically active members of the RISC complex.

Differences between Small Interfering RNA (siRNA) and Micro RNA (miRNA):

Micro RNA [miRNA] is a short (about 21 to 23 nucleotides) single-stranded RNA molecule that is now recognized as playing an important role in gene regulation. It has some similarities and some differences with small interfering RNA (siRNA).

Both miRNA and siRNA have gene regulation functions, but there are slight differences. The miRNA may be slightly shorter [21-23 nucleotides] than siRNA (20 to 25 nucleotides). The miRNA is single-stranded, while siRNA is formed from two complementary strands. The two kinds of RNAs regulate genes in slightly different ways.

The miRNA attaches to a piece of messenger RNA (mRNA)—which is the master template for building a protein – in a non-coding part at one end of the molecule. This acts as a signal to prevent translation of the mRNA into a protein. siRNA, on the other hand, attaches to a coding region of mRNA, and so it physically blocks translation.

TABLE 20.1. Comparison of small interference RNA (siRNA) and microRNA (miRNA)

S.No.	*Particulars*	*siRNA*	*miRNA*
1.	Function	Gene regulation	Gene regulation
2.	Size in nucleotides	20-25	21-23
3.	Strands	Two complementary	Single
4.	Attachment to mRNA	Coding region	Non-Coding region
5.	Effect on translation	Translation is blocked	Translation is blocked

Roles of Micro RNA [miRNA]:

The miRNAs play important role in gene regulation. Micro RNAs arc also expected to be useful in detection of various diseases and their treatment in the years ahead.

The role of miRNAs in gene regulation and disease detection are briefly discussed as follows:

1. Gene Regulation:

The important cellular function of miRNAs is related to gene regulation. The miRNA is attached to the mRNA at a specific point and inhibits protein translation. In other words, the miRNA complex blocks the protein translation machinery. This is thought to be the primary mode of action of plant miRNAs.

In such cases, the formation of the double-stranded RNA through the binding of the miRNA leads to the degradation of the mRNA transcript. It is also believed' that miRNA can prevent translation without causing degradation of the mRNA.

2. Micro RNA and Diseases:

The discovery of miRNA has opened up new areas of research. Now miRNA-based diagnostics and therapeutics are getting increasing importance. The miRNA technology is expected to help in diagnosis and treatment of serious diseases like cancer, heart diseases and diseases related to the nervous system.

This will also help in reclassification of different types of cancers. Thus miRNA technology has wide applications in several areas such as cardiac research; virology, cell biology in general and plant biology. Studies on miRNA expression profiling demonstrated that expression levels of specific miRNAs changed in diseased human hearts, pointing to their involvement in cardiomyopathies. Furthermore, studies on specific miRNAs in animal models have identified distinct roles for miRNAs both during heart development and under pathological conditions, including the regulation of key factors important for cardio genesis, the hypertrophic growth response, and cardiac conductance. Similarly, several miRNAs have been found to have links with some types of cancer.

Concept of RNA Interference Technology

RNA (RNAi) interference is an exciting field in biotechnology. This is a new approach for achieving shut down of target gene. In RNA interference method only a few double-stranded RNA (dsRNA) molecule will be employed per cell to silence the expression of target gene.

Therefore, it is invaluable research tool for down regulation of gene expression particularly in mammalian cells. It is several thousand times powerful than antisense technology. RNA interference was first discovered in worms in 1998. RNA interference operates in plants and other organisms like fungi and mammals. It is also a powerful regulatory process in viral defence and for transposon silencing in plans and animals. Presently, this technique is widely used for medical purpose in silencing of disease related genes.

Intense study on molecular mechanism of RNAi showed that dsRNA molecules are triggered in the partial study of gene silencing. Like antisense oligonucleotides, SiRNA or RNAi can be produced by chemical synthesis and delivered exogenously to cells or SiRNA can be induced by transfection of plasmids that express SiRNA endogenously under the control of RNA pol III promoters. When double-stranded RNA is enter/available in cells, it is recognised and cut into

pieces (21-23 bp) by an enzyme called Dicer. It is a member of RNase III family in ATP dependent progressive manner. These 21-23 bp long with short overhangs is known as short interference RNA or SiRNA. In the next step of duplex SiRNA are the unwind by a helicase activity assemble with special proteins to form what is known as RNA induced silencing complex (RISC). This is otherwise called as sequence-specific mRNA eating machine. Unwinding of SiRNA takes place using ATP and one of the strands recognise complementary mRNA of target gene in the cell. Consequently double-stranded RNA is formed. This dsRNA from mRNA then becomes a substrate for Dicer cleavage activity, which results in the destruction of mRNA and formation of new SiRNA.

Degradation of RNA:

The concentration of an mRNA is a function of both its rate of synthesis and its rate of degradation. For this reason, if two genes are transcribed at the same rate, the steady-state concentration of the corresponding mRNA that is more stable will be higher than the concentration of the other. The stability of an mRNA also determines how rapidly synthesis of the encoded protein can be shut down. For a stable mRNA, synthesis of the encoded protein persists long after transcription of the gene is repressed. Most bacterial mRNAs are unstable, decaying exponentially with a typical half-life of a few minutes. For this reason, a bacterial cell can rapidly adjust the synthesis of proteins to accommodate changes in the cellular environment. Most cells in multicellular organisms, on the other hand, exist in a fairly constant environment and carry out a specific set of functions over periods of days to months or even the lifetime of the organism (nerve cells, for example). Accordingly, most mRNAs of higher eukaryotes have half-lives of many hours. However, some proteins in eukaryotic cells are required only for short periods of time and must be expressed in bursts. For example, certain signaling molecules called cytokines which are involved in the immune response of mammals, are synthesized and secreted in short bursts. Similarly, many of the transcription factors that regulate the onset of the S phase of the cell cycle, such as c-Fos and c-Jun, are synthesized for brief periods only. Expression of such proteins occurs in short bursts because transcription of their genes can be rapidly turned on and off and their mRNAs have unusually short half-lives, on the order of 30minutes or less.

Cytoplasmic mRNAs are degraded by one of the pathways shown in Figure 12-29. For most mRNAs, the length of the poly(A) tail gradually decreases with time through the action of a deadenylating nuclease. When it is shortened sufficiently, PABPI molecules can no longer bind and stabilize interaction of the 5_ cap and initiation factors. The exposed cap then is removed by a decapping enzyme, and the unprotected mRNA is degraded by a 5’ → 3’exonuclease. Removal of the poly(A) tail also makes mRNAs susceptible to degradation by cytoplasmic exosomes containing3’ → 5’ exonucleases. The 5’→ 3’ exonucleases predominate in yeast, and the 3’→ 5’ exosome apparently predominates in mammalian cells. For mRNAs degraded in these deadenylation-dependent pathways, the rate at which they are deadenylated controls the rate at which they are degraded. The rate of deadenylation varies inversely with the frequency of translation initiation for an mRNA: the higher the frequency of initiation, the slower the rate of deadenylation. This relation probably is due to the reciprocal interactions between initiation

factors and PABPI that stabilize the binding of PABPI to the poly(A) tail, thereby protecting it from the deadenylation exonuclease. Many short-lived mRNAs in mammalian cells contain multiple, sometimes overlapping, copies of the sequence AUUUA in their 3' untranslated region. Specific RNA-binding proteins have been found to bind to these 3' AU-rich sequences. Recent experiments suggest that the bound proteins interact with a deadenylating enzyme and with the exosome, thereby promoting the rapid deadenylation and subsequent 3' → 5' degradation of these mRNAs. In this mechanism, the rate of mRNA degradation is uncoupled from the frequency of translation. Thus mRNAs containing the AUUUA sequence can be translated at high frequency, yet also degraded rapidly, allowing the encoded proteins to be expressed in short bursts. Some mRNAs are degraded in pathways that do not involve significant deadenylation. In one of these, mRNAs are decapped before the poly(A) tail is shortened extensively. It appears that certain mRNA sequences make the cap sensitive to the decapping enzyme, but the precise mechanism is unclear. In the other alternative pathway, mRNAs first are cleaved internally by endonucleases. The RNA-induced silencing complex (RISC) discussed earlier is an example of such an endonuclease. The fragments generated by internal cleavage then are degraded by exonucleases.

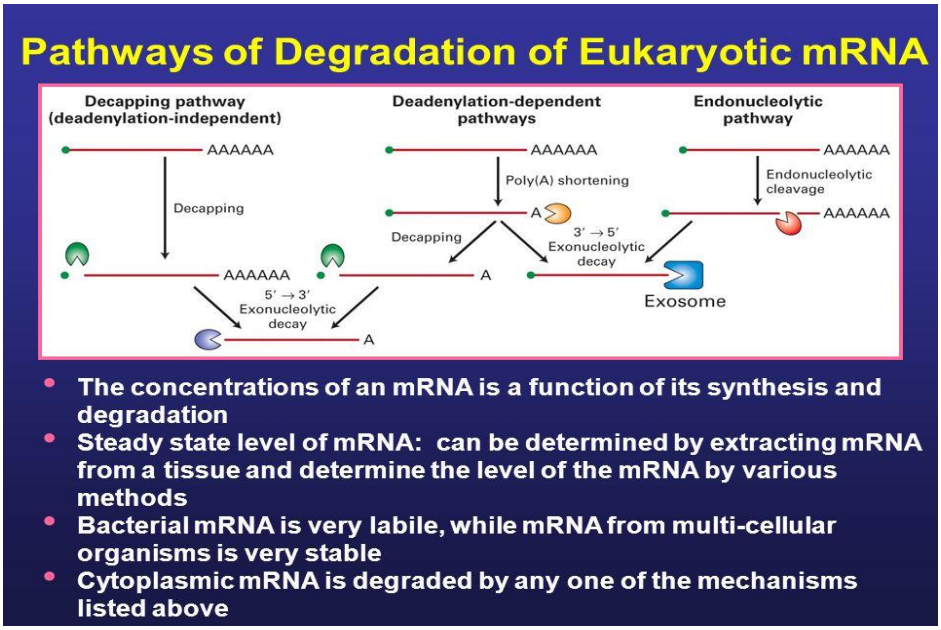


Fig: Pathways for degradation of eukaryotic mRNAs. In the deadenylation-dependent (*middle*) pathways, the poly(A) tail is progressively shortened by a deadenylase (orange)until it reaches a length of 20 or fewer A residues at which the interaction with PABPI is destabilized, leading to weakened interactions between the 5' cap and translation initiation factors. The deadenylated mRNA then may either (1) be decapped and degraded by a 5' → 3' exonuclease or (2) be degraded by a3' → 5' exonuclease in cytoplasmic exosomes. Some mRNAs(*right*) are cleaved internally by an endonuclease, and the fragments degraded by an exosome. Other mRNAs are decapped before they are deadenylated, and then degraded by a 5' → 3' exonuclease.

Probable Questions:

1. How 5’ capping is doe in eukaryotic mRNA?
2. How ply A tailing is done in eukaryotic mRNA?
3. What is intron ? How it is evolved?
3. What is spliceosome? How it removes introns from hnRNA ?
4. What is self splicing of RNA ? Describe the process.
5. What is exon skipping?
6. Describe different types of RNA editing.
7. How nascent rRNAs are processed?
8. How nascent tRNAs are processed?
9. What is ribozyme? How it is used in protein synthesis.
10. What is alternative splicing? How it enhances proteome diversity?
11. How mature miRNA are produced? Explain with diagram.
12. Compare siRNA and miRNA?
13. How RNAs are degraded? Describe in detail.

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.

Unit-III

Transport across the nuclear envelope - structure of nuclear membrane and nuclear pore complexes, processes of nuclear import and export and their regulation.

Objective: In this unit we will discuss nuclear membrane structure and export through nuclear pore. We will also discuss about translation process, tRNA structure and modification, RNA degradation and accuracy during aminoacylation.

Structure of Nuclear Membrane:

The nuclear membrane or karyotheca form an envelope-like structure around the nuclear contents and is commonly known as nuclear envelope. The nuclear membrane in higher plant and animal disappears in late prophase during mitosis and re-forms around the daughter chromosomes during telophase. In lower eukaryotes, the nuclear envelope remains intact throughout mitosis.

It separates nucleus from cytoplasm and functions to facilitate and regulate nucleocytoplasmic interaction. The light microscope provides little information about the nuclear envelope. Under electron microscope the nuclear envelope in the interphase or prophase stage appears to consist of two concentric membranes, viz., inner nuclear membrane and outer nuclear membrane.

Each membrane is about 75 to 90 A thick and lipoproteinous in nature. The outer and inner membranes are separated by perinuclear space of 100-170 A^0. The inter-membrane space is known as perinuclear cisternae. The inner membrane defines the content of nucleus itself and it contains specific proteins that act as binding sites for the nuclear lamina.

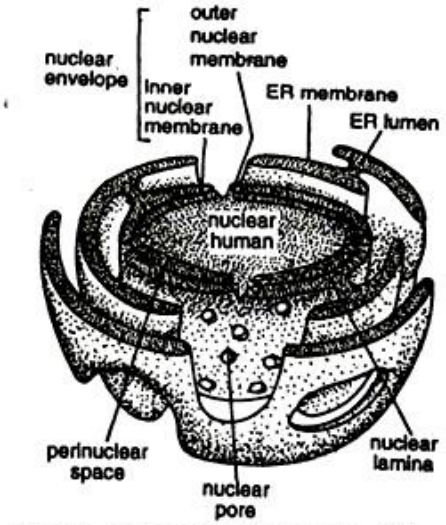


Fig. 9.4: A three-dimensional sketch of the double-membrane envelope that surrounds the nucleus. The nuclear envelope is penetrated by nuclear pores and is continuous with the endoplasmic reticulum.

The outer membrane is rough due to presence of ribosomes (25 nm in diameter) attached with it. The ribosomes are engaged in protein synthesis. The proteins made on these ribosomes are transported into the space between the inner and outer nuclear membrane.
In many cells, the outer nuclear membrane is continuous with rough endoplasmic reticulum. The space between the inner and outer nuclear membrane is continuous with the lumen or inner cavity of the rough endoplasmic reticulum.

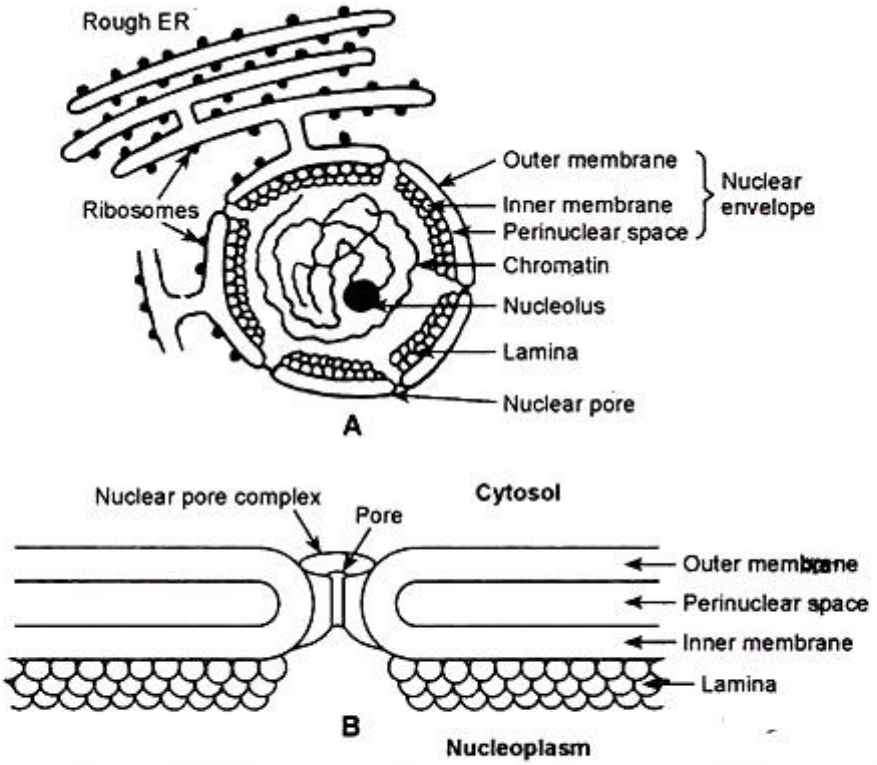


Fig. 2.22. Diagrammatic representation of a nucleus and a nuclear pore complex. **A.** Nuclear envelope connected with ER (out side) and associated with lamina on the inside. **B.** Part of nuclear envelope showing connection of outer and inner membrane at the pore. Lamina is adjacent to the nucleoplasm side of inner membrane.

Under the electron microscope the nuclear envelope appears to consist of two membranes, the outer and the inner nuclear membranes, separated by a perinuclear space of 20 nm (. Each of the two membranes of the nuclear envelope appears to have trilaminar unit membrane structure of 7 – 10 nm thickness. The outer nuclear membrane communicates with endoplasmic reticulum at several points and has ribosomes on the outer side.
The nuclear envelope is perforated by many circular apertures called nuclear pores. Each nuclear pore shows the presence of an electron- dense ring or cylinder called the annulus. The actual opening of the nuclear pore is thus confined to the cavity of the annulus. The annulus extends both into the cytoplasm and the nucleoplasm. The annulus typically consists of eight subunits arranged in radial symmetry around the periphery of the pore. The subunits have been variously interpreted as micro-cylinders, filaments, spheres or ovoid's.
A central ribonucleoprotein granule of 10-15 nm size may be present in some pore complexes and may be absent in adjacent ones. On the inner side of the nuclear envelope of many cell

types is present fibrous material which has been called the fibrous lamina which extends into the nucleoplasm.

Function of Nuclear Envelope:

The nuclear envelope is an interface between nucleus and the cytoplasm. It serves to separate the genetic component of the cell (the chromosomes) from the protein-synthesis machinery (ribosome and ER). It thus provides protection to DNA against the mutagenic effects of cytoplasmic enzymes. It is concerned with nucelocytoplasmic exchange, attachment of structural elements to the cytoplasm, attachment of nuclear components, contribution to other cell membranes and electron transport activity.

Assembly and Disassembly of Nuclear Membrane:

From the light microscopic observation, it is known that the nuclear envelope disappears in late prophase during cell division when the chromatin condenses into chromosome and reappears around the daughter chromosomes during telophase. Disappearance and reappearance of nuclear membrane during cell division are also correlated with disassembly and assembly of nuclear lamina.

Immunofluorescence studies have shown that at the onset of prophase the lamins start to dis-assembly and appear in the cytoplasm. Disassembly of the lamina takes place due to de-polymerisation. Biochemical studies have shown that the de-polymerisation of lamina is due to phosphorylation. The phosphorylation reaction is catalysed by an enzyme, lamin kinase. This whole process, in turn, causes the nuclear envelope to disassembly into a number of small vesicles that disperse into the cytoplasm. The electron microscope, however, shows that lamin B remains attach with these vesicles, whereas lamins A and C are de-polymerised to small oligomer and dispersed into the cytoplasm.

During telophase, after the daughter chromosomes have separated and begin to de-condense, nuclear assembly is started. This process induces the de-phosphorylation and polymerisation of lamins into a fibrous network and simultaneously small nuclear vesicles fuse with each other. Polymerised lamin B associated with fused vesicle in turn binds with polymerised lamins A and C and the whole process makes a bridge between the chromatin and fused small vesicle of nuclear envelope.

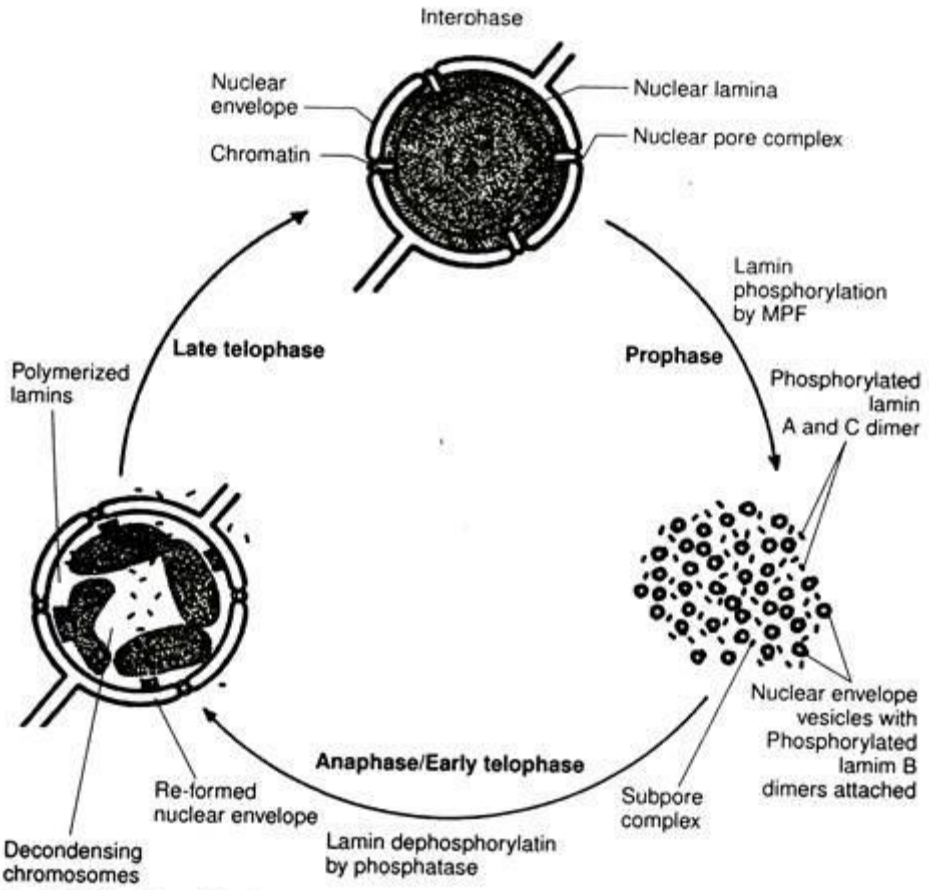


Fig. 9.7: Overview of depolymerisation and polymerisation of nuclear lamins during the cell cycle.

Structure of Nuclear Pore Complex :

The nuclear pore is a large complex structure of 125 million Daltons with 120 nm diameter and 50 nm thickness. Electron micrograph has shown that nuclear pore complexes have an eight-fold symmetry. Pore complex consists of annuli and a structure is formed from a set of large protein granules arranged in octagonal patterns.

The hole in the centre of each complex often appears to be plugged by a large central granule. Eight radial spokes also extends from plug to rings (Fig. 2.67 & 2.68).

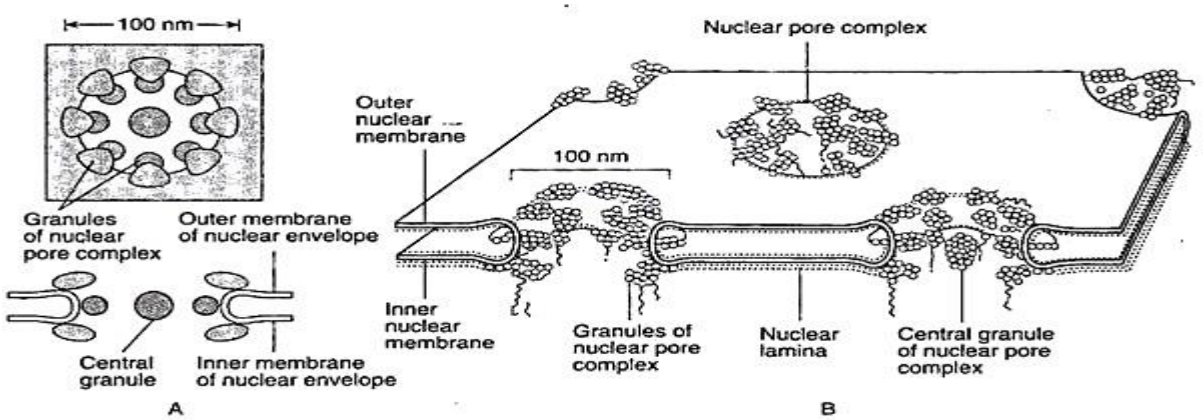


Fig. 2.67: The arrangement of the nuclear pore complexes in the nuclear envelope. A. A top view and a central vertical section. The "central granule" is seen in some pores but not in others; these granules may be a part of the pore, or they may be large complexes caught in transit through it. B. Three-dimensional sketch of a small region of the nuclear evelope (from Roy and De)

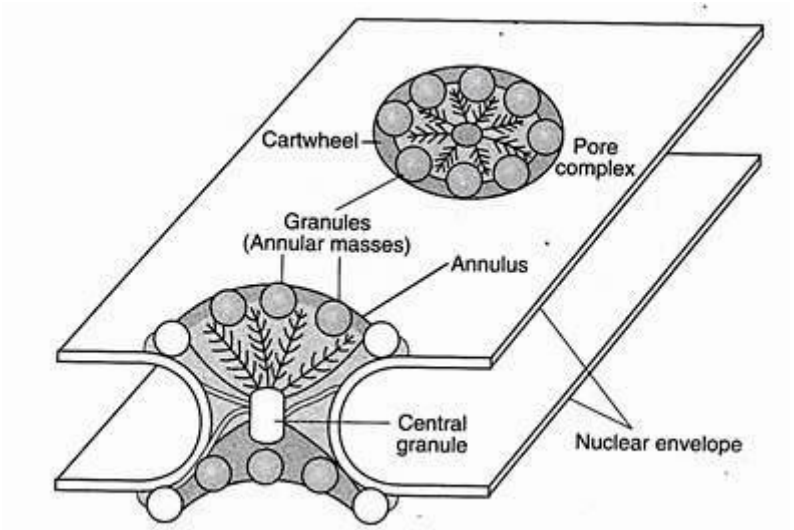


Fig. 2.68: The pore complex, Franke (1970)

During 1990s, significant progress has been made towards better understanding of the structure and function of tine nuclear pore. New pore protein have been identified, cloned, mutants isolated and detailed mechanism of nucleocytoplasmic transfer has been proposed.
Further, the pore has been reconstituted in vitro, a number of signal sequences and one or more signal sequence receptors have been identified and a new 'basket like structure' has been attached to the inner side of the nuclear pore.

It consists of four separate elements:

(i) Scaffold, which included majority of the pore,
(ii) Transporter, the central hub which carries out active transport (both import and export) of proteins and RNAs,
(iii) Short thick filaments attached to the cytoplasmic side of the pore,
(iv) A basket attached to the nucleocytoplasmic side of the pore (Fig. 2.69).
The scaffold is a stack of three closely apposed rings — cytoplasmic ring, nucleocytoplasmic ring and a central ring of thick spokes. The spokes of central ring are attached to the transporter on the inner side and to the nucleocytoplasmic and cytoplasmic rings on the outer

side. Interspersed between the spokes are aqueous channels, 9 nm wide, which allow diffusion of proteins and metabolites between the nucleus and the cytoplasm.

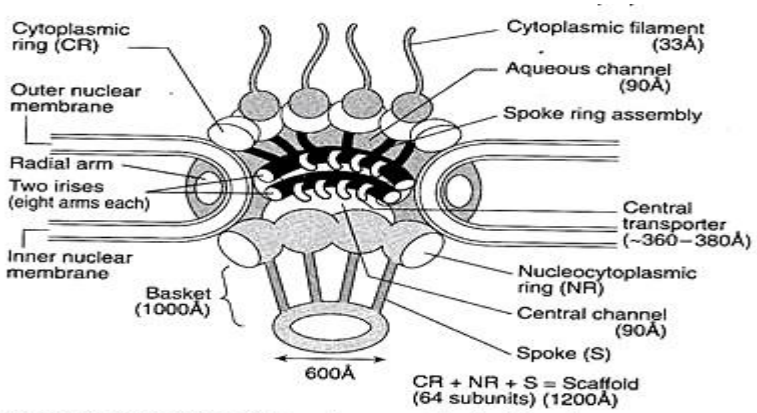


Fig. 2.69: Detailed structure of the nuclear pore, showing iris and transporter

The transporter is a proteinaceous ring, 36-38 nm in diameter and consists of two irises of eight arms each. The two irises are assumed to be stacked atop one another and open sequentially, each like the diaphragm of a camera, to let a nuclear protein or RNA pass through from the nucleus to the cytoplasm. On the cytoplasmic side of the pore, thick filaments of 3.3 nm in diameter, extend into the cytoplasm. On the nuclear side, a large basket like structure is found, which consists of eight filaments of 100 nm long, extending from nucleocytoplasmic ring of the pore and meeting a smaller ring of 60 nm in diameter within the nucleus. This basket plays an important role in RNA export.
The function of nuclear pore complex is the nucleocytoplasmic transport mediated through a number of proteins, called nucleoporin (NUP). The nuclear pore complex has a passive diffusion channel and also can diffuse many substances by active process using energy or signal sequence mediated by carrier molecules.

(a) Import of nuclear proteins through nuclear pore involves the formation of NLS- Protein-lmportin complex (NLS = nuclear localization sequence).

(**b)** Export of RNA from the nucleus across the pore is mediated through NES-Rev protein (NES = nuclear export sequence).

(c) Export followed by reimport of 5SrRNA and UsnRNA occurs through the nuclear pore by the protein with NES like sequence.

Transport across Nuclear envelope:

The entry and exit of large molecules from the cell nucleus is tightly controlled by the nuclear pore complexes (NPCs). Although small molecules can enter the nucleus without regulation macromolecules such as RNA and proteins require association with transport factors like karyopherins called importins to enter the nucleus and exportins to exit.
Protein that must be imported to the nucleus from the cytoplasm carry nuclear localization signals (NLS) that are bound by importins. A NLS is a sequence of amino acids that acts as a tag.

They are diverse in their composition and most commonly hydrophilic, although hydrophobic sequences have also been documented. Proteins, transfer RNA, and assembled ribosomal subunits are exported from the nucleus due to association with exportins, which bind signaling sequences called nuclear export signals (NES). The ability of both importins and exportins to transport their cargo is regulated by the small Ras related GTPase, Ran.
GTPases are enzymes that bind to a molecule called guanosine triphosphate (GTP) which they then hydrolyze to create guanosine diphosphate (GDP) and release energy. Ran is in a different conformation depending on whether it is bound to GTP or GDP. In its GDP bound state, Ran is capable of binding karyopherins (importins and exportins). Importins release cargo upon binding to RanGTP, while exportins must bind RanGTP to form a ternary complex with their export cargo. The dominant nucleotide binding state of Ran depends on whether it is located in the nucleus (RanGTP) or the cytoplasm (RanGDP).

Nuclear import:

Importin proteins bind their cargo in the cytoplasm, after which they are able to interact with the nuclear pore complex and pass through its channel. Once inside the nucleus, interaction with Ran-GTP causes a conformational change in the importin that causes it to dissociate from its cargo. The resulting complex of importin and Ran-GTP then translocates to the cytoplasm, where a protein called Ran Binding Protein (RanBP) separates Ran-GTP from importin. Separation allows access to a GTPase activating protein (GAP) that binds Ran-GTP and induces the hydrolysis of GTP to GDP. The Ran-GDP produced from this process now binds the nuclear transport factor NUTF2 which returns it to the nucleoplasm. Now in the nucleus, the Ran-GDP interacts with a guanine nucleotide exchange factor (GEF) which replaces the GDP with GTP, resulting again in Ran-GTP, and beginning the cycle anew.

Nuclear export:

Nuclear export roughly reverses the import process; in the nucleus, the exportin binds the cargo and Ran-GTP and diffuses through the pore to the cytoplasm, where the complex dissociates. Ran-GTP binds GAP and hydrolyzes GTP, and the resulting Ran-GDP complex is restored to the nucleus where it exchanges its bound ligand for GTP. Hence, whereas importins depend on RanGTP to dissociate from their cargo, exportins require RanGTP in order to bind to their cargo. A specialized mRNA exporter protein moves mature mRNA to the cytoplasm after post-transcriptional modification is complete. This translocation process is actively dependent on the Ran protein, although the specific mechanism is not yet well understood. Some particularly commonly transcribed genes are physically located near nuclear pores to facilitate the translocation process. TRNA export is also dependent on the various modifications it undergoes, thus preventing export of improperly functioning tRNA. This quality control mechanism is important due to tRNA's central role in translation, where it is involved in adding amino acids to a growing peptide chain. The tRNA exporter in vertebrates is called exportin-t. Exportin-t binds directly to its tRNA cargo in the nucleus, a process promoted by the presence of RanGTP. Mutations that affect tRNA's structure inhibit its ability to bind to exportin-t, and consequentially, to be exported, providing the cell with another quality control

step. As described above, once the complex has crossed the envelope it dissociates and releases the tRNA cargo into the cytosol.

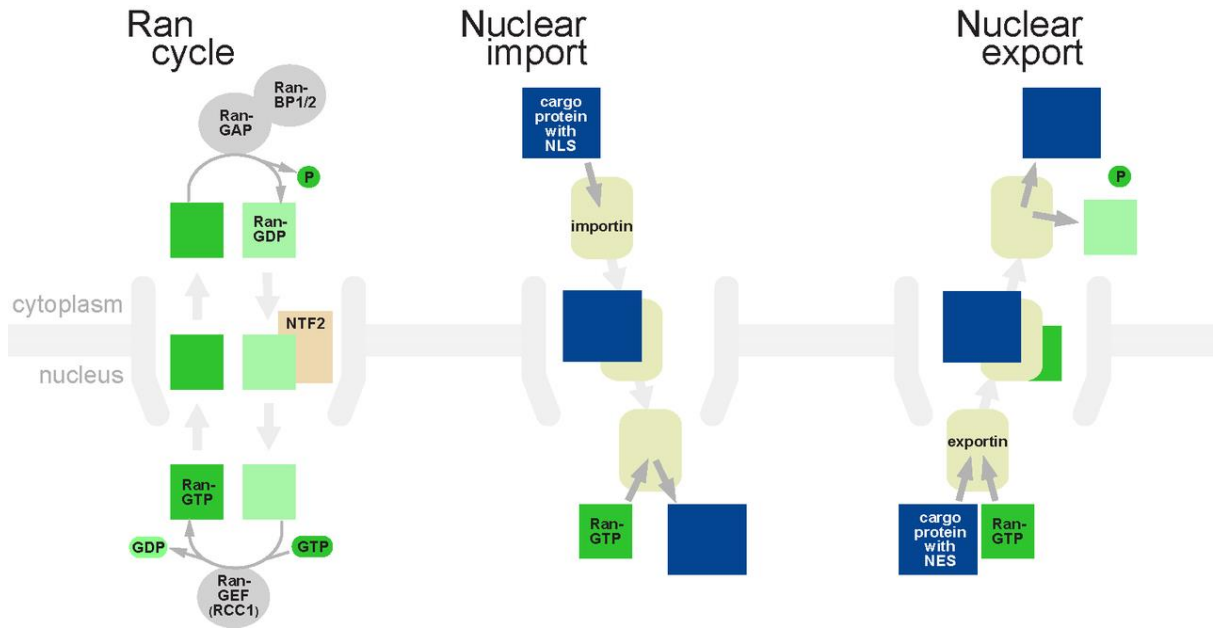


Figure : Macromolecules, such as RNA and proteins, are actively transported across the nuclear membrane in a process called the Ran-GTP nuclear transport cycle.

Protein Shuttling:

Many proteins are known to have both NESs and NLSs and thus shuttle constantly between the nucleus and the cytosol. In certain cases one of these steps (i.e., nuclear import or nuclear export) is regulated, often by post-translational modifications. Protein shuttling can be assessed using a heterokaryon fusion assay

Regulation of nuclear transport:

The nuclear envelope establishes an essential regulatory barrier, which eukaryotic cells can use to control cellular processes such as gene expression and cell cycle progression. Thus, the dynamic compartmentalization of proteins between the nucleus and the cytoplasm can be utilized to spatially and temporally regulate protein function. Use of this nucleocytoplasmic compartmentalization as a method for regulating cellular processes requires rapid, selective, and highly regulated nuclear transport.

All macromolecules that move into and out of the nucleus are transported through nuclear pore complexes, large proteinaceous channels that are embedded in the nuclear envelope. Soluble factors are required to recognize, target, and transport most macromolecules through the nuclear pores. The best characterized nuclear transport process occurs via receptor recognition of classic nuclear localization signals (NLSs)1 on protein cargoes targeted for nuclear import (3, 6). These classic NLS cargo proteins are recognized in the cytoplasm by a heterodimeric receptor composed of importin/karyopherin α and β. Importin α recognizes and binds the NLS, and importin β translocates the trimeric import complex through the nuclear pore (3, 5). Delivery into the nucleus is dependent on the small GTPase Ran, which governs the

interactions between the nuclear transport receptors and macromolecular cargoes and thus confers directionality to nucleocytoplasmic transport. Once inside the nucleus, the cargo is delivered and the transport receptors are recycled to the cytoplasm.

Classic NLSs are typified by a single cluster of basic amino acids (monopartite) or two clusters of basic amino acids separated by a 10- to 12-amino acid linker (bipartite). The prototypical monopartite NLS is that of the SV40 large-T antigen (PKKKRKV) and the prototypical bipartite NLS is that of nucleoplasmin (KRPAATKKAGQAKKKK). Recent studies that have examined the structural and energetic contributions made by individual amino acid residues within the sequence have refined our definition of a functional NLS and have thereby allowed easier identification of potential NLS signals within a protein sequence

The dynamic compartmentalization of NLS-containing proteins requires regulated changes in the relative import and/or export rates of a protein. In numerous cases, transport of an NLS cargo into the nucleus is regulated by phosphorylation . There are at least three ways in which phosphorylation could regulate protein import into the nucleus: 1) phosphorylation could cause a conformational change in the protein, which reveals or masks an NLS sequence; 2) phosphorylation could cause the release or binding of an NLS masking protein; or 3) phosphorylation could directly modulate the affinity of an NLS for the import receptor importin α. Examples of each of these mechanisms include the growth regulatory protein STAT (signal transducers and activators of transcription) where phosphorylation causes the protein to dimerize creating an NLS, the p65 subunit of NF-κB where phosphorylation of NF-κB-bound I-κB leads to degradation of I-κB and unmasking of the NF-κB NLS, and the v-Jun oncoprotein where phosphorylation may directly modulate interactions with the NLS receptor . Direct regulation of the interaction between an NLS cargo and importin α (regulation method 3 above) requires that the phosphorylation site is within or adjacent to the NLS. In contrast, the other two forms of regulation, methods 1 and 2 above, could be due to phosphorylation at any site within the protein. Each of these modes of regulation is non-exclusive, and thus a combination could regulate the nuclear transport of a particular cargo. Here we will focus on the direct modulation of importin α binding by phosphorylation.

A number of important regulatory proteins contain phosphorylation sites located within or adjacent to classic NLS sequences. Recent studies have shown that transport into and out of the nucleus correlates with regulated phosphorylation at these sites. For example, p53, the adenomatous polyposis protein (APC), and the *Saccharomyces cerevisiae* Swi6 protein all have documented phosphorylation sites within or proximal to their NLS sequences. Phosphorylation of these sites is associated with cytoplasmic protein localization and, conversely, hypophosphorylation correlates with nuclear protein localization. Thus, phosphorylation appears to represent an important mechanism to regulate the nuclear transport and consequently the function of an NLS cargo protein. Furthermore, the proximal position of the phosphorylation site to the NLS sequence suggests that the nuclear import of these cargoes may be modulated by directly regulating the binding affinity of the NLS for the NLS receptor.

Importin α structural studies revealed the determinants for specific recognition of classic NLS sequences by the NLS-binding pocket of importin α. Structures of the NLS binding domain of *S. cerevisiae* importin α (amino acid residues 89–530) bound to various NLS peptides showed that this domain of importin α consists of ten helical repeats known as armadillo motifs. These

armadillo motifs form a concave NLS binding groove, which is lined by conserved tryptophan and asparagine residues and surrounded by acidic amino acids. This structure creates specific binding pockets for NLS cargoes that combine both hydrophobic interactions and electrostatic interactions with the positively charged residues of the NLS. These observations suggest that the addition of a negatively charged phosphate group proximal to an NLS could decrease binding of the NLS to importin α by disrupting the electrostatic interactions. Furthermore, a similarly positioned negative group within any classic NLS sequence may change its binding affinity for importin α and thus modulate the intracellular localization of a cargo protein.

Recent studies that correlate differences in the rate of import with changes in the phosphorylation state of NLS-containing proteins have not examined the change in binding affinity between the NLS and its receptor using a quantitative assay. Furthermore, although one of these studies solved the co-crystal structure of a phosphorylated peptide containing an NLS sequence and non-autoinhibited importin α, the phosphorylation site was 14 amino acids upstream of the NLS sequence. In contrast, in our analyses the phosphorylation sites examined are located one amino acid upstream of the NLS. Thus, it is not known to what extent phosphorylation proximal to or within an NLS changes the binding affinity for importin α. Furthermore, it is important to determine if this change in affinity is sufficient to account for the observed changes in protein localization. A complete understanding of phosphorylation-mediated regulation of nuclear import by modulation of the interaction between an NLS and importin α requires a quantitative model for the import of a cargo that correlates the in vitro interaction energies with the in vivo localization of a protein.

Probable Questions:

1. Describe structure of nuclear envelop with diagram.
2. Write the functions of nuclear envelope?
3. How nuclear membrane assembled and disassembled?
4. Describe structure of nuclear pore complex.
5. How nuclear export has occurred?
6. How nuclear import has occurred?
7. How nuclear transport is regulated?

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.

UNIT-IV

Translational machinery and translational control - energetics of amino acid polymerization, tRNAs and their modifications, aminoacyl tRNA synthetases, accuracy during aminoacylation of tRNA, regulation of initiation of translation in eukaryotes, elongation and its control,inhibitors of translations

Objective: In this unit we will discuss about Translational machinery and translational control - energetics of amino acid polymerization, tRNAs and their modifications, aminoacyl tRNA synthetases, accuracy during aminoacylation.

Translation (Protein Synthesis):

Proteins are giant molecules formed by polypeptide chains of hundreds to thousands of amino acids. These polypeptide chains are formed by about twenty kinds of amino acids. An amino acid consists of a basic amino group ($-NH_2$) and an acidic carboxyl group (-COOH). Different arrangement of amino acids in a polypeptide chain makes each protein unique. Proteins are fundamental constituents of protoplasm and building material of the cell.

They take part in the structural and functional organization of the cell. Functional proteins like enzymes and hormones control the metabolism, biosynthesis, energy production, growth regulation, sensory and reproductive functions of the cell. Enzymes are catalysts in most of the biochemical reactions. Even the gene expression is controlled by enzymes. The replication of DNA and transcription of RNA is controlled by the proteinous enzymes.

Components of Protein Synthesis:

Protein synthesis is governed by the genetic information carried in the genes on DNA of the chromosomes.

I. Amino Acids:

Proteins are the polymers of amino acids. Therefore, amino acids form the raw material for protein synthesis. The proteins of living organisms need about 20 amino acids as building blocks or monomers. These are available in the cytoplasmic matrix as an amino acid pool.

II. DNA as Specificity Control:

A cell, in order to maintain its own special characteristics, must manufacture proteins exactly similar to those present already in it. Thus, protein synthesis requires specificity control to provide instructions about the exact sequence in which the given numbers and kinds of amino acids should be linked to get the desired polypeptides. The specificity control is exercised by DNA through mRNA sequences of 3 consecutive nitrogenous bases in the DNA double helix form the biochemical or genetic code. Each base triplet codes for a specific amino acid. Since

the DNA is more or less stable, the proteins formed in a cell are exactly like the preexisting proteins.

III. RNAs:

RNA molecule is a long, un-branched, single-stranded polymer of ribonucleotides (Fig. 7.12). Each nucleotide unit is composed of three smaller molecules: a phosphate group, a 5- carbon ribose sugar, and a nitrogen-containing base. The bases in RNA are adenine, guanine, uracil and cytosine. The various components are linked up as in DNA. There are three types of RNA in every cell: messenger RNA or mRNA, ribosomal RNA or rRNA and transfer RNA or tRNA. The three types of RNAs are transcribed from different regions of DNA template, RNA chain is complementary to the DNA strand which produces it. All the three kinds of RNAs play a role in protein synthesis.

(a) mRNA:

The DNA, that controls protein synthesis, is located in the chromosomes within the nucleus, whereas the ribosomes, on which the protein synthesis actually occurs, are placed in the cytoplasm. Therefore, some sort of agency must exist to carry instructions from the DNA to the ribosomes. This agency does exist in the form of mRNA.

The mRNA carries the message (information) from DNA about the sequence of particular amino acids to be joined to form a polypeptide, hence its name. It is also called informational RNA or template RNA. The mRNA forms about 5% of the total RNA of a cell. Its molecule is linear and the longest of all the three RNA types. Its length is related to the size of the polypeptide to be synthesized with its information.

There is a specific mRNA for each polypeptide. Because of the variation of size in mRNA population in a cell, the mRNA is often called heterogeneous nuclear RNA, or hnRNA:

In eukaryotes, mRNA carries information for one polypeptide only It is monocistronic (monogenic) because it is transcribed from a single cistron (gene) and has a single initiator codon and a single terminator codon.

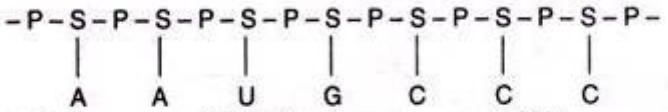


Fig. 7.12 A segment of RNA showing arrangement of its components.

Bacterial mRNA often carries information for more than one polypeptide chains. Such a mRNA is said to be polycistronic (polygenic) because it is transcribed from many contiguous (adjacent) genes. A polycistronic mRNA has an initiator codon and a terminator codon for each polypeptide to be formed by it.

(b) tRNA:

The tRNA has many varieties. Each variety carries a specific amino acid from the amino acid pool to the mRNA on the ribosomes to form a polypeptide, hence its name. The tRNAs form about 15% of the total RNA of a cell. Its' molecule is the smallest of all the RNA types.

Structure of tRNA:

The universally accepted 2-dimensional model of tRNA is the "clover leaf model" formed due to pairing of short complementary sequences and the formation of unpaired loops (Fig. 4.2). The 5′ P end terminates usually into guanine (G), while the 3′ OH end always terminates into a 5′ CCA3′ sequence. Amino acid is carried on the 3′-end, associated with the adenine (A), general structure of clover leaf model is described below.

(1) The 3′-end terminates into 5'CCA3′ sequence that is always unpaired. The terminal A residue is the site at which the amino acid is bound covalently.

(2) Starting from the 3′-end, after the 3′ ACC5′ sequence, the region includes few paired bases.

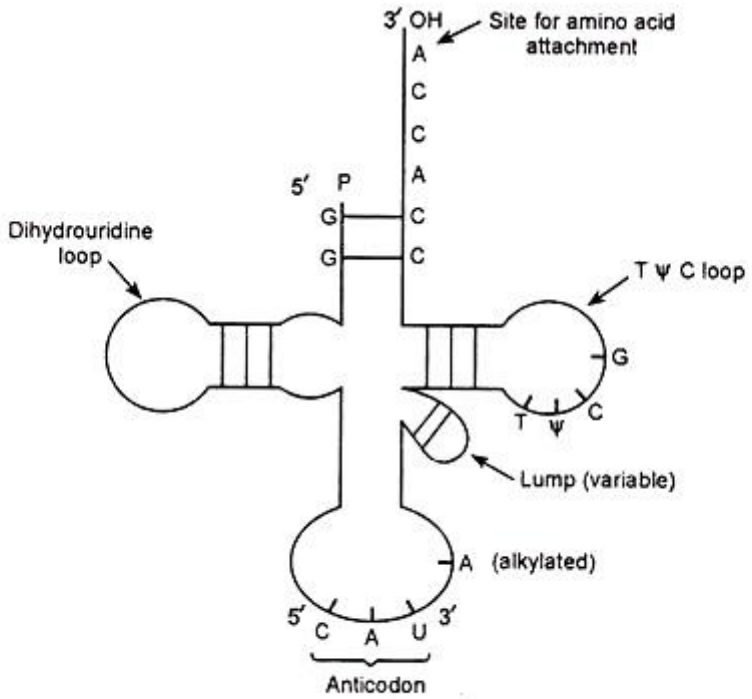


Fig. 4.2. Diagrammatic representation of the clover leaf model of transfer RNA. (This is a model of $tRNA^{met}$).

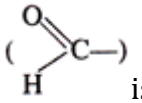
(3) Then comes the first loop containing 7 unpaired bases. This loop is designated as "TΨ C loop" because it always contains a sequence 5′ ribothymidine- pseudouridine- cytidine 3′. This loop is involved in binding to ribosome.

(4) After the "5′ -T Ψ C-3′ loop", in the 5′ direction, there occurs a loop of variable size, called the extra loop or the "lump". The lump may contain 3 to 21 bases.

(5) The third loop contains 7 unpaired bases and it has the "anticodon." Anticodon consists of 3 bases. At the 3′ -end of the anticodon, there is a purine (A or G) while at the 5′ -end, there is always a uracil (U). At the time of protein synthesis, anticodon pairs with its complementary "codon" on mRNA.

(6) The fourth loop is larger than others and contains 8-12 unpaired bases. It is designated as "D-loop" because it is rich in dihydrouridine (UH_2). The enzyme aminoacylsynthetase binds to this loop.

Initiator Transfer RNA:

Marcker and Sanger in 1964 isolated two different types of methionine specific tranfer RNAs. One type, designated as $tRNA_f^{met}$ carries the amino acid N-formyl methionine, while the other type, designated as $tRNA^{met}$ carries the normal methionine. After the formation of aminoacyl-tRNA (methionyl-tRNA), the formate group  is added to methionine to form N-formyl methionine, by the enzyme transformylase. The formate group comes from N-formyltetrahydrofolic acid.

The formate group protects the polypeptide chain from peptide bond formation with other amino acids. In prokaryotes the initiator amino acid is the N-f-methionine, while in eukaryotes, it is the normal methionine. The initiator tRNA in eukaryotes also differs from the $tRNA^{met}$ and is designated as $_tRNA_f^{met}$.

tRNAs and their modifications:

Transfer RNAs (tRNAs), also called **"soluble RNAs"** (sRNAs) are small molecules varying from 75 to 100 nucleotides in length. They carry amino acids and bring them to ribosomes to form the polypeptide chain. There are 20 amino acids but the number of tRNA types is greater because some amino acids are adopted by more than one form of tRNAs.

Transfer RNAs are designated by the amino acid abbreviation in super script, such as, $tRNA^{gly}$ (tRNA for glycine). Different tRNA adopting the same amino acid are designated by number in subscript, e.g. $tRNA_1^{leu}$, $tRNA_2^{leu}$. Such tRNAs are called **"iso-accepting tRNAs".** Transfer RNAs are produced on DNA templates and the reaction is catalysed by DNA dependent RNA polymerase. Precursor tRNA is produced first and then processing is made to produce the mature tRNA. During processing, modifications of some bases occurs, such as dihydrouridine (UH_2), pseudouridine (Ψ), inosine (I), 7-methylguanosine (m^7 G), acetyl cytidine (acC), N^2-dimethyl guanosine (m_2G) and 1- methyl adenosine (m^1A) etc. About 10-15% of the bases in tRNA are modified.

(c) rRNA:

The rRNA molecule is greatly coiled. In combination with proteins, it forms the small and large subunits of the ribosomes, hence its name. It forms about 80% of the total RNA of a cell. The rRNA also seems to play some general role in protein synthesis.

(IV) Ribosomes:

Ribosomes serve as the site for protein synthesis. The small and large subunits of ribosomes occur separately when not involved in protein synthesis. The two sub units form association (join) when protein synthesis starts, and undergo dissociation (separate) when protein synthesis stops. Many ribosomes line up on the mRNA chain during protein synthesis. Such a group of active ribosomes is called a polyribosome, or simply a polysome.

In a polysome, the adjacent ribosomes are about 340 Å apart. The number of ribosomes in a polysome is related to the length of the mRNA molecule, which reflects the length of the polypeptide to be synthesized. It has been established that polypeptides are synthesized at the

polysomes and not at the single free ribosomes as held earlier. This is true for both prokaryotes and eukaryotes as well as for the cell organelles such as mitochondria and plastids.

A ribosome has two binding sites for tRNA molecules. One is called A (acceptor or aminoacyl) site and the other is termed P (peptidyl) site. These sites span across the large and small subunits of the ribosome (Fig. 7.13). The A site receives the tRNA-amino acid complex. From P site, the tRNA leaves after leaving its amino acid to the forming polypeptide. However, the first tRNA-amino acid complex directly enters the P site of the ribosome.

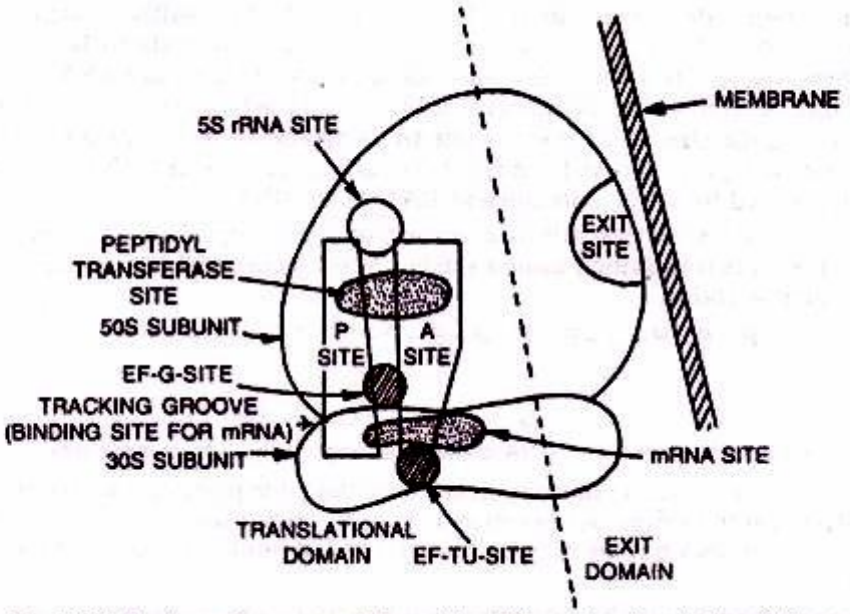


Fig. 7.13. A schematic representation of the different functional sites of ribosomes.

The function of the ribosome is to hold in position the mRNA, tRNA and the associated enzymes controlling the process until a peptide bond forms between the adjacent amino acids.

Mechanisms of Protein Synthesis:

In prokaryotes, the RNA synthesis (transcription) and protein synthesis (translation) take place in the same compartment as there is no separate nucleus. But in eukarytoes, the RNA synthesis takes place in the nucleus while the protein synthesis takes place in the cytoplasm. The mRNA synthesized in the nucleus is exported to cytoplasm through nucleopores.

First, Francis Crick in 1955 suggested and later Zemecnik proved that prior to their incorporation into polypeptides, the amino acids attach to a special adaptor molecule called tRNA. This tRNA has a three nucleotide long anticodon which recognizes three nucleotide long codon on mRNA.

Role of Ribosomes in Protein synthesis:

Ribosome is a macromolecular structure that directs the synthesis of proteins. A ribosome is a multicomponent, compact, ribonucleoprotein particle which contains rRNA, many proteins and enzymes needed for protein synthesis. Ribosome brings together a single mRNA molecule and

tRNAs charged with amino acids in a proper orientation so that the base sequence of mRNA molecule is translated into amino acid 1 sequence of polypeptides.
Ribosome is a nucleoprotein particle having two subunits. These two subunits lie separately but come together for the synthesis of polypeptide chain. In E. coli ribosome is a 70S particle having two subunits of 30S and 50S. Their association and dissociation depends a upon the concentration of magnesium.

$$\underset{\text{Subunits of ribosome}}{30S} + 50S \underset{\text{Low magneisum conc.}}{\overset{\text{High magnesium conc.}}{\rightleftharpoons}} 70S \text{ ribosome}$$

Small subunit of ribosome contains the decoding centre in which charged tRNAs decode o the codons of mRNA. Large subunit contains peptidyl transferase centre, which forms the peptide bonds between successive amino acids of the newly synthesized peptide chain.

Both 30S and 50S subunits consist of ribosomal RNA (rRNA) and proteins. The mRNA binds to the 16S rRNA of smaller subunit. Near its 5′-end mRNA binds to the 3′-end of 16S rRNA.
The main role of ribosome is the formation of peptide bond between successive amino acids of the newly synthesized polypeptide chain. The ribosome has two channels in it. The linear mRNA enters and escapes through one channel, which has the decoding centre. This channel is accessible to the charged tRNAs. The newly synthesized polypeptide chain escapes through the other channel.

Direction of Translation:

Each protein molecule has an -NH_2 end and -COOH end. Synthesis begins at amino end and ends at carboxyl end. The mRNA is translated in 5 → 3′ direction from amino to carboxyl end. Synthesis of mRNA from DNA transcription also occurs in 5′ → 3′ direction.

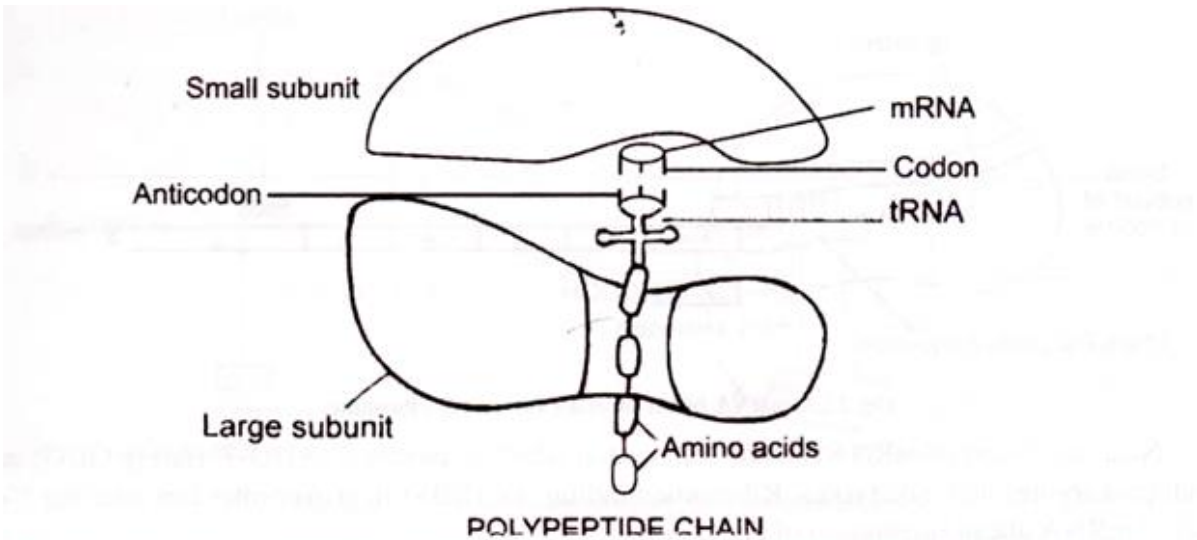


Fig. 12.1. Ribosome showing two subunits and position of mRNA and tRNA. The nascent polypeptide chain passes through a channel.

Steps of Protein Synthesis

Protein biosynthesis involves following major steps:

(i) Activation of Amino Acids:

Amino acid reacts with ATP to form amino acid -AMP complex and pyrophosphate. The reaction is catalyzed by a specific amino acid-activating enzyme called aminoacyl- tRNA synthetase in the presence of Mg^{2+}. There is a separate aminoacyl – tRNA synthetize enzyme for each kind of amino acid. Much of the energy released by the separation of phosphate groups from ATP is trapped in the amino acid — AMP complex.

The complex remains temporarily associated with the enzyme. The amino acid-AMP-enzyme complex is called an activated amino acid (Fig. 7.14). The pyrophosphate is hydrolysed to 2Pi, driving the reaction to the right.

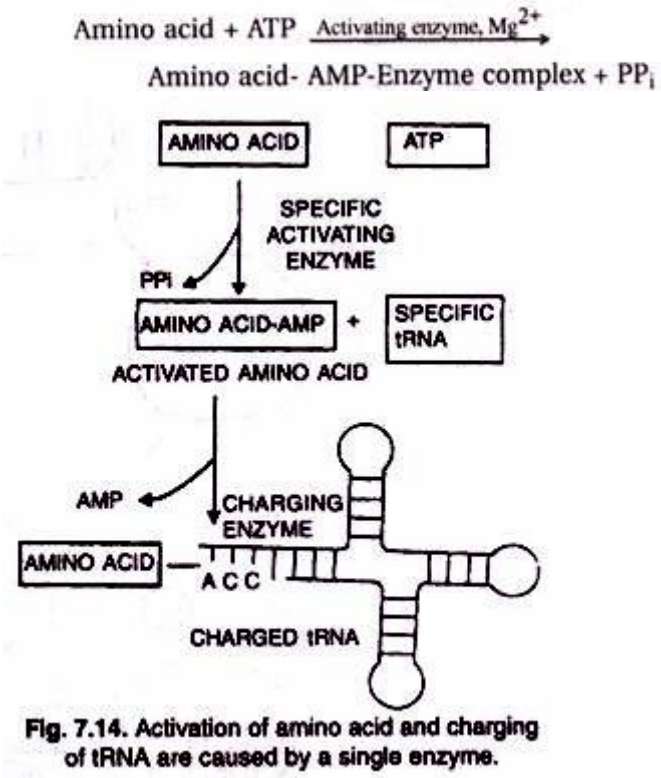
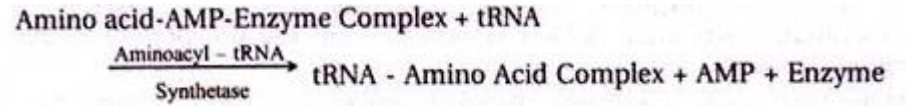


Fig. 7.14. Activation of amino acid and charging of tRNA are caused by a single enzyme.

(ii) Charging of tRNA:

The amino acid-AMP- enzyme complex joins to the amino acid binding site of its specific tRNA, where its -COOH group bonds to – OH group of the terminal base triplet CCA. The reaction is catalyzed by the same aminoacyl-tRNA synthetase enzyme. The resulting tRNA-amino acid complex is called a charged tRNA (Fig. 7.14). AMP and enzyme are freed. The freed enzyme can activate and attach another amino acid molecule to another tRNA molecule. The energy released by change of ATP to AMP is retained in the amino acid-tRNA complex. This energy is later used to drive the formation of peptide bond when amino acids link together on ribosomes.



The tRNA-amino acid complex moves to the site of protein synthesis, the ribosome.

(iii) Activation of Ribosomes:

The small and the large subunits of ribosomes must be joined together for protein synthesis. This is brought about by mRNA chain. The latter joins the small ribosomal subunit by first codon through base pairing with appropriate sequence on rRNA. The combination of the two is called initiation complex (Fig. 7.15). The large subunit later joins the small subunit, forming active ribosome. Activation of ribosome by mRNA requires proper concentration of Mg^{2+} (0.001 Molar conc.)

(iv) Assembly of Amino Acids (Polypeptide Formation):

The events in protein synthesis are better known in bacteria than in eukaryotes. Although these are thought to be similar in the two groups, some differences do occur. The following description refers mainly to protein synthesis in bacteria on the 70S ribosomes. Polypeptide formation involves 3 events: initiation, elongation and termination of amino acid chain.

(a) Initiation of Polypeptide Chain:

The mRNA chain has at its 5′ end an "initiator" or "start" codon (AUG) that signals the start of polypeptide formation. This codon lies close to the P site of the ribosome. The amino acid formyl-methionine (methionine in eukaryotes) initiates the process. It is carried by tRNA having UAC anticodon which bonds to AUG initiator codon of mRNA by hydrogen bonds.

Initiation factors (IF 1, IF 2 and IF 3) and GTP promote the initiation process. The large ribosomal subunit now joins the small subunit to complete the ribosome. At this stage, GTP is hydrolyzed to GDP. The ribosome has formylmethionine-bearing tRNA (tRNA f^{Met}) at the P site (Fig. 7.15). Later, the formylmethionine is changed to normal methionine by the enzyme deformylase. If not required, methionine is later separated from the polypeptide chain by a proteolytic enzyme amino peptidase. Initiation factors are used again to start new chains. As already established, translation of the codons of mRNA takes place in the 5′ – 3′ direction, thus P site and A site on the ribosomes recognize the polarity of the mRNA chain.

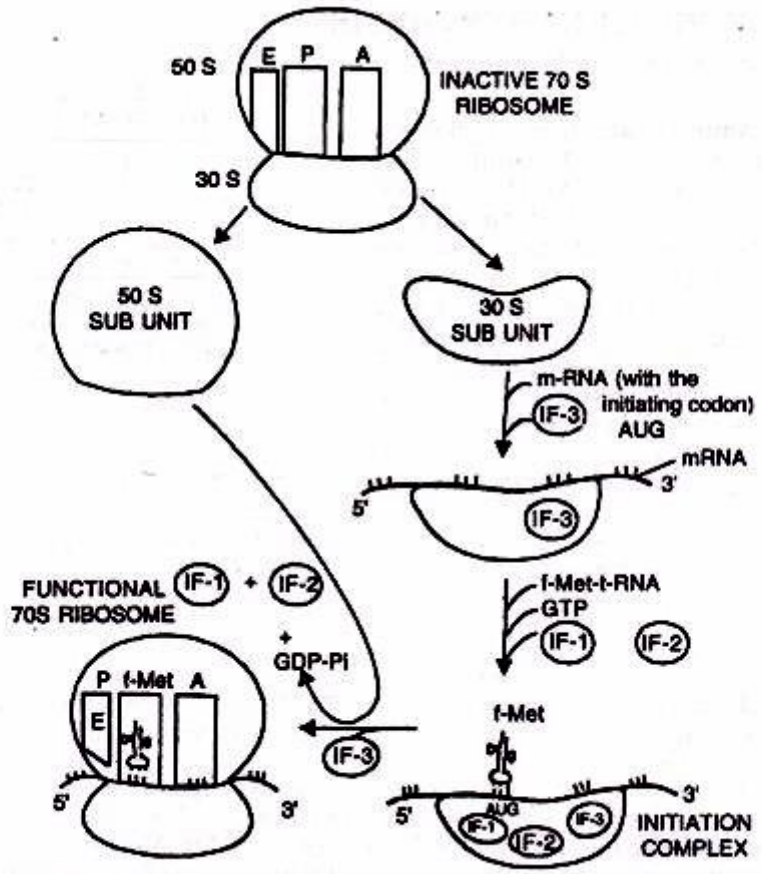


Fig. 7.15. Formation of initiation complex and assembly of the functional 70 S ribosome.

(b) Elongation Phase:

At this point fmet- $tRNA_f^{met}$ molecule in the 70S initiation complex occupies the P site on the ribosome. The other site for a tRNA molecule, i.e. the A site, is empty. The fmet-$tRNA_f^{met}$ is positioned in such a way that its anticodon pairs with the initiating AUG (or GUG) codon on mRNA. The reading frame is specified by this interaction and by pairing of the adjoining purine-rich sequence to a pyrimidine-rich sequence in 16S rRNA. The elongation cycle in the protein synthesis begins with the insertion of an aminoacyl tRNA into the empty A site on the ribosome. The species of tRNA to be inserted depends upon the mRNA codon that is present in the A site. The complementary aminoacyl tRNA is transferred to the A site by a non-ribosomal specific cytoplasmic protein, called the elongation factor T (EF-T) that binds to the aminoacyl tRNA. The factor EF-T contains two subunits, EF-Ts and EF-Tu. EF-Tu like IF2 contains a bound guanyl nucleotide and cycles between a GTP and a GDP. If the codon matches the anticodon, GTP is hydrolysed, positioning the aminoacyl tRNA in the A site and GDP bound with EF-Tu dissociates from the ribosome. A second elongation factor EF-Ts joins the EF-Tu complex and GDP is displaced from the complex forming a EF-Tu-Ts complex.

Finally, GTP binds to the EF-Tu- EF-Ts complex, releasing EF-Ts. EF-Tu containing bound GTP is ready to pick up another aminoacyl tRNA and deliver to the A site of the ribosome. This GTP-GDP cycle keeps repeating. It should be noted that EF-Tu does not recognise the fmet-tRNA initiator, hence the initiator tRNA is not delivered to the A site. On the contrary before fmet-tRNAet, like all other aminoacyl tRNAs, can bind to EF-Tu. This explains why internal AUG codons are not read by initiator tRNA. It has been observed that rapid binding of EF-Tu to an

activated aminoacyl-tRNA prevents hydrolysis, but after the formation of H'-Tu-GTP-tRNA complex, a time lag of several milliseconds allows the codon-anticodon mismatches to diffuse away (before OTP hydrolysis).

Peptide Bond Formation and Translocation:

Once the initiator fmet-tRNA occupies the P site and the next aminoacyl-tRNA occupies the A site, a peptide bond between the adjacent amino acids is formed by an enzyme, peptidyl transferase belonging to the 50S subunit. The active site of the peptidyl transferase is the 23 S rRNA. The uncharged $tRNA_f^{met}$ occupies the P site and the dipeptide formed is attached to the second tRNA occupying the A site following the formation of a peptide bond. The product of the first peptide bond formation is called dipeptidyl-tRNA bound to the A site.

The next step of the elongation cycle is translocation, which requires a third elongation factor EF-G (also called translocase) causing hydrolysis of GTP.

Three important movements occur:

(1) The fmet-tRNA which is now uncharged leaves the P site,

(2) The second tRNA with bound dipeptide is moved to the P site, and

(3) mRNA moves a distance of three nucleotides.

After translocation, the A site is opened up to accept the incoming aminoacyl-tRNA to match the next codon, now positioned at the A site for the next round of elongation (Fig. 7.16). The factor EF-Tu delivers the next aminoacyl-tRNA for the empty A site.

The accuracy of protein synthesis depends on having the correct aminoacyl-tRNA in the A site when the peptide bond is formed, hence the incoming aminoacyl-tRNA is meticulously scrutinized so that its anticodon is complementary and matches the codon at the A site. A mismatch aminoacyl-tRNA may bind with two or three nucleotides of a codon only temporarily, but will leave the A site before a peptide bond is formed. It takes a few milliseconds for the ribosome to decide if the incoming aminoacyl-tRNA is the correct one or not and the time lag is determined by GTPase site of EF-Tu. A peptide bond cannot be formed until EF-Tu is released from the aminoacyl-tRNA and the process requires hydrolysis of GTP to GDP and Pi.

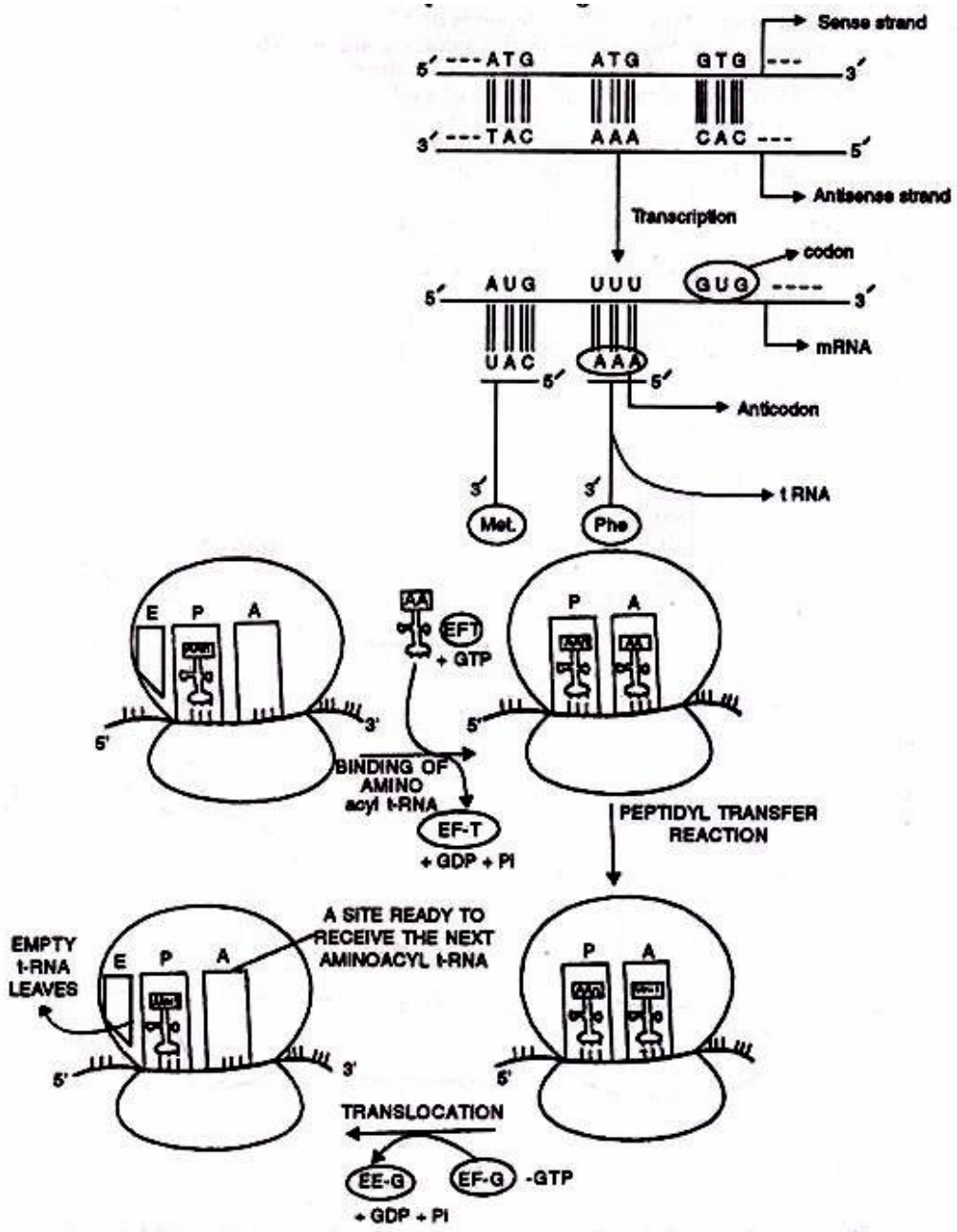


Fig. 7.16. The elongation process of peptide chain. AA n+1 denotes a formyl methionyl peptide.

Termination Phase:

Two conditions are necessary for termination of protein synthesis. One is the presence of a stop codon that signals the chain elongation to terminate, and the other is the presence of release factors (RF) which recognise the chain terminating signal. There are three terminating codons, UAA, UGA and UAG for which tRNAs do not exist. Termination of polypeptide chain is signalled by one of these codons in the mRNA. Behind all this complexity is the fact that after the polypeptide chain has reached its full length, its carboxyl end is still bound to its tRNA adapter. Termination must, therefore, involve the splitting of the terminal tRNA. Release of the peptidyl tRNA from the ribosome is promoted by three specific release factors, RF1, RF2 and RF3. RF1 recognises triplets UAA and UAG, while RF2 recognises UAA and UGA. The third factor RF3 does not possess any release activity of its own, but it binds to OTP and stimulates

the binding of RF1 and RF2 with the ribosome. In E. coli, 16S rRNA is essential in reading the stop codon. The release factors bind to stop codon to cause a shift of the polypeptidyl-tRNA from A to P site (Fig. 7.17). Whether OTP hydrolysis is required for chain termination is not yet firmly established, although the RF3, which appears to enhance RF1 and RF2 binding with ribosome, does not bind to OTP. The ester bond between the polypeptide chain and the last tRNA is then hydrolysed. Binding of RF to the terminating codon causes water to act as the acceptor of the growing peptide and not another amino acid on a tRNA Release of the polypeptide chain is followed by dissociation of mRNA and tRNA. Subsequently dissociation of 30S and SOS ribosome subunits takes place with concomitant binding of IF3 to 30S subunit to prevent reassembly in the absence of mRNA and fmet-tRNA.

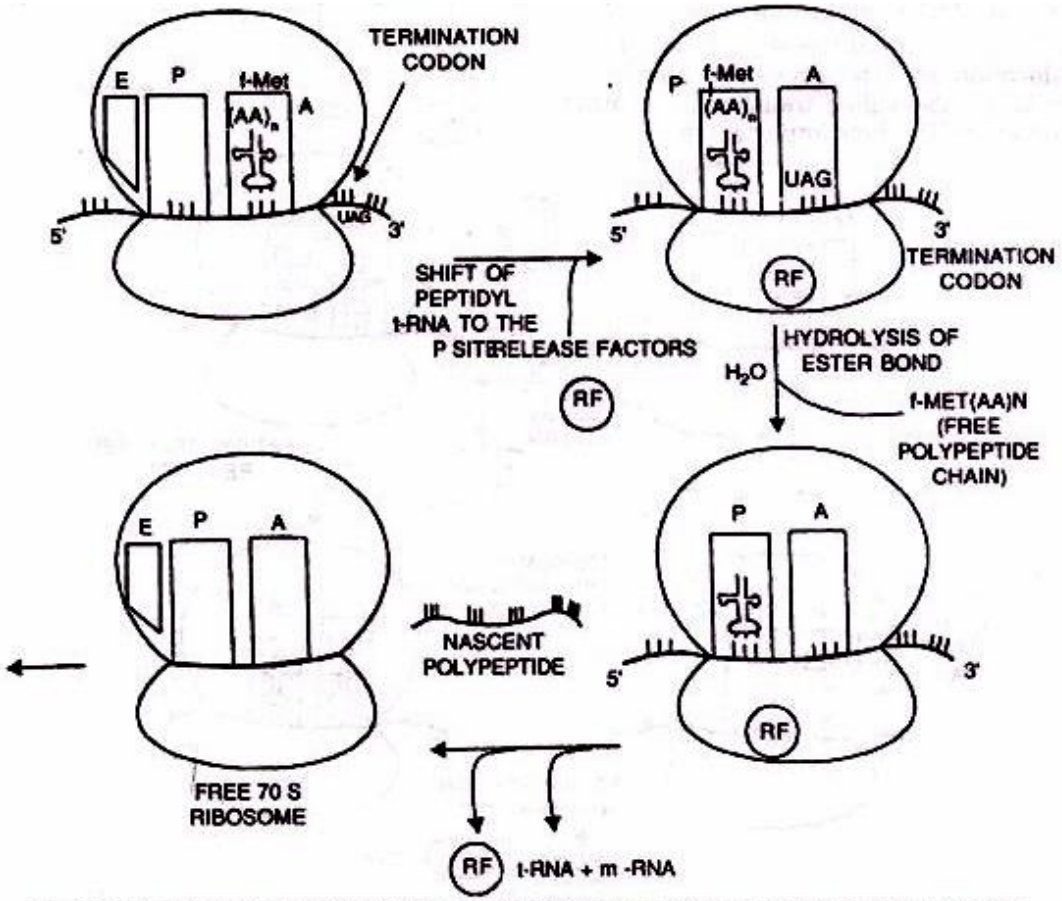


Fig. 7.17. Termination of protein synthesis. Steps showing the release of completed chain. The termination codon (UAG) at the A site alongwith the release factors helps in the termination process.

Polyribosome or Polysome:

A single mRNA molecule can be read simultaneously by several ribosomes. A polyribosome or polysome consists of several ribosomes attached to the same RNA. The number of ribosomes in a polysome depends upon the length of mRNA.

A fully active mRNA has one ribosome after every 80 nucleotides. There may be about 50 ribosomes in a polycistronic mRNA of prokaryotes. Ribosomes move along mRNA in 5′ 3′ direction. There is a gradual increase in the size of polypeptide chain as the ribosomes move along mRNA towards its 3′-end. Polypeptide chain starts near the 5′-end and is completed near the 3′-end.

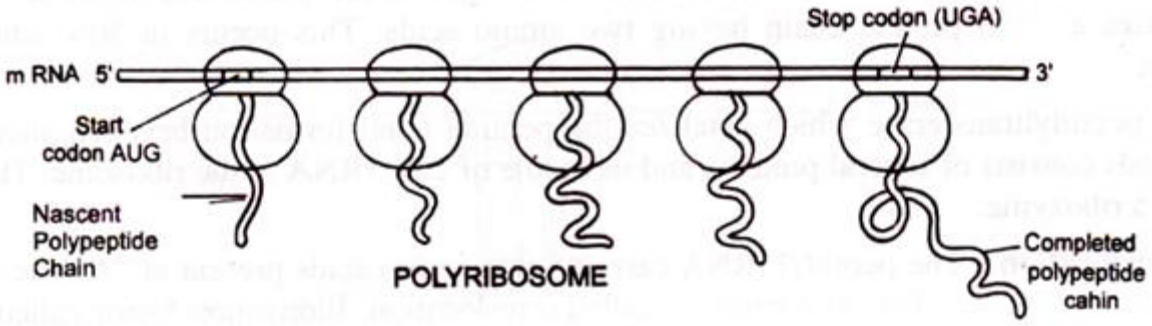


Fig. 12.9. Polyribosome

The ribosomes closest to the 5′-end of mRNA have the smallest polypeptide chain, while ribosomes nearest to the 3′-end have longest chain. Polysome increases the rate of protein synthesis tremendously. In bacteria protein is synthesized at the rate of about 20 amino acids per second.

Simultaneous Transcription and Translation in Prokaryotes:

In prokaryotes, all components of transcription and translation are present in the same compartment. The mRNA molecule is synthesized in 5′ → 3′ direction and protein synthesis also occurs in 5′ → 3′ direction. In this way mRNA molecule while still under synthesis has a free 5′-end whose other end is still under synthesis.

Ribosomes bind at free 5′-end and start protein synthesis. In this way the free end (5′-end) of mRNA starts the process of protein synthesis while still attached to DNA. This is called Coupled Transcription and Translation. This increases the speed of protein synthesis. After the protein synthesis is completed, the degradation of mRNA molecule by nucleases also starts at 5′-end and proceeds in 5′ → 3′ direction.

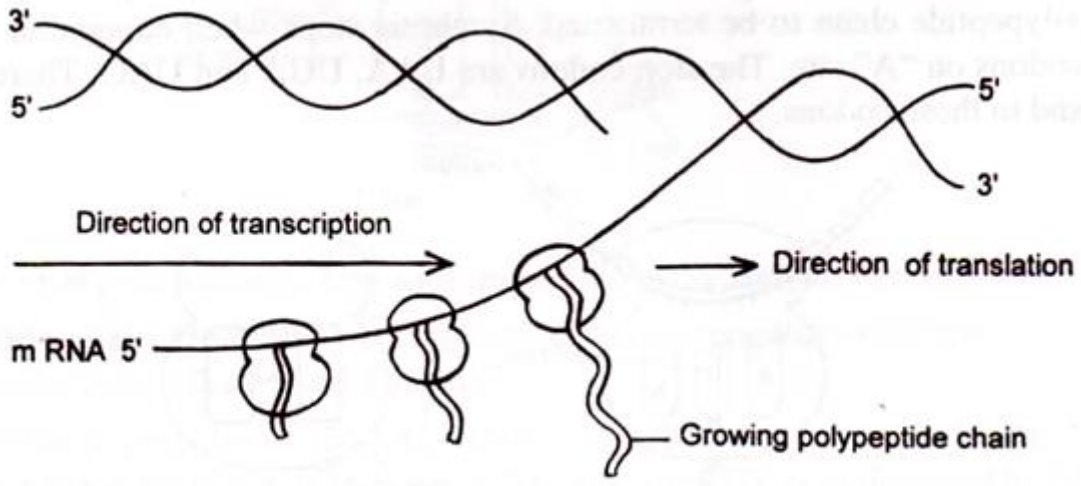


Fig. 12.10. Coupled transcription and translation.

Protein Synthesis in Eukaryotes:

Protein synthesis in eukaryotes is basically similar to that of prokaryotes except some differences. The ribosomes in eukaryotes are of 80S having 40S and 60S subunits. In eukaryotes the initiating amino acid is methionine and not f-methionine as in the case of prokaryotes. A special tRNA binds methionine to start codon AUG. This tRNA is called $tRNAi^{Met}$. This is distinct from $tRNA^{Met}$ which binds amino acid methionine to any other internal position in the polypeptide.

There is no Shine-Dalgarno sequence in eukaryotic mRNA to function as ribosome binding site. Between 5′-end and AUG codon of mRNA there is a sequence of bases called cap. Small subunit of ribosome scans the mRNA in 5′ → 3′ direction until it comes across 5′- AUG-3′ codon. This process is called scanning. Initiation factors also closely associated with 3′-end of mRNA through its poly-A tail. Initiation factors circularize mRNA by its poly-A tail. In this way poly-A tail also contributes to the translation of mRNA. Eukaryotic mRNAs are monocistsonic and encode a single polypeptide, therefore have a single open reading frame.

There are ten initiation factors in eukaryotes. They are elF (eukaryotic intiation factors) are elFI, eIF2, eIF3, eIF4A, eIF4B, eIF4C, eIF4D, eIF4F, eIF5, eIF6. There are two elongation factors in eukaryotes like prokaryotes. They are eEFl (similar to EF-Tu) and eEF2 (similar to EF-G). Eukaryotes have only one release factor eRF which requires GTP termination of protein synthesis. It recognizes all the three stop codons.

In eukaryotes the mRNA is synthesized in the nucleus, then processed, modified and passed on into the cytoplasm through nucleopores. The protein synthesis takes place in the cytoplasm. The mRNA in prokaryotes is very unstable and its life span is of a few minutes only. The mRNA of eukaryotes is quite stable and has a longer life span extending upto several days.

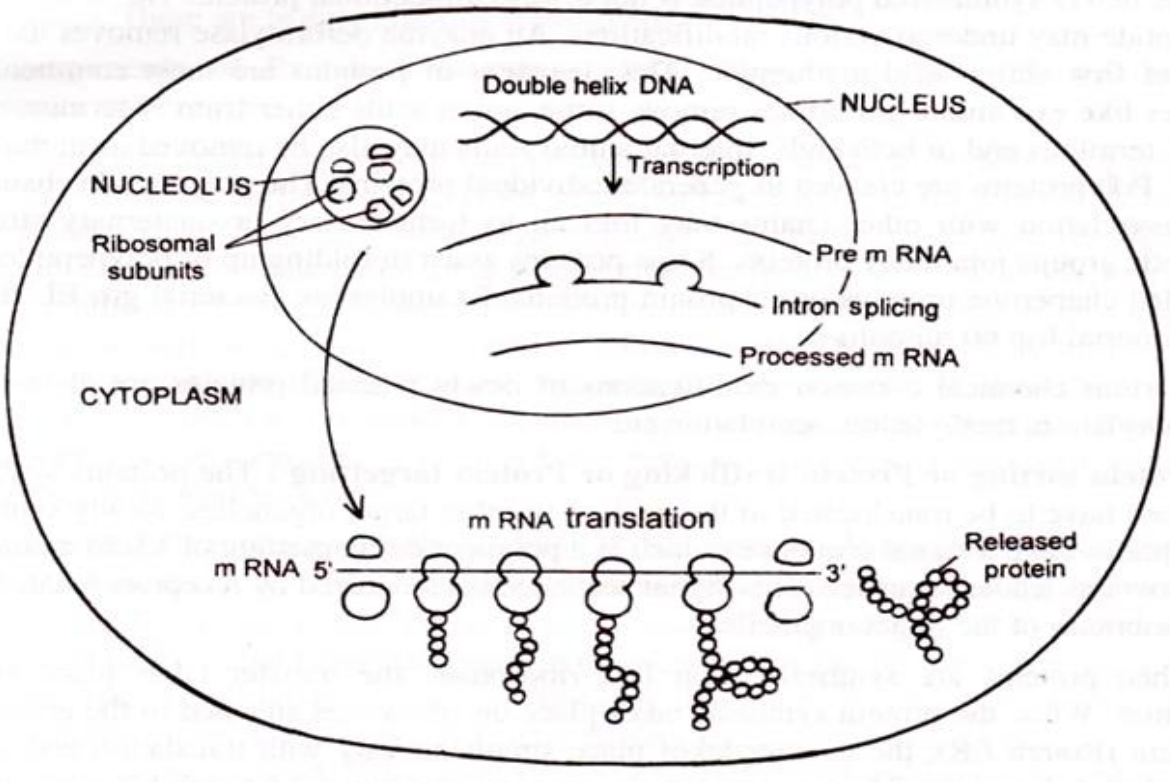


Fig. 12.11. Transcription, mRNA processing and mRNA translation.

Protein Synthesis on Bound Ribosomes:

Ribosomes occur in free state in the cytoplasm as well as bound to the outer surface of endoplasmic reticulum called rough endoplasmic reticulum (RER). The attachment of ribosomes to ER occurs after the protein synthesis starts. Whether the ribosomes synthesize protein on free or attached state depends upon the type of proteins to be synthesized by ribosomes. Most of the proteins which remain in free state in the cytoplasm are synthesized by free ribosomes.

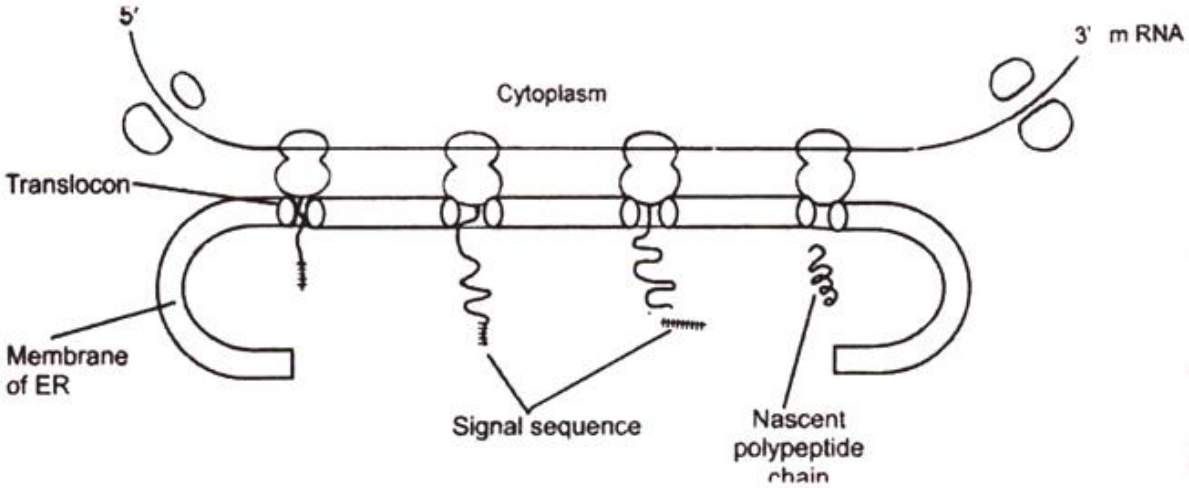


Fig. 12.12. Protein synthesis on ribosome bound on ER (RER).

Proteins synthesized by ribosomes on ER enter into the lumen of cisternae of ER from where they may enter into golgi apparatus where they are glycosylated and form secretary granules and many of them enter lysosomes.

Modification of Folding of Released Polypeptides:

DNA molecule specifies only the primary structure while folding and other modifications controlled by proteins themselves. The newly synthesized polypeptide is not always a functional protein. The newly released polypeptide may undergo various modifications. An enzyme deformylase removes the formyl group of first amino acid methionine. The cleavages of proteins are most common. Some enzymes like exo-amino-peptidases remove some amino acids either from N-terminus end or from C-terminus end or both ends. Internal amino acids may also be removed as in the case of insulin. Polyproteins are cleaved to generate individual proteins. The polypeptide chain singly or in association with other chains may fold up to form tertiary or quaternary structures. Prosthetic groups join many proteins. Some proteins assist in folding up of polypeptides. They are called chaperone proteins or chapronin proteins. Examples are Bacterial gro EL (E. coli), mitochondrial hsp60 mitonin. Various chemical common modifications of newly released proteins are glycosylation, phosphorylation, methylation, acetylation etc.

Protein Sorting or Protein Trafficking or Protein Targeting:

The proteins synthesized in the cell have to be translocated to the nucleus or other target organelles. Newly synthesized polypeptides have a signal sequence (which is a polypeptide) consisting of 13-36 amino acids. It is known as leader sequence. This signal sequence is recognized by receptors located within the membranes of the target organelles.

When proteins are synthesized on free ribosomes, the transfer takes place after the translation. When the protein synthesis takes place on ribosomes attached to the endoplasmic reticulum (Rough ER), the transfer takes place simultaneously with translation and is called co-translational transfer. The proteins which enter into the lumen of rough ER may enter into golgi apparatus, from where they may enter secretary lysosomes. The signal sequence is degraded by protease enzymes. Once all these proteins are assembled into their proper place, they provide the proper biochemical machinery, which keeps the cell feeding, locomoting, multiplying and alive.

Aminoacyl tRNA synthetases:

An aminoacyl-tRNA synthetase (aaRS or ARS), also called tRNA-ligase, is an enzyme that attaches the appropriate amino acid onto its tRNA. It does so by catalyzing the esterification of a specific cognate amino acid or its precursor to one of all its compatible cognate tRNAs to form an aminoacyl-tRNA. In humans, the 20 different types of aa-tRNA are made by the 20 different aminoacyl-tRNA synthetases, one for each amino acid of the genetic code.

This is sometimes called "charging" or "loading" the tRNA with the amino acid. Once the tRNA is charged, a ribosome can transfer the amino acid from the tRNA onto a growing peptide, according to the genetic code. Aminoacyl tRNA therefore plays an important role in RNA translation, the expression of genes to create proteins. As genetic efficiency evolved in higher organisms, 13 new domains with no obvious association with the catalytic activity of aaRSs genes have been added

Mechanism:

The synthetase first binds ATP and the corresponding amino acid (or its precursor) to form an aminoacyl-adenylate, releasing inorganic pyrophosphate (PP_i). The adenylate-aaRS complex then binds the appropriate tRNA molecule's D arm, and the amino acid is transferred from the aa-AMP to either the 2'- or the 3'-OH of the last tRNA nucleotide (A76) at the 3'-end.

The mechanism can be summarized in the following reaction series:

1. **Amino Acid + ATP → Aminoacyl-AMP + PP_i**
2. **Aminoacyl-AMP + tRNA → Aminoacyl-tRNA + AMP**

Summing the reactions, the highly exergonic overall reaction is as follows:

- **Amino Acid + tRNA + ATP → Aminoacyl-tRNA + AMP + PP_i**

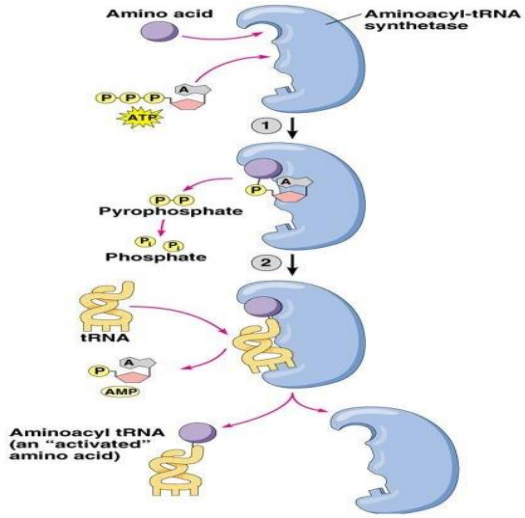


Figure: Steps involves in Amino acyl tRNA synthetase action

Some synthetases also mediate an editing reaction to ensure high fidelity of tRNA charging. If the incorrect tRNA is added (aka. the tRNA is found to be improperly charged), the aminoacyl-tRNA bond is hydrolysed. This can happen when two amino acids have different properties even if they have similar shapes—as is the case with Valine and Threonine.
The accuracy of aminoacyl-tRNA synthetase is so high that it is often paired with the word "superspecificity" when it is compared to other enzymes that are involved in metabolism. Although not all synthetases have a domain with the sole purpose of editing, they make up for it by having specific binding and activation of their affiliated amino acids. Another contribution to the accuracy of these synthetases is the ratio of concentrations of aminoacyl-tRNA synthetase and its cognate tRNA. Since tRNA synthetase improperly acylates the tRNA when the synthetase is overproduced, a limit must exist on the levels of aaRSs and tRNAs in vivo.

Classes of aminoacyl tRNA synthetase:

There are two classes of aminoacyl tRNA synthetase, each composed of ten enzymes:[3][4]

- Class I has two highly conserved sequence motifs. It aminoacylates at the 2'-OH of a terminal adenosine nucleotide on tRNA, and it is usually monomeric or dimeric (one or two subunits, respectively).
- Class II has three highly conserved sequence motifs. It aminoacylates at the 3'-OH of a terminal adenosine on tRNA, and is usually dimeric or tetrameric (two or four subunits, respectively). Although phenylalanine-tRNA synthetase is class II, it aminoacylates at the 2'-OH.

The amino acids are attached to the hydroxyl (-OH) group of the adenosine via the carboxyl (-COOH) group. Regardless of where the aminoacyl is initially attached to the nucleotide, the 2'-*O*-aminoacyl-tRNA will ultimately migrate to the 3' position via transesterification.

Structures of aminoacyl tRNA synthetase:

Both classes of aminoacyl-tRNA synthetases are multidomain proteins. In a typical scenario, an aaRS consists of a catalytic domain (where both the above reactions take place) and an anticodon binding domain (which interacts mostly with the anticodon region of the tRNA and ensures binding of the correct tRNA to the amino acid). In addition, some aaRSs have additional RNA binding domains and editing domains[5] that cleave incorrectly paired aminoacyl-tRNA molecules.

The catalytic domains of all the aaRSs of a given class are found to be homologous to one another, whereas class I and class II aaRSs are unrelated to one another. The class I aaRSs have the ubiquitous Rossmann fold and have the parallel beta-strands architecture, whereas the class II aaRSs have a unique fold made up of antiparallel beta-strands. The alpha helical anticodon binding domain of Arginyl, Glycyl and Cysteinyl-tRNA synthetases is known as the DALR domain after characteristic conserved amino acids.

Aminoacyl-tRNA synthetases have been kinetically studied, showing that Mg2+ ions play an active catalytic role and therefore aaRs have a degree of magnesium dependence. Increasing the Mg2+ concentration leads to an increase in the equilibrium constants for the aminoacyl-tRNA synthetases' reactions. Although this trend was seen in both class I and class II synthetases, the magnesium dependence for the two classes are very distinct. Class II synthetases have two or three (more frequently three) Mg2+ ions, while class I only requires one Mg2+ ion. Beside their lack of overall sequence and structure similarity, class I and class II synthetases feature different ATP recognition mechanisms. While class I binds via interactions mediated by backbone hydrogen bonds, class II uses a pair of arginine residues to establish salt bridges to its ATP ligand. This oppositional implementation is manifested in two structural motifs, the Backbone Brackets and Arginine Tweezers, which are observable in all class I and class II structures, respectively. The high structural conservation of these motifs suggest that they must have been present since ancient times.

Accuracy during aminoacylation of tRNA:

The aminoacyl-tRNA synthetases carry out two important functions in protein synthesis: information transfer and chemical activation. The information transfer involves matching amino acids (AA) with cognate tRNA according to the rules of the genetic code. The chemical activation involves formation of a high energy ester bond between the carboxyl group of an AA and a hydroxyl of the 3′-terminal adenosine of tRNA, with an aminoacyl adenylate as intermediate.

$$\mathbf{AARS + AA + ATP = AARS\text{-}AA\text{-}AMP + PPi}$$

$$\mathbf{AARS\text{-}AA\text{-}AMP + tRNAAA = AARS + AA{-}tRNAAA + AMP}$$

Aminoacyl adenylate formation is the least accurate step in the tRNA aminoacylation pathway. Some AA [e.g., Met vs. homocysteine (Hcy); Ile vs. Val and Hcy; Leu vs. Hcy; Val vs. Cys and Thr; Ala vs. Gly; Lys vs. ornithine (Orn); or Thr vs. Ser] are so similar that AARSs misactivate them at frequencies exceeding the frequency of translational errors, forming AARS-bound

noncognate aminoacyl adenylates. Noncognate adenylates are directly or indirectly destroyed by the editing function of an AARS. Editing can occur by two alternative pathways: pretransfer, by hydrolysis of the noncognate aminoacyl adenylates or post-transfer, by the hydrolysis of the mischarged tRNA. Because of the fast dissociation of aminoacyl-tRNA from AARS, post-transfer editing can contribute only a factor of ~2 to selectivity. This may explain the more widespread use of pretransfer editing pathways by AARSs. Overall, editing improves the AA selectivity of an AARS by a factor > 100. Consequently, nonprotein AAs Hcy or Orn are not transferred to tRNA. Some noncognate AAs are transferred to tRNA with low efficiency. For instance, IleRS promotes one misacylation of $tRNA^{Ile}$ with Val per 350,000 correct acylations with Ile. ValRS promotes one misacylation of $tRNA^{Val}$ with Ile and Thr per 5,000 and 350,000 correct acylations with Val, respectively. LysRS catalyzes one misacylation of $tRNA^{Lys}$ with Arg, Thr, Met, Leu, Ala, Cys or Ser per 1,600, 16,000, 32,000, 132,000, 265,000, 560,000 or 750,000 correct acylations with Lys, respectively. Some AARS, such as TyrRS, CysRS, ArgRS, AspRS and SerRS bind the cognate AA so much more tightly than their competitors that they do not need to edit. Overall, the accuracy of tRNA aminoacylation is greater than the accuracy of subsequent steps of translation on ribosomes.

Energetics of amino acid polymerization:

Protein synthesis itself has a higher cost than previously estimated. While transcription of the amine acid mRNA codon requires six ATP per amine acid and activation to amine acyl-tRNA requires another two ATP, there is disagreement about the total energy cost of mRNA translation. In addition to the one ATP per peptide required for capping the 5-prime end of the peptide, at least one GTP per peptide bond is required for initiation, two GTP per bond for elongation, and one GTP per peptide for termination. However, recent evidence suggests that one additional molecule of GTP is required for chain elongation (Schimmel, 1993), and hydrolysis of an additional GTP might be required during initiation and/or at the termination step as well. If hydrolysis of these additional high-energy bonds is proven correct, the net energy cost of protein synthesis alone will increase significantly from estimates made a decade ago.

In addition, there are a variety of other costs that are difficult to estimate. All of the additional sequences that are involved in, for example, synthesizing ''pre-proteins" and "pre-pro-proteins" and the costs of alternate splicing are not easy to quantify. In a sense, synthesizing and then removing these unused peptide sequences is wasted energy unless some as-yet-unknown energy advantage is discovered for this process. Similarly, the cost of synthesizing nonessential amino acids that are required for protein synthesis and the costs of posttranslational modifications are not known with certainty.

Further, the folding (Hartl, 1996) and the movement of the synthesized proteins to their sites of action (Rothman and Wieland, 1996) are highly energy-dependent processes. ATP-dependent mechanisms are required for polypeptide chain folding by heat-shock protein 70 and the chaperonin families of molecular chaperones (Hartl, 1996). Translocation across the membrane of the rough endoplasmic reticulum is an energy-dependent process, as is each transport step to the cis, medial, and trans Golgi compartments (Rothman and Wieland, 1996).

n addressing the potential energy costs of regulatory proteins, one is struck by the fact that, although these proteins do not represent much in the way of mass compared with structural proteins, their turnover rates are very high. Therefore, they may represent a significant energy drain. It is now known, for example, that an immense number of processes are controlled by reversible enzymatic phosphorylation/dephosphorylation reactions. This central mechanism of regulatory control occurs in every cell at an unimaginable number of times per minute. What the actual net energy cost of these regulatory events is to the whole body is only speculative, but likely significant.

Signal transduction processes, including those mediated by the more than 100 known members of the protein kinase family, and the energy-dependent costs of the second messenger families, for example the phosphatidyl and inositol kinases, are additional energy costs of regulatory protein metabolism. Furthermore, the energy costs of posttranslational modifications such as protein glycosylation are additional. The addition of each nucleotide sugar costs the hydrolysis of one uridine triphosphate bond. Approximately 10 percent of proteins are glycosylated. Each protein averages between 2 and 5 sugar chains, and each sugar chain averages about 12 sugars. Precisely how to tally this cost in the overall sum of the total daily energy cost of protein turnover, is not known, but the cost is potentially large.

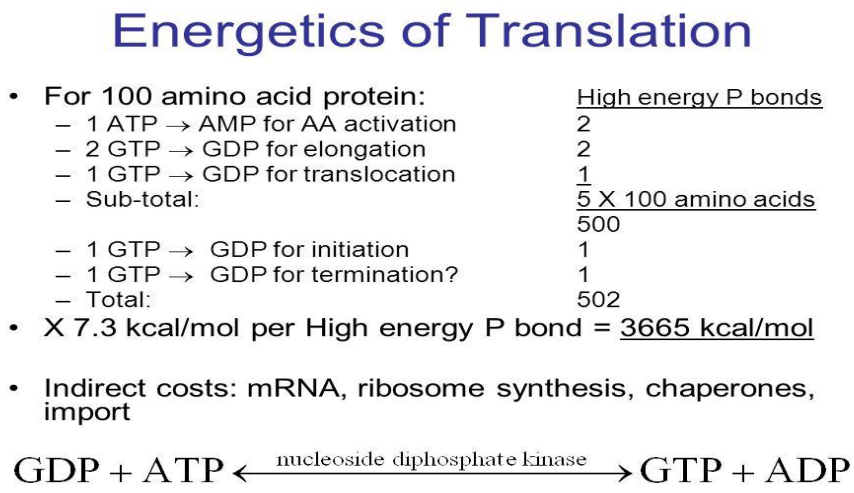


Figure: Energetics of Translation

Inhibitors of Protein Synthesis:

There are many chemicals, both synthetic as well as those obtained from different sources like fungi, which bind to the components of translation machinery and arrest the translation process. Most of them are antibacterial agents or antibiotics that act exclusively on bacteria and are thus powerful tools in the hands of man to combat various infectious diseases. Most of antibiotics are inhibitors of translation machinery.

a. Puramycin: It binds at "A" site on ribosome. This causes pre-mature termination of polypeptide chain.

b. Kirromycin: It inhibits the elongation factor EF-Tu.

c. Fusidic acid: It inhibits the elongation factor EF-G.

d. Tetracycline: It attacks "A" site on ribosome and prevents the binding of aminoacyl- tRNA.

e. Chloramphenicol: It blocks the peptidyl transfer reaction

f. Erythromycin: It binds the polypeptide exit channel of ribosome, therefore blocks the exit of growing polypeptide chain, thus stops the translation process.

g. Streptomycin and Neomycin: These inhibit the binding of $tRNA^{fMet}$ to the "P" site.

Inhibitors in Eukaryotes:

Diphtheria toxin is a toxin produced by *Corynebacterium diphtheriae*. This causes modification of eukaryotic elongation factor.

Probable Questions:

1. What are the main components of translation?
2. Write down the structure of tRNA.
3. Discuss the role of ribosome in protein translation.
4. Describe the initiation process of translation.
5. How termination occurs in protein translation
6. Describe the elongation phase of translation.
7. What is polysome complex?
8. How amino acyl tRNA works?
9. describe different inhibitors of protein translation.
10. describe different classes of aminoacyl tRNA synthetase.
11. Describe energetics of protein translation.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwa

UNIT-V

Basic biology of cloning vectors: plasmids, phages, single stranded DNA vectors, high capacity vectors, retroviral vectors, expression vectors and other advanced vectors in use. Gene cloning strategies: methods of transforming E. coli with rDNA; methods of selection and screening of transformed cells; construction of genomic and cDNA libraries; phage display

Objective: In this unit we will discuss about different types of Restriction endonuclease with their property and mechanism of actions. Also we will discuss about restriction mapping, DNA fingerprinting and DNA foot printing methods. We will also discuss about PCR and methyl interference assay.

Restriction Endonuclease:

Restriction endonuclease enzymes occur naturally in bacteria as a chemical weapon against the invading viruses. They cut both strands of DNA when certain foreign nucleotides are introduced in the cell. Endonucleases break strands of DNA at internal positions in random manner.

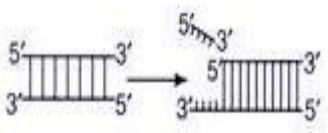
These are also known as molecular scissors, used for cutting of DNA. The cutting of DNA at specific locations became possible with the discovery of molecular scissors, i.e., restriction enzymes. In the year 1963, the two enzymes responsible for restricting the growth of bacteriophage in E. coli were isolated. One of these added methyl groups to DNA, while the other cuts the DNA. Later was termed as restriction endonucleases.

The first restriction endonuclease was isolated by Smith Wilcox and Kelley in 1968 was Hind II. It was found that it always cuts DNA molecules at a particular point by recognising a specific sequence of six base pairs known as recognition sequence for Hind II. Today, more than 900 restriction enzymes have been isolated from over 230 bacterial strains each of which recognise different recognition sequences.

Differences between Exonucleases and Endonucleases

Exonucleases:

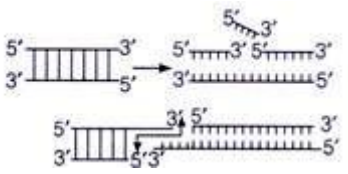
1. These nucleases cleave base pairs of DNA at their terminal ends
2. They act on single – strand of DNA or gaps in double –stranded DNA. They do not cut RNA



Endonucleases:

1. They cleave DNA at any Pont except the terminal ends

2. They cleave one strand (figure below) or both strands (figure below) of double – stranded DNA. They may cut RNA



Each restriction endonuclease recognizes a specific palindromic nucleotide sequences in the DNA. Palindrome in the DNA is a groups of letters that forms the same words when read both forward and backward. For example, the following sequences read the same on the two strands in 5′ → 3′ direction as well as 3′ → 5′ direction.

5′ —— GAATTC —— 3′
3′ —— CTTAAG —— 5′

Types of Restriction Enzymes:

1. Restriction enzyme Type I:

These enzymes interact with an unmodified recognition sequence in double-stranded DNA and then attach to long DNA molecule. After travelling for distance between 1000 to 5000 nucleotides the enzymes cleaves only one strand of the DNA at an apparently random site, and creates a gap of about 75 nucleotides in length.

Acid soluble oligonucleotides removed from the gap are released. A second enzyme molecule is needed to cleave the remaining strand of DNA. The cofactors for the enzyme are Mg^{2+} ions, ATP and S-adenosyl- methionine. This kind of enzyme is not useful for genetic engineering, because its cleavage sites are non-specific.

Type I restriction enzyme can simultaneously hold two different sites on DNA creating a loop in nucleic acid. This enzyme consists of three types of subunits. The Eco K enzyme, for example, has the structure R_2M_2S. The R subunit is responsible for restriction and the M subunit for methylation. The binding of enzyme to DNA may be succeeded by either restriction or modification and this property is characterized by S subunit.

2. Restriction enzyme Type II:

These enzymes recognise a particular target sequence in a double-stranded DNA molecule. They cleave the polynucleotide chain within or near that sequence to give rise to distinct DNA fragments of defined length and sequence. They require Mg^{2+} ions for the action (i.e., restriction). Type II enzymes are used for gene manipulation studies.

3. Restriction enzyme Type III:

These enzymes cleave double-stranded DNA at well-defined sites. They require ATP, Mg^{2+} ions and have very partial requirement for S-adenosyl-methionine for restriction. They have intermediate properties between Type I and Type II REs.

Naming of Restriction Endonuclease Enzymes:

About 350 types of restriction endonucleases have been isolated from more than 200 bacterial strains. Large number of these enzymes require a system of uniform nomenclature. A system based on the proposals of Smith and Nathans (1973) has been followed for the most part.

Table 55.2. Characteristics of restriction endonuclease enzymes.

Characteristics	Type I	Type II	Type III
1. Restriction and modification activities	Single multifunctional enzyme	Separate endonuclease and methylase	Separate enzymes with a subunit is common
2. Protein structure of enzyme	Three different subunits	Simple	Two different subunits
3. Requirements for restriction	ATP, Mg^{2+} S–adenosyl-Mg^{2+} methionine		ATP, Mg^{2+} (S-adenosyl methionine)
4. Sequence of host specificity sites	*Eco* B: TGAN*$_8$TGGT *Eco* K: AA N*$_6$ GTGC	Usually rotational symmetry	*Eco* P1: AGACC *Eco* P15: CAGCAG
5. Cleavage sites	Possibly random at least 1000 bp from host specificity site	At or near host specificity site	24-26 bp to 3' of host specificity site
6. Enzymatic turnover	No	Yes	Yes
7. DNA translocation	Yes	No	No
8. Site of methylation	Host specificity site	Host specificity site	Host specificity site

Naming exercise of RE enzymes is based on following rules:

1. Each RE enzyme is named by a three-letter code.
2. The first letter of this code is derived from the first epithet (first letter of name) of the genus name. It is printed in italics.
3. The second and third letters are from the first two letters of its species name. They are also printed in italics.
4. This is followed by the strain number. If a particular strain has more than one restriction enzyme, these will be identified by Roman numerals as I, II, III, etc.

For example, the enzyme Eco RI was isolated from the bacterium Escherichia (E) coli (co) strain RY13 (R) and it was the first endonuclease (I). R also indicates antibiotic resistant plasmid of the bacterium. Likewise, Hind II from Haemophilus influenzae strain Rd and Bgl I from Bacillus *globigii*. A few restriction endonuclease enzymes and their sources are given in Table 55.3.

Table 55.3. A few restriction endonuclease enzymes, their sources and recognition sites.

Name of the restriction endonuclease enzyme	Source (Microorganisms)	Recognition sequence and cleavage site
1. *Aat* II	*Acetobacter aceti*	GACGT ↓ C
2. *Bcl* I	*Bacillus Caldoyticus*	T ↓ GATCA
3. *Cvn* I	*Chromatium vinosum*	CC ↓ TNAGG
4. *Eco* RI	*Escherichia coli* RY13	G ↓ AATTC
5. *Eco* RII	*Escherichia coli R*245	↓ CCTGG
6. *Hind* II	*Haemophilus influenzae* Rd	GTP_y ↓ PuAC*
7. *Hind* III	*Haemophilus influenzae* Rd	A ↓ AGCTT
8. *Kpn* I	*Klebsiella pneumoniae* OK	GGTAC ↓ C
9. *Nop* I	*Nocardia opaca*	G ↓ TC GAC
10. *Nsp* B II	*Nostoc*	C(A/C)G ↓ C(T/G)G
11. *Sfa* I	*Streptococcus faecalis*	GG ↓ CC
12. *Sal* PP	*Streptomyces albus*	GTGCA ↓ G
13. *Sal* I	*Streptomyces alubs* G	G ↓ TC GAC
14. *Xor* II	*Xanthomonas oryzae*	CGAT ↓ CG
15. *Xba* I	*Xanthomonas badrii*	T ↓ CTAGA

*Pu = purine Py = pyrimidine

Target Sites of Restriction Endonuclease Enzymes:

A restriction endonuclease enzyme of type H recognises a specific recognition site (base sequence) on the DNA and makes a cut at this site only. These target sites are 4 to 6 nucleotides long (Fig. 55.3). They exhibit palindromic symmetry, i.e., nucleotide pair sequences are same reading forward or backward from a central axis of symmetry, like the nonsense phrase-AND 'MADAM DNA'.

The term palindromic has also been applied to sequences such as:

5′-AGCCGA—

3′-TCGGCT— both of which are palindromic strands.

X-ray crystallography of RE enzyme-DNA complex indicate that endonuclease acts as a dimer of identical subunits and that the palindromic nature of target sequence reflects the two fold rotational symmetry of the dimeric protein.

Nature of Cut Ends:

Two types of cut ends of DNA, namely blunt or flush ends and sticky or cohesive ends, are produced by the restriction endonuclease s. The nature of these cut ends generated by the REs are very important in designing the gene cloning experiments.

1. Blunt cut ends:

In case of the blunt cut end, the enzyme (e.g., Haelll, Smal) makes a simple double-stranded cut in the middle of the recognition sequence. Thus the blunt ends or flush ends are formed. The RE Hae III makes a cut in the 5′-GGCC-3′ target site as shown in Fig. 55.3.

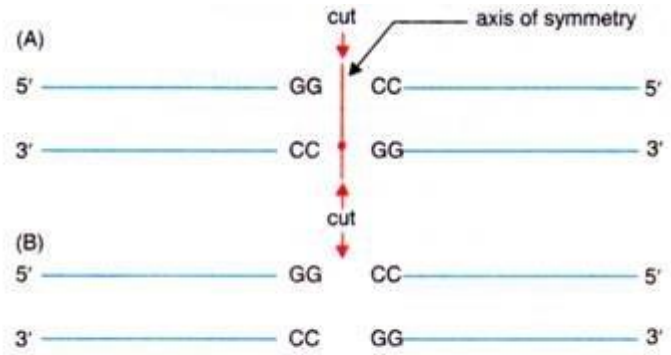


Fig. 55.3. A-The target site for restriction endonuclease *Hae* III. Arrows indicate the sites where the cuts will be affected. B– The target site after the cut. As a result of this the ends of cut DNA are blunt, that is protruding unpaired bases.

The utility of generation of blunt end cuts during the joining of DNA fragments is that any pair of ends may be joined together irrespective of sequence. This is especially useful for those researchers who are interested to join two defined sequences without introducing any additional material between them. Table 55.4 shows certain blunt-end restriction sites.

Table 55.4. Examples of certain blunt end restriction sites and restriction endonucleases.

Enzyme	Recognition site	Cleavage product
1. *Hae* III	–GG ↓ CC–	–GG CC–
2. *Hae* III	–CC ↑ GG–	–CC GG–
3. *Sma* I	–CCC ↓ GGG–	–CCC GGG–
4. *Sma* I	–GGG ↑ CCC–	–GGG CCC–

2. Sticky-or Cohesive ends:

Many restriction enzymes (e.g., Eco RI, Bam HI, and Hind III) make staggered, single-stranded cuts, producing short single-stranded projections at each end of the cleaved DNA, called sticky ends. Since the restriction sites are symmetrical, so that both strands have the same sequence when read in the 5′ to 3′ direction. Thus, such staggered cuts will generate identical single-stranded projections on the either site of the cut (Fig. 55.4).

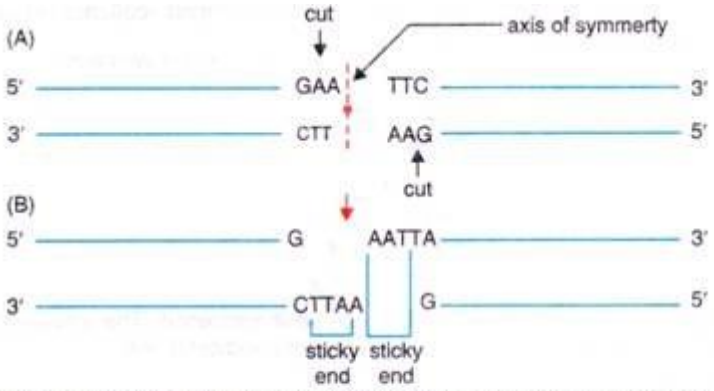


Fig. 55.4. A–Diagram showing the recognition site of the restriction endonuclease *Eco* RI. This enzyme makes a cut in the sugar-phosphate backbone between the nucleotides with bases guanine and adenine on both the polynucleotides with bases guanine and adenine on both polynucleotide strands (shown by arrows). B–This leads to the formation of DNA fragments with sticky ends.

Table 55.5. Examples of certain sticky– or cohesive-end restriction enzymes and sites.

Enzyme	Recognition site	Cleavage product
1. *Eco* RI	–G ↓ AATTC–	–G AATTC–
2. *Eco* RI	–CTTAA ↑ G–	–CTTAA G–
3. *Hind* III	–A ↓ AGCTT–	–A AGCTT–
4. *Hind* III	–TTCG A ↑ A–	–TTCGA A–

These ends are not only identical, but complementary, and will base pair with each other; they are therefore, known as cohesive or sticky ends. Because of specificity of restriction enzymes, every copy of a given DNA molecule will give the same set of fragments when cleaved with a particular enzyme. Different DNA molecules will in general, give different sets of fragments when treated with the same enzyme. The table 55.5 shows sticky or cohesive-end restriction enzymes and sites:

Host Controlled Restriction and Modification:

Certain strains of bacteria are immune to bacteriophages. This phenomenon is called host controlled restriction. This restriction is due to these restriction endonuclease enzymes (e.g., Eco RI) which could recognise and split specific loci in the foreign DNA. Thus these enzymes prevent or restrict the survival of foreign DNA in the host. This is analogous to an immune system.

All restriction sites in host chromosome of a bacterium are protected from its own restriction endonuclease enzyme due to a modification system. This system helps in preventing suicidal self-degradation.

Such modification occurs by methylation of specific bases in the recognition sequence of the endonuclease. The enzymes involved in such modification are called methyltransferases. These enzymes methylate adenine (i.e., adds a methyl group to the base) in the N6 position and cytosine either in N5 or W position and produce 6 methyl adenine and 5 or 4 methyl cytosine

respectively. Unmodified foreign DNA entering the cell is degraded by the host restriction system. As both the enzymes, i.e., methyltransferases and endonucleases recognise the restriction site, they are together called as restriction and modification system.

Star Activity:

Various REs show , the star activity when they exhibit relaxation in specificity of sequence under non-optimal conditions. In such condition, endonuclease enzymes even recognise other alternative base instead of a specific base. The following factors are known to alter the DNA recognition sequence for several to alter the DNA recognition sequence for several endonuclease enzymes: non-ideal strength buffers, high glycerol concentration (more than 5% v\v) and high enzyme concentration.

Isoschizomers:

Isoschizomers are restriction endonuclease enzymes which are isolated from different organisms but recognize identical base sequences in the DNA. For example, Asp 718 and Kpn I have identical recognition sites-

Source of Asp718 is Achromobacter species 718; source of Kpn I is Klebsiella pneumoniae OK 8. Some pairs of isochizomers cut their target at different places (e.g., Sma I, Xma I).

Use of Restriction Endonuclease Enzymes in Genetic Engineering:

In gene cloning experiments, DNA molecules have to be cut in a very precise and reproducible manner. Restriction endonuclease enzymes play an important role in cutting the desired gene as well as cleaving the vector.

1. Cutting the gene:

The required DNA fragment from a large DNA molecule should be cleaved in a precise manner for further genetic manipulations. A particular restriction endonuclease enzyme can recognize and bind to specific base sequence of the DNA and then will cleave it. It is highly reproducible and can be programmed according to DNA sequences of required gene and particular endonuclease enzymes identifying and cleaving it.

2. Cutting the vectors:

The function of a vector DNA molecule is to carry a gene of interest to a second organism where it can express it (i.e., can produce a gene specific product). During this technique the DNA to be cloned is integrated with the plasmid. Hence each vector molecule should be cleaved with same restriction site at a single position to open the circular form so that the new DNA fragment can be inserted at these complementary sites.

If foreign DNA is introduced into E. coli host, it may be attacked by restriction endonucleases active in a host cell. Because restriction phenomenon provides a natural defence against

invasion by foreign DNA, it is usual to employ a K restriction deficient E. coli K12 strain as a host in transformation with newly created recombinant DNA molecules. This will eliminate the chance that the incoming sequence will be restricted.

Restriction Enzymes as Tools:

Recognition sequences typically are only four to twelve nucleotides long. Because there are only so many ways to arrange the four nucleotides—A,C,G and T–into a four or eight or twelve nucleotide sequence, recognition sequences tend to "crop up" by chance in any long sequence. Furthermore, restriction enzymes specific to hundreds of distinct sequences have been identified and synthesized for sale to laboratories.

As a result, potential "restriction sites" appear in almost any gene on chromosome. Meanwhile, the sequences of some artificial plasmids include a "linker" that contains dozens of restriction enzyme recognition sequences within a very short segment of DNA. So no matter the context in which a gene naturally appears, there is probably a pair of restriction enzymes that can cut it out, and which will produce ends that enable the gene to be spliced into a "plasmid". Another use of restriction enzymes can be to find specific SNPs. If a restriction enzyme can be found such that it cuts only one possible allele of a section of DNA (that is, the alternate nucleotide of the SNP causes the restriction site to no longer exist within the section of DNA), this restriction enzyme can be used to genotype the sample without completely sequencing it. The sample is first run in a restriction digest to cut the DNA, and then gel electrophoresis is performed on this digest.

If the sample is homozygous for the common allele, the result will be two bands of DNA, because the cut will have occurred at the restriction site. If the sample is homozygous for the rarer allele, the sample will show only one band, because it will not have been cut. If the sample is heterozygous at that SNP, there will be three bands of DNA. This is an example of restriction mapping.

Other important enzymes which are used in Recombinant DNA Technology:

1. Alkaline Phosphatases:

The enzyme alkaline phosphatase (AP) catalyses the removal of the 5′-terminal phosphate residues from nucleic acids (RNA, DNA and ribo- and deoxyribonucleotide triphosphates). This enzyme is isolated from bacteria (BAP) or calf intestine (CAP). This enzyme is a dimeric glycoprotein with a molecular weight 14,000. It is made up of two identical or similar subunits each with a molecular weight of 6900. It is a zinc-containing enzyme with four atoms of Zn^{2+} per molecule.

Uses of Alkaline Phosphatase Enzyme:

1. Linearized cloning vectors can be prevented from recircularizing by dephosphorylation with alkaline phosphatase enzyme.
2. The free 5′-OH can be phosphorylated with polynucleotide kinase and Υ^{-32P} ATP to produce 32P end labelled nucleic acid.

3. AP enzyme is used for mapping and DNA fingerprinting studies.

2. Taq Polymerase:

Taq polymerases are DNA dependent DNA polymerase from Thermus aquaticus, primarily used for synthesis of longer stretches of DNA. Thermus aquaticus is the source for this enzyme which is an extreme thermophile, living in hot springs.

Due to its thermophillic habitats, it shows extreme resistant to high temperature, which exhibits peak of activity at temperature optimum of 72°C. Taq polymerase consists of single polypeptide chain with a molecular weight of 90,000 daltons. This enzyme is particularly useful when reactions need to take place at high temperature, denaturation of high G + C content into DNA. Thermostable property of this enzyme is significant in the polymerase chain reaction, where it is used to extend primers in the process of repeated rounds of heating, cooling and complete the synthesis of new DNA strand.

3. Nucleases:

Nucleases are DNases and RNases. DNases digest both the strands of DNA. It can hydrolyse each DNA strand independently in presence of Mg^{+} ion. DNases are used in the purification of DNA by eliminating contaminated DNA. It is also widely used in foot printing and nick translation.

RNases cleaves phosphodiester bond between the two adjacent nucleotides. RNase cleaves the bond next to uracil and guanine. Aspergillus and bovine pancreas are the chief source for RNase. The RNase H is widely employed in cDNA preparation where it removes mRNA from RNA-DNA hybrid. Removal of polyA tail requires RNase.

4. Reverse Transcriptase Enzyme:

The enzyme reverse transcriptase is isolated from avian myeloblastos virus (AMV). It is an RNA-dependent DNA polymerase. The enzyme requires DNA primer complementary to the RNA template, as well as presence of Mg^{2+} or Mn^{2+} for initiation of transcription. Reverse transcriptase mediates the conversion of genetic information present in single- stranded molecule of RNA into a double-stranded molecule of DNA.

Until recently, it was known that the genetic information's of DNA pass to protein through mRNA. During 1960s, Temin and coworkers postulated that in certain cancer causing animal viruses which contain RNA as genetic material, transcription of cancerous genes (on RNA into DNA) takes places most probably by DNA polymerase directed by viral RNA. Then DNA is used as template for synthesis of many copies of viral RNA in a cell. In 1970, S. Mizutani, H.M. Temin and D. Baltimore discovered that information can pass back from RNA to DNA.

They found that retroviruses (possessing RNA) contain RNA dependent DNA polymerase which IS also called reverse transcriptase. This process produces single- stranded DNA which in turn functions as template for complementary chain of DNA. Reverse transcriptase enzyme has two subunits. The enzymatically active forms of the purified enzyme are α, β and αβ. The molecular weight of the α-subunit is 68000 and that of β-subunit is 92,000. The mature α-β form is the most active form of AMV reverse transcriptase enzyme. It has several enzymatic roles such as RNA-directed DNA polymerase action, DNA dependent RNA polymerase activity

and RNase-H activity. The α- subunit of reverse transcriptase contains the polymerase activity. It also has the RNase- H activity during which degradation of RNA in DNA: RNA hybrids takes place. Such a sort of exonucleolytic activity of RNase-H enzyme can proceed either from the 5′- or 3′- terminus.

Uses of Reverse Transcriptase Enzyme:

a. The in vitro synthesis of cDNA from mRNA and other RNA molecule using reverse transcriptase enzyme has become a very important technique in the field of molecular biology.

b. DNA-dependent DNA polymerase activity of reverse transcriptase enzyme is responsible for second-strand formation in cDNA synthesis. Such a polymerising activity of reverse transcriptase is inhibited by the addition of actinomycin-D.

c. The reverse transcriptase enzyme mediates the conversion of genetic information present in single-stranded molecule of RNA into a double-stranded molecule of DNA.

5. DNA Ligating Enzyme (Ligases):

Ligases acts as a key player in genetic engineering experiment for its role as molecular suture, where it facilitates joining of DNA fragments. The main source of this enzyme is the T4 phage virus. The two DNA pieces are efficiently joined by DNA ligase.

The joining of DNA fragments requires ATP for T4 DNA ligase or NAD^+ in the case of E. coli ligase. Joining of DNA fragments is accomplished by forming a covalent bond between the 5′ phosphoryl of one strand and 3′ hydroxyl of the adjacent strand. Thus, it catalyses the end to end joining of DNA duplex at the base paired end. During sealing process, two phosphodiester bonds are formed by T4 DNA ligase. The stability of joined fragments is due to the formation of 3′ to 5′ phosphotodiester linkage between cohesive ends. The blunt ends produced by certain restriction enzymes may hinder the sealing process (Fig. 13.1). Hence, it is estimated to use ligase at high strength to accomplish sealing process.

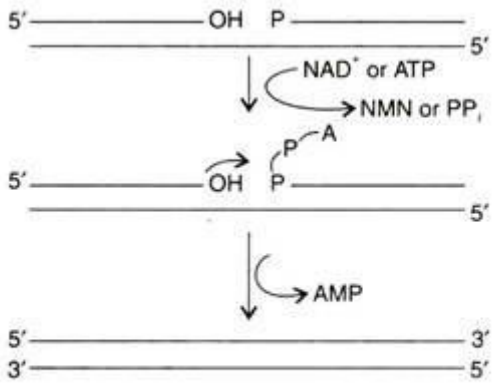


Fig. 13.1 Mechanism of DNA ligation

6. DNA Polymerase and the Klenow Fragment:

The DNA polymerase that is generally utilized is either the DNA Pol I from E. coli or the T4 DNA polymerase encoded by the phage gene. The E. coli enzyme is basically a proof-reading and repairing enzyme. It is composed of 3 subunits each with a specific activity. They are: 5′-3′ polymerase, 3′-5′ exonuclease and 5′-3′ exonuclease.

The enzyme is useful for synthesizing short length of a DNA strand, especially by the nick translation method. The 5-3′ exonuclease activity may be deleted, this edited enzyme is referred to as the Klenow fragment. The T4 DNA Pol possesses, like the klenow fragment, only the polymerase and proofreading (3′-5′ exonuclease) functions.

7. Phosphonucleotide Kinase Enzyme:

The enzyme phosphonucleotide kinase catalyses the transfer of the terminal phosphate group of ATP to the 5′-hydroxylated terminal of DNA or RNA. This enzyme is frequently used to end-label the nucleic acids with ^{32}P (i.e., it adds the phosphate back to 5′-termini of DNA).

This can be accomplished by any method among following:

a. Forward reaction:

Transfer of labelled ϒ-phosphate form (ϒ-^{32}P)-ATP to the free 5′-hydroxyl group of substrate-

$5' - OH - DNA + [^{-32}P]\ ATP5'^{32} \rightarrow PO - DNA + ADP$

Substrate lacking a free 5′-hydroxyl requires prior dephosphorylation by alkaline phosphatase.

b. Exchange reaction:

In the initial step, the terminal 5′-phosphate is transferred from substrate to ADP present in the reaction mixture. Then, the labelled ϒ-phosphate from [ϒ-^{32}P]-ATP is transferred to free hydroxyl group of substrate.

$$5' - PO - DNA + ADP \rightarrow 5'HO - DNA + ATP$$

$$5' - HO - DNA + [\Upsilon^{-32}\ P] - ATP \rightarrow 5'^{32}\ PO - DNA + ATP$$

Uses of Polynucleotide Kinase Enzyme:

The enzyme polynucleotide kinase is used to label 5′-termini of DNA and RNA with [ϒ-^{32}P]-ATP by phosphorylation of 5′-hydroxyl groups or by the exchange reaction. This 5′-terminal labelling is used in mapping of restriction sites, DNA or RNA fingerprinting, hybridization studies and sequence analysis of DNA.

8. S1 Nuclease Enzyme:

The S1 nuclease enzyme is single- strand specific endonuclease which cleaves DNA to release 5′-mono and 5′-oligonucleotides. Normally, double- stranded DNA, double- stranded RNA and DNA-RNA hybrids are resistant to action of S1 nuclease enzyme.

However, very large amounts of S1 nuclease enzyme can completely hydrolyze double-stranded nucleic acids. The enzyme hydrolyzes single stranded regions in duplex DNA such as loops and gaps. S1 nuclease enzyme can also cleave single stranded areas of super helical DNA

at torsional stress points where DNA may be unpaired or weakly hydrogen bonded. Once the super-helical DNA is nicked, S1 nuclease enzyme can cleave the second strand near the nick to generate linear DNA. S1 nuclease enzyme is a monomeric protein with 3800 Dalton molecular weight. It requires Zn^{2+} for its activity and is relatively stable against denaturing reagents such as urea, SDS and formamide. The optimum pH requirement lies between 4 to 4.5.

Uses of S1 Nuclease Enzyme:

1. S1 nuclease enzyme is used to analyse DNA-RNA hybrid structures to map transcripts.

2. It can be used to remove singles stranded tails from DNA fragments to produce blunt ends.

3. Hair pin loop structures formed during synthesis of double-stranded cDNA is digested by this enzyme.

4. S1 nuclease enzyme is also used for DNA mapping, called SI nuclease mapping Turner.

9. Ribonuclease:

Generally RNase A and RNase T1 enzymes are used in genetic engineering techniques. Both enzymes cleave the phosphodiester bond between adjacent ribonucleotides. RNase A cleaves next to uracil (U) and cytosine (C) in such a way that phosphate remains with these pyrimidines. The nucleotide present on the other side of phosphate is dephosphorylated. RNase A enzyme is isolated from the bovine pancreas.
RNase T1 cleaves specifically next to guanine. The phosphate group at the 3′ end of the nucleotide remains with the cut end. This enzyme is isolated from *Aspergillus oryzae.*

10. Ribonuclease H (RNase H):

The enzyme RNase H is an endoribonuclease that degrades the RNA portion of the RNA- DNA hybrids. RNase H enzyme cuts the RNA into short fragments.

Applications of RNase H:

a. RNase H is the key enzyme in the cDNA cloning technique. In this case, it is used to remove the mRNA from the RNA-DNA hybrid.

b. RNase H enzyme is used to detect the presence of RNA-DNA hybrid.

c. RNase H enzyme is used to remove poly (A) tails on mRNA.

11. Deoxyribonuclease I (Dnase I):

The enzyme DNase I is an endonuclease enzyme which digests either single or double-stranded DNA, producing a mixture of mononucleotides and oligonucleotides. DNase I hydrolyses each strand of double-stranded DNA independently and at random. Addition of Mg^{2+} to reaction mixture ensures random cleavage while addition of Mn^{2+} gives cleavage

nearly at the same place on both strands. DNase enzyme is obtained mostly from bovine pancreas.

Uses of DNase I Enzyme:

DNase 1 enzyme is useful for a variety of applications including nick translation, DNA foot printing, bisulphite mediated mutagenesis and RNA purification.

12. Terminal Deoxynucleotidyl Transferase Enzyme:

The enzyme deoxynucleotidyl transferase catalyses the repetitive addition of monodeoxynucleotide units from a deoxynucleoside triphosphate to the terminal 3′-hydroxyl group of a DNA molecule. This enzyme has a molecular weight of 32000 and consists of two subunits each with a molecular weight of 26500 and 8000. This enzyme is isolated from calf thymus.

Uses of Terminal Transferase Enzyme:

1. The enzyme terminal transferase is used to add homopolymer tails of DNA fragments. Using a technique called homopolymer tailing, sticky ends can be built up on blunt-ended DNA molecules. For examples, one preparation of DNA could be treated with the enzyme terminal transferase in the presence of dATP, resulting in the addition of a poly (dA) chain to each DNA strand. There is another preparation of DNA which provides 3 tails of poly (T) using same enzyme with TTP. When both types of DNA preparations DNA fragments with poly A tails and DNA fragments with poly T tails, are mixed, there takes place base pairing between complementary sticky ends, which could then be ligated. One advantage of this method is that ligation does not take place between fragments from the same DNA preparation.

2. Terminal transferase enzyme is used for 3′-end labelling of DNA fragments
3. Terminal transferase enzyme is also used for the addition of single nucleotides to the 3- end of DNA for in vitro mutagenesis.

Restriction Mapping:

A description of restriction endonuclease cleavage sites within a piece of DNA is referred to as a restriction map. Such a map is usually generated as the first step in characterizing an unknown DNA, and a prerequisite to manipulating it for other purposes. Restriction enzymes that cleave DNA infrequently (e.g. those with 6 bp recognition sites) are relatively inexpensive are used to produce at a map (Chakraborty, Pandey, et.al., 2006). Restriction sites are specific recognition sites where enzymes known as endonucleases cleave the DNA. e.g. EcoRI cuts at GAATTC (Gale, 2003). When discovered in archaea and bacteria, these enzymes were part of the defence mechanism of such organisms, limiting the foreign DNAs to act upon the cell. Theses enzymes will defend cells by digesting invading DNA into small, non-functional pieces. Thus this is where the name “restriction enzyme” comes from; the function of the enzyme, i.e. the ability of the enzyme to restricting access to the cell (Carroll, Griffiths, et.al., 2008). Restriction maps show the relative location of a selection of restriction sites along linear or circular DNA.

Restriction mapping involves a series of restriction enzymes digesting the DNA and then separating the resultant fragments by agarose gel electrophoresis. The patterns of fragments that are produced by restriction enzyme digestion determine the distance between restriction enzyme sites; this is how information about the structure of an unknown piece of DNA can be obtained (Champness & Snyder, 2007).

Techniques of Restriction Mapping:

There are several methods for restriction mapping; the most straightforward being the digesting of samples of the plasmid with a set of individual and pairs of those enzymes; these digests are then "run out" on an agarose gel to determine sizes of the fragments generated.
Consider to illustrate these ideas, a plasmid that contains a 3000 base pair (bp) fragment of unknown DNA. Immediately flanking the unknown DNA within the vector are unique recognition sites for the enzymes Kpn I and BamH I. Then, consider digestions with Kpn I and BamH I separately. In essence, single digests are used to determine which fragments are in the unknown DNA, and double digests to order and orient the fragments correctly (Chakraborty, Pandey, et.al., 2006). If a DNA fragment is labelled with a radioisotope on one end only, this can directly reveal where the cleavage sites are located as by partially digesting the fragment with restriction enzymes, labelled fragments are generated (Chakraborty, Pandey, et.al., 2006).
If the sequence is known, any number of computer programmes can be used to build up a map. It is simply a matter of feeding the sequence into the programme which will then search the sequence for dozens of restriction enzyme recognition sites and build you a map (Chakraborty, Pandey, et.al., 2006).

Uses and Applications of Restriction Mapping:

Restriction map information is important for many techniques used to manipulate DNA; one application being the cutting large pieces of DNA into smaller fragments for allowing it to be sequenced. Another application is to use restriction mapping to compare DNA fragments without having any information of their nucleotide sequence (Gale, 2003).

Restriction mapping has contributed immensely towards our knowledge of vectors and plasmids (OUP, 1995). It has also contributed heavily to our ability to genetically engineer organisms and recombinant DNA technology where an organism's genes are manipulated indirectly; examples of this include the generation of synthetic human insulin using modified bacteria and the production of erythropoietin in hamster ovary cells, amongst many more (Banting, 1929).
Industries like medicine, agriculture etc. also use this technique for the production of several medically useful substances like the hepatitis-B vaccine, human interferon and human growth hormone. Identifying the sequences with restriction mapping has allowed for plants to produce their own pesticides ant to perform nitrogen fixation by genetically adjusting the plant species. Bacteria capable of biodegrading oil have been produced using this technique for the use in oil-spill clean ups. The technique of restriction mapping has its applications in the field of gene knock out experiments in mice as well as identification of gene before introduction into a foreign organism to make transgenic pigs and cats. Similarly, we have been able to express

several medicinal proteins in bacterial systems using restriction techniques; the most famous examples are insulin(Banting 1929).

Is restriction mapping still useful?

The process of restriction mapping is simple and easy. It can be carried out in 1-2 days. The advancements in the field of computing have enabled automated softwares to virtually analyze the sequence by identifying the restriction sites. Restriction mapping is a helpful tool for experiments where sequencing can be out of budget or not necessary. It can be used to determine whether a gene has been cloned into the plasmid. It is a much better technique for relatively short segments of DNA.

Technologies That Succeeded Restriction Mapping.

One major method that has replaced restriction mapping is the High-Throughput Sequencing and Genotyping, which is to facilitate the unraveling of genetic information across the large and diverse collection of animals, plants and microbes. This has been very useful in cases where DNA does not contain any known restriction sites, or DNA, which contains sites for enzymes, which are not commercially available. It is also advisable to send the sample for sequencing when the sample is very small (Mitchelson, 2007).

Restriction fragment length polymorphisms (RFLP) are variations in DNA fragment-banding patterns (from different individuals of a species) of electrophoresed restriction digests of DNA (Appa Rao, Mohan, et.al., 1994).

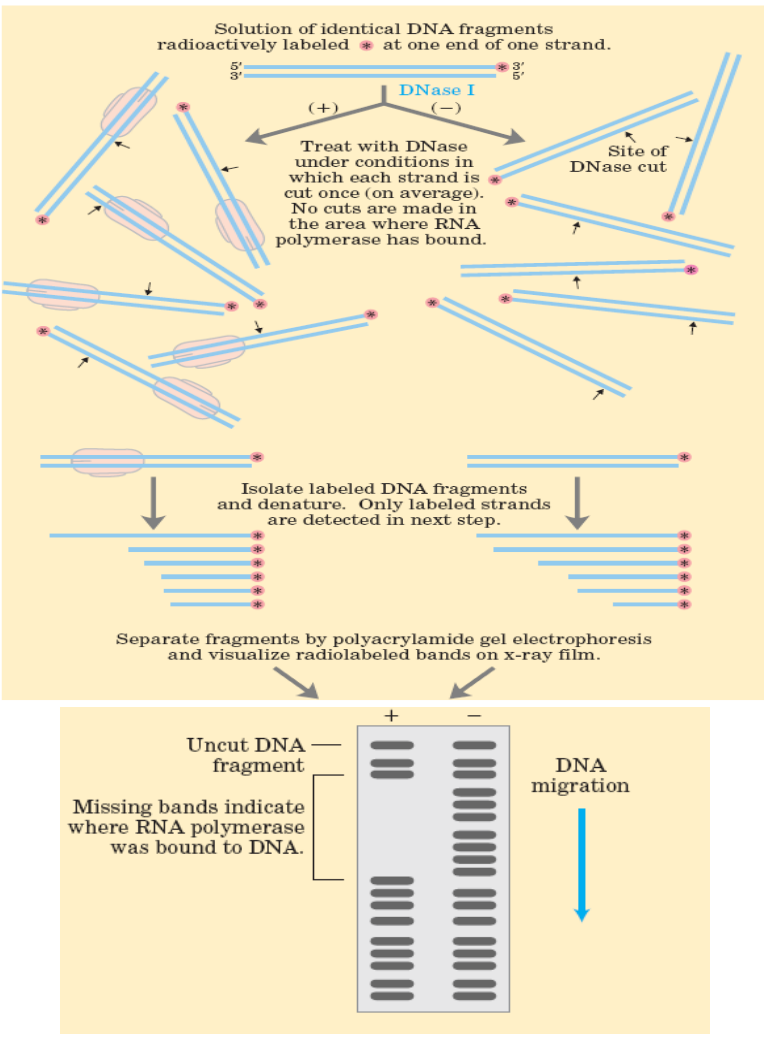
Random amplification of polymorphic DNA (RAPD) is a molecular marker technique using PCR with arbitrary primers for amplifying anonymous stretches of DNA (Chang & Meyerowitz, 1991). Southern blotting is a method of detection of specific DNA sequences in DNA samples. A southern blot combines the transfer of electrophoresis-separated DNA fragments to filter membranes and subsequent fragment detection by probe hybridization (Bignon, Roux-Dosseto, et.al., 1990). With regards to in vitro enzymatic amplification of DNA, the polymerase chain reaction (PCR) has developed into one of the most promising methods allowing widespread applications in DNA cloning, sequencing and mutagenesis related studies (Appa Rao, Mohan, et.al., 1994).

Conclusion:

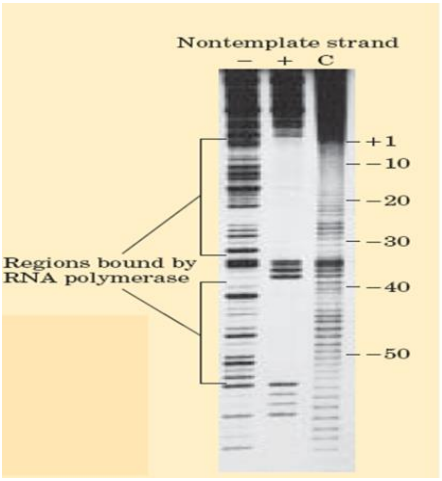
Restriction mapping is a technology to identify the unknown genes without sequencing. It has enabled to industry of biotechnology to its riches now. Though there are many techniques that replaced restriction mapping, it is still in use for low-cost academic exercises and other experiments. It has contributed heavily to our knowledge of DNA manipulation studies. It is an essential tool in comparing DNA fragments together. Many of the technologies like, RFLP, RAPD, PCR, HTGS have replaced restriction mapping nowadays though it would be immature to underestimate the value of restriction mapping.

DNA Footprinting

Footprinting**,** a technique derived from principles used in DNA sequencing, identifies the DNA sequences bound by a particular protein. Researchers isolate a DNA fragment thought to contain sequences recognized by a DNA-binding protein and radiolabel one end of one strand. They then use chemical or enzymatic reagents to introduce random breaks in the DNA fragment (averaging about one per molecule). Separation of the labeled cleavage products (broken fragments of various lengths) by high-resolution electrophoresis produces a ladder of radioactive bands. In a separate tube, the cleavage procedure is repeated on copies of the same DNA fragment in the presence of the DNA-binding protein. The researchers then subject the two sets of cleavage products to electrophoresis and compare them side by side. A gap ("footprint") in the series of radioactive bands derived from the DNA protein sample, attributable to protection of the DNA by the bound protein, identifies the sequences that the protein binds. The precise location of the protein binding site can be determined by directly sequencing copies of the same DNA fragment and including the sequencing lanes (not shown here) on the same gel with the footprint. Footprinting results for the binding of RNA polymerase to a DNA fragment containing a promoter. The polymerase covers 60 to 80 bp; protection by the bound enzyme includes the -10 and -35 region



Footprint analysis of the RNA polymerase–binding site on a DNA fragment. Separate experiments are carried out in the presence (+) and absence (-) of the polymerase.



Footprinting results of RNA polymerase binding to the lac promoter. In this experiment, the 5'end of the nontemplate strand was radioactively labeled. Lane C is a control in which the labeled DNA fragments were cleaved with a chemical reagent that produces a more uniform banding pattern.

Methyl interference assay:

The methylation interference assay is an analytical method that is used to determine which nucleotides in a DNA molecule are important for protein binding. This method provides information on where protein binding sites are and what proteins bind to a specific DNA molecule. Thus providing insight into how a gene may be regulated.

History and Significance:

The methylation interference assay was developed from the Maxam-Gilbert DNA sequencing method which also uses the methylation of guanine and adenine nucleotides followed by cleavage of these modified residues by piperidine. This technique was originally used to determine which factors were involved in modulating expression of the immunoglobin gene but has been used to study a wide range of problems from disease susceptibility to muscle function. It is especially useful due to its ability to resolve single nucleotides and is one of the highest resolution methods used to investigate sequence specific DNA-protein interaction.

Technique

Initially DNA fragments are ^{32}P end labeled after which guanine and adenine nucleotides are methylated by treatment with dimethyl sulfate. Treatment with this chemical generally results in 1 modified base per DNA molecule. Next the protein of interested is incubated with the methylated DNA fragment. If the DNA is modified at nucleotides that are involved in protein binding, the protein will be unable to bind to the DNA. However, if methylation has occurred at nucleotides that are not important to protein binding, the protein will be able to bind to the DNA. DNA fragments that are bound by protein and those that are not bound by protein can be

separated via electrophoretic mobility shift assay. Other common techniques used to separate bound and unbound DNA fragments include filter binding and immunoprecipitation. Next the DNA molecules are treated with an enzyme, such as piperidine, which will cleave the DNA molecules into smaller fragments. Piperidine cleaves DNA at modified bases. Finally the fragments are run on a denaturing gel along with a sequencing ladder. Because each DNA molecule is only methylated at a single position the sequencing ladder can be used to determine which nucleotide had been modified. Cleavage fragments that are generated from unbound DNA will differ from the fragments generated from DNA that was bound by protein. Thus the fragments generated from bound or unbound DNA can be used to determine the nucleotides that are important in protein binding.

Gene Amplification: Polymerase Chain Reaction (PCR):

PCR provides a simple and ingenious method for exponential amplification of specific DNA sequences by in vitro DNA synthesis, i.e., this technique has made it possible to synthesize large quantities of DNA fragments without cloning it.

Kary Mullis in 1985 developed the technique based on the use of an enzyme which is named as Taq DNA polymerase. The PCR technique has now been automated and is carried out by a specially designed machine.

Technique:

The technique involves the following three steps (Fig. 22.16):

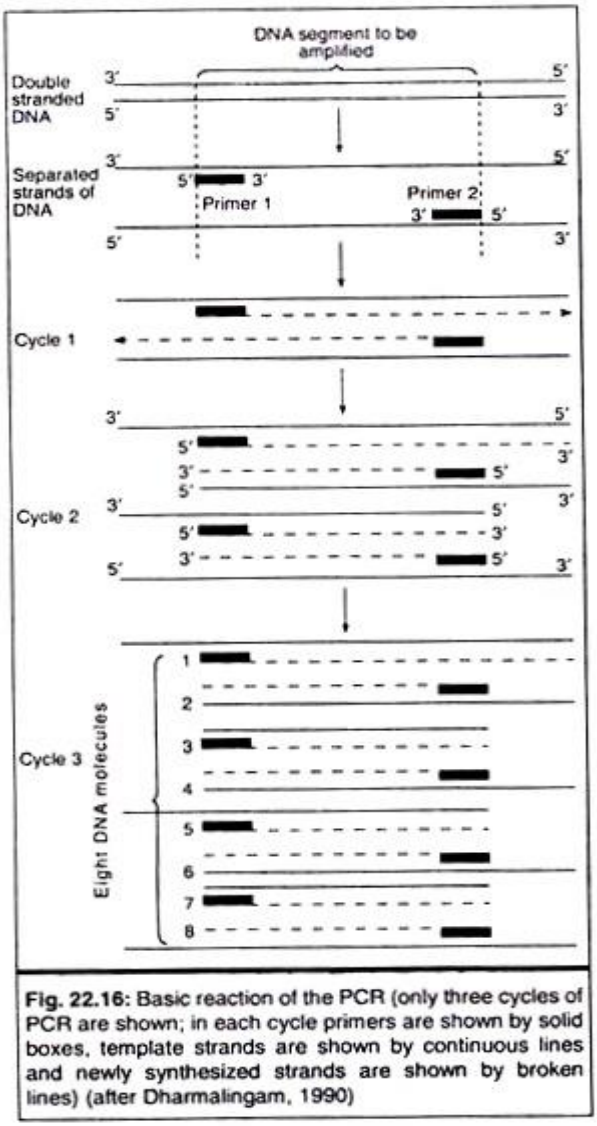


Fig. 22.16: Basic reaction of the PCR (only three cycles of PCR are shown; in each cycle primers are shown by solid boxes, template strands are shown by continuous lines and newly synthesized strands are shown by broken lines) (after Dharmalingam, 1990)

i. Denaturation of DNA Fragment:

The target DNA containing sequence to be amplified is heat denatured (around 94°C for 15 sec) to separate its complementary strands, this process is called melting of target DNA.

ii. Annealing of Primers:

Primers are added in excess and the temperature is lowered to about 68°C for 60 sec., as a result the primers form the hydrogen bonds and anneal to the DNA on both sides of the DNA sequence.

iii. Primer Extension:

Finally different nucleoside triphosphate (dATP, dGTP, dCTP, dTTP) and a thermo-stable DNA polymerase (Taq polymerase from Thermus aquaticus and Vent polymerase from *Thermococcus litoralis*) are added to the reaction mixture, it helps in polymerization process of primers and, therefore, extends the primers (at 68°C) resulting in synthesis of multiple copies of target DNA sequence.

After completion of all these steps in one cycle, again the second cycle is repeated following the same process. If 20 such cycles occur, then about one million copies of target DNA sequence are produced. Recently this technology has been improved much more, where instead of Taq polymerase the rTth polymerase is used which transcribe RNA to DNA, and thereafter amplify the DNA.

Modified Forms of PCR:

The conventional PCR is the symmetrical PCR technique. There are some other modified forms of PCR which are used for various purposes:

a. AP-PCR (Arbitrarily Primed Polymerase Chain Reaction):

It requires only a single primer of relatively much smaller length compared to the primers used in PCR. This technique is used for DNA profiling, in animal and plant biotechnology as well as in forensic medicine.

b. Asymmetrical PCR:

Target sequences of one strand may be amplified in several orders of magnitude more as compared to its complementary strand. This approach is particularly useful for generating single stranded DNA fragment to be used for sequencing of DNA.

c. I-PCR (Inverted Polymerase Chain Reaction):

In this method it allows the amplification of DNA flanking a known DNA sequence, the primers are facing outwards. Using the inverse PCR, the unknown sequences flanking known sequences can be readily amplified.

d. RT-PCR (Reverse Transcriptase Polymerase Chain Reaction):

Although the PCR amplification is generally performed on the DNA template but using this technique the RNA also can be used for amplification. This technique is particularly useful for studying the expression of genes and for monitoring the obscure species of mRNA.

e. Nested PCR:

Nested PCR primers are ones that are internal to the first primer pair. The larger fragments produced by the first round of PCR is used as the template for the second PCR. This technique eliminates any spurious non-specific amplification products.

Advantages of Polymerase Chain Reaction:

PCR is so sensitive that DNA sequences present in an individual cell can be amplified. The isolation and amplification of a specific DNA sequence by PCR is faster and less technically difficult than traditional cloning methods using recombinant DNA techniques.

Application of PCR in Biotechnology: PCR has many fold applications.

1. The amplification of gene fragments as fast alternative of cloning:

(a) Inserts of bacterial plasmids can be amplified with primers.

(b) DNA from known sequence can be obtained by designing primers.

(c) PCR helps in identification of homologous sequences from related organisms.

(d) Using RT-PCR the 3′ end of cDNA can be amplified (RACE: Rapid Amplification of cDNA Ends).

(e) Reverse PCR helps to know the flanking sequences of a known DNA clone.

2. Modification of DNA Fragments:

Site directed mutagenesis using oligonucleotides as PCR primers provides a powerful approach to study structure-function relation.

3. Diagnosis of Pathogenic Microorganism:

DNA from the infected parts of a person or animal may be subjected to PCR with primer specific gene of the pathogen and diagnosis can be done on amplification of DNA.

4. DNA Analysis of Archaeological Specimens:

As DNA is relatively stable and remain intact for a long period of time, PCR can help in analysis of DNA from those embedded materials.

5. Detection of Mutation Relevant for Inherited Diseases:

Any point mutation, a deletion or an insertion and expanded tandem trinucleotide repeat can be detected by PCR. Somatic mutations in oncogenes or tumour repressor genes can also be detected by PCR with primers flanking the insertions or deletions.

6. Analysis of Genetic Markers for Forensic Applications, for paternity testing and for the mapping of hereditary traits.

(a) Amplification of SSR.

(b) RAPD (Random Amplified Polymorphic DNA) with arbitrary, often short (10 bp) primers.

7. Species-Specific Amplification of DNA Segments between interspersed repeat elements (IRS) using the primer based on the SINE sequence (Short Interspersed Nuclear Elements).

8. Genetic Engineering using PCR:

Using PCR we can incorporate alteration or mutation in the ultimate product by choice altering, removing or adding sequences to the primer at the 5′ end. By recombinant PCR technique, it is possible to join two DNA fragments at a specific site through complementary overlaps (This technique is termed as splicing). By synthesizing two mutagenic primers,

spanning the internal site to be changed, it is possible to introduce mutations within a fragment.

Probable Questions:

1. Differentiate among Type I, II and III restriction endonuclease.
2. What is star activity? Give example.
3. Define isoschizomer. Give example.
4. How nomenclature of a type II RE is done?
5. Explain blunt cut and staggered cut of RE.
6. Hoe RE are used in genetic engineering ?
7. Explain the role of alkaline phosphatase in RDT.
8. Explain the role of DNA ligase in RDT.
9. Explain the role of Reverse transcriptase in RDT.
10. Explain the role of S1 nuclease in RDT.
11. Explain the role of RNase H in RDT.
12. What is restriction mapping? How it is performed?
13. What is DNA footprinting? What is its significance?
13. What is methyl interference assay. Explain the procedure.
14. Define PCR. Describe different steps of PCR.
15. Discuss application of PCR in biotechnology.
16. Explain nested PCR, asymmetrical PCR, AP-PCR, RT-PCR in brief.

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.
9. Primrose SB and Twyman RM. (2006) Principles of Gene Manipulation and Genomics, 7th edition. Blackwell Publishing, Oxford, U.K.
10. Sambrook J and Russell D. (2001) Molecular Cloning-A Laboratory Manual. 3rd edition. Cold Spring Harbor Laboratory Pres

Unit-VI

Manipulating genes in animals: gene transfer to animal cells, genetic manipulation of animals, transgenic technology, application of recombinant DNA technology; genetically modified organisms: gene knockouts, mouse disease models, gene silencing, gene therapy, somatic and germ- line therapy

Objective: In this unit you will learn about various techniques which are used for introduction of foreign genes into animal cells. You will also have an idea on GMO and their uses. Gene silencing methods including knock out mice and gene therapy methods will also be discussed in this unit.

Gene transfer methods in Animal cells:

There are several methods available by which foreign gene (transgene) can be transferred into animal cells.

Method - 1. Calcium Chloride ($CaCl_2$) Mediated DNA Transfer:

This is used for the transformation of prokaryotic host cells.

Principle:

In the process of transformation all bacterial cells cannot uptake the exogenous DNA molecule. Those who are capable to take are called competent cells. So our aim in this step is to make bacterial cells more competent so that the possibility of transferring of the recombinant DNA into the host cell increases to a higher fold. $CaCl_2$ makes the cell wall of the bacteria more permeable to the exogenous DNA and thus increases the competence of the host cell.

Procedure:

Growing E. Coli cells are isolated and suspended in 50 mM $CaCl_2$ at a concentration of 108-1010 cells/ml. The cells may be incubated for 12- 24 hr. to increase the frequency of transformation. The recombinant DNA is then added.

Efficient transformation takes only a few minutes and the cells are plated on a suitable medium for the selection of transformed clones. The frequency of transformed cells is 106-107 per mg of plasmid DNA; this is about one transformation per 10,000 plasmid molecules.

The transformed cells are suitably diluted and spread thinly on a suitable medium so that each cell is well separated and produces a separate colony. Generally, the medium is so designed that it permits only the transformed cells to divide and produce colonies. This frequency can be further improved by using special *E. Coli* strains, e.g., SK1590, SK1592, X1766, etc.

Method - 2. Rubidium Chloride Mediated DNA Transfer:

The rubidium chloride method is a variant of the calcium chloride method that offers some-what higher competency. The process followed is same as before but just the $CaCl_2$is replaced with $RbCl_2$. This is also used in the transformation of the prokaryotic host cell.

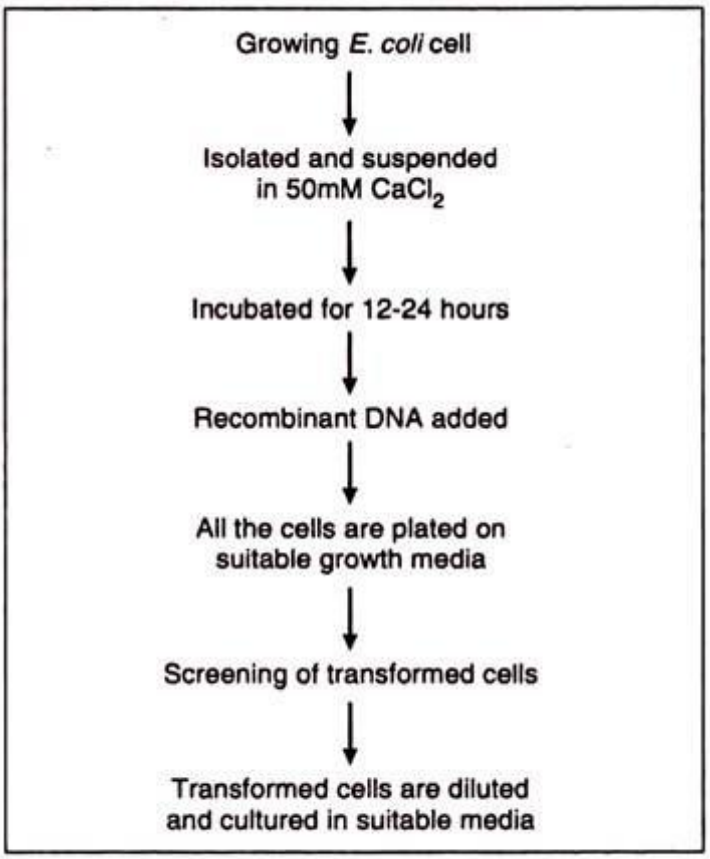


Fig. 5.5: Calcium chloride ($CaCl_2$) mediated DNA transfer

Method – 3: Electroporation:

Electroporation or electro-permeabilization is the process of applying electrical field to a living cell for a brief duration of time in order to create microscopic pores in the plasma membrane called electro-pores. This technique is used for transferring the recombinant DNA molecule into wide range of hosts starting from bacteria to plant (plant protoplasts) and animal cells.

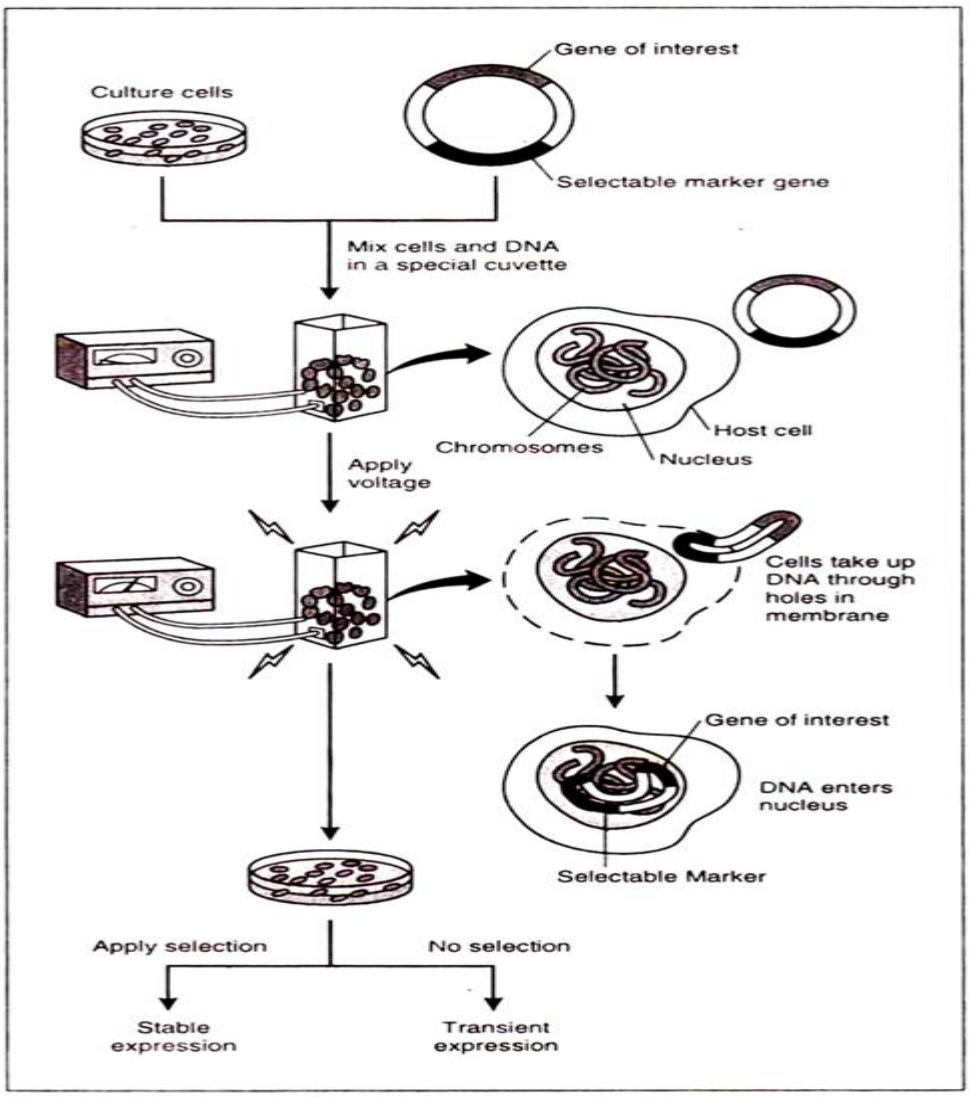


Fig. 5.6: Gene transfer by electroporation

Principle:

The phospholipid molecules of the plasma membrane are not static. When we apply electric field to them their kinetic energy increases resulting in the increase in the membrane permeability at certain points. This is exactly where we see the formation of electro-pores. The recombinant DNA can pass through these transient pores before they close.

Procedure:

In this process cells are mixed with the recombinant DNA and the mixture is placed in a small chamber with electrodes connected to a specialized power supply. Then a brief electric impulse is discharged across the electrodes, which makes pores (holes) in the plasma membrane.

These pores remain for some time and are again resealed themselves. Recombinant DNA enters the cell which are removed and plated in fresh selective medium. The process of selection is then applied to isolate cells carrying recombinant DNA.

Method # 4. Liposome Encapsulation (Lipofection):

This technique is found very successful in the transfection of plant protoplasts and animal host cells.

Principle:

Liposomes are microscopic vesicles developed in a laboratory environment. Each liposome is a spherical ball like structure made up of phospholipid bilayers with a hollow central space, allowing liposomes to interact directly with cells.

A liposome can fuse with the cell membrane of the taken host cell and can deliver its content to it. The recombinant DNA enclosed in the liposome vesicles penetrates into the protoplast of the host cell.

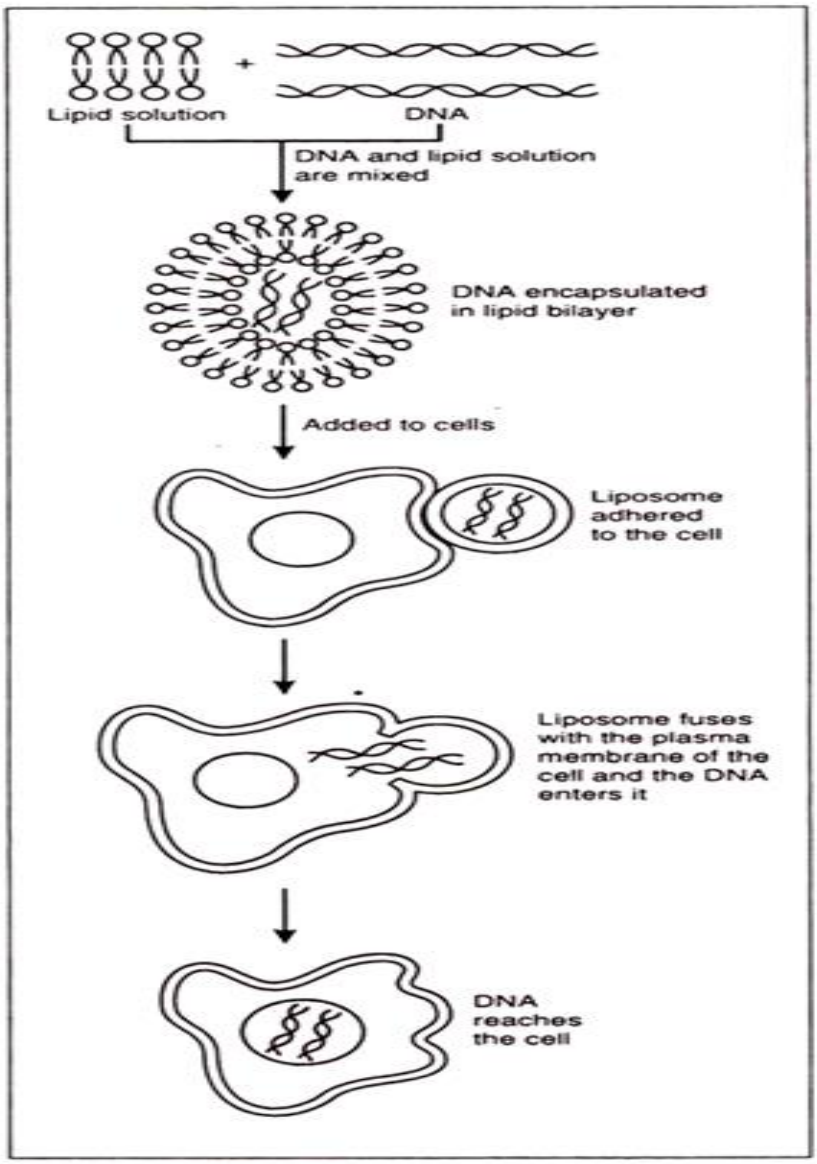


Fig. 5.7: Lipofection

Procedure:

In this technique the recombinant DNA, which is negatively charged at a near neutral pH because of its phosphodiester backbone, is mixed with the lipid molecules with positively charged (cationic) head groups. The lipid molecules form a bilayer around the recombinant DNA molecules. This results in the formation of liposomes which are further mixed with the host cells. Most eukaryotic cells are negatively charged at their surface, so the positively charged liposomes interact with the cells. Cells take up the lipid-recombinant DNA complexes, and some of the transfected DNA enters the nucleus.

Method -5. Microinjection:

This is the direct introduction of the recombinant DNA into the host cell. This technique has been used successfully with both plan and animal cells. In this procedure the cell is held on a glass capillary by gentle suction.

The microinjection needle is made by drawing out a heated glass capillary to a fine point. Using a micromanipulator (a mechanical device for fine control of the capillary) the needle has been inserted into the nucleus of the host cell.

One obvious disadvantage is that this technique is labour-intensive and not suitable for primary cloning procedures where large numbers of recombinants are required. However, in certain specialised cases it is an excellent method for targeting DNA delivery once a suitable re-combinant has been identified and developed to the point where microinjection is feasible.

Method # 6. Biolistic Particle Delivery System:

A gene gun or a biolistic particle delivery system is a device which can directly bombard small particles coated with the recombinant DNA on the nucleus of the target cell. This technique is often simply referred to as bio-ballistics or biolistics and has been successfully used in the transfection of both plant and animal cells.

In this technique the recombinant DNA is coated with microscopic tungsten particles known as micro-projectiles, which are then accelerated on a macro-projectile by firing a gunpowder charge or by using compressed gas to drive the macro-projectile.

At one end of the 'gun' there is a small aperture that stops the macro-projectile but allows the micro-projectiles to pass through. When directed at cells, these micro-projectiles carry the DNA into the cell and, in some cases, stable transformation will occur.

Method # 7. Calcium Phosphate Co-Precipitation:

This technique is used for the transfection of plant and mostly animal cells. The recombinant DNA is mixed with calcium chloride in a phosphate buffer at neutral pH. This results in the formation of recombinant DNA-calcium phosphate complex which appears as a thin precipitate. This precipitate is then added to the host cell.

The precipitate is taken up by the cell by the process of phagocytosis. The recombinant DNA enters the nucleus and integrates into the host's genome. The transfection efficiency can be increased by exposing the host cell to 10-20% glycerol or Dimethyl sulfoxide (DMSO).

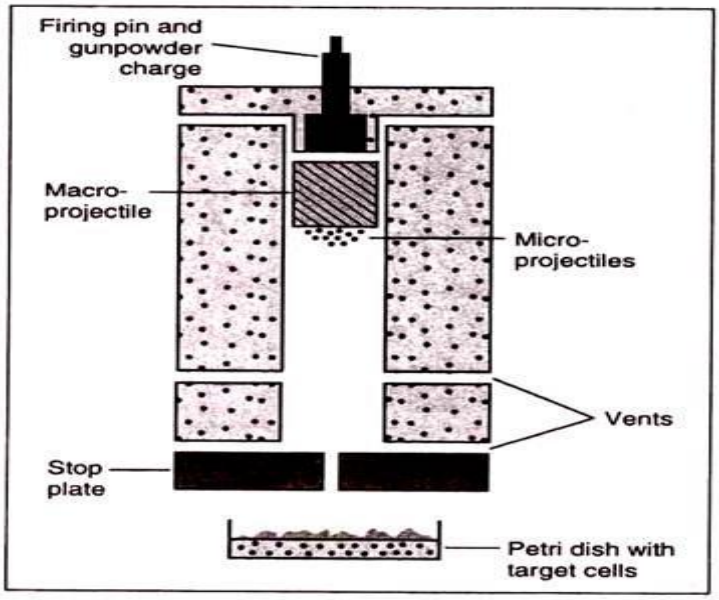


Fig. 5.9: Biolistic particle delivery system

Method - 8. Sonoporation:

Sonoporation, or cellular sonication, is the use of sound (typically ultrasonic frequencies) for the transfer of recombinant DNA into the target host cell. This process has been successfully used in a wide range of host cells starting from bacteria to plant and animal cells.
This employs the acoustic waves to increase the permeability of the plasma membrane. Taking the advantage of this situation the recombinant DNA enters the host cell.

Method - 9. Optical Transfection:

Optical Transfection is the process of introducing nucleic acids into cells using light. This has been successful in transfecting animal cells. In this technique the plasma membrane of the host cell is exposed to the highly focused laser beam for a small amount of time (typically tens of milliseconds to seconds), generating a transient pore on the membrane called photo-pore. Through the photo-pore the recombinant DNA can enter the host cell.

Method - 10. Impalefection:

Impalefection is a method of gene delivery using Nano materials, such as carbon Nano fibres, carbon nanotubes, nanowires, etc. This technique is used for the transfection of plant and animal cells. In this technique needle-like nanostructures are synthesized perpendicularly to the surface of a substrate.Recombinant DNA is attached to the nanostructure surface. A chip with arrays of these needles is then pressed against cells or tissue.

Method - 11. Magnetofection:

Magnetofection, or Magnet assisted transfection is a method, which uses magnetic force to deliver recombinant DNA into target host cells. Nucleic acids are first associated with magnetic nanoparticles. Then, application of magnetic force drives the nucleic acid particle complexes towards and into the target host cells, where the cargo is released. This has been successfully used to transfect the plant and animal cells.

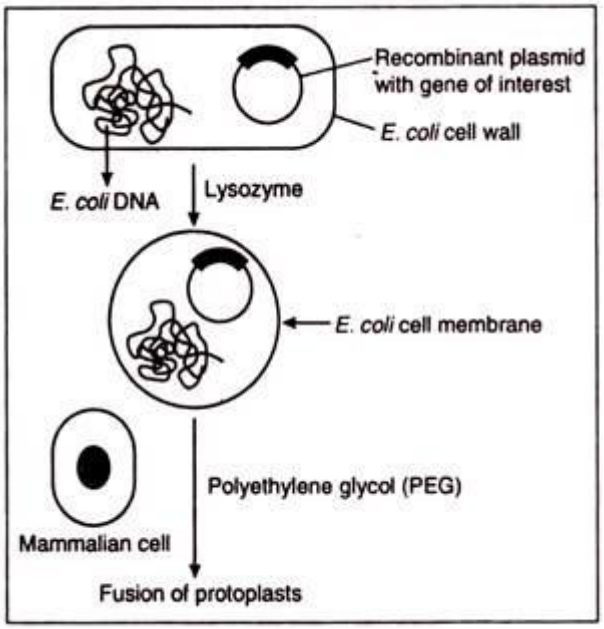


Fig. 5.10: Gene transfer by protoplast fusion

Method - 12. Protoplast Fusion:

This technique is used for introducing gene of interest into plant and animal cells. In this technique first we transfer the recombinant DNA into a bacterial cell then dissolve its cell wall by treating it with lysozyme. After this we fuse the host protoplast with the bacterial cell (lacking cell wall) by the help of polyethylene glycol (PEG). The transfected cells are then selected by suitable methods.

Method - 13. Virus Mediated Gene Transfer:

In other way the gene can be packed into a virus and allow it to infect the host cell without harming it in any way. This method can be used both for the transformation of prokaryotic host cell as well as transfection of eukaryotic host cells. In the case of bacterial host cells the recombinant DNA can be packed into the empty head of a specially designed bacteriophage (e.g., lambda phage) and allow the virion to infect the host cell.

Similarly, while transfecting the plant host cells we can follow the similar strategy by using plant viruses like Caulimo virus and Gemini virus. In the case of animals, retrovirus infection of embryos has been used for the production of transgenic mice.

This virus has been found to be an efficient vector system for animals. The virus carrying the gene of interest transfers it into the genome of embryonic cells leading to its integration and production of transgenic animals.

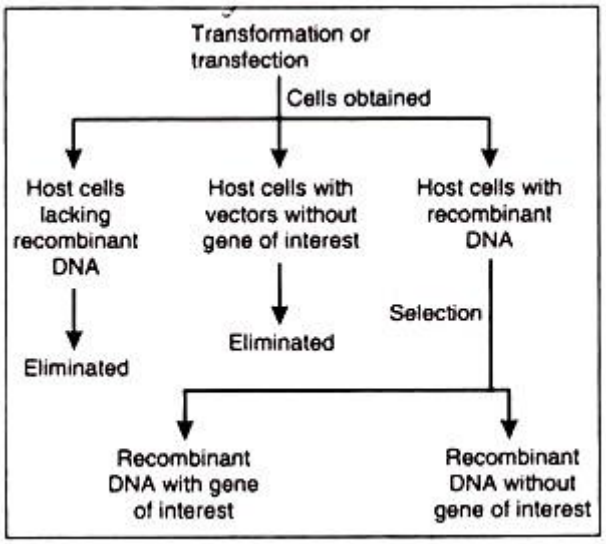


Fig. 5.11: The necessity of selection procedure

Application of Recombinant DNA Technology:

Major applications of recombinant DNA technology are:

1. Medical Diagnosis of Disease
2. Gene Therapy
3. Production of Vaccines through rDNA Technology
4. Cloning
5. Genetically Modified Crops
6. Sustainable Agriculture.

Application - 1. Medical Diagnosis of Disease:

rDNA technology acts as a tool to diagnose the diseases. This involves the construction of probes (short, single strands of radioactive or fluorescent DNA, used to identify the complementary DNA). These probes are used to identify the infectious agents, such as Salmonella (food poisoning), Staphylococcus (pus), HIV, hepatitis virus, etc. With the help of this technique the infected child can also be identified. This can be done by testing the DNA of prospective parents for any genetic disorder, i.e. they are not carrier of a disorder.

Few examples of this technique are as follows:

a. Phenylketonuria:

In this, phenylalanine fails to get converted into tyrosine. This causes disturbances in metabolism resulting in mental retardation. It is possible to cure this disease by using rDNA technique in early periods of pregnancy.

b. Thalassemia Genes:

In this, syntheses of alpha and beta globin chains are reduced and the excess chains precipitate and cause haemolytic anaemia and spleen enlargement. Human globin genes have been identified and sequenced. Alpha and beta globin genes are closely linked. Human globin gene has also been developed and cloned. Still lot of work needs to be done to cure this disease.

c. Haemophilia Gene:

It is a sex linked disease in human where blood clotting does not take place normally due to the absence of clotting factor VIII C. By using gene cloning technique, the clotting factor VIII C gene was cloned to express in mammalian cell lines and produce the protein VIII C responsible for blood clotting.

Application - 2. Gene Therapy:

Gene therapy means to change a faulty gene with a normal, healthy gene. Gene therapy can be used to correct a rare disease, like sickle cell anaemia, which is caused by single mutation and killer diseases such as Severe Combined Immuno Deficiency (SCID). Gene therapy is used to produce recombinant therapeutic bio chemicals such as insulin, somatotropin, somatostatin, interferon, human blood clotting factor VIII, etc.

Several protocols have been developed for expression and introduction of genes in humans, but the clinical efficiency has to be demonstrated conclusively. Success of gene therapy is dependent on the development of better gene transfer vector for sustained, long-term expression of foreign gene as well as better understanding of gene physiology of human disease.

There are two gene transfer strategies:

a. The in vivo approach which involves introduction of genes directly into the target organ of an individual. This is done in patients, therefore called as patient therapy.

b. Ex vivo approach where the cells are isolated for gene transfer in vitro followed by transplantation of genetically modified cells back into the patients.

Application - 3. Production of Vaccines through rDNA Technology:

Human Insulin:

For this the gene of interest is picked up from a human cell. Plasmid from E. coli is taken and by using restriction enzyme it is cut to create sticky ends. Now the gene of interest (insulin gene) and plasmid are joined by DNA ligase. This is now known as rDNA. This rDNA is now inserted in the plasmid free E.coli. Multiplication of rDNA starts growth in medium. Clones of genetically engineered bacteria are used to extract recombinant insulin. In the same way by using hepatitis B virus and Agrobacterium tumefaciens, hepatitis B vaccine and edible vaccine can be generated.

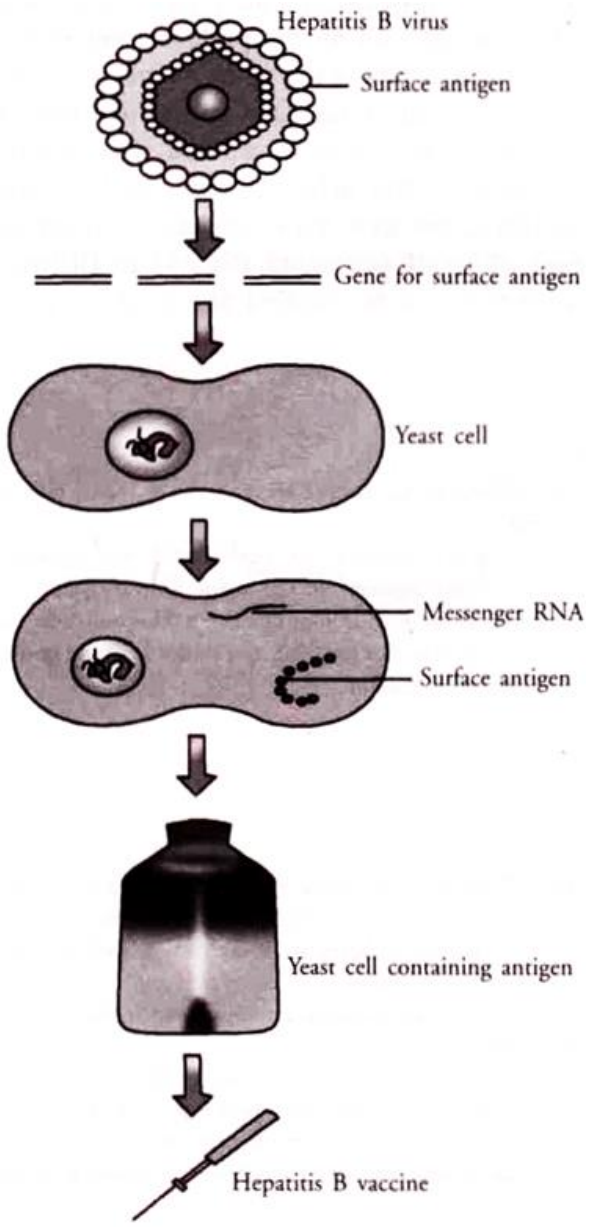


Fig. 2 Formation of injectable hepatitis-B vaccine.

Application - 4. Cloning:

Cloning means to create a carbon copy or identical copy of single parent. This word is related to only the living world and not to non-living world, where we can find thousands of copies of one object like number of photocopies of the same document. In nature asexually reproducing organisms produce clones.

For example, amoeba reproducing by binary fission produce two daughter amoebae which are clones. In human beings, monozygotic, identical twins are clones. They are the result of separation of the two cells of zygote which are in double cell stage. The most famous example of cloning is Dolly sheep.

A. Microbial Cloning:

Once the microbial cells are modified or genetically altered they are cloned on a growth medium. In a few days there are millions of clones generated. Each one is the copy of a single parent. Table 2 shows the genetically modified microbes and their applications.

Table 2 Applications of genetically modified microbes.

Microbes	Applications
Escherichia coli (gut bacterium)	Production of human insulin, human growth factor, interferons, interleukin and so on
Bacillus thuringiensis (soil bacterium)	Production of endotoxin (Bt toxin), highly potent, safe and biodegradable insectide for plant protection
Rhizobium meliloti (bacterium)	Nitrogen fixation by incorporating *'nif'* gene in cereal crops
Pseudomonas fluorescence (bacterium)	Prevention of frost damage to the plants (e.g. strawberries) on which it grows
Pseudomonas putida (bacterium)	Scavenging of oil spills by digesting hydrocarbons of crude oil
Bacterial strains capable of accumulating heavy metal	Bioremediation (cleaning of pollutants in the environment)
Trichoderma (fungus)	Production of enzyme chitinases for biocontrol of fungal diseases in plants

B. Cell Cloning:

This technique is based on the fact that certain cells are totipotent, i.e. they are not differentiated. This phenomenon is seen both in plants and animals. When seen in plants it is called totipotency and when seen in animals it is known as pluriopotency. Almost all the plants show totipotency, whereas in animals pluriopotency is seen in fertilised egg and stem cells in blastocyst. Cells showing pluriopotency can be differentiated into nerve cells, kidney cells and even heart cells.

C. Plant Cloning:

The growth areas of plant, i.e. root and shoot tip are used in plant cloning. This is used to multiply those plants which are agronomically (crop plants) important. Plants which are useful to horticulturist (orchids, gladiolies, etc.) are multiplied at a very fast pace. By gene manipulation we can have drought, disease, insect and pest resistant varieties. We can also have herbicide tolerant variety. Genetically modified food can also be produced like Vitamin A rich rice (Golden Rice), Lysin rich pulse, etc.

Application - 5. Genetically Modified Crops:

A crop which bears a foreign gene of desired function of other organism and expresses itself is called genetically modified crop (GM crop) or transgenic crop. In the last 20 years, considerable progress has been made on isolation, characterisation and introduction of novel genes into plants. In the year 2002, transgenic plants were cultivated on around 587 million hectare land in the world. The number of farmers involved in this was 5.5 million. Transgenic

crop plants have many beneficial traits such as pest and insect resistance, weed control, improved oil quality, herbicide tolerance, delayed fruit ripening, etc.

Two main advantages of transgenic crops are:

a. Any gene can be transferred from any organism
b. Change in genotype can be controlled as only the desired gene is introduced.
In contrary to this, when conventional method like hybridisation is used only those genes can be used which are found in such species and along with desirable change undesirable genes are also added.

The effects of introduction of foreign gene of interest are as follows:
a. Existing biosynthetic pathway gets modified so that a new end product is obtained.

b. It produces a protein that is the product of interest.

c. It produces a protein that on its own produce the desired phenotype.

d. It prevents the expression of an already existing gene.

Application - 6. Sustainable Agriculture:

Human population is ever increasing and because of this the major challenge for agricultural scientists is to increase the food production at almost the same pace. This is not possible by the conventional methods used for increasing and improving the yield. In recent years, it has been understood that biotechnology can play a major role to overcome this problem.

Table 4 Transgenic plants that produce antigens.

Transgenic plants	Active against bacterial/viral pathogens	Antigens (vaccines)
Potato	*E.coli*	Heat labile enterotoxin β-subunit
	Vibrio cholerae	Cholera toxin β-subunit
Tobacco	Hepatitis B virus	Hepatitis B surface antigen (HBsAg)
	Vibrio cholerae	Cholera toxin β-subunit
Tomato	Rabies virus	Rabies virus glycoprotein

Earlier emphasis was laid on the use of pesticides and fertilisers. Later it was realised that use of pesticides and fertilisers led to environmental pollution. Due to this reason such practices cannot be continued for indefinite period of time. The only way to overcome this is to switch over to sustainable agriculture.
Sustainable agriculture means to use the resources in judicious manner where we do not exploit it to support the present requirement and take care to leave behind sufficient resources for the coming generation. Care should also be taken to use renewable resources. Development which minimises the use of non-renewable resources results in minimising the environmental exploitation.

Through genetic manipulation crop production and usefulness of the products can be increased. Due to the introduction of agricultural biotechnology we have plants which have not one but many improved traits. The only problem which is faced while commercialising these varieties is that with the passage of time their expression is reduced. To overcome this problem scientists cross different transgenic lines that have improved breeding material. The progeny so formed undergoes selfing to give rise to varieties which have desired characters. By the above technique we can have traits that are all of same size, colour, weight and shape. Their nutritional value can also be improved keeping intact their time of ripening.

Genetically Modified Organisms:

A genetically modified organism (GMO) is any organism whose genetic material has been altered using genetic engineering techniques (i.e., a genetically engineered organism). GMOs are used to produce many medications and genetically modified foods and are widely used in scientific research and the production of other goods. The term GMO is very close to the technical legal term, 'living modified organism', defined in the Cartagena Protocol on Biosafety, which regulates international trade in living GMOs (specifically, "any living organism that possesses a novel combination of genetic material obtained through the use of modern biotechnology").A more specifically defined type of GMO is a "transgenic organism." This is an organism whose genetic makeup has been altered by the addition of genetic material from an unrelated organism. This should not be confused with the more general way in which "GMO" is used to classify genetically altered organisms, as typically GMOs are organisms whose genetic makeup has been altered without the addition of genetic material from an unrelated organism. The first genetically modified mouse was created in 1974, and the first plant was produced in 1983.

Applications of GMOs:

GMOs are used in biological and medical research, production of pharmaceutical drugs,[experimental medicine (e.g. gene therapy and vaccines against the Ebola virus), and agriculture (e.g. golden rice, resistance to herbicides), with developing uses in conservation. The term "genetically modified organism" does not always imply, but can include, targeted insertions of genes from one species into another. For example, a gene from a jellyfish, encoding a fluorescent protein called GFP, or green fluorescent protein, can be physically linked and thus co-expressed with mammalian genes to identify the location of the protein encoded by the GFP-tagged gene in the mammalian cell. Such methods are useful tools for biologists in many areas of research, including those who study the mechanisms of human and other diseases or fundamental biological processes in eukaryotic or prokaryotic cells.

Controversy:

There is controversy over GMOs, especially with regard to their use in producing food. The dispute involves buyers, biotechnology companies, governmental regulators, nongovernmental organizations, and scientists. The key areas of controversy related to GMO food are whether GM food should be labelled, the role of government regulators, the effect of GM crops on health and the environment, the effect on pesticide resistance, the impact of GM crops for farmers,

and the role of GM crops in feeding the world population. In 2014, sales of products that had been labelled as non-GMO grew 30 percent to $1.1 billion.
There is a scientific consensus that currently available food derived from GM crops poses no greater risk to human health than conventional food, but that each GM food needs to be tested on a case-by-case basis before introduction. Nonetheless, members of the public are much less likely than scientists to perceive GM foods as safe. The legal and regulatory status of GM foods varies by country, with some nations banning or restricting them, and others permitting them with widely differing degrees of regulation.
No reports of ill effects have been proven in the human population from ingesting GM food. Although labelling of GMO products in the marketplace is required in many countries, it is not required in the United States and no distinction between marketed GMO and non-GMO foods is recognized by the US FDA. In a May 2014 article in The Economist it was argued that, while GM foods could potentially help feed 842 million malnourished people globally, laws such as the one passed in Vermont, to require labelling of foods containing genetically modified ingredients, could have the unintended consequence of interrupting the process of spreading GM technologies to impoverished countries that suffer with food security problems.
The Organic Consumers Association, and the Union of Concerned Scientists, and Greenpeace stated that risks have not been adequately identified and managed, and they have questioned the objectivity of regulatory authorities. Some health groups say there are unanswered questions regarding the potential long-term impact on human health from food derived from GMOs, and propose mandatory labelling or a moratorium on such products. Concerns include contamination of the non-genetically modified food supply, effects of GMOs on the environment and nature, the rigor of the regulatory process, and consolidation of control of the food supply in companies that make and sell GMOs, or concerns over the use of herbicides with glyphosate.

Regulation of GMOs:

The regulation of genetic engineering varies widely by country. Countries such as the United States, Canada, Lebanon and Egypt use substantial equivalence as the starting point when assessing safety, while many countries such as those in the European Union, Brazil and China authorize GMO cultivation on a case-by-case basis. Many countries allow the import of GM food with authorization, but either do not allow its cultivation (Russia, Norway, Israel) or have provisions for cultivation, but no GM products are yet produced (Japan, South Korea). Most countries that do not allow for GMO cultivation do permit research.
One of the key issues concerning regulators is whether GM products should be labelled. Labelling of GMO products in the marketplace is required in 64 countries. Labelling can be mandatory up to a threshold GM content level (which varies between countries) or voluntary. A study investigating voluntary labelling in South Africa found that 31% of products labelled as GMO-free had a GM content above 1.0%.In Canada and the USA labelling of GM food is voluntary, while in Europe all food (including processed food) or feed which contains greater than 0.9% of approved GMOs must be labelled. There is a scientific consensus that currently available food derived from GM crops poses no greater risk to human health than conventional food, but that each GM food needs to be tested on a case-by-case basis before introduction.

Nonetheless, members of the public are much less likely than scientists to perceive GM foods as safe. The legal and regulatory status of GM foods varies by country, with some nations banning or restricting them, and others permitting them with widely differing degrees of regulation. There is no evidence to support the idea that the consumption of approved GM food has a detrimental effect on human health. Some scientists and advocacy groups, such as Greenpeace and World Wildlife Fund, have however called for additional and more rigorous testing for GM food.

Examples:

a. Microbes:

Bacteria were the first organisms to be modified in the laboratory, due to the relative ease of modifying their genetics. They continue to be important model organisms for experiments in genetic engineering. In the field of synthetic biology, they have been used to test various synthetic approaches, from synthesizing genomes to creating novel nucleotides. These organisms are now used for several purposes, and are particularly important in producing large amounts of pure human proteins for use in medicine. Genetically modified bacteria are used to produce the protein insulin to treat diabetes. Similar bacteria have been used to produce biofuels, clotting factors to treat haemophilia, and human growth hormone to treat various forms of dwarfism.

In 2017 researchers genetically modified a virus to express spinach defensin proteins. The virus was injected into orange trees to combat citrus greening disease that had reduced orange production 70% since 2005. In addition, various genetically engineered micro-organisms are routinely used as sources of enzymes for the manufacture of a variety of processed foods. These include alpha-amylase from bacteria, which converts starch to simple sugars, chymosin from bacteria or fungi, which clots milk protein for cheese making, and pectinesterase from fungi, which improves fruit juice clarity.

b. Plants:

Transgenic plants have been engineered for scientific research, to create new colours in plants, and to create different crops. In research, plants are engineered to help discover the functions of certain genes. One way to do this is to knock out the gene of interest and see what phenotype develops. Another strategy is to attach the gene to a strong promoter and see what happens when it is over expressed. A common technique used to find out where the gene is expressed is to attach it to GUS or a similar reporter gene that allows visualisation of the location.' After thirteen years of collaborative research, an Australian company – Florigene, and a Japanese company – Suntory, created a blue rose (actually lavender or mauve) in 2004.The genetic engineering involved three alterations – adding two genes, and interfering with another. One of the added genes was for the blue plant pigment delphinidin cloned from the pansy. The researchers then used RNA interference (RNAi) technology to depress all colour production by endogenous genes by blocking a crucial protein in colour production, called dihydroflavonol 4-reductase (DFR), and adding a variant of that protein that would not be blocked by the RNAi but that would allow the delphinidin to work. The roses are sold in Japan, the United States, and Canada. Florigene has also created and sells lavender-coloured

carnations that are genetically engineered in a similar way. Simple plants and plant cells have been genetically engineered for production of biopharmaceuticals in bioreactors as opposed to cultivating plants in open fields. Work has been done with duckweed Lemna minor, the algae *Chlamydomonas reinhardtii* and the moss *Physcomitrella patens*. An Israeli company, Protalix, has developed a method to produce therapeutics in cultured transgenic carrot and tobacco cells. Protalix and its partner, Pfizer, received FDA approval to market its drug Elelyso, a treatment for Gaucher's disease, in 2012.

c. Invertebrates:

Fruit flies

In biological research, transgenic fruit flies (Drosophila melanogaster) are model organisms used to study the effects of genetic changes on development. Fruit flies are often preferred over other animals due to their short life cycle, low maintenance requirements, and relatively simple genome compared to many vertebrates.

Mosquitoes

In 2010, scientists created "malaria-resistant mosquitoes" in the laboratory. The World Health Organization estimated that malaria killed almost one million people in 2008. Genetically modified male mosquitoes containing a lethal gene have been developed to combat the spread of dengue fever and the Zika virus. Aedes aegypti mosquitoes, the single most important carrier of dengue fever and the Zika virus, were reduced by 80% in a 2010 trial of these GM mosquitoes in the Cayman Islands and by 90% in a 2015 trial in Bahia, Brazil. In comparison, the Florida Keys Mosquito Control District has achieved only 30–60% population reduction with traps and pesticide spraying. In 2016 FDA approved a genetically modified mosquito intervention for Key West, Florida. UK firm Oxitec proposed the release of millions of modified male (non-biting) mosquitoes to compete with wild males for mates. The males are engineered so that their offspring die before maturing, helping to eradicate mosquito-borne disease. Final approval was to be based on a local referendum to be held in November. Andrea Crisanti, a molecular biologist at Imperial College in London is working on ways to stop the *A. gambiae* mosquito from transmitting disease.

Bollworms

A strain of *Pectinophora gossypiella* (Pink bollworm) has been genetically engineered to express a red fluorescent protein. This allows researchers to monitor bollworms that have been sterilized by radiation and released to reduce bollworm infestation. The strain has been field tested for over three years and has been approved for release.

Cnidaria

Cnidaria such as Hydra and the sea anemone *Nemato stellavectensis* are attractive model organisms to study the evolution of immunity and certain developmental processes. An important technical breakthrough was the development of procedures for generation of stable transgenic hydras and sea anemones by embryo microinjection.

d. Chordates :

Fishes:

Genetically modified fish are used for scientific research and as pets, and are being considered for use as food and as aquatic pollution sensors.GM fish are widely used in basic research in genetics and development. Two species of fish, zebrafish and medaka, are most commonly modified because they have optically clear chorions (membranes in the egg), rapidly develop, and the 1-cell embryo is easy to see and microinject with transgenic DNA.

The GloFish is a patented brand of genetically modified (GM) fluorescent zebrafish with bright red, green, and orange fluorescent color. Although not originally developed for the ornamental fish trade, it became the first genetically modified animal to become publicly available as a pet when it was introduced for sale in 2003. They were quickly banned for sale in California.

GM fish have been developed with promoters driving an over-production of "all fish" growth hormone for use in the aquaculture industry to increase the speed of development and potentially reduce fishing pressure on wild stocks. This has resulted in dramatic growth enhancement in several species, including salmon, trout and tilapia. AquaBounty Technologies, a biotechnology company working on bringing a GM salmon to market, claims that their GM AquAdvantage salmon can mature in half the time as wild salmon. AquaBounty applied for regulatory approval to market their GM salmon in the US, and was approved in November 2015. On 25 November 2013 Canada approved commercial scale production and export of GM Salmon eggs but they are not approved for human consumption in Canada. Several academic groups have been developing GM zebrafish to detect aquatic pollution. The lab that originated the GloFish discussed above originally developed them to change colour in the presence of pollutants, to be used as environmental sensors. A lab at University of Cincinnati has been developing GM zebrafish for the same purpose, as has a lab at Tulane University. Recent research on pain in fish has resulted in concerns being raised that genetic-modifications induced for scientific research may have detrimental effects on the welfare of fish.

Amphibians:

Genetically modified frogs are used for scientific research and are widely used in basic research including genetics and early development. Two species of frog, *Xenopus laevis* and *Xenopus tropicalis*, are most commonly used. GM frogs are also being used as pollution sensors, especially for endocrine disrupting chemicals.

Mammals:

Genetically modified mammals are an important category of genetically modified organisms. Ralph L. Brinster and Richard Palmiter developed the techniques responsible for transgenic mice, rats, rabbits, sheep, and pigs in the early 1980s, and established many of the first transgenic models of human disease, including the first carcinoma caused by a transgene. The process of genetically engineering animals is a slow, tedious, and expensive process. However, new technologies are making genetic modifications easier and more precise. The first transgenic (genetically modified) animal was produced by injecting DNA into mouse embryos then implanting the embryos in female mice. Genetically modified animals currently being

developed can be placed into six different broad classes based on the intended purpose of the genetic modification:

1. to research human diseases (for example, to develop animal models for these diseases);
2. to produce industrial or consumer products (fibres for multiple uses);
3. to produce products intended for human therapeutic use (pharmaceutical products or tissue for implantation);
4. to enrich or enhance the animals' interactions with humans (hypo-allergenic pets);
5. to enhance production or food quality traits (faster growing fish, pigs that digest food more efficiently);
6. to improve animal health (disease resistance)

Gene Knockouts:

This technique is used to study the function of the gene by making it nonfunctional. The principle behind this is simple. Under normal condition a particular gene is doing its function by the production of RNA from transcription and, finally the protein by translation. By inactivating the gene (gene knockout), we are able to switch-off the gene and the phenotype of the organism can be studied in the absence of the product made from that particular gene.

There are various ways to knockout a gene, by disrupting the gene in the genome, by deleting the whole or part of the gene, or by inserting an additional DNA in the gene, which act as an insulator in the transcription. These methods of Gene knockout are now becoming very powerful tools in the study of the genome and also the function of individual genes.

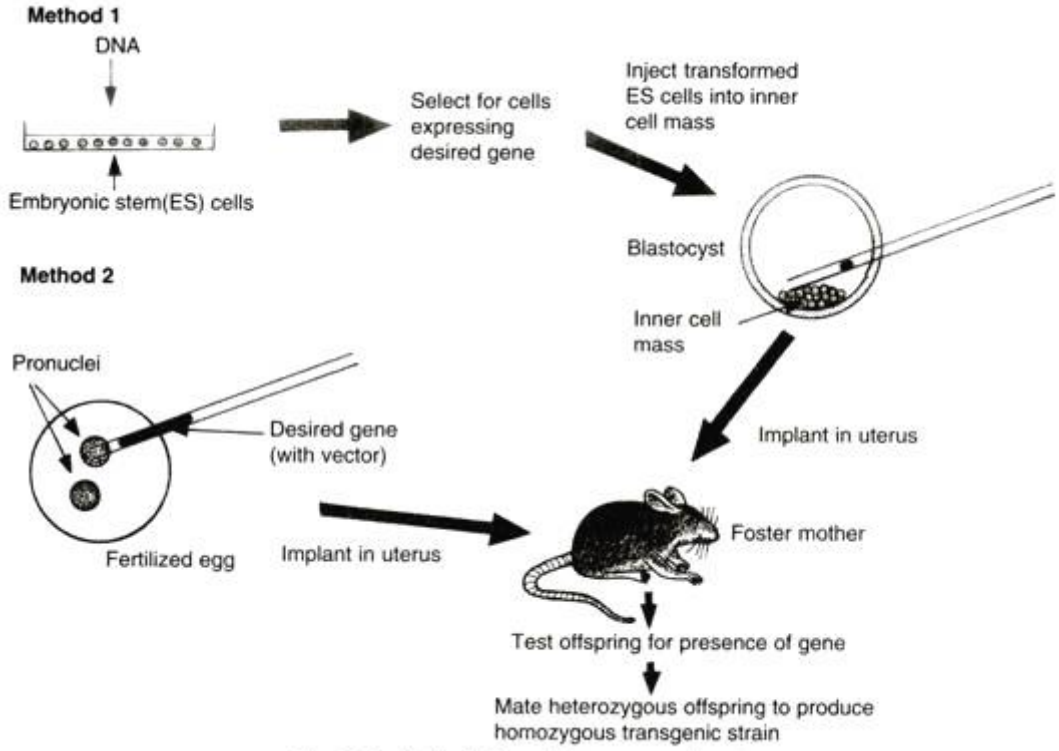


Fig. 18.4. Method of producing transgenic mice.

Mouse disease models:

Because of their phylogenetic relatedness and physiological similarity to humans, the ease of maintaining and breeding them in the laboratory, and the availability of many inbred strains, house mice, Mus musculus, have long served as models of human biology and disease. Genomic studies have highlighted the striking genetic homologies between the two species. These studies, together with the development of methods for the creation of transgenic, knockout, and knock-in mice, have provided added impetus and powerful tools for mouse research, and have led to a dramatic increase in the use of mice as model organisms. Studies on mice have contributed immeasurably to our understanding of human biology. All too often, however, mice respond to experimental interventions in ways that differ markedly from humans. Endostatin, the anticancer drug alluded to in the epigraph, is but one of many treatments that cure cancer in mice but have limited effectiveness in humans. Indeed, the majority of oncology drugs that enter clinical trials never reach the marketplace. There are many reasons for the high failure rate of drug development, but the limitations of the animal models used in drug testing are an important factor. Many substances that are carcinogens in mice are not carcinogenic in humans—and vice versa. Moreover, mouse strains that were created to mimic human genetic diseases frequently have phenotypes that differ from their human counterparts. Because of the assumption that mice will serve as reliable models for humans, differences between the two species are often greeted with surprise as well as dismay. But these differences should not elicit surprise; indeed, they should be expected. The lineages leading to modern rodents and primates are thought to have diverged from a common ancestral species that lived some 85 million years ago. Since that time, species in these lineages evolved in and became adapted to very different environments.

The use of model organisms in biological research is based on the concept of unity in biology, a concept expressed most famously in Jacques Monod and François Jacob's aphorism, "Anything found to be true of E. coli must also be true of elephants" . But biology is characterized by diversity as well as unity; evolution is "descent with modification" . The art of choosing model organisms involves recognizing the properties of these organisms that they are likely to share with organisms of other species—especially, for biomedical research, humans. Monod and Jacob were concerned with genetic regulatory mechanisms and other basic biological processes that must have arisen very early in the evolutionary history of living organisms and so are similar in bacteria and in mammals. Mice have served and will continue to serve as valuable models for the study of basic biological processes that, in Wimsatt's terms, became developmentally entrenched before the rodent and primate lineages diverged and have been conserved during the separate evolutionary histories of mice and humans.

Studies of the immune system highlight both the value of mouse research in elucidating common features of mammalian biology as well as the limitations of translating this research in areas in which humans are likely to differ from mice. Research on mice has contributed greatly to our knowledge of the adaptive immune system; mouse research has led to the discovery of the major histocompatibility complex genes and the T cell receptor, and to our understanding of the regulation of antibody synthesis and many other features of the immune system. But there are many differences between the mouse and human immune systems, such that much research on immunological diseases in mice is not transferable to humans, and many immunologists are now calling for a return to the study of human immunology. From an

evolutionary perspective, this is understandable. The adaptive immune system evolved in jawed fish some hundreds of million years before the evolution of mammals. Many features of this ancestral immune system, including antigen recognition, generation of antibody diversity, clonal selection, and immunological tolerance, are critical for survival and have been maintained in most or all of the descendants of these early vertebrates. On the other hand, species differences in the mechanisms for the maintenance of memory T cells must have evolved in response to the evolution of different life spans. Moreover, specific features of the immune system evolve rapidly, as host species coevolve with their pathogens and commensal microbiota . Since humans and mice harbour different sets of pathogens and microbiomes, it is not surprising that host–pathogen and host–microbiome coevolution has led to differences between the human and mouse immune systems.

The fact that the highly conserved mammalian genome can give rise to a wide variety of different species indicates that the relationships between genotype and phenotype differ among mammalian species. Comparisons between mice and humans are invaluable for understanding the developmental mechanisms that lead to such different genotype–phenotype relationships. Some of the genetic differences between mice and humans are differences in coding sequences, which give rise to proteins with different properties. For example, mouse haemoglobin has a lower affinity for O2 than does human haemoglobin, which facilitates the dissociation of O2 from haemoglobin in peripheral tissues and helps to support the higher metabolic rate in mice. Perhaps more importantly, however, are differences in the genetic or epigenetic regulation of gene expression in these species. The expression of potassium channel genes in the heart exemplifies these differences. Mice have a heart rate of ~600 beats/min, while humans have a resting heart rate of ~70 beats/min. This difference in heart rate entails that the cardiac action potential be much shorter in mice than in humans. Indeed, the repolarization phase of the cardiac action potential, which is due to outward K+ currents, is much shorter in mice. This difference is due to different contributions of various K+ currents, which in turn are presumably due to differences in expression of K+ channel genes in the two species. Evolved differences in the regulation of gene expression are important because they may involve the rewiring of gene (or protein) networks. Gene networks in mice and humans have similar numbers of nodes (genes) but the connectivity of the nodes in these networks, and the relationships between genes and phenotypes, differ between the two species. The different network architectures and different genotype–phenotype relationships between mice and humans mean that the relationships between genotype and disease are also likely to differ in these two species. Perturbations of gene and protein networks by environmental manipulation as well as by mutation are likely to have different effects on diseases as well as on other phenotypes in mice than in humans. In short, mice are problematic models for understanding human disease. There are other good reasons to pursue research on mice. Although house mice are not a major source of human disease, they can transmit lymphocytic choriomeningitis virus and perhaps other pathogens to humans, and other rodent species are important reservoirs for zoonoses. Research on mice may yield information that will help to prevent or ameliorate these diseases. Finally, mice should be studied for their own sake, to understand their biology and to maintain the health of pet mice, laboratory mice, and wild mice.

Unfortunately, despite the many attempts to translate the results of mouse research to humans, we still cannot specify in advance which research in mice is likely to benefit or shed light on human biology and health. For the most part, we have only anecdotal information about studies in mice that translated to humans and those that did not. We need more systematic collection, reporting and analysis of mouse research (and research on other "model organisms") to figure out what works and what does not. Until we have that information, we need to be more critical in pursuing mouse research and in making claims about the applicability of this research to humans.

In addition to problems resulting from the evolved differences between mice and humans, other aspects of mouse research have compromised the value of this research and have further complicated the extrapolation of mouse research to humans. Thus, e.g., laboratory mice are often housed at temperatures below their thermoneutral zone, and as a result are cold-stressed, sleep deprived, and hypertensive. The biology of laboratory mice may also be affected by their housing in same-sex groups and their lack of opportunities for physical exercise. Although mice are often used as models of diseases of aging, for logistical and financial reasons most mouse research is carried out on young animals. And although mouse cells are more sensitive to oxygen damage than are human cells, cell culture studies are often carried out in 20% oxygen, which is non-physiological and is more damaging to mouse cells than to human cells. Finally, there are no agreed upon standards for the design, analysis, or publication of mouse research (or research with other model organisms). The statistical analysis of studies of mice and other animals is often substandard, and there may be important publication biases because negative results may not get published. All of these problems need to be addressed before studies on mice can be properly interpreted and extrapolated to humans.

Despite all of the documented differences between mice and humans, and despite the history of "errors in translation" in the application of research on mice to humans, reports of research on mice are frequently accompanied by unwarranted and misleading claims, such as "Furthering our understanding of mouse X should provide novel insights into human Y." Such claims raise false hopes and are ultimately self-defeating, in that they waste resources and increase public skepticism concerning the value of biomedical research. Indeed, the problems of translating research on mice and other model organisms to humans have led a number of scientists to question the value of this research . Furthermore, critical discussions of animal experimentation are routinely distorted by "animal rights" activists to support their belief that this experimentation should be stopped. These intrusions, however unwelcome, should not stifle discussion. For reasons mentioned above, research on mice (and other species) is essential and should be supported. This research should, however, be designed and interpreted with appropriate appreciation of the evolved differences as well as the similarities between *M. musculus* and *H. sapiens*.

Gene Silencing :

Gene silencing is the regulation of gene expression in a cell to prevent the expression of a certain gene. Gene silencing can occur during either transcription or translation and is often used in research. In particular, methods used to silence genes are being increasingly used to produce therapeutics to combat cancer and diseases, such as infectious diseases and neurodegenerative disorders.

Gene silencing is often considered the same as gene knockdown. When genes are silenced, their expression is reduced. In contrast, when genes are knocked out, they are completely erased from the organism's genome and, thus, have no expression. Gene silencing is considered a gene knockdown mechanism since the methods used to silence genes, such as RNAi, CRISPR, or siRNA, generally reduce the expression of a gene by at least 70% but do not completely eliminate it. Methods using gene silencing are often considered better than gene knockouts since they allow researchers to study essential genes that are required for the animal models to survive and cannot be removed. In addition, they provide a more complete view on the development of diseases since diseases are generally associated with genes that have a reduced expression.

Transcriptional Gene Silencing (TGS):

Transcriptional gene silencing (TGS) is a gene silencing mechanism in which inactivation of (trans) gene specific RNA synthesis takes place. These are predominantly observed in transgenic plants containing multiple copies of homologous transgene or endogenous gene.

TGS is characterized by the sequence specific DNA methylation at promoter region. TGS is believed to occur through direct DNA interaction between silencer transgene locus and other loci exhibit homologous sequence in their promoter region. In transgenic plants, it has been demonstrated that de-novo methylation of the transgene promoter sequence results in transcriptional inactivation. The pairing of DNA-DNA probably results in methylation and in turn occurs gene silencing. Although methylation at promoter region is significance of TGS, it is however, presumed that methylation do not alone suppress transcription. DNA methylation probably induces the transcriptional silencing through chromatin components. The link between DNA methylation and transcriptional inactivation is aided by repressive protein MeCP2 which specifically binds to transgene region. Some proteins of the repressive complex initiates de novo DNA methylation. Increase in methylation is followed by acquiring condensed chromatin structure by transcriptionally silent transgene. This type of chromatin remodelling is believed to be responsible for maintenance of repressive status and in the propagation of non-symmetrical methylation pattern in plants.

Post Transcriptional gene silencing:

Production of dsRNA from inverted transgene repeats and single copy transgene triggers PTGS. It was proposed that dsRNA would be act as template for an RNA-directed RNA polymerase. At least four models have been proposed to explain induction of PTGS. First model is known as threshold model explains multicopy insertion or highly expressed single gene in the threshold concentration.

According to this model, plant cells exhibit surveillance system that can detect mRNA expressed above as certain concentration and acts as templates for RDRP. The viral resistant plants contains high number of transgene copies which results in the transcription of high amount of transgene mRNA. This high concentration of mRNA is easily sensed by plant cells and initiate sequence specific process of transgene mRNA degradation. Second model transgenes are inserted as multi copies. Which trigger ectopic pairing and in turn induce methylation. The whole process is culminated in premature termination of transcription.

Production of this type of RNA known as aberrant RNA. The truncated RNA transcript acts as template for RDRP. The third model envisages of inverted repeat transgene integration duplex formed by the mRNA (self-complementation) is template for RDRP (Fig. 20.3). Analysis of transgene organisation reveals that silencing occurs in plants containing inverted-repeats. (IR) of the transgene. The palindromic sequence formed by IR is able to activate production of aberrant RNA (abRNA) from endogenous homologous via ectopic pairing. Recent studies on PTGS of nitrate reductase (NR) genes strongly suggested that the active transcriptional status of NR genes is indispensable for co-suppression of NR (trans) in transgenic tobacco plant.

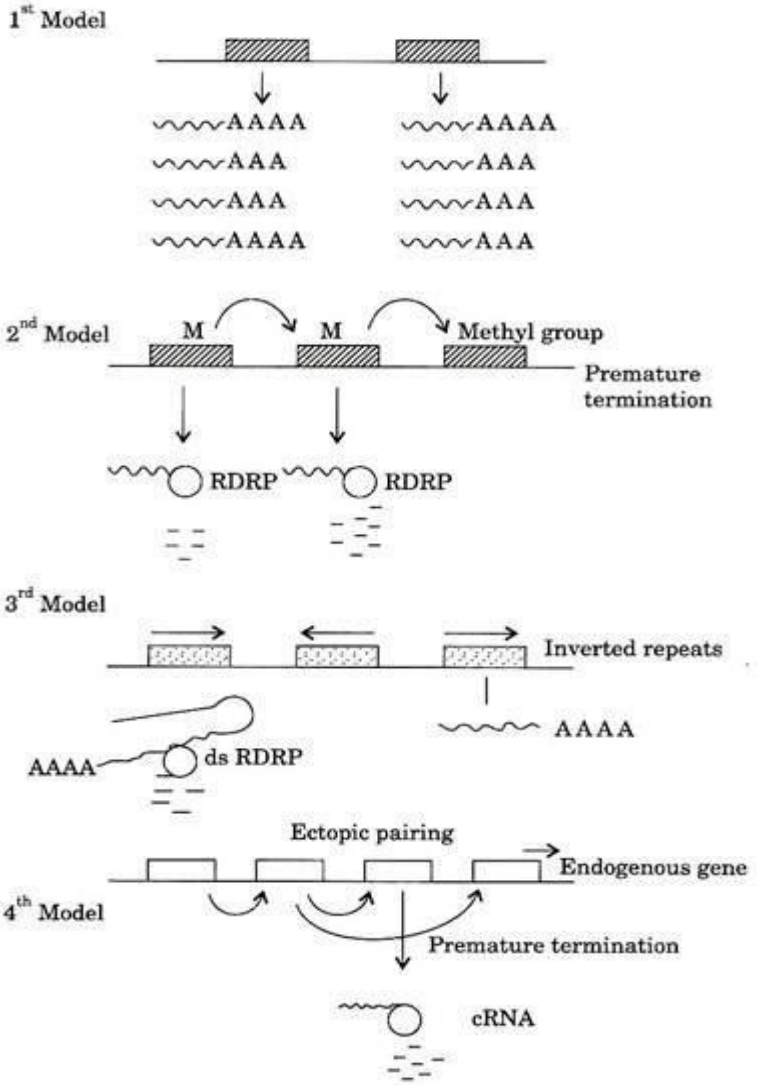


Fig. 20.3 Various models for PTGS

In the last and forth model modification (or methylation) is done to the endogenous gene by ectopic pairing and methylate transgene in inverted repeat and an endogenous gene. Thus modified endogenous gene produce truncated mRNA, which acts as template for RDRP.

Once the transgene mRNA produced. They are potentiating to silence same sense endogenous gene transcript as viral genomic RNA. This is due to plant encoded RNA dependent RNA

polymerase which produces small RNA molecules (cRNA) is complementary for transgenic mRNA. The small cRNA potentiate the degradation of the target RNA. Hybridization occurs between their two and target for their degradation mediated by dsRNA or the endonucleases that cleaves ssRNA, which is present adjacent to the dsRNA duplex. The co-suppression takes place by sequence-specific RNA degradation within cytoplasm suggested that PTGS related RNA degradation takes place within cytoplasm.

The possibility of gene silencing by dsRNA has been studied for both virus resistance and co-suppression. In one of the case studies plant expressing transgene sense mRNA derived from virus crossed with plant expressing antisense mRNA of the transgene. Each parent is actually susceptible to virus infection because it contains single hemizygous transgene. All progeny that inherent both sense and to antisense transgenes are resistant to the virus. The PTGS absolutely requires dsRNA and therefore recognise only mRNA that has larger regions of self-complementary. These are produced by read through transcription of transgene in an inverted repeat configuration. The small cRNA produced by RDRP can easily escape the nucleus to potentiate the degradation of cytoplasmic RNA (Fig. 20.4).

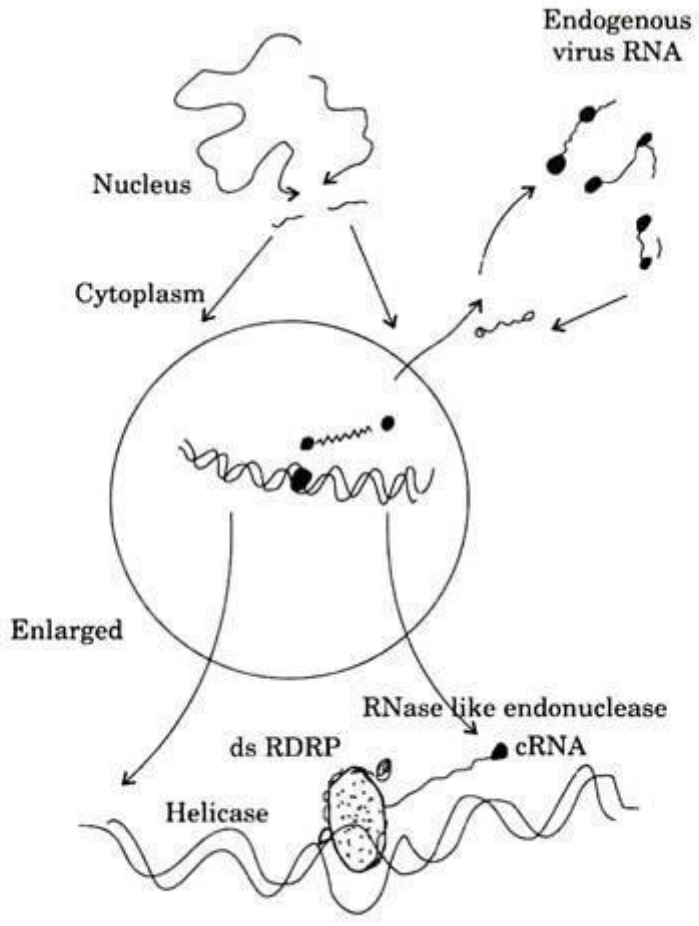


Fig. 20.4 Model for dsRNA induced PTGS (After waterhouse *et al*., 1998, Proc. Natl. Acad. Sci. USA.)

Execution of post-transcriptional gene silencing of endogenous or reporter gene have been described in transgenic plants containing sense or antisense transgenes. In these transgenic plants, either co-suppression or antisense appears to be induction of surveillance system with-in plant that specifically degrades both transgene and target RNA.

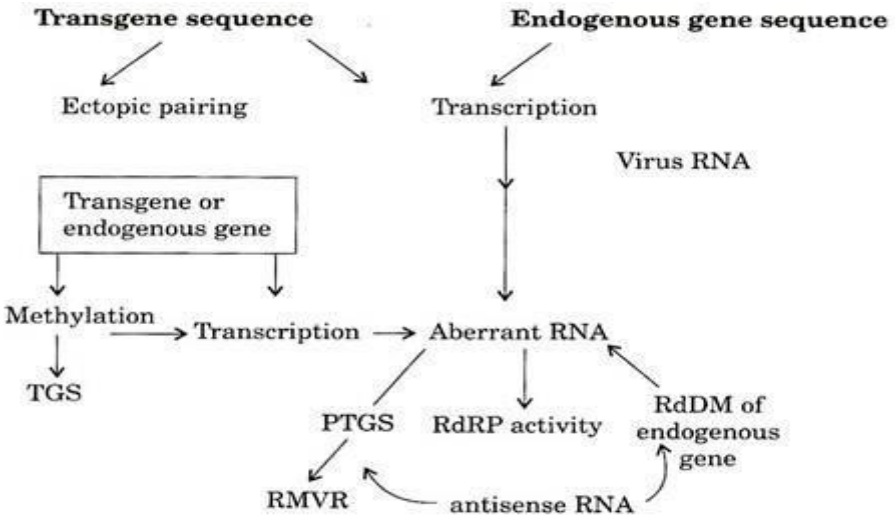


Fig. 20.5 Gene Silencing mechanism of TGS and PTGS

Table 20.1 Relationship between various types of homology-dependent gene silencing and host defence response to invasive sequence (After Kooter *et al.*, 1999)

Types of HDGS	Cell compartment	Invasive sequence	Molecular effect	Possible triggers
TGS	Nucleus	Transposable elements (TE)	DNA hyper methylation	DNA-DNA pairing : IRS
		Retroelement	DNA hyper methylation	(Abb) RNA-DNA
		viroids	DNA hyper	Sequence incompa-
		bacterial DNA	methylation	tibility
		viruses		
		RNA genome or		
PTGS	Cytoplasm	RNA replication intermediate of DNA genome	RNA turn over	(Ab) RNA or over-express RNA

Gene Therapy:

There are two types of gene therapies:

I. Ex vivo gene therapy:

This involves the transfer of genes in cultured cells (e.g., bone marrow cells) which are then reintroduced into the patient.

II. In vivo gene therapy:

The direct delivery of genes into the cells of a particular tissue is referred to as in vivo gene therapy

Type - I. Ex Vivo Gene Therapy:

The ex vivo gene therapy can be applied to only selected tissues (e.g., bone marrow) whose cells can be cultured in the laboratory. The technique of ex vivo gene therapy involves the following steps (Fig. below).

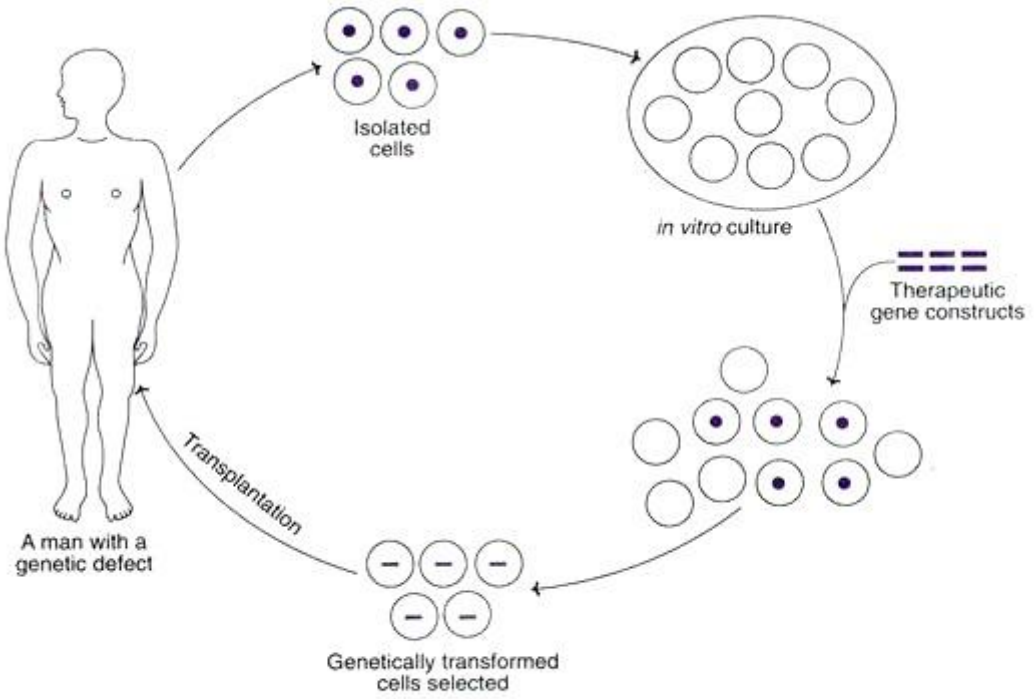


Fig. 13.2 : The procedure for ex vivo gene therapy.

1. Isolate cells with genetic defect from a patient.

2. Grow the cells in culture.

3. Introduce the therapeutic gene to correct gene defect.

4. Select the genetically corrected cells (stable trans-formants) and grow.

5. Transplant the modified cells to the patient.

The procedure basically involves the use of the patient's own cells for culture and genetic correction, and then their return back to the patient. This technique is therefore, not associated with adverse immunological responses after transplanting the cells. Ex vivo gene therapy is efficient only, if the therapeutic gene (remedial gene) is stably incorporated and continuously expressed. This can be achieved by use of vectors.

Vectors in Gene Therapy:

The carrier particles or molecules used to deliver genes to somatic cells are referred to as vectors. The important vectors employed in ex vivo gene therapy are listed below and briefly described next.

i. Viruses
ii. Human artificial chromosome
iii. Bone marrow cells.

i. Viruses:

The vectors frequently used in gene therapy are viruses, particularly retroviruses. RNA is the genetic material in retroviruses. As the retrovirus enters the host cell, it synthesizes DNA from RNA (by reverse transcription). The so formed viral DNA (referred to as provirus) gets incorporated into the DNA of the host cell.

The proviruses are normally harmless. However, there is a tremendous risk, since some of the retroviruses can convert normal cells into cancerous ones. Therefore, it is absolutely essential to ensure that such a thing does not happen.

Making retroviruses harmless:

Researchers employ certain biochemical methods to convert harmful retroviruses to harmless ones, before using them as vectors. For instance, by artificially removing a gene that encodes for the viral envelope, the retrovirus can be crippled and made harmless. This is because, without the envelope, retrovirus cannot enter the host cell. The production of a large number (billions) of viral particles can be achieved, starting from a single envelope defective retrovirus
.

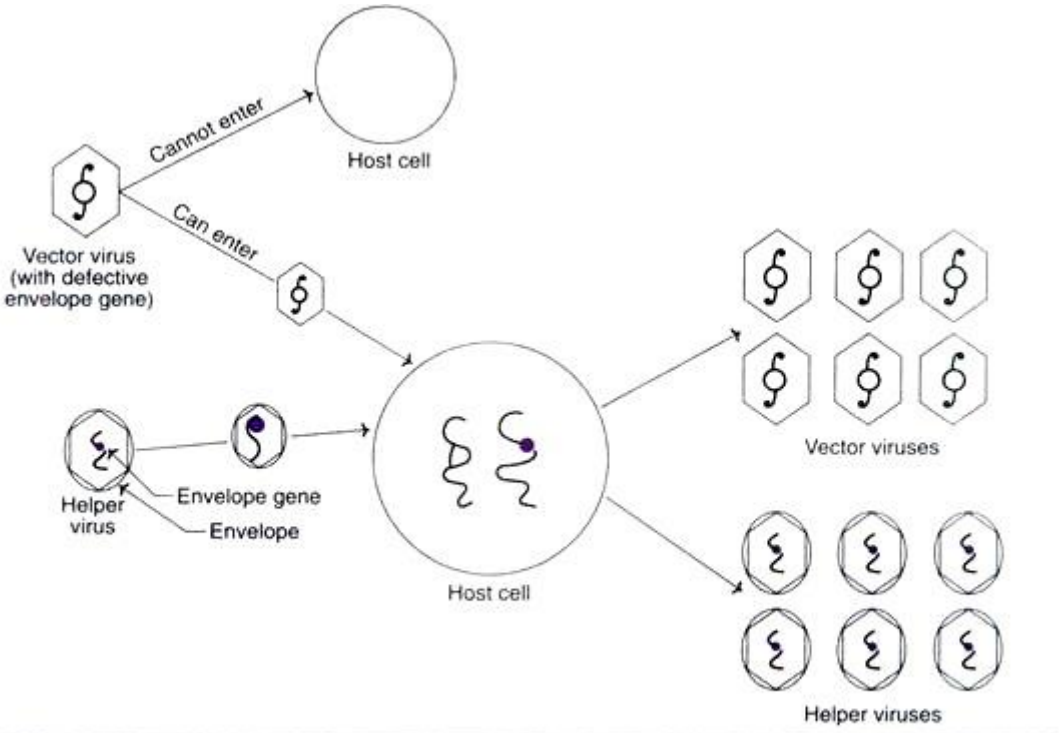


Fig. 13.3 : Large scale production of vector viruses by using helper viruses.

This is made possible by using helper viruses which contain normal gene for envelope formation. Along with the helper virus, the vector (with defective envelope gene) can enter the host cell and both of them multiply. By repeated multiplication in host cells, billions of vector and helper viruses are produced.

The vector viruses can be separated from the helper viruses and purified. Isolation of vector viruses, totally free from helper viruses, is absolutely essential. Contamination of helper viruses is a big threat to the health of the patients undergoing gene therapy.

Retroviruses in gene therapy:

The genetic map of a typical retrovirus is depicted in Fig. 13.4A. In general, the retrovirus particle has RNA as a genome organized into six regions. It has a 5′-long terminal repeat (5′-LTR), a non-coding sequence required for packaging RNA designated as psi (Ψ), a gene gag coding for structural protein, a gene pol that codes for reverse transcriptase, a gene env coding for envelope protein and a 3-LTR sequence.

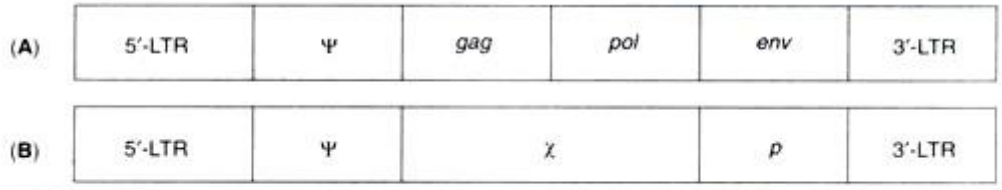


*Fig. 13.4 : A retrovirus used in gene therapy. (**A**) General map of a typical retrovirus (**B**) Gene map of a modified retrovirus for use in gene therapy (LTR-Long terminal repeat; Ψ-Packaging signal sequence;* gag-*Coding sequence for structural protein;* pol-*Coding sequence for reverse transcriptase;* env-*Envelope protein coding sequence; χ-Therapeutic gene; p–Promoter gene).*

For use of a retrovirus as a vector, the structural genes gag and pol are deleted. These genes are actually adjacent to Ψ region. In addition, a promoter gene is also included (Fig. 13.4B). This vector design allows the synthesis of cloned genes. A retroviral vector can carry a therapeutic DNA of maximum size of 8 kb.

A retroviral vector DNA can be used to transform the cells. However, the efficiency of delivery and integration of therapeutic DNA are very low. In recent years, techniques have been developed to deliver the vector RNA to host cells at a high frequency. For this purposes, packaged retroviral RNA particles are used. This technique allows a high efficiency of integration of pharmaceutical DNA into host genome.

Several modified viral vectors have been developed in recent years for gene therapy. These include oncoretrovirus, adenovirus, adeno-associated virus, herpes virus and a number of hybrid vectors combining the good characters of the parental vectors.

Murine leukaemia viruses in gene therapy:

This is a retrovirus that causes a type of leukaemia in mice. It can react with human cells as well as the mouse cells, due to a similarity in the surface receptor protein. Murine leukaemia virus (MLV) is frequently used in gene transfer.

AIDS virus in gene therapy

It is suggested that the human immunodeficiency virus (HIV) can be used as a vector in gene transfer. But this is bound to create public uproar. Some workers have been successful in creating a harmless HIV (crippled HIV) by removing all the genes related to reproduction. At the same time, the essential genes required for gene transfer are retained. There is a distinct advantage with HIV when compared with MLV. MLV is capable of bringing out gene transfer only in dividing cells. HIV can infect even non-dividing cells (e.g., brain cells) and do the job of gene transfer effectively. However, it is doubtful whether HIV can ever be used as a vector.

ii. Human Artificial Chromosome:

The details of human artificial chromosome (HAC) are described elsewhere .HAC is a synthetic chromosome that can replicate with other chromosomes, besides encoding a human protein. As already discussed above, use of retroviruses as vectors in gene therapy is associated with a heavy risk. This problem can be overcome if HAC is used. Some success has been achieved in this direction.

iii. Bone Marrow Cells:

Bone marrow contains totipotent embryonic stem (ES) cells. These cells are capable of dividing and differentiating into various cell types (e.g., red blood cells, platelets, macrophages, osteoclasts, B- and T-lymphocytes). For this reason, bone marrow transplantation is the most widely used technique for several genetic diseases.

And there is every reason to believe that the genetic disorders that respond to bone marrow transplantation are likely to respond to ex vivo gene therapy also. For instance, if there is a gene mutation that interferes with the function of erythrocytes (e.g., sickle-cell anaemia), bone

marrow transplantation is done. Bone marrow cells are the potential candidates for gene therapy of sickle-cell anaemia. However, this is not as simple as theoretically stated.

Selected Examples of Ex Vivo Gene Therapy:

Therapy for Adenosine Deaminase Deficiency:

The first and the most publicized human gene therapy was carried out to correct the deficiency of the enzyme adenosine deaminase (ADA). This was done on September 14, 1990 by a team of workers led by Blaese and Anderson at the National Institute of Health, USA (The girl's name is Ashanti, 4 years old then).

Severe combined immunodeficiency (SCID):

This is rare inherited immune disorder associated with T-lymphocytes, and (to a lesser extent) B-lymphocytes dysfunction. About 50% of SCID patients have a defect in the gene (located on chromosome 20, and has 32,000 base pairs and 12 exons) that encodes for adenosine deaminase. In the deficiency of ADA, deoxyadenosine and its metabolites (primarily deoxyadenosine 5′-triphosphate) accumulate and destroy T-lymphocytes.

T-Lymphocytes are essential for body's immunity. Besides participating directly in body's defence, they promote the function of B-lymphocytes to produce antibodies. Thus, the patients of SCID (lacking ADA) suffer from infectious diseases and die at an young age. Previously, the children suffering from SCID were treated with conjugated bovine ADA, or by bone marrow transplantation.

Technique of therapy for ADA deficiency:

The general scheme of gene therapy adopted for introducing a defective gene in the patient has been depicted in Fig 13.2. The same procedure with suitable modifications can also be applied for other gene therapies.

A plasmid vector bearing a pro-viral DNA is selected. A part of the pro-viral DNA is replaced by the ADA gene and a gene (G 418) coding for antibiotic resistance, and then cloned. The antibiotic resistance gene will help to select the desired clones with ADA gene. A diagrammatic representation of the treatment of ADP deficient patient is depicted in Fig. below.

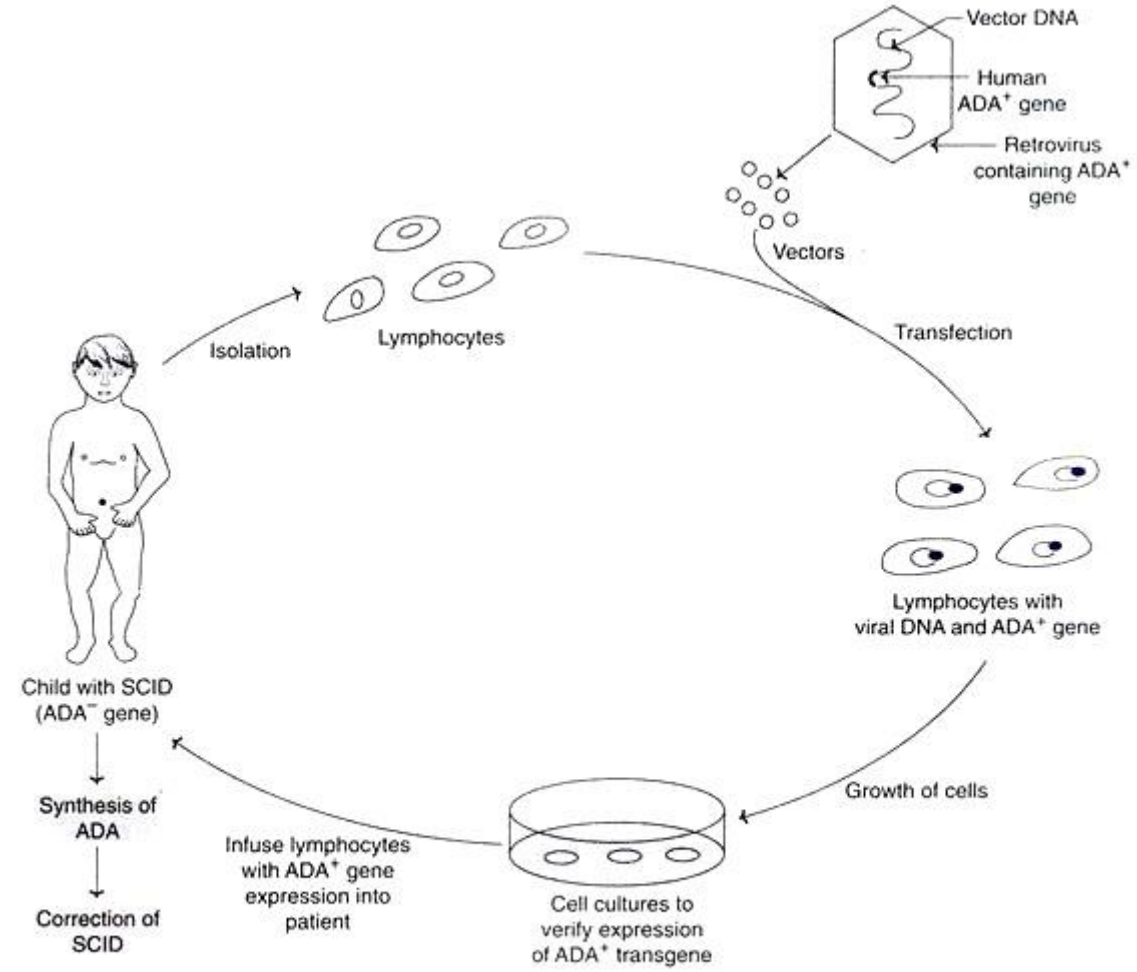


Fig. 13.5 : Treatment of adenosine deaminase (ADA) deficient patient by somatic ex vivo gene therapy (SCID-Severe combined immunodeficiency disease).

Circulating lymphocytes are removed from a patient suffering from ADA deficiency. These cells are transfected with ADA gene by exposing to billions of retroviruses carrying the said gene. The genetically-modified lymphocytes are grown in cultures to confirm the expression of ADA gene and returned to the patient. These lymphocytes persist in the circulation and synthesize ADA. Consequently, the ability of the patient to produce antibodies is increased. However, there is a limitation. The lymphocytes have a short life span (just live for a few months), hence the transfusions have to be carried out frequently.

Transfer of ADA gene into stem cells:

In 1995, ADA gene was transferred into the stem cells, obtained from the umbilical cord blood, at the time of baby's delivery. Four days after birth, the infant received the modified cells back. By this way, a permanent population of ADA gene producing cells was established.

Therapy for Familial Hypercholesterolemia:

The patients of familial hypercholesterolemia lack the low density lipoprotein (LDL) receptors on their liver cells. As a result, LDL cholesterol is not metabolised in liver. The accumulated LDL- cholesterol builds up in the circulation, leading to arterial blockage and heart diseases.

Attempts are being made by gene therapists to help the victims of familial hypercholesterolemia. In fact, there is some success also. In a woman, 15% of the liver was removed. The hepatocytes were transduced with retroviruses carrying genes for LDL receptors. These genetically modified hepatocytes were infused into the patient's liver.

The hepatocytes established themselves in the liver and produced functional LDL-receptors. A significant improvement in the patient's condition, as assessed by estimating the lipid parameters in blood, was observed. Further, there were no antibodies produced against the LDL-receptor molecules, clearly showing that the genetically modified liver cells were accepted.

Therapy for Lesch-Nyhan Syndrome:

Lesch-Nyhan syndrome is an inborn error in purine metabolism due to a defect in a gene that encodes for the enzyme hypoxanthine-guanine phosphoribosyl transferase (HCPRT). In the absence of HGPRT, purine metabolism is disturbed and uric acid level builds up, resulting in severe gout and kidney damage. The victims of Lesch- Nyhan syndrome exhibit symptoms of mental retardation, besides an urge to bite lips and fingers, causing self-mutilation.

By using retroviral vector system, HGPRT producing genes were successfully inserted into cultured human bone marrow cells. The major problem in humans is the involvement of brain. Experiments conducted in animals are encouraging. However, it is doubtful whether good success can be achieved by gene therapy for Lesch-Nyhan syndrome in humans, in the near future.

Therapy for Haemophilia:

Haemophilia is a genetic disease due lack of a gene that encodes for clotting factor IX. It is characterized by excessive bleeding. By using a retroviral vector system, genes for the synthesis of factor IX were inserted into the liver cells of dogs. These dogs no longer displayed the symptoms of haemophilia.

Ex Vivo Gene Therapy with Non-Autologous Cells:

The ex vivo gene therapies described above are based on the transplantation of genetically modified cells for the production of desired proteins. However, there are several limitations in using the patient's own cells (autologous cells) for gene therapy. These include lack of enough cells from target tissues, defective uptake of genes and their inadequate expression. To overcome these problems, attempts are on to develop methods to use non-autologous cells (i.e., cells from other individuals or animals). The outline of the procedure is briefly described below.

Tissue-specific cells capable of growing in culture are selected. These include fibroblasts from skin, hepatocytes from liver, and myoblasts from muscle and astrocytes from brain. These cells

are cultured and genetically modified with the therapeutic gene. They are then encapsulated in artificial membrane composed of a synthetic polymer (e.g., polyether sultone, alginase-poly L-lysine-alginate). The polymeric membranes are non-immunogenic, therefore the patient can accept non-autologous encapsulated cells. Further, being semipermeable in nature, these membranes allow the nutrients to enter in, and the encoded protein (by the therapeutic gene) to pass out. Experiments conducted in animals have shown some encouraging results for using non-autologous cells in gene therapy. The encapsulated cells were found to proliferate and produce the required protein. However, the success has been very limited in human trials.

Type - II. In Vivo Gene Therapy:

The direct delivery of the therapeutic gene (DNA) into the target cells of a particular tissue of a patient constitutes in vivo gene therapy (see Fig. below). Many tissues are the potential candidates for this approach. These include liver, muscle, skin, spleen, lung, brain and blood cells. Gene delivery can be carried out by viral or non- viral vector systems. The success of in vivo gene therapy mostly depends on the following parameters

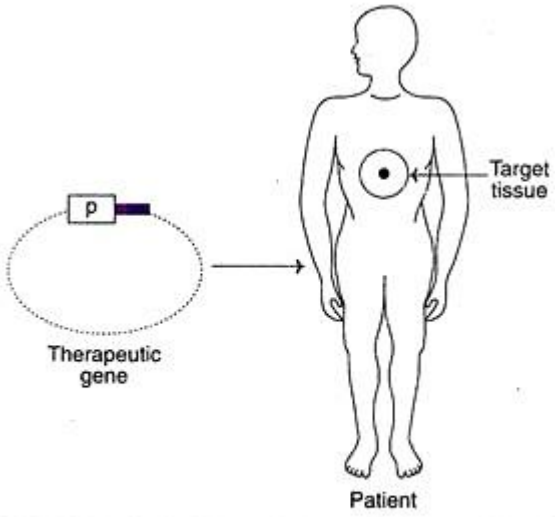


Fig. 13.6 : Diagrammatic representation of in vivo gene therapy. (p-Promoter gene specific for therapeutic gene)

i. The efficiency of the uptake of the remedial (therapeutic) gene by the target cells.
ii. Intracellular degradation of the gene and its uptake by nucleus.
iii. The expression capability of the gene.
In vivo gene therapy with special reference to gene delivery systems (viral, non-viral) with suitable examples is described.

Gene Delivery by Viruses:

Many viral vector systems have been developed for gene delivery. These include retroviruses, adenoviruses, adeno-associated viruses and herpes simplex virus.

Retrovirus vector system:

Replication defective retrovirus vectors that are harmless are being used. A plasmid in association with a retrovirus, a therapeutic gene and a promoter is referred to as plasmovirus. The plasmovirus is capable of carrying a DNA (therapeutic gene) of size less than 3.4 kb. Replication defective virus particles can be produced from the plasmovirus.

As such, for the delivery of genes by retroviral vectors, the target cells must be in a dividing stage. But majority of the body cells are quiescent. In recent years, viral vectors have been engineered to infect non-dividing cells. Further, attempts are on to include a DNA in the retroviral vectors (by engineering env gene) that encodes for cell receptor protein. If this is successfully achieved, the retroviral vector will specifically infect the target tissues.

Adenoviral vector system:

Adenoviruses (with a DNA genome) are considered to be good vectors for gene delivery because they can infect most of the non-dividing human cells. A common cold adenovirus is a frequently used vector. As the target cells are infected with a recombinant adenovirus, the therapeutic gene (DNA) enters the nucleus and expresses itself.

However, this DNA does not integrate into the host genome. Consequently, adenoviral based gene therapy required periodic administration of recombinant viruses. The efficiency of gene delivery by adenoviruses can be enhanced by developing a virus that can specifically infect target cells. This is possible by incorporating a DNA encoding a cell receptor protein.

Adeno-associated virus vector system:

Adeno-associated virus is a human virus that can integrate into chromosome 19. It is a single-stranded, non-pathogenic small DNA virus (4.7 kb). As the adeno-associated virus enters the host cell, the DNA becomes double- stranded, gets integrated into chromosome and expresses.

Adeno-associated viruses can serve as good vectors for the delivery of therapeutic genes. Recombinant viruses are created by using two plasmids and an adenovirus (i.e., helper virus) by a special technique. Some attempts were made to use therapeutic genes for the treatment of the human diseases-hemophilia (for production of blood clotting factor IX) and cystic fibrosis (for synthesis of cystic fibrosis trans membrane regulator protein) by employing adeno-associated viruses.

Therapy for cystic fibrosis:

Cystic fibrosis (CF) is one of the most common (frequency 1: 2,500) and fatal genetic diseases. It is characterized by the accumulation of sticky, dehydrated mucus in the respiratory tract and lungs. Patients of CF are highly susceptible to bacterial infections in their lungs and most of them die before reaching the age of thirty.

In the normal persons the chloride ions of the cells are pushed out through the participation of a protein called cystic fibrosis trans membrane regulator (CFTR). In the patients of cystic fibrosis, the CFTR protein is not produced due to a gene defect. Consequently, the chloride ions concentrate within the cells which draw water from the surroundings. As a result, the respiratory tract and the lungs become dehydrated with sicky mucus, an ideal environment for bacterial infections.

As the defective gene for cystic fibrosis was identified in 1989, researchers immediately started working on gene therapy for this disease. Adenoviral vector systems have been used, although the success has been limited. The major drawback is that the benefits are short-lived, since the adenoviruses do not integrate themselves into host cells. Multiple administration of recombinant adenovirus caused immunological responses that destroyed the cells.

By using adeno-associated virus vector system, some encouraging results were reported in the gene therapy of CF. In the phase I clinical trials with CF patients, the vector persisted for about 70 days and some improvement was observed in the patients. Some researchers are trying to insert CF gene into the developing fetal cells (in experimental animals such as mice) to produce CFTR protein. But a major breakthrough is yet to come.

Herpes simplex virus vector system:

The retroviruses and adenoviruses employed in in vivo gene therapy are engineered to infect specific target cells. There are some viruses which have a natural tendency to infect a particular type of cells. The best example is herpes simplex virus (HSV) type I, which infects and persists in non-dividing nerve cells. HSV is a human pathogen that causes (though rarely) cold sores and encephalitis.

These are a large number of diseases (metabolic, neurodegenerative, immunological, tumors) associated with nervous system. HSV is considered as an ideal vector for in vivo gene therapy of many nervous disorders. The HSV has a double-stranded DNA of about 152 kb length as its genome. About 30 kb of HSV genome can be replaced by a cloned DNA without loss of its basic characteristics (replication, infection, packaging etc.). But there are some technical difficulties in dealing with large-sized DNAs in genetic engineering experiments. Some modified HSV vectors with reduced genomic sizes have been developed.

Most of the work on the gene therapy, related to the use of HSV as a vector, is being conducted in experimental animals. And the results are quite encouraging. HSV vectors could deliver therapeutic genes to the brain and other parts of nervous system. These genes are well expressed and maintained for long periods. More research, however, is needed before going for human trials. If successful, HSV may help to treat many neurodegenerative syndromes such as Parkinson's disease and Alzheimer's disease by gene therapy.

Gene Delivery by Non-Viral Systems:

There are certain limitations in using viral vectors in gene therapy. In addition to the prohibitive cost of maintaining the viruses, the viral proteins often induce inflammatory responses in the host. Therefore, there is a continuous search by researchers to find alternatives to viral vector systems.

Pure DNA constructs:

The direct introduction of pure DNA constructs into the target tissue is quite simple. However, the efficiency of DNA uptake by the cells and its expression are rather low. Consequently, large quantities of DNA have to be injected periodically. The therapeutic genes produce the proteins in the target cells which enter the circulation and often get degraded.

Lipoplexes:

The lipid-DNA complexes are referred to as lipoplexes or more commonly liposomes. They have a DNA construct surrounded by artificial lipid layers. A large number of lipoplexes have been prepared and used. They are non-toxic and non-immunogenic.

The major limitation with the use of lipoplexes is that as the DNA is taken up by the cells, most of it gets degraded by the lysosomes. Thus, the efficiency of gene delivery by lipoplex is very low. Some clinical trials using liposome-CFTR gene complex showed that the gene expression was very short-lived.

DNA-molecular conjugates:

The use of DNA-molecular conjugates avoids the lysosomal breakdown of DNA. Another advantage of using conjugates is that large-sized therapeutic DNAs (> 10 kb) can be delivered to the target tissues. The most commonly used synthetic conjugate is poly-L-lysine, bound to a specific target cell receptor. The therapeutic DNA is then made to combine with the conjugate to form a complex .

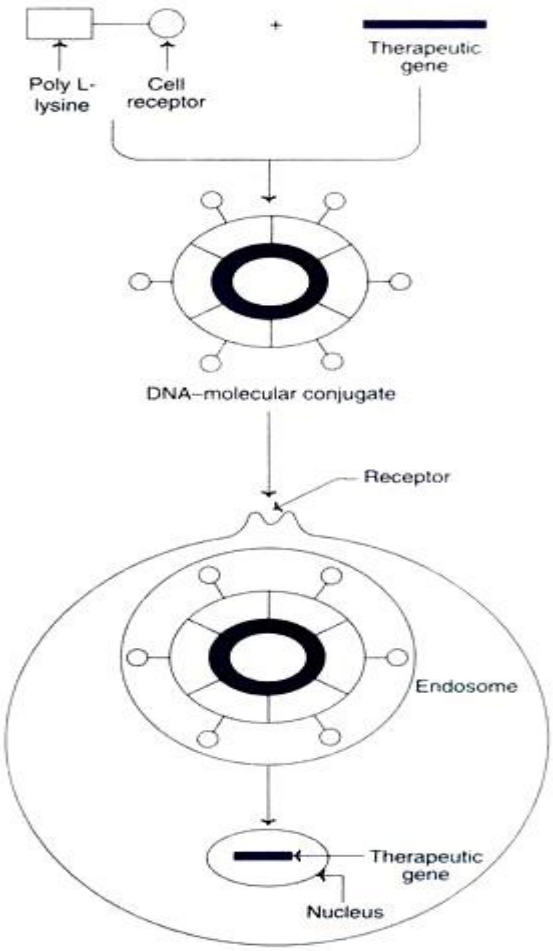


Fig. 13.7 : DNA-molecular conjugate in the delivery of therapeutic gene.

This DNA molecular conjugate binds to specific cell receptor on the target cells. It is engulfed by the cell membrane to form an endosome which protects the DNA from being degraded. The DNA released from the endosome enters the nucleus where the therapeutic gene is expressed.

Human artificial chromosome:

Human artificial chromosome (HAC) which can carry a large DNA one or more therapeutic genes with regulatory elements is a good and ideal vector. Studies conducted in cell cultures using HAC are encouraging. But the major problem is the delivery of the large-sized chromosome into the target cells. Researchers are working to produce cells containing genetically engineered HAC. There exists a possibility of encapsulating and implanting these cells in the target tissue.

Efficiency of gene delivery by non-viral vectors:

Although the efforts are continuously on to find suitable non-viral vectors for gene delivery, the success has been very limited. This is mainly due to the following two reasons.

1. The efficiency of transfection is very low.
2. The expression of the therapeutic gene is for a very short period, consequently there is no effective treatment of the disease.

Gene Therapy Strategies for Cancer:

Cancer is the leading cause of death throughout the world, despite the intensive treatment strategies (surgery, chemotherapy, radiation therapy). Gene therapy is the latest and a new approach for cancer treatment. Some of the developments are briefly described hereunder.

Tumour necrosis factor gene therapy:

Tumour necrosis factor (TNF) is a protein produced by human macrophages. TNF provides defence against cancer cells. This is brought out by enhancing the cancer-fighting ability of tumour- infiltrating lymphocytes (TILs), a special type of immune cells.

The tumour-infiltrating lymphocytes were transformed with a TNF gene (along with a neomycin resistant gene) and used for the treatment of malignant melanoma (a cancer of melanin producing cells usually occurs in skin). TNF as such is highly toxic, and fortunately no toxic side effects were detected in the melanoma patients injected with genetically altered TILs with TNF gene. Some improvement in the cancer patients was observed.

Suicide gene therapy:

The gene encoding the enzyme thymidine kinase is often referred to as suicide gene, and is used for the treatment of certain cancers. Thymidine kinase (TK) phosphorylates nucleosides to form nucleotides which are used for the synthesis of DNA during cell division. The drug ganciclovir (GCV) bears a close structural resemblance to certain nucleosides (thymidine). By mistake, TK phosphorylates ganciclovir to form triphosphate-GCV, a false and unsuitable nucleotide for DNA synthesis. Triphosphate-GCV inhibits DMA polymerase (Fig. 13.8).

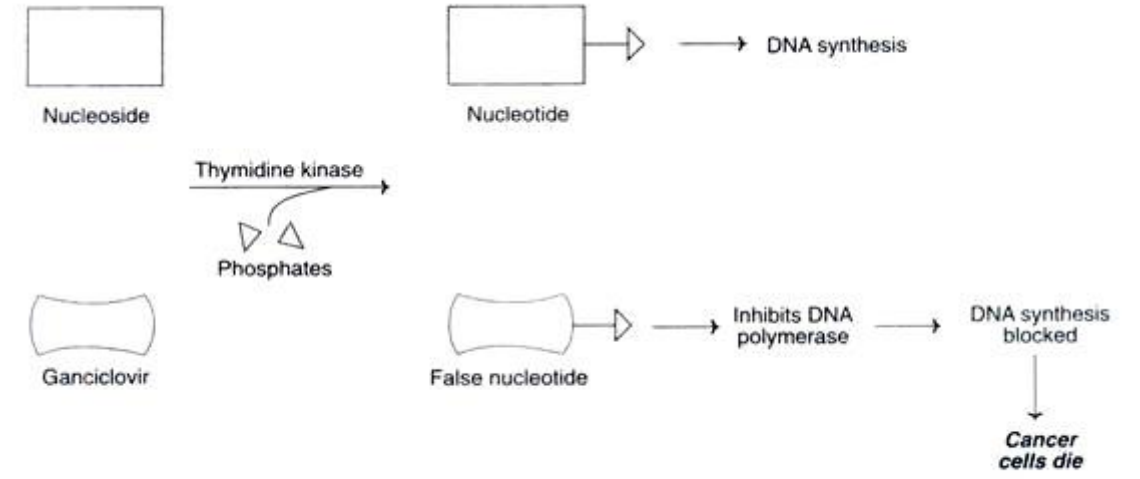


Fig. 13.8 : The action of ganciclovir mediated by thymidine kinase to inhibit the growth of cancer cells.

The result is that the elongation of the DNA molecule abruptly stops at a point containing the false nucleotide (of ganciclovir). Further, the triphosphate-GCV can enter and kill the neighbouring cancer cells, a phenomenon referred to as bystander effect. The ultimate result is that the cancer cells cannot multiply, and therefore die. Thus, the drug ganciclovir can be used to kill the cancer cells.

Ganciclovir is frequently referred to as a pro-drug and this type of approach is called pro-drug activation gene therapy. Ganciclovir has been used for treatment of brain tumours (e.g., glioblastoma, a cancer of glial cells in brain), although with a limited success.

In the suicide gene therapy, the vector used is herpes simplex virus (HSV) with a gene for thymidine kinase (TK) inserted in its genome. Normal brain cells do not divide while the brain tumour cells go on dividing unchecked. Thus, there is a continuous DNA replication in tumour cells. By using GCV-HSVTK suicide gene therapy, some reduction in proliferating tumour cells was reported. Several new strategies are being developed to increase the delivery of HSVTK gene to all the cells throughout a tumour.

Two-gene cancer therapy:

For treatment of certain cancers, two gene systems are put together and used. For instance, TK suicide gene (i.e., GCV-HSVTK) is clubbed with interleukin-2 gene (i.e. a gene promoting immunotherapy). Interleukin-2 produced mobilizes immune response. It is believed that certain proteins are released from the tumour cells on their death.

These proteins, in association with immune cells, reach the tumour and initiate immunological reactions directed against the cancer cells. Two-gene therapies have been carried out in experimental animals with colon cancer and liver cancer, and the results are encouraging.

Gene replacement therapy:

A gene named p^{53} codes for a protein with a molecular weight of 53 kilo Daltons (hence p^{53}). p^{53} is considered to be a tumour-suppressor gene, since the protein it encodes binds with DNA and inhibits replication. The tumour cells of several tissues (breast, brain, lung, skin, bladder,

colon, bone) were found to have altered genes of p^{53}(mutated p^{53}), synthesizing different proteins from the original.
These altered proteins cannot inhibit DNA replication. It is believed that the damaged p^{53} gene may be a causative factor in tumour development. Some workers have tried to replace the damaged p^{53}gene by a normal gene by employing adenovirus vector systems .There are some encouraging results in the patients with liver cancer.

Probable Questions:

1. Describe two methods which are used in gene transfer in animal cells?
2. What is lipofection? Describe the procedure?
3. What is gene gun? How biolistic method is used for gene transfer?
4. What is sonoporation?
5. What is optical transfection?
6. What is magnetofection?
7. Write down different applications in Recombinant DNA Technology?
8. What are the applications of GMO ?
9. What is knock out mice? What are its importance in genetic engineering?
10. Explain transcriptional gene silencing?
11. What is post transcriptional gene silencing?
12. What is ex vivo gene therapy? Explain.
13. What is in vivo gene therapy? Explain.

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.
9. Primrose SB and Twyman RM. (2006) Principles of Gene Manipulation and Genomics, 7th edition. Blackwell Publishing, Oxford, U.K.

UNIT-VII

Genome manipulation-CRISPR-Cas9 System

Objective: In this unit we will discuss about Genome manipulation-CRISPR-Cas9 System

Introduction:

Genome editing (also called gene editing) is a group of technologies that give scientists the ability to change an organism's DNA. These technologies allow genetic material to be added, removed, or altered at particular locations in the genome. Several approaches to genome editing have been developed. A well-known one is called CRISPR-Cas9, which is short for clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9. The CRISPR-Cas9 system has generated a lot of excitement in the scientific community because it is faster, cheaper, more accurate, and more efficient than other genome editing methods.

CRISPR-Cas9 was adapted from a naturally occurring genome editing system that bacteria use as an immune defence. When infected with viruses, bacteria capture small pieces of the viruses' DNA and insert them into their own DNA in a particular pattern to create segments known as CRISPR arrays. The CRISPR arrays allow the bacteria to "remember" the viruses (or closely related ones). If the viruses attack again, the bacteria produce RNA segments from the CRISPR arrays that recognize and attach to specific regions of the viruses' DNA. The bacteria then use Cas9 or a similar enzyme to cut the DNA apart, which disables the virus.

Researchers adapted this immune defence system to edit DNA. They create a small piece of RNA with a short "guide" sequence that attaches (binds) to a specific target sequence in a cell's DNA, much like the RNA segments bacteria produce from the CRISPR array. This guide RNA also attaches to the Cas9 enzyme. When introduced into cells, the guide RNA recognizes the intended DNA sequence, and the Cas9 enzyme cuts the DNA at the targeted location, mirroring the process in bacteria. Although Cas9 is the enzyme that is used most often, other enzymes (for example Cpf1) can also be used. Once the DNA is cut, researchers use the cell's own DNA repair machinery to add or delete pieces of genetic material, or to make changes to the DNA by replacing an existing segment with a customized DNA sequence.

Genome editing is of great interest in the prevention and treatment of human diseases. Currently, genome editing is used in cells and animal models in research labs to understand diseases. Scientists are still working to determine whether this approach is safe and effective for use in people. It is being explored in research and clinical trials for a wide variety of diseases, including single-gene disorders such as cystic fibrosis, hemophilia, and sickle cell disease. It also holds promise for the treatment and prevention of more complex diseases, such as cancer, heart disease, mental illness, and human immunodeficiency virus (HIV) infection.

Ethical concerns arise when genome editing, using technologies such as CRISPR-Cas9, is used to alter human genomes. Most of the changes introduced with genome editing are limited to somatic cells, which are cells other than egg and sperm cells (germline cells). These changes are isolated to only certain tissues and are not passed from one generation to the next. However, changes made to genes in egg or sperm cells or to the genes of an embryo could be passed to future generations. Germline cell and embryo genome editing bring up a number of ethical challenges, including whether it would be permissible to use this technology to enhance normal human traits (such as height or intelligence). Based on concerns about ethics and safety, germline cell and embryo genome editing are currently illegal in the United States and many other countries.

Importance

The CRISPR/Cas 9 technique is one of a number of gene-editing tools. Many favour the CRISPR/Cas9 technique because of its high degree of flexibility and accuracy in cutting and pasting DNA. One of the reasons for its popularity is that it makes it possible to carry out genetic engineering on an unprecedented scale at a very low cost. How it differs from previous genetic engineering techniques is that it allows for the introduction or removal of more than one gene at a time. This makes it possible to manipulate many different genes in a cell line, plant or animal very quickly, reducing the process from taking a number of years to a matter of weeks. It is also different in that it is not species-specific, so can be used on organisms previously resistant to genetic engineering.

The technique is already being explored for a wide number of applications in fields ranging from agriculture through to human health. In agriculture it could help in the design of new grains, roots and fruits. Within the context of health it could pave the way to the development of new treatments for rare metabolic disorders and genetic diseases ranging from haemophilia through to Huntingdon's disease. It is also being utilised in the creation of transgenic animals to produce organs for transplants into human patients. The technology is also being investigated for gene therapy. Such therapy aims to insert normal genes into the cells of people who suffer from genetic disorders such as cystic fibrosis, haemophilia or Tay Sachs. Several start-up companies have been founded to exploit the technology commercially and large pharmaceutical companies are also exploring its use for drug discovery and development purposes. The importance of the CRISPR/Cas9 was recognised with the awarding of the Nobel Prize in Chemistry to Jennifer Doudna and Emmanuel Charpentier on 7th October 2020. What is missed in the awarding of the Prize is the significant role that many others, including Virginijus Siksnys, played in helping to bring about the development of gene editing.

Discovery:

In 1987 a Japanese team of scientists at Osaka University noticed a strange pattern of DNA sequences in a gene belonging to Escherichia coli, a microbe that lives in the gut. It appeared that the gene had five short repeating segments of DNA separated by short non-repeating 'spacer' DNA sequences. All five repeating segments had identical sequences composed of 29

bases, the building blocks of DNA. By contrast each of the 'spacer' sequences had their own unique sequence, composed of 32 bases. Microbiologists had never seen such a pattern before. By the end of the 1990s, however, they had begun to discover, with the aid of new improvements to DNA sequencing, that this pattern was prevalent in many different microbe species.

So common was the pattern that it was given its own name: 'clustered regularly inter-spaced short palindromic repeats' or CRISPR for short. The term was coined by a team of Dutch scientists led by Rudd Jansen at Utrecht University, in 2002, who the same year noted that another set of sequences always accompanied the CRISPR sequence. This second set of sequences they dubbed 'Cas genes', an abbreviation for CRISPR-associated genes. The Cas genes appeared to code for enzymes that cut DNA. By 2005 three scientific teams had independently worked out that the 'spacer' sequences between the CRISP sequences shared similarities with the DNA of viruses and hypothesised that it could be a tool in the defence mechanism of bacteria.

Knowledge about how the CRISPR/Cas 9 system worked was opened up by some experiments conducted in 2007 by scientists at Danisco, a Danish food manufacturer later acquired by DuPont. The team infected a milk-fermenting microbe, Streptococcus thermophilius, with two virus strains. Many of these bacteria were killed by the viruses, but some survived and went on to produce offspring also resistant to the viruses. On further investigation it appeared that the microbes were inserting DNA fragments from the viruses into their 'spacer' sequences and that they lost resistance whenever the new 'spacer' sequences were cut out.

In 2008 Eugene Koonin and colleagues at the National Center for Biotechnology Information in Bethesda, Maryland, demonstrated for the first time how the CRISPR/Cas 9 mechanism worked. Whenever bacteria confront an invader, such as a virus, they copy and incorporate its DNA segments into their genome as 'spacers' between the short DNA repeats in CRISPR. The segments in the 'spacers' provide a template for the bacteria's RNA molecules to recognise any future DNA of an incoming virus and help guide the Cas 9 enzyme to cut it up so as to disable the virus.

Four years later, in August 2012, a small team of scientists led by Jennifer Doudna, University California Berkeley, and Emmanuelle Charpentier, University of Umea, published a paper showing how to harness the natural CRISPR-Cas9 system as a tool to cut any DNA strand in a test tube. Shortly before this another researcher, Virginijus Siksnys at Vilnus University, independently submitted a paper to Cell, elucidating the potential of CRISPR-CAS9 for gene editing in a paper. The editor of Cell rejected the manuscript without sending it out for review. Siksys eventually had his paper was published in the Procceedings of the National Academic of Sciences in September 2012. A year later, in January 2013, a number of researchers at different laboratories published papers within a few weeks of each other demonstrating how the CRISPR/Cas 9 system could be used to edit genomes in human cells. This included teams led by Doudna, Feng Zhang at MIT-Harvard Broad Institute, and George Church at Harvard Medical School.

A number of changes are now underway to improve the accuracy and efficiency of the CRISPR-Cas 9 technique. A key breakthrough has been the development of new Cas9 fusion proteins to act as base editors. The fusion proteins make it possible to convert cytosine to uracil without cutting DNA. Uracil is subsequently transformed into thymine through DNA replication or repair. The first base editors were generated in 2016 by Alexis Komor and colleagues in the laboratory of David Liu at Harvard University.

Application:

The CRISPR/Cas 9 system was first exploited by Danisco in 2008. The company used it to improve the immunity of bacterial cultures against viruses and many food manufacturers now use the technology to produce cheese and yoghurt. Since then the technology has been used to delete, insert and modify DNA in human cells and other animal cells grown in petri dishes. Scientists are also using it to create transgenic animals such as mice, rats, zebrafish, pigs and primates. Between 2014 and 2015 scientists reported the successful use of CRISPR/Cas 9 in mice to eliminate muscular dystrophy and cure a rare liver disease, and to make human cells immune to HIV. It is also being investigated in conjunction with pluripotent stem cells to provide human organs from transgenic pigs. Such work is directed towards helping solve some of the shortage of human organs for transplant operations and overcome some of the side-effects caused by organ transplantation such as graft-versus host disease. The technology is also being investigated as a means to genetically engineer insects so as to wipe out insect-borne diseases such as malaria, transmitted by mosquitoes, and lyme disease, transmitted by ticks.

Issues:

In April 2015 a Chinese group reported the first application of CRISPR/Cas9 to (non-viable) human embryos. This development, together with the decreasing costs of the technology have triggered a major bioethical debate about how far the technology should be used. The technology faces two major issues.

The first issue is a philosophical dilemma. It centres on the extent to which CRISPR/Cas9 should be used to alter 'germ-line' cells - eggs and sperm - which are responsible for passing genes on to the next generation. While it will take many more years before the technology will be viable to use to create designer babies, a public debate has already begun on this issue. So great is the fear that some scientists, including some who helped pioneer CRISPR/Cas9, have called for a moratorium on its use in germ-line cells.

The second issue is one of safety. One of the major problems is that the technology is still in its infancy and knowledge about the genome remains very limited. Many scientists caution that the technology still needs a lot of work to increase its accuracy and make sure that changes made in one part of the genome do not introduce changes elsewhere which could have unforeseen consequences. This is a particularly important issue when it comes to the use of the technology for applications directed towards human health. Another critical issue is that once

an organism, such as a plant or insect, is modified they are difficult to distinguish from the wild-type and once released into the environment could endanger biodiversity.

Policy-makers are still debating about what limitations to put on the technology. In April 2015 the US National Institutes of Health issued a statement indicating that it will not fund any research that uses genome editing tools such as CRISPR in human embryos. Meanwhile, the UK's Human Fertilisation and Embryology Authority, under whose remit such research would fall, has indicated that the CRISPR/Cas9 technology can be used on human-animal hybrid embryos under 14 days old. Any researcher working in this area would need to first get a license from the Authority. Other leading UK research councils have indicated that they support the continued use of CRISPR/Cas9 and other genome editing tools in preclinical research.

As regulators debate what restrictions to enforce with CRISPR/Cas9, the technology has become the subject of a major patent dispute. The first application to patent the technology was filed by DuPont in March 2007 (WO/2007/025097). This covers the use of the technology to develop phage resistant bacterial strains for food production, feeds, cosmetics, personal care products and veterinary products. Since then three heavily financed start-up biotechnology companies and half a dozen universities have filed patents. Two major competing patent claims have been filed in the US. The first, filed on 25 May 2015, is grounded in the work led by Jennifer Doudna at the University of California, Berkeley, and Emmanuelle Charpentier, originally at the University of Vienna and now at the Helmholz Centre for Infectious Research in Germany. The application has 155 claims and covers numerous applications for a variety of cell types (US Patent Application No. PCT/US2013/032589). The second, was filed by MIT-Harvard Broad Institute on 12 December 2012 for the work of Feng Zhang which focused on the use of CRISPR/Cas9 for genome editing in eukaryotic cells. It was given fast-track status and was granted on 15 April 2014 (US Patent No. 8,697,359). In April 2015 Charpentier and the Universities of California and Vienna filed a challenge to the patent with the US Patent and Trademark Office. It will take several years for the patent dispute to be settled. The legal wranglings over patents is unlikely to affect the use of CRISPR for basic research because the technology is available through an open-source repository. However, it could have an impact on clinical applications using the technique.

Genome editing is a type of genetic engineering in which DNA is deliberately inserted, removed, or modified in living cells. The name CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) refers to the unique organization of short, partially repeated DNA sequences found in the genomes of prokaryotes. CRISPR and its associated protein (Cas-9) is a method of adaptive immunity in prokaryotes to defend themselves against viruses or bacteriophages. Japanese scientist Ishino and his team accidentally found unusual repetitive palindromic DNA sequences interrupted by spacers in Escherichia coli while analyzing a gene for alkaline phosphatase first discovered CRISPR in 1987. However, they did not ascertain its biological function. In 1990, Francisco Mojica identifies similar sequences in other prokaryotes and he named CRISPR, yet the functions of these sequences were a mystery. Later on in 2007, a CRISPR was experimentally conferred as a key element in the adaptive immune system of

prokaryotes against viruses. During the adaptation process, bacterial cells become immunized by the insertion of short fragments of viral DNA (spacers) into a genomic region called the CRISPR array. Hence, spacers serve as a genetic memory of previous viral infections. The CRISPR defense mechanism protects bacteria from repeated viral attacks via three basic stages: adaptation (spacer acquisition), crRNA synthesis (expression), and target interference. CRISPR loci are an array of short repeated sequences found in chromosomal or plasmid DNA of prokaryotes. Cas gene is usually found adjacent to CRISPR that codes for nuclease protein (Cas protein) responsible to destroy or cleave viral nucleic acid.

Before the discovery of CRISPR/Cas-9, scientists were relied on two gene-editing techniques using restriction enzymes, zinc finger nucleases (ZFN) and Transcription activator-like effector nucleases (TALENs). ZFN has a zinc finger DNA binding domain used to bind a specific target DNA sequence and a restriction endonuclease domain used to cleave the DNA at the target site. TALENs are also composed of DNA binding domain and restriction domain like ZFN but their DNA binding domain has more potential target sequence than the ZFN gene-editing tool. In both cases, the difficulty of protein engineering, being expensive, and time-consuming were the major challenges for researchers and manufacturers. The development of a reliable and efficient method of a gene-editing tool in living cells has been a long-standing goal for biomedical researchers. After figuring out the CRISPR mechanism in prokaryotes, scientists understood that it could have beneficial use in humans, plants, and other microbes. It was in 2012 that Doudna, J, and Charpentier, E discovered CRISPR/Cas-9 could be used to edit any desired DNA by just providing the right template. Since then, CRISPR/Cas-9 becomes the most effective, efficient, and accurate method of genome editing tool in all living cells and utilized in many applied disciplines. Thus, this review aims to discuss the mechanisms of genome editing mediated by CRISPR/Cas-9 and to highlight its recent applications as one of the most important scientific discoveries of this century, as well as the current barriers to the transformation of this technology.

Components of CRISPR/Cas-9

Based on the structure and functions of Cas-proteins, CRISPR/Cas system can be divided into Class I (type I, III, and IV) and Class II (type II, V, and VI). The class I systems consist of multi-subunit Cas-protein complexes, while the class II systems utilize a single Cas-protein. Since the structure of type II CRISPR/Cas-9 is relatively simple, it has been well studied and extensively used in genetic engineering. Guide RNA (gRNA) and CRISPR-associated (Cas-9) proteins are the two essential components in CRISPR/Cas-9 system. The Cas-9 protein, the first Cas protein used in genome editing was extracted from *Streptococcus pyogenes* (SpCas-9). It is a large (1368 amino acids) multi-domain DNA endonuclease responsible for cleaving the target DNA to form a double-stranded break and is called a genetic scissor. Cas-9 consists of two regions, called the recognition (REC) lobe and the nuclease (NUC) lobe. The REC lobe consists of REC1 and REC2 domains responsible for binding guide RNA, whereas the NUC lobe is composed of RuvC, HNH, and Protospacer Adjacent Motif (PAM) interacting domains. The RuvC and HNH domains are used to cut each single-stranded DNA, while PAM interacting domain confers PAM

specificity and is responsible for initiating binding to target DNA. Guide RNA is made up of two parts, CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA). The crRNA is an 18–20 base pair in length that specifies the target DNA by pairing with the target sequence, whereas tracrRNA is a long stretch of loops that serve as a binding scaffold for Cas-9 nuclease. In prokaryotes, the guide RNA is used to target viral DNA, but in the gene-editing tool, it can be synthetically designed by combining crRNA and tracrRNA to form a single guide RNA (sgRNA) in order to target almost any gene sequence supposed to be edited.

Mechanisms of CRISPR/CAS-9 Genome Editing

The mechanism of CRISPR/Cas-9 genome editing can be generally divided into three steps: recognition, cleavage, and repair. The designed sgRNA directs Cas-9 and recognizes the target sequence in the gene of interest through its 5'crRNA complementary base pair component. The Cas-9 protein remains inactive in the absence of sgRNA. The Cas-9 nuclease makes double-stranded breaks (DSBs) at a site 3 base pair upstream to PAM.[14] PAM sequence is a short (2–5 base-pair length) conserved DNA sequence downstream to the cut site and its size varies depending on the bacterial species. The most commonly used nuclease in the genome-editing tool, Cas-9 protein recognizes the PAM sequence at 5'-NGG-3' (N can be any nucleotide base). Once Cas-9 has found a target site with the appropriate PAM, it triggers local DNA melting followed by the formation of RNA-DNA hybrid, but the mechanism of how Cas-9 enzyme melts target DNA sequence was not clearly understood yet. Then, the Cas-9 protein is activated for DNA cleavage. HNH domain cleaves the complementary strand, while the RuvC domain cleaves the non-complementary strand of target DNA to produce predominantly blunt-ended DSBs. Finally, the DSB is repaired by the host cellular machinery.

Double-Stranded Break Repair Mechanisms :

Non-homologous end joining (NHEJ), and homology-directed repair (HDR) pathways are the two mechanisms to repair DSBs created by Cas-9 protein in CRISPR/Cas-9 mechanism. NHEJ facilitates the repair of DSBs by joining DNA fragments through an enzymatic process in the absence of exogenous homologous DNA and is active in all phases of the cell cycle. It is the predominant and efficient cellular repair mechanism that is most active in the cells, but it is an error-prone mechanism that may result in small random insertion or deletion (indels) at the cleavage site leading to the generation of frameshift mutation or premature stop codon. HDR is highly precise and requires the use of a homologous DNA template. It is most active in the late S and G2 phases of the cell cycle. In CRISPR-gene editing, HDR requires a large amount of donor (exogenous) DNA templates containing a sequence of interest. HDR executes the precise gene insertion or replacement by adding a donor DNA template with sequence homology at the predicted DSB site.

Applications of CRISPR/CAS-9

In just a few years of its discovery, the CRISPR/Cas-9 genome editing tool has already being explored for a wide number of applications and had a massive impact on the world in many areas including medicine, agriculture, and biotechnology. In the future, researchers hope that this technology will continue to advance for treating and curing diseases, develop more nutritious crops, and eradicating infectious diseases.[18] Highlights for some of the recent CRISPR/Cas-9 applications and clinical trials being investigated are discussed below.

Role in Gene Therapy :

More than 6000 genetic disorders have been known so far. But the majority of the diseases lack effective treatment strategies. Gene therapy is the process of replacing the defective gene with exogenous DNA and editing the mutated gene at its native location. It is the latest development in the revolution of medical biotechnology. From 1998 to August 2019, 22 gene therapies including the novel CRISPR/Cas-9 have been approved for the treatment of human diseases.

Since its discovery in 2012, CRISPR/Cas-9 gene editing has held the promise of curing most of the known genetic diseases such as sickle cell disease, β-thalassemia, cystic fibrosis, and muscular dystrophy. CRISPR/Cas-9 for targeted sickle cell disease (SCD) therapy and β-thalassemia have been also applied in clinical trials. SCD is an autosomal recessive genetic disease of red blood cells, which occurs due to point mutation in the β-globin chain of hemoglobin leading to sickle hemoglobin (HbS). During the deoxygenation process, HbS polymerization leads to severe clinical complications like hemolytic anemia. Either direct repairing the gene of hemoglobin S or boosting fetal γ-globin are the two main approaches that CRISPR/Cas-9 is being used to treat SCD. However, the most common method used in a clinical trial is based on the approach of boosting fetal hemoglobin. First bone marrow cells are removed from patients and the gene that turns off fetal hemoglobin production, called B-cell Lymphoma 11A (BCL11A) is disabled with CRISPR/Cas-9. Then, the gene-edited cells are infused back into the body. BCL11A is a 200 base pair gene found on chromosome 2 and its product is responsible to switch γ-globin into the β-globin chain by repressing γ-globin gene expression. Once this gene is disabled using CRISPR/Cas-9, the production of fetal hemoglobin containing γ-globin in the red blood cells will increase, thereby alleviating the severity and manifestations of SCD.

Scientists have been also investigating CRISPR/Cas-9 for the treatment of cystic fibrosis. The genetic mutation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene decreases the structural stability and function of CFTR protein leading to cystic fibrosis. CFTR protein is an anion channel protein regulated by protein kinase-A, located at the apical surface of epithelial cells of the lung, intestine, pancreas, and reproductive tract. Although there is no cure for cystic fibrosis, symptom-based therapies (such as antibiotics, bronchodilators, and

mucus thinning medications) and CFTR modulating drugs have become the first-line treatments to relieve symptoms and reduce the risk of complications. Currently, gene manipulation technologies and molecular targets are also being explored. The use of CRISPR/Cas-9 technology for genome editing has great potential, although it is in the early stages of development. In 2013, researchers culture intestinal stem cells from two cystic fibrosis patients and corrected the mutation at the CFTR locus resulting in the expression of the correct gene and full function of the protein. Since then, the potential utility of the application of CRISPR/Cas-9 for cystic fibrosis was established. Furthermore, Duchenne muscular dystrophy (DMD), which is caused by a mutation in the dystrophin gene and characterized by muscle weakness, has been successfully corrected by CRISPR/Cas-9 in patient-induced pluripotent stem cells. Despite considerable efforts, the treatment available for DMD remains supportive rather than curative. Currently, several therapeutic approaches (gene therapy, cell therapy, and exon skipping) have been investigated to restore the expression of dystrophin in DMD muscles. Deletion/excision of intragenic DNA and removing the duplicated exon by CRISPR/Cas-9 are the new and promising approaches in correcting the DMD gene, which restores the expression of dystrophin protein.

Moreover, the latest researches show that the CRISPR/Cas-mediated single-base editing and prime editing systems can directly install mutations in cellular DNA without the need for a donor template. The CRISPR/Cas-base editor and prime editor system do not produce DSB, which reduces the possibility of indels that are different from conventional Cas-9. So far, two types of base editors have been developed: cytosine base editor (CBE) and adenine base editor (ABE). The CBE is a type of base editor composed of cytidine deaminase fused with catalytically deficient or dead Cas-9 (dCas-9). It is one of the novel gene therapy strategies that can produce precise base changes from cytidine (C) to thymidine (T). However, the target range of the CBE base editor is still restricted by PAM sequences containing G, T, or A bases. Recently, a more advanced fidelity and efficiency base editor called nNme2-CBE (discovered from *Neisseria meningitides*) with expanded PAM compatibility for cytidine dinucleotide has been developed in both human cells and rabbits embryos. The ABE uses adenosine deaminase fused to dCas-9 to correct the base-pair change from adenosine (A) to guanosine (G). Overall, single-base editing through the fusion of dCas-9 to cytidine deaminase or adenosine deaminase is a safe and efficient method to edit point mutations. But both base editors can only fix four-transition mutations (purine to purine or pyrimidine to pyrimidine). To overcome this shortcoming, the most recent member of the CRISPR genome editing toolkit called Prime Editor (PE) has been developed to extend the scope of DNA editing beyond the four types of transition mutations.[43] PE contains Cas-9 nickase fused with engineered reverse transcriptase and multifunctional primer editing guide RNA (pegRNA). The pegRNA recognizes the target nucleotide sequence; the Cas-9 nickase cuts the non-complementary strand of DNA three bases upstream from the PAM site, exposing a 3'-OH nick of genomic DNA. The reverse transcriptase then extends the 3' nick by copying the edit sequence of pegRNA. Hence, PE not only corrects all 12 possible base-to-base transitions, and transversion mutations but also small insertion and deletion mutations in genetic disorders.

Therapeutic Role of CRISPR/Cas-9 :

The first CRISPR-based therapy in the human trial was conducted to treat patients with refractory lung cancer. Researchers first extract T-cells from three patient's blood and they engineered them in the lab through CRISPR/Cas-9 to delete genes (*TRAC*, *TRBC*, and *PD-1*) that would interfere to fight cancer cells. Then, they infused the modified T-cells back into the patients. The modified T-cells can target specific antigens and kill cancer cells. Finally, no side effects were observed and engineered T-cells can be detected up to 9 months of post-infusion. CRISPR/Cas-9 gene-editing technology could also be used to treat infectious diseases caused by microorganisms. One focus area for the researchers is treating HIV, the virus that leads to AIDS. In May 2017, a team of researchers from Temple University demonstrated that HIV-1 replication can be completely shut down and the virus eliminated from infected cells through excision of HIV-1 genome using CRISPR/Cas-9 in animal models. In addition to the approach of targeting the HIV-genome, CRISPR/Cas-9 technology can also be used to block HIV entry into host cells by editing chemokine co-receptor type-5 (*CCR5*) genes in the host cells. For instance, an in vitro trial conducted in China reported that genome editing of CCR5 by CRISPR/Cas-9 showed no evidence of toxicity (infection) on cells and they concluded that edited cells could effectively be protected from HIV infection than unmodified cells.

Role in Agriculture :

As the world population continues to grow, the risk of shortage in agricultural resources is real. Hence, there is a need for new technologies for increasing and improving natural food production. CRISPR/Cas-9 is an existing addition to the field since it has been used to genetically modify foods to improve their nutritional value, increase their shelf life, make them drought-tolerant, and enhance disease resistance. There are generally three ways that CRISPR is solving the world's food crisis. It can restore food supplies, help plants to survive in hostile conditions, and could improve the overall health of the plants.

Role in Gene Activation and Silencing :

Beyond genome editing activity, CRISPR/Cas-9 can be used to artificially regulate (activate or repress) a certain target of a gene through advanced modification of Cas-9 protein. Researchers had performed an advanced modified Cas-9 endonuclease called dCas-9 nuclease by inactivating its HNH and RuvC domains. The dCas-9 nuclease lacks DNA cleavage activity, but its DNA binding activity is not affected. Then, transcriptional activators or inhibitors can be fused with dCas-9 to form the CRISPR/dCas-9 complex. Therefore, catalytically inactive dCas-9 can be used to activate (CRISPRa) or silence (CRISPRi) the expression of a specific gene of interest. Moreover, the CRISPR/dCas-9 can be also used to visualize and pinpoint where specifically the gene of interest is located inside the cell (subcellular localization) by fusing a marker such as Green Fluorescent Proteins (GFP) with

dCas-9 enzyme. This enables site-specific labeling and imaging of endogenous loci in living cells for further utilization.

Challenges for CRISPR/Cas-9 Application

Despite its great promise as a genome-editing system CRISPR/Cas-9 technology had hampered by several challenges that should be addressed during the process of application. Immunogenicity, lack of a safe and efficient delivery system to the target, off-target effect, and ethical issues have been the major barriers to extend the technology in clinical applications. Since the components of the CRISPR/Cas-9 system are derived from bacteria, host immunity can elicit an immune response against these components. Researchers also found that there were both pre-existing humoral (anti-Cas-9 antibody) and cellular (anti-Cas-9 T cells) immune responses to Cas-9 protein in healthy humans. Therefore, how to detect and reduce the immunogenicity of Cas-9 protein is still one of the most important challenges in the clinical trial of the system.

Safe and effective delivery of the components into the cell is essential in CRISPR/Cas-9 gene editing. Currently, there are three methods of delivering the CRISPR/Cas-9 complex into cells, physical, chemical, and viral vectors. Non-viral (physical and chemical) methods are more suitable for ex vivo CRISPR/Cas-9-based gene editing therapy. The physical methods of delivering CRISPR/Cas-9 can include electroporation, microinjection, hydrodynamic injection, and so on. Electroporation applies a strong electric field to the cell membrane to temporarily increase the permeability of the membrane, allowing the CRISPR/Cas-9 complex to enter the cytoplasm of the target cell. However, the main limitation of this method is that it causes significant cell death. Microinjection involves injecting the CRISPR/Cas-9 complex directly into cells at the microscopic level for rapid gene editing of a single cell. Nevertheless, this method also has several disadvantages such as cell damage, which is technically challenging and is only suitable for a limited number of cells. The hydrodynamic injection is the rapid injection of a large amount of high-pressure liquid into the bloodstream of animals, usually using the tail vein of mice. Although this method is simple, fast, efficient, and versatile, it has not yet been used in clinical applications due to possible complications. The chemical methods of CRISPR/Cas-9 delivery involve lipid and polymer-based nanoparticles. Lipid nanoparticles/liposomes are spherical structures composed of lipid bilayers membrane and is synthesized in aqueous solutions using Lipofectamine-based reagents. The positively charged liposomes encapsulated with negatively charged nucleic acids thereby facilitate the fusion of the complex across the cell membrane into cells. Polymeric nanoparticles, such as polyethyleneimine and poly-L-lysine, are the most widely used carriers of CRISPR/Cas-9 components. Like lipid nanoparticles, polymer-based nanoparticles can also transverse the complex in the membrane through endocytosis.

Viral vectors are the natural experts for in vivo CRISPR/Cas-9 delivery. Vectors, such as adenoviral vectors (AVs), adeno-associated viruses (AAVs), and lentivirus vectors (LVs) are currently being widely used as delivery methods due to their higher delivery efficiency relative

to physical and chemical methods. Among them, AAVs are the most commonly used vectors due to their low immunogenicity and non-integration into the host cell genome compared to other viral vectors. However, the limited virus cloning capacity and the large size of the Cas-9 protein remain a major problem. One strategy to tackle this hurdle is to package sgRNA and Cas-9 into separate AAVs and then co-transfect them into cells. Recent methods employ a smaller strain of Cas-9 from *Staphylococcus aureus* (SaCas-9) instead of the more commonly used SpCas-9 to allow packaging of sgRNA and Cas-9 in the same AAVs Lately, the development of extracellular vesicles (EVs), for the in vivo delivery of CRISPR/Cas-9 to avoid some of the limitations of viral and non-viral methods has shown a great potential.

The designed sgRNA will mismatch to the non-target DNA and can result in nonspecific, unexpected genetic modification, which is called the off-target effect. The CRISPR/Cas-9 target efficiency is determined by the 20-nucleotide sequences of sgRNA and the PAM sequences adjacent to the target genome. It has been shown that more than three mismatches between the target sequence and the 20-nucleotide sgRNA can result in off-target effects. The off-target effect can possibly cause harmful events such as sequence mutation, deletion, rearrangement, immune response, and oncogene activation, which limits the application of the CRISPR/Cas-9 editing system for therapeutic purposes. To mitigate the possibility of CRISPR/Cas-9 off-target effect, several strategies have been developed, such as optimization of sgRNA, modification of Cas-9 nuclease, utilization of other Cas-variants, and the use of anti-CRISPR proteins. Selecting and designing an appropriate sgRNA for the targeted DNA sequence is an important first step to reduce the off-target effect. When designing sgRNA, strategies such as GC content, sgRNA length, and chemical modifications of sgRNA must be considered. Generally speaking, studies revealed that GC content of between 40% and 60%, truncated (short length of sgRNA), and incorporation of 2'-O-methyl-3'-phosphonoacetate in the sgRNA ribose-phosphate backbone are the preferred methods to increase genome editing efficiency of CRISPR/Cas-9. Modifying the Cas-9 protein to optimize its nuclease specificity is another way to reduce off-target effects. For instance, mutating either one of the catalytic residues of Cas-9 nuclease (HNH and RuvC) will convert the Cas-9 into nickase that could only generate a single-stranded break instead of a blunt cleavage. It has been reported that the use of the inactivated RuvC domain of Cas-9 with sgRNA can reduce the off-target effect by 100 to 1500 times.[70] Moreover, the nuclease Cas-12a (previously known as Cpf1) is a type V CRISPR/Cas system that provides high genome editing efficiency. Unlike the CRISPR/Cas-9 system, CRISPR/Cas-12a can process pre-crRNA into mature crRNA without tracrRNA, thereby reducing the size of plasmid constructs. The Cas-12a protein recognizes a T-rich (5'-TTTN) PAM sequence instead of 5'-NGG and provides high accuracy at the target gene loci than Cas-9.[69] Recently, the use of multicomponent Class I CRISPR proteins, such as CRISPR/Cas-3 and CRISPR/Cas-10 provides better genome editing efficiency than Cas-9.[72] The Cas-3 is an ATP-dependent nuclease/helicase that can delete a large part of DNA from the target site without prominent off-target effect. For instance, the DMD gene were repaired by Cas-3-mediated system in induced pluripotent stem cell. The Cas-10 protein does not require the PAM sequence and can identify sequences even in the presence of point mutation. Anti-CRISPR (Acr) proteins are phage derived small proteins that inhibit the activity of CRISPR/Cas system. They are a recently discovered method to reduce off-target effects of CRISPR/Cas-9. From Acr proteins, AcrIIA4 specifically targets Cas-9 nuclease. AcrIIA4

mimics DNA and binds to the Cas-9 site, making impossible to perform further cleavage in area outside the target region. Furthermore, CRISPR/Cas-9 gene editing has been challenged by ethics and safety all over the world. Since the technology is still in its infancy and knowledge about the genome is limited, many scientists restrain that it still needs a lot of work to increase its accuracy and make sure that changes made in one part of the genome do not have unforeseen consequences, especially in the application towards human trials.

Conclusions:

CRISPR/Cas-9 system in nature is used to protect prokaryotes from invading viruses by recognizing and degrading exogenous genetic elements. CRISPR/Cas-9 gene editing is adopted from acquired immunity in prokaryotes and consists of two elements: guide RNA used to locate (bind) the target DNA to be edited and Cas-9, a protein that essentially cuts the DNA at the location identified by guide RNA. The fundamental part of the CRISPR/Cas-9 gene-editing process is the identification of the target gene that determines the phenotype of interest and designing the guide RNA. Now it becomes a new era in molecular biology and has countless roles ranging from basic molecular researches to clinical applications. Although tremendous efforts have been made, there are still some challenges to rub in the practical applications and various improvements are needed to overcome obstacles in order to assure its maximum benefit while minimizing the risk.

Probable Questions:

1. What is genome editing?
2. What are the importances of CRISPR-Cas9 system?
3. How CRISPR-Cas9 system was discovered?
4. Discuss components of CRISPR-Cas9 system.
5. Discuss mechanism of CRISPR-Cas9 system.
6. Discuss therapeutic role of CRISPR-Cas9 system.
7. Discuss role of CRISPR-Cas9 system in agriculture.
8. Discuss role of CRISPR-Cas9 system in gene activation and silencing.
9. Briefly discuss challenges of CRISPR-Cas9 system.

Suggested Readings:

1. Ormond KE(1), Mortlock DP(2), Scholes DT(3), Bombard Y(4), Brody LC(5), Faucett WA(6), Garrison NA(7), Hercher L(8), Isasi R(9), Middleton A(10), Musunuru K(11), Shriner D(12), Virani A(13), Young CE(3). Human Germline Genome Editing. Am J Hum Genet. 2017 Aug 3;101(2):167-176. PubMed: 28777929. Free full-text available from PubMed Central: PMC5544380.

2. Gupta RM, Musunuru K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. J Clin Invest. 2014 Oct;124(10):4154-61. doi: 10.1172/JCI72992. Epub 2014 Oct 1. Review. PubMed: 25271723. Free full-text available from PubMed Central: PMC4191047.

3. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-78. doi:10.1016/j.cell.2014.05.010. Review. PubMed: 24906146. Free full-text available from PubMed Central: PMC4343198.

4. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell. 2017 Apr 20;169(3):559. doi:10.1016/j.cell.2017.04.005. PubMed: 28431253.

UNIT-VIII

Polymerase chain reaction– methods and applications

Objectives: In this unit we will discuss about the Polymerase chain reaction process, its applications, and types.

Concept of PCR:

The polymerase chain reaction (PCR) provides a simple and ingenious method for exponentially amplification of specific DNA sequences by in vitro DNA synthesis. This technique was developed by Kary Mullis at Cetus Corporation between 1983 and 1985.

This technique has made it possible to synthesize large quantities of a DNA fragment without cloning it. The details of PGR techniques and its mechanism are described by Erlich (1989) in his edited book 'PCR Technology'. The PCR technique has now been automated and is carried out by a specially designed machine.

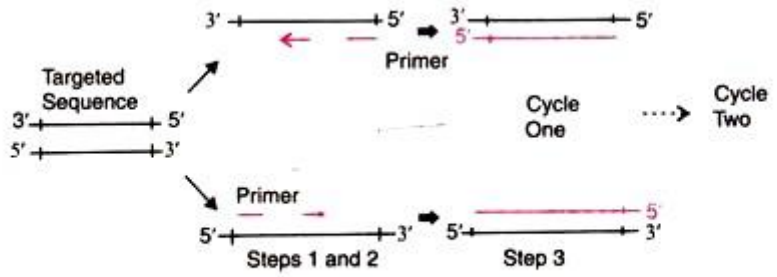


Fig. 6.11 : The working system of PCR. Cycle two follows the steps of cycle one.

Steps of PCR:

]

The PCR includes the following three essential steps to amplify a specific DNA sequence

(i) Melting of Target DNA:

The target DNA containing sequence (between 100 and 5,000 base) to be amplified is heat denatured (around 94°C for 15 seconds) to separate its complementary strands (step 1). This process is called melting of target DNA.

(ii) Annealing of Primers:

The second step is the annealing of two oligonucleotide primers to the denatured DNA strands. Primers are added in excess and the temperature lowered to about 68°C for 60 seconds;

consequently the primers form hydrogen bonds i.e. anneal to the DNA on both sides of the DNA sequence (step 2).

(iii) Primer Extension:

Finally, deoxynucleotide triphosphates (dATP, dGTP, dCTP, dTTP) and a thermo-stable DNA polymerase are added to the reaction mixture. The DNA polymerase accelerates the polymerization process of primers and, therefore, extends the primers (at 68°C) resulting in synthesis of copies of target DNA sequence (step 3). Only those DNA polymerases which are thermo-stable i.e. function at the high temperature are employed in PCR technique.

For this purpose the two popular enzymes, Taq polymerase (of a thermophilic bacterium Thermus aquaticus) and the vent polymerase (from Thermococcus litoralis) are used in PCR technology. These enzymes exhibit relative stability at DNA-melting replenishment after each cycle of synthesis. Also it reduces the cost of PCR and allows automated thermal cycling. However, after completion of step 3 (of one cycle) the targeted sequences on both strands are copied and four strands are produced. Now, the three step cycle (first cycle) is repeated which yields 8 copies from four strands.

Similarly, the third cycle produces 16 strands. This cycle is repeated about 50 times. Theoretically, 20 cycles (each of three steps) will produce about one million copies of the target DNA sequence, and 30 cycles will produce about one billion copies. In each cycle the newly synthesized DNA strands serve as targets for subsequent DNA synthesis as the three steps are repeated upto 50 times.

For the working of PCR about 10-100 Pico moles of primers are required. The concentration of target DNA can be about 10^{-20} to 10^{-15} M (or 1 to 10^5 DNA copies per ml). The PCR machine can carry-out 25 cycles and amplify DNA 10^5 times in 75 minutes.
The PCR technology has been improved in recent years. RNA can also be efficiently used in PCR technology. The rTth DNA polymerase can also be used instead of the Taq polymerase. The rTth polymerase will transcribe RNA to DNA, thereafter amplify the DNA. Therefore, cellular RNA and RNA viruses may be studied when they are present in small quantities.

Modified Forms of PCR:

The conventional PCR is the symmetrical PCR technique. There are some other modified forms of PCR which are used for various purposes:

AP-PCR (Arbitrarily Primed Polymerase Chain Reaction):

It requires only a single primer of relatively much smaller length compared to the primers used in PCR. This technique is used for DNA profiling, in animal and plant biotechnology as well as in forensic medicine.

Asymmetrical PCR:

Target sequences of one strand may be amplified in several orders of magnitude more as compared to its complementary strand. This approach is particularly useful for generating single stranded DNA fragment to be used for sequencing of DNA.

IPCR (Inverted Polymerase Chain Reaction):

In this method it allows the amplification of DNA flanking a known DNA sequence, the primers are facing outwards. Using the inverse PCR, the unknown sequences flanking known sequences can be readily amplified.

RT-PCR (Reverse Transcriptase Polymerase Chain Reaction):

Although the PCR amplification is generally performed on the DNA template but using this technique the RNA also can be used for amplification. This technique is particularly useful for studying the expression of genes and for monitoring the obscure species of mRNA

Nested PCR:

Nested PCR primers are ones that are internal to the first primer pair. The larger fragments produced by the first round of PCR is used as the template for the second PCR. This technique eliminates any spurious non-specific amplification products.

Application of PCR in Biotechnology:

PCR has many fold applications.

1. The amplification of gene fragments as fast alternative of cloning:

(a) Inserts of bacterial plasmids can be amplified with primers.

(b) DNA from known sequence can be obtained by designing primers.

(c) PCR helps in identification of homologous sequences from related organisms.

(d) Using RT-PCR the 3′ end of cDNA can be amplified (RACE: Rapid Amplification of cDNA Ends).

(e) Reverse PCR helps to know the flanking sequences of a known DNA clone.

2. Modification of DNA Fragments:

Site directed mutagenesis using oligonucleotides as PCR primers provides a powerful approach to study structure-function relation.

3. Diagnosis of Pathogenic Microorganism:

DNA from the infected parts of a person or animal may be subjected to PCR with primer specific gene of the pathogen and diagnosis can be done on amplification of DNA.

4. DNA Analysis of Archaeological Specimens:

As DNA is relatively stable and remain intact for a long period of time, PCR can help in analysis of DNA from those embedded materials.

5. Detection of Mutation Relevant for Inherited Diseases:

Any point mutation, a deletion or an insertion and expanded tandem trinucleotide repeat can be detected by PCR. Somatic mutations in oncogenes or tumour repressor genes can also be detected by PCR with primers flanking the insertions or deletions.

6. Analysis of Genetic Markers for Forensic Applications, for paternity testing and for the mapping of hereditary traits.

(a) Amplification of SSR.

(b) RAPD (Random Amplified Polymorphic DNA) with arbitrary, often short (10 bp) primers.

7. Species-Specific Amplification of DNA Segments between interspersed repeat elements (IRS) using the primer based on the SINE sequence (Short Interspersed Nuclear Elements).

8. Genetic Engineering using PCR:

Using PCR we can incorporate alteration or mutation in the ultimate product by choice altering, removing or adding sequences to the primer at the 5′ end. By recombinant PCR technique, it is possible to join two DNA fragments at a specific site through complementary overlaps (This technique is termed as splicing). By synthesizing two mutagenic primers, spanning the internal site to be changed, it is possible to introduce mutations within a fragment.

Advantages of Polymerase Chain Reaction:

PCR is so sensitive that DNA sequences present in an individual cell can be amplified. The isolation and amplification of a specific DNA sequence by PCR is faster and less technically difficult than traditional cloning methods using recombinant DNA techniques.

The top six applications are: (1) PCR in Clinical Diagnosis (2) PCR in DNA Sequencing (3) PCR in Gene Manipulation and Expression Studies (4) PCR in Comparative Studies of Genomes (5) PCR in Forensic Medicine and (6) PCR in Comparison with Gene Cloning.

1. PCR in Clinical Diagnosis:

The specificity and sensitivity of PCR is highly useful for the diagnosis of various diseases in humans. These include diagnosis of inherited disorders (genetic diseases), viral diseases, bacterial diseases etc. The occurrence of genetic diseases frequently identified by restriction fragment length polymorphism (RFLP) can be employed only when there is a mutation resulting in a detectable change in the length of restriction fragment. Many genetic diseases occur without the involvement of RFLP. For all such disorders, PCR technique is a real boon, as

it provides direct information of DNA. This is done by amplification of DNA of the relevant region, followed by the direct analysis of PCR products.

Prenatal diagnosis of inherited diseases:
PCR is employed in the prenatal diagnosis of inherited diseases by using chorionic villus samples or cells from amniocentesis. Thus, diseases like sickle-cell anemia, p-thalassemia and phenylketonuria can be detected by PCR in these samples.

Diagnosis of retroviral infections:
PCR from cDNA is a valuable tool for diagnosis and monitoring of retroviral infections, e.g., HIV infection.

Diagnosis of bacterial infections:
PCR is used for the detection of bacterial infection e.g., tuberculosis by Mycobacterium tuberculosis.

Diagnosis of cancers:
Several virally-induced cancers (e.g., cervical cancer caused by human papilloma virus) can be detected by PCR. Further, some cancers which occur due to chromosomal translocation (chromosome 14 and 18 in follicular lymphoma) involving known genes are identified by PCR.

PCR in sex determination of embryos:
Sex of human and livestock embryos fertilized in vitro, can be determined by PCR, by using primers and DNA probes specific for sex chromosomes. Further, this technique is also useful to detect sex — linked disorders in fertilized embryos.

2. PCR in DNA Sequencing:

As the PCR technique is much simpler and quicker to amplify the DNA, it is conveniently used for sequencing. For this purpose, single-strands of DNA are required. In asymmetric PCR, preferential amplification of a single-strand is carried out. In another method, strand removal can be achieved by digesting one strand (usually done by exonuclease by its action on 5′-phosphorylated strand).

3. PCR in Gene Manipulation and Expression Studies:

The advantage with PCR is that the primers need not have complementary sequences for the target DNA. Therefore, the sequence of nucleotides in a piece of the gene (target DNA) can be manipulated and amplified by PCR.

By using this method, coding sequence can be altered (thereby changing amino acids) to synthesize protein of interest. Further, gene manipulations are important in understanding the effects of promoters, initiators etc., in gene expression.

PCR is important in the study of mRNAs, the products of gene expression. This is carried out by reverse transcription PCR.

4. PCR in Comparative Studies of Genomes:

The differences in the genomes of two organisms can be measured by PCR with random primers. The products are separated by electrophoresis for comparative identification. Two genomes from closely related organisms are expected to yield more similar bands. For more details, refer the technique random amplified polymorphic DNA.

PCR is very important in the study evolutionary biology, more specifically referred to as phylogenetic. As a technique which can amplify even minute quantities of DNA from any source (hair, mummified tissues, bone, or any fossilized material), PCR has revolutionized the studies in palaentology and archaelogy. The movie 'Jurassic Park' has created public awareness of the potential applications of PCR!

5. PCR in Forensic Medicine:

A single molecule of DNA from any source (blood strains, hair, semen etc.) of an individual is adequate for amplification by PCR. Thus, PCR is very important for identification of criminals.

6. PCR in Comparison with Gene Cloning:

PCR has several advantages over the traditional gene cloning techniques .These include better efficiency, minute quantities of starting material (DNA), cost-effectiveness, minimal technical skill, time factor etc. In due course of time, PCR may take over most of the applications of gene cloning.

Real-Time PCR is a technique used to monitor the progress of a PCR reaction in real-time. At the same time, a relatively small amount of PCR product (DNA, cDNA or RNA) can be quantified. Real-Time PCR is based on the detection of the fluorescence produced by a reporter molecule which increases, as the reaction proceeds.

- Real-Time PCR is also known as a **quantitative polymerase chain reaction (qPCR)**, which is a laboratory technique of molecular biology based on the **polymerase chain reaction (PCR).**

- qPCR is a powerful technique that allows exponential amplification of DNA sequences.

- A PCR reaction needs a pair of primers that are complementary to the sequence of interest. Primers are extended by the DNA polymerase.

- The copies produced after the extension, so-called amplicons, are re-amplified with the same primers leading thus to exponential amplification of the DNA molecules.

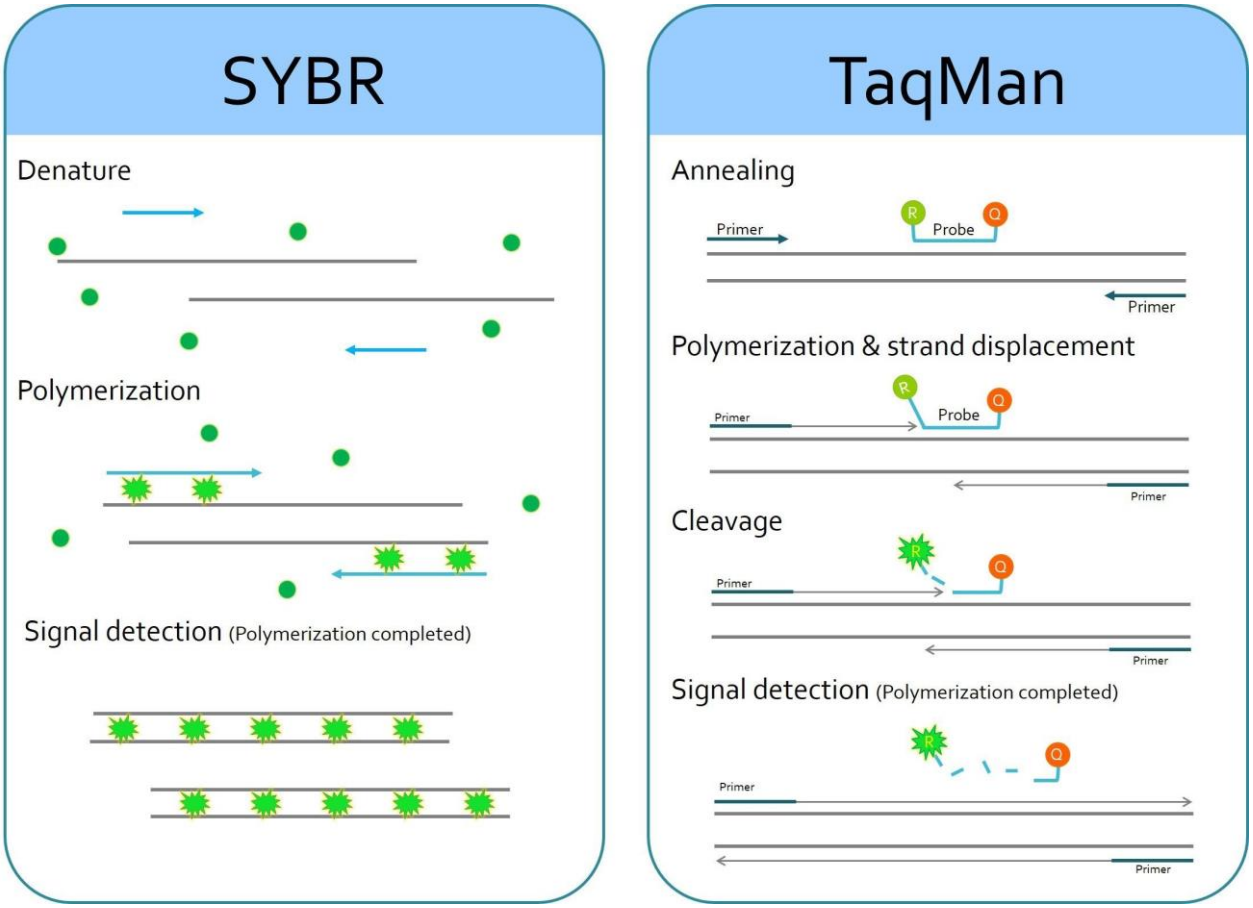
- After amplification, however, gel electrophoresis is used to analyze the amplified PCR products and this makes conventional PCR time consuming; since the reaction must finish before proceeding with the post-PCR analysis. Real-Time PCR overcomes this problem.

- The term “real-time” denotes that it can monitor the progress of the amplification when the process is going on in contrast to the conventional PCR method where analysis is possible only after the process is completed.

Principle of Real-Time PCR:

This same principle of amplification of PCR is employed in real-time PCR. But instead of looking at bands on a gel at the end of the reaction, the process is monitored in “real-time”. The reaction is placed into a real-time PCR machine that watches the reaction occur with a camera or detector.

Although many different techniques are used to monitor the progress of a PCR reaction, all have one thing in common. They all link the amplification of DNA to the generation of fluorescence which can simply be detected with a camera during each PCR cycle. Hence, as the number of gene copies increases during the reaction, so does the fluorescence, indicating the progress of the reaction.



The working procedure can be divided into two steps:

A. Amplification

1. Denaturation

High temperature incubation is used to "melt" double- stranded DNA into single strands and loosen secondary structure in single-stranded DNA. The highest temperature that the DNA polymerase can withstand is typically used (usually 95°C). The denaturation time can be increased if template GC content is high.

2. Annealing

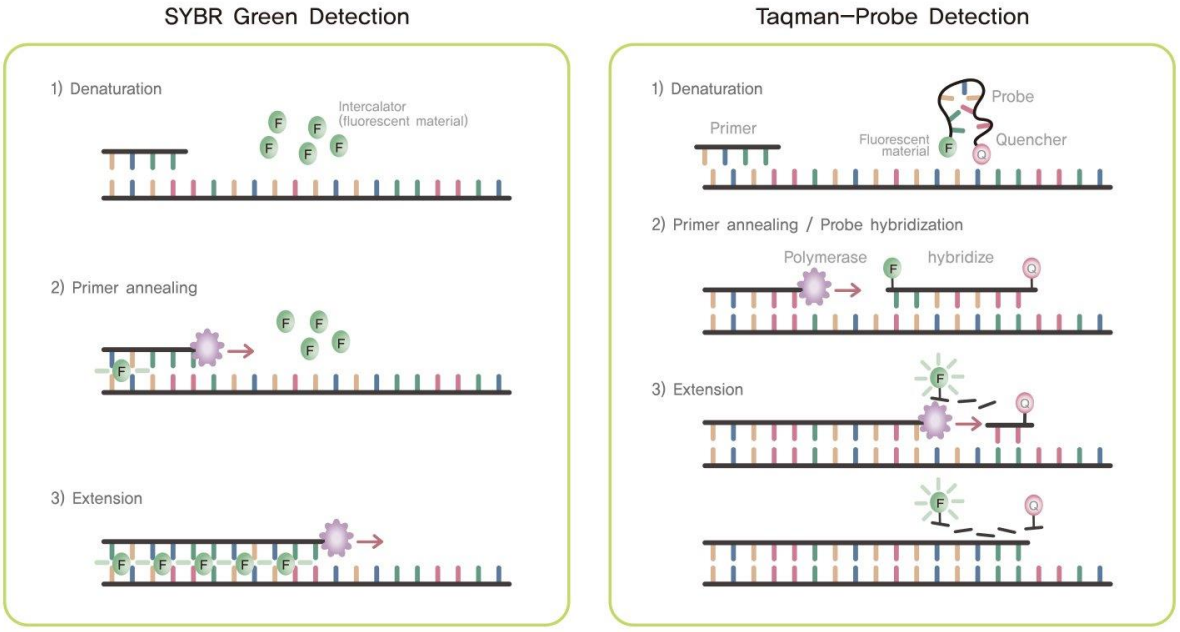
During annealing, complementary sequences have an opportunity to hybridize, so an appropriate temperature is used that is based on the calculated melting temperature (Tm) of the primers(5°C below the Tm of the primer).

3. Extension

At 70-72°C, the activity of the DNA polymerase is optimal, and primer extension occurs at rates of up to 100 bases per second. When an amplicon in real-time PCR is small, this step is often combined with the annealing step using 60°C as the temperature.

B. Detection

- The detection is based on fluorescence technology.
- The specimen is first kept in proper well and subjected to thermal cycle like in the normal PCR.
- The machine, however, in the Real Time PCR is subjected to tungsten or halogen source that lead to fluoresce the marker added to the sample and the signal is amplified with the amplification of copy number of sample DNA.
- The emitted signal is detected by an detector and sent to computer after conversion into digital signal that is displayed on screen.
- The signal can be detected when it comes up the threshold level (lowest detection level of the detector).



There are many different markers used in Real Time PCR but the most common of them include:

1. **Taqman probe.**
2. **SYBR Green.**

1. Taqman Probe

- It is a hydrolysis probe which bear a reporter dye, often fluorescein (FAM) at its 5' end and a quencher tetramethylrhodamine (TAMRA), attached to the 3' end of the oligonucleotide.
- Under normal conditions, the probe remain coiled on itself bringing the fluorescence dye near the quencher, which inhibits or quenches of fluorescent signal of the dye.
- The oligonucleotide of the Taqpolymerase has a homologous region with the target gene and thus when the target sequence is present in the mixture, it bind with the sample DNA.
- As the taqpolymerase start to sunthesize new DNA strand in the extension stage, it causes degradation of the probe by 5' end nuclease activity and the fluorescein is separated from the quencher as a result of which fluorescence signal is generated.
- As this procedure continues, in each cycle the number of signal molecule increases, causing the increase in fluorescence which is positively related with the amplification of the target.

2. SYBR Green

- This is a dye that emits prominent fluorescent signal when it binds at the minor groove of DNA, nonspecifically.
- Other fluorescent dyes like Ethidium Bromide or Acridine Orange can also be used but SYBR Green is better used for its higher signal intensity.
- SYBR Green is more preferred than the Taqman Probe as it can provide information about each cycle of amplification as well as about the melting temperature which is not obtained from the Taqman probe.
- However, its disadvantage is the lack of specificity as compared to Taqman Probe.

Advantages of Real-Time PCR:

It has many advantages over the normal PCR:

- It gives a look in to the reaction that is help to decide which reactions have worked well and which have failed.
- The efficiency of the reaction can be precisely calculated.
- There is no need to run the PCR product out on a gel after the reaction as the melt curve analysis serve the purpose.
- The real-time PCR data can be used to perform truly quantitative analysis of gene expression. In comparison, old fashioned PCR was only ever semi-quantitative at best.
- Faster than normal PCR.
- Less complexity at the quantification of sample.etc.

Thus, unlike the ordinary preparative PCR, Real Time PCR allows the success of multiple PCR reaction to be determined automatically after only a few cycles, without separate analysis of each reaction, and avoids the problem of “false negatives”.

Probable Questions:

1. What are the different steps of PCR? Give suitable diagram?
2. What is Tm? How it affects annealing temperature of a PCR ?
3. How primers are designed?
4. What is reverse transcriptase PCR? What is its importance?
5. What is real time PCR? Why it is so called?
6. Discuss role of PCR in modern biology.
7. How PCR can be used in clinical diagnosis?
8. Write working procedure of Real time PCR.
9. Discuss SYBR-Green method of real time PCR.
10. Discuss Taqman probe method of real time PCR.
11. What are the advantages of real time PCR over normal PCR?

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.

UNIT-IX

Basic Concepts of Microscopy: Magnification, Resolution, Limit of Resolution, Chromatic Aberrations. Types of microscopies: Bright Field Microscopy, Dark Field Microscopy, Phase Contrast Microscopy and Differential Interference Contrast Microscopy: Fluorescent Microscopes, Confocal microscopy

Objective: In this unit we will discuss about basic concepts of microscopy and different aspects of microscopy. We will also discuss about different types of microscopy such as Bright Field Microscopy, Dark Field Microscopy, Phase Contrast Microscopy, Fluorescent Microscopes, Confocal microscopy.

Introduction: A microscope is an optical instrument that is used for magnifying objects too small to be seen by naked eye. Investigations or studies of cell architecture by means of the microscope are called microscopy, and the person who pursuing the study is called microscopist.

History of Microscopy:

1590: The two Janssen brothers of Holland, Francis Janssen and Zacharias Janssen, who were spectacle makers built the first operational light microscope.

1611: Kepler built the first compound microscope.

1665: Robert Hooke developed the first laboratory microscope which has a magnification of 14-42 X. He observed small pores in sections of cork that he called cells.

1674: Leeuwenhoek discovered protozoa by his self built microscope with magnification of 270 X. He discovered bacteria for the first time 9 years later.

1905: Zsigmondy invented dark-field microscopy.

Properties of Microscope:

A microscope has dual property i.e. magnification and resolution. The usefulness of a microscope depends not so much on the degree of magnification but rather on the resolution. Resolution has nothing to do with the magnification.

(a) Magnification:

Magnification or magnifying power of a microscope is the degree of increase in size of optical image over the actual size of object being viewed.

Magnification = Size of retinal image seen with microscope/Size of retinal image with naked eye

Magnification of microscope is calculated by multiplying the magnification of the objective lens with that of the eye piece (ocular lens). For example, the magnification of eye piece is 10X and the magnification of objective lens is 40X, then the microscope magnifies the object by 10×40= 400 times i.e. magnification is 400X.

The human eye has no power of magnification, so magnifying glasses maybe used to magnify images up to about 10 times. A light microscope in which combination of lens used has a magnification of 100-2000 X. For higher magnification over400X, oil immersion lens can be used in which cedar wood oil placed between objective and the coverslip increase the light gathering properties of the lens.

Units of Measurement used in Microscopy:

1 metre (m) = 10^2 cm = 10^3 mm = 10^6 mm = 10^9 nm = 10^{10} A
1 centimeter (cm) = 1/100 metre (m) = 0.4 inch

1 millimetre (mm) =1/1000 metre = 0.001 m – 10^{-3} m = 10^{-3} mm = 10^6 nm = 10^7 A 1 micrometre (mm)** = 1 /1000 mm = 0.001mm = 10^{-3}mm = 10^{-6} m = 10^3 nm = 10^4 A 1 nanometer (nm) = 1/1000 mm = 0.001mm = 10^{-3} mm = 10^{-6} mm = 10^{-9} m = 10 A
1 angstrom $(A)^+$ = 1/10 nm = 0.1 nm = 10^{-1} nm = 10^{-7}mm = 10^{-10}m
Micrometers were formerly known as microns (μ), and nanometers as millimicrons (mμ).

The Angstrom is not an accepted measurement in the International system of Units. It is included here, however, because it was widely used in microscopey in the past.

(b) Resolution or Resolving power:

Resolution (= resolving power or resolving limit) of an optical device (eye or microscope) is its ability to distinguish between two very closely placed objects as separate objects. The resolving power of a microscope depends on (i) Wavelength of light (A) and (it) numerical- aperture (NA) of the lens system used. Resolution of a microscope can be calculated by Abbe Equation, after the name of German Physicist Ernst Abbe in 1876.

Resolution = $0.61\lambda/NA = 0.61\lambda/N \sin\theta$

0.61 = a trigonometric constant

λ = Wavelength of light used; 450-750 mm for visible light used in compound microscope,

Blue light has shortest wavelength (λ =450nm) gives maximum resolution. Therefore, blue filter blue light commonly used in microscopy.

NA – N Sin θ; where N is the refractive index of the medium (usually air or oil) between the specimen and objective lens. For air N = 1.0 and for immersion oil N = 1.5.

θ or a- half angle of the cone of light entering the objective lens from the specimen. The maximum value of 6 for the best objective lens is 70° (Sin 70° = 0.94). The resolution of light microscope, using air and blue light, will be Lm= 0.61x 450nm/ 1.0x 0.94= 292nm or – 0.3μm

If oil and blue light used, then Lm = 0.61x 450nm/ 1.0x 0.94= 194nmor – 0.2 μm

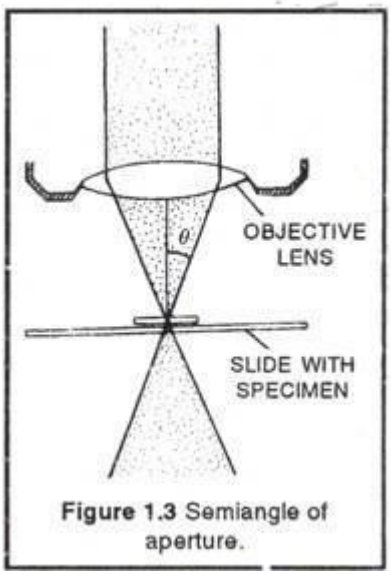


Figure 1.3 Semiangle of aperture.

Thus, light microscope can never resolve two closer particles less than about 0.2 nm apart, no matter how many times the image is magnified. The resolution of electron microscope is about 0.0005 μm whereas the human eye is about l00 μm. It should be noted that lower the value of Lm Higher will be resolution, which can be done by changing A, N or 6. Resolution will increase with a decrease in A and with an increase in NA; i.e.Lm is inversely proportional to A and Lm is proportional to NA. The numerical aperture (NA) is the light collecting ability of lens.

A. Light Microscope:

A light microscope is a biology laboratory instrument or tool, that uses visible light to detect and magnify very small objects, and enlarging them.They use lenses to focus light on the specimen, magnifying it thus producing an image. The specimen is normally placed close to the microscopic lens.Microscopic magnification varies greatly depending on the types and number of lenses that make up the microscope. Depending on the number of

lenses, there are two types of microscopes i. e Simple light microscope (it has low magnification because it uses a single lens) and the Compound light microscope (it has a higher magnification compared to the simple microscope because it uses at least two sets of lenses, an objective lens, and an eyepiece). The lenses are aligned in that, they can be able to bend light for efficient magnification of the image.

The functioning of the light microscope is based on its ability to focus a beam of light through a specimen, which is very small and transparent, to produce an image. The image is then passed through one or two lenses for magnification for viewing. The transparency of the specimen allows easy and quick penetration of light. Specimens can vary from bacterial to cells and other microbial particles.

Principle of Light Microscope:

As mentioned earlier, light microscopes visualize an image by using a glass lens and magnification is determined by, the lens's ability to bend light and focus it on the specimen, which forms an image. When a ray of light passes through one medium into another, the ray bends at the interface causing **refraction.** The bending of light is determined by the **refractive index**, which is a measure of how great a substance slows the speed of light. The direction and magnitude of the bending of the light are determined by the refractive indexes of the two mediums that form the interface.
A medium with a lower refractive index such as glass to air, it normally speeds up the light penetration and making light bend away from the normal and when light is passed through a medium with a greater refractive index such as air to glass, it normally slows down and bends towards the normal, perpendicularly to the surface.
If an object is put between these two mediums i.e between water and air, in this case, a prism, the prism will bend the light at an angle. This is how the microscopic lenses work, they bend the light at an angle. The lens (convex) on receiving the light rays, it focuses the rays at a specific point known as the **focal point (F-point)**. The measure of distance from the center of the lens and the focal point is known as the **focal length.** A microscope uses lenses whose strength is predetermined, in that, the strength of a lens is directly related to the focal length i.e short focal length magnifies objects more than lenses with a long focal length. Microscopy works strictly with a factor of resolution whereby resolution being the ability of a lens to be able to differentiate small objects that are closely packed together. The resolution of a light microscope is determined by a **numerical aperture** of its lens system and by the wavelength of the light it employs; a numerical aperture a definition of the light wavelengths produced when the specimen is illuminated. A minimum distance (d) between two objects that distinguishes then to be two separate entities, determined by the wavelengths of the light can be calculated by an Abbe equation using the wavelength of the light that illuminated the specimen (Lambda, **λ**) and the numerical aperture (NA, n sin θ) i.e. **$d=0.5\,\lambda/n\sin\theta$**

Types of light microscopes (optical microscope):

With the evolved field of Microbiology, the microscopes used to view specimens are both simple and compound light microscopes, all using lenses. The difference is simple light microscopes use a single lens for magnification while compound lenses use two or more lenses for magnifications. This means that a series of lenses are placed in an order such that, one lens magnifies the image further than the initial lens.

I. Brightfield Light Microscope (Compound light microscope):

This is the most basic optical Microscope used in microbiology laboratories which produces a dark image against a bright background. Made up of two lenses, it is widely used to view plant and animal cell organelles including some parasites such as *Paramecium* after staining with basic stains.Its functionality is based on being able to provide a high-resolution image, which highly depends on the proper use of the microscope. This means that an adequate amount of light will enable sufficient focusing of the image, to produce a quality image.
It is also known as a compound light microscope.

It is composed of:

- Two lenses which include the **objective lens** and the **eyepiece or ocular lens**.
- Objective lens is made up of six or more glasses, which make the image clear from the object
- The condenser is mounted below the stage which focuses a beam of light onto the specimen. It can be fixed or movable, to adjust the quality of light, but this entirely depends on the microscope.
- They are held together by a sturdy metallic curved back used as an **arm** and a stand at the bottom, known as the **base**, of the microscope. The arm and the base hold all the parts of the microscope.
- The stage where the specimen is placed, allowing movement of the specimen around for better viewing with the flexible knobs and it is where the light is focused on.
- Two focusing knobs i.e the fine adjustment knob and the coarse adjustment knob, found on the microscopes' arm, which can move the stage or the nosepiece to focus on the image. the sharpen the image clarity.
- It has a **light illuminator** or a **mirror** found at the base or on the microbes of the nosepiece.
- The nosepiece has about three to five objective lenses with different magnifying power. It can move round to any position depending on the objective lens to focus on the image.

- An aperture diaphragm also is known as the contrast, which controls the diameter of the beam of light that passes through the condenser, in that,

when the condenser is almost closed, the light comes through to the center of the condenser creating high contrast. But when the condenser is widely open, the image is very bright with very low contrast.

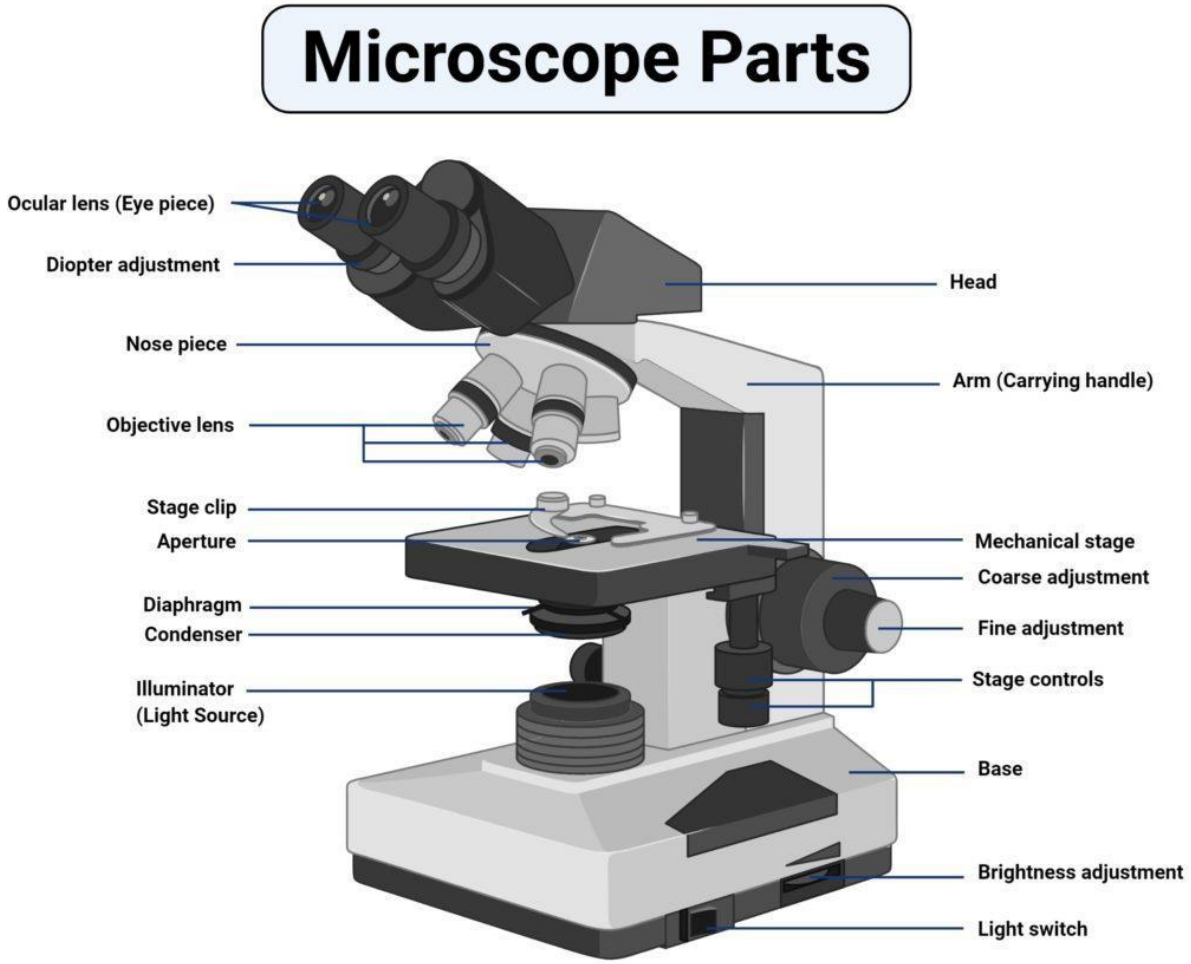


Figure: Parts of a microscope, Image Copyright © Sagar Aryal, www.microbenotes.com

Magnification by Bright field Microscope (Compound light microscope):

During visualization, the objective lens remains parfocal which means, when the objective lens is changed, the image still remains in focus. The objective lens plays a major role in focusing the image on the condenser forming an enlarged clear image within the microscope, which is then further magnified by the eyepiece to a primary image.

What is seen in the microscope as an enlarged clear image of the specimen is known as the virtual image. To calculate the magnification, multiply the objective and eyepiece objective magnification together. The magnification is standard, i.e not too high nor too low, and therefore depending on the magnification power of the lenses, it will range between 40X and 100X.

Calculation of magnification = Magnification of objective lens/magnification of the eyepiece lens. The objective lens plays a vital role in not only enlarging the image but also making it clear for viewing, a feature known as **resolution.** Resolution according to Prescott, is the ability of a lens to separate or distinguish between small objects closely linked together. Whereas the eyepiece magnifies the image at the end of the viewing, its magnification range is lower than that of the objective lens at 8X-12X (10X standard) and that of the objective lens at 40X-100X, magnification, and resolution of the microscope is highly dependant on the objective lens.

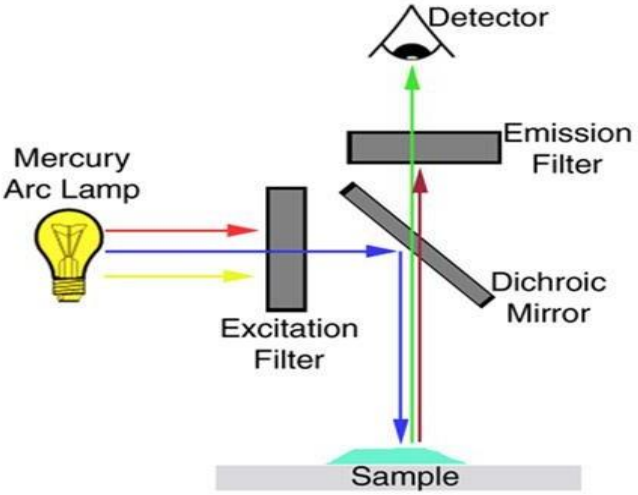
Applications of the Bright Field Light Microscope (Compound light microscope):
Vastly used in Microbiology this microscope is used to view fixed and live specimens, that have been stained with basic stains. This gives contrast for easy visibility under the microscope. Therefore it can be used to identify basic bacteria cells and parasitic protozoans such as *Paramecium.*

II. Fluorescent Microscopes:

A fluorescence microscope is an optical microscope that uses fluorescence and phosphorescence instead of, or in addition to, reflection and absorption to study properties of organic or inorganic substances.
Fluorescence is the emission of light by a substance that has absorbed light or other electromagnetic radiation while phosphorescence is a specific type of photoluminescence related to fluorescence. Unlike fluorescence, a phosphorescent material does not immediately re-emit the radiation it absorbs. The fluorescence microscope was devised in the early part of the twentieth century by August Köhler, Carl Reichert, and Heinrich Lehmann, among others.

Principle of fluorescence microscope :

Most cellular components are colourless and cannot be clearly distinguished under a microscope. The basic premise of fluorescence microscopy is to stain the components with dyes. Fluorescent dyes, also known as fluorophores or fluorochromes, are molecules that absorb excitation light at a given wavelength (generally UV), and after a short delay emit light at a longer wavelength. The delay between absorption and emission is negligible, generally on the order of nanoseconds.
The emission light can then be filtered from the excitation light to reveal the location of the fluorophores. Fluorescence microscopy uses a much higher intensity light to illuminate the sample. This light excites fluorescence species in the sample, which then emit light of a longer wavelength. The image produced is based on the second light source or the emission wavelength of the fluorescent species rather than from the light originally used to illuminate, and excite, the sample.



Working

Light of the excitation wavelength is focused on the specimen through the objective lens. The fluorescence emitted by the specimen is focused on the detector by the objective. Since most of the excitation light is transmitted through the specimen, only reflected excitatory light reaches the objective together with the emitted light.

Forms

The "fluorescence microscope" refers to any microscope that uses fluorescence to generate an image, whether it is a more simple set up like an epifluorescence microscope, or a more complicated design such as a confocal microscope, which uses optical sectioning to get better resolution of the fluorescent image. Most fluorescence microscopes in use are epifluorescence microscopes, where excitation of the fluorophore and detection of the fluorescence are done through the same light path (i.e. through the objective).

Typical components of a fluorescence microscope are:

a. Fluorescent dyes (Fluorophore)

A fluorophore is a fluorescent chemical compound that can re-emit light upon light excitation. Fluorophores typically contain several combined aromatic groups, or plane or cyclic molecules with several π bonds.

Many fluorescent stains have been designed for a range of biological molecules. Some of these are small molecules that are intrinsically fluorescent and bind a biological molecule of interest. Major examples of these are nucleic acid stains like DAPI and Hoechst, phalloidin which is used to stain actin fibers in mammalian cells.

b. A light source

Four main types of light sources are used, including xenon arc lamps or mercury-vapor lamps with an excitation filter, lasers, and high- power LEDs. Lasers are mostly used for complex fluorescence microscopy techniques, while xenon lamps, and mercury lamps, and LEDs with a dichroic excitation filter are commonly used for wide- field epifluorescence microscopes.

c. The excitation filter

The exciter is typically a bandpass filter that passes only the wavelengths absorbed by the fluorophore, thus minimizing the excitation of other sources of fluorescence and blocking excitation light in the fluorescence emission band.

d. The dichroic mirror

A dichroic filter or thin-film filter is a very accurate colour filter used to selectively pass light of a small range of colours while reflecting other colours.

e. The emission filter.

The emitter is typically a bandpass filter that passes only the wavelengths emitted by the fluorophore and blocks all undesired light outside this band – especially the excitation light.
By blocking unwanted excitation energy (including UV and IR) or sample and system autofluorescence, optical filters ensure the darkest background.

Advantages of fluorescence microscope:

1. Fluorescence microscopy is the most popular method for studying the dynamic behaviour exhibited in live-cell imaging.
2. This stems from its ability to isolate individual proteins with a high degree of specificity amidst non-fluorescing material.
3. The sensitivity is high enough to detect as few as 50 molecules per cubic micrometer.
4. Different molecules can now be stained with different colours, allowing multiple types of the molecule to be tracked simultaneously.
5. These factors combine to give fluorescence microscopy a clear advantage over other optical imaging techniques, for both in vitro and in vivo imaging.

Limitations of fluorescence microscope:

1. Fluorophores lose their ability to fluoresce as they are illuminated in a process called photobleaching. Photobleaching occurs as the fluorescent molecules accumulate chemical damage from the electrons excited during fluorescence.
2. Cells are susceptible to phototoxicity, particularly with short-wavelength light. Furthermore, fluorescent molecules have a tendency to generate reactive chemical species when under illumination which enhances the phototoxic effect.
3. Unlike transmitted and reflected light microscopy techniques fluorescence

microscopy only allows observation of the specific structures which have been labelled for fluorescence.

III. Confocal Microscope:

Optical sections are produced in the laser scanning confocal microscope by scanning the specimen point by point with a laser beam focussed in the specimen, and using a spatial filter, usually a pinhole (or a slit), to remove unwanted fluorescence from above and below the focal plane of interest (Fig. 4.11). The power of the confocal approach lies in the ability to image structures at discrete levels within an intact biological specimen. There are two major advantages of using the LSCM in preference to conventional epifluorescence light microscopy. Glare from out-of-focus structures in the specimen is reduced and resolution is increased both laterally in the X and the Y directions (0.14 mm) and axially in the Z direction (0.23 mm). Image quality of some relatively thin specimens, for example, chromosome spreads and the leading lamellipodium of cells growing in tissue culture (<0.2 mm thick) is not dramatically improved by the LSCM whereas thicker specimens such as fluorescently labelled multicellular embryos can only be imaged using the LSCM.

For successful confocal imaging, a minimum number of photons should be used to efficiently excite each fluorescent probe labelling the specimen, and as many of the emitted photons from the fluorochromes as possible should make it through the light path of the instrument to the detector. The LSCM has found many different applications in biomedical imaging. Some of these applications have been made possible by the ability of the instrument to produce a series of optical sections at discrete steps through the specimen (Fig. 4.12). This series of optical sections collected with a confocal microscope are all in register with each other, and can be merged together to form a single projection of the image (Z projection) or a 3D representation of the image (3D reconstruction).

Multiple-label images can be collected from a specimen labelled with more than one fluorescent probe using multiple laser light sources for excitation (Fig. 4.13, see also colour section). Since all of the images collected at different excitation wavelengths are in register it is relatively easy to combine them into a single multicoloured image. Here any overlap of staining is viewed as an additive colour change. Most confocal microscopes are able to routinely image three or four different wavelengths simultaneously. The scanning speed of most laser scanning systems is around one full frame percentage second. This is designed for collecting images from fixed and brightly labelled fluorescent specimens. Such scan speeds are not optimal for living specimens, and laser scanning instruments are available that scan at faster rates for more optimal live cell imaging. In addition to point scanning, swept field scanning rapidly moves a mm thin beam of light horizontally and vertically through the specimen.

Optical sectioning: Many images collected from relatively thick specimens produced

using epifluorescence microscopy are not very clear. This is because the image is made up of the optical plane of interest together with contributions from fluorescence above and below the focal plane of interest. Since the conventional epifluorescence microscope collects all of the information from the specimen, it is often referred to as a wide field microscope. The 'out-of-focus fluorescence' can be removed using a variety of optical and electronic techniques to produce optical sections. The term optical section refers to a microscope's ability to produce sharper images of specimens than those produced using a standard wide field epifluorescence microscope by removing the contribution from out-of-focus light to the image, and in most cases, without resorting to physically sectioning the tissue. Such methods have revolutionised the ability to collect images from thick and fluorescently labelled specimens such as eggs, embryos and tissues. Optical sections can also be produced using high-resolution DIC optics, micro computerised tomography (CT) scanning or optical projection tomography. However, currently by far the most prevalent method is using some form of confocal or associated microscopical approach.

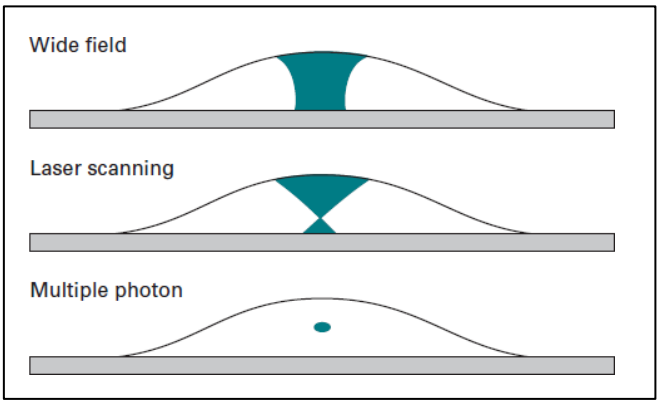


Figure: Illumination in a wide field, a confocal and a multiple photon microscope. The diagram shows a schematic of a side view of a fluorescently labelled cell on a coverslip. The shaded green areas in each cell represent the volume of fluorescent excitation produced by each of the different microscopes in the cell. Conventional epifluorescence microscopy illuminates throughout the cell. In the LSCM fluorescence illumination is throughout the cell but the pinhole in front of the detector excludes the out-of-focus light from the image. In the multiple photon microscope, excitation only occurs at the point of focus where the light flux is high enough.

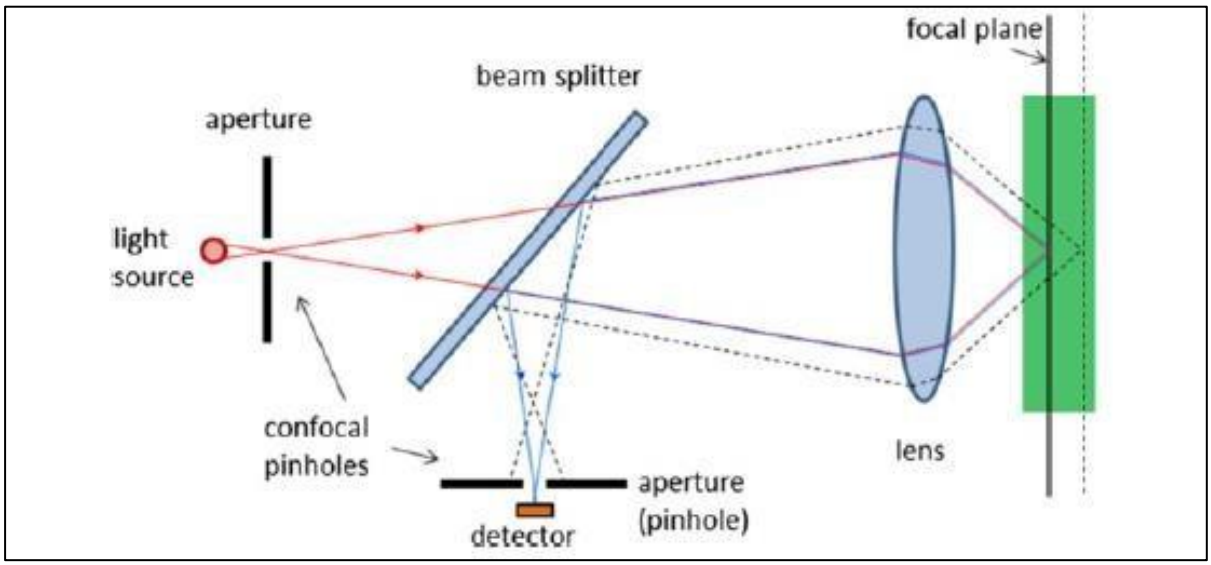


Figure: The principle of confocal microscopy. Only light reflected by structures very close to focal plane can be detected.

IV. Phase Contrast Microscope:

In recent years, remarkable advances have been made in the study of living cells (unstained) by the development of special optical techniques such as phase contrast and interference are highly transparent to visible light and they cause phase changes in transmitted radiations microscopy.

The biological specimens:

The phase contrast microscope has the same resolving power as the ordinary light microscope but it permits visualization of different parts of the cell due to differences in their refractive index (Refractive index is defined as the ratio of the velocity of light in a vacuum to its velocity in a transmitting medium).

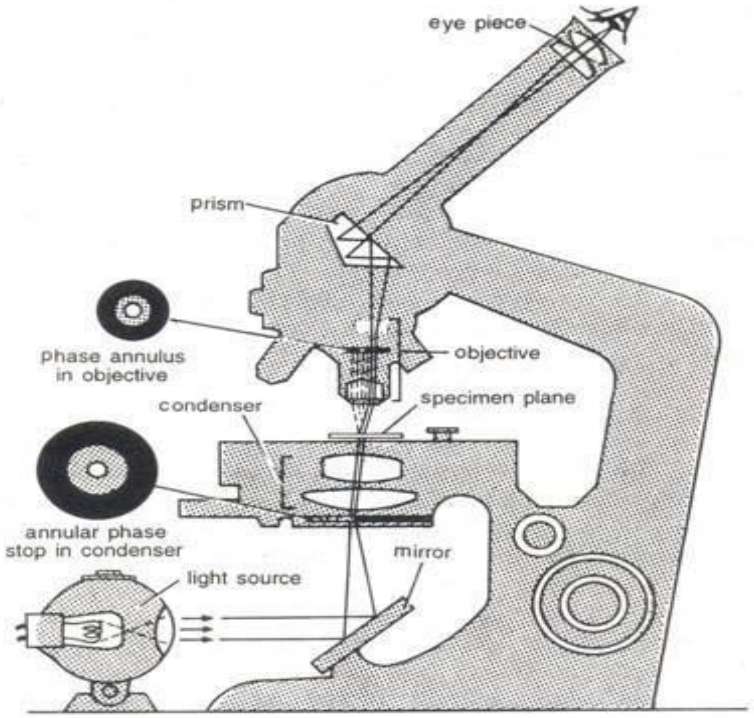


Fig. 2. The light path in a phase contrast microscope.

Because light is transmitted through a structure at a velocity inversely proportional to the refractive index of the structure, light waves emerging from structures with different refractive index will be out of phase with one another. The phase contrast microscope is able to convert these differences in phase to differences in light intensity, producing an image with good contrast. The phase-contrast microscope utilizes interference between two beams of light.

In the phase contrast microscope, the small phase differences are intensified. The most lateral light passing through the objective lens of the microscope is advanced or retarded by an additional l/4th wavelength (1/4λ.) with respect to the central light passing through the medium around the object, by an annular phase plate that introduces a 1/4 wavelength variation in the back focal plane of the objective.

In addition an annular diaphragm is placed in the substage condenser. The phase effect results from the interference between the direct geometric image given by the central part of the objective and the lateral diffracted image, which has been retarded or advanced to a total of 1/2 wavelength. In bright or negative contrast, the two sets of rays are added and the object appears brighter than the surroundings. In dark or positive contrast, the two sets of rays are subtracted making the image of the object darker than the surroundings. Because of this interference, the minute phase changes within the object are amplified and intensified.

A transparent object thus appears in various shades of gray, depending upon the thickness of the object and the difference between the refractive indices of the object and the medium. Phase microscopy is used to observe living cells and tissues. It is particularly valuable for observing the cells cultured in vitro during mitosis. In addition an annular diaphragm is placed in the substage condenser. The phase effect results from the

interference between the direct geometric image given by the central part of the objective and the lateral diffracted image, which has been retarded or advanced to a total of 1/2 wavelength. In bright or negative contrast, the two sets of rays are added and the object appears brighter than the surroundings. In dark or positive contrast, the two sets of rays are subtracted making the image of the object darker than the surroundings. Because of this interference, the minute phase changes within the object are amplified and intensified.

A transparent object thus appears in various shades of grey, depending upon the thickness of the object and the difference between the refractive indices of the object and the medium. Phase microscopy is used to observe living cells and tissues. It is particularly valuable for observing the cells cultured in vitro during mitosis.

Principle of the phase-contrast microscope:

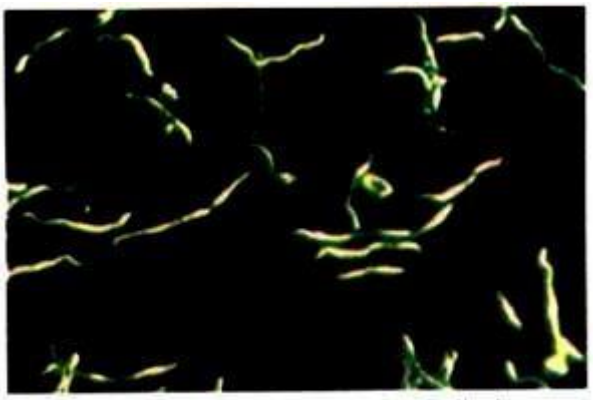
This is to convert small phase differences into differences in contrast that can be detected visually. An annular phase plate is placed in the objective of the microscope and an annular diaphragm is placed in the condenser as shown in the figure 2. As light is transmitted through the lenses, some of the rays pass through in a direct path while others are diffracted laterally. Diffracted light rays are, thus out of phase with the direct light, and an image of strong contrasts is produced. The annular diaphragm illuminates the object with a narrow cone of light, and the annular phase plate produces a variation of 1/4 A. between the diffracted lateral light and the direct light. The phase effect is the result of interference between the direct image in the centre of the objective and the diffracted lateral image. If the diffracted image is retarded, negative contrast results, whereas if it is advanced, positive contrast results. When the refractive index of the medium is greater than that of the object, the object is dark, and when the refractive index of the medium is less than that of the object, the object is bright.

V. Dark Field Microscope:

In this type of microscopy, a dark back ground is produced against which objects are brilliantly illuminated. For this purpose the light microscope is equipped with a special kind of condenser that transmits a hollow core of light from the source of illumination. Thus, if the aperture of condenser is allowed to open completely, and a dark field stop inserted below the condenser, the light rays reach the objects form a hollow core.

Any object within this beam of light will reflect some light into the objective and will be visible. This method of illuminating an object where the object appears self-illuminous against a dark field, called dark-field illustration.

The condensers used are Abbe condenser, paraboloid condenser and cardoid condenser. Dark field microscopy is particularly valuable for the examination of unstained microorganisms suspended in fluid wet mount and hanging drop operations.



Treponema pallidum, the spirochete that causes syphilis; dark-field microscopy.

Probable Questions:

1. Write the basic principle of fluorescent microscope?
2. Describe different components of fluorescent microscope.
3. What are the advantages of fluorescent microscope ?
4. What are the limitations of fluorescent microscope ?
5. Describe basic principle of phase contrast microscope.
6. Describe basic principle of Dark field microscope.
7. Describe basic principle of confocal microscope.

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

UNIT-X

Electronic Imaging Systems- Electron Microscopy, TEM Vs. SEM. Different fixation and staining techniques for EM, freeze-etch and freeze-fracture methods for EM, image processing methods in microscopy

Objective: In this unit we will discuss about Electronic Imaging Systems- Electron Microscopy, TEM Vs. SEM. Different fixation and staining techniques for EM, freeze-etch and freeze-fracture methods for EM, image processing methods in microscopy.

Electron Microscopy:

In 1931 Knoll and Ruska, German scientists discovered electron microscopy. Von Borries and Ruska (1938) in Berlin constructed first practical electron microscope. The commercial instrument first came in around 1940.In electron microscope the source of illumination is electron beam. The construction and principle of electron microscope are easily related to those of light microscope. The range of wave length of visible light used in light microscope is 4000 Å – 7800 Å, while with an electron microscope employing 60-80 KV electron, the wave length is only 0.05 Å.

In the instrument as shown in the figure, the electron gun generates electron beam. These electrons are concentrated by other components of electron gun producing a fast moving narrow beam of electron. Electrons are focused by electromagnetic lenses. Electromagnetic lens consists of wire en- cased in soft iron casing. When electric current is passed through the coil, it generates an electromagnetic field through which electrons are focused.

There are three general types of electromagnetic lenses. The one is placed between the source of illumination and the specimen. This focuses the beam of electron on specimen functions in a similar manner as that of light microscope. The other two lenses are on the opposite side of specimen which magnify the image in similar fashion as objective and ocular in light microscope.

General Principle of EM:

The fundamental principle of EM is similar to those of LM. In EM, a high velocity beam of electrons (instead of light) is used to travel in a vacuum tube. The beam of electrons is focused by a series of electromagnetic lenses analogous to the condenser, objective and eye piece lenses of the light microscope.

The object is placed between the condenser and objective. The magnified image of the object is formed on the fluorescent screen or on photographic film rather than being observed through eye piece. Since the image produced by electrons does not have the colour, the electron micrograph always has shades of black, grey and white.The objects under examination must be extremely thin and are treated with chemicals or dyes to enhance the contrast as such the live objects cannot be studied. Techniques like negative staining, shadow casting and tracers are commonly used to increase the contrast. Theoretically, the maximum resolution of the EM is 0.005 nm which is less than the diameter of a single atom, or 40,000 times the resolution of the light microscope and 2 million times that of the naked eye. However, the practical resolution of modern EM is of 0.1 nm (1 A).

Construction of an Electron Microscope:

An electron microscope consists of an electric gun, microscope column, electromagnetic coils, a fluorescent screen and some other accessories described below:

(a) The electron gun is located at the top of the body of microscope. It is the source of electrons. It is made up of a tungsten filament surrounded by a negatively biased shield with an aperture. The electron beam is drawn off through this aperture.

(b) The microscope column or central column is made up of an evacuated metal tube. It protects the person operating the microscope from X-rays that are generated when the electrons strike the surface of the metal tube.

(c) The electromagnetic coils or lenses include projector coils, objective arid condenser. In each coil, the coils of electric wire are wound on a hollow metallic cylinder. The magnetic field, produced by passing the electric current through the magnetic coil, functions as a magnifying lens.

(d) The fluorescent screen is used for observing the magnified image of the object. It remains coated with a chemical which, on being excited, forms the image as on the screen of television.

(e) Some other essential accessories of the electron microscope include high voltage transformers (for developing high voltage current for the electron gun and electromagnetic coils), vacuum pumps (for maintaining high vacuum inside the microscope column), a water cooling system (for prevention from overheating of various parts), a circulating pump, a refrigeration plant and also a filter system.

All these parts require elaborate arrangements and contribute to the massive size of the electron microscope. The image formation in this microscope occurs by the scattering of electrons. The electrons strike the atomic nuclei and get dispersed. These dispersed electrons

form the electron image. By projecting on a fluorescent screen or photographic film, this electron image is converted into a visible image of the object. The electron beam in this microscope is made by accelerating electrons through a potential difference of from 1-1500 kV in an electron gun.

Only dried specimens are studied by electron microscope. Living cells cannot be studied with this microscope because they possess water which causes large scale scattering of electrons. Ultrathin sections (10-50 nm thickness), which are more than 200 times thinner than those routinely used for light microscopy, are cut for electron microscopy. These are cut with the help of diamond or glass knives of an ultra-microtome.

Electron microscopes are of two types:

(1) Transmission electron microscope, and

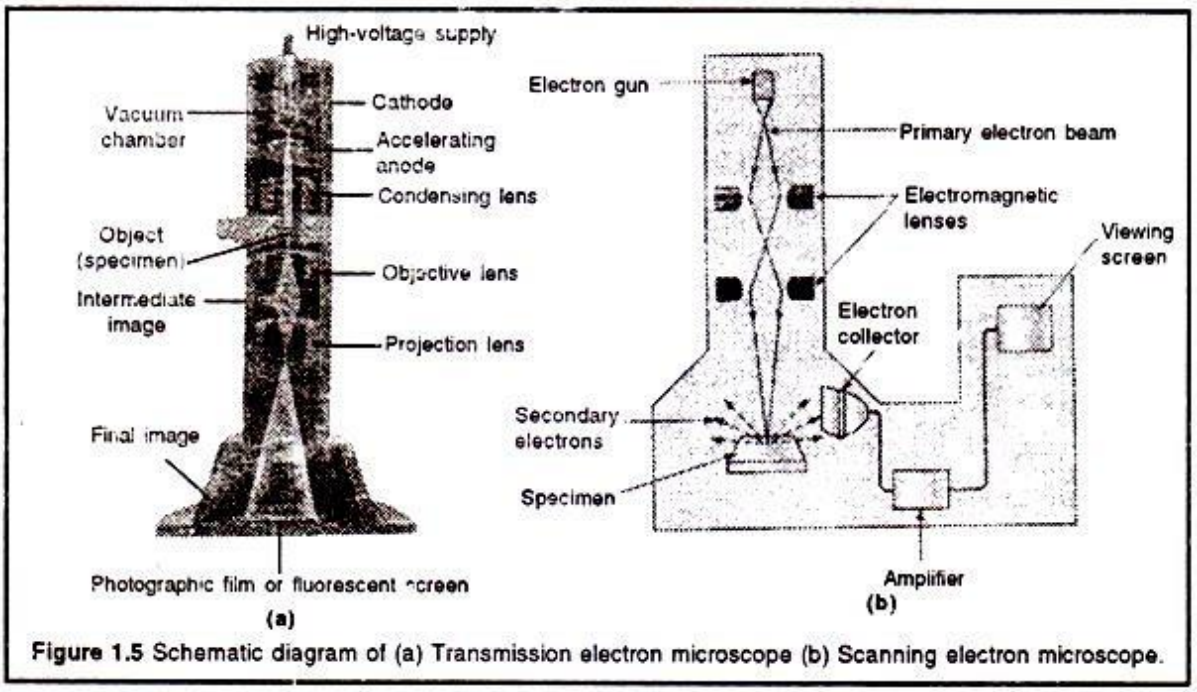


Figure 1.5 Schematic diagram of (a) Transmission electron microscope (b) Scanning electron microscope.

2) Scanning electron microscope.

a. Transmission Electron Microscope:

This microscope forms an image of the specimen by the electrons that have passed through the specimen (Fig. 21.2). The components of the specimen that scatter electrons appear dark and are called **"electrons dense".** The part that have less ability of electron scattering appear light.

The electron scattering ability of the element with higher atomic number, such as, uranium, lead etc. is greater than those of lower atomic numbers. The biological molecules are composed of the elements with comparatively low atomic number, viz., hydrogen, carbon, nitrogen, oxygen, phosphorus and sulphur. These elements have poor electron scattering ability. Therefore, biological molecules are stained with metals of high atomic

number such as uranium, lead and osmium.

The material is fixed in osmium tetrachloride, $KMnO_4$ or phosphotungstic acid. The fixed tissues are then embedded in hard plastic resin. Ultra-microtome is used to cut ultrathin sections (50-100 nm) of the material. These sections are examined under the electron microscope. Intact organelles and viruses are not sectioned. Followings are some techniques used to observe the materials by electron microscope.

Structural parts of a TEM- The structural parts of a TEM are as follows :

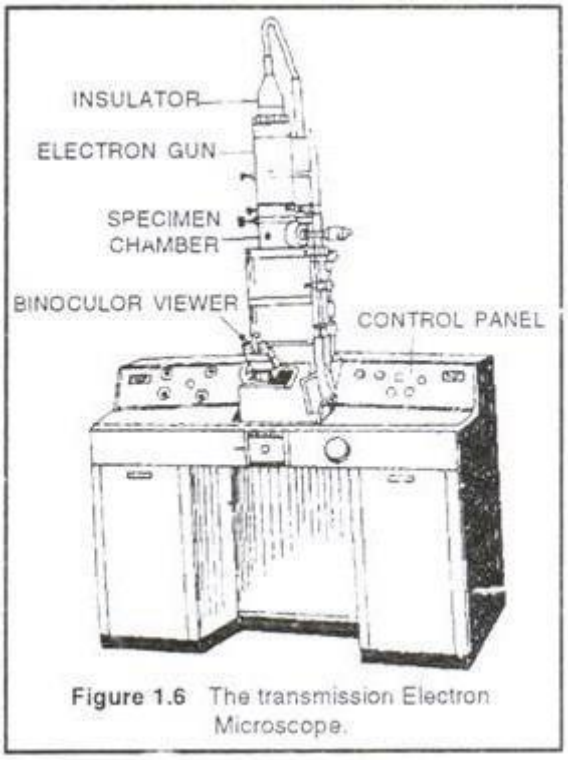


Figure 1.6 The transmission Electron Microscope.

(a) Electron gun:

It consists of a tungsten filament or cathode that emits electrons on receiving high voltage electric current (50,000-100,000 volts). Near the top of the tube is an anode which attracts electrons.

(b) Ray tube (Microscope Column):

It is a high vacuum metal tube (2mt. high) through which electrons travel.

(c) Condense lens:

It is the electromagnetic coil which focuses the electron beam in the plane of the specimen.

(d) Objective lens:

It is the electromagnetic coil which produces the first magnified image formed by the objective lens and produces the final image.

(e) Projector lens*:*

It is also an electromagnetic coil which further magnifies the first image formed by the objective lens and produces the final image. Each electromagnetic coil has a coil of wire encased by a soft iron casing.

(f) Fluorescent Screen or Photographic Film:

Since unaided human eye cannot observe electrons, therefore, a fluorescent screen is used for viewing the final image of the specimen. The final image can be captured on photographic film and die photograph obtained is called an electron micrograph.

Preparation of material for TEM:

The material to be studied under electron microscope must be well preserved, fixed, completely dehydrated, ultrathin and impregnated with heavy metals that sharpen the difference among various organelles. The material is preserved by fixation with glutaraldehyde and then with osmium tetroxide. The fixed material is dehydrated and then embedded in plastic (epoxy resin) and sectioned with the help of diamond or glass razor of ultra-microtome. The sections are ultrathin about 50-100 nm thick. These sections are placed on a copper grid and exposed to electron dense materials like lead acetate, uranylacetate, palladium vapours, phosphotungstate etc. Now the sections can be viewed in the TEM. The coating with electron dense material enables the specimen to withstand electric bombardment.

Shadow Casting:

This technique is used to un-sectioned materials e.g., viral particles. The sample dried on a film supported by a grid is placed in an evaporation chamber. The chamber is evacuated. Heavy metal atoms projected from a glowing filament impinge at a predetermined angle on the film (Fig. 21.4). The metal is deposited as a uniform electron opaque layer on the film. The metal is deposited on one side of the specimen, while the other side lacks the deposition.

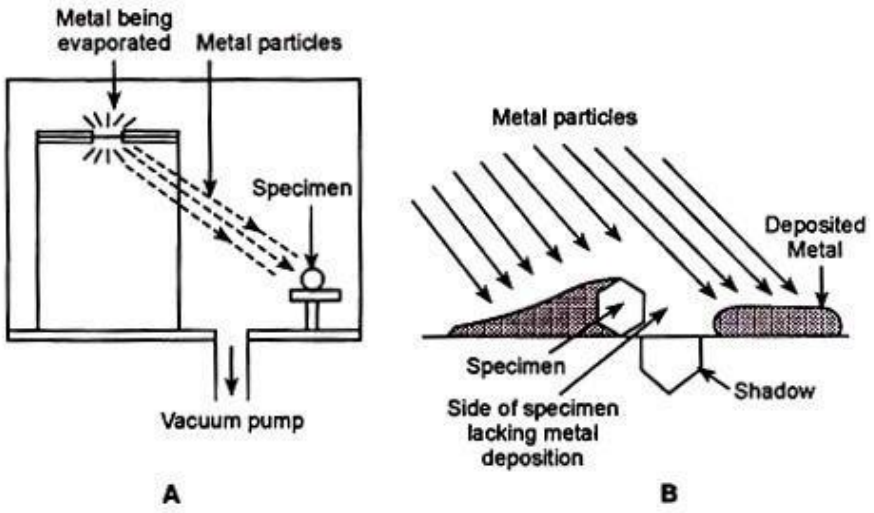


Fig. 21.4. Diagram showing method of shadow casting. **A.** Evaporation chamber. **B.** Shadow of specimen under electron microscope.

Examination under the electron microscope shows the "shadow" of the specimen in the place lacking the deposited metal. The size and shape of shadow provide the information on the 3- dimensional shape of the material.

Positive Staining:

The specimen is stained and than the excess of stain is removed. It gives an unstained background and stained object. Certain viruses can be stained by salts that become absorbed selectively. For example uranyl acetate stains the viral nucleic acid and other components. Abs conjugated to ferritin (electron opaque molecule) stains the protein.

Negative Staining:

Negative staining can be used to study the viral particles and organelles. The viral particles are mixed with salt, such as sodium phosphotungstate which is highly opaque to electrons. The mixture is spread on a carbon membrane and dried.
The regions of the particles which are not penetrated by salt form electron lucent area on an opaque background. Details of the surface structure are revealed by perpetration between protruding parts of the salt.

The Whole Mount Technique:

This technique is also used for the un-sectioned materials, but it does not involve staining or heavy metal deposition. The scattering of electrons from the object produces the image.

The Freeze-etch Technique:

By this technique, a unique picture of cells is viewed, especially where the membrane is involved. The cell is broken along and across the membranes and therefore, it shows the four views of the biological membrane, viz., protoplasmic surface, exoplasmic surface, protoplasmic fracture faces, and exoplasmic fracture faces. The technique does not involve fixatives, stains and embedding agents and therefore, the cell structure is not deformed.
The material to be studied is frozen in liquid Freon in a vacuum. The cell function is instantly arrested due to rapid freezing. The frozen material becomes very hard, and when struck by a knife, it is broken along the lines of membranes.
Water is evaporated by placing the broken material in a vacuum. Water loss causes the "etching effect" i.e., details become much clearer. A heavy metal (e.g., platinum) is used for shadowing the fractured surface, and a replica is prepared by using a carbon film. A strong acid is used to remove the tissue and to leave the metal replica. This metal replica is viewed with the electron microscope.

b. Scanning Electron Microscope:

Scanning electron microscopes combine the mechanism of electron microscopy and television. SEM became commercially available in early 1960's and the researchers were Knoll, Von Ardenne, Zworytein etc.
In SEM, electrons are not transmitted through the very thin specimen from below but impinge on its surface from above. The specimen may be opaque and of any manageable thickness and size. If the specimen is an electron conductor, it needs only to be held on an appropriate support. If it is non-conductor, it is allowed to dry but if moist, freeze dried in liquid nitrogen is necessary. The specimen is then coated with metal vapour (gold) in vacuum. The electrons originate at high energy (20,000 V) from a hot tungsten or lanthanum hexoboride cathode "gun". These electrons are sharply focused, adjusted and narrowed by an arrangement of magnetic fields.
Instead of forming a broad inverted cone of rays, in SEM a needle sharp probe (about 5 – 10 mm in diameter) is made. This primary beam (probe) acts only as an exciter of image forming secondary electrons emerging from the surface of the specimen. The probe scans the specimen like that on a blank TV screen. The probe can impinge on depth and heights with equal speed and accuracy giving great depth of field and producing images with three dimensions. Images are elicited from wherever the probe strikes the metal coated areas of the specimen. Magnification is the ratio of final image to the diameter of area scanned.
Any of the secondary electrons with sufficient energy can emerge from the surface. Those that emerge not too far from the point of impact of the probe can be used to form an image. The useful secondary electrons are magnetically deflected to a collector or

detector. Here, they produce a signal that represents at any single moment, only 5-10 mm area or spot of impingement of the probe on the specimen. The successive signals from the collector are amplified and transmitted to a cathode ray (TV) tube. The scanning beam and TV tube beam are synchronized.

The image scan by the eye on TV screen is thus the sequence of signals representing in araster pattern, the successive areas traversed by the primary probe beam. Exposure may range from a few second to one-half hour or more. The TV image may be photographed, video taped or processed in motion on a computer. This microscope shows 3-dimensional surface architecture of cells and organelles. The present day scanning electron microscopes have the resolution power of 10 nm which is less than the resolution power of transmission electron microscopes. However, this resolving power can be increased by making further improvements. In this system (Fig. 21.5), a beam of electrons is used that moves back and forth across the specimen by a canning coil.

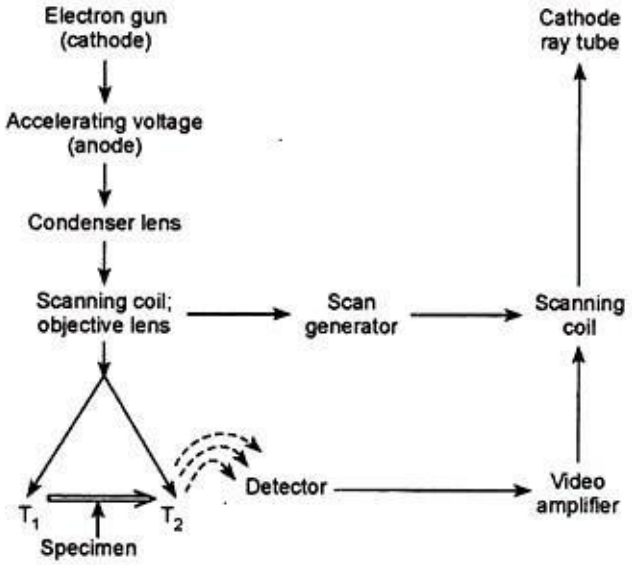


Fig. 21.5. Schematic view of the working system of a scanning electron microscope. The electron beam illuminates different points in the specimen at different times, T_1 and T_2. Movement of this beam is synchronized with beam in cathode ray tube by a scan generator. Detector picks the scattered electrons from specimen and modulates the beam in the cathode ray tube; an the scattered electrons from specimen and modulates the beam in the cathode ray tube; an image of the specimen is formed.

It illuminates different points on the surface of specimen at different times. The scan generator synchronizes the movement of this beam in a cathode ray rube (television tube). Electrons are deflected from the specimen and are picked up by a detector that modulates the beam in the cathode ray tube. A 3-dimensional structure of the surface of the cell or organelle is obtained.

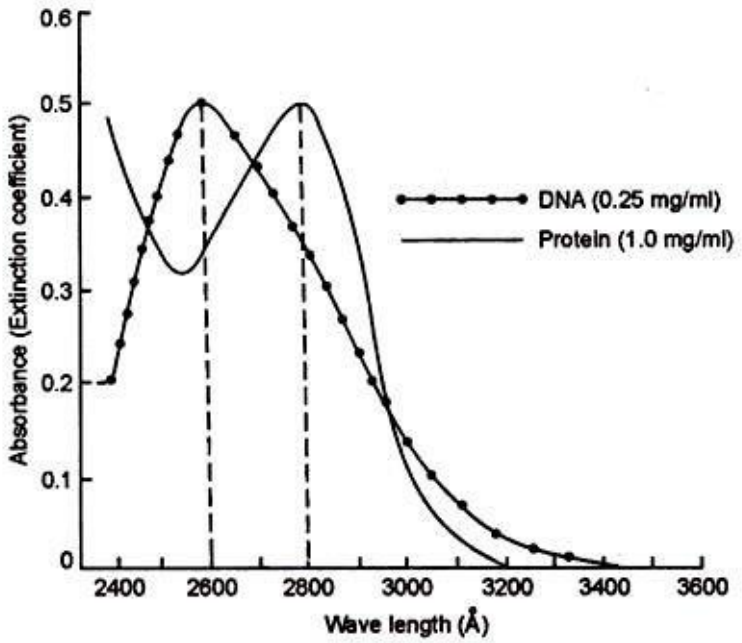


Fig. 21.6. Degree of ultraviolet absorption by DNA and proteins at different wavelengths (10Å = 1 nm).

Difference between Light Microscope vs electron Microscope:

Light Microscope :

1. Visible light is used in this microscope.
2. Source of illumination is situated at the bottom.
3. For magnification in this microscope the lens system consists of glass lenses.
4. The lenses are ocular, objective and condenser
5. The image is either seen with the eye or recorded on a photographic film with a camera in this microscope.

Electron Microscope:

2. Electrons are used in this microscope.
3. Source of illumination is situated at the top in this microscope.
4. The lens system consists of electromagnetic coils in this microscope.
5. This microscope has projector coils, an objective and a condenser.
6. The image in an electron microscope is either recorded on a fluorescent scent screen or recorded on a photographic film.

Difference between TEM and SEM:

Both SEM (scanning electron microscope/microscopy) and TEM (transmission electron microscope/microscopy) refer both to the instrument and the method used in electron microscopy. There are a variety of similarities between the two. Both are types of electron microscopes and give the possibility of seeing, studying, and examining small, subatomic particles or compositions of a sample. Both also use electrons (specifically, electron beams), the negative charge of an atom. Also, both samples in use are required to be "stained" or mixed with a particular element in order to produce images. Images produced from these instruments are highly magnified and have a high resolution. However, an SEM and TEM also share some differences. The method used in SEM is based on scattered electrons while TEM is based on transmitted electrons. The scattered electrons in SEM are classified as backscattered or secondary electrons. However, there is no other classification of electrons in TEM.

The scattered electrons in SEM produced the image of the sample after the microscope collects and counts the scattered electrons. In TEM, electrons are directly pointed toward the sample. The electrons that pass through the sample are the parts that are illuminated in the image. The focus of analysis is also different. SEM focuses on the sample's surface and its composition. On the other hand, TEM seeks to see what is inside or beyond the surface. SEM also shows the sample bit by bit while TEM shows the sample as a whole. SEM also provides a three-dimensional image while TEM delivers a two-dimensional picture.

In terms of magnification and resolution, TEM has an advantage compared to SEM. TEM has up to a 50 million magnification level while SEM only offers 2 million as a maximum level of magnification. The resolution of TEM is 0.5 angstroms while SEM has 0.4 nanometers. However, SEM images have a better depth of field compared to TEM produced images. Another point of difference is the sample thickness, "staining," and preparations. The sample in TEM is cut thinner in contrast to a SEM sample. In addition, an SEM sample is "stained" by an element that captures the scattered electrons.In SEM, the sample is prepared on specialized aluminium stubs and placed on the bottom of the chamber of the instrument. The image of the sample is projected onto the CRT or television-like screen. On the other hand, TEM requires the sample to be prepared in a TEM grid and placed in the middle of the specialized chamber of the microscope. The image is produced by the microscope via fluorescent screens. Another feature of SEM is that the area where the sample is placed can be rotated in different angles. TEM was developed earlier than SEM. TEM was invented by Max Knoll and Ernst Ruska in 1931. Meanwhile, SEM was created in 1942. It was developed at a later time due to the complexity of the machine's scanning process.

Draw backs of EM:

(i) It is complicated and costly.

(ii) There is risk of radiation leak,

(iii) It requires very high voltage electric current.

(iv) A cooling system is required,

(v) The specimen or object has to be given special treatment including complete dehydration.

Probable Questions:

1. Describe basic components of a electron microscope.
2. What is shadow casting in electron microscopy?
3. What is positive and negative staining in electron microscopy?
4. Write five differences between light microscope and electron microscope.
5. Differentiate between TEM and SEM.
6. Describe the working principle of FACS with suitable diagram.

Suggested Readings:

3. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
4. Biophysical Chemistry by Upadhyay and Upadhyay.

UNIT-XI

Basic concept of flow cytometry

Objective: In this unit we will discuss about fluorescent activated cell sorting or FACS.

Basic principles of flow cytometry:

Flow cytometry is used when there is a need to profile a large number of different cell types in a population. The cells are separated on the basis of differences in size and morphology. Additionally, fluorescently-tagged antibodies that target specific antigens on the cell surface can be used to identify and segregate various sub-populations.
The basic steps include passing the cells through a narrow channel, such that each cell is illuminated by a laser one at a time. A series of sensors then detect the refracted or emitted light, and this data is integrated and compiled to generate information about the sample.

Flow cytometry principle - How the fluorescence activated cell sorting (FACS) work ?

Flow cytometry is a technique to identify and isolate cells from a mixture of other cells using fluorescence activity. Flow cytometry was developed by Fulwyler in 1965. Till today it is used for research in cell biology. In that technique cell sorting and cell counting was done by using laser light technology.
There are different steps involved in a process of flow cytometer; First step is Flow of cell in that liquid containing cells i.e. liquid stream is passing single file through light beam of laser light for sensing. Second step is that measuring system, which commonly used for measurement of conductivity, Optical system containing Mercury and Xenon lamp resulting in light signal. The third step is to detection of light scattering, in that step light signal are converted analogue to digital signal with the help of Analogue to digital conversion system. It will detect Forward scatter light (FSC) and Side scatter light as well as fluorescence signal from light in to electrical signal that can be processed by computer. The fourth step is that analysis of signal by computer, in that collecting of data from sample using cytometer this collecting of data is termed as Acquisition. This acquisition is carried out by computer connected to the flow cytometer software. This software handle the digital interface with cytometer, it is able to adjusting the parameter required for the voltage compensation. It is also monitor initial sample analysis.
Fluorescence labelled antibodies was developed for clinical research. In modern instrument contain multiple laser and fluorescence detector, currently in industrial instrument ten laser and 18 fluorescence detector. More number of detector and laser

allow for multiple antibody labelling and identify a target population by their marker. In certain instrument can even take digital image of individual cell for the analysis.

The data is to be analysed by data generated by computer either in histogram or dot plot. Computer analysis give automated population identification, this automated identification could potentially help finding of rare hidden population. Fluorescence-activated cell sorting (FACS) is a specialized type of flow cytometry. It provides a method for sorting a heterogeneous mixture of biological cells into two or more containers, one cell at a time, based upon the specific light scattering and fluorescent characteristics of each cell. It is a useful scientific instrument as it provides fast, objective and quantitative recording of fluorescent signals from individual cells as well as physical separation of cells of particular interest. The technique was expanded by Len Herzenberg, who was responsible for coining the term FACS.

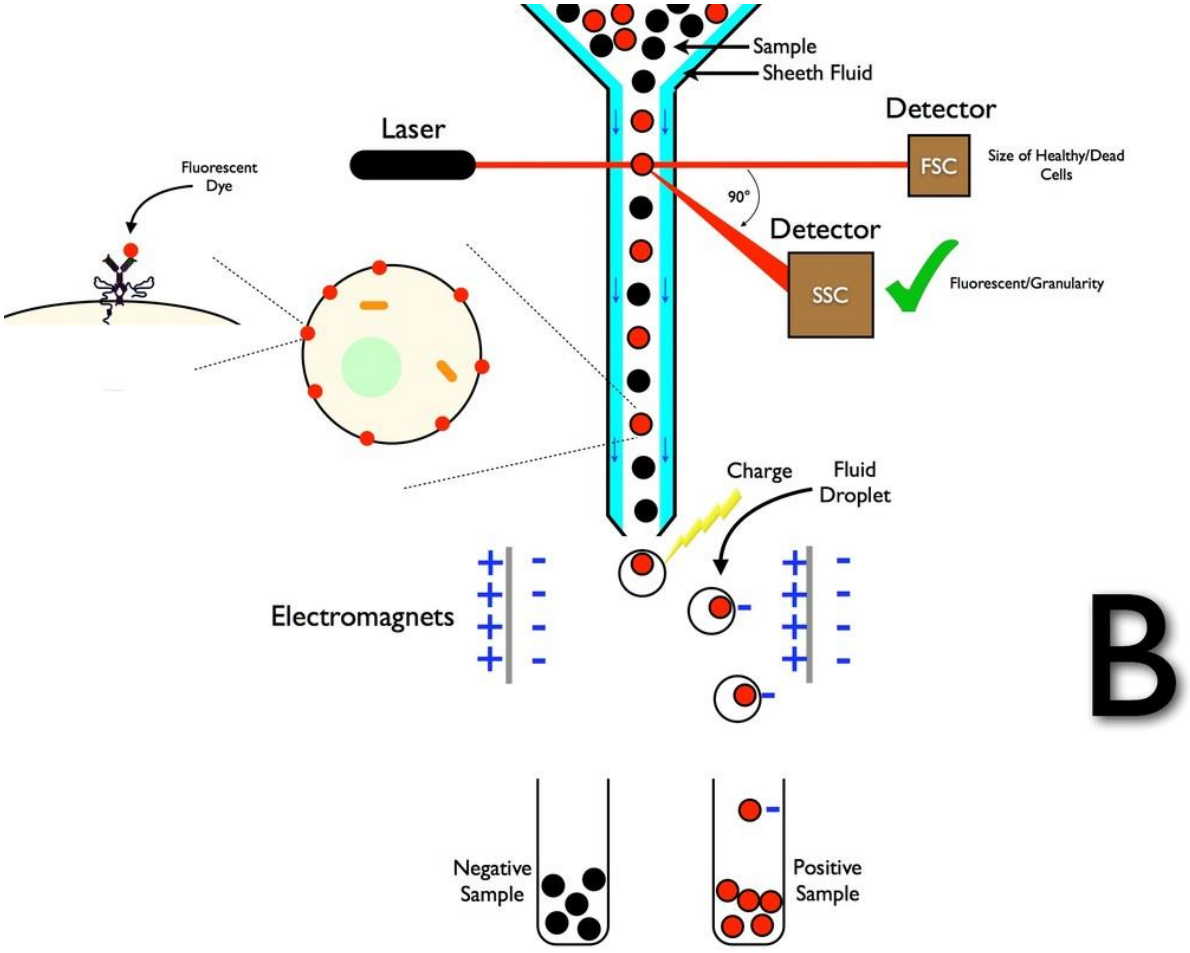


Fig- Fluorescence-Activated Cell Sorting (FACS) principle

The cell suspension is entrained in the center of a narrow, rapidly flowing stream of liquid. The flow is arranged so that there is a large separation between cells relative to their diameter. A vibrating mechanism causes the stream of cells to break into individual droplets. The system is adjusted so that there is a low probability of more than one cell per droplet. Just before the stream breaks into droplets, the flow passes through a fluorescence measuring station where the fluorescent character of interest of each cell is measured. An electrical charging ring is placed just at the point where the stream breaks into droplets. A charge is placed on the ring based on the immediately prior fluorescence intensity measurement, and the opposite charge is trapped on the droplet as it breaks from the stream. The charged droplets then fall through an electrostatic deflection system that diverts droplets into containers based upon their charge. In some systems, the charge is applied directly to the stream, and the droplet breaking off retains charge of the same sign as the stream. The stream is then returned to neutral after the droplet breaks off The technology has applications in a number of fields, including molecular biology, pathology, immunology, plant biology and marine biology It has broad application in medicine (especially in transplantation, hematology, tumor immunology and chemotherapy, prenatal diagnosis, genetics and sperm sorting for sex preselection). Also, it is extensively used in research for the detection of DNA damage, caspase cleavage and apoptosis In marine biology, the autofluorescent properties of photosynthetic plankton can be exploited by flow cytometry in order to characterise abundance and community structure. In protein engineering, flow cytometry is used in conjunction with yeast display and bacterial display to identify cell surface-displayed protein variants with desired properties

What can flow cytometry be used to measure?

Flow cytometry helps to analyse several parameters of a cell simultaneously. Some of these parameters are described below:

Functional analysis

This method can determine biological activity inside cells, such as the generation of reactive oxygen species, mitochondrial membrane changes during apoptosis, phagocytosis rates in labelled bacteria, native calcium content, and changing metal content in response to drugs, etc.

Determining cell viability

This method can also be used to assess cell viability after the addition of pathogenic organisms or drugs. Any breach in cell membrane integrity can be determined using dyes

that can enter the punctured cell membrane. Fluorescent probes such as bis-oxonol can bind to proteins present on the cell membrane, allowing for the identification of various stages of necrosis.

Measuring apoptosis and necrosis

Apoptosis or programmed cell death is accompanied by characteristic changes in cell shape, loss of structures, cell detachment, condensation of the cytoplasm, cell shrinkage, phagocytosis of cellular residues and changes in the nuclear envelope.
Some of the biochemical changes include proteolysis, DNA denaturation, cell dehydration, protein cross-linking, and a rise in the free calcium ions. These physical and biochemical changes can be detected using flow cytometry.
Oncosis is a necrotic event where the cell starts to swell rather than shrink. This leads to rupture of the plasma membrane and release of proteolytic enzymes that can also damage the surrounding tissues. These changes in the plasma membrane and cell shape can be assessed using flow cytometry.

Cell cycle analysis:

The amount of DNA present in the nucleus varies during each phase of the cell cycle. This variation in DNA content can be assessed using fluorescent dyes that bind to DNA or monoclonal antibodies, which can allow the detection of antigen expression.
Other factors including the content of cell pigments such as chlorophyll, DNA copy number variation, intracellular antigens, enzymatic activity, oxidative bursts, glutathione, and cell adherence can similarly be measured using this method.

Data interpretation

Each cell that passes through the laser light is detected as a separate event. Also, different types of detected light: forward-scatter, side-scatter, and specific wavelengths of fluorescence emission, is assigned a distinct channel. The data for each of these events is plotted independently and can be represented by two methods: histograms and dot-plots.
Histograms compare a single parameter, where intensity is plotted on one axis and the number of events is plotted on a separate axis. Dot-plots can compare more than one parameter simultaneously, where each event is displayed as a single point and the intensity values of two or three channels are represented on the various axis.
In this scenario, events that have similar intensities cluster together in the dot plot. While dot-plots can compare multiple parameters together, histograms are easier to read and understand. In many cases, dot-plots and histograms are not mutually exclusive, and in many flow cytometry experiments both types of graphs are plotted to represent and assess multi-parametric data.

Electrostatic Cell Sorting:

A major application of flow cytometry is to separate cells according to subtype or epitope expression for further biological studies. This process is called cell sorting or FACSTM analysis. After the sample is hydro-dynamically focused, each particle is probed with a beam of light. The scatter and fluorescence signal is compared to the sort criteria set on the instrument.

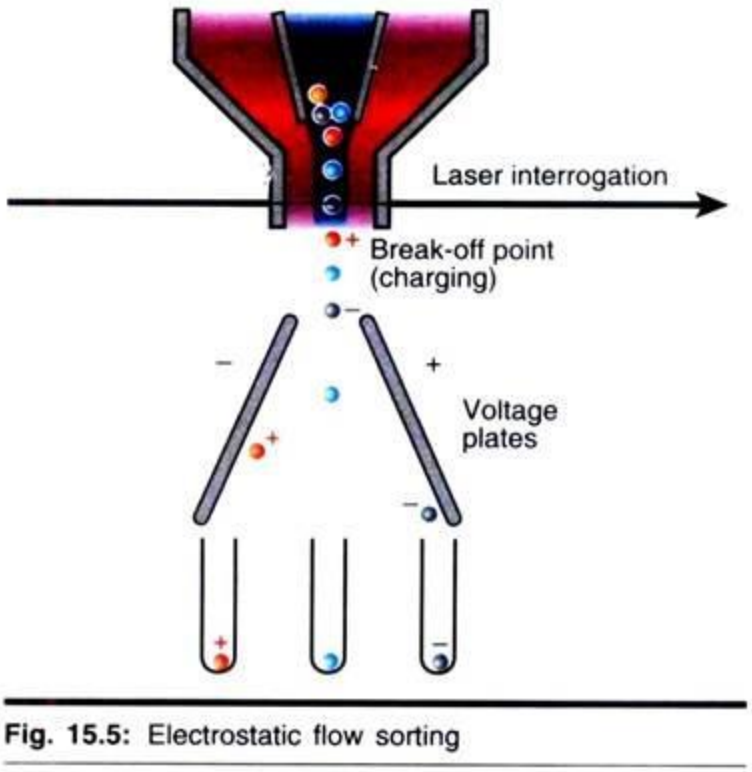


Fig. 15.5: Electrostatic flow sorting

If the particle matches the selection criteria, the fluid stream is charged as it exits the nozzle of the fluidics system. Electrostatic charging actually occurs at a precise moment called the 'break-off point', which describes the instant droplet containing the particle of interest separated from the stream. To prevent the break-off point happen-ing at random distances from the nozzle and to maintain consistent droplet sizes, the nozzle- is vibrated at high frequency. The droplets eventually pass through a strong electrostatic field, and are deflected left or right based on their charge (Fig. 15.4). The speed of flow sorting depends on several factors including particle size and the rate of droplet formation. A typical nozzle is between 50-70 μm in diameter and, depending on the jet velocity from it, can produce 30,000-100,000 droplets per second, which is ideal for accurate sorting. Higher jet

velocities risk the nozzle becoming blocked and will also decrease the purity of the preparation.

Sample Preparation:

Single cells must be suspended at a density of 105-107 cells/ml to prevent the narrow bores of the flow cytometer and its tubing from clogging up. The concentration also influences the rate of flow sorting, which typically progresses at 2000-20,000 cells/second. However, higher sort speeds ma; decrease the purity of the preparation.

Phosphate buffered saline (PBS) is a common suspension buffer and the most straightforward samples for flow cytometry include non-adherent cells from culture, water-borne micro-organisms, bacteria and yeast. Even whole blood is easy to use – red cells are usually removed by a simple lysis step; it is then possible to quickly identify lymphocytes, granulocytes and monocytes by their FSC/SSC characteristics.

However, researchers may also wish to analyze cells from solid tissues, e.g., liver or tumours. In order to produce single cells, the solid material must be disaggregated. This can be done either mechanically or enzymatically. Mechanical disaggregation is suitable for loosely bound structures, e.g., adherent cells from culture, bone marrow and lymphoid tissue. It involves passing a suspension of chopped tissue through a fine-gauge needle several times, followed by grinding and sonication as necessary.

Enzymes are used to disrupt protein-protein interactions and the extracellular matrix that hold cells together. Their action is dependent on factors including pH, temperature and co-factors, so care must be taken when choosing an enzyme. For example, pepsin works optimally between pH 1.5-2.5 but the acidic conditions would damage cells if left un-neutralized for too long, and cell surface antigens of interest may be lost.

Chelators like EDTA and EGTA can remove divalent cations responsible for maintaining cell function and integrity but their presence may inhibit certain enzymes, for instance, collagenase re quires Ca^{2+} for activity. Enzymatic and mechanical disaggregation is often a trial and error process to optimize the isolation of the epitope under investigation.
To study intracellular components, e.g. cytokines by flow cytometry, the plasma membrane of the cell must be permeabilized to allow dyes or antibody molecules through while retaining the cell's overall integrity. Low concentrations (up to 0.1%) of non-ionic detergents like saponin are suitable. In summary, the method for sample preparation will depend on the starting material and the nature of the epitope.

Optics and Detection: After hydrodynamic focusing, each particle passes through one or more beams of light. Light scattering or fluorescence emission (if the particle is labelled with a fluorochrome) provides information about the particle's properties. The laser and the arc lamp are the most commonly used light sources in modern flow cytometry. Lasers produce a single wavelength of light (a laser line) at one or more discrete frequencies (coherent light). Arc lamps tend to be less expensive than lasers and exploit the colour emissions of an ignited gas within a sealed tube. However, this produces unstable incoherent light of a mixture of wavelengths, which needs subsequent optical filtering. Light that is scattered in the forward direction, typically up to 20″ offset from the laser beam's axis, is collected by a lens known as the forward scatter channel (FSC). The FSC intensity roughly equates to the particle's size and can also be used to distinguish between cellular debris and living cells. Light measured approximately at a 90° angle to the excitation line is called side scatter. The side scatter channel (SSC) provides information about the granular content within a particle. Both FSC and SSC are unique for every particle, and a combination of the two may be used to differentiate different cell types in a heterogeneous sample.

Fluorescence measurements taken at different wavelengths can provide quantitative and qualitative data about fluorochrome-labelled cell surface receptors or intracellular molecules such as DNA and cytokines. Flow cytometers use separate fluorescence (FL) channels to detect light emitted. The number of detectors will vary according to the machine and its manufacturer. Detectors are either silicon photodiodes or photomultiplier tubes (PMTs). Silicon photodiodes are usually used to measure forward scatter when the signal is strong. PMTs are more sensitive instruments and are ideal for scatter and fluorescence readings.

The specificity of detection is controlled by optical filters, which block certain wavelengths while transmitting (passing) others. There are three major filter types. 'Long pass' filters allow through light above a cut-off wavelength, 'short pass' permit light below a cut-off wavelength and 'band pass' transmit light within a specified narrow range of wavelengths (termed a bandwidth).

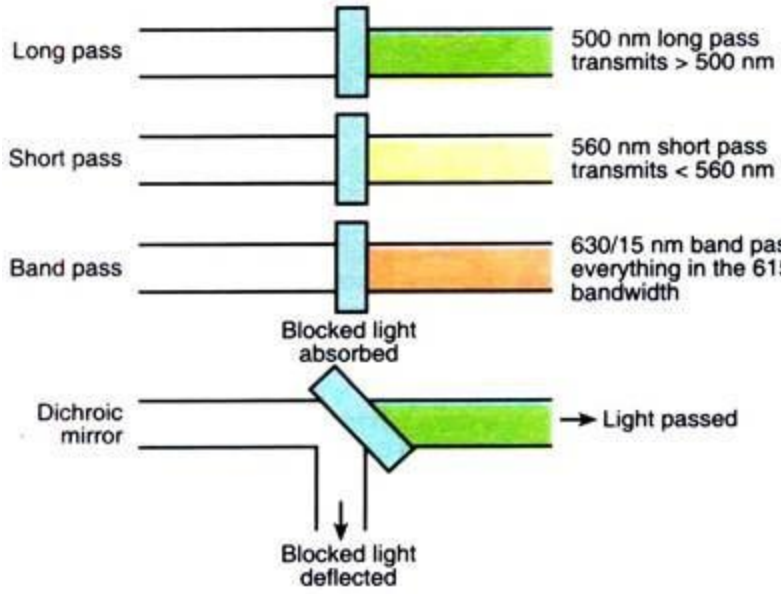


Fig. 15.3: Different types of optical filters

All these filters block light by absorption. When a filter is placed at a 45° angle to the oncoming light it becomes a dichroic filter/mirror. As the name suggests, this type of filter performs two functions; first, to pass specified wavelengths in the forward direction and, second, to deflect blocked light at a 90° angle. To detect multiple signals simultaneously, the precise choice and order of optical filters will be an important consideration.

Signal Processing:

When light hits a photo detector a small current (a few microamperes) is generated. Its associated voltage has an amplitude proportional to the total number of light photons received by the detector. This voltage is then amplified by a series of linear or logarithmic amplifiers, and by analog to digital convertors (ADCs), into electrical signals large enough (5-10 volts) to be plotted graphically.

Log amplification is normally used for fluorescence studies because it expands weak signals and compresses strong signals, resulting in a distribution that is easy to display on a histogram Linear scaling is preferable where there is not such a broad range of signals, e.g., in DNA analysis.

The measurement from each detector is referred to as a 'parameter', e.g., forward scatter, side scatter or fluorescence. The data acquired in each parameter are known as the 'events' and refer to the number of cells displaying the physical feature or marker of interest.

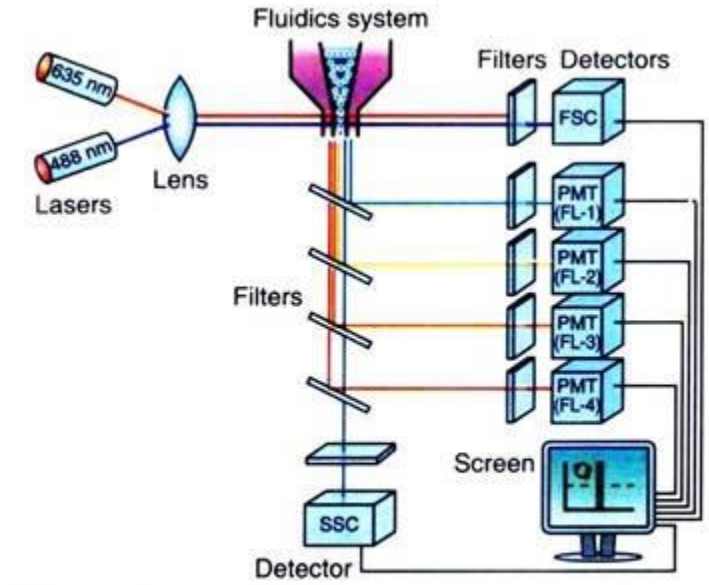


Fig. 15.4: Schematic overview of a typical flow cytometer set-up

By treating appropriately processed blood samples with a fluorescently labelled antibody and performing flow cytometric analysis, one can obtain the following information:

i. How many cells express the target antigen as an absolute number and also as a percentage of cells passing the beam. For example, if one uses a fluorescent antibody specific for an antigen present on all T cells, it would be possible to determine the percentage of T cells in the total white blood cell population. Then, using the cell-sorting capabilities of the flow cytometer, it would be possible to isolate the T-cell fraction of the leukocyte.

ii. The distribution of cells in a sample population according to antigen densities as determined by fluorescence intensity. It is thus possible to obtain a measure of the distribution of antigen density within the population of cells that possess the antigen. This is a powerful feature of the instrument, since the same type of cell may express different levels of antigen depending upon its developmental or physiological state.

iii. The size of cells. This information is derived from analysis of the light-scattering properties of members of the cell population under examination. Flow cytometry also makes it possible to analyze cell populations that have been labelled with two or even three different fluorescent antibodies. For example, if a blood sample is reacted with a fluorescein-tagged antibody specific for T cells, and also with a phycoerythrin-tagged

antibody specific for B cells, the percentages of B and T cells may be determined simultaneously with a single analysis.

Numerous variations of such "two-colour" analyses are routinely carried out, and "three-colour" experiments are common. Aided by appropriate software, highly sophisticated versions of the flow cytometer can even perform "five-colour" analyses.

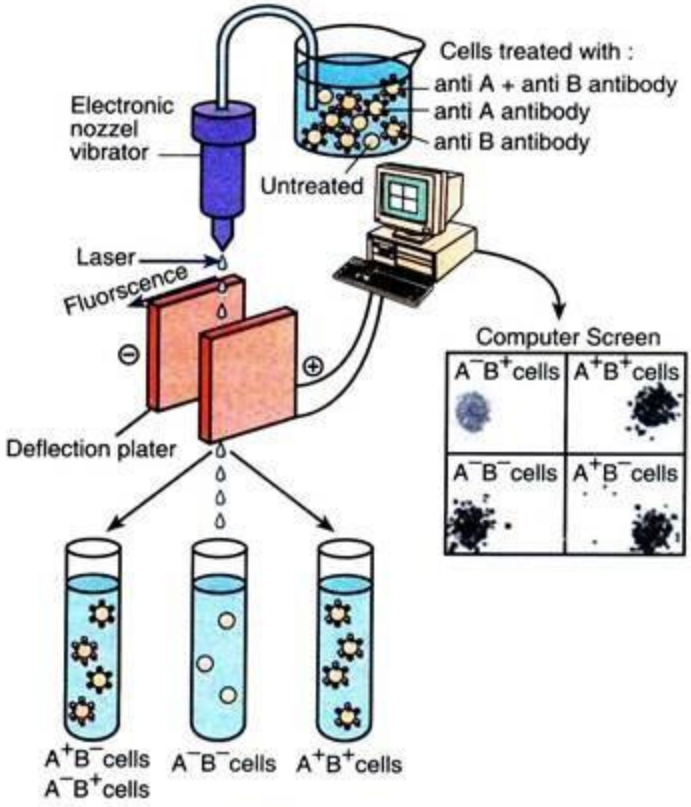


Fig. 15.1: Separation of fluorochrome-labelled cells with the flow cytometer. In the example shown, a mixed cell population is stained with two antibodies, one specific for surface antigen A and the other specific for surface antigen B. The anti-A antibodies are labelled with fluorescein (green) and the anti-B antibodies with rhodamine (red). The stained cells are loaded into the sample chamber of the cytometer. The cells are expelled, one at a time, from a small vibrating nozzle that generates microdroplets, each containing no more than a single cell. As it leaves the nozzle, each droplet receives a small electrical charge, and the computer that controls the flow cytometer can detect exactly when a drop generated by the nozzle passes through the beam of laser light that excites the fluorochrome. The intensity of the fluorescence emitted by each droplet that contains a cell is monitored by a detector and displayed on a computer screen. Because the computer tracks the position of each droplet, it is possible to determine when a particular droplet will arrive between the deflection plates. By applying a momentary charge to the deflection plates when a droplet is passing between them, it is possible to deflect the path of a particular droplet into one or another collecting vessel. This allows the sorting of a population of cells into subpopulations having different profiles of surface markers.

Immunophenotyping:

All normal cells express a variety of cell surface markers, dependent on the specific cell type and degree of maturation. However, abnormal growth may interfere with the natural expression of markers resulting in overexpression of some and under-representation of others.

Flow cytometry can be used to immunophenotype cells and thereby distinguish between healthy and diseased cells. It is unsurprising that today immunophenotyping is one of the major clinical applications of flow cytometry, and is used to aid the diagnosis of myelomas, lymphomas and leukemia's. It can also be used to monitor the effectiveness of clinical treatments.

The differences between the blood profiles of a healthy individual and one suffering from leukemia, for instance, are very dramatic. This can be seen from the FSC vs. SSC plots in Fig. 15.17. In the healthy person the cell types are clearly defined, whereas blood from a leukemia patient is abnormal and does not follow the classic profile.

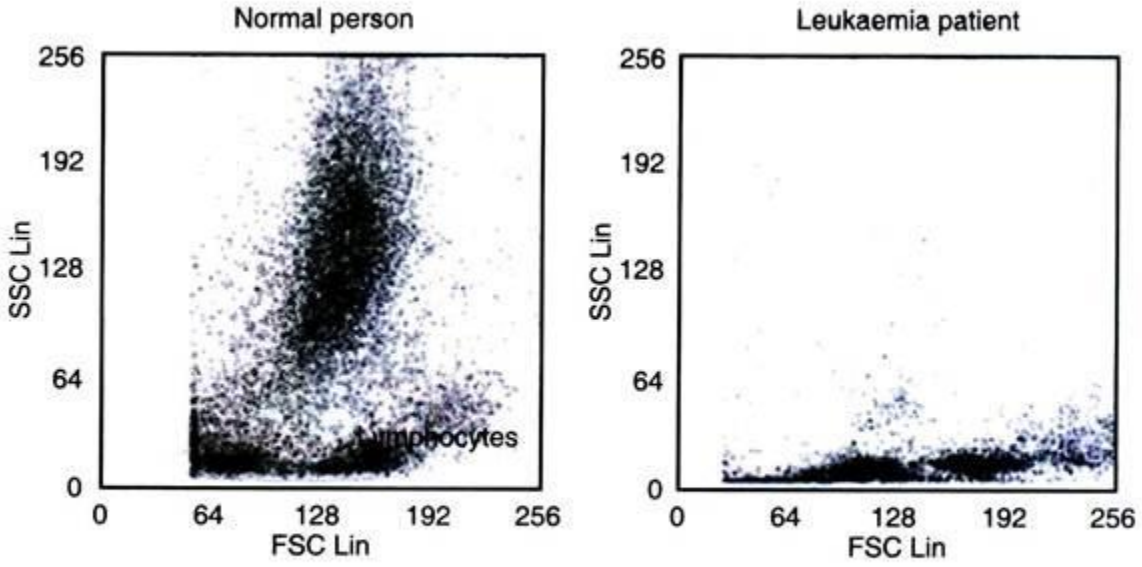


Fig. 15.17: Immunophenotyping

Testing the patient's lymphocytes for specific cell surface markers also reveals more about the condition.

CD3:

A normal person has a significant proportion of CD3-positive lymphocytes. In the patient with leukemia, staining for CDS is absent.

CD20:

In the leukemia patient there are a large number of cells staining positive for CD20. In the healthy person only a few stain positive.

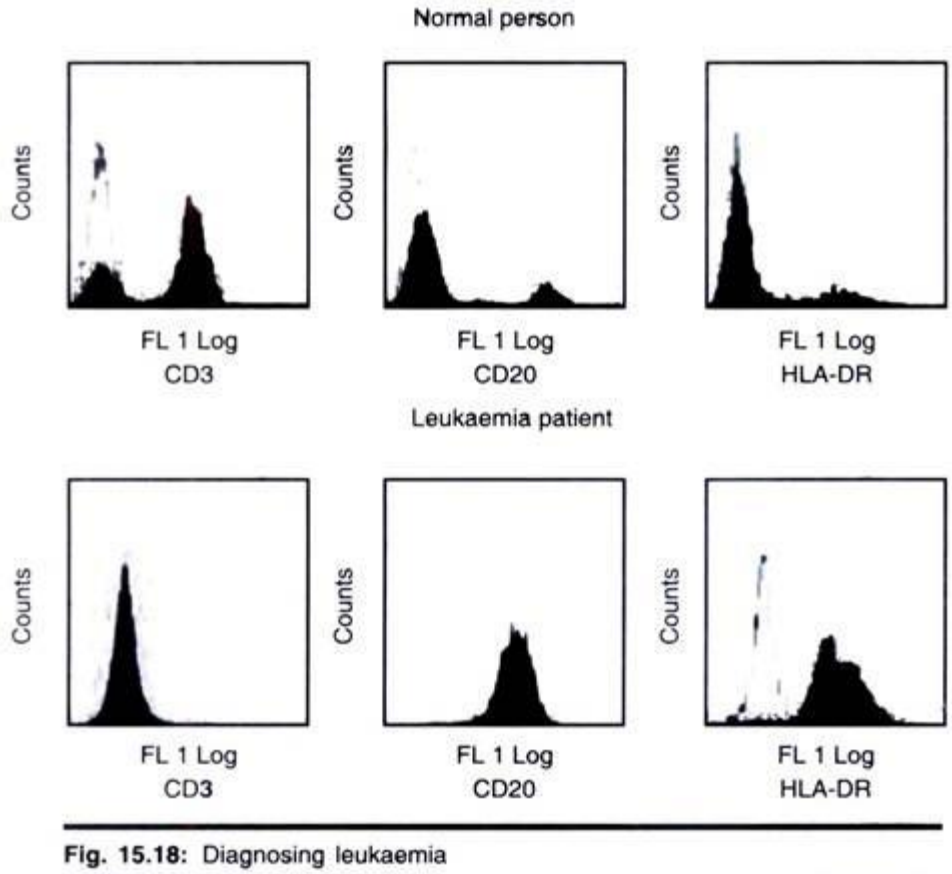


Fig. 15.18: Diagnosing leukaemia

HLA-DR:

The leukemia patient is HLA-DR-positive. In the normal person only a small number of cells stain positive. Being CD3-negative, CD20-positive and HLA-DR-positive, a clinician could diagnose with certainty that this patient is suffering from a B cell lineage leukemia or lymphoma. The precise classification of disease may be determined using further antibodies.

Probable Questions:

1. Describe the working principle of FACS with suitable diagram.
2. Write different utility of FACS.
3. How sample preparation is done in FACS?
4. Discuss about immunotyping.
5. How signals are detected in FACS?
6. Discuss about optics and detection system of FACS.

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

Unit-XII

Cell fractionation methods: a) Preparative Ultracentrifugation b) Gradient Centrifugation

Objective: In this unit we will discuss different cell fractionation methods such as preparative ultracentrifugation and gradient centrifugation

Basic Principle of Centrifugation Technique:

The principle of the centrifugation technique is to separate the particles suspended in liquid media under the influence of a centrifugal field. These are placed either in tubes or bottles in a rotor in the centrifuge. Particles differing in sizes, shape and density are separated as their sedimentation rate is different.

The centrifugal force is generated by rotating the rotor of the centrifuge at a high speed. Besides normal and high speed centrifuge there is a very high speed centrifuge known as Ultracentrifuge, which is developed by Theodor Svedberg in 1940.

This instrument is designed to produce centrifugal forces up to several hundred thousand times which can separate and purify subcellular organelles, proteins, nucleic acids and several macromolecules. Thus the ultracentrifuge has opened up a new line in many types of fundamental studies in Cell Biology, Biochemistry and Molecular Biology.

The rate of sedimentation of a particle in a centrifugal force can be shown in the following:

$$\frac{dx}{dt} = \frac{2r^2(P_p - P_m)}{gn} \qquad (1)$$

where $\frac{dx}{dt}$ = the rate at which the particle moves towards the centrifugal field
r = radius of the particle
P_p = densities of the particle
P_m = densities of the medium
n = viscosity of the suspending medium
g = centrifugal force

Now, the particles will not move if the densities of the particles and the medium are equal. If the densities of the particles are greater than the medium, they will move toward the bottom of the centrifuge tube while they will remain at the top of the tube, if the particles are lighter than the medium.

The centrifugal force produced by the centrifuge is measured by the gravity units as:

$$G = w^2 r \qquad (2)$$

where G = centrifugal field
w = angular velocity of the rotor
r = radial distance of the particle from the axis of the rotation

Again, angular velocity of the rotor is known as in the formula:

$$w = \frac{2\text{rev/min}}{60} \qquad (3)$$

Putting the value of w in the equation (2), the centrifugal field (G) will be

$$G = \frac{4^2\ (\text{revolution/min})^2\ r}{60 \times 60}$$

$$\therefore\ G = \frac{4^2\ (\text{rev/min})^2\ r}{3,600}$$

Which is generally shown as a multiple of g, i.e., gravitational field of the earth. Sometimes it is also expressed as R.C.F (Relative Centrifugal Field) which is the ratio of the weight of the particle in the centrifugal field to the weight of the same particle acted on by gravity. On the basis of this principle of separation, particles are separated depending on their densities, size, centrifugal force, time of separation etc. Different cell components are separated in the following order—whole cells and cell debris first followed by nuclei, plastids, mitochondria, lysosomes, microsomes, fragments of endoplasmic reticulum and ribosome.
The method of separation becomes complicated when the particles are not spherical, which requires some complicated formula for calculation. In case of Ordinary rotors as used in the preparative centrifuge, the centrifugal field does not remain uniform, because the radial dimension of a particle will vary according to the position in the centrifuge tube (r_{min} and r_{max}). The particle will have a greater centrifugal field as it is further away from the axis of rotation. This occurs both in the fixed angle and swing- away rotor (Fig. 7.1). Hence, the centrifugal field is calculated from the average radius of rotation ($r_{av.}$) of the column of liquid in the centrifuge tube. The details of maximum and the method of calculation of relative centrifugal field (R.C.F.) are generally given in the manual of the centrifuge.

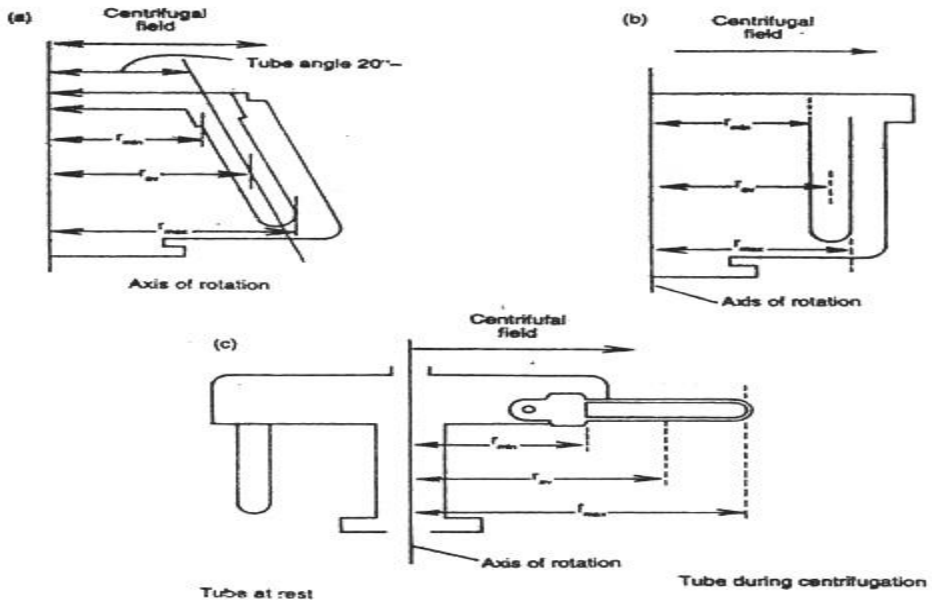


Fig. 7.1: Diagrammatic representation of: (a) fixed angle; (b) vertical tube and (c) swinging bracket.

The sedimentation rate of a particle can also be expressed as sedimentation coefficient (s) which is the sedimentation rate per unit of centrifugal field. The sedimentation values depend on the solvent-solute systems. As the sedimentation coefficient of many of macromolecules is very small, the basic unit is taken as $10\sim^{13}$ seconds and is designated as Svedberg unit (S). For example, the Ribosomal RNA showing sedimentation values as 5 x 106^{-13} seconds is said to be 5S (5 Svedberg units).

Sedimentation coefficients of some of the macromolecules are shown in Table 7.1.

Table 7.1: Sedimentation coefficients of some macromolecules

Macromolecules	Range of sedimentation coefficients in S
Soluble proteins	2 to 25 S
Nucleic acids	5 to 100 S
Ribosome	20 to 200 S
Viruses	40 to 1,000 S
Lysosomes	4,000 S
Mitochondria	20×10^3 S to 70×10^3 S
Nuclei	$4,000 \times 10^3$ S to $40,000 \times 10^3$ S

Types of Centrifugation Technique:
There are generally 4 types of centrifuges:

(1) Clinical Bench Centrifuges,

(2) High Speed Refrigerated Centrifuges,

(3) Continuous flow Centrifuges, and

(4) Ultracentrifuges.

The last type can again be classified into two types:

i. Preparative and

ii. Analytical Ultracentrifuges.

1. Clinical Bench Type Centrifuge:

These are the most simple type of centrifuge used in many laboratories for routine type of work, particularly for sedimenting yeast cells, blood cells or any coarse and medium particles. The maximum speed of this type of centrifuge is between 4,000 to 6,000 r.p.m. with a 'g' of 3,000 to 7,000.Sometimes the cooling arrangement can also be made in this type of centrifuge. Now some 'Microfuges' are available where maximum speed of 8,000 to 13,000 r.p.m. with 'g' value of about 10,000 can be made using small Eppendorf tubes.

2. High Speed Refrigerated Centrifuge:

These instruments are used to isolate organelles, to purify and isolate soluble proteins, microorganisms etc. with a speed of about 25,000 r.p.m. having 'g' value of 60,000. Both fixed angle and swing-out rotors can be used here. But this speed of the instrument is not sufficient for centrifuging ribosomes and viruses.

3. Continuous Flow Centrifuge:

It is also one type of high speed centrifuge where the rotor is slightly modified or specially designed one. In this type there is a continuous flow of the medium in the centrifuge tube. Here the cells or particles axe sedimented against the wall and the excess medium or liquid

comes out through the exit tube. Cells can be harvested continuous from a large volume of the culture medium.

4. Ultracentrifuges:

(a) Preparative Ultracentrifuge:

This is a type of instrument where actual isolation, purification of macromolecules or cell organelles can be done. It operates in refriger-ated condition under vacuum to avoid frictional resistance of the rotor caused by the spinning air.

Cellular and subcellular fractionation techniques are indispensable methods used in biochemical research. Although the proper separation of many subcellular structures is absolutely dependent on preparative ultracentrifugation, the isolation of large cellular structures, the nuclear fraction, mitochondria, chloroplasts or large protein precipitates can be achieved by conventional high-speed refrigerated centrifugation. Differential centrifugation is based upon the differences in the sedimentation rate of biological particles of different size and density. Crude tissue homogenates containing organelles, membrane vesicles and other structural fragments are divided into different fractions by the stepwise increase of the applied centrifugal field. Following the initial sedimentation of the largest particles of a homogenate (such as cellular debris) by centrifugation, various biological structures or aggregates are separated into pellet and supernatant fractions, depending upon the speed and time of individual centrifugation steps and the density and relative size of the particles. To increase the yield of membrane structures and protein aggregates released, cellular debris pellets are often rehomogenised several times and then recentrifuged. This is especially important in the case of rigid biological structures such as muscular or connective tissues, or in the case of small tissue samples as is the case with human biopsy material or primary cell cultures.

The differential sedimentation of a particulate suspension in a centrifugal field is diagrammatically shown in Fig. 3.4a. Initially all particles of a homogenate are evenly distributed throughout the centrifuge tube and then move down the tube at their respective sedimentation rate during centrifugation. The largest class of particles forms a pellet on the bottom of the centrifuge tube, leaving smaller-sized structures within the supernatant. However, during the initial centrifugation step smaller particles also become entrapped in the pellet causing a certain degree of contamination. At the end of each differential centrifugation step, the pellet and supernatant fraction are carefully separated from each other. To minimise cross-contamination, pellets are usually washed several times by resuspension in buffer and recentrifugation under the same conditions. However, repeated washing steps may considerably reduce the yield of the final pellet fraction, and

are therefore omitted in preparations with limiting starting material. Resulting supernatant fractions are centrifuged at a higher speed and for a longer time to separate medium-sized and small-sized particles. With respect to the separation of organelles and membrane vesicles, crude differential centrifugation techniques can be conveniently employed to isolate intact mitochondria and microsomes.

Density-gradient centrifugation:

To further separate biological particles of similar size but differing density, ultracentrifugation with preformed or self-establishing density gradients is the method of choice. Both rate separation or equilibrium methods can be used. In Fig. 3.4b, the preparative ultracentrifugation of low- to high-density particles is shown. A mixture of particles, such as is present in a heterogeneous microsomal membrane preparation, is layered on top of a preformed liquid density gradient. Depending on the particular biological application, a great variety of gradient materials are available. Caesium chloride is widely used for the banding of DNA and the isolation of plasmids, nucleoproteins and viruses. Sodium bromide and sodium iodide are employed for the fractionation of lipoproteins and the banding of DNA or RNA molecules, respectively.

Various companies offer a range of gradient material for the separation of whole cells and subcellular particles, e.g. Percoll, Ficoll, Dextran, Metrizamide and Nycodenz. For the separation of membrane vesicles derived from tissue homogenates, ultra-pure DNase, RNase and protease-free sucrose represents a suitable and widely employed medium for the preparation of stable gradients. If one wants to separate all membrane species spanning the whole range of particle densities, the maximum density of the gradient must exceed the density of the most dense vesicle species. Both step gradient and continuous gradient systems are employed to achieve this. If automated gradient makers are not available, which is probably the case in most undergraduate practical classes, the manual pouring of a stepwise gradient with the help of a pipette is not so time-consuming or difficult. In contrast, the formation of a stable continuous gradient is much more challenging and requires a commercially available gradient maker. Following pouring, gradients are usually kept in a cold room for temperature equilibration and are moved extremely slowly in special holders so as to avoid mixing of different gradient layers. For rate separation of subcellular particles, the required fraction does not reach its isopycnic position within the gradient. For isopycnic separation, density centrifugation is continued until the buoyant density of the particle of interest and the density of the gradient are equal.

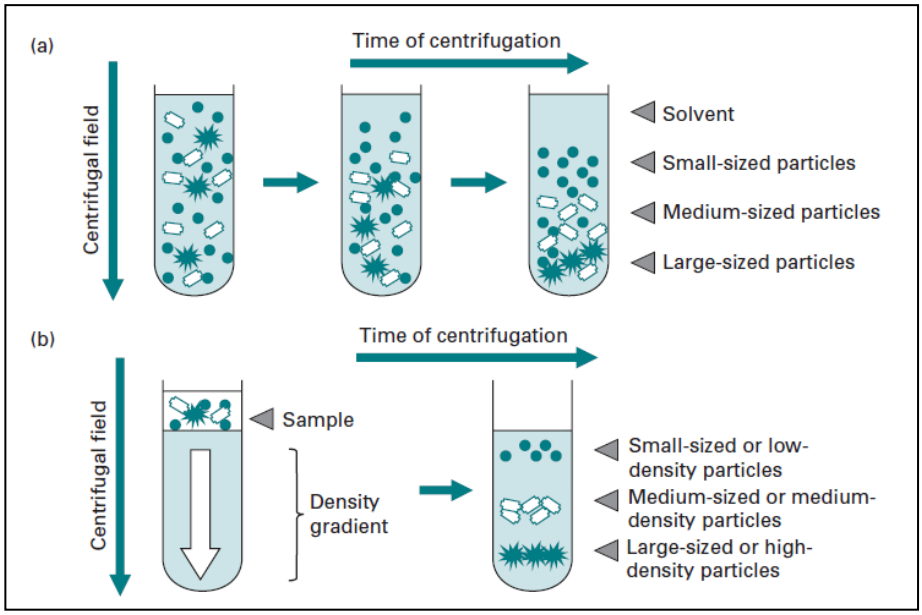


Fig. 3.4 Diagram of particle behaviour during differential and isopycnic separation. During differential sedimentation (a) of a particulate suspension in a centrifugal field, the movement of particles is dependent upon their density, shape and size. For separation of biological particles using a density gradient (b), samples are carefully layered on top of a preformed density gradient prior to centrifugation. For isopycnic separation, centrifugation is continued until the desired particles have reached their isopycnic position in the liquid density gradient. In contrast, during rate separation, the required fraction does not reach its isopycnic position during the centrifugation run.

Practical applications of preparative centrifugation:

To illustrate practical applications of differential centrifugation, density gradient ultracentrifugation and affinity methodology, the isolation of the microsomal fraction from muscle homogenates and subsequent separation of membrane vesicles with a differing density is described, the isolation of highly purified sarcolemma vesicles outlined, and the sub-fractionation of liver mitochondrial membrane systems shown. Skeletal muscle fibres are highly specialised structures involved in contraction and the membrane systems that maintain the regulation of excitation–contraction coupling, energy metabolism and the stabilisation of the cell periphery are diagrammatically shown in Fig. 3.5a. The surface membrane consists of the sarcolemma and its invaginations, the transverse tubular membrane system. The transverse tubules may be subdivided into the non-junctional region and the triad part that forms contact zones with the terminal cisternae of the sarcoplasmic reticulum. Motor neuron-induced depolarisation of the sarcolemma travels

into the transverse tubules and activates a voltage-sensing receptor complex that directly initiates the transient opening of a junctional calcium release channel. The membrane system that provides the luminal ion reservoir for the regulatory calcium cycling process is represented by the specialised endoplasmic reticulum. It forms membranous sheaths around the contractile apparatus whereby the longitudinal tubules are mainly involved in the uptake of calcium ions during muscle relaxation and the terminal cisternae provide the rapid calcium release mechanism that initiates muscle contraction. Mitochondria are the site of oxidative phosphorylation and exhibit a complex system of inner and outer membranes involved in energy metabolism.

For the optimum homogenisation of tissue specimens, mincing of tissue has to be performed in the presence of a biological buffer system that exhibits the right pH value, salt concentration, stabilising co-factors and chelating agents. The optimum ratio between the wet weight of tissue and buffer volume as well as the temperature (usually 4 ^{o}C) and presence of a protease inhibitor cocktail is also essential to minimize proteolytic degradation. Prior to the 1970s, researchers did not widely use protease inhibitors or chelating agents in their homogenisation buffers. This resulted in the degradation of many high-molecular-mass proteins. Since protective measures against endogenous enzymes have been routinely introduced into subcellular fractionation protocols, extremely large proteins have been isolated in their intact form, such as 427 kDa dystrophin, the 565 kDa ryanodine receptor, 800 kDa nebulin and the longest known polypeptide, of 2200 kDa, named titin. Commercially available protease inhibitor cocktails usually exhibit a broad specificity for the inhibition of cysteine proteases, serine-proteases, aspartic-proteases, metallo-proteases and amino-peptidases. They are used in the micromolar concentration range and are best added to buffer systems just prior to the tissue homogenisation process. Depending on the half-life of specific protease inhibitors, the length of a subcellular fractionation protocol and the amount of endogenous enzymes present in individual fractions, tissue suspensions might have to be replenished with a fresh aliquot of a protease inhibitor cocktail. Protease inhibitor kits for the creation of individualised cocktails are also available and consist of substances such as trypsin inhibitor, E-64, aminoethyl-benzenesulfonyl fluoride, antipain, aprotinin, benzamidine, bestatin, chymostatin, E-aminocaproic acid, N-ethylmaleimide, leupeptin, phosphoramidon and pepstatin. The most commonly used chelators of divalent cations for the inhibition of degrading enzymes such as metallo-proteases are EDTA and EGTA.

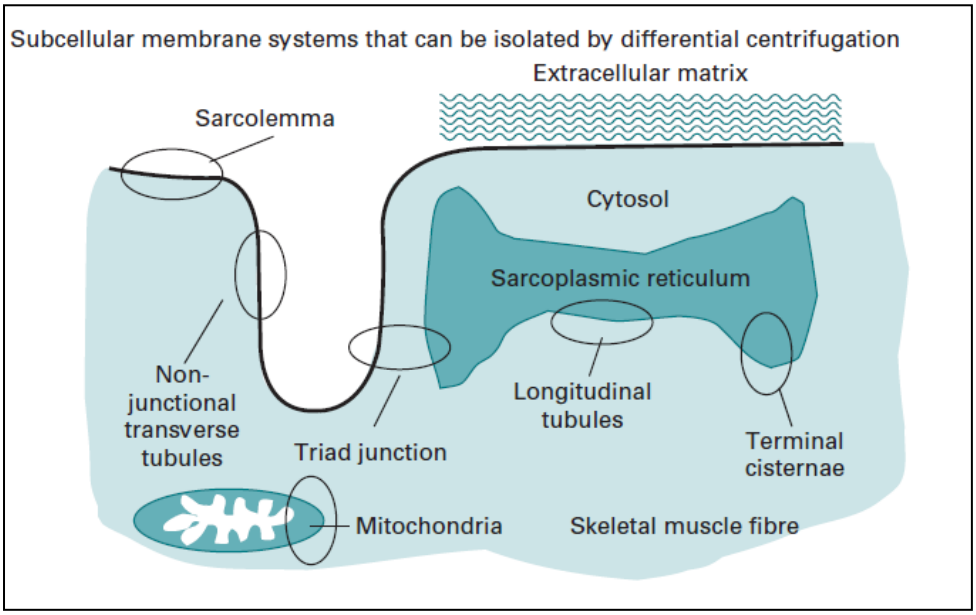


Figure: Scheme of the fractionation of skeletal muscle homogenate into various subcellular fractions. Shown is a diagrammatic presentation of the subcellular membrane system from skeletal muscle fibres.

(b) Analytical Ultracentrifuge:

This instrument has many applications in the fundamental studies of macromolecules show-ing the molecular weight, purity and shape of the material. It runs at a speed of about 70-80,000 r.p.m. with about 500,000 g and consists of a specially designed rotor in a special rotor chamber which remains under vacuum at low temperature.

There is an arrangement of a special optical system to determine the concentration distributions within the sample during centrifugation. There are two special optical cells on the rotor, known as the Ana-lytical cell and the Counterpoise cell (Fig. 7.2). There are two holes (Reference holes) in the counterpoise cell for the calibration of distances in the analytical cell. The rotor chamber has an upper and lower lens and the upper lens is joined with a camera lens which emits lights on the photographic plate. Light from the light source comes through the bottom.

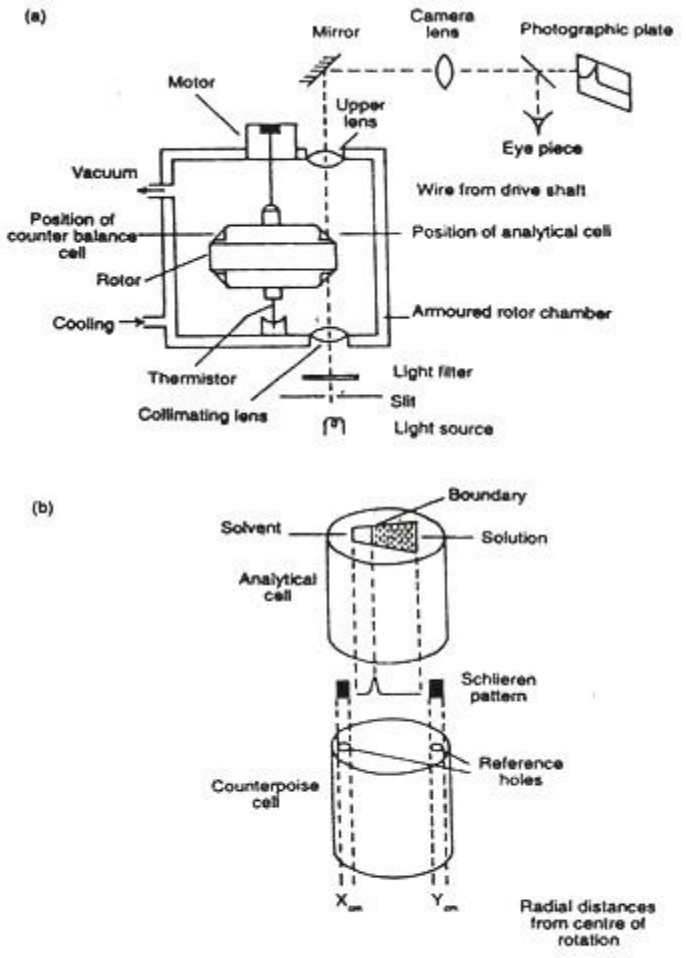


Fig. 7.2: Schematic representation of Analytical ultracentrifuge (a) and (b). Details of Analytical and counterpoise cell.

The principle of monitoring in this system is done either through the ultraviolet absorption system or by noting the differences in the refrac-tive index. If the concentration is uniform, light passes through it without any deviation. But if the light passes through a solution of different density zones, it is refracted at the boundary between these zones.

By measuring the re-fractive index (Fig. 7.3) between the reference solvent and the solution, the concentration of solute at any point can be measured. In recent models, the photographic plate system has been replaced by electronic scanning system which can directly measure and plot the concentration of the sample at all points in the analytical cell.

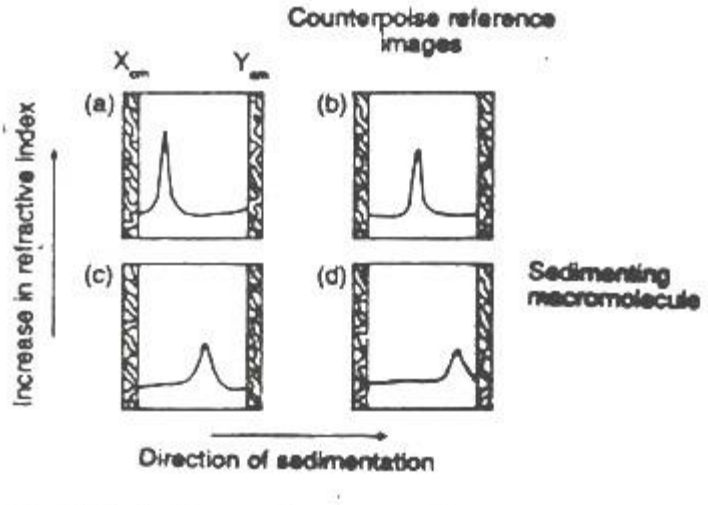


Fig. 7.3: Sedimentation pattern in the screening system of Analytical ultracentrifuge.

Centrifuge tubes are either manufactured in hard glass or with polypropylene, polycarbon-ate, stainless steel and nylon materials. Generally, centrifugation is done to separate the particles or cells on the basis of their size, length or mass. But in some cases, separation is done on the basis of the density of the particles. When the shape (size) and density of some macromolecules are same then these macromolecules can be separated from each other according to mass. This type of sepa-ration through centrifugation is known as Rate Zonal Centrifugation [Fig. 7.4(a)]. As the name signifies, different-sized molecules will occupy different zones in a centrifuge tube after cen-trifugation.

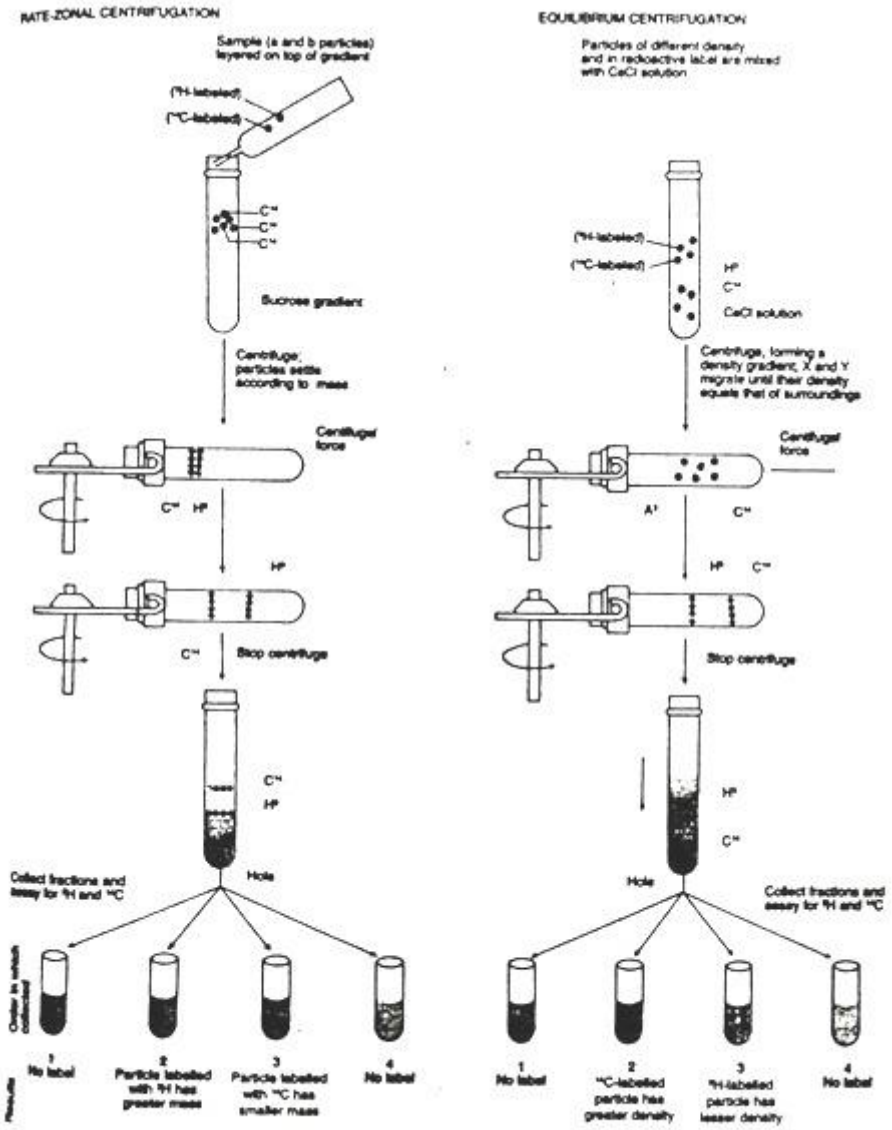


Fig. 7.4: Separation process: (a) by Rate-Zonal centrifugation, and (b) by Equilibrium centrifugation

This separation of macromolecules at different zones is stabilized by using Sucrose gradient. So, this technique is also known as Sucrose Density Gradient Centrifugation. This is used for separating all types of particles and organelles.

In analytical ultracentrifuge the rate of sedimentation can be measured by taking photographs of the moving boundaries of sedimenting particles. To separate particles of different densities, they are labelled with H3 or C14 and then centrifuged in Sucrose density gradient. Particles labelled with H3 have the greater mass and sediment faster than those labelled with C14. The second type of density gradient cen-trifugation is known as Equilibrium Density Gradient Centrifugation [Fig. 7.4(b)] where the density gradient is formed during centrifuga-tion. The material used in this process is the aqueous solution of

Caesium chloride (CsCl). After a run in the ultracentrifuge, it will form 0.92g/ml heavier at the bottom than at the top permitting the separation of particles which differ in density by even a fraction of 0.02g/ml. The densities of protein, DNA and RNA are 1.3, 1.6 to 1.7 and 1.75 to 1.8 g/ml, respec-tively. The same chemical CsCl can separate the different macromolecules like DNA and RNA due to the fact that Cs+ binds to DNA at phosphate groups, while it binds to RNA both at phosphates and at the hydroxyl groups of sugar thus increasing the density of RNA more than that of DNA.

Different isotopes are used with different labelling precursors to alter their densities which will be helpful for the separation of these macromolecules on the basis of density. Gradient forming chemicals commonly used are Caesium and Rubidium chloride, Sucrose, some proteins and polysaccharides, colloidal sil-ica (Percoll, Ludox), Metrizamide, Nycodenz, Renograffin etc. (Table 7.2). Of them, Sucrose is most commonly used in Density Gradient centrifugation because of being very viscous even at 10% concentrations. Ficoll (copolymer of Sucrose and epichlorhydrin) is used for the separation of whole cells and cellular organelles.

Table 7.2: Use of some common gradient chemicals

Chemicals	Ionic strength	Density of aqueous solution $g.cm^{-3}$	Common uses
Caesium chloride	High	1.91	Separating DNA, nucleoproteins, viruses, isolation of plasmid
Caesium sulphate	High	2.01	Separation DNA, RNA, etc.
Sodium bromide	High	1.53	Fractionation of Lipoproteins
Glycerol	Non-ionic	1.26	Separation of membrane fragments, protein
Sucrose	Non-ionic	1.32	Separation of subcellular particles, proteins, nucleic acids etc.
Ficoll	Non-ionic	1.17	Separation of cells, nucleic acids, organelles etc.
Dextran	Non-ionic	1.13	Cells, microsomes etc.
Percoll	Non-ionic	1.30	Cells and organelles
Metrizamide	Non-ionic	1.46	Cells, organelles, nuclei membrane
Nycodenz	Non-ionic	1.42	Cells, organelles, membranes, nucleoproteins, viruses etc.

CsCl solutions are used for Equilibrium Den-sity Gradient or isopycnic separation of nucleic acids. After density gradient centrifugation, gener-ally the visual bands separating the particles are collected with the help of the hypodermic needle or syringe. Sometimes the centrifuge tube is punctured at the base by a fine needle. As the drops of the liquid come out through the needle they may be collected and analysed using ultraviolet spectrophotometer.

Preparative vs Analytical Centrifugation:

Centrifugation is a separation method in which the application of the centrifugal force sediments or fractionates a heterogeneous mixture. Microcentrifugation, high-velocity centrifugation, and ultracentrifugation are the three types of centrifugation based on the volume and speed used in the process. Among other centrifugation methods, ultracentrifugation uses the maximum angular velocity. The rotating speed of ultracentrifugation is as high as 1 000 000 g. Thus, ultracentrifugation is used to isolate small particles such as ribosomes, proteins, and viruses. Preparative and analytical centrifugationare the two types of ultracentrifugation methods. The **main difference** between preparative and analytical centrifugation **is that** preparative centrifugation is used in pelleting small materials such as membranes, organelles, viruses, **DNA and RNA** whereas analytical centrifugation is used to determine the mass and shape of **macromolecules** such as protein complexes and rate of sedimentation of molecules**.**

Probable Questions:

1. Describe basic principle of centrifugation technique.
2. What is preparative centrifugation? Explain.
3. What is density gradient centrifugation? Explain.
4. Describe practical applications of preparative centrifugation.
5. Compare preparative vs analytical centrifugation.

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

Unit-XIII

Separation of Cell Constituents by Chromatography: ion exchange; gel filtration and HPLC

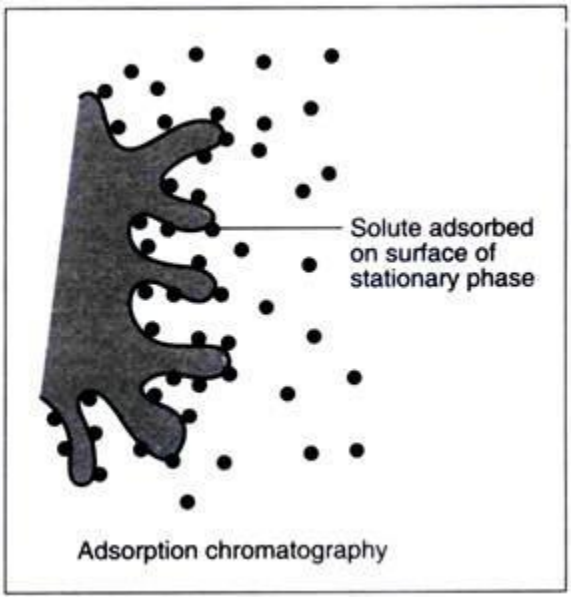
Objective: In this unit you will learn how cell constituents van be separated by the process of Chromatography

Chromatgraphy:

Chromatography was originally introduced by Tswett in 1906, a Polish Botanist, for separation of different colour pigments present in the plant extract. Amino acids can also be separated from one another by partition chromatography.

Adsorption Chromatography:

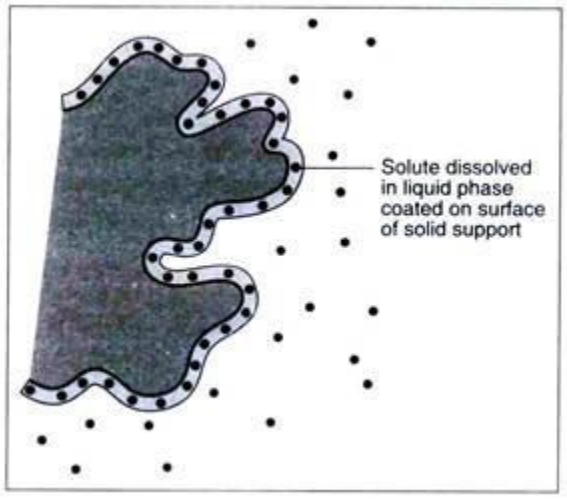
Adsorption chromatography is probably one of the oldest types of chromatography around. It utilizes a mobile liquid or gaseous phase that is adsorbed onto the surface of a stationary solid phase. The equilibration between the mobile and stationary phases accounts for the separation of different solutes.



Adsorption chromatography

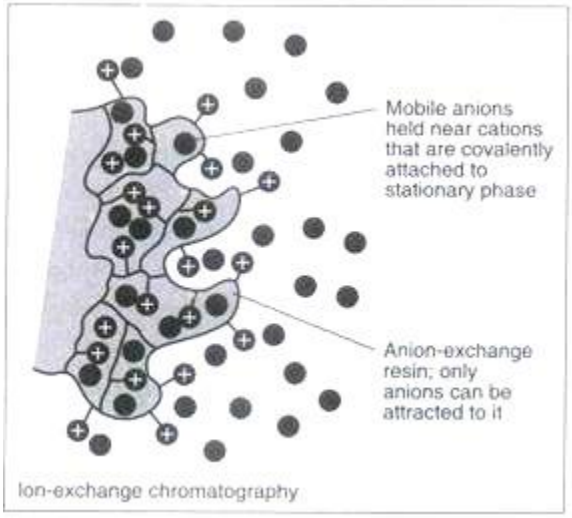
Partition Chromatography:

This form of chromatography is based on a thin film formed on the surface of a solid support by a liquid stationary phase. Solute equilibrates between the mobile phase and the stationary liquid.



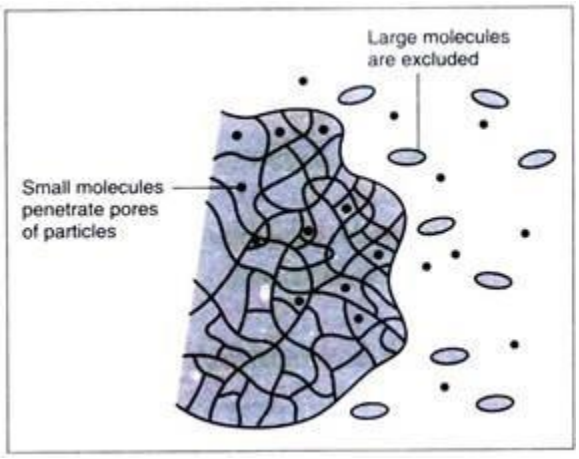
Ion Exchange Chromatography:

In this type of chromatography, the use of a resin (the stationary solid phase) is used to covalently attach anions or cations onto it. Solute ions of the opposite charge in the mobile liquid phase are attracted to the resin by electrostatic forces.



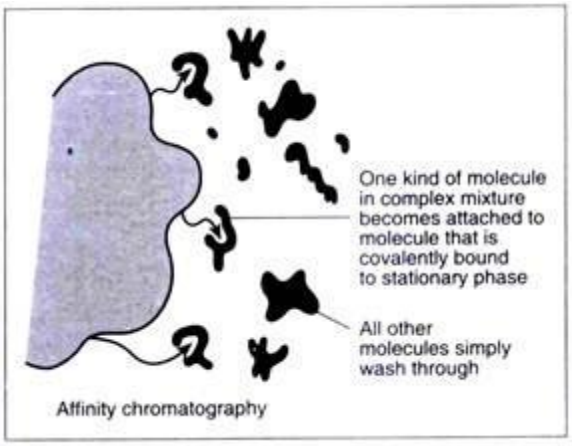
Molecular Exclusion Chromatography:

Also known as gel permeation or gel filtration, this type of chromatography lacks an attractive interaction between the stationary phase and solute. The liquid or gaseous phase passes through a porous gel which separates the molecules according to its size. The pores are normally small and exclude the larger solute molecules, but allow smaller molecules to enter the gel, causing them to flow through a larger volume. This causes the larger molecules to pass through the column at a faster rate than the smaller ones.



Affinity Chromatography:

This is the most selective type of chromatography employed. It utilizes the specific interaction between one kind of solute molecule and a second molecule that is immobilized on a stationary phase. For example, the immobilized molecule may be an antibody to some specific protein. When solute containing a mixture of proteins are passed by this molecule, only the specific protein is reacted to this antibody, binding it to the stationary phase. This protein is later extracted by changing the ionic strength or pH.



Chromatography might be of different types, such as:

(a) Paper,

(b) Thin-layer,

(c) Column

The amino acids are separated between a stationary and a mobile phase. In paper chromatography (Fig. 2.2), a drop of amino acid mixture is placed on a filter paper and allowed to dry. The paper is then kept in contact with a suitable solvent which is allowed to flow over the dried drop slowly either by capillary action alone (ascending chromatography) or in combination with gravitational force (descending chromatography). As the solvent moves, it carries along with it the individual amino acids. Suitable tests are then applied to localise the individual amino acids which have been found to be carried away to a characteristic distance from the original place of application. The ratio of the distance travelled by the compound to the distance covered by the solvent on paper is called R_F value of the compound.

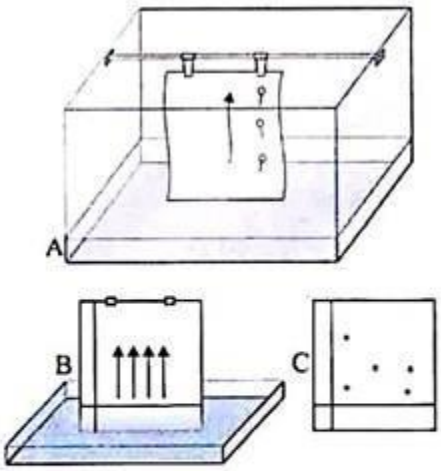


Fig. 2.2 Separation of amino acid mixture by paper chromatography. (A) placed in a suitable solvent the substances migrate upwards. (B & C), separation achieved by turning the paper 90° and by using a different solvent

To have a clearer separation, an improved method, the two-dimensional chromatography (Fig. 2.3) has been developed. Here after suitable chromatographic procedure, the paper is allowed to re-chromatograph at a right angle to the first one.

In thin-layer chromatography (TLC), a suitable adsorbent like alumina, cellulose powder, etc., is spread on glass plates which are then used as 'paper in chromatographic separation. This technique has some special advantages in separation.

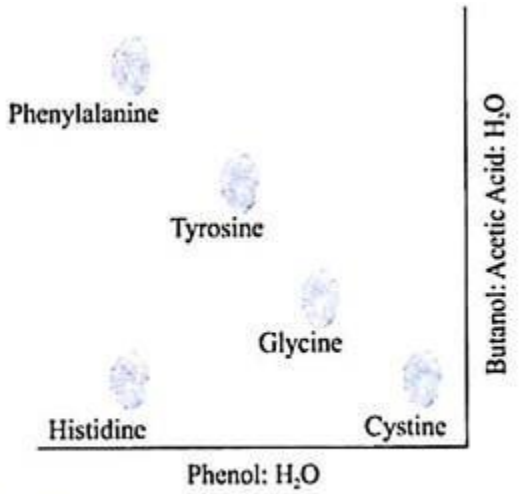


Fig. 2.3 Two-dimensional paper chromatography of some amino acids.

In column chromatography or ion-exchange chromatography, various solid adsorbents like starch, cellulose or an ion- pig. 2.4 exchange resins are placed as columns in glass tubes (Fig. 2.4). Different amino acids (or other substances) are adsorbed on the column are then eluted with suitable solvents.

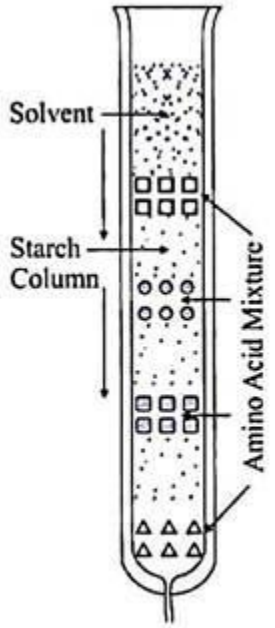


Fig. 2.4 Column chromatographic separation of amino acid mixture

A. Partition Chromatography:

This term covers liquid-liquid partition chromatography (Paper chromatography and thin layer chromatography). Gel chromatography, Gas liquid chromatography.

Paper Chromatography of Amino Acid Mixture:

Introduction:

In biology and chemistry, it is most often necessary to separate components of a mixture which are very similar and are difficult to separate by chemical or physical methods. Chromatography and electrophoresis arc two powerful modern methods utilised for such purpose.

Tswett (1906) the Russian biologist, first appreciated the possibilities of chromatography and he put the term "Chromatography". Consden. Gorden, and Martin (1944) described paper chromatography in which separations were done mainly by partition.

Principle:

The separation of components of a mixture by a chromatographic system depends on multiple containing the mixture of compounds is put on a strip of filter paper and allowed to dry. A partition process. Small differences in partitioning of each component of a mixture are multiplied many fold. The greater such differences the greater is the ease of separation.

A small drop of solution suitable solvent (mixture of two solvents) is allowed to flow along the filter paper over this spot. The substances in the initial spot are extracted by the flowing solvent and carried forward along the filter paper to a distance which appears related to their partition coefficient between the free and bound solvent phases of the filter paper. After the solvent has run for a suitable distance along the paper, the paper is removed, dried and subjected to suitable tests to locate the various compounds. R_f value is defined as the ratio of the distance travelled by the component to the distance covered by the solvent. R_f value depends on the nature of the solvent, the temperature, and the presence of other substances.

Procedure:

Whatmann filter paper is cut into 35×15 cm sheet.

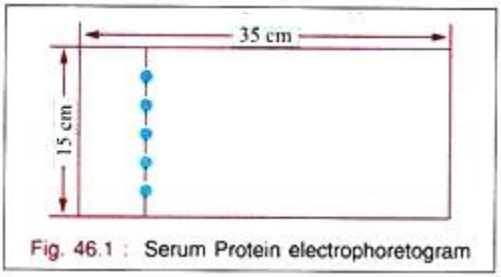


Fig. 46.1 : Serum Protein electrophoretogram

A pencil line is drawn about 3 cm above the shorter edge of the paper and 5 points are marked at equal spacing leaving 2.5 cm from the two edges. On the middle point, a mixture of four amino acids is applied with the help of a fine capillary. It is dried with the help of hot air blown by a hair dryer.

Again, another small quantity of the mixture is spotted at the same place and dried. This process is repeated 2 to 3 times more. On the other points, the individual amino acids of the mixture are similarly applied. The positions of these amino acids are marked with a pencil.

The solvent for developing the chromatogram is a mixture of n-butanol: acetic acid: water (12:3: 5) respectively. This mixture is freshly prepared. If a chromatographic tank is not available, specimen jar (about 40 x 15 cm) with fitting lid can very well serve the purpose.

In such ajar, put about 100 to 150 ml of the solvent mixture and replace the lid so that the lid is airtight. In case of doubts, apply Vaseline to the lid to avoid leakage. In an hour, the inside atmosphere will be saturated with the solvent vapour. Now fold the paper in which sample has been applied in the shape of a cylinder and tie the opposing ends of the paper with staples or thread. Open the lid and place this folded paper in upright position in the jar, the pencil line lower most and about a centimeter above the solvent. Replace the lid. The paper should stand absolutely vertically. Leave the chromatogram to develop for 10 to 15 hours or earlier if the solvent has ascended quite near the upper margin of the paper.

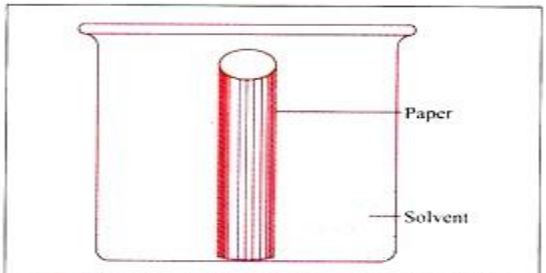


Fig. 46.2 : The chromatographic jar containing the solvent and the paper

Take out the paper at the desired time, cut the stitches and let it dry completely in the air. After the paper has dried thoroughly, the location reagent (0.2 per cent ninhydrin in acetone) is sprayed uniformly on the paper with the help of an all glass sprayer.
The paper is then allowed to dry first in the air and then in a hot air oven at 105°C for 3 minutes. Purple coloured amino acid spots are seen on the chromatogram. Identify the amino acids in the mixture with the help of spots produced by known amino acids.

The R_f values of the amino acids can be calculated as:

R_f = Distance travelled by particular amino acid/Distance covered by the solvent from the point of origin

Reagents:

a. N-butanol (Chromatographic grade).
b. Glacial acetic acid.
c. Solvent mixture of butanol, acetic acid, and distilled water in the proportion of (12 : 3 : 5)
d. Ninhydrin : 0.2 per cent solution of ninhydrin in acetone. It is prepared just before use.

B. Thin Layer Chromatography (TLC):

Introduction:

In recent years, thin layer chromatography has been developed. This technique consists of a thin layer of absorbents (silica gel, alumina. Kiselguhr or cellulose) on a glass plate or plastic sheet. Since absorbent is used, it is also termed as absorption chromatography.
This technique also provides superior results than that of paper chromatography. The spots are more compact with better resolution and the run is comparatively of shorter duration. Therefore, quicker run is possible.

Preparation of Plates:

Chromatographic plates (20 x 20 cm) of 200 μ thickness are prepared by using a suspension of 30 grams of silica gel G in 63 ml of 0.1 M Na_2CO_3 solution by shaking vigorously for 90 seconds. Only these plates are used which appear to be uniform in both transmitted and reflected light. These plates are activated at 110°C for 30 minutes immediately prior to use.

Table 46.1 : Electrophoretic pattern changes in some conditions. '—' mark indicate no change

Disease	Albumin	α_1-globulin	α_2-globulin	β-globulin	γ-globulin
Normal (gm/100ml)	3.5 to 5.5	0.1 to 0.4	0.4 to 0.8	0.5 to 1.0	0.7 to 1.5
Infancy	—	high	high	—	—
Pregnancy	low	high	high	high	—
Acute infection	—	high	high	—	—
Chronic infection	low	high	high	—	high
Cirrhosis	low	high	—	—	high
Nephrotic syndrome	low	high	very high	—	low
Myelomatosis	—	—	—	—	high
Hypogammaglobinemia	—	—	—	—	very low

Separate band occassionally seen in α and β position

Procedure:

Samples (5-100 (JL) are applied as a spot of less than 5 mm diameter on the lower right corner of the plates under a stream of warm air. Plates are first developed in a standard Brinkmann developing chamber previously saturated with the vapour of the solvent mixture with chloroform: methanol: acetic acid: water (250 : 74 : 19 : 3 ; V/V). When the solvent front migrates about 15 cm, plates are dried in air for 15 minutes and develop in the second dimension (90° rotation clockwise) with chloroform: methanol: 7 M. ammonium hydroxide (230 : 90 : 15 ; V/V). The solvent front is again allowed to move about 15 cm.

Developed plates are then dried in air for 5 minutes and exposed to iodine vapour in a sealed chamber for 30 to 60 seconds. The pale yellow areas are quickly outlined using a dental probe and the plates are exposed to air until the iodine has evaporated from the spots. When a permanent record of developed plates is desired, plates are sprayed lightly with 10 N. H_2SO_4 and then heated at 110°G for 15 minutes. The silica gel in each spot is scraped with the aid of a sharp edged polyethylene blade on paper. The weighing papers are then transferred to a 12 ml conical centrifuge tube and eluted by different solvents for estimation by photoelectric colorimeter.

Discussion:

The constituents of the mixture of amino acids, and the constituents of neutral lipids and phospholipids are separated and estimated in a short time.

C. Gel Chromatography:

(i) This type of Chromatography uses a porous gel. The dry gel particles are first allowed to take up the chosen solvent. This is accompanied by swelling; the liquid taken up constitutes the stationary phase.

(ii) These swollen particles are when made into a column with the same solvent, the spaces between them are filled by solvents. The mobile phase is the "void volume" (Vo); the gel particles are sponge like and the channels within them are of similar diameter (pore size) for a particular grade of gel.

(iii) Molecules can enter the stationary phase only within the gel if their diameter is less than this pore-size. Thus small molecules have the whole fluid volume (bed volume), but molecules above this limit are confined to the void volume and are rapidly washed through the column as mobile phase.

(a) This type of separation is usually performed using the pressure of a small head of the liquid phase to force this through the column. By per-fusing the mobile phase at much higher pressures the separation can be carried out quickly with high resolution.

(b) Dense column packing gives high resolution and high pressure metering systems maintain extremely reproducible solvent flows.

(c) This equipment consists of a column, a solvent delivery system, a sample injector and a detection. The column is of stainless steel of 10-50 cm long and 2-5 mm internal diameter. The solvent delivery is effected by solvent pumps which generate pressure at several thousand Psi and direct a pulse-free delivery of the solvent.

Constant pressure and constant volume pumps are used for the purpose. The sample injection is done with a syringe. The detectors may be ultraviolet with fixed wavelength Hg lamp 254 nm, refractive index detectors of deflection type, fluorescence detectors.

(d) This method is highly applicable for the separation of carbohydrates, proteins, peptides, amino acids, vitamins, steroids, neuropeptides, hormones and drugs. Separations are achieved very quickly. Hence this method is indispensable in laboratories where advance research and sophisticated analyses are performed.

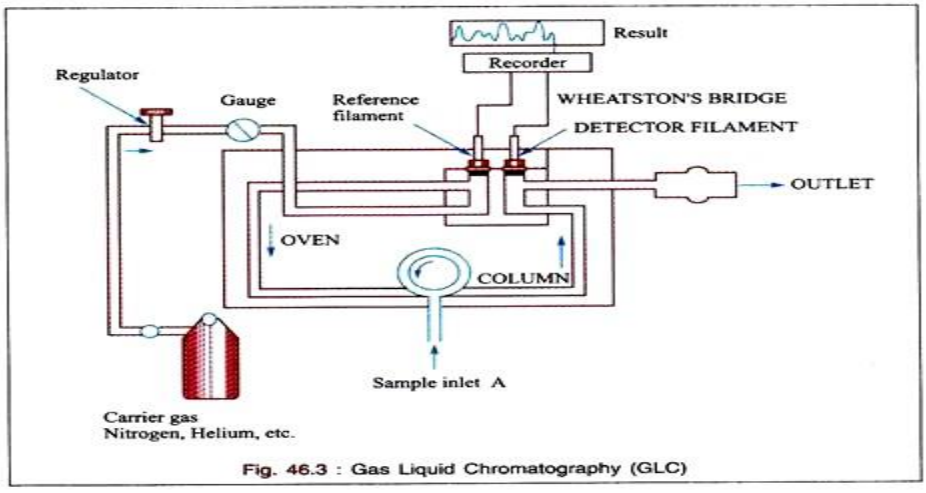


Fig. 46.3 : Gas Liquid Chromatography (GLC)

D. Gas Liquid Chromatography (GLC):

(i) The substances to be separated are carried as vapours in an inert gas like nitrogen, argon or helium over liquids when there is partitioning of the substances between the gas and the liquid. The liquids used are silicone, oils, lubricating greases, etc. held in inert solids like di-aomaceous earth or ground firebrick.

(ii) Glass or metal tube 1-2 metres long and of 0.2 – 2 cm diameter can be used.

(iii) If fatty acids are to be separated, they are first converted to methyl esters which are easily evaporated. The vapours are swept constantly by nitrogen through the tube containing the liquid phase at a temperature of 170-225°C, so that the vapours of the esters may remain as vapours. Separation takes places owing to the partitioning of the esters between the gas and the liquid. The ingredients can be identified by physical, chemical means or flame ionisation.

E. Ion-Exchange Chromatography:

(i) Ion-exchange resins are nothing but cross-linked polymers. The polymers must have negligible solubility but be porous enough for the ions to diffuse freely through it.

(ii) Ion-exchange resins are of cation and anion exchangers. Strong cation-exchange resins contain sulphuric acid groups ($-SO_3^-$), weak ones carboxylic acid groups ($-COO^-$), whereas strong anion-exchange resins have -N (R_1 R_2 R_3) and weak ones — N(R_1 R_2).

(iii) The most important resins are polystyrene resins formed by condensation of styrene (vinyl benzene) and divinyl benzene. Acidic or basic groups are introduced before or after polymerizing.

(iv) Strong alkaline cellulose treated with chloroacetic acid introduces the carboxy methyl group to give the weak cation-exchange resin carboxymethyl-cellulose (CM-cellulose) while condensation with 2-chlorotriethylamine gives the weak anion-exchange diethylaminoethyl-cellulose (DEAE-cellulose). Cellulose ion-ex-change materials are specially suitable for protein separations.

(v) The resins can be looked on as insoluble acids or bases which form insoluble salts shown below.

$$H^+ - Resin^- + Na^+ \rightarrow Na^+ - Resin^- + H^+ \text{ (Cation-exchanger)}$$

$$esin^+ - OH^- + CI^- \rightarrow Resin^+ - CI^- + OH^- \text{ (Anion-exchanger)}$$

The more strongly acidic the ion-change resin, the greater is the ionisation of the acidic group and the lower the pH at which it will exchange.

(vi) Ion-exchange resins have been widely used for the separation of amino acids and peptides. These have also been used to separate organic weak cations or anions from inorganic salts – ion-exchange resulting.
(vii) Mixed beds of anion and cation exchange resins have the property of replacing cations and anions of any salt in water by equivalent amounts of H^+ and OH^- respectively. This process is used in the preparation of "deionised" water in the laboratory. Non-ionic contaminants are not removed.

F. Gel Filtration Chromatography:

Gel Filtration Chromatography or Size exclusion chromatography (SEC) is a chromatographic method in which particles are separated based on their size, or in more technical terms, their hydrodynamic volume. It is usually applied to large molecules or macromolecular complexes such as proteins and industrial polymers. When an aqueous solution is used to transport the sample through the column, the technique is known as gel filtration chromatography.
The name gel permeation chromatography is used when an organic solvent is used as a mobile phase. The main application of gel filtration chromatography is the fractionation of proteins and other water-soluble polymers, while gel permeation chromatography is used to analyse the molecular weight distribution of organic-soluble polymers. Either technique should not be confused with gel electrophoresis, where an electric field is used to "pull" or "push" molecules through the gel depending on their electrical charges.
SEC is a widely used technique for the purification and analysis of synthetic and biological polymers, such as proteins, polysaccharides and nucleic acids. Biologists and biochemists typically use a gel medium—usually polyacrylamide, dextran or agarose—and filter under low pressure. Polymer chemists typically use either a silica or cross-linked polystyrene medium under a higher pressure. These media are known as the stationary phase.
The advantage of this method is that the various solutions can be applied without interfering with the filtration process, while preserving the biological activity of the particles to be separated. The technique is generally combined with others that further separate molecules by other characteristics, such as acidity, basicity, charge, and affinity for certain compounds.
The underlying principle of SEC is that particles of different sizes will elute (filter) through a stationary phase at different rates. This result in the separation of a solution of particles based on size, provided that all the particles are loaded simultaneously or near simultaneously, particles of the same size should elute together.

This is usually achieved with an apparatus called a column, which consists of a hollow tube tightly packed with extremely small porous polymer beads designed to have pores of different sizes. These pores may be depressions on the surface or channels through the bead. As the solution travels down the column some particles enter into the pores. Larger particles cannot enter into as many pores. The larger the particles, the less overall volume to traverse over the length of the column, and the faster the elution.

The filtered solution that is collected at the end is known as the eluent. The void volume consists of any particles too large to enter the medium, and the solvent volume is known as the column volume. In real life situations, particles in solution do not have a constant, fixed size, resulting in the probability that a particle which would otherwise be hampered by a pore may pass right by it. Also, the stationary phase particles are not ideally defined; both particles and pores may vary in size.

Elution curves, therefore, resemble Gaussian distributions. The stationary phase may also interact in undesirable ways with a particle and influence retention times, though great care is taken by column manufacturers to use stationary phases which are inert and minimize this issue.

Like other forms of chromatography, increasing the column length will tighten the resolution, and increasing the column diameter increases the capacity of the column. Proper column packing is important to maximize resolution: an over packed column can collapse the pores in the beads, resulting in a loss of resolution. An under packed column can reduce the relative surface area of the stationary phase accessible to smaller species, resulting in those species spending less time trapped in pores.

In simple manual columns the eluent is collected in constant volumes, known as fractions. The more similar the particles are in size, the more likely they will be in the same fraction and not detected separately. More advanced columns overcome this problem by constantly monitoring the eluent. The collected fractions are often examined by spectroscopic techniques to determine the concentration of the particles eluted. Three common spectroscopy detection techniques are refractive index (RI), evaporative light scattering (ELS), and ultraviolet (UV). When eluting spectroscopically similar species (such as during biological purification) other techniques may be necessary to identify the contents of each fraction.

The elution volume decreases roughly linearly with the logarithm of the molecular hydrodynamic volume (often assumed to be proportional to molecular weight). Columns are often calibrated using 4-5 standard samples (e.g., folded proteins of known molecular weight) to determine the void volume and the slope of the logarithmic dependence. This calibration may need to be repeated under different solution conditions.

Applications:

1. Proteomics:

SEC is generally considered a low resolution chromatography as it does not discern similar species very well, and is, therefore, often reserved for the final "polishing" step of a purification. The technique can determine the quaternary structure of purified proteins which have slow exchange times, since it can be carried out under native solution conditions, preserving macromolecular interactions.

SEC can also assay protein tertiary structure as it measures the hydrodynamic volume (not molecular weight), allowing folded and unfolded versions of the same protein to be distinguished. For example, the apparent hydrodynamic radius of a typical protein domain might be 14a and 36A for the folded and unfolded forms respectively.

SEC allows the separation of these two forms as the folded form will elute much later due to its smaller size. Alternatively, folded and unfolded versions of the same metalloproteinase can be separated according to their different isoelectric points by using quantitative preparative native continuous polyacrylamide gel electrophoresis (QPNC-PAGE).

2. Polymer synthesis:

SEC can be used as a measure of both the size and the polydispersity of a synthesised polymer; that one is able to find distribution of sizes of polymer molecules. If standards of a known size are run previously, then a calibration curve can be created to determine the sizes of polymer molecules of interest. Alternatively, techniques such as light scattering and/or viscometry can be used online with SEC to yield absolute molecular weights that do not rely on calibration with standards of known molecular weight. Due to the difference in size of two polymers with identical molecular weights, the absolute determination methods are generally more desirable. A typical SEC system can quickly (in about half an hour) give polymer chemists information on the size and polydispersity of the sample.

G. High Performance Liquid Chromatography:

High-performance liquid chromatography (HPLC) is a form of column chromatography used frequently in biochemistry and analytical chemistry. It is also sometimes referred to as high-pressure liquid chromatography. HPLC is used to separate components of a mixture by using a variety of chemical interactions between the substance being analysed (analyte) and the chromatography column.

In isocratic HPLC, the analyte is forced through a column of the stationary phase (usually a tube packed with small round particles with a certain surface chemistry) by pumping a liquid (mobile phase) at high pressure through the column. The sample to be analysed is

introduced in a small volume to the stream of mobile phase and is retarded by specific chemical or physical interactions with the stationary phase as it traverses the length of the column.

The amount of retardation depends on the nature of the analyte, stationary phase and mobile phase composition. The time at which a specific analyte elutes (comes out of the end of the column) is called the retention time and is considered a reasonably unique identifying characteristic of a given analyte. The use of pressure increases the linear velocity (speed) giving the components less time to diffuse within the column, leading to improved resolution in the resulting chromatogram. Common solvents used include any miscible combinations of water or various organic liquids (the most common are methanol and acetonitrile). Water may contain buffers or salts to assist in the separation of the analyte components, or compounds such as Trifluoroacetic acid which acts as an ion pairing agent. A further refinement to HPLC has been to vary the mobile phase composition during the analysis; this is known as gradient elution. A normal gradient for reverse phase chromatography might start at 5% methanol and progress linearly to 50% methanol over 25 minutes, depending on how hydrophobic the analyte is.

The gradient separates the analyte mixtures as a function of the affinity of the analyte for the current mobile phase composition relative to the stationary phase. This partitioning process is similar to that which occurs during a liquid-liquid extraction but is continuous, not step-wise. In this example, using a water/methanol gradient, the more hydrophobic components will elute (come off the column) under conditions of relatively high methanol; whereas the more hydrophilic compounds will elute under conditions of relatively low methanol. The choice of solvents, additives and gradient, depends on the nature of the stationary phase and the analyte. Often a series of tests are performed on the analyte and a number of generic runs may be processed in order to find the optimum HPLC method for the analyte — the method which gives the best separation of peaks.

Types of HPLC:

1. Normal phase chromatography:

Normal phase HPLC (NP-HPLC) was the first kind of HPLC chemistry used, and separates analytes based on polarity. This method uses a polar stationary phase and a non-polar mobile phase, and is used when the analyte of interest is fairly polar in nature. The polar analyte associates with and is retained by the polar stationary phase.

Adsorption strengths increase with increase in analyte polarity, and the interaction between the polar analyte and the polar stationary phase (relative to the mobile phase) increases the elution time. The interaction strength not only depends on the functional groups in the analyte molecule, but also on steric factors, and structural isomers are often

resolved from one another. Use of more polar solvents in the mobile phase will decrease the retention time of the analytes while more hydrophobic solvents tend to increase retention times. Particularly polar solvents in a mixture tend to deactivate the column by occupying the stationary phase surface. This is somewhat particular to normal phase because it is most purely an adsorptive mechanism (the interactions are with a hard surface rather than a soft layer on a surface). NP-HPLC had fallen out of favour in the 1970s with the development of reversed-phase HPLC because of a lack of reproducibility of retention times as water or protic organic solvents changed the hydration state of the silica or alumina chromatographic media. Recently it has become useful again with the development of HILIC bonded phases which utilize a partition mechanism which provides reproducibility.

2. Reverse phase chromatography:

Reversed phase HPLC (RP-HPLC) consists of a non-polar stationary phase and a moderately polar mobile phase. One common stationary phase is a silica which has been treated with RMe_2SiCl, where R is a straight chain alkyl group such as $C_{18}H_{37}$ or C_8H_{17}. The retention time is, therefore, longer for molecules which are more non-polar in nature, allowing polar molecules to elute more readily.

Retention time is increased by the addition of polar solvent to the mobile phase and decreased by the addition of more hydrophobic solvent. Reversed phase chromatography is so commonly used that it is not uncommon for it to be incorrectly referred to as "HPLC" without further specification.

RP-HPLC operates on the principle of hydrophobic interactions which result from repulsive forces between a relatively polar solvent, the relatively non-polar analyte, and the non-polar stationary phase. The driving force in the binding of the analyte to the stationary phase is the decrease in the area of the non-polar segment of the analyte molecule exposed to the solvent. This hydrophobic effect is dominated by the decrease in free energy from entropy associated with the minimization of the ordered molecule-polar solvent interface. The hydrophobic effect is decreased by adding more non-polar solvent into the mobile phase. This shifts the partition coefficient such that the analyte spends some portion of time moving down the column in the mobile phase, eventually eluting from the column.

The characteristics of the analyte molecule play an important role in its retention characteristics. In general, an analyte with a longer alkyl chain length results in a longer retention time because it increases the molecule's hydrophobicity.

Very large molecules, however, can result in incomplete interaction between the large analyte surface and the alkyl chain. Retention time increases with hydrophobic surface area which is roughly inversely proportional to solute size. Branched chain compounds elute more rapidly than their corresponding isomers because the overall surface area is decreased.

Apart from mobile phase hydrophobicity, other mobile phase modifiers can affect analyte retention. For example, the addition of inorganic salts causes a linear increase in the surface tension of aqueous solutions, and because the entropy of the analyte-solvent interface is controlled by surface tension, the addition of salts tends to increase the retention time. Another important component is pH since this can change the hydrophobicity of the analyte. For this reason most methods use a buffering agent, such as sodium phosphate, to control the pH. An organic acid such as formic acid or most commonly trifluoro-acetic acid is often added to the mobile phase. These serve multiple purposes: they control pH, neutralize the charge on any residual exposed silica on the stationary phase and act as ion pairing agents to neutralize charge on the analyte. The effect varies depending on use but generally improves the chromatography.

Reversed phase columns are quite difficult to damage compared with normal silica columns; however, many reverse phase columns consist of alkyl derivatized silica particles and should never be used with aqueous bases as these will destroy the underlying silica backbone. They can be used with aqueous acid but the column should not be exposed to the acid for too long, as it can corrode the metal parts of the HPLC equipment.

The metal content of HPLC columns must be kept low if the best possible ability to separate substances is to be retained. A good test for the metal content of a column is to inject a sample which is a mixture of 2, 2′- and 4, 4′- bipyridine. Because the 2,2′- bipyridine can chelate the metal it is normal that when a metal ion is present on the surface of the silica the shape of the peak for the 2,2′-bipyridine will be distorted, tailing will be seen on this distorted peak.

Parameters:

There are different parameters upon which the separation by HPLC relies. **Some of the important parameters are discussed below:**

1. Internal diameter:

The internal diameter (ID) of an HPLC column is a critical aspect that determines quantity of analyte that can be loaded onto the column and also influences sensitivity. Larger columns are usually seen in industrial applications such as the purification of a drug product for later use. Low ID columns have improved sensitivity and lower solvent consumption at the expense of loading capacity.

i. Larger ID columns (over 10 mm) are used to purify usable amounts of material because of their large loading capacity.

ii. Analytical scale columns (4.6 mm) have been the most common type of columns, though smaller columns are rapidly gaining popularity. They are used in traditional quantitative analysis of samples and often use a UV-Vis absorbance detector.

iii. Narrow-bore columns (1-2 mm) are used for applications when more sensitivity is desired either with special UV-Vis detectors, fluorescence detection or with other detection methods like liquid chromatography-mass spectrometry
iv. Capillary columns (under 0.3 mm) which are used almost exclusively with alternative detection means such as mass spectrometry. They are usually made from fused silica capillaries, rather than the stainless steel tubing that larger columns employ.

2. Particle size:

Most traditional HPLC is performed with the stationary phase attached to the outside of small spherical silica particles (very small beads). These particles come in a variety of sizes with 5 μm beads being the most common. Smaller particles generally provide more surface area and better separations, but the pressure required for optimum linear velocity increases by the inverse of the particle diameter squared.
This means that changing to particles that are half as big in the same size of column will double the performance, but increase the required pressure by a factor of four. Larger particles are more often used in non-HPLC applications such as solid-phase extraction.

3. Pore size:

Many stationary phases are porous to provide greater surface area. Small pores provide greater surface area while larger pore size has better kinetics especially for larger analytes. For example, a protein which is only slightly smaller than a pore might enter the pore but not easily leave once inside.

4. Pump pressure:

Pumps vary in pressure capacity, but their performance is measured on their ability to yield a consistent and reproducible flow rate. Pressure may reach as high as 6000 lbf/in^2 (~40 MPa, or about 400 atmospheres). Modern HPLC systems have been improved to work at much higher pressures, and therefore, be able to use much smaller particle sizes in the columns (< 2 micrometres).
These "Ultra High Performance Liquid Chromatography" systems or UHPLCs can work at up to 15,000 lbf/in^2 (-100 MPa or about 1000 atmospheres). Note that the term "UPLC", sometimes found instead is a trademark of Waters Corporation and not the name for the technique in general.

Probable Questions:

1. Write basic principle of chromatography.

2. Define stationary phase and mobile phase.

3. Define absorption and partition chromatography.

4. How amino acids can be separated by paper chromatography ?

5. Write basic principle of gel filtration chromatography.

6. What is TLC? How this method can differentiate components of a mixture?

7. Write basic principle of ion exchange chromatography.

8. What are the applications of gel filtration chromatography.

9. How HPLC can resolve molecules?

10. What is forward phase and reverse phase HPLC?

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

UNIT-XIV

Separation of Cell Constituents by Electrophoresis: PAGE, SDS-PAGE (One and Two dimentional)

Electrophoresis: The term electrophoresis describes the migration of a charged particle under the influence of an electric field. Many important biological molecules, such as amino acids, peptides, proteins, nucleotides and nucleic acids, possess ionisable groups and, therefore, at any pH, exist in solution as electrically charged species either as cations (+) or anions (-). Under the influence of an electric field charged particles will migrate either to the cathode or to the anode, depending on the nature of their net charge.

In order to understand fully how charged species separate, it is necessary to look at some simple equations related to electrophoresis. When a potential difference (voltage) is applied across the electrodes, it generates a potential gradient, E, which is applied voltage, V, divided by the distance, d, between the electrodes. When this potential gradient E is applied, the force on a molecule bearing a charge of q coulombs is Eq newton's. It is this force that derives a charged molecule towards an electrode. However, there is also a factional resistance that retards the movement of this charged molecule. This frictional force is a measure of the hydrodynamic size of the molecule, the shape of the molecule, the pore size of the medium in which electrophoresis is taking place and the viscosity of the buffer.

The velocity, v, of a charged molecule in electrical field is, therefore, given by the equation: $v = Eq/f$, where f is the frictional coefficient.

More commonly the term electrophoretic mobility (μ) of an ion is used, which is the ratio of the velocity of the ion to field strength (v/E). When a potential difference is applied, therefore, molecules with different overall charges will begin to separate owing to their different electrophoretic nobilities.

Even molecules of similar charges will begin to separate if they have different molecular sizes, since they will experience different frictional forces. As will be seen below, some forms of electrophoresis rely almost totally on the different charges on the molecules to effect separation, whilst other methods exploit differences in molecular size and, therefore, encourage frictional effects to bring about separation.

In short, electrophoretic mobility can be given as:
Mobility = (voltage) (charge)/ (frictional coefficient)

Provided the electric field is removed before the molecules in the sample reach the electrodes, the components will have been separated according to their electrophoretic mobility. Electrophoresis is thus an incomplete form of electrolysis. The separated samples are then located by staining with an appropriate dye or by autoradiography if the sample is radiolabelled.
The current in the solution between the electrodes is conducted mainly by the buffer ions, a small proportion being conducted by the sample ions. Ohm's law expresses the relationship between current (I), voltage (V) and resistance (R):
$V/I = R$.

It, therefore, appears that it is possible to accelerate an electrophoretic separation by voltage, which would result in a corresponding increase in the current flowing. The distance migrated by the ions will be proportional to both current and time. However, this would ignore one of the major problems for most common forms of electrophoresis, namely the generation of the heat.

During electrophoresis, the power (W, watts) generated in the supporting medium is given by: $W = I^2R$.

Most of this power generated is dissipated as heat. Heating of the electrophoretic medium has the following effects:
i. An increased rate of diffusion of sample and buffer ions leading to broadening of the separated samples.
ii. The formation of convection currents, which leads to mixing of the separated samples.
iii. Thermal instability of the samples that are rather sensitive to heat. This may include denaturation of the proteins (e.g., the loss of enzyme activity).
iv. A decrease of buffer viscosity, and hence a reduction in the resistance of the medium.

If a constant voltage is applied, the current increases during electrophoresis owing to the decrease in the resistance and the rise in current increases the heat output still further. For this reason, workers often use a stabilized power supply, which provides constant power and thus eliminates fluctuations in heating.
Constant heat generation is, however, a problem. One may think that electrophoresis may be run at a very low power (low current) to overcome any heating problem, but this can lead to poor separation as a result of the increased amount of the diffusion resulting from a long separation times. Compromise conditions, therefore, have to be found with reasonable

power settings, to give acceptable separation times, and an appropriate cooling system, to remove liberated heat. While such systems work fairly well, the effects of heating are not always totally eliminated. For example, for electrophoresis carried out in cylindrical tubes or in slab gels, although heat is generated uniformly through the medium, heat is removed only from the edges, resulting in a temperature gradient within the gel, the temperature at the centre of the gel being higher than that at the edges.

Since the warmer fluid at the centre is less viscous, electrophoretic mobility are, therefore, greater in the central region (electrophoretic mobilities increase by about 2% for each 1°C rise in the temperature), and electrophoretic zones develop a bowled shape, with the zone centre migrating faster than the edges. A final factor that can effect electrophoresis separation is the phenomenon of electroendoosmosis (also known as electro-osmotic flow), which is due to the presence of charged groups on the surface of the support medium. For example, paper has some carboxyl groups present, agarose (depending on the purity grade) contains sulphate group and the surface of the glass walls used in capillary electrophoresis contains silanol (Si-OH) groups.

Mechanism of electroendoosmosis has been explained in details in the section covering capillary electrophoresis, although the principle is the same for any support medium that has charged groups on it. In short in a fused silica capillary tube, above a pH value of about 3, silanol groups on the silica capillary walls will ionise, generating negatively charged sites. It is these charges that generate electroendoosmosis.

The ionized silanol groups create an electrical double layer, or region of charge separation, at the capillary wall/electrolyte interface. When a voltage is applied, cations in the electrolyte near the capillary wall migrate towards the cathode, pulling electrolyte solution with them. This creates a net electrosmotic flow towards the cathode.

However, the introduction of the use of gels as a support medium led to a rapid improvement in methods for analyzing macromolecules. The earliest gel system to be used was the starch gel and, although this has some uses, the vast majority of electrophoretic techniques used nowadays involve either agarose gel or polyacrylamide gel.

Agarose Gel Electrophoresis:

Agarose is a linear polysaccharide (average relative molecular mass about 12000) made up of the basic repeat unit agarobiose, which comprises alternating units of the galactose and 3, 6- anhydrogalactose (Fig. 8.4).

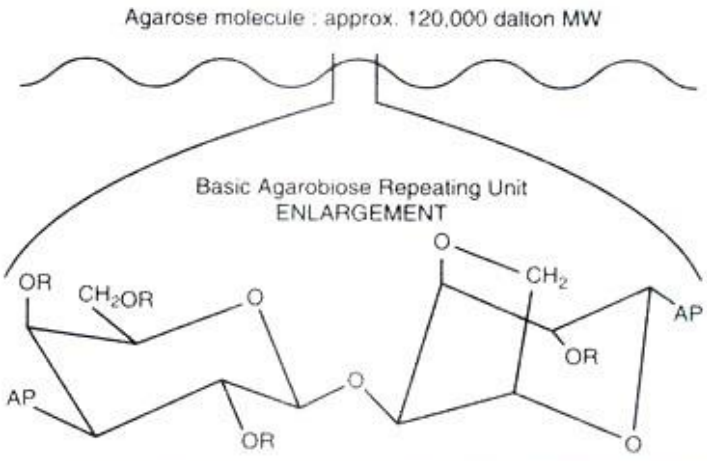


Fig. 8.4: Structure of agarose

Agarose is one of the components of agar that is mixture of polysaccharides isolated from certain seaweeds. Agarose is usually used at a concentration between 1% and 3%. Agarose gels are formed by suspending dry agarose in aqueous buffer, then boiling the mixture until a clear solution is formed. This is poured and allowed to cool to room temperature to form a rigid gel.

The gelling properties are attributed to both inter- and intermolecular hydrogen bonding within and between the long agarose chains. This cross-linked structure gives the gel good anti-conventional properties. The pore size in the gel is controlled by the initial concentration of agarose; large pore sizes are formed from low concentration and smaller pore sizes are formed from higher concentrations.

Although essentially free from charge, substitution of the alternating sugar residues with carboxyl, methoxyl, pyruvate and specially sulphate group occur to varying degrees. This substitution can result in electroendoosmosis during electrophoresis and ionic interactions between the gel and sample in all uses, both unwanted effects. Agarose is, therefore, sold in very different purity grades, based on sulphate concentration — the lower the sulphate content the higher the purity.

Agarose gels are used for the electrophoresis of both proteins and nucleic acids. For proteins, the pore sizes of a 1% agarose gel are large relative to the sizes of proteins. Agarose gels are therefore used in techniques such as Immunoelectrophoresis or flat bed

isoelectric focusing, where the proteins are required to move unhindered in the gel matrix according to their native charge.

Such large pore gels are also used to separate much larger molecules such as DNA or RNA, because the pore sizes in the gel are still large enough for DNA or RNA molecule to pass through the gel. Now, however, the pore size and molecule size are more comparable and frictional effects begin to play a role in the separation of these molecules.

A further advantage of using agarose is the availability of low melting temperature agarose (62-65°C). As the name suggests, these gels can be re-liquefied by heating to 65°C and thus, for example, DNA samples separated in a gel can be cut out of the gel, returned to solution and recovered.

Table 8.1: Range for fractionation of different compounds in different agarose percentages

Agarose Concentration	Protein Fractionation Range* (kDa)	Polysaccharide Fractionation* (kDa)	Nucleic Acid Exclusion Limit (bp)
1.0%	1,000 to 150,000	1,000 to 150,000	(> 3,000)
2.0%	80 to 40,000	90 to 20,000	1,340
4.0%	50 to 15,000	40 to 5,000	860
6.0%	10 to 5,000	10 to 1,000	180

Owing to the poor elasticity of agarose gels and the consequent problems of removing them from small tubes, the gel rod system is sometimes used, since acrylamide gel is not used. Horizontal slab gels are invariably used for isoelectric focusing or immunoelectrphoresis in agarose. Horizontal gels are also used routinely for DNA and RNA gels, although vertical systems have been used by some workers.

Table 8.2: Approximate gel concentrations for separation of DNA linear fragments of various sizes

Gel Concentration	Separation Range (bp)
agarose - 0.3%	60,000 - 5,000
agarose - 0.7%	20,000 - 800
agarose - 0.9%	7,000 - 500
agarose - 1.2%	6,000 - 400
agarose - 1.5%	4,000 - 200
agarose - 2.0%	3,000 - 100
agarose - 4.0%	500 - 10
acrylamide - 4%	1,000 - 800
acrylamide - 10%	500 - 25
acrylamide - 20%	50 - 1

Agarose Gel Electrophoresis of Nucleic Acids:

Nucleic acids are polymers composed of individual nucleotide units. The units are connected via phosphate diester linkages of the backbone sugars. The net effect of these linkages is to give the polymers a net negative charge. From the earliest days of electrophoresis it has been axiomatic that molecules carrying an electrical charge will migrate in an electrical field in a predictable manner.

When subjected to an electrical field, a molecule carrying a net negative charge will migrate toward the positive pole and a molecule with a net positive charge will migrate toward the negative pole. In a semi-solid matrix like agarose, the equation describing mobility can be re-interpreted, at least heuristically, by defining gas gel density or concentration and r as the length of the molecule.

Thus, when they are placed in the semi-solid matrix of a gel, nucleic acids will migrate toward the positive pole in a predictable and reproducible manner that can be described as a negative exponential function of length. That is to say, shorter molecules will migrate faster and longer molecules will migrate slower. Indeed, in the case of a nucleic acid in a gel in an electrical field, every other element of the migration expression is a constant and mobility is completely determined by molecular length.

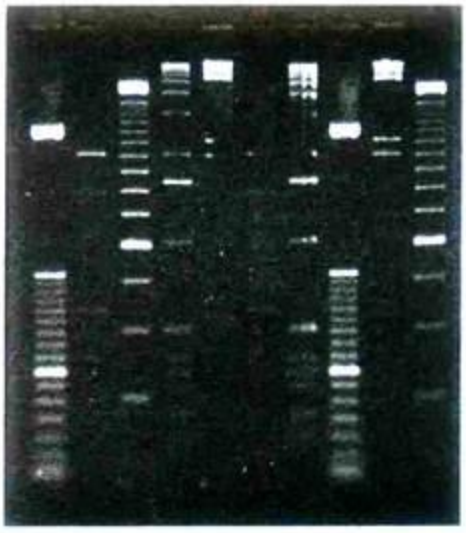


Fig. 8.6: DNA separated on 0.8% agarose gel (*Courtesy*: Dr D Dash, Deptt. of Biochemistry, IMS, BHU)

As a matter of practice, it is difficult to accurately resolve double-stranded nucleic acids smaller than about 100 bases in an agarose gel because the sieving properties of agarose are not fine enough. On the other end of the scale, molecules longer than about 25,000 bp but shorter than around 2,000,000 bp will all run at the same rate. This is called limiting mobility.

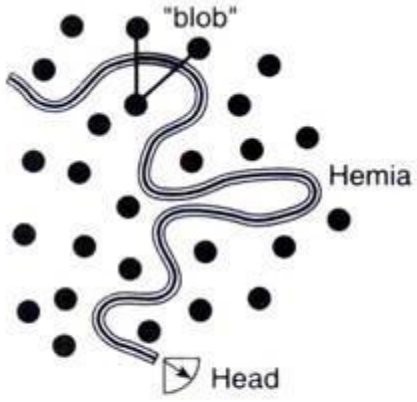


Fig. 8.7: Biased reptation

Nucleic acid molecules longer than 2,000,000 bp will not even enter an agarose gel. Thus, the effective size range for agarose gel electrophoresis of double stranded nucleic acids is

between 100 bp and 25,000 bp. In this range the behaviour of the molecule is precise and predictable. This behaviour is shown in Fig. 8.8. As can be seen there is minimal separation of the larger fragments but resolution improves as the fragments get smaller.

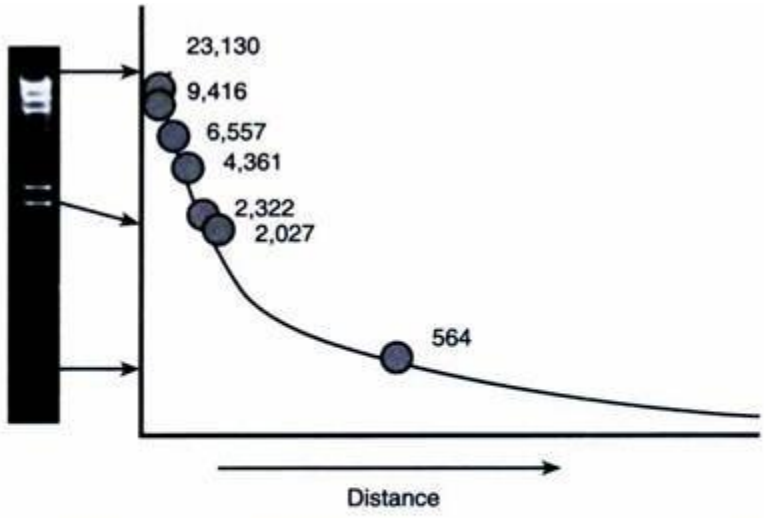


Fig. 8.8: Electrophoretic mobility of DNA restriction fragments in an agarose gel. The DNA is from bacteriophage lambda digested with the restriction enzyme *Hind III*. The graph shows migration distance in centimetres by size of the restriction fragment in base pairs.

While this phenomenon has been known for many years, what was not known was how the nucleic acid molecules actually moved in the gel matrix. In the 1980s a theory was put forward that nucleic acids migrated through the gel much the same way that a snake moves. That is, the leading edge moves forward and pulls the rest of the molecule with it.

In this model, as the molecule gets longer resistance being pulled along increases. This resistance is further increased by the interaction of the molecule with the gel matrix. The increase in resistance is non-linear. This model, called “biased reptation”, is sufficient to explain all of the behaviour of a nucleic acid in a semi-solid matrix (Lerman et al, 1982; Lumpkin et al., 1985).

In 1989 a group at the University of Washington put this theory to the test. They filmed DNA molecules moving through an agarose gel. Their films showed both reptation and the nucleic acid/ gel matrix interactions (Smith et al., 1989).

In the mid-1980s a number of methods were developed to electrophoretically analyse nucleic acid molecules in the limiting mobility size range. The solution involved artificially introducing a size dependent mobility on nucleic acid molecules by altering the electrophoretic field. The first such alteration involved simply switching the polarity of the field in a regular pattern. Carle et al. (1986) showed that periodic reversals of polarity would induce the molecules to make U-turns in the gel.

Even at very large sizes, this turning would permit separation of molecules. In the length of time the molecules were reversed was about one-third the time they were oriented forward, for example, three seconds forward and one second back, molecules as large as 2,000,000 bp could be resolved in a standard agarose gel in a few hours. The first practical demonstration of this method, called Field Inversion Gel Electrophoresis (FIGE), was to completely resolve intact yeast chromosomes. Since then, a variety of methods, collectively termed pulsed-field gel electrophoresis, have been developed.

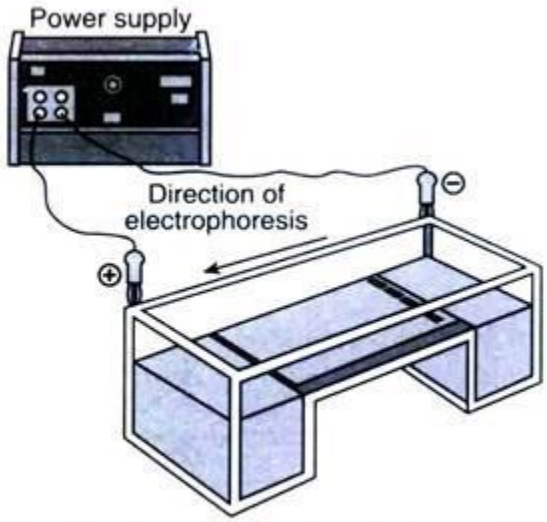


Fig. 8.9: A typical horizontal gel electrophoresis system

Many laboratories routinely use 0.8% gels, which are suitable for separating DNA molecules in the range 0.5- 10 kb. Since agarose gels separate DNA according to size, the M_r of a DNA fragment may be determined from its electrophoretic mobility by running a number of standard DNA markers of known M_r on the same gel. This is most conveniently achieved by running a sample of bacteriophage λ DNA (49 kb) that has been cleaved with a restriction enzyme such as EcoRI. Since the base sequence of λ DNA is known, and the cleavage sites for EcoRl are known, this generates fragments of accurately known size.

DNA gels are invariably run as horizontal, submarine or submerged gels; so named because such a gel is totally immersed in buffer. Agarose, dissolved in gel buffer by boiling, is poured onto a glass or plastic plate, surrounded by a wall of adhesive tape or a plastic frame to provide a gel about 3 mm in depth. Loading wells are formed by placing a plastic well-forming template or comb in the poured gel solution, and removing this comb once the gel is set.

The gel is placed in electrophoresis tank, covered with buffer, and samples loaded by directly injecting the sample into the wells. Samples are prepared by dissolving them in a buffer solution that contains sucrose, glycerol or Ficoll, which makes the solution dense and allows it to sink to the bottom of the well. A dye such as bromophenol blue is also

included in the sample solvent; it makes it easier to see the sample that is being loaded and also acts as a marker of the electrophoresis front. No stacking gel is needed for the electrophoresis of DNA because the mobilities of DNA molecules are much greater in the well than in the gel, and therefore, all the molecules in the well pile up against the gel within a few minutes of the current being turned on, forming a tight band at the start of the run.

General purpose gels are approximately 25 cm long and 12 cm wide, and are run at a voltage gradient of about 1.5 V cm^{-1} overnight. A higher voltage would cause excessive heating. For rapid analyses that do not need extensive separation of DNA molecules, it is common to use mini-gels that are less than 10 cm long. In this way information can be obtained in 2-3 h.

Fig. 8.10: Horizontal gel electrophoresis system (*Courtesy:* Dr D Dash, Deptt. of Biochemistry, IMS, BHU)

Once the system has been run, the DNA in the gel needs to be stained and visualized. The reagent most widely used is the fluorescent dye ethidium bromide. The gel is rinsed gently in a solution of ethidium bromide (0.5 μg cm^{-1}) and then viewed under ultraviolet light (300 nm wavelength). Ethidium bromide is a cyclic planar molecule that binds between the stacked base-pairs of DNA (i.e., it intercalates).

The ethidium bromide concentration, therefore, builds up at the site of DNA bands and under ultraviolet light the DNA bands fluoresce orange-red. As little as 10 ng of DNA can be visualized as a 1 cm wide band. It should be noted that extensive viewing of DNA with ultraviolet light can result in damage of the DNA by nicking and base-pair dimerization.

This is of no consequence if the gel is only to be viewed, but obviously viewing of the gel should be kept to a minimum if the DNA is to be recovered. It is essential to protect one's eye by wearing goggles when ultraviolet light is used. If viewing of gel under ultraviolet is

carried out for long periods, a plastic mask that covers the whole face should be used to avoid 'sunburn'.

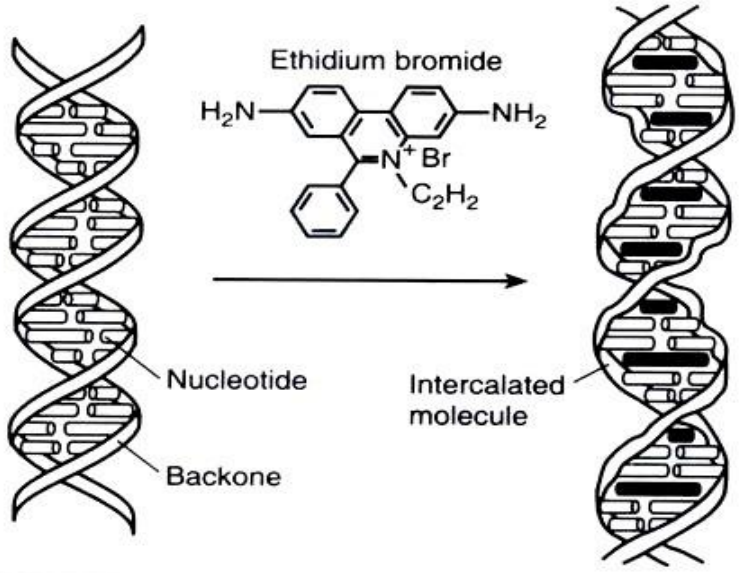


Fig. 8.11: Interaction of ethidium bromide with DNA molecule

DNA Sequencing Gels:

Although agarose gel electrophoresis of DNA is a 'workhorse' technique for the molecular biologists, a different form of electrophoresis has to be used when DNA sequences are to be determined. Whichever DNA sequencing method is used, the final analysis usually involves separating single-stranded DNA molecules shorter than about 1000 nt and differing in size by only 1 nt.

To achieve this it is necessary to have a small-pored gel and so acrylamide gels are used instead of agarose. For example, 3.5% polyacrylamide gels are used to separate DNA in the range 80-1000 nt and 12% gels to resolve fragments of between 20-100 nt. If a wide range of sizes is being analysed it is often convenient to run a gradient gel, for example, from 3.5% to 7.5%. Sequencing gels are run the presence of denaturing agents, urea and form amide. Since it is necessary to separate DNA molecules that are very similar in size, DNA sequencing gels tend to be very long (100 cm) to maximize the separation achieved. A typical DNA sequencing gel is shown in Fig. 8.12.

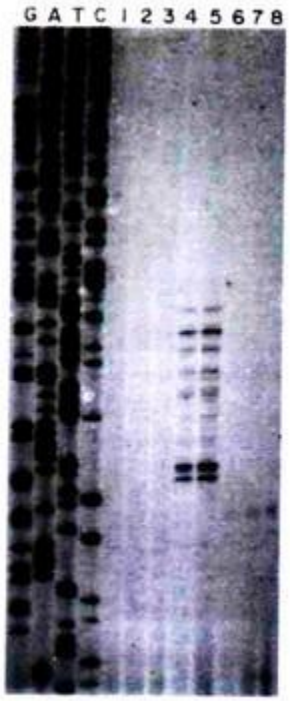


Fig. 8.12: DNA sequence gel

As mentioned above, electrophoresis in agarose can be used as a preparative method for DNA.

The DNA bands of interest can be cut out of the gel and the DNA recovered by:

(a) Electro elution,
(b) Macerating the gel piece in buffer, centrifuging and collecting the supernatant; or
(c) If low melting point agarose is used, melting the gel piece and diluting with buffer.
In each case, the DNA is finally recovered by precipitation of the supernatant with ethanol.

Polyacrylamide Gels:

Electrophoresis in acrylamide gels is frequently referred to as PAGE, being an abbreviation for Polyacrylamide gel electrophoresis. Cross-linked polyacrylamide gels are formed from the polymerization of acrylamide monomer in the presence of smaller amounts of N-N-methylenebisacrylamide (normally referred to as bisacrylamide) (Fig. 8.17). Note that bis-acrylamide is basically two units of acrylamide molecules linked by a methylene group, and is used as a cross-linking agent.

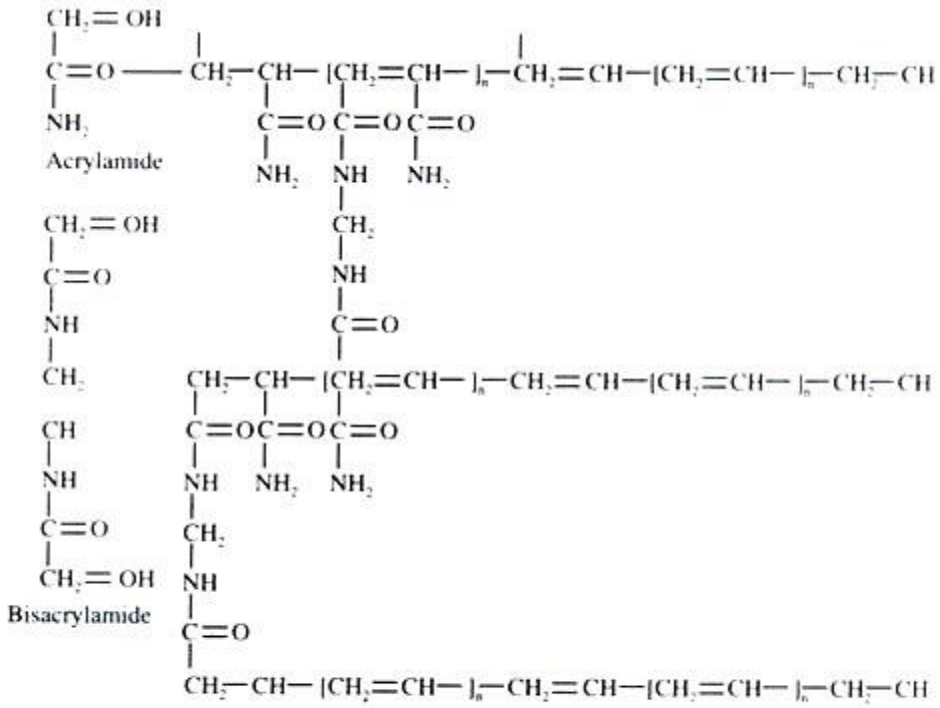


Fig. 8.17: Chemical structure of acrylamide, bisacrylamide, and polyacylamide gel

Acrylamide monomer is polymerized in a head-to-tail fashion into long chains and occasionally a bisacrylamide molecule is built into the growing chains, thus introducing a second site for chain extension. Proceeding in this way a cross-linked matrix of fairly well-defined structure is formed, the polymerization of acrylamide is an example of free radical catalysis, and is initiated by the addition of ammonium per-sulphate and the base N, N, N', N'-tetra-methylenediamine (TEMED). TEMED catalyses the decomposition of per-sulphate ion to give a free radical (i.e., a molecule with an unpaired electron):

$$S_2O_8^{2-} + e^- \rightarrow SO_4^{2-} + SO_4^{\bullet}$$

If this free radical is represented as R* (where the dot represents an unpaired electron) and M as an acrylamide monomer molecule, then the polymerization can be represented as follow:

$$R^{\bullet} + M \rightarrow RM^{\bullet}$$

$$RM^{\bullet} + M \rightarrow RMM^{\bullet}$$

$$RMM^{\bullet} + M \rightarrow RMMM^{\bullet}, \text{ etc.}$$

Free radicals are highly reactive because of the presence of an unpaired electron that needs to be paired with another electron to stabilize the molecule. R*, therefore, reacts with M, forming a single bond by sharing its unpaired electron with one from the outer shell of the monomer molecule.

In this way long chains of acrylamide are built up, being cross-linked by the introduction of the occasional bisacrylamide molecule into growing chain oxygen mops up free radicals and, therefore, all gel solutions are normally degassed (the solutions are briefly placed under vacuum to remove loosely dissolved air) prior to use.

The degassing of the gel solution also serves a second purpose. The polymerization of acrylamide is an exothermic reaction (i.e., heat is liberated) and the warming up of the gel as it sets can liberate air bubbles that become trapped in the polymerized gel. The degassing step prevents this possibility.

Photo-polymerisation is an alternative method that can be used to polymerize acrylamide gels. The ammonium per-sulphate and TEMED are replaced by riboflavin and when the gel is poured it is placed in front of a bright light for 2-3 h. Photodecomposition of riboflavin generates a free radical that initiates polymerization.

Acrylamide gels are defined in terms of the total percentage of acrylamide present, and the pore size in the gel can be varied by changing the concentrations of both the acrylamide and bisacrylamide. Acrylamide gels can be made with a content of between 3% and 30% acrylamide. Thus low percentage gels (e.g., 4%) have large pore sizes and are used, for example, in the electrophoresis of proteins, when free movement of the proteins by electrophoresis is required without any noticeable frictional effect and for another example, in flat-bed isoelectric focusing or the stacking gel system of an SDS-polyacrylamide gel.

Low percentage acrylamide gels are also used to separate DNA. Gels containing between 10% to 20% acrylamide are used in techniques such as SDS-gel electrophoresis, where the smaller pore size now introduces a sieving effect that contributes to the separation of proteins according to their size. Proteins were originally separated on polyacrylamide gels that were polymerized in glass tubes, approximately 7 mm in diameter and about 10 cm in length. The tubes were easy to load and run, with minimum apparatus requirements. However, only one sample could be run per tube and, because conditions of separation could vary from tube to tube, comparison between different samples was not always accurate.

The later introduction of vertical gel slabs allowed running of up to 20 samples under identical conditions in a single run. Vertical slabs are now used routinely for both analysis of proteins and for the separation of DNA fragments during DNA sequence analysis. Although some workers prepare their own acrylamide gels, others purchase commercially available ready-made gels for techniques such as SDS-PAGE, native gels and isoelectric focusing (IEF).

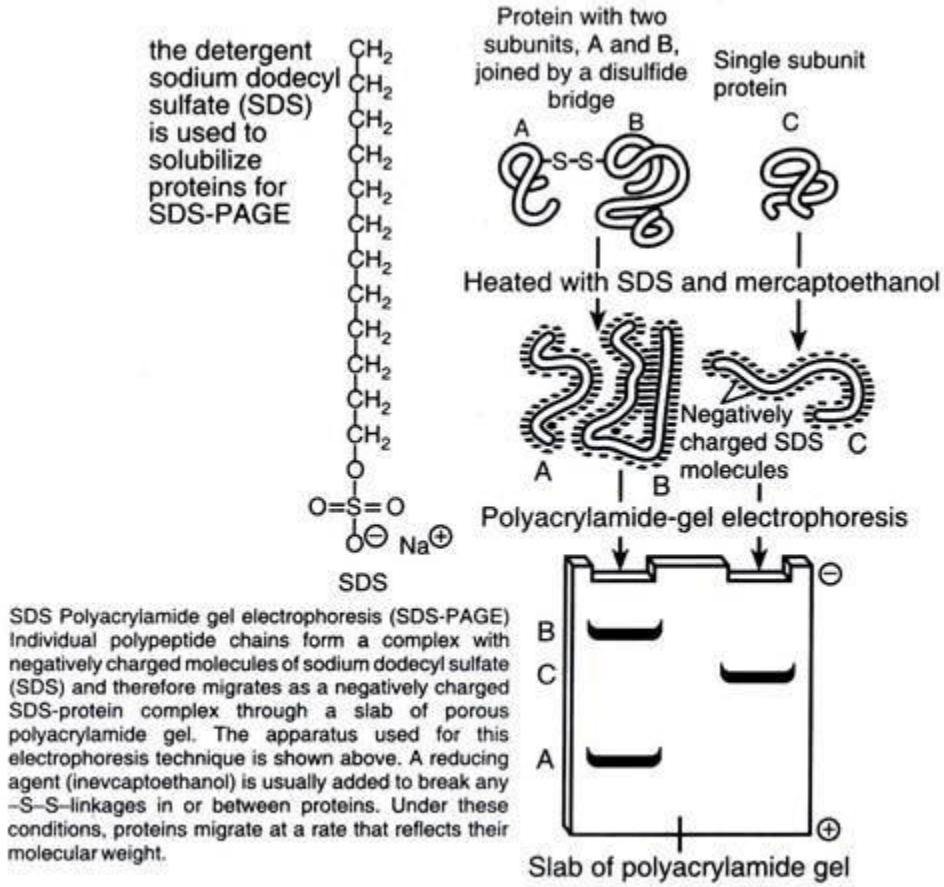
Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis:

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) is the most widely used method for analyzing protein mixtures qualitatively. It is particularly useful for monitoring protein purification and, as the method is based on the separation of proteins according to size, it can also be used to determine the relative molecular mass of proteins. SDS (CH_3-$(CH_2)_{10}$-CH_2OSO_3-Na^+) is an anionic detergent.

Samples to be run on SDS-PAGE are firstly boiled for 5 min in sample buffer containing β-mercaptoethanol and SDS. The beta β-mecaptoethnol reduces any disulphide bridges present that are holding together the protein tertiary structure, and the SDS binds strongly to, and denatures, the protein. Each protein in the mixture is, therefore, fully denatured by this treatment and opens up into a rod-shaped structure with a series of negatively charged SDS molecules along the polypeptide chain.

On average, one SDS molecule binds for every two amino acid residues. The original native charge on the molecule is, therefore, fully swamped by the negatively charged SDS molecules. The rod-like structure remains, as any rotation that tends to fold up the protein chain would result in repulsion between negative charges on different parts of the protein chain, returning the conformation back to rod-shape.

The sample buffer also contains an ionisable tracking dye, usually bromophenol blue, that allows the electrophoretic run to be monitored, and sucrose or glycerol, which gives the sample solution density thus allowing the sample to settle easily through the electro-phoresis buffer to bottom when injected into the loading well.



SDS Polyacrylamide gel electrophoresis (SDS-PAGE) Individual polypeptide chains form a complex with negatively charged molecules of sodium dodecyl sulfate (SDS) and therefore migrates as a negatively charged SDS-protein complex through a slab of porous polyacrylamide gel. The apparatus used for this electrophoresis technique is shown above. A reducing agent (inevcaptoethanol) is usually added to break any –S–S–linkages in or between proteins. Under these conditions, proteins migrate at a rate that reflects their molecular weight.

Fig. 8.18: Principle of SDS-PAGE

Before loading to gel the protein samples are lysed to denature them by boiling them in lysis buffer for 10 min. Lysis buffer contains buffer to maintain pH along with glycerol so as to make the protein heavy to properly settle in wells. Lysis buffer also have SDS to impart negative charge, DTT or β-mercaptoethanol to reduce the disulphide bonds in proteins.

Once the samples are all loaded, a current is passed through the gel. The samples to be separated are not in fact loaded directly into the main separating gel. When the main separating gel (normally about 5 cm long) has been poured between the glass plates and allowed to set, a shorter (approximately 0.5 cm) stacking gel is poured on top of the separating gel and it is into this gel that the wells are formed and the proteins loaded.

The purpose of this stacking gel is to concentrate the protein sample into a sharp band before it enters the main separating gel. This is achieved by utilizing the differences in ionic strength and pH between the electrophoresis buffer and the stacking gel buffer and involves a phenomenon known as isotachophoresis.

The stacking gel has a very large pore size (4% acrylamide), which allows the proteins to move freely and concentrate, or stack, under the effect of the electric field. The band-sharpening effect relies on the fact that negatively charged glycinate ions (in the electrophoresis buffer) have lower electrophoretic mobility than do the protein—SDS complexes, which, in turn, have lower mobility than the chloride ions (CI^-) of the loading buffer and the stacking gel buffer.

When the current is switched on, all the ionic species have to migrate at the same speed as CI^- only if they are in a region of high field strength. Field strength is inversely proportional to conductivity, which is proportional to concentration. The result is that the three species of interest adjust their concentration so that [CI^-] > [protein-SDS] > [glycinate].

There is only a small quantity of protein-SDS complexes, so they concentrate in a very tight band between glycinate and CI^- boundaries. Once the gycinate reaches the separating gel it becomes more fully ionized in the higher pH environment and its mobility increases. (The pH of the stacking gel is 6.8, that of the separating gel is 8.8).

Thus, the interface between glycinate and CI^- leaves behind the protein-SDS complex's, which are left to electrophorese at their own rates. The negatively charged protein-SDS complexes now continue to move towards the anode, and as because they have the same charge per unit length, they travel into the separating gel under the applied electric field with the same mobility.

However, as they pass through the separating gel the proteins separate, owing to the molecular sieving properties of the gel. Quite simply the smaller the protein the more easily it can pass through the pores of the gel, whereas large proteins are successively retarded by frictional resistance due to the sieving effect of the gels. Being a small molecule, the bromophenol blue dye is totally un-retarded and, therefore, indicates the electrophoresis front.

When the dye reaches the bottom of the gel, the current is turned off, and the gel is removed from between the glass plates and shaken in an appropriate stain solution and then washed in destain solution. The destain solution removes unbound background dye from the gel, leaving stained proteins visible as blue bands on a clear background.

A typical minigel would take about 1 h to prepare and set, 40 min to run at 200 V and have a 1 h staining time with Coomassie Brilliant Blue. Upon de-staining, strong protein bands would be seen in the gel within 10-20 min, but overnight de-staining is needed to completely remove all background stain. Vertical slab gel are invariably run, since this allows up to 10n different sample to be loaded onto a single gel. A typical SDS-polyacrylamide gel is shown in Fig. 8.19.

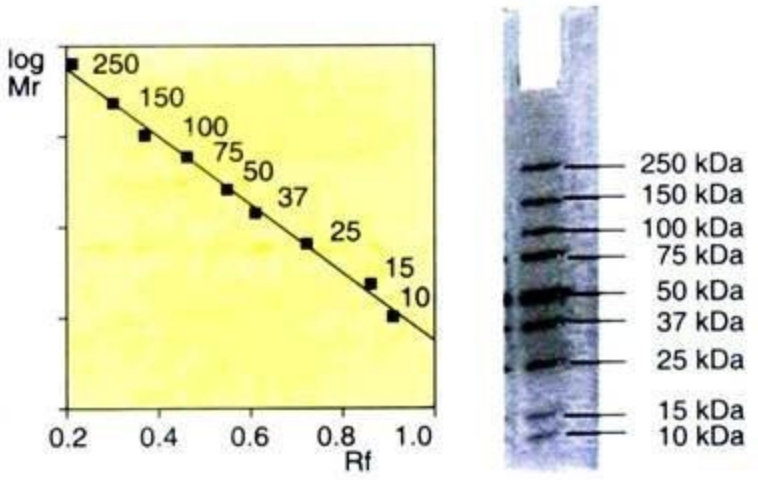


Fig. 8.19: Coomassie stained SDS-PAGE gel with separated starnds of molecular weight marker (right) and their relative mobility (left)

Typically, the separating gel used is a 15% polyacrylamide gel. This gives a gel a certain pore size in which proteins of relative molecular mass (M_r) 10,000 move through the gel relatively unhindered, whereas proteins of M_r 10, 00,000 can only just enter the pores of this gel. Gels of 15% polyacrylamide are, therefore, useful for separating proteins in the range M_r 1, 00,000 to 10,000.

However, a protein of M_r 1, 50,000, for example, would be unable to enter a 15% gel. In this case a larger-pored gel (e.g., a 10% or even 7.5% gel) would be used so that the protein could now enter the gel and be stained and identified. It is obvious; therefore, that the choice of gel to be used depends on the size of the proteins in the range of different percentage acrylamide gel is shown in Table 8.3.

Table 8.3: Percent of acrylamide required for separation of different molecular weight proteins		
% Acrylamide in resolving gel		Separation size range ($M_r \times 10^{-3}$)
Single percentage:	5%	36-200
	7.5%	24-200
	10%	14-200
	12.5%	14-100*
	15%	14-60*
Gradients:	5-15%	14-200
	5-20%	10-200
	10-20%	10-150

This shows, for example, that in a 10% polyacrylamide gel proteins greater than 200 kDa in mass cannot enter the gel, whereas proteins with relative molecular mass (M_r) in the range 200,000 to 15,000 will separate. Proteins of M_r 15,000 are too small to experience the sieving effect of the gel matrix, and all run together as a single band at the electrophoresis front.

The M_r of a protein can be determined by comparing its mobility with those of a number of standard proteins of known M_r that are run on a same gel. Plotting a graph of distance moved against log M_r for each of the standard proteins, a calibration curve can be constructed. The distance moved by the protein of unknown M_r is then measured, and then its log M_r and hence M_r can be determined from the calibration curve.

SDS-gel electrophoresis is often used after each step of a purification protocol to assess the purity or otherwise of the sample. A pure protein should give a single band on an SDS-polyacrylamide gel, unless the molecule is made up of two unequal subunits. In the latter case, two bands, corresponding to the two subunits, will be seen.

Since only sub-microgram amounts of the protein are needed for the gel, very little material is used in this form of purity assessment and at the same time a value for the relative molecular mass of the protein can be determined on the same gel run (as described above), with no more material being used.

Two-Dimensional Gel Electrophoresis:

Two-dimensional electrophoresis (2-D electrophoresis) is a powerful and widely used method for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples. This technique sorts proteins according to two independent properties in two discrete steps: the first-dimension step, isoelectric focusing (IEF), separates proteins according to their isoelectric points (pi); the second-dimension step, SDS-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their molecular weights (M_r, relative molecular weight).

Each spot on the resulting two-dimensional array corresponds to a single protein species in the sample. Thousands of different proteins can thus be separated, and information such as the protein pi, the apparent molecular weight, and the amount of each protein are obtained.

Two-dimensional electrophoresis was first introduced by P. H. O'Farrell and J. Klose in 1975. In the original technique, the first-dimension separation was performed in carrier ampholyte-containing polyacrylamide gels cast in narrow tubes. A. Gorg and colleagues developed the currently employed 2-D technique, where carrier ampholyte-generated pH gradients have been replaced with immobilized pH gradients and tube gels replaced with gels supported by a plastic backing.

A large and growing application of 2-D electrophoresis is "proteome analysis." The analysis involves the systematic separation, identification, and quantification of many proteins simultaneously from a single sample. Two-dimensional electrophoresis is used in this technique due to its unparalleled ability to separate thousands of proteins simultaneously.

Two-dimensional electrophoresis is also unique in its ability to detect post- and co-translational modifications, which cannot be predicted from the genome sequence. Applications of 2-D electrophoresis include proteome analysis, cell differentiation, and detection of disease markers, monitoring therapies, drug discovery, cancer research, purity checks, and micro-scale protein purification.

The 2-D process begins with sample preparation. Proper sample preparation is absolutely essential for a good 2-D result. The next step in the 2-D process is IPG (Isoelectric pH gradient) strip rehydration. IPG strips are provided dry and must be rehydrated with the appropriate additives prior to IEF (Immunoelectrophoresis).

First-dimension IEF is performed on a flatbed system at very high voltages with active temperature control. Next, strip equilibration in SDS-containing buffer prepares the sample for the second-dimension separation. Following equilibration, the strip is placed on the second-dimension gel for SDS-PAGE. The final steps are visualization and analysis of the resultant two-dimensional array of spots. 2-D electrophoresis begins with 1-D electrophoresis but then separates the molecules by a second property in direction 90 degrees from the first. In 1-D electrophoresis, proteins (or other molecules) are separated in one dimension, so that all the proteins/molecules will lie along a lane but be separated from each other by a property (e.g. isoelectric point). The result is that the molecules are

spread out across a 2-D gel. Because it is unlikely that two molecules will be similar in two distinct properties, molecules are more effectively separated in 2-D electrophoresis than in 1-D electrophoresis. The two dimensions that proteins are separated into using this technique can be isoelectric point, protein complex mass in the native state, and protein mass.

To separate the proteins by isoelectric point is called isoelectric focusing (IEF). Thereby, a gradient of pH is applied to a gel and an electric potential is applied across the gel, making one end more positive than the other. At all pHs other than their isoelectric point, proteins will be charged. If they are positively charged, they will be pulled towards the more negative end of the gel and if they are negatively charged they will be pulled to the more positive end of the gel. The proteins applied in the first dimension will move along the gel and will accumulate at their isoelectric point; that is, the point at which the overall charge on the protein is 0 (a neutral charge).

For the analysis of the functioning of proteins in a cell, the knowledge of their cooperation is essential. Most often proteins act together in complexes to be fully functional. The analysis of this sub organelle organisation of the cell requires techniques conserving the native state of the protein complexes. In native polyacrylamide gel electrophoresis (native PAGE), proteins remain in their native state and are separated in the electric field following their mass and the mass of their complexes respectively. To obtain a separation by size and not by net charge, as in IEF, an additional charge is transferred to the proteins by the use of coomassie or lithium dodecyl sulfate (LDS). After completion of the first dimension the complexes are destroyed by applying the denaturing SDS-PAGE in the second dimension, where the proteins of which the complexes are composed of are separated by their mass.

Before separating the proteins by mass, they are treated with sodium dodecyl sulfate (SDS) along with other reagents (SDS-PAGE in 1-D). This denatures the proteins (that is, it unfolds them into long, straight molecules) and binds a number of SDS molecules roughly proportional to the protein's length. Because a protein's length (when unfolded) is roughly proportional to its mass, this is equivalent to saying that it attaches a number of SDS molecules roughly proportional to the protein's mass. Since the SDS molecules are negatively charged, the result of this is that all of the proteins will have approximately the same mass-to-charge ratio as each other. In addition, proteins will not migrate when they have no charge (a result of the isoelectric focusing step) therefore the coating of the protein in SDS (negatively charged) allows migration of the proteins in the second dimension (NB SDS is not compatible for use in the first dimension as it is charged and a nonionic or zwitterionic detergent needs to be used). In the second dimension, an electric potential is again applied, but at a 90 degree angle from the first field. The proteins will be attracted to the more positive side of the gel proportionally to their mass-to-charge ratio. As previously explained, this ratio will be nearly the same for all proteins. The proteins' progress will be slowed by frictional forces. The gel therefore acts like a molecular sieve when the current is

applied, separating the proteins on the basis of their molecular weight with larger proteins being retained higher in the gel and smaller proteins being able to pass through the sieve and reach lower regions of the gel.
The result of this is a gel with proteins spread out on its surface. These proteins can then be detected by a variety of means, but the most commonly used stains are silver and coomassie staining. In this case, a silver colloid is applied to the gel. The silver binds to cysteine groups within the protein. The silver is darkened by exposure to ultra-violet light. The darkness of the silver can be related to the amount of silver and therefore the amount of protein at a given location on the gel. This measurement can only give approximate amounts, but is adequate for most purposes.

Molecules other than proteins can be separated by 2D electrophoresis. In supercoiling assays, coiled DNA is separated in the first dimension and denatured by a DNA intercalator (such as ethidium bromide or the less carcinogenic chloroquine) in the second. This is comparable to the combination of native PAGE /SDS-PAGE in protein separation.

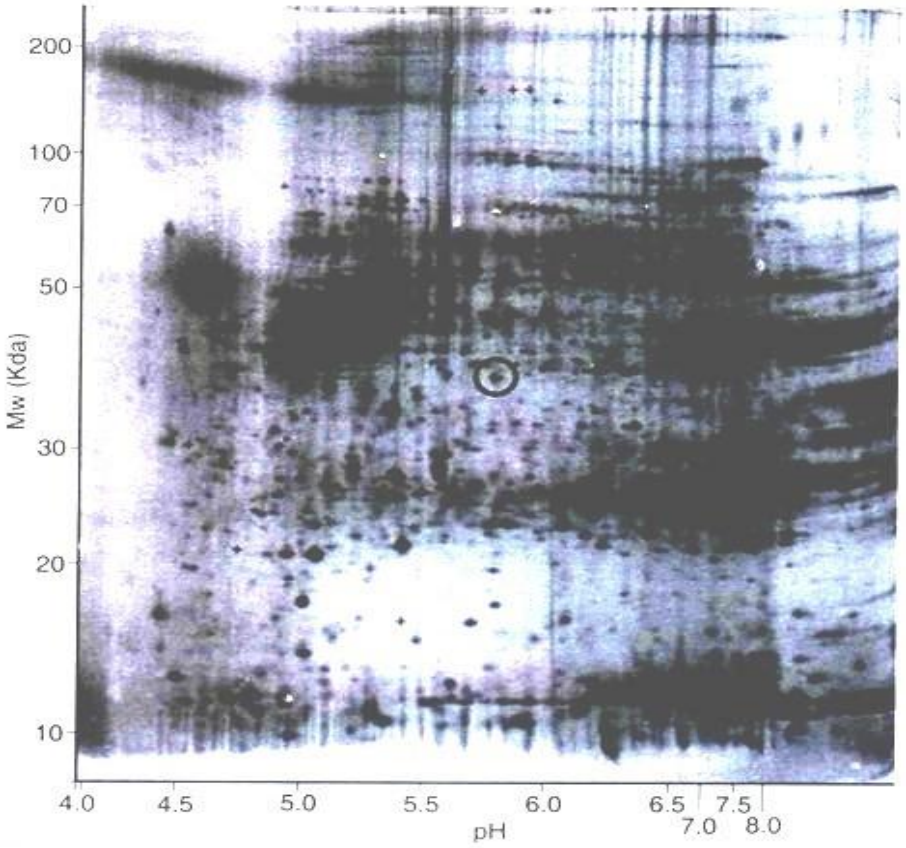


Fig. 8.28: A 2D gel (*Courtsey*: Dr D Dash, Deptt. of Biochemistry, IMS, BHU)

Silver staining is the most sensitive non-radioactive method (below 1 ng). Silver staining is a complex, multi-step process utilizing numerous reagents for which quality is critical. It is, therefore, often advantageous to purchase these reagents in the form of a dedicated kit, in which the reagents are quality assured specifically for the silver-staining application. By omitting glutardialdehyde from the sensitizer and formaldehyde from the silver nitrate solution the method becomes compatible with mass spectrometry analysis, however, at the expense of sensitivity.

Coomassie staining, although 50 to 100-fold less sensitive than silver staining, is a relatively simple method and more quantitative than silver staining. Coomassie blue is preferable when relative amounts of protein are to be determined by densitometry. Colloidal staining methods are recommended, because they show the highest sensitivity, down to 100 ng/protein spot.

Negative Zinc—Imidazole staining has a detection limit of approx. 15 ng protein/spot and is well compatible with mass spectrometry, but it is a poor quantification technique. Fluorescent labelling and fluorescent staining with dyes have a sensitivity in-between colloidal Coomassie and Silver Staining.

These techniques require fluorescence scanners, but they are compatible with mass spectrometry and show a wide dynamic range for quantification. Apart from staining, Second-dimension gels can be blotted onto a nitrocellulose or PVDF membrane for immunochemical detection of specific proteins or chemical micro sequencing.

Preserving the Gels:

The gels are optimally stored in sheet protectors after soaking them in 10% v/v glycerol for 30 min. Un-backed gels are shrunk back to their original sizes by soaking them in 30% (v/v) methanol or ethanol/4% glycerol until they match their original sizes. For autoradiography the gels are dried onto strong filter paper with a vacuum drier or in-between two sheets of wet cellophane sealed at ends.

Further Analysis of Protein Spots:

a. Picking the spots:

Robotic systems are available that automatically picks selected protein spots from stained or de-stained gels using a pick list from the image analysis, and transfers them into micro-plates for further analysis.

b. Digestion of the proteins:

The gel plugs are automatically digested in the computer controlled Digester; the supernatant peptides are mixed with MALDI matrix material and spotted onto MALDI slides using robotic spotter.

c. MALDI-ToF mass spectrometry:

In the MALDI-ToF mass spectrometer, a laser beam is fired into the dried peptide-matrix spots for ionization of the peptides. After accurate determination of the peptide masses, databases are searched for identification of the original proteins.

Probable Questions:

1. Write basic principle of electrophoresis.
2. How agarose can resolute DNA molecule.
3. What is Polyacrylamide? How it is used in separation of protein?
4. Why SDS is used in SDS-PAGE?
5. What is the principle of SDS-PAGE?
6. What is resolving gel and stacking gel?
7. What is isoelectric focussing? How it can help to separate different proteins?
8. What is 2D PAGE? Why it is more authentic than 1D AGE in protein separation?
9. How sequencing gel are prepared?
10. How separated spots are analysed in 2D PAGE?

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

UNIT-XV

Spectroscopy: UV- spectroscopy and Circular Dichroism

Objective: In this unit we will discuss spectroscopy with special emphasis to spectrophotometer. We will also discuss circular dichroism. .

Principle of Spectroscopy:

Spectroscopy is the study of the interaction of electromagnetic radiation with matter. When matter is energized (excited) by the application of thermal, electrical, nuclear or radiant energy, electromagnetic radiation is often emitted as the matter relaxes back to its original (ground) state.

The spectrum of radiation emitted by a substance that has absorbed energy is called an emission spectrum and the science is appropriately called emission spectroscopy.

Another approach often used to study the interaction of electromagnetic radiation with matter is one whereby a continuous range of radiation (e.g., white light) is allowed to fall on a substance; then the frequencies absorbed by the substance are examined.

The resulting spectrum from the substance contains the original range of radiation with dark spaces that correspond to missing, or absorbed frequencies. This type of spectrum is called an absorption spectrum. In spectroscopy the emitted or absorbed radiation is usually analyzed, i.e., separated into the various frequency components and the intensity is measured by means of an instrument called spectrometer.

The resultant spectrum is mainly a graph of intensity of emitted or absorbed radiation versus wavelength or frequency. There are in general three types of spectra: continuous, line, and band. The sun and heated solids produce continuous spectra in which the emitted radiation contains all frequencies within a region of the electromagnetic spectrum. A rainbow and light from a light bulb are examples of continuous spectra.

Line spectra are produced by excited atoms in the gas phase and contain only certain frequencies, all other frequencies being absent. Each chemical element of the periodic chart has a unique and, therefore, characteristic line spectrum. Band spectra are produced by excited molecules emitting radiation in groups of closely spaced lines that merge to form bands. These categories of emission and absorption spectra contain tremendous amounts of useful information about the structure and composition of matter. Spectroscopy is a powerful and sensitive form of chemical analysis, as well as a method of probing electronic and nuclear structure and chemical bonding. The key to interpreting this spectral information is the knowledge that certain atomic and molecular processes involve only certain energy ranges. Fig. 12.1 shows the regions of the electromagnetic spectrum and the associated energy transitions that occur in atomic and molecular processes.

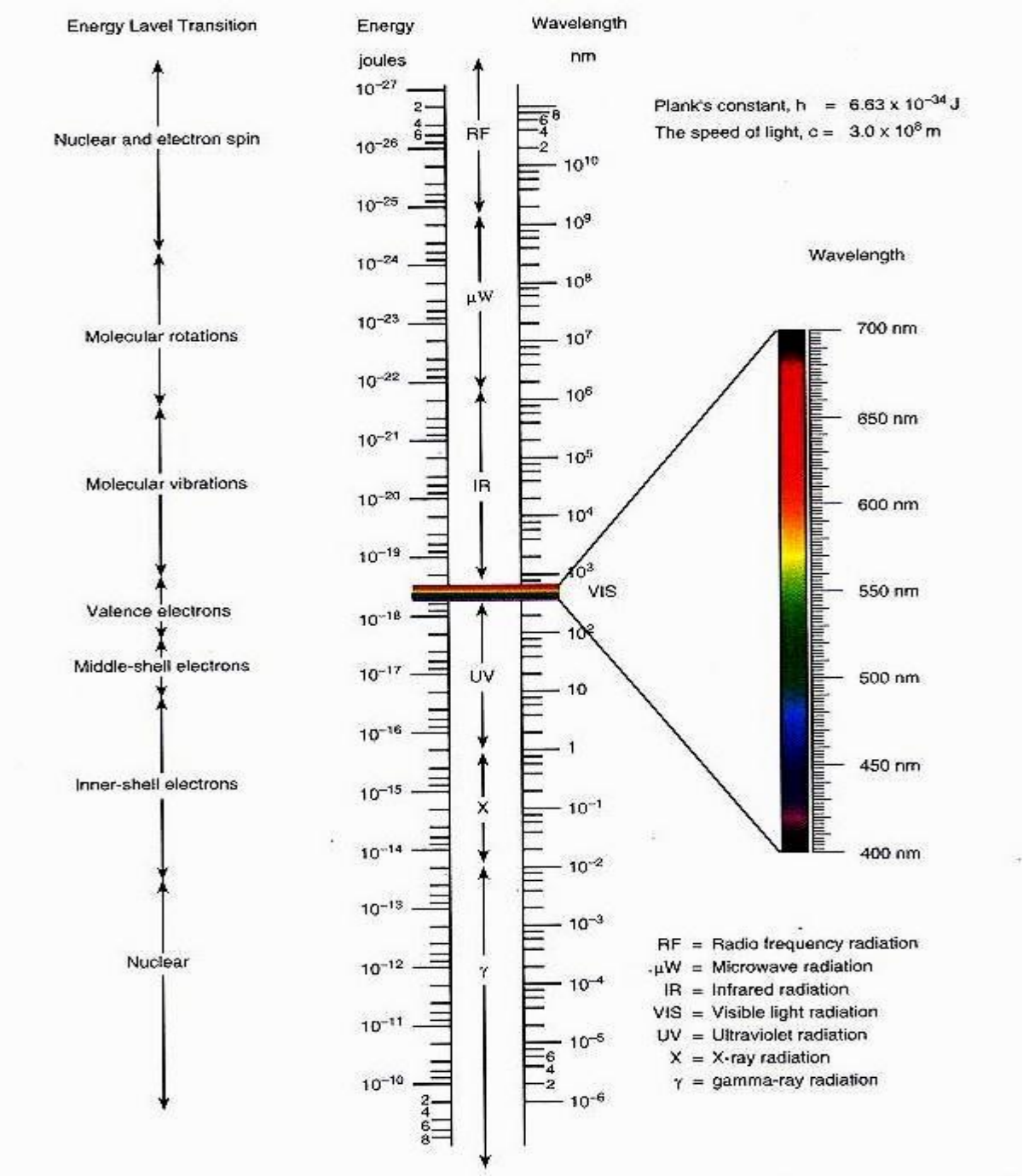


Fig. 12.1: The electromagnetic radiation ruler

Much of the scientific knowledge of the structure of the universe, from stars to atoms, is derived from interpretations of the interaction of radiation with matter. One example of the power of these techniques is the determination of the composition, the velocities, and the evolutionary dynamics of stars.

The source of the incredible amount of energy produced by the sun is nuclear fusion reactions going on within the hot interior (temperature 40 × 106 K). Two fusion cycles, the carbon cycle and the proton cycle, convert hydrogen nuclei into helium nuclei via heavier nuclei, such as carbon 12 and nitrogen 14. The enormous radiation of energy from the hot core seethes outwards by convection. This radiation consists of the entire electromagnetic spectrum as a continuous spectrum. Towards the surface of the sun (the photosphere), the different elements all absorb at their characteristic frequencies. The radiation that shoots into space toward the earth is a continuous emission spectrum with about 22,000 dark absorption lines present in it (Fraunhofer lines), of which about 70% have been identified. These absorption lines, i.e., missing frequencies, prove that more than 60 terrestrial elements are certainly present in the sun.

Spectrophotometer:

A **spectrophotometer** can be located in many studies, biology, chemistry, and industrial laboratories. The spectrophotometer is utilized for research and data evaluation in different scientific fields.

Some of the major fields in which a spectrophotometer is employed are physics, molecular biology, chemistry, and biochemistry labs. Generally, the title refers to Ultraviolet-Visible (UV-Vis) Spectroscopy. What a spectrophotometer does is transmit and receive light. The spectrophotometer is utilized to evaluate samples of test material by passing light by means of the sample and studying the intensity of the wavelengths.

Different samples modify the light in numerous distinct ways and this allows researchers to obtain much more facts about the check content, by viewing the change in light conduct as it passes by way of the sample. These final results must be precise or the researcher will just be throwing away time making use of a flawed instrument. The only way to make sure accuracy is by executing a spectrophotometer calibration.

Quantification of light absorption: The chance for a photon to be absorbed by matter is given by an extinction coefficient which itself is dependent on the wavelength l of the photon. If light with the intensity I0 passes through a sample with appropriate transparency and the path length (thickness) d, the intensity I drops along the pathway in an exponential manner. The characteristic absorption parameter for the sample is the extinction coefficient a, yielding the correlation $I= I_0\ e^{-\alpha d}$ .The ratio $T= I/I_0$ is called transmission. Biochemical samples usually comprise aqueous solutions, where the substance of interest is present at a molar concentration c.

$$\lg \frac{I_0}{I} = \lg \frac{1}{T} = \varepsilon \times c \times d = A \qquad (12.2)$$

where $[d] = 1\ \text{cm}$, $[c] = 1\ \text{mol dm}^{-3}$, and $[\varepsilon] = 1\ \text{dm}^3\ \text{mol}^{-1}\ \text{cm}^{-1}$. ε is the molar absorption coefficient (also molar extinction coefficient) ($\alpha = 2.303 \times c \times \varepsilon$). A is the absorbance of the sample, which is displayed on the spectrophotometer.

Beer–Lambert Law:

The **Beer-Lambert law (or Beer's law)** is the linear relationship between absorbance and concentration of an absorbing species. The general Beer-Lambert law is usually written as:

$$A = a(\lambda) * b * c$$

where A is the measured absorbance, $a(\lambda)$ is a wavelength-dependent absorptivity coefficient, **b** is the path length, and **c** is the analyte concentration. When working in concentration units of molarity, the Beer-Lambert law is written as:

$$A = \varepsilon * b * c$$

where ε is the wavelength-dependent molar absorptivity coefficient with units of $M^{-1}\ cm^{-1}$. Data are frequently reported in percent transmission ($I/I_0 * 100$) or in absorbance [$A = \log (I/I_0)$]. The latter is particularly convenient. [common coefficients of near-ultraviolet absorption bands of some amino acids and nucleotides]

Sometimes the extinction coefficient is given in other units; for example,

$$A = E^{1\%} * b * c$$

where the concentration C is in gram per 100 ml of solution. This useful when the molecular weight of the solute is unknown or uncertain.

The Beer–Lambert law is valid for low concentrations only. Higher concentrations might lead to association of molecules and therefore cause deviations from the ideal behaviour. Absorbance and extinction coefficients are additive parameters, which complicates determination of concentrations in samples with more than one absorbing species. Note that in dispersive samples or suspensions scattering effects increase the absorbance, since the scattered light is not reaching the detector for readout. The absorbance recorded by the spectrophotometer is thus overestimated and needs to be corrected.

Deviations from the Beer–Lambert law:

According to the Beer–Lambert law, absorbance is linearly proportional to the concentration of chromophores. This might not be the case any more in samples with high absorbance. Every spectrophotometer has a certain amount of stray light, which is light received at the detector but not anticipated in the spectral band isolated by the monochromator. In order to obtain reasonable signal-to-noise ratios, the intensity of light at the chosen wavelength (Il) should be 10 times higher than the intensity of the stray light (Istray). If the stray light gains in intensity, the effects measured at the detector have

nothing or little to do with chromophore concentration. Secondly, molecular events might lead to deviations from the Beer–Lambert law. For instance, chromophores might dimerize at high concentrations and, as a result, might possess different spectroscopic parameters.

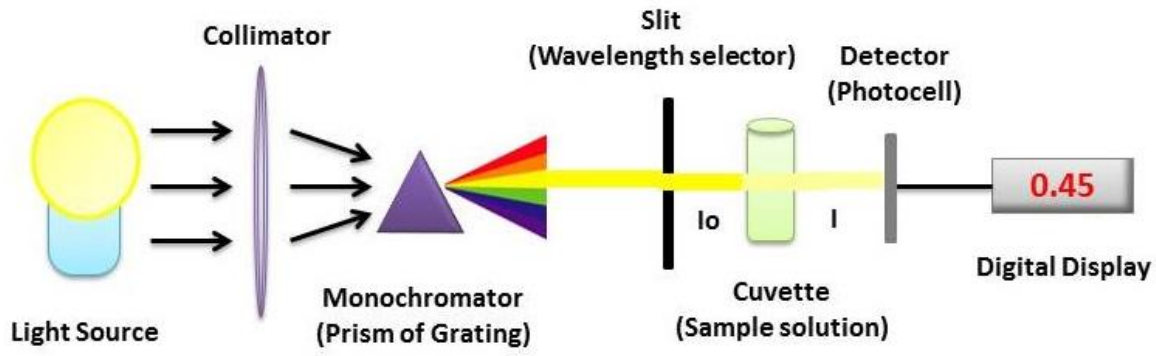
Definition: In chemistry, **spectrophotometry** is the quantitative measurement of the reflection or transmission properties of a material as a function of wavelength.[2] It is more specific than the general term electromagnetic spectroscopy in that spectrophotometry deals with visible light, near-ultraviolet, and near-infrared, but does not cover time-resolved spectroscopic techniques.

Principle

The Spectrophotometer is a much more refined version of a colorimeter. In a colorimeter, filters are used which allow a broad range of wavelengths to pass through, whereas in the spectrophotometer a prism (or) grating is used to split the incident beam into different wavelengths. By suitable mechanisms, waves of specific wavelengths can be manipulated to fall on the test solution. The range of the wavelengths of the incident light can be as low as 1 to 2nm. *The spectrophotometer* is useful for measuring the absorption spectrum of a compound, the absorption of light by a solution at each wavelength. This is the basic Principle of spectrophotometry in biochemistry.

Spectrophotometer Instrumentation

The essential components of spectrophotometer instrumentation include:



Basic Instrumentation of a Spectrophotometer

1. A Stable and cheap radiant energy source
2. A monochromator, to break the polychromatic radiation into component wavelength (or) bands of wavelengths.
3. Transport vessels (cuvettes), to hold the sample
4. A Photosensitive detector and an associated readout system

1. Radiant Energy Sources

Materials that can be excited to high energy states by a high-voltage electric discharge (or) by electrical heating serve as excellent radiant energy sources.

a. **Sources of Ultraviolet radiation:** The most commonly used sources of UV radiation are the hydrogen lamp and the deuterium lamp. Xenon lamp may also be used for UV radiation, but the radiation produced is not as stable as the hydrogen lamp.

b. **Sources of Visible radiation:** "Tungsten filament" lamp is the most commonly used source for visible radiation. It is inexpensive and emails continuous radiation in the range between 350 and 2500nm. "Carbon arc" which provides more intense visible radiation is used in a few commercially available instruments.

c. Sources of IR radiation: "Nernst Glower" and "Global" are the most satisfactory sources of IR radiation. Global is more stable than the nearest flower.

3. Wavelength Selectors: Wavelength selectors are of two types.

1. **Filters:** "Gelatin" filters are made of a layer of gelatin, coloured with organic dyes and sealed between glass plates.

2. **Monochromators:** A monochromator resolves polychromatic radiation into its individual wavelengths and isolates these wavelengths into very narrow bands. The essential components of a monochromator are.
   - Entrance slip-admits polychromatic light from the source
   - Collimating device–Collimates the polychromatic light onto the dispersion device.
   - Wavelength resolving device like a PRISM (or) a GRATING
   - A focusing lens (or) a mirror
   - An exit slip–allows the monochromatic beam to escape.

The kinds of resolving element are of primary importance

- **PRISMS**
- **GRATINGS**

PRISMS:

A prism disperses polychromatic light from the source into its constituent wavelengths by virtue of its ability to reflect different wavelengths to a different extent;
The degree of dispersion by the prism depends on upon

- The optical angle of the Prism (usually 60^0)
- The material of which it is made

Two types of Prisms are usually employed in commercial instruments. Namely, 60^0 cornu quartz prism and 30^0 Littrow Prism.

GRATINGS:

Gratings are often used in the monochromators of spectrophotometers operating ultraviolet, visible and infrared regions.

3. Sample Containers

Sample containers are also one of the parts of Spectrophotometer instrumentation. Samples to be studied in the ultraviolet (or) visible region are usually glasses (or) solutions and are put in cells known as "CUVETTES".
Cuvettes meant for the visible region are made up of either ordinary glass (or) sometimes Quartz. Most of the spectrophotometric studies are made in solutions, the solvents assume prime importance.
The most important factor in choosing the solvent is that the solvent should not absorb (optically transparent) in the same region as the solute.

4. Detection Devices

Most detectors depend on the photoelectric effect. The current is then proportional to the light intensity and therefore a measure of it. Important requirements for a detector including

- High sensitivity to allow the detection of low levels of radiant energy
- Short response time
- Long-term stability
- An electric signal which easily amplified for a typical readout apparatus.

5. Amplification And Readout

Radiation detectors generate electronic signals which are proportional to the transmitted light. These signals need to be translated into a form easy to interpret. This is accomplished by using amplifiers, Ammeters, Potentiometers and Potentiometric recorders.

The above 5 major parts are the major part of Spectrophotometer instrumentation. Now let us see the Applications of Spectrophotometer.

Spectrophotometer Applications

How to use the spectrophotometer? There are uses of spectrophotometry in biochemistry which are listed below:

1. **Qualitative Analysis**

The visible and UV spectrophotometer may be used to identify classes of compounds in both the pure state and in biological preparations. This is done by plotting absorption spectrum curves. Absorption by a compound in different regions gives some hints to its structure.

2. Quantitative Analysis

Spectrophotometer uses in the Quantitative analysis of Biochemistry practicals. Quantitative analysis method developing for determining an unknown concentration of a species by absorption spectrometry.

Most of the organic compounds of biological interest absorb in the UV-visible range of the spectrum.

Thus, several important classes of biological compounds may be measured semi-quantitatively using the UV-visible spectrophotometer. Nucleic acids at 254nm protein at 280nm provide good examples of such use.

The absorbance at 280nm by proteins depends on their “Tyrosine” and “Tryptophan” content.

- **Estimation of Proteins by Lowry method**
- **Estimation of Tyrosine by Folin-Ciocalteau Method**
- **Estimation of Blood Glucose level by Folin-Wu method**

3. Enzyme Assay:

This is the basic application of spectrophotometry. This assay is carried out most quickly and conveniently when the substrate (or) the product is color (or) absorbs light in the UV range.

Eg 1: Lactate Dehydrogenase (LDH)

Lactate + NAD^+ ↔ Pyruvate + NADH + H^+

- The LDH is engaged in the transfer of electrons from lactate to NAD^+.
- The products of the reaction are pyruvate, NAD, and a proton
- One of the products, NADH, **absorbs radiation in the UV range at 340 nm** while its oxidized counterpart, NAD^+ does not.
- The reaction in the forward direction can be followed by measuring the increment in the light absorption of the system at 540nm in a spectrophotometer.

Eg 2: Pyruvate Kinase

Phosphoenolpyruvate + ADP ↔ Pyruvate + ATP

Pyruvate + NADH + H^+ ↔ Lactate + NAD^+

We have added a large excess of NADH to the system, the system now absorbs at 340nm. According to the above-given reactions, each molecule of Pyruvate formed in the reaction, **a molecule of NADH is oxidized to NAD+** in the second reaction when the system converts pyruvate to locate. Since NAD+ does not absorb at 340nm, the absorbance goes on decreasing with increased pyruvate generation. Such measurements are known as "Coupled assays".

4. Molecular Weight Determination

Molecular weights of amine picrates, sugars and much aldehyde and ketone compounds have been determined by this method. Molecular weights of only small molecules may be determined by this method.

1. **Study of Cis-Trans Isomerism:** Geometrical isomers differ in the spatial arrangement of groups about a plane, the absorption spectra of the isomers also differs. The trans-isomer is usually more elongated than its cis counterpart. Absorption spectrometry can be utilized to study Cis-Trans isomerism.
2. **Control of Purification:** Impurities in a compound can be detected very easily by spectrophotometric studies. "Carbon disulfide" impurity in carbon tetrachloride can be detected easily by measuring absorbance at 318nm where carbon sulfide absorbs. A lot many commercial solutions are routinely tested for purity spectroscopically.

5. Other Physiochemical Studies:

Spectrophotometry (UV-VIS) has been used to study the following physiochemical phenomena:

- Heats of formation of molecular addition compound and complexes in solution
- Determination of the empirical formula
- Formation constants of complexes in solution
- Hydration equilibrium of carbonyl compounds
- Association constants of weak acids and bases in organic solvents
- Protein-dye interactions
- Chlorophyll-Protein complexes
- Vitamin-A aldehyde–Protein complex
- Determination of reaction rates
- Dissociation constants of acids and bases
- Association of cyanine dyes

Circular Dichroism

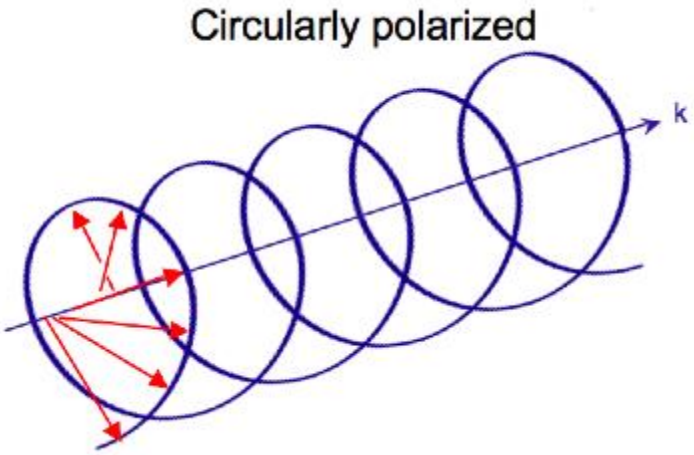
Introduction

Circular Dichroism, an absorption spectroscopy, uses circularly polarized light to investigate structural aspects of optically active chiral media. It is mostly used to study biological molecules, their structure, and interactions with metals and other molecules. Circular Dichroism (CD) is an absorption spectroscopy method based on the differential absorption of left and right circularly polarized light. Optically active chiral molecules will preferentially absorb one direction of the circularly polarized light. The difference in absorption of the left and right circularly polarized light can be measured and quantified. UV CD is used to determine aspects of protein secondary structure. Vibrational CD, IR CD, is used to study the structure of small organic molecules, proteins and DNA. UV/Vis CD investigates charge transfer transitions in metal-protein complexes.

Circular Polarization of Light

Electromagnetic radiation consists of oscillating electric and magnetic fields perpendicular to each other and the direction of propagation. Most light sources emit waves where these fields oscillate in all directions perpendicular to the propagation vector. Linear polarized light occurs when the electric field vector oscillates in only one plane. In circularly polarized light, the electric field vector rotates around the propagation axis maintaining a constant magnitude. When looked at down the axis of propagation the vector appears to trace a circle over the period of one wave frequency (one full rotation occurs in the distance equal to the wavelength). In linear polarized light the direction of the vector stays

constant and the magnitude oscillates. In circularly polarized light the magnitude stays constant while the direction oscillates.



As the radiation propagates the electric field vector traces out a helix. The magnetic field vector is out of phase with the electric field vector by a quarter turn. When traced together the vectors form a double helix. Light can be circularly polarized in two directions: left and right. If the vector rotates counterclock wise when the observer looks down the axis of propagation, the light is left circularly polarized (LCP). If it rotates clockwise, it is right circularly polarized (RCP). If LCP and RCP of the same amplitude, they are superimposed on one another and the resulting wave will be linearly polarized.

Interaction with Matter

As with linear polarized light, circularly polarized light can be absorbed by a medium. An optically active chiral compound will absorb the two directions of circularly polarized light by different amounts

$$\Delta A = Al - Ar$$

This can be extended to the Beer-Lambert Law. The molar absorpitivty of a medium will be different for LCP and RCP. The Beer-Lambert Law can be rewritten as

$$A = (\varepsilon l - \varepsilon r)cl$$

The difference in molar absorptivity is also known as the molar circular dichroism

$$\Delta\varepsilon = \varepsilon l - \varepsilon r$$

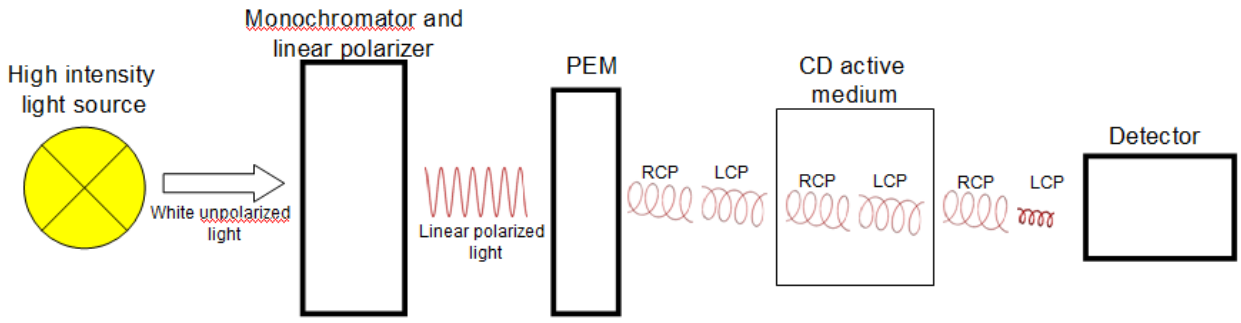
The molar circular dichroism is not only wavelength dependent but also depends on the absorbing molecules conformation, which can make it a function of concentration, temperature, and chemical environment.

Any absorption of light results in a change in amplitude of the incident wave; absorption changes the intensity of the light and intensity of the square of the amplitude. In a chiral medium the molar absorptivities of LCP and RCP light are different so they will be absorbed by the medium in different amounts. This differential absorption results in the LCP and RCP having different amplitudes which means the superimposed light is no longer linearly polarized. The resulting wave is elliptically polarized.

Applications

Instrumentation

Most commercial CD instruments are based on the modulation techniques introduced by Grosjean and Legrand. Light is linearly polarized and passed through a monochromator. The single wavelength light is then passed through a modulating device, usually a photoelastic modulator (PEM), which transforms the linear light to circular polarized light. The incident light on the sample switches between LCP and RCP light. As the incident light swtches direction of polarization the absorption changes and the differention molar absorptivity can be calculated.



Biological molecules

The most widely used application of CD spectroscopy is identifying structural aspects of proteins and DNA. The peptide bonds in proteins are optically active and the ellipticity they exhibit changes based on the local conformation of the molecule. Secondary structures of proteins can be analyzed using the far-UV (190-250 nm) region of light. The ordered α-

helices, β-sheets, β-turn, and random coil conformations all have characteristic spectra. These unique spectra form the basis for protein secondary structure analysis. It should be noted that in CD only the relative fractions of residues in each conformation can be determined but not specifically where each structural feature lies in the molecule. In reporting CD data for large biomolecules it is necessary to convert the data into a normalized value that is independent of molecular length. To do this the molar ellipticity is divided by the number of residues or monomer units in the molecule.

The real value in CD comes from the ability to show conformational changes in molecules. It can be used to determine how similar a wild type protein is to mutant or show the extent of denaturation with a change in temperature or chemical environment. It can also provide information about structural changes upon ligand binding. In order to interpret any of this information the spectrum of the native conformation must be determined. Some information about the tertiary structure of proteins can be determined using near-UV spectroscopy. Absorptions between 250-300 nm are due to the dipole orientation and surrounding environment of the aromatic amino acids, phenylalanine, tyrosine, and tryptophan, and cysteine residues which can form disulfide bonds. Near-UV techniques can also be used to provide structural information about the binding of prosthetic groups in proteins.

Metal containing proteins can be studied by visible CD spectroscopy. Visible CD light excites the d-d transitions of metals in chiral environments. Free ions in solution will not absorb CD light so the pH dependence of the metal binding and the stoichiometry can be determined. Vibrational CD (VCD) spectroscopy uses IR light to determine 3D structures of short peptides, nucleic acids, and carbohydrates. VCD has been used to show the shape and number of helices in A-, B-, and Z-DNA. VCD is still a relatively new technique and has the potential to be a very powerful tool. Resolving the spectra requires extensive *ab initio* calculations, as well as, high concentrations and must be performed in water, which may force the molecule into a nonnative conformation.

Probable Questions:

1. What is the basic principle of spectroscopy?
2. Explain Lambert-Beer law.
3. What is molar extinction coefficient? How it is calculated?
4. Describe different components of a spectrophotometer?
5. Describe different applications of spectrophotometer?
6. What are the applications of CD?
7. Write the basic principles of CD?

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

Unit-XVI

Blotting Methods: Southern, Northern & Western Blotting

Objective: In this unit we will discuss different types of blotting techniques such as Southern Blotting for DNA, Northern Blotting for RNA and Western Blotting for Proteins.

Southern Blot Definition

Southern blot is the process of transfer of DNA fragments that are separated by electrophoresis onto a membrane for immobilization and identification. Southern blotting has been adopted as a routine procedure for the analysis of DNA samples for different applications.

The technique was discovered by Edwin Southern, and the technique was named after him. The technique later gave rise to other techniques like Western Blotting and Northern Blotting that are based on the same principle. The most basic form of the technique is used to determine the size of a DNA fragment from a complex mixture of genomic DNA. The technique is also relatively quantitative and can be used to determine the number of copies of a segment present in a genome.

Southern Blotting can be modified based on the choices of the membrane, transfer buffer, and method. The most commonly used membrane is the nitrocellulose membrane, as it is robust and can be reprobed a number of times. Similarly, the original protocol of the southern blotting utilizes the use of radioactive probes; however, other labeling systems utilizing fluorescence and chemiluminescence. Southern blotting has been modified in a number of ways to better serve the application and has been made more complex and efficient.

Principle of Southern Blot

The principle of southern blotting is similar to the blotting technique involving the transfer of biomolecules from a membrane to another for detection and identification. The DNA to be analyzed is digested with restriction enzymes and fractionated by size by the process of agarose gel electrophoresis. The DNA strands are denatured by alkaline treatment and are transferred to a nylon or nitrocellulose membrane by the blotting process.

The strands on the membrane are immobilized on the surface by baking or UV irradiation. The DNA sequences on the membrane can be detected by the process of hybridization.

Hybridization reactions are specific as the probes used bind to target fragments consisting of complementary sequences. The probes used are labelled with different components that can be visualized by different methods depending on the type of probes used.

Procedure of Southern Blot

a. Restriction digestion of DNA

About 10 μg of the extracted genomic DNA is digested with the appropriate restriction enzyme in a microcentrifuge tube.

The tube is incubated overnight at 37°C. In some cases, the tubes are heated in a water bath at 65°C for 20 minutes after the incubation to denature the restriction enzymes. To the tubes, 10μl of the DNA sample buffer is added, and the mixture is poured on agarose gel for electrophoresis.

b. Electrophoresis

The percentage and size of the gel are determined based on the size of the DNA fragments to be separated. The gel is then prepared accordingly. The electrophoresis buffer is prepared with ethidium bromide and poured into the tank in a way that is a few millimeters above the gel support. The gel cast is prepared along with a comb with teeth to form wells that can hold the sample volume. Once the comb is in place, the gel is slowly poured into the cast. Once the gel has set, the comb is removed, and the gel is placed on the tank. Running buffer is added to the tank to cover the gel. The samples are prepared by adding loading buffer and carefully pipetted into the wells. The tank is connected to the power supply and allowed to run overnight.

c. Denaturation

The gel is removed from the electrophoresis apparatus and placed in a glass tray with 500 ml denaturation buffer (1.5 M NaCl and 0.5 M NaOH) for 45 minutes at room temperature. The denaturation buffer is poured off and replaced with a neutralization buffer. The gel is allowed to soak for 1 hour while slowly rotating on a platform rotator.

d. Blotting

An oblong sponge that is slightly larger than the gel is placed on a glass dish which is filled with SSC to leave the soaked sponge about half-submerged in the buffer. Three pieces of Whatman 3mm paper are cut the same size as the sponge. These are placed on the sponge and wet with SSC.

The gel is placed on the filter paper and squeezed out to remove bubbles by rolling a glass pipette over the surface. A nylon membrane, just large enough to cover the surface of the gel is placed on top of the gel. The membrane is further flooded with SSC, and few sheets of filter paper are placed on top of it. Finally, a glass plate is laid on top of the structure to hold everything in place. The DNA transfer is allowed to occur overnight.

e. Baking/ Immobilization

The nylon membrane is removed from the blotting structure and attached to a vacuum or regular oven at 80°C for 2-3 hours. The DNA strands on the membrane can also be immobilized by exposing the membrane to ultraviolet radiation.

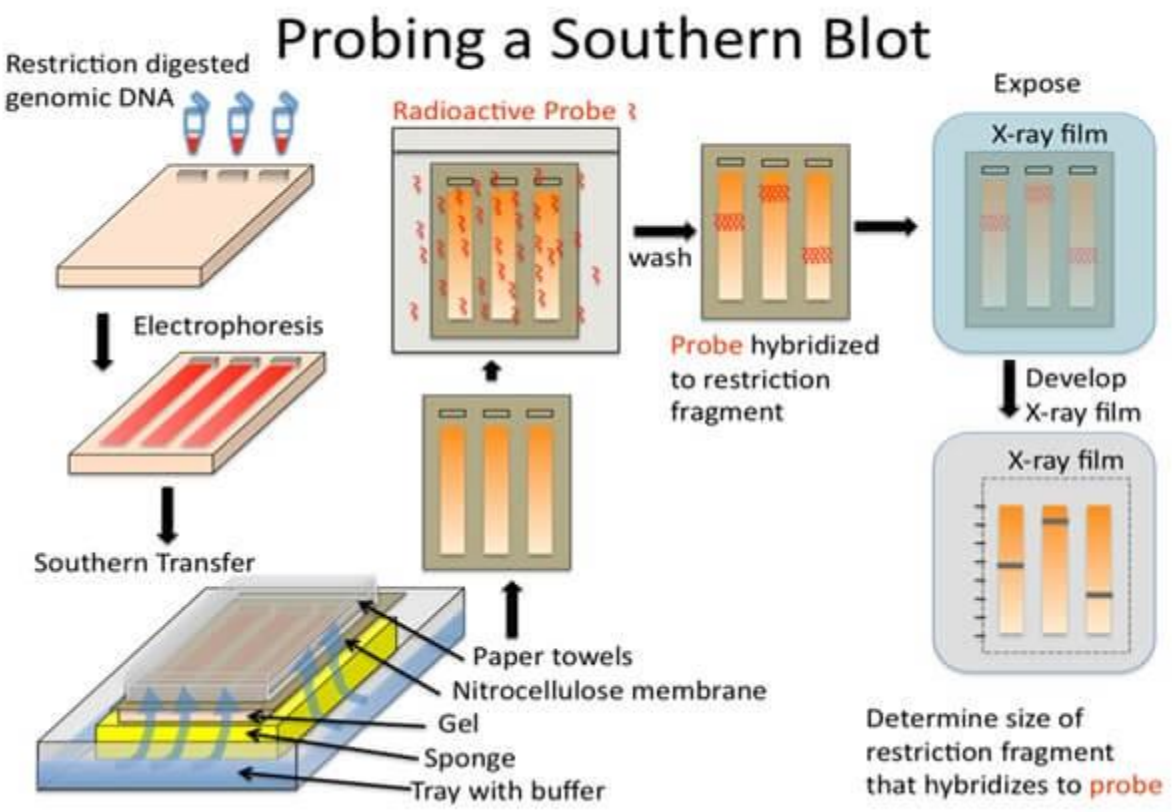
f. Hybridization

The membrane is exposed to the hybridization probe, which can either be a DNA fragment or an RNA segment with a specific sequence that detects the target DNA. The probe nucleic acid is labeled so that it can be detected by incorporating radioactivity or tagging the molecules with fluorescent or chromogenic dye. The conditions during the process are chosen in a way that the probe hybridizes the target DNA with a complementary sequence on the membrane. The hybridization is followed by washing with a buffer to remove the probe that is bound nonspecifically or remain unbound so that only labeled probes remain bound to the target sequence.

g. Detection

The hybridized regions on the membrane can be detected via autoradiography by placing the nylon membrane in contact with a photographic film. The images indicate the position of the hybridized DNA molecules, which can be used to determine the length of the fragments by comparing them with the marker DNA molecules of known length.

Similarly, the images also provide information about the number of the hybridizing fragments and their size. If a fluorescent or a chromogenic dye is used, these can be visualized on X-ray film or by the development of color on the membrane.



Result Interpretation of Southern Blot

The results of a Southern blot are observed in the form of bands on the membrane. The size of the DNA fragments can be determined by comparing their relative size with the DNA bands of known lengths.

Applications of Southern Blot

Southern blotting has many applications in the field of gene discovery, mapping, evolution, and diagnostic studies. The technique can be used for DNA analysis to detect point mutations and other structural rearrangements in the DNA sequences.

The method also allows the determination of molecular weights of the restriction fragments, which helps in the analysis of such fragments. Since the technique enables the detection of a particular DNA segment, it can be used in personal identification via fingerprinting. It can be used in disease diagnosis as well as prenatal diagnosis of genetic diseases.

Limitations of Southern Blot

The method is costly as it requires expensive equipment and reagents as compared to other tests.

It is a complex process consisting of multiple steps. The process is also labor intensive that requires trained personnel.

It is a time-consuming process that can be replaced by other faster processes like Polymerase Chain Reaction.

It is a semi-quantitative process that only provides estimated sizing of the DNA fragments.

Southern blotting is not a suitable method for detecting mutations at the base-pair level.

The sample requires a large amount of sample and higher quality of DNA via superior isolation methods.

NORTHERN BLOTTING:

Northern blot is a technique based on the principle of blotting for the analysis of specific RNA in a complex mixture. The technique is a modified version of the Southern Blotting, which was discovered for the analysis of DNA sequences.

The detection of certain sequences of nucleic acids extracted from different types of biological samples is essential in molecular biology, which makes blotting techniques imperative in the field. The principle is identical to southern blotting except for the probes used for the detection as northern blotting detects RNA sequences. This technique provides information about the length of the RNA sequences and the presence of variations in the sequence. Even though the technique is primarily focused on the identification of RNA sequences, it has also been used for the quantification of RNA sequences.

Since the discovery of the technique, several modifications have been made in the technique for the analysis of mRNAs, pre-mRNAs, and short RNAs. Northern blotting was employed as the primary technique for the analysis of RNA fragments for a long time; however, new, more convenient, and cost-effective techniques like RT-PCR have slowly replaced the technique.

Principle of Northern Blot

The principle of the northern blot is the same as all other blotting technique that is based on the transfer of biomolecules from one membrane to another.

The RNA samples are separated on gels according to their size by gel electrophoresis. Since RNAs are single-stranded, these can form secondary structures by intermolecular base pairing. The electrophoretic separation of the RNA segments is thus performed under denaturing conditions. The separated RNA fragments are then transferred to a nylon membrane. Nitrocellulose membrane is not used as RNA doesn't bind effectively to the membrane.

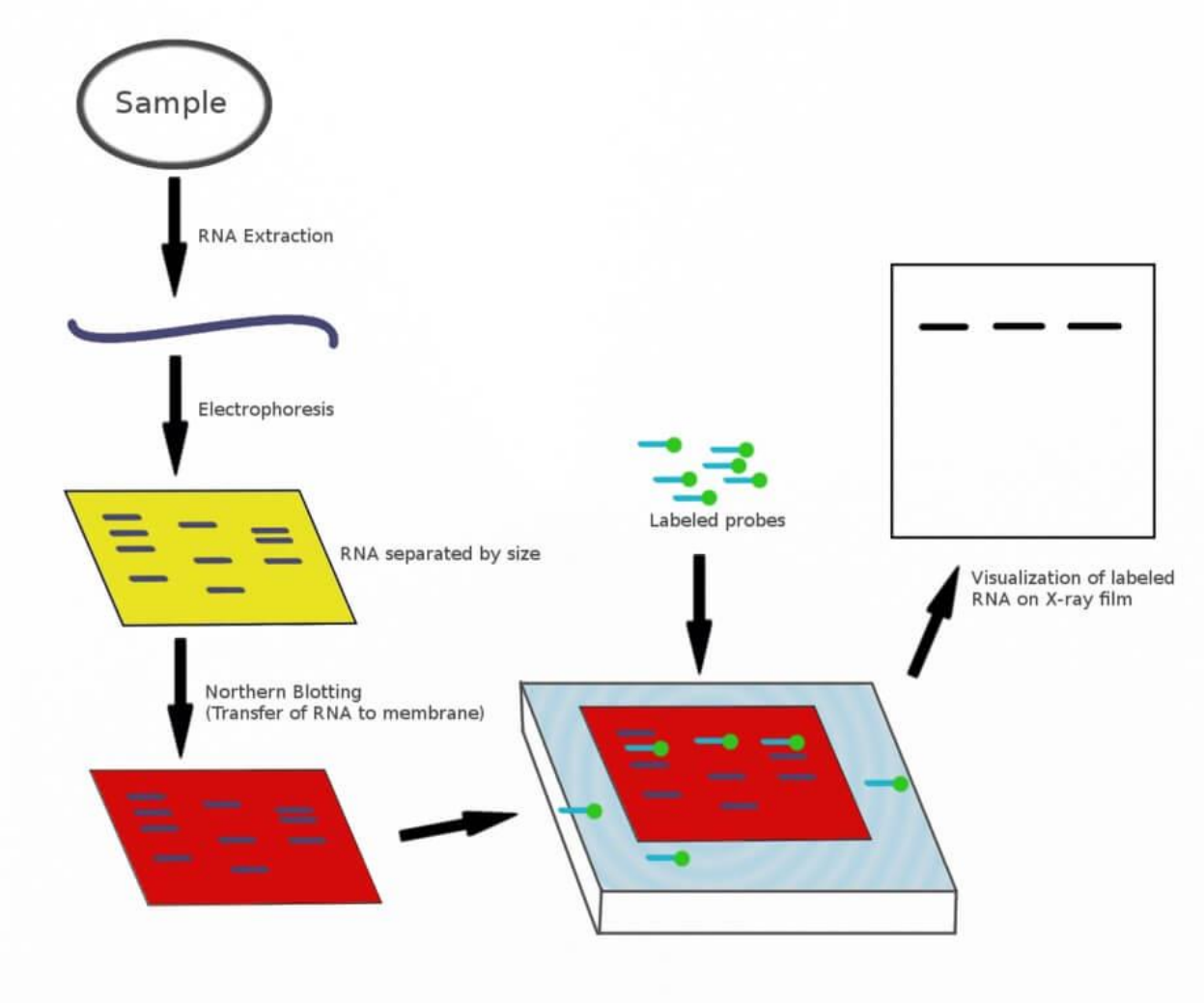
The transferred segments are immobilized onto the membrane by fixing agents. The RNA fragments on the membrane are detected by the addition of a labeled probe complementary to the RNA sequences present on the membrane. The hybridization forms the basis of the detection of RNA as the specificity of hybridization between the probe, and the RNA allows the accurate identification of the segments. Northern blot utilizes size-dependent separation of RNA segments and thus can be used to determine the sizes of the transcripts.

Procedure/Steps of Northern Blot

a. Separation of RNA on a denaturing gel

The RNA gel solution is prepared by adding formaldehyde to the agarose solution. The cast is assembled, and the prepared denaturing gel is poured into the cast. As the gel begins to set, a comb with appropriate teeth is added to form wells. Once the gel is set, the comb is removed, and the gel is equilibrated with a running buffer for 30 minutes before running.

15 μg RNA sample is mixed with an equal volume of RNA loading buffer. Three μg of RNA markers are added in the same volume of RNA loading buffer. The samples are incubated at 65°C on a heating block for about 12-15 minutes. The samples are loaded to the equilibrated gel, and the first row of wells is filled with RNA markers. The gel is then run at 125V for about 3 hours.



b. Transfer of RNA from gel to the nylon membrane

A nylon membrane is cut that is larger than the size of the denaturing gel, and a filter paper with the same size as the nylon membrane is also prepared. Once the electrophoresis process is complete, the RNA gel is removed from the tank and rinsed with water. An oblong sponge that is slightly larger than the gel is placed on a glass dish, and the dish is filled with SSC to a point so as to leave the soaked sponge about half-submerged in the buffer. A few pieces of Whatman 3mm papers are placed on top of the sponge and are wetted with SSC buffer. The gel is then placed on top of the filter paper and squeezed out to remove air bubbles by rolling a glass pipette over the surface. The nylon membrane prepared is wetted with distilled water on an RNase-free dish for about 5 minutes. The wetted membrane is placed on the surface of the gel while avoiding any air bubbles formation.

The surface is further flooded with SSC, and a few more filter papers are placed on top of the membrane. A glass plate is placed on top of the structure in order to hold everything in place. The structure is left overnight to obtain an effective transfer.

c. Immobilization

Once the transfer is complete, the gel is removed and rinsed with SSC, and allowed to dry. The membrane is placed between two pieces of filter paper and baked in a vacuum oven at 80°C for 2 hours. In some cases, the membrane can be wrapped in a UV transparent plastic wrap and irradiates for an appropriate time on a UV transilluminator.

d. Hybridization

The DNA or RNA probes to be used are to be labeled to a specific activity of >108 dpm/μg, and unincorporated nucleotides are to be removed.

The membrane carrying the immobilized RNA is wetted with SSC. The membrane is placed in a hybridization tube with the RNA-side-up, and 1 ml of formaldehyde solution is added.

The tube is placed in the hybridization oven and incubated at 42°C for 3 hours. If the probe used is double-stranded, it is denatured by heating in a water bath or incubator for 10 minutes at 100°C. The desired volume of the probe is pipette into the hybridization tube and further incubated at 42°C. The solution is poured off, and the membrane is washed with a wash solution. The membrane is then observed under autoradiography.

Result Interpretation of Northern Blot

The RNA bands are observed under radiography in the form of bands. The distance of the bands from the markers can be used to determine the length and semi quantification of the RNA fragments.

Applications of Northern Blot

The technique can be used for the identification and separation of RNA fragments collected from different biological sources.

Northern blotting is used as a sensitive test for the detection of transcription of DNA fragments that are to be used as a probe in Southern Blotting.

It also allows the detection and quantification of specific mRNAs from different tissues and different living organisms.

Northern blotting is used as a tool for gene expression studies related to overexpression of cancer-causing genes, and gene expression during transplant rejects.

Northern blotting has been used as a molecular tool for the diagnosis of diseases like Crohn's disease.

The process is used as a method for the detection of viral microRNAs that play important roles in viral infection.

Limitations of Northern Blot

Northern blotting has a lower sensitivity as compared to other modern techniques like RT-PCR and nuclease protection assays.

The method requires a large amount of sample RNA, and these should be of high quality.

The technique is time-consuming and complex, especially in cases where multiple probes are to be added.

WESTERN BLOTTING:

Western blot, also known as immunoblotting, is the process of separating proteins and identifying them in a complex biological sample. The use of polyacrylamide gel electrophoresis is a prerequisite for western blotting in order to separate proteins prior to their identification. The process of western blotting involves the transfer of proteins separated by SDS PAGE into an absorbent membrane. The proteins can then be identified on the membrane by different means. Western blotting has revolutionized the field of immunology with the use of antibody probes against membrane-bound proteins.

The immunodetection of proteins has a wide application in biochemistry and other sciences as it can detect and characterize a multitude of proteins. The sensitivity of the process depends on the efficiency of transfer retention of proteins during processing and the final detection. Western blotting or protein blotting depends on the specificity of interaction between the protein of interest and the probe used for the detection of the protein. Unlike Southern blotting that utilizes radio-labeled nucleic acid probes, western blotting usually uses a second antibody tagged with an enzyme.

Western blotting has a number of advantages over other similar techniques as the process only requires the use of a small amount of reagents, and the same protein transfer can be used for multiple analyses.

Principle of Western Blot

The principle of western blotting is the interaction between the proteins and the probes used for the detection of the proteins. The proteins used for western blotting are separated by gel electrophoresis to obtain them on a gel matrix.

The proteins are then transferred to a nitrocellulose or polyvinylidene fluoride (PVDF) membrane, where they are immobilized. The transfer of the protein is known as blotting. The protein on the membrane can either be detected by the use of a reporter-labeled primary antibody directed against the protein or a reporter-labeled secondary antibody directed at the primary antibody. The reporter or probe present on the antibody can be an enzyme that produces a color reaction or a luminescent signal at the antigen-antibody binding site that produces a fluorescent signal in the presence of a particular substrate. The signal or color generated by the probe requires a detection system that is appropriate for the signal or intensity generated.

Procedure of Western Blot

The process of western blotting consists of the following steps;

1. Sample Preparation

The most commonly used samples for western blot are cell lysates which are collected by the process of extraction. The extraction can be achieved by different means like mechanical destruction, chemical extraction, or the use of enzymes.

The extraction of often performed at cold temperature in the presence of protease inhibitors in order to prevent the denaturation of the proteins.

2. Gel Electrophoresis

The protein sample is diluted with the sample buffer and is heated and shaken for 10 minutes at 70°C. The sample is then centrifuged at 5000g. The gel case is removed from the pouch and is placed in the buffer tank against the rubber seal with the gel walls facing the inside of the tank reservoir. The running buffer is poured onto the upper reservoir while ensuring that no buffer leakage occurs on the lower tank.

Each of the wells is then loaded with an equal volume of heat-denatured sample, and one of the lanes is reserved for the protein ladder. The lid is placed on the tank, and it is connected to the power supply. The run is allowed to run at 200 V constant for 50 minutes.

3. Protein Transfer

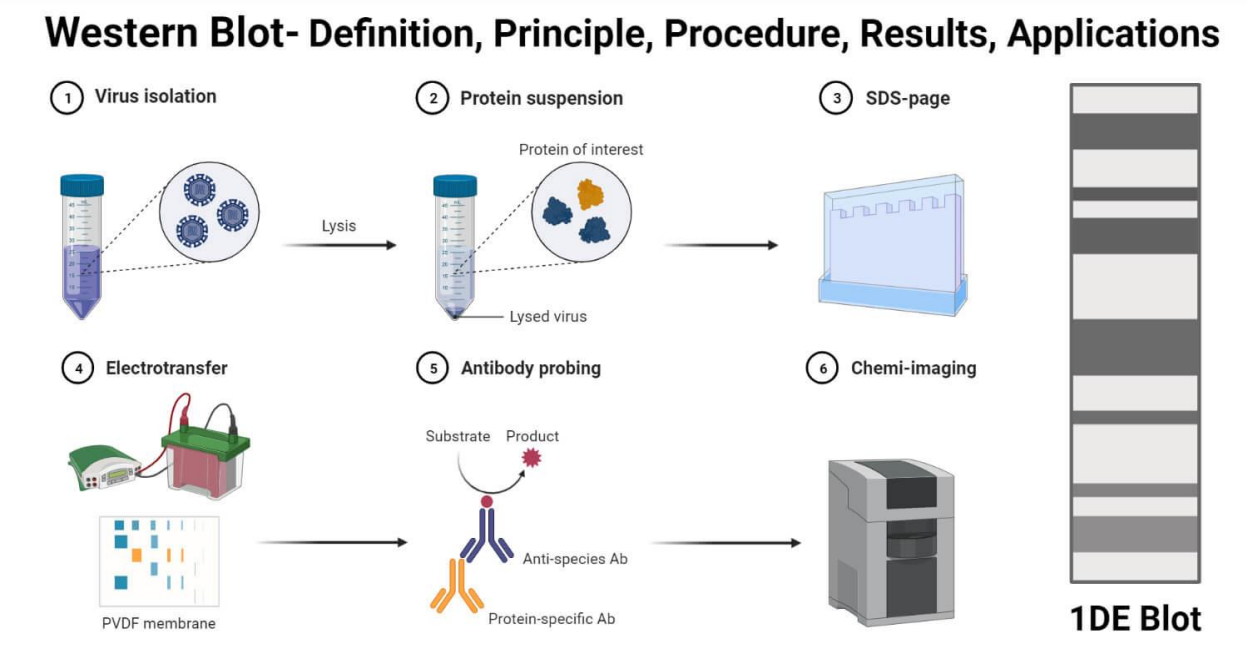
The transfer buffer is prepared by adding 10% methanol to the buffer. The transfer case is taken and laid out. It is then covered with a transfer buffer. A foam sponge is taken and laid on the backside, over which goes the filter paper. These should be placed to ensure that both of them are wet and slightly submerged. The gel is taken out from the tank and placed on the wet filter paper. The nitrocellulose membrane is wet with the transfer buffer and is placed on top of the gel in a way that there are no bubbles between the gel and the membrane.

The transfer case is placed into the transfer tank, which is further filled with transfer buffer. The tank is then connected to power at 100V for 1 hour. Once the transfer is complete, the transfer case is removed, and the nitrocellulose membrane is removed from the gel.

4. Immunodetection

The membrane is washed with Tris-buffered saline for 5 minutes in a Petri dish. The 10% nonfat dry milk is mixed with the Tris buffer, and the membrane is covered with the mixture for 30 minutes at room temperature. The membrane is washed with the Tris buffer to remove any excess mixture remaining on the membrane. With the help of the forceps, the membrane is transferred to a new Petri dish onto which the primary antibody is added.

The membrane with the antibody is incubated for 3 hours at room temperature. The membrane is washed after incubation with the Tris buffer. The membrane is transferred again to a new Petri dish, where a secondary HRP-conjugated antibody is added. The membrane is incubated for 1 hour. The concentration of secondary antibodies often remains at 1 µg/ml, but this also depends on the dilution. The membrane is washed again with Tris buffer to remove excess antibodies from the surface. The membrane is incubated with the substrate for 5 minutes, and the observation is made.



Result Interpretation of Western Blot

The result of western blotting depends on the type of probes used during the process. If an enzyme-conjugated secondary antibody is used, the reaction between the substrate and the enzyme produces a color. The soluble dye is converted into an insoluble form, resulting in a different color on the membrane. In order to stop the development of a blot, the dye is removed by washing the membrane. The protein levels can then be evaluated by spectrophotometry.

Applications of Western Blot

Western blotting is an excellent method with high sensitivity in order to detect a particular protein even in low quantity. Western blotting has been used in the clinical diagnosis of different diseases. The confirmatory test for HIV involves a western blot by detecting anti-HIV antibodies in the serum. The technique has been used to quantify proteins and other gene products in gene expression studies.

Since western blotting detects the proteins by their size and ability to bind to the antibody, it is appropriate for evaluating the protein expressions in cells and further analysis of protein fractions during protein purification. Western blotting is also used for the analysis of different biomarkers like growth factors, cytokines, and hormones.

Limitations of Western Blot

Since it is a very sensitive process, any imbalance in the process can affect the results of the entire process.

In some cases, no bands or erroneous bands might be observed due to the insufficient transfer of the proteins.

The test can only be used as a semi-quantitative test as the estimation is not always precise.

The process is time-consuming and complex, thus can only be performed by well-trained personnel.

Western blotting can only be performed for proteins if the primary antibodies for the proteins are available.

Some antibodies might exhibit off-target effects by interacting with more than one protein in the sample.

The technique is a costly process with the cost of antibodies and expensive detection methods.

Small proteins might not be retained by the membrane, whereas larger proteins are difficult to transfer to the membrane.

Probable Questions:

1. Write down the basic principle of Southern blotting?
2. Describe different steps of Southern Blotting with a suitable diagram.
3. What are the applications of Southern Blotting?
4. Write down the basic principle of Northern blotting?
5. Describe different steps of Northern Blotting with a suitable diagram.
6. What are the applications of Northern Blotting?
7. Write down the basic principle of Western blotting?
8. Describe different steps of Western Blotting with a suitable diagram.
9. What are the applications of Western Blotting?
10. Compare Southern, Northern and Western Blotting techniques.

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

UNIT-XVII

RFLP, RAPD AND AFLP TECHGNIQUES

Objective: In this unit we will discuss about various DNA Fingerprinting methods such as RFLP, RAPD and AFLP.

Introduction:

Different types of molecular markers are used to understand and ascertain relationship in different organisms/individuals as well as to detect or diagnose character. These markers are to locate certain characteristics on the gel (banding pattern) which can be used to detect a specific character/defect in the genome. Unlike genetic mapping, physical mapping is not to locate the genes/characters on a genome, but to create a unique pattern by processing the genomic DNA.

There are several molecular markers available which are used depending upon the objective of the work and facilities available at the centre. Use of these markers to create maps (e.g., electrophoretic patterns) of an organism is known as 'physical mapping'. Molecular markers used in physical mapping are described below. New technologies are also developed simultaneously to resolve biological problems and help legal proceedings.

Restriction Fragment Length Polymorphism (RFLP):

RFLP is a method used by molecular biologists to follow a particular sequence of DNA as it is passed on to other cells. RFLPs can be used in many different settings to accomplish different objectives. RFLPs can be used in paternity cases or criminal cases to determine the source of a DNA sample. RFLPs can be used to determine the disease status of an individual. RFLPs can be used to measure recombination rates which can lead to a genetic map with the distance between RFLP loci measured in centiMorgans.

RFLP, as a molecular marker, is specific to a single clone/restriction enzyme combination. It is a difference in homologous DNA sequences that can be detected by the presence of fragments of different lengths after digestion of the DNA samples in question with specific restriction endonucleases. Most RFLP markers are co-dominant (both alleles in heterozygous sample will be detected) and highly locus-specific.

An RFLP probe is a labelled DNA sequence that hybridizes with one or more fragments of the digested DNA sample after they were separated by gel electrophoresis, thus revealing a unique blotting pattern characteristic to a specific genotype at a specific locus. Short,

single- or low-copy genomic DNA or cDNA clones are typically used as RFLP probes. The RFLP probes are frequently used in genome mapping and in variation analysis (genotyping, forensics, paternity tests, hereditary disease diagnostics, etc.) (Fig. 21.5).

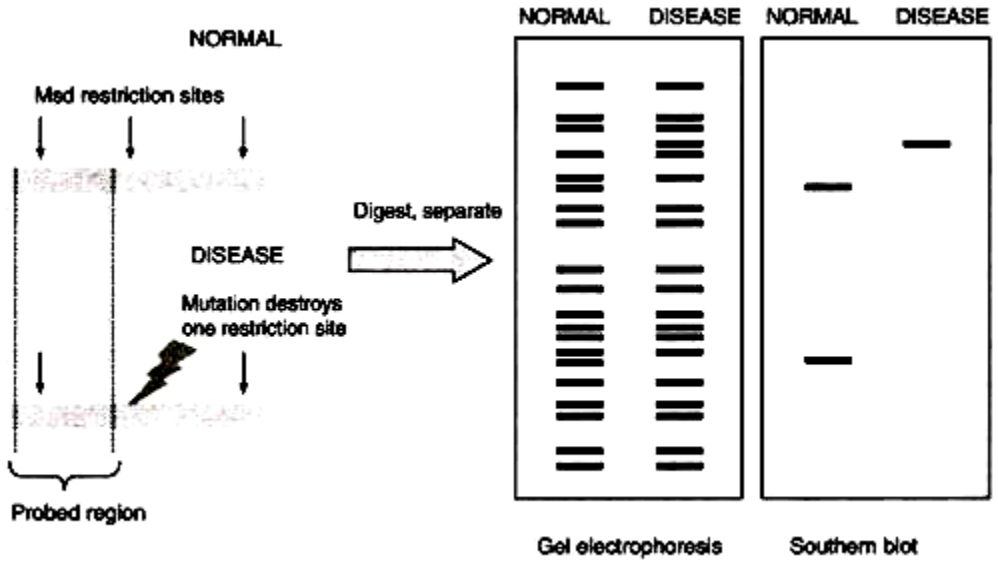


Fig. 21.5. Disease diagnosis by RFLP.

i. Procedure:

Usually, DNA from an individual specimen is first extracted and purified. Purified DNA may be amplified by polymerase chain reaction (PCR), The DNA is then cut into restriction fragments using suitable endonucleases, which only cut the DNA molecule where there are specific DNA sequences, termed recognition sequence or restriction sites that are recognized by the enzymes.

These sequences are specific to each enzyme, and may be either four, six, eight, ten or twelve base pairs in length. The more base pairs there are in the restriction site, the more specific it is and the lower the probability that it will find a place to be cut. The restriction fragments are then separated according to length by agarose gel electrophoresis. The resulting gel may be enhanced by Southern blotting. Alternatively, fragments may be visualized by pre-treatment or post-treatment of the agarose gel, using methods such as ethidium bromide staining or silver staining respectively.

RFLPs have provided valuable information in many areas of biology, including: screening human DNA for the presence of potentially deleterious genes (Fig. 21.6). Providing evidence to establish the innocence of or a probability of the guilt of, a crime suspect by DNA “fingerprinting”. The distance between the locations cut by restriction enzymes (the restriction sites) varies between individuals, due to insertions, deletions or trans-versions.

This causes the length of the fragments to vary, and the position of certain amplicons differs between individuals (thus polymorphism). This can be used to genetically tell

individuals apart. It can also show the genetic relationship between individuals, because children inherit genetic elements from their parents. Mitochondrial DNA RFLP analyses can lead to the determination of maternal relationships.

Fragments may also be used to determine relationships among and between species by comparison of the resulting haplotypes (abridged for 'haploid genotype'). RFLP is a technique used in marker assisted selection. Terminal Restriction Fragment Length Polymorphism (TRFLP or sometimes T-RFLP) is a molecular biology technique initially developed for characterizing bacterial communities in mixed-species samples. The technique has also been applied to other groups including soil fungi.

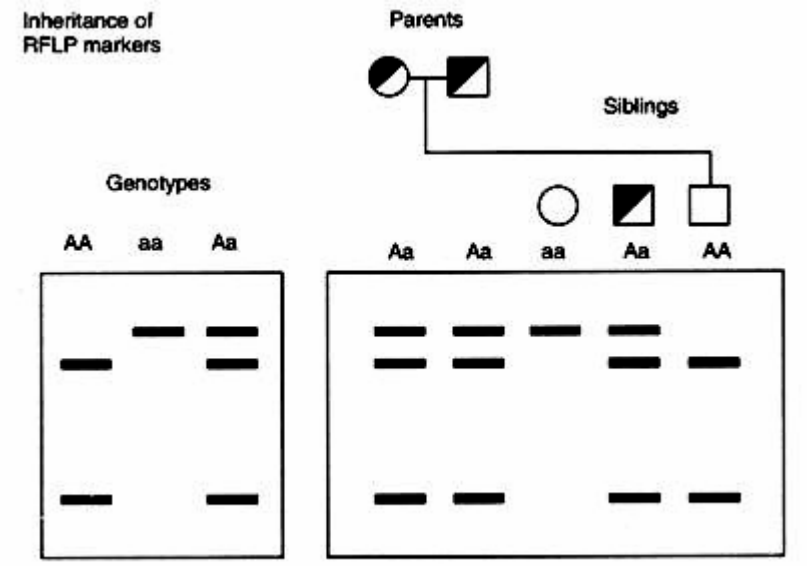


Fig. 21.6. Inheritance of RFLP markers.

The technique works by PCR amplification of DNA using primer pairs that have been labelled with fluorescent tags. The PCR products are then digested using RFLP enzymes and the resulting patterns visualized using a DNA sequencer. The results are analyzed either by simply counting and comparing bands or peaks in the TRFLP profile, or by matching bands from one or more TRFLP runs to a database of known species.

ii. Measurement of distance between two RFLP loci:

To calculate the genetic distance between two loci, you need to be able to observe recombination. Traditionally, this was performed by observing phenotypes but with RFLP analysis, it is possible to measure the genetic distance between two RFLP loci whether they are a part of genes or not. Let's look at a simple example in fruit flies. Two RFLP loci with two RFLP bands possible at each locus (Fig. 21.7).

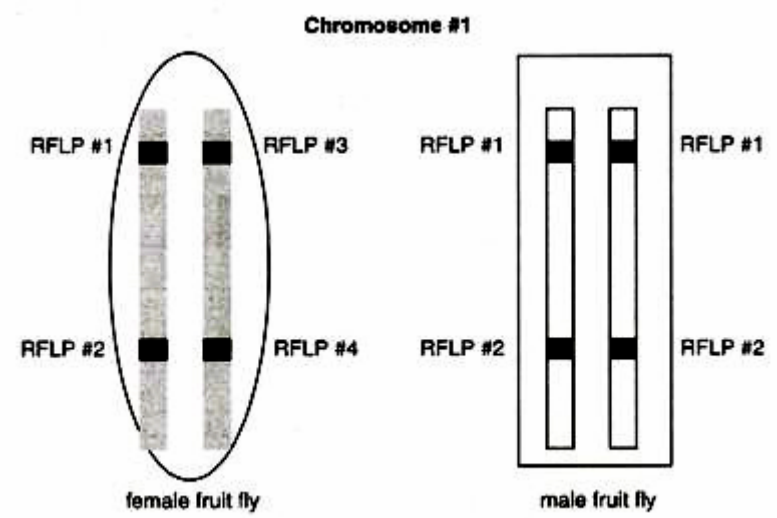


Fig. 21.7. RFLP loci on male and female chromosome.

These loci are located on the same chromosome for the female (left) and the male (right). The upper locus can produce two different bands called 1 and 3. The lower locus can produce bands called 2 or 4. The male is homozygous for band 1 at the upper locus and 2 for the lower locus. The female is heterozygous at both loci. Their RFLP banding patterns can be seen on the Southern blot below (Fig. 21.8).

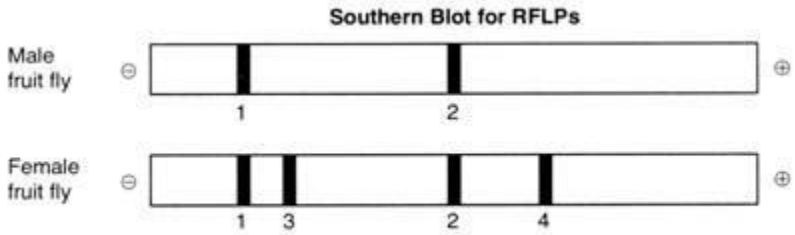


Fig. 21.8. RFLP loci on gel of male and female chromosome.

The male can only produce one type of gamete (1 and 2) but the female can produce four different gametes. Two of the possible four are called parental because they carry both RFLP bands from the same chromosome; 1 and 2 from the left chromosome or 3 and 4 from the right chromosome. The other two chromosomes are recombinant because recombination has occurred between the two loci and thus the RFLP bands are mixed so that 1 is now linked to 4 and 3 is linked to 2.

Type of chromatid	Alleles
Parental	RFLP 1 and 2
Parental	RFLP 3 and 4
Recombinant	RFLP 1 and 4
Recombinant	RFLP 3 and 2

When these two flies mate, the frequency of the four possible progeny can be measured and from this information, the genetic distance between the two RFLP loci (upper and lower) can be determined (Fig. 21.9).

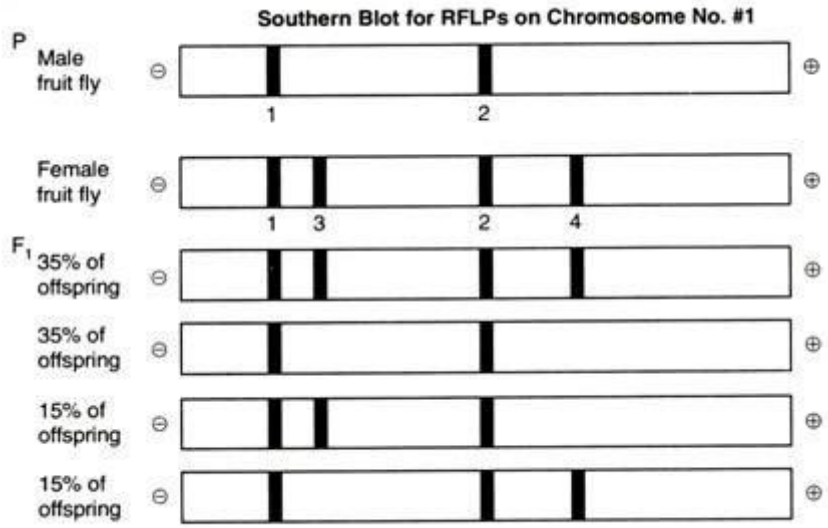


Fig. 21.9. Southern blot for RFLP's on chromosome no. 1.

In this example, 70% of the progeny were produce from parental genotype eggs and 30% were produced by recombinant genotype eggs. Therefore, these two RFLP loci are 30 centiMorgans apart from each other.

iii. PCR-RFLP:

Isolation of sufficient DNA for RFLP analysis is time consuming and labour intensive. However, PCR can be used to amplify very small amounts of DNA, usually in 2-3 hours, to the levels required for RFLP analysis. Therefore, more samples can be analysed in a shorter time. An alternative name for the technique is Cleaved Amplified Polymorphic Sequence (CAPS) assay.

iv. Limitations:

RFLP is a multistep procedure involving restriction enzymatic cleavage, electrophoresis, southern blotting and detection of specific sequences. It is a time consuming process.

Random Amplified Polymorphic DNA (RAPD):

This technique can be used to determine taxonomic identity, assess kinship relationships, detect inter-specific gene flow, analyse hybrid speciation, and create specific probes. Advantages of RAPDs include suitability for work on anonymous genomes, applicability to work where limited DNA is available, efficiency and low expense. It is also useful in distinguishing individuals, cultivars or accessions. RAPDs also have applications in the identification of asexually reproduced plant varieties for forensic or agricultural purposes, as well as ecological ones.In RAPD by using different primers, molecular characters can be generated that are diagnostic at different taxonomic levels. This is really a stripped-down version of PCR but uses a single sequence in the design of the primer (i.e., two primers are still needed for PCR: the same primer is used at either end).

The primer may be designed specifically, but could be chosen randomly and is used to amplify a series of samples which will include both the material of interest as well as other control samples with which the experimental material needs to be compared. Choice of primer length will be critical to the determination of band complexity in the resulting amplification pattern. Eventually a particular probe will be found that is able to distinguish between the sample of interest and those that are different.

i. Procedure:

Unlike traditional PCR analysis, RAPD (pronounced 'rapid') does not require any specific knowledge of the DNA sequence of the target organism: the identical 10-mer primers will or will not amplify a segment of DNA, depending on positions that are complementary to the primers' sequence. For example, no fragment is produced if primers annealed too far apart or 3′ ends of the primers are not facing each other.

Therefore, if a mutation has occurred in the template DNA at the site that was previously complementary to the primer, a PCR product will not be produced, resulting in a different pattern of amplified DNA segments on the gel (Fig. 21.10). RAPD is an inexpensive yet powerful typing method for many bacterial species

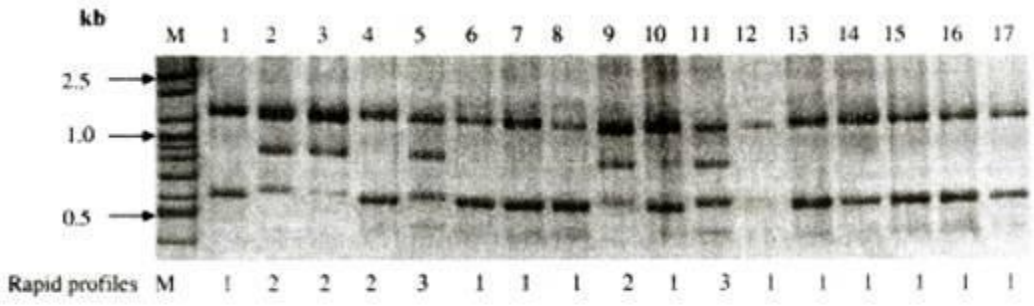


Fig. 21.10. Silver-stained polyacrylamide gel showing three distinct RAPD profiles generated by primer OPE15 for *Haemophilus ducreyi* isolates from different countries. Selecting the right sequence for the primer is very important because different sequences will produce different band patterns and possibly allow for a more specific recognition of individual strains.

RAPD amplification products can be either variable (polymorphic) or constant (non-polymorphic). In a RAPD analysis of several individuals within a species, and species within a genus, constant fragments diagnostic for a genus may be identified, as well as fragments which are polymorphic between species of the genus. RAPDs can be applied to analyse fusion of genotypes at different taxonomic levels. At the level of the individual, RAPD markers can be applied to parentage analysis, while at the population level, RAPD can detect hybrid populations, species or subspecies.

The detection of genotype hybrids relies on the identification of diagnostic RAPD markers for the parental genotypes under investigation. However RAPD markers tend to underestimate genetic distances between distantly related individuals, for example in inter-specific comparisons. It is wise to be cautious when using RAPD for taxonomic studies above the species level. Conventional RFLP techniques are ill-suited for the analysis of paternity and estimation of reproductive success in species with large offspring clutches, because of the need to determine paternity for each individual offspring. RAPD fingerprinting provides a ready alternative for such cases. Synthetic offspring may be produced by mixing equal amounts of the DNA of the mother and the potential father. The amplification products from the synthetic offspring should ideally contain the full complement of bands that appear in any single offspring of these parents.

ii. Limitations of RAPD:

1. Nearly all RAPD markers are dominant, i.e., it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies). Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely.

2. PCR is an enzymatic reaction, therefore the quality and concentration of template DNA, concentrations of PCR components, and the PCR cycling conditions may greatly influence

the outcome. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed laboratory protocols to be reproducible.
3. Mismatches between the primer and the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret.

Amplification Fragment Length Polymorphism (AFLP):

Amplified Fragment Length Polymorphism (AFLP) is a polymerase chain reaction (PCR) based genetic fingerprinting technique that was developed in the early 1990's by Keygene. AFLP can be used in the fingerprinting of genomic DNA of varying origins and complexities. The amplification reaction is rigorous, versatile and robust, and appears to be quantitative.

While AFLP is capable of producing very complex fingerprints (100 bands where RAPD produces 20), it is a technique that requires DNA of reasonable quality and is more experimentally demanding. AFLP uses restriction enzymes to cut genomic DNA, followed by ligation of complementary double stranded adaptors to the ends of the restriction fragments.

A subset of the restriction fragments are then amplified using 2 primers complementary to the adaptor and restriction site fragments. The fragments are visualized on denaturing polyacrylamide gels either through auto-radiographic or fluorescence methodologies.

i. Procedure:

AFLP-PCR is a highly sensitive method for detecting polymorphisms in DNA. The technique was originally described by Vos and Zabeau in 1993. The procedure of this technique is divided into three steps (Fig. 21.11):

1. Digestion of total cellular DNA with one or more restriction enzymes that cuts frequently (Msel, 4 bp recognition sequences) and one that cuts less frequently (EcoRI, 6 bp recognition sequence). The resulting fragments are ligated to end-specific adaptor molecules.
2. Selective amplification of some of these fragments with two PCR primers that have corresponding adaptor and restriction site specific sequences.
3. Electrophoretic separation of amplicons on a gel matrix, followed by visualisation of the band pattern.

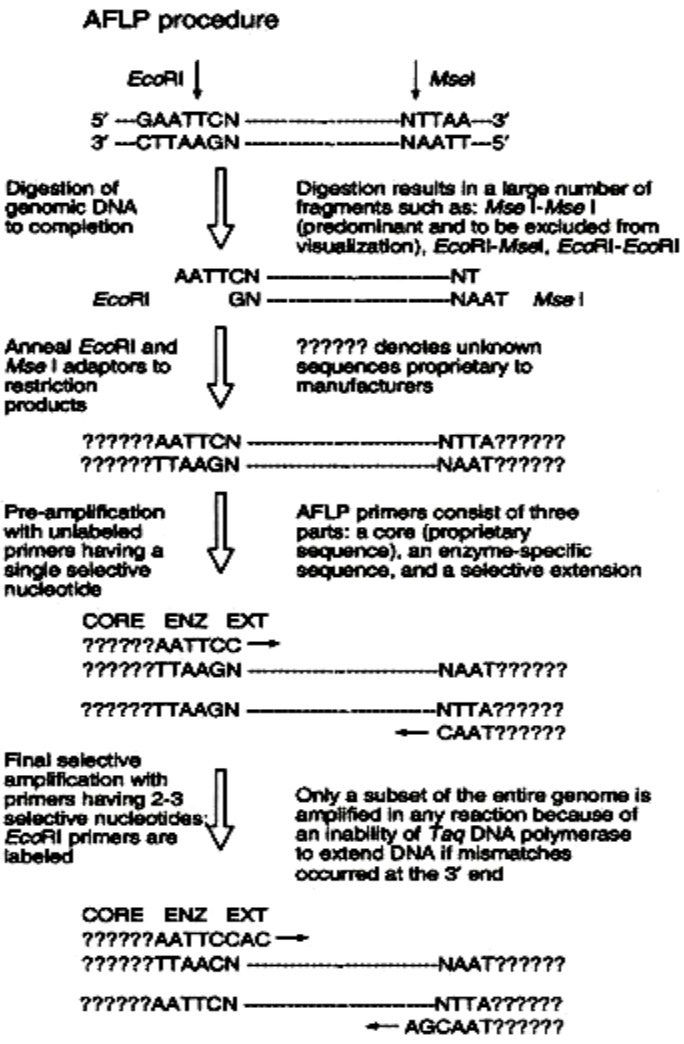


Fig. 21.11. AFLP procedure.

In a second, "selective", PCR, using the products of the first as template, primers containing two further additional bases, chosen by the user, are used. The EcoRI-adaptor specific primer used bears a label (fluorescent or radioactive). Gel electro-phoretic analysis reveals a pattern (fingerprint) of fragments representing about 1/4000th of the EcoRl-Msel fragments.

AFLP's, can be co-dominant markers, like RFLP's. Co-dominance results when the polymorphism is due to sequences within the amplified region. Yet, because of the number of bands seen at one time, additional evidence is needed to establish that a set of bands result from different alleles at the same locus.

If, however, the polymorphism is due to presence/absence of a priming site, the relationship is dominance. The non-priming allele will not be detected as a band. Compared to RAPD, fewer primers should be needed to screen all possible sites. AFLP can be used for mapping, fingerprinting and genetic distance calculation between genotypes. The

advantage of AFLP is its high multiplexity and therefore the possibility of generating high marker densities.
One limitation of the AFLP technique is that fingerprints may share few common fragments when genome sequence homology is less than 90%. Therefore, AFLP cannot be used in comparative genomic analysis with hybridization-based probes or when comparing genomes that are evolving rapidly such as those of some microbes. Conversely, very homogeneous genomes may not be suitable for AFLP analysis.

ii. Limitations of AFLP:

1. Proprietary technology is needed to score heterozygotes and homozygotes. Otherwise, AFLP must be dominantly scored.
2. Developing locus-specific markers from individual fragments can be difficult.
3. Need to use different kits adapted to the size of the genome being analysed.

Probable Questions:

1. What is RFLP ? Describe the procedure.
2. What is RAPD ? Describe the procedure.
3. How RFLP can be used in genetic mapping?
4. How RAPD can be used in genetic mapping?
5. How distance between two RFLP loci can be determined?
6. What are the limitations of RAPD ?
7. Describe procedure of AFLP .
8. What are the limitations of AFLP?

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal

UNIT-XVIII

PESTICIDE FORMULATION

Objective: In this unit we will discuss about different methods of pesticide formulation and their uses.

Introduction: Pesticide chemicals in their "raw" or unformulated state are not usually suitable for pest control. These concentrated chemicals and active ingredients may not mix well with water, may be chemically unstable, and may be difficult to handle and transport. For these reasons, inert substances, such as clays and solvents, are added to improve application effectiveness, safety, handling, and storage. Inert ingredients do not possess pesticidal activity and are added to serve as a carrier for the active ingredient. The percentage of inert ingredients is listed in the formulation or designates them as "other ingredients" on the labels of the formulation. There are several inert substances, such as petroleum distillates and xylene, which have a specific statement identifying their presence in the formulation.

Glossary related to pesticide formulation:

Abrasive: Capable of wearing away or grinding down another object.

Active ingredient (a.i.): The substance in a pesticide product that is intended to kill, repel, or otherwise control a target pest.

Adjuvant: Chemical that is either premixed in the pesticide formulation or added to the spray tank to improve mixing, application or to enhance pesticidal activity.

Carrier: The primary material used to allow a pesticide to be dispersed effectively. For example, talc in a dust formulation.

Diluent: Anything used to dilute a pesticide.

Emulsifier: Agent that helps to prevent an emulsion from separating.

Emulsion: A mixture of two or more liquids that is not soluble in one another. For example, oil droplets dispersed in water.

Formulation: A mixture of active ingredient combined during manufacture with inert ingredients.

Inert ingredients: All materials in the pesticide formulation other than the active ingredient. They are added to dilute the pesticide or to make it safer, more effective, and easier to measure, mix, apply, and handle. Some inert ingredients may be toxic or hazardous to people.

Insoluble: Does not dissolve in liquid.

Phytotoxicity: Injury to plants.

Soluble: Able to dissolve in another substance, usually a liquid.

Solvent: A liquid, such as water, kerosene, xylene, or alcohol, that will dissolve a pesticide to form a solution.

Suspension: A substance that contains undissolved particles mixed throughout a liquid.

Volatile: Evaporating rapidly; turning easily into a gas or vapor.

The mixture of active and inert ingredients is called a pesticide formulation.

This formulation may consist of:

- The pesticide active ingredient that controls the target pest.
- The carrier, such as an organic solvent or mineral clay.
- Adjuvants, such as stickers and spreaders.
- Other ingredients, such as stabilizers, dyes and chemicals that improve or enhance pesticidal activity.

Usually a formulated product is mixed with water or oil for final application. Most baits, granules, gels, and dusts, however, are ready for use without additional dilution.

A single active ingredient often is sold in several kinds of formulations. Abbreviations are frequently used to describe the formulation (e.g., WP for wettable powders); how the pesticide is used (e.g., TC for termiticide concentrate); or the characteristics of the formulation (e.g., ULV for an ultra-low-volume formulation). For example, an 80% SP contains 80 percent by weight of active ingredient and is a “soluble powder.” If it is in a 10-pound bag, it contains 8 pounds of a.i. and 2 pounds of inert ingredient. Liquid formulations indicate the amount of a.i. in pounds per gallon. For example, 1E means 1 pound, and 4E

means 4 pounds of the a.i. per gallon in an emulsifiable concentrate formulation. If more than one formulation is available for pest control, the best one is chosen for the job.

Types of formulations:

1) Liquid Formulations

Liquid formulations are generally mixed with water, but in some instances crop oil, diesel fuel, kerosene, or some other light oil are also mixed as a carrier. Common liquid pesticide formulations are –

a) Emulsifiable Concentrates (EC or E)

An emulsifiable concentrate formulation usually contains a liquid active ingredient, one or more petroleum-based solvents (which give EC formulations their strong odor), and an agent—known as an emulsifier—that allows the formulation to be mixed with water to form an emulsion. Upon mixing with water, they take on a "milky" appearance Most ECs contain between 25% and 75% (2–8 pounds) active ingredient per gallon. ECs are among the most versatile formulations. They are used against agricultural, ornamental and turf, forestry, structural, food processing, livestock, and public health pests. They are adaptable to many types of application equipment including portable sprayers, hydraulic sprayers, low-volume ground sprayers, mist blowers, and low-volume aircraft sprayers.

Advantages of emulsifiable concentrates:

- Relatively easy to handle, transport, and store
- Little agitation required; will not settle out or separate when equipment is running
- Not abrasive
- Will not plug screens or nozzles
- Little visible residue on treated surfaces

Disadvantages of emulsifiable concentrates:

- High a.i. concentration makes it easy to overdose or underdose through mixing or calibration errors.
- Easily absorbed through skin of humans or animals.
- Solvents may cause rubber or plastic hoses, gaskets, and pump parts and surfaces to deteriorate.
- May cause pitting or discoloration of painted finishes.
- Flammable—should be used and stored away from heat or open flame.
- May be corrosive.

b) Solutions (S) :

Some pesticide active ingredients dissolve readily in a liquid carrier, such as water or a petroleum-based solvent. When mixed with the carrier, they form a solution that does not settle out or separate. Formulations of these pesticides usually contain the active ingredient, the carrier, and one or more other ingredients.

c) Ready-to-Use Low Concentration Solutions (RTU) :

Low-concentrate RTU formulations are ready to use and require no further dilution before application (Figure 6). They consist of a small amount of active ingredient (often 1% or less per unit volume) dissolved in an organic solvent. They usually do not stain fabrics nor have unpleasant odours. They are especially useful for structural and institutional pests and for household use. Major disadvantages of low-concentrate formulations include limited availability and high cost per unit of active ingredient.

d) Ultra-Low Volume (ULV) :

These concentrates may approach 100% active ingredient. They are designed to be used "as is" or to be diluted with only small quantities of a specified carrier. They are used at rates of no more than 1/2 gallon per acre. These special purpose formulations are used mostly in outdoor applications, such as in agricultural, forestry, ornamental, and mosquito control programs.

Advantages of ultra-low-volume formulations:

- Relatively easy to transport and store Remain in solution; little agitation required.
- Not abrasive to equipment.
- Will not plug screens and nozzles.
- Leave little visible residue on treated surfaces.

Disadvantages:

- Difficult to keep pesticide on target—high drift hazard.
- Specialized equipment required.
- Easily absorbed through skin of humans or animals.
- Solvents may cause rubber or plastic hoses, gaskets, and pump parts and surfaces to deteriorate.

• Calibration and application must be done very carefully because of the high concentration of active ingredient

e) Invert Emulsions :

An invert emulsion contains a water-soluble pesticide dispersed in an oil carrier. Invert emulsions require a special kind of emulsifier that allows the pesticide to be mixed with a large volume of petroleum-based carrier, usually fuel oil. Invert emulsions aid in reducing drift. With other formulations, some spray drift results when water droplets begin to evaporate before reaching target surfaces; as a result, the droplets become very small and light. Because oil evaporates more slowly than water, invert emulsion droplets shrink less; therefore, more pesticide reaches the target. The oil helps to reduce runoff and improves rain resistance. It also serves as a sticker-spreader by improving surface coverage and absorption. Because droplets are relatively large and heavy, it is difficult to get thorough coverage on the undersides of foliage. Invert emulsions are most commonly used along rights-of-way where drift to susceptible nontarget plants or sensitive areas can be a problem.

2. Flowables (F)/Liquids (L)

A flowable or liquid formulation combines many of the characteristics of emulsifiable concentrates and wettable powders. Manufacturers use these formulations when the active ingredient is a solid that does not dissolve in either water or oil. The active ingredient, impregnated on a substance such as clay, is ground to a very fine powder. The powder is then suspended in a small amount of liquid. The resulting liquid product is quite thick (Figure 8). Flowables and liquids share many of the features of emulsifiable concentrates, and they have similar disadvantages. They require moderate agitation to keep them in suspension and leave visible residues similar to those of wettable powders. Flowables/liquids are easy to handle and apply. Because they are liquids, they are subject to spilling and splashing. They contain solid particles, so they contribute to abrasive wear of nozzles and pumps. Flowable and liquid suspensions settle out in their containers. Always shake them thoroughly before pouring and mixing. Because flowable and liquid formulations tend to settle, manufacturers package them in containers of 5 gallons or less to make remixing easier.

3. Aerosols (A)

These formulations contain one or more active ingredients and a solvent. Most aerosols contain a low percentage of active ingredients. There are two types of aerosol formulations: the ready-to-use type commonly available in pressurized, sealed containers and those products used in electric- or gasoline-powered aerosol generators that release the formulation as a “smoke” or “fog.” Ready-to-use aerosols are usually small, self-contained units that release the pesticide when the nozzle valve is triggered. The pesticide is driven through a fine opening by an inert gas under pressure, creating fine droplets. These

products are used in greenhouses, in small areas inside buildings, or in localized outdoor areas. Commercial models, which hold 5–10 pounds of pesticide, are usually refillable.

Advantages :

- Ready to use
- Portable
- Easily stored
- Convenient way to buy a small amount of a pesticide
- Retain potency over fairly long time

Disadvantages:

- Practical for only very limited uses
- Risk of inhalation injury
- Hazardous if punctured, overheated, or used near an open flame
- Difficult to confine to target site or pest

Formulations for smoke or fog generators are aerosol formulations but not under pressure. They are used in machines that break the liquid formulation into a fine mist or fog (aerosol) using a rapidly whirling disk or heated surface. These formulations are used mainly for insect control in structures such as greenhouses and warehouses and for mosquito and biting fly control outdoors.

4. Liquid Baits

An increasing number of insecticides and rodenticides are being formulated as liquid baits. Liquid rodenticides are mixed with water and placed in bait stations designed for these products. They have two major benefits. Liquid rodenticides are effective in controlling rodents, especially rats, in areas where they cannot find water. They are also effective in areas of poor sanitation where readily available food renders traditional baits ineffective.

Liquid insecticide baits are used primarily by the structural pest control industry for controlling ants and, to a lesser extent, cockroaches. They are packaged as ready-to-use, sugar-based liquids placed inside bait stations. Liquid insecticide ant baits have a number of advantages. They are very effective against certain species of sugar-feeding ants. These ants typically accept and transfer liquid baits into the ant colonies. However, some ants will not feed on liquid baits. Liquid baits also must be replaced often.

5. Dry or Solid Formulations

Dry formulations can be divided into two types: ready-to-use and concentrates that must be mixed with water to be applied as a spray. This section will present more detailed information about the common dry or solid pesticide formulations.

6. Dusts (D)

Most dust formulations are ready to use and contain a low percentage of active ingredients (usually 10% or less by weight), plus a very fine, dry inert carrier made from talc, chalk, clay, nut hulls, or volcanic ash. The size of individual dust particles varies. A few dust formulations are concentrates and contain a high percentage of active ingredients. These concentrates are mixed with dry inert carriers before applying. Dusts are always used dry and can easily drift to non target sites. They are widely used as seed treatments and sometimes for agricultural applications. In structures, dust formulations are used in cracks and crevices and for spot treatments to control insects such as cockroaches. Insects ingest poisonous dusts during grooming or absorb the dusts through their outer body covering. Dusts also are used to control lice, fleas, and other parasites on pets and livestock.

Advantages of dust formulations :

- Most are ready to use, with no mixing .
- Effective where moisture from a spray might cause damage.
- Require simple equipment.
- Effective in hard-to-reach indoor areas.

Disadvantages:

- Easily drift off target during application.
- Residue easily moved off target by air movement or water.
- May irritate eyes, nose, throat, and skin.
- Will not stick to surfaces as well as liquids.
- Dampness can cause clogging and lumping.
- Difficult to get an even distribution of particles on surfaces.

Special dusts, known as tracking powders, are used for monitoring and controlling rodents and insects. For rodent control, the tracking powder consists of finely ground dust combined with a stomach poison. Rodents walk through the dust, pick it up on their feet and fur, and ingest it when they clean themselves. Tracking powders are useful when bait acceptance is poor because of an abundant, readily available food supply. Nontoxic powders, such as talc or flour, often are used to monitor and track the activity of rodents in buildings.

7. Baits (B)

A bait formulation is an active ingredient mixed with food or another attractive substance. The bait either attracts the pests or is placed where the pests will find it. Federal regulations require that certain rodenticide baits must be contained in tamper-resistant bait stations. Pests are killed by eating the bait that contains the pesticide. The amount of active ingredient in most bait formulations is quite low, usually less than 5%.

Baits are used inside buildings to control ants, roaches, flies, other insects, and rodents. Outdoors they sometimes are used to control snails, slugs, and insects such as ants and termites. Their main use is for control of vertebrate pests such as rodents, other mammals, and birds.

Advantages of baits :

- Entire area need not be covered because pest goes to bait
- Control pests that move in and out of an area

Disadvantages:

- Can be attractive to children and pets
- May kill domestic animals and nontarget wildlife outdoors
- Pest may prefer the crop or other food to the bait
- Dead vertebrate pests may cause odor problem
- Other animals may be poisoned as a result of feeding on the poisoned pests
- If baits are not removed when the pesticide becomes ineffective, they may serve as a food supply for the target pest or other pests
- Laws require that outdoor, above-ground placement of certain rodenticide bait products be contained in tamper-resistant bait stations

Pastes and gels are mainly used in the pest control industry for ants and cockroaches. Insecticides formulated as pastes and gels are now the primary formulations used in cock-roach control. They are designed to be injected or placed as either a bead or dot inside small cracks and crevices of building elements where insects tend to hide or travel. Two basic types of tools are used to apply pastes and gels: syringes and bait guns. The applicator forces the bait out of the tip of the device by applying pressure to a plunger or trigger.

8. Granules (G)

Granular formulations are similar to dust formulations except granular particles are larger and heavier. The coarse particles are made from materials such as clay, corncobs, or walnut shells. The active ingredient either coats the outside of the granules or is absorbed into them. The amount of active ingredient is relatively low, usually ranging from less than 1 to 15 percent by weight.

Granular pesticides are most often used to apply chemicals to the soil to control weeds, fire ants, nematodes, and insects living in the soil or for absorption into plants through the roots. Granular formulations are sometimes applied by airplane or helicopter to minimize drift or to penetrate dense vegetation. Once applied, granules release the active ingredient slowly. Some granules require soil moisture to release the active ingredient. Granular formulations also are used to control larval mosquitoes and other aquatic pests. Granules are used in agricultural, structural, ornamental, turf, aquatic, right-of-way, and public health (biting insect) pest control operations.

Advantages of granular formulations:

- Ready to use, no mixing
- Drift hazard is low, and particles settle quickly
- Little hazard to applicator; no spray, little dust
- Weight carries the formulation through foliage to soil or water target
- Simple application equipment needed, such as seeders or fertilizer spreaders
- May break down more slowly than WPs or ECs because of a slow-release coating

Disadvantages:

- Often difficult to calibrate equipment and apply uniformly
- Will not stick to foliage or other uneven surfaces
- May need to be incorporated into soil or planting medium
- May need moisture to activate pesticide
- May be hazardous to non target species, especially waterfowl and other birds that mistakenly feed on the seed-like granules
- May not be effective under drought conditions because the active ingredient is not released in sufficient quantity to control the pest.

9. Pellets (P or PS)

Most pellet formulations are very similar to granular formulations; the terms often are used interchangeably. In a pellet formulation, however, all the particles are the same weight and shape. The uniformity of the particles allows use with precision application equipment. A few fumigants are formulated as pellets; some may be referred to as tablets.

However, these are clearly labelled as fumigants. Do not confuse them with nonfumigant pellets.

10. Wettable Powders (WP or W)

Wettable powders are dry, finely ground formulations that look like dusts. They usually must be mixed with water for application as a spray. A few products, however, may be applied either as a dust or as a wettable powder; the choice is left to the applicator. Wettable powders contain 5%–95% active ingredient by weight, usually 50% or more. The particles do not dissolve in water. They settle out quickly unless constantly agitated to keep them suspended. Wettable powders are one of the most widely used pesticide formulations. They can be used for most pest problems and in most types of spray equipment where agitation is possible. Wettable powders have excellent residual activity. Because of their physical properties, most of the pesticide remains on the surface of treated porous materials such as concrete, plaster, and untreated wood. In such cases, only the water penetrates the material.

Advantages of wettable powders:

- Easy to store, transport, and handle
- Less likely than ECs and other petroleum-based pesticides to cause unwanted harm to treated plants, animals, and surfaces
- Easily measured and mixed
- Less skin and eye absorption than ECs and other liquid formulations

Disadvantages:

- Inhalation hazard to applicator while measuring and mixing the concentrated powder
- Require good and constant agitation (usually mechanical) in the spray tank or will quickly settle out if the agitator is turned off
- Abrasive to many pumps and nozzles, causing them to wear out quickly

11. Soluble Powders (SP or WSP)

Soluble powder formulations look like wettable powders. However, when mixed with water, soluble powders dissolve readily and form a true solution. After they are mixed thoroughly, no additional agitation is necessary. The amount of active ingredient in soluble powders ranges from 15% to 95% by weight; it usually is more than 50%. Soluble powders have all the advantages of wettable powders and none of the disadvantages except the inhalation hazard during mixing. Few pesticides are available in this formulation because few active ingredients are readily soluble in water.

12. Water-Dispersible Granules (WDG) or Dry Flowables (DF)

Water-dispersible granules, also known as dry flowables, are like wettable powders except instead of being dustlike, they are formulated as small, easily measured granules. Water-dispersible granules must be mixed with water to be applied. Once in water, the granules break apart into fine particles similar to wettable powders. The formulation requires constant agitation to keep them suspended in water. The percentage of active ingredient is high, often as much as 90 percent by weight. Water-dispersible granules share many of the same advantages and disadvantages of wettable powders except:

- They are more easily measured and mixed
- Because of low dust, they cause less inhalation hazard to the applicator during handling

13. Other Formulations

Other formulations include chemicals that cannot be clearly classified as liquid or as dry/solid pesticide formulations. These are -

a. Microencapsulated Materials (M or ME)

Manufacturers cover liquid or dry pesticide particles in a plastic coating to produce a microencapsulated formulation. Microencapsulated pesticides are mixed with water and sprayed in the same manner as other sprayable formulations. After spraying, the plastic coating breaks down and slowly releases the active ingredient. Microencapsulated materials have several advantages:

- Highly toxic materials are safer for applicators to mix and apply
- Delayed or slow release of the active ingredient prolongs its effectiveness, allowing for fewer and less precisely timed applications
- The pesticide volatilizes more slowly; less is lost from the application site

Microencapsulated materials offer fewer hazards to the skin than ordinary formulations. Microencapsulated materials, however, pose a special hazard to bees. Foraging bees may carry microencapsulated materials back to their hives because they are about the same size as pollen grains. As the capsules break down, they release the pesticide, poisoning the adults and brood.

Breakdown of the microencapsulated materials to release the pesticide sometimes depends on weather conditions. Under certain conditions, the microencapsulated materials may break down more slowly than expected. This could leave higher residues of pesticide active ingredient in treated areas beyond normal restricted-entry or harvest intervals with the potential to injure fieldworkers. For this reason, regulations require long restricted-entry intervals for some microencapsulated formulations.

b. Water-Soluble Packets (WSB or WSP)

Water-soluble packets reduce the mixing and handling hazards of some highly toxic pesticides. Manufacturers package precise amounts of wettable powder or soluble powder formulations in a special type of plastic bag. When you drop these bags into a filled spray tank, they dissolve and release their contents to mix with the water. There are no risks of inhaling or contacting the undiluted pesticide as long as you do not open the packets. Once mixed with water, however, pesticides packaged in water-soluble packets are no safer than other diluted pesticides.

c. Attractants

Attractants include pheromones, sugar and protein syrups, yeasts, and rotting meat. Pest managers use these attractants in various types of traps (Figure 19). Attractants also can be combined with pesticides and sprayed onto foliage or other items in the treatment area.

d. Impregnated Products

Manufacturers impregnate (saturate) pet collars, livestock ear tags, adhesive tapes, plastic pest strips, and other products with pesticides (Figure 20). These pesticides evaporate over time, and the vapors provide control of nearby pests. Some paints and wood finishes have pesticides incorporated into them to kill insects or retard fungal growth. Fertilizers also may be impregnated with pesticides.

e. Repellents

Various types of insect repellents are available in aerosol and lotion formulations. People apply these to their skin or clothing or to plant foliage to repel biting and nuisance insects. You can mix other types of repellents with water and spray them onto ornamental plants and agricultural crops to prevent damage from deer, dogs, and other animals.

f. Animal Systematics

Systemic pesticides protect animals against fleas and other external blood-feeding insects as well as against worms and other internal parasites. A systemic animal pesticide is one that is absorbed and moves within the animal. These pesticides enter the animal's tissues after being applied orally or externally. Oral applications include food additives and premeasured capsules and liquids. External applications involve pour-on liquids, liquid sprays, and dusts.

g. Fumigants

Fumigants are pesticides that form gases or vapours toxic to plants, animals, and microorganisms. Some active ingredients are formulated, packaged, and released as gases; others are liquids when packaged under high pressure and change to gases when they are

released. Other active ingredients are volatile liquids when enclosed in an ordinary container and, therefore, are not formulated under pressure. Others are solids that release gases when applied under conditions of high humidity or in the presence of water vapor. Fumigants are used for structural pest control, in food and grain storage facilities, and in regulatory pest control at ports of entry and at state and national borders. In agricultural pest control, fumigants are used in soil, greenhouses, granaries, and grain bins.

Advantages of fumigants:

- Toxic to a wide range of pests
- Can penetrate cracks, crevices, wood, and tightly packed areas such as soil or stored grains
- Single treatment usually kills most pests in treated area

Disadvantages of fumigants:

- The target site must be enclosed or covered to prevent the gas from escaping
- Nonspecific in that they are highly toxic to humans and all other living organisms
- Require the use of specialized protective equipment, including respirators specifically approved for use with fumigants
- Require the use of specialized application equipment

h. Adjuvants

Adjuvants are substances used with a pesticide to enhance performance. By themselves, they do not possess pesticidal activity. Adjuvants may be added to the product at the time of formulation or by the applicator to the spray mix just prior to treatment. Adjuvants include surfactants, compatibility agents, antifoaming agents and spray colorants (dyes), and drift control agents.

Care should be taken when selecting an adjuvant. Pesticide performance can differ depending on what type of adjuvant is used. The pesticide label will state if specific surfactants are required and the amount (%) of active ingredient it must contain.

Probable Questions:

1. What are the advantages and disadvantages of emulsifiable pesticides?
2. What is aerosol pesticide? What are its advantages and disadvantages ?
3. What are the advantages and disadvantages of dust pesticide formulation ?
4. What are the advantages and disadvantages of bait pesticide formulation ?
5. What are the advantages and disadvantages of granular pesticide formulation ?
6. What are the advantages and disadvantages of wettable pesticide formulation ?
7. What are the advantages and disadvantages of fumigant pesticide formulation ?
8. What is adjuvant? How it is used as pesticide formulation ?

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

Unit-XIX

Database search tool; Sequence alignment and database searching; Computational tools and biological databases, NCBI, EBL, Sequence similarity tools; Blast and FASTA

Objective: In this unit we will discuss about different search tools which are used for sequence alignment analysis.

Definition of Bioinformatics:

Bioinformatics is currently defined as the study of information content and information flow in biological systems and processes. It serves as the bridge between observations (data) in diverse biologically-related disciplines and the derivations of understanding (information) about how the systems or processes function and subsequently the application (knowledge). Though Hwa Lim, Father of Bioinformatics, coined the word 'bio/informatique' in 1987, but Temple Smith used the term 'Bioinformatics' in 1991. In Silico Biology, a new area of Biology, has been developed in recent years because of generation of data in the field of genetics at an unprecedented exponential rate; the management and use of which requires the increasing use of computers and the relevant software.

Computational Biology, another term often used interchangeably with bioinformatics, although the former typically focuses on algorithm development and specific computational methods, while the latter focuses more on hypothesis testing and discovery in the biological domain.

Systems Biology, another area of research, emerged due to availability of enormous amount of molecular data and bioinformatics tools creating unprecedented opportunities to assemble and integrate this data into networks of genes, proteins and bio-chemical pathways. Bioinformatics involves collection, storage, retrieval and analysis of biological data that has a lot of applications in pharmaceutical, agricultural and food industries, and in molecular genetics research. Biological data are generated from various genome sequencing projects, obtained by different techniques like ONA sequencing (genome and EST), 2D gel electrophoresis, mass spectroscopy (MS, MALDI, LC-MS), protein crystallization, microarrays (e.g., cDNA, oligos, peptide), molecular markers (e.g., RFLP, RAPD, AFLP, SNP). Thus bioinformatics is an interface of biological sciences, mathematics, physical sciences and computer sciences, i.e., the integrated field of biology and information technology.

Major public domain bioinformatics facilities are (Fig. 19.2):

(i) Institutes

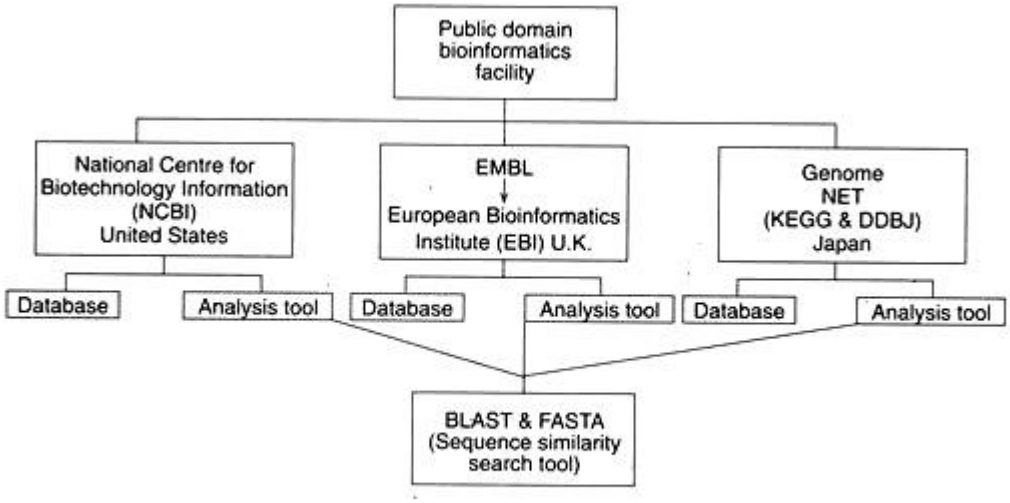


Fig. 19.2: Major publicly available databases and data mining tools

(a) NCBI – National Centre for Biotechnology Information, USA.

(b) EBI – European Bioinformatics Institute, UK.

(c) SIB – Swiss Institute of Bioinformatics, Switzerland.

(d) Genome NET (KEGG & DDBJ), Japan.

ii. Websites:

Some important websites commonly used for bioinformatics are depicted in Table 19.1.

Table 19.1: Some important websites commonly used for bioinformatics (From P. K. Gupta)

Subject	Source	Link
Nucleic acid sequence	Gen Bank	http://www.ncbi.nih.gov:80enterz/query/fcgi?bd-Nucleotide
Genome sequence	SRS at EMBL/FBI	
	Entrez Genome	http://srs.cbiac.uk
	TIGR database	http://www.ncbi.nlm.nin.gov:80/entrez/query.bd=Genome
Protein sequence	GenBank	http//www./tigr.org/tbl/
	SWISS-PORT at ExPASY	http://www.ncbi.nlm.nin.gov:80/entrez/query.fcgi?bd=Protein
	PIR	http://www.expasy.ch/spro/
Protein structure	Protein Data Bank	http://www.ndrf.georgetown.edu
Post translational modifications	RESID	http://www.rcsb.org/pdb/
Biochemical and biophysical information	ENZYME	http://www.ndrf.georgetwon.edu/pirwww/search/textresid.html
	BIND	http://www.expasy.ch/enzyme
Biochemical pathways	Path DB	http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Structure
	KEGG	http://www//ncgr.org/software/pathdb
	WIT	http://www.genome.ad.anl.jp/eegg/
Microarray	Gene Expression Links	http://www.wit.mcs.anl.gov/WIT2/
Other interesting sites	European Bioinformatics	http//industry.ebi.ac.uk/valan/MicroArray
	Institute	http://www.ebi.ac.uk
	DNA Database of Japan	http://www.nig.oc.jp/home.html

iii. Databases:

Bioinformatics is involved in storing the sequence information in different nucleic acid and protein databases which can be assessed by people all over the world through network technology.

The major protein databases are:

PDB, SWISS-PROT, PROSITE, ExPASy, PIR, PRINTS, BLOCKS, PRODOM, Pfam, Inter Pro.

Nucleic acid databases:

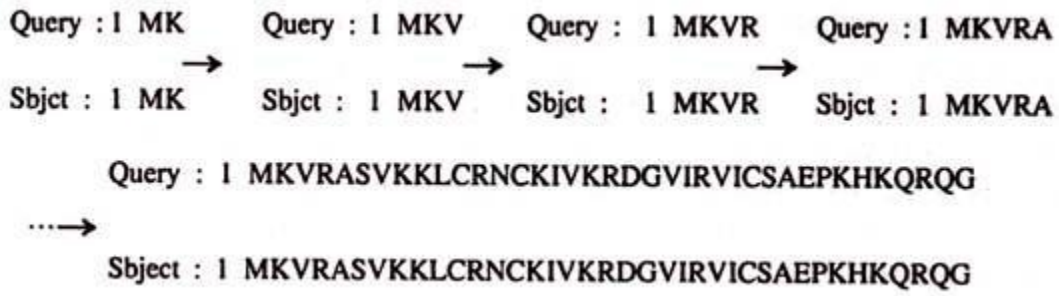
The major nucleic acid databases are: Gen Bank, DDBJ, Ref Seq, dbEST, NDB, CSD, EMBL.

Principles of Sequence Similarity Searches:

The characterization of any new DNA or protein sequence starts with a database search to find out whether homologs of this gene (protein) are available, and in what detail. Clearly,

looking for a matching sequence is quite straightforward. Take the first letter of the query sequence, search for its first occurrence in the database, and then check if the second letter of the query is the same in the subject.

If the two letters match, check the third, then the fourth, and continue this comparison to the end of the query. If the match for second letter fails, the search for another occurrence of the first letter will be done, and so on. This will identify all the sequences in the database that are identical to the query sequence (or include it).



Here we looked only for sequences that exactly match the query. To find sequences with the exclusion of the first letter, the same analysis may be conducted with the fragments starting from the second letter of the original query, then from the third one, and so on.

Query 1:1 KVRASVKKLCRNCKIVKRDGVIRVICSAEPKHKQRQG

Query 2:1 VRASVKKLCRNCKIVKRDGVIRVICSAEPKHKQRQG

Query 3:1 RASVKKLCRNCKIVKRDGVIRVICSAEPKHKQRQG

Query 4:1 ASVKKLCRNCKIVKRDGVIRVICSAEPKHKQRQG

These searches, at higher scale, become time-consuming. Finding close relatives would lead to additional conceptual and technical problems. Next, assume that sequences that are 99% identical are definitely homologous. Then, what is the threshold to consider sequences not to be homologous:50% identity, 33%, or perhaps 25% ? The example of two lysozymes shows that sequences with as low as low as 8% identity may belong to orthologous proteins and perform the same function.
Following the information theory of C E Shannon [The Mathematical Theory of Communication, 1949], we can calculate the information content of nucleic acids and of protein. If we use 2-bits (0 or 1 constitute a bit), we can encode 4 units of information (00, 01, 10, 11) which is sufficient to represent one base position in the DNA or RNA.

However, two bases (4-square) are not sufficient to code for the 20 amino acids that are used to constitute the various protein molecules. If we take three bases (4-cube), it gives us a code space of 64 which is more than the requisite 20. This redundancy leads to many codons for each amino acid, error-correcting codes and third place specialties (such as stop codon: TAA, TAG, TGA).

Another aspect is the execution of the "Central Dogma." This is interesting in that it leads to introduction of noise from such sources as vector sequences, heterologous sequences, rearranged & deleted sequences, repetitive element contamination, frame shift errors and sequencing errors or natural polymorphism.

As a matter of fact, all the four nucleotides, A, T, C, and G, are found in the database with approximately the same frequencies and have roughly the same probability of mutating one into another. As a result, DNA-DNA comparisons are largely based on simple text matching, which makes them fairly slow and not particularly sensitive, although a variety of heuristics have been devised to overcome this.

In Contrast, Amino Acid Sequence Comparisons have Several Distinct Advantages, which, at least Potentially, Lead to a Much Greater Sensitivity:

(i) There are 20 amino acids but only four bases. Hence, an amino acid match carries with it > 4 bits of information as opposed to only two bits for a nucleotide match. Thus, statistical significance can be established for much shorter sequences in protein comparisons than in nucleotide comparisons,

(ii) There is redundancy of the genetic code. Almost one-third of the bases in coding regions are under a weak (if any) selective pressure and represent noise, which adversely affects the sensitivity of the searches,

(iii) Nucleotide sequence databases are much larger than protein databases because of the vast amounts of non-coding sequences coming out of eukaryotic genome projects, and this further lowers the search sensitivity,

(iv) Probably most importantly, unlike in nucleotide sequence, the likelihoods of different amino acid substitutions occurring during evolution are substantially different, and taking this into account greatly improves the performance of database search methods. Given all these advantages, comparisons of any coding sequences are typically carried out at the level of protein sequences ; even when the goal is to produce a DNA- DNA alignment (e.g. for analysis of substitutions in silent codon positions), it is usually first done with protein sequences, which are then replaced by the corresponding coding sequences. Direct nucleotide sequence comparison is indispensable only when non-coding regions are analysed.

The laboratory-based as well as research-based sequencing and other types of information relating to the nucleic acids and the proteins are collected as bioinformatics databases in two broad categories: central repository (such as NCBI for nucleotide sequences, Swiss-Prot and PDB for protein sequences, and the smaller ones like Flybase, MGD for mouse genome and RGD for rat genome etc) and combined/secondary databases (such as KEGG for pathway and genome, prosite for annotated protein etc.). The databases are of the most sophisticated type in the computer world and hence require organizational as well as voluntary support for maintenance and upkeep. In fact, the databases are not mere collection of sequences. For example, the PDB (Protein Data Bank) is the single largest worldwide repository for three-dimensional structures of large biological molecules and as early September 2006, it stores 38620 structures.

Thus it houses the sequence, atomic coordinates, derived geometric data, secondary structure content as well as annotations about protein literature references. The PDB was established with 7 structures in 1971 and in 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) was assigned to manage its affairs at Brookhaven National Laboratory.

Substitution Scores and Substitution Matrices:

The fact that each of the 20 standard protein amino acids has its own unique properties means that the likelihood of the substitution of each particular residue for another residue during evolution should be different. Generally, the more similar the physico-chemical properties of two residues, the greater is the chance that the substitution will not have an adverse effect on the protein's function and, accordingly, on the organism's fitness.

Hence, in sequence comparisons, such a substitution should be penalized less than a replacement of amino acid residue with one that has dramatically different properties. This is an oversimplification, because the effect of a substitution depends on the structural and functional environment where it occurs. But, in general, we do not have a priori knowledge of the location of a particular residue in the protein structural and functional environment where it occurs, and even with such knowledge, incorporating it in a database search algorithm is an extremely complex task.

Thus, a generalized measure of the likelihood of amino acid substitutions is applied so that each substitution is given an appropriate value or score (weight) to be used in sequence comparisons. The score for a substitution between amino acids i and j can be expressed by the following intuitively plausible formula, which shows how likely is a particular substitution, given the frequencies of each the two residues in the analysed database:

$$S_{ij} = K \ln (q_{ij}/p_i p_j) \quad (I)$$

where K is a coefficient, q_{ij} is the observed frequency of the given substitution, and p_i, p_j are the background frequencies of the respective residues. Obviously, here the product $p_i p_j$ is the expected frequency of the substitution and, if $q_{ij} = p_i p_j$ ($S_{ij} = 0$), the substitution occurs

just as often as expected. In practice, the scores used are scaled such that the expected score for aligning a random pair of amino acid sequences is negative.

There are two fundamental ways to design a substitution score matrix, i.e. a triangular table containing 210 numerical score values for each pair of amino acids, including identities (diagonal elements of the matrix). As in many other situations in computational biology, the first approach works abolition, whereas the second one is empirical.

One ab initio approach calculates the score as the number of nucleotide substitutions that are required to transform a codon for one amino acid in a pair into a codon for the other. In this case, the matrix is obviously unique (as long as alternative genetic codes are not considered) and contains only four values, 0, 1,2, or 3. Accordingly, this is a very coarse grain matrix that is unlikely to work well. The other ab initio approach assigns scores on the basis of similarities and differences in the physico-chemical properties of amino acids.

Under this approach, the number of possible matrices is infinite, and they may have as fine a granularity as desirable, but a degree of arbitrariness is inevitable because our understanding of protein physics is insufficient to make informed decisions on what set of properties "correctly" reflects the relationships between amino acids. Empirical approaches, which came first, attempt to derive the characteristic frequencies of different amino acid substitutions from actual alignments of homologous protein families. In other words, these approaches strive to determine the actual likelihood of each substitution occurring during evolution. Obviously, the outcome of such efforts critically depends on the quantity and quality of the available alignments, and even now, any alignment database is far from being complete or perfectly correct.

Furthermore, simple counting of different types of substitutions will not suffice if alignments of distantly related proteins are included because, in many cases, multiple substitutions might have occurred in the same position, Ideally, one should construct the phylogenetic tree for each family, infer the ancestral sequence for each internal node, and then count the substitutions exactly. This is not practicable in most cases, and various shortcuts need to be taken.

Several solutions to these problems have been proposed, each resulting in a different set of substitution scores. The first substitution matrix, constructed by Dayhoff and Eck (1968), was based on an alignment of closely related proteins, so that the ancestral sequence could be deduced and all the amino acid replacements could be considered occurring just once.

This model was then extrapolated to account for more distant relationships, which resulted in the PAM series of substitution matrices. PAM (Accepted Point Mutaion) is a unit of evolutionary divergence of protein sequences, corresponding to one amino acid change per 100 residues. Thus, for example, the PAM30 matrix is supposed to apply to proteins that differ, on average, by 0.3 change per aligned residue, whereas PAM250 should reflect evolution of sequences with an average of 2.5 substitution per position.

Accordingly, the former matrix should be employed for constructing alignments of closely related sequences, whereas the latter is useful in database searches aimed at detection of

distant relationships. Using an approach similar to that of Dayhoff, combined with rapid algorithms for protein sequence clustering and alignment, Jones, Taylor, and Thornton produced the series of the so-called JTT matrices, which are essentially and update of the PAMS. The PAM and JTT matrices, however, have limitations arising out of the fact that they have been derived from alignments of closely related sequences and extrapolated to distantly related ones. This extrapolation may not be fully valid because the underlying evolutionary model might not be adequate, and the trends that determine sequence divergence of closely related sequences might not apply to the evolution at larger distances. In 1992, Steven and Jorja Henikoff developed a series of substitution matrices using conserved ungapped alignments of related proteins from the BLOCKS database. The use of these alignments offered three important advantages over the alignments used for constructing the PAM matrices.

First, the BLOCKS collection obviously included a much larger number and, more importantly, a much greater diversity of protein families than the collection that was available to Dayhoff and coworkers in the 1970's. Second, coming from rather distantly related proteins, BLOCKS alignments better reflected the amino acid changes that occur over large phylogenetic distances and thus produced substitution scores that represented sequence divergence in distant homologs directly, rather than through extrapolation.

Third, in these distantly related proteins, BLOCKS included only the most confidently aligned regions, which are likely to best represent the prevailing evolutionary trends. These substitution matrices, named the BLOSUM (= BLOCKS Substitution Matrix) series, were tailored to particular evolutionary distances by ignoring the sequences that had more than a certain percent identity. In the BLOSUM62 matrix, for example, the substitution scores were derived from the alignments of sequences that had no more than 62% identity, the substitution scores of the BLOSUM45 matrix were calculated from the alignments that contained sequences with no more than 45% identity.

Accordingly, BLOSUM matrices with high numbers, such as BLOSUMSO, are best suited for comparisons of closely related sequences (it is also advisable to use BLOSUMSO for database searches with short sequences), whereas low-number BLOSUM matrices, such as BLOSUM45, are better for distant relationships. In addition to the general purpose PAM, JTT, and BLOSUM Series, some specialized substitution matrices were developed, for example, for integral membrane proteins, but they never achieved comparable recognition.

Several early studies found the PAM matrices based on empirical data consistently resulted in greater search sensitivity than any of the ab initio matrices. An extensive empirical comparison showed that: (i) BLOSUM matrices consistently outperformed PAMs in BLAST searches and (ii) on average, BLOSUM62 performed best in the series ; this ; this matrix is currently used as the default in most sequence database searches. It is remarkable that, so far, empirical matrices have consistently outperformed those based on theory, either physico-chemical or evolutionary. This perhaps points out that we do not yet have an adequate theory to describe protein evolution.

Statistics of Protein Sequence Comparison:

Let us consider the same protein sequence (E. coli RpsJ) as above

Query ” 1 MKVRASVKKLCRNCKIVKRDGVIRVICSAEPKHKQRQG 38

and check how many times segments of this sequence of different lengths are found in the database (we chose fragments starting from the second position in the sequence because nearly every protein in the database starts with a methionine). Not unexpectedly, we find that the larger the fragment, the smaller the number of exact matches in the database.

With the decrease in the number of database hits, the likelihood that these hits are biologically relevant, i.e. belong to homologs of the query protein, increases. Thus, 13 of the 23 occurrences of the string KVRASV and all 8 occurrences of the string KVRASVK are from RpsJ orthologs.

The number of occurrences of a given string in the database can be roughly estimated as follows. The probability of matching one amino acid residue is 1/20 (assuming equal frequencies of all 20 amino acids in the database ; this not being the case, the probability is slightly greater). The probability of matching two residues in a row is then $(1/20)^2$, and the probability of matching n residues is $(1/20)^n$. Given that the protein database currently contains $N \sim 2 \infty 10^8$ letters, one should expect a string of n letters to match approximately $N \infty (1/20)^n$ times.

Searching for perfect matches is the simplest but insufficient form of sequence database search. However, it is important as one of the basic steps in currently used search algorithms.

Further, the goal of a search is to find homologs, including distant homologs where only a small fraction of the amino acid residues are identical or even similar. Even in close homologs, a region of high similarity is usually flanked by dissimilar regions like in the following alignment of E. coli RpmJ with its ortholog from Vibrio cholerae.

```
E. coli RpmJ :  1  MKVRASVKKLCR—NCKIVKRDGVIRVICSAEPKHKQRQG
                   MKV +S+K      +C+IVKR G + VIC + P + K Q
Vibrio VC0879 : 1  MKVLSSLKSAKNRHPDCQIVKRRGRLYVICKSNPRFKAVQR
```

In this example, the region of highest similarity is in the middle of the alignment, but including the less conserved regions on both sides improves the overall score.

Further along the alignment, the similarity almost disappears so that inclusion of additional letters into the alignment would not increase the overall score or would even decrease it. Such fragments of the alignment of two sequences whose similarity score cannot be improved by adding or trimming any letters, are referred to as high-scoring segment pairs (HSPs). For this approach to work, the expectation of the score for random sequences must be negative, and the scoring matrices used in database searches are scaled accordingly.

So, instead of looking for perfect matches, sequence comparisons programs actually search for HSPs. Once a set of HSPs is found, different methods, such as Smith-Waterman, FASTA, or BLAST, deal with them in different fashions.

However, the principal issue that any database search method needs to address is identifying those HSPs that are unlikely to occur by chance and, by inference, are likely to belong to homologs and to be biologically relevant. This problem has been solved by Samuel Karlin and Stephen Altschul, who showed that maximal HSP scores follow the extreme value distribution. Accordingly, if the lengths of the query sequence (m) and the database (n) are sufficiently high, the expected number of HSPs with a score of at least S is given by the formula

$$E = Kmn2^{-\lambda s} \text{ (II)}$$

Here, S is the so-called raw score calculated under a given scoring system, and K and λ are natural scaling parameters for the search space size and the scoring system, respectively. Normalizing the score according to the formula:

$$S' = (\lambda S - \ln K)/\ln 2 \text{ (III)}$$

gives the bi score, which has a standard unit accepted in information theory and computer science. Then,

$$E = mn2^{-S'} \text{ (IV)}$$

and, since it can be shown that the number of random HSPs with score _ S' is described by Poisson distribution, the probability of finding at least one HSP with bit score _ S' is

$$P = 1 - e^{-E} \text{ (V)}$$

Equation (V) links two commonly used measures of sequence similarity, the probability (P-value) and expectation (E-value). For example, if the score S is such that three HSPs with this score (or greater) are expected to be found by chance, the probability of finding at least one such HSP is $(1 - e^{-3})$, ~ 0.95.

By definition, P-values vary from 0 to 1, whereas E-values can be much greater than 1. The BLAST programs report E- values, rather than P-values, because E-values of, for example, 5 and 10 are much easier to comprehend than P-values of 0.993 and 0.99995. However, for E < 0.01, P-value and E-value are nearly identical.

The product mn defines the search space, a critically important parameter of any database search. Equations (II) and (IV) codify the intuitively obvious notion that the larger the search space, the higher the expectation of finding an HSP with a score greater than any given value. There are two corollaries of this that might take some more time in getting used to: (i) the same HSP may come out statistically significant in a small database and not significant in a large database; with the natural growth of the database, any given alignment becomes less and less significant (but by no means less important because of

that) and (ii) the same HSP may be statistically significant in a small protein (used as a query) and not significant in a large protein.

Clearly, one can easily decrease the E-value and the P-value associated with the alignment of the given two sequences by lowering n in equation (II), i.e. by searching a smaller database. However, the resulting increase in significance is false, although such a trick can be useful for detecting initial hints of subtle relationships that should be subsequently verified using other approaches.

It is the experience of the author that the simple notion of E (P)-value is often misunderstood and interpreted as if these values applied just to a single pairwise comparison (i.e., if an E-value of 0.001 for an HSP with score S is reported, then, in a database of just a few thousand sequences, one expects to find a score > S by chance).

It is critical to realize that the size of the search space is already factored in these E-values, and the reported value corresponds to the database size at the time of search (thus, it is certainly necessary to indicate, in all reports of sequence analysis, which database was searched, and desirably, also on what exact date).

The Karlin-Altschul statistics has been rigorously proved to apply only to sequence alignments that do not contain gaps, whereas statistical theory for the more realistic gapped alignments remains an open problem. However, extensive computer simulations have shown that these alignments also follow the extreme value distribution to a high precision ; therefore, at least for all practical purposes, the same statistical formalism is applicable.

Sequence Alignment and Similarity Search:
The Basic Alignment Concepts and Principal Algorithms:

The similarity searches air at identifying the homologs of the given query protein (or nucleotide) sequences in the database. In principle, the only way to identify homologs is by aligning the query sequence against all the sequences in the database (some important heuristics that allow an algorithm to skip sequences that are obviously unrelated to the query are discussed below), sorting these hits based on the degree of similarity, and assessing their statistical significance that is likely to be indicative of homology. Let's briefly discuss alignment methods first.

It is important to make a distinction between a global (i.e. full-length) alignment and a local alignment, which includes only parts of the analysed sequences (subsequences). Although, in theory, a global alignment is best for describing relationships between sequences, in practice, local alignments are of more general use for two reasons: (i) it is common that only parts of compared proteins are homologous (e.g. they share one conserved domain, whereas other domains are unique), and (ii) often, only a portion of the sequence is conserved enough to carry a detectable signal, whereas the rest have diverged beyond

recognition. Optimal global alignment of two sequences was first implemented in the Needleman-Wunsch algorithm, which employs dynamic programming.

Sequence Database Search Algorithms:

Smith-Waterman:

Any pairwise sequence alignment method in principle can be used for database search in a straightforward manner. All that needs to be done is to construct alignments of the query with each sequence in the database, one by one, rank the results by sequence similarity, and estimate statistical significance.

The classic Smith-Waterman algorithm is a natural choice for such an application, and it has been implemented in several database search programs, the most popular one being SSEARCH written by William Pearson and distributed as part of the FASTA package. It is currently available on numerous servers around the world.

The major problem preventing SSEARCH and other implementations of the Smith-Waterman algorithm from becoming the standard choice for routine database searches is the computational cost, which is orders of magnitude greater than it is for the heuristic FASTA and BLAST methods.

Since extensive comparisons of the performance of these methods in detecting structurally relevant relationships between proteins failed to show a decisive advantage of SSEARCH, the fast heuristic methods dominate the field. Nevertheless, on a case- by-case basis, it is certainly advisable to revert to full Smith-Waterman search when other methods do not reveal a satisfactory picture of homologous relationship for a protein of interest. A modified, much faster version of the Smith-Waterman algorithm has been implemented in the MPSRCH program.

FASTA:

FASTA, introduced in 1988 by William Pearson and David Lipman, was the first database search program that achieved search sensitivity comparable to that of Smith-Waterman but was much faster. FASTA looks for biologically relevant global alignments by first scanning the sequence for short exact matches called “words” ; a word search is extremely fast.

The idea is that almost any pair of homologous sequences is expected to have at least one short word in common. Under this assumption, the great majority of the sequences in the database that do not have common words with the query can be skipped without further examination with a minimal waste of computer time. The sensitivity and speed of the database search with FASTA are inversely related and depend on the “k-tuple” variable, which specifies the word size ; typically, searches are run with k = 3, but, if high sensitivity at the expense of speed is desired, one may switch to k = 2.

Subsequently, Pearson introduced several improvements to the FASTA algorithm, which are implemented in the FASTA3 program.

BLAST:

Basic Local Alignment Search Tool (BLAST) is the most widely used method for sequence similarity search ; it is also the fastest one and the only one that relies on a complete, rigorous statistical theory. Like FASTA and in contrast to the Smith-Waterman algorithm, BLAST employs the word search heuristics to quickly eliminate irrelevant sequences, which greatly reduces he search time. The program initially searches for a word of a given length W (usually 3 amino acids or 11 nucleotides) that scores at least T When compared to the query using a given substitution matrix.

Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding exceeding the threshold of S. The W and T parameters dictate the speed and sensitivity of the search, which can thus be varied by the user. The original version of BLAST (known as BLAST 1.4) produced only ungapped local alignments, for which rigorous statistical theory is available. Although this program performed well for many practical purposes, it repeatedly demonstrated lower sensitivity than the Smith-Waterman algorithm and the FASTA program, at least when run with the default parameters. The new generation of BLAST makes alignments with gaps, for which extensive simulations have demonstrated the same statistical properties as proved for ungapped alignments.

The BLASTX, TBLASTN, and TBLASTX programs are used when either the query or the database or both are uncharacterized sequences and the location of protein-coding regions is not known. These programs translate the nucleotide sequence of the query in all six possible frames and run a protein sequence comparison analogous to that in BLASTP.

A version of gapped BLAST, known as WU-BLAST, with a slightly different statistical model, which, in some cases, may lead to a greater search sensitivity, is supported by Waren Gish at Washington University in St. Louis. Recently, the BLAST suite was supplemented with BLAST2 sequences, a tool for comparing just two nucleotide or protein sequences. Because of its speed, high selectivity, and flexibility, BLAST is the first choice program in any situation when a sequence similarity search is required, and importantly, this method is used most often as the basis for genome annotation. Therefore, we may consider the practical aspects of BLAST use in some detail. Before that, however, we need to introduce some additional concepts that are critical for protein sequence analysis.

Analysis and Interpretation of BLAST Results:

In spite of the solid statistical foundation, including composition-based statistics, BLAST searches inevitably produce both false positives and false negatives. The main cause for the appearance of false positives, i.e. database hits that have “significant” E- values but, upon

more detailed analysis, turn out not to reflect homology, seems to be subtle compositional bias missed by composition-based statistics or low-complexity filtering.

The reason why false negatives are inevitable is, in a sense, more fundamental: in many cases, homologs really have low sequence similarity that is not easily captured in database searches and, even if reported, may not cross the threshold of statistical significance. In an iterative procedure like PSI-BLAST, both the opportunities to detect new and interesting relationships and the pitfalls are further exacerbated. Beyond the (conceptually) straightforward issues of selectivity and sensitivity, functional assignments based on database search results require careful interpretation if we want to extract the most out of this type of analysis while minimizing the chance of false predictions. Below we consider both the issues of search selectivity and sensitivity and functional interpretation.

No cut-off value is capable of accurately partitioning the database hits for a given query into relevant ones, indicative of homology, and spurious ones. By considering only database hits with very high statistical significance (e.g. $E < 10^{10}$) and applying composition-based statistics, false positives can be eliminated for the overwhelming majority of queries, but the price to pay is high: numerous homologs, often including those that are most important for functional interpretation, will be missed.

This brief discussion certainly cannot cover all "trade secrets" of sequence analysis. However, the above seems to be sufficient to formulate a few rules of thumb that help a researcher to extract maximal amount of information from database searches while minimizing the likelihood of false "discoveries".

Probable Questions:

1. Define Bioinformatics.
2. Name major Institutes related to bioinforatics study.
3. What is gap opening penalty and gap extension penalty?
4. How ideal sequence alignment can be achieved?
5. What is FASTA ? State its importance in Bioinformatics .
6. What is BAST? State its importance in Bioinformatics .
7. What is substitution score?
8. What is substitution matrix ?
9. Describe Smith Waterman algorithm.
10. What is Trace back matrix ?

Suggested Readings:

1. Bioinformatics – Principles and Applications – Ghosh and Mullick
2. Bioinformatics by R. Sundaralingam & V Kumarsen – Saras Publication
3. Bioinformatics – Sequence and Genome analysis by David. W. Mount.
4. Introduction to Bioinformatics by Arthur. M. Lesk.

UNIT-XX

Methods for analysis of gene expression at RNA and protein level, large scale expression, such as micro array based techniques

DNA Microarray:

The DNA Microarray technology is used to determine the level of expression of many thousands of genes simultaneously. This new approach is used not for individual genetic loci, rather, for the analysis of genome-wide patterns of gene expression. Using DNA microarrays, it is possible to estimate the relative level of gene expression of each gene in the genome.

The DNA microarray or chip is a high density grid system, consisting of a flat solid substrate about the size of a postage stamp that can be used to detect hybridisation of target DNA under appropriate conditions. The chip contains 10,000 to 100,000 distinct spots, from 75 to 150 μm in diameter.

The spacing between spots on an array is usually 100 to 200 μm. Each spot contains a different immobilised DNA sequence that can be hybridised with DNA (or RNA) from a large number of different cells. Two types of chips are currently available: one, in which oligonucleotides have been synthesised directly on the chip, one nucleotide at a time, by automated procedures.

These chips have hundreds of thousands of spots per array; second, chips in which double-stranded DNA sequences of 500 to 5000 base pairs have been deposited through drops by capillary action from miniaturized devices mounted on the movable head of a robotic workstation. These chips have tens of thousands of spots per array. The surface onto which DNA is spotted is critically important. The ideal surface immobilizes the target DNAs, and is compatible with stringent probe hybridisation conditions.

The procedure shown (Fig. 24.1) depicts only 6 spots in a chip, each of which contains a DNA sequence that serves as a probe for a different gene. Experimental cells are used for the extraction of cellular mRNA, and a control sample of mRNA from another source. The samples are subjected to reverse transcription to obtain DNA strands. In the experimental material, the primer for reverse transcription is tagged with a green fluorescent label, while primers of the control material receive red fluorescent label. After the DNA strands have been obtained in sufficient quantity, the fluorescent samples are mixed and hybridised with the DNA in the spots in the chip. The hybridisation is competitive because the two samples were mixed.

Therefore, the density of red and green strands bound to the chip is proportional to the concentration of red or green molecules in the mixture. Genes that are over-expressed in the experimental sample relative to the control will have more green strands hybridised to the spot, whereas those that are under-expressed in the experimental sample relative to the control will have more of red strands hybridised to the spot.

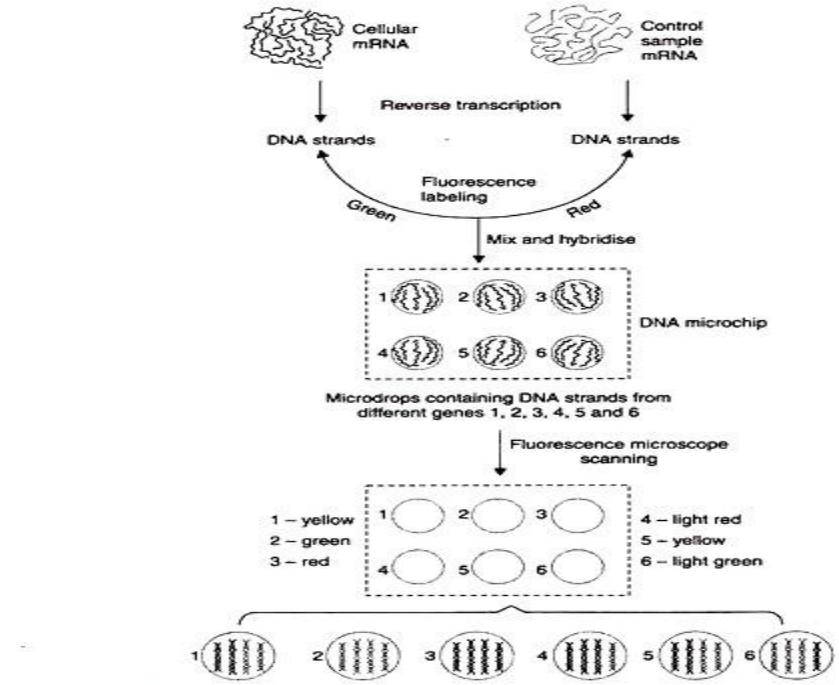


Fig. 24.1 Procedure for DNA microarrays. Six dried microdrops are introduced into a DNA microchip. Each drop contains immobilised DNA strands from a different gene numbered 1 to 6. These are hybridised with fluorescence-labelled DNA samples obtained by reverse transcription of cellular mRNA (green) and red labelled control DNA sample. Competitive hybridisation of green (experimental) and red (control) label is proportional to the *relative* abundance of each mRNA in the sample. The intensity of red and green fluorescence is analysed by microscopy and interpreted as overexpression, underexpression and equal expression of the gene depending on the intensity of red, green, orange, yellow-green and yellow fluorescence.

The intensity of fluorescence is viewed by placing the chip under a laser scanning microscope or a fluorescence microscope that scans each pixel, which is the smallest discrete unit in a visual image. The intensity of fluorescent label is recorded. The signals are synthesised to produce a signal value for each spot in the microarray.

The signals indicate the relative levels of gene expression through colour. Green or yellow green indicate over-expression in experimental sample, while red or orange indicates under-expression in experimental sample. Yellow indicates equal expression in both experimental and control samples. DNA microarray technology is useful for study of large number of cells growing under different conditions, at different developmental stages, or at different stages of a disease. Besides detection of gene expression, this technology can be used to detect mutations and polymorphisms, to map genomic DNA clones, and to compare the gene expression pattern in normal and diseased tissues

Fabrication:

Microarrays can be fabricated using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using pre-made masks, photolithography using dynamic micro-mirror devices, ink-jet printing, or electrochemistry on microelectrode arrays.
DNA microarrays can be used to detect RNAs that may or may not be translated into active proteins. Scientists refer to this kind of analysis as "expression analysis" or expression profiling. Since there can be tens of thousands of distinct probes on an array, each microarray experiment can accomplish the equivalent number of genetic tests in parallel. Arrays have, therefore, dramatically accelerated many types of investigations. The use of microarrays for gene expression profiling was first published in 1995 (Science) and the first complete eukaryotic genome {Saccharomyces cerevisiae) on a microarray was published in 1997 (Science).

1. Spotted Microarrays*:*

In spotted microarrays (or two-channel or two-colour microarrays), the probes are oligonucleotides, cDNA or small fragments of PCR products that correspond to mRNAs and are spotted onto the microarray surface. This type of array is typically hybridized with cDNA from two samples to be compared (e.g., diseased tissue versus healthy tissue) that are labelled with two different fluorophores (e.g., Rhodamine (Cyanine 5, red) and Fluorescein (Cyanine 3, green)). The two samples are mixed and hybridized to a single microarray that is then scanned in a microarray scanner to visualize fluorescence of the two fluorophores. Relative intensities of each fluorophore are then used to identify up-regulated and down-regulated genes in ratio-based analysis. Absolute levels of gene expression cannot be determined in the two-colour array, but relative differences in expression among different spots (= genes) can be estimated with some oligonucleotide arrays.

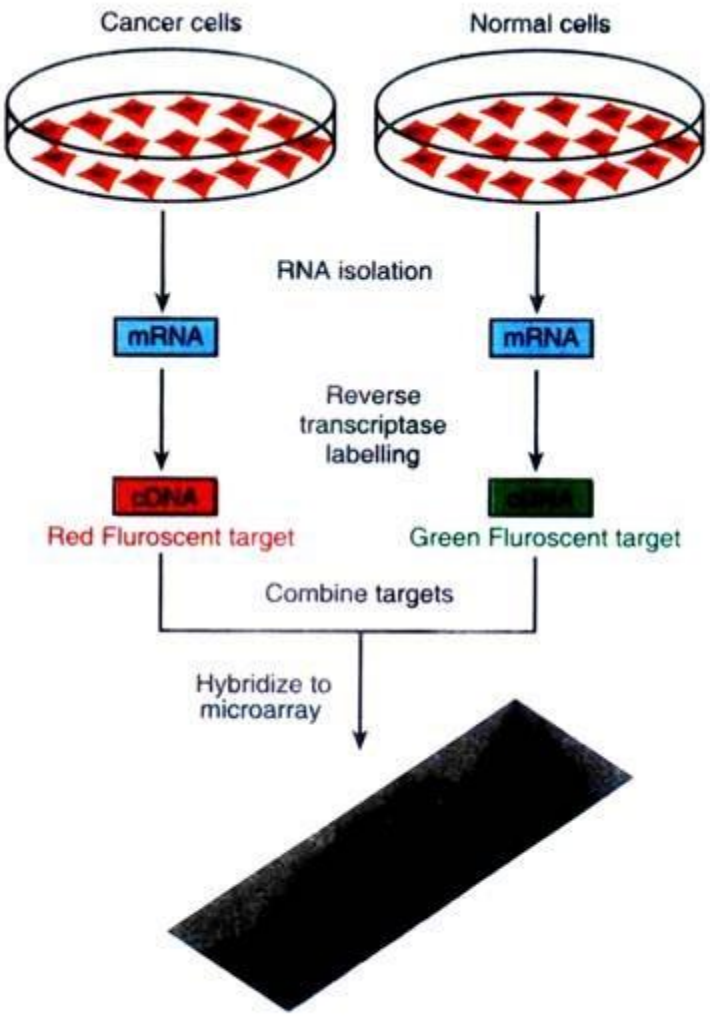


Fig. 16.2: Diagram of typical dual-colour microarray experiment

2. Oligonucleotide Microarrays:

In oligonucleotide microarrays (or single-channel microarrays), the probes are designed to match parts of the sequence of known or predicted mRNAs. There are commercially available designs that cover complete genomes from companies such as GE Healthcare, Affymetrix, Ocimum Bio-solutions, or Agilent. These microarrays give estimations of the absolute value of gene expression and, therefore, the comparison of two conditions requires the use of two separate micro- arrays.

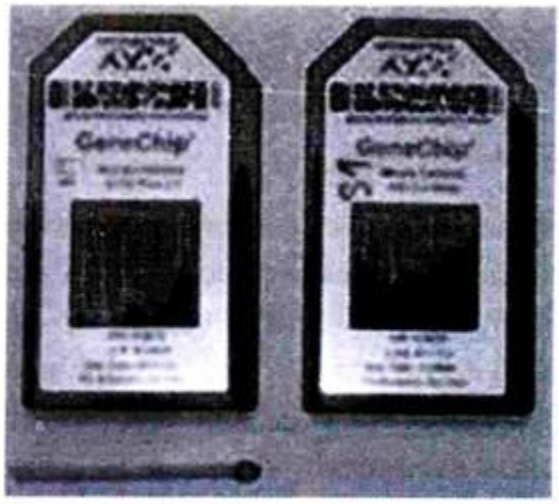


Fig. 16.3: Two Affymetrix chips

Oligonucleotide Arrays can be either produced by piezoelectric deposition with full length oligonucleotides or in situ synthesis. Long Oligonucleotide Arrays are composed of 60-mers, or 50-mers and are produced by ink-jet printing on a silica substrate. Short Oligonucleotide Arrays are composed of 25-mer or 30-mer and are produced by photolithographic synthesis (Affymetrix) on a silica substrate or piezoelectric deposition (GE Healthcare) on an acrylamide matrix. More recently, Maskless Array Synthesis from NimbleGen Systems has combined flexibility with large numbers of probes. Arrays can contain up to 390,000 spots, from a custom array design. New array formats are being developed to study specific pathways or disease states for a systems biology approach.
Oligonucleotide microarrays often contain control probes designed to hybridize with RNA spike-ins. The degree of hybridization between the spike-ins and the control probes is used to normalize the hybridization measurements for the target probes.

Genotyping Microarrays:

DNA microarrays can also be used to read the sequence of a genome in particular positions. SNP microarrays are a particular type of DNA microarrays that are used to identify genetic variation in individuals and across populations.
Short oligonucleotide arrays can be used to identify the single nucleotide polymorphisms (SNPs) that are thought to be responsible for genetic variation and the source of susceptibility to genetically caused diseases. Generally termed genotyping applications, DNA microarrays may be used in this fashion for forensic applications, rapidly discovering or measuring genetic predisposition to disease, or identifying DNA-based drug candidates.

These SNP microarrays are also being used to profile somatic mutations in cancer, specifically loss of heterozygosity events and amplifications and deletions of regions of DNA. Amplifications and deletions can also be detected using comparative genomic hybridization, or aCGH, in conjunction with microarrays, but may be limited in detecting novel Copy Number Polymorphisms, or CNPs, by probe coverage.
Re-sequencing arrays have also been developed to sequence portions of the genome in individuals. These arrays may be used to evaluate germ line mutations in individuals, or somatic mutations in cancers. Genome tiling arrays include overlapping oligonucleotides designed to blanket an entire genomic region of interest. Many companies have successfully designed tiling arrays that cover whole human chromosomes.

Microarrays and Bioinformatics:

1. Experimental Design:

Due to the biological complexity of gene expression, the considerations of experimental design that are discussed in the expression profiling article are of critical importance if statistically and biologically valid conclusions are to be drawn from the data.

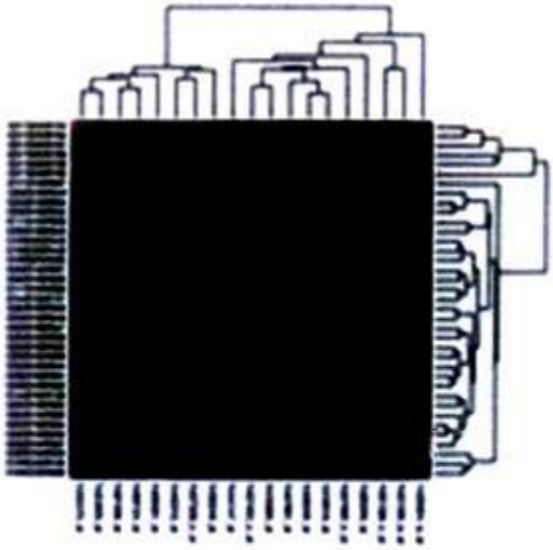


Fig. 16.4: Gene expression values from microarray experiments can be represented as heat maps to visualize the result of data analysis

There are three main elements to consider when designing a microarray experiment.

First, replication of the biological samples is essential for drawing conclusions from the experiment.
Second, technical replicates (two RNA samples obtained from each experimental unit) help to ensure precision and allow for testing differences within treatment groups. The technical replicates may be two independent RNA extractions or two aliquots of the same extraction.
Third, spots of each cDNA clone or oligonucleotide are present at least as duplicates on the microarray slide, to provide a measure of technical precision in each hybridization. It is critical that information about the sample preparation and handling is discussed in order to help identify the independent units in the experiment as well as to avoid inflated estimates of significance.

2. Standardization:

The lack of standardization in arrays presents an interoperability problem in bioinformatics, which hinders the exchange of array data. Various grass-roots open-source projects are attempting to facilitate the exchange and analysis of data produced with non-proprietary chips.

a. The "Minimum Information about a Microarray Experiment" (MIAME) checklist helps define the level of detail that should exist and is being adopted by many journals as a requirement for the submission of papers incorporating microarray results. MIAME describes the minimum required information for complying experiments, but not its format. Thus, as of 2007, whilst many formats can support the MIAME requirements there is no format which permits verification of complete semantic compliance.
b. The "MicroArray Quality Control (MAQC) Project" is being conducted by the FDA to develop standards and quality control metrics which will eventually allow the use of MicroArray data in drug discovery, clinical practice and regulatory decision-making.
3. The MicroArray and Gene Expression (MAGE) group is working on the standardization of the representation of gene expression data and relevant annotations.

3. Statistical Analysis:

The analysis of DNA microarrays poses a large number of statistical problems, including the normalization of the data. There are dozens of proposed normalization methods in the published literature; as in many other cases where authorities disagree, a sound conservative approach is to try a number of popular normalization methods and compare the conclusions reached; how sensitive are the main conclusions to the method chosen?
From a hypothesis-testing perspective, the large number of genes present on a single array means that the experimenter must take into account a multiple testing problem; even if the statistical P-value assigned to a given gene indicates that it is extremely unlikely that differential expression of this gene was due to random rather than treatment effects, the

very high number of genes on an array makes it likely that differential expression of some genes represents false positives or false negatives.
Statistical methods tailored to microarray analyses have recently become available that assess statistical power based on the variation present in the data and the number of experimental replicates, and can help minimize type I and type II errors in the analyses.
A basic difference between microarray data analysis and much traditional biomedical research is the dimensionality of the data. A large clinical study might collect 100 data items per patient for thousands of patients. A medium-size microarray study will obtain many thousands of numbers per sample for perhaps a hundred samples. Many analysis techniques treat each sample as a single point in a space with thousands of dimensions, then attempt by various techniques to reduce the dimensionality of the data to something humans can visualize.

4. Relation between Probe and Gene:

The relation between a probe and the mRNA that it is expected to detect is problematic. On the one hand, some mRNAs may cross-hybridize probes in the array that are supposed to detect another mRNA. On the other hand, probes that are designed to detect the mRNA of a particular gene may be relying on genomic EST information that is incorrectly associated with that gene.

Public Databases of Microarray Data:

Database	Microarray Experiment Sets	Sample Profiles	As of Date
Gene Expression Omnibus - NCBI	5366	134669	April 1, 2007
Stanford Microarray database	12742	?	April 1, 2007
UNC Microarray database	~31	2093	April 1, 2007
MUSC database	~45	555	April 1, 2007
ArrayExpress at EBI	1643	136	April 1, 2007
caArray at NCI	41	1741	November 15, 2006

Online Microarray Data Analysis Programs and Tools:

Several Open Directory Project categories list online microarray data analysis programs and tools:

i. Bioinformatics: Online Services:
Gene Expression and Regulation at the Open Directory Project

ii. Gene Expression:
Databases at the Open Directory Project

iii. Gene Expression:
Software at the Open Directory Project

iv. Data Mining:
Tool Vendors at the Open Directory Project

v. Bio-conductor:
Open source and open development software project for the analysis and comprehension of genomic data

vi. Genevestigator:
Web-based database and analysis tool to study gene expression across large sets of tissues, developmental stages, drugs, stimuli, and genetic modifications.

DNA Chip and Development study:

The mechanistic basis of metazoan development represents one of the unsolved mysteries of biology: how does a single fertilized egg, through successive cell divisions and differentiation events, mature into an adult organism? The fruitfly *Drosophila melanogaster* has been a pioneering model organism for geneticists and developmental biologists for many decades. Drosophilogists have been quick to exploit the power of genome-wide expression profiling using DNA microarrays. One notable example is the study of the expression of 4028 genes analysed in wild-type flies throughout *Drosophila* development during 66 sequential time periods. These included sampling RNA at fertilization, embryonic, larval and pupal periods as well as the first 30 days of adulthood. Each experimental sample was compared with a common reference sample, allowing the relative abundance of any transcript to be determined at every developmental stage. The analysis of such a huge amount of data conventionally proceeds

by the use of algorithms that group or cluster genes according to similarity in their expression profiles.

The analysis of the *Drosophila* dataset revealed that, despite the use of whole animals, it was possible to discern expression profiles in specific organs, as well as those associated with particular biological processes. For example, one cluster of 23 genes included eight known to be expressed in terminally differentiated muscle. The profile of this cluster has two peaks of expression, one coinciding with the larval stage and a second with adult muscle development. Initiation of larval muscle development is regulated by the basic helix–loop–helix (bHLH) transcription factor Twist, which induces expression of *dMef 2*, which itself encodes a MADS box transcription factor regulating the transcription of muscle differentiation genes. Crucially, this muscle-specific regulatory hierarchy was recapitulated in the microarray data: the peak of *twist* expression preceded that of *dMef 2*, which preceded transcription of genes in a muscle differentiation cluster. Moreover, 15 of the 23 genes in this latter cluster contained pairs of predicted dMEF2-binding sites. Similar clusters were identified revealing coordinate expression profiles associated with particular biochemical and cellular functions, including mitochondrial proteins, components of the 26S proteasome complex and cytoskeletal/neuronal factors.

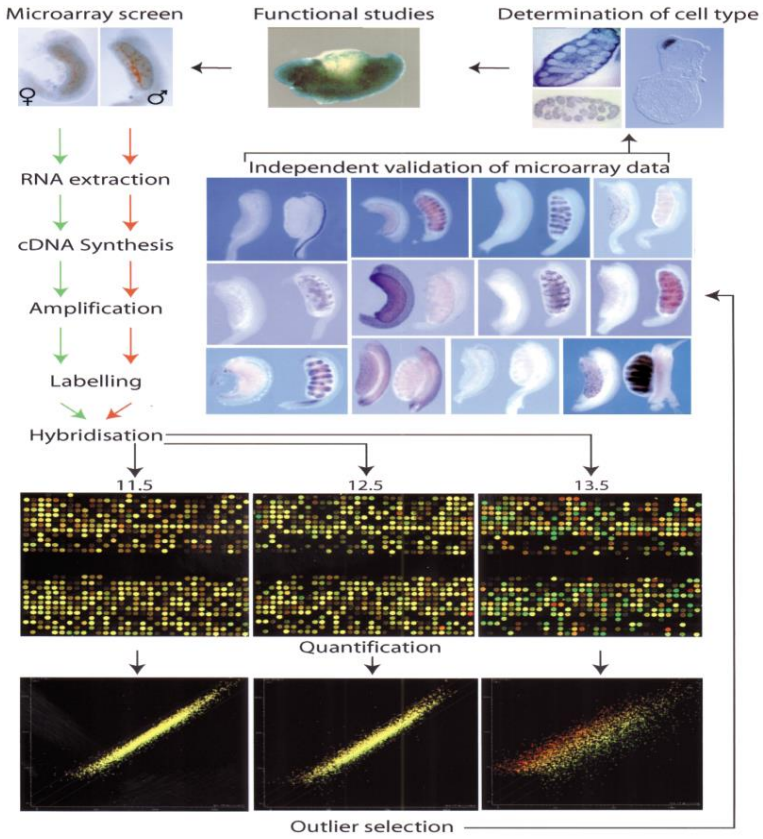


Fig 5: A screen for genes expressed in a sexually dimorphic fashion during mouse gonad and mesonephros development using DNA microarrays.

Global transcriptional information during morphogenesis was also readily available: the vast majority of genes (>88%) that exhibit transcriptional modulation during the stages analysed are expressed during the first 20 h of development, before the end of embryogenesis. A total of 2103 changed during embryogenesis, 445 changed during larval life, 646 during the pupal stage and 118 during adult life. The transcript levels of only 16 genes changed significantly during the adult time period sampled. These data suggest a strong association between modulation of transcriptional activity and morphogenesis.
The pioneering experiments in invertebrates suggest that the notion that gene expression profiles alone do not reveal biological function needs to be re-examined. Surveying gene expression under a wide range of conditions and tissues, in wild-type and mutant animals, seems to transform the significance of data that on a smaller scale would be considered descriptive. Of course, for any individual gene residing within a 'functionally loaded' cluster the task remains to determine the phenotypic consequences of its mutation. Yet perhaps such experiments should be seen as complementing our understanding of that gene's function developed by other means, rather than being exclusively definitive thereof: particularly given the high frequency with which no clear phenotype is observed after mutagenesis. To infer function from gene expression profiles is not mere speculation if the design of the experiment and the complexity of the dataset permit otherwise

Studies were done in the mammalian embryo at a genome-wide level throughout its development, in a manner reminiscent of those discussed in invertebrates. Given the widespread accessibility of microarray technology today, the observations and analysis are very complex, involving references to both technical and 'cultural' issues. The most common technical problem concerns the small amounts of RNA available from standard dissections of mammalian embryos. By 'cultural', we mean the familiarity that developmental biologists have with in situ hybridization (ISH), their relative lack of familiarity with microarrays and the common attitude that descriptions of gene expression patterns support only speculation about function. However, these remarks are equally applicable to developmental biologists using flies and worms as a model. The use of DNA microarrays to examine mammalian development is a small but rapidly growing field of study. It is currently dominated by the exploitation of arrays to perform screens for molecules involved in particular developmental processes.

Systematic genome-wide studies of mammalian development using microarrays stand out due to their rarity. Studies in mouse and the analysis of the expression of 18 816 mouse genes in 49 different embryonic and adult tissues, permitted some clustering of genes pertinent to the development of specific tissues, such as the central nervous system. However, the limited number of embryonic samples, totalling 11, means that this study falls short of providing a transcriptional profile of mouse development. Perhaps due to the relative complexity of the mammalian embryo, more familiar are studies aimed at profiling

expression at specific embryonic stages or in specific embryonic tissues, including (without attempting to be comprehensive): 12.5 days post coitum (dpc) mouse placenta, mouse retina, mouse lung, mouse mammary gland, preimplantation mouse embryos, mouse hippocampus, and mouse B cells. Developmental biologists have also been quick to adapt familiar techniques for the purposes of exploiting microarray technology, including the use of cell line models and organ cultures too.

Applications of these Arrays include:

1. mRNA or gene expression profiling:

Monitoring expression levels for thousands of genes simultaneously is relevant to many areas of biology and medicine, such as studying treatments, disease, and developmental stages. For example, microarrays can be used to identify disease genes by comparing gene expression in diseased and normal cells.

2. Comparative genomic hybridization (Array CGH):

Assessing large genomic rearrangements.

3. SNP detection arrays:

Looking for single nucleotide polymorphism in the genome of populations.

4. Chromatin immunoprecipitation (ChIP) studies:

Determining protein binding site occupancy throughout the genome, employing ChlP-on-chip technology.

Probable Questions:

1.How phage display technique can be used for protein protein interaction assay?

2. What is yeast two hybrid system? How it is used to determine protein protein interaction?

3. What is DNA chip?

4. State the basic principle of DNA microarray analysis.

5. What is spotted microarray? Explain.

6. What is oligonucleotide microarray? Explain .

7. How DNA chip can be used in developmental study?

8. What are the applications of DNA microarray?

Suggested readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.

DISCLAIMER: This Self Learning Material (SLM) has been compiled from various authentic books, Journals articles, e-journals and other web sources.