POST-GRADUATE DEGREE PROGRAMME (CBCS)

IN

ZOOLOGY

(M.Sc. Programme)

SEMESTER-IV

Molecular Biology, Biotechnology, Tools and Technique

ZCORT-412

Self-Learning Material

DIRECTORATE OF OPEN AND DISTANCE LEARNING
UNIVERSITY OF KALYANI
KALYANI-741235, WEST BENGAL

CONTENT WRITER:

Sl. No.	Name of Content Writer	Designation	Unit
1.	Dr. Laishram Pradeep Kumar Singh	Assistant Professor of Zoology, University of Kalyani	Units I-III
2.	Dr. Anupam Podder	Assistant Professor of Zoology, DODL, University of Kalyani	Units IV, V, IX, X, XVI, XIX & XX
3.	Dr. Asmita Samadder	Assistant Professor of Zoology, University of Kalyani	Units VI & VIII
4.	Dr. Arunima Biswas	Assistant Professor of Zoology, University of Kalyani	Units XI & XIII
5.	Dr. Sudeshna Banerjee	Assistant Professor of Zoology, DODL, University of Kalyani	Units VII, XII, XIV, XV, XVII, XVIII & XX

Acknowledgements:

The author thankfully acknowledges all the faculty members of Department of Zoology, University of Kalyani for their academic contribution and valuable suggestions regarding the preparation of Self Learning Material.

May, 2025

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

Printed by East India Photo Composing Centre, 209A, Bidhan Sarani, Kolkata-700006

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of Distance and reaching the unreached students are the three fold 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 endeavor. 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.) Kallol Paul, Hon'ble Vice-Chancellor, University of Kalyani**, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it with in 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 wouldespecially 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

List of PGBOS members

1)	Prof. Debjani Nath, Professor and Head, Dept.of Zoology, University of Kalyani	Chairperson
2)	Prof. Kausik Mondal, Professor, Dept. of Zoology, University of Kalyani	Member
3)	Dr. Arunima Biswas, Assistant Professor, Dept.of Zoology, University of Kalyani	Member
4)	Professor Anilava Kaviraj, Retired Professor, University of Kalyani	External nominated member
5)	Professor Subir Chandra Dasgupta, Retired professor (WBSES) MA college, Honorary Emeritus professor, Dept of Zoology, Ramkrishna mission Vidyamandir, Belur math, Howrah711202	External nominated member
6)	Dr. Sudeshna Banerjee, Assistant Professor of Zoology, DODL, University of Kalyani	Member
7)	Dr. Anupam Podder, Assistant Professor of Zoology, DODL, University of Kalyani	Member
8)	Director, DODL, University of Kalyani	Convener

Molecular Biology, Biotechnology, Tools and Technique
ZCORT-412

Module	Unit	Content	Credit	Page No.
Part-B: Tools and Technique	**XV**	Separation of Cell Constituents: i) Chromatography: ion exchange; gel filtration and HPLC ii) Electrophoresis-PAGE, SDS-PAGE (One and Two dimensional)	**3**	288-310
	XVI	Methods for analysis of gene expression at RNA and protein level, large scale expression, such as micro array-based techniques		311-319
	XVII	Spectroscopy: UV- spectroscopy, Circular Dichroism, surface plasma resonance methods		320-330
	XVIII	Blotting Methods: Southern, Northern & Western blotting. RFLP, RAPD and AFLP techniques		331-351
	XIX	Pesticide formulation		352-357
	XX	Database search tool; Sequence alignment and database searching; Computational tools and biological databases, NCBI, EMBL, PDB, Sequence similarity tools; Blast and FASTA Phylogenetic analysis with the program PHYLIP, DISTANCES, and GROWTREE. Basics of designing a microarray, image analysis and normalization, annotations		358-375

Unit-I

Transcriptional 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.

INTRODUCTION

Most of the genetic information carried by DNA specifies the sequence of amino acids in proteins. But this DNA does not direct the synthesis of proteins directly, instead producing RNA as an intermediary. When the cell needs a particular protein, the nucleotide sequence of the appropriate portion of the DNA molecule in a chromosome is first copied into RNA (a process called *transcription*). These RNA copies of segments of the DNA sequence are then used to direct the synthesis of the protein (a process called *translation*). The genetic information in cells thereby flows from DNA to RNA to protein. All cells, from bacteria to humans, express their genetic information in this way—a principle so fundamental that it is termed the *central dogma* of molecular biology (Figure. 1.1)

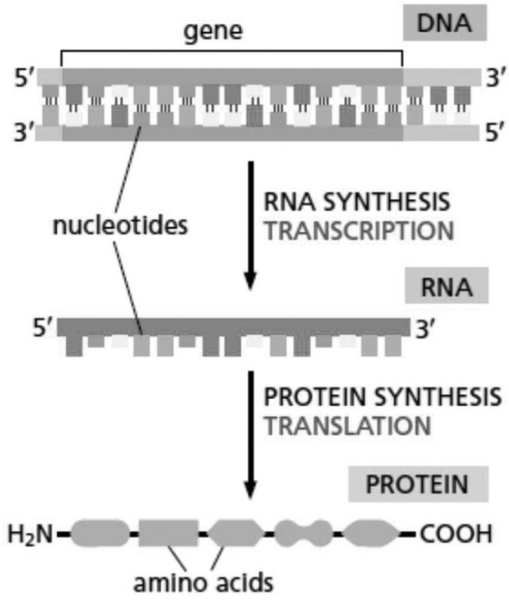


Figure 1.1: Genetic information flow as Central dogma of molecular biology

Gene expression is a multistep process that proceeds from transcription to translation, and it may involve posttranslational effects on protein structure and function. As shown in **Figure 1.2**, gene regulation can occur at any of these steps in the pathway of gene expression.

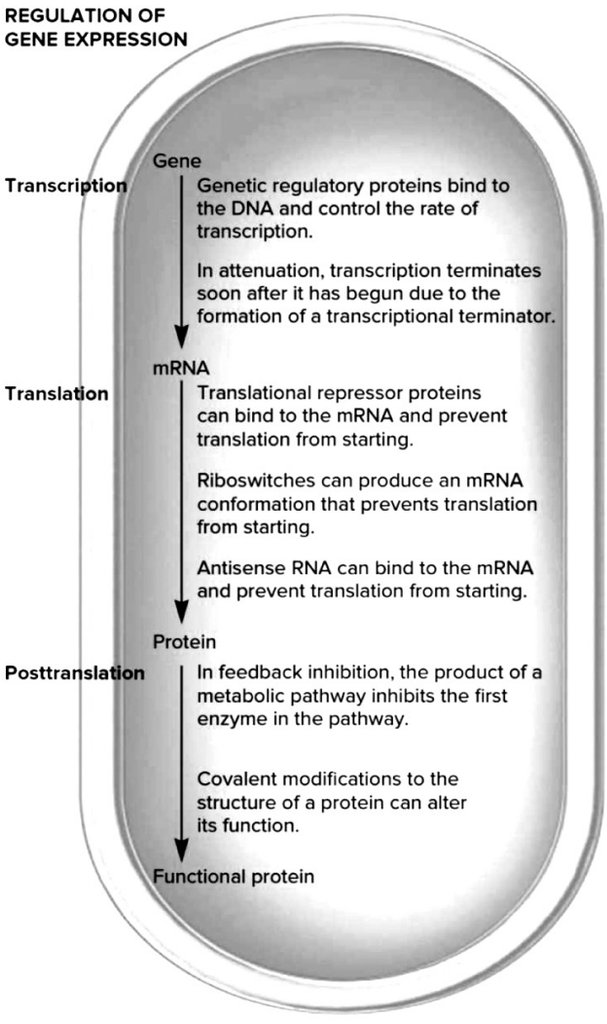


Fig 1.2: Multistep process of Gene regulation

Gene regulation is the phenomenon in which the level of gene expression can vary under different conditions. In comparison, unregulated genes have essentially constant levels of expression in all conditions over time. Unregulated genes are also called **constitutive genes.** Frequently, constitutive genes encode proteins that are continuously needed for the survival of the bacterium. In contrast, the majority of genes are **regulated** so that the proteins they encode can be produced at the proper times and in the proper amounts.

A key benefit of gene regulation is that the encoded proteins are produced only when they are required. Therefore, the cell avoids wasting valuable energy making proteins it does not need. From the viewpoint of natural selection, this enables an organism such as a bacterium to compete as efficiently as possible for limited resources.

1. TRANSCRIPTION

1.1 Overview

The first step in the transfer of information from DNA to protein is to produce an RNA strand whose nucleotide sequence matches the nucleotide sequence of a DNA. segment. Because this process is reminiscent of transcribing (copying) written words, the synthesis of RNA is called *transcription.* The DNA is said to be transcribed into

RNA, and the RNA is called a **transcript**. Information encoded in DNA is transferred to the RNA transcript by the complementary pairing of DNA and RNA bases. Consider the transcription of a chromosomal segment that constitutes a gene. First, the two strands of the DNA double helix separate locally to form a **transcription bubble**. One of the separated strands acts as a template for RNA synthesis and is called the **template strand** (or noncoding strand) and the other strand is called the **non-template strand** (or coding strand). The resulting RNA sequence is complementary to the template strand and identical (except for the use of uracil in place of thymine) to the non-template strand. When DNA sequence is cited in the scientific literature, the sequence of the non-template strand is almost always given because it is the same as the RNA sequence.

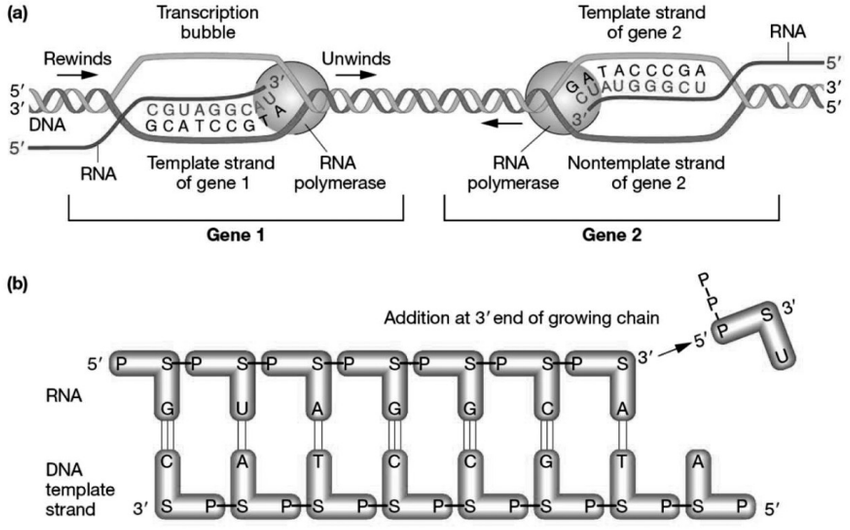


Figure 1.3: Overview of transcription

Across the genome, both DNA strands may be used as templates, but in any one gene, only one strand is used. Starting at the 3′ end of the template strand, ribonucleotides form base pairs by hydrogen bonding with their complementary DNA nucleotides. The ribonucleotide A pairs with T in the DNA, C with G, G with C, and U with A.

Each ribonucleotide is positioned opposite its complementary nucleotide by the enzyme **RNA polymerase**. This enzyme moves along the DNA template strand in the 3′-to-5′ direction forming **phosphodiester bonds** that covalently link aligned ribonucleotides to build RNA in the 5′-to-3′ direction, as shown in Figures 3. As the RNA strand is progressively lengthened, the 5′ end is displaced from the DNA template and the transcription bubble closes behind RNA polymerase. Multiple RNA polymerases, each synthesizing an RNA molecule, can move along a gene at the same time.

DNA-dependent RNA polymerases require Mg2 and all four rNTPs (ATP, GTP, UTP, and CTP) as substrates for the polymerization reaction. The chemistry and mechanism of RNA synthesis closely resemble those of DNA synthesis (Figure 3.a). RNA polymerase extends an RNA strand by adding ribonucleotide units to the 3'-hydroxyl end, building RNA in the 5' → 3' direction. The 3'-hydroxyl group makes a nucleophilic attack on the ⟨-phosphate of the incoming rNTP, with the concomitant release of pyrophosphate. As noted

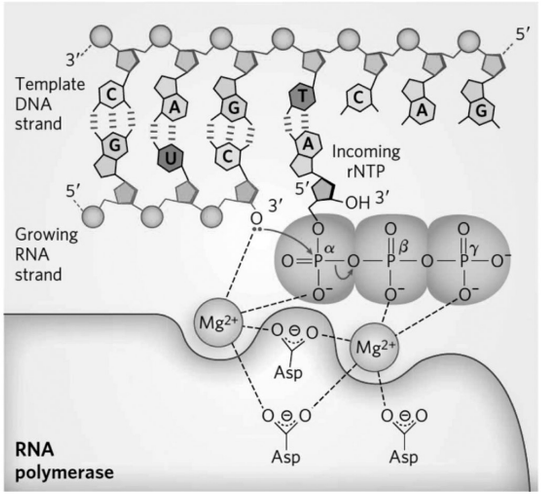


Figure 1.4: Transcriptional regulation by Regulatory Transcription factor

above, only one of the two DNA strands serves as template. The template DNA is copied in the 3′ → 5′ direction (antiparallel to the new RNA strand), just as in DNA replication. Each nucleotide in the newly formed RNA is selected by Watson-Crick base pairing: U residues and not T residues, as in DNA are inserted in the RNA to pair with A residues in the DNA template, G residues are inserted to pair with C residues, and so on (see Figure 3.2). Base-pair geometry may also play a role in nucleotide selection and the resulting fidelity of the polymerase reaction.

1.2 Stages of transcription:

Genes are segments of DNA embedded in extremely long DNA molecules Because the DNA of a chromosome is a continuous unit, the transcriptional machinery must be directed to the start of a gene to begin transcription, continue transcribing the length of the gene, and finally stop transcribing at the end of the gene. These three distinct stages of transcription are called initiation, elongation, and termination, respectively. The following describes the stages of transcription in bacteria and eukaryotic cells.

1.3.1 Transcription initiation in bacteria

In bacteria, RNA polymerase usually binds to a specific DNA sequence called a promoter, located close to the start of the transcribed region. Promoters are an important part of the transcriptional regulatory region of a gene (Figure 4). The first transcribed base is called the *initiation site* or the transcription start site. The promoter is referred to as upstream of the initiation site because it is located ahead of the initiation site (5′ of the gene). A downstream site is located later in the direction of transcription. Nucleotide positions upstream of the initiation site are indicated by a negative (–) sign and those downstream by a positive (+) sign. By convention, the first DNA base to be transcribed is numbered +1. Because the same RNA polymerase binds to the promoter sequences of these different genes, similarities among the promoters are not surprising. In particular, two regions of great similarity appear in virtually every case. These regions have been termed the –35 (minus 35) and –10 (minus 10) regions because they are located 35 base pairs and 10 base pairs, respectively, upstream of the first transcribed base. The –35 and –10 regions from different genes do not have to be identical to perform a similar function. Nonetheless, it is possible to arrive at a sequence of nucleotides, called a **consensus sequence**, that is in agreement with most sequences (Figure 4). An RNA polymerase holoenzyme binds to the DNA at this point, then unwinds the DNA double helix and begins the synthesis of an RNA molecule. The region between the start of transcription and the start of translation is referred to as the **5′ untranslated region (5′ UTR)**. The bacterial RNA polymerase that scans the DNA for a promoter sequence is called the **RNA polymerase holoenzyme.**

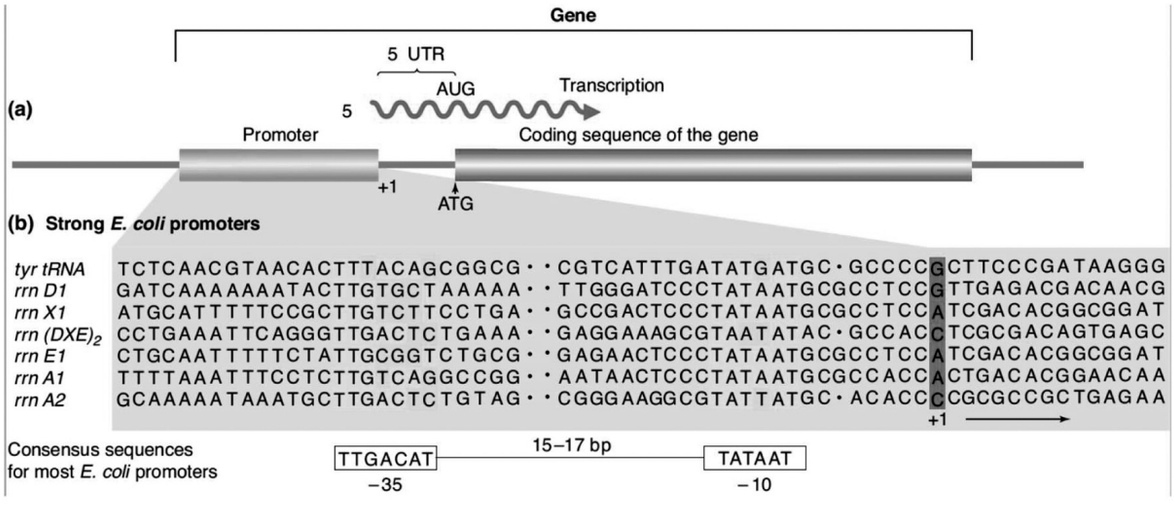


Figure 1.5: Promoter sequence of *E. coli* of (a) typical gene of and (b) different type of gene

This multi-protein complex is composed of a five-subunit **core enzyme** (two subunits of α, one of β, one of β2 , and one of ω) plus a subunit called sigma factor (σ). The two α subunits help assemble the enzyme and promote interactions with regulatory proteins, the β subunit is active in catalysis, the β 2 subunit binds DNA, and the ω subunit has roles in assembly of the holoenzyme and the regulation of gene expression. The σ subunit binds to the –10 and –35 regions, thus positioning the holoenzyme to initiate transcription correctly at the start site (Figure 4). The σ subunit also has a role in separating the DNA strands around the "10 region so that the core enzyme can bind tightly to the DNA in preparation for RNA synthesis. After the holoenzyme is bound, transcription initiates and the σ subunit dissociates (Figure 5). The core enzyme then elongates through the gene.

1.3.2 Transcription elongation in bacteria

As the RNA polymerase moves along the DNA, it unwinds the DNA ahead of it and rewinds the DNA that has already been transcribed (Figure 5). In this way, it maintains a region of single-stranded DNA, called a transcription bubble, within which the template strand is exposed. In the bubble, RNA polymerase monitors the binding of a free ribonucleoside triphosphate to the next exposed base on the DNA template and, if there is a complementary match, adds it to the chain (Figure 5). Energy for the addition of a nucleotide is derived from breaking a phosphate bond. RNA polymerase synthesizes RNA at a rate of 50 to 90 nucleotides per second. Within this range, slower rates of synthesis may provide time for the RNA to fold properly and for translation to synchronize with transcription. Inside the transcription bubble, the last 8 to 9 nucleotides added to the RNA chain form an RNA–DNA hybrid by complementary base pairing with the template strand. As the RNA chain lengthens at its 3′ end, the 5′ end is further extruded from the polymerase. The complementary base pairs are broken at the point of exit, leaving the extruded region of RNA single-stranded.

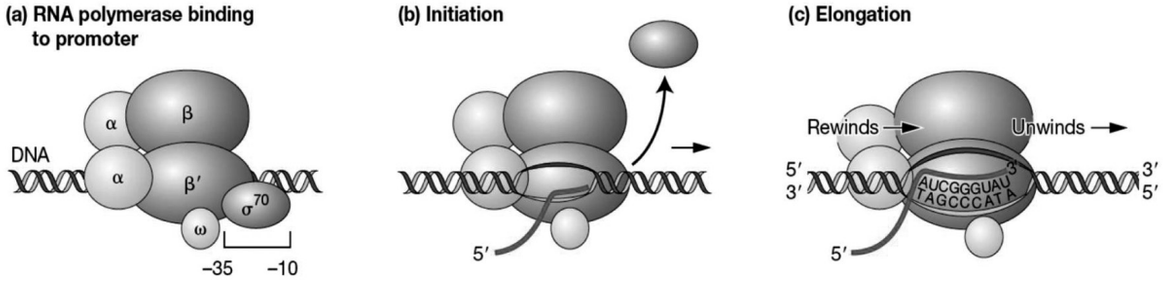


Figure 1.6: Transcription Initiation and Elongation in bacteria

1.3.3. Transcription termination in bacteria

Transcription continues beyond the protein-coding segment of a gene, creating a 3′ untranslated region (3′ UTR) at the end of the transcript. Elongation proceeds until RNA polymerase recognizes special nucleotide sequences that act as a signal to stop transcription and release RNA polymerase and the nascent (i.e., newly synthesized) RNA from the template.

There are two major types of termination mechanisms in *E. coli* (and other bacteria), **factor-independent termination** (also called intrinsic or rho-independent) and **Rho-dependent termination** (also called factor-dependent).

Factor-independent termination occurs after the transcription of a GC-rich stretch followed by an A-rich stretch in the template strand. In the RNA, the GC-rich sequence is self-complementary and forms a 7- to 20-base-pair stem loop followed by a 7- to 8-nucleotide U-rich tract. Stem loops are also called hairpins because they resemble the metal clips used to hold a person's hair in place Normally, in the course of transcription elongation, RNA polymerase will pause if the short RNA–DNA hybrid in the transcription bubble is weak and will backtrack to stabilize the hybrid. The strength of the hybrid is determined by the relative number of 3-hydrogen bond G"C and C"G base pairs compared with 2-hydrogen bond A"T and U"A base pairs. In the factor-independent mechanism, the polymerase is believed to pause after synthesizing the U's (U"A forms a weak RNA–DNA hybrid). However, the backtracking polymerase encounters the hairpin. This roadblock sets off the release of RNA from the polymerase and the polymerase from the DNA template.

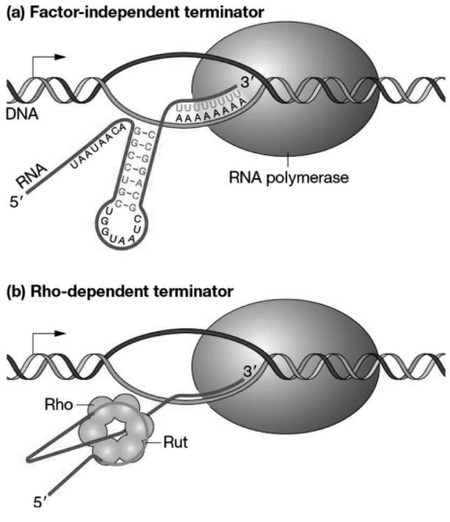


Figure 1.7: Transcription Termination in Bacteria (a) Factor independent (b) Factor dependent

In the Rho-dependent termination mechanism, a protein called Rho factor recognizes nucleotide sequences in the RNA that act as a termination signal for RNA polymerase. RNAs with Rho-dependent termination signals do not have the string of U residues at their 3′ end and usually do not have a hairpin (Figure 6). Instead, they have a sequence of about 50–90 nucleotides that is rich in C residues and poor in G residues and includes an upstream segment called a Rut (Rho utilization) site. Rut sites are located upstream (recall that upstream means 5′ of) from sequences at which the RNA polymerase tends to pause. Rho is a homo-hexamer consisting of six identical subunits that has **helicase** activity. Helicases use energy from ATP hydrolysis to move along a nucleic acid and unwind nucleic acid helices. Once bound at the Rut site, Rho travels toward the 3′ end of the transcript. When Rho encounters a paused RNA polymerase, it unwinds the RNA-DNA hybrid within the transcription bubble, dissociating the RNA and terminating transcription. Thus, Rho-dependent termination entails binding of Rho to the Rut site, pausing of RNA polymerase, and Rho-mediated release of the RNA from RNA polymerase.

1.4. Transcription in Eukaryotes

Transcription in eukaryotic organisms, including humans, is similar to transcription in bacteria in that they retain many of the events associated with initiation, elongation, and termination.

Transcription of nuclear genes occurs within the nucleus. In keeping with the greater complexity of the eukaryotic system, the transcriptional machinery is more complex. For example, in prokaryotes, apparently only one RNA polymerase transcribes all the different species of RNA. However, in higher forms, there are different RNA polymerases with separate functions (Table 1.1). Unlike the presence of relatively few accessory proteins for assembly of transcription initiation complex in prokaryotes, in eukaryotes a large number of transcription accessory factors are known.

Type of Polymerase	Transcription product	Location
RNA Polymerase I	18S and 28S rRNA	Nucleolus
RNA Polymerase II	hnRNA, small nuclear RNAs	Nucleus
RNA Polymerase III	t-RNAs, 5.6S r-RNA, small cytoplasmic RNAs	Nucleus

Table 1.1: Type of RNA polymerase in eukaryotes and their functions

In eukaryotes, different classes of genes have different promoter regions. The promoter region of the rRNA genes is recognised specifically by the transcription

factors associated with RNA Polymerase I. The transcription factors associated with RNA Polymerase II specifically interact with promoter regions of genes producing heterogeneous RNA. It may be pointed out here that while in most cases the promoter region does not get transcribed, the promoter region of some genes transcribed by the RNA Polymerase III gets partially transcribed. That is to say that the promoter region extends to the downstream region as well. The promoter for RNA Polymerase II comprises of a start point, a TATA like sequence and sequences recognised by the transcription factors (Figure 7).

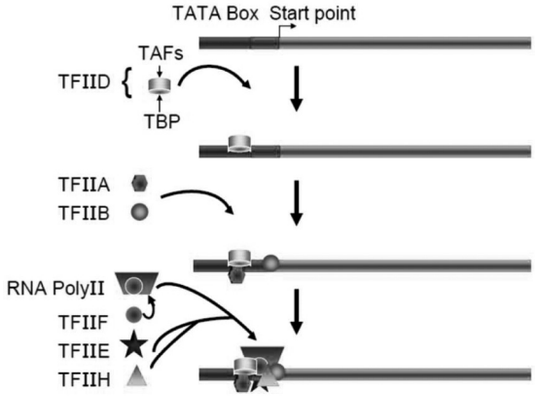


Figure 1.8. Model for assembly of Initiation complex at promoter site of RNA polymerase II in eukaryotes

The core promoter element includes the start site and the TATA box and is sufficient for initiation of transcription. The core promoter is about 100 bases long. The TATA box is located about 25-30 bp upstream of the start point. It may be further up-stream for lower eukaryotes like yeast. TATA binding proteins bind to the TATA box. Some core promoters may also include an initiator element (Inr). Inr elements seem to be involved in recruitment of transcriptional machinery. Interestingly, there may be core promoters with both TATA and Inr elements (composite promoters), TATA or Inr elements singly or may not contain either.

RNA polymerases I, II, and III cannot recognize promoter sequences on their own. However, unlike bacteria, where promoters are recognized by σ factor as an integral part of the RNA polymerase holoenzyme, eukaryotic promoters are recognized by each polymerase with its own set of **general transcription factors (GTFs)** that first bind specific sequences in the promoter and then bind the RNA polymerase. Nevertheless, the mechanisms in bacteria and eukaryotes are conceptually similar. In both cases, the information that defines a promoter is provided by short DNA sequences located near the transcription start site, and the sequences are bound by proteins that associate with RNA polymerase and position it at the correct site to start transcription.

1.4.1 Transcription initiation in eukaryotes

RNA polymerase II is responsible for transcribing all mRNAs as well as numerous ncRNAs, including snRNAs involved in splicing and miRNAs involved in mRNA decay and inhibition of translation. Transcription by RNA polymerase II is the most complex transcriptional system in eukaryotes because of the large number of gene targets with unique expression patterns. RNA polymerase II promoters, which are somewhat arbitrarily defined as sequences located within 100 base pairs of the transcription start site, contain a variety of promoter elements, a few of which are relatively common. About 25 percent of promoters in yeast and humans contain a **TATA box**, a sequence element so-named because the nucleotide sequence TATA appears in the consensus sequence TATA AAA. In animals, the TATA box is located about 30 nucleotides upstream of the transcription start site (i.e., "30), but in yeast its location is more variable (between "50 and "125). Another common promoter element is the initiator (Inr), which is located right at the transcription start site in about 40 percent of genes. Collectively, only about 50 percent of RNA polymerase II genes contain a TATA box and/or an Inr. This predicts the existence of other promoter elements. Computational analyses that searched for common sequences surrounding the transcription start sites of RNA polymerase II genes have identified additional promoter elements, including the downstream promoter element (DPE), which is located at about +25, and the TFIIB recognition element (BRE), which is located at about "40. However, since many genes lack all of the known promoter elements, it is likely that promoter elements remain to be discovered.

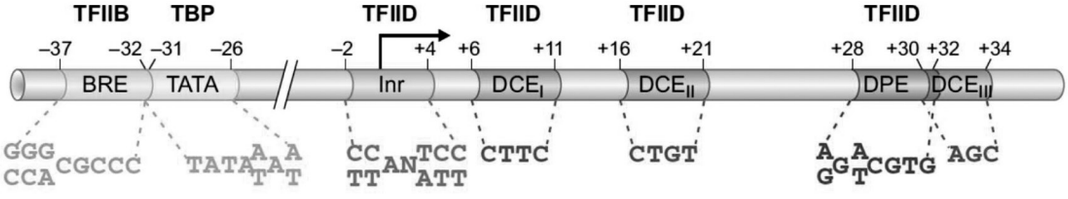


Fig 1.9: Polymerase II Promoter of Eukaryotes

All of the proteins that bind RNA polymerase II promoter elements are subunits of GTFs. Transcription factor IIB (TFIIB) binds the BRE, and TFIID binds the other promoter elements. The TFIID complex contains TBP (the same protein involved in RNA polymerase I transcription) and about 15 TBP-associated factors (TBP). TBP binds the TATA box, and TAFs bind the Inr and DPE. Binding of TFIID at a promoter is the first step in the sequential assembly of other GTFs and RNA polymerase II. TFIID binding instructs assembly of TFIIA and TFIIB, followed by TFIIF and RNA polymerase II as a

pre-assembled complex and ending with addition of TFIIE and TFIIH. The assemblage of GTFs and RNA polymerase II constitutes the **preinitiation complex (PIC)**, which serves to position RNA polymerase II at the transcription start site, generate the transcription bubble, and position the DNA in the active site of RNA polymerase II. TFIIA stabilizes the binding of TFIIB and TFIID at the promoter. TFIIH, which is recruited to the promoter by TFIIE, contains proteins with helicase activity that unwind the DNA into two strands to form the transcription bubble. Lastly, TFIIF places the promoter DNA in a position in RNA polymerase II that is appropriate for DNA unwinding and initiation of transcription at the start site. After transcription has been initiated, RNA polymerase II dissociates from most of the GTFs to elongate the RNA transcript. Some GTFs, including TFIID, remain at the promoter to attract the next RNA polymerase II. In this way, multiple RNA polymerase II molecules can simultaneously synthesize transcripts from a single gene.

A high, regulated levels of transcription in vivo require, additionally, transcriptional regulatory proteins, the Mediator complex, and nucleosome modifying enzymes (Figure 9). One reason for these additional requirements is that the DNA template in vivo is packaged into chromatin This condition complicates binding to the promoter of polymerase and its associated factors.

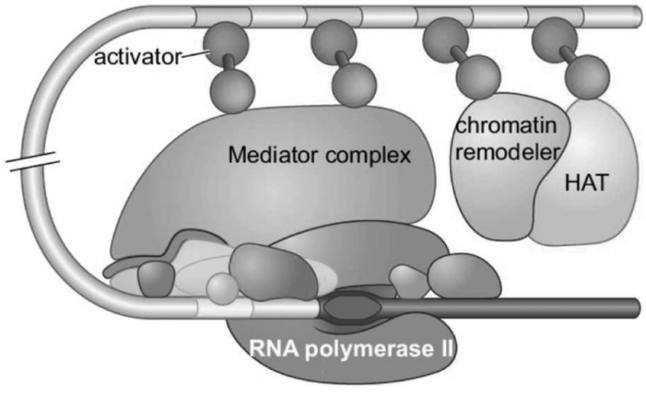


Figure 1.10: Assembly of the preinitiation complex (PIC) in promoter of eukaryotes

Transcriptional regulatory proteins called **activators** help recruit polymerase to the promoter, stabilizing its binding there. This recruitment is mediated through interactions between DNA-bound activators, chromatin-modifying and -remodeling factors, and parts of the transcription machinery. One such interaction is with the Mediator complex (hence, its name). Mediator is associated with the basic transcription machinery, most likely touching the CTD "tail" of the large polymerase subunit through one surface, while presenting other surfaces for interaction with DNA-bound activators.

In addition, Mediator aids initiation by regulating the CTD kinase in TFIIH. The need for nucleosome modifiers and remodelers also differs at different promoters or even at the same promoter under different circumstances. When and where required, these complexes are also typically recruited by the DNA-bound activators, or sometimes by regulatory RNAs.

1.4.2. RNA polymerase II transcription elongation

Shortly after transcription initiation, phosphorylation of RNA polymerase II by a protein kinase in TFIIH helps coordinate the processing of mRNAs as they are being transcribed. The carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II contains the sequence YSPTSPS (tyrosine-serine-proline-threonine-serine-proline-serine) tandemly repeated 26 times in yeast and 5′ times in humans. Phosphorylation of the serine in position 5 of the repeat (S5) by TFIIH serves as a signal for the binding of enzymes that cap the 5′ end of the mRNA (Figure 10).

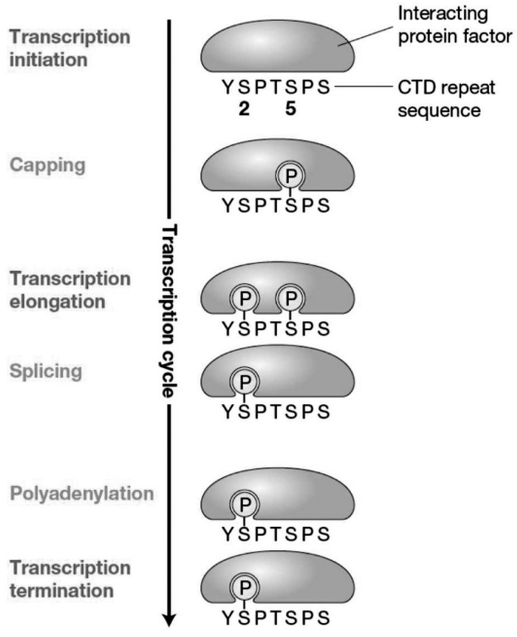


Figure 1.11: Pattern of amino acid modification in RNA polymerase II CTD

The CTD is located near the site where nascent RNA emerges from RNA polymerase II, so it is in an ideal place to orchestrate the binding and release of proteins needed to process the nascent transcript while RNA synthesis continues. Post-translational modification of S5 and other amino acids in the CTD change as RNA polymerase II transcribes through a gene, creating different binding sites for other processing factors as well as factors that regulate transcription elongation and termination. Transcription

initiation is not a green light that liberates RNA polymerase II to completely transcribe a gene unimpeded. In fact, transcription continues to be regulated all along the length of a gene. For example, for a large fraction of human genes, transcription elongation is temporarily stopped (i.e., paused) about 50 base pairs downstream of the transcription start site. Pausing of this type is caused by the protein factors NELF (negative elongation factor) and DSIF (DRB sensitivity-inducing factor) and relieved by P-TEFb (positive transcription elongation factor b). To release paused RNA polymerase II into productive elongation, P-TEFb phosphorylates NELF and DSIF. NELF dissociates from the elongation complex, and DSIF travels along with RNA polymerase II and functions as a positive elongation factor. P-TEFb also phosphorylates the RNA polymerase II CTD on serine 2 (S2) within the YSPTSPS repeats (see Figure 10), which serves as a signal for the binding of factors involved in processing of the pre-mRNA and transcription termination.

1.4.3. Transcription termination in eukaryotes

Transcription termination for the three RNA polymerases occurs by different mechanisms. Elongating RNA polymerase I is stopped by protein factors bound at specific DNA sequences called terminator elements and is released from DNA by other factors. In contrast, RNA polymerase III terminates elongation and dissociates from DNA after the synthesis of a poly(U) stretch, similar to factor-independent termination in bacteria.

Two models have been proposed for transcription termination by RNA polymerase II- the **torpedo model** and the **allosteric model (Figure 11)**. The models are conceptually similar to Rho-dependent and factor-independent mechanisms, respectively, in *E. coli*, but different factors are involved. Both RNA polymerase II termination models couple 3′ -end formation to termination. The 3′ ends of mRNAs are determined by cleavage of the pre-mRNA and addition of a poly(A) tail to the new 3′ end. In the torpedo termination model (Figure 10a), RNA polymerase II continues to transcribe past the site of cleavage, the pre-mRNA is cleaved, and the new 5′-monophosphorylated end that is formed is a substrate for a 5′ -to-3′ exonuclease called Xrn2, which digests the RNA one nucleotide at a time until eventually reaching RNA polymerase II and causing it to dissociate from DNA. Xrn2 is positioned to act in termination through its association with the CTD phosphorylated on serine 2.

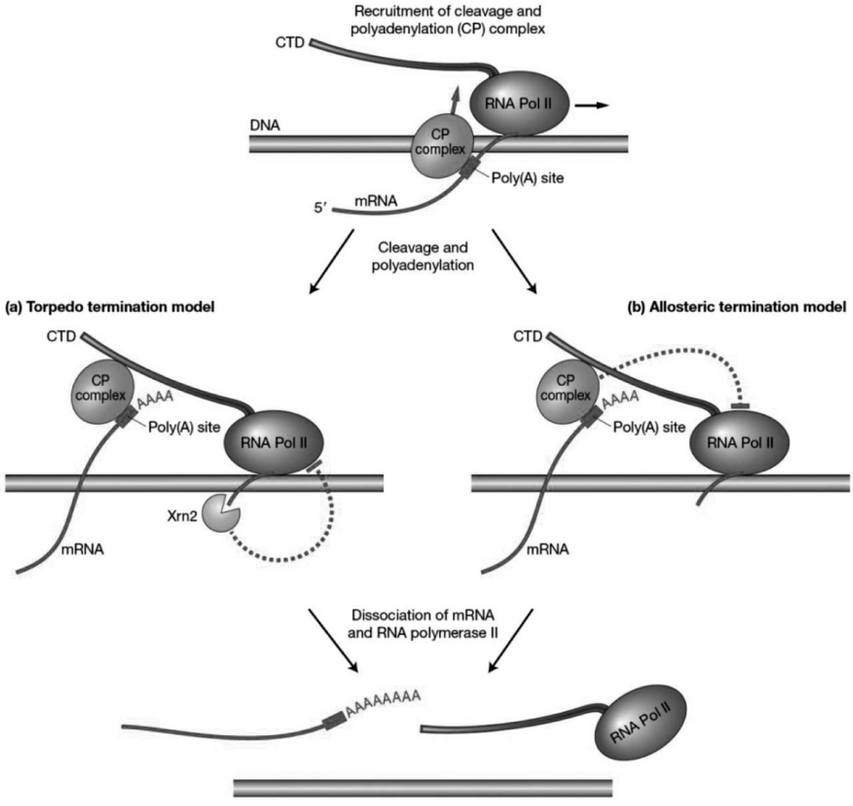


Figure 1.12: Transcription termination model in eukaryotes

In the allosteric termination model (Figure 10b), transcription through the site of cleavage causes elongation factors to dissociate, leading to a conformational change within the active site of RNA polymerase II and its release from DNA. In this model, it remains to be determined how RNA polymerase II senses passage through the site of cleavage and how this leads to dissociation of elongation factors.

2. GENE EXPRESSION REGULATION

Gene expression can be regulated at many different points in the synthesis of a functional RNA or protein. Transcription initiation is the most widely used regulatory point in both bacteria and eukaryotes, as this is the least costly way to control a gene. Initiation of transcription occurs at the very beginning of the synthetic pathway, before the investment in energy needed to make either RNA or protein. Nevertheless, mechanisms to regulate gene expression are found at virtually every point along the biosynthetic pathway. The points of regulation, shown in Figure 12, include (1) transcription initiation, (2) RNA processing, (3) RNA stability, (4) protein synthesis, (5) protein modification, (6) protein transport, and (7) protein degradation. It is obviously

important for cells to use their resources efficiently and not to waste energy synthesizing gene products that they do not need in a particular growth environment. But just as critical as efficiency is adaptability and thus control: cells must be able to respond rapidly to changes in the need for various gene products. Some regulatory mechanisms are directed at mRNA or even at the protein products of mRNA translation. Such pathways provide a means of rapidly altering the levels of active proteins in response to the cell's needs. Over the course of evolution, cells and organisms with such capability have won out over those that may have been more efficient but less able to adapt to changing conditions.

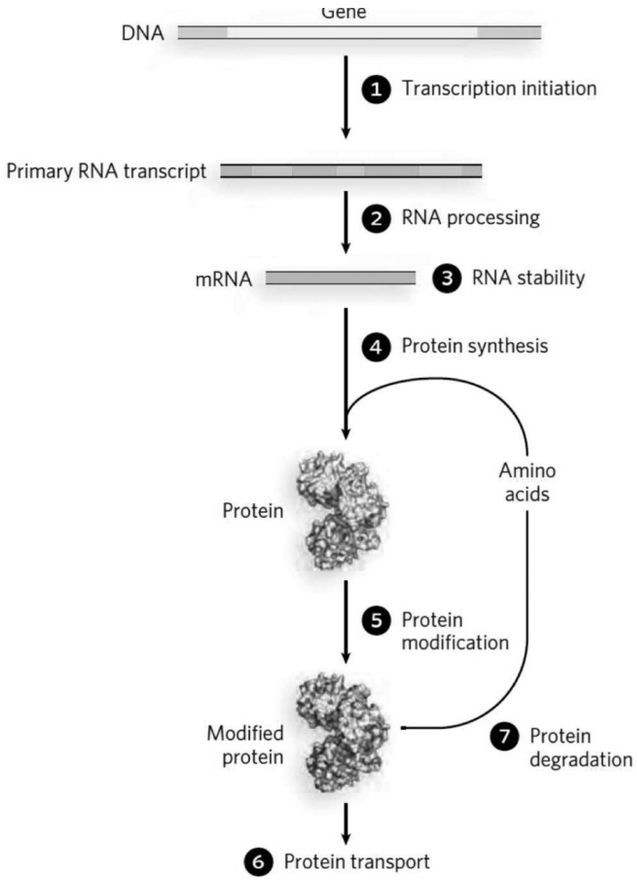


Figure 1.13. Levels of gene expression regulation

Thus, gene regulation involves a fine balance between efficiency and adaptability. New and surprising regulatory mechanisms continue to be discovered, and newly discovered posttranscriptional and translational regulatory processes are proving to be highly important, especially in eukaryotes. This chapter presents some general principles of gene regulation that are common to the mechanisms used by both bacteria and eukaryotes.

2.1. Positive and Negative Regulation through Activators and Repressors

The most basic mechanism for regulation of transcription initiation is encoded in the DNA sequence of the promoter. RNA polymerase has different intrinsic affinities for promoters of different sequence. In the absence of other controls, these differences in promoter strength correlate with the efficiency with which the genes are transcribed. Genes for products that are required at all times, such as the enzymes of central

metabolic pathways, are expressed at a nearly constant level. These genes are often referred to as housekeeping genes, and unvarying expression is called constitutive gene expression. Although housekeeping genes are expressed constitutively, the expression levels of different housekeeping genes vary widely. For these genes, the RNA polymerase–promoter interaction strongly influences the rate of transcription initiation; with differences in promoter sequences, the cell can synthesize the appropriate level of each housekeeping gene product.

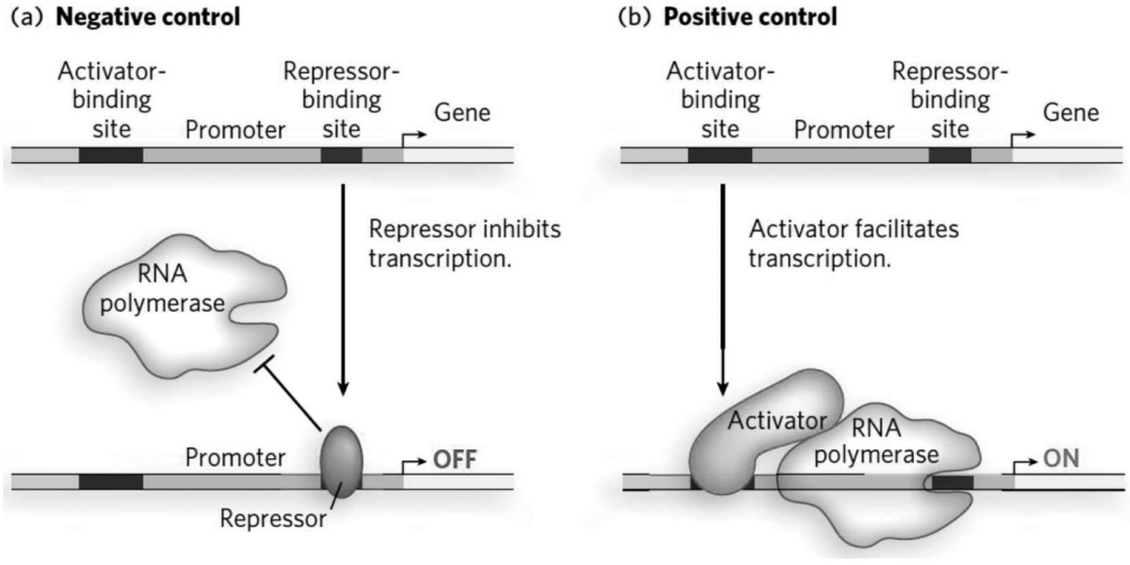


Figure 1.14(a) and (b): Negative and Positive transcriptional regulation

When the level of a gene product rises and falls with a cell's changing needs, this is known as regulated gene expression. Activation is an increase in expression and repression is a decrease in expression of a gene in response to a change in environmental conditions. The mechanisms of gene activation and repression, in both bacteria and eukaryotes, require the assistance of transcription factors (also called transcription regulators), proteins that alter the affinity of the RNA polymerase for the promoter. Transcription factors that enhance gene expression are called activators, and those that reduce expression are called repressors. regulators act by binding to specific DNA sequences known as regulatory sites. A gene is said to be under positive regulation when binding of an activator protein promotes or increases expression of that gene. Conversely, a gene is under negative regulation when binding of a repressor protein prevents or decreases expression. Thus, positive and negative regulation refer to the type of regulatory protein involved: the bound protein either facilitates or inhibits transcription (Figure 13).

2.2. Transcriptional gene expression regulation in bacteria.

The regulation of gene expression has been extensively studied in bacteria, particularly in *E. coli*. Geneticists have learned that highly efficient genetic mechanisms have evolved in these organisms to turn transcription of specific genes on and off, depending on the cell's metabolic need for the respective gene products. Not only do bacteria respond to changes in their environment, but they also regulate gene activity associated with a variety of cellular activities, including the replication, recombination, and repair of their DNA, and with cell division.

As early as 1900, it was shown that when lactose (a galactose and glucose-containing disaccharide) is present in the growth medium of yeast, the organisms synthesize enzymes required for lactose metabolism. When lactose is absent, synthesis diminishes to a basal level. Soon thereafter, investigators generalized that bacteria adapt to their environment, producing certain enzymes only when specific chemical substrates are present. These are now referred to as inducible enzymes. In contrast, enzymes that are produced continuously, regardless of the chemical makeup of the environment, are called constitutive enzymes.

More recent investigation has revealed a contrasting system, whereby the presence of a specific molecule *inhibits* gene expression. Such molecules are usually end products of anabolic biosynthetic pathways. For example, utilizing a multistep metabolic pathway, the amino acid tryptophan can be synthesized by bacterial cells. If a sufficient supply of tryptophan is present in the environment or culture medium, then it is inefficient for the organism to expend energy to synthesize the enzymes necessary for tryptophan production. A mechanism has therefore evolved whereby tryptophan plays a role in repressing the transcription of mRNA needed for producing tryptophan-synthesizing enzymes. In contrast to the inducible system controlling lactose metabolism, the system governing tryptophan expression is said to be repressible.

Regulation, whether of the inducible or repressible type, may be under either negative or positive control. Under negative control, genetic expression occurs *unless it is shut off by some form of a regulator molecule*. In contrast, under positive control, transcription occurs *only if a regulator molecule directly stimulates RNA production*. In theory, either type of control or a combination of the two can govern inducible or repressible systems.

The enzymes involved in lactose digestion and tryptophan synthesis are under negative control.

2.2.1. Operon model of Bacteria

In bacteria, genes that code for enzymes with related functions (for example, the set of genes involved with lactose metabolism) tend to be organized in clusters on the bacterial chromosome, and transcription of these genes is often under the coordinated control of a single regulatory region. Such clusters, including their adjacent regulatory sequences, are called **operons.** The location of the regulatory region is almost always upstream (5′) of the gene cluster it controls. Because the regulatory region is on the same strand as those genes, we refer to it as a ***cis*-acting site.** *Cis*-acting regulatory regions bind molecules that control transcription of the gene cluster. Such molecules are called ***trans*-acting factors.** Events at the regulatory site determine whether the genes are transcribed into mRNA and thus whether the corresponding enzymes or other protein products may be synthesized from the genetic information in the mRNA. Binding of a *trans*-acting element at a *cis*-acting site can regulate the gene cluster either negatively (by turning off transcription) or positively (by turning on transcription of genes in the cluster). In this section, we discuss how transcription of such bacterial gene clusters is co-ordinately regulated.

The discovery of a regulatory gene and a regulatory site that are part of the gene cluster was paramount to the understanding of how gene expression is controlled in the system. Neither of these regulatory elements encodes enzymes necessary for lactose metabolism—the function of the three genes in the cluster. As illustrated in Figure 14, the three structural genes and the adjacent regulatory site constitute the **lactose (*lac*) operon.** Together, the entire gene cluster functions in an integrated fashion to provide a rapid response to the presence or absence of lactose.

Genes coding for the primary structure of an enzyme are called **structural genes.** There are three structural genes in the *lac* operon. The *lacZ* gene encodes β-**galactosidase,** an enzyme whose primary role is to convert the disaccharide lactose to the monosaccharides glucose and galactose.

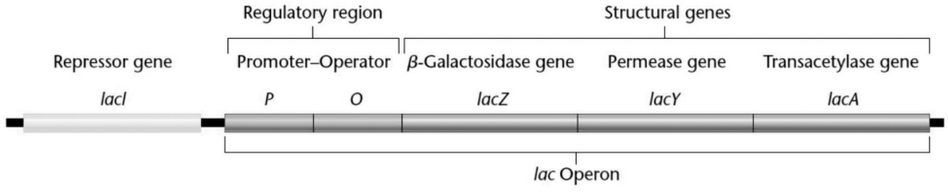


Figure 1.15: Overview of gene and regulatory units of lactose metabolism

This conversion is essential if lactose is to serve as the primary energy source in glycolysis. The second gene, *lacY,* specifies the primary structure of permease, an enzyme that facilitates the entry of lactose into the bacterial cell. The third gene,

lacA, codes for the enzyme transacetylase. While its physiological role is still not completely clear, it may be involved in the removal of toxic by-products of lactose digestion from the cell. Knowledge of their close linkage led to another discovery relevant to what later became known about the regulation of structural genes: All three genes are transcribed as a single unit, resulting in a so-called *polycistronic mRNA* (recall that *cistron* refers to the part of a nucleotide sequence coding for a single gene). This results in the coordinate regulation of all three genes, since a single-message RNA is simultaneously translated into all three gene products.

2.2.2. Negative Control of Lac Operon

Around 1960, Jacob and Monod proposed a hypothetical mechanism involving negative control that they called the **operon model,** in which a group of genes is regulated and expressed together as a unit. As we saw in Figure 14, the *lac* operon they proposed consists of the *Z, Y,* and *A* structural genes, as well as the adjacent sequences of DNA referred to as the *operator region*. They argued that the *lacI* gene regulates the transcription of the structural genes by producing a **repressor molecule** and that the repressor is **allosteric,** meaning that the molecule reversibly interacts with another molecule, undergoing both a conformational change in three-dimensional shape and a change in chemical activity. Figure 15 illustrates the components of the *lac* operon as well as the action of the *lac* repressor in the presence and absence of lactose. Jacob and Monod suggested that the repressor normally binds to the DNA sequence of the operator region. When it does so, it inhibits the action of RNA polymerase, effectively repressing the transcription of the structural genes [Figure 15]. However, when lactose is present, this sugar binds to the repressor and causes an allosteric (conformational) change.

The change alters the binding site of the repressor, rendering it incapable of interacting with operator DNA [Figure 15].

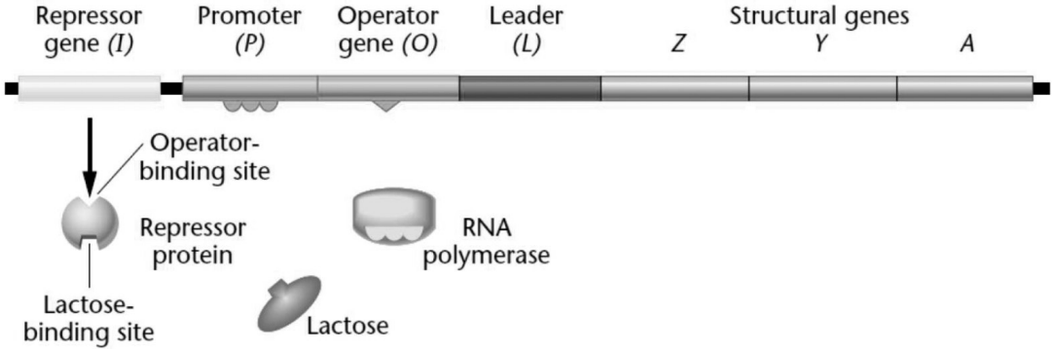


Figure 1.16: Lac operon in presence or absence of lactose

In the absence of the repressor–operator interaction, RNA polymerase transcribes the structural genes, and the enzymes necessary for lactose metabolism are produced. Because transcription occurs only when the repressor *fails* to bind to the operator region, regulation is said to be under *negative control*. To summarize, the operon model invokes a series of molecular interactions between proteins, inducers, and DNA to explain the efficient regulation of structural gene expression. In the absence of lactose, the enzymes encoded by the genes are not needed and the expression of genes encoding these enzymes is repressed. When lactose is present, it indirectly induces the activation of the genes by binding with the repressor. If all lactose is metabolized, none is available to bind to the repressor, which is again free to bind to operator DNA and to repress transcription.

2.2.3. Positive Control of Lac Operon

The role of β-galactosidase is to cleave lactose into its components, glucose and galactose. Then, to be used by the cell, the galactose, too, must be converted to glucose. Given that glucose is the preferred carbon source for *E. coli,* it would not be energetically efficient for a cell to induce transcription of the *lac* operon, since what it really needs—glucose—is already present.

Still another molecular component, called the **catabolite-activating protein (CAP),** is involved in diminishing the expression of the *lac* operon when glucose is present. This inhibition is called **catabolite repression.**

When the *lac* repressor is bound to the inducer (lactose), the *lac* operon is activated, and RNA polymerase transcribes the structural genes. transcription is initiated as a result of the binding that occurs between RNA polymerase and the nucleotide sequence of the **promoter region,** found upstream (5′) from the initial coding sequences. Within the *lac* operon, the promoter is found between the *I* gene and the operator region (*O*) (see Figure 15). Careful examination has revealed that polymerase binding is never very efficient unless CAP is also present to facilitate the process.

In the absence of glucose and under inducible conditions, CAP exerts positive control by binding to the CAP site, facilitating RNA-polymerase binding at the promoter, and thus transcription. Therefore, for maximal transcription of the structural genes to occur, the repressor must be bound by lactose (so as not to repress operon expression) *and* CAP must be bound to the CAP-binding site. In the presence of glucose still another molecule, **cyclic adenosine monophosphate. (cAMP),** upon which CAP binding is dependent. *In order to bind to the promoter, CAP must be bound to cAMP*. The level of

cAMP is itself dependent on an enzyme, **adenyl cyclase,** which catalyzes the conversion of ATP to cAMP. The role of glucose in catabolite repression is now clear. It inhibits the activity of adenyl cyclase, causing a decline in the level of cAMP in the cell. Under this condition, CAP cannot form the cAMP–CAP complex essential to the positive control of transcription of the *lac* operon (Figure 16).

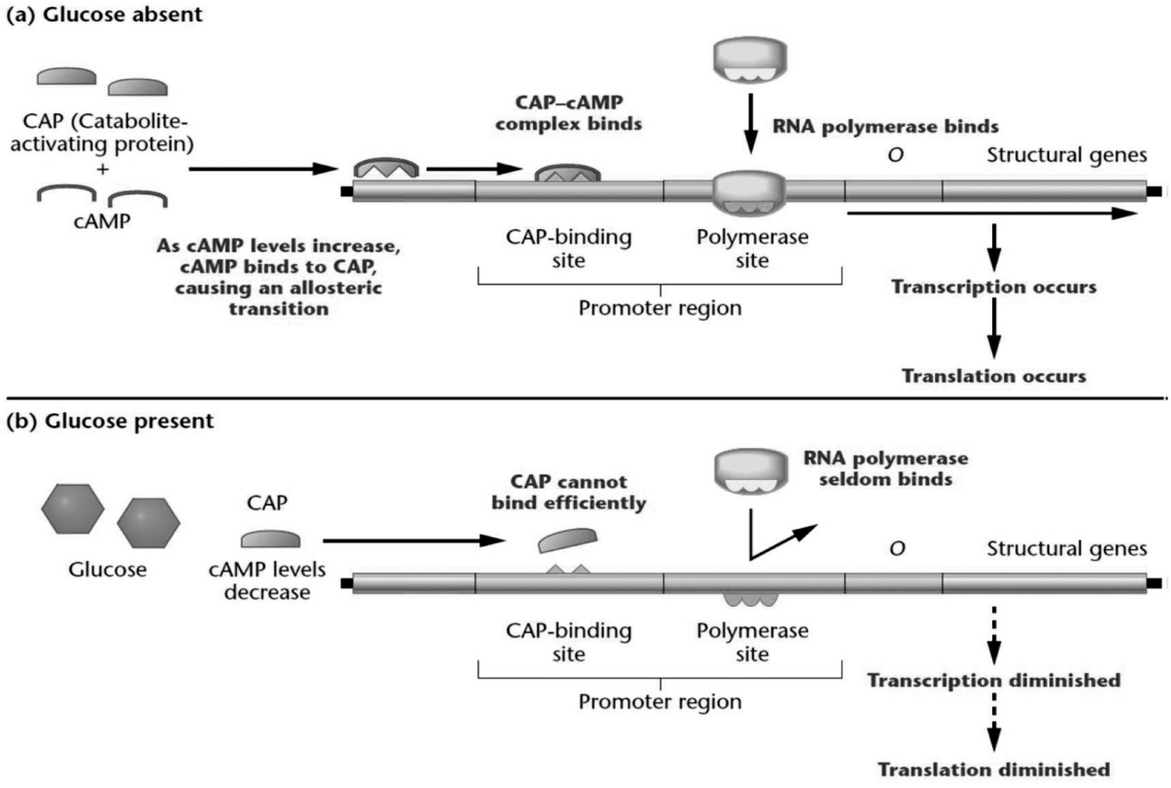


Figure 1.17. Catabolite repression and positive regulation of lac operon.

The cAMP–CAP complex, when bound to DNA, bends it, causing it to assume a new conformation. neither cAMP–CAP nor RNA polymerase has a strong tendency to bind to *lac* promoter DNA, nor does either molecule have a strong affinity for the other. However, when both are together in the presence of the *lac* promoter DNA, a tightly bound complex is formed, an example of what is called **cooperative binding.** In the case of cAMP–CAP and the *lac* operon, this phenomenon illustrates the high degree of specificity (Figure 16).

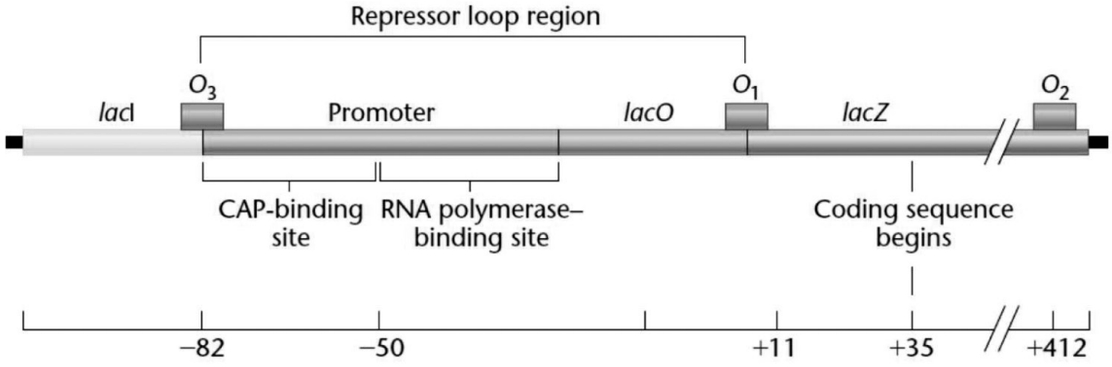


Figure 1.18. Operator of lac operon

Crystallographic studies show that the actual region of repressor binding of the primary operator, *O*1, consists of 21 base pairs. Two other auxiliary operator regions have been identified, as shown in Figure 17. One, *O*2, is 401 base pairs downstream from the primary operator, within the *lacZ* gene. The other, *O*3, is 93 base pairs upstream from *O*1, just beyond the CAP site. *In vivo*, all three operators must be bound for maximum repression. Binding by the repressor simultaneously at two operator sites distorts the conformation of DNA, causing it to bend away from the repressor. When a model is created to demonstrate dual binding of operators *O*1 and *O*3 [Figure 17], the 93 base pairs of DNA that intervene must jut out, forming what is called a **repression loop.**

2.3.1. The Tryptophan (*trp*) Operon in *E. coli*

Although the process of induction had been known for some time, it was not until 1953 that Monod and colleagues discovered a repressible operon. When grown in minimal, wild-type *E. coli* produce the enzymes necessary for the biosynthesis of amino acids as well as many other essential macromolecules. Focusing his studies on the amino acid tryptophan and the enzyme

Tryptophan synthetase, Monod discovered that if tryptophan is present in sufficient quantity in the growth medium, the enzymes necessary for its synthesis are not produced. It is energetically advantageous for bacteria to repress expression of genes involved in tryptophan synthesis when ample tryptophan is present in the growth medium. Further investigation showed that a series of enzymes encoded by five contiguous genes on the *E. coli* chromosome are involved in tryptophan synthesis. These genes are part of an operon, and in the presence of tryptophan, all are coordinately repressed and none of the enzymes are produced.

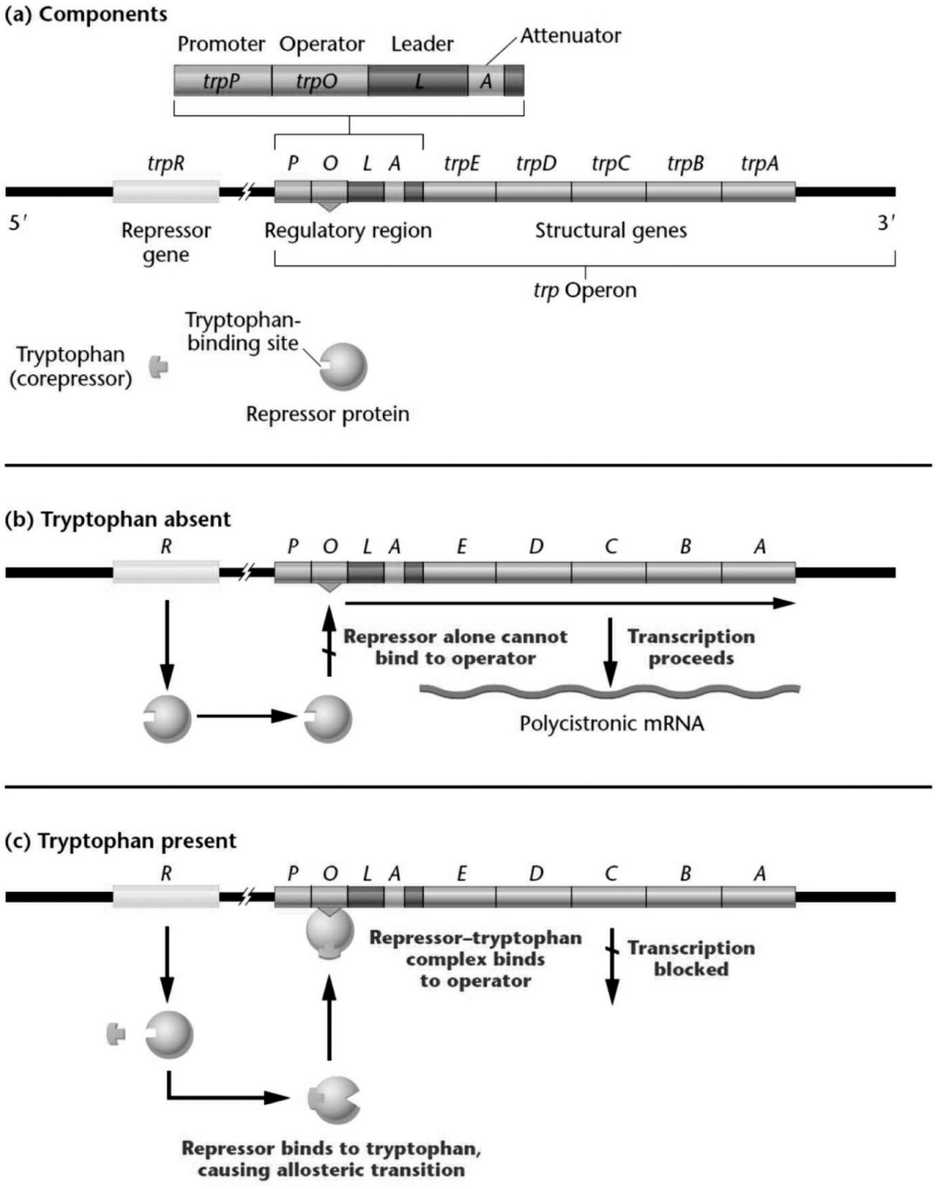


Figure 1.19: Tryptophan operon

To account for repression, Jacob and Monod suggested the presence of a *normally inactive repressor* that alone cannot interact with the operator region of the operon. However, the repressor is an allosteric molecule that can bind to tryptophan. When tryptophan is present, the resultant complex of repressor and tryptophan attains a new conformation that binds to the operator, repressing transcription (Figure 18).

Thus, when tryptophan, the end product of this anabolic pathway, is present, the system is repressed and enzymes are not made. Since the regulatory complex inhibits transcription of the operon, this repressible system is under negative control. And

as tryptophan participates in repression, it is referred to as a **corepressor** in this regulatory scheme.

The entire *trp* operon has now been well defined, as shown in Figure 18. The five contiguous structural genes (*trpE*, *D*, *C*, *B*, and *A*) are transcribed as a polycistronic message directing translation of the enzymes that catalyze the biosynthesis of tryptophan. As in the *lac* operon, a promoter region (*trpP*) represents the binding site for RNA polymerase, and an operator region (*trpO*) binds the repressor. In the absence of binding, transcription is initiated within the *trpP–trpO* region and proceeds along a **leader sequence** 162 nucleotides prior to the first structural gene (*trpE*). Within that leader sequence, still another regulatory site has been demonstrated, called an *attenuator*. This regulatory unit is an integral part of this operon's control mechanism.

2.3.2 Attenuation

Gene regulation in bacteria can also occur through the interactions of regulatory molecules with specific regions of a nascent mRNA, after transcription has been initiated. The binding of these regulatory molecules alters the secondary structure of the mRNA, leading to premature transcription termination or repression of translation. Three types of regulation involving RNA are known-*attenuation*, *riboswitches*, and *small noncoding RNAs*, abbreviated in bacteria as *sRNAs*. These types of regulation fine-tune levels of gene expression in bacteria.

Charles Yanofsky, Kevin Bertrand, and their colleagues observed that, when tryptophan is present and the *trp* operon is repressed, initiation of transcription still occurs at a low level but is subsequently terminated at a point about 140 nucleotides along the transcript. They called this process **attenuation,** as it "weakens or impairs" expression of the operon (Figure 19). In contrast, when tryptophan is absent or present in very low concentrations, transcription is initiated but is *not* subsequently terminated, instead continuing beyond the leader sequence into the structural genes. Based on these observations, Yanofsky and colleagues presented a model to explain how attenuation occurs (Figure 19). They proposed that the initial DNA sequence that is transcribed gives rise to an mRNA sequence that has the potential to fold into two mutually exclusive stem-loop structures referred to as "hairpins." If tryptophan is scarce, an mRNA secondary structure referred to as the **antiterminator hairpin** is formed. Transcription proceeds past the antiterminator hairpin region, and the entire mRNA is subsequently produced. Alternatively, in the presence of excess tryptophan, the mRNA structure that is formed is referred to as a **terminator hairpin,** and transcription is almost always terminated prematurely, just beyond the attenuator.

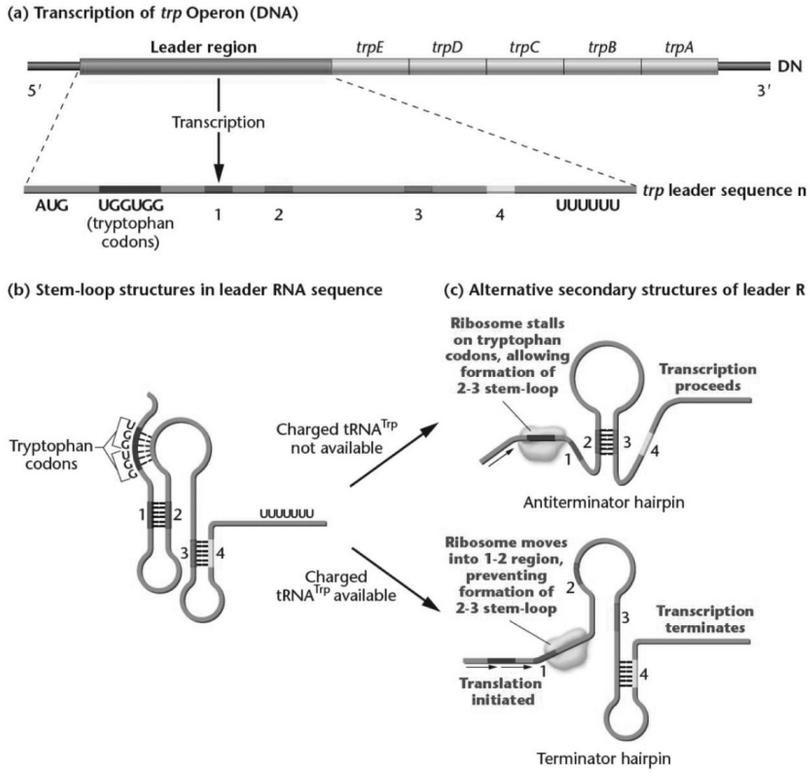


Figure 1.20: Attenuation model of Tryptophan operon

A key point in Yanofsky's model is that the transcript of the leader sequence in the 5′ untranslated region (5′-UTR) of the mRNA must be translated for the antiterminator hairpin to form. This leader transcript contains two triplets (UGG) that encode tryptophan, and these are present just downstream of the initial AUG sequence that signals the initiation of translation by ribosomes. When adequate tryptophan is present, charged tRNA-Trp is present in the cell, whereby ribosomes translate these UGG triplets, proceed through the attenuator, and allow the *terminator hairpin* to form. The terminator hairpin signals RNA polymerase to prematurely terminate transcription, and the operon is not transcribed.

If cells are starved of tryptophan, charged tRNA-Trp will be unavailable and ribosomes will "stall" during translation of the UGG triplets. The presence of ribosomes in this region of the mRNA interferes with the formation of the terminator hairpin but allows the formation of the anti-terminator hairpin within the leader transcript. As a result, transcription proceeds, leading to expression of the entire set of structural genes. Many other bacterial operons use attenuation to control gene expression. These include operons that encode enzymes involved in the biosynthesis of amino acids such as threonine, histidine, leucine, and phenylalanine. As with the *trp* operon, attenuation occurs in a leader sequence that contains an attenuator region.

3.1. Transcriptional gene expression regulation in eukaryotes

Eukaryotic cells, like bacteria, express only a subset of their genes at any given time. Through gene regulation, bacteria are able to adapt to environmental changes and respond to signaling molecules and viral assaults. Eukaryotes, too, must respond to their environment and external stimuli. But in addition, multicellular eukaryotes must manage complex pathways of cell division and differentiation that give rise to the multitude of cell types required for organismal development. Developmental programs are extremely precise—it is critical that each protein influencing cellular differentiation is active at the right time and in the right place—and any deviation from the program can have drastic consequences. Many of the genes needed for development are so critical that if mutation renders them nonfunctional, the embryo dies before the organism is fully formed. Yet, even though the needs of a eukaryote are more complex than those of a bacterium, basic principles of gene regulation are still the key to all of these processes.

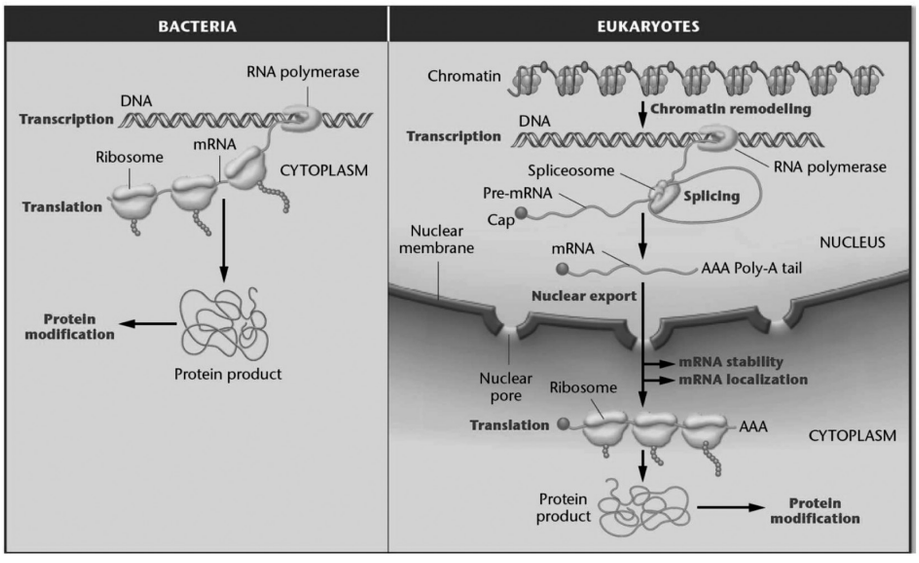


Figure 1.21: A comparison of gene regulation in bacteria and eukaryotes

Many bacterial genes and operons are regulated at the level of transcription initiation (Figure 20). This is true in eukaryotes as well, and as we'll see, many eukaryotic regulatory mechanisms build on those used in bacteria. However, there is a fundamental difference in bacterial and eukaryotic regulation of transcription. The **transcriptional ground state**, the inherent activity of promoters and transcription machinery in vivo in the absence of regulatory mechanisms, is not the same in bacteria and eukaryotes. In bacteria, the transcriptional ground state is nonrestrictive. In other words, RNA polymerase generally has access to every promoter and can bind and initiate transcription at some level of efficiency in the absence of activators or repressors. In contrast, eukaryotic genes contain strong promoters that are generally inactive in the

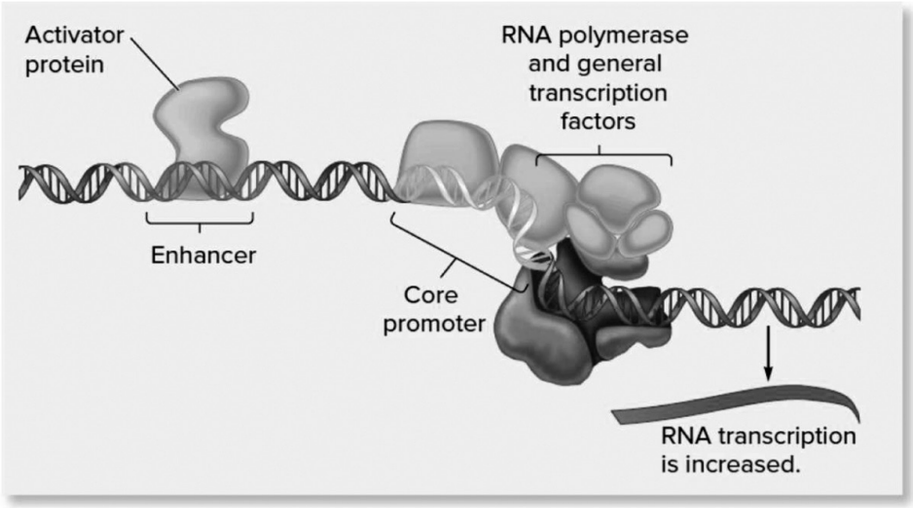
absence of regulatory proteins; that is, the transcriptional ground state in eukaryotes is restrictive.

Crucial differences in DNA packaging and cell structure give rise to at least four important distinguishing features of regulation of gene expression in eukaryotes. First, access to eukaryotic promoters is restricted by the structure of chromatin, and transcriptional activation is associated with many changes in chromatin structure in the transcribed region. Second, although eukaryotic cells have both positive and negative regulatory mechanisms, positive mechanisms predominate in all systems investigated so far; given that the transcriptional ground state is restrictive, virtually every eukaryotic gene requires activation. Third, eukaryotic cells have larger, more complex, multiprotein regulatory networks than bacteria. And finally, transcription in the nucleus is separated from translation in the cytoplasm, in both space and time. As a result, posttranscriptional control plays a larger role in controlling gene expression in eukaryote (Figure 20).

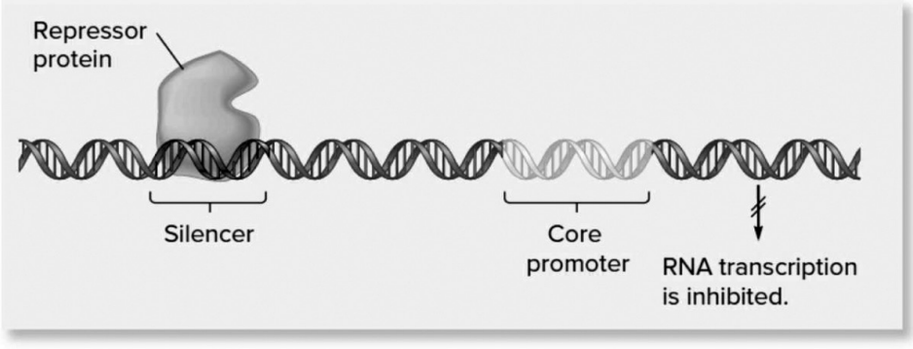
3.2. Regulatory Transcription Factor

The term **transcription factor** is broadly used to describe a category of proteins that influence the ability of RNA polymerase to transcribe a given gene. Two types 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 any transcription to occur. In addition, eukaryotic cells possess a diverse array of **regulatory transcription factors** that serve to regulate the rate of transcription of target genes. 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, regulatory elements,** or **regulatory sequences.** 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 (**Figure 21**). 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.



(a) Gene activation



(b) Gene repression

Figure 1.22: Transcriptional regulation by Regulatory Transcription factor

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. At the level of transcription, the following are factors that commonly 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 change the structure of chromatin in a way that inhibits transcription.

3.2.1. Structural Features of Regulatory Transcription Factor

In recent years, the molecular structures of transcription factor proteins have become an area of intense research. These 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, the structurally similar region is called a **motif.**

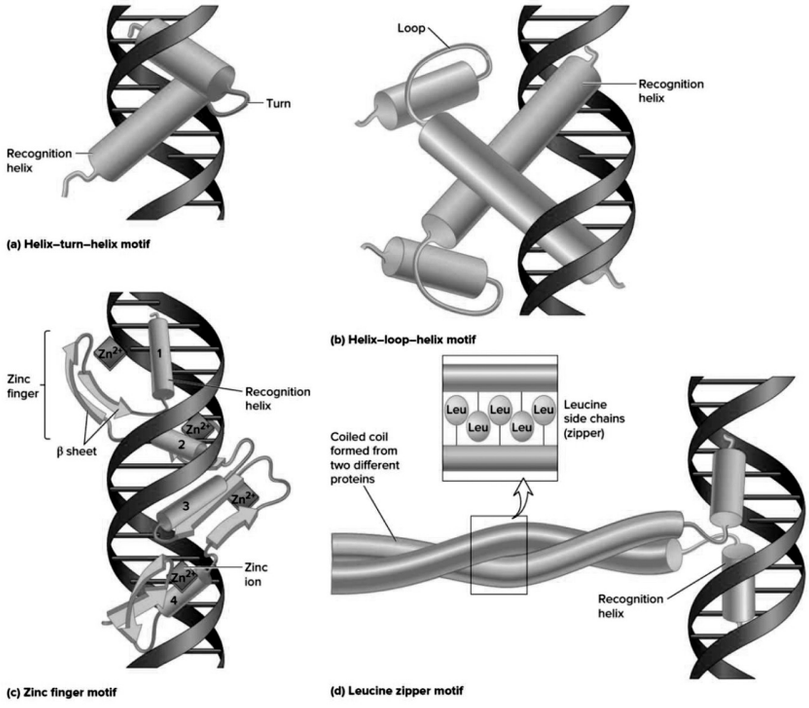


Figure 1.23: Structural motif of Transcription factors

Figure 23 depicts several different domain structures found in transcription factor proteins. The protein secondary structure known as an α helix occurs frequently in transcription factors. Why is the α helix common in such proteins? The explanation is that 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 (Figure 22a, b). Recall that the major groove is a region of the DNA double helix where the nucleotide bases are in contact with the water in cellular fluid. Hydrogen bonding between the amino acid side chains in an α helix and the nucleotide bases in the DNA is one way that a transcription factor binds 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. zinc finger motif is composed of one α helix and two β sheets that are held together by a zinc (Zn^{2+}) metal ion (Figure 22c). The zinc finger can also recognize DNA sequences within the major groove. A second interesting feature of certain motifs is that they promote protein dimerization. The leucine zipper (Figure 22d) and helix-loop-helix motif (see Figure 22b) mediate protein dimerization. For example, Figure 22d depicts the dimerization and DNA binding of two proteins that have several leucine amino acids (a zipper). Alternating leucines in both proteins interact ("zip up"), resulting in protein dimerization. Two identical transcription factor proteins may come together to form a **homodimer,** or two different transcription factors can form a **heterodimer.** The dimerization of transcription factors can be an important way to modulate their function.

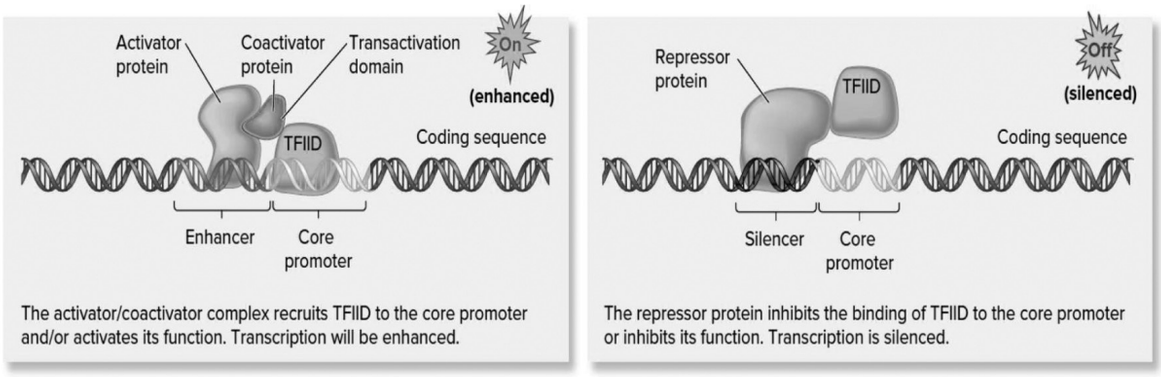
3.2.2 Enhancers and Silencers

When the binding of a regulatory transcription factor to a regulatory element increases transcription, the regulatory element is known as an **enhancer**. Such elements can stimulate transcription 10- to 1000-fold, a phenomenon known as up regulation. Alternatively, regulatory elements that serve to inhibit transcription are called **silencers**, and their action is called down regulation. Many regulatory elements are orientation-independent, or bidirectional. This means that the regulatory element functions in the forward or reverse direction. This enhancer is also bound by a regulatory transcription factor and enhances transcription even when it is rotated 180° and oriented in the reverse direction Striking variation is also found in the location of regulatory elements relative to a gene's promoter. Regulatory elements are often located in a region within 200 base pairs (bp) upstream from the core promoter. However, they can be quite distant, even 100,000 bp away, yet exert strong effects on the ability of RNA polymerase to initiate transcription at the core promoter! Regulatory elements were first discovered by Susumu Tonegawa and coworkers in the 1980s. While studying genes that play a role in immunity, these researchers identified a region that is far away from the core promoter, but is needed for high levels of transcription to take place. In some cases, regulatory elements are located

downstream from the promoter site and may even be found within introns, the noncoding parts of genes. As you may imagine, the variation in the orientation and location of regulatory elements profoundly complicates the efforts of geneticists to identify those elements that affect the expression of any given gene.

3.2.3 Regulation via TFIID

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 can enhance the ability of TFIID to initiate transcription. One possibility is that activator proteins might help TFIID bind to the TATA box, or they might enhance the function of TFIID in a way that facilitates its ability to recruit 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. This type of activation is shown in **Figure 24a**. Coactivators typically contain a **transactivation domain** that promotes the activation of RNA polymerase, often by interacting with general transcription factors. In contrast, repressors inhibit the function of TFIID. They could exert their effects by preventing the binding of TFIID to the TATA box (**Figure 24b**) 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

Figure 1.24: Regulation via TFIID

3.2.3. Regulation via Mediator

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 name **mediator** refers to the observation that this complex 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 via phosphorylation of the carboxyl-terminal domain (CTD).

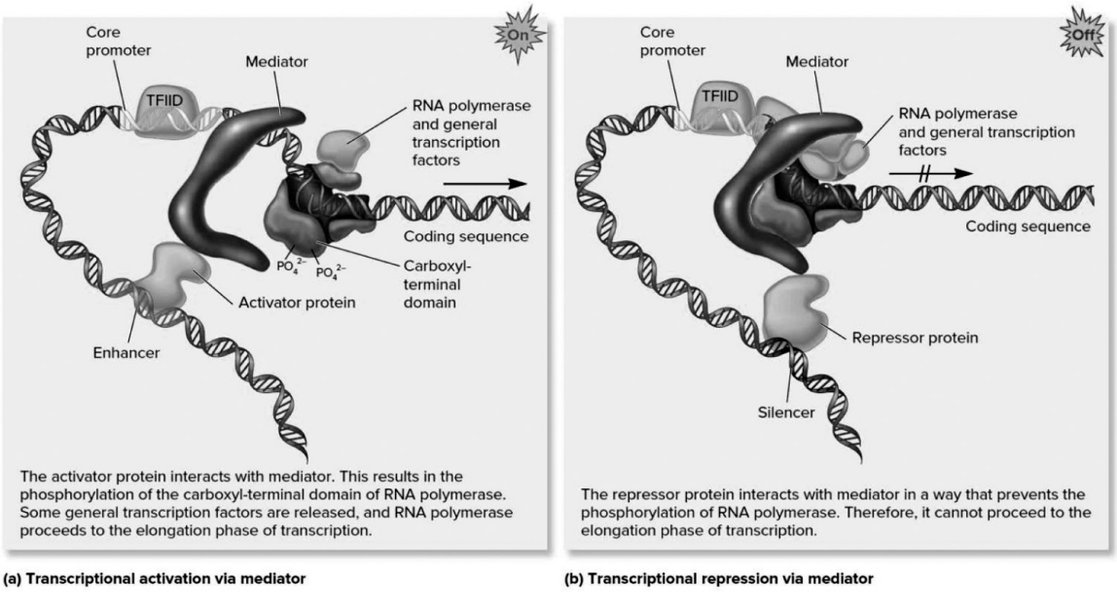


Figure 1.25: Regulation via Mediator

Transcriptional activators stimulate the ability of mediator to cause the phosphorylation of the carboxyl-terminal domain, thereby facilitating the switch between the initiation and elongation stages. In contrast, repressors have the opposite effect. In the example shown in **Figure 24a**, 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. Alternatively, a repressor protein may prevent mediator from allowing RNA polymerase to proceed to the elongation phase of transcription (Figure 24b).

3.3 Regulation via Changes in Chromatin Structure

The genomic DNA of eukaryotes wraps around small basic proteins called histones to form nucleosomes, the building blocks of chromatin (Figure 25). The transcription machinery must necessarily deal with chromatin structure in order to access particular genes. As a result, eukaryotic genes are generally expressed at low levels-or not at all-in the absence of regulatory proteins.

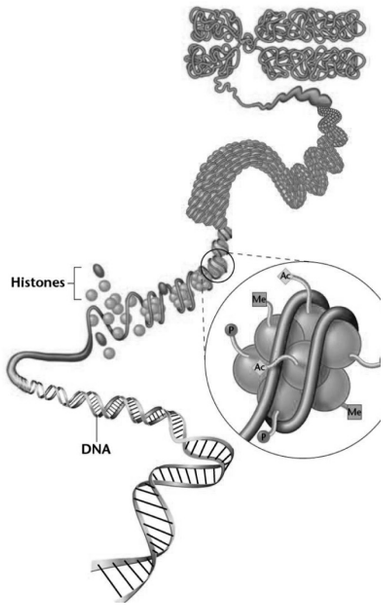


Figure 1.26: Histone in the nucleosome and the tail modification.

Chromatin structure is controlled and altered by at least three interrelated mechanisms: ATP-dependent changes in nucleosome positioning on the DNA, posttranslational chemical modifications of histone proteins, and substitution of specialized histone variants into chromatin.

Nucleosome remodeling complexes use ATP to shift nucleosomes along the DNA. Active promoters contain open regions with nucleosomes positioned away from the promoter region, allowing access to transcription factors. Chromatin remodeling involves the repositioning or removal of nucleosomes on DNA, brought about by **chromatin remodeling complexes.** Chromatin remodeling complexes are large multi-subunit enzymes that use the energy of ATP hydrolysis to move and rearrange nucleosomes. Repositioning of nucleosomes makes regions of the chromosome accessible to transcription regulatory proteins and RNA polymerase (Figure 26). Eukaryotes have multiple families of chromatin remodelers. Though their names may differ depending on the species, common families of chromatin-remodeling complexes are referred to as the SWI/SNF-family, the ISWI-family, the INO80-family, and the Mi-2-family.

The N-terminal region of each histone extends beyond the nucleosome, forming a tail. Amino acids in these tails can be covalently modified in several ways. Some posttranslational modifications of histones, including acetylation by histone acetyltransferases (HATs), result in a decondensing of chromatin and provide access to DNA-binding factors; proteins containing a bromodomain bind acetylated histones and facilitate opening of the chromatin structure.

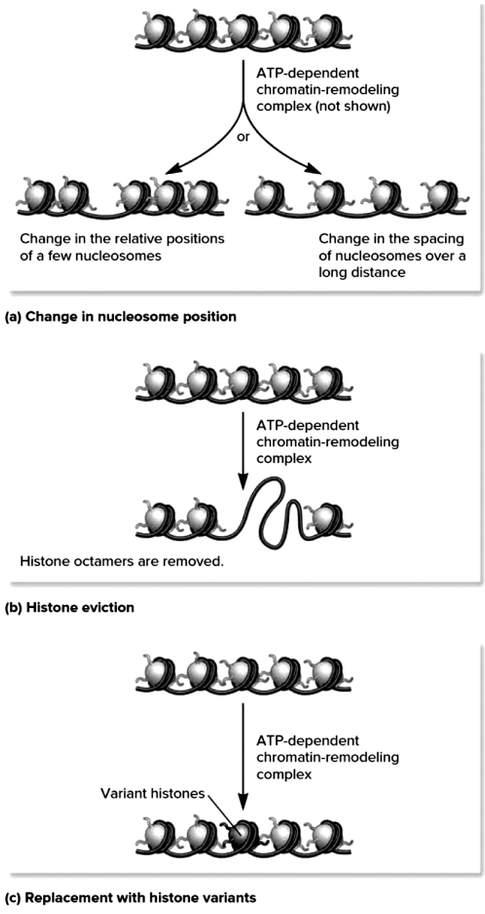


Figure 1.27. ATP-dependent Chromatin remodeling

Several sets of proteins are involved in the process. These include proteins that add chemical groups to histones ("writers"), proteins that interpret those modifications ("readers"), and proteins that remove those chemical groups ("erasers"). Some of these histone-modifying proteins are listed in Table 1.2.

Type	Number Identified
Writers	**78**
Protein methyltransferases	
Histone acetyltransferases	
Readers	**156**
Tudor domain-containing proteins	
MBT domain-containing proteins	
Chromodomain-containing proteins	
Erasers	**42**
Histone deacetylases	
Lysine demethylases	

Table 1.2: Histone modifying proteins

Over 20 different chemical modifications can be made to histones, but the major changes include the addition of acetyl, methyl, and phosphate groups (Figure 27). Such additions alter the structure of chromatin, making genes on nucleosomes with modified histones accessible or inaccessible for transcription.

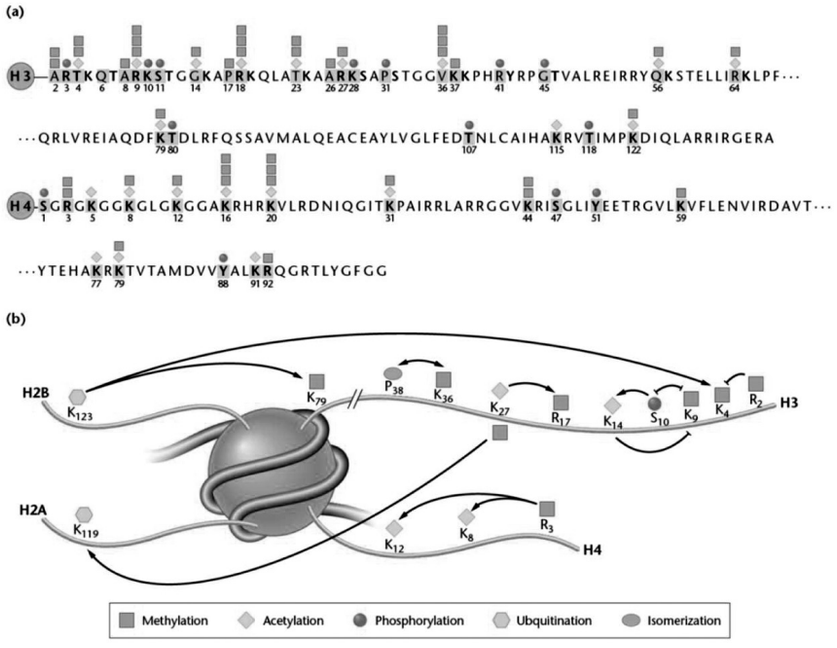


Figure 1.28: Possible N-terminal histone tail modification of H3 and H4

Histone acetylation, for example, relaxes the grip of histones on DNA and makes genes available for transcription. Furthermore, acetylation is reversible. Removing (erasing) acetyl groups contributes to changing chromatin from an "open" configuration to a "closed" state, thereby silencing genes by making them unavailable for transcription. Histone modifications occur at specific amino acids in the N-terminal tail of histones 2A, 2B, H3, and H4. Figure 27b shows some modifications commonly found on histones H3 and H4. Many combinations of histone modifications are possible within and between histone molecules, and the sum of their complex patterns and interactions is called the **histone code**. The basic idea behind a histone code is that reversible enzymatic modification of histone amino acids (by writers and erasers) recruits nucleoplasmic proteins (readers) that either further modify chromatin structure or regulate transcription. Thus, H3K27me3 represents a trimethylated lysine at position 27 from the N-terminus of histone H3. The roles of some histone modifications in regulating gene expression are shown in Table 2. Specific combinations of histone modifications and interactions between modified amino acids within and between histones control the transcriptional status of a chromatin region. For example, whether

or not H3K9 will be methylated is controlled by modifications made elsewhere on the protein. On one hand, if H3S10 is phosphorylated, methylation of the adjacent amino acid H3K9 is inhibited. On the other hand, if H3K14 is deacetylated, methylation of H3K9 is facilitated. Methylation of histones H3K4 and H3K36 is associated with transcriptional activation, while demethylation of H3K4, H3K9, and H3K27 is associated with gene repression. The histone code can be extremely complex. Considering only the addition of one, two, or three methyl groups to amino acids in H3, there are about 280 billion combinations. When all possible modifications of all histones are considered, the number of possible combinations is truly astronomical. Much work lies ahead to identify all their epigenetic roles.

Alternatively, histone modifications cause chromatin to become tightly closed to transcription. For example, methylated histones are bound by proteins containing chromodomains, and these proteins help condense the chromatin. Chromatin structure is also modulated by several histone variants. These proteins are homologous to the common histones and can take their place in nucleosomes, but they also contain amino acid extensions that have a variety of functional consequences.

3.4 DNA Methylation

Another type of chromatin modification that plays a role in gene regulation in some eukaryotes is the enzyme-mediated addition or removal of methyl groups to or from bases in DNA. DNA methylation in eukaryotes most often occurs at position 5 of cytosine (5-methylcytosine), causing the methyl group to protrude into the major groove of the DNA helix. Methylation occurs most often on the cytosine of CG doublets in DNA, usually on both strands:

5′ - ^{m}CpG - 3′

3′ - GpC^{m} - 5′

Methylatable CpG sequences are not randomly distributed throughout the genome, but tend to be concentrated in CpG-rich regions, called CpG islands, which are often located in or near promoter regions. Roughly 70 percent of human genes have a CpG island in their promoter sequence. Evidence suggests that DNA methylation represses gene expression (Figure 28). For example, large transcriptionally inert regions of the genome, such as the inactivated X chromosome in female mammalian cells, are often heavily methylated. Conversely, blocking methylation of genes on the inactivated X chromosome leads to their expression. By what mechanism might methylation affect gene regulation? Data from *in vitro* studies suggest that methylation can repress

transcription by inhibiting the binding of transcription factors to DNA. Methylated DNA may also recruit repressive chromatin remodeling complexes and HDACs to gene-regulatory regions It is important to know that while cytosine methylation is clearly an important mechanism for gene regulation in some eukaryotes, it is not uniformly true for all eukaryotes. For example, while DNA methylation is an important gene regulatory mechanism in humans, mice, and many plants, DNA methylation is absent in yeast and the roundworm *Caenorhabditis elegans*.

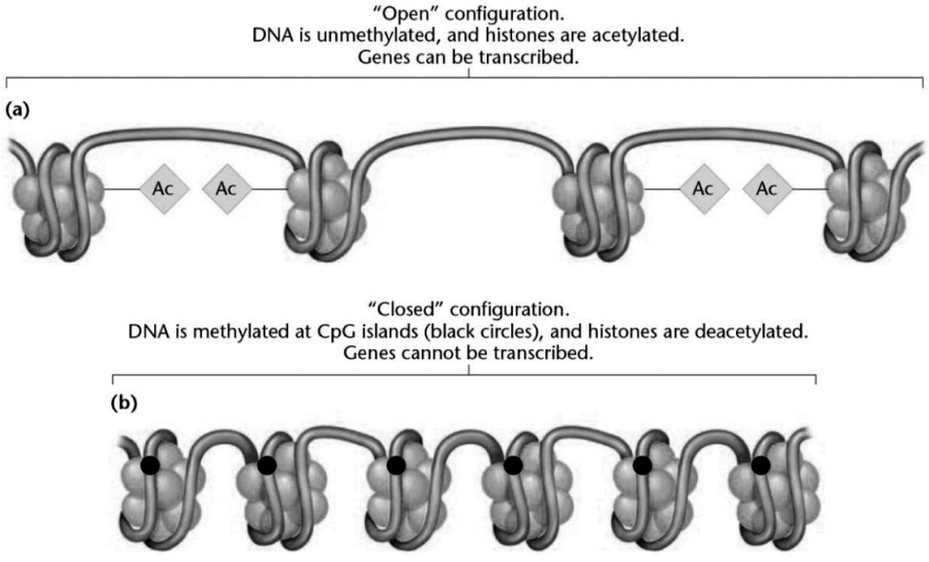


Figure 1.29: DNA methylation of CpG and chromatin structure

3.5 Noncoding RNAs

Recent advances have highlighted the roles of noncoding RNAs, including microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) in gene regulation. miRNAs (20–30 nucleotides) generally act by the RNA interference pathway to inhibit translation or induce degradation of targeted mRNAs, and gene regulation at this post-transcriptional level. Global analyses of transcription in human cells have also focused attention on the roles of lncRNAs (more than 200 nucleotides) in gene regulation. More than 50,000 lncRNAs are encoded in the human genome. Although the functions of the majority of this unexpectedly large number of lncRNAs remain to be determined, it appears that at least one important function of lncRNAs is to regulate gene expression at the transcriptional level. Many lncRNAs act as scaffolds that form complexes with proteins that modify chromatin and recruit these chromatin-modifying complexes to their target sites. A good example is provided by the *Xist* lncRNA, which mediates

X chromosome inactivation in mammals The *Xist* lncRNA binds protein complexes that include histone deacetylases, Polycomb proteins, and DNA methylases, and recruits these complexes to its target site by binding to a chromatin-associated protein (Figure). The resulting histone deacetylation, methylation of histone H3 lysine 27, and DNA methylation lead to transcriptional silencing of the inactive X chromosome. Several lncRNAs involved in genomic imprinting also act to repress their target genes by forming complexes with Polycomb proteins.

Other lncRNAs form complexes with a histone H3 lysine 9 methylase, which also induces transcriptional silencing. Importantly however, not all lncRNAs repress transcription of their target genes-some lncRNAs are associated with chromatin modifiers that activate transcription, such as histone acetylases and enzymes that methylate histone H3 lysine 4. Thus, lncRNAs can associate with a number of different chromatin-modifying enzymes and can function as either repressors or activators of their target genes.

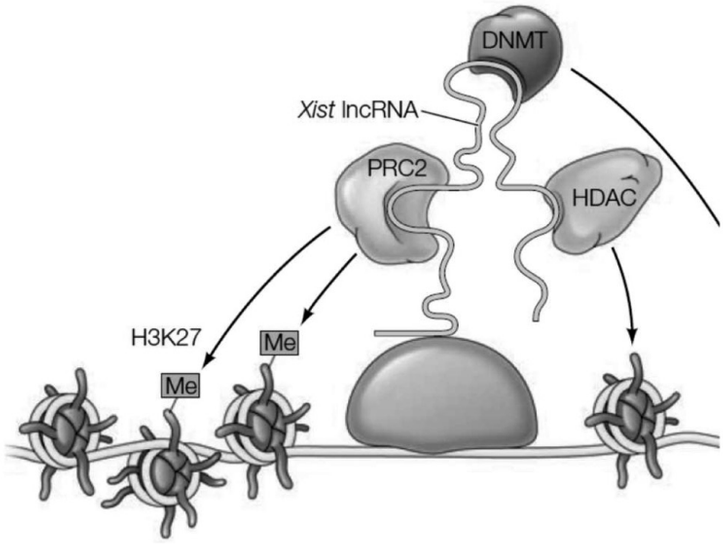


Figure 1.30: Xist lncRNA

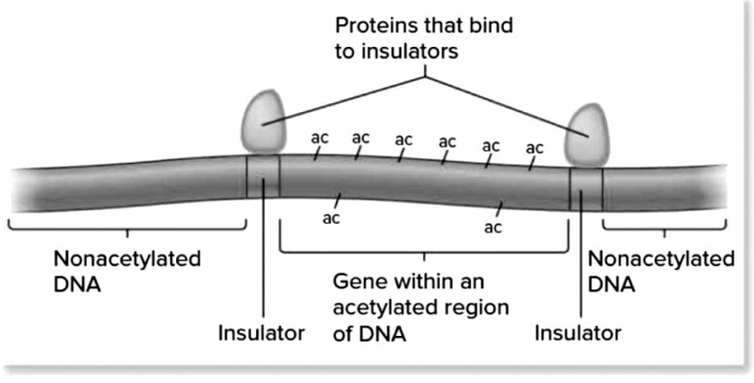
3.6 Regulations by Insulators

In eukaryotes, adjacent genes usually exhibit different patterns of gene regulation. Because eukaryotic gene regulation can occur over long distances, a bewildering aspect of such regulation is its ability to control a particular gene but not affect neighbouring genes. An **insulator** is a segment of DNA that functions as a boundary between two genes. An insulator is so named because it protects, or "insulates," a gene from the regulatory effects of a neighbouring gene.

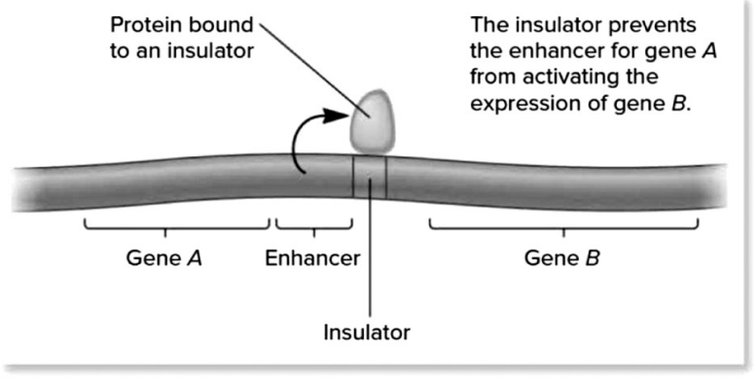
Insulators typically perform two roles. One role is to act as a barrier to chromatin-remodeling complexes or histone-modifying enzymes. As an example, let's consider the effects of histone deacetylase, which removes acetyl groups from core histone proteins, thereby favouring a closed chromatin conformation that is transcriptionally silent. Histone deacetylase may act over a long region of chromatin.

In **Figure 30a**, a gene is found in a chromosomal region in which most of the histones are not acetylated due to the action of histone deacetylase. However, this gene is flanked by two insulators that allow the region where the gene is located to be highly acetylated and transcriptionally activated. In this example, the insulators act as barriers to the action of histone deacetylase. The mechanisms that enable insulators to act as barriers are not well understood.

However, such insulators often bind proteins that recruit histone modifying enzymes or chromatin-remodeling complexes to the region. For example, an insulator could bind a protein that recruits histone acetyltransferase, which would favor the acetylation of core histones.



(a) Insulators as a barrier to changes in chromatin structure



(b) Insulator that blocks the effects of a neighboring enhancer

Figure 1.31: Role of Insulators

A second role of insulators is to block the effects of enhancers that exert their effects on neighbouring genes. As discussed earlier in this chapter, enhancers may stimulate the expression of genes that are relatively far away. To prevent an enhancer of one gene from activating the expression of an adjacent gene, an insulator may be located between them. In the example shown in **Figure 30b**, the enhancer can activate the expression of gene *A* but the protein bound to the insulator prevents the enhancer from exerting its effects on gene *B*. In other words, gene *B* is insulated from the effects of this enhancer.

Probable Questions

1. Explain the structure and function of the lac operon in *E. coli*, highlighting the mechanisms of both negative and positive transcriptional regulation.
2. Describe the stages of transcription in both prokaryotes and eukaryotes. How do initiation, elongation, and termination differ between these two domains of life?
3. Discuss the roles of general and regulatory transcription factors in eukaryotic transcription, including how they interact with promoter elements such as TATA box, Inr, DPE, and BRE.
4. What are the molecular mechanisms behind Rho-dependent and Rho-independent termination of transcription in bacteria? How do these compare with eukaryotic RNA polymerase II termination models?
5. Illustrate the function of the Mediator complex in eukaryotic transcription regulation. How does it bridge regulatory proteins and the RNA polymerase II machinery?
6. Define chromatin remodeling and histone code hypothesis. How do histone modifications influence transcriptional activation or repression in eukaryotic cells?
7. Describe the attenuation mechanism of the trp operon in *E. coli*. How does this regulatory method fine-tune gene expression in response to tryptophan levels?
8. Compare and contrast the roles of enhancers and silencers in transcriptional gene regulation. How does their position and orientation influence gene expression?
9. Explain the role of DNA methylation in eukaryotic gene regulation. Include in your discussion how CpG islands and methyl-binding proteins contribute to transcriptional silencing.
10. What is combinatorial control in eukaryotic transcriptional regulation? Provide examples of how multiple factors work together to regulate the expression of a single gene.

Suggested Readings

1. Anthony J. F. Griffiths - Introduction to Genetic Analysis
2. Lodish Harvey, Berk Arnold, Kaiser Chris et. al. - Molecular Cell Biology
3. Robert Brooker – Genetics Analysis and Principles
4. William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino, Darrell Killian - Concepts of Genetics
5. James D. Watson, Tania A. Baker, Stephen P. Bell Molecular Biology of Gene
6. Bruce Alberts, Rebecca Heald, Alexander Johnson, David Morgan, M - Molecular Biology of the Cell
7. Janet Iwasa, Gerald Karp, Wallace Marshall - Cell biology
8. David L. Nelson, Michael M. Cox, Aaron A. Hoskins - Lehninger Principles of Biochemistry

Unit-II

Post-transcriptional gene control: Regulation of Pre-mRNA Processing; Splicing, Types of introns and their splicing, evolution of introns, catalytic RNA.

INTRODUCTION

Most genes are regulated at the first step in gene expression, transcription, by regulation of the assembly of the transcription preinitiation complex on a promoter DNA sequence and of transcription elongation in the promoter-proximal region. Once transcription has been initiated, synthesis of the encoded RNA requires that RNA polymerase transcribe the entire gene and not terminate transcription prematurely. Moreover, the initial primary transcripts produced from eukaryotic genes must undergo various processing reactions to yield the corresponding functional RNAs. For mRNAs, the 5′ cap structure necessary for translation must be added, introns must be spliced out of pre-mRNAs, and the 3′ end must be polyadenylated (Figure 1). Once formed in the nucleus, mature, functional RNAs are

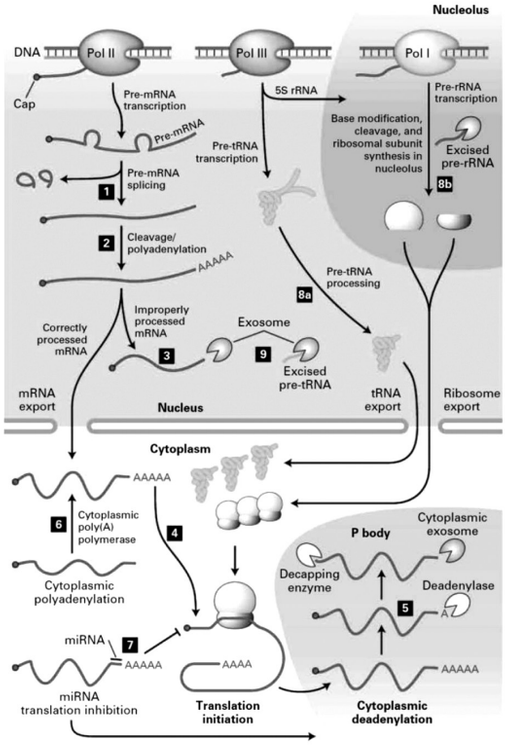


Figure 2.1: Overview of RNA Processing and Post-transcriptional gene control

exported to the cytoplasm as components of ribonucleoproteins. Both the processing of RNAs and their export from the nucleus offer opportunities for further regulation of gene expression after the initiation of transcription.

The vast amount of sequence data on human mRNAs expressed in different tissues and at various times during embryogenesis and cellular differentiation has revealed that some 95 percent of human genes give rise to alternatively spliced mRNAs. These **alternatively spliced** mRNAs encode related proteins with differences in their sequences that are limited to specific functional domains. In many cases, alternative RNA splicing is regulated to meet the need for a specific protein isoform in a specific cell type. Given the complexity of pre-mRNA splicing, it is not surprising that mistakes are occasionally made, giving rise to mRNA precursors with improperly spliced exons. However, eukaryotic cells have evolved RNA surveillance mechanisms that prevent the export of incorrectly processed RNAs to the cytoplasm or lead to their degradation if they are exported.

Additional control of gene expression can occur in the cytoplasm. In the case of protein-coding genes, for instance, the amount of protein produced depends on the stability of the corresponding mRNAs in the cytoplasm and the rate of their translation. For example, during an immune response, lymphocytes communicate by secreting polypeptide hormones called cytokines that signal neighbouring lymphocytes through cytokine receptors that span their plasma membranes. It is important for lymphocytes to synthesize and secrete cytokines in short bursts. This is possible because cytokine mRNAs are extremely unstable; consequently, the concentration of these mRNAs in the cytoplasm falls rapidly once their synthesis is stopped. In contrast, mRNAs encoding proteins required in large amounts that function over long periods, such as ribosomal proteins, are extremely stable, so that multiple polypeptides are transcribed from each mRNA.

Just as pre-mRNA processing, nuclear export, and translation are regulated, so is the cellular localization of many, if not most, mRNAs, so that newly synthesized protein is concentrated where it is needed. Particularly striking examples of this type of regulation occur in the nervous systems of multicellular animals. Some neurons in the human brain generate more than a thousand separate synapses with other neurons. During the process of learning, synapses that fire more frequently than others increase in size many times, while other synapses made by the same neuron do not.

Another type of gene regulation involves micro-RNAs (miRNAs), which regulate the translation and stability of specific target mRNAs in multicellular animals and plants. Analyses of these short miRNAs in various human tissues indicate that about 1900

miRNAs are expressed in the multiple types of human cells. Although some have recently been discovered to function through inhibition of target-gene expression in the appropriate tissue and at the appropriate time in development, the functions of the vast majority of human miRNAs are unknown and are the subject of a growing new area of research.

If most miRNAs do indeed have significant functions, miRNA genes constitute an important subset of the 25,000 or so human genes. A closely related process, called RNA interference (RNAi), leads to the degradation of viral RNAs in infected cells and the degradation of transposon-encoded RNAs in many eukaryotes. This discovery is of tremendous significance to biological researchers because it is possible to design short interfering RNAs (siRNAs) to inhibit the translation of specific mRNAs experimentally by a process called *RNA knockdown*. This method makes it possible to inhibit the function of any desired gene, even in organisms that are not amenable to classical genetic methods for isolating mutants. We refer to all the mechanisms that regulate gene expression following transcription as *post-transcriptional gene control* (Figure 1). Because the stability and translation rate of an mRNA contribute to the amount of protein expressed from a gene, these post-transcriptional processes are important components of gene control. Indeed, the protein output of a gene is regulated at every step in the life of an mRNA, from the initiation of its synthesis to its degradation. Thus, genetic regulatory processes act on RNA as well as on DNA.

1. Processing of Eukaryotic Pre-mRNA

Eukaryotic cells convert the initial primary transcript synthesized by RNA polymerase II into a functional mRNA three major events: *5′ capping*, *3′ cleavage and polyadenylation*, and *RNA splicing.* Adding these specific modifications to the 5′ and 3′ ends of the pre-mRNA protect it from enzymes that quickly digest uncapped RNAs generated by RNA processing, such as spliced-out introns and RNA transcribed downstream from a polyadenylation site. Thus the 5′ cap and 3′ poly(A) tail distinguish pre-mRNA molecules from the many other kinds of RNAs in the nucleus. Pre-mRNA molecules are bound by nuclear proteins that function in mRNA export to the cytoplasm. Prior to nuclear export, introns must be removed to generate the correct coding region of the mRNA.

The pre-mRNA processing events of capping, polyadenylation, and splicing occur in the nucleus as the nascent mRNA precursor is being transcribed. Thus, pre-mRNA processing is *co-transcriptional*. As the RNA emerges from the surface of RNA polymerase II, its 5′ end is immediately modified by the addition of the 5′

cap structure found on all mRNAs (Figure 2). As the nascent pre-mRNA continues to emerge from the surface of the polymerase, it is immediately bound by members of a complex group of RNA-binding proteins that assist in RNA splicing and export

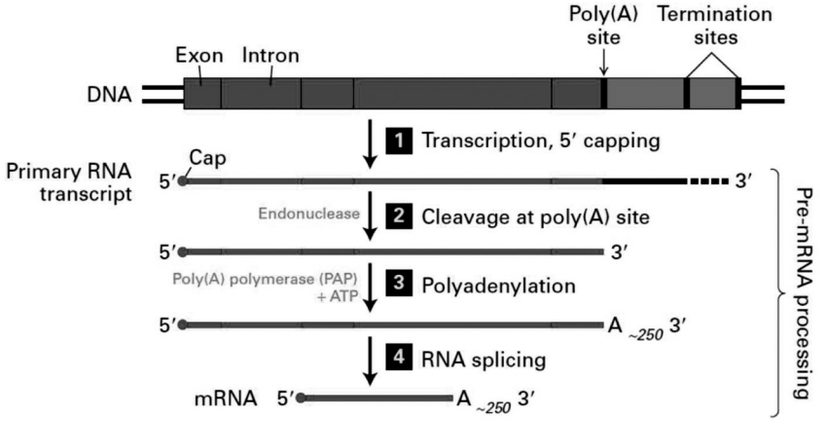


Figure 2.2: Overview of mRNA processing in eukaryotes

of the fully processed mRNA through nuclear pore complexes into the cytoplasm. Some of these proteins remain associated with the mRNA in the cytoplasm, but most either remain in the nucleus or shuttle back into the nucleus shortly after the mRNA is exported to the cytoplasm. Cytoplasmic RNA-binding proteins are exchanged for the nuclear ones. Consequently, mRNAs never occur as free RNA molecules in the cell, but are always associated with proteins as **ribonucleoprotein** (**RNP**) **complexes**, first as nascent *pre-mRNPs* that are capped and spliced as they are transcribed. Then, following cleavage and polyadenylation, they are referred to as *nuclear mRNPs*. Following the exchange of proteins that accompanies export to the cytoplasm, they are called *cytoplasmic mRNPs*.

1.1 Capping of 5′ Nascent RNAs

RNAs synthesized by RNA polymerase II, including mRNAs and snRNAs, are modified at their 5′ end by addition of a methylated guanine nucleotide, **7-methylguanosine** (**m^7G**), more commonly referred to as a **cap** because it covers the "head" of the RNA The 5′ cap is added during transcription when the RNA is about 25 nucleotides long and has just emerged from the exit channel of RNA polymerase II (Figure 5). The process of adding a cap (capping) involves the sequential action of three enzymes RNA triphosphatase removes the gamma phosphate from the first nucleotide in the RNA chain, guanylyltransferase uses GTP as a substrate and links GMP to the first nucleotide by an unusual 5′, 5′ -triphosphate linkage (GpppN, where N is the first nucleotide in the mRNA), and 7-methyltransferase adds a methyl group

(CH_3) from *S*-adenosylmethionine (SAM) to the N-7 position of the newly added guanine base. Through their interaction with the CTD of RNA polymerase II, these enzymes are in position to act on RNAs early in the transcription process (Figure 3.

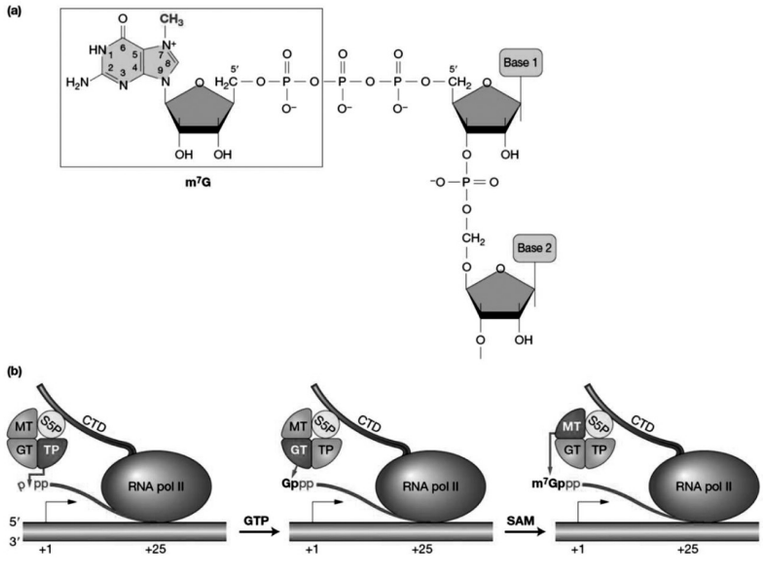


Figure 2.3. Capping of mRNA

Caps serve multiple functions. They protect RNAs from decay by exonucleases, which often require 5′-phosphates to recognize their substrates. Caps on mRNAs also serve as a binding site for proteins such as the cap binding complex (CBC) that mediate subsequent events, including splicing, polyadenylation, and nuclear export, by interacting with processing and export factors. The CBC is also critical for the first round of translation, while another cap binding protein, eIF4E, is required for subsequent rounds of translation.

1.2. Polyadenylation

Like the 5′ end, the 3′ end of an mRNA is also processed to protect it from decay and to promote translation. Processing at the 3′ end consists of two events: cleavage, which cuts the mRNA away from the transcribing RNA polymerase II, and polyadenylation, which adds 50–250 adenosine (A) residues to the end of the cleaved mRNA. Sequence elements within the 3′ UTR determine where cleavage occurs.

In humans, the highly conserved six-nucleotide (hexanucleotide) sequence AAUAAA is located 10–30 nucleotides upstream of the cleavage site, also known as the poly(A) site (Figure 8-17a). A less well conserved U-rich or GU-rich downstream sequence element (DSE) is located 20–40 nucleotides downstream of the poly(A) site. Cleavage

at the poly(A) site often occurs after a CA or UA (Figure 4). The AAUAAA is important for both cleavage and polyadenylation because it is bound by a protein complex called Cleavage and Polyadenylation Specificity Factor (CPSF), which contains the endonuclease enzyme that executes the cleavage step. CPSF also recruits poly(A) polymerase (PAP), the enzyme that uses ATP as a substrate to add a string of A's onto the 3′-OH of the mRNA, which is referred to as a poly(A) tail. PAP is an unusual RNA polymerase because, unlike DNA polymerases and other RNA polymerases, it does not copy a nucleic acid template.

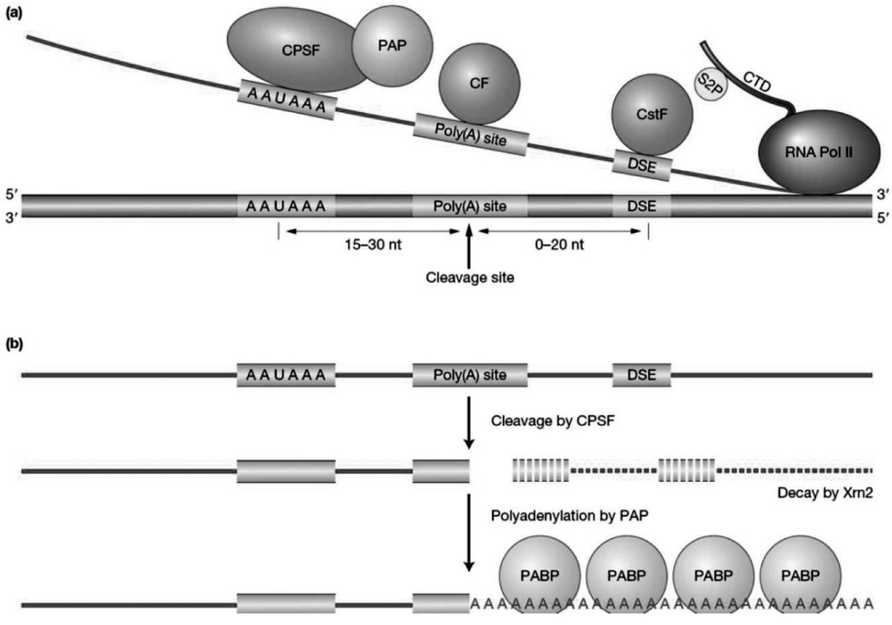


Figure 2.4: Polyadenylation of mRNA

In addition to CPSF, the cleavage step involves CstF (Cleavage Stimulatory Factor), which binds the DSE and helps determine the site of cleavage that is bound by Cleavage Factors I and II (CFI and CFII). During its synthesis, the poly(A) tail is bound by poly(A) binding protein (PABP), which in the cytoplasm protects the mRNA from decay by exonucleases and promotes translation by interacting with the translation machinery.

1.3. Splicing

Transcription copies the DNA sequence of protein-coding genes into mRNA, yet sequence comparison of most pairs of human mRNAs and genes shows that they are different: large stretches of DNA sequence are transcribed into RNA and later removed from the RNA. In 1977, the laboratories of Philip Sharp and Richard Roberts independently discovered this process of mRNA splicing, which removes segments of mRNA called introns and links together the remaining segments called exons. The

cutting out of introns and the joining of exons is called splicing because it resembles the way in which movie film is cut and rejoined to delete a specific segment.

The splicing mechanism

After the discovery of exons and introns, researchers turned their attention to the mechanism of mRNA splicing. Because splicing must occur with single nucleotide precision to maintain the information that directs translation, the intron-containing mRNA precursor (pre-mRNA) must hold the information that points the splicing machine called the **spliceosome** where to act. Researchers hypothesized that the information would be provided by sequences at the boundaries between exons and introns. In fact, alignments of boundary sequences for many pre-mRNAs revealed that almost all introns begin with GU and end with AG (Figure 5a). In addition, there is high conservation of intron and exon nucleotides adjacent to the GU and AG. The GU and AG sequence elements define the **5′ splice site** and **3′ splice site,** respectively, where cuts are made by the spliceosome to remove the intron. In addition, a third conserved sequence called the **branch point** is located 15–45 nucleotides upstream of the 3′ splice site. An invariant adenosine within the branch point participates in the first catalytic step of splicing. The existence of conserved nucleotide sequences at splice sites and the branch point suggested that components of the spliceosome are directed to act at specific places in pre-mRNAs by binding to these sequences.

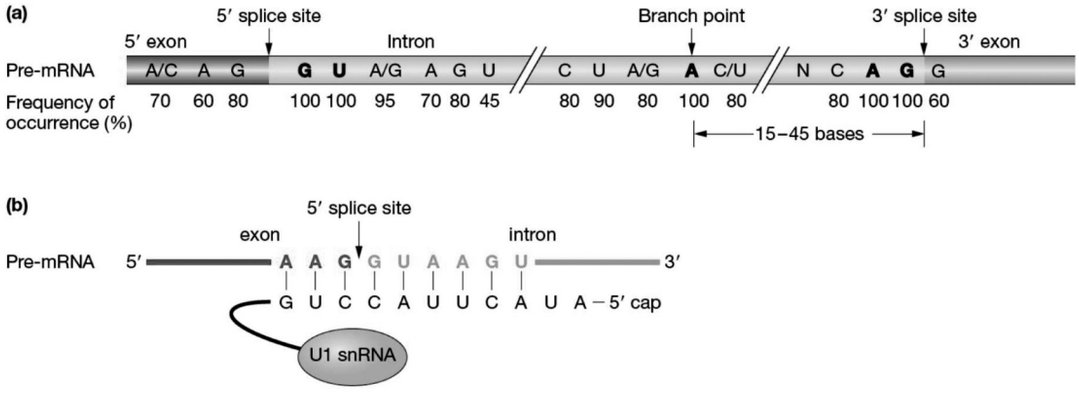


Figure 2.5: Splice Site of pre-mRNA

A nuclear RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs), pronounced “snurps,” that are comprised of a small nuclear RNA (snRNA) 100–200 nucleotides in length that serves as a scaffold for binding several proteins. The sequence at the 5′ end of the snRNA named U1 has extensive complementarity to the sequence at 5′ splice sites, suggesting that the U1 snRNA identifies 5′ splice sites by base pairing (Figure 5b).

In addition to the U1 snRNP, the spliceosome contains U2, U4, U5, and U6 snRNPs as well as many proteins that have conserved functions in eukaryotes from yeast to humans. The splicing reaction begins with stepwise recognition of pre-mRNA sequence elements (Figure 6a). First, U1 binds the 5′ splice site and U2 binds the branch point, with the U2 snRNA base pairing to nucleotides across the branch point, except for the key adenosine. Spliceosome assembly is completed by entry of the U4, U5, and U6 snRNPs as a preassembled tri-snRNP complex. At this point, the spliceosome undergoes several conformational changes to become catalytically active. The U1 and U4 snRNPs are released from the spliceosome, the U6 snRNP base pairs to the 5′ splice site, and the U5 snRNP base pairs to both exon sequences, placing the splice sites in close proximity.

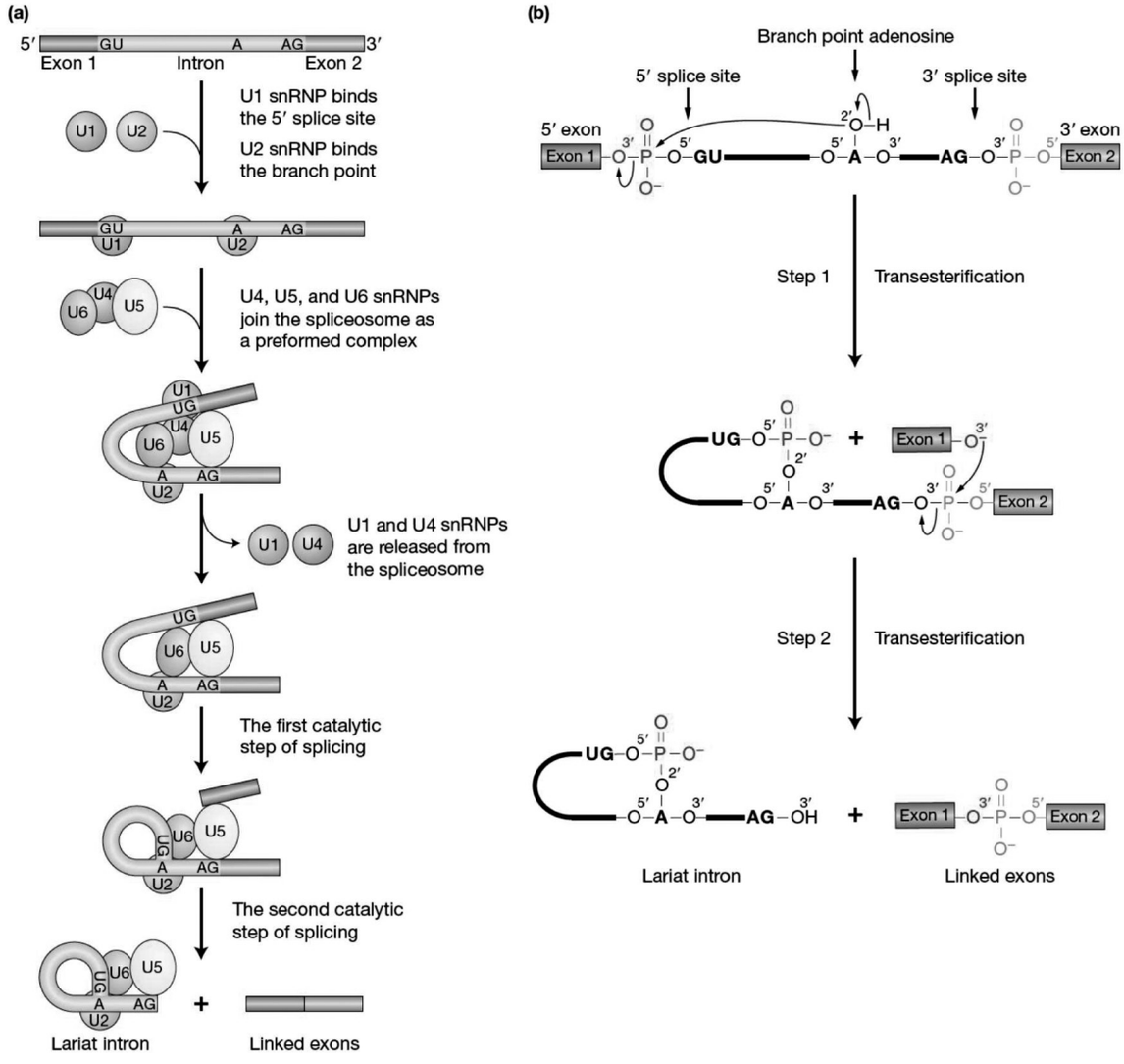


Figure 2.6: Spliceosomes and mechanism of pre-mRNA Splicing

Splicing, then, takes place by means of two transesterification reactions (Figure 6b). The first step of the reaction involves nucleophilic attack by the 2′-OH of the unpaired branch point adenosine at the phosphodiester bond at the 5′ splice site, which cuts the pre-mRNA between the 5′ exon and the intron and produces an intron with a loop structure called a lariat because it resembles the shape of a cowboy's lariat (lasso). The second step of the reaction involves nucleophilic attack by the 3′-OH (hydroxyl) of the 5′ exon at the phosphodiester bond at the 3′ splice site, which covalently links together the 5′ and 3′ exons and frees the intron as a lariat. Lastly, the U2, U5, and U6 snRNPs are released from the excised lariat and participate in another cycle of splicing along with previously released U1 and U4 snRNPs. This process is repeated for each intron in a pre-mRNA.

1.4. Types of Introns

There are several classes of introns (Table 2.1). The GT-AG (or GU-AG) introns described above are by far the most frequent in eukaryotic nuclear genes. The AT-AC (or AU-AC) introns are extremely similar to the GT-AG introns except for their different intron boundary sequences. They are processed in an almost identical manner, by a different, but closely related, set of splicing factors.

Class of Intron	Location of Genes
GT-AG (or GU-AG) introns	eukaryotic nucleus (common)
AT-AC (or AU-AC) introns	eukaryotic nucleus (rare)
Group I introns	organelles, prokaryotes (rare), rRNA in lower eukaryotes
Group II introns	organelles (of plants and fungi), some prokaryotes
Group III introns	organelles
Twintrons	organelles
Pre-tRNA introns	tRNA of eukaryotic nucleus
Archeal introns	archaebacterial tRNA and rRNA

Table 2.1: Types of Introns

Group I introns are **self-splicing**. The RNA itself provides the catalytic activity and thus acts as an RNA enzyme or ribozyme. No proteins are required for splicing. Folding of the RNA into a series of base-paired stem and loop structures is needed for ribozyme activity. The 3D structure is folded so as to bring the two splice sites together and to strain the bonds that will be broken. The reaction pathway starts with the guanosine of any of GMP, GDP, or GTP attacking the 5′ splice site (Figure. 7a) and cutting the exon and intron apart.

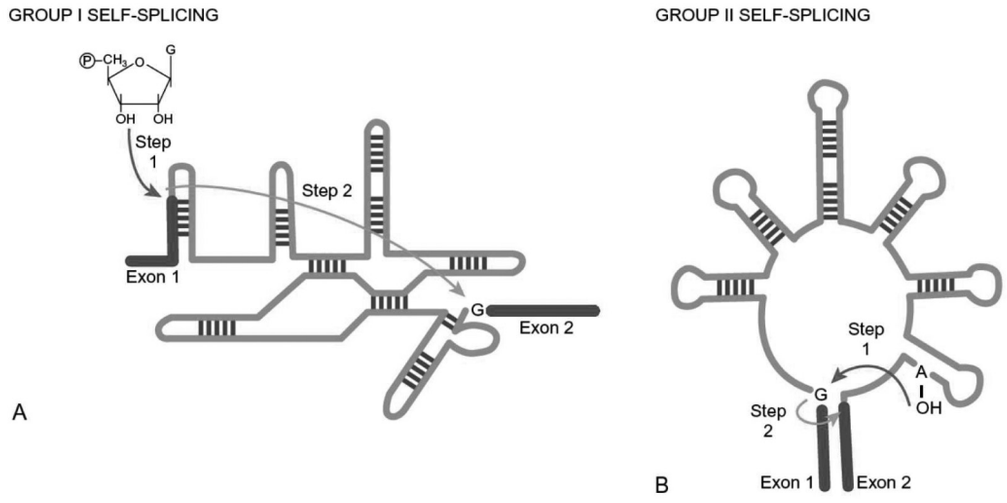


Figure 2.7: Group I and Group II Introns Splicing

Note that the guanosine nucleotide is free in solution and is not part of any RNA. The free exon-3'-OH then reacts with the downstream splice site. Group I introns include those in the rRNA of lower eukaryotes, such as the single-celled, ciliated freshwater protozoan, *Tetrahymena*. However, most are found in genes of mitochondria and chloroplasts. Occasional cases occur in bacteria and bacteriophages.

Group II introns are found in the organelles of fungi and plants and occasional examples occur in prokaryotes. Group III introns are found in organelles. Both classes are also self-splicing. However, the reaction is started by attack of an internal adenosine (not a free nucleotide as in Group I introns) (Figure. 7). This results in a lariat structure being formed, as in the typical nuclear pre-mRNA introns described above. These three types of intron may thus have a common evolutionary origin. Group III introns are similar to Group II introns, but are smaller and have a somewhat different 3D structure.

Twintrons are complex arrangements in which one intron is embedded within another. They consist of two or more Group I, Group II, or Group III introns. Since introns are embedded within other introns, they must be spliced out in the correct order, innermost first, rather like dealing with parentheses in algebra. Archeal introns are found in tRNA and rRNA and are similar in some respects to eukaryotic pre-tRNA introns. No complex splicing occurs; no snurps are needed and no ribozymes are involved. The tRNA and rRNA precursors fold up into their normal 3D structures with the intron forming a loop. This loop is cut out by a ribonuclease and the ends joined by an RNA ligase. Their stable 3D structures hold the two halves of the tRNA and rRNA molecules together during cleavage and ligation, and there is no need for extra factors such as snRNPs for recognition or processing.

1.5 Catalytic RNA

Catalytic RNAs are broadly separated into two classes based on their size and reaction mechanisms. The large catalytic RNAs consist of RNase P, and the group I and group II introns. These molecules range in size from a few hundred nucleotides to around 3000. They catalyze reactions that generate reaction intermediates and products with 3′ hydroxyls and 5′ phosphates. The small catalytic RNAs include the hammerhead, the hairpin (or paperclip), hepatitis delta and vs RNA. These molecules range in size from <“35 to <“155 nucleotides. They use the 2′ hydroxyl of the ribose sugar as a nucleophile, and they generate products with a 2′, 3′-cyclic phosphate and a 5′ hydroxyl.

The relationship between the size and the reaction mechanism of these molecules has raised intriguing questions about their origins and evolution. It may be that the reaction mechanism and the size of the large ribozymes are needed to bring often

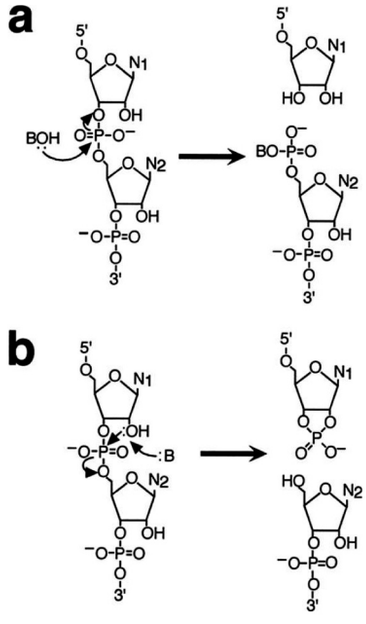


Figure 2.8: Catalytic RNA mechanism (a) large catalytic RNA (b) small catalytic RNA

very distal elements of the substrate into close proximity. The small, self-cleaving, RNAs are not faced with this constraint and perhaps this permitted them to evolve smaller catalytic centers. It remains possible, however, that the relationship between the size and reaction mechanism is simply fortuitous.

With one exception, all these RNAs catalyze reactions that modify themselves. Hence, they cannot be considered true enzymes or catalysts. The exception is RNase P, which processes the 5′ end of tRNA precursors. It is the only known example of a naturally occurring RNA-based enzyme. However, all these molecules can be converted, with some

clever engineering, into true RNA enzymes that modify other RNAs *in trans* without becoming altered themselves and sometimes termed as **Ribozyme**.

All known ribozymes have an absolute requirement for a divalent cation, which is generally Mg^{2+}. Some, notably within the large catalytic RNAs, require divalent cations for proper assembly of the tertiary structures as well. On this basis, catalytic RNAs are considered to be metalloenzymes, and a general two-metal-ion reaction mechanism has been proposed for the large catalytic RNAs, based on analogy with the properties of protein metalloenzymes. The role of divalent cations for the small catalytic RNAs is less clear, but they are generally considered to be essential for catalysis

1.6. Evolution of Introns

The origin of introns remains one of the mysteries of modern biology. Because many exons correspond to protein structural domains, it has been argued that introns are relics of early genes, in which sequences were stitched together randomly from shorter segments and only those with useful functions were maintained. According to this line of thinking, bacteria lack introns because competition for rapid growth, and consequent genome streamlining, led to intron loss from all but a few, rare bacterial and bacteriophage genes. Alternatively, introns could have arisen more recently on the evolutionary timeline. Analyses of intron positions in related genes from many different species, made possible with the arrival of online whole- genome sequence databases in the 1990s, show that in many cases, the introns and their positions within genes are not conserved. This might mean that introns were introduced relatively late in the evolution of modern genomes, or that introns are highly mobile.

The group II introns found in bacteria and in mitochondrial and chloroplast DNA are examples of mobile introns. Like retrotransposons, the introns encode proteins with both endonuclease and reverse transcriptase activities, allowing them to splice themselves back into DNA. In a transposition process termed **retrohoming**, the encoded protein forms a complex with the intron RNA after the intron is spliced from the primary transcript (Figure 9). Normally, the intron moves from one copy (allele) of a gene to an identical site in another copy of the same gene that lacks the intron. The initial insertion steps reprise the splicing mechanism, but in reverse. Once the RNA strand has been integrated into the DNA, the endonuclease cleaves the opposite DNA strand, and the inserted RNA is copied by the reverse transcriptase function associated with the endonuclease. The RNA is removed and replaced by DNA, converting the RNA intron to an inserted segment of DNA.

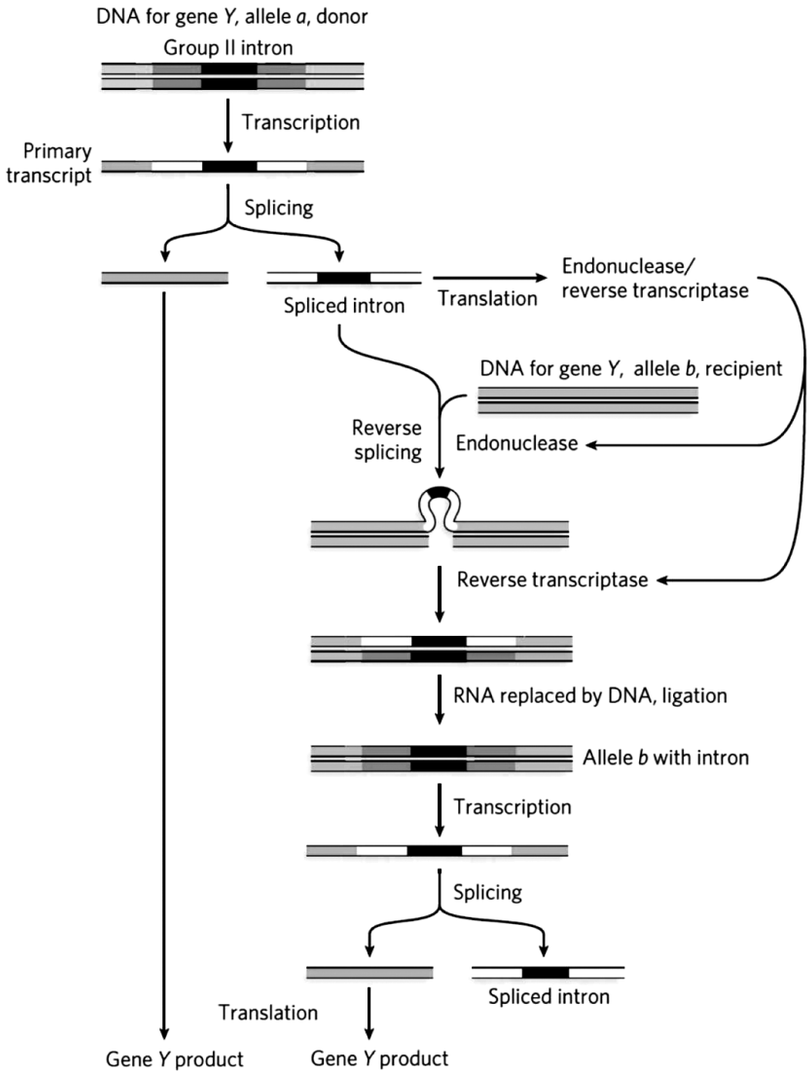


Figure 2.9: Origin of Introns

Over time, every copy of a particular gene in a population may acquire the intron. Much more rarely, the intron may insert itself into a new location in an unrelated gene. If this event does not kill the host cell, it can lead to the evolution and distribution of an intron in a new location. These mobile group II introns are thought to be the evolutionary precursors of the more widespread (and nonmobile) group II introns found in many eukaryotic genes.

Probable Questions:

1. Describe the different stages of eukaryotic pre-mRNA processing and explain how each stage contributes to mRNA stability and functionality.

2. What are the types of introns found in eukaryotic and prokaryotic genomes? Compare their splicing mechanisms with suitable examples.
3. Illustrate the mechanism of pre-mRNA splicing in eukaryotes with emphasis on splice site recognition, spliceosome assembly, and catalysis.
4. Discuss the role and significance of the 5′ cap and the 3′ poly(A) tail in mRNA metabolism and gene expression regulation.
5. What are ribozymes? Classify catalytic RNAs and explain their mechanism of action with reference to their evolutionary significance.
6. Explain how alternative splicing contributes to proteome diversity. What are the implications of this process in cellular differentiation and disease?
7. Define RNA surveillance mechanisms and discuss how they prevent the export or translation of aberrant mRNAs.
8. Describe the structure and function of microRNAs (miRNAs) and their role in post-transcriptional gene regulation. How do they differ from siRNAs?
9. Trace the evolutionary origin of introns. What evidence supports the hypothesis that group II introns are the precursors of modern spliceosomal introns?
10. How does mRNA localization contribute to spatial regulation of gene expression in multicellular organisms? Provide specific examples, particularly from neuronal systems.

Suggested Readings

1. Lodish Harvey, Berk Arnold, Kaiser Chris et. al. - Molecular Cell Biology
2. Robert Brooker – Genetics Analysis and Principles
3. William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino, Darrell Killian - Concepts of Genetics
4. James D. Watson, Tania A. Baker, Stephen P. Bell Molecular Biology of Gene
5. Bruce Alberts, Rebecca Heald, Alexander Johnson, David Morgan, M - Molecular Biology of the Cell
6. Janet Iwasa, Gerald Karp, Wallace Marshall - Cell biology
7. David L. Nelson, Michael M. Cox, Aaron A. Hoskins - Lehninger Principles of Biochemistry
8. Anthony J. F. Griffiths - Introduction to Genetic Analysis

Unit-III

Post transcriptional gene control : Alternative splicing and proteome diversity, micro-RNA and other non-coding RNAs.

1. Alternative Splicing and Proteome Diversity

RNA splicing involves the removal of noncoding introns from a pre-mRNA to form a mature mRNA. However, the pre-mRNAs of many eukaryotic genes may be spliced in alternative ways to generate different spliceforms that include or omit different exons. This process, known as **alternative splicing**, enables a single gene to encode more than one variant of its protein product. These variants, known as isoforms, differ in the amino acids encoded by differentially included or excluded exons. Isoforms of the same gene may have different functions. Even small changes to the amino acid sequence of a protein may alter the active site of an enzyme, modify the DNA-binding specificity of a transcription factor, or change the localization of a protein within the cell. Thus, alternative splicing is important for the regulation of gene expression.

1.1 Types of Alternative Splicing

There are many different ways in which a pre-mRNA may be alternatively spliced (Figure 10). One example involves **cassette exons**—such exons may be excluded from

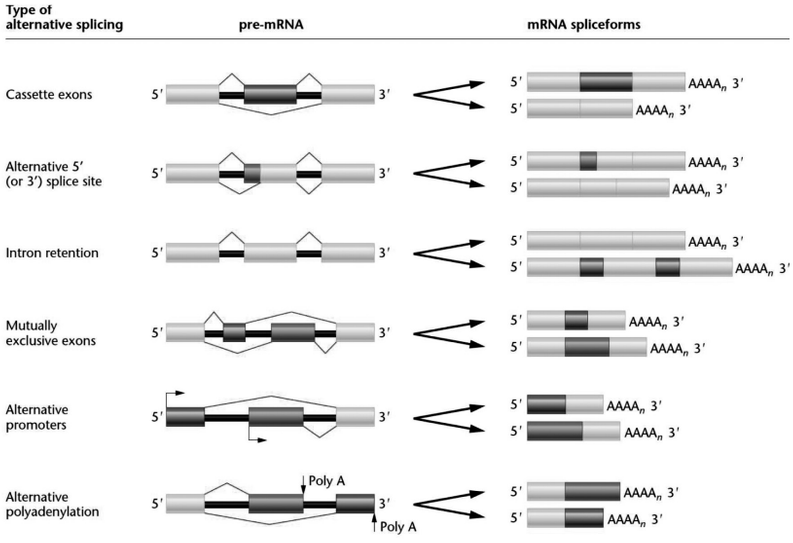


Figure 3.1: Types of alternative Splicing

the mature mRNA by joining the 3' end of the upstream exon to the 5' end of the downstream exon. Skipping of cassette exons is the most prevalent type of alternative splicing in animals, accounting for nearly 40 percent of the alternative splicing events.

In slightly over a quarter of the alternative splicing events in animals, splicing occurs at an **alternative splice site** within an exon that may be upstream or downstream of the normally used splice site. While some of these splice events are likely "noise," or errors in splice site selection by the spliceosome, as we will see below, some instances of alternative splice site usage are important regulatory events.

Intron retention is the most common type of alternative splicing event in plants, fungi, and protozoa, but is rare in mammals. In some cases, introns, which are normally noncoding sequences, are included in the mature mRNAs and are translated, producing novel isoforms. In other cases, intron retention serves to negatively regulate gene expression at the posttranscriptional level; such mRNAs are degraded or are retained in the nucleus. In rare cases, splicing is co-regulated for a cluster of two or more adjacent exons such that inclusion of one exon leads to the exclusion of the others in the same cluster. The use of these so-called **mutually exclusive exons** allows for swapping of protein domains encoded by different exons.

Pre-mRNAs with different 5' and 3' ends may be produced from the same gene due to different transcription initiation and termination sites. Some genes have **alternative promoters,** so they have more than one site where transcription may be initiated. Transcription from alternative promoters produces pre-mRNAs with different 5' exons, which may be alternatively spliced to downstream exons. Tissue-specific expression of isoforms may result from different transcription factors recognizing different promoters of a gene in different tissues. Spliceforms with different 3' ends are produced by **alternative polyadenylation.** The polyadenylation signal is a sequence that directs transcriptional termination and addition of a poly-A tail. Thus, when a polyadenylation signal is transcribed, transcription is soon terminated, and any downstream exon sequences are omitted. However, when an exon containing a polyadenylation signal is skipped, downstream exons are included, and a downstream polyadenylation signal will be used. While alternative polyadenylation may produce spliceforms with different coding sequences, it also specifies different 3' untranslated regions (UTRs) that are important for other posttranscriptional regulatory events.

1.2 Alternative Splicing and the Proteome diversity

Alternative splicing increases the number of proteins that can be made from each gene. As a result, the number of proteins that an organism can make—its **proteome**—

may greatly exceed the number of genes in the genome. Alternative splicing is found in plants, fungi, and animals but is especially common in vertebrates, including humans. Deep sequencing of RNA from human cells suggests that over 95 percent of human multi-exon genes undergo alternative splicing. While not all of these splicing events affect protein-coding sequences, it is clear that alternative splicing contributes greatly to human proteome diversity.

How many different polypeptides can be produced through alternative splicing of the same pre-mRNA? One answer to that question comes from research on the ***Dscam* gene** in *Drosophila melanogaster*. During nervous system development, neurons must accurately connect with each other. Even in *Drosophila*, with only about 250,000 neurons, this is a formidable task. Neurons have cellular processes called axons that form connections with other nerve cells. The *Drosophila Dscam* gene encodes a protein that guides axon growth, ensuring that neurons are correctly wired together. The mature *Dscam* mRNA contains 24 exons; however, the pre-mRNA includes different alternative options for exons 4, 6, 9, and 17 (Figure 11). There are 12 alternatives for exon 4; 48 alternatives for exon 6; 33 alternatives for exon 9; and 2 alternatives for exon 17. The number of possible combinations that could be formed in this way suggests that, theoretically, the *Dscam* gene can produce 38,016 different proteins.

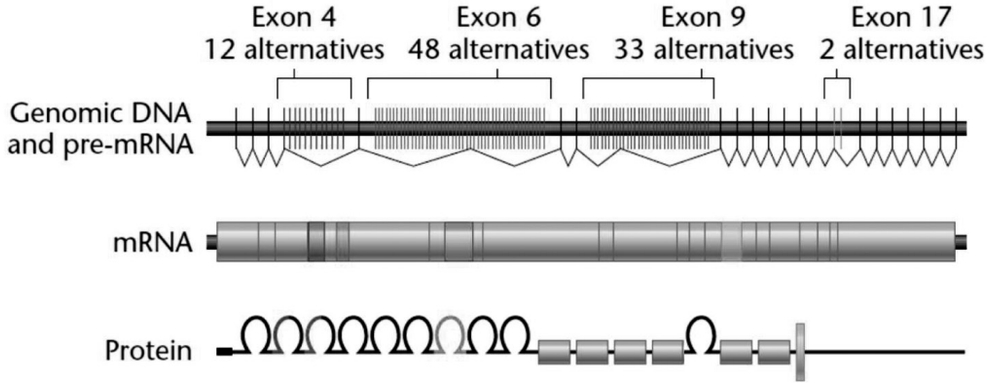


Figure 3.2: Alternative splicing of *Drosophila Dscam* gene mRNA

Although this is an impressive number of isoforms, does the *Drosophila* nervous system require all these alternatives? Recent research suggests that it does. Each neuron expresses a different subset of Dscam protein isoforms. In addition, *in vitro* studies show that each Dscam isoform binds to the same isoform, but not to others. Even a small change in amino acid sequence reduces or eliminates the binding between two Dscam molecules. *In vivo* studies show that cells expressing the same Dscam isoforms interact with each other. Therefore, it appears that the diversity of Dscam isoforms provides a molecular identity tag for each neuron, helping guide axons to the correct target and

preventing miswiring of the nervous system. The *Drosophila* genome contains about 14,000 protein coding genes, but the *Dscam* gene alone encodes 2.5 times that many proteins. Because alternative splicing is far more common in vertebrates, the suite of proteins that can be produced from the human genome may be astronomically high. A large-scale mass spectrometry study of the human proteome found that the 20,000 protein-coding genes in the human genome can produce at least 290,000 different proteins.

2. Micro RNA and other non-coding RNAs

In addition to messenger RNA (mRNA), genome transcription produces several classes of noncoding RNAs (ncRNAs) which are transcribed from DNA but not translated into proteins. The ncRNAs related to epigenetic regulation include two groups: (1) short ncRNAs (less than 31 nucleotides) and (2) long ncRNAs (greater than 200 nucleotides). Both types of ncRNAs have several roles, including the formation of heterochromatin, histone modification, site-specific DNA methylation, and gene silencing. They are also important in epigenetic regulatory networks.

There are three classes of short ncRNAs: miRNAs (microRNAs), siRNAs (short interfering RNAs), and piRNAs (piwi-interacting RNAs). miRNAs and siRNAs are transcribed as precursor molecules about 70–100 nucleotides long that contain a double-stranded stem-loop and single-stranded regions. After several processing steps that shorten the RNAs to lengths of 20–25 ribonucleotides, these RNAs act as repressors of gene expression. The origin of piRNAs is unclear, but they interact with proteins to form RNA-protein complexes that participate in epigenetic gene silencing in germ cells. Long noncoding RNAs (lncRNAs) share many properties with mRNAs; they often have 5' caps, 3' poly-A tails, and are spliced. What distinguishes lncRNAs from coding (mRNA) transcripts is the lack of an extended open reading frame that codes for the insertion of amino acids into a polypeptide. The discovery of lncRNAs was a by-product of the Human Genome Project. Genomic sequencing identified several thousand RNA genes that were not protein coding. More recent studies using RNA sequencing have identified over 14,000 lncRNA genes in the human genome. lncRNA loci are often classified by their relationship to nearby protein-coding genes. *Antisense lncRNA genes* partially overlap protein-coding genes and are transcribed in the opposite direction to the protein coding gene [Figure 12(a)]. *Intronic lncRNA genes* are located within introns, and their transcription does not overlap with the adjacent exons of protein-coding genes [Figure 12(b)]. *Bidirectional lncRNA genes* use the promoter of a protein-coding gene but are transcribed in the opposite direction [Figure 12(c)]. *Intergenic lncRNA genes* are discrete transcription units located outside protein-coding genes [Figure 12(d)].

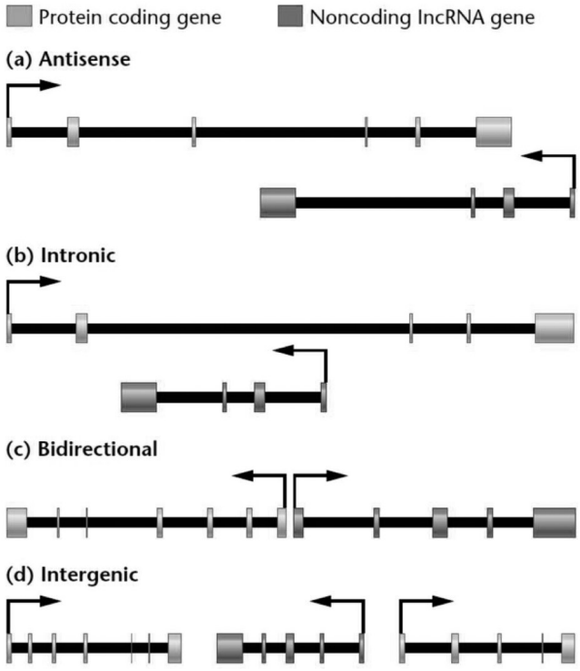


Figure 3.3: Four class of lncRNA loci.

2.1. Mechanism of gene regulation by non-coding RNA

In spite of the large number of lncRNA loci and differences in gene organization and transcriptional pattern, it is clear that lncRNAs share some common properties: (1) they form RNA-protein complexes with many different chromatin regulators, (2) they deliver these complexes to specific locations in the genome, and (3) they participate in chromatin remodeling, interact with transcription factors, and carry out other, as yet unidentified mechanisms of gene regulation.

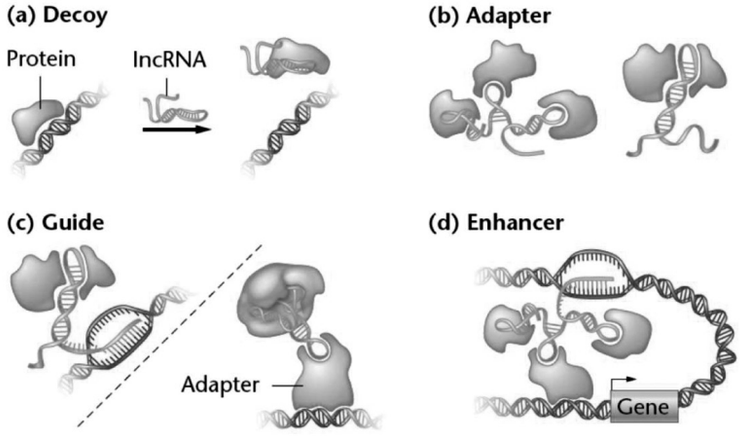


Figure 3.4. Four mechanisms of lncRNA action.

Research to date suggests that lncRNAs can interact with selected regions of the genome by several mechanisms. First, lncRNAs can act as *decoys* to bind transcription factors and prevent them from interacting with target genes [Figure 13(a)]. lncRNAs can also serve as platforms or *adapters*, to bring two or more proteins together to form a functional complex [Figure 13(b)]. For example, the lncRNA HOTAIR binds two protein complexes that allow their coordinated action to methylate H3K27 and demethylate H3K4me2, a combination of histone modifications that silences target genes.

lncRNAs can also target chromatin remodeling enzyme complexes that are involved in gene silencing [Figure 13(c)]. In this capacity, the lncRNAs serve as *guides* to target gene silencing in an allele-specific fashion, highlighting two of the basic functions of lncRNAs. More recently, it has been proposed that lncRNA guidance occurs through another mechanism via the looping of chromosomal regions in a model similar to the way *enhancers* bring proteins to the upstream regulatory sequences of a gene [Figure 13(d)].

lncRNAs are found in the nucleus and the cytoplasm and, through a variety of mechanisms, are involved in both transcriptional and post-transcriptional regulation of gene expression. As epigenetic initiators, lncRNAs bind to chromatin-modifying enzymes and direct their activity to specific regions of the genome. At these sites, the lncRNAs direct chromatin modification, altering the pattern of gene expression. In summary, epigenetic modifications alter chromatin structure by several mechanisms: DNA methylation, reversible covalent modification of histones, and action of short and long RNAs, all without changing the sequence of genomic DNA. This suite of epigenetic changes creates an epigenome that, in turn, can regulate normal development and generate changes in gene expression as a response to environmental signals.

2.2. Micro RNA and gene regulation

In the 1980s and 1990s, numerous experiments were under way, on a variety of organisms, to use DNA or RNA oligonucleotides complementary, or antisense, to an mRNA to block protein expression. The hypothesis was that base pairing between the antisense oligonucleotide and the target mRNA would prevent recognition by the translation machinery, or lead to degradation of the hybrid complex, or both. Experiments in plants, however, showed that many transgenic plants containing an artificial gene encoding an antisense RNA failed to suppress expression of the corresponding endogenous gene. Furthermore, in both plants and nematodes, “control” experiments in which the sense strand, rather than the antisense strand, of RNA was introduced into cells often showed just as much suppression of the targeted gene as did experiments using the antisense

strand. Careful analysis of these phenomena in the nematode system by Craig Mello and Andrew Fire revealed a fascinating explanation for these puzzling observations. The observed RNAi in *C. elegans* resulted from the presence of small amounts of double-stranded RNA that contaminated the preparations of sense or antisense RNA injected into the worms.

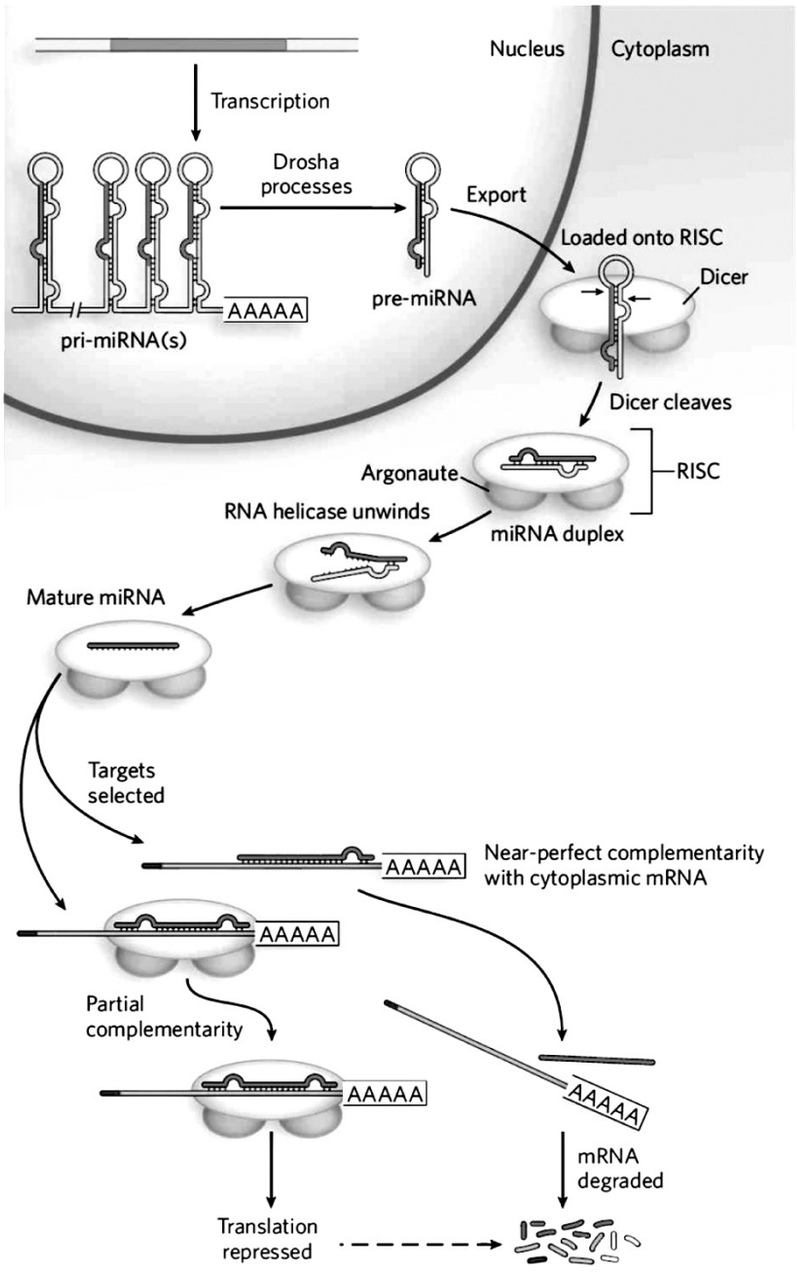


Figure 3.5. miRNAs and gene regulation

Further experiments in plants, nematodes, fruit flies, and mammals revealed many endogenous, or naturally occurring, small RNAs—**microRNAs (miRNAs)**—that correspond to sequences in cellular mRNAs. In fact, hundreds of different miRNAs have been identified in higher eukaryotes. They are transcribed by Pol II, or in some cases Pol III, as **primary miRNA transcripts (pri-miRNAs)** with one or more sets of internally complementary sequences that can fold to form hairpin-like structures. The pri-miRNAs are cleaved by the nuclear endonuclease Drosha, a member of the ribonuclease III family of enzymes, to produce shortened hairpins—60 to 70 nucleotides long—with a 5' phosphate and a two-nucleotide 3' overhang (Figure 14). These partially processed **precursor miRNAs**

(pre-miRNAs) then bind to export receptor proteins and are transported from the nucleus to the cytoplasm for further processing.

Once in the cytoplasm, pre-miRNAs are cleaved by Dicer, another ribonuclease III family member, to generate a 20- to 2′-nucleotide miRNA paired with its complementary sequence. Dicer is part of a larger complex that includes a protein called Argonaute. The overall complex is called the **RNA-induced silencing complex (RISC)**.

After cleavage, the miRNA is unwound, and the unneeded strand is discarded. The strand complementary to the target is delivered to particular mRNAs. Note that complementarity between the miRNA and the targeted mRNA is typically imperfect, with one or more mismatched or unmatched bases within the duplex. These mismatches usually occur two to eight nucleotides downstream from the 5' end of the miRNA. The nucleotides at the 5' end are called the "seed" region and must be perfectly base paired for efficient miRNA targeting. In animals, the resulting miRNA-mRNA-protein complex somehow triggers RISC to inhibit translation of the bound mRNA, through a process that probably blocks translation initiation. In plants, miRNAs typically induce RISC-mediated cleavage of the targeted mRNA, leading to subsequent degradation.

Probable Questions

1. Explain the concept of alternative splicing and discuss how it contributes to proteome diversity in eukaryotic organisms. Include specific types of alternative splicing and their significance.
2. Describe the structural and functional role of isoforms generated through alternative splicing. How can small changes in exon inclusion influence protein function?
3. Discuss the types of alternative splicing events observed in animals and plants. How do these variations affect gene regulation at the post-transcriptional level?
4. Using the Drosophila Dscam gene as an example, illustrate how alternative splicing can generate extensive protein diversity. What is the functional relevance of this diversity in neural development?
5. Compare and contrast the mechanisms of alternative promoter usage and alternative polyadenylation in generating transcript diversity. How do they contribute to tissue-specific gene expression?
6. What are non-coding RNAs (ncRNAs), and how are they classified? Discuss their roles in epigenetic regulation and chromatin remodeling.

7. Describe the major classes of long non-coding RNAs (lncRNAs) and explain their mechanisms of gene regulation with examples. How do they interact with chromatin modifiers?
8. Explain the biogenesis and mechanism of action of microRNAs (miRNAs). How do miRNAs regulate gene expression in animals versus plants?
9. Outline the process by which pri-miRNAs are processed into mature miRNAs. Highlight the roles of Drosha, Dicer, and RISC in this pathway.
10. Evaluate the functional implications of imperfect base pairing between miRNAs and their target mRNAs in animals. What is the importance of the "seed region" in miRNA targeting?

Suggested Readings

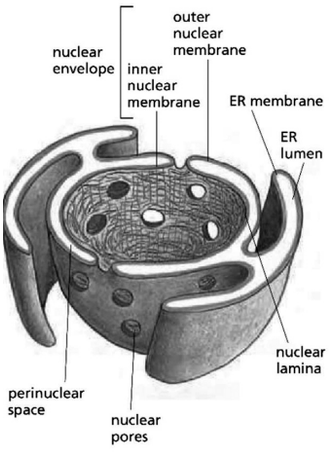
1. Anthony J. F. Griffiths - Introduction to Genetic Analysis
2. Lodish Harvey, Berk Arnold, Kaiser Chris et. al. - Molecular Cell Biology
3. Robert Brooker – Genetics Analysis and Principles
4. William S. Klug, Michael R. Cummings, Charlotte A. Spencer, Michael A. Palladino, Darrell Killian - Concepts of Genetics
5. James D. Watson, Tania A. Baker, Stephen P. Bell Molecular Biology of Gene
6. Bruce Alberts, Rebecca Heald, Alexander Johnson, David Morgan, M - Molecular Biology of the Cell
7. Janet Iwasa, Gerald Karp, Wallace Marshall - Cell biology
8. David L. Nelson, Michael M. Cox, Aaron A. Hoskins - Lehninger Principles of Biochemistry

Unit-IV

RNA Transport, Translation, and Stability of RNA - Structure of Nuclear Membrane and Nuclear Pore Complexes (NPCs), Processes of Nuclear Import and Export and Their Regulation (Including RNA Transport), Degradation of RNA

1. Structure of Nuclear Membrane and Nuclear Pore Complexes (NPCs)

The nuclear envelope (NE) is a highly specialized double-membrane structure that encapsulates the eukaryotic genome. It comprises two concentric membranes: the outer nuclear membrane (ONM) and the inner nuclear membrane (INM). The ONM is continuous with the rough endoplasmic reticulum (RER) and hosts ribosomes, whereas the INM is lined by the nuclear lamina, a dense fibrillar network composed primarily of lamin proteins. These intermediate filaments provide structural support and help anchor chromatin to the NE. The space between the ONM and INM, called the perinuclear space, is contiguous with the lumen of the ER. The NE functions not only as a physical barrier separating the nucleoplasm from the cytoplasm but also plays critical roles in genome organization, gene regulation, and mechanotransduction. Inserted at regular intervals across the NE are large protein assemblies called nuclear pore complexes (NPCs). These serve as regulated gateways for bidirectional nucleocytoplasmic transport. Each NPC is an elaborate structure with a molecular mass of ~120 MDa, composed of ~30 different nucleoporins (Nups), each present in multiple copies including scaffold proteins, channel nucleoporins, and nuclear basket filaments, which form a selective and dynamic barrier for macromolecular trafficking. It acts as a double membrane barrier separating the nucleoplasm from the cytoplasm. It contains large, cylindrical channels known as Nuclear Pore Complexes (NPCs) (~30 nm diameter), which facilitate the bidirectional transport of proteins and RNAs. While small molecules (<40–60 kDa) can diffuse freely, larger macromolecules, including RNAs, require active transport mediated by specific transporter proteins. In a typical mammalian cell, the NE contains approximately 3000–4000 NPCs, each capable of transporting ~1000 macromolecules per second in both directions.



The nuclear envelope.

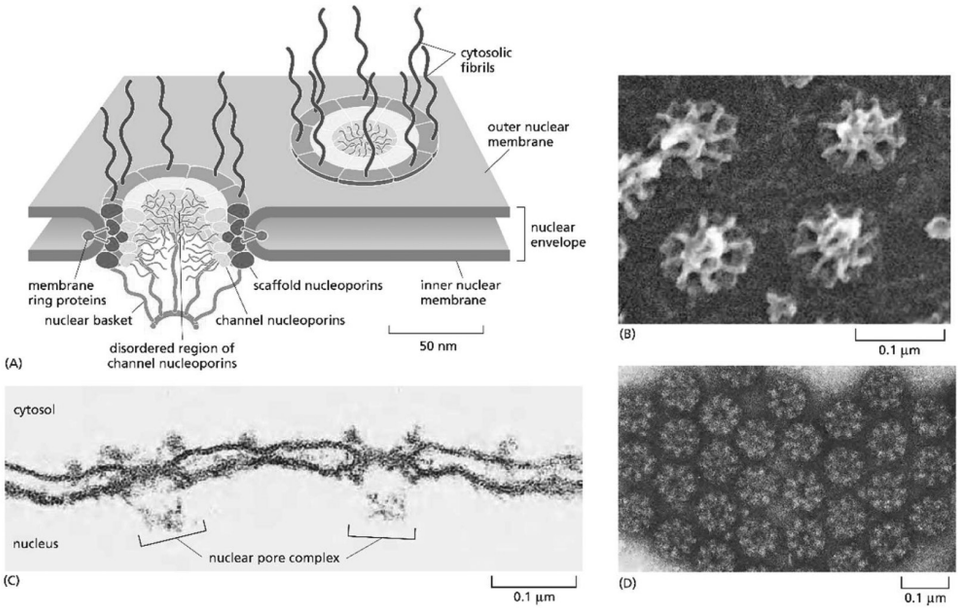


Figure 4.1

The arrangement of NPCs in the nuclear envelope. (A) In a vertebrate NPC, nucleoporins are arranged with striking eightfold rotational symmetry. In addition, immunoelectron microscopic studies show that the proteins that make up the central portion of the NPC are oriented symmetrically across the nuclear envelope, so that the nuclear and cytosolic sides look identical. The eightfold rotational and twofold transverse symmetry explains how such a huge structure can be formed from only about 30 different proteins: many of the nucleoporins are present in 8, 16, or 32 copies. Based on their approximate localization in the central portion of the NPC, nucleoporins can be classified into (1) transmembrane ring proteins that span the nuclear envelope and anchor the NPC to the envelope; (2) scaffold nucleoporins that form layered ring structures. Some scaffold nucleoporins are membrane-bending proteins that stabilize the sharp membrane curvature where the nuclear envelope is penetrated; and (3) channel nucleoporins that line a central pore. In addition to folded domains that anchor the proteins in specific places, many channel nucleoporins contain extensive unstructured regions, where the polypeptide chains are intrinsically disordered. The central pore is filled with a tangled mesh of these disordered domains that blocks the passive diffusion of large macromolecules. The disordered regions contain a large number of phenylalanine-glycine (FG) repeats. Fibrils protrude from both the cytosolic and the nuclear sides of the NPC. By contrast to the twofold transverse symmetry of the NPC core, the fibrils facing the cytosol and nucleus are different: on the nuclear side, the fibrils converge at their distal end to form a basketlike structure. The precise arrangement of individual nucleoporins in the assembled NPC is still a matter of intense debate, because atomic resolution analyses have been hindered by the sheer size and flexible nature of the NPC, and by difficulties in purifying sufficient amounts of homogeneous material. A combination of electron microscopy, computational analyses, and crystal structures of nucleoporin subcomplexes has been used to develop the current models of the NPC architecture. (B) A scanning electron micrograph of the nuclear side of the nuclear envelope of an oocyte (see also Figure 9–52). (C) An electron micrograph showing a side view of two NPCs *(brackets)*; note that the inner and outer nuclear membranes are continuous at the edges of the pore. (D) An electron micrograph showing face-on views of negatively stained NPCs. The membrane has been removed by detergent extraction. Note that some of the NPCs contain material in their center, which is thought to be trapped macromolecules in transit through these NPCs. (A, adapted from A. Hoelz, E.W. Debler and G. Blobel, *Annu. Rev. Biochem.* 80:613–643, 2011. With permission from Annual Reviews; B, from M.W. Goldberg and T.D. Allen, *J. Cell Biol.* 119:1429–1440, 1992. With permission from The Rockefeller University Press; C, courtesy of Werner Franke and Ulrich Scheer; D, courtesy of Ron Milligan.)

NPCs exhibit eightfold rotational symmetry and can be structurally divided into several distinct subcomplexes:

- **Transmembrane Nups:** These span the nuclear envelope and anchor the NPC into the membrane.
- **Scaffold Nups:** These form concentric rings that provide architectural integrity and create the central transport channel.
- **Channel Nups:** These contain phenylalanine-glycine (FG) repeats, forming a gel-like, selectively permeable matrix within the central channel.

The FG-repeat domains create an entropic barrier that permits passive diffusion of molecules <40–60 kDa while excluding larger macromolecules unless they are chaperoned by transport receptors. This meshwork resembles a "kelp bed, dynamically shifting to allow receptor-bound cargo through the central pore. The cytoplasmic face of the NPC features long fibrils extending into the cytoplasm, while the nuclear face has a basket-like structure formed by thin filaments that converge distally. These structures not only facilitate molecular trafficking but also mediate interactions with regulatory components, including components involved in mRNA surveillance and export.

2. Processes of Nuclear Import and Export and Their Regulation

a) Nuclear Import

Nuclear import primarily involves proteins containing nuclear localization signals (NLSs)—typically short sequences rich in lysine and arginine residues. These signals are recognized by importins, a family of nuclear transport receptors. The process is as follows:

1. Recognition: Importin-α binds the NLS-bearing cargo, and importin-β docks onto the NPC by interacting with FG-repeat Nups.
2. Translocation: The importin–cargo complex traverses the central NPC channel by repeatedly binding and unbinding to FG repeats in a stochastic manner.
3. Release: Once inside the nucleus, Ran-GTP binds importin-β, inducing a conformational change that releases the cargo.
4. Recycling: The importins and Ran-GTP return to the cytoplasm, where Ran-GTP is hydrolyzed by RanGAP (GTPase-activating protein), allowing the importins to engage in another round of transport.

This mechanism ensures directionality due to the asymmetric distribution of Ran-GTP (high in nucleus) and Ran-GDP (high in cytoplasm).

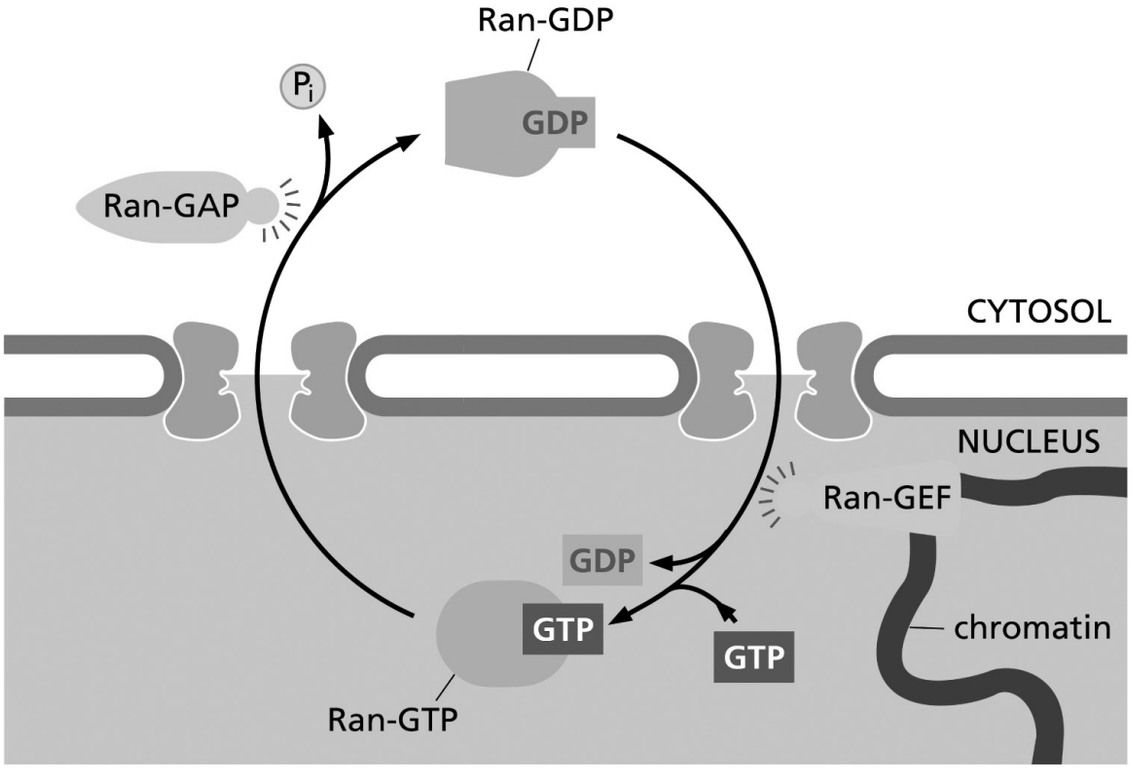


Figure 4.2: The compartmentalization of Ran-GDP and Ran-GTP.

Localization of Ran-GDP in the cytosol and Ran-GTP in the nucleus results from the localization of two Ran regulatory proteins: Ran GTPase activating protein (Ran-GAP) is located in the cytosol and Ran guanine nucleotide exchange factor (Ran-GEF) binds to chromatin and is therefore located in the nucleus. Ran-GDP is imported into the nucleus by its own import receptor, which is specific for the GDP-bound conformation of Ran. The Ran-GDP receptor is structurally unrelated to the main family of nuclear transport receptors. However, it also binds to FG- repeats in NPC channel nucleoporins.

b) Nuclear Export

Conversely, nuclear export uses nuclear export signals (NESs) and exportins to shuttle proteins and RNAs out of the nucleus. Exportins form a trimeric complex with NES-bearing cargo and Ran-GTP in the nucleus. After crossing the NPC, Ran-GTP is hydrolyzed in the cytoplasm, releasing the cargo and exportin.

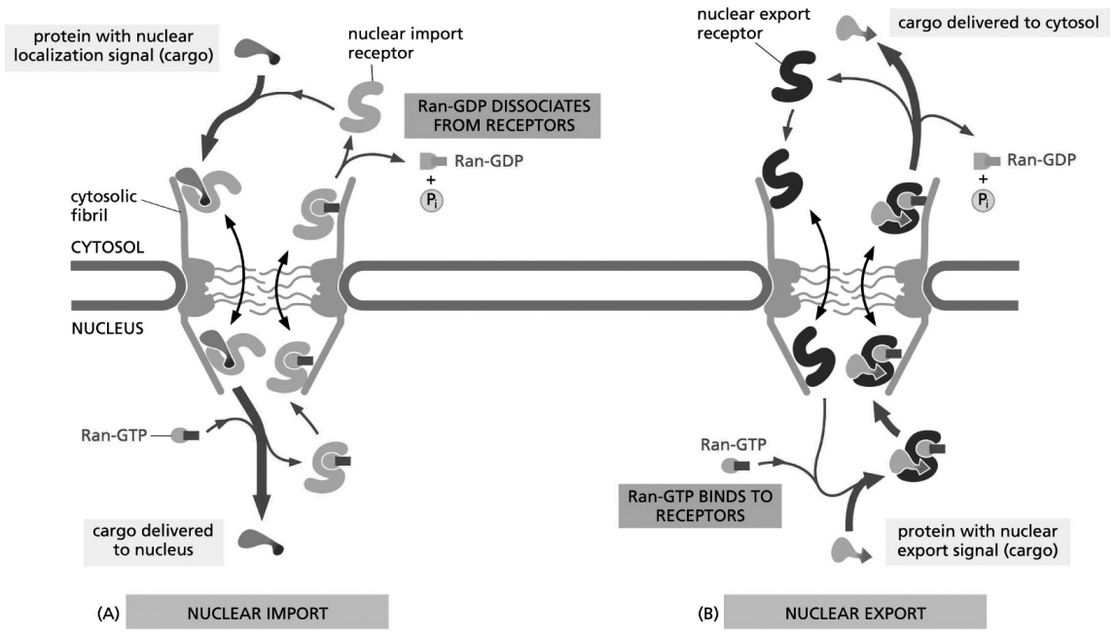
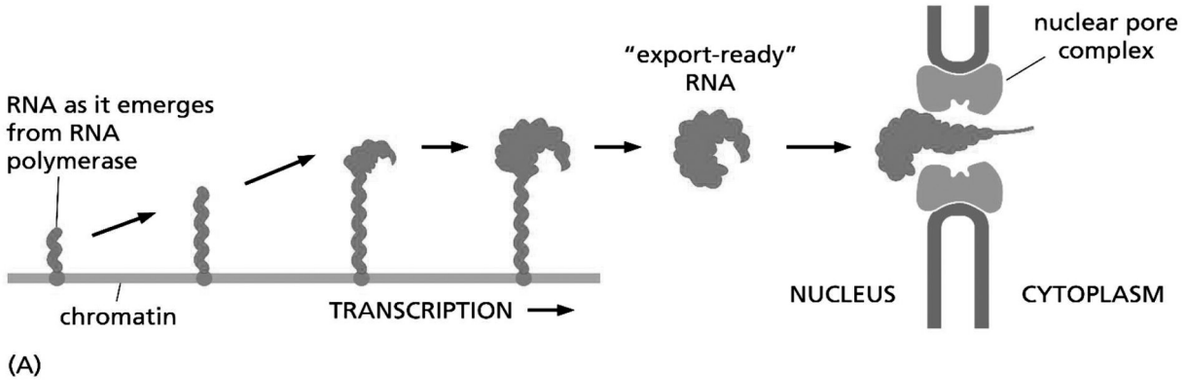


Figure 4.3

c) Introduction to RNA Transport

The transport of RNA across the nuclear envelope is a critical post-transcriptional regulatory step that determines the fate of an mRNA transcript. It ensures that only fully processed, mature messenger RNAs (mRNAs) are exported to the cytoplasm for translation, while unprocessed or aberrant transcripts are retained and degraded within the nucleus. This highly regulated process involves various classes of ribonucleoprotein (RNP) complexes and an ensemble of specialized proteins working in concert with structural elements like the nuclear pore complex (NPC).



Transport of a large mRNA molecule through the nuclear pore complex.

Figure 4.4

d) Export of mRNA Through the NPC

Unlike protein export, mRNA export is Ran-independent and involves a specialized export machinery. Fully processed mRNAs in the nucleus are packaged into large complexes called messenger Ribonucleoprotein particles (mRNPs). These include:

- **hnRNP (heterogeneous nuclear RNP) proteins**
- **Exon Junction Complexes (EJCs)**
- **RNA Export Factors (REF)**
- **SR proteins (Serine/Arginine-rich proteins)**

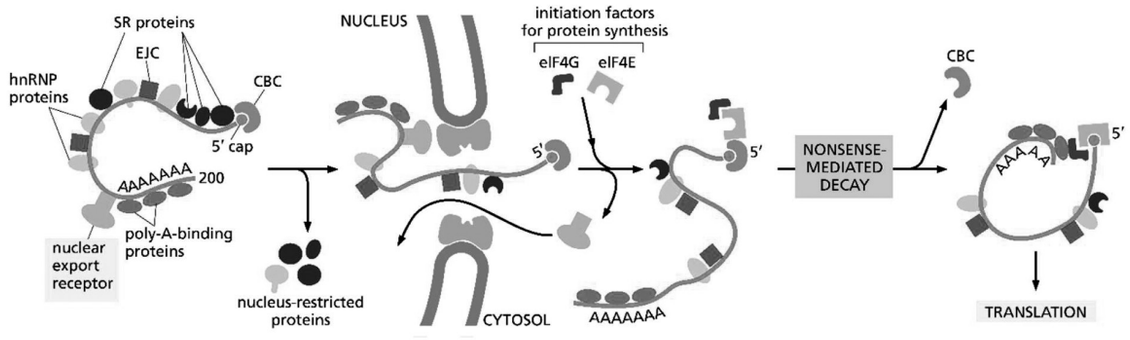


Figure 4.5

These mRNPs are exported by a heterodimeric exporter complex comprising:

- **NXF1 (Nuclear Export Factor 1)**
- **NXT1 (Nuclear Export Transporter 1)**

NXF1 binds mRNPs via REF and SR proteins, which are associated with **exonic splicing enhancers,** thus coupling RNA splicing to export. This interaction facilitates the docking of mRNPs to the NPC's central channel, followed by their translocation into the cytoplasm.

e) Directionality and mRNP Remodeling During Export

Directionality of export is maintained through mRNP remodeling, a process where:

- Nuclear mRNP proteins dissociate and are replaced by cytoplasmic proteins.
- **Dbp5,** a **DEAD-box RNA helicase** located on the cytoplasmic NPC filaments, facilitates the removal of NXF1/NXT1 and other nuclear proteins from the mRNP.

In the cytoplasm:

- **CBC (Cap Binding Complex)** at the 5′ cap is replaced by **eIF4E (Eukaryotic Initiation Factor 4E).**

- **PABPN1 (Nuclear Poly(A)-Binding Protein)** is replaced by **PABPC1 (Cytoplasmic Poly(A)-Binding Protein).**

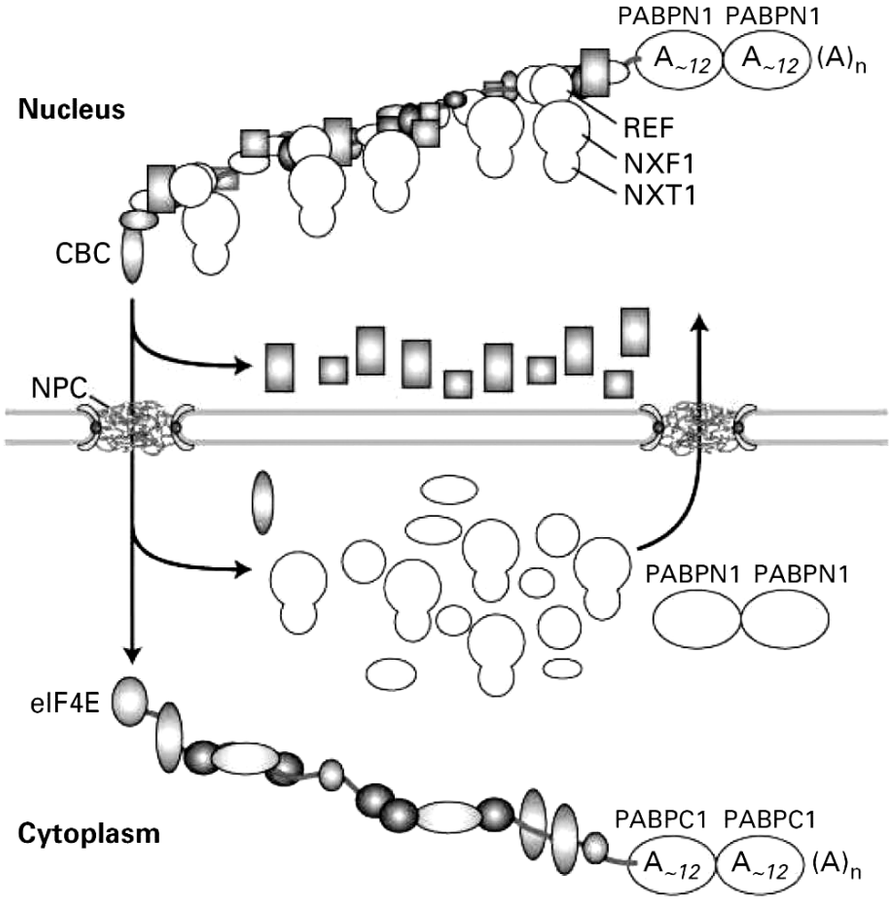


Figure 4.6

Remodeling of mRNPs during nuclear export.

Some mRNP proteins (rectangles) dissociate from nuclear mRNP complexes before their export through an NPC. Others (ovals) are exported through the NPC with the mRNP, but dissociate from it in the cytoplasm and are shuttled back into the nucleus through an NPC. In the cytoplasm, translation initiation factor eIF4E replaces CBC bound to the 5′ cap, and PABPC1 replaces PABPN1.

f) Regulation via Phosphorylation of SR Proteins

Phosphorylation status of SR proteins regulates nuclear export:

- In yeast, the **SR protein Npl3** binds to mRNA when phosphorylated.
- After **3'-end cleavage and polyadenylation,** Npl3 is dephosphorylated by **Glc7**, enabling its interaction with NXF1/NXT1.

- In the cytoplasm, **Sky1 kinase** re-phosphorylates Npl3, triggering its dissociation from the mRNP.

This phosphorylation-dephosphorylation cycle ensures export of only correctly processed mRNAs and prevents premature export.

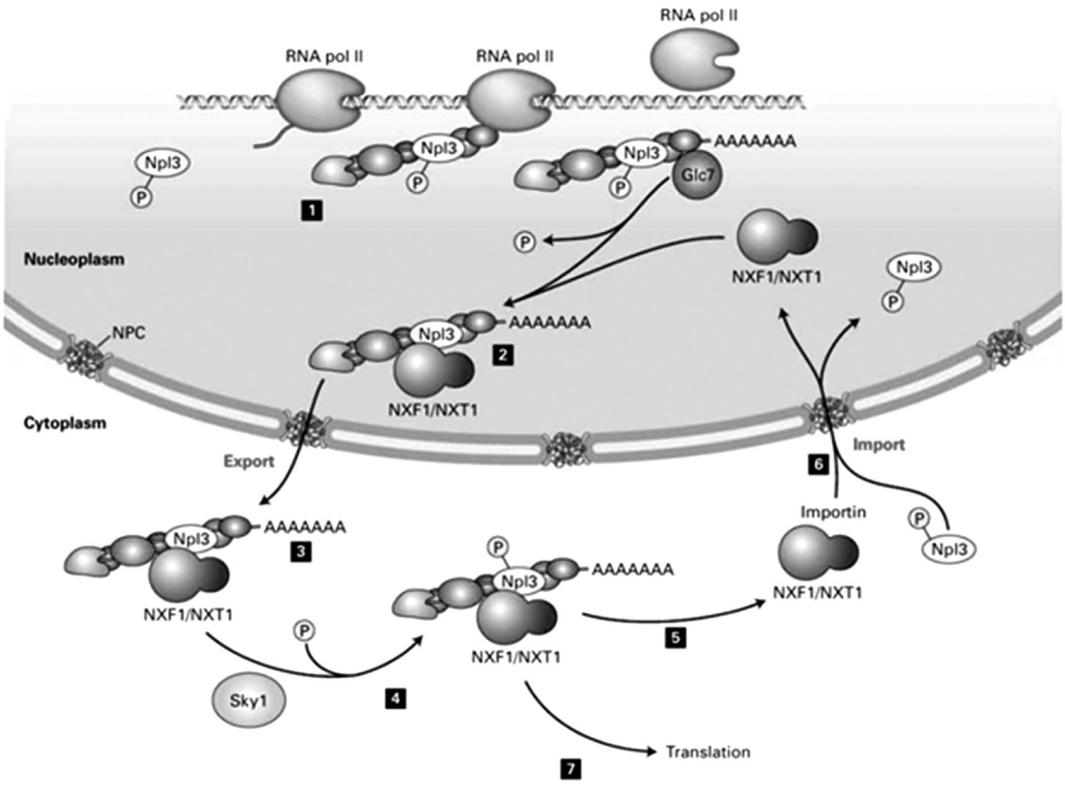


Figure 4.7

Reversible phosphorylation and direction of mRNP nuclear export. Step 1: The yeast SR protein Npl3 binds nascent pre-mRNAs in its phosphorylated form. Step 2: When polyadenylation has occurred successfully, the Glc7 nuclear phosphatase dephosphorylates Npl3, promoting the binding of the mRNP exporter, NXF1/NXT1. Step 3: The mRNP exporter allows diffusion of the mRNP complex through the central channel of the nuclear pore complex (NPC). Step 4: The cytoplasmic protein kinase Sky1 phosphorylates Npl3 in the cytoplasm, causing step 5 dissociation of the phosphorylated Npl3 from the mRNP exporter, probably through the action of an RNA helicase associated with NPC cytoplasmic filaments step 6. The mRNA transporter and phosphorylated Npl3 are transported back into the nucleus through NPCs. Step 7 Transported mRNA is available for translation in the cytoplasm. See E. Izaurralde, 2004, *Nat. Struct. Mol. Biol.* **11**:210–212; see also W. Gilbert and C. Guthrie, 2004, *Mol. Cell* **13**:201–212.

g) Quality Control: Prevention of Export of Unprocessed pre-mRNA

Only fully processed mRNAs are exported. **Spliceosomes** bound to pre-mRNA prevent export:

- Mutations in splice sites that still allow binding of **small nuclear RNPs (snRNPs)** but inhibit splicing result in nuclear retention.
- A protein component of the **NPC nuclear basket** ensures this retention.

This is critical to prevent the synthesis of truncated or defective proteins.

h) Experimental Visualization: Balbiani Ring mRNAs

Studies in Chironomus tentans larvae, using Balbiani ring (BR) genes in salivary glands, have shown:

- BR mRNAs (~75 kb) form coiled hnRNPs, which remodel and uncoil during export.
- The 5′ end of the mRNA leads the export through the NPC.
- The exported mRNAs rapidly associate with ribosomes in the cytoplasm, confirming directional export.

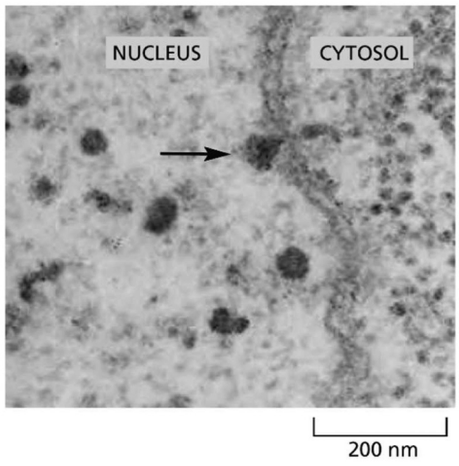


Figure 4.8: Electron microscope micrographs of an unusually large and abundant insect RNA, called the Balbiani Ring mRNA

i) Special Cases: Viral mRNA Export – HIV Rev Protein

In HIV, unspliced and singly spliced mRNAs contain incomplete splice sites, and their export is mediated by:

- The **Rev protein**, which binds to a **Rev Response Element (RRE)** in the viral RNA.
- Rev contains a **leucine-rich Nuclear Export Signal (NES)** and utilizes **exportin 1** instead of NXF1/NXT1.

This allows bypassing of cellular mRNA quality control, enabling export and translation of viral mRNAs that would otherwise be retained.

j) Key Points of Regulation

- mRNP export is regulated to ensure fidelity of gene expression.
- Transport requires proper **RNA processing (splicing, capping, polyadenylation)**.

- **Nuclear retention** and **exosomal degradation** prevent accumulation of aberrant mRNAs.
- Some viruses exploit or bypass these quality control mechanisms for their replication.

Component/Protein	Function
NXF1/NXT1	Main mRNA export receptor
REF (RNA Export Factor)	Adapter protein bound near exon-exon junctions
SR Proteins	Splicing factors and mRNA export enhancers
Dbp5	RNA helicase aiding in mRNP remodeling in cytoplasm
CBC & eIF4E	Cap-binding proteins (nuclear vs. cytoplasmic)
PABPN1 & PABPC1	Poly(A)-binding proteins (nuclear vs. cytoplasmic)
Npl3	SR protein regulating export via phosphorylation cycle
Exportin 1	Transports Rev protein-bound HIV RNAs
Exosomes	Degrade defective RNAs retained in nucleus

Table 4.1: Key Components in RNA Export

3. Degradation of RNA

a) Introduction to RNA Degradation

Messenger RNA (mRNA) degradation is a pivotal aspect of **post-transcriptional gene regulation** that determines both the lifetime and translational capacity of an mRNA transcript in the cytoplasm. The concentration of an mRNA in a cell is ultimately governed by two key factors: the rate at which it is synthesized and the rate at which it is degraded. Rapid mRNA turnover is especially important in the regulation of proteins involved in transient cellular responses such as cytokine secretion, cell cycle progression, and developmental signaling.

b) General Pathways of mRNA Degradation in the Cytoplasm

Three major cytoplasmic degradation pathways for eukaryotic mRNAs are outlined as follows:

(i) Deadenylation-dependent mRNA decay

This is the most prevalent pathway in eukaryotic cells.

- **Step 1:** Progressive shortening of the **poly(A) tail** by the **deadenylase complex,** which reduces the ability of **PABPC1 (Poly(A)-Binding Protein, Cytoplasmic 1)** to stabilize mRNA.
- **Step 2:** When the tail becomes too short, the **5' cap structure (m^7GpppN)** is removed by the **decapping complex DCP1/DCP2**.

- **Step 3:** The now unprotected mRNA is rapidly degraded by:
  - **XRN1** – a **5' → 3' exonuclease**
  - **Exosomes – 3' → 5' exonuclease complexes,** more dominant in mammalian cells

(ii) Deadenylation-independent decapping pathway

Some mRNAs are directly decapped without prior shortening of the poly(A) tail.

- This process is regulated by elements in the 5' untranslated region (UTR).
- Example: mRNA of **Rps28B**, where the **Edc3 (Enhancer of Decapping 3)** protein recruits the decapping complex DCP1/DCP2, leading to rapid degradation by **XRN1**.

(iii) Endonucleolytic cleavage pathway

- In this pathway, internal cleavage of the mRNA by endonucleases occurs.
- The resulting fragments are degraded by cytoplasmic exosomes and XRN1.

This mechanism is typical in **RNA interference (RNAi)** mediated by **siRNAs (small interfering RNAs)**.

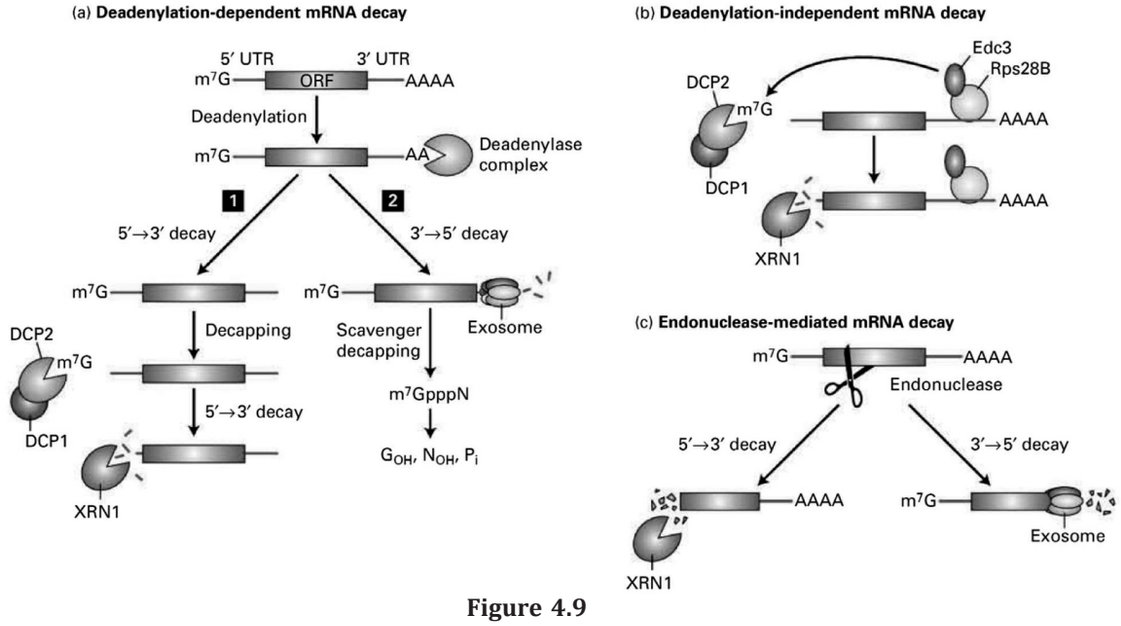


Figure 4.9

Pathways for degradation of eukaryotic mRNAs. (a) In the most common pathway of mRNA degradation, the deadenylation-dependent pathway, the poly(A) tail is progressively shortened by a deadenylase complex until it reaches a length of 20 or fewer A residues, at which point the interaction between PABPC1 and the remaining poly(A) is destabilized, leading to weakened interactions between the 5′ cap and translation initiation factors (see Figure 5-23). The deadenylated mRNA then may either (1) be decapped by the DCP1/DCP2 deadenylation complex and degraded by XRN1, a 5′ →3′ exonuclease, or (2) be degraded by 3′→5′ exonucleases in cytoplasmic exosomes. (b) Other mRNAs are decapped before they are deadenylated and then degraded by the XRN1 5′→3′ exonuclease. In the example shown from yeast, an RNA-binding protein Rps28B binds a sequence in the 3′-UTR of its own mRNA, which then interacts with Edc3 (*e*nhancer of *d*ecapping 3). Edc3 then recruits the DCP1/2 decapping enzyme to the mRNA, auto regulating expression of Rps28B. (c) Some mRNAs are cleaved internally by an endonuclease and the fragments degraded by a cytoplasmic exosome and the XRN1 exonuclease. See N. L. Garneau, J. Wilusz, and C. J. Wilusz, 2007, *Nat. Rev. Mol. Cell Biol.* **8**:113.

c) RNA Stability and Translational Regulation

(iv) Inverse Relationship Between Translation and Degradation

There is an **inverse correlation** between the **frequency of translation initiation** and **mRNA degradation**:

- Actively translating mRNAs are **more stable** due to protection by **translation initiation factors** and **PABPC1**.
- Translation initiation promotes **cap–tail interactions**, stabilizing the mRNA.

(v) AU-Rich Elements (AREs) and mRNA Instability

Certain mRNAs, particularly those encoding **cytokines** and **transcription factors**, possess **AU-rich elements (AREs)** in their **3' UTR**.

- These AREs are recognized by **ARE-binding proteins**, which:
  - Interact with **deadenylating enzymes**
  - Recruit the **exosome**
- Result: **Rapid deadenylation and decay** even when translation is active

This mechanism allows **short bursts of gene expression** while maintaining tight control.

d) Role of P-Bodies in mRNA Decay

Processing Bodies (P-bodies) are specialized cytoplasmic granules involved in:

- **Translational repression**
- **mRNA degradation**

P-bodies contain:

- **DCP1/DCP2 decapping enzymes**
- **XRN1 exonuclease**
- **RNA-binding proteins** and regulatory molecules

mRNAs with **inhibited translation** or those targeted by **miRNAs (microRNAs)** are often **sequestered into P-bodies** where they may be:

- Degraded
- Temporarily stored

e) Post-transcriptional Modifications and RNA Degradation

(i) m^6A (N6-methyladenosine) modification

- The **most prevalent RNA modification** in eukaryotic mRNA.
- Found near **stop codons** and **3' UTRs**
- m^6A-modified RNAs are:
  - **Less stable**
  - Targeted for degradation via **P-body association**

(ii) microRNA (miRNA)-Mediated Degradation

Mechanism:

- **miRNAs (~2′ nt)** are processed and loaded into **RISC (RNA-Induced Silencing Complex)** with **Argonaute (AGO) proteins.**
- RISC binds imperfectly to 3′ UTR of target mRNAs:
  - **Inhibits translation**
  - Recruits mRNA to P-bodies for degradation

Multiple miRNA–RISC complexes bound to a single mRNA enhances degradation likelihood.

f) RNA Interference (RNAi) and siRNA-Mediated Cleavage

siRNAs (Small Interfering RNAs):

- Derived from **double-stranded RNAs (dsRNAs)** processed by **Dicer**
- Loaded into **RISC**, where they **base-pair perfectly** with target mRNA

Outcome:

- **AGO2**, the catalytic component of RISC, cleaves the mRNA at a specific site (between nt 10–11 of siRNA).
- Cleaved fragments are rapidly degraded by **XRN1** and **exosomes**.

g) Regulation by Cytoplasmic Polyadenylation

- In **early embryos** and **neurons**, mRNAs are stored with **short poly(A) tails** and translationally repressed.

- **Cytoplasmic Polyadenylation Element Binding Protein (CPEB)** and **Maskin** block translation.
- Upon signal (e.g., **progesterone** in oocytes or **synaptic activity** in neurons), **CPEB is phosphorylated** and recruits:
    - **Cytoplasmic PAP (Poly(A) Polymerase)**
    - **PABPC1**
- Result: Poly(A) tail is extended, and translation is initiated.

h) Conclusion

mRNA degradation is a dynamic and multifactorial process essential for:

- Fine-tuning gene expression
- Ensuring cellular homeostasis
- Providing temporal control during development, immune response, and stress

Regulatory pathways like **miRNA targeting, ARE-mediated decay,** and **cytoplasmic deadenylation** integrate signals from the cellular environment to control the fate of transcripts post-transcriptionally.

Probable Questions:

1. Describe the structural organization and functional components of the nuclear envelope and nuclear pore complexes (NPCs). How do these structures facilitate selective nucleocytoplasmic transport?
2. Explain the step-by-step mechanism of nuclear import of proteins, including the roles of nuclear localization signals (NLS), importins, Ran-GTP, and FG-nucleoporins.
3. Discuss the molecular mechanism of mRNA export through the nuclear pore complex. How does the NXF1/NXT1 export pathway differ from the Ran-GTP-dependent export of proteins?
4. Elaborate on the role of SR proteins and their phosphorylation cycle in regulating mRNA export. How does this post-translational modification ensure export fidelity?
5. Describe how the nuclear pore complex contributes to the quality control of mRNA export. What mechanisms prevent the export of unprocessed or aberrant pre-mRNAs?

6. Explain the role of the Balbiani ring mRNA model in studying mRNA export. What insights have been gained about the structure and dynamics of mRNP remodeling?
7. Outline the degradation pathways of cytoplasmic mRNA in eukaryotic cells. Include the deadenylation-dependent, deadenylation-independent, and endonucleolytic cleavage pathways.
8. Discuss the regulatory role of AU-rich elements (AREs) in mRNA stability. How do ARE-binding proteins contribute to rapid transcript turnover?
9. Describe the functions of P-bodies in post-transcriptional regulation. What components are found in P-bodies, and how do they contribute to mRNA silencing and decay?
10. Explain the mechanisms of miRNA- and siRNA-mediated mRNA degradation. How does RNA interference contribute to post-transcriptional gene silencing in eukaryotic cells?

Suggested Readings:

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). Molecular biology of the cell (6th ed.). Garland Science.
2. Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Martin, K. C., Yaffe, M., & Amon, A. (2021). Molecular Cell Biology (9th ed.). W. H. Freeman.
3. Berg, J. M., Tymoczko, J. L., Gatto, G. J., & Stryer, L. (2022). Biochemistry (10th ed.). W. H. Freeman

Unit-V

RNA Transport, Translation and stability 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

A. Translational Machinery and Translational Control – Energetics of Amino Acid Polymerization

1. Introduction to Translational Machinery

The **translational machinery** is a highly complex, evolutionarily conserved system that interprets the information encoded in mRNA and catalyzes the synthesis of polypeptides. It comprises:

- **Ribosomes** (rRNA + protein complexes)
- **Transfer RNAs (tRNAs)** – adaptors between amino acids and codons
- **Aminoacyl-tRNA synthetases** – charge tRNAs with their corresponding amino acids
- **Messenger RNA (mRNA)** – the coding template
- **Protein translation factors** – initiation, elongation, and termination helpers

This machinery must ensure **speed, specificity, and fidelity** to enable the synthesis of complex proteins while minimizing translational errors.

2. Directionality and Fidelity of Translation

Protein synthesis progresses from the **amino (N-) terminus to the carboxyl (C-) terminus**, while the mRNA is read in the **5′ to 3′ direction**, one codon (three nucleotides) at a time.

The system must strike a balance between **speed and accuracy**:

- In *E. coli*, translation occurs at **~20 amino acids per second**.

- Error rates must be extremely low. An error rate of $\mathbf{10^{-3}}$ is sufficient for 300-residue proteins (~33 kDa), while $\mathbf{10^{-4}}$ is required for accurate synthesis of 1000-residue proteins (~110 kDa).

Table 5.1: Accuracy of protein synthesis

Frequency of inserting an incorrect amino acid	Probability of synthesizing an error-free protein		
	Number of amino acid residues		
	100	300	1000
10^{-2}	0.366	0.049	0.000
10^{-3}	0.905	0.741	0.368
10^{-4}	0.990	0.970	0.905
10^{-5}	0.999	0.997	0.990

Note: The probability p of forming a protein with no errors depends on n, the number of amino acids, and ε, the frequency of insertion of a wrong amino acid: $p = (1 - \varepsilon)^n$.

3. Energetics of Amino Acid Polymerization

The polymerization of amino acids into a polypeptide chain is an **energetically expensive** process, involving multiple ATP and GTP hydrolysis steps.

A. Aminoacylation of tRNA (Activation)

- Each amino acid is first **activated** by its **aminoacyl-tRNA synthetase** in two steps:
- **Amino acid + ATP (Adenosine Triphosphate) → Aminoacyl-AMP + PPi**
- **Aminoacyl-AMP + tRNA → Aminoacyl-tRNA + AMP (Adenosine Monophosphate)**
- This reaction consumes **2 ATP equivalents**:
- One ATP is used for amino acid activation.
- The hydrolysis of **pyrophosphate (PPi)** to 2P*i* drives the reaction forward.

B. Peptide Bond Formation

- The **peptidyl transferase center** (a ribozyme within the **28S rRNA** of the 60S ribosomal subunit in eukaryotes) catalyzes the nucleophilic attack of the α-amino group of the A-site aminoacyl-tRNA on the carbonyl carbon of the P-site peptidyl-tRNA.

- This reaction is **thermodynamically favourable** because of the **high-energy ester bond** in the aminoacyl-tRNA.

C. Elongation and Translocation

- For each amino acid added:
- **1 GTP (Guanosine Triphosphate)** is hydrolyzed by **eEF1A (or EF-Tu in bacteria)** during the delivery of aminoacyl-tRNA to the A site.
- **1 GTP** is hydrolyzed by **eEF2 (or EF-G)** to drive **ribosomal translocation**.

4. Total Energy Cost Per Amino Acid

Step	Molecule Consumed	High-Energy Phosphate Bonds Used
Amino acid activation	ATP $\rightarrow$ AMP + PPi	2 ATP equivalents
Elongation (tRNA delivery)	GTP $\rightarrow$ GDP + Pi	1 GTP
Translocation	GTP $\rightarrow$ GDP + Pi	1 GTP
Total per amino acid	—	**4 high-energy phosphate bonds**

This energy expenditure underscores the **cellular investment in translational fidelity**, reinforcing the importance of both **accuracy and efficiency**.

5. Translational Control Mechanisms

While Stryer primarily discusses **bacterial systems**, in eukaryotes, **translational control** is exerted at:

- **Initiation** (e.g., via phosphorylation of eIF2α or mTOR regulation)
- **Elongation** (e.g., eEF2 phosphorylation halts translocation under stress)
- **Termination** and **ribosome recycling**

These regulatory mechanisms fine-tune protein synthesis in response to environmental conditions, stress, nutrient availability, and developmental signals.

6. Structural Optimization: The tRNA–Ribosome Axis

The **tRNA** molecule is structurally optimized for translation:

- L-shaped 3D conformation
- Anticodon loop interacts with mRNA in the **decoding center**
- Aminoacyl-CCA tail fits into the **peptidyl transferase center**

- Allows **accurate base pairing**, **codon–anticodon recognition**, and **energetic optimization** of peptide bond formation

7. Conclusion

The energetics of amino acid polymerization is a tightly coordinated process requiring:

- **Energy investment** (4 high-energy phosphate bonds per amino acid)
- **Structural machinery** (ribosome, tRNA, rRNA)
- **Enzymatic specificity and proofreading** (by aminoacyl-tRNA synthetases)
- **Regulation** (via GTPases and signaling pathways)

Together, these components ensure **high-fidelity protein synthesis**, essential for cell survival, adaptation, and function.

B. *t*RNAs and Their Modifications

Introduction: The Adaptor Molecule in Translation

Transfer RNA (tRNA) is central to translation—the process of decoding an mRNA sequence into a polypeptide chain. Serving as the critical adaptor between the nucleotide codons of mRNA and the corresponding amino acids, tRNAs ensure that the correct amino acid is inserted at each step during protein synthesis. Each tRNA is specific for one amino acid and contains an **anticodon** loop that base-pairs with the complementary codon on mRNA.

Structure of tRNA: The Cloverleaf and L-shaped Models

All tRNAs share a highly conserved secondary and tertiary structure that enables them to interact uniformly with ribosomes, mRNAs, aminoacyl-tRNA synthetases, and elongation factors.

1. Primary Features of tRNA

Each tRNA molecule:

- Is a single-stranded RNA consisting of **73–93 ribonucleotides**, ~25 kDa in mass.
- Folds into a **cloverleaf secondary structure**, with four key arms:
    - **Acceptor Stem (3′ CCA end):** Attachment site for the amino acid.
    - **TψC Loop:** Contains ribothymidine (T), pseudouridine (ψ), and cytidine (C); involved in ribosomal binding.

- **DHU Loop:** Contains **dihydrouridine**, important for synthetase recognition.
- **Anticodon Loop:** Contains the **anticodon triplet**, which pairs with mRNA codons.

2. Tertiary Structure

- The cloverleaf folds into a **compact L-shaped three-dimensional structure**.
- The **CCA acceptor stem** and the **anticodon loop** lie at opposite ends of the L, allowing:
  - The **anticodon** to interact with mRNA.
  - The **amino acid** to be incorporated into the growing polypeptide chain at the peptidyl transferase center.

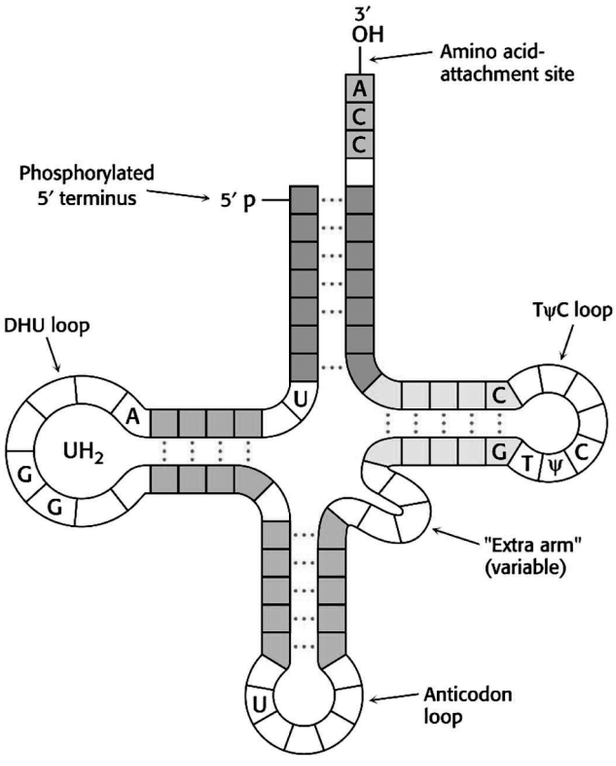


Figure 5.1: General structure of tRNA molecules.
Comparison of the base sequences of many tRNAs reveals a number of conserved features.

Chemical Modifications in tRNAs: Types and Functional Roles

Transfer RNAs are among the most **chemically modified RNAs** in cells. Each tRNA contains **7 to 15 unusual or modified nucleotides**, generated post-transcriptionally via **enzymatic modification**.

Common Modified Bases:

- **Inosine (I):** Derived from **adenosine** by deamination; found in the anticodon loop, promotes **wobble base pairing**.
- **Dihydrouridine (UH,):** A **reduced uracil**, increases flexibility in the DHU loop.
- **Ribothymidine (T):** Found in the TψC loop.
- **Pseudouridine (ψ):** Isomer of uridine, stabilizes tRNA structure.
- **Methylguanosine (mG), Dimethylguanosine (m^2G), Methylinosine (mI), Methylcytidine (mC)**

Functions of Modifications:

- **Stabilize secondary and tertiary structure**
- Prevent unwanted base-pairing or misfolding
- Enhance tRNA recognition by **aminoacyl-tRNA synthetases**
- Increase translational fidelity by improving codon–anticodon interactions
- Allow **wobble base pairing**, enabling a single tRNA to recognize multiple codons

The Wobble Hypothesis and Codon Recognition

The **wobble hypothesis**, proposed by Francis Crick and validated experimentally, explains how some tRNAs can recognize more than one codon.

Key Principles:

- The **first two bases** of the codon form **strict Watson–Crick base pairs** with the corresponding anticodon.
- The **third base** of the codon (the "wobble" position) allows **non-Watson–Crick pairings**, especially with **inosine** in the anticodon.

Inosine (I) can pair with:

- **U, C, or A** in the third codon position

Biological Advantage:

- Reduces the number of tRNAs needed.
- Increases translation efficiency while preserving accuracy.

tRNA Identity and Recognition by Synthetases

Each tRNA must be **specifically recognized** by its corresponding **aminoacyl-tRNA synthetase**, which charges it with the correct amino acid.

tRNA Identity Determinants:

- **Anticodon sequence**
- **Acceptor stem (including the CCA 3′ end)**
- **Modified bases in loops (e.g., DHU loop, TψC loop)**
- **Overall tertiary conformation**

Structural Adaptations:

- Recognition sites vary for different synthetases.
- Many tRNAs share similar architecture but contain **unique identity elements** ensuring high fidelity in aminoacylation.

Conclusion

Transfer RNAs are not just passive carriers of amino acids; they are **sophisticated molecular adaptors** equipped with intricate structural and chemical features that enable the decoding of the genetic code with remarkable precision. The extensive post-transcriptional modifications present in tRNAs enhance their structural stability, functional specificity, and accuracy in translation—features essential for maintaining the integrity of the proteome in all living organisms.

C. Aminoacyl tRNA Synthetases

Introduction

Aminoacyl tRNA synthetases (aaRSs) are a class of enzymes essential for interpreting the genetic code. These enzymes are responsible for catalyzing the covalent attachment of amino acids to their respective **tRNAs**, forming **aminoacyl-tRNAs (charged tRNAs)**. This step is crucial as it determines the **fidelity** of translation—ensuring that the correct amino acid is assigned to each codon.

These enzymes are not only vital in translation but are also considered to be the **true translators** of the genetic code, as they establish the connection between the nucleotide codon (via tRNA anticodon) and the corresponding amino acid.

Mechanism of Action: Two-Step Reaction

The formation of aminoacyl-tRNA occurs through a **two-step enzymatic reaction**:

Step 1: Activation

- The amino acid is **activated** by reacting with **ATP**, forming **aminoacyl-AMP** (an aminoacyl adenylate) and **pyrophosphate (PPi)**.

Amino acid + ATP → Aminoacyl-AMP + PP*i*

Step 2: Transfer to tRNA

- The **aminoacyl group** is then transferred from aminoacyl-AMP to the **3′ end** of the corresponding tRNA (specifically to the **2′ or 3′ hydroxyl group** of the terminal adenosine in the CCA tail), forming **aminoacyl-tRNA** and releasing AMP.

Aminoacyl-AMP + tRNA → Aminoacyl-tRNA + AMP

Overall Reaction:

Amino acid + ATP + tRNA → Aminoacyl-tRNA + AMP + 2P*i*

Energetics of the Reaction

Although the **ΔG°′** of the aminoacylation is approximately zero, the reaction is driven forward by the **hydrolysis of pyrophosphate (PPi)** to two inorganic phosphates (2P*i*), rendering the overall process **highly exergonic and irreversible**.

Hence, the formation of each aminoacyl-tRNA consumes the **equivalent of two ATP molecules**.

Structural Specificity and Substrate Recognition

Each aminoacyl-tRNA synthetase must exhibit **high specificity** for:

- Its **cognate amino acid**
- Its corresponding **tRNA**

Misacylation (charging the wrong amino acid) can lead to translation errors and the production of faulty proteins, which is why many synthetases have evolved proofreading functions (discussed in the next section).

Two Classes of Aminoacyl-tRNA Synthetases

Aminoacyl-tRNA synthetases are evolutionarily divided into **two distinct structural and functional classes**:

Class I	Class II
Acylate the **2′-OH** of the terminal adenosine	Acylate the **3′-OH** of the terminal adenosine
Typically, **monomeric**	Typically, **dimeric or tetrameric**
Approach tRNA from the **minor groove**	Approach tRNA from the **major groove**
Examples: Arg, Cys, Gln, Glu, Ile, Leu, Met, Trp, Tyr, Val	Examples: Ala, Asn, Asp, Gly, His, Lys, Phe, Pro, Ser, Thr

These two classes represent independent evolutionary lineages and exhibit **distinct tRNA-binding architectures**, with the CCA arm of tRNA adopting different conformations depending on the enzyme class.

Substrate Recognition Strategies

Each aaRS uses unique strategies to recognize its amino acid. For example:

- **Threonyl-tRNA synthetase** distinguishes threonine from similar amino acids (valine, serine) using a **zinc ion** in the active site and a **hydrogen-bonding aspartate** to detect the hydroxyl group of threonine.
- Methyl groups in valine and serine are excluded due to steric clashes or lack of interaction with the zinc-binding site.

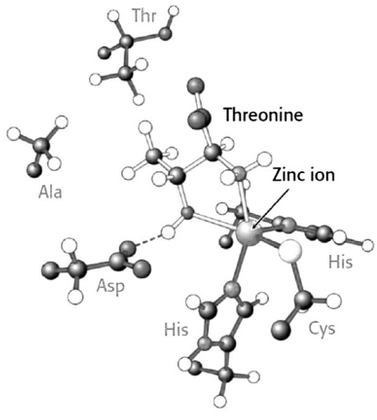


Figure 5.2: Active site of threonyl-tRNA synthetase. *Notice* that the amino acid-binding site includes a zinc ion (green ball) that binds threonine through its amino and hydroxyl groups.

This specificity ensures correct aminoacylation in ~99.99% of cases, which is essential given that even a 0.1% error rate can compromise protein fidelity.

Aminoacylation Site vs. Editing Site

Many aaRSs possess a **proofreading site** in addition to the aminoacylation active site:

- The **activation site** rejects amino acids **larger** than the correct one.
- The **editing site** hydrolyzes tRNAs mischarged with **smaller or incorrect** amino acids.

This **"double-sieve mechanism"** improves fidelity by allowing:

- **Discrimination at the aminoacylation site** based on size/chemistry
- **Correction at the editing site** through hydrolytic cleavage of mischarged tRNAs (e.g., Ser-tRNA^{Thr}).

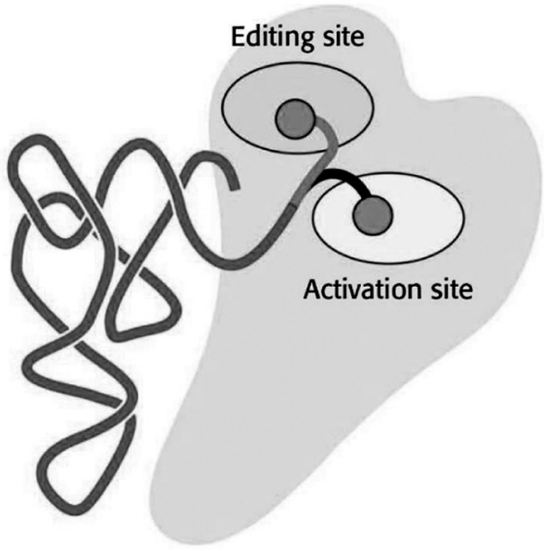


Figure 5.3: Editing of aminoacyl-tRNA.
The flexible CCA arm of an aminoacyl-tRNA can move the amino acid between the activation site and the editing site. If the amino acid fits well into the editing site, the amino acid is removed by hydrolysis.

tRNA Recognition Mechanisms

Synthetases identify their correct tRNA partners by recognizing:

- **Anticodon loop sequences**
- **Acceptor stem conformation**
- **Modified bases in loops (e.g., DHU, TψC)**
- **Tertiary structure and identity elements**

For example, threonyl-tRNA synthetase interacts with both the **anticodon (CGU)** and the **acceptor stem** to ensure specificity.

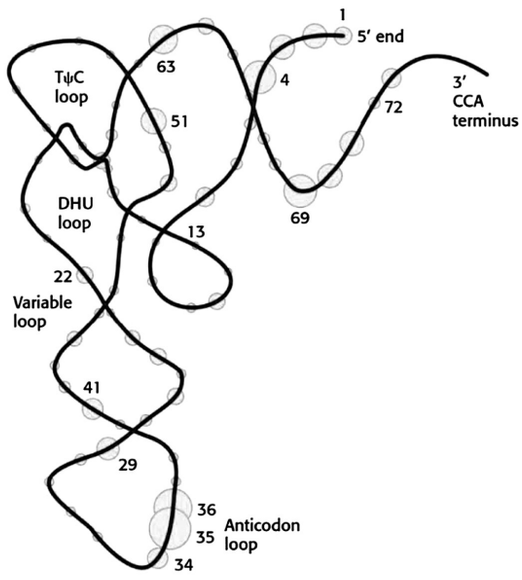


Figure 5.4: Recognition sites on tRNA. Circles represent nucleotides, and the sizes of the circles are proportional to the frequency with which they are used as recognition sites by aminoacyl-tRNA synthetases. The numbers indicate the positions of the nucleotides in the base sequence, beginning from the 5′ end of the tRNA molecule. [Information from M. Ibba, and D. Söll, *Annu. Rev. Biochem.* 69:617–650, 1981, p. 636.]

Evolutionary Implications

The existence of two distinct synthetase classes with different structural approaches may reflect a **coevolutionary need to recognize both faces of the tRNA molecule**, allowing unique identity signals for all 20 amino acids. This diversity likely reflects evolutionary pressure to **maximize specificity while minimizing cross-reactivity**.

Conclusion

Aminoacyl tRNA synthetases are **master gatekeepers of genetic fidelity**. Their dual capabilities—**charging the correct amino acid** and **proofreading mischarged tRNAs**—highlight their indispensable role in ensuring the accuracy of protein synthesis. The existence of structurally and functionally distinct enzyme classes further emphasizes their evolutionary optimization to match the complexity of the genetic code.

D. Accuracy During Aminoacylation of tRNA

Introduction

The accuracy of protein synthesis critically depends on the **correct pairing of amino acids with their respective tRNAs**. This process is catalyzed by **aminoacyl-tRNA synthetases**, and errors in this step would lead to the **misincorporation of amino acids into proteins**, potentially resulting in dysfunctional or deleterious proteins.

Despite the structural similarity among some amino acids, the error rate for amino acid attachment to tRNA is **extremely low (~10^{-4} to 10^{-5})**. This level of precision is achieved through a **multi-tiered selection mechanism**, including **substrate discrimination** and **post-transfer editing**, embedded within the synthetase enzymes.

1. The Specificity Challenge

Each aminoacyl-tRNA synthetase must:

- Correctly identify one amino acid from a pool of 20.
- Precisely match it with its **cognate tRNA** (or multiple isoacceptor tRNAs).

Example: **Threonyl-tRNA synthetase** must distinguish threonine from structurally similar amino acids:

- **Valine**: differs by a single methyl group.
- **Serine**: has the same hydroxyl group but lacks the methyl group.

2. Structural Basis for Amino Acid Discrimination

Synthetases utilize **unique structural motifs** and **cofactor binding sites** for selective recognition:

- **Threonyl-tRNA synthetase** employs a **zinc ion** coordinated by two **histidines** and one **cysteine**, which interacts specifically with the **hydroxyl group** of threonine.
- **Valine**, lacking the hydroxyl group, cannot form the required hydrogen bonds.
- An **aspartate** residue additionally stabilizes the correct side chain.

Despite this high specificity, occasional **misactivation** occurs, particularly with **serine**, whose structure is partially similar to threonine.

3. Proofreading and Editing Mechanisms

To mitigate errors from initial misactivation, many synthetases incorporate **editing (proofreading) domains**:

Double-Sieve Mechanism

- **Activation site** acts as a **coarse sieve**, excluding **larger** non-cognate amino acids due to steric hindrance.
- **Editing site** acts as a **fine sieve**, hydrolyzing mischarged tRNAs containing **smaller** or structurally close non-cognate amino acids.

Example: Serine vs. Threonine

- Serine, occasionally mischarged onto $tRNA^{Thr}$, is cleaved by a **dedicated hydrolytic site** in the enzyme.
- Correctly charged **Thr-$tRNA^{Thr}$** does **not fit** into the editing site and thus escapes hydrolysis.

4. Intramolecular Translocation: From Activation to Editing

The **flexible CCA arm** of the tRNA plays a vital role in enabling the editing process:

- After mischarging, the CCA arm can **swing** the aminoacyl group from the **synthetase active site** into the **editing pocket**.
- If the amino acid **fits** the editing site (e.g., serine), it is cleaved; if not (e.g., threonine), it remains attached.

This mechanism allows **error correction without releasing the tRNA**, thus maintaining efficiency and preventing misacylated tRNAs from entering the translation machinery.

5. Fidelity Achieved Through Binding Energy Discrimination

Not all aminoacyl-tRNA synthetases require a separate editing site:

- Some, such as **tyrosyl-tRNA synthetase**, distinguish between tyrosine and phenylalanine **solely via binding affinity**.
- Tyrosine binds $\sim 10^4$ times more tightly to the synthetase due to the presence of a hydroxyl group on its aromatic ring, which is absent in phenylalanine.

Hence, **editing evolves only when substrate similarity necessitates it**.

6. tRNA Recognition: A Key to Accuracy

In addition to amino acid discrimination, synthetases must recognize:

- The correct **tRNA structure**, involving:
- **Anticodon base sequences**
- **Acceptor stem conformation**
- **Unique modified bases**

Errors in tRNA recognition can also be prevented by **synergistic contacts** across the tRNA, enhancing the **specificity of the enzyme–tRNA complex**.

7. Evolutionary Adaptation of Accuracy

High-fidelity aminoacylation evolved to balance:

- **Speed of protein synthesis**
- **Biological cost of translation errors**

The typical accuracy of 1 error in 10,000 charging events reflects an **evolutionarily optimized equilibrium**, minimizing deleterious effects while maintaining translational efficiency.

Conclusion

The accuracy of aminoacylation is a cornerstone of translational fidelity. Through a combination of **steric discrimination**, **structural adaptation**, **editing sites**, and **dynamic proofreading**, aminoacyl-tRNA synthetases ensure that each tRNA is correctly charged with its corresponding amino acid. These mechanisms exemplify molecular precision and serve as the first checkpoint in the accurate flow of genetic information from nucleic acids to proteins.

E. Initiation of Translation and Its Regulation in Eukaryotes

1. Introduction to Translation Initiation

Initiation is the first and most **regulatory-intensive phase** of translation. It defines the correct **reading frame** and determines where **protein synthesis begins** on the mRNA. This process differs significantly between **prokaryotes** and **eukaryotes**, with eukaryotes employing a **more elaborate mechanism** involving **more initiation factors**, **cap-dependent scanning**, and **strict regulation** through signaling pathways.

2. Key Molecular Components in Eukaryotic Initiation

Eukaryotic initiation involves:

- **mRNA** with a 5′-cap (mw GpppN) and 3′ poly(A) tail
- **40S and 60S ribosomal subunits**, forming the 80S ribosome
- **Initiator tRNAi^{Met}** (not formylated as in bacteria)
- A suite of **eukaryotic initiation factors (eIFs)**

3. The Initiation Process: Cap-Dependent Scanning Model

Step 1: Formation of the 43S Pre-Initiation Complex

- Begins with the binding of **eIF2-GTP** to **Met-tRNAi^{Met}**.
- This complex associates with the **40S ribosomal subunit** along with other factors (eIF1, eIF1A, eIF3, eIF5).

Step 2: mRNA Recruitment and Circularization

- mRNA binds to **eIF4F complex**, which includes:
- **eIF4E** (cap-binding protein)
- **eIF4G** (scaffold protein, binds eIF3 and PABP)
- **eIF4A** (RNA helicase)
- **Poly(A)-binding protein (PABP)** binds to the 3′ poly(A) tail and interacts with eIF4G, promoting **circularization of mRNA**, enhancing translational efficiency.

Step 3: Scanning for Start Codon (AUG)

- The **43S complex** binds near the 5′ cap and **scans in the 5′ → 3′ direction** using ATP hydrolysis by eIF4A to unwind secondary structures.
- Upon encountering an AUG within a **favourable Kozak consensus** sequence (e.g., GCCRCCAUGG), scanning halts.

Step 4: Formation of the 48S Initiation Complex

- AUG recognition triggers GTP hydrolysis by **eIF2**, mediated by **eIF5**.
- Initiation factors dissociate, making room for the 60S subunit.

Step 5: Joining of the 60S Subunit

- Catalyzed by **eIF5B-GTP**.
- GTP hydrolysis by eIF5B triggers the release of eIF5B and eIF1A.
- Final result: **80S initiation complex** with Met-tRNAi^{Met} in the **P site**, A and E sites vacant.

While these details are standard in eukaryotic systems, equivalent initiation steps in **bacteria** with initiator tRNAf^{Met}, are **formylated**.

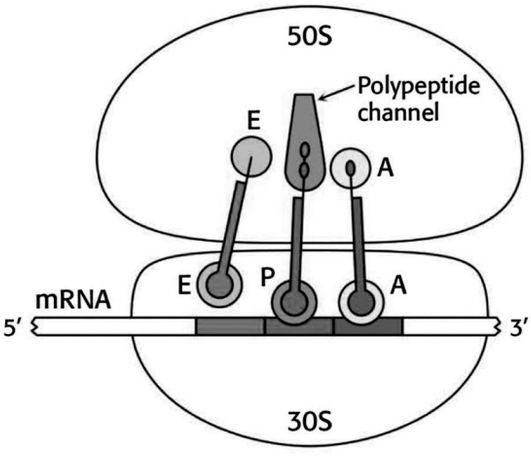


Figure 5.5 : An active ribosome. This schematic representation shows the relations among the key components of the translation machinery.

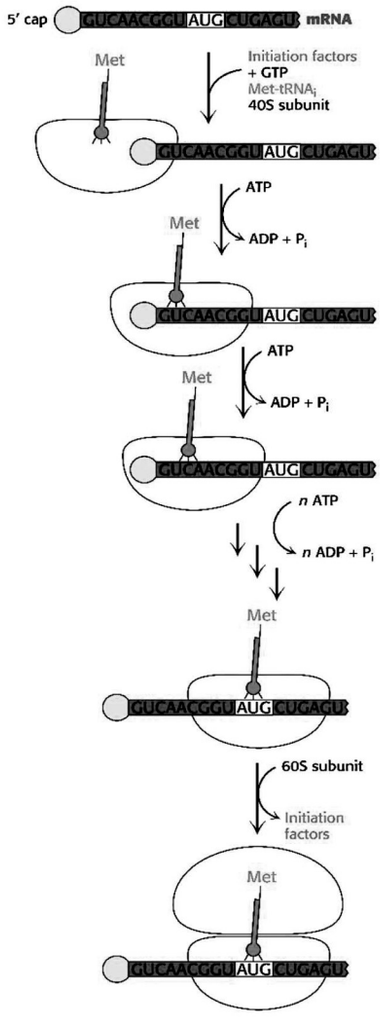


Fig. 5.6 : Eukaryotic translation initiation. In eukaryotes, translation initiation starts with the assembly of a complex on the 5′ cap that includes the 40S subunit and Met-tRNA$_i$. Driven by ATP hydrolysis, this complex scans the mRNA until the first AUG is reached. The 60S subunit is then added to form the 80S initiation complex.

4. Differences Between Eukaryotic and Prokaryotic Initiation

Feature	Eukaryotes	Prokaryotes
Ribosome Size	80S (40S + 60S)	70S (30S + 50S)
Initiator tRNA	Met-tRNAi^{Met} (non-formylated)	fMet-tRNAf^{Met} (formylated)
mRNA Recognition	5′ cap and scanning	Shine–Dalgarno sequence base-pairs with 16S rRNA
Number of Initiation Factors	≥12 (eIFs)	3 (IF1, IF2, IF3)
Regulatory Complexity	High (integrated with signaling)	Low

5. Regulation of Eukaryotic Translation Initiation

Eukaryotic translation initiation is a **key control point** for gene expression, especially under conditions like **stress, starvation, viral infection, or mitogenic stimulation**. Regulation primarily occurs at two levels:

A. Phosphorylation of eIF2α

- **eIF2** must be in its **GTP-bound form** to initiate translation.
- Under stress, **eIF2α is phosphorylated** by stress kinases (e.g., PERK, PKR, GCN2), converting eIF2 into a **competitive inhibitor** of eIF2B (the guanine nucleotide exchange factor).
- Result: **Global suppression of translation**, sparing select mRNAs with **internal ribosome entry sites (IRES)**.

B. Control by mTOR Pathway

- **mTOR (mechanistic Target of Rapamycin)** regulates translation by modulating the availability of:
- **eIF4E** via **4E-BP** (eIF4E-binding protein)
- **S6K**, which activates ribosomal protein S6 and translation of 5′ TOP mRNAs
- **Nutrient abundance** → mTOR active → **Translation ON**
- **Nutrient deficiency** → mTOR suppressed → **Translation OFF**

6. Special Cases: Internal Ribosome Entry Sites (IRES)

Under specific conditions (e.g., stress, apoptosis, viral infection), some eukaryotic mRNAs use **cap-independent initiation** via **IRES elements**:

- Ribosome binds **directly to the IRES** without the need for eIF4E.
- Used by both **viral** (e.g., picornavirus) and **host stress-response** genes (e.g., c-Myc, Apaf-1).

7. Clinical Relevance

Defects in initiation regulation are linked to various diseases:

- **Cancer**: Overactive mTOR/eIF4E promotes tumor growth.
- **Neurodegeneration**: Dysregulation of eIF2α phosphorylation impairs memory and cognition.

- **Viral infections**: Many viruses hijack host initiation machinery for their own translation.

Conclusion

Translation initiation in eukaryotes is a **highly controlled, energy-intensive** process that integrates signals from nutrient status, growth factors, stress responses, and cellular needs. This complex orchestration ensures accurate and context-dependent protein synthesis, making initiation one of the most strategically regulated steps in gene expression.

F. Translation Elongation and Its Control in Eukaryotes

1. Overview of Elongation Phase

Once the **80S initiation complex** has formed in eukaryotes, the process of **elongation** begins—this is the stage at which amino acids are sequentially added to the growing polypeptide chain. The core objectives of elongation are to:

- Accurately match each codon on the mRNA with the correct **aminoacyl-tRNA**
- Catalyze the formation of **peptide bonds**
- Translocate the ribosome along the mRNA by **three nucleotides**

This process is both **highly efficient** and **tightly regulated**, and it is carried out with the help of **elongation factors** and **GTP hydrolysis** to maintain directionality and fidelity.

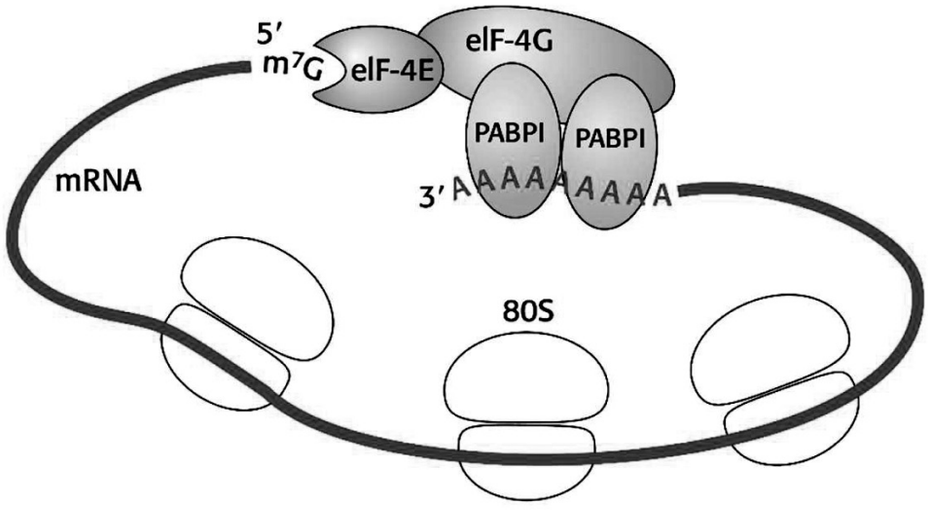


Figure 5.7 : Protein interactions circularize eukaryotic mRNA.

2. Key Ribosomal Sites in Elongation

Generally, the ribosome has **three tRNA binding sites** involved in elongation:

- **A site (Aminoacyl site):** Entry point for aminoacyl-tRNA
- **P site (Peptidyl site):** Holds the growing polypeptide chain attached to tRNA
- **E site (Exit site):** Where deacylated tRNA is released

These sites are strategically arranged to ensure accurate codon-anticodon interactions and smooth progression of the translation process.

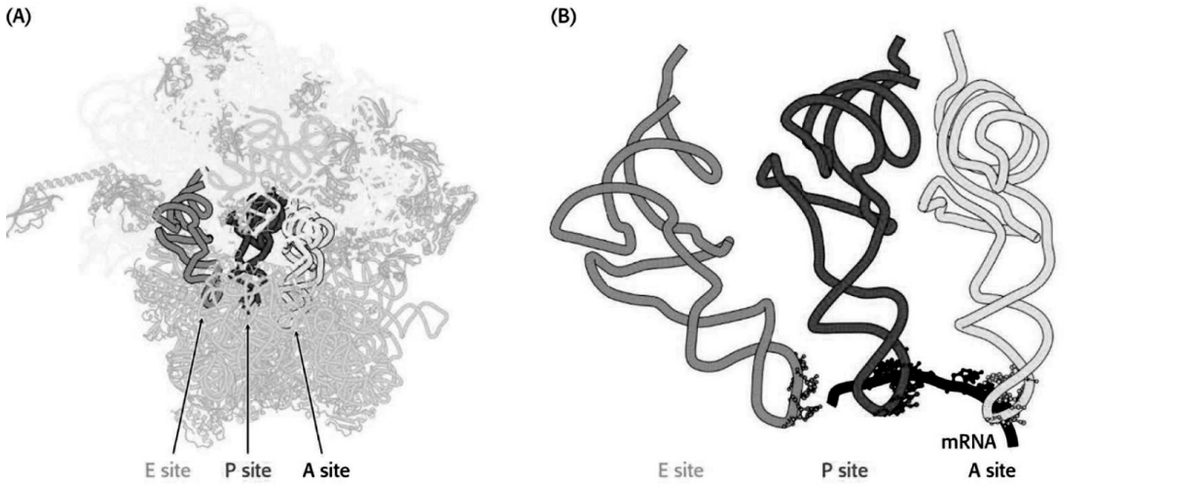


Figure 5.8 Transfer RNA-binding sites. (A) Three tRNA-binding sites are present on the 70S ribosome. They are called the A (for aminoacyl), P (for peptidyl), and E (for exit) sites. Each tRNA molecule contacts both the 30S and the 50S subunit. (B) The tRNA molecules in sites A and P are base-paired with mRNA. [(B) Drawn from 1JGP. pdb.]

3. Elongation Cycle:

Step 1: Codon Recognition

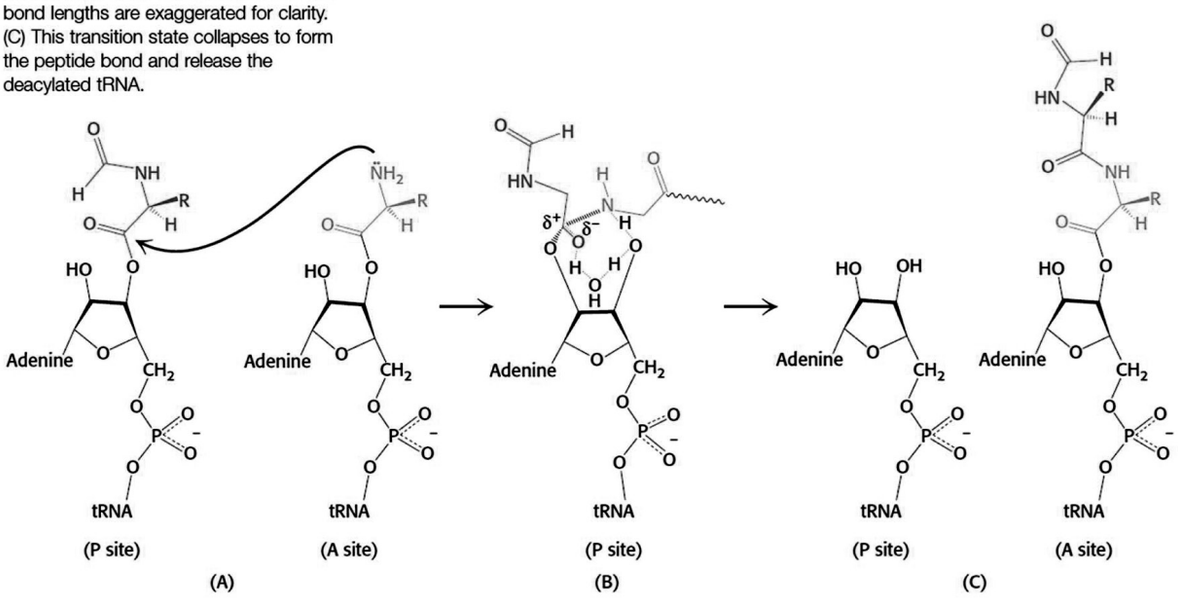
- A **charged aminoacyl-tRNA**, escorted by **eEF1A-GTP** (eukaryotic homolog of bacterial EF-Tu), binds to the **A site** of the ribosome.
- Codon-anticodon base pairing is monitored by the **small subunit rRNA**, especially conserved nucleotides such as **A1492, A1493, and G530** in prokaryotes (16S rRNA; conceptually equivalent in eukaryotes with 18S rRNA).
- Correct pairing triggers **GTP hydrolysis by eEF1A**, followed by the release of eEF1A-GDP.

Step 2: Peptide Bond Formation

- The **amino group** of the aminoacyl-tRNA in the A site attacks the **carbonyl carbon** of the peptidyl-tRNA in the P site.

- This reaction is catalyzed by the **peptidyl transferase center (PTC)**, a ribozyme function of **28S rRNA** in the large subunit (60S).
- Result: The growing peptide chain is transferred to the A-site tRNA.

FIGURE 5.9 : Peptide-bond formation. (A) The amino group of the aminoacyl-tRNA attacks the carbonyl group of the ester linkage of the peptidyl-tRNA. (B) An eight-membered transition state is formed. Note: Not all atoms are shown and some bond lengths are exaggerated for clarity. (C) This transition state collapses to form the peptide bond and release the deacylated tRNA.



Step 3: Translocation

- The ribosome advances **one codon (3 nucleotides)** in the 5′ → 3′ direction.
- This step is mediated by **eEF2-GTP** (homologous to bacterial EF-G).
- GTP hydrolysis drives conformational changes that:
  - Move the deacylated tRNA from the P site to the E site
  - Shift the peptidyl-tRNA from the A site to the P site
  - Prepare the A site for the next incoming aminoacyl-tRNA

After translocation, the **E site *t*RNA is released**, and the cycle repeats.

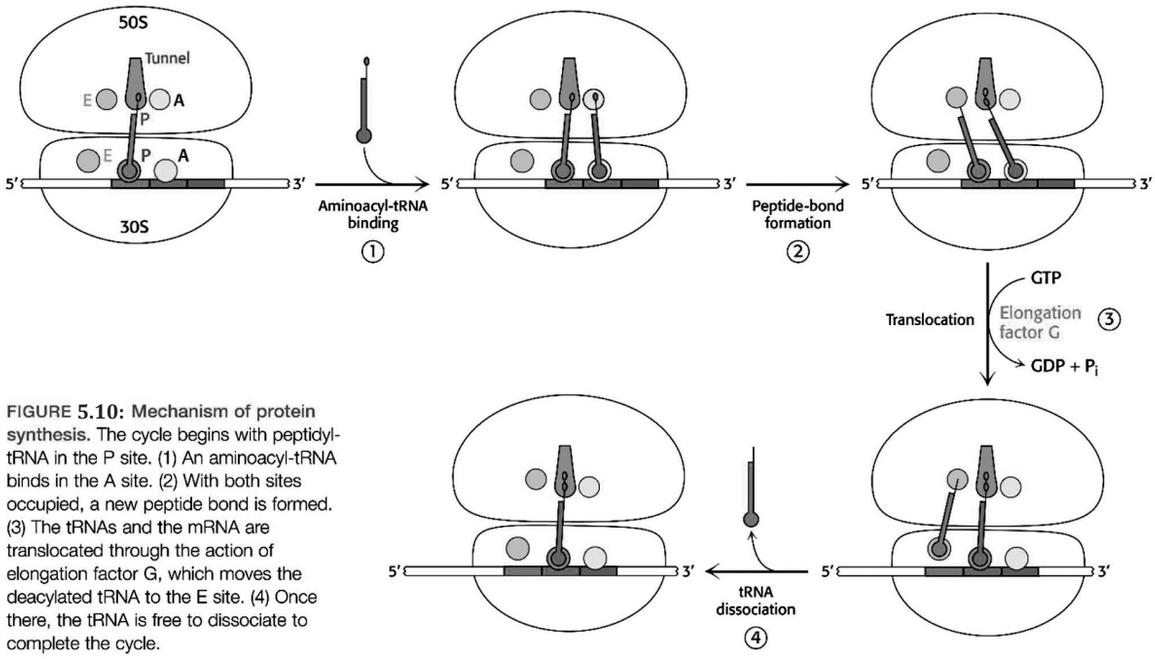


FIGURE 5.10: Mechanism of protein synthesis. The cycle begins with peptidyl-tRNA in the P site. (1) An aminoacyl-tRNA binds in the A site. (2) With both sites occupied, a new peptide bond is formed. (3) The tRNAs and the mRNA are translocated through the action of elongation factor G, which moves the deacylated tRNA to the E site. (4) Once there, the tRNA is free to dissociate to complete the cycle.

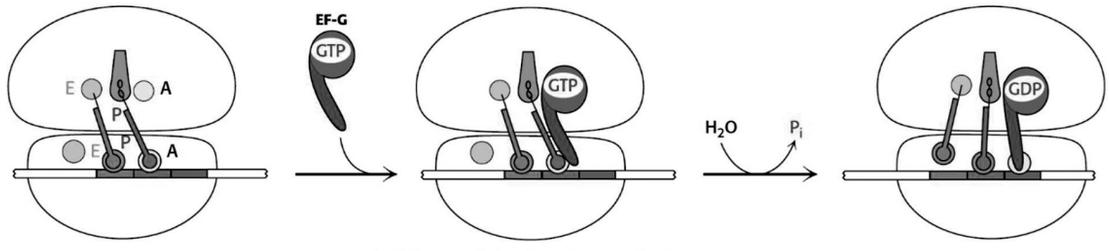


FIGURE 5.11: Translocation mechanism. In the GTP form, EF-G binds to the A site on the 50S subunit. This binding stimulates GTP hydrolysis, inducing a conformational change in EF-G that forces the tRNAs and mRNA to move through the ribosome by a distance corresponding to one codon.

4. Energetics of Elongation

Each elongation cycle consumes **two molecules of GTP**:

- One for **eEF1A-mediated delivery**
- One for **eEF2-mediated translocation**

Together with the two ATP equivalents used during aminoacylation, the total energy cost per peptide bond is **four high-energy phosphate bonds**.

5. Proofreading and Fidelity Control

The ribosome ensures **high translational accuracy** through:

- **Kinetic proofreading:** Only correctly paired aminoacyl-tRNAs trigger efficient GTP hydrolysis and accommodation into the A site.
- **Conformational checkpoints:** Incorrect codon-anticodon pairings are rejected before peptide bond formation.
- **rRNA surveillance:** Conserved residues in **18S rRNA** interact with the minor groove of the codon–anticodon helix, enhancing fidelity.

6. Regulation of Translation Elongation in Eukaryotes

While elongation is generally a **high-speed, processive event**, it is subject to **physiological and pathological regulation**:

A. Phosphorylation of eEF2

- **eEF2** can be **phosphorylated by eEF2 kinase** in response to:
- Nutrient deprivation
- Stress
- Hypoxia
- Phosphorylated eEF2 is **inactive**, thus **slowing translation**.
- Acts as a **translational brake**, allowing cells to conserve resources.

B. Ribosome Pausing and Stalling

- Specific **secondary structures** or **codon usage bias** in mRNAs can induce ribosome **stalling**.
- This is exploited in:
- Regulation of **upstream open reading frames (uORFs)**
- **Co-translational folding** of nascent polypeptides
- **Quality control mechanisms** like **No-Go Decay (NGD)**

C. MicroRNA-Mediated Elongation Repression

- Some **miRNAs** can inhibit elongation rather than initiation by binding the **coding region** and impeding ribosomal progression.
- Leads to **ribosome drop-off** or **P-body sequestration**.

7. Medical and Biotechnological Relevance

- **Cancer cells** often exhibit **hyperactive elongation**, with upregulated eEF1A and eEF2.
- **Antibiotics (e.g., cycloheximide)** target the elongation step in **eukaryotic ribosomes**.
- **Toxins** such as **diphtheria toxin** inactivate eEF2 via **ADP-ribosylation**, shutting down translation.

Conclusion

Translation elongation in eukaryotes is a **highly dynamic**, **regulated**, and **precision-governed** process. It balances the need for rapid protein synthesis with stringent fidelity and resource control. Elongation factors, GTP hydrolysis, and ribosomal conformational changes work synergistically to advance the ribosome along mRNA, while regulatory pathways ensure that elongation proceeds only when appropriate to the cell's physiological state.

G. Termination of Translation

1. Introduction to Termination

Termination of translation is the final step in the elongation cycle, where the ribosome recognizes a **stop codon** and releases the completed polypeptide chain. This process is **codon-directed**, **non-enzymatic**, and requires the coordinated action of **protein release factors (RFs)** that mimic the structure of tRNA but do not carry amino acids. Accurate and efficient termination is essential to ensure the **fidelity of gene expression** and to prevent the production of truncated or misfolded proteins.

2. Stop Codons and Signal Recognition

The genetic code includes **three termination codons**:

- **UAA (Ochre)**
- **UAG (Amber)**
- **UGA (Opal)**

These codons do **not correspond to any aminoacyl-tRNA** and are thus interpreted by **protein release factors** that recognize them and catalyze **peptidyl-tRNA hydrolysis**.

3. Termination Machinery in Prokaryotes vs. Eukaryotes

Feature	Prokaryotes	Eukaryotes
Main Release Factors	RF1 (UAA, UAG), RF2 (UAA, UGA)	eRF1 (recognizes all stop codons)
GTPase Supporting Factor	RF3	eRF3
Ribosome Size	70S (50S + 30S)	80S (60S + 40S)
Termination Complex	Peptidyl-tRNA + RF + GTP + RF3	Peptidyl-tRNA + eRF1/eRF3-GTP

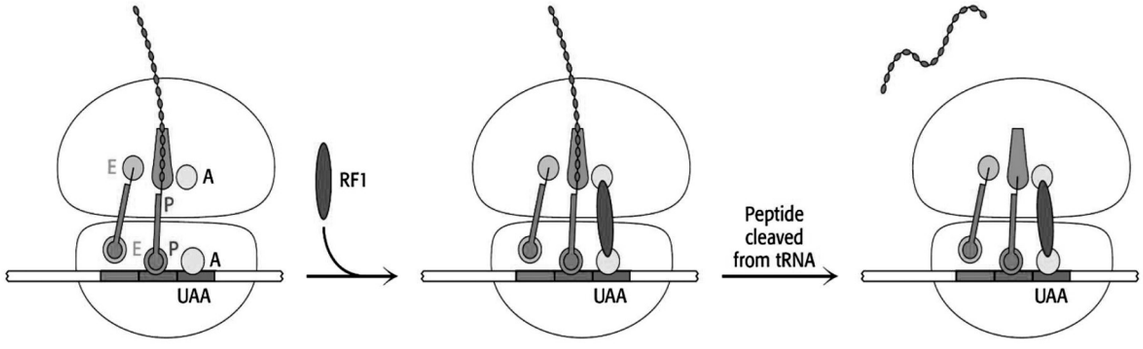


FIGURE 5.12: Termination of protein synthesis. A release factor recognizes a stop codon in the A site and stimulates the release of the completed protein from the tRNA in the P site.

4. Mechanism of Termination in Eukaryotes

Step 1: Stop Codon Recognition

- When a **stop codon** enters the **A site**, no tRNA can base-pair.
- **eRF1**, a **tRNA mimic**, binds to the stop codon in the A site.
- **eRF1 structurally resembles tRNA**, with domains that recognize the stop codon and catalyze hydrolysis.

Step 2: Peptidyl-tRNA Hydrolysis

- **eRF3**, a **GTP-binding protein**, forms a complex with eRF1 and the ribosome.
- GTP hydrolysis by eRF3 drives conformational changes that:
- Stabilize eRF1 at the A site
- Induce **hydrolysis of the ester bond** between the polypeptide and tRNA at the P site
- The released **polypeptide chain** exits through the **ribosomal exit tunnel**.

Step 3: Ribosome Recycling

- After peptide release:
- **eRF1 and eRF3** dissociate
- Additional factors such as **ABCE1 (an ATPase)** and **ribosome recycling factors (RRF)** assist in splitting the ribosome into **40S and 60S subunits**
- Deacylated tRNA and mRNA are released

Although **ABCE1** is not covered in Stryer, its role is crucial in **ribosome disassembly and turnover** in eukaryotic systems.

5. Structural Insights and Fidelity of Termination

- eRF1 binds **all three stop codons** with similar efficiency using **conserved motifs (e.g., NIKS, GGQ)**.
- The **GGQ motif** (Gly-Gly-Gln) in the catalytic domain is critical for **positioning water** to hydrolyze the peptidyl-tRNA linkage.
- This reaction occurs at the **peptidyl transferase center (PTC)** in the **28S rRNA** of the large subunit, confirming its **ribozyme activity**.

6. Clinical and Evolutionary Relevance

A. Readthrough and Suppression

- In certain contexts, **stop codons may be bypassed** ("readthrough") by near-cognate tRNAs or induced by **mutations** or **viral elements**.
- This can result in **C-terminal extensions** or translational recoding.
- In therapeutics, small molecules like **Ataluren** promote **readthrough** to treat diseases like **Duchenne muscular dystrophy** (caused by nonsense mutations).

B. Premature Termination Codons (PTCs)

- Found in many genetic diseases (e.g., β-thalassemia, cystic fibrosis)
- Lead to **nonfunctional proteins** or **nonsense-mediated decay (NMD)**
- The interplay between **termination efficiency and mRNA surveillance** determines disease severity.

7. Termination vs. Ribosome Rescue

Under **stress** or **defective mRNA conditions** (e.g., no stop codon or truncated mRNA), the ribosome may **stall**. Eukaryotic cells employ **rescue mechanisms** involving:

- **Dom34 and Hbs1** (homologs of eRF1 and eRF3) that recognize stalled ribosomes and trigger recycling.
- **No-Go Decay (NGD)** and **Non-Stop Decay (NSD)** to eliminate faulty transcripts.

Conclusion

Termination of translation is a **precisely orchestrated and highly conserved process** essential for releasing completed proteins and resetting the translational machinery for future use. The involvement of **release factors that mimic tRNA**, the catalytic role of **ribosomal RNA**, and the regulation by **GTPases** highlight the intricate design of this essential process. Understanding termination also reveals key checkpoints for **therapeutic interventions**, particularly in **genetic disorders** and **translational recoding strategies**.

H. Inhibitors of Translation

1. Introduction

The process of translation is a complex and highly conserved mechanism involving ribosomes, tRNAs, aminoacyl-tRNA synthetases, elongation factors, and other auxiliary proteins. Due to its essential role in protein biosynthesis, **translation is a prime target for both natural and synthetic inhibitors**, many of which are used as **antibiotics**, **anticancer agents**, or **molecular probes**.

Translation inhibitors function by interfering with one or more steps of the translational cycle—**initiation, elongation, translocation, or termination**—and can act **selectively** on prokaryotic or eukaryotic ribosomes based on structural differences.

2. Major Translation Inhibitors and Their Mechanisms

The table 5.1 below summarizes key inhibitors, their molecular targets, and the affected organisms:

Inhibitor	Target / Mechanism	Effect on Translation	Target Organism
Streptomycin	Binds to 30S subunit	Causes misreading of codons and inhibits initiation	Prokaryotes
Tetracycline	Binds 30S subunit A site	Blocks aminoacyl-tRNA entry into the A site	Prokaryotes

Chloramphenicol	Binds 50S subunit	Inhibits peptidyl transferase activity	Prokaryotes
Erythromycin	Binds 50S subunit near the exit tunnel	Blocks translocation by inhibiting ribosome movement	Prokaryotes
Puromycin	Structural analog of aminoacyl-tRNA	Premature chain termination by acting as a peptidyl acceptor	Both Prokaryotes and Eukaryotes
Cycloheximide	Binds 60S subunit	Inhibits peptidyl transferase activity	Eukaryotes
Diphtheria Toxin	Inactivates eEF2 by ADP-ribosylation	Halts translocation in elongation	Eukaryotes
Ricin	Removes an adenine from 28S rRNA	Disrupts ribosome function by depurinating a critical nucleotide	Eukaryotes

3. Detailed Mechanisms of Selected Inhibitors

A. Streptomycin

- An aminoglycoside antibiotic.
- Binds to the **16S rRNA** of the **30S subunit**, distorting its structure.
- Leads to **incorrect tRNA selection**, causing **translation errors** and eventual inhibition of initiation.
- Clinical use: **Tuberculosis** and other bacterial infections.
- **Resistance** may arise from **modifying enzymes** or ribosomal mutations.

B. Tetracycline

- Binds reversibly to the **30S A site**.
- Prevents **aminoacyl-tRNA binding**, halting elongation.
- Used against **Gram-positive and Gram-negative bacteria**.

C. Chloramphenicol

- Inhibits the **peptidyl transferase center** on the **50S subunit**.
- Competitively blocks the **peptidyl-tRNA and aminoacyl-tRNA interaction**.
- Affects **mitochondrial ribosomes** in humans due to structural similarity with bacterial ribosomes.
- Can cause **bone marrow suppression**.

D. Erythromycin

- A **macrolide antibiotic**.
- Binds to the **50S ribosome** near the **peptide exit tunnel**, preventing **elongation and translocation**.
- Used in respiratory and skin infections.

E. Puromycin

- Mimics the 3′ end of an **aminoacyl-tRNA**.
- Enters the **A site**, accepts the growing polypeptide chain, and terminates elongation **prematurely**.
- Non-selective: affects both **prokaryotic and eukaryotic ribosomes**.
- Used as a **molecular biology tool** to monitor translation.

F. Cycloheximide

- Eukaryotic-specific inhibitor.
- Blocks **peptidyl transferase activity** on the **60S subunit**.
- Halts **elongation**.
- Widely used in research to study **protein half-lives**, **ribosome profiling**, and **translation stress**.

G. Diphtheria Toxin

- A **bacterial exotoxin** (produced by *Corynebacterium diphtheriae*).
- Catalyzes **ADP-ribosylation of eEF2**, rendering it inactive.
- eEF2 is essential for **translocation**.
- Results in **complete inhibition of protein synthesis**, leading to **cell death**.
- Mechanism exploits **elongation factor specificity** in eukaryotic systems.

H. Ricin

- A **plant-derived toxin** from *Ricinus communis* (castor bean).
- **Depurinates an adenine** in the **28S rRNA** of the **eukaryotic large subunit**.
- Irreversibly damages the **sarcin–ricin loop**, inactivating the ribosome.
- Among the most **potent biological toxins** known.

4. Selectivity: Why Do Some Inhibitors Target Only Prokaryotes?

The difference in **ribosome structure** (70S in prokaryotes vs. 80S in eukaryotes) accounts for **selective toxicity**:

- Antibiotics like **tetracycline, chloramphenicol, and erythromycin** exploit **prokaryote-specific rRNA sequences and ribosomal proteins**.
- **Mitochondria**, which have **prokaryotic-like ribosomes**, may be **accidentally targeted** by these antibiotics, explaining certain **side effects** (e.g., chloramphenicol-induced aplastic anemia).

5. Biomedical Applications of Translation Inhibitors

- **Antibiotic therapy**: First-line and second-line treatments for bacterial infections.
- **Cancer research**: Cycloheximide and puromycin used to study **protein synthesis rates**, **tumour proliferation**, and **ribosome biogenesis**.
- **Toxin-based therapies**: Modified diphtheria or ricin used in **targeted immunotoxins** for cancer therapy.
- **Ribosome profiling**: Puromycin and cycloheximide help in **translational landscape mapping** of cells.

6. Resistance Mechanisms

- **Efflux pumps**: Eject antibiotics from bacterial cells (e.g., tetracycline resistance).
- **Ribosomal mutations**: Alter binding sites (e.g., streptomycin resistance).
- **Enzymatic modification**: Bacterial enzymes inactivate drugs (e.g., aminoglycoside-modifying enzymes).

Understanding these mechanisms helps in designing **next-generation antibiotics** and **combination therapies**.

Conclusion

Translation inhibitors serve as powerful tools for both **clinical medicine** and **molecular biology research**. Their specificity for different steps of translation, and even for different domains of life, underscores the **evolutionary divergence of ribosomes** and the **conserved mechanisms of protein synthesis**. In an age of rising antimicrobial resistance, these molecules remain at the forefront of therapeutic innovation.

Probable Questions:

1. Discuss the energetics of amino acid polymerization during translation. Explain the energy costs associated with each step of peptide bond formation.
2. Describe the structural features and post-transcriptional modifications of tRNAs. How do these features contribute to translational fidelity?
3. Explain the role of aminoacyl-tRNA synthetases in ensuring the accuracy of translation. Differentiate between Class I and Class II synthetases with suitable examples.
4. What mechanisms are employed by aminoacyl-tRNA synthetases to prevent misacylation? Discuss the double-sieve model in detail.
5. Illustrate the process of translation initiation in eukaryotes. How is it regulated by phosphorylation events and the mTOR signaling pathway?
6. Outline the steps involved in eukaryotic translation elongation. How is fidelity maintained during codon recognition and peptide bond formation?
7. What are the regulatory mechanisms that modulate translation elongation in eukaryotic cells under stress conditions? Discuss the role of eEF2 phosphorylation.
8. Describe the molecular events involved in the termination of translation in eukaryotes. How do release factors mediate peptide release and ribosome recycling?
9. Explain how internal ribosome entry sites (IRES) function as an alternative mechanism of translation initiation. What is their biological significance?
10. Enumerate the major inhibitors of translation and their specific targets. How do structural differences between prokaryotic and eukaryotic ribosomes account for inhibitor specificity?

Suggested Readings:

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2008). Molecular biology of the cell (6th ed.). Garland Science.
2. Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Martin, K. C., Yaffe, M., & Amon, A. (2021). Molecular Cell Biology (9th ed.). W. H. Freeman.
3. Berg, J. M., Tymoczko, J. L., Gatto, G. J., & Stryer, L. (2022). Biochemistry (10th ed.). W. H. Freeman

Unit-VI
Basic recombinant DNA techniques

Recombinant DNA technology is a technique used in genetic engineering that involves the identification, isolation and insertion of gene of interest into a vector such as a plasmid or bacteriophage to form a recombinant DNA molecule and production of large quantities of that gene fragment or product encoded by that gene.

Cutting and joining DNA molecules:

Many bacteria have enzymes that recognize specific DNA sequences (usually 4 or 6 nucleotides) and then cut the double stranded DNA helix at this sequence (Fig. 6.1). These enzymes are called site-specific restriction endonucleases, or more simply "restriction enzymes", and they naturally function as part of bacterial defenses against viruses and other sources of foreign DNA. To cut DNA at known locations, researchers use restriction enzymes that have been purified from various bacterial species, and which can be purchased from various commercial sources. These enzymes are usually named after the bacterium from which they were first isolated. For example, EcoRI and EcoRV are both enzymes from E. coli. EcoRI cuts double stranded DNA at the sequence GAATTC, but note that this enzyme, like many others, does not cut in exactly the middle of the restriction sequence (Figure 8.4.88.4.8). The ends of a molecule cut by EcoRI have an overhanging region of single stranded DNA, and so are sometimes called sticky-ends. On the other hand, EcoRV is an example of an enzyme that cuts both strands in exactly the middle of its recognition sequence, producing what are called blunt-ends, which lack overhangs.

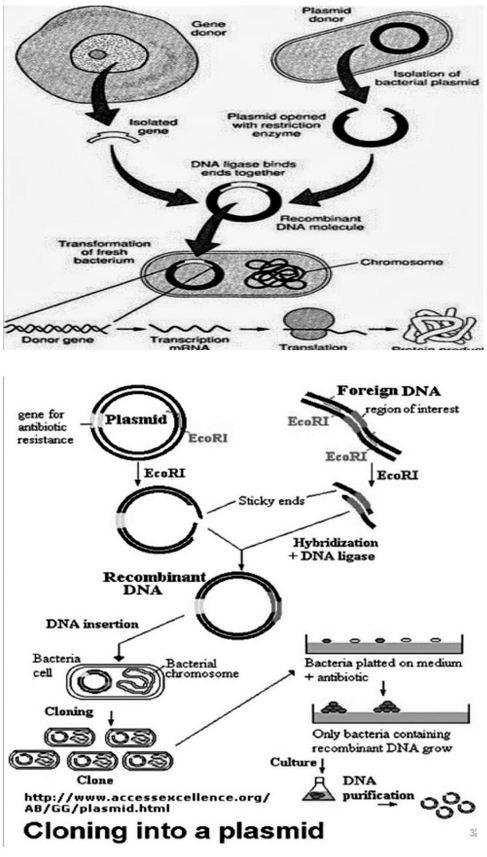


Figure 6.1

The process of **DNA ligation** occurs when DNA strands are covalently joined, end-to-end through the action of an enzyme called **DNA ligase**. Sticky-ended molecules

with complementary overhanging sequences are said to have **compatible ends**, which facilitate their joining to form recombinant DNA. Likewise, two blunt-ended sequences are also considered compatible to join together, although they do not ligate together as efficiently as sticky-ends. Note: sticky-ended molecules with non-complementary sequences will not ligate together with DNA ligase. Ligation is therefore central to the production of recombinant DNA, including the insertion of a double stranded DNA fragment into a plasmid vector. Progress in any scientific discipline is dependent on the availability of technologies, tools and methods that extend the range and sophistication of experiments, which may be performed. Over the last 30 years or so, this has been demonstrated in a spectacular way by the emergence of genetic engineering. This field has grown rapidly to the point where, in many laboratories around the world, it is now routine practice to isolate a specific DNA fragment from the genome of an organism, determine its base sequence and assess its function. The technology is also now used in many other applications, including forensic analysis of scene-of-crime samples, paternity disputes, medical diagnosis, genome mapping and sequencing and the biotechnology industry. What is particularly striking about the technology of gene manipulation is that it is readily accessible to individual scientists with well-established laboratory protocols and methodologies.

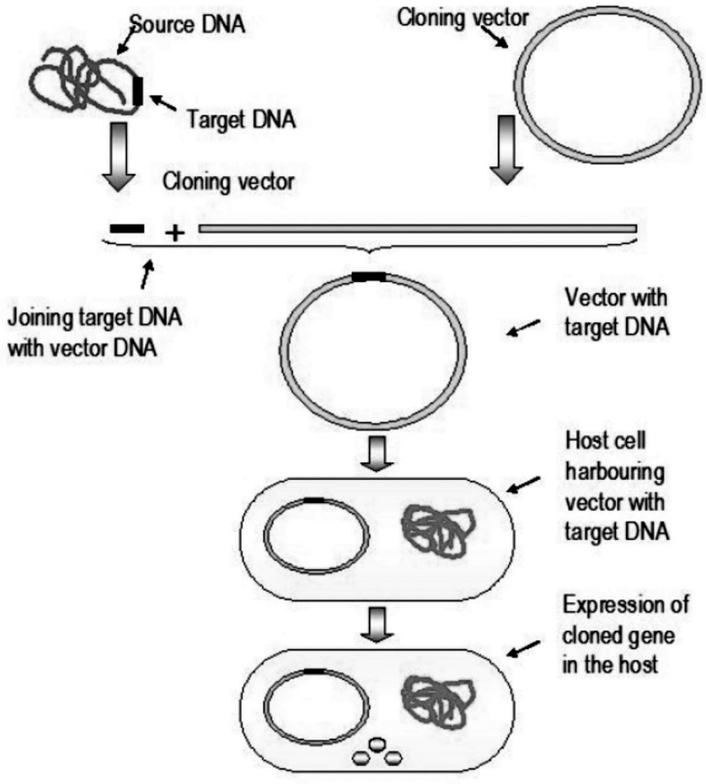


Figure 6.2: Basic steps involved in a rDNA technology

Although there are many, diverse and complex techniques involved, the basic principles of gene manipulation are reasonably simple. The premise on which the technology is based is that genetic information, encoded by DNA and arranged in the form of genes, is a resource that can be manipulated in various ways to achieve certain goals in both pure and applied science by adopting the following steps: 1) generating DNA fragments using restriction endonucleases, 2) ligation strategies, 3) cloning strategies, 4) modes of introduction of cloned vectors into hosts and 5) selection strategies for cloned genes

Generating DNA fragments using restriction endonucleases

Restriction endonucleases are enzymes that recognize specific sequences within duplex DNA molecules and cut the DNA at or near these sites. More than 500 different restriction endonucleases have been discovered. These enzymes can be grouped into three types viz. Type I, II and III. For practical purposes, the Type I and III restriction enzymes are not much used in rDNA technology. The real precision scissors are the Type II enzymes. Type II restriction endonucleases recognize and cut DNA within particular sequences of tetra, penta, hexa or hepta nucleotides which have an axis of rotational symmetry. In the following examples, different restriction enzymes cut the DNA at specific sequences as indicated by arrows.

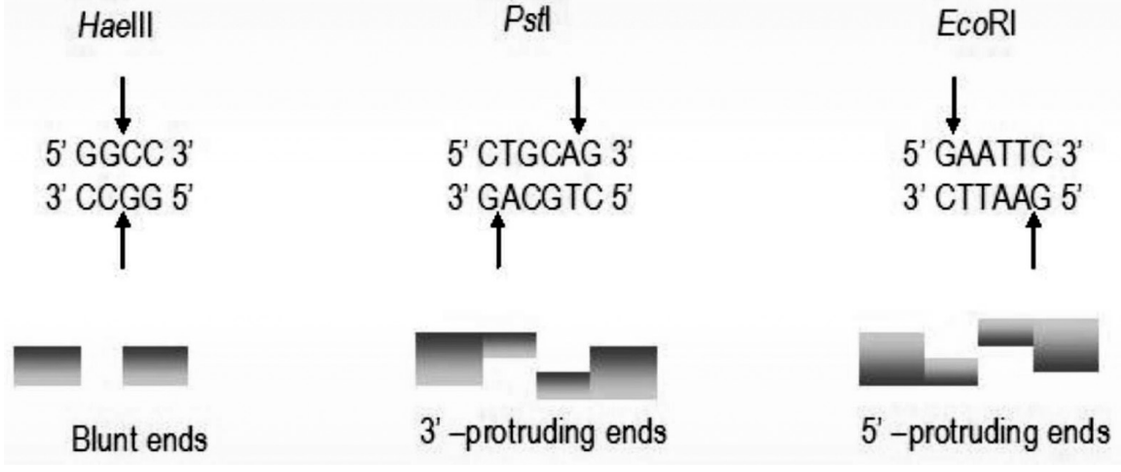


Figure 6.3

Among the restriction enzymes, some enzymes cut the DNA molecules to give **blunt end** fragments otherwise termed as **flush end** DNA fragments and some others produce DNA molecules where one of the strands will have protruding 5' or 3' termini. These fragments are called fragments with **cohesive ends** or **sticky ends**. The majority of the recognition sequences for restriction endonucleases are **palindromic**, that is the sequence is the same if read from 5' to 3' from both complementary strands.

The sites of cut made by endonucleases are called **target sites** or **cleavage sites** and the number of these sites in a DNA molecule depends on the size of the DNA, its base composition and the GC content of the recognition site. The number and size of the fragments generated by a restriction enzyme depends on the frequency of occurrence of the target site in the DNA to be cut. Assuming a DNA molecule with a 50 percent G+C content and a random distribution of four bases, a restriction enzyme recognizing a particular tetranucleotide sequence will be able to cut the DNA molecules into fragments at once in every 4^4 (i.e. 256) nucleotide pairs. If the enzyme is having the property of making cuts in hexanucleotide sequences means, the cuts will be made at every 4^6 (i.e. 4096) nucleotide pairs and an eight nucleotide recognition sequence 4^8 (65536) base pairs.

Restriction enzymes that have the same recognition sequences can be isolated from different bacterial species. Such enzymes are called **isoschizomers**. An example is provided by **Mbol**(Moraxella boris) and **Sau3A** (Staphyhcoccusaureus), both of which recognize the sequence **GATC**. Furthermore, some restriction enzymes generate cohesive ends that can reanneal with identical termini produced by other enzymes. For instance, DNA cleaved with BamHI (GGATCC) has compatible ends with DNA cleaved with BglII, MboI, Sau3A, etc.

The number of restriction fragments made by an enzyme would be reduced if there is a methylation of restriction sites. In some cases, the DNA recognition by an enzyme will not be altered by methylation and enzymes of this nature are said to be **enzymes with star activity** e.g. EcoRl, BarnHI and Sal1.

Ligation strategies

In rDNA technology, sealing discontinuities in the sugar-phosphate chains, otherwise called as ligation, is vital step. This process is catalyzed by DNA ligase by repairing broken phophodiester bonds. During ligation, the enzyme's activity is influenced by factors such as 1) substrate specificity, 2) temperature and 3) salt concentration

Substrate specificity

The DNA ligase isolated from E. coli and T4 differ not only in their cofactor requirements but also in their substrate specificities. The physiological substrate for both enzymes is breakage point at phosphodiester bond between neighbouring 3' hydroxyl and 5' phosphate ends still held together by an intact complementary strand. Another substrate for either enzyme contains the open and staggered phosphodiester bonds formed through reassociation of the protruding termini of different DNA molecules generated by digestion with certain type II restriction endonucleases.

Among the two ligases, T4 DNA ligase is capable of ligating nicks in the RNA chains of double stranded RNA-DNA hybrids and annealing RNA termini with DNA strands. One of the most remarkable properties of T4 DNA ligase, which distinguishes from the bacterial DNA ligase, is its ability to accomplish blunt end ligation of double stranded DNA molecules.

Temperature

Reaction temperature is another important parameter which influences ligase activity. The optimal ligation temperature is 37°C, but at this temperature the hydrogen bond joint between the sticky end is unstable. The optimum temperature for ligating the cohesive termini is therefore a compromise between the rate of enzyme action and association of the termini and has been found by experiment to be in the range 4 -15°C.

Concentration

At constant ionic strength, preferences for intra-or-intermolecular reactions depend on the length of DNA fragments and the DNA concentration. The smaller the DNA fragment at a given DNA concentration, more is the intramolecular reactions, leading to circularization of DNA. At constant lengths, the probability of circularization increase with decreasing concentrations.

Ligation methods

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process which has been extensively used to create artificial recombinants. If the termini of DNA fragments are not compatible, there are other methods to ligate the fragments.

Cohesive end ligation

The cohesive end ligation is possible when both the foreign DNA to be cloned and the vector DNA possess the same molecular ends. The compatible sticky ends have been generated by cleavage with the same enzyme on the same recognition sequences of both foreign DNA and vector DNA. Using DNA ligase, these molecules can be ligated without any problem.

Very often it is necessary to ligate DNA fragments with different and non-compatible ends, or blunt ends with either staggered 3’ or 5’ ends. Incompatible DNA fragments with recessed ends can be ligated by modifying their ends by any one of the following methods viz., (i) filling in recessed 3’ termini and (ii) renewal of 5’ protruding termini.

Blunt end ligation

The E. coli DNA ligase will not catalyze blunt end ligation except under special reaction conditions of macromolecular crowding. The unique property of T4 DNA ligase was used to ligate DNA fragments with blunt ends involving short decameric oligonucleotides called **linkers**.

Using linkers

Short oligonucleotides (decamers) which contain sites for one or more restriction enzymes are used to facilitate the ligation process among the DNA fragments with blunt ends.

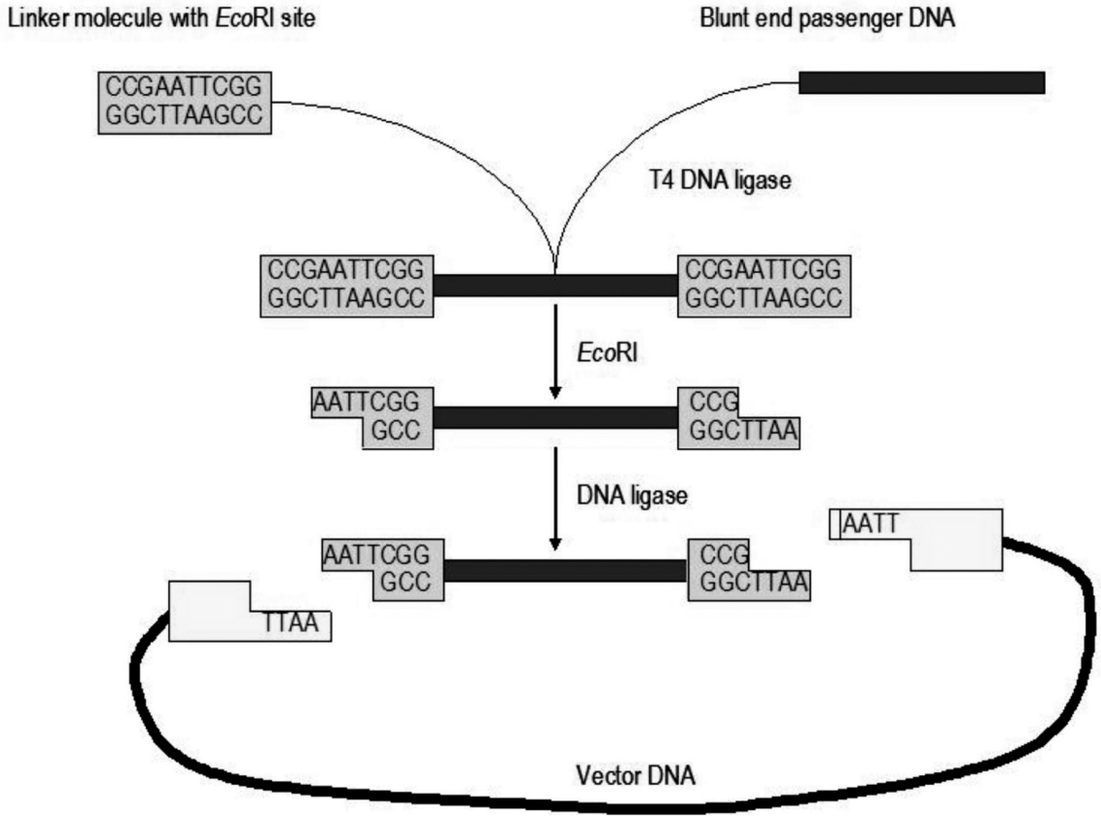


Figure 6.4: Joining of blunt end DNA to a vector using linkers

The linker molecules can be ligated to both ends of the foreign DNA to be cloned and then treated with restriction endonuclease to produce sticky end fragments which can be incorporated into a vector molecule that has been cut with the same restriction endonuclease. Insertion by means of the linker creates restriction sites at each end of the foreign DNA, and thus enables the foreign DNA excised and recovered after cloning and amplification in the host bacterium.

Using adaptors

The other strategy adopted for ligating DNA fragments with blunt ends is using **adaptors.** The adaptor molecules are synthetic deoxynucleotides that can be used to join two incompatible cohesive ends, two blunt ends or a combination of both. Such adaptors are of several types viz., **preformed**, **conversion** and **single stranded adaptors**.

Preformed adaptors

Preformed adaptors are short DNA duplexes with at least one cohesive end. The problem of internal cleavage of the insert DNA can be overcome by using a preformed adaptor that will introduce a new restriction site. For example, an adaptor having BamHI cohesive ends and sites HpaII and SmaI can be attached to passenger DNA and inserted into a BamHI in vector. After cloning, passenger DNA can be excised from the hybrid by using any one of the enzymes that recognize the restriction sites within the adaptor region.

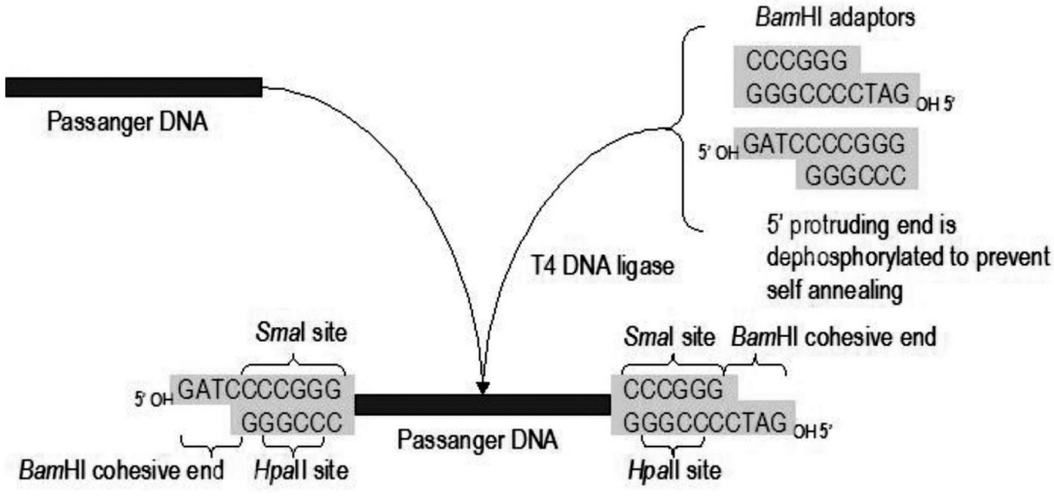


Figure 6.5

Use of preformed adaptors

Conversion adaptors

Conversion adaptors are synthetic oligonucleotides bearing different cohesive restriction termini. Such adaptors enable vector molecules that have been cleaved with one endonuclease to be joined to passenger fragments that have been cleaved with another. Often these adaptors contain internal restriction sites that permit recovery of the passenger fragment, for example, the EcoRI-BamHI adaptor contains a site for XhoI.

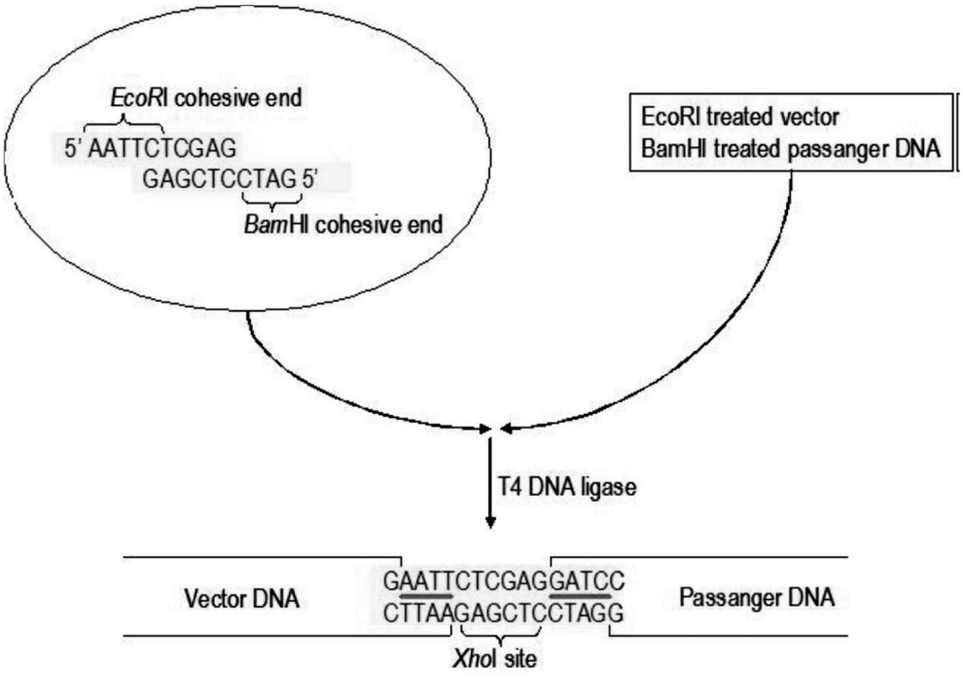


Figure 6.6

Single stranded adaptors

Single stranded adaptors can be used to make 3'-protruding cohesive ends compatible with 5' protruding ends. Such adaptors permit the insertion of passenger fragments into sites on vectors from which they would otherwise be precluded because of incompatible cohesive ends

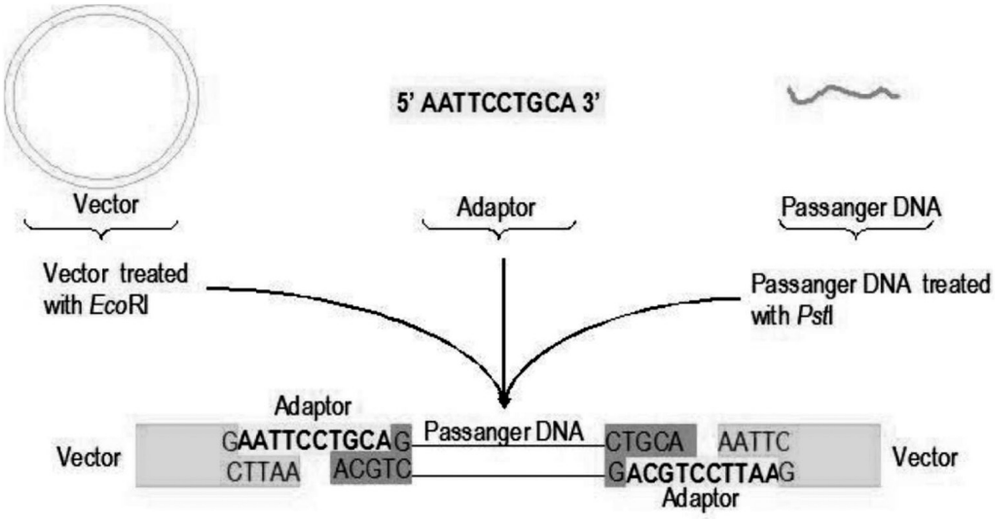


Figure 6.7

Homopolymer tailing

Homopolymer tailing is the other method adopted to clone blunt DNA molecules, especially cDNA molecules.

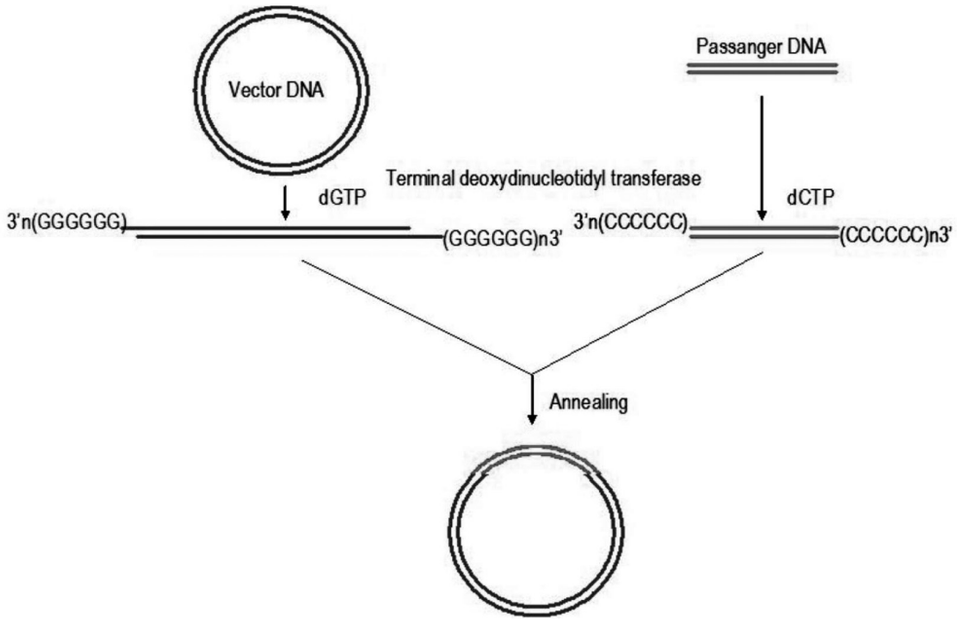


Figure 6.8

The addition of several nucleotides of single type to the 3' blunt end of DNA molecule is catalyzed by the enzyme **terminal deoxynucleotidyltransferase**. The terminal transferase permits the addition of complementary homopolymer tails (50 to 150 dA or dT long and about 20 dG or dC long) to 3' end of plasmid vector and passenger DNA. These tails can reanneal to form open circular hybrid molecules, which can be ligated in vitro or more commonly in vivo following transformations to produce functional recombinant molecules.

Restriction modification systems:

Restriction enzymes have been identified in the early 1950s of the past century and have quickly become key players in the molecular biology of DNA. Forty years ago, the scientists whose pioneering work had explored the activity and sequence specificity of these enzymes, contributing to the definition of their enormous potential as tools for DNA characterization, mapping and manipulation, were awarded the Nobel Prize. In this short review, we celebrate the history of these enzymes in the light of their many different uses, as these proteins have accompanied the history of DNA for over 50 years representing active witnesses of major steps in the field.

The restriction enzyme is a protein produced by bacteria that cleaves the DNA at specific sites. This site is known as the restriction site.

The restriction enzymes protect the live bacteria from bacteriophages. They recognize and cleave at the restriction sites of the bacteriophage and destroy its DNA.

Restriction enzymes are important tools for genetic engineering. They can be isolated from the bacteria and used in the laboratories.

The restriction enzymes recognize short and specific nucleotide sequences in the DNA known as the recognition sequences. When the restriction enzyme recognizes a DNA sequence, it hydrolyzes the bond between adjacent nucleotide and cuts through the DNA molecule.

The bacteria prevents its own DNA sequences from degradation by the addition of the methyl group at the adenine or cytosine bases within the recognition sequence with the help of enzyme methylases.

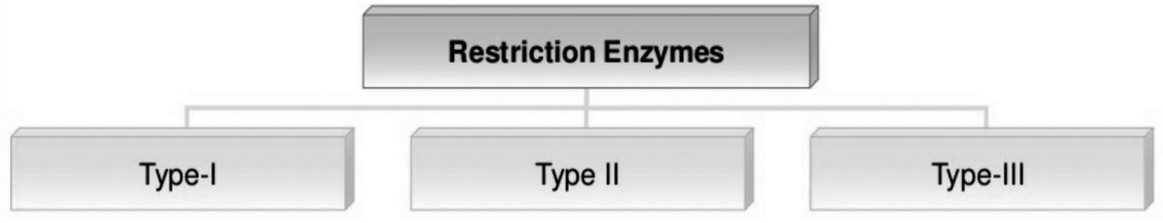


Figure 6.9

Type I Restriction Enzyme:

Do not generate specific fragments, require the Prescence of Mg^{2+}, ATP and S-adenosyl methionine; the latter activates the enzyme and tracks along the DNA for a variable distance before breaking. e.g. EcoKI.

2. Type II Restriction Enzyme:

Type II Restriction Enzyme cut within or immediately adjacent to target sequence and generates specific fragments. Mg^2 is essential for this enzyme, ATP and adenosyl methionine are not required as well. e.g. EcoRI.

3. Type III Restriction Enzyme:

Type III Restriction Enzyme requires ATP and S-adenosyl methionine for cleavage. It make breaks in DNA 25 bp away from recognition site. e.g. EcoP151.

Apart from this classification restriction endonuclease can also be classified into following types:

Blunt End Ligation

Endonucleases and exonucleases are the restriction enzymes used in molecular techniques. They cut the desired DNA portion. Usually, a straight cut creates blunt ends or non-overhanging ends. These ends can be joined using a DNA ligase enzyme. The joining of two blunt ends is called blunt end ligation. This does not need matching or complementary ends for ligation. e.g. **EcoRV, HaeIII, AluI**, and **SmaI** are examples of restriction enzymes that produce blunt ends.

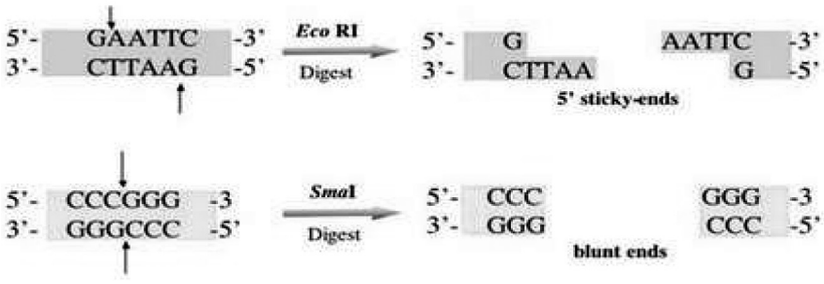


Figure 6.10

Sticky End Ligation

Likewise, the restriction enzymes creating a staggered cut leads to two sticky ends or overhanging ends. The ligation between two overhanging ends with matching or complementary bases is called sticky end ligation. This sticky end ligation is more efficient than blunt end ligation. Thus, it is the most desired process in cloning techniques. e.g. **EcoRI, BamHI**, and **HindIII** are examples of restriction enzymes that produce sticky ends.

The discovery of restriction enzymes arose from work with bacteriophage (viruses which infect bacteria). Restriction systems allow bacteria to monitor the origin of incoming DNA and to destroy it if it is recognized as foreign. Restriction endonucleases recognize specific sequences in the incoming DNA and cleave the DNA into fragments, either at specific sites or more randomly. When the incoming DNA is a bacteriophage genome, the effect is to reduce the efficiency of plating, i.e. to reduce the number of plaques formed in plating tests. The phenomena of restriction and modification were well illustrated and studied by the behaviour of phage λ on two E. coli host strains. If a stock preparation of phage λ, for example, is made by growth upon E. coli strain C and this stock is then titred upon E. coli C and E. coli K, the titres observed on these two strains will differ by several orders of magnitude, the titre on E. coli K being the lower. The phage are said to be restricted by the second host strain (E. coli K). When those phage that do result from the infection of E. coli K are now replated on E. coli

K they are no longer restricted; but if they are first cycled through E. coli C they are once again restricted when plated upon E. coli K (See figure). Thus the efficiency with which phage λ plates uponaparticular host strain depends upon the strain on which it was last propagated. This non-heritable change conferred upon the phage by the second host strain (E. coli K) that allows it to be replated on that strain without further restriction is called modification.

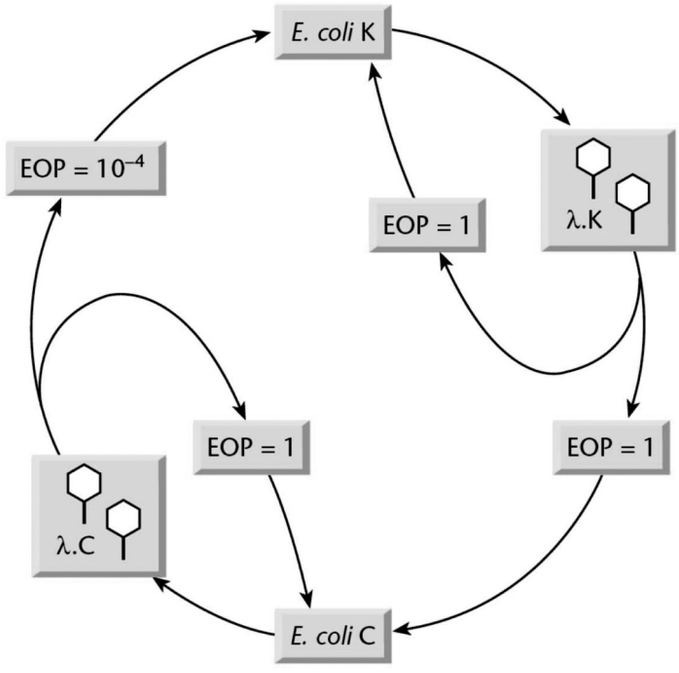


Figure 6.11

The restricted phages adsorb to restrictive hosts and inject their DNA normally. When the phage arelabelled with 3′, it is apparent that their DNA is degraded soon after injection and the endonuclease that is primarilyresponsible for this degradation is called a restriction endonuclease or restriction enzyme. The restrictive host must, of course, protect its own DNA from the potentially lethal effects of the restriction endonuclease and so its DNA must be appropriately modified. Modification involves methylation of certain bases at a very limited number of sequences within DNA, which constitute the recognition sequences for the restriction endonuclease. This explains why phage that survive one cycle of growth upon the restrictive host can subsequently reinfect that host efficiently; their DNA has been replicated in the presence of the modifying methylase and so it, like the host DNA, becomes methylated and protected from the restriction system. Although phage infection has been chosen as our example to illustrate restriction and modification, these processes can occur whenever DNA is transferred from one bacterial strain to another. λ. In 1978 Werner Arber, Hamilton

Smith and Daniel Nathans, were awarded a Nobel Prize for the discovery of restriction enzymes and their application to problems of molecular genetics.

Nomenclature of Restriction enzyme

The discovery of a large number of restriction and modification systems called for a uniform system of nomenclature. A suitable system was proposed by Smith and Nathans. The simplified version of this is in use today. The key features are:

- The species name of the host organism is identified by the first letter of the genus name and the first two letters of the specific epithet to generate a three letter abbreviation. This abbreviation is always written in italics.
- Where a particular strain has been the source then this is identified.
- When a particular host strain has several different R-M systems, these are identified by roman numerals.

Some examples are given in table below. Homing endonucleases are named in a similar fashion except that intron-encoded endonucleases are given the prefix 'I-' (e.g. I-CeuI) and intein endonucleases have the prefix 'PI-' (e.g. Pl-PspI). Where it is necessary to distinguish between the restriction and methylating activities, they are given the prefixes 'R' and 'M', respectively, e.g. R.SmaI and M.SmaI.

Table 6.1: Recognition sequences and cutting sites for some restriction endonucleases

Enzyme	Recognition sequence	Cutting sites	Ends
*Bam*HI	5′-GGATCC-3′	G↓GATCC CCTAG↑G	5′
*Eco*RI	5′-GAATTC-3′	G↓AATTC CTTAA↑G	5′
*Hae*III	5′-GGCC-3′	GG↓CC CC↑GG	Blunt
*Hpa*I	5′-GTTAAC-3′	GTT↓AAC CAA↑TTG	Blunt
*Pst*I	5′-CTGCAG-3′	CTGCA↓G G↑ACGTC	3′
*Sau*3A	5′-GATC-3′	↓GATC CTAG↑	5′
*Sma*I	5′-CCCGGG-3′	CCC↓GGG GGG↑CCC	Blunt
*Sst*I	5′-GAGCTC-3′	GAGCT↓C	3′

		C↑T C G A G	
*Xma*I	5′-CCCGGG-3′	C↓C C G G G	5′
		G G G C C↑C	

Note: The recognition sequences are given in single-strand form, written 5′→3′. Cutting sites are given in double-stranded form to illustrate the type of ends produced by a particular enzyme; 5′ and 3′ refer to 5′- and 3′-protruding termini, respectively. The point at which the phosphodiester bonds are broken is shown by the arrow on each strand of the recognition sequence.

Various enzymes used in Recombinant DNA technology:

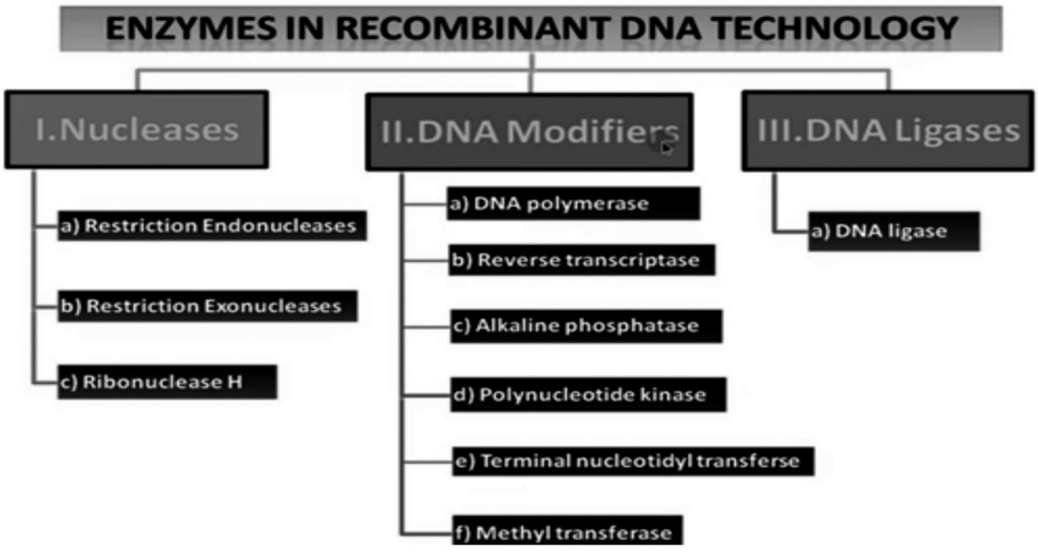


Figure 6.12

1. Nucleases(Nucleic acid degrading enzymes):

i. **Restriction Endonuclease:** Also known as molecular scissors. Makes cut at specific sites within DNA called as restriction sites or recognition sequences. E.g. EcoRI, BamHI, HindIII, and SmaI.

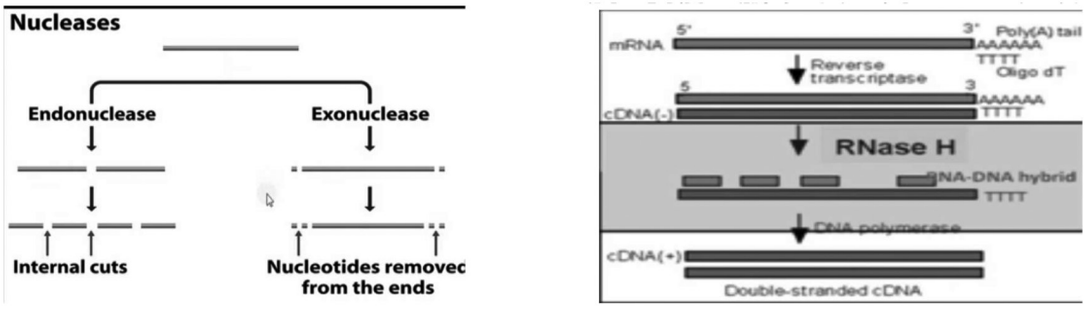


Figure 6.13

ii. Restriction Exonuclease: It removes nucleotides from the ends. E.g. EcoRI, HindII.

iii. Ribonuclease H: Rnase H is a RNA degrading enzyme that selectively removes mRNA from RNA-DNA hybrids to separate the cDNA to synthesize second strand.

2. DNA Modifiers:

i. DNA Polymerase: DNA polymerase is a ubiquitous enzyme that synthesizes complementary DNA strands according to the template DNA in living cells. Multiple enzymes have been identified from each organism, and the shared functions of these enzymes have been investigated. In addition to their fundamental role in maintaining genome integrity during replication and repair, DNA polymerases are widely used for DNA manipulation in vitro, including DNA cloning, sequencing, labeling, mutagenesis, and other purposes. The fundamental ability of DNA polymerases to synthesize a deoxyribonucleotide chain is conserved. However, the more specific properties, including processivity, fidelity (synthesis accuracy), and substrate nucleotide selectivity, differ among the enzymes.

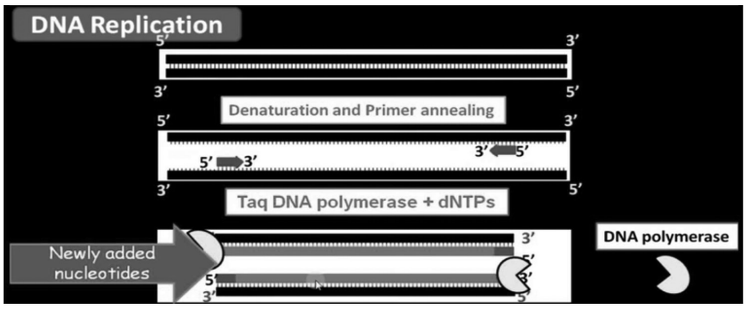


Figure 6.14

ii. Reverse Transcriptase: Reverse Transcriptase is an enzyme that converts RNA into DNA, commonly found in retroviruses like HIV. It is used in molecular biology research to create complementary DNA strands from RNA templates, allowing for the amplification of RNA sequences similar to DNA.

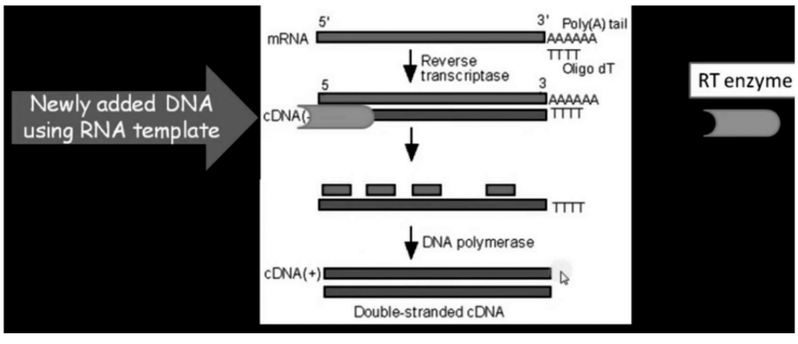


Figure 6.15

iii. **Alkaline Phosphatase:** Alkaline phosphatase plays a crucial role in recombinant DNA technology by preventing the self-ligation of DNA, which is undesirable during cloning. It removes 5' phosphate groups from DNA, making it unable to re-circularize or ligate to itself after restriction enzyme digestion. This allows for the efficient insertion of foreign DNA into vectors. In radioactive labelling using P^{32}, phosphate group at the 5' end is removed and replaced with P^{32} using Alkaline Phosphatase.

iv. **Polynucleotide Kinase:** Polynucleotide kinase (PNK) is an enzyme that catalyzes the transfer of the gamma-phosphate from ATP to the 5'-hydroxyl terminus of polynucleotides, including DNA and RNA. **It's uses are**: to phosphorylate the 5' end of dephosphorylated DNA(vector DNA) or RNA and used in radiolabelling as well, it transfers radioactive P^{32} from ATP to dephosphorylate 5' end of DNA or RNA.

v. **Terminal nucleotidyl transferase:** catalyzes the elongation of single-stranded DNA (ssDNA) by adding dNTPs to the 3′-end of a primer, facilitated by metal ions. This reaction is irreversible because of the lack of exchange between free pyrophosphate and nucleoside triphosphate, of which elongation is the primary controlling factor. It is used to make homopolymer sticky or cohesive tails at 3' end of DNA fragment which helps in joining fragments with blunt ends and also used to make radioactive DNA probes by end labelling.

vi. **Methyl transferase:** A Methyltransferase is an enzyme that is responsible for transferring methyl groups to various proteins, phospholipids, and nucleotides in mammals, playing a crucial role in cellular functions such as DNA and RNA synthesis, methylation, and gene expression. It methylate and protects end of DNA from action of restriction enzyme.

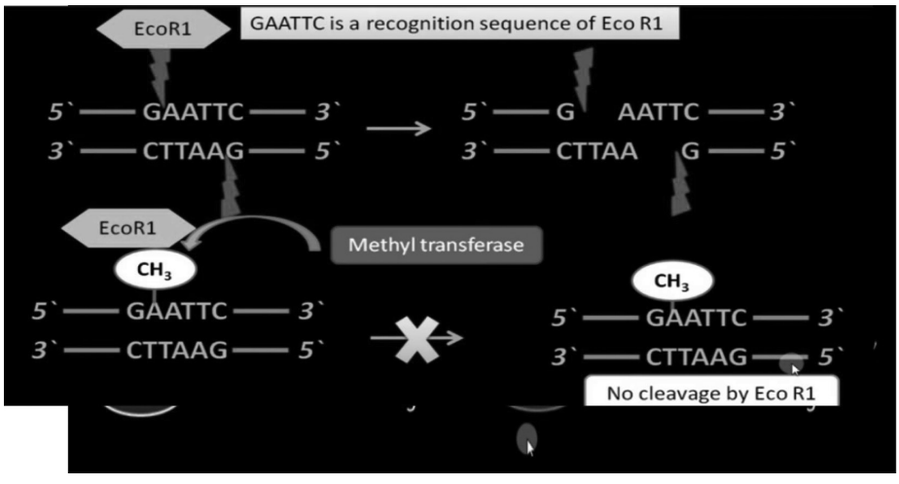


Figure 6.16

3. DNA Ligases:

It is joining enzyme that joins the ends of 2 ds DNA molecule, the process is known as ligation. The bond formed here is called phosphodiester bond. Between the 5'P of a nucleotide of one DNA fragment and the 3'OH end of another and it requires ATP and NAD^+ for its activity.

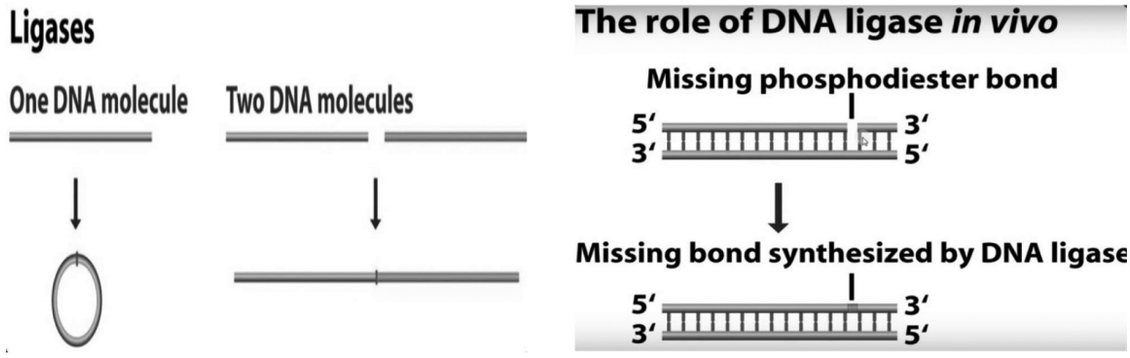


Figure 6.17

Restriction Maps and Mapping Techniques:

Most pieces of DNA will have recognition sites for various restriction enzymes, and it is often beneficial to know the relative locations of some of these sites. The technique used to obtain this information is known as **restriction mapping**. This involves cutting a DNA fragment with a selection of restriction enzymes, singly and in various combinations. The fragments produced are run on an agarose gel and their sizes determined. From the data obtained, the relative locations of the cutting sites can be worked out. A fairly simple example can be used to illustrate the technique, as outlined below. Let us say that we wish to map the cutting sites for the restriction enzymes BamHI, EcoRI and PstI,and that the DNA fragment of interest is 15 kb in length. Various digestions are carried out, and the fragments arising from these are analyzed and their sizes determined. The results obtained are shown in Table T. As each of the single enzyme reactions produces two DNA fragments, we can conclude that the DNA has a single cutting site for each enzyme. The double digests enable a map to be drawn up, and the triple digest confirms this. Construction of the map is outlined in Fig. 4.4.

Table 6.2: Digestion of a 15 kb DNA fragment with three restriction enzymes

*Bam*HI	*Eco*RI	*Pst*I	*Bam*HI + *Eco*RI	*Bam*HI + *Pst*I	*Eco*RI + *Pst*I	*Bam*HI + *Eco*RI + *Pst*I
14	12	8	11	8	7	6
1	3	7	3	6	5	5
			1	1	3	3
						1

Note: Data shown are lengths (in kb) of fragments that are produced on digestion of a 15 kb DNA fragment with the enzymes *Bam*HI, *Eco*RI and *Pst*I. Single, double and triple digests were carried out as indicated.

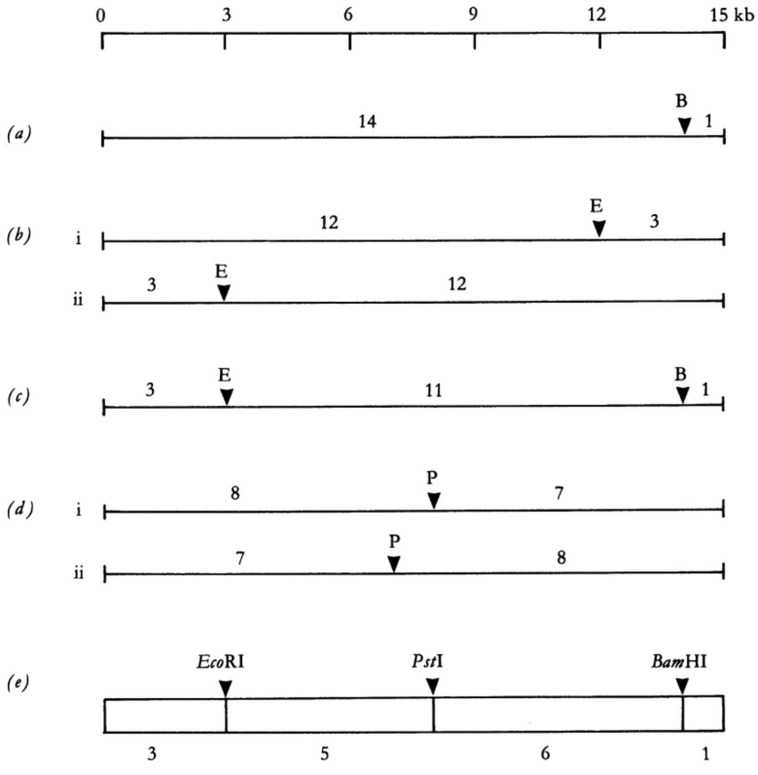


Figure 6.18

Restriction mapping. (a) The 15 kb fragment yields two fragments of 14 and 1 kb when cut with BamHI. (b) The EcoRI fragments of 12 and 3 kb can be orientated in two ways with respect to the BamHI site. The BamHI/EcoRI double digest gives fragments of 11, 3 and 1 kb, and therefore the relative positions of the BamHI and EcoRI sites are as shown in (c). Similar reasoning with the orientation of the 8 and 7 kb PstI fragments (d) gives the final map (e).

Nucleic acid probes:

The power of nucleic acid hybridization lies in the fact that complementary sequences will bind to each other with a very high degree of fidelity. In practice this depends on the degree of **homology** between the hybridizing sequences, and usually the aim is to use a probe that has been derived from the same source as the target DNA. However, under certain conditions, sequences that are not 100% **homologous** can be used to screen for a particular gene, as may be the case if a probe from one organism is used to detect clones prepared using DNA from a second organism. Such **heterologous** probes have been extremely useful in identifying many genes from different sources.

There are three main types of DNA probe: (1) cDNA, (2) genomic DNA, and (3) oligonucleotides. Alternatively, RNA probes can be used if these are suitable. The availability of a particular probe will depend on what is known about the target gene sequence. If a cDNA clone has already been obtained and identified, the cDNA can be used to screen a genomic library and isolate the gene sequence itself. Alternatively, cDNA may be made from mRNA populations and used without cloning the cDNAs. This is often used in what is known as '**plus/minus' screening**. If the clone of interest contains a sequence that is expressed only under certain conditions, probes may be made from mRNA populations from cells that are expressing the gene (the plus probe) and from cells that are not expressing the gene (the minus probe). By carrying out duplicate hybridizations, the clones can be identified by their different patterns of hybridisation with the plus and minus probes. Although this method cannot usually provide a definitive identification of a particular sequence, it can be useful in narrowing down the range of candidate clones.

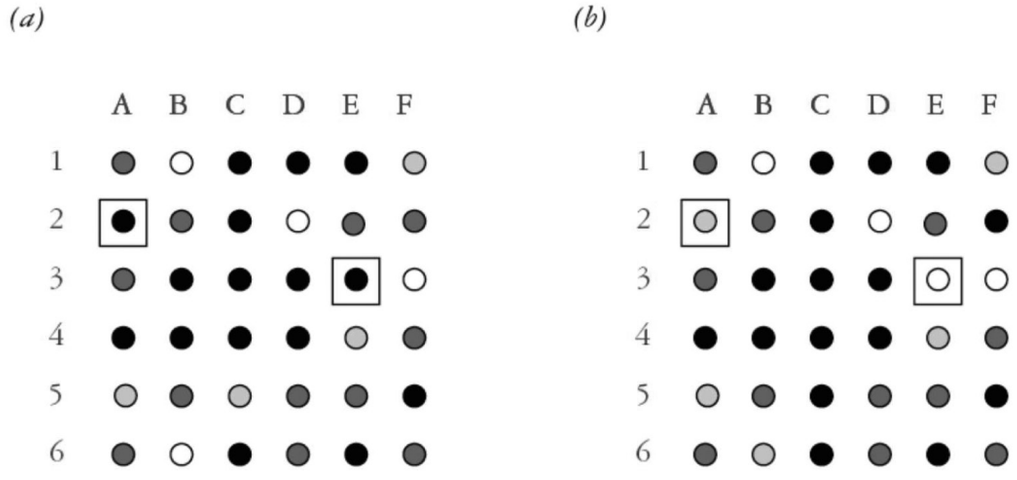


Figure 6.19

Screening by the 'plus/minus' method. In this example clones have been spotted on duplicate filters in a 6 × 6 array. (a) One copy of the filter has been hybridised with a radioactive cDNA probe prepared from cells expressing the gene of interest (the + probe). (b) The other copy of the filter has been hybridised with a probe prepared from cells that are not expressing the gene (the – probe). Some clones hybridise weakly with both probes (e.g. B1), some strongly with both (e.g. C1). Some show a strong signal with the – probe and a weaker signal with the + probe (e.g. C5). Clones A2 and E3 (boxed) show a strong signal with the + probe and a weaker signal with the – probe. These would be selected for further analysis.

Genomic DNA probes are usually fragments of cloned sequences that are used either as heterologous probes or to identify other clones that contain additional parts of the gene in question. This is an important part of the techniques known as **chromosome walking** and **chromosome jumping** and can enable the identification of overlapping sequences which, when pieced together, enable long stretches of DNA to be characterized.

The use of oligonucleotide probes is possible where some amino acid sequence data are available for the protein encoded by the target gene. Using the genetic code, the likely gene sequence can be derived and an oligonucleotide synthesized. The degenerate nature of the genetic code means that it is not possible to predict the sequence with complete accuracy, but this is not usually a major problem if the least degenerate sequence is used. As we saw when looking at primer design for PCR, a mixed probe can be synthesized that covers all the possible sequences by varying the base combinations at degenerate wobble positions. Alternatively, inosine can be used in highly degenerate parts of the sequence. The great advantage of oligonucleotide probes is that only a short stretch of sequence is required for the probe to be useful and, thus, genes for which clones are not already available can be identified by sequencing peptide fragments and constructing probes accordingly.

A major development in clone identification procedures has appeared hand-in-hand with genome sequencing projects. Assuming that the genome sequence for your target organism is available, a computer can be used to search for any particular sequence in the genome. This type of 'experiment' has become established as a very powerful method, and is sometimes referred to as doing experiments **'in silico'** (as in a silicon chip in the computer, rather than in vivo or in vitro). In some cases this might even remove the need for cloning at all, as the information might be used to prepare PCR primers so that the target fragment could then be amplified specifically. For a more conventional approach, the sequence data could be used to prepare an oligonucleotide probe to isolate the cloned fragment from a cDNA or genomic DNA library.

When a suitable probe has been obtained, it can be labelled with a radioactive isotope such as 32P. This produces a radiolabelled fragment of high specific activity that can be used as an extremely sensitive screen for the gene of interest. Alternatively, non-radioactive labelling methods such as fluorescent tags may be used if desired.

Blotting techniques:

Although a clone may have been identified and its restriction map determined, this information in itself does not provide much of an insight into the fine structure of the cloned fragment and the gene that it contains. Ultimately the aim may be to obtain the gene sequence, but it may be that it is not appropriate to begin sequencing straight away. If, for example, a 20 kb fragment of genomic DNA has been cloned in aλ replacement vector, and the area of interest is only 2 kb in length, then much effort would be wasted by sequencing the entire clone. In many experiments it is therefore essential to determine which parts of the original clone contain the regions of interest. This can be done by using a variety of methods based on blotting nucleic acid molecules onto membranes and hybridizing with specific probes. Such an approach is in some ways an extension of clone identification by colony or plaque hybridization, with the refinement that information about the structure of the clone is obtained.

The first blotting technique was developed by Ed Southern, and is eponymously known as **Southern Blotting**. In this method fragments of DNA, generated by restriction digestion, are subjected to agarose gel electrophoresis. The separated fragments are then transferred to a nitrocellulose or nylon membrane by a 'blotting' technique. The original method used capillary blotting, as shown in Fig. 8.9. Although other methods such as vacuum blotting and electroblotting have been devised, the original method is still used extensively because it is simple and inexpensive. Blots are often set up with whatever is at hand, and precarious-looking versions of the blotting apparatus are a common sight in many laboratories.

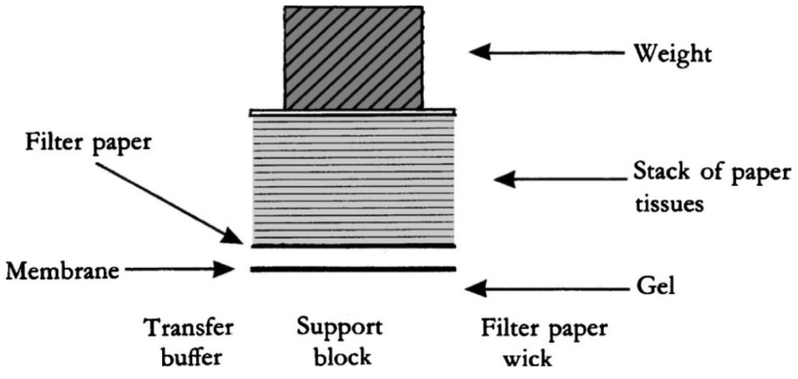


Figure 6.20

Blotting apparatus. The gel is placed on a filter paper wick and a nitrocellulose or nylon filter placed on top. Further sheets of filter paper and paper tissues complete the setup. Transfer buffer is drawn through the gel by capillary action, and the nucleic acid fragments are transferred out of the gel and onto the membrane.

When the fragments have been transferred from the gel and bound to the filter, it becomes a replica of the gel. The filter can then be hybridised with a radioactive probe in a similar way to colony or plaque filters. As with all hybridisation, the key is the availability of a suitable probe. After hybridisation and washing, the filter is exposed to X-ray film and an autoradiograph prepared, which provides information on the structure of the clone.

Although Southern blotting is a very simple technique, it has many applications and has been an invaluable method in gene analysis. The same technique can also be used with RNA, as opposed to DNA, and in this case it is known as **Northern blotting**. It is most useful in determining hybridisation patterns in mRNA samples and can be used to determine which regions of a cloned DNA fragment will hybridise to a particular mRNA. However, it is more often used as a method of measuring transcript levels during expression of a particular gene.

There are further variations on the blotting theme. If nucleic acid samples are not subjected to electrophoresis but are spotted onto the filters, hybridisation can be carried out as for Northern and Southern blots. This technique is known as **dot blotting** and is particularly useful in obtaining quantitative data in the study of gene expression. In some variants of the apparatus the nucleic acid is applied in a slot rather than as a dot. Not surprisingly, this is called **slot blotting**. The final technique is known as **Western blotting**, which involves the transfer of electrophoretically separated protein molecules to membranes. Often used with SDS-PAGE (polyacrylamide gel electrophoresis under denaturing conditions), Western blotting can identify proteins specifically if the appropriate antibody is available. The membrane with the proteins fixed in position is probed with the antibody to detect the target protein in a similar way to immunological screening of plaque lifts from expression libraries. Sometimes Western blots can be useful to measure the amount of a particular protein in the cell at any given time. By comparing with other data (such as amount of mRNA, and/or enzyme activity) it is possible to build up a picture of how the expression and metabolic control of the protein is regulated.

BLOTTING METHOD:

Southern blot:

A Southern blot is a method used in molecular biology for detection of a specific DNA sequence in DNA samples. Southern blotting combines transfer of electrophoresis-separated DNA fragments to a filter membrane and subsequent fragment detection by probe hybridization. The method is named after the British biologist Edwin Southern who first published it in 1975. Other blotting methods (i.e., western blot northern blot, eastern blot, southwestern blot) that employ similar principles, but using RNA or protein, have later been named in reference to Edwin Southern's name. As the label is eponymous, Southern is capitalised, as is conventional for proper nouns. The names for other blotting methods may follow this convention, by analogy.

Principle:

Procedure for identifying specific sequences of DNA, in which fragments separated on a gel are transferred directly to a second medium on which assay by hybridization may be carried out.

Southern blotting is the transfer of DNA fragments from an electrophoresis gel to a membrane support, resulting in immobilization of the DNA fragments, so the membrane carries a semipermanent reproduction of the banding pattern of the gel. After immobilization, the DNA can be subjected to hybridization analysis (UNIT 10.6B), enabling bands with sequence similarity to a labeled probe to be identified. In addition to standard uses, immunologists have long used Southern blotting to identify genes that have undergone somatic rearrangements. When setting up a Southern transfer, choices must be made between different types of membrane, transfer buffer, and transfer method. The most popular membranes are made of nitrocellulose, uncharged nylon, or positively charged nylon. Although these materials have different properties, the three types of membranes are virtually interchangeable for many applications. The main advantage of nylon membranes is that they are relatively robust and so can be reprobed ten or more times before falling apart. Nitrocellulose membranes are fragile and can rarely be reprobed more than three times; however, these are still extensively used, as they give lower backgrounds with some types of hybridization probe. The properties and advantages of the different membranes are discussed more fully in the commentary.

The basic protocol describes Southern blotting via upward capillary transfer of DNA from an agarose gel onto a nylon or nitrocellulose membrane, using a high-salt transfer buffer to promote binding of DNA to the membrane. With the high-salt buffer, the DNA

becomes bound to the membrane during transfer but is not permanently immobilized. Immobilization is achieved by UV irradiation (for nylon) or baking (for nitrocellulose). A support protocol describes how to calibrate a UV transilluminator for optimal UV irradiation of a nylon membrane. The first alternate protocol details transfer using nylon membranes and an alkaline buffer, and is primarily used with positively charged nylon membranes. The advantage of this combination is that no post-transfer immobilization step is required, as the positively charged membrane binds DNA irreversibly under alkaline transfer conditions. The method can also be used with neutral nylon membranes, but less DNA will be retained. The second alternate protocol describes a transfer method based on a different transferstack setup. The traditional method of upward capillary transfer of DNA from gel to membrane described in the first basic and alternate protocols has certain disadvantages, notably the fact that the gel can become crushed by the weighted filter papers and paper towels that are laid on top of it. This slows down the blotting process and may reduce the amount of DNA that can be transferred. The downward capillary method described in the second alternate protocol is therefore more rapid than the basic protocol and can result in more complete transfer.

Southern blotting onto a nylon or nitrocellulose membrane with high-salt buffer:

This procedure is specifically designed for blotting an agarose gel onto an uncharged or positively charged nylon membrane. With the minor modifications detailed in the annotations, the same protocol can also be used with nitrocellulose membranes. The protocol is divided into three stages. First, the agarose gel is pretreated by soaking in a series of solutions that depurinate, denature, and neutralize the DNA and gel matrix. The second stage is the transfer itself, which occurs by upward capillary action. Finally, the DNA is immobilized on the membrane by UV irradiation (for nylon) or baking (for nitrocellulose). Materials DNA samples to be analyzed 0.25 M HCl Denaturation solution: 1.5 M NaCl/0.5 M NaOH (store at room temperature) Neutralization solution: 1.5 M NaCl/0.5 M Tris.Cl, pH 7.0 (store at room temperature) 20× and 2× SSC Oblong sponge slightly larger than the gel being blotted Whatman 3MM filter paper sheets Nylon or nitrocellulose membrane UV-transparent plastic wrap (e.g., Saran Wrap) for nylon membranes UV transilluminator or UV light box (e.g., Stratagene Stratalinker) for nylon membranes Additional reagents and equipment for restriction endonuclease digestion and agarose gel electrophoresis CAUTION: Wear gloves from step 2 of the protocol onward to protect your hands from the acid and alkali solutions and to protect the membrane from contamination.

Preparation of the gel:

1. Digest the DNA samples with appropriate restriction enzyme(s), electrophorese in an agarose gel with appropriate DNA size markers, stain with ethidium bromide, and photograph with a ruler laid alongside the gel so that band positions can later be identified on the membrane. The gel should contain the minimum agarose concentration needed to resolve bands in the area of interest and should be ≤7 mm thick, preferably less. The amount of DNA that must be loaded depends on the relative abundance of the target sequence that will subsequently be sought by hybridization probing.
2. Rinse the gel in distilled water and place in a clean glass dish containing ~10 gel volumes of 0.25 M HCl. Shake slowly on a platform shaker for 30 min at room temperature. This step results in partial depurination of the DNA fragments, which in turn leads to strand cleavage. The length reduction improves the transfer of longer molecules. The step is not necessary with PCR products 5 kb is not required. The time needed for the acid treatment depends on the concentration and thickness of the gel. To check that the treatment has been sufficient, watch the xylene cyanol and bromophenol blue dyes. When these change color to green and yellow respectively, the gel has equilibrated with the acid. Adequate depurination takes a further 10 min.
3. Pour off the HCl and rinse the gel with distilled water. Add ~10 vol denaturation solution and shake as before for 20 min. Replace with fresh denaturation solution and shake for a further 20 min. Denaturation unzips the DNA to give single-stranded molecules that have unpaired bases and are suitable for subsequent hybridization analysis.
4. Pour off the denaturation solution and rinse the gel with distilled water. Add ~10 vol neutralization solution, shake as before for 20 min, then replace with fresh neutralization solution and shake for a further 20 min. The aim of the neutralization step is to bring the gel pH down to 9.0, carry out a third washing in neutralization solution. Neutralization is less critical with a nylon membrane but should still be carried out.

Set up the transfer:

5. Using Figure 10.6.1 as a guide, place an oblong sponge, slightly larger than the gel, in a glass or plastic dish (if necessary, use two or more sponges placed side by side). Fill the dish with enough 20× SSC to leave the soaked sponge about

half-submerged in buffer. The sponge forms the support for the gel. Any commercial sponge will do; before a sponge is used for the first time, it should be washed thoroughly with distilled water to remove any detergents that may be present. As an alternative, use a solid support with wicks made out of Whatman 3MM filter paper. An electrophoresis tank should not be used, as the high-salt transfer buffer will corrode the electrodes.

6. Cut three pieces of Whatman 3MM paper to the same size as the sponge. Place these on the sponge and wet them with 20 × SSC.
7. Place the gel on the filter paper and squeeze out air bubbles by rolling a glass pipet over the surface.
8. Cut four strips of plastic wrap and place over the edges of the gel. This is to prevent the buffer from "short-circuiting"—i.e., so that it flows through rather than around the gel.
9. Cut a piece of nylon membrane just large enough to cover the exposed surface of the gel. Pour distilled water ~0.5 cm deep in a glass dish and wet the membrane by placing it on the surface of the water. Allow the membrane to submerge, then leave for 5 min. If a nitrocellulose membrane is being used, submerge in distilled water; replace the water with 20 × SSC and leave for 10 min. Avoid handling nylon and nitrocellulose membranes even with gloved hands—use clean blunt-ended forceps instead.
10. Place the wetted membrane on the surface of the gel. Try to avoid getting air bubbles under the membrane; remove any that appear by carefully rolling a glass pipet over the surface.

    For blots of substantial amounts of plasmid or other very-low-complexity DNA, it is important to lay the membrane down precisely the first time, as detectable transfer can take place almost immediately, and moving the membrane may cause "mobile" bands as streaks on the autoradiograph.
11. Flood the surface of the membrane with 20 × SSC. Cut five sheets of Whatman 3MM paper to the same size as the membrane and place these on top of the membrane.
12. Cut paper towels to the same size as the membrane and stack these on top of the Whatman 3MM papers to a height of ~4 cm.
13. Lay a glass plate on top of the structure and place a weight on top to hold everything in place. Leave overnight. The weight should be sufficient to compress the paper towels to ensure good contact throughout the stack. Excessive weight, however, will crush the gel and retard transfer. An overnight transfer is

sufficient for most purposes. Extend the transfer time if the gel concentration is >1%, or transfer of fragments >20 kb is desired. Make sure that the reservoir of 20× SSC does not run dry during the transfer. If the gel contains large amounts of DNA (see annotation to step 10), high transfer efficiencies may not be required and shorter transfer times (<“1 hr) may be used.

Disassemble the transfer pyramid:

14. Remove the paper towels and filter papers and recover the membrane. Mark in pencil the position of the wells on the membrane and ensure that the up-down and back-front orientations are recognizable. Pencil is preferable to pen, as ink marks may wash off the membrane during hybridization. An alternative for a nylon membrane only is to cut slits with a razor blade to mark the positions of the wells (do not do this before transfer or the buffer will short-circuit). The best way to record the orientation of the membrane is by making an asymmetric cut at one corner.

15. Rinse the membrane in 2× SSC, then place it on a sheet of Whatman 3MM paper and allow to dry. The rinse has two purposes: to remove agarose fragments that may adhere to the membrane and to leach out excess salt. The gel can be restained with ethidium bromide to assess the efficiency of transfer.

Immobilize the DNA:

16. Wrap the membrane in UV-transparent plastic wrap, place DNA-side-down on a UV transilluminator (254-nm wavelength) and irradiate for the time determined from the support protocol. CAUTION: Exposure to UV irradiation is harmful to the eyes and skin. Wear suitable eye protection and avoid exposure of bare skin. Nitrocellulose membranes should not be UV irradiated. Instead, place between two sheets of Whatman 3MM paper and bake under vacuum for 2 hr at 80°C. This results in noncovalent attachment of the DNA to the membrane; the vacuum is needed to prevent the membrane from igniting. UV crosslinking is recommended for a nylon membrane as this leads to covalent attachment and enables the membrane to be reprobed several times. The membrane must be completely dry before UV crosslinking; check the manufacturer’s recommendations. A common procedure is to bake for 30 min at 80°C prior to irradiation. Plastic wrap is used to protect the membrane during irradiation, but it must be UV-transparent. A UV light box can be used instead of a transilluminator (follow manufacturer’s instruction).

17. Store membranes dry between sheets of Whatman 3MM paper for several months at room temperature. For long-term storage, place membranes in a desiccator at room temperature or 4^0C.

NORTHERN BLOT:

The **northern blot**, or RNA blot is a technique used in molecular biology research to study gene expression by detection of RNA (or isolated mRNA) in a sample. With northern blotting it is possible to observe cellular control over structure and function by determining the particular gene expression rates during differentiation and morphogenesis, as well as in abnormal or diseased conditions. Northern blotting involves the use of electrophoresis to separate RNA samples by size, and detection with a hybridization probe complementary to part of or the entire target sequence. The term 'northern blot' actually refers specifically to the capillary transfer of RNA from the electrophoresis gel to the blotting membrane. However, the entire process is commonly referred to as northern blotting. The northern blot technique was developed in 1977 by James Alwine, David Kemp, and George Stark at Stanford University with contributions from Gerhard Heinrich. Northern blotting takes its name from its similarity to the first blotting technique, the Southern blot, named for biologist Edwin Southern. The major difference is that RNA, rather than DNA, is analyzed in the northern blot.

Principle:

The **northern blot**, or RNA **blot**, is a technique used in molecular biology research to study gene expression by detection of RNA (or isolated mRNA) in a sample.

Procedure:

A general blotting procedure starts with extraction of total RNA from a homogenized tissue sample or from cells. Eukaryotic mRNA can then be isolated through the use of oligo (dT) cellulose chromatography to isolate only those RNAs with a poly(A) tail. RNA samples are then separated by gel electrophoresis. Since the gels are fragile and the probes are unable to enter the matrix, the RNA samples, now separated by size, are transferred to a nylon membrane through a capillary or vacuum blotting system.

Capillary blotting system setup for the transfer of RNA from an electrophoresis gel to a blotting membrane.

A nylon membrane with a positive charge is the most effective for use in northern blotting since the negatively charged nucleic acids have a high affinity for them. The transfer buffer used for the blotting usually contains formamide because it lowers the

annealing temperature of the probe-RNA interaction, thus eliminating the need for high temperatures, which could cause RNA degradation. Once the RNA has been transferred to the membrane, it is immobilized through covalent linkage to the membrane by UV light or heat. After a probe has been labeled, it is hybridized to the RNA on the membrane. Experimental conditions that can affect the efficiency and specificity of hybridization include ionic strength, viscosity, duplex length, mismatched base pairs, and base composition. The membrane is washed to ensure that the probe has bound specifically and to prevent background signals from arising. The hybrid signals are then detected by X-ray film and can be quantified by densitometry. To create controls for comparison in a northern blot, samples not displaying the gene product of interest can be used after determination by microarrays or RT-PCR.

Gels:

RNA run on a formaldehyde agarose gel to highlight the 28S (top band) and 18S (lower band) ribosomal subunits. The RNA samples are most commonly separated on agarose gels containing formaldehyde as a denaturing agent for the RNA to limit secondary structure. The gels can be stained with ethidium bromide (EtBr) and viewed under UV light to observe the quality and quantity of RNA before blotting. Polyacrylamide gel electrophoresis with urea can also be used in RNA separation but it is most commonly used for fragmented RNA or micro RNAs. An RNA ladder is often run alongside the samples on an electrophoresis gel to observe the size of fragments obtained but in total RNA samples the ribosomal subunits can act as size markers. Since the large ribosomal subunit is 28S (approximately 5kb) and the small ribosomal subunit is 18S (approximately 2kb) two prominent bands appear on the gel, the larger at close to twice the intensity of the smaller.

Probes:

Probes for northern blotting are composed of nucleic acids with a complementary sequence to all or part of the RNA of interest, they can be DNA, RNA, or oligonucleotides with a minimum of 25 complementary bases to the target sequence. RNA probes (riboprobes) that are transcribed in vitro are able to withstand more rigorous washing steps preventing some of the background noise.Commonly cDNA is created with labelled primers for the RNA sequence of interest to act as the probe in the northern blot. The probes must be labelled either with radioactive isotopes (32P) or with chemiluminescence in which alkaline phosphatase or horseradish peroxidase (HRP) break down chemiluminescent substrates producing a detectable emission of light. The chemiluminescent labelling can occur in two ways: either the probe is attached

to the enzyme, or the probe is labeled with a ligand (e.g. biotin) for which the ligand (e.g., avidin or streptavidin) is attached to the enzyme (e.g. HRP). X-ray film can detect both the radioactive and chemiluminescent signals and many researchers prefer the chemiluminescent signals because they are faster, more sensitive, and reduce the health hazards that go along with radioactive labels. The same membrane can be probed up to five times without a significant loss of the target.

WESTERN BLOTTING:

The western blot called the protein immunoblot is a widely used analytical technique used to detect specific proteins in a sample of tissue homogenate or extract. It uses gel electrophoresis to separate native proteins by 3-D structure or denatured proteins by the length of the polypeptide. The proteins are then transferred to a membrane (typically nitrocellulose or PVDF), where they are stained with antibodies specific to the target protein. The gel electrophoresis step is included in Western blot analysis to resolve the issue of the cross-reactivity of antibodies.

This method is used in the fields of molecular biology, immunogenetics and other molecular biology disciplines. A number of search engines, such as CiteAb, are available that can help researchers find suitable antibodies for use in Western Blotting. Other related techniques include dot blot analysis, immunohistochemistry and immunocytochemistry where antibodies are used to detect proteins in tissues and cells by immunostaining, and enzyme-linked immunosorbent assay (ELISA).

Steps:

Tissue preparation:

Samples can be taken from whole tissue or from cell culture. Solid tissues are first broken down mechanically using a blender (for larger sample volumes), using a homogenizer (smaller volumes), or by sonication. Cells may also be broken open by one of the above mechanical methods. However, virus or environmental samples can be the source of protein.

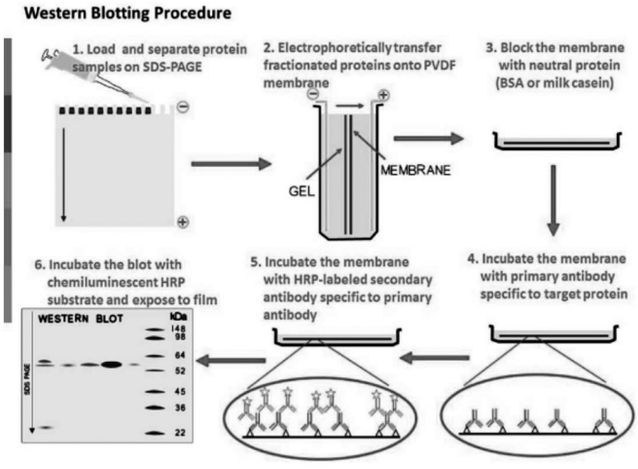


Figure 6.21

Assorted detergents, salts, and buffers may be employed to encourage lysis of cells and to solubilize proteins. Protease and phosphatase inhibitors are often added to prevent the digestion of the sample by its own enzymes. Tissue preparation is often done at cold temperatures to avoid protein denaturing and degradation.

A combination of biochemical and mechanical techniques – comprising various types of filtration and centrifugation – can be used to separate different cell compartments and organelles.

Gel electrophoresis:

The proteins of the sample are separated using gel electrophoresis. Separation of proteins may be by isoelectric point (pI), molecular weight, electric charge, or a combination of these factors. The nature of the separation depends on the treatment of the sample and the nature of the gel. This is a very useful way to identify a protein.

SDS-PAGE (SDS polyacrylamide gel electrophoresis) maintains polypeptides in a denatured state once they have been treated with strong reducing agents to remove secondary and tertiary structure and thus allows separation of proteins by their molecular weight. Sampled proteins become covered in the negatively charged SDS and move to the positively charged electrode through the acrylamide mesh of the gel.

Samples are loaded into *wells* in the gel. One lane is usually reserved for a *marker*or *ladder*, a commercially available mixture of proteins having defined molecular weights, typically stained so as to form visible, coloured bands. When voltage is applied along the gel, proteins migrate through it at different speeds dependent on their size.

Transfer:

In order to make the proteins accessible to antibody detection, they are moved from within the gel onto a membrane made of *nitrocellulose or polyvinylidene difluoride (PVDF*). The primary method for transferring the proteins is called electroblotting and uses an electric current to pull proteins from the gel into the PVDF or nitrocellulose membrane. As a result of either "blotting" process, the proteins are exposed on a thin surface layer for detection. Protein binding is based upon hydrophobic interactions, as well as charged interactions between the membrane and protein.

The uniformity and overall effectiveness of transfer of protein from the gel to the membrane can be checked by staining the membrane with Coomassie Brilliant Blue or Ponceau S dyes. Ponceau S is the more common of the two, due to its higher sensitivity and water solubility.

Blocking:

Blocking of non-specific binding is achieved by placing the membrane in a dilute solution of protein – typically 3-5% Bovine serum albumin (BSA) or non-fat dry milk (both are inexpensive) in Tris-Buffered Saline (TBS). The protein in the dilute solution attaches to the membrane in all places where the target proteins have not attached.

Detection:

During the detection process the membrane is "probed" for the protein of interest with a modified antibody which is linked to a reporter enzyme; when exposed to an appropriate substrate, this enzyme drives a colourimetric reaction and produces a color. Primary antibodyAfter blocking, a dilute solution of primary antibody (generally between 0.5 and 5 micrograms/mL) is incubated with the membrane under gentle agitation. Typically, the solution is composed of buffered saline solution with a small percentage of detergent, and sometimes with powdered milk or BSA. The antibody solution and the membrane can be sealed and incubated together for anywhere from 30 minutes to overnight. It can also be incubated at different temperatures, with higher temperatures being associated with more binding, both specific and non-specific. Secondary antibody After rinsing the membrane to remove unbound primary antibody, the membrane is exposed to another antibody, directed at a species-specific portion of the primary antibody. The secondary antibody is usually linked to biotin or to a reporter enzyme such as alkaline phosphatase or horseradish peroxidase. This means that several secondary antibodies will bind to one primary antibody and enhance the signal. Most commonly, a horseradish peroxidase-linked secondary is used to cleave a chemiluminescent agent, and the reaction product produces luminescence in proportion to the amount of protein. A sensitive sheet of photographic film is placed against the membrane, and exposure to the light from the reaction creates an image of the antibodies bound to then blot. The enzyme can be provided with a substrate molecule that will be converted by the enzyme to a coloured reaction product that will be visible on the membrane.

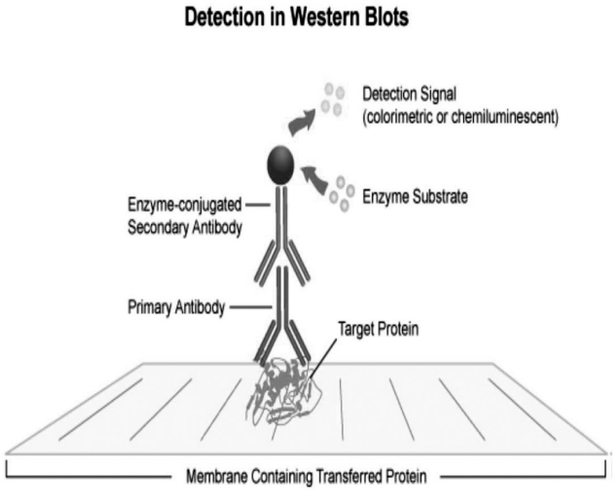


Diagram 2: *Illustration of detection in Western Blots.*

Figure 6.22

DNA fingerprinting:

The original DNA fingerprinting technique was devised in 1985 by Alec Jeffreys, who realised that the work he was doing on sequences within the myoglobin gene could have wider implications. The method is based on the fact that there are highly variable regions of the genome that are specific to each individual. These are **mini-satellite** regions, which have a variable number of short repeated-sequence elements known as **variable number tandem repeats** (**VNTRs**). Within the VNTR there are core sequence motifs that can be identified in other polymorphic VNTR loci, and also sequences that are restricted to the particular VNTR. The arrangements of the VNTR sequences, and the choice of a suitable probe sequence, are the key elements that enable a unique 'genetic fingerprint' to be produced.

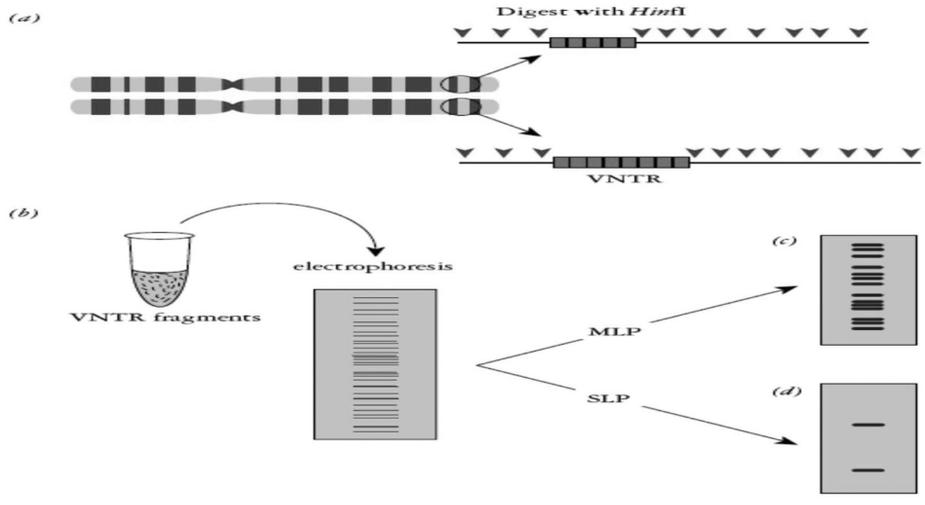


Figure 6.23

The first requirement is to isolate DNA and prepare restriction fragments for electrophoresis. If an enzyme is used that does not cut the core sequence, but cuts frequently outside it, and then the VNTR is effectively isolated. For human DNA the enzyme HinfI is often used. By using a probe that hybridises to the core sequence, and carrying out the hybridisation under low stringency, polymorphic loci that bind the probe can be identified. The probe in this case is known as a **multi-locus probe**, as it binds to multiple sites. This generates a pattern of bands that is unique – the 'genetic fingerprint'. If probes with sequences that are specific for a particular VNTR are used (**single-locus probes**), a more restricted fingerprint is produced, as there will be two alleles of the sequence in each individual, one maternally derived and one paternally derived. An overview of the basis of the technique is shown in Figure below.

Genetic fingerprinting of minisatellite DNA sequences. (a) A chromosome pair, with one minisatellite (VNTR) locus highlighted. In this case the locus is heterozygous

for VNTR length. Cutting with HinfI effectively isolates the VNTR. (b) The VNTR fragments produced (from many loci) are separated by electrophoresis and blotted. Challenging with a multi-locus probe (MLP) produces the 'bar code' pattern shown in (c). If a single-locus probe (SLP) is used, the two alleles of the specific VNTR are identified.

In forensic analysis, the original DNA profiling technique has now been largely replaced by a PCR-based method that amplifies parts of the DNA known as **short tandem repeats** (**STRs**, also known as **microsatellites**). These are repeats of 2, 3, 4, or 5 base pairs. A major advantage over minisatellite (VNTR) repeats is that STRs are found throughout the genome; thus, better coverage is achieved than with minisatellites. The PCR overcomes any problems associated with the tiny amounts of sample that are often found at the crime scene. The reaction is set up to amplify the loci involved — usually 3 or 4 are sufficient if the loci are selected carefully to optimise the information generated. By using fluorescent labels and automated DNA detection equipment (similar to the genome sequencing equipment a DNA profile can be generated quickly and accurately.

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

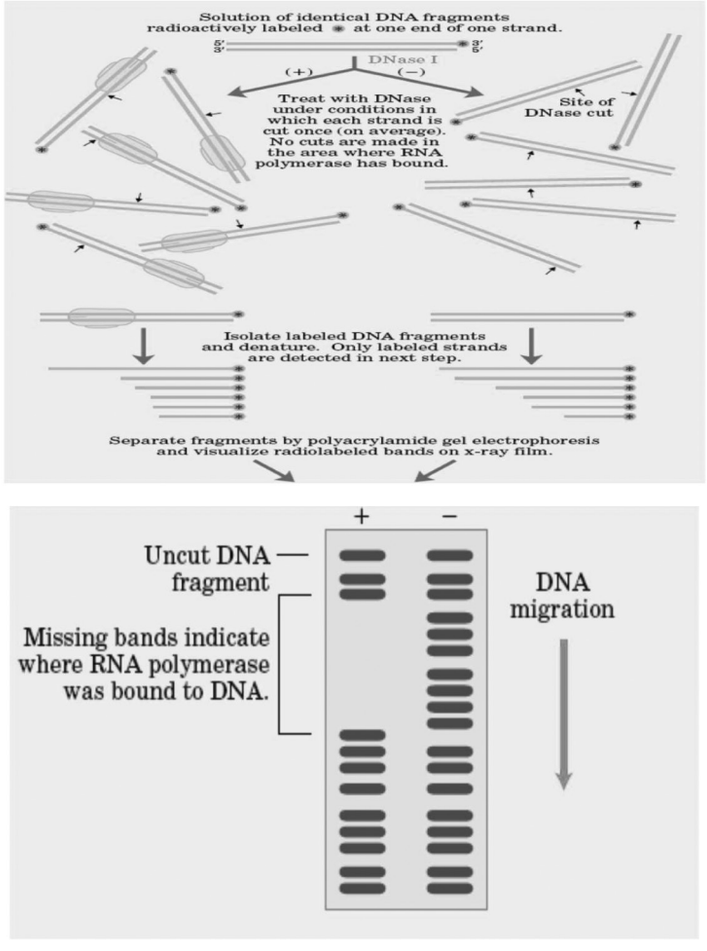


Figure 6.24

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.

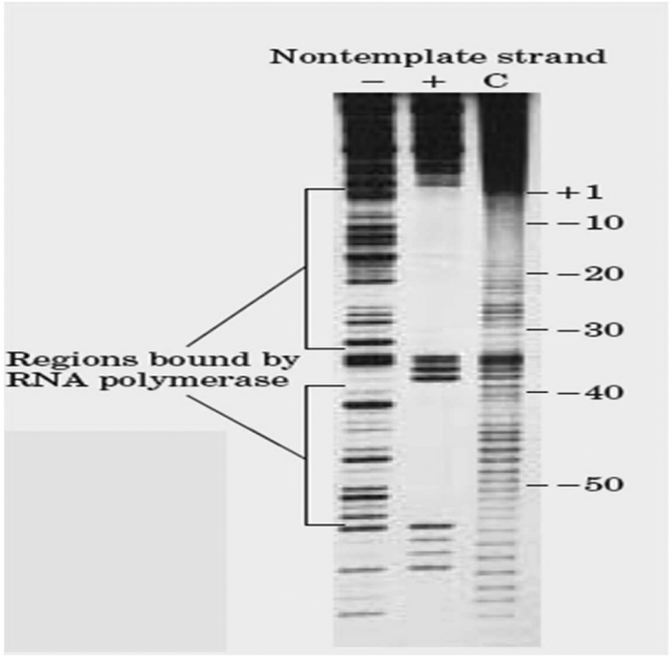


Figure 6.25

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.

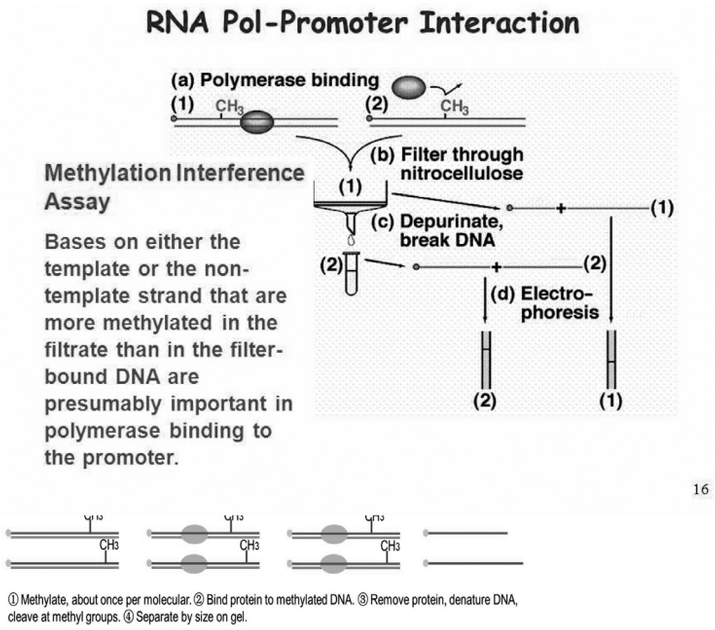


Figure 6.26

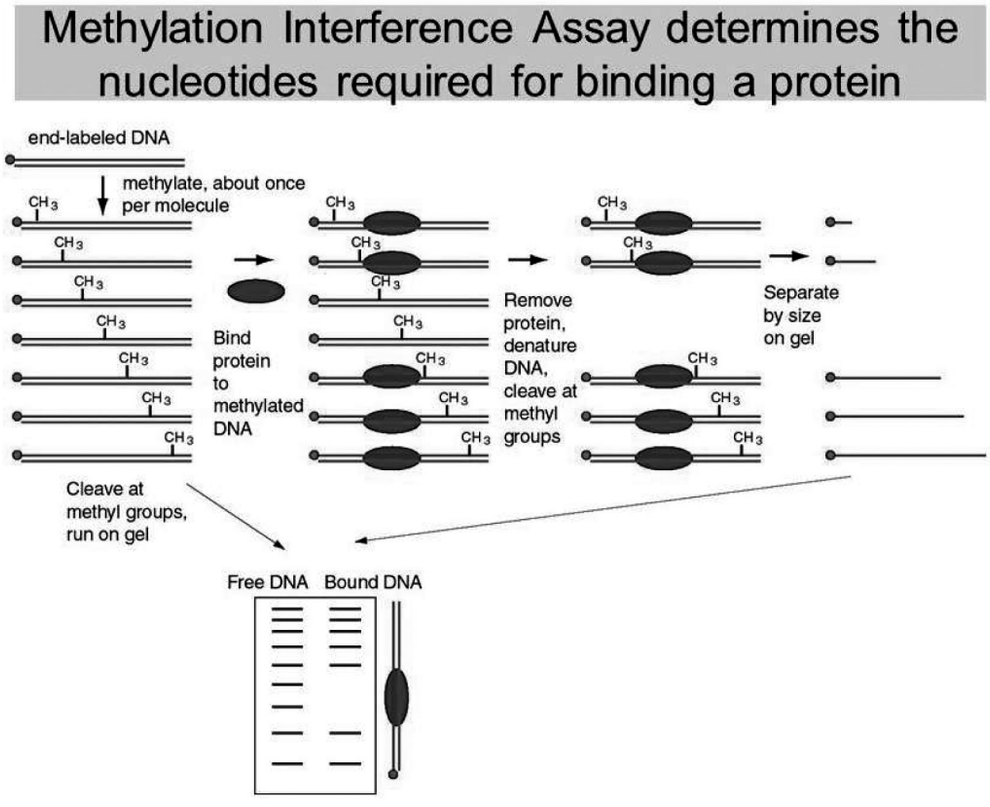


Figure 6.27

Declaration: All contents prepared for this SLM have been taken from several molecular biology and recombinant DNA technology books, research articles, online book chapters, Wikipedia and other sources from google

Probable Questions

1. Explain the role and mechanism of Type II restriction endonucleases in recombinant DNA technology. How are they different from Type I and Type III enzymes?
2. Discuss the various ligation strategies used in recombinant DNA technology, emphasizing on cohesive end ligation, blunt end ligation, and the use of linkers and adaptors.
3. Describe the Southern blotting technique in detail. Explain the principle, steps, and applications in molecular biology.
4. What are nucleic acid probes? Discuss the different types of probes and their applications in clone identification and gene expression analysis.
5. Outline the procedure and principle of DNA fingerprinting. How has the method evolved with the advent of STR analysis and PCR-based techniques?
6. Compare and contrast Northern blotting and Western blotting techniques with respect to their principle, methodology, and applications.
7. Explain the significance of restriction modification systems in bacteria. How do these systems contribute to the bacterial defense against phage infection?
8. Describe the process of restriction mapping. How can single and double digestions with restriction enzymes be used to determine the relative positions of recognition sites on a DNA molecule?
9. Explain the methylation interference assay technique. How does it help in identifying DNA-protein interaction sites and gene regulation elements?
10. List the various enzymes used in recombinant DNA technology and explain the role of each in the manipulation of nucleic acids.

Unit VII

Polymerase chain reaction–methods and applications

Objective: In this unit we will discuss about methods and applications of Polymerase chain reaction.

Introduction:

The polymerase chain reaction (PCR) is an enzymatic process that allows for the detection of specific genes within an environmental DNA sample. This technique is used to amplify a precise fragment of DNA from a complex mixture of starting material usually termed as template DNA. The purpose of a PCR is to make a huge number of copies of a gene. It was originally invented by Kary Mullis in 1985 in California and got the Nobel Prize in 1993.

Principle

The basic principle of PCR is that the double stranded DNA molecule, when heated to a high temperature, separate yielding two single-stranded DNA molecules. The single stranded DNA molecules can easily be copied with the help of a DNA polymerase and nucleosides resulting in the duplication of original DNA molecules. By repeating these events, multiple copies of the original DNA molecule can be generated.

Requirements

i) A thermal cycler (an instrument having a microprocessor-controlled temperature cycling) - heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction.

ii) DNA segment to be amplified

iii) Two primers, which are oligonucleotides (about 10-18 nucleotides long), oriented with their ends facing each other so that DNA synthesis can occur between them. These short artificial DNA fragments containing sequences complementary to the target region, that are complementary to the 3' (three prime) ends of each of the sense(forward) and the 5' end of the anti-sense (reverse) strand of the DNA target.

iv) The enzyme Taq polymerase (a DNA polymerase) which is stable at high temperature. Taq polymerase enzyme puts the free nucleotides together. It starts at the 3'end of the primer, and uses the complementary DNA strain as a template.

v) $MgCl_2$ - cofactor of the enzyme.

vi) dNTPs (deoxy nucleoside triphosphate: dATPs, dTTPs, dGTPs, dCTPs)

vii) Buffer solution, maintains pH and ionic strength of the reaction solution suitable for the activity of the enzyme.

viii) Pipettes (1-10 µl, 5-50 µl, 20-200 µl, and 100-1000 µl) and aerosol barrier pipette tips

ix) PCR tubes (0.2 ml or 0.5 ml)

x) Master mix tubes (1.5 ml microcentrifuge tubes)

xi) Sterile, nuclease-free water: must be present for the reaction to work.

xii) Gloves

Procedure

Typically, PCR consists of a series of 25-40 repeated temperature changes, called cycles, each cycle of PCR includes steps for template denaturation, primer annealing and primer extension.

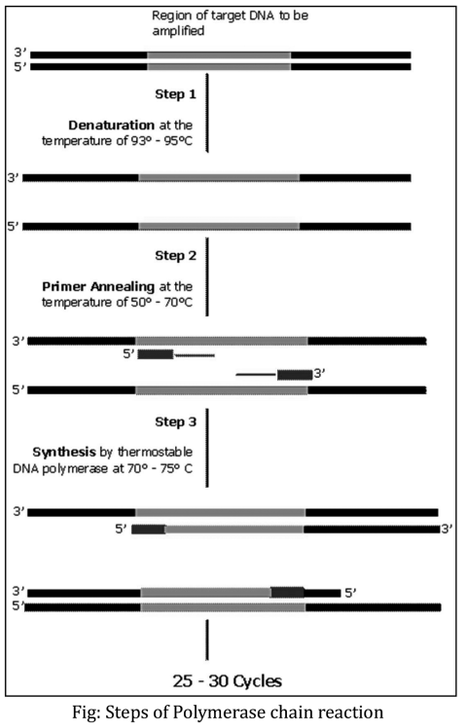


Fig: Steps of Polymerase chain reaction
[Source: https://www.biology.lifeeasy.org/2481/steps-involved-in-pcr-polymerase-chain-reaction-process]

Figure 7.1

i) Denaturation: The DNA, from which a segment is to be amplified, is mixed with an excess of the two primer molecules, all the four kinds of dNTPs, $MgCl_2$ and Taq polymerase in a reaction mixture. The reaction mixture is heated to a high temperature (94-96ºC) so that the DNA molecule is denatured i.e. the two strands of DNA duplex get separated. Each strand of the target DNA then acts as a template for DNA synthesis.

ii) Annealing: The mixture is then cooled by lowering the temperature upto 55-65°C for 30 sec-1min allowing annealing of the primers to the single-stranded DNA template. Annealing occurs due to presence of complementary sequences located at the 3' ends of the template DNA. The

annealing temperature is about 3-5 °C below the Tm of the primers used ("Tm" refers to the melting temperature, the temperature at which half of a DNA duplex (primer-template) dissociates and becomes single-stranded). Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA formation.

iii) Extension: In this step, the temperature is so adjusted that the Taq polymerase becomes active. Synthesis of new DNA strand begins in between the primers, dNTPs and Mg^{2+}. The optimum temperature for this polymerization is kept at 72°C.

At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to be amplified. At its optimum temperature, the DNA polymerase will polymerize a thousand bases per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

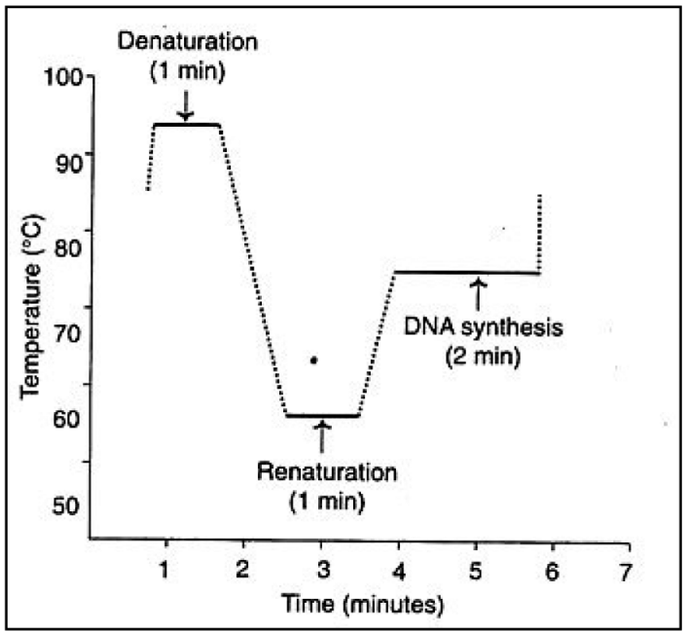


Fig. 7.2: The three stages in each cycle of PCR in relation to temperature and time (each cycle takes approximately 3-5 minutes)
[Source: https://www.biologydiscussion.com/biotechnology/polymerase-chain-reaction/polymerase-chain-reaction-techniques-and-variations/11081]

Factors Affecting PCR Specificity and Efficiency

The process of optimizing a specific PCR can be a challenging and time-consuming task because it involves numerous parameters. These parameters include factors such as;

i. The quality and concentration of the DNA template,

ii. The design and concentration of primers

iii. The concentration of magnesium ions

iv. The concentration of the four deoxynucleotides

v. The type of PCR buffer systems used

vi. The selection and concentration of DNA Polymerase

vii. The PCR thermal cycling conditions

Applications for PCR:

1. Medical Diagnostics:

- **Infectious Disease Detection:** PCR is used to detect trace amounts of viral or bacterial DNA/RNA in patient samples, enabling the diagnosis of infectious diseases like COVID-19, HIV, and hepatitis.
- **Genetic Disease Testing:** PCR helps analyze gene sequences and mutations associated with genetic disorders like cystic fibrosis and thalassemia, crucial for developing diagnostics and treatments.
- **Monitoring Gene Therapy:** PCR can be used to monitor the gene in gene therapy.

2. Forensic Science:

- **DNA Fingerprinting:** PCR amplifies tiny DNA samples from crime scenes to identify suspects through DNA fingerprinting and establish forensic evidence.
- **Identifying Human Remains:** PCR can be used to identify human remains and develop genetic fingerprints from crime scene samples.

3. Genetic Research:

- **Cloning and Sequencing DNA:** PCR is used to amplify and clone specific DNA regions for further analysis, including sequencing.

- **Analysing Genetic Mutations:** PCR can detect mutations in genes, aiding in the study of genetic diseases and their inheritance patterns.
- **Studying Gene Expression:** PCR, particularly RT-PCR (Reverse Transcription PCR), helps analyze gene expression levels by quantifying mRNA transcripts.
- **Studying Genetic Diseases:** PCR is used to test for genetic mutations associated with various diseases.

Significance of Polymerase Chain Reaction

Polymerase chain reaction (PCR) is a technique that enables the detection of small amounts of DNA. It has significantly impacted various fields, including genomics, forensics, diagnostics, archaeology, and palaeontology. The following are some of the key points that highlight the importance of PCR:

i. Polymerase chain reaction has played an essential role in various fields of molecular biology and genomics, including molecular cloning, DNA sequencing, and gene expression analysis.

ii. It has revolutionised forensic investigations, allowing the identification of criminals from tiny traces of DNA evidence left at crime scenes.

iii. PCR has also transformed clinical microbiology and infectious disease management by enabling the diagnosis of infectious diseases through the detection of pathogen DNA signatures.

iv. It has facilitated paternity testing, genetic fingerprinting, and DNA databasing for forensic identification.

v. Polymerase chain reaction has enabled the DNA barcoding of species, catalysing metagenomics studies of environmental microbial diversity.

vi. PCR has democratised access to genomic information by enabling the generation of vast DNA data from minute samples

Advantages of PCR:

- **High Sensitivity:**

  PCR can detect even a small amount of target DNA, making it useful for identifying pathogens or analyzing genetic material from limited samples.

- **Rapid Results:**

  PCR assays can provide results much faster than traditional culturing methods, which can take days or weeks.

- **Versatility:**

  PCR can be used for a wide range of applications, including pathogen detection, genetic analysis, and forensic science.

- **Automated and High-Throughput:**

  PCR technology is often automated and can be used to screen a large number of samples quickly.

- **Cost-Effective:**

  PCR can be more cost-effective than traditional methods, especially when used selectively.

Disadvantages of PCR:

- **Potential for Contamination:**

  PCR is highly sensitive, so even a small amount of contaminating DNA can lead to false-positive results.

- **Specificity Issues:**

  PCR primers may anneal to sequences that are similar, but not identical, to the target DNA, leading to false positives.

- **Need for Prior Sequence Knowledge:**

  PCR requires knowledge of the sequence of the target DNA to design primers, limiting its use to known pathogens or genes.

- **False Negatives:**

  PCR can produce false-negative results due to inhibitors in the sample, poor DNA recovery, or ineffective release of microbial DNA content.

- **Lack of Novel Information:**

  PCR can only amplify and target specific DNA sequences targeted by the primers, providing limited information and inability to detect novel DNA sequences.

- **Cost: This is a expensive process.** the cost of reagents and equipment can be significant, especially for complex assays.

  But PCR can be cost-effective in some cases also.

Limitations of PCR

1. Contamination of the reaction mixture of PCR even by a single gene can result into impure product of PCR reaction
2. If diagnosis of male specific genes is to be done, female workers are required otherwise cells of skin of male workers can also cause contamination.
3. PCR cannot be performed directly on food samples and processing of sample is required to extract DNA to be studied. DNA recovery in complex food samples (milk, oysters) may be only few percent. Food substances copurified with bacterial DNA may reduce amplification efficiency. If cell growth is allowed to increase DNA, target DNA recovery probability increases and interference by food substances reduces but time for analysis is then increased.
4. Dead cells may contain amplifiable DNA; hence positive PCR does not mean viable cells in sample. Thus, pasteurized milk may give positive PCR and technique loses using reverse transcriptase and then PCR can directly detect mRNA and can be significant to differentiate between live and dead cells.
5. PCR is most useful for the amplification of DNA segments less than 2 kb in length.

Probable questions:

1. What is the PCR and what is it being used for in this exercise?
2. What are oligonucleotide primers, and why are they necessary in the PCR, i.e., what function do they serve?
3. What is Taq polymerase and where does it come from?
4. What is Tm? How it affects annealing temperature of a PCR?
5. How primers are designed?
6. Why must the DNA polymerase used in PCRs be thermostable?
7. When the polymerase chain reaction is used to amplify DNA in vitro, the components of the reaction mixture are subjected to three different temperature settings during each reaction cycle. These temperatures cause DNA to be denatured, to anneal and then to extend. Explain what is occurring during each of these steps.
8. Why is amplification of 16S ribosomal DNA likely to yield a product that can be used in the identification/classification of unknown bacteria?

9. What are the advantages of PCR?
10. State the limitations of PCR.
11. State the disadvantages of PCR.
12. What is the significance of 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.
8. https://vajiramandravi.com/upsc-exam/polymerase-chain-reaction/
9. https://www.news-medical.net/life-sciences/Polymerase-Chain-Reaction-Applications.aspx
10. https://egyankosh.ac.in/bitstream/123456789/96170/1/Unit-8.pdf
11. https://pmc.ncbi.nlm.nih.gov/articles/PMC4102308/

Unit-VIII
Basic biology of cloning vectors

A cloning vector is a DNA molecule, typically a plasmid or a bacteriophage, that is used to carry foreign DNA into a host cell for the purpose of cloning and replicating that DNA. These vectors are designed to replicate independently of the host cell's chromosomal DNA, allowing for the creation of multiple copies of the inserted foreign DNA.

Plasmid:

Plasmids are naturally occurring components of the bacterial and some eukaryotic cells and therefore have capacity to remain and propagate stably in the host cell. Choices of plasmid developing vectors depend upon various factors such high copy numbers, stability and compatibility with other DNA molecules in the host cells. The following is the discussion on some plasmid-based vectors using for cloning in prokaryotic systems.

i. pBR vectors: pBR322 is a plasmid and was one of the first widely used *E. coli* cloning vectors. Created in 1977 in the laboratory of Herbert Boyer at the University of California, San Francisco, it was named after Francisco Bolivar Zapata, the postdoctoral researcher and Raymond L. Rodriguez. The p stands for "plasmid," and BR for "Bolivar" and "Rodriguez."

pBR322 is 4361 base pairs in length and has two antibiotic resistance genes – the gene *bla* encoding the ampicillin resistance (Amp^R) protein, and the gene *tetA* encoding the tetracycline resistance (Tet^R) protein. It contains the origin of replication of pMB1, and the *rop* gene, which encodes a restrictor of plasmid copy number. The plasmid has unique restriction sites for more than forty restriction enzymes. Eleven of these forty sites lie within the Tet^R gene. There are two sites for restriction enzymes HindIII and ClaI within the promoter of the Tet^R gene. There are six key restriction sites inside the Amp^R gene. The source of these antibiotic resistance genes are from pSC101 for Tetracycline and RSF2124 for Ampicillin.

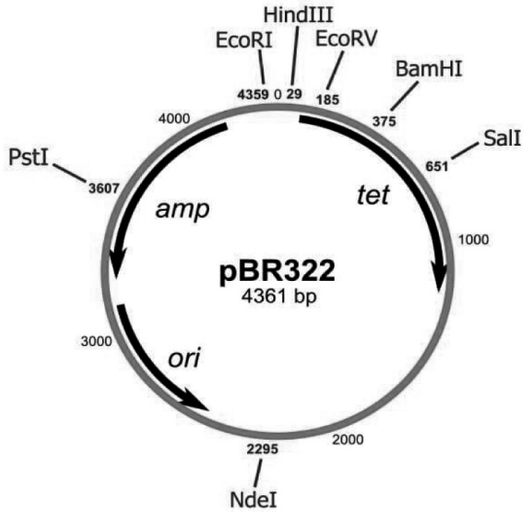


Figure 8.1

The circular sequence is numbered such that 0 is the middle of the unique EcoRI site and the count increases through the Tet^R gene. If we have to remove ampicillin for instance, we must use restriction endonuclease or molecular scissors against PstI and then pBR322 will become anti-resistant to ampicillin. The same process of Insertional Inactivation can be applied to Tetracycline. The Amp^R gene is penicillin beta-lactamase. Promoters P1 and P3 are for the beta-lactamase gene. P3 is the natural promoter, and P1 is artificially created by the ligation of two different DNA fragments to create pBR322. P2 is in the same region as P1, but it is on the opposite strand and initiates transcription in the direction of the tetracycline resistance gene.

Early cloning experiments may be conducted using natural plasmids such the ColE1 and pSC101. Each of these plasmids may have its advantages and disadvantages. For example, the ColE1 plasmid and its derivatives have the advantage of higher copy number and allow for chloramphenicol amplification of plasmid to produce a high yield of plasmid, however screening for immunity to colicin E1 is not technically simple. The plasmid pSC101, a natural plasmid from *Salmonella panama*, confers tetracycline resistance which allows for simpler screening process with antibiotic selection, but it is a low copy number plasmid which does not give a high yield of plasmid. Another plasmid, RSF 2124, which is a derivative of ColE1, confers ampicillin resistance but is larger.

Many other plasmids were artificially constructed to create one that would be ideal for cloning purpose, and pBR322 was found to be most versatile by many and was therefore the one most popularly used. It has two antibiotic resistance genes, as selectable markers, and a number of convenient unique restriction sites that made it suitable as a cloning vector. The plasmid was constructed with genetic material from 3 main sources – the tetracycline resistance gene of pSC101, the ampicillin resistance gene of RSF 2124, and the replication elements of pMB1, a close relative of the ColE1 plasmid.

A large number of other plasmids based on pBR322 have since been constructed specifically designed for a wide variety of purposes. Examples include the pUC series of plasmids. Most expression vectors for extrachromosomal protein expression and shuttle vectors contain the pBR322 origin of replication, and fragments of pBR322 are very popular in the construction of intraspecies shuttle or binary vectors and vectors for targeted integration and excision of DNA from chromosome.

ii. pUC vector:

A pUC vector is a commonly used plasmid cloning vector in E. coli, particularly pUC18 and pUC19. They are derived from the pBR322 plasmid and are characterized by high copy number, typically 500-600 copies per cell. This high copy number, along with their smaller size compared to pBR322, facilitates efficient replication and clon-

ing. They have a high copy number, meaning they are present in many copies within each host cell, leading to high levels of plasmid DNA. They typically contain an ampicillin resistance gene, allowing for selection of transformed bacteria in the presence of ampicillin. pUC vectors often have the lacZ gene, allowing for blue-white screening to distinguish between recombinant and non-recombinant colonies. The lacZ gene encodes beta-galactosidase, and inserting DNA into a specific site within the lacZ gene disrupts its function, resulting in white colonies on plates with X-gal and IPTG. pUC vectors have multiple restriction enzyme sites where foreign DNA can be inserted.

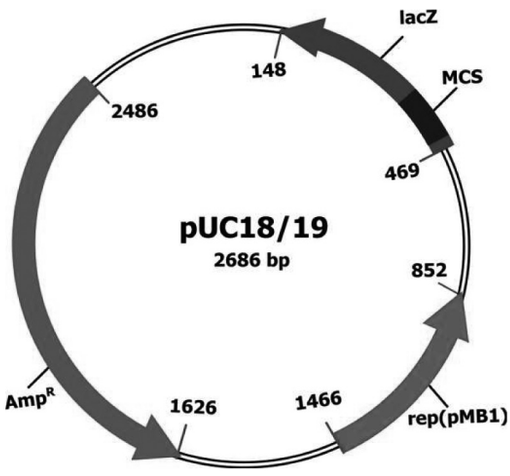


Figure 8.2

pUC vectors are designed for efficient and convenient cloning and manipulation of DNA in E. coli. They are widely used in research and biotechnology due to their high copy number, selectable marker, and the ability to perform blue-white screening, which allows for easy identification of recombinant colonies.

Phages:

Bacteriophages are a group of viruses that used bacterial cells as host and reproduce by infecting bacterial cells. Like all other viruses bacteriophages are also consisting of a protein coat that provides protection to the genome. Most of the phages are DNA viruses that code for several genes responsible for virus replication. A major problem of using plasmids as a cloning vector is the size of the gene of interest is small. However, bacteriophages help to overcome this problem, as larger genes can be transferred by using bacteriophage as a cloning vector.

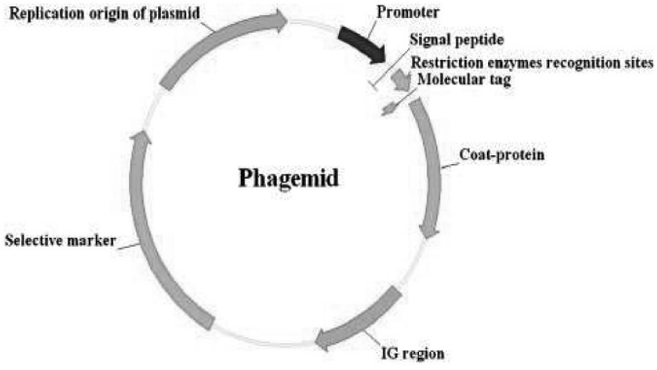


Figure 8.3

i. Lytic and Lysogenic Phages:

The pattern of infection of different bacteriophages is mostly needed three steps to complete. Firstly, the attachment of the phage at a specific site of bacterial takes place. The phage is then injected into the DNA inside the bacterial cell. In the second

step, the genes present in the DNA of the phage codes several enzymes which helps to replicate the phage DNA. Finally, other genes also code several proteins by which the capsid is formed and the new phage particle is assembled inside the capsid. This step is followed by lysis of the cell and release of the phages from the cell. In several bacteriophages, this cycle occurs rapidly and it is called a lytic cycle. On the other hand, in some phages, the phage DNA got incorporated in the host genome after entering the host cell and remains there for several cell divisions. The phage DNA which has incorporated in the bacterial DNA is called prophage and it is quiescent. The bacteria which carry the phage DNA remain as same as any other uninfected cells. This type of infection is called a lysogenic cycle. The phage DNA finally released from the host cell by entering into the lytic cycle. The infection cycle of lambda phage is an excellent example of a lysogenic cycle. In the case of cloning, the lysogenic cycle is most important, because, the phage DNA can carry a specific gene that is further incorporated in the host genome. As a result, the host genome will code a specific gene carries by the phage. Therefore, lambda phages are mostly used as a cloning vector.

ii. Lambda phage:

The lambda phage is a typical example of bacteriophage which has a polyhedral head and tail. The DNA is protected by the polyhedral structure, and the tail interacts with the specific site of bacteria for attachment. Lambda phage usually infects *E. coli.* The DNA of lambda phage is 49 kb of size. Gene mapping and DNA sequencing analysis provide all the information regarding the genetic arrangement of the phage DNA. According to the data gathered from these analyses, the genes of similar functions are clustered together in the genome. To illustrate, the genes responsible for the formation of the capsid are clustered together at the left-hand side of the polyhedral head while the genes responsible for the incorporation of the phage DNA with the host genome clustered together at the middle.

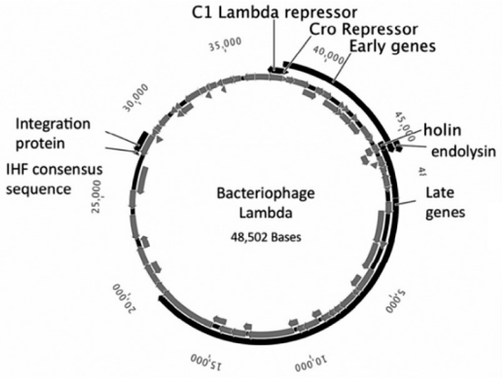


Figure 8.4

iii. M-13: Filamentous phage:

M-13 is a type of bacteriophage, and it is totally different from the lambda phages. The structure is filamentous, the genome size is very small (6407 nucleotides), single-stranded and remain in a circular formation. Due to the small size of DNA, the

phage genome can carry only a few recombinant genes. The M-13 phage DNA enters inside an *E. coli* cell via pilus. Inside the cell, the phage DNA codes for specific enzymes that synthesize a complementary strand of DNA. This newly formed double-stranded DNA molecule then proliferates into several copies. During the bacterial cell division, each daughter cell contains copies of phage DNA and the phage DNA continues to replicate inside the daughter cells. Finally, new phage particles are assembled and released from the bacterial cell. M-13 phages are also used as a cloning vector. The genome size is 10 kb which is enough to insert new genes. Apart from that, the modification of single-strand DNA to double-strand DNA inside the host makes it unique.

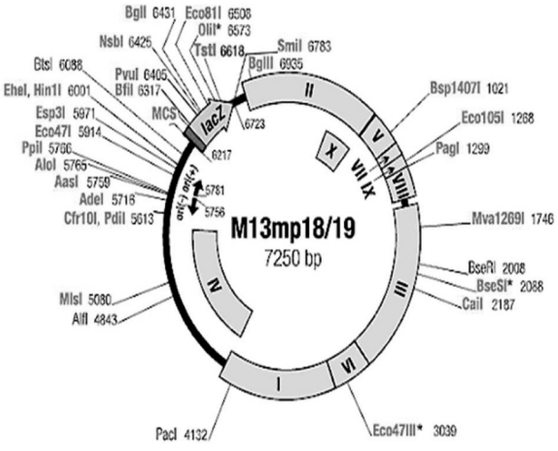


Figure 8.5

Single stranded DNA vectors:

Single-stranded DNA (ssDNA) vectors are specialized DNA molecules used in molecular biology for various applications, primarily involving cloning, expression, and sequencing. They are characterized by having a single strand of DNA, unlike the double-stranded DNA (dsDNA) typically found in cells. ssDNA vectors, like filamentous phages (M13, fd, f1), are used to isolate and replicate DNA fragments for sequencing or other genetic manipulations. While less common than dsDNA vectors for gene expression, ssDNA can be used in specific systems, such as those involving DNA-mediated gene regulation in mammalian cells. ssDNA can be used to induce targeted recombination in cells, facilitating the introduction of specific DNA sequences or gene editing. Circular ssDNA can be used as a programmable vector for gene expression, particularly in systems that require switchable or dynamic control of gene activity. Examples of ssDNA Vectors are as follows:

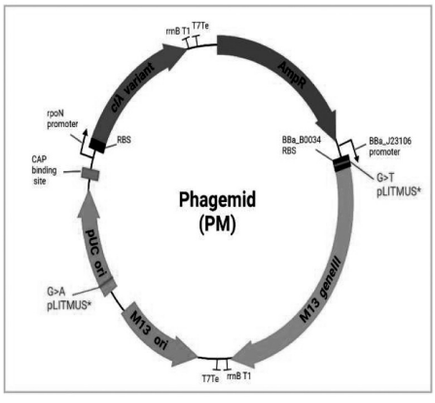


Figure 8.6

i. **M13 phage:** A widely used filamentous phage vector for cloning and sequencing.

ii. **fd phage:** Another filamentous phage vector similar to M13.

iii. **pUC118, pUC119:** These plasmids are designed to produce ssDNA when co-transformed with a helper phage.

iv. **Phagemids:** A phagemid vector is a hybrid DNA cloning vector that combines the properties of both plasmids and bacteriophages. It contains a plasmid's replication origin and a bacteriophage's packaging sequence, allowing it to be replicated like a plasmid and packaged into bacteriophage particles. This unique combination makes phagemids useful for various applications, including phage display and generating single-stranded DNA. They replicate autonomously like plasmids and can be packaged into bacteriophage particles, allowing for the production of phage displaying the desired protein. **Phagemids offer advantages over plasmids**, **such as the ability to produce single-stranded DNA and the potential for phage display.** e.g. pEMBL8.

Advantages of using ssDNA vectors:

- **Ease of Sequencing:** ssDNA is readily sequenced using Sanger sequencing due to its single-stranded nature.
- **High Yield of ssDNA:** Phage-based vectors can produce large quantities of ssDNA for various applications.
- **Versatility:** ssDNA vectors can be used for cloning, sequencing, expression, and targeted recombination.
- **Dynamic Regulation:** Circular ssDNA (CssDNA) offers the potential for switchable or dynamic gene regulation in living cells.

High-capacity vectors:

High-capacity vectors are specialized cloning vectors designed to carry and replicate large DNA fragments, ranging from tens of kilobases to megabases. These vectors are essential for constructing genomic libraries, performing physical mapping of genomes, and studying large genes.

i. Cosmid: Cosmid vectors are hybrid cloning vectors used in molecular biology, particularly for cloning large DNA fragments. They are constructed by combining the features of plasmids and bacteri-

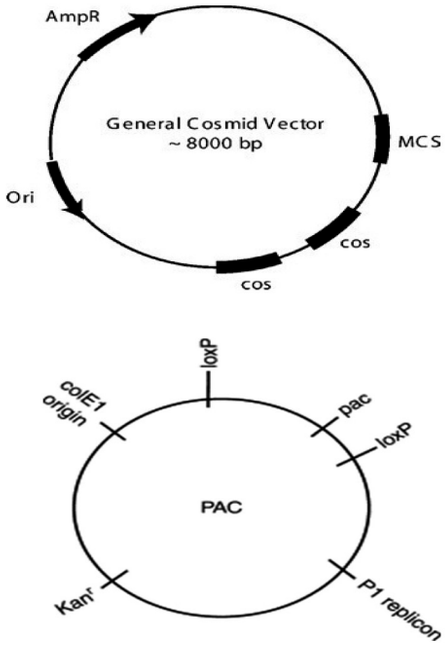


Figure 8.7

ophages, specifically by inserting the cos site (cohesive end site) from a lambda phage into a plasmid. This allows cosmids to be packaged into phage particles and transferred into bacterial cells, similar to phage vectors, but they also replicate autonomously like plasmids. Cosmids are typically used for constructing genomic libraries and cloning large genes due to their ability to accommodate DNA fragments up to 45 kb in size.

ii. PAC: PAC (P1-derived artificial chromosome) vectors are a type of cloning vector used to clone and propagate large DNA fragments in bacteria. They are constructed by combining features of P1 bacteriophage DNA with those of bacterial artificial chromosomes (BACs), allowing them to carry large DNA inserts (up to **150 kbp**). PACs are valuable for genome mapping and construction of genomic libraries.

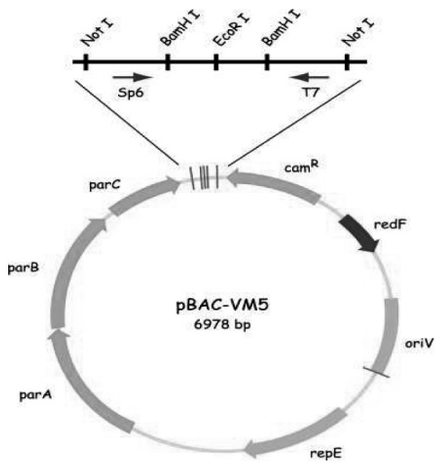


Figure 8.8

iii. BAC: A bacterial artificial chromosome (BAC) is an engineered DNA molecule used to clone DNA sequences in bacterial cells (for example, E. coli). BACs are often used in connection with DNA sequencing. Segments of an organism's DNA, ranging from **100,000** to about **300,000** base pairs, can be inserted into BACs. The BACs, with their inserted DNA, are then taken up by bacterial cells. As the bacterial cells grow and divide, they amplify the BAC DNA, which can then be isolated and used in sequencing DNA.

A large piece of DNA can be engineered in a fashion that allows it be propagated as a circular artificial chromosome in bacteria—so-called bacterial artificial chromosome, or BAC. Each BAC is a DNA clone containing roughly 100 to 300 thousand base pairs of cloned DNA. Because the BAC is much smaller than the endogenous bacterial chromosome, it is straightforward to purify the BAC DNA away from the rest of the bacteria cell's DNA, and thus have the cloned DNA in a purified form. This and other powerful features of BACs have made them extremely useful for mapping and sequencing mammalian genomes.

Advantages of Bacterial Artificial Chromosome:

- **Large insert capacity:** BACs can hold inserts of up to 300 kb, allowing for the cloning and stable maintenance of significant DNA fragments.
- **Stability:** BACs are highly stable in bacterial hosts, minimizing the risk of insert deletion or recombination.

- **Suitable for large genome studies:** BACs are widely used in whole genome sequencing projects and for constructing genomic libraries of organisms with large genomes.
- **Accurate gene expression:** BACs can minimize site of integration effects and allow for more accurate expression of endogenous genes.
- **Easy manipulation:** BACs can be easily manipulated and modified compared to YACs (Yeast Artificial Chromosomes).

Disadvantages of Bacterial Artificial Chromosome:

- **Complex manipulation and propagation:** Working with BACs can be more time-consuming and labor-intensive than with simpler vectors like plasmids.
- **Challenge in selecting clones:** Finding the right clone from a BAC library can be difficult.
- **Low gene expression:** BACs may exhibit lower gene expression rates in some host cells compared to other vectors.

BACs are widely used in genome sequencing projects, creating large DNA libraries, physical mapping of genomes, and developing infectious clone technologies. The stability of BAC clones is a key advantage, making them suitable for long-term storage and manipulation of large DNA fragments.

iv. YAC: Yeast artificial chromosomes (YACs) are genetically engineered chromosomes derived from the DNA of the yeast. It is a human-engineered DNA molecule used to clone DNA sequences in yeast cells. They are the products of a recombinant DNA cloning methodology to isolate and propagate very large segments of DNA in a yeast host.By inserting large fragments of DNA, the inserted sequences can be cloned and physically mapped using a process called chromosome walking. The amount of DNA that can be cloned into a YAC is, on average, from **200 to 500 kb**. However, as much as **1 Mb** can be cloned into a YAC.

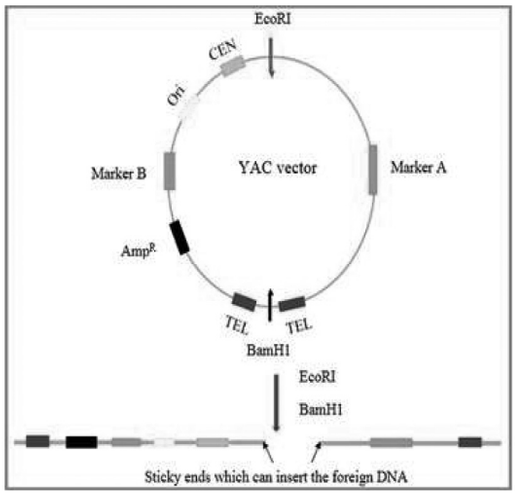


Figure 8.9

A yeast artificial chromosome cloning vector consists of two copies of a yeast telomeric sequence (telomeres are the sequences at the ends of chromosomes), a yeast centromere, a yeast ars (an autonomously replicating sequence where DNA replication begins), and appropriate selectable markers.

The yeast artificial chromosome, which is often shortened to YAC, is an artificially constructed system that can undergo replication. The design of a YAC allows extremely large segments of genetic material to be inserted. Subsequent rounds of replication produce many copies of the inserted sequence, in a genetic procedure known as cloning.

Advantages of Yeast Artificial Chromosomes:

- **Large insert capacity:**Yeast artificial chromosomes (YACs) provide the largest insert capacity of any cloning system.
- **post-translational modification.:**Yeast expression vectors, such as YACs, YIPs (yeast integrating plasmids), and YEPs (yeast episomal plasmids), have advantageous over bacterial artificial chromosomes (BACs). They can be used to express eukaryotic proteins that require post-translational modification.
- **Stability:** A major advantage of cloning in yeast, a eukaryote, is that many sequences that are unstable, underrepresented, or absent when cloned into prokaryotic systems, remain stable and intact in YAC clones.
- **Reintroduction:** It is possible to reintroduce YACs intact into mammalian cells where the introduced mammalian genes are expressed and used to study the functions of genes in the context of flanking sequences.

Limitations of Yeast Artificial Chromosomes:

A problem encountered in constructing and using YAC libraries is that they typically contain clones that are chimeric, i.e., contain DNA in a single clone from different locations in the genome. YAC clones frequently contain deletions, rearrangements, or non-contiguous pieces of the cloned DNA. As a result, each YAC clone must be carefully analysed to be sure that no rearrangements of the DNA have occurred.

- **Lower efficiency:** The efficiency of cloning is low (about 1000 clones are obtained per microgram of vector and insert DNA).
- **Less stability:** YACs have been found to be less stable than BACs.
- **Low yield:** The yield of YAC DNA isolated from a yeast clone containing a YAC is quite low.

- **Lower quantity:** The YAC DNA is only a few percents of the total DNA in the recombinant yeast cell. It is difficult to obtain even 1 μg of YAC DNA.
- **Complicated cloning process:** The cloning of YACs is too complicated to be carried out by a lone researcher.

Retroviral vectors:

Traditionally, the retrovirus is regarded as an enemy to be overcome. However, for the past two decades retroviruses have been harnessed as vehicles for transferring genes into eukaryotic cells, a process known as transduction. During this time, the technology has moved from being a scientific laboratory tool to a potential clinical molecular medicine to be used in gene therapy. This review explains the strategy for harnessing the retrovirus life cycle, the scientific research and clinical applications of this methodology, and its limitations, as well as possible future developments.

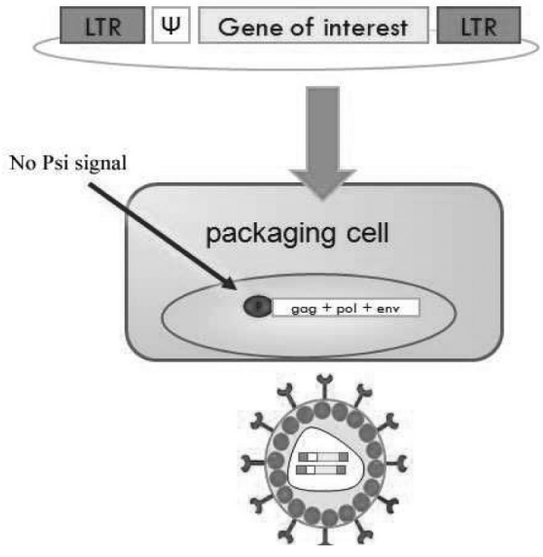


Figure 8.10

Viral techniques of gene transfer harness the method of entry and integration with the host genome used by the wild-type organism. They provide a useful alternative to non-viral methods, which have low transfer efficiencies in certain cell types—for example, primary culture and epithelial cells. The most frequently used vectors are adenoviral or retroviral; however, others include adeno associated virus (AAV), herpes virus, vaccinia virus, and several RNA viruses.

Although adenoviral vectors are useful in transient assays, retroviral vectors stably integrate into the dividing target cell genome so that the introduced gene is passed on and expressed in all daughter cells.

The retrovirus consists of two copies of a single stranded RNA genome with sequences known as gag, pol, and env, which encode viral structural and catalytic proteins. These are surrounded by a glycoprotein envelope. At the onset of infection, the surface glycoprotein envelope interacts with receptors on the surface of the target cell to gain entry. When inside the cell, the single stranded viral genome is converted into linear double stranded DNA by a virus encoded reverse transcriptase. As the target cell

undergoes mitosis, the viral DNA integrates with the target cell DNA—at which point it is known as a provirus. It is this proviral DNA that is manipulated to form retroviral vectors for gene transfer. The provirus then undergoes transcription and translation with the rest of the genome, resulting in the assembly of new viral particles that bud off the surface of the target cell to infect others cells.

A retroviral vector consists of proviral sequences that can accommodate the gene of interest, to allow incorporation of both into the target cells. The vector also contains viral and cellular gene promoters, such as the CMV promoter, to enhance expression of the gene of interest in the target cells. The most important advance in vector technology has been use of the packaging cell.

Expression Vectors:

Expression vectors are actually the plasmids that allow the expression of the foreign DNA. Organization and expression of the eukaryotic genome are nowadays studied in vivo as it gives us the live telecast of working with eukaryotic cells. There are different eukaryotic vectors that can be brought into use for studying the expression of the eukaryotic genome. But there are some vectors which are commonly used such as yeast, animal and plant. In expression vectors it is actually the sequences more commonly called regulatory sequences that allow finished products that is protein to be obtained by means of common pathway of transcription followed by translation of the genes. Most of the drugs containing protein products manufactured by the pharmaceutical industry are made by using expression vectors only. Appropriate selection of vector for maintaining it within the host is an important part of any expression system. A promoter within an expression system can be either regulated or constituted (unregulated) one. Stable proteins are obtained if the promoter is constituted but the protein of interest that is, desired one can be obtained by making use of vector containing regulated promoter. Also most optimal combination of cell density and specific protein can be obtained in short time by making use of regulated promoter. One can easily isolate as well as purify a protein if it is exported outside the cell. Protein modifications are required for proteins produced by recombinant technology are available only in eukaryotic cells. Common host for eukaryotic expression includes yeast, insect cells and mammalian cells. Eukaryotic expression vectors are similar to prokaryotic expression vectors by many ways such as promoter, transcription, transcription and translation signal sequences.

Shuttle vectors are first propagated in bacteria and then transferred to eukaryotic cells for expression as it contains prokaryotic sequences.Yeast can be grown easily on both small scale and large scale and it is also considered as a safe organism. Therefore

it is used in pharmaceuticals for use in human without any approval from the government. It mostly secretes protein in a very small quantity but subsequently engineered to produce recombinant proteins which can be purified easily.

VECTORS FOR EXPRESSION OF PROTEINS IN "YEAST":

Different varieties of vectors are use for use with yeast and so they are classified into three main classes:

1. Plasmid vectors like yeast shuttle vectors.
2. Vectors that integrate into the yeast chromosome but this approach is not used mostly as it gives just the single copy of the cloned gene and is also lost in large scale production.
3. Yeast artificial chromosomes- it is however not so convenient for use as expression vectors.

Highest level of expressions can be obtained by yeast episomal plasmid but they are unstable in large cultures. Insulin, blood coagulation factors, several growth factors and several virus proteins are now produced using *Saccharomyces cerevesiae.*

(i) Yeast Episomal Plasmids (YEps):

They are first constructed by Beggs in 1978 using a naturally occurring yeast plasmid. It is 2μm long that is 6.3 kb and is found in many strain of *Saccharomyces cerevesiae* and has no known function. There are 50-100 copies per cell which base pairs to two unique regions each with a pair of genes transcribed from a divergent promoter. The plasmid may replicated autonomously or integrate with the chromosome. They have been extensively used in the production of either intra- extracellular heterologous proteins. They form the basis of several cloning techniques. It has transformation frequency of 103-105 transformants per μg DNA. They are actually fragment of yeast nuclear DNA and E.coli vector pMB9. There are two stages of preparation

1. Vector pMB9 and 2 μm is cut with EcoRI and then ligated.
2. Nuclear yeast DNA digested with Pst-I is inserted in yeast hybrid.

The advantages are as follows:

1. HCN (50-100).
2. High transformation frequency.
3. High stability.
4. Very useful for studying complementation.
5. Readily recovered from yeast.

The disadvantages are as follows:

1. Recombinant vectors have been developed from this plasmid but are unstable.
2. Novel recombinants are generated in vivo by recombination with endogenous 2μm plasmid.

(ii) Yeast Integrative Plasmid (YIps):

It is bacterial plasmid that can insert itself into DNA of one of the yeast chromosome. Genes integrated into yeast chromosomes are less liable to be lost by the cell as it divides than are the genes on the plasmid. Although transformation efficiency of yeast integrative plasmid is low and the copy number is one, it has proved to be useful in yeast genetics. Yeast integrative are used for placing DNA segments within yeast genome. They can be replicated and maintained in E.coli but not in yeast. It has a transformation frequency of 104 transformants per μg DNA.

The advantages are as follows:

1. Gives most stable maintenance of cloned genes.
2. It behaves as an ordinary genetic marker.
3. Useful for surrogate genetics of yeast like to introduce deletions, inversions etc.

The disadvantages are as follows:

1. Low transformation frequency.
2. Chromosomal DNA needs to be cut with restriction endonuclease for recovering yeast but it doesn't cleave original vector containing cloned gene.

(iii) Yeast Replicative Plasmid (YRps):

These plasmids were constructed by Struhl et al in 1979. It consists of 1.4 Kb yeast DNA fragment containing the trp1 yeast gene inserted into EcoRI site of pBR322. They remain as independent plasmids and do not integrate. Their copy number is 1-20 per cell. They carry autonomous replicating sequences (ARS) that allow them to replicate when the cell divides. These vectors have chromosomal replication origins and give rise to high frequency of transformants that is 104 μg DNA. The resulting transformants are highly unstable.

The advantages are as follows:

1. Can be easily recovered from yeast.
2. It has high copy number.
3. It has high transformation frequency.
4. Very useful for complementation studies.
5. It can be integrated into the chromosome.

The disadvantages are as follows:

Instability of transformants

(iv) Yeast Centromere Plasmids (YCps)

These are the plasmids that contain sequences around the centromere region of chromosomes and chromosomal replication region is similar to yeast replicative plasmids that is it possess autonomous replicating sequences (ARS). It has a transformation frequency of 104 transformants per µg DNA They shows three characteristics of chromosomes in yeast cells and they are as follows:

They are mitotically stable in absence of selective pressure

They segregate in Mendelian manner during meiosis

They are found in LCN that is, 1-2/cell in the host.

Advantage- they can be stably maintained.

Disadvantages are as follows:

1. It has low copy number
2. Wild type yeast cells are force to maintain multiple Yeast centromere plasmids bearing independent selectable markers, the cells grow and cell viability is reduced indicating effect from presence of excess chromosome.

(v) Yeast Artificial Chromosome (YAC):

All the autonomous vectors like YEp, YIp, YCp, YRp exist in yeast as circular DNA molecules, thus none of these resemble normal yeast chromosome that have a linear structure. Also the ends of all chromosomes of yeast have telomeres as that of linear eukaryotic chromosomes. It has two telomeres one of the left and other on the right and thus prevents degradation and are required for chromosomal replication. Origin

of replication that is "ori" site on the plasmid is the site where the DNA replication begins. The presence of yeast centromere helps in proper segregation of chromosome. Chromosome won't get pulled into new cells during cell division without the absence of centromere regions. Selectable markers are also present in some yeast artificial chromosomes. They are present as single copy per cell. In actual, they are hybrids of bacterial plasmid DNA and yeast DNA. They are grown in yeast.

Advantages are as follows:

1. Have large carrying capacity
2. It is highly stable
3. It is very efficient because it mimics the chromosome as it has a sequence that functions as origin of DNA replication, centromeric and telomeric sequences.
4. Large genes such as that for muscular dystrophy can be cloned in linear manner.

Disadvantages are as follows:

1. It is an inefficient system
2. Very few clones can be prepared per µg of DNA
3. Resolution is difficult once introduced into yeast cell
4. Cannot be mapped by standard techniques

Gene cloning strategies: Methods of transforming E.coli and other cells with rDNA:

In vitro packaging of phage or cosmid DNA : The phage or cosmid vectors can be introduced into the host cells by packaging the DNA into phage capsids. The important events in the *in vitro* packaging are as follows: 1) encapsidating the DNA with the phage head precursor and capsid protein and 2) incorporating the products of genes responsible for head and tail. The principle of packaging *in vitro* is to supply the ligated recombinant DNA with high concentrations of phage head precursor, packaging proteins and phage tails. A principal difficulty of this *in vitro* packaging system is that the endogenous DNA from the phage lysates competes with exogenously added recombinant DNA.

***In vitro* packaging of phage DNA:**

Most bacteriophage cloning vectors have been constructed from the bacteriophage

λchromosome. The complete 48,502 nucleotide-pair sequence of the wild type genome is known along with the functions of all of its genes. The central one-third of the λ chromosome contains genes that are required for lysogeny, but not for lytic growth.

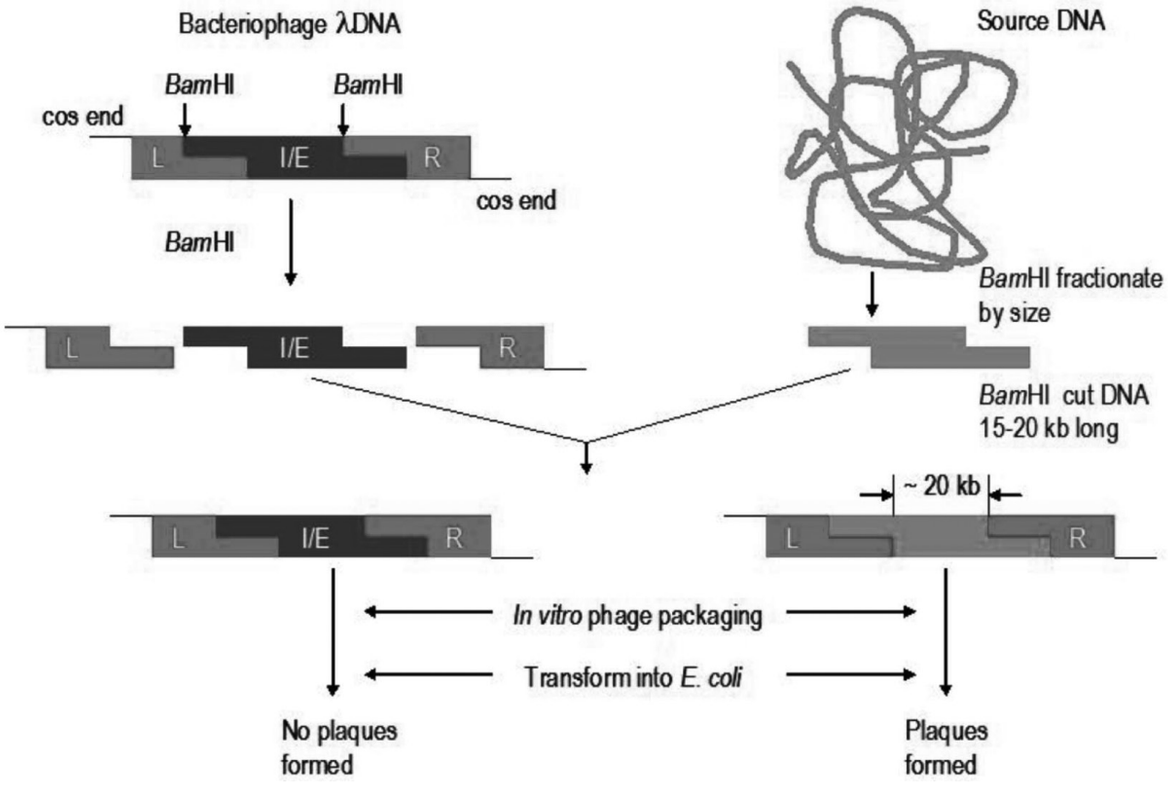


Figure 8.11

Thus the central part (about 15 kb in length) of the λ chromosome can be excised withrestriction enzymes and replaced with foreign DNA. The resulting recombinant DNA molecules can be packaged in phage heads *in vitro.* The phage particles can inject the recombinant DNA molecules into *E. coli* cells, where they will replicate and produce clones of the recombinant DNA molecules.

In vitro packaging of cosmid DNA:

Some eukaryotic genes are larger than 15 kb in size and cannot be cloned intact in eitherplasmid or lambda phage cloning vectors. For this and other reasons, scientists have developed vectors that can accommodate larger DNA insertions. The first such vectors, called **cosmids** (for *cos* site and plasmid), were hybrids between plasmids and the phage λ chromosome. *Cos* stands for cohesive site, in reference to 12 base complementary single stranded termini in the mature λ chromosome.

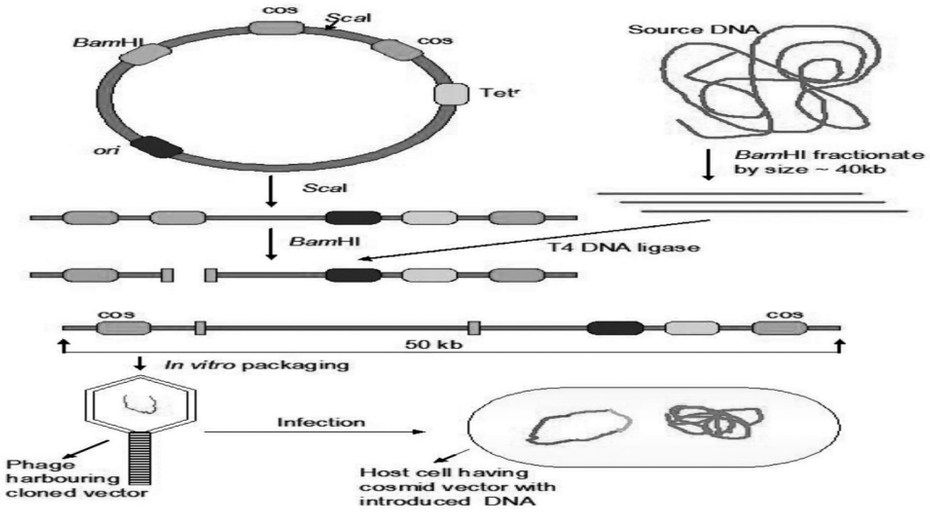


Figure 8.12

The cos site is recognized by the λ DNA packaging apparatus, which makes staggered cuts at this site during packaging to produce complementary cohesive ends of the mature lambda chromosome. Cosmids combine the key advantages of plasmid and phage vectors; they possess1) the plasmid's ability to replicate autonomously in *E. coli* cells and 2) the *in vitro* packaging capacity of λ chromosome. The cosmid vector is capable of accommodating inserts of 35 – 45 kb in size.

Modes of introduction of cloned vectors into hosts:

The foreign DNA materials are incorporated into the hosts by exploiting phenomena *viz.,* **transformation, conjugation, transduction** and **transfection**.

Transformation: Transformation is the process by which microorganisms take up naked DNA and subsequently acquire an altered genotype. Transformation has been reported in various bacteria, in yeast and in some filamentous fungi. This process involves the binding of DNA to competent cells, uptake of the DNA and its establishment within recipients (either as a replicon or byrecombination with resident replicon). The transforming DNA may be of chromosomal or plasmid origin. The ability to introduce plasmids into cells by transformation is an important prerequisite forgene cloning *in vitro*. Accordingly, transformation systems are now being developed for a wide range of microorganisms that have not previously been explored genetically.

Conjugation: Conjugation is the process of gene transfer that requires cellular contact between donors and recipients. The plasmids in bacteria having the property of conjugation are called conjugative plasmids. These plasmids have *tra* gene which control

the transfer of genes. Besides *tra* genes, some conjugative plasmids possess *mob* gene which transfers some of the non-conjugative plasmids that are resident in the same cell. Nature of conjugation varies with gram negative and gram-positive bacteria.

Transduction: Transduction is the process of gene transfer that is mediated by a bacteriophage. There are two types of transductions *viz.,* generalized transduction and specialized transduction. In generalized transduction random parts of donor DNA are packaged in phage capsids and transferred to recipient cells. In specialized transduction, specific regions of bacterial DNA in covalent union with phage DNA are packaged and transduced. Both types of transductions are normally mediated by temperate phages, although certain virulent phages can mediate generalized transduction. The amount of bacterial DNA that can be transduced by a phage can be increased by using phage derivatives that carry deletions in non-essential regions of the phage genome.

Transfection: The uptake and infection of cells with naked viral DNA is generally termed transfection.

Expression of cloned genes: The expression of cloned DNA in the form of functional protein depends on the factors *viz.,* 1) transcription of the appropriate gene, 2) efficient translation of the mRNA and 3) posttranslational processing and compartmentalization of the synthesized protein. A failure to perform correctly in any one of these processes results in the failure of a given gene to be expressed. However, the expression of cloned genes can be improved by constructing the vectors which give improved transcription of inserts and positioning cloned inserts in the correct translational reading frame.

Selection of recombinants:

Making recombinant molecules is a game with very long odds against success. Even when the bits of DNA have been joined up and inserted into cells, only very few cells out of many tens of thousands will contain the recombinant molecule and all the technical expertise in the world is no use whatsoever unless one can find the cell that contains the recombinant DNA. Techniques for selecting the few valuable cells from the mass of useless ones are thus of paramount importance. In the following section the most commonly used method for the selection of recombinant DNA is discussed.

Directional selection: The phenotypes conferred by the cloned genes on the host are used as markers of selection. All useful vector molecules carry a selectable genetic marker or have a genetically selectable property. Plasmid vectors generally possess drug resistance or nutritional markers and in phage vectors the plaque formation itself is the selectable property.

Insertional inactivation: The technique depends upon homologous recombination between DNA cloned and the host genome. If the cloned sequence lacks both promoter and sequences encoding essential regions of the carboxyl terminus of the protein, recombination with homologous genomic sequences will cause gene disruption and produce a mutant genotype. On the other hand, if the cloned fragment contains appropriate transcriptional and translational signals, homologous recombination will result in synthesis of a functional mRNA transcript, and no mutant phenotype will be observed.

Colony hybridization : Various recombinant detection methods employing hybridization with DNA isolated and purified from the transformed cells have been developed of which screening procedure to detect DNA sequences in transformed colonies by hybridization *insitu* with radioactive probe RNA remains as the best choice. In this method, the colonies to be screened are at first replica plated onto a nitrocellulose filter disc that has been placed on the surface of an agar plate prior to inoculation. A reference set of these colonies on the master plate is retained. The filter paper bearing the colonies is lysed and their DNA are denatured. The filter paper is then treated with protease K to remove protein and leave denatured DNA bound to the nitrocellulose, for which it has a high, affinity in the form of a DNA print of the colonies. The DNA is fixed firmly by baking at 80°C.

Selection of recombinant clones by colony hybridization: The labeled probe is hybridized to this DNA, and the result of this hybridization is monitored by autoradiography. A colony whose DNA print gives positive autoradiographic result can then be picked from the reference plate.

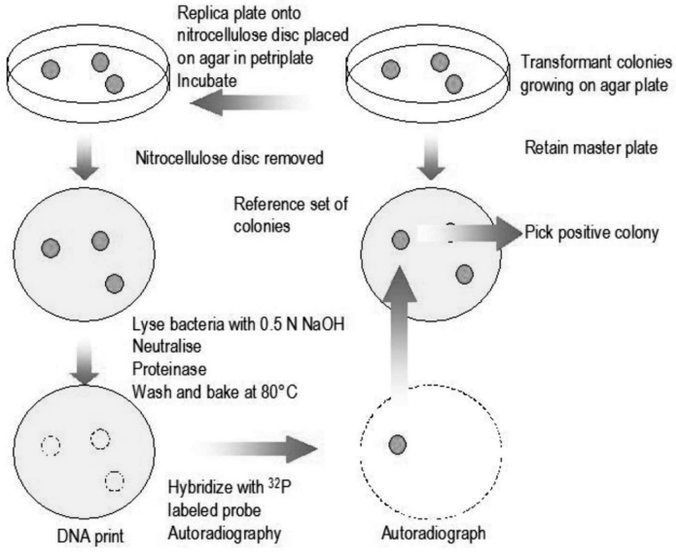


Fig. 8.13: Selection of recombinant clones by colony hybridization

The labeled probe is hybridized to this DNA, and the result of this hybridization is monitored by autoradiography. A colony whose DNA print gives positive autoradiographic result can then be picked from the reference plate.

Immunological screening: Immunological screening involves the use of antibodies that specifically recognizeantigenic determinants on the polypeptide synthesized by a target clone. This is one of the versatile expression cloning strategies, because it can be applied to any protein for which an antibody is available. Furthermore, the protein need not be functional.

Antigen-antibody complex formation in the immunological screening: In the immunological screening method, first transformed cells are plated on agar in a conventional petridish. A replica plate must also be prepared because subsequent procedures kill these colonies. The bacterial colonies are then lysed in one of a number of ways. The lysisreleases the antigens from positive colonies. A sheet of polyvinyl that has been coated with the appropriate antibody (unlabelled) is applied to the surface of the plate, whereupon the antigen complexes with bound antibody (IgGunlabelled). The sheet is removed and exposed to ^{125}I-labelledIgG. The ^{125}IgG can react with the bound antigen *via* antigenic determinants at sites other than those involved in the initial binding of the antigen to IgG coated sheet. Washing the sheet and making an autoradiographic image detect positivelyreacting colonies. The required clones can then be recovered from the replica plate.

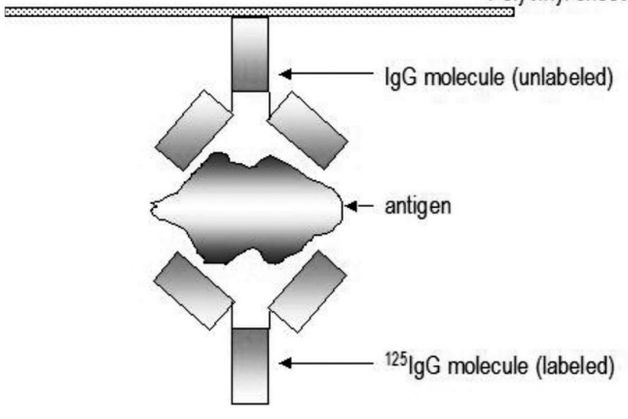


Fig. 8.14: Antigen-antibody complex formation in the immunological screening.

In the immunological screening method, first transformed cells are plated on agar in a conventional petridish. A replica plate must also be prepared because subsequent procedures kill these colonies. The bacterial colonies are then lysed in one of a number of ways. The lysis releases the antigens from positive colonies. A sheet of polyvinyl that has been coated with the appropriate antibody (unlabelled) is applied to the surface

of the plate, whereupon the antigen complexes with bound antibody (IgG unlabelled). The sheet is removed and exposed to ^{125}-I labelled IgG. The ^{125}IgG can react with the bound antigen via antigenic determinants at sites other than those involved in the initial binding of the antigen to IgG coated sheet. Washing the sheet and making an autoradiographic image detect positively reacting colonies. The required clones can then be recovered from the replica plate.

The generalized overview of various cloning strategies and various applications of rDNA technology are schematically represented below.

Cloning strategies:

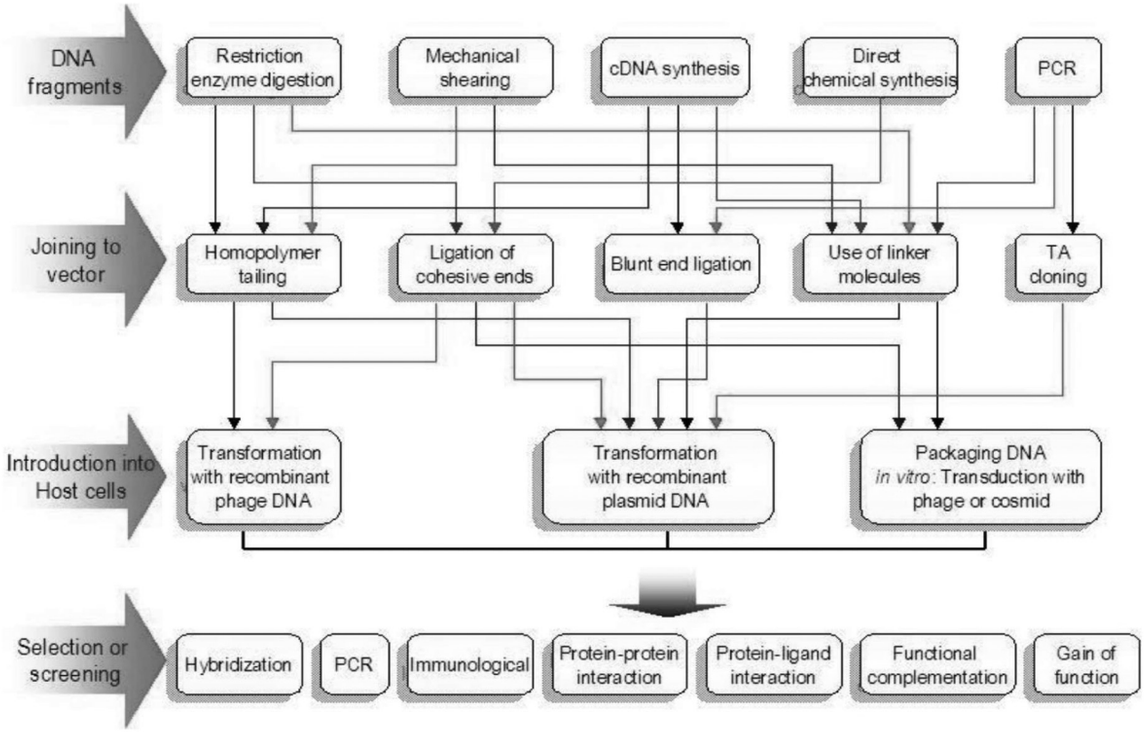


Figure 8.15

Construction of Genomic DNA Library:

DNA Library or Gene Library is simply the collection of DNA fragments cloned into vectors and stored within host organisms. They contain either the entire genome of a particular organism or the genes that are expressed at a given time.

The genome is vast and complex. To understand the entire genome or study specific genes, it is important to study it in smaller and more manageable fragments. DNA

libraries make the entire genome accessible in small fragments. These libraries are used to identify, isolate, and study particular genes of interest.

The development of molecular biology technologies including recombinant DNA (rDNA) technology, cloning vectors, and techniques for transforming bacteria laid the foundation for the construction of DNA libraries.

Types of DNA Library:

Genomic information can be obtained by two primary methods. Based on this, DNA libraries are divided into two types: genomic and cDNA library.

1. Genomic Library:

A genomic DNA library is a collection of DNA fragments that represent all genetic information of an organism. This includes both coding and noncoding regions of the DNA.

Genomic libraries are suitable for a wide range of applications, including genome mapping and comparative genomics. It allows the study of regulatory elements and noncoding sequences that are important in gene expression and regulation.

Genomic libraries also have certain disadvantages including the complexity and size of handling large DNA fragments and the resource-intensive process of creating and maintaining such libraries.

Construction of Genomic Library:

A. Isolation of Genomic DNA:

- The construction of a genomic library begins with the isolation of genomic DNA from the organism of interest.
- Genomic DNA can be isolated using different methods such as cell lysis, protein digestion, and phenol-chloroform extraction.
- The isolated DNA represents the entire genome of the organism and contains both coding and non-coding regions.

B. Fragmentation of Genomic DNA:

- The isolated genomic DNA obtained is often too large to be cloned directly into vectors. So, it needs to be fragmented into smaller fragments suitable for cloning.

- This fragmentation can be achieved using physical methods such as sonication, mechanical shearing, or enzymatic methods involving restriction enzymes.

C. Cloning:

- The fragmented genomic DNA is then cloned into a suitable vector. Some common vectors used for genomic library construction include plasmids, bacteriophages, bacterial artificial chromosomes (BACs), and yeast artificial chromosomes (YACs).
- Bacterial vectors are suitable for smaller DNA fragments, while YACs are used for larger DNA fragments.
- The vectors are treated with restriction enzymes to create sticky ends that are compatible with the fragmented DNA.
- DNA ligase enzyme is used to bind the DNA fragments to the vector, creating recombinant DNA molecules.

D. Transformation:

- The recombinant vectors containing the cloned genomic DNA fragments are transformed into a suitable host organism, usually *E. coli* and yeast.
- The transformed host cells take up the recombinant vectors and are cultured on agar plates containing selective media to allow the growth of colonies containing the recombinant DNA. These colonies form a genomic DNA library.

Applications of genomic libraries:

- Genomic libraries are essential for constructing physical maps of genomes which helps in understanding the layout and structure of genes.
- These libraries are particularly useful for studying non-coding regions, regulatory elements, and gene sequences that may not be expressed.
- Genomic libraries from different species can be compared to study their evolutionary relationships.
- Genomic libraries also help identify genetic mutations and disease-associated genes which is important for understanding the genetic basis of diseases.

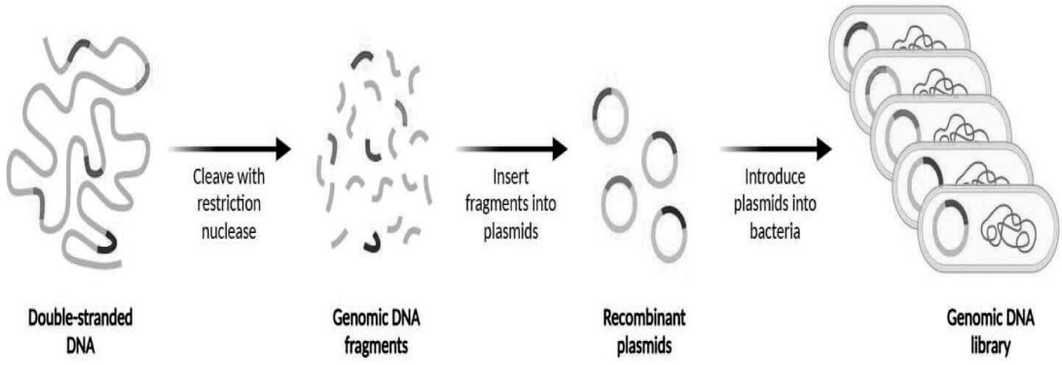


Figure 8.16

2. cDNA Library:

- A cDNA (complementary DNA) library is a collection of cDNA molecules derived from mRNA.
- Unlike genomic libraries, cDNA libraries represent only the expressed genes of an organism, excluding non-expressed genomic regions such as introns and other noncoding sequences.
- cDNA libraries are useful for studying gene expression, protein functions, and producing recombinant proteins. Since cDNA libraries exclude noncoding regions, they provide a more focused view of the expressed genetic information.
- The disadvantages of the cDNA library include the limitation of studying gene regulation due to the absence of regulatory sequences, limited gene diversity, and bias toward highly expressed genes.
- cDNA libraries are specifically created from eukaryotes to study the expressed genes. Prokaryotes do not contain introns. Therefore, creating cDNA libraries for prokaryotes is generally not necessary, as their genomic DNA directly corresponds to their mRNA.

Construction of cDNA Library:

1. Isolation of mRNA:

- Construction of a cDNA library starts with the isolation of mRNA from eukaryotic cells.
- mRNA is isolated and purified using methods such as column purification. Column purification uses oligomeric dT nucleotide-coated resins that bind only mRNA with a poly-A tail.

- Eukaryotic mRNA contains a poly-A tail at the 3' end. When a cell lysate is passed through the poly-T column, the poly-A tails of mRNA molecules bind to the oligo-dT sequences.
- This retains the mRNA in the column while all other molecules that do not have poly-A tails pass through the column and are discarded. Finally, the mRNA molecules are separated from the oligo-dT sequences using an eluting buffer.

2. Synthesis of cDNA:

- After isolating and purifying mRNA molecules, the next step is the synthesis of cDNA molecules from the isolated mRNA.
- At first, a short oligo-dT primer is annealed to the 3' poly-A tails of the mRNA molecule, which initiates the synthesis of the first DNA. Using reverse transcriptase enzyme, the primer is extended to form an RNA-DNA duplex.
- RNase H enzyme degrades the mRNA strand in the mRNA-DNA hybrid, leaving small RNA fragments that are used as primers for the synthesis of the second DNA strand.
- DNA polymerase I synthesizes the second DNA strand in segments removing the RNA primers.
- DNA ligase enzyme seals the nicks between the newly synthesized DNA fragments, resulting in the formation of a double-stranded cDNA copy of the mRNA.

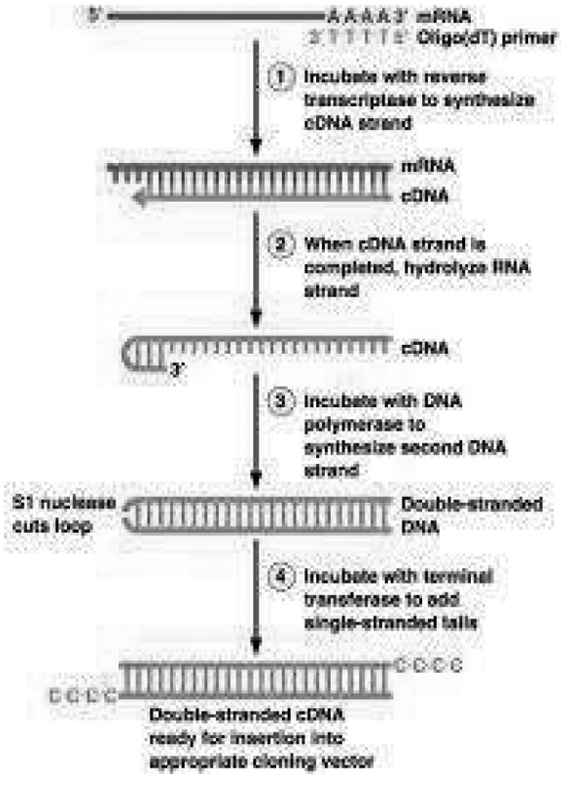


Figure 8.17

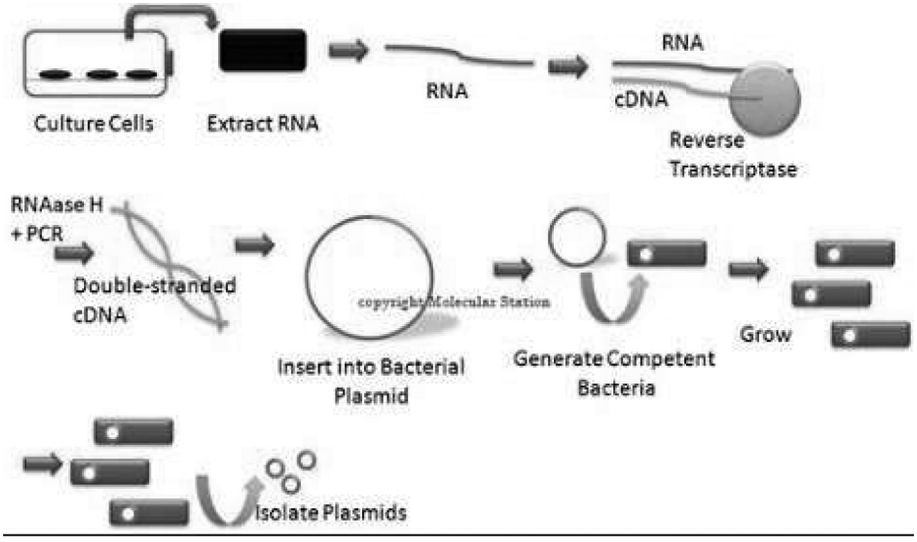
3. cDNA Cloning:

- The next step is the ligation of cDNA molecules into suitable vectors for cloning.
- Since cDNA has blunt ends, restriction site linkers or adapters need to be added to the ends of the cDNA molecules to make them compatible with the vector DNA. These linkers contain recognition sites for restriction enzymes. The linkers are then digested using restriction enzymes.

- The most commonly used vectors for cloning cDNA are plasmid and phage vectors. The suitable vectors are cut with the same restriction enzyme as the cDNA.
- The cDNA molecules are joined with the vector DNA which creates recombinant DNA molecules.

4. Transformation:

- The recombinant DNA molecules are transformed into host cells that can be cultured to produce colonies containing the cloned cDNA inserts.
- To select host cells that have successfully taken up the recombinant DNA, the transformed cells are cultured on agar plates containing a selective medium. The selective medium contains antibiotics that inhibit the growth of untransformed cells.
- The presence of a selectable marker gene on the vector ensures that only cells containing the recombinant DNA survive and form colonies on the selective medium. The resulting colonies form the cDNA library.



colonies form the cDNA library.

Figure 8.18

Applications of cDNA libraries:

- cDNA libraries are useful for studying actively expressed genes in different tissues under specific conditions. cDNA library allows the identification and cloning of expressed genes.

- cDNA libraries can be used to compare gene expression profiles between different species which is useful in the study of evolutionary biology.
- cDNA libraries are also used to produce recombinant proteins.
- These libraries are also used to identify and study the expression of disease-related genes, which can help in the development of diagnostic markers.

Screening of DNA Library:

- After the construction of gene libraries, the screening process for both cDNA and genomic libraries is important to identify and isolate the recombinant clones containing the DNA inserts of interest.
- There are several screening methods. The most commonly used screening method for gene libraries is hybridization-based screening. This method uses labeled DNA or RNA probes that are complementary to the target DNA sequence in the gene library.
- In hybridization-based screening, the colonies or plaques in the gene libraries are transferred onto a solid membrane such as a nitrocellulose membrane.
- The membrane is then probed with radiolabeled probes to identify clones containing the target sequence. The probe will hybridize with its complementary sequence in the target DNA.
- Positive clones are identified based on the presence of hybridization signals, which are visualized using autoradiography or fluorescence detection.
- Other commonly used screening methods include PCR screening, immunological assay, and sequencing-based screening.

Strategies of Expressing cloned genes:

Effectively express cloned genes, researchers use various strategies. These include carefully selecting a suitable host organism and engineering the vector DNA to ensure correct protein production and sufficient amounts. This involves optimizing the coding and regulatory sequences within the vector. The chosen host and vector must also be compatible for transformation and successful expression of the cloned gene.

Here's a more detailed look at the strategies:

1. Choosing a Host and Vector:

- **Host Selection:** The host organism must be compatible with the vector and capable of replicating and expressing the cloned gene efficiently. Common choices include bacteria (like E. coli), yeast, and mammalian cells.
- **Vector Selection:** The vector (e.g., a plasmid) needs to contain appropriate elements for replication, expression, and selection in the chosen host. These elements include an origin of replication, selectable markers (like antibiotic resistance genes), and promoter sequences.

2. Engineering the Vector DNA:

- **Promoter and Regulatory Sequences:** Ensuring the cloned gene is under the control of a strong and inducible promoter is crucial for efficient protein expression. Inducible promoters allow for controlled protein production when needed.
- **Ribosomal Binding Site (RBS):** The RBS is a sequence in mRNA that is recognized by ribosomes and is essential for translation (protein synthesis).
- **Coding Sequence Optimization:** The coding sequence of the cloned gene may need to be optimized for the chosen host's codon usage preferences to ensure efficient translation.
- **Signal Peptides:** For proteins that need to be secreted, a signal peptide sequence may need to be added to the vector.

3. Transformation and Selection:

- **Transformation:** The recombinant DNA (containing the cloned gene and vector) is introduced into the host cells using methods like transformation or transduction.
- **Selection:** Once the host cells are transformed, selection methods are used to identify cells that successfully took up the recombinant DNA. This can be done using selectable markers like antibiotic resistance genes.

4. Expression and Purification:

- **Protein Expression:** Once the cells containing the cloned gene are identified, they are grown and induced to express the protein product.
- **Protein Purification:** The expressed protein is then purified from the host cells using techniques like centrifugation, chromatography, and other methods.

Phage Display:

Phage display was first developed by G. Smith in 1985 as a method of presenting polypeptides on the surface of lysogenic filamentous bacteriophages. In phage display technique, a gene encoding a protein of interest is inserted into a phage coat protein gene, causing the phage to display the protein on the outside. And containing the gene for the protein inside, resulting in a connection between genotype and phenotype.

These displaying phages can be screened for other proteins, peptides or DNA sequences, in order to detect interaction between the displayed protein and those other molecules. So, it has become one of the most powerful and widely used laboratory technique for the study of protein-protein, protein-peptide and protein-DNA interactions.

In this way, large libraries of proteins can be screened and amplified in the process called in vitro selection, which is analogous to natural selection. Phage display is also an effective way for producing large amounts of peptides, proteins and antibodies.

Phage display technology has evolved into an extremely versatile and powerful platform for protein engineering.

2. Phage Display Protocol:

In general, there are 5 steps for phage display technology stated as below:

STEP 1: Construct phage display library:

Recombinant DNA technology is used to incorporate foreign cDNA into viral DNA. Different sets of genes are inserted into the genomes of multiple phages. Spliced into gene for a coat protein, so that the protein will be displayed on the outside of phage particles, and these separate phages will only display one protein, peptide, or antibody.

Collections of these phages can comprise libraries, such as antibody phage library, protein phage library, or random phage library.

STEP 2: Binding:

These libraries are exposed to selected targets and only some phages will interact with targets. The target is for which specific ligands planned to be identified such as immobilized protein, cell surface protein or vascular endothelium.

STEP 3: Washing:

Unbound phages can be washed away, and only those which showing affinity for the receptors was left.

STEP 4: Elution:

Recovery of the target bound phage by elution.

STEP 5: Amplification:

Eluted phages showing specificity are used to infect new host cells for amplification, or direct bacterial infection and amplification of the recovered phage.

Back to step 1, repeated cycle 2-3 times for stepwise selection of best binding sequence. After that, Enrichment and purification of the phage repertoire by precipitation methods to increasing the phage titer.

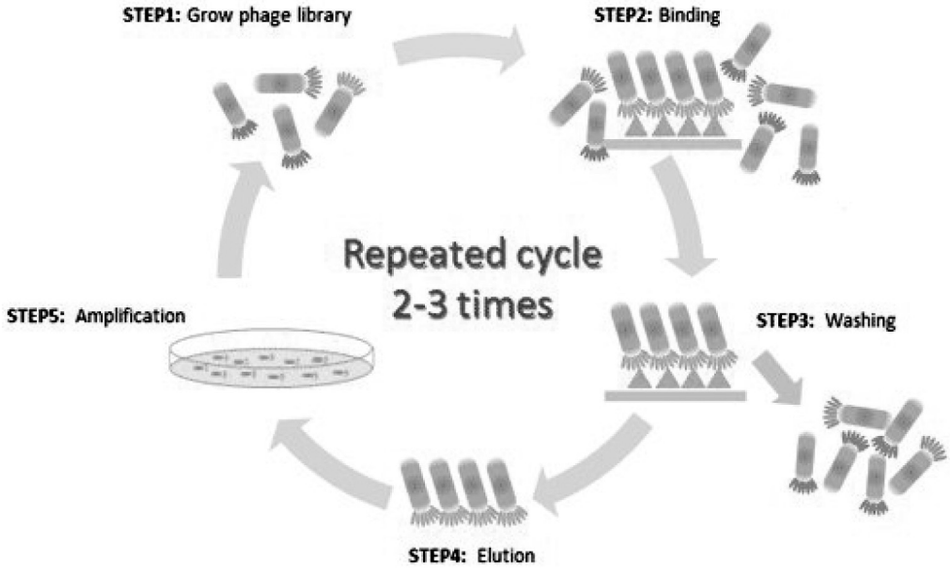


Figure 8.19

3. Classification of Phage Display Systems:

The most common bacteriophages used in phage display are E.coli filamentous bacteriophages (f1, fd, M13) , though T4, T7, and λ phage have also been used.

3.1 E.coli filamentous bacteriophages (f1, fd, M13)

A filamentous bacteriophage is a type of phage, defined by its filament-like or rod-like shape. Filamentous phages usually contain a genome of single-stranded DNA and infect Gram-negative bacteria. The family of Ff, M13, fd, and f1 are vital phages which have utility in phage display among which M13 phage is the most generally used.

M13 bacteriophage has a cylindrical shape with a length of 880nm and a diameter of 6nm. It encapsulates a single-strand genome that encodes five different capsid pro-

teins which comprise two groups, major coat proteins (pVIII) and minor coat proteins (pVII, pIX, pVI and pIII).

E. coli filamentous bacteriophages are commonly used for phage display. Most antibodies and peptides are displayed at phage proteins pIII and pVIII, which constructed pIII and pVIII display system. Moreover, hybrid phage system enables displaying large proteins with all five M13 coat proteins as N-terminal fusions with pIII, pVIII, pVII and pIX, and also as C-terminal fusions with pVI, pIII, and pVIII.

pIII is the protein that determines the infectivity of the virion. It consists of 406 amino acid residues and occurs at the phage tip in 3 to 5 copies. An advantage of using pIII rather than pVIII is that pIII allows for monovalent display when using a phagemid combined with a helper phage. Moreover, pIII allows for the insertion of larger protein sequences (>100 amino acids) and is more tolerant to it than pVIII.

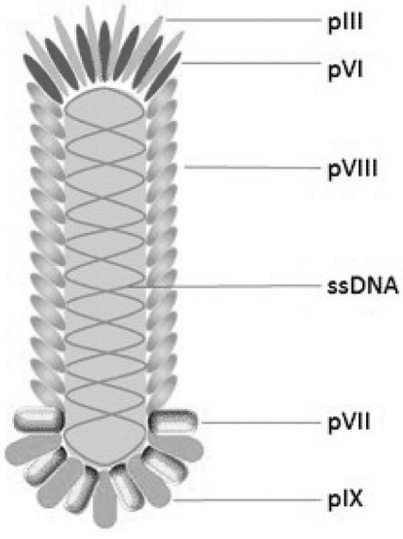


Figure 8.20

pVIII is the main coat protein of Ff phages, which is expressed by gene 8 and occurs in 2700 copies. Therefore it is used to enhance detection signal when phage displayed antibody associates with antigen. Peptides are usually fused to the N-terminus of Pviii which are usually 6-8 amino acids long . This makes the use of this protein unfavorable for the discovery of high affinity binding partners. Moreover modifications of pVIII are made to increase the efficiency of display onto pVIII[9], and now there has been great progress.

pVI has been widely used for the display of cDNA libraries, which is an attractive alternative to the yeast-2-hybrid method for the discovery of interacting proteins and peptides due to its high throughput capability pVI has been used preferentially to pVIII and pIII for the expression of cDNA libraries because one can add the protein of interest in the C-terminus of pVI without greatly affecting pVI's role in phage assembly.

pVII and **pIX**, located to the phage tip opposite that of pIII, may both complement current phage display systems and be used as alternative scaffolds for display and selection to further improve phage display as the ultimate combinatorial engineering platform.

3.2 T4 Phage

T4 phage (Enterobacteria phage T4) is a bacteriophage that infects E.coli bacteria. It is a member of the T-even phages, a group including enterobacteriophages T2 and T6. T4 phage is a relatively large phage, at approximately 90 nm wide and 200 nm long. It's double-stranded DNA genome is about 169 kbps long and encodes 289 proteins. T4 phage is built with three essential proteins: gp23, which forms the hexagonal capsid lattice; gp24, which forms pentamers at eleven of the twelve vertices; and gp20, which forms the unique dodecameric portal vertex through which DNA enters during packaging and exits during infection.

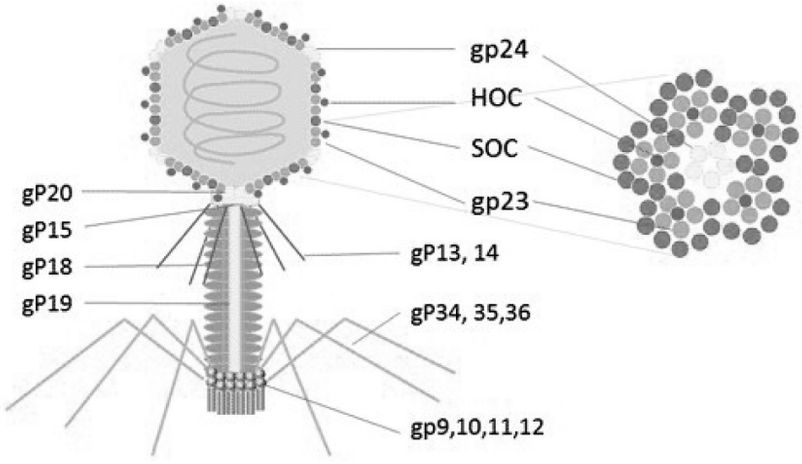


Figure 8.21

In addition to the essential capsid proteins, gp23, gp24, and gp20, the T4 capsid is decorated with two nonessential outer capsid proteins: HOC (highly antigenic outer capsid protein) and SOC (small outer capsid protein). Both HOC and SOC are dispensable, and bind to the capsid after the completion of capsid assembly. Null (amber or deletion) mutations in either or both the genes do not affect phage production, viability, or infectivity.

HOC and SOC are dispensable T4 capsid proteins that can be used for phage display of multiple copies of peptides and proteins. The phage T4 HOC, SOC bipartite display system is attractive for the expression of cDNA and display of peptides or proteins at high copy numbers on the phage capsid surface[12]. It could be applied to cDNA expression, displays larger proteins in high copy number and inserts into stop codon on the C-terminal of SOC protein that occurs in 810 copies or N-terminal of HOC protein that occurs in 155 copies.Therefore, the phage T4 dual site display emerges as a powerful method with an enhanced immune response in animals for research and development of immunological products.

3.3 T7 Phage

T7 phage is an icosahedral virus of the Podoviridae family and has a linear double stranded (ds) DNA genome. Similar to T4, bacteriophage T7 possesses a head and tail structure. The icosahedral head, where T7 conserves its dsDNA genome, is composed of 415 copies of capsid gp10, arranged as 60 hexamers on the surface and 11 pentamers at the vertices. There exist two isoforms of major coat protein gp10, gp10A and gp10B, in a nine-to-one ratio, resulted from natural translational frame shift at amino acid 341.

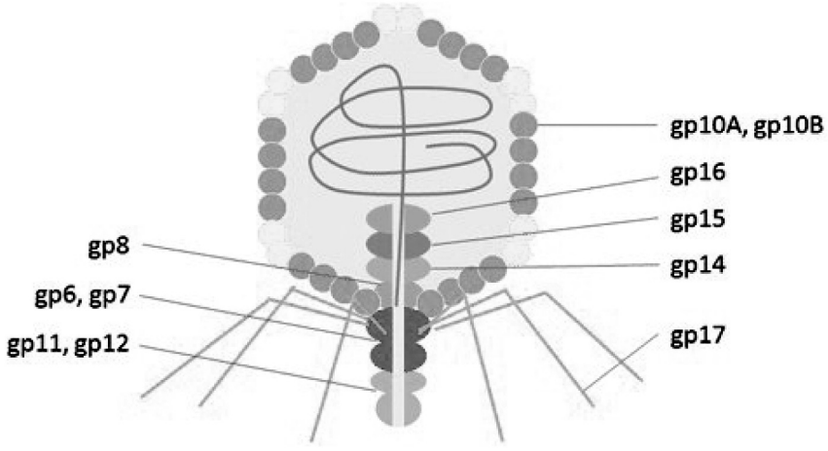


Figure 8.22

The minor protein gp10B results from a frame shift in the end of the gene that makes the capsid protein 5′ residues longer. Fusion proteins are displayed on protein gp10B C-terminally of the 5′ extra residues. So that it can avoid problems associated with steric hindrance. T7 phage particles exhibit high stability under various extreme conditions, including high temperature and low pH, which facilitates effective high throughput affinity elutriation. Recent applications of the T7 phage display system have been instrumental in uncovering mechanisms of molecular interaction, particularly in the fields of antigen discovery, vaccine development, protein interaction, and cancer diagnosis and treatment.

3.4 lambda phage

lambda phage (Enterobacteria phage λ, coliphage λ) is a bacterial virus which infects the bacterial species *E. coli*. It was discovered by Esther Lederberg in 1950. The lambda phage has an icosahedral head. The main structure of the shell is built from the major coat protein gpE (415 copies) and is stabilized by the capsid protein gpD (402–420 copies) . The head is linked to a flexible helical tail constructed by disks of the major tail protein gpV.

Both the tail protein gpV and the head protein gpD have been used for phage display. Initially, the lambda foo vector was constructed for the C-terminal display on gpV, with a low display level that made it suitable for capturing high-affinity interactions . Later, systems were developed for the display of peptides N-terminally or C-terminally to the major coat protein gpD . The lambda phage has been engineered to display efficiently multiple copies of peptides or even large protein domains providing a powerful tool for screening libraries of peptides, proteins and cDNA.

4. Applications of Phage Display

Nowadays, phage display as a rapidly developing technology has been used in a wide range of applications in different research areas including epitope mapping, identification of new receptors & ligands, in vitro protein evolution, drug discovery, and so on. Some of the most successful applications of phage display are explained in the following sections.

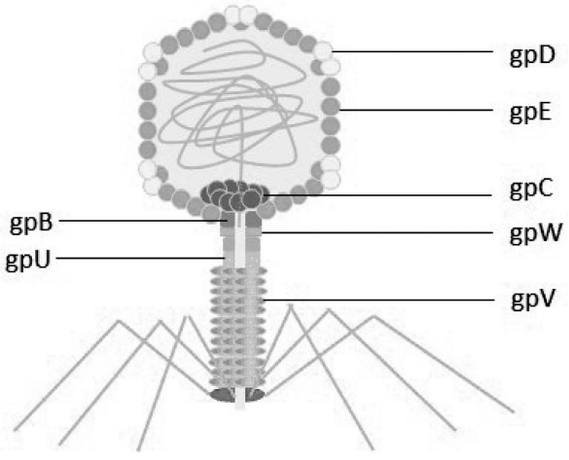


Figure 8.23

4.1 Epitope mapping and mimicking

Upon encountering antigen, host humoral immunity activates and triggers production of antibodies which directed against foreign protein epitopes. Knowledge of these protein epitopes is pivotal in understanding the pathogenesis of pathogen infections and in developing diagnostic reagents, therapeutic antibodies, and effective vaccines.

Peptide phage display libraries are useful tools for identification of continuous or linear epitopes involving in interaction with antibody. Isolation and identification of mimotopes from peptide phage libraries is powerful approach to improve immunological studies in order to design and develop vaccine candidate.

4.2 Identification of new receptors & ligands

Phage display method using either gene-specific libraries, or random peptide libraries, provides a powerful technique for an approach to epitope identification. The technique can identify amino acids on protein antigens that are critical of antibody binding.

The random peptide sequence was displayed on the surface of the phage to obtain the phage display polypeptide library. Polypeptides identifying specific cells were ob-

tained by differential screening using cells as screening targets. By studying the polypeptide sequence, we can further to obtain the receptor protein expressed specifically on the cell surface.

4.3 Protein-protein interaction studies

Protein–protein interactions mediate essentially all biological processes. A detailed understanding of these interactions is thus a major goal of modern biological chemistry. Phage display is a potent and versatile method for studying protein-protein interaction. It can be applied to a wide range of protein interaction partners and used in a number of applications, especially in mapping intracellular interactions of the distinct protein domain.

The polypeptide library displayed by bacteriophages is composed of random short peptide sequences of specific length. Short peptide sequences can be obtained by affinity screening for the random library by target proteins (such as receptors). Sequence analysis and synthesis of corresponding short peptides, and then we can study the interaction between two proteins.

4.4 Recombinant Antibody Production

Recombinant antibodies are useful tools for therapy, diagnosis and research. With major developments in molecular biology, numerous display technologies have been successfully introduced for recombinant antibody production. Even so, antibody phage display still remains the gold standard for recombinant antibody production. Its success is mainly attributed to the robust nature of phage particles allowing for automation and adaptation to modifications.

a. **Production of gene fragment**: This phase involves animal immunization with the desired antigen and then isolation of B lymphocytes, mRNA extraction and cDNA synthesis. The synthesized cDNA contain genetic information of all antibodies targeting various antigens and consist of approximately 109 to 1011 lymphocyte clones.

b. **Cloning of gene fragments in the phagemid vectors**: Genes related to the different clones of antibodies are digested with restriction enzymes, clone into phagemid vectors and then display on the surface of phages. Sequence diversity of the fragments at this step leads to optimum isolation of antibodies in the later steps. The phagemid vectors need helper phage to package and exit from the bacterial cells and enter to the medium.

c. **Selection of specific phages**: After cloning of the fragments into phagemid vectors, variety of clones of antibodies display on the surface of the phage. Selections of specific clone that recognize the antigen (target of interest) perform by biopanning. Then the phage carrying specific antibody can be isolated from non-specific phages due to antigen–antibody binding properties.

d. **Screening**: Isolation of antibodies with high affinity to target is the main aim of this step. Screening is performed using different methods like: immunoassay, immunocytochemistry, active isolation of cells due to their fluorescent properties and immunoblotting.

4.5 Protein directed evolution

Phage display technology as a selection based system is an attractive method for evolution of new biological drugs. Directional transformation protein refers to use cassette mutation, error-prone PCR method to mutant protein or a particular cod sequence structure domain. Proteins or structural domain mutations present library will be display in phage surface. Then we can obtain the required have directional change of phage clones by affinity screening. The primary structure can be derived from the sequence of DNA that can be used to screen more cytokines receptors ability, new enzyme inhibitors, DNA transcription factors in combination with the new sites, new cytokine antagonists, new enzymes and enhance biological active protein.

4.6 Drug discovery

Peptides as biologically active molecules in hormones, neurotransmitters, cytokines, antigens, and growth factors are involved in a wide variety of biological processes. So, peptides are extensively used as therapeutic and diagnostic agents in the medical fields such as oncology, endocrinology, urology, and obstetrics. The peptide phage libraries with presenting a huge number of different peptides mimicking the genuine epitopes play a key role in the development of new therapeutic peptides. Until now, several peptide drugs have been developed using phage display technology.

4.7 In vitro diagnostic

At present, phage display library technology is broadly employed to examine host-pathogen interaction, development of disease diagnostic markers, and identification of vaccine candidates and novel antibodies against pathogenic targets.

The cell-surface antigens in pathogens as molecular binding sites are suitable targets for vaccine development, since these antigens can affect the bacterial division, replication, and virulence of pathogens. There are two distinct strategies for application of phage display technology in infectious diseases area: In the pathogen-targeted phage display, molecular targets such as cell replication enzymes or host-pathogen virulence factors are targeted for screening; while in the cell-based screening method, bacterial whole cell is employed as target for screening. The cell-based screening with live pathogens can target the native cell surface proteins. This screening method is suitable when target antigen is unstable under immobilization conditions.

Declaration: All contents prepared for this SLM have been taken from several molecular biology and recombinant DNA technology books, research articles, online book chapters, Wikipedia and other sources from google**.**

Probable Questions

1. Describe the structure and features of pBR322 and explain its significance as a cloning vector.
2. Compare and contrast the properties and applications of plasmid vectors and phage vectors in recombinant DNA technology.
3. Explain the mechanism of action and advantages of using single-stranded DNA vectors like M13 in cloning experiments.
4. Discuss the structure, components, and applications of high-capacity vectors such as BACs and YACs.
5. What are the key strategies used for the expression of cloned genes in yeast? Describe the different types of yeast expression vectors.
6. Outline the different methods of introducing recombinant DNA into host cells and compare their mechanisms.
7. Explain the principle and steps involved in the construction of a genomic DNA library and its applications.
8. Discuss the cDNA library construction process, highlighting how it differs from genomic library construction and its specific uses.
9. Describe the immunological screening method for selection of recombinant clones and discuss its advantages.
10. What is phage display technology? Describe its protocol, different phage systems used, and various applications in research and medicine.

Unit-IX

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

1. Manipulating Genes in Animals: Gene Transfer to Animal Cells

Introduction

Gene transfer to animal cells is a cornerstone of genetic engineering, enabling the study of gene function, the development of disease models, and the production of therapeutic proteins. This process involves introducing foreign genetic material into animal cells to alter their genetic makeup, leading to transient or stable expression of the introduced genes.

Methods of Gene Transfer

1.1 Physical Methods

1.1.1 Microinjection

- *Definition*: Direct injection of DNA into the nucleus of a cell using a fine glass micropipette.
- *Application*: Commonly used for creating transgenic animals by injecting DNA into fertilized oocytes.
- *Advantages*: High precision; allows for targeted delivery.
- *Limitations*: Technically demanding; low efficiency; potential for damage to the embryo.

1.1.2 Electroporation

- *Definition*: Application of an electrical field to cells to increase membrane permeability, facilitating DNA uptake.

- *Application*: Suitable for both in vitro and in vivo gene transfer.
- *Advantages*: Simple and rapid; applicable to various cell types.
- *Limitations*: Cell viability can be affected; optimization required for different cell types.

1.1.3 Biolistics (Gene Gun)

- *Definition*: Delivery of DNA-coated microscopic particles (usually gold or tungsten) into cells using high-velocity propulsion.
- *Application*: Used in tissues where other methods are less effective.
- *Advantages*: Useful for hard-to-transfect cells; no need for vectors.
- *Limitations*: Potential for tissue damage; lower efficiency compared to viral methods.

1.2 Chemical Methods

1.2.1 Calcium Phosphate Precipitation

- *Definition*: DNA is precipitated with calcium phosphate and then taken up by cells via endocytosis.
- *Application*: One of the earliest methods used for gene transfer in cultured cells.
- *Advantages*: Cost-effective; simple to perform.
- *Limitations*: Variable efficiency; not suitable for all cell types.

1.2.2 Lipofection

- *Definition*: Use of liposomes (lipid vesicles) to encapsulate DNA and facilitate its fusion with the cell membrane.
- *Application*: Widely used for transfecting mammalian cells.
- *Advantages*: High efficiency; low toxicity.
- *Limitations*: Transient expression; may not be effective for all cell types.

1.3 Viral Methods

1.3.1 Retroviral Vectors

- Definition: Use of modified retroviruses to deliver genetic material into host cells.
- *Application*: Stable integration into the host genome; used in gene therapy.

- *Advantages*: Long-term expression; efficient for dividing cells.
- *Limitations*: Risk of insertional mutagenesis; limited to dividing cells.

1.3.2 Adenoviral Vectors

- *Definition*: Use of adenoviruses to deliver DNA without integrating into the host genome.
- *Application*: Transient expression in both dividing and non-dividing cells.
- *Advantages*: High transduction efficiency; broad host range.
- *Limitations*: Immune response; transient expression.

1.3.3 Lentiviral Vectors

- *Definition*: Subclass of retroviruses capable of infecting non-dividing cells.
- *Application*: Stable gene transfer in a wide range of cell types.
- *Advantages*: Efficient integration; long-term expression.
- *Limitations*: Complex production; biosafety concerns.

Applications of Gene Transfer

- **Functional Genomics**: Understanding gene function by overexpressing or silencing specific genes.
- **Gene Therapy**: Treating genetic disorders by correcting defective genes.
- **Biopharmaceutical Production**: Producing therapeutic proteins in animal cells.
- **Transgenic Animal Models**: Creating animals with specific genetic modifications for research.

Challenges and Considerations

- **Efficiency**: Varies depending on the method and cell type.
- **Safety**: Potential for insertional mutagenesis and immune responses.
- **Ethical Concerns**: Especially relevant in germline modifications.

2. Genetic Manipulation of Animals: Transgenic Technology

Introduction

Transgenic technology involves the introduction of foreign genes (transgenes) into an organism's genome, resulting in the expression of new traits. In animals, this technology has revolutionized research, agriculture, and medicine by enabling the development of models for human diseases, improved livestock, and biopharmaceutical production.

Methods of Creating Transgenic Animals

2.1 DNA Microinjection

- *Process*: Direct injection of DNA into the pronucleus of a fertilized egg.
- *Application*: Commonly used in mice to create transgenic lines.
- *Advantages*: Direct method; relatively straightforward.
- *Limitations*: Low efficiency; random integration; potential for mosaicism.

2.2 Embryonic Stem (ES) Cell-Mediated Gene Transfer

- *Process*: Genetic modification of ES cells followed by incorporation into embryos.
- *Application*: Allows for targeted gene modifications (knockouts/knockins).
- *Advantages*: Precise genetic alterations; selection of modified cells.
- *Limitations*: Limited to species with established ES cell lines.

2.3 Somatic Cell Nuclear Transfer (SCNT)

- *Process*: Transfer of a nucleus from a genetically modified somatic cell into an enucleated oocyte.
- *Application*: Cloning of transgenic animals; used in livestock.
- *Advantages*: Allows for precise genetic modifications; applicable to various species.
- *Limitations*: Technical complexity; ethical concerns.

2.4 Sperm-Mediated Gene Transfer (SMGT)

- *Process*: Binding of DNA to sperm cells, which then deliver the DNA during fertilization.
- *Application*: Experimental method for generating transgenic animals.

- *Advantages*: Simple; non-invasive.
- *Limitations*: Low efficiency; inconsistent results.

2.5 Viral Vector-Mediated Gene Transfer

- **Process**: Involves the use of genetically engineered viruses (such as retroviruses, lentiviruses, adenoviruses) to deliver transgenes into early embryos or germ cells.
- **Application**: Widely used in animal models for gene therapy research, functional genomics, and disease modeling.
- **Advantages**: High efficiency of gene delivery; some vectors integrate into the host genome for long-term expression.
- **Limitations**: May induce immune responses or cause insertional mutagenesis; biosafety concerns require stringent containment.

2.6 CRISPR/Cas9-Based Genome Editing

- **CRISPR** (Clustered Regularly Interspaced Short Palindromic Repeats) and **Cas9** (CRISPR-associated protein 9) have revolutionized transgenic technology.
- **Process**:
- A guide RNA (gRNA) is designed to target a specific DNA sequence.
- Cas9 enzyme creates a double-stranded break at the targeted site.
- The cell's natural repair mechanisms (NHEJ or HDR) either disrupt or replace the gene.
- **Applications**:
- Precise genome editing for gene knockouts, knockins, or corrections.
- Modeling human genetic diseases in animals.
- **Advantages**: High specificity, scalability, and efficiency.
- **Limitations**: Off-target effects; delivery challenges in some organisms.

Applications of Transgenic Animals

2.6.1 Biomedical Research

- *Disease Models*: Transgenic mice expressing human genes are used to model cancer, neurodegenerative diseases (e.g., Alzheimer's), cardiovascular diseases, etc.

- *Pharmacological Testing*: Testing drug efficacy and toxicity on genetically modified animals mimicking human physiology.

2.6.2 Biopharming

- Transgenic animals are engineered to produce therapeutic proteins (e.g., insulin, antithrombin) in milk, eggs, or blood.
- **Example**: Transgenic goats producing antithrombin III in milk, used to treat blood clotting disorders.

2.6.3 Agriculture

- Enhanced traits such as faster growth (e.g., transgenic salmon), disease resistance, and improved nutritional content (e.g., enriched milk composition).
- Genetic modifications in livestock to improve reproduction, meat quality, or milk yield.

2.6.4 Xenotransplantation

- Use of transgenic pigs whose organs are genetically modified to reduce immunological rejection in humans.

2.6.5 Environmental and Conservation Applications

- Engineering animals to serve as biosensors for environmental pollutants.
- Potential role in reviving endangered species through genetic restoration.

Ethical, Regulatory, and Biosafety Issues

Ethical Concerns

- Welfare of genetically modified animals.
- Concerns over 'playing God' and interfering with natural processes.
- Ethical boundaries between therapeutic use and enhancement.

Regulatory Aspects

- Oversight by regulatory bodies such as:
  - **U.S. Food and Drug Administration (FDA)**
  - **European Medicines Agency (EMA)**
  - **Indian Department of Biotechnology (DBT)**

- Guidelines ensure:
    - Risk assessment of transgene effects.
    - Environmental impact evaluation.
    - Humane treatment of animals.

Biosafety Measures

- Risk of horizontal gene transfer.
- Strict containment and monitoring required to prevent unintentional release.
- Need for public awareness and engagement in policy decisions.

Future Directions in Transgenic Technology

- **Improved Precision**: Advancements in base editing and prime editing for even more accurate gene modifications.
- **Multi-Gene Editing**: Simultaneous editing of multiple loci using multiplex CRISPR systems.
- **Synthetic Biology Integration**: Designing entirely new gene circuits or synthetic organisms.
- **Translational Research**: More robust humanized animal models for clinical testing.

3. Transgenic Technology

Introduction

Transgenic technology involves the introduction of foreign genes (transgenes) into an organism's genome, resulting in the expression of new traits. This technology has revolutionized various fields, including agriculture, medicine, and research, by enabling the development of organisms with desirable characteristics.

Definition

A transgenic organism is one that carries a gene or genes that have been artificially inserted instead of the organism acquiring them through reproduction. The inserted gene sequence (known as the transgene) may come from another unrelated organism or from the same species. The process of creating transgenic organisms is known as transgenesis.

Methods of Creating Transgenic Organisms

- **Microinjection**: Direct injection of DNA into the nucleus of a fertilized egg. This method is commonly used in creating transgenic mice.
- **Embryonic Stem Cell Transformation**: Involves introducing the desired gene into embryonic stem cells, which are then incorporated into an embryo.
- **Viral Vectors**: Utilizes modified viruses to deliver the transgene into the host genome.
- **Electroporation**: Application of an electrical field to cells to increase the permeability of the cell membrane, allowing DNA to enter.
- **Gene Gun (Biolistics)**: DNA-coated microscopic particles are physically shot into target cells.
- **CRISPR/Cas9 Technology**: A genome editing tool that allows for precise modifications at specific locations in the DNA.

Applications of Transgenic Technology

Agriculture

- **Enhanced Crop Traits**: Development of crops with improved yield, pest resistance, and tolerance to environmental stresses.
- **Nutritional Enhancement**: Biofortification of crops to address nutritional deficiencies, such as Golden Rice enriched with Vitamin A.

Medicine

- **Pharmaceutical Production**: Transgenic animals producing human proteins, such as insulin or clotting factors, in their milk.
- **Gene Therapy**: Correcting defective genes responsible for disease development.
- **Disease Models**: Creation of transgenic animals that mimic human diseases for research purposes.

Industrial Applications

- **Bioremediation**: Transgenic organisms engineered to degrade environmental pollutants.
- **Biofactories**: Production of enzymes and other industrially relevant proteins.

Ethical and Biosafety Considerations

The creation and use of transgenic organisms raise several ethical and biosafety concerns:

- **Environmental Impact**: Potential for transgenes to spread to wild populations.
- **Animal Welfare**: Concerns about the well-being of transgenic animals.
- **Human Health**: Unintended effects of consuming genetically modified organisms (GMOs).

Regulatory frameworks and risk assessment protocols are essential to address these concerns and ensure the safe application of transgenic technology.

4. Applications of Recombinant DNA Technology

Introduction

Recombinant DNA (rDNA) technology involves combining DNA molecules from different sources into one molecule to create new genetic combinations. This technology has become a cornerstone in biotechnology, with applications spanning medicine, agriculture, and industry.

Medical Applications

- **Production of Therapeutic Proteins**: rDNA technology enables the production of human insulin, growth hormones, and clotting factors in microbial systems.
- **Vaccine Development**: Creation of recombinant vaccines, such as the hepatitis B vaccine, which are safer and more effective.
- **Gene Therapy**: Correcting genetic defects by introducing functional genes into a patient's cells.
- **Diagnostic Tools**: Development of DNA probes and PCR-based tests for detecting genetic disorders and pathogens.

Agricultural Applications

- **Genetically Modified Crops**: Development of crops with enhanced traits, such as pest resistance, herbicide tolerance, and improved nutritional content.
- **Animal Husbandry**: Creation of transgenic animals with desirable traits, such as faster growth rates or disease resistance.
- **Biopesticides and Biofertilizers**: Engineering microorganisms to produce substances that protect plants or enhance soil fertility.

Industrial Applications

- **Enzyme Production**: Microorganisms engineered to produce enzymes used in detergents, food processing, and textile manufacturing.
- **Biodegradable Plastics**: Production of polyhydroxyalkanoates (PHAs) by genetically modified bacteria as an alternative to petroleum-based plastics.
- **Biofuels**: Engineering microbes to produce ethanol and other biofuels from renewable resources.

Environmental Applications

- **Bioremediation**: Use of genetically modified organisms to clean up oil spills, heavy metals, and other environmental contaminants.
- **Biosensors**: Development of organisms that can detect and signal the presence of environmental pollutants.

Ethical and Safety Considerations

The application of rDNA technology raises several ethical and safety issues:

- **Gene Flow**: Risk of genetically modified genes spreading to non-target species.
- **Biodiversity**: Potential impact on natural ecosystems and species diversity.
- **Human Health**: Concerns about allergenicity and long-term health effects of GMOs.

Regulatory agencies have established guidelines and assessment protocols to evaluate and manage these risks effectively.

5. Genetically Modified Organisms: Gene Knockouts

Introduction

Gene knockout technology is a pivotal tool in molecular biology and genetics, allowing researchers to study the function of specific genes by observing the effects of their inactivation. This technique involves the deliberate inactivation or "knocking out" of a gene to assess its role in the organism's physiology and development.

Definition

A **gene knockout** refers to a genetic technique in which an organism is engineered to carry genes that have been made inoperative. This is achieved by disrupting or replacing the target gene with an artificial piece of DNA, rendering it nonfunctional. The

resulting organism is termed a **knockout organism**, and when applied to mice, it is referred to as a **knockout mouse**.

Methods of Gene Knockout

- **Homologous Recombination in Embryonic Stem Cells**: This traditional method involves introducing a DNA construct into embryonic stem (ES) cells, where it replaces or disrupts the target gene through homologous recombination. The modified ES cells are then injected into blastocysts to develop chimeric mice, which can pass the knockout gene to their offspring.
- **CRISPR/Cas9 Technology**: The **Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)** and **CRISPR-associated protein 9 (Cas9)** system has revolutionized gene editing by allowing precise, efficient, and cost-effective gene knockouts. By designing specific guide RNAs, researchers can direct the Cas9 enzyme to introduce double-strand breaks at desired genomic locations, leading to gene disruption during the repair process.

Applications of Gene Knockout

- **Functional Genomics**: Understanding gene function by observing phenotypic changes resulting from gene inactivation.
- **Disease Modeling**: Creating animal models that mimic human diseases for research and drug development.
- **Drug Target Validation**: Identifying and validating potential targets for therapeutic intervention.
- **Agricultural Improvements**: Developing crops with desirable traits by knocking out genes responsible for unfavorable characteristics.

Challenges and Considerations

- **Compensatory Mechanisms**: Other genes may compensate for the knocked-out gene, masking its function.
- **Lethality**: Knocking out essential genes can result in embryonic lethality, complicating studies.
- **Off-Target Effects**: Especially with CRISPR/Cas9, unintended edits can occur, necessitating thorough validation.

6. Mouse Disease Models

Introduction

Mouse models are indispensable in biomedical research due to their genetic, biological, and behavioral similarities to humans. Genetically modified mice, including knockout and transgenic models, have been instrumental in elucidating disease mechanisms and testing therapeutic interventions.

Why Mice?

- **Genetic Similarity**: Mice share approximately 95% of their genes with humans.
- **Short Reproductive Cycle**: Rapid breeding allows for the study of multiple generations in a short time.
- **Ease of Genetic Manipulation**: Techniques like CRISPR/Cas9 facilitate precise genetic modifications.

Types of Mouse Disease Models

- **Knockout Models**: Mice with specific genes inactivated to study gene function and disease mechanisms.
- **Transgenic Models**: Mice engineered to carry additional genes, often human genes, to study gene expression and function.
- **Humanized Mice**: Mice that carry human genes, cells, tissues, or organs, providing more accurate models for human diseases.

Applications in Disease Research

- **Cancer**: Mouse models have been developed to study various cancers, including breast, colon, and prostate cancers, aiding in understanding tumor biology and testing treatments.
- **Neurodegenerative Diseases**: Models for Alzheimer's, Parkinson's, and Huntington's diseases help in studying disease progression and potential therapies.
- **Metabolic Disorders**: Mice are used to study diabetes, obesity, and other metabolic syndromes, providing insights into disease mechanisms and treatment options.
- **Infectious Diseases**: Mouse models have been crucial in studying diseases like HIV/AIDS, tuberculosis, and COVID-19, contributing to vaccine and therapeutic development.

Limitations

- **Species Differences**: Not all human diseases can be accurately replicated in mice due to physiological differences.
- **Ethical Considerations**: The use of animals in research raises ethical concerns, necessitating strict regulatory oversight.

7. Gene Silencing

Introduction

Gene silencing refers to the regulation of gene expression in a cell to prevent the expression of a certain gene. This process can occur at various stages, including transcriptional and post-transcriptional levels, and is essential for controlling gene expression, defending against viruses, and maintaining genome stability.

Mechanisms of Gene Silencing

1. **Transcriptional Gene Silencing (TGS):** Involves modifications to DNA or histones that prevent transcription. Key mechanisms include:
    - **DNA Methylation:** Addition of methyl groups to DNA, typically at cytosine bases, leading to gene repression.
    - **Histone Modification:** Chemical changes to histone proteins, such as methylation or deacetylation, resulting in a more compact chromatin structure and reduced gene expression.
2. **Post-Transcriptional Gene Silencing (PTGS):** Occurs after transcription and involves the degradation of mRNA or inhibition of translation. Key pathways include:
    - **RNA Interference (RNAi):** A process where small RNA molecules, such as small interfering RNA (siRNA) and microRNA (miRNA), guide the degradation of target mRNA or inhibit its translation.
    - **Antisense Oligonucleotides (ASOs):** Short, synthetic strands of nucleotides that bind to complementary mRNA sequences, blocking translation or promoting degradation.

RNA Interference (RNAi): The Central Mechanism of PTGS

RNA interference (RNAi) is a conserved biological process where small RNA molecules inhibit gene expression by neutralizing targeted mRNA molecules. The two primary small RNAs involved are:

- **Small Interfering RNAs (siRNAs):** Typically, exogenous double-stranded RNAs processed within the cell.
- **MicroRNAs (miRNAs):** Endogenously encoded RNAs that form hairpin structures and are processed into mature miRNAs.

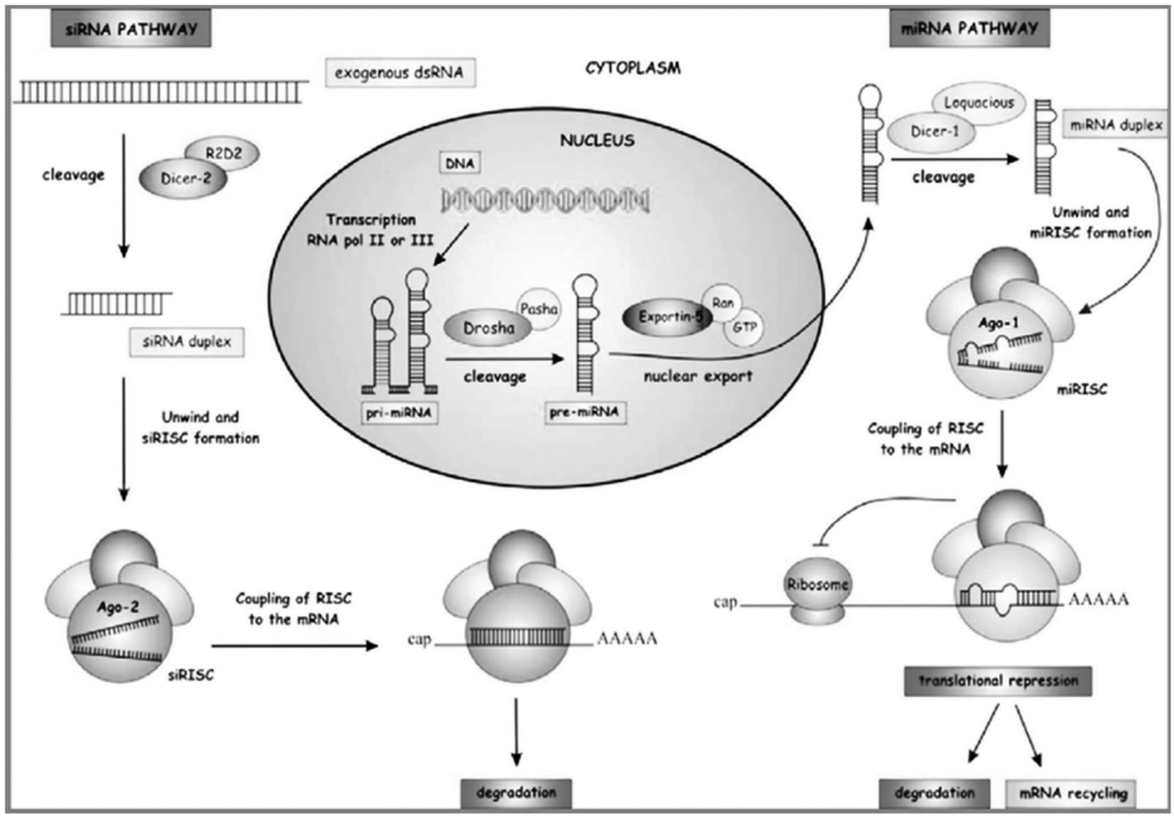


Figure 9.1

Step-by-Step Mechanism of RNAi-Mediated Gene Silencing

A. siRNA Pathway

1. **Initiation:**
    - Introduction of long double-stranded RNA (dsRNA) into the cytoplasm, either exogenously or through endogenous transcription.
2. **Dicer Processing:**
    - The enzyme **Dicer**, an RNase III family endonuclease, cleaves the dsRNA into 21–23 nucleotide siRNA duplexes with 2-nucleotide 3' overhangs.

3. **RISC Assembly:**
    - The siRNA duplex is incorporated into the **RNA-Induced Silencing Complex (RISC)**.
    - Within RISC, the duplex is unwound; the **passenger strand** is degraded, and the **guide strand** remains bound.
4. **Target Recognition:**
    - The guide strand directs RISC to complementary sequences on target mRNA molecules through base pairing.
5. **mRNA Cleavage:**
    - The **Argonaute (Ago)** protein within RISC cleaves the target mRNA at the site of complementarity, leading to its degradation and preventing translation.

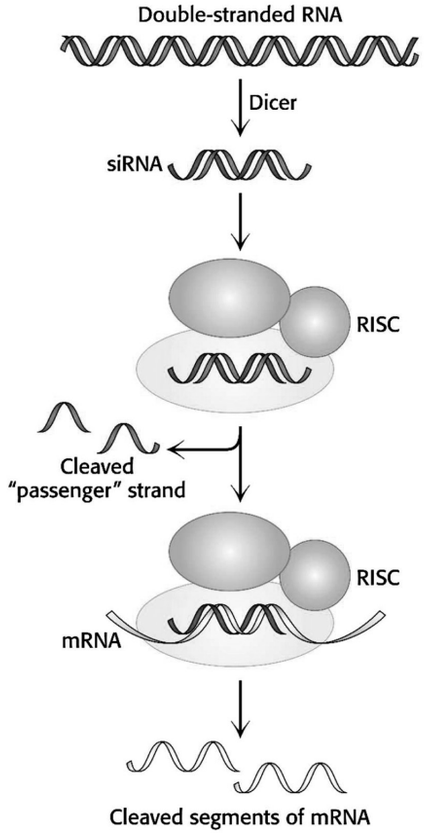


Figure 9.2

Fig. RNA interference mechanism. A double-stranded RNA molecule is cleaved into 21-bp fragments by the enzyme Dicer to produce siRNAs. These siRNAs are incorporated into the RNA-induced silence complex (RISC), where the single-stranded RNAs guide the cleavage of mRNAs that contain complementary sequences.

B. miRNA Pathway

1. **Transcription:**
    - miRNA genes are transcribed by RNA polymerase II into primary miRNA (pri-miRNA) transcripts.
2. **Nuclear Processing:**
    - The **Drosha-DGCR8** complex processes pri-miRNA into precursor miRNA (pre-miRNA), a ~70 nucleotide hairpin structure.
3. **Nuclear Export:**
    - Pre-miRNA is exported to the cytoplasm via **Exportin-5**.
4. **Cytoplasmic Processing:**
    - **Dicer** further processes pre-miRNA into a ~2′ nucleotide miRNA duplex.
5. **RISC Loading:**
    - Similar to siRNA, the miRNA duplex is loaded into RISC, the passenger strand is discarded, and the guide strand remains.
6. **Target Recognition and Silencing:**
    - The guide strand directs RISC to partially complementary sequences in target mRNAs, leading to translational repression or mRNA destabilization.

Applications of Gene Silencing

- **Therapeutic Applications:** Gene silencing techniques are being explored for treating various diseases, including cancer, viral infections, and genetic disorders, by downregulating harmful gene expressions.
- **Functional Genomics:** Researchers use gene silencing to study gene function by observing the effects of reduced gene expression.
- **Agricultural Biotechnology:** Gene silencing is employed to develop crops with desirable traits, such as pest resistance or improved nutritional content.

Challenges and Considerations

- **Off-Target Effects:** Ensuring specificity to avoid unintended gene silencing.
- **Delivery Mechanisms:** Developing efficient and safe methods to deliver gene silencing agents to target cells.

- **Ethical and Regulatory Aspects:** Addressing concerns related to gene manipulation and ensuring compliance with regulatory standards.

8. Gene Therapy: Somatic and Germline Therapy

Introduction

Gene therapy involves the introduction, removal, or alteration of genetic material within a person's cells to treat or prevent disease. It holds the potential to address the root causes of genetic disorders by correcting defective genes.

Types of Gene Therapy

1. **Somatic Gene Therapy:**
    - **Definition:** Targets non-reproductive (somatic) cells, meaning changes are not heritable.
    - **Applications:** Used to treat diseases such as severe combined immunodeficiency (SCID), certain cancers, and hemophilia.
    - **Methods:** Commonly employs viral vectors to deliver therapeutic genes to affected tissues.
2. **Germline Gene Therapy:**
    - **Definition:** Involves modifications to reproductive (germ) cells, making changes heritable.
    - **Applications:** Potential to prevent inherited diseases by correcting mutations in embryos or gametes.
    - **Ethical Considerations:** Raises significant ethical and safety concerns, leading to widespread regulatory restrictions.

Delivery Methods

- **Viral Vectors:** Utilize modified viruses to deliver genetic material into cells.
- **Non-Viral Methods:** Include techniques like electroporation, liposomes, and nanoparticles to introduce genes without using viruses.

Challenges and Ethical Considerations

- **Safety:** Risks of immune reactions, insertional mutagenesis, and unintended effects.

- **Efficacy:** Ensuring long-term expression and functionality of the introduced gene.
- **Ethical Issues:** Concerns about germline modifications, consent, and potential for non-therapeutic enhancements.

Regulatory Landscape

Gene therapy is subject to rigorous regulatory oversight to ensure safety and ethical compliance. Agencies like the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have established frameworks for the evaluation and approval of gene therapy products.

Feature	Somatic Gene Therapy	Germline Gene Therapy
Definition	Introduction or alteration of genes in **non-reproductive (somatic) cells.**	Introduction or alteration of genes in **reproductive (germ) cells** (sperm, egg, embryo).
Heredity	Changes are **not passed on** to the next generation.	Changes are **heritable** and passed on to future generations.
Target Cells	Targets **body cells** (e.g., bone marrow cells, liver cells, lung cells).	Targets **gametes** (sperm or eggs) or early-stage embryos.
Purpose	To **treat or cure** a genetic disorder in an individual.	To **prevent** a genetic disorder from being transmitted to offspring.
Current Status	**Widely researched** and **clinically applied** in some countries.	**Largely prohibited** for clinical use in humans due to ethical concerns.
Ethical Issues	Fewer ethical concerns; generally considered acceptable if safe and effective.	Major ethical concerns related to human enhancement, eugenics, and consent of future generations.
Risk Factors	Limited to the treated individual; lower overall risk for the population.	Risks may affect the entire lineage; unknown long-term effects.

Feature	Somatic Gene Therapy	Germline Gene Therapy
Examples of Application	Treatment of **cancer, cystic fibrosis, severe combined immunodeficiency (SCID).**	Theoretical prevention of inherited diseases like **Tay-Sachs disease** or **Huntington's disease.**
Regulatory Approval	Approved therapies exist (e.g., **Luxturna** for inherited retinal disease).	Not approved for clinical use in most countries; permitted only in experimental settings under strict regulations.
Ethical Bodies/ Regulations	Regulated by bodies like **FDA (U.S.), EMA (Europe).**	Strictly monitored and restricted by international guidelines (e.g., **Oviedo Convention**).

Probable Questions:

1. Describe the different methods used for gene transfer into animal cells. Compare their advantages and limitations.
2. Explain the process of creating transgenic animals through embryonic stem cell-mediated gene transfer and its applications.
3. Discuss the mechanism of RNA interference (RNAi) in gene silencing, highlighting the roles of siRNA and miRNA.
4. Evaluate the applications of recombinant DNA technology in medicine, agriculture, industry, and environmental management.
5. Differentiate between somatic gene therapy and germline gene therapy with examples, and critically assess the associated ethical concerns.
6. Illustrate the use of CRISPR/Cas9 technology in genome editing and its impact on functional genomics and disease modeling.
7. Define gene knockout and describe the methods used to generate knockout mice. What are the applications and limitations of this approach?
8. Discuss the role of transgenic animal models in biomedical research, citing specific examples of disease models and their significance.
9. What are the biosafety and regulatory measures associated with the use of genetically modified organisms (GMOs)? Why are these important?

10. Describe the step-by-step mechanism of post-transcriptional gene silencing via RNA interference. How is this pathway exploited therapeutically?

Suggested Readings:

1. Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nature Reviews Genetics*, 6(6), 507–512. DOI: 10.1038/nrg1619
2. Gonçalves GAR, Paiva RMA. Gene therapy: advances, challenges and perspectives. Einstein (Sao Paulo). 2017 Jul-Sep;15(3):369-375. doi: 10.1590/S1679-45082017RB4024. PMID: 29091160; PMCID: PMC5823056.
3. Clark, A. J. (1998). The mammary gland as a bioreactor: Expression, secretion and processing of recombinant proteins. Journal of Applied Animal Research, 32(1), 1-17.
4. Cong, L. et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823
5. Gordon, J. W., & Ruddle, F. H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. Science, 214(4526), 1244-1246. DOI: 10.1126/science.6272397
6. Houdebine, L. M. (2000). Transgenic animal bioreactors. Transgenic Research, 11(6), 577–590.
7. Rosenthal, N., & Brown, S. (2007). The mouse ascending: perspectives for human-disease models. Nature Cell Biology, 9(9), 993–999. DOI: 10.1038/ncb437
8. Capecchi, M. R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nature Reviews Genetics, 6(6), 507–512. DOI: 10.1038/nrg1619
9. Perlman, R. L. (2016). Mouse models of human disease: An evolutionary perspective. Evolution, Medicine, and Public Health, 2016(1), 170–176. DOI: 10.1093/emph/eow014
10. Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157(6), 1262–1278.
11. Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), 1262–1278. DOI: 10.1016/j.cell.2014.05.010

12. Muramatsu, T., Nakamura, A., & Park, H. M. (1998). Basics and Applications of Gene Transfer in Animal Cells and Tissues for Production and Therapeutics. *Journal of Mammalian Ova Research*, 15(1), 1-9.

13. Niemann, H., & Kues, W. A. (2007). Transgenic farm animals: An update. Reproduction, Fertility and Development, 19(6), 762–770. DOI: 10.1071/RD07022

14. Perlman, R. L. (2016). Mouse models of human disease: An evolutionary perspective. *Evolution, Medicine, and Public Health*, 2016(1), 170–176. DOI: 10.1093/emph/eow014

15. Recillas-Targa F. Gene transfer and expression in mammalian cell lines and transgenic animals. Methods Mol Biol. 2004;267:417-33. doi: 10.1385/1-59259-774-2:417. PMID: 15269440.

16. Friedmann, T., & Roblin, R. (1972). Gene therapy for human genetic disease? Science, 175(4025), 949–955. DOI: 10.1126/science.175.4025.949

17. Rosenthal, N., & Brown, S. (2007). The mouse ascending: perspectives for human-disease models. *Nature Cell Biology*, 9(9), 993–999. DOI: 10.1038/ncb437

18. Twyman, R. M. (1998). *Gene Transfer to Animal Cells*. Springer Science & Business Media. ISBN: 9781859962046.

Unit-X:
Genome manipulation-CRISPR-Cas9 System

1. Introduction and Basic Concepts

The CRISPR-Cas9 system (***C***lustered ***R***egularly ***I***nterspaced ***S***hort ***P***alindromic ***R***epeats - **CRISPR**; CRISPR-associated protein 9 - Cas9) has revolutionized molecular biology by enabling precise and programmable genome editing across organisms. Originally discovered as a natural bacterial immune system, CRISPR provides adaptive immunity to prokaryotes against viruses and plasmids.

In biotechnology, CRISPR-Cas9 has been engineered to introduce site-specific double-stranded breaks (DSBs) in DNA, facilitating targeted modifications such as gene knockouts, insertions, point mutations, and gene corrections.

2. Fundamental Principles of CRISPR-Cas9

2.1. Natural Biological Role

- CRISPR arrays and Cas genes are found in ~50% of sequenced bacterial genomes.
- When a bacterium survives a phage infection, it captures fragments of phage DNA, inserting them into its genome between palindromic repeats.
- Upon re-infection, CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) guide the Cas9 enzyme to cleave the invading phage DNA by base pairing with the complementary sequence.

2.2. Engineered System for Genome Editing

- Minimal components:
    - A single-guide RNA (sgRNA) engineered from crRNA and tracrRNA fusion.
    - A Cas9 protein engineered for specific DNA cleavage.
- Together, sgRNA and Cas9 form a complex that recognizes and cleaves DNA at a desired location in a genome.

3. Detailed Mechanism of CRISPR-Cas9 System

The functioning of the CRISPR-Cas9 system for genome editing involves five major steps:

3.1. Guide RNA (gRNA) Design

- gRNA contains:
    - A scaffold sequence that binds Cas9.
    - A targeting sequence (~20 nucleotides) complementary to the genomic site.

3.2. Cas9 and gRNA Complex Formation

- Cas9 binds to the gRNA scaffold, forming a ribonucleoprotein complex capable of scanning DNA for matching sequences.

3.3. Target DNA Recognition

- The complex binds DNA at a site adjacent to a Protospacer Adjacent Motif (PAM), typically NGG for *Streptococcus pyogenes* Cas9.

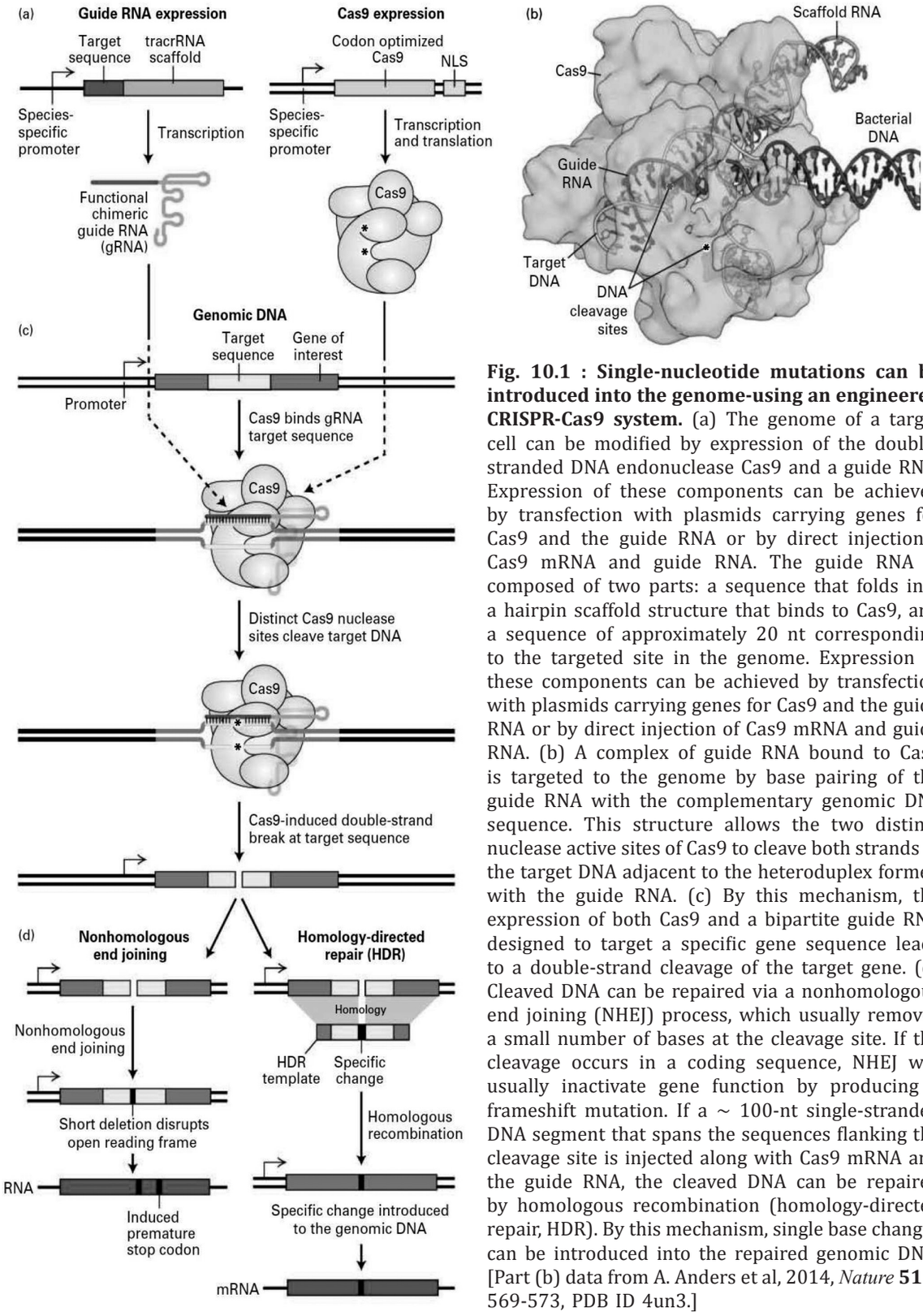


Fig. 10.1 : Single-nucleotide mutations can be introduced into the genome-using an engineered CRISPR-Cas9 system. (a) The genome of a target cell can be modified by expression of the double-stranded DNA endonuclease Cas9 and a guide RNA. Expression of these components can be achieved by transfection with plasmids carrying genes for Cas9 and the guide RNA or by direct injectionof Cas9 mRNA and guide RNA. The guide RNA is composed of two parts: a sequence that folds into a hairpin scaffold structure that binds to Cas9, and a sequence of approximately 20 nt corresponding to the targeted site in the genome. Expression of these components can be achieved by transfection with plasmids carrying genes for Cas9 and the guide RNA or by direct injection of Cas9 mRNA and guide RNA. (b) A complex of guide RNA bound to Cas9 is targeted to the genome by base pairing of the guide RNA with the complementary genomic DNa sequence. This structure allows the two distinct nuclease active sites of Cas9 to cleave both strands of the target DNA adjacent to the heteroduplex formed with the guide RNA. (c) By this mechanism, the expression of both Cas9 and a bipartite guide RNA designed to target a specific gene sequence leads to a double-strand cleavage of the target gene. (d) Cleaved DNA can be repaired via a nonhomologous end joining (NHEJ) process, which usually removes a small number of bases at the cleavage site. If the cleavage occurs in a coding sequence, NHEJ will usually inactivate gene function by producing a frameshift mutation. If a ~ 100-nt single-stranded DNA segment that spans the sequences flanking the cleavage site is injected along with Cas9 mRNA and the guide RNA, the cleaved DNA can be repaired by homologous recombination (homology-directed repair, HDR). By this mechanism, single base changes can be introduced into the repaired genomic DNA. [Part (b) data from A. Anders et al, 2014, *Nature* **513**-569-573, PDB ID 4un3.]

3.4. DNA Cleavage

- Cas9 has two endonuclease domains:
  - HNH domain: Cleaves the DNA strand complementary to the gRNA.
  - RuvC domain: Cleaves the non-complementary DNA strand.
- A double-stranded break (DSB) is introduced precisely at the target site.

3.5. DNA Repair Pathways

Following cleavage, two repair pathways are engaged:

- Non-Homologous End Joining (NHEJ):
  - Error-prone.
  - Leads to insertions or deletions (indels) causing frameshift mutations and gene knockout.
- Homology-Directed Repair (HDR):
  - Requires a donor template.
  - Allows precise gene insertion or correction.

4. Applications of CRISPR-Cas9

CRISPR-Cas9 has broad applications across biological, medical, agricultural, and industrial domains:

Application Area	Details
Gene Knockout	Disrupting gene function by inducing frameshift mutations through NHEJ.
Gene Correction	Correcting point mutations via HDR (e.g., correcting Crygc mutation in mouse zygotes).
Transgenic Animal Models	Creating precise genetic models (e.g., disease models in mice, zebrafish).
Functional Genomics	High-through put CRISPR screens to study gene functions systematically.
Therapeutic Genome Editing	Trials for treating diseases like β-thalassemia, sickle cell anemia, and hereditary blindness.
Agricultural Engineering	Developing crops with improved traits (e.g., drought resistance, higher yields).

CRISPRi and CRISPRa: Overview

CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) are advanced gene regulation techniques derived from the CRISPR-Cas9 system. Unlike traditional CRISPR-Cas9, which introduces double-stranded breaks in DNA, CRISPRi and CRISPRa utilize a catalytically inactive Cas9 protein (dCas9) to modulate gene expression without altering the DNA sequence.

- CRISPRi: Represses gene expression by blocking transcription initiation or elongation.
- CRISPRa: Enhances gene expression by recruiting transcriptional activators to gene promoters.

These tools allow for precise, reversible, and programmable control of gene expression, facilitating functional genomics studies and therapeutic applications.

Mechanisms of Action

CRISPRi: Gene Repression

1. dCas9 Binding: A single-guide RNA (sgRNA) directs dCas9 to a specific DNA sequence near the transcription start site (TSS) of the target gene.
2. Transcriptional Blockade: dCas9 physically obstructs RNA polymerase binding or progression, effectively silencing gene expression.
3. Enhanced Repression: Fusion of dCas9 with repressor domains, such as KRAB (Krüppel-associated box), can recruit chromatin-modifying enzymes to establish a repressive chromatin state, leading to more robust and long-term gene silencing.

CRISPRa: Gene Activation

1. dCas9 Targeting: sgRNA guides dCas9 to a region upstream of the TSS of the gene of interest.
2. Recruitment of Activators: dCas9 is fused to transcriptional activator domains (e.g., VP64, p65, Rta) or engineered systems like the SunTag or SAM (Synergistic Activation Mediator) to recruit the transcriptional machinery.
3. Transcriptional Enhancement: The recruited activators facilitate the assembly of the transcriptional complex, increasing gene expression from the endogenous locus.

5. CRISPR-Cas9 for Genome Manipulation

CRISPR-Cas9 has become the gold standard tool for genome manipulation because:

- It is programmable: Just by changing the gRNA sequence, a different gene can be targeted.
- It is efficient: Genome editing occurs at high frequencies.
- It is simple: Two components (Cas9 + gRNA) are sufficient.
- It is versatile: Works in almost all model organisms (mice, zebrafish, fruit flies, humans).

Key genome manipulations:

- Knockouts (via NHEJ).
- Knock-ins, point mutations, gene insertions (via HDR).
- Multiplexed Editing (multiple genes edited simultaneously).

6. Limitations and Challenges

Despite its advantages, CRISPR-Cas9 is not without limitations:

Limitation	Description
Off-target Effects	Cleavage at unintended sites due to imperfect gRNA binding.
Delivery Efficiency	Difficulty in delivering CRISPR components into some cells/ tissues.
Immune Responses	Cas9 derived from bacteria could trigger immune reactions in humans.
HDR Efficiency	Homology-Directed Repair is less efficient than NHEJ, especially in non-dividing cells.
Ethical Concerns	Human germline editing raises serious ethical issues globally.

7. Future Prospects

Research is underway to address current challenges:

- Development of high-fidelity Cas9 variants (e.g., eSpCas9, SpCas9-HF1) to minimize off-target effects.
- Base editing technologies that modify single nucleotides without DSBs.
- Prime editing allowing versatile edits without templates.

- In vivo gene therapies targeting diseases directly in patients.
- CRISPRi (CRISPR interference) and CRISPRa (CRISPR activation) for gene repression and activation without DNA cleavage.

Probable Questions:

1. Explain the natural biological role of the CRISPR-Cas9 system in prokaryotes. How has this system been adapted for genome editing in biotechnology?
2. Describe the detailed mechanism of the CRISPR-Cas9 system in genome editing, outlining each of the five key steps involved.
3. Differentiate between Non-Homologous End Joining (NHEJ) and Homology-Directed Repair (HDR) in the context of CRISPR-mediated genome editing. Provide examples of their applications.
4. Discuss the structure and functional role of single-guide RNA (sgRNA) in the CRISPR-Cas9 system. How does it guide the Cas9 protein to specific genomic locations?
5. Illustrate the various applications of the CRISPR-Cas9 system in medical, agricultural, and industrial biotechnology with relevant examples.
6. What are CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa)? Compare their mechanisms and potential applications in gene expression regulation.
7. List and explain the major limitations and challenges associated with the CRISPR-Cas9 system. How are researchers attempting to overcome these limitations?
8. Discuss the ethical concerns surrounding the use of CRISPR-Cas9 technology, particularly in the context of human germline editing.
9. Evaluate the future prospects of genome manipulation using CRISPR-Cas9, including recent advancements like base editing and prime editing.
10. Describe the mechanisms by which dCas9 is used in CRISPRi and CRISPRa systems for gene silencing and activation. How do fusion domains enhance their effectiveness?

Suggested Readings:

1. Hsu PD, Lander ES, Zhang F. 2014. "Development and applications of CRISPR-Cas9 for genome engineering." *Cell*, 157(6): 1262–1278. DOI: 10.1016/j.cell.2014.05.010.

2. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.” *Nature*, 533(7603): 420–424. DOI: 10.1038/nature17946.
3. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR. 2019. “Search-and-replace genome editing without double-strand breaks or donor DNA.” *Nature*, 576(7785): 149–157. DOI: 10.1038/s41586-019-1711-4.

Unit-XI
Introduction to Microscopy

1. What is a Microscope?

A **microscope** is an instrument used to observe objects that are too small to be seen clearly with the naked eye. It allows scientists and students to explore the structure of cells, microorganisms, tissues, and even molecules.

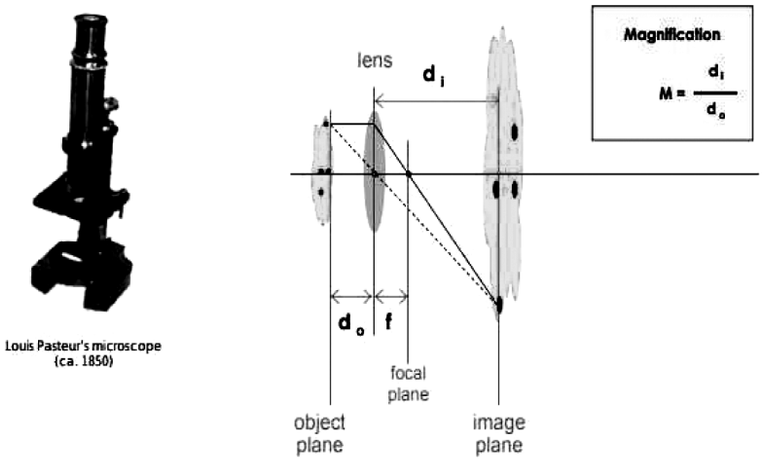


Fig. 11.1: Basic Plan of a Microscope

Microscopy is the science of using microscopes to view samples and objects.

2. Why Do We Use Microscopes?

- **To study the structure of cells and tissues**.
- **To diagnose diseases** by observing pathogens or abnormal cells.
- **To examine microorganisms**, like bacteria and viruses.
- **To conduct scientific research** in biology, materials science, and medicine.
- **To understand complex structures** at microscopic and even nanoscopic levels.

3. Important Terms in Microscopy

Magnification

- This is how much larger an object appears when viewed through a microscope.
- It is the **ratio** of the image size to the actual size of the object.

- **Example**: If an object is magnified 100 times (100x), it appears 100 times larger than its real size.
- The **Numerical Aperture (NA)** is a key parameter in microscopy that determines the resolving power and light-gathering ability of a microscope objective lens.
- **Definition:** Numerical Aperture is defined by the equation: $NA=n.\sin(\theta)$

  Where:

  n = Refractive index of the medium between the lens and the specimen (e.g., air, water, oil)

  θ = Half of the angular aperture of the objective lens (i.e., the maximum angle of light that can enter the lens)

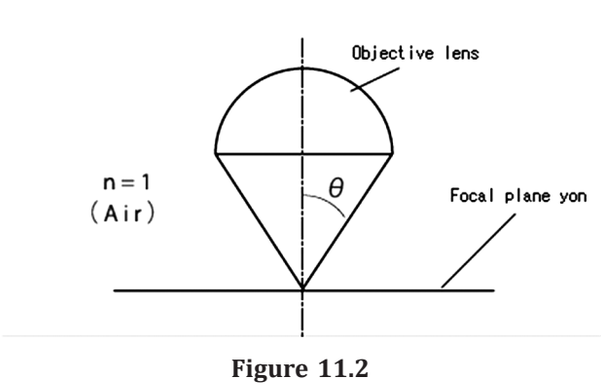


Figure 11.2

Importance of Numerical Aperture:

- **Resolution:** Higher NA allows for greater resolution, meaning finer details can be distinguished. The resolving power of a microscope is inversely proportional to NA.
- **Brightness:** Objectives with higher NA collect more light, resulting in brighter images.
- **Depth of Field:** Higher NA decreases the depth of field (focuses on a thinner optical slice), which is useful for high-resolution imaging of flat specimens.

Types:

- **Dry objectives:** Typically have NA values up to ~0.95 (using air as the medium, $n \approx 1$).
- **Oil immersion objectives:** Can have NA values greater than 1 (using oil with $n \approx 1.5$), allowing for enhanced resolution.

Resolving Power (Resolution)

- The ability of a microscope to **distinguish two points as separate**.
- Higher resolution means finer detail can be seen.
- It's different from magnification – you can magnify a blurry image, but it won't make it clearer without good resolution.

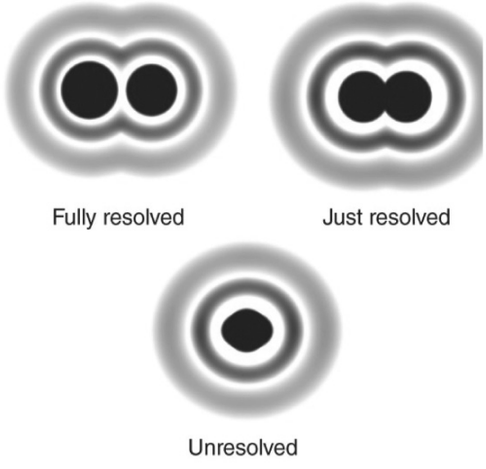


Figure 11.3

Limit of Resolution

- The **smallest distance** between two points that can still be distinguished as separate.
- Determined by the **wavelength of light** (for light microscopes) or **electrons** (for electron microscopes).

4. How is Magnification Determined?

For **compound light microscopes**:

Total Magnification = Objective Lens Magnification × Eyepiece Lens Magnification

Example:

- Eyepiece = 10x
- Objective lens = 40x
- Total magnification = 10 × 40 = **400x**

5. Types of Microscopes / Microscopy

A. Light Microscopy

Uses visible light to illuminate specimens.

- **Bright-field Microscopy**: Most common, uses transmitted light.
- **Phase-contrast Microscopy**: Enhances contrast in transparent specimens.
- **Fluorescence Microscopy**: Uses fluorescent dyes and UV light.

B. Electron Microscopy

Uses beams of electrons for much higher resolution.

- **Transmission Electron Microscope (TEM)**:
  - Passes electrons through thin slices.
  - Shows detailed internal structure.
  - Very high resolution.
- **Scanning Electron Microscope (SEM)**:
  - Scans the surface with electrons.
  - Provides 3D images of surfaces.

C. Other Advanced Types

- **Confocal Microscopy**: Uses lasers for sharp, 3D images of thick specimens.
- **Scanning Probe Microscopy (like AFM)**: Detects surface properties at the atomic level.
- **Super resolution Microscopy**

Bright-field Microscopy

Definition

Bright-field microscopy is the **simplest and most commonly used type of light microscopy**. It works by passing visible light through a specimen, where contrast is created by the **absorption of light** in dense areas of the sample.

Principle

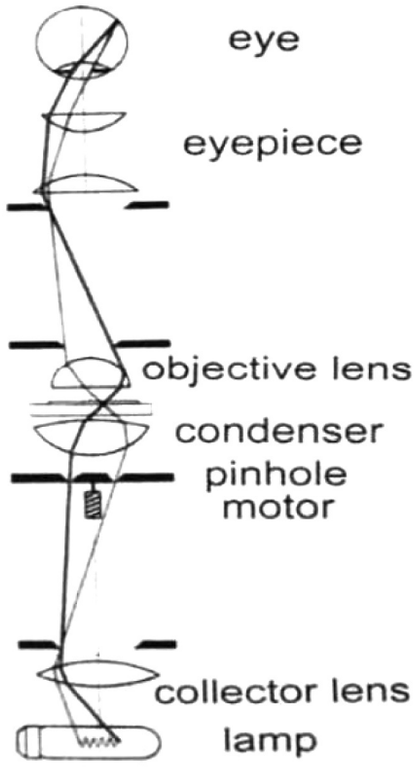


Figure 11.4

In bright-field microscopy:

- A **light source** illuminates the specimen from below.
- Light passes through the specimen and is gathered by the **objective lens**.
- The specimen appears **dark or colored against a bright background**, depending on its density or staining.

Key Components

- **Light source** (usually a halogen lamp or LED)
- **Condenser lens** (focuses light on the specimen)
- **Objective lenses** (magnify the image)
- **Eyepiece lens** (further magnifies and allows viewing)

Sample Preparation

- Most specimens are **stained** to enhance contrast.
- **Thin and transparent** specimens are ideal.
- Common stains: **methylene blue**, **crystal violet**, **Gram stain**.

Advantages

- Simple and easy to use.
- Inexpensive and widely available.
- Good for stained specimens (e.g., bacteria, tissue sections).
- Can be used for **live cells**, though with limited contrast.

Limitations

- **Low contrast** for unstained, transparent samples.
- **Limited resolution** compared to advanced microscopy.
- Cannot provide detailed views of internal ultrastructure.

Applications

- **Routine laboratory work**
- **Teaching and education**
- **Microbiology** (e.g., Gram staining of bacteria)
- **Histology** (studying tissue sections)
- **Cell biology** (observing basic cell structure)

Dark-field Microscopy

Definition

Dark-field microscopy is a **light microscopy technique** that enhances the contrast of **unstained, transparent specimens** by making them appear **bright against a dark background**.

Principle

- Instead of directly illuminating the specimen, **light is directed at an angle**.
- Only **scattered light** from the specimen enters the objective lens.
- The background remains **dark**, and only the parts of the specimen that **scatter light** appear bright.

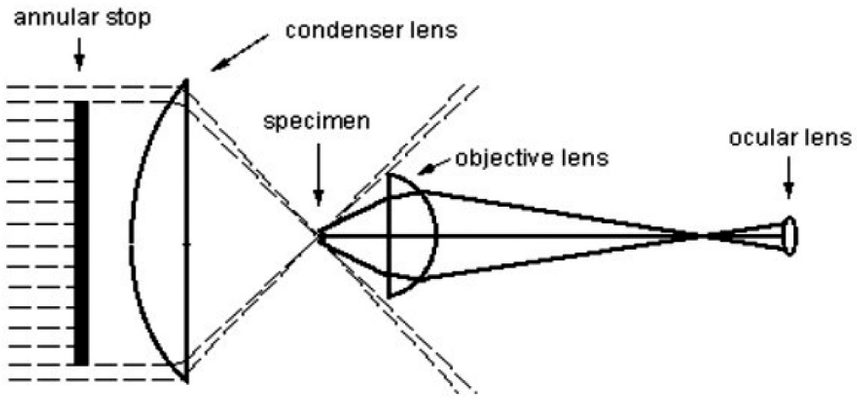


Figure 11.5

Key Idea: Light that passes directly through the specimen does **not** enter the objective—only scattered light does.

Key Components

- **Dark-field condenser**: blocks central light and directs light at an oblique angle.
- **Standard objective and eyepiece lenses**.
- No staining is needed, preserving **live and delicate samples**.

Advantages

- Excellent for **observing live, unstained organisms**.
- Enhances visibility of **thin or transparent specimens** (e.g., flagella, spirochetes).
- Non-destructive: samples stay alive and unaltered.

Limitations

- Lower resolution compared to other advanced methods.
- Bright objects can **easily scatter too much light**, causing glare.
- Requires **precise alignment** of optical components.
- Not ideal for thick specimens (causes blurry images).

Applications

- **Microbiology**: to observe **bacteria with poor staining properties** like *Treponema pallidum* (causes syphilis).
- **Aquatic biology**: studying live **protozoa and plankton**.
- Observing **motility and shape** of live microorganisms.

Phase-Contrast Microscopy

Definition

Phase-contrast microscopy is a powerful optical technique that allows for the visualization of **transparent and colorless specimens—without the need for staining**. It converts small differences in the **phase of light waves** passing through different parts of a specimen into **differences in light intensity (contrast)**.

Why is it Important?

- Many biological specimens like **living cells**, organelles, and microorganisms are **transparent and nearly invisible** under bright-field microscopy.
- **Staining can kill or alter live cells**, so phase-contrast provides a **non-destructive way** to observe them in their natural state.
- It allows researchers to study **live-cell dynamics**, **cell division**, **motility**, and **internal organelle movement** in real time.

Principle

- Light passing through a specimen experiences **changes in phase** due to differences in thickness or refractive index.
- These phase shifts are **not visible** to the human eye.
- The microscope uses a **phase plate and annular diaphragm** to transform these invisible phase changes into visible differences in brightness and contrast.

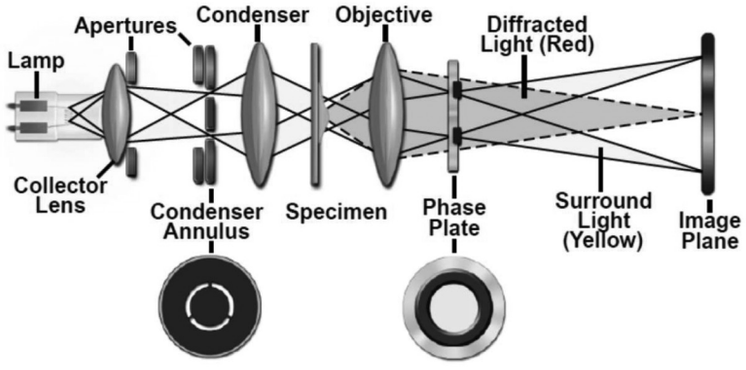


Figure 11.6

Advantages

- Ideal for viewing **live, unstained specimens**.
- Enhances contrast in **internal structures of cells** (e.g., nuclei, vacuoles).
- No need for dyes or stains, preserving cell viability.

Limitations

- Can produce **halo or shading artifacts** around specimens.
- Not suitable for **thick samples** (can result in blurry images).

Applications

- **Cell biology** – observing live cells and organelles.
- **Microbiology** – viewing bacteria, protozoa, and plankton.
- **Developmental biology** – studying embryos and tissues in real time.

DIC Microscopy (Differential Interference Contrast)

(Also called Nomarski Interference Microscopy)

DIC is a **refined version of phase-contrast microscopy** that also uses differences in refractive index to create contrast but does so using **polarized light** and optical prisms.

- DIC produces **high-contrast, 3D-like images** of transparent specimens.
- It eliminates the **halo effect** often seen in phase contrast.
- The image has **depth and shadowing**, making it look more **textured and realistic**.

☑ Best For: Observing fine details in living cells, membranes, and organelles with a more natural, relief-like appearance.

Feature	Bright-field Microscope	Phase Contrast Microscope
Contrast Generation	Based on **light absorption**; contrast comes from stains or natural pigments	Based on **differences in refractive index**; converts phase shifts into contrast
Use of Stains	Often requires **staining** to see transparent specimens	**No staining needed**; can view live, transparent cells clearly
Image Appearance	Specimens appear **dark/colored on a bright background**	Specimens appear with **enhanced contrast and internal details**
Live Cell Observation	Not ideal; staining may **kill cells**	Ideal for **live cell imaging** without altering or killing them
Artifacts	Fewer artifacts when properly stained	Can produce **halo or shading artifacts** around objects
Special Components	Standard condenser and objective lenses	Requires **phase plate and annular diaphragm**

Best Suited For General observation of **stained tissue sections, slides** Observing **live, unstained cells, organelles, and microorganisms**

Feature	Bright-field Microscope	Phase Contrast Microscope
Contrast Generation	Based on **light absorption;** contrast comes from stains or natural pigments	Based on **differences in refractive index;** converts phase shifts into contrast
Use of Stains	Often requirs **staining** to see transparent specimens	**No staining need;** can view live, transparent cells clearly
Image Appearance	Specimens appear **dark/colored on a bright background**	Specimens appear with **enhanced contrast and internal details**
Live Cell Observation	Not ideal; staining may **kill cells**	Ideal for **live cell imaging** without altering or killing them
Artifacts	Fewer artifacts when properly stained	Can produce **halo or shading artifacts** around objects
Special Components	Standard condenser and objective lenses	Requires **phase plate and annular diaphragm**
Best Suited For	General observation of **stained tissue sections, slides**	Observing **live, unstained cells, organelles, and microorganisms**

Fluorescence Microscopy

Fluorescence microscopy is a powerful imaging technique that uses fluorescence to generate an image. It relies on the ability of certain substances, called fluorophores, to absorb light at a specific wavelength and emit it at a longer wavelength. This technique is especially useful in biological and biomedical sciences for studying structures and processes at the cellular and molecular levels.

Components of a Fluorescence Microscope:

1. **Light Source** – Typically a high-intensity lamp (e.g., mercury or xenon lamp) or lasers, providing excitation light.
2. **Excitation Filter** – Selects the specific wavelength of light that excites the fluorophore.
3. **Dichroic Mirror (Beamsplitter)** – Reflects the excitation light toward the sample and transmits emitted light toward the detector.
4. **Objective Lens** – Focuses the light on the sample and collects the emitted fluorescence.

5. **Emission Filter** – Allows only the emitted fluorescence light to reach the detector, blocking the excitation light.
6. **Detector** – Usually a camera or photomultiplier tube that captures the fluorescent image.

Advantages:

- **High specificity** due to fluorophore-target binding.
- **Real-time imaging** of dynamic processes in living cells.
- **Enhanced contrast** in comparison to traditional light microscopy.
- Allows **multi-labeling**, enabling visualization of multiple targets simultaneously.

Applications:

- **Cell biology** – Observing organelles, proteins, and intracellular dynamics.
- **Medical diagnostics** – Identifying pathogens, cancer markers, or genetic abnormalities.
- **Neuroscience** – Tracing neural pathways and synaptic activity.
- **Genetics and molecular biology** – In situ hybridization and gene expression studies.

Confocal Microscopy

Confocal microscopy is an advanced optical imaging technique that improves upon conventional fluorescence microscopy by providing sharper, high-resolution images with better depth selectivity. It uses point illumination and a spatial pinhole to eliminate out-of-focus light, enabling the collection of clear images from specific focal planes within thick specimens.

Components of a Confocal Microscope:

1. **Laser Light Source** – Provides precise, intense, and monochromatic excitation light.
2. **Beam Splitter (Dichroic Mirror)** – Directs laser light to the specimen and transmits emitted fluorescence to the detector.
3. **Scanning System (Galvanometer Mirrors)** – Moves the laser beam point-by-point across the sample.
4. **Objective Lens** – Focuses the laser onto the specimen and collects emitted light.

5. **Pinhole Aperture** – Blocks out-of-focus light, allowing only in-focus light to reach the detector.
6. **Photodetector (PMT or CCD)** – Captures the emitted light to generate high-resolution images.
7. **Computer and Imaging Software** – Reconstructs and processes optical sections to form 3D images.

Advantages Over Conventional Fluorescence Microscopy:

- **Higher resolution** and **contrast** by eliminating background fluorescence.
- **Optical sectioning** allows the imaging of specific layers within thick specimens.
- **3D reconstruction** of samples through serial imaging of optical slices.
- **Reduced photobleaching** outside the focal plane.
- Ideal for **live-cell imaging** with improved clarity.

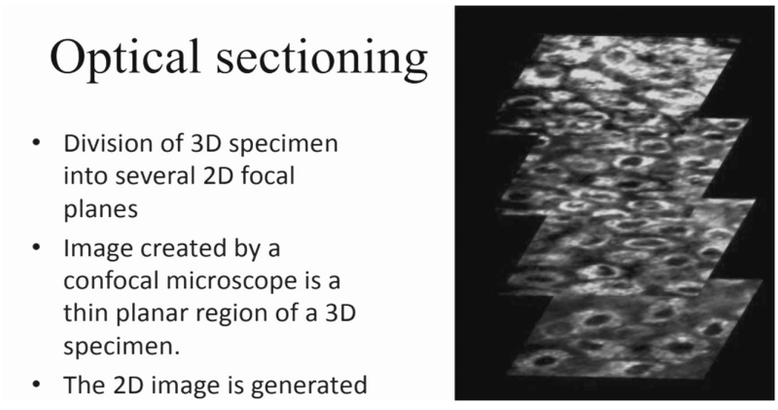


Figure 11.7

Applications:

- **Cell and tissue imaging** – Visualizing subcellular structures and complex tissues.
- **Developmental biology** – Studying embryo development in 3D.
- **Neuroscience** – Mapping neural networks and brain structures.
- **Cancer research** – Identifying cellular morphology and tumor microenvironments.
- **Material science** – Investigating surface features and microstructures.

Introduction to Electron Microscopy

Electron microscopy is a cutting-edge imaging technique that uses a beam of electrons instead of light to visualize specimens. Electrons are used because they have **much shorter wavelengths** than visible light, which allows for **significantly higher resolution** and the ability to reveal **ultrastructural details** at the nanometer and even atomic scale.

Why Use Electrons?

- According to the wave-particle duality principle, electrons can behave like waves.
- The **wavelength of electrons** (especially when accelerated) is **thousands of times shorter** than that of visible light.
- Shorter wavelength = **greater resolving power**, allowing observation of much **smaller structures** than is possible with light microscopy.

How It Helps in Microscopy:

- Enables visualization of **fine cellular structures**, such as organelles, membranes, and even molecules.
- Reveals **surface textures and internal structures** with exceptional clarity.
- Crucial for fields that require **nanometer or sub-nanometer resolution**, such as virology, materials science, and nanotechnology.

Differences from Light Microscopy:

Feature	Light Microscopy	Electron Microscopy
Radiation Used	Visible light	Electron beam
Resolution	~200 nm (limited by light)	~0.1-1 nm (much higher)
Lenses	Glass lenses	Electromagnetic lenses
Imaging Medium	Air or immersion oil	Vacuum (to prevent electron scattering)
Sample Prepartion	Simple; live imaging possible	Complex; usually requires fixation
Color Imaging	Natural or stained colors	Black and white (false color possible)

Types of Electron Microscopy

There are two main types of electron microscopy, each serving distinct purposes and offering different kinds of imaging:

1. Transmission Electron Microscopy (TEM)

Principle:

- In TEM, a **high-energy beam of electrons** is transmitted **through a very thin specimen**.
- Electrons interact with the sample, and the transmitted electrons are used to form an image.

Key Features:

- **Internal structures** of cells and materials can be observed in **great detail**.
- Produces **2D images** with **very high resolution** (up to ~0.1 nm).
- Requires **ultra-thin sample sections** and often involves staining with heavy metals for contrast.

Applications:

- Studying **cellular ultrastructure** (organelles, viruses, ribosomes).
- Examining **crystal structures** and **nanomaterials**.
- Investigating **macromolecular complexes** in biology.

2. Scanning Electron Microscopy (SEM)

Principle:

- In SEM, an electron beam **scans the surface** of a specimen.
- Electrons interact with atoms on the surface, producing **secondary electrons** that are collected to form an image.

Key Features:

- Provides detailed **3D-like surface images**.
- Lower resolution than TEM (~1–10 nm), but excellent **depth of field**.
- Samples are usually coated with a thin conductive layer (e.g., gold) to enhance imaging.

Applications:

- Observing **surface morphology** of cells, insects, materials.
- Used in **forensics, materials science, and semiconductor inspection**.
- Studying **topography and texture** of samples.

Summary Table:

Feature	TEM	SEM
Image Type	2D (internal)	3D-like (surface)
Resolution	Higher (~0.1 nm)	Lower (~1–10 nm)
Sample Requirement	Very thin slices	Surface intact, conductive coating
Focus	Internal structure	Surface details
Applications	Biology, virology, nanotech	Surface analysis, materials science

1. Transmission Electron Microscopy (TEM) Sample Preparation

Purpose: To create ultra-thin sections that allow electrons to pass through the specimen for internal structural imaging.

Steps:

1. **Fixation:**
   - Preserves cell structure using chemical fixatives like **glutaraldehyde** (for proteins) and **osmium tetroxide** (for lipids).
2. **Dehydration:**
   - Water is removed using a series of ethanol or acetone washes to avoid distortion during embedding.
3. **Embedding:**
   - The specimen is infiltrated with a **resin** (e.g., epoxy) to support it for ultrathin sectioning.
4. **Sectioning:**
   - Thin slices (typically **50–100 nm**) are cut using an **ultramicrotome** with a diamond or glass knife.
5. **Staining:**
   - Sections are stained with heavy metals like **uranyl acetate** and **lead citrate** to enhance contrast.
6. **Mounting:**
   - The sections are placed on **copper grids** for observation in the TEM.

2. Scanning Electron Microscopy (SEM) Sample Preparation

Purpose: To prepare the **surface** of the specimen for electron scanning while preserving its 3D structure.

Steps:

1. **Fixation:**
   - Similar to TEM, usually using **glutaraldehyde** to preserve surface features.
2. **Dehydration:**
   - Done through a graded ethanol or acetone series.
3. **Drying:**
   - **Critical point drying** is commonly used to avoid collapse or shrinkage due to surface tension during air drying.
4. **Mounting:**
   - The specimen is mounted on an **aluminum stub** using conductive adhesive (e.g., carbon tape).
5. **Coating (if non-conductive):**
   - A thin layer of **conductive material** (e.g., gold, palladium, platinum, or carbon) is sputter-coated to prevent charging under the electron beam.

How an Image is Generated by a Scanning Electron Microscope (SEM):

A Scanning Electron Microscope (SEM) generates highly detailed images of a sample's surface by scanning it with a focused beam of electrons.

Image Formation Process:

1. Electron Beam Generation:

   An electron gun emits a beam of high-energy electrons. This beam is focused into a fine probe using electromagnetic lenses.

2. Scanning the Surface:

   The electron beam is scanned in a raster pattern across the sample's surface.

3. Electron-Sample Interaction:

   When the beam hits the sample, it interacts with the atoms on the surface, producing various signals:

   - Secondary Electrons (SE): Ejected from the sample's surface; used to create high-resolution images showing surface topography.
   - Backscattered Electrons (BSE): Reflected from the sample; give information about composition (heavier elements appear brighter).
   - X-rays: Emitted due to inner shell electron transitions; used for elemental analysis (via EDS – Energy Dispersive X-ray Spectroscopy).

4. Signal Detection:

   Detectors collect the emitted secondary or backscattered electrons. The intensity of these signals varies depending on surface features and composition.

5. Image Display:

   The detected signals are converted into a grayscale image where contrast represents differences in electron emission. The image builds up line-by-line as the beam scans the sample.

Electron Beam & Sample Interactions

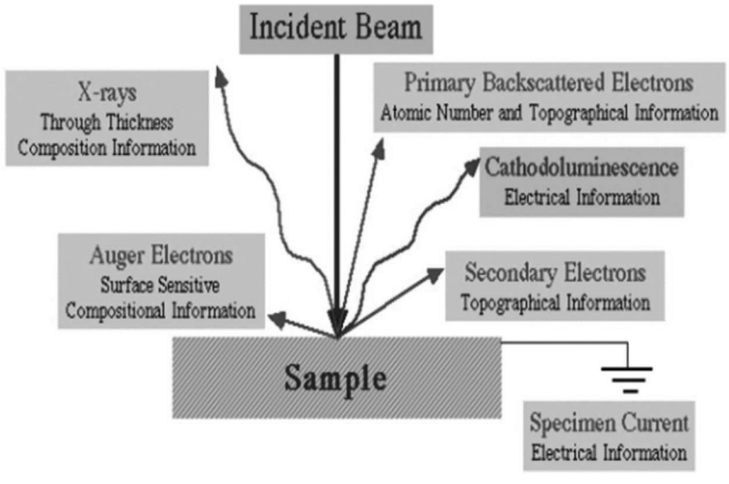


Figure 11.8

Key Differences:

Feature	TEM	SEM
Sample Thickness	Ultra-thin (50-100 nm)	Bulk sample with intact surface
Sectioning Required	Yes	No
Coating Needed	Usually not (unless insulating resin)	Yes, If specimen is non-conductive
Imaging	Internal structure	Surface morphology

Note on Cryo-TEM (Cryogenic Transmission Electron Microscopy):

Cryo-TEM is a powerful imaging technique used to observe biological specimens and other samples at cryogenic (very low) temperatures. Unlike conventional TEM, where samples may be damaged by the electron beam or preparation methods, Cryo-TEM preserves the native structure of the sample by rapidly freezing it in vitreous (non-crystalline) ice.

Key Features:

- **Preservation of Native State:** Samples are flash-frozen, typically using liquid ethane, to prevent the formation of ice crystals and maintain their original structure.
- **No Staining Required:** Since the technique relies on contrast created by electron scattering in frozen water, it avoids chemical stains that might alter the sample.
- **High Resolution:** Enables visualization of molecular and even atomic structures, especially in protein complexes, viruses, and nanoparticles.
- **Single Particle Analysis:** A major application of Cryo-TEM is the reconstruction of 3D structures from thousands of 2D projections of identical particles.

High pressure freezing and freeze substitution

Chemical fixation introduces a number of artifacts due to factors such as slow diffusion, selective reactions between fixatives and cell components and osmolarity differences between fixative and specimen. Subsequently, proteins can cluster together as a result of crosslinking, membranes become "wobbly" and antigenicity is affected. To avoid or substantially minimize the occurrence of these artifacts, a sample can be vitrified in liquid nitrogen and under very high pressure (over 2000 bar) instead of being fixed chemically. This results in instant and simultaneous immobilization of all cell components without ice crystal formation and cell component disruption.

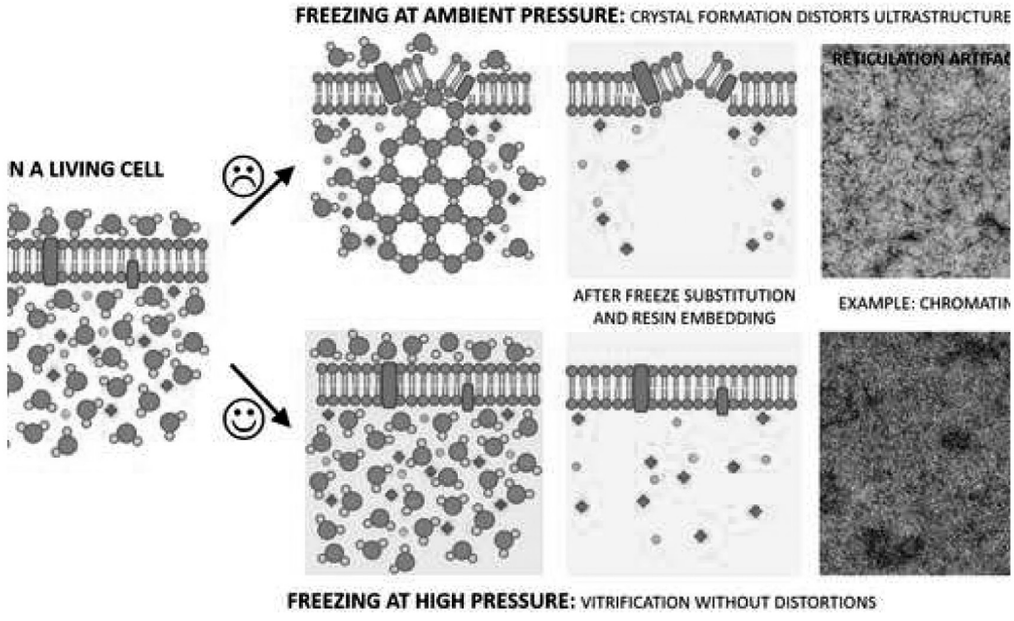


Figure 11.9: Difference between freezing at ambient pressure and high-pressure vitrification.

Sample size for high pressure freezing is limited as vitrification is effective only to the depth of 200 µm. Therefore, a very small and thin piece of tissue or a very small volume (several microlitres) of cell suspension is required. A solution of cryoprotectant, most commonly an animal serum albumin, is used to surround the sample during the freezing to prevent drying out and ice crystal formation and optimise heat transfer.

Following the freezing, vitrified water is removed with solvent containing one or a combination of the following components to increase contrast and additionally fix the specimen: a heavy metal stain (most commonly uranyl acetate), formaldehyde, glutaraldehyde, osmium tetroxide, tannic acid. This process is known as free substitution (Fig. 2). The solvent is finally replaced with liquid resin and cured at sub-zero temperature using UV light to avoid heat damage to epitopes.

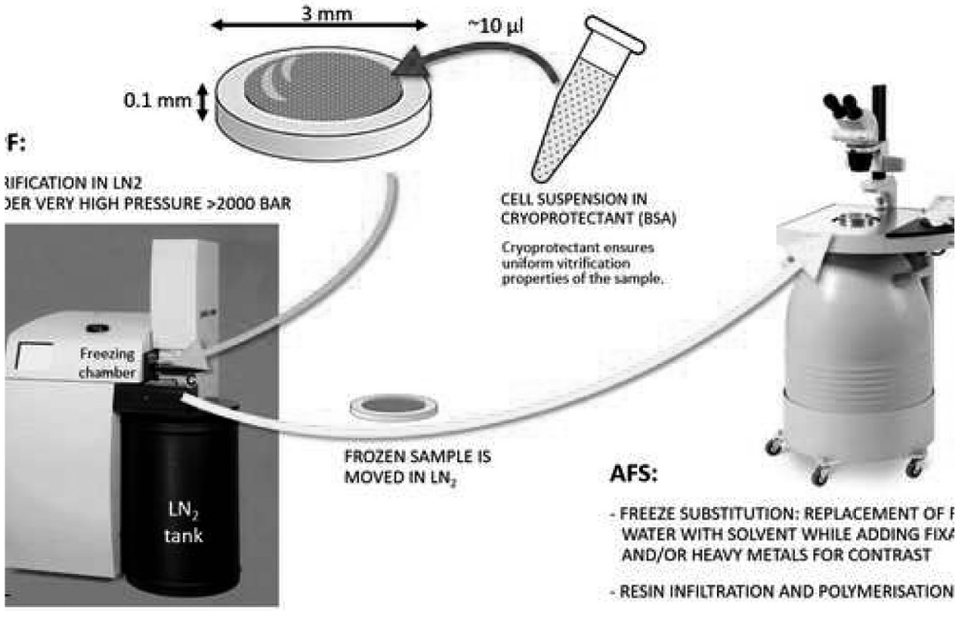


Figure 11.10: High pressure freezing and freeze substitution workflow.

Applications:

- Structural biology (e.g., proteins, viruses)
- Nanotechnology
- Materials science
- Drug discovery and development

Cryo-TEM has revolutionized structural biology, especially after advancements in detector technology and image processing software, earning it the nickname "the resolution revolution." In 2017, the Nobel Prize in Chemistry was awarded to scientists who developed the technique, underscoring its significance in modern science.

Probable Questions

1. Explain the principle and importance of Numerical Aperture (NA) in microscopy. How does it influence resolution, brightness, and depth of field?
2. Compare and contrast Bright-field, Phase-contrast, and Differential Interference Contrast (DIC) microscopy in terms of principle, image quality, components, advantages, and applications.
3. Describe the working principle of Fluorescence Microscopy. What are its key components, advantages, and major applications in biological research?
4. Discuss the differences between Transmission Electron Microscopy (TEM) and Scanning Electron Microscopy (SEM) with respect to image type, resolution, sample preparation, and applications.
5. Define resolving power in microscopy. How is it different from magnification, and what factors determine the resolution limit of a microscope?
6. Explain the process of sample preparation in TEM and SEM. Why are steps like fixation, dehydration, and coating crucial for obtaining quality images?
7. What is Cryo-TEM? How does it overcome the limitations of conventional chemical fixation? Describe the workflow of high-pressure freezing and freeze substitution.
8. Illustrate the working mechanism of Phase-contrast Microscopy. Why is it particularly suited for observing live, unstained biological specimens?
9. Enumerate and explain the types of light microscopy techniques. In what scenarios would you prefer one over the others? Provide suitable examples.
10. How does Confocal Microscopy differ from conventional fluorescence microscopy? Describe the principles, components, advantages, and 3D imaging capabilities of confocal microscopy.

Unit XII

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 Electron Microscopy, TEM Vs. SEM. Different fixation and staining techniques for EM, freeze-etch and freeze-fracture methods for EM, and image processing methods in microscopy.

Electron Microscopy

Introduction

Electron Microscopes were developed due to the limitations of Light Microscopes which are limited by the physics of light to a resolution of about 0.2 micrometers. In the early 1930's this theoretical limit had been reached and there was a scientific desire to see the fine details of the interior structures of organic cells (nucleus, mitochondria...etc.). This required 10,000x plus magnification which was just not possible using Light Microscopes. The Transmission Electron Microscope (TEM) was the first type of Electron Microscope to be developed and is patterned exactly on the Light Transmission Microscope except that a focused beam of electrons is used instead of light to "see through" the specimen. It was Ernst Ruska and Max Knoll, a physicist and an electrical engineer, respectively, from the University of Berlin, who created the first electron microscope in 1931. Ernst Ruska later received Nobel Prize for his work in 1986. Conventional transmission electron microscope (TEM) today can achieve a resolution of 0.05nm.

Principle of Electron Microscope

Electron microscopy is used when the greatest resolution is required, and when the living state can be ignored. The images produced in an electron microscope reveal the ultrastructure of cells. There are two different types of electron microscope – the transmission electron microscope (TEM) and the scanning electron microscope (SEM).

In the TEM, electrons that pass through the specimen are imaged. In the SEM electrons that are reflected back from the specimen (secondary electrons) are collected, and

the surfaces of specimens are imaged. The equivalent of the light source in an electron microscope is the electron gun. When a high voltage of between 40 000 and 100 000 volts (the accelerating voltage) is passed between the cathode and the anode, a tungsten filament emits electrons. The negatively charged electrons pass through a hole in the anode forming an electron beam. The beam of electrons passes through a stack of electromagnetic lenses (the column). Focussing of the electron beam is achieved by changing the voltage across the electromagnetic lenses. When the electron beam passes through the specimen some of the electrons are scattered while others are focussed by the projector lens onto a phosphorescent screen or recorded using photographic film or a digital camera.

The electrons have limited penetration power which means that specimens must be thin (50–100 nm) to allow them to pass through. Thicker specimens can be viewed by using a higher accelerating voltage, for example in the high-voltage electron microscope (HVEM) which uses 1 000 000 V accelerating voltage or in the intermediate voltage electron microscope (IVEM) which uses an accelerating voltage of around 400 000 V. Here stereo images are made by collecting two images at 8–10 _ tilt angles. Such images are useful in assessing the 3D relationships of organelles within cells when viewed in a stereoscope or with a digital stereo projection system.

Applications of electron microscopy

i. Biology and Biomedical Sciences for ultrastructural analysis and study of cellular and subcellular structures such as organelles, membranes, and macromolecular complexes, pathology (diagnosis of various diseases by examining tissue samples at high resolution, allowing for detailed observations of cellular abnormalities), virology (visualization and characterization of viruses and viral structures, viral morphology and replication mechanisms).

ii. Forensic Science to analyze trace evidence such as fibers, hairs, paint chips, and gunshot residue, providing valuable information for criminal investigations.

iii. Nanotechnology for characterization of nanoscale devices such as nanowires, quantum dots, and nanotubes, nanostructure imaging to visualize and analyze the structure, composition, and behavior of nanomaterials, facilitating the design and optimization of nanotechnology-based applications.

iv. Material Science for characterization of nanomaterials, analysis of crystal structure, examination of surface morphology, aiding in the study of surface properties and phenomena like corrosion, adhesion, and catalysis.

v. Mineralogy to study the crystal structure, composition, and texture of minerals, aiding in mineral identification and characterization.

vi. Petrology in the analysis of rock samples to understand their formation processes, mineralogy, and geochemical properties.

vii. Electronics and Semiconductor Industry for the visualization and analysis of semiconductor device structures, defects, and interfaces, crucial for the development and improvement of electronic devices.

There are two types of electron microscopes, with different operating styles:

1. TEM [Transmission Electron Microscope]
2. SEM [Scanning Electron Microscope]

TEM [Transmission Electron Microscope]

In TEM, electrons pass through the specimen to form an image of internal structure of the specimen. It requires ultrathin sections for electrons to penetrate with ease. The specimen is placed along the direction of the electron beam.

TEM has three working parts which include:

1. Electron gun
2. Image producing system
3. Image recording system

Electron gun

- This is the part of the Transmission Electron Microscope responsible for producing electron beams.
- Electrons are produced by a cathode that is a tungsten filament that is V-shaped and it is normally heated. The tungsten filament is covered by a control grid known as a Wehnelt cylinder made up of a central hole which lies columnar to the tube. The cathode lies on top of or below the cylindrical column hole. The cathode and the control grid are negatively charged with an end of the anode which is disk-shaped that also has an axial hole.
- When electrons are transmitted from the cathode, they pass through the columnar aperture (hole) to the anode at high voltage with constant energy, which is efficient for focusing the specimen to produce an accurately defined image.

- It also has the condenser lens system which works to focus the electron beam on the specimen by controlling the energy intensity and the column hole of the electron gun. The TEM uses two condenser lenses to converge the beam of electrons to the specimen. The two condenser lens each function to produce an image i.e., the first lens which has strong magnification, produces a smaller image of the specimen, to the second condenser lens, directing the image to the objectives.

Image- Producing system

- It's made up of the objective lens, a movable stage or holding the specimen, intermediate and projector lenses. They function by focusing the passing electrons through the specimen forming a highly magnified image.
- The objective has a short focal length of about 1-5mm and it produces an intermediate image from the condenser which are transmitted to the projector lenses for magnification.
- The projector lenses are of two types, i.e the intermediate lens which allows great magnification of the image and the projector lens which gives a generally greater magnification over the intermediate lens.

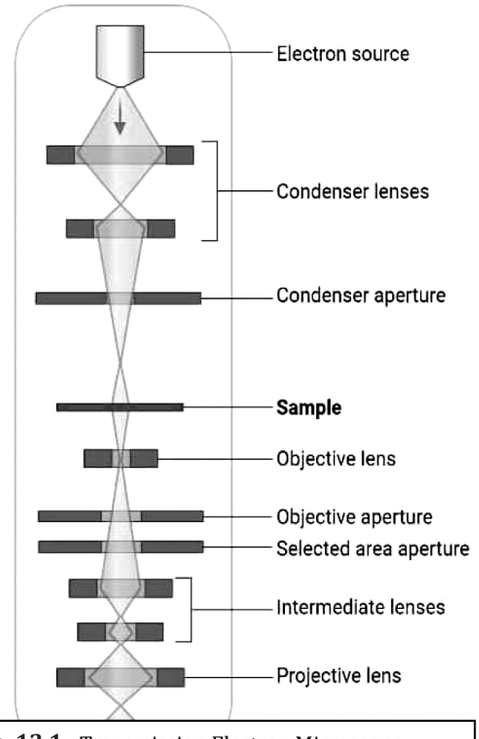


Fig. 12.1: Transmission Electron Microscope [Source: https://microbenotes.com/transmission-electron-microscope-tem/]

- To produce efficient high standard images, the objectives and the projector lenses need high power supplies with high stability for the highest standard of resolution.

Image-Recording System

- It's made up of the fluorescent screen used to view and to focus on the image. They also have a digital camera that permanently records the images captured after viewing.

- They have a vacuum system that prevents the bombardment or collision of electrons with air molecules disrupting their movement and ability
- to focus. A vacuumed system facilitates the straight movement of electrons to the image.
- The vacuumed system is made up of a pump, gauge, valves and a power supply.
- The image that is formed is called a monochromatic image, which is greyish or black and white. The image must be visible to the human eye, and therefore, the electrons are allowed to pass through a fluorescent screen fixed at the base of the microscope.
- The image can also be captured digitally and displayed on a computer and stored in a JPEG or TIFF format. During the storage, the image can be manipulated from its monochromatic state to a colored image depending on the recording apparatus e.g. use of pixel cameras can store the image in colour.
- The presence of colored images allows easy visualization, identification, and characterization of the images.

Preparation of specimen for ultrastructure imaging

A conventional protocol to prepare samples for ultrastructure imaging involves the following steps.

1. **Primary fixation with aldehydes (proteins)**

   During this step proteins and, to a lesser extent, other cell molecules, become crosslinked by formaldehyde and/or glutaraldehyde molecules. Small mammals can be fixed by perfusion, whereby the fixative is introduced via the vascular system. Other samples need to be fixed by immersion and the specimen needs to be dissected no thicker than 1 mm in at least one direction.

2. **Secondary fixation with osmium tetroxide (lipids)**

   This step ensures that lipids, for example the phospholipids forming membranes, are preserved and are not extracted during dehydration. During the fixation a black insoluble precipitate is formed on the membranes, creating membrane contrast.

3. **Tertiary fixation and contrasting with uranyl acetate**

   Uranyl acetate is a heavy metal salt which binds to proteins, lipids and nucleic acids, providing additional contrast. Some authors believe it also has fixative properties. Samples can be incubated en bloc in a solution of uranyl acetate

before dehydration but the stain can also be applied to the sectioned specimen before lead staining.

4. **Dehydration series with solvent (ethanol or acetone)**

   A fixed specimen is dehydrated by incubation in a series of ethanol or acetone solutions. Solvent concentration is increased gradually so that water is removed gently, without causing artefacts, mainly shrinkage.

5. **Resin infiltration and embedding**

   Following dehydration, the solvent is replaced with a gradually increasing concentration of liquid resin (typically epoxy resin for ultrastructure studies). The specimen is placed in a mold filled with liquid resin and cured into a hard block using heat or UV light. After this, a sample can be stored indefinitely.

6. **Sectioning and mounting sections on specimen grids**

   A specimen embedded in hardened resin can be sectioned extremely thinly, at less than 100 μm. This allows for the electron beam to pass from the electron gun through the specimen to the detector. The sections are mounted on specimen grids which fit into microscope sample holder.

7. **Contrasting (post staining)**

   Biological specimens are naturally not very electron opaque as they are composed of atoms with low atomic numbers and the beam passes through them easily. To increased ample contrast, the sections can be post-stained with lead citrate. This heavy metal salt, similarly to osmium tetroxide and uranyl acetate, binds to cell components and scatters the incident beam electrons. The areas of specimen section which scatter electrons more are recorded as darker pixels, which stand out against the brighter background.

Application of TEM

i. TEM is used to gather information on samples' topography, morphology, atomic composition, and crystal structures. This information is used to study materials at molecular level, analyzing nanoscale structures in the materials.

ii. It is also applied in manufacturing sectors such as semiconductors and nanotechnology.

iii. Forensic science is another rich application area for TEM. In forensics, close materials analysis can provide evidence for law enforcement and justice processes. For example, TEM can be used to analyze gunshot residue, blood found

in clothing fibres or items left at crime scenes, and other biological substances that could be used to find DNA and pinpoint the identity of people at the scene.

iv. TEM has a significant role in scientific research. In life sciences, for example, TEM devices are used to study the nature and mechanisms of diseases more elaborately which increase the knowledge and helps to find the ways to combat them.

Advantages of TEM

i. Highest magnification is achieved by TEM over any other magnifying technology. Because of this, TEM devices can give insight into materials' elements and compound structures.

ii. The images from TEM surveys are of high quality and contain large amounts of detail. This detail includes information about materials' nano surface features, shape, size, and molecular structure.

Disadvantages of TEM

i. TEM devices are large, unwieldy, and very expensive.

ii. Installing them usually requires bespoke groundworks and often climate-controlled environments and highly efficient technologist is required for installing and operating this microscope.

iii. Sample preparation is relatively laborious, requiring scientists to obtain ultra-thin samples of target materials. Sample preparation can also lead to artifacts appearing in the sample, leading to erroneous survey results.

SEM [Scanning Electron Microscope]

Instrumentation

SEM is a very complex structure with a variety of components operating in it to analyze the data of the specimen surface. The essential components in the SEM constitute an electron gun, condenser and objective lens, specimen stage, secondary electron detector, image display, recording, and vacuum system.

1. **Electron gun:** The electron gun is used to produce the electron beam that mostly uses thermionic emission from the cathode source (tungsten filament). The filament is heated to a very high temperature (2800K) and the emitted

thermoelectric is focused through a metal plate which acts like an anode. This is done in order to focus the current of the electron beam at the desired point.

2. **Condenser and objective lens:** These lenses are used to enable the adjustment of the diameter of the electron beam. A condenser lens helps in strengthening the electron beam and adjusting the diameter of the electron beam when it passes through this lens. An objective lens is used to focus the electron beam onto the specimen surface and it determines the final diameter of the electron beam.
3. **Specimen stage:** This acts as the supporting base of the specimen which stably supports the specimen by moving smoothly in vertical, horizontal, and rotational ways.
4. **Secondary electron detector:** This is used to detect the secondary electrons emitted from the specimens and it is placed above the objective lens. Magnetic fields are utilized in detecting secondary electrons.
5. **Image display and recording:** The output signals obtained from the secondary electron detector are amplified and sent to the display unit. Initially, cathode ray tubes were used for the display units however now the liquid crystal display is being used. Recording of these images is obtained in digital format.
6. **Vacuum system:** The electron optical system and the specimen chamber should be in vacuum conditions and hence the components are evacuated by diffusion pumps. In the case of an oil-free environment then, turbo molecular pumps are used.

Application of SEM

i. SEM is used to determine high-resolution images of the material's shapes and determine the chemical compositions by acquiring the elemental maps.

ii. In biological science, large objects such as insects and animal tissues through to small objects such as bacteria are studied by SEM. SEM can be used in entomology, archaeology, plant science, cell research, and taxonomy among other topics.

iii. In medical science, SEM is used to compare blood cells and tissue samples to determine the cause of illness. It is also used to study the medicine and its effect on patients as well as for researching and developing new treatments.

iv. In forensic science police laboratories use SEM to examine and compare evidence, such as metal fragments, paint, inks, hair and fibres to provide evidence of a person's guilt or innocence.

Advantages of SEM

i. Application is easy and used to analysis the surface of material.

ii. The image processing and data analysis of the topographical characteristics of any material specimen can be done easily.

Disadvantages of SEM

i. SEM can damage samples, particularly those that are sensitive to electrons, such as organic or biological materials.

ii. SEM is a powerful technique but delicate tool to use, requiring specific skills and ongoing training.

iii. SEM samples must be solid and vacuum-compatible. However, higher pressures can be used for imaging of vacuum-sensitive samples that are nonconductive and volatile.

iv. Samples that are strong insulators must be coated—usually with gold or carbon—before testing.

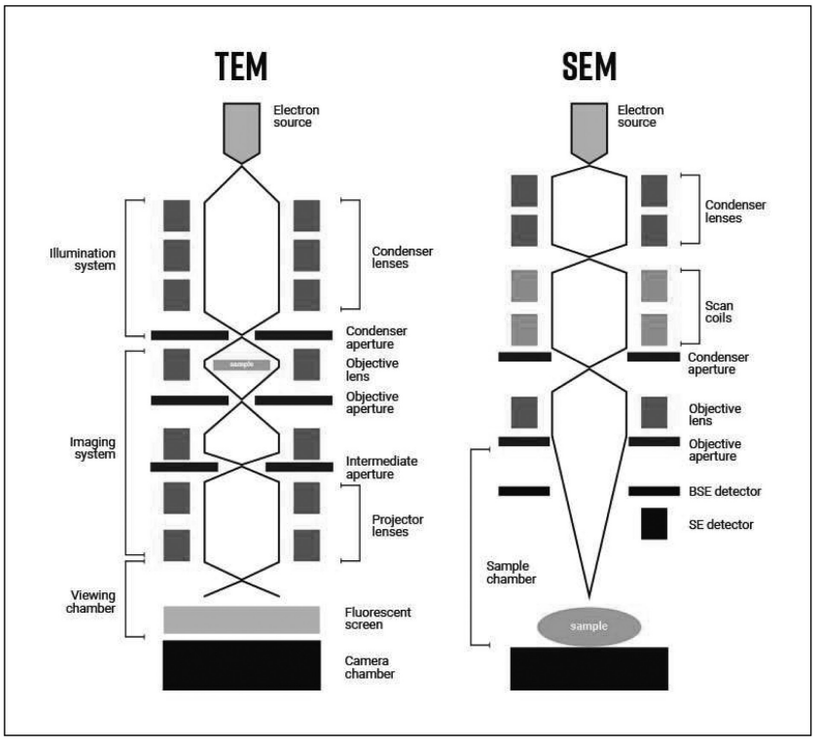


Figure 12.2: Schematic diagram of TEM vs SEM column design, highlighting the major components (SEM vs TEM)

[Source: https://www.nanoscience.com/]

Summarizing the key differences between SEM and TEM

	SEM	TEM
Signals analyzed	Backscattered electrons, secondary electrons	Transmitted electrons
Image formation mechanism	The beam is scanned across the sample surface and signals are collected in a pixel-by-pixel fashion	Static beam illuminates the sample and a projection image of the transmitted beam is acquired in a single frame
Sample thickness	Any (limited by sample chamber)	<100 nm for most materials
Accelerating voltage	~1-30 kV	~30-300 kV
Type of information in the image	Surface compositional contrast, surface morphology, surface topography (can have 3D appearance)	Phase or diffraction contrast image through sample's surface and internal structure (2D appearance)
Max magnification	Up to 1-2 million times	50 million times or more
Typical field of view	A few μm to several mm's	A few nm's up to several μm
Spatial resoltuion	~1-2 nm	<1 Å
Size of instrument	Smaller, desktop models and floor model options	Larger and taller, takes up entire room
Cost	Less expensive	More expensive
Time-to-image	Fast	Slow

- **Different fixation and staining techniques for EM**

Primary fixation of specimens. Prior to submitting samples for electron microscopy processing, it is imperative to use the proper fixation method and fixative to ensure adequate preservation of the targeted cellular structures. Fixative solutions should be freshly prepared, and ideally used the same day as prepared, and no longer than next-day use. Two major types of fixations are perfusion and immersion. Perfusion fixation involves flushing blood from a specimen using a saline solution and then flowing fresh fixative through a major blood vessel to reach the target tissue before excising the tissue (Glauert et al, 1999).

For tissue that is sensitive to oxygen deprivation, such as kidney, brain, or heart, perfusion fixation is the preferred method. In many cases, the excised tissue is further fixed by immersion. Immersion fixation involves immediate submersion of small pieces of tissue in fresh fixative solution (Glauert et al, 1999). If the tissue is not sensitive to being removed from the blood supply, such as skin or cells from cell culture, immersion

fixation is suitable and easier to perform. For tissue, the procedure begins by mincing the tissue into small pieces (less than 1 mm in at least two dimensions; 0.5 mm is even better) in a drop of fixative. For cells, the fixative can be added directly to the culture plates (adherent cells) or conical tubes containing suspended cells. Samples are then kept in a volume of fixative twenty times greater than the volume of tissue for one hour at room temperature and then kept at 4°C overnight. Following fixation, the sample should be washed with a buffered solution no longer than 24 hours after initial fixation. Failure to properly fix tissue affects the quality of processing downstream, and even the best post-fixation staining cannot recover lost structural integrity.

Fixative solutions containing aldehydes are preferentially used for primary fixation in EM studies.

Glutaraldehyde is the most popular choice, as its irreversible reaction with amine groups, especially lysine, allows for strong crosslinking of proteins and thus best preserves the ultrastructure of cells the slow diffusion rate of glutaraldehyde in tissues is often insufficient to properly fix sensitive tissues. Dense tissue such as kidney and heart benefit from a fixative solution combining glutaraldehyde with paraformaldehyde, such as Karnovsky's fixative (Karnovsky, 1965) at 2% glutaraldehyde and 2% paraformaldehyde.

The **paraformaldehyde** rapidly penetrates tissue, allowing the glutaraldehyde more time to stabilize ultrastructure. For immunocytochemical studies, such as those employing colloidal gold, using a fixative with a higher percentage of paraformaldehyde relative to glutaraldehyde is advantageous, as paraformaldehyde is less reactive than glutaraldehyde.

1. The TOU method

The TOU method of post-fixation staining is the standard TEM protocol we use for a variety of cell and tissue preparations. The protocol is to first incubate glutaraldehyde-fixed tissue in **T**annic acid, **O**smium tetroxide, then **U**ranyl acetate prior to dehydration and embedding. It is generally a cost-effective method for obtaining high-contrast images that preserve the high-resolution character of EM images. It is also suitable for standard SEM imaging and array tomography.

Tannic acid acts as both a fixative and a mordant. As a fixative, it penetrates slowly but strongly fixes connective tissues, such as elastic fibers, and is thought to crosslink both proteins and carbohydrates, though the mechanism is not fully known (Glauert et al, 1999). It also aids in the preservation of membranes (Hart et al, 2024). As a mordant, it enhances contrast in imaging by increasing the uptake of heavy metals such as UA and lead stains (Glauert et al, 1999).

OsO4 is a potent secondary fixative and stain that is at the heart of ultrastructural EM analyses. Its use has been documented since the early days of electron microscopy, and no reagent since has been deemed suitable to replace it. While the chemistry of OsO4 is not fully understood, the tetroxide is known to react with lipids, proteins, and carbohydrates, forming cross-links that stabilize cytoplasmic structures and surface membranes (Glauert et al, 1999). Its slow penetration makes it a poor primary fixative, but its use as a secondary fixative for aldehydes is critical in the preservation of membranes since aldehydes form cross-links between proteins but not lipids. It is typically used at a 1% concentration in buffered solution. OsO4 is a highly toxic reagent in all routes of exposure and should only be handled inside an operational fume hood while wearing a lab coat and nitrile gloves at all times.

Uranyl acetate is an important fixative and stain for studying ultrastructure by EM. Its fixative properties are especially notable in membranes, where it binds to and stabilizes phosphate groups in the membrane phospholipids, but it also fixes DNA and RNA. Further, UA reduces lipid extraction from membranes during dehydration (Glauert et al, 1999). As a stain, it enhances the contrast of membranes, organelles, and extracellular proteins. Samples are typically incubated in a 1% aqueous solution for 1 hour at room temperature. Since UA is light-sensitive, covering the samples with foil during incubation is recommended to prevent the UA from precipitating out of solution.

2. The OTO method

The OTO method of post-fixation staining is also used for a variety of cell and tissue preparations in our core, as it results in more contrasted and stabilized tissue, especially membranes, than TOU. The order in this method is **O**smium tetroxide-potassium ferricyanide, **T**hio-carbohydrazide, then **O**smium tetroxide. It has also been shown to provide enhanced staining in tissues with higher lipid concentrations (Seligman et al 1966; Guyton et al 1985).

The purpose of OsO4 in post-fixation staining has already been discussed. The addition of potassium ferricyanide to OsO4 in the first step of this protocol primarily enhances the staining of glycogen in tissues, but it also provides additional staining of membranes (Glauert et al, 1999). The chemistry is not fully understood, but it is thought that ferricyanide acts as an oxidizing agent on reduced osmium and other oxidizable sites produced by primary fixation with aldehydes (Glauert et al, 1999).

Similar to tannic acid in the TOU protocol, TCH acts as a mordant between the two OsO4 incubations. It forms a “bridge” between the two osmium steps in the OTO

protocol, enhancing membrane staining (Seligman et al, 1966). It also stains carbohydrates (Seligman et al, 1965).

• Freeze-etch and freeze-fracture methods for EM

Freeze fracture describes the technique of breaking a frozen specimen to reveal internal structures. Freeze etching is the sublimation of surface ice under vacuum to reveal details of the fractured face that were originally hidden. A metal/carbon mix enables the sample to be imaged in a SEM (block-face) or TEM (replica). It is used to investigate for instance cell organelles, membranes, layers and emulsions. The technique is traditionally used for biological applications but started to develop significance in physics and material science.

Freeze-fracture methods for EM

i. The specimens are subjected to ultra-rapid freezing immobilizing the cellular components immediately.

ii. The freezing is carried out in the presence of cryoprotective agents such as glycerol to prevent ice crystal formation.

iii. The frozen specimen is then fractured at liquid nitrogen cooled temperatures under high vacuum. The fractured specimen splits into two halves conferring two faces called by convention the PF-face (plasma fracture-face) and EF-face (extracellular fracture-face).

iv. The fracture is irregular and occurs along lines of weakness like the plasma membrane or surfaces of organelles.

v. The resultant fractured surface is then shadowed with platinum vapor from an angle that confers three-dimensional detail to the cast. This is followed by evaporation of carbon vertically onto the surface to produce carbon replica.

vi. The organic material is digested away during the procedure leaving a replica.

vii. The carbon-metal replica is trimmed to proper size retrieved onto a standard EM grid and examined under electron microscope. The replica is actually a template like impression of the distribution of particulars in the original specimen.

viii. Freeze-fracture preparations are examined by transmission electron microscopy and their major contribution to high resolution morphologic studies is their unique representation of structure/function elements of cell membranes.

ix. The electron beam readily passes through the portions of the replica containing the carbon but is absorbed by the areas containing platinum.

x. The resulting images which have a three-dimensional impact are considered different from those obtained with sectional materials.

xi. This technique is used specially for the investigation of cell membranes and their specializations and has contributed considerably to the understanding of cellular form to related cell function.

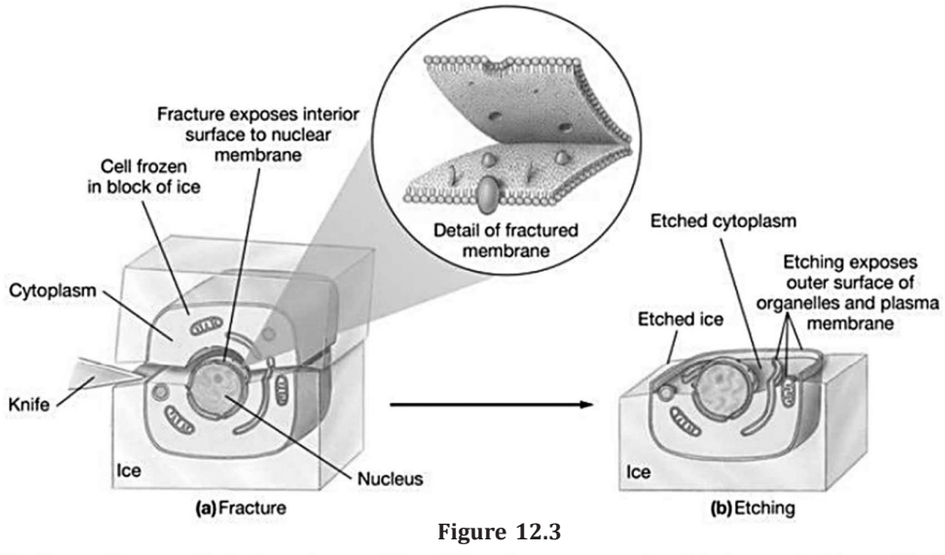


Figure 12.3

The Freeze-Fracture Technique. In steps (a) and (b), a frozen eucaryotic cell is fractured with a cold knife. Etching by sublimation is depicted in (c). Shadowing with platinum plus carbon and replica formation are shown in (d) and (e).

Source: https://www.deshbandhucollege.ac.in/pdf/resources/1588005857_LS-VI

Freeze-etch methods for EM

i. Freeze fracture replication by itself is an extremely valuable technique but Russell Steer (early 1970s) made even more informative by including a step called freeze etching.

ii. In this step, the frozen fractured specimen is exposed to vacuum at an elevated temperature for a few minutes while still in the cold chamber.

iii. As result a layer of ice evaporates (sublimates) from the exposed surface, preserving the structure of the specimen.

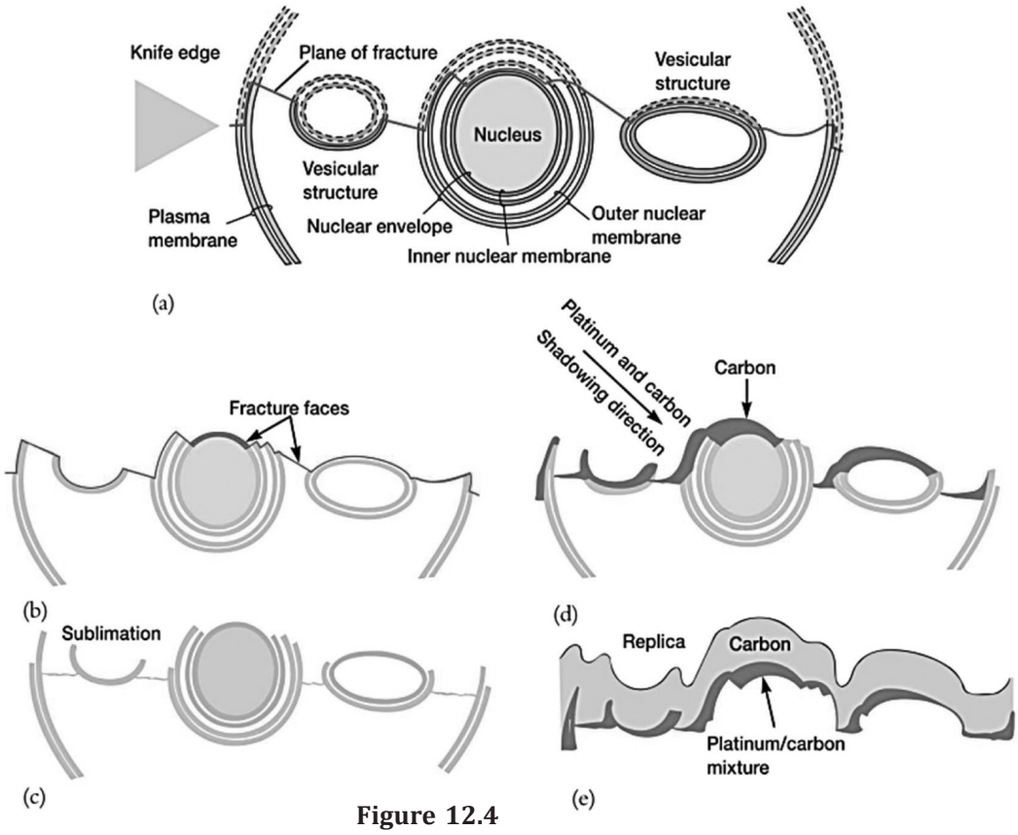


Figure 12.4

The Freeze-Etching Technique. In steps **(a)** and **(b)**, a frozen eucaryotic cell is fractured with a cold knife. Etching by sublimation is depicted in **(c)**. Shadowing with platinum plus carbon and replica formation are shown in **(d)** and **(e)**.

Source: https://www.deshbandhucollege.ac.in/pdf/resources/1588005908_LS-VI-Practical_Study_of_Microscopic_techniques-2.pdf

iv. Now the surface of the structure is coated with ultrathin (2-7 nm) heavy metal coating followed by thick (15-20 nm) carbon coating to stabilize metal film and to elate a metallic replica that reveals the external and internal surfaces of the cellular membrane.

v. The technique delivers very high resolution and can be used to reveal the structure and distribution of macromolecular complexes such as those of cytoskeleton.

vi. This technique has been successfully employed to determine the presence of tight or occluding junctions where membrane glycoproteins bind cells together. The technique serves as the only way to determine the presence of such junctions.

vii. This technique has also been employed to study the inter-membrane structures.

viii. It also allows the visualization and detailed analysis of the function of specific proteins in bacteria and viruses.

ix. The technique is further improved by John Heuser as **deep-etching technique** with even higher resolution (about 2 nm). Greater amount of surface ice is removed which enables the visualization of cellular organelles.

Image processing methods in microscopy

- **Image formation in SEM**

An image in SEM is obtained by scanning the fine focused electron beam over the surface of the specimen and the simultaneous registration of the signals from the detectors. At each point of the specimen the electrons of the beam interact with the specimen. A number of phenomena occur as the result of the elastic and inelastic scattering of the primary electrons. If in the case of elastic scattering the scattering angle exceeds 90°, the electron is said to be **backscattered** and may emerge from the specimen close to the point where it entered. The efficiency of elastic scatter events increases with the atomic number of the specimen. A region containing elements with a high atomic number will produce more backscattered electrons than a region with a low atomic number. Therefore, chemical phases can be recognized in backscattered electron images based on atomic number differences.

Since the backscattered electrons are coming from deeper in the specimen, the interaction volume is much larger than the beam diameter and, as a consequence, the resolution in the backscattered image is at most of the order of 200 nm.

If in the case of inelastic scattering the final stage of the transitions of the electrons lies above the vacuum level of the solid and if the excited atomic electron has enough energy to reach the surface of the specimen, it may be emitted as a **secondary electron**.

The secondary electrons can only escape from a very shallow depth. The intensity of the secondary electron emission is little influenced by the composition of the specimen but is highly dependent on the orientation of the sample surface with the respect to the detector. This makes that they provide important topographical information about the surface of the specimen. The low exit depth allows the resolution of the order of 5-20 nm to be reached.

Backscattered electrons are detected by a set of 2 solid state detectors, mounted close above the specimen. The electron beam passes through the hole and the backscattered electrons hit the detector and produce a current. The signals which are obtained by both detectors, can be combined into two types of images, topographical and compositional.

The topographical image originates from the difference in the incident and backscattered angle and can be obtained by subtracting the signals from both detectors.

The secondary electrons are detected by a scintillator-photomultiplier combination which is known as the Everhart-Thornley detector (Kiss,1988). The secondary electrons are collected by a grid. The electrons which pass through the collector grid, are accelerated to the scintillator and basically, they generate photons interacting with the scintillator. An image in SEM is obtained by scanning the fine focused electron beam over the surface of the specimen and the simultaneous registration of the signal from a detector.

The signal which was formed in the detector, is suitably amplified and used to modulate the intensity of a cathode ray tube (CRT) which is scanned in synchronism with the electron beam and, thus, a SEM image is formed. At the exit of the head amplifier the video signal is normally proportional to the number of electrons recorded. This signal can be used not only to modulate the intensity of a CRT, but also can be converted to a digital form and the image can be stored digitally. This is done by an analogue-to-digital converter (Reimer, 1985; Russ, 1990) which is usually connected to a computer and, therefore, the digitized image can be directly transferred to the computer. A digital image is represented as a two-dimensional data array where each data point is called a picture element or pixel.

A digitized SEM image consists of pixels where the intensity of each pixel is proportional to the number of the backscattered (in a backscattered electron image) or secondary (in a secondary electron image) electrons, emitted from the corresponding point on the surface of a specimen. Such images are called **gray level images.**

• Image formation in TEM

If a specimen is sufficiently thin and the energy of the incident electrons is high enough, most of the incident electrons will pass through the specimen with little or no energy loss and can be 'visualized' on a fluorescent screen or registered on a photographic plate located below the specimen. The TEM image obtained in this way is

called a **conventional electron micrograph** and is, basically, the result of the *scattering, diffraction and absorption of the electrons* as they pass through the specimen.

Different regions of the specimen scatter the electrons by various degrees where the extent of the scattering depends on the local elemental composition and the local thickness and density of the specimen. The TEM is used to investigate ultra-thin samples, typically less than 200 nm. The image resolution depends on the sample thickness and the aberrations of the lenses and is typically of the order of 1-2 nm.

Inelastically scattered electrons are indistinguishable from the unscattered electrons and are also imaged into the final image plane. Modern transmission electron microscopes are equipped with the imaging electron energy spectrometer (filter of the prism/mirror/prism type) which allows to work in the electron spectroscopic imaging (ESI) mode so that the electrons of selected energy loss can be visualized. The magnetic field of the prism disperses the electron beam according to the energy of the electrons and the electrons which did not loss energy, are deflected over 90°, reflected by the electrostatic field of the mirror and deflected again by the magnetic field into the optical axis of the microscope.

Electrons which have suffered energy losses are slower, so the magnetic field of the prism deflects them at larger angles. After the spectrometer they move outside the optical axis and are caught by the spectrometer slit. Only the unscattered electrons (electrons which suffered no energy loss) reach the final image plane and such imaging is called zero-loss electron spectroscopic imaging (ZLESI). Zero-loss filtering not only increases the contrast, also a better comparison with simulated images is possible. To image with electrons which have lost an energy “E the accelerating voltage E of the primary electron beam is increased to E + “E and the beam electrons which have lost energy “E in the specimen, enter the spectrometer with an energy E. They stay on the optical axis of the microscope and pass through the spectrometer slit into the final image plate. Thus, by varying the accelerating voltage, electrons of selected energy loss can be used for imaging. The choice between zero-loss filtering and the use of an energy-loss window depends on the specimen and the information wanted.

Probable questions

1. Write the principle of Electron microscope?
2. State the advantages and disadvantages of SEM and TEM?
3. Write the differences between TEM and SEM.

4. Write the application of Electron microscoipe.
5. State the preparation of specimen for ultrastructure imaging.
6. What is OTO method of fixation and staining techniques for Electron microscope?
7. Which chemicals are used in OTO method?
8. What is the role of Osmium tetroxide in sample preparation for staining techniques for Electron microscope?
9. Role the role of tannic acid in sample preparation for staining techniques for Electron microscope?
10. Describe TOU method of fixation and staining techniques for Electron microscope?
11. Role of Glutaraldehyde in in sample preparation for Electron microscope?
12. Describe the freeze-fracture methods for EM.
13. What is secondary electron? What is the function of it?
14. What is CRT?
15. What is grey level image?
16. Discuss the image processing methods in SEM?
17. Discuss the image processing methods in TEM?

Suggested Reading:

1. Wilson and Walker's Principles and Techniques of Biochemistry and Molecular Biology 2018, ISBN: 9781316677056
2. Biochemistry Laboratory: Modern Theory and Techniques, Latest Edition, Rodney Boyer, Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.ISBN-10: 0-13-604302-X, ISBN-13: 978-0-13-604302-7

References

i. https://nios.ac.in/media/documents/dmlt/HC/Lesson-19.pdf

ii. https://microbenotes.com/electron-microscope-principle-types-components-applications-advantages-limitations/

iii. https://analyticalscience.wiley.com/do/10.1002/imaging.6389

iv. https://www.britannica.com/technology/scanning-electron-microscope

v. https://www.thermofisher.com/blog/materials/what-is-sem-scanning-electron-microscopy-explained/

vi. https://www.jove.com/v/5656/scanning-electron-microscopy-sem

vii. https://academic.oup.com/mt/article/32/2/37/7651167

viii. A Guide to Fixation and Staining of Biological Samples for Electron Microscopy, Rachel C Hart, Darian T Carroll, Maria Vinogradova, Evan S Krystofiak, *Microscopy Today*, Volume 32, Issue 2, March 2024, Pages 37–40, https://doi.org/10.1093/mictod/qaae009

ix. AM Glauert and PR Lewis, Biological Specimen Preparation for Transmission Electron Microscopy, (1999) Princeton University Press

x. AM Seligman et al., J Cell Biol 30 (1966) https://doi.org/10.1083/jcb.30.2.424.

xi. JR Guyton et al., Arteriosclerosis 5 (1985) https://doi.org/10.1161/01.ATV.5.6.644.

xii. AM Seligman et al., J Histochem Cytochem 13 (1965) https://doi.org/10.1177/13.8.629.

xiii. MJ Karnovsky, J Cell Biol 27 (1965).

xiv. https://www.deshbandhucollege.ac.in/pdf/resources/1588005857_LS-VI Practical_Study_of_Microscopic_techniques-1.pdf

xv. https://www.deshbandhucollege.ac.in/pdf/resources/1588005908_LS-VI-Practical_Study_of_Microscopic_techniques-2.pdf

xvi. https://users.ncsa.illinois.edu/kindr/phd/PART1.PDF

xvii. K. Kiss, Problem Solving with Microbeam Analysis, Elsevier, Budapest, 1988

xviii. L. Reimer, Scanning Electron Microscopy: Physics of Image Formation and Microanalysis, Springer-Verlag, Berlin, 1985

xix. Russ, Computer-Controlled Microscopy: The Measurement and Analysis of Images, Plenum Press, New York, 1990.

Unit-XIII
Flow Cytometry

Flow cytometry is a technology that simultaneously measures and then analyzes multiple physical characteristics of single particles, usually cells, as they flow in a fluid stream through a beam of light. The properties measured include a particle's relative size, relative granularity or internal complexity, and relative fluorescence intensity.

These characteristics are determined using an optical-to-electronic coupling system that records how the cell or particle scatters incident laser light and emits fluorescence.

A flow cytometer is made up of three main systems: **fluidics, optics, and electronics.**

- The fluidics system transports particles in a stream to the laser beam for interrogation.
- The optics system consists of lasers to illuminate the particles in the sample stream and optical filters to direct the resulting light signals to the appropriate detectors.
- The electronics system converts the detected light signals into electronic signals that can be processed by the computer. **For some instruments equipped with a sorting feature, the electronics system is also capable of initiating sorting decisions to charge and deflect particles.**

In the flow cytometer, particles are carried to the laser intercept in a fluid stream. Any suspended particle or cell from 0.2–150 micrometers in size is suitable for analysis.

Cells from solid tissue must be desegregated before analysis. The portion of the fluid stream where particles are located is called the sample core. When particles pass through the laser intercept, they scatter laser light. Any fluorescent molecules present on the particle fluoresce. The scattered and fluorescent light is collected by appropriately positioned lenses. A combination of beam splitters and filters steers the scattered and fluorescent light to the appropriate detectors. The detectors produce electronic signals proportional to the optical signals striking them. List mode data are collected on each particle or event. The characteristics or parameters of each event are based on its light scattering and fluorescent properties. The data are collected and stored in the computer. This data can be analyzed to provide information about subpopulations within the sample.

Flow cytometry is routinely used in basic research, clinical practice, and clinical trials. Uses for flow cytometry include:

- Cell counting
- Cell sorting
- Determining cell characteristics and function
- Detecting microorganisms
- Biomarker detection
- Protein engineering detection
- Diagnosis of health disorders such as blood cancers

Figure 13.1: Flow cytometer

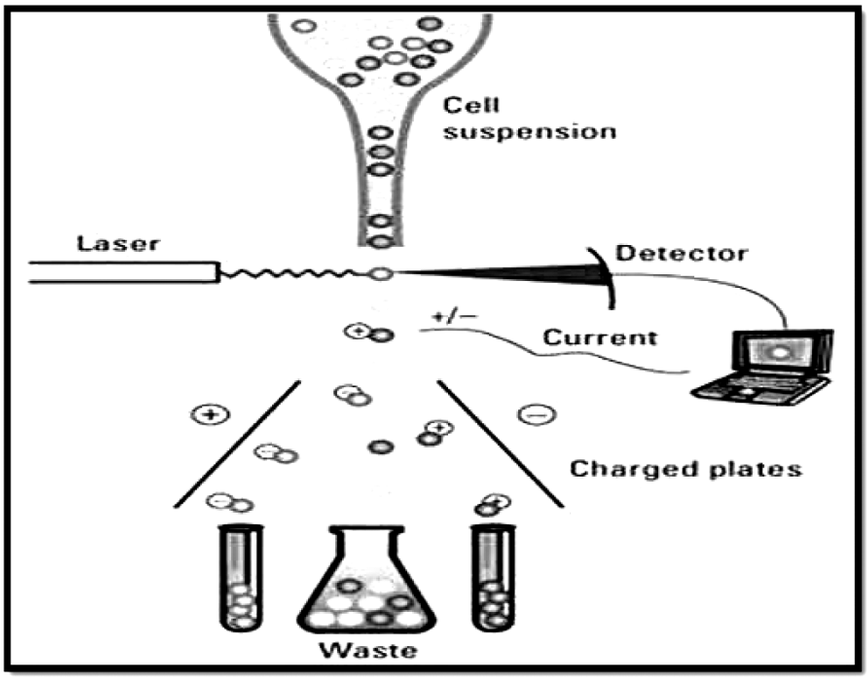


Figure 13.2: What happens inside a flow cytometer

Fluidics

The purpose of the fluidics system is to transport particles in a fluid stream to the laser beam for interrogation. For optimal illumination, the stream transporting the particles should be positioned in the center of the laser beam. In addition, only one cell or particle should move through the laser beam at a given moment. To accomplish this, the sample is injected into a stream of sheath fluid within the flow chamber. The flow chamber in a benchtop cytometer is called a flow cell and the flow chamber in a stream-in-air cytometer is called a nozzle tip. The design of the flow chamber causes the sample core to be focused in the center of the sheath fluid where the laser beam will then interact with the particles. Based on principles relating to laminar flow, the sample core remains separate but coaxial within the sheath fluid. The flow of sheath fluid accelerates the particles and restricts them to the center of the sample core. This process is known as hydrodynamic focusing.

The sample pressure and the sheath fluid pressure are different from each other. The sample pressure is always greater than the sheath fluid pressure. The sample pressure regulator controls the sample flow rate by changing the sample pressure relative to the sheath pressure.

Optics

The optics part deals with the laser light which falls on the particles and what happens to them as it strikes the single-file particle. Light scattering occurs when a particle deflects incident laser light. The extent to which this occurs depends on the physical properties of a particle, namely its size and internal complexity. Factors that affect light scattering are the cell's membrane, nucleus, and any granular material inside the cell. Cell shape and surface topography also contribute to the total light scatter. Forward-scattered light (FSC) is proportional to cell-surface area or size. FSC is a measurement of mostly diffracted light and is detected just off the axis of the incident laser beam in the forward direction by a photodiode (Figure). FSC provides a suitable method of detecting particles greater than a given size independent of their fluorescence and is therefore often used in immunophenotyping to trigger signal processing. Side-scattered light (SSC) is proportional to cell granularity or internal complexity. SSC is a measurement of mostly refracted and reflected light that occurs at any interface within the cell where there is a change in refractive index (Figure 1). SSC is collected at approximately 90 degrees to the laser beam by a collection lens and then redirected by a beam splitter to the appropriate detector.

The optical system consists of excitation optics and collection optics. The excitation optics consist of the laser and lenses that are used to shape and focus the laser beam. The collections optics consist of a collection lens to collect light emitted from the particle–laser beam interaction and a system of optical mirrors and filters to route specified wavelengths of the collected light to designated optical detectors. Optical Bench and optical Filters (Long pass filter and short pass filter) form the Optical system.

Electronics

Light signals are generated as particles pass through the laser beam in a fluid stream. These light signals are converted to electronic signals (voltages) by photodetectors and then assigned a channel number on a data plot. There are two types of photodetectors in BD flow cytometers: photodiodes and photomultiplier tubes (PMTs). The photodiode is less sensitive to light signals than the PMTs and thus is used to detect the stronger FSC signal. PMTs are used to detect the weaker signals generated by SSC and fluorescence. A voltage pulse is created when a particle enters the laser beam and starts to scatter light or fluoresce. Once the light signals, or photons, strike one side of the PMT or the photodiode, they are converted into a proportional number of electrons that are multiplied, creating a greater electrical current. The electrical current travels to the amplifier and is converted to a voltage pulse. The highest point of the pulse occurs when the particle is in the center of the beam and the maximum amount of scatter or fluorescence is achieved. As the particle leaves the beam, the pulse comes back down to the baseline.

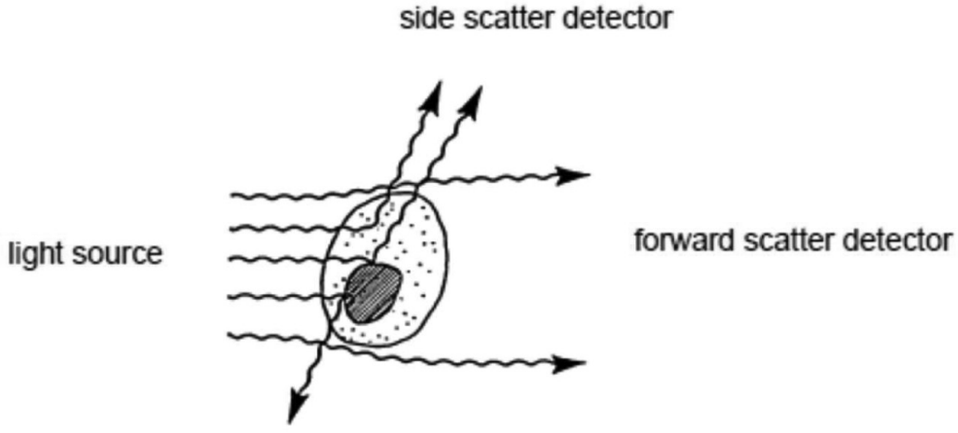


Figure 13.3: Light Scatter in a cell

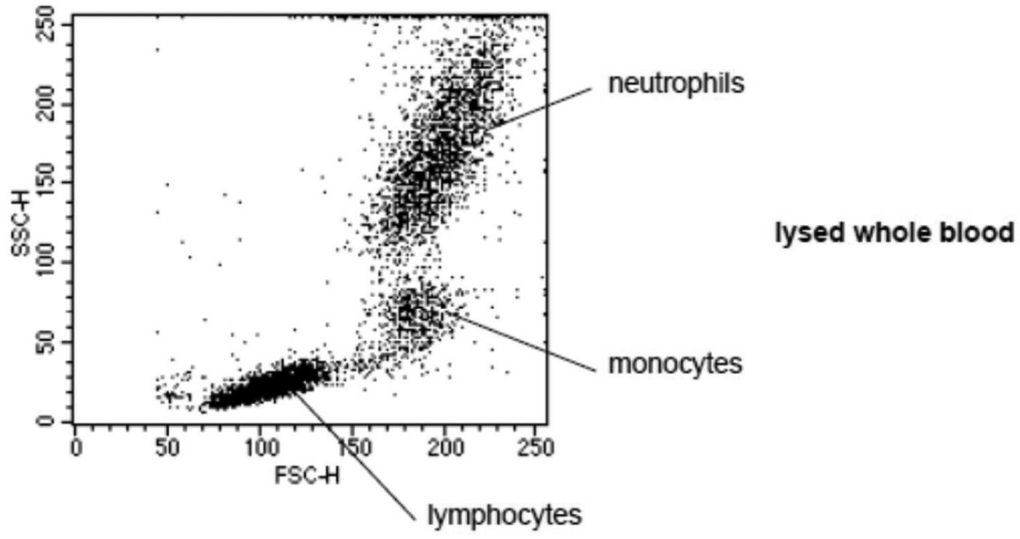


Figure 13.4: Cell population based on FSC and SSC

Data Collection and Display

Once light signals have been converted to electronic pulses and then converted to channel numbers by the ADC, the data must be stored by the computer system. Flow cytometric data is stored according to a standard format, the flow cytometry standard (FCS) format, developed by the Society for Analytical Cytology. According to the FCS standard, a data storage file includes a description of the sample acquired, the instrument on which the data was collected, the data set, and the results of data analysis. A single cell analyzed for four parameters (FSC, SSC, FITC, and PE fluorescence) generates 8 bytes of data. When multiplied by the approximately 10,000 events collected for a single sample, an FCS data file typically contains 80 kB of data. Once a data file has been saved, cell populations can be displayed in several different formats. A single parameter such as FSC or FITC (FL1) can be displayed as a single parameter histogram, where the horizontal axis represents the parameter's signal value in channel numbers and the vertical axis represents the number of events per channel number (Figure 3). Each event is placed in the channel that corresponds to its signal value. Signals with identical intensities accumulate in the same channel. Brighter signals are displayed in channels to the right of the dimmer signals. Two parameters can be displayed simultaneously in a plot. One parameter is displayed on the x-axis and the other parameter is displayed on the y-axis.

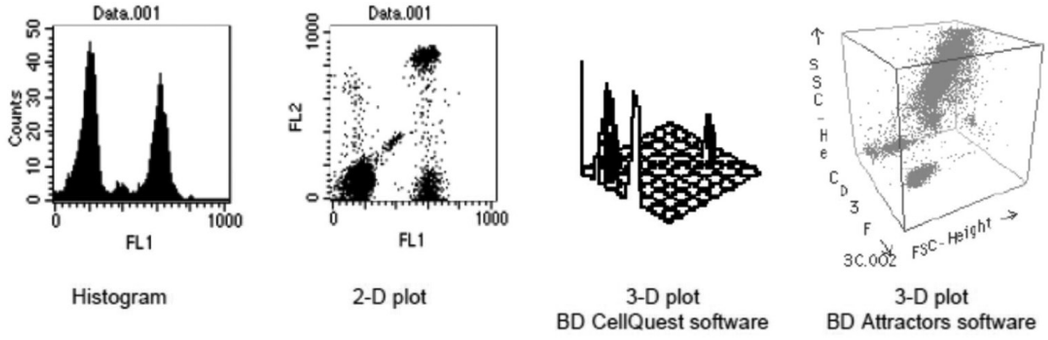


Figure 13.5: Different display plots

Correlated measurements of FSC and SSC can allow for differentiation of cell types in a heterogeneous cell population. Major leucocyte subpopulations can be differentiated using FSC and SSC (Figure 2).

Gating

The data generated by flow cytometer can be plotted in a single dimension, to produce a histogram, or in two-dimensional dot plots or even in three dimensions. The regions on these plots can be sequentially separated, based on fluorescence intensity, by creating a series of subset extractions, termed "gates." Specific gating protocols exist for diagnostic and clinical purposes especially in relation to hematology. Individual single cells are often distinguished from cell doublets or higher aggregates by their "time-of-flight" (denoted also as a "pulse-width") through the narrowly focused laser beam

The plots are often made on logarithmic scales. Because different fluorescent dyes' emission spectra overlap, signals at the detectors have to be compensated electronically as well as computationally. Data accumulated using the flow cytometer can be analyzed using software. Once the data is collected, there is no need to stay connected to the flow cytometer and analysis is most often performed on a separate computer. This is especially necessary in core facilities where usage of these machines is in high demand.

Fluorescence

- The narrow bandwidth of the lasers used to count the cells in flow cytometry lends the process easily to the use of fluorescence to further categorise the cells.
- Fluorescence is the emission of radiation following excitation by a higher energy of radiation. For example, when fluorescein isothiocyanate (FITC) is excited by

light of a wavelength of 495nm (blue light) it emits light at a wavelength of 519nm (green light).

- Fluorescent dyes can be conjugated to monoclonal antibodies which bind to proteins characteristic of certain cell types. This allows cells to be fluorescently labeled according to the proteins on their surfaces
- Many flow cytometers use an argon laser which produces light at 488nm. This not only excites FITC, but also several other fluorogens which emit light at different wavelengths. This means that several different labels can be detected if the correct filters are used in conjunction with the light detectors.

Data display

- The output of the detectors can be displayed in a number of different ways:
- *Single Channel* – generally expressed as a number of "events" (ie. cells) over the signal from a single detector. eg. if DNA is stained with a fluorescent marker such as propidium iodide, a population of cells may appear as:

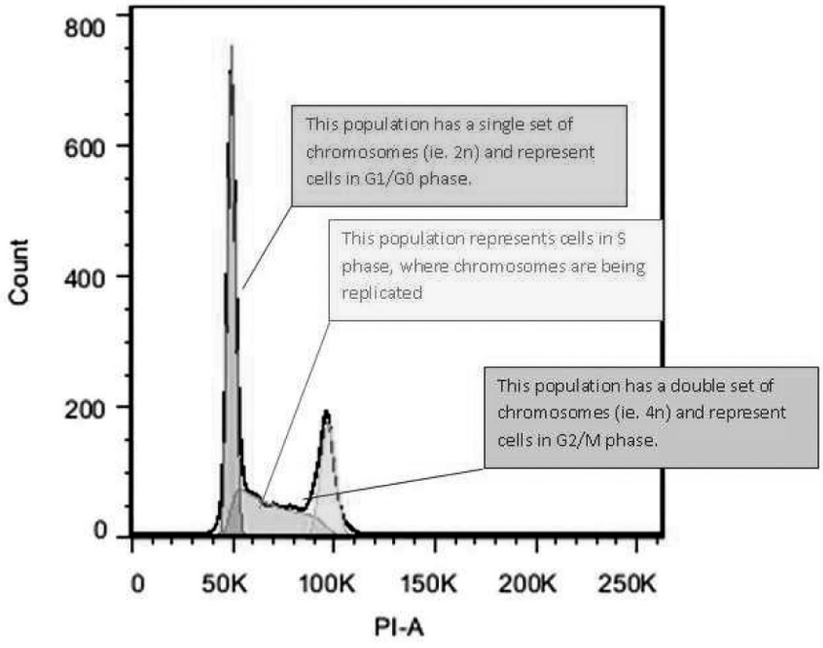


Figure 13.6: Histogram peaks of a Cell cycle analysis

- *2 Channel* – expressed as a plot of the output from one detector with a filter for a single wavelength, against that of a second detector with a filter for a different wavelength. eg. in the following plot, a population of immune cells have

been labelled with antibodies against the membrane protein CD3 (conjugated to FITC) and CD19 (conjugated to PE).

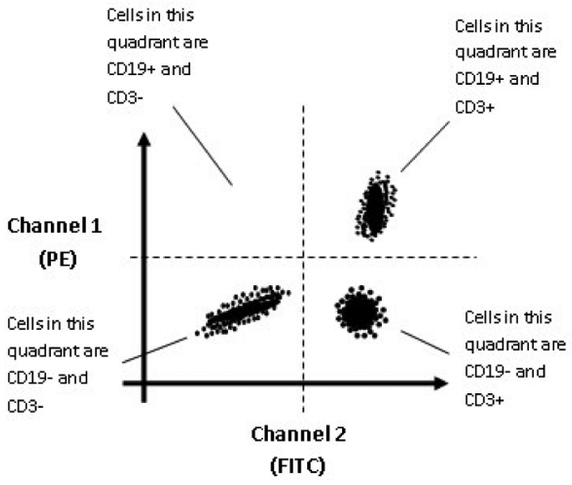


Figure 13.7: Dot-plot analysis

Cell sorting

- In conventional laser flow cytometry, cells which pass through the flow cell go to waste. In FACS (fluorescence assisted cell sorting), the characteristics of the cells determined in the flow cell may be used as a criteria to divert the cell to a collection chamber.
- In order to sort cells, a set of criteria (a "sorting gate") needs to be established which divided the cells into discrete groups. The four quadrants in the previous figure could be used to establish a sorting gate for these populations of cells.
- As the cells pass through the flow cell, signals from each of the detectors are passed through to the processor, which makes a decision based on the characteristics of the cell and how these fit into the established sorting gate.
- Depending on the model of the FACS unit, cell segregation may be achieved by:
    - Mechanical Means eg. a collecting tube swings into the flow stream to intercept the cells and divert them to the appropriate collection vessel.
    - Electrostatic Means eg. the flow stream is broken into a series of discrete drops through vibration, with each drop containing a single cell. Depending on the characteristics of the cell, each drop is given a charge, and charged plates are used to divert the drops to the collection vessel.

Applications of Flow Cytometry / FACS

1. **Immunophenotyping**
   - Identification and classification of immune cell subsets based on surface and intracellular markers.
   - Essential for diagnosing hematological malignancies (e.g., leukemia, lymphoma).
2. **Cell Cycle Analysis**
   - Quantification of DNA content using dyes like propidium iodide.
   - Assessment of cell populations in G0/G1, S, and G2/M phases.
   - Used in cancer research and drug response evaluation.
3. **Apoptosis and Cell Death Assays**
   - Detection of early and late apoptotic cells using markers like Annexin V and 7-AAD.
   - Differentiates necrosis from apoptosis.
4. **Intracellular Cytokine Staining**
   - Measures cytokine production at the single-cell level.
   - Important in immunology, vaccine development, and infection biology.
5. **Detection of Rare Cell Populations**
   - Identification and isolation of rare events like circulating tumor cells (CTCs), stem cells, or fetal cells in maternal blood.
   - Critical in cancer diagnostics and prenatal testing.
6. **Stem Cell Research and Sorting**
   - Characterization and purification of hematopoietic and mesenchymal stem cells.
   - Useful for regenerative medicine and cell therapy studies.
7. **Functional Assays**
   - Measurement of calcium flux, reactive oxygen species (ROS), phagocytosis, or cell proliferation.
   - Real-time functional analysis of immune responses or cellular stress.

8. **FACS-Based Cell Sorting**
    - High-purity sorting of live cells based on fluorescent markers.
    - Enables downstream applications like RNA sequencing, culture, or gene editing.
9. **Vaccine and Infection Studies**
    - Monitoring antigen-specific T-cell or B-cell responses.
    - Evaluating efficacy and immunogenicity of vaccines.
10. **Transfection and Gene Expression Studies**
    - Sorting transfected cells (e.g., GFP-positive) for gene function analysis.
    - Quantifying protein expression levels post-transfection.
11. **Cancer Immunotherapy Monitoring**
    - Tracking immune checkpoint expression (PD-1, CTLA-4).
    - Assessing immune cell infiltration in tumor microenvironments.
12. **Minimal Residual Disease (MRD) Detection**
    - High-sensitivity detection of residual malignant cells post-treatment.
    - Crucial for prognosis and treatment adjustment in leukemia.
13. **Microbiome and Pathogen Analysis**
    - Characterization and sorting of microbial populations using fluorescent in situ hybridization (FISH).
    - Used in environmental microbiology and infectious disease research.
14. **Allergy and Hypersensitivity Testing**
    - Basophil activation test (BAT) using flow cytometry to identify allergen-specific responses.
15. **Quality Control in Biomanufacturing**
    - Monitoring cell viability, identity, and phenotype in bioproduction processes (e.g., CAR-T manufacturing).

Probable Questions

1. Describe the principle of flow cytometry. How are cells analyzed and what physical characteristics are measured?
2. Explain the components and working of the fluidics system in a flow cytometer. How does hydrodynamic focusing aid in single-cell analysis?
3. Discuss the optical system in flow cytometry. Differentiate between forward scatter (FSC) and side scatter (SSC) and explain how they relate to cell properties.
4. Illustrate the role of fluorescence in flow cytometry. How do fluorochrome-conjugated antibodies facilitate cellular identification and classification?
5. Explain the process of signal detection and data processing in the electronics system of a flow cytometer. How are light signals converted into analyzable data?
6. What is gating in flow cytometry? Discuss its importance in identifying specific cell populations and elaborate on the concept of time-of-flight.
7. Compare and contrast histograms and dot plots in data visualization. How do these formats help in interpreting flow cytometric data?
8. Describe the mechanism of Fluorescence-Activated Cell Sorting (FACS). How are cells physically separated based on their fluorescence profiles?
9. Enumerate and explain at least five major applications of flow cytometry in clinical and research settings.
10. Discuss the significance of flow cytometry in immunophenotyping and minimal residual disease (MRD) detection. How do these applications contribute to diagnostics and therapeutic monitoring?

Unit XIV

Cell Fractionation Methods: i) Ultracentrifugation ii) Gradient Centrifugation

Objective: In this unit we will discuss about Cell Fractionation Methods which includes Ultracentrifugation and Gradient Centrifugation

Ultracentrifugation

Ultracentrifugation is a specialized technique used to spin samples at exceptionally high speeds. Current ultracentrifuges can spin to as much as 150 000 rotations per minute (rpm) (equivalent to 1 000 000 g) (Biocompare, 2019b). However, extreme centrifugal forces may cause overheating, so to avoid sample damage, ultracentrifuges are equipped with vacuum systems that keep a constant temperature in the centrifuge's rotor (Biocompare, 2019b).

Principle

- The working of an ultracentrifuge is based on the sedimentation principle, which states that the denser particles settle down faster when compared to less dense particles under gravity.
- However, the sedimentation of particles under gravity would take a larger amount of time, and that is why an additional force is applied to aid the sedimentation process.
- In an ultracentrifuge, the sample is rotated about an axis, resulting in a perpendicular force, called centrifugal force, that acts on different particles on the sample.
- The larger molecules move faster, whereas the smaller molecules move slower.
- At the same time, denser molecules are moved outwards to the periphery of the tubes whereas the less dense molecules are rotated towards the center of the tube.
- Once the process is completed, the larger and more dense particles settle down, forming pellets at the bottom of the tube. In comparison, the smaller

and less dense particles remain either in the suspended in the supernatant or float on the surface.

Parts of Ultracentrifuge

I. Ultracentrifuges are provided with a variety of parts and components perform different functions.

II. The rotors are an essential part of any ultracentrifuge. Ultracentrifuges use all three types of rotors, namely, vertical rotors, swinging bucket rotors, and fixed-angle rotors.

III. The swinging bucket rotor is the most commonly used rotor in ultracentrifuge because this yields the highest concentration of particles. This is because the direction of centrifugal force in vertical rotors aligns with the position of the tubes.

IV. The drive is the power unit which spins the rotor holding cells or tubes which contain the solution of molecules or suspension of particles.

V. Besides the interchangeable electric drive and rotors, it includes analytical rotors accommodating up to four cells, temperature devices with range and control from 0° to 40°, cells with thicknesses covering a factor of 10 in the sensitivity, and wedge quartz windows permitting several cells to be used simultaneously.

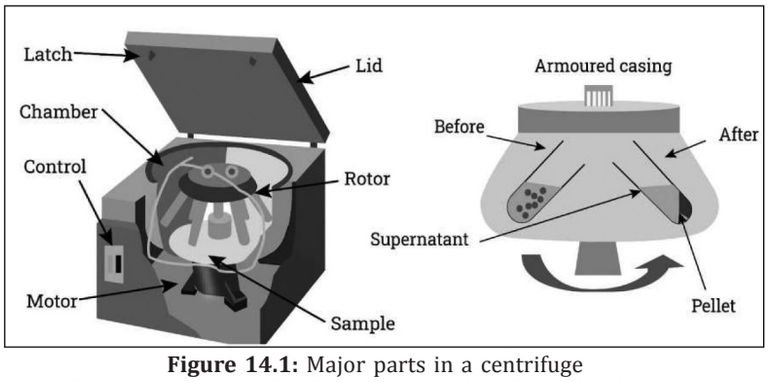


Figure 14.1: Major parts in a centrifuge
Source: https://www.iqsdirectory.com/articles/centrifuge.html

VI. The temperature system allows the control of temperature in the system as heat production is common during the operation of ultracentrifuge at high speeds.

VII. Analytical ultracentrifuge also has a two-dimensional comparator with a printing desk calculator, interference, and absorption optical systems. The optical systems are necessary for the real-time analysis of molecules.

VIII. Additionally, a gradient-forming device, hand refractometers, and a recording spectrophotometer with flow cell and fraction collector might also be present in a preparative ultracentrifuge.

Procedure/ Steps of Ultracentrifuge

I. The process of operating an ultracentrifuge might differ depending on whether it is an analytical ultracentrifuge or a preparative ultracentrifuge. In general, the following are the steps to be followed while performing analytical ultracentrifugation:

II. Small sample sizes (20-120 mm3) are taken in analytical cells to be placed inside the ultracentrifuge.

III. The ultracentrifuge is then operated so that the centrifugal force causes a migration of the randomly distributed biomolecules through the solvent radially outwards from the center of rotation.

IV. The distance of the molecules from the center is determined through the Schlieren optical system.

V. A graph is drawn from the solute concentration versus the squared radial distance from the center of rotation, based on which the molecular mass is determined.

The following is the procedure for the operation of a preparative ultracentrifuge:

1. Density gradient centrifugation

- ✓ In preparative ultracentrifuge using density gradient, a density gradient has to be prepared. For this, a layer of less concentrated sucrose is applied over a layer of more concentrated sucrose which creates a sucrose density gradient. Some ultracentrifuges come with gradient-forming devices that can make the gradient by themselves.
- ✓ Then the sample is placed in the centrifuge tube over the gradient, and then the tubes are placed in the racks of the rotors of the ultracentrifuges.
- ✓ The temperature and time are set in the ultracentrifuge before the process is started.
- ✓ The lid is closed, and the process is started.
- ✓ The particles travel through the gradient until they reach a point at which their density matches the density of the surrounding medium.

- ✓ The fractions are removed and separated, obtaining the particles as isolated units.

2. Differential centrifugation

- ✓ In differential centrifugation, the sample is homogenized in the medium containing buffer.
- ✓ The sample is then placed in the centrifuge tube, which is operated at a particular centrifugal force for a specific time at a particular temperature.
- ✓ By the end of this operation, a pellet will be formed at the bottom of the tube, which is separated from the supernatant.
- ✓ The supernatant is added to a new centrifuge tube where it is centrifuged at another speed for a particular time and particular temperature.
- ✓ Again, the supernatant is separated from the pellets formed.
- ✓ These steps are continued until all particles are separated from each other.
- ✓ The particles can then be identified by testing for indicators that are unique to the specific particles.

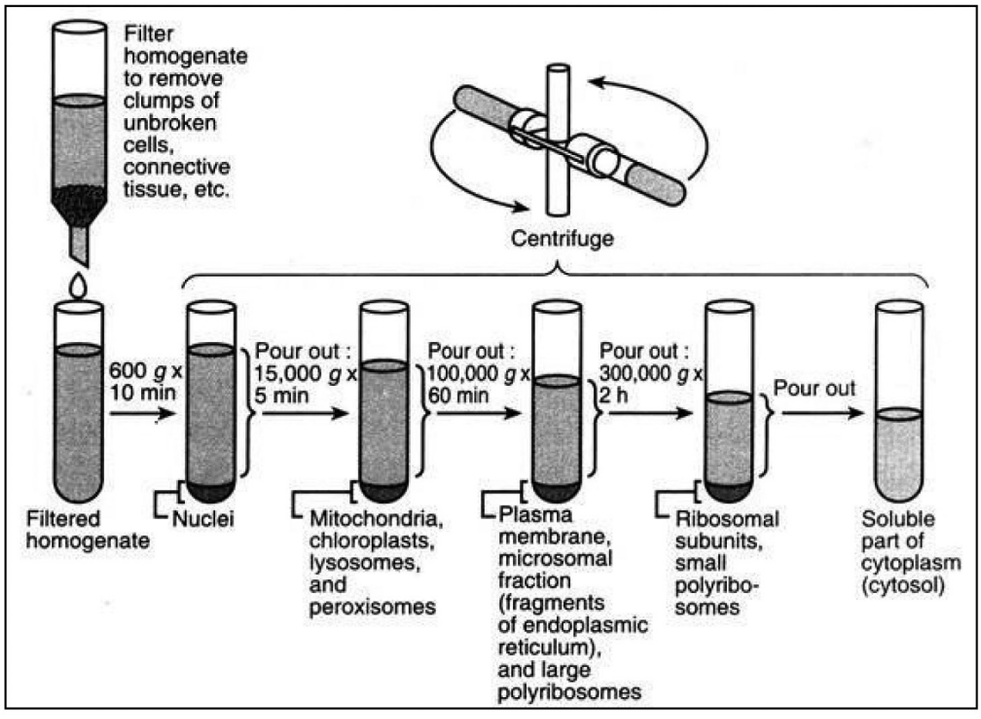


Fig. 14.2: Differential centrifugation
Source: https://conductscience.com/ultracentrifugation-basics-and-applications/?srsltid=AfmBOoowYdaN27yteB6KgrCwc4KOrxKrkRAh4h2rvsAzW0kyTMNPJBCd

Advantages of Ultracentrifugation

1. Ultracentrifugation is applied to biology, chemistry, material science, and others fields of sciences.
2. Ultracentrifugation extended the limits of biology research to the subcellular level, by allowing the isolation of particles as small as ribosomes, subcellular organelles, membranes, and ribonucleic acids.
3. With the advent of analytical ultracentrifugation, sub microscopic world is also discovered. It helps to characterize molecular size, shape, and structure.

Limitations of Ultracentrifugation

Ultracentrifugation has its own limitations. These include:

1. Low sample yield – In preparative ultracentrifugation, samples must be washed several times after spinning, to ensure that there is no cross-contamination between fractions. Samples for preparative centrifugation are usually limited in size (e.g., tissues) or volume (e.g., cell suspensions or blood). In every wash step that a sample is subjected to, there is a loss of material, and thus, after an ultracentrifugation protocol, the yield can be very low.
2. Ultracentrifugation is still a time-consuming process, and it can take up to several hours to fractionate all the components of a single mixture.
3. Ultracentrifuges are extremely expensive devices, which require constant maintenance and efficient laboratory technician.

Gradient Centrifugation

Density gradient centrifugation is a centrifugation process where samples are placed into a centrifuge, but the end goal is not to sort them by size. Spinning from the centrifuge causes more dense particles to move to the bottom of the tube because these particles have more mass and are carried further by their inertia. Less dense particles then settle higher within the sample. This process creates a sorted solution layered by particle density from least to most, top to bottom.

This method was used in the famous experiment, which proved that DNA is semi-conservative by using different isotopes of nitrogen.

Principles of Density Gradient Centrifugation

Each particle has a specific set of physical characteristics; the properties of its biological makeup that can be used for separation and isolation. Density gradient centrifugation focuses on two characteristics—size and density.

The length of time required for this process depends upon the particles' size. Larger particles will reach their position of stability earlier, whereas smaller particles take longer to pass through the larger particle zone and assume a position deeper in the gradient.

Uses of Density gradient centrifugation

- Density gradient centrifugation can be applied for the purification of large volumes of biomolecules.
- It can even be used for the purification of different viruses which aids their further studies.
- This technique can be used both as a separation technique and the technique for the determination of densities of various particles.

Probable questions:

1. What is ultracentrifugation?
2. State the principle of ultracentrifugation.
3. Name the parts of ultracentrifugation. Describe the ultracentrifugation structure.
4. What is the function of rotor in ultracentrifugation?
5. Describe the operating process of Density gradient centrifugation.
6. Mention the limitations of ultracentrifugation.
7. What are the advantages of ultracentrifugation application?
8. Describe the principle of Density gradient centrifugation
9. State the uses of Density gradient centrifugation.

Suggested Reading:

1. Wilson and Walker's Principles and Techniques of Biochemistry and Molecula Biology 2018, ISBN: 9781316677056

2. Biochemistry Laboratory: Modern Theory and Techniques, Latest Edition, Rodney Boyer, Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.ISBN-10: 0-13-604302-X, ISBN-13: 978-0-13-604302-7

References

1. https://microbenotes.com/ultracentrifuge/
2. (2019a). Laboratory Centrifuges. Retrieved November, 2019, from https://www.biocompare.com/Lab-Equipment/Laboratory-Centrifuges/
3. (2019b). Ultracentrifuges. from https://www.biocompare.com/Lab-Equipment/10155-Benchtop-Ultracentrifuge/
4. https://conductscience.com/ultracentrifugation-basics-and-applications/?srsltid=AfmBOoowYdaN27yteB6KgrCwc4KOrxKrkRAh4h2rvsAzW0kyTMNPJBCd
5. https://microbenotes.com/ultracentrifuge/
6. https://www.akadeum.com/blog/density-gradient-centrifugation/?srsltid=AfmBOoqQ8NuZUv20Dn2ivNhCYA2VC6qbPJbSqhEPJWPxohY2C8FZBS-b&cn-reloaded=1

UNIT XV

Separation of Cell Constituents: i) Chromatography: ion exchange; gel filtration and HPLC. ii) Electrophoresis-PAGE, SDS-PAGE (One and Two dimensional)

Objective: In this unit we will discuss about separation of cell constituents which includes Chromatography (ion exchange; gel filtration and HPLC) and Electrophoresis-PAGE, SDS-PAGE (One and Two dimensional).

What is chromatography?

Chromatography is a technique for separating the components, or solutes, of a mixture on the basis of the relative amounts of each solute distributed between a moving fluid stream, called the *mobile phase*, and a contiguous *stationary phase*. The mobile phase may be either a liquid or a gas, while the stationary phase is either a solid or a liquid.

The discovery of chromatography, however, is generally attributed to the Russian botanist Mikhail S. Tsvet (also spelled Tswett), because in 1901 he recognized the physicochemical basis of the separation and applied it in a rational and organized way to the separation of plant pigments, particularly the carotenoids and the chlorophylls.

Principle of Chromatography

•Distribution of analytes between phases:

The distribution of analytes between phases can often be described quite simply. An analyte is in equilibrium between the two phases;

$$A_{mobile} \leftrightarrow A_{stationary}$$

The equilibrium constant, K, is termed the partition coefficient, defined as the molar concentration of analyte in the stationary phase divided by the molar concentration of the analyte in the mobile phase. The time between sample injection and an analyte peak reaching a detector at the end of the column is termed the retention time (t_R). Each analyte in a sample will have a different retention time. The time taken for the mobile phase to pass through the column is called t_M.

A term called the retention factor, k', is often used to describe the migration rate of an analyte on a column. You may also find it called the capacity factor. The retention factor for analyte A is defined as

$$k_A' = (t_R - t_M)/t_M.$$

t_R and t_M are easily obtained from a chromatogram. When an analytes retention factor is less than one, elution is so fast that accurate determination of the retention time is very difficult. High retention factors (greater than 20) mean that elution takes a very long time. Ideally, the retention factor for an analyte is between one and five.

We define a quantity called the selectivity factor, α, which describes the separation of two species (A and B) on the column

$$a = k_B'/k_A'$$

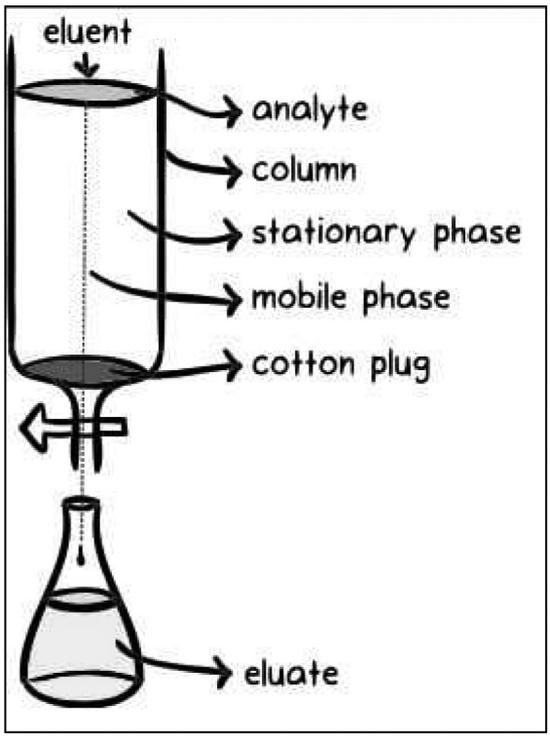


Fig. 15.1
Image Source: Khan Academy

When calculating the selectivity factor, species A elutes faster than species B. The selectivity factor is always greater than one.

Three components thus form the basis of the chromatography technique.

1. **Stationary phase:** This phase is always composed of a "solid" phase or "a layer of a liquid adsorbed on the surface solid support".

2. **Mobile phase:** This phase is always composed of "liquid" or a "gaseous component."
3. **Separated molecules** The type of interaction between the stationary phase, mobile phase, and substances contained in the mixture is the basic component effective on the separation of molecules from each other.

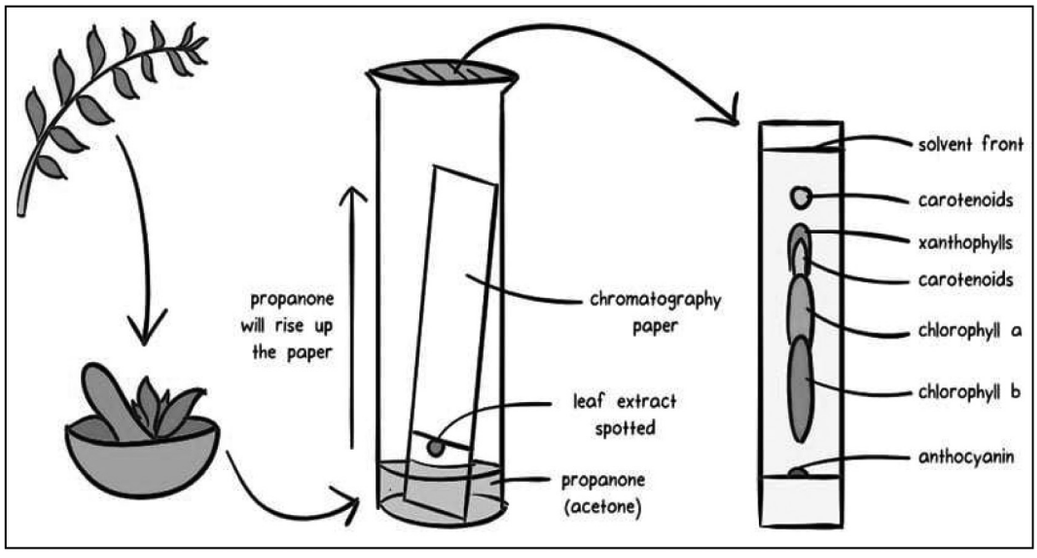


Image Source: Khan Academy

Fig. 15.2

Various technical terminologies of use in the topic:

i. The analyte is the substance which is to be separated during chromatography.
ii. Analytical chromatography is used to determine the existence and possibly also the concentration of analyte(s) in a sample.
iii. A bonded phase is a stationary phase that is covalently bonded to the support particles or to the inside wall of the column tubing.
iv. A chromatogram is the visual output of the chromatograph. In the case of an optimal separation, different peaks or patterns on the chromatogram correspond to different components of the separated mixture.
v. A chromatograph is equipment that enables a sophisticated separation, e.g., gas chromatographic or liquid chromatographic separation.
vi. The effluent is the mobile phase leaving the column.
vii. An immobilized phase is a stationary phase which is immobilized on the support particles, or on the inner wall of the column tubing.

viii. The mobile phase is the phase which moves in a definite direction. It may be a liquid (LC and CEC), a gas (GC), or a supercritical fluid (supercritical-fluid chromatography, SFC).

ix. Preparative chromatography is used to nondestructively purify sufficient quantities of a substance for further use, rather than analysis.

x. The retention time is the characteristic time it takes for a particular analyte to pass through the system (from the column inlet to the detector) under set conditions.

xi. The sample is the matter analysed in chromatography. It may consist of a single component or it may be a mixture of components. When the sample is treated in the course of an analysis, the phase or the phases containing the analytes of interest is/are referred to as the sample whereas everything out of interest separated from the sample before or in the course of the analysis is referred to as waste.

xii. The solute refers to the sample components in partition chromatography.

xiii. The solvent refers to any substance capable of solubilizing other substance, especially the liquid mobile phase in LC.

Applications of Chromatography

Pharmaceutical sector

- To identify and analyze samples for the presence of trace elements or chemicals.
- Separation of compounds based on their molecular weight and element composition.
- Detects the unknown compounds and purity of mixture.
- In drug development.

Chemical industry

- In testing water samples and also checks air quality.
- HPLC and GC are very much used for detecting various contaminants such as polychlorinated biphenyl (PCBs) in pesticides and oils.
- In various life sciences applications.

Food Industry

- In food spoilage and additive detection
- Determining the nutritional quality of food

Forensic Science

- In forensic pathology and crime scene testing like analyzing blood and hair samples of crime place.

Molecular Biology Studies

- Various hyphenated techniques in chromatography such as EC-LC-MS are applied in the study of metabolomics and proteomics along with nucleic acid research.
- HPLC is used in Protein Separation like Insulin Purification, Plasma Fractionation, and Enzyme Purification and also in various departments like Fuel Industry, biotechnology, and biochemical processes.

Ion exchange chromatography

Introduction

Ion exchange chromatography can be defined as a reversible process in which ions of similar charged are exchanged between solid and liquid. The solid is known as an ion exchanger. It is an adsorption chromatography, a useful and popular method for separation of a mixture of similar charged substances into pure components. It is also known as cation anion exchange chromatography.

In this process, two types of exchangers i.e., cationic and anionic exchangers can be used.

1. **Cationic exchangers** possess negatively charged group, and these will attract positively charged cations. These exchangers are also called "Acidic ion exchange" materials, because their negative charges result from the ionization of acidic group.
2. **Anionic exchangers** have positively charged groups that will attract negatively charged anions. These are also called "Basic ion exchange" materials.

- Ion exchange chromatography is most often performed in the form of column chromatography. However, there are also thin-layer chromatographic methods that work basically based on the principle of ion exchange.

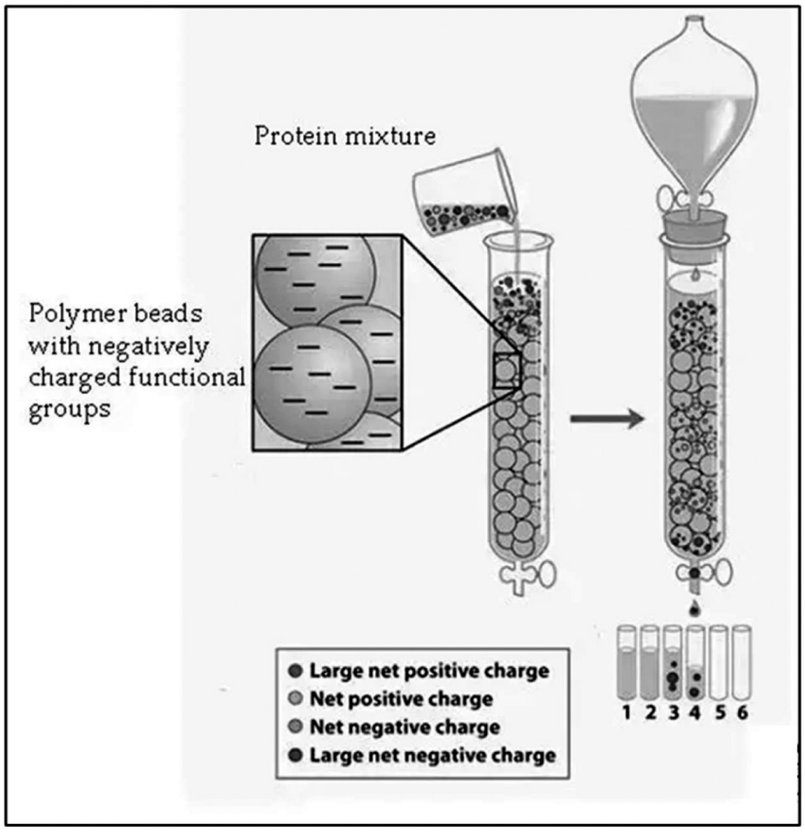


Fig. 15.3 : Ion exchange chromatography technique

Source: Lehninger Principle of Biochemistry, 5th Edition @2008

Principle of ion exchange chromatography

This form of chromatography relies on the attraction between oppositely charged stationary phase, known as an ion exchanger, and analyte. The ion exchangers basically contain charged groups covalently linked to the surface of an insoluble matrix. The charged groups of the matrix can be positively or negatively charged. When suspended in an aqueous solution, the charged groups of the matrix will be surrounded by ions of the opposite charge. In this "ion cloud", ions can be reversibly exchanged without changing the nature and the properties of the matrix.

Procedure of ion exchange chromatography

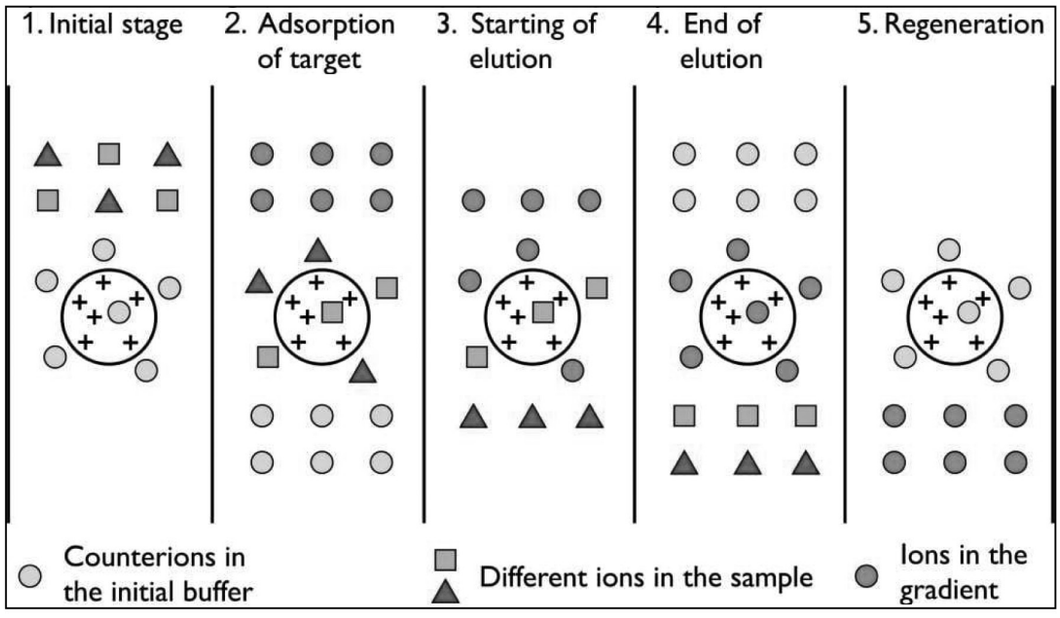


Fig. 15.4

- Ion exchange separations are carried out mainly in columns packed with an ion-exchanger.
- These ionic exchangers are commercially available. They are made up of styrene and divinyl benzene. Example. DEAE-cellulose is an anionic exchanger, CM-cellulose is a cationic exchanger.
- The choice of the exchanger depends upon the charge of particle to be separated. To separate anions "Anionic exchanger" is used, to separate cations "Cationic exchanger" is used.
- First the column is filled with ion exchanger then the sample is applied followed by the buffer. The tris-buffer, pyridine buffer, acetate buffer, citrate and phosphate buffers are widely used.
- The particles which have high affinity for ion exchanger will come down the column along with buffers.
- In next step using corresponding buffer separates the tightly bound particles.
- Then these particles are analyzed spectroscopically.

Instrumentation of ion exchange chromatography

Typical IC instrumentation includes: pump, injector, column, suppressor, detector and recorder or data system.

1. Pump

The IC pump is considered to be one of the most important components in the system which has to provide a continuous constant flow of the eluent through the IC injector, column, and detector.

2. Injector

Sample introduction can be accomplished in various ways. The simplest method is to use an injection valve. Liquid samples may be injected directly and solid samples need only to be dissolved in an appropriate solvent. Injectors should provide the possibility of injecting the liquid sample within the range of 0.1 to 100 ml of volume with high reproducibility and under high pressure (up to the 4000 psi).

3. Columns

Depending on its ultimate use and area of application, the column material may be stainless steel, titanium, glass or an inert plastic such as PEEK. The column can vary in diameter from about 2mm to 5 cm and in length from 3 cm to 50 cm depending on whether it is to be used for normal analytical purposes, microanalysis, high speed analyses or preparative work.

Guard column is placed anterior to the separating column. This serves as a protective factor that prolongs the life and usefulness of the separation column. They are dependable columns designed to filter or remove particles that clog the separation column.

4. Suppressor

The suppressor reduces the background conductivity of the chemicals used to elute samples from the ion-exchange column which improves the conductivity measurement of the ions being tested. IC suppressors are membrane-based devices which are designed to convert the ionic eluent to water as a means of enhancing the sensitivity.

5. Detectors

Electrical conductivity detector is commonly use.

6. Data system

In routine analysis, where no automation is needed, a pre-programmed computing integrator may be sufficient. For higher control levels, a more intelligent device is necessary, such as a data station or minicomputer.

Applications

1. Softening of hard water: Hardness of water is due to the presences of Ca^{2+}, Mg^{2+} and other divalent ions may be removed by passing the hard water through the cation exchanger charged with Na+ ions.
2. Purification of organic compounds: Many natural products extracted in water have been found to contain ions originally present in water. Those ions can be removed by using ion exchange process.
3. Ion exchange methods can be used to separate the complex mixture of 18 amino acids obtained by the acid hydrolysis of proteins.
4. It is used for purification and recovery of antibiotics, vitamins, alkaloids, hormones and other chemicals of pharmaceutical importance during their manufacturing process.
5. Biochemical separations like some drugs or metabolites from blood, urine or other biological fluids.
6. It is used for the measurement of drugs and their metabolites in serum and urine, for residue analysis in food raw material.
7. Additives such as vitamins and preservatives in food and beverages are also measured by this process.

Gel filtration Chromatography

Introduction

Gel filtration chromatography is used to separate proteins, peptides and oligo-nucleotides on the basis of their size only. It is also known as size exclusion chromatography. In gel filtration, as sample moves through a bed of porous beads, molecules of different size diffuse into the beads to greater or lesser extent. Small molecules enter further into the beads and thus, elute slowly. In comparison to that, large molecules enter less or move directly through the void space and thus, eluting fast. Gel Filtration Chromatography is widely applied for molecular size analysis, separation of components of a mixture, as well as for desalting and buffer exchange.

Principle

The separation of molecules on the basis of their molecular size and shape is achieved by gel filtration chromatography. It uses the molecular sieve properties of

various porous resins. Large molecules that are completely excluded from the pores pass through the void space, interstitial spaces between the resin particles and thus, they elute first. While smaller molecules get distributed between the mobile phase inside and outside the beads and therefore pass through the column at a slower rate. Thus, they elute in the last. Trapping of the mobile phase inside the beads solely depends upon the porosity of the resin beads and the size of the molecules to be separated (figure 1). Thus, the distribution of a molecule in a gel filtration column of crosslinked beads is determined by the total volume of mobile phase, both inside and outside the beads. In case of gel filtration chromatography, the distribution coefficient of a molecule between the inner and outer mobile phase is a function of its molecular size. If a molecule is sufficiently large to completely exclude from the mobile phase between the beads, then K = 0. On the other hand, if a molecule is so small that it accesses to the innermost mobile phase deep in the beads, then K = 1. For others, value of K will vary between 0 and 1. This variation of K between 0 and 1 makes it possible to separate various molecules present in a mixture within a narrow molecular size range. Figure 1: Gel filtration chromatography

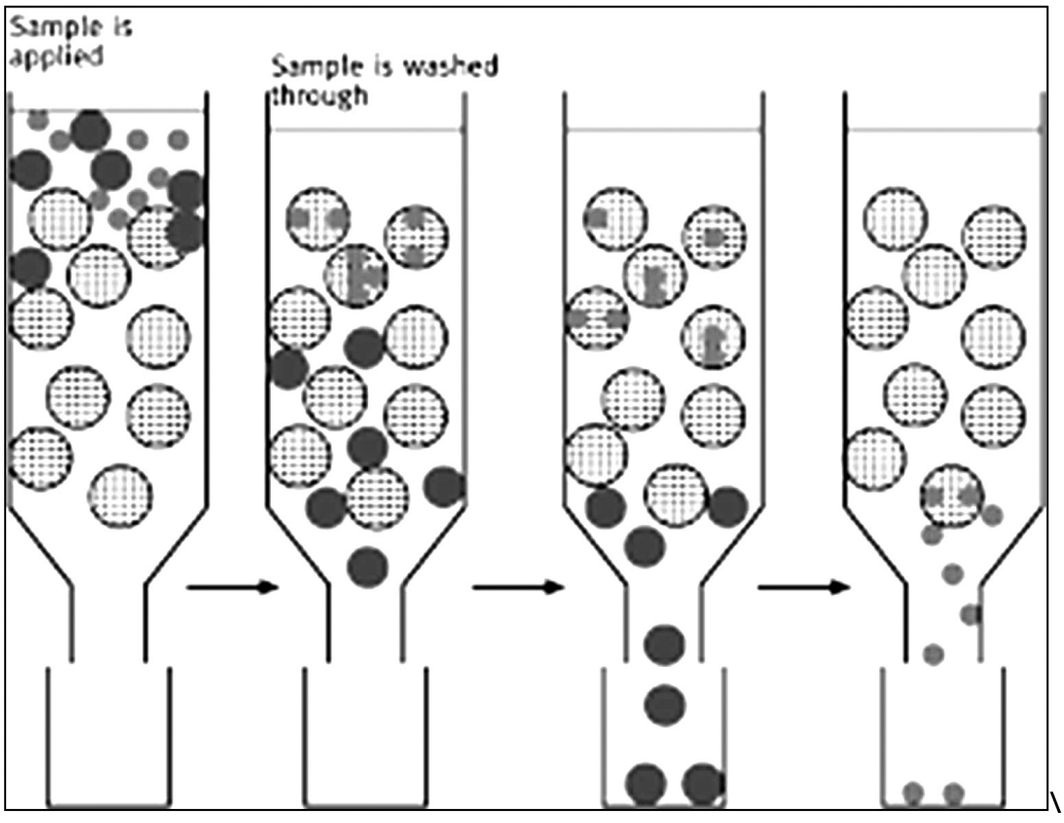


Fig. 15.5: Gel filtration chromatography [Source: MBL Life Science]

Steps in Gel Filtration Chromatography

I. Spherical particles of gel filtration medium are packed into a column.

II. The sample is applied to the column.

III. Buffer (mobile phase) and sample move through the column.

IV. Molecules diffuse in and out of the pores of the matrix (also described as the partitioning of the sample between the mobile phase and the stationary phase).

V. Smaller molecules move further into the matrix and so stay longer on the column.

VI. As buffer passes continuously through the column, molecules that are larger than the pores of the matrix are unable to diffuse into the pores and pass through the column.

VII. Smaller molecules diffuse into the pores and are delayed in their passage down the column.

VIII. Separation occurs at different intervals which are followed by detection of components.

Applications of Gel Filtration Chromatography

i. Gel filtration plays a key role in the purification of enzymes, polysaccharides, nucleic acids, proteins, and other biological macromolecules.

ii. Gel filtration can also be used to facilitate the refolding of denatured proteins by careful control of changing buffer conditions.

iii. It is used in protein fractionation experiments.

iv. Gel filtration technique is also used in molecular weight determination.

v. Separation of sugar, proteins, peptides, rubbers, and others on the basis of their size.

vi. Can be used to determine the quaternary structure of purified proteins.

HPLC (High-performance liquid chromatography)

Introduction:

The analytical technique of High-Performance Liquid Chromatography (HPLC) is used extensively throughout the pharmaceutical industry. It is used to provide information on the composition of drug-related samples (McPolin, 2009). It can separate, identify, and quantify the compounds that are present in any sample that can dissolve in liquid. HPLC is the most accurate analytical method widely used for the quantitative as well as qualitative analysis of drug products and used for the determination of drug product stability (Bhardwaj, 2015). High pressure should apply to have an eluent flow through the column due to the physical properties of HPLC (Varhadi, 2020). HPLC is recognized from traditional liquid chromatography because operational pressures are fundamentally higher (50bar to 350 bar). The small sample amount of isolated in scientific HPLC, column section measurements are 2.1nm to 4.6nm distance across, and 30 nm to 250nm in length. Additionally, the HPLC segment is made with the smaller sorbent particles (2µm to 50µm in a normal molecule size). This gives HPLC determine high resolving power (the capacity is recognized components) while isolating mixtures which makes its prominent chromatographic method (Varhadi, 2020).

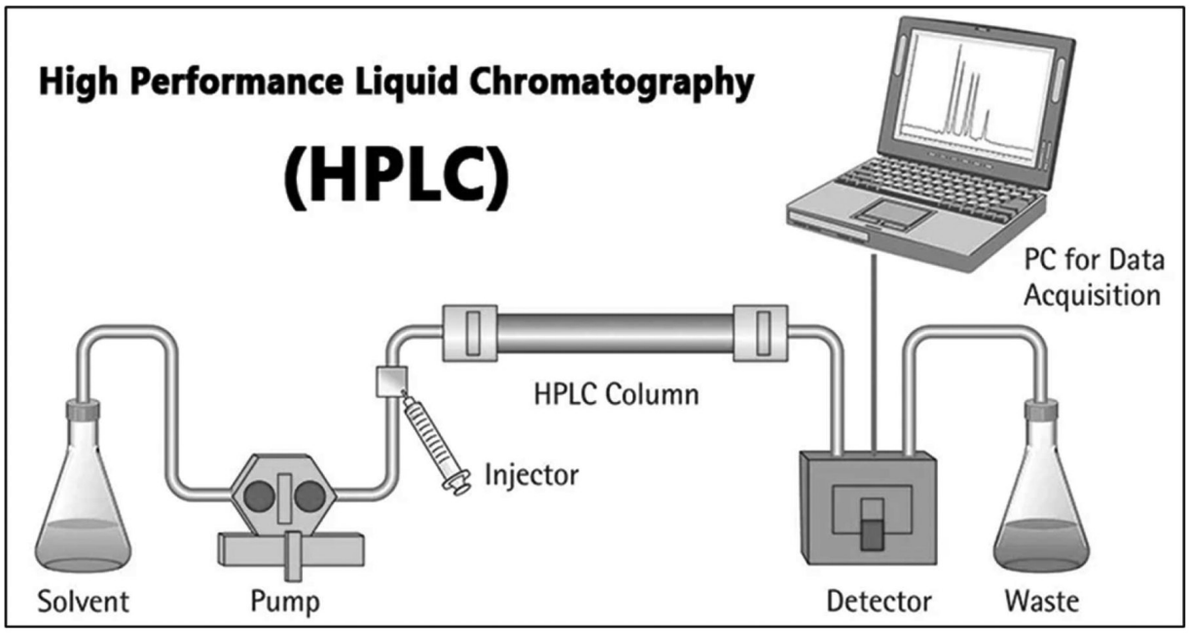


Fig. 15.6

Source: Sartorius AG.

HPLC Principle

- The purification takes place in a separation column between a stationary and a mobile phase.
- The stationary phase is a granular material with very small porous particles in a separation column.
- The mobile phase, on the other hand, is a solvent or solvent mixture which is forced at high pressure through the separation column.
- Via a valve with a connected sample loop, i.e. a small tube or a capillary made of stainless steel, the sample is injected into the mobile phase flow from the pump to the separation column using a syringe. Subsequently, the individual components of the sample migrate through the column at different rates because they are retained to a varying degree by interactions with the stationary phase.
- After leaving the column, the individual substances are detected by a suitable detector and passed on as a signal to the HPLC software on the computer.
- At the end of this operation/run, a chromatogram in the HPLC software on the computer is obtained. The chromatogram allows the identification and quantification of the different substances.

Applications of HPLC

The HPLC has developed into a universally applicable method so that it finds its use in almost all areas of chemistry, biochemistry, and pharmacy.

- Analysis of drugs
- Analysis of synthetic polymers
- Analysis of pollutants in environmental analytics
- Determination of drugs in biological matrices
- Isolation of valuable products
- Product purity and quality control of industrial products and fine chemicals
- Separation and purification of biopolymers such as enzymes or nucleic acids
- Water purification
- Pre-concentration of trace components
- Ligand-exchange chromatography

- Ion-exchange chromatography of proteins
- High-pH anion-exchange chromatography of carbohydrates and oligosaccharides

Limitations of HPLC

- Cost: Despite its advantages, HPLC can be costly, requiring large quantities of expensive organics.
- Complexity
- HPLC does have low sensitivity for certain compounds, and some cannot be detected as they are irreversibly adsorbed.
- Volatile substances are better separated by gas chromatography.

Electrophoresis

Introduction

The term electrophoresis describes the migration of a charged particle under the influence of electric field (electro-charged particle and phoresis-movement). This technique is used to separate DNA fragments (or other macromolecules, such as RNA and proteins) based on their size and charge. Electrophoresis involves running a current through a gel containing the molecules of interest. Based on their size and charge, the molecules will travel through the gel in different directions or at different speeds, allowing them to be separated from one another.

The only difference with respect to SDS-PAGE is that PAGE does not uses detergent for denaturating protein before loading on to gel. PAGE takes place under non-denaturating condition and in native gels. Separation of proteins is based on electrophoretic mobilities of protein in gels and sieving effect of gels.

PAGE (Polyacrylamide gel electrophoresis)

Polyacrylamide gel electrophoresis (PAGE) is a powerful tool widely used in biochemistry, forensic chemistry, genetics, molecular biology and biotechnology to separate biological macromolecules, usually proteins or nucleic acids, according to their electrophoretic mobility. Polyacrylamide gel with small pores helps to examine smaller molecules better since the small molecules can enter the pores and travel through the gel while large molecules get trapped at the pore openings.

The most commonly used form of polyacrylamide gel electrophoresis is the Sodium Dodecyl Sulphate (SDS) Polyacrylamide gel electrophoresis (SDS-PAGE) used mostly for the separation of proteins. SDS–polyacrylamide gel electrophoresis (SDS–PAGE) is the most widely used method for analysing protein mixtures qualitatively. It is particularly useful for monitoring protein purification and, because 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.

Principle:

When a potential difference is applied across the electrodes of a horizontal electrophoretic tank containing agarose gel and biomolecules (such as nucleic acids) are loaded, then they get separated according to their molecular size (bigger molecules have more molecular size and smaller molecules have small molecular size) and move to their respective electrodes. Here the agarose gel acts as a sieve.

As in a sieve the large particles stay above and the particles which are smaller than the pore size passes through it, similarly in the gel the larger and the bulky molecules stay behind whereas the smaller molecules move faster and quickly towards their respective electrodes.

This process may be imagined like a running competition. The one who is thinner and have a flexible body will be at the ending point sooner than the one who is fat and bulky.

Steps Involved in Polyacrylamide Gel Electrophoresis (PAGE)

1. Sample preparation

Samples may be any material containing proteins or nucleic acids.

- The sample to analyze is optionally mixed with a chemical denaturant if so desired, usually SDS for proteins or urea for nucleic acids.
- SDS is an anionic detergent that denatures secondary and non–disulfide–linked tertiary structures, and additionally applies a negative charge to each protein in proportion to its mass. Urea breaks the hydrogen bonds between the base pairs of the nucleic acid, causing the constituent strands to anneal. Heating the samples to at least 60 -95° C further promotes denaturation.

- A tracking dye may be added to the solution. This typically has a higher electrophoretic mobility than the analytes to allow the experimenter to track the progress of the solution through the gel during the electrophoretic run.

2. Preparation of polyacrylamide gel

- The gels typically consist of acrylamide, bis acrylamide, the optional denaturant (SDS or urea), and a buffer with an adjusted pH.
- The ratio of bisacrylamide to acrylamide can be varied for special purposes, but is generally about 1 part in 35. The acrylamide concentration of the gel can also be varied, generally in the range from 5% to 25%.
- Lower percentage gels are better for resolving very high molecular weight molecules, while much higher percentages of acrylamide are needed to resolve smaller proteins,
- Gels are usually polymerized between two glass plates in a gel caster, with a comb inserted at the top to create the sample wells.
- After the gel is polymerized, the comb is removed and the gel is ready for electrophoresis.

3. Electrophoresis

- Various buffer systems are used in PAGE depending on the nature of the sample and the experimental objective.
- The buffers used at the anode and cathode may be the same or different.
- An electric field is applied across the gel, causing the negatively charged proteins or nucleic acids to migrate across the gel away from the negative and towards the positive electrode (the anode).
- Depending on their size, each biomolecule moves differently through the gel matrix: small molecules more easily fit through the pores in the gel, while larger ones have more difficulty.
- The gel is run usually for a few hours, though this depends on the voltage applied across the gel.
- After the set amount of time, the biomolecules will have migrated different distances based on their size.
- Smaller biomolecules travel farther down the gel, while larger ones remain closer to the point of origin.

- Biomolecules may therefore be separated roughly according to size, which depends mainly on molecular weight under denaturing conditions, but also depends on higher-order conformation under native conditions.

4. Detection

- Following electrophoresis, the gel may be stained (for proteins, most commonly with Coomassie Brilliant Blue or autoradiography; for nucleic acids, ethidium bromide; or for either, silver stain), allowing visualization of the separated proteins, or processed further (e.g., Western blot).
- After staining, different species biomolecules appear as distinct bands within the gel.
- It is common to run molecular weight (size) markers of known molecular weight in a separate lane in the gel to calibrate the gel and determine the approximate molecular mass of unknown biomolecules by comparing the distance travelled relative to the marker

Application of PAGE

The applications of PAGE are diverse and include:

- Protein purification and characterization
- Identification of protein isoforms and variants
- Analysis of protein-protein interactions
- Determination of protein molecular weight
- Detection and quantification of protein contaminants in food and pharmaceuticals.

SDS-PAGE [Sodium Dodecyl Sulfate (SDS) polyacrylamide gel electrophoresis (PAGE)]

Sodium Dodecyl Sulfate (SDS) polyacrylamide gel electrophoresis is mostly used to separate proteins accordingly by size. This is one of the most powerful techniques to separate proteins on the basis of their molecular weight.

Principle:

This technique uses anionic detergent Sodium Dodecyl Sulfate (SDS) which dissociates proteins into their individual polypeptide subunits and gives a uniform negative charge along each denatured polypeptide. SDS also performs another important task.

It forces polypeptides to extend their conformations to achieve similar charge: mass ratio. SDS treatment therefore eliminates the effects of differences in shape so that chain length, which reflects their molecular mass, is the sole determinant of migration rate of proteins in the process of electrophoresis.

When these denatured polypeptides are loaded at the cathode end of an electrophoretic tank having polyacrylamide gel (as the supporting media) and subjected to an electric field, then we get clear bands of proteins arranged in an decreasing order of their molecular mass from the cathode to anode.

The rate of movement is influenced by the gel's pore size and the strength of electric field. In SDS- PAGE the vertical gel apparatus is mostly used. Although it is used to separate proteins on a routine basis, SDS-PAGE can also be used to separate DNA and RNA molecules.

Instrumentation:

•Physical Apparatus:

a. Electrophoretic Apparatus:

Vertical horizontal tank with electrodes, gel cassettes, Teflon spacers, clips, pipette or syringe, comb, acrylic cover.

b. Power Supply:

A power supply of 100-200 volts is needed. This is ideal for running and transferring protein resolving gels.

c. Staining Box:

These are trays in which the gels are stained and made up of clear plastics. These are resistant to most organic dyes, silver and other stains.

•Chemical Components:

(a) Supporting Media:

SDS-PAGE acrylamide is used as the supporting medium. It is a white crystalline powder, when acrylamide dissolves in water, it undergoes polymerization reaction to form a net-like structure called polyacrylamide. Polyacrylamide is a polymer (CF_2CH-$CONH_2$-) formed from acrylamide subunits that can also be readily cross-linked.

This type of electrophoresis has a discontinuous system of gel, i.e., we have two different systems of gels present in the electrophoretic tank physically placed one over another.

These are as follows:

Resolving Gel:

This is also called separating or running gel. The separating gel constitutes about $2/3^{rd}$ of the length of gel plate and is prepared by 5-10% of acrylamide. The pores in this gel (which is formed after the polyacrylamide is cross- linked) are numerous and smaller in diameter which impacts sieving property to this gel.

Stacking gel:

Stacking gel is poured on the top of resolving gel and a gel comb is inserted which forms the well. It is the upper layer of gel and constitutes $1/3^{rd}$ of the gel plate. The percentage of acrylamide is chosen depending on the size of protein that one wishes to identify or probe in the sample.

The smaller the known weight, the higher the percentage that should be used. Generally, the percentage of acrylamide in stacking gel is 2-5%. It is highly porous and devoid of molecular sieving action.

(b) Buffer:

Two types of buffers are used in SDS-PAGE. The lower reservoir (which has the running gel) has amine buffers. It is adjusted by using HCl. The upper reservoir (which has stacking gel) also has amine buffers but its pH is slightly above that of running gel buffer and is adjusted with glycine instead of HCl.

(c) Dissociating Agent:

SDS is the most common dissociating agent used to denature native proteins to individual polypeptides. When a protein mixture is heated to 100°C in presence of SDS, the detergent wraps around the polypeptide backbone.

It binds to polypeptides in a constant weight ratio of 1.4 g/g of polypeptide. In this process, the intrinsic charges of polypeptides becomes negligible when compared to the negative charges contributed by SDS. Thus, polypeptides after treatment become rod-like structure possessing uniform charge density, that is the same net negative charge per unit length.

(d) Stains:

The stains are used to see the bands of separated proteins after the process of electrophoresis. Coomassie Brilliant Blue R-250 (CBB) is the most popular protein stain. It is an anionic dye, which binds with proteins non-specifically.

Proteins in the gel are fixed by acetic acid and simultaneously stained. The excess dye incorporated in the gel can be removed by de-staining with the same solution but without the dye. The proteins are detected as blue bands on a clear background.

Procedure of SDS-PAGE:

The solution of proteins to be analysed is first mixed with SDS, an anionic detergent, an anionic detergent which denatures secondary structure. Besides addition of SDS, proteins may optionally be boiled in the presence of a reducing agent, such as Di-Thio-Threitol (DTT) or 2-mercaptoethanol, which further denatures the proteins by reducing disulfide linkages, thus overcoming some forms of tertiary protein folding, and breaking up quaternary protein structure (Oligomeric subunits).

This is known as reducing SDS-PAGE, and is most commonly used. Non-reducing SDS-PAGE (no boiling and no reducing agent) may be used when native structure is important in further analysis (e.g., enzyme activity, shown by the use of zymograms). The denatured proteins are subsequently loaded into the wells of stacking gel flooded with stacking buffer.

This end is connected with the cathode of power supply. Then an electric current is applied across the gel, causing negatively charged proteins to migrate across the gel towards anode. After crossing the stacking gel, denatured proteins enter the running gel which has its own buffer system (running buffer).

Depending on their size, each protein will move differently through the gel matrix: short proteins will more easily fit through the pores in the gel, while larger ones will have more difficulty.

After the separation is over the gel is gently taken out and transferred to the staining box and treated with the staining dye, e.g., CBB R-250. Excess of stains are removed by de-staining using acetic acid solution. The bands appear to be blue stained which are then analysed according to the need of the experiment.

Application:

SDS-PAGE has many applications. It is mostly used for following purposes:

1. It is used to measure the molecular weight of the molecules.
2. It is used to estimate the size of the protein.

3. It is used to compare the polypeptide composition of different structures.
4. It is used to estimate the purity of the proteins.
5. It is used in Western Blotting and protein ubiquitination.
6. Analyzing the size and number of polypeptide subunits.
7. To analyze post-translational modifications.

Advantages of SDS-PAGE:

1. Mobility of the molecules is high and separation is rapid.
2. All the proteins are negatively charged; therefore, all migrate towards anode.
3. The proteins treated with SDS fixed dyes are better than the native proteins.
4. SDS solubilizes all proteins, including very hydrophobic and even denatured proteins.

Two-dimensional Gel Electrophoresis

This technique has evolved by combining the resolving power of IEF and SDSPAGE. In this protein sample first subjected to IEF, and then the IEF gel is transferred on to the top of the pre-casted SDS gel to perform PAGE, hence it is named as 2-D. IEF and SDS-PAGE separate proteins based on their isoelectric point and molecular weight, both these properties are unique for any protein. Thus 2-D has got an added advantage over any other electrophoretic technique with regard to separation of proteins. The resolving power of 2-D is so high that a mixture containing 5000 proteins can be separated in to individual proteins spots. 2-D is widely used to analyze crude protein samples. In simple words, 2-D can be assumed as performing IEF in horizontal direction and SDS-PAGE in vertical direction for the same protein sample.

Probable questions:

1. Define chromatography? What is the basic principle of chromatography?
2. Discuss the basic principle of thin layer chromatography.
3. What is mobile phase?
4. What is stationary phase?
5. Mention the principle of Ion exchange chromatography.

6. State the applications of Ion exchange chromatography.
7. State the principle of Gel filtration chromatography.
8. When is gel filtration chromatography applied to separate molecules?
9. Illustrate the steps of gel filtration chromatography.
10. What is HPLC?
11. How is HPLC performed to separate molecules?
12. Describe the principle of HPLC.
13. State the application of SDS-PAGE.
14. Illustrate the steps of PAGE.
15. Illustrate the steps and principles of SDS-PAGE.
16. What is stacking dye?

Suggested Reading:

1. Wilson and Walker's Principles and Techniques of Biochemistry and Molecular Biology 2018, ISBN: 9781316677056
2. Biochemistry Laboratory: Modern Theory and Techniques, Latest Edition, Rodney Boyer, Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.ISBN-10: 0-13-604302-X, ISBN-13: 978-0-13-604302-7

References:

1. https://chromatography.conferenceseries.com/events-list/applications-of-chromatography
2. http://www.biologydiscussion.com/biochemistry/chromatography-techniques/top-12-types-of-chromatographic-techniques-biochemistry/12730
3. https://www.iptsalipur.org/wp-content/uploads/2020/08/BP701T_IMA_V.pdf
4. https://microbenotes.com/ion-exchange-chromatography/
5. https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S001174BS/P001204/M011038/ET/1479291507P9M16eTextOct12.pdf
6. https://microbenotes.com/gel-filtration-chromatography/

7. http://kirschner.med.harvard.edu/files/protocols/GE_gelfiltration.pdf
8. Wilson, K., Walker, J. (2018). Principles and Techniques of Biochemistry and Molecular Biology (8 eds.). Cambridge University Press: New York.
9. S. D. Varhadi et al. A Short Review on: Definition, Principle and Applications of High-Performance Liquid Chromatography, Ijppr.Human, 2020; Vol. 19 (2): 628-634
10. Oona McPolin, an Introduction to HPLC for Pharmaceutical Analysis. 2009: 1-9
11. Santosh Kumar Bhardwaj, K. Dwivedi and D.D Agarwal, a Review: HPLC Method Development and Validation.2015; 5(4):76-81.
12. https://microbenotes.com/high-performance-liquid-chromatography-hplc/
13. https://www.biotechnologynotes.com/electrophoresis/electrophoresis-meaning-definition-and-classification-with-diagram/293

Unit-XVI

Methods for analysis of gene expression at RNA and protein level, large scale expression, such as micro array-based techniques

1. Introduction

The understanding of gene expression patterns at both the RNA and protein levels is fundamental to molecular biology, biotechnology, and biomedical research. Studying gene expression provides insights into the functional elements of the genome, mechanisms of disease, and therapeutic targets. With the advancement of high-throughput technologies, large-scale analyses like microarrays allow for simultaneous monitoring of thousands of genes.

2. Methods for Analysis of Gene Expression at RNA Level

2.1 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

Principle:

Reverse Transcription Polymerase Chain Reaction (RT-PCR) quantifies mRNA levels by first converting RNA into complementary DNA (cDNA) using **Reverse Transcriptase (RT)**, followed by amplification through **Polymerase Chain Reaction (PCR)**.

Steps:

- Extraction of high-quality total RNA.
- Reverse transcription of RNA into cDNA.
- Amplification of cDNA using gene-specific primers.
- Visualization via gel electrophoresis (for semi-quantitative RT-PCR).

Application:

Detection of specific gene transcripts; semi-quantitative comparison.

2.2 Quantitative Real-Time PCR (qRT-PCR)

Principle:

qRT-PCR allows real-time monitoring of PCR amplification using fluorescent dyes like **SYBR Green** or probe-based detection systems like **TaqMan probes**.

Steps:

- cDNA synthesis from RNA.
- Amplification with real-time detection.
- Fluorescence correlates with the amount of amplified product.

Advantages:

- Highly sensitive and quantitative.
- Capable of detecting even low-abundance transcripts.

Applications:

- Validation of gene expression from microarray or RNA-seq data.
- Diagnostic applications.

2.3 Northern Blotting

Principle:

Northern blotting detects specific RNA molecules among a mixture through separation via **gel electrophoresis**, transfer to a membrane, and hybridization with labeled probes.

Steps:

- RNA separation by gel electrophoresis.
- Transfer to a membrane (usually nylon or nitrocellulose).
- Hybridization with complementary labeled probe (radioactive or non-radioactive).

Applications:

- Determination of transcript size.
- Expression profiling across tissues or developmental stages.

Limitations:

- Labor-intensive.
- Less sensitive compared to qRT-PCR.

2.4 RNA Sequencing (RNA-seq)

Principle:

RNA-seq provides comprehensive transcriptome profiling by sequencing cDNA fragments derived from RNA.

Steps:

- RNA isolation and fragmentation.
- Conversion to cDNA.
- Library preparation and sequencing.
- Data analysis through bioinformatics pipelines.

Advantages:

- Unbiased detection of known and novel transcripts.
- Quantitative and qualitative information.

Applications:

- Discovery of novel RNA species.
- Differential gene expression analysis.
- Isoform identification.

3. Methods for Analysis of Gene Expression at Protein Level

3.1 Western Blotting

Principle:

Western blotting detects specific proteins via gel electrophoresis, transfer to a membrane, and identification using antibodies.

Steps:

- Protein extraction.
- SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis).
- Transfer to a membrane.
- Immunodetection using primary and secondary antibodies.

Applications:

- Quantification and verification of protein expression.
- Post-translational modification studies.

3.2 Enzyme-Linked Immunosorbent Assay (ELISA)

Principle:

ELISA quantitatively measures proteins (e.g., cytokines, hormones) using antigen-antibody interactions and a detectable enzyme reaction.

Types:

- Direct ELISA.
- Indirect ELISA.
- Sandwich ELISA.

Applications:

- High-throughput protein quantification.
- Diagnostics and biomarker discovery.

3.3 Immunohistochemistry (IHC)

Principle:

IHC detects proteins in tissue sections using antibodies conjugated with enzymes or fluorophores.

Applications:

- Localization of protein expression in tissues.
- Diagnostic pathology (e.g., cancer biomarkers).

4. Large-Scale Expression Analysis: Microarray-Based Techniques

4.1 DNA Microarray

Principle:

Microarrays measure the expression levels of thousands of genes simultaneously by hybridizing labelled cDNA to complementary probes immobilized on a solid surface.

Steps:

- RNA extraction and conversion to labelled cDNA.
- Hybridization to a chip with DNA probes.
- Washing and scanning.
- Data analysis.

Applications:

- Comparative gene expression analysis.
- Genotyping.
- Biomarker identification.

Types:

- cDNA microarrays.
- Oligonucleotide microarrays (e.g., Affymetrix GeneChips).

Advantages:

- High-throughput.
- Relatively inexpensive compared to RNA-seq.

Limitations:

- Limited to known sequences.
- Lower dynamic range than RNA-seq.

4.2 Data Analysis in Microarray

Key steps:

- Background correction.
- Normalization (e.g., Robust Multi-array Average (RMA)).
- Statistical testing (e.g., t-tests, ANOVA for differential expression).
- Pathway and gene ontology enrichment analysis.

Software Tools:

- R and Bioconductor packages (e.g., limma, affy).
- GeneSpring.
- Partek Genomics Suite.

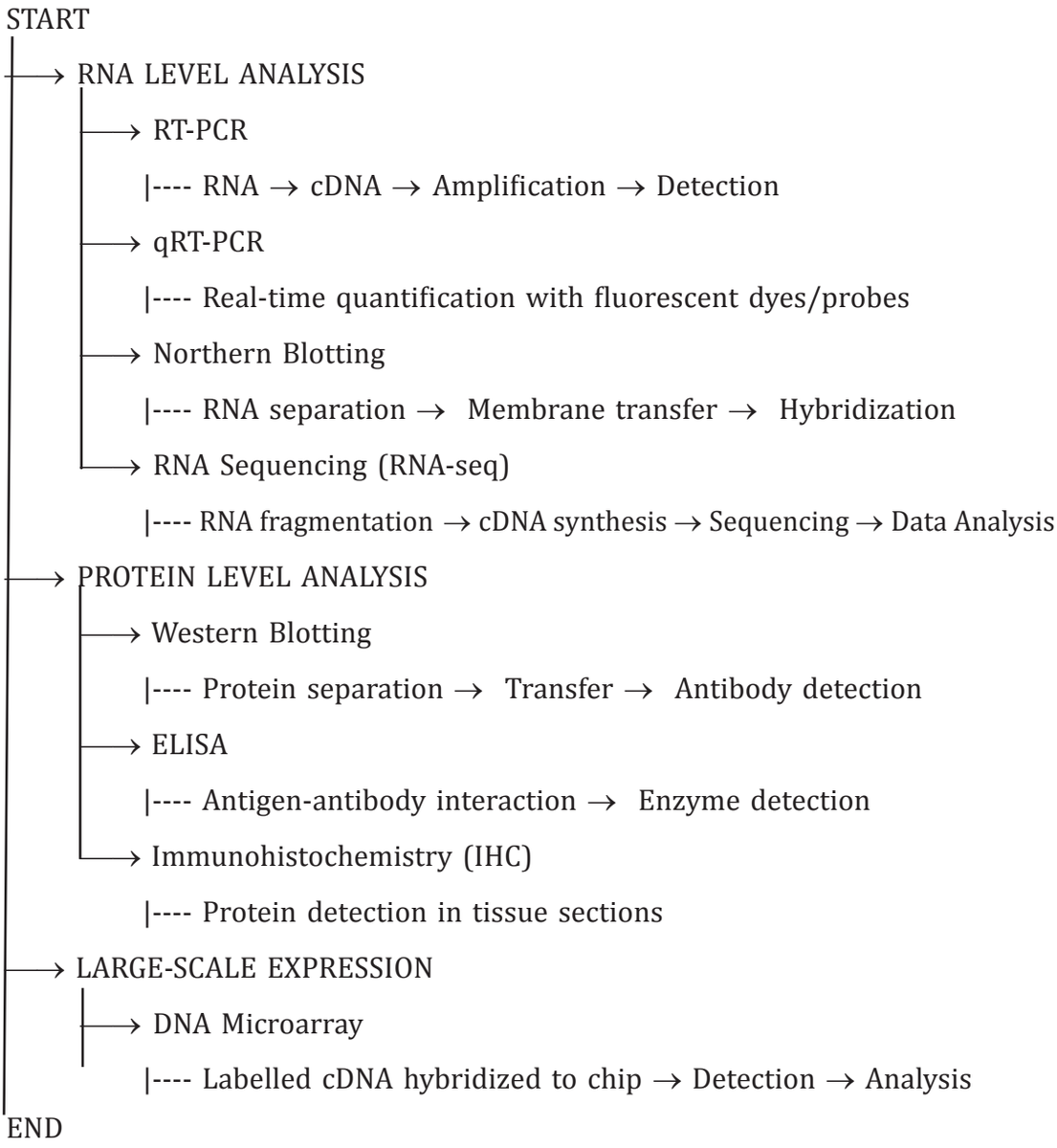
5. Conclusion

The analysis of gene expression at RNA and protein levels forms a cornerstone of modern molecular biology. Techniques such as qRT-PCR, RNA-seq, Western blotting, ELISA, and immunohistochemistry have provided powerful means to dissect molecular pathways. Furthermore, large-scale platforms like microarrays have revolutionized the

ability to simultaneously monitor the activity of thousands of genes, greatly enhancing our understanding of complex biological systems.

Understanding the principles, methodologies, and applications of these techniques is crucial for students and researchers aiming to conduct experimental research or interpret gene expression studies in the context of disease, development, and biotechnology.

6. Conceptual Flowchart: Analysis of Gene Expression at RNA and Protein Level



7. Tabularized Summary of Methods for Gene Expression Analysis

Level	Technique	Principle	Steps	Applications	Reference
RNA	RT-PCR	Reverse transcription of RNA to cDNA, followed by PCR amplification	RNA extraction → cDNA syntehsis → PCR amplification → Detection	Detection of gene expression	Green & Sambrook (2012)
RNA	qRT-PCR	Real-time PCR quantification using fluorescent probes	RNA extraction → cDNA synthesis → Real-time amplification → Data interpretation	Quantitative analysis of transcripts	Nolan et al. (2006)
RNA	Northern Blotting	RNA separation and hybridization with labeled probe	Gel electrophoresis → Membrane transfer → Probe hybridization → Detection	Transcript size analysis, gene expression profiling	Alwine et al. (1977)
RNA	RNA-seq	High-throughput sequencing of RNA-derived cDNA	RNA fragmentation → cDNA synthesis → Library prep → Sequencing → Analysis	Comprehensive transcriptome profiling	Wang et al. (2009)
Protein	Western Blotting	Antibody-based detection of proteins	Protein extraction → SDS-PAGE → Transfer → Antibody binding → Detection	Protein expression analysis, PTM detection	Mahmood & Yang (2012)
Protein	ELISA	Quantitative measurement of proteins via antigen-antibody binding	Coating → Blocking → Antibody binding → Detection	High-throughput protein quantification, diagnostics	Engvall & Perlmann (1971)
Protein	Immuno-histo-chemistry (IHC)	Detection of proteins in tissue sections	Tissue processing → Antibody binding → Visualization	Tissue-specific protein expression, diagnostics	Ramos-Vara (2005)

Level	Technique	Principle	Steps	Applications	Reference
Large-Scale	DNA Microarray	Hybridization of labeled cDNA to probes on a chip	RNA extraction → cDNA labeling → Hybridization → Scanning → Analysis	Simultaneous gene expression profiling, biomarker discovery	Lockhart & Winzeler (2000)

Probable Questions

1. Describe the principle, steps, and applications of Reverse Transcription Polymerase Chain Reaction (RT-PCR) in gene expression analysis.
2. Differentiate between qRT-PCR and traditional RT-PCR in terms of principle, sensitivity, and application.
3. Explain the method of RNA sequencing (RNA-seq) for transcriptome profiling. How does it compare to microarray analysis?
4. Discuss the principle and procedural steps involved in Northern blotting. What are its advantages and limitations in RNA expression analysis?
5. Explain the process of Western blotting. How is it used for detecting protein expression and post-translational modifications?
6. Write a detailed note on the types and principles of ELISA. How does it serve as a tool for protein quantification?
7. Describe the immunohistochemistry (IHC) technique and its role in tissue-specific protein expression analysis.
8. Outline the workflow of DNA microarray analysis and explain how it enables large-scale gene expression profiling.
9. Discuss the importance and steps of data analysis in microarray experiments, including normalization and statistical testing.
10. Compare and contrast RNA-level and protein-level gene expression analysis techniques. Provide examples of when each is preferred.

Suggested Readings

1. Green, M. R., & Sambrook, J. (2012). *Molecular Cloning: A Laboratory Manual* (4th ed.). Cold Spring Harbor Laboratory Press.
2. Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols, 1*(3), 1559–1582. https://doi.org/10.1038/nprot.2006.236
3. Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977). Method for detecting specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences, 74*(12), 5350–5354. https://doi.org/10.1073/pnas.74.12.5350
4. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57–63. https://doi.org/10.1038/nrg2484
5. Mahmood, T., & Yang, P. C. (2012). Western blot: technique, theory, and trouble shooting. *North American Journal of Medical Sciences, 4*(9), 429–434. https://doi.org/10.4103/1947-2714.100998
6. Engvall, E., & Perlmann, P. (1971). Enzyme-linked immunosorbent assay (ELISA). *Journal of Immunology, 109*(1), 129–135.
7. Ramos-Vara, J. A. (2005). Technical aspects of immunohistochemistry. *Veterinary Pathology, 42*(4), 405–426. https://doi.org/10.1354/vp.42-4-405
8. Lockhart, D. J., & Winzeler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature, 405*(6788), 827–836. https://doi.org/10.1038/35015701
9. Smyth, G. K. (2009). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology, 3*(1), Article3. https://doi.org/10.2202/1544-6115.1027

Unit XVII

Spectroscopy: UV- spectroscopy, Circular Dichroism, surface plasma resonance methods

Objective: In this unit we will discuss about Spectroscopy which includes UV- spectroscopy. Circular Dichroism and surface plasma resonance methods will be discussed also.

Introduction

Spectroscopy is the branch of science dealing with the study of interaction of electromagnetic radiation with matter like atoms and molecules. The interaction of Electromagnetic Radiations (EMR) with matter gives rise to two types of spectra namely atomic spectra and molecular spectra. Atomic spectrum arises from the transition of electrons from one energy level to another due to changes of energy in the atom. Molecular spectrum involves transition of electrons between rotational and vibrational energy levels in addition to electronic transition. Therefore, molecular spectrum is much more complicated than the atomic spectrum. Molecular Spectroscopy provides a clear image of how diatomic and polyatomic molecules interact by looking at the Frequency, Wavelength, Wave number, Energy, and molecular process also provides most useful information regarding the shape and size of molecules, the bond angles, bond lengths, strength of bonds and bond dissociation energies. Hence molecular spectroscopy is of great use in determining the structure and constitution of compounds.

- **Electromagnetic spectrum**

The field of spectroscopy deals with the study of absorption or emission of electromagnetic radiations. We are familiar with the number of different types of electromagnetic radiations such as cosmic rays, X-rays, Ultra violet light, visible light, infra-red radiations, radio waves radar waves etc. All the electromagnetic radiations travel with the same velocity i, e.3x1010 cm/sec but differ from one another in the wave length of their waves. The arrangement of all types of electromagnetic radiations in the order of their increasing wavelength or decreasing frequency is called the electromagnetic spectrum.

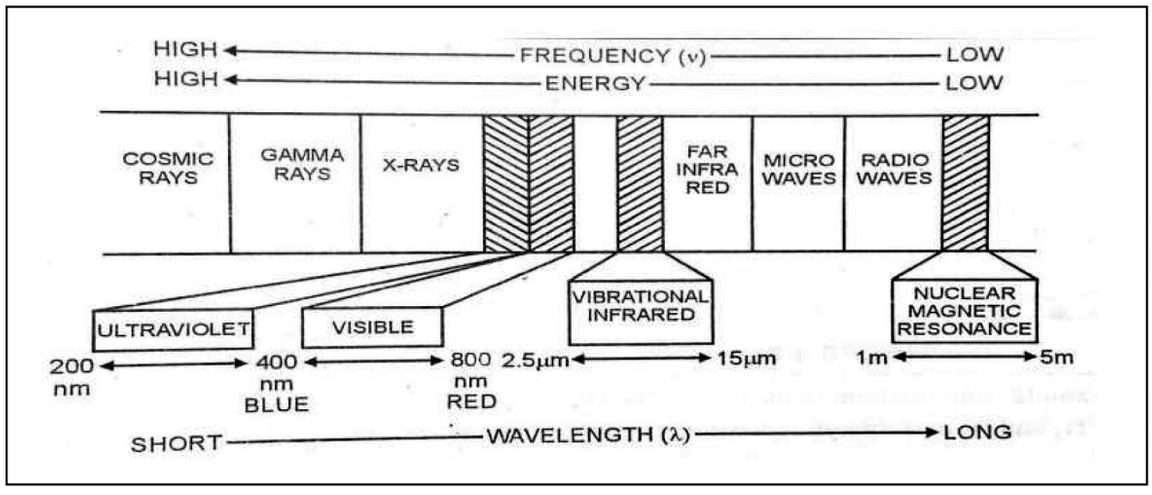


Fig. 17.1

- **Principles of UV spectroscopy**

UV spectroscopy is an important tool in analytical chemistry. The other name of UV (UltraViolet) spectroscopy is electronic spectroscopy as it involves the promotion of the electrons from the ground state to the higher energy or excited state. It is type of absorption spectroscopy in which light of ultra-violet region (200-400 nm.) is absorbed by the molecule. Absorption of the ultra-violet radiations results in the excitation of the electrons from the ground state to higher energy state. The energy of the ultra-violet radiation that is absorbed is equal to the energy difference between the ground state and higher energy states. The most favoured transition is from the highest occupied molecular orbital (HOMO) to lowest unoccupied molecular orbital (LUMO).

UV spectroscopy obeys the Beer-Lambert law, which states that: when a beam of monochromatic light is passed through a solution of an absorbing substance, the rate of decrease of intensity of radiation with thickness of the absorbing solution is proportional to the incident radiation as well as the concentration of the solution. The expression of Beer-Lambert law is:

$A = \log (I_0/I) = \varepsilon . c.l$

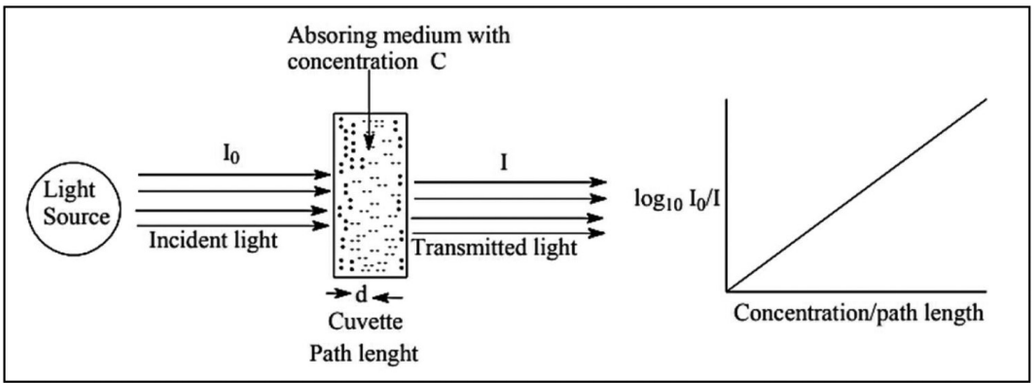


Fig. 17.2

- **The Beer–Lambert Law**

Wilhelm Beer and Johann Lambert independently proposed that at a given wavelength, the absorbance of a sample depends on the amount of absorbing species that the light encounters as it passes through a solution of the sample. In other words, absorbance depends on both the concentration of the sample and the length of the light path through the sample.

The relationship among absorbance, concentration, and length of the light path is known as the Beer–Lambert law and is given by

$A = \varepsilon cl$

Where

A = absorbance of the sample = logI0/I

I_0 = intensity of the radiation entering the sample

I = intensity of the radiation emerging from the sample

c = concentration of the sample, in moles/Liter

l = length of the light path through the sample, in centimetres

ε = molar absorptivity (Liter mol-1 cm-1)

From the Beer-Lambert law it is clear that greater the number of molecules capable of absorbing light of a given wavelength, the greater the extent of light absorption. This is the basic principle of UV spectroscopy.

- **Working of UV spectroscopy**

Most of the modern UV spectrometers consist of the following parts:

I. **Light Source**- Tungsten filament lamps and Hydrogen-Deuterium lamps are most widely used and suitable light source as they cover the whole UV region. Tungsten filament lamps are rich in red radiations; more specifically they emit the radiations of 375 nm, while the intensity of Hydrogen-Deuterium lamps falls below 375 nm. Now day's xenon lamps are being used.

II. **Monochromator**- Monochromators generally composed of prisms and slits. The most of the spectrophotometers are double beam spectrophotometers. The radiation emitted from the primary source is dispersed with the help of rotating prisms. The various wavelengths of the light source which are separated by the prism are then selected by the slits such the rotation of the prism results in a series of continuously increasing wavelength to pass through the slits for recording purpose. The beam selected by the slit is monochromatic and further divided into two beams with the help of another prism.

III. **Sample and reference cells**- One of the two divided beams is passed through the sample solution and second beam is passé through the reference solution. Both sample and reference solution are contained in the cells. These cells are made of either silica or quartz. Glass can't be used for the cells as it also absorbs light in the UV region.

IV. **Detector**- Generally two photocells serve the purpose of detector in UV spectroscopy. One of the photocells receives the beam from sample cell and second detector receives the beam from the reference. The intensity of the radiation from the reference cell is stronger than the beam of sample cell. This results in the generation of pulsating or alternating currents in the photocells.

V. **Amplifier**- The alternating current generated in the photocells is transferred to the amplifier. The amplifier is coupled to a small servometer. Generally current generated in the photocells is of very low intensity, the main purpose of amplifier is to amplify the signals many times so we can get clear and recordable signals.

VI. **Recording devices**- Most of the time amplifier is coupled to a pen recorder which is connected to the computer. Computer stores all the data generated and produces the spectrum of the desired compound.

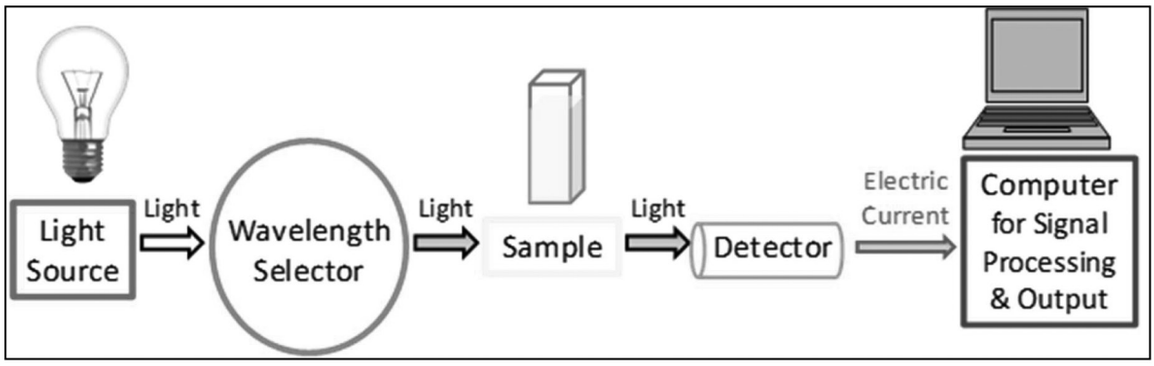


Fig. 17.3: A simplified schematic of the main components in a UV-Vis spectrophotometer.
[Source: https://www.technologynetworks.com/analysis/articles/uv-vis-spectroscopy-principle-strengths-and-limitations-and-applications-349865]

Terminology of UV-visible spectroscopy

A. Chromophore: The multiple bonded groups in the organic molecules due to which they owe their colour are called chromophore (Chrom = colour, phore = bearing/ possessing). The examples of chromophores are as follow:

B. Auxochrome: An auxochrome are auxiliary groups which interacts with the chromophore and cause a bathochromic shift i.e. they do not possess their own colour but intensify the colour bearing capacity of chromophore if attached with it.

Viz; NH_2, NHR, NR_2, -OH, $-OCH_3$ –R, -SH etc.

The unique property of auxochrome grouping is its ability to provide additional opportunity for charge delocalization and provide smaller energy increments for transition to excited states. The delocalization in molecules is increased due to lone pair of electrons in most of the auxochrome or due to other electronic displace like hyper conjunction. Attachment of auxochrome to chromophore always accuse bathochromic shift i.e. absorption takes place towards longer wavelength region. Some examples of auxo chromic effects are as follow

C. Bathochromic shift (Red-shift): Shift of absorption of EMR towards longer wavelength (λ) or low energy (low frequency) region in UV-Visible spectroscopy is known as bathochromic shift or Red-shift. The shift of absorption of peak position is due to 1. Attachment of auxochrome to chromophore 2. Increased conjugation 3. Change in polarity of solvent

D. Hypochromic shift (Blue-shift): A shift of absorption of EMR towards shorter wavelength (λ) or high energy (high frequency) region in UV-Visible spectroscopy is known as hypochromic shift or blue-shift. This may be caused by

1. Detachment of auxochrome

2. Decreased conjugation

3. Change in solvent polarity

viz acetone absorb at 279 nm in hexane while in water the λmax value of acetone is 265 nm. This shift results from H….. bonding which lowers the energy of n-orbital. This can also be produced when auxochrome is attached to double bonds where n = electrons are available viz; C=O.

E. Hyperchromic effect: It is effect leading to uncreased absorption intensity. It is related to bathochromic shift and molar absorptivity coefficient (ε_{max} increases)

F. Hypochromic effect: It is effect leading to decreased absorption intensity. It is related to hypochromic shift (ε_{max} decreases)

G. Isobestic point: A point common to all curves produced in the spectra of a compound taken at various pH values is called isobestic point.

- **Advantages of UV Spectrometers**

1. The biggest advantage for chemists and astronomers who use UV- spectrometers is the accuracy of the device. Even small UV-VIS spectrometers can give extremely accurate readings, which is crucial when you are preparing chemical solutions or recording the movement of celestial bodies.
2. UV spectrometers are easy to use.
3. Small quantity samples can be analysed.
4. Most UV spectrometers simple to operate, there is little chance of a UV spectrometer being used improperly.

- **Disadvantages of UV-VIS Spectrometers**

1. The main disadvantage of using a UV spectrometer is the time it takes to prepare to use as setup is key for the use of UV spectrometers.
2. You must clear the area of any outside light, electronic noise, or other outside contaminants that could interfere with the spectrometer's reading.
3. Accuracy of UV spectrometer is so high that a small bit of outside light or vibration from a small electronic device could interfere with the results.

Circular dichroism (CD)

Circular dichroism (CD) is a spectroscopy technique that measures the absorption difference between left and right circularly polarized light. By symmetry, this asymmetric absorption can only occur for asymmetric molecules, meaning chiral molecules. It is also in the study of large biological molecules where it finds its most important applications. Ultraviolet CD has found a particular application for empirically assessing protein secondary structure, enabling a quick determination of whether a protein is primarily α-helix, β-sheet, or unfolded. Vibrational CD, in the infrared, is especially useful for determining the structures of small molecules. CD has also been explored theoretically for over a century, culminating in a thorough understanding that is sufficient for accurately predicting many CD spectra from molecular structures, although these calculations typically require substantial computation (Andrews et al. 2020).

- **Principle of Operation**

In the CD spectrometer the sample is places in a cuvette and a beam of light is passed through the sample. The light (in the present context all electromagnetic waves will be refer to as light) coming from source is subjected to circular polarization, meaning that its plane of polarization is made to rotate either clockwise (right circular polarization) or anti-clockwise (left circular polarization) with time while propagating.

The sample is, firstly irradiated with left rotating polarized light, and the absorption is determined. A second irradiation is performed with right polarized light. Now, due to the intrinsic asymmetry of chiral molecules, they will interact with circularly polarized light differently according to the direction of rotation there is going to be a tendency to absorb more for one of rotation directions.

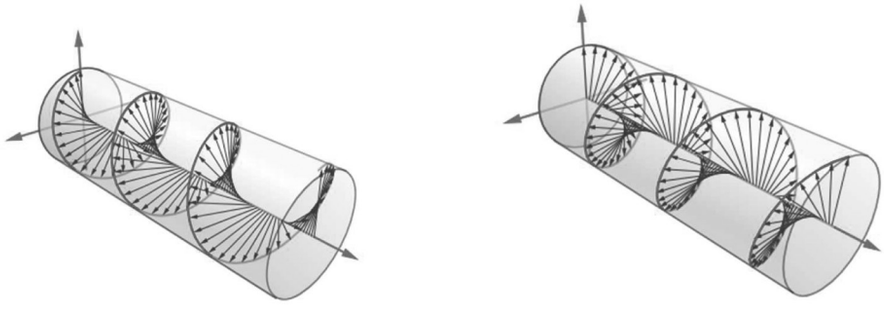


Fig. 17.4: Schematic representation of (a) right circularly polarized and (b) left circularly polarized light. (CC BY 4.0; Ümit Kaya via LibreTexts)

The difference between absorption of left and right circularly polarized light is the data, which is obtained from 7.7.2, where εL and εR are the molar extinction coeffi-

cients for left and right circularly polarized light, c is the molar concentration, l is the path length, the cuvette width (in cm). The difference in absorption can be related to difference in extinction, Δε, by 7.7.3.

$$A = \varepsilon cl$$

$$\Delta A = AL - AR = (\varepsilon L - \varepsilon R)cl \quad (7.7.2)$$

$$\Delta\varepsilon = \varepsilon L - \varepsilon R \quad (7.7.3)$$

Usually, due to historical reasons the CD is reported not only as difference in absorption or extinction coefficients but as degree of ellipticity, [θ]. The relationship between [θ] and Δε is given by 7.7.4.

$$[\theta] = 3{,}298\Delta\varepsilon \quad (7.7.4)$$

Since the absorption is monitored in a range of wavelengths, the output is a plot of [θ] versus wavelength or Δε versus wavelength. Figure 17.5 shows the CD spectrum of Δ–[Co(en)3]Cl3.

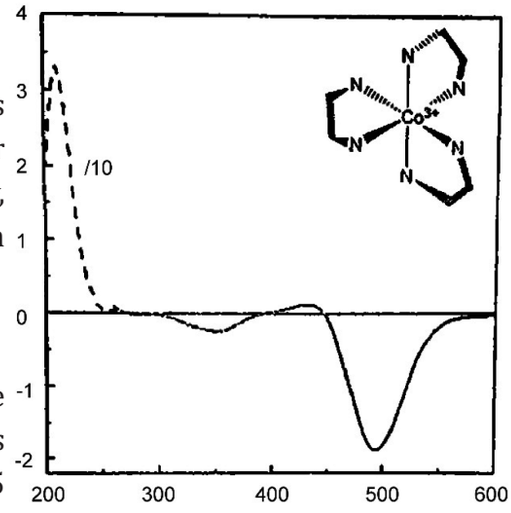


Fig. 17.5: CD spectrum of Δ–$[Co(en)_3]Cl_3$.

- **Application**

A primary use of CD is in analysing the secondary structure or conformation of macromolecules, particularly proteins as secondary structure is sensitive to its environment, temperature or pH, circular dichroism can be used to observe how secondary structure changes with environmental conditions or on interaction with other molecules.

Surface plasma resonance methods

Surface Plasmon Resonance (SPR) is an optical technique used to measure molecular interactions in real time. This technique is used for detecting the interaction of two different molecules in which one is mobile and one is fixed on a thin gold film (Schuck, 1997). SPR can occur when plane-polarized light hits a metal film under total internal reflection conditions. SPR signal is directly dependent on the refractive index of the medium on the sensor chip.

Principle of Surface Plasmon Resonance (SPR)

SPR is an optical phenomenon that offers a non-invasive, label-free means of monitoring binding interactions between an immobilized molecule and injected analyte

in real time. An optical detection process occurs when a polarized light of certain wavelength and angle strikes on gold plated prism. Total internal reflection creates and strikes an electrically conducting gold layer at the interface between media of different refractive index: the glass of a sensor surface (high refractive index) and a buffer (low refractive index). The free electrons at the surface of the biochip absorb incident light photons and convert them into surface plasmon waves. The effect of the plasmon is to create a very strong oscillating electric field near the surface, which is called the evanescent wave (Fig 1). The index of refraction of a surface bound layer is proportional to the concentration of bound molecules. It can determine the kinetic on and off rates for the interaction of a biomolecule with a ligand in real time. The change of the incident angle required for SPR is defined as SPR response in the unit of response unit (RU). Here we explain the SPR system by using Biacore system. Biacore system is a label free system to study the bimolecular interaction assay by using surface Plasmon resonance technology in real time.

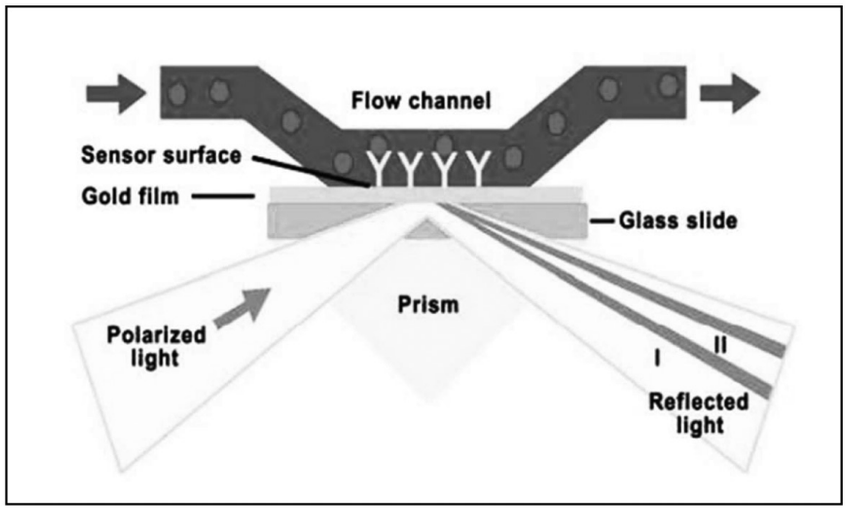


Fig. 17.6: Optical illustration of SPR [Source: https://epgp.inflibnet.ac.in]

Application of SPR:

1. SPR measures the direct binding between analyte and target which is why the target does not need to be an enzyme, and no substrate is needed.
2. The impact of various co-factors on analyte-target interaction can be tested.
3. Low molecular mass fragment compounds (100–300 Da) tend to demonstrate low binding affinity; thus, the compounds require screening at a high concentration. This is better tolerated in the SPR platform than in many biochemical assays.

4. SPR can be used for antibody affinity determination, determination of kinetic parameters, epitope mapping, binding specificity, and cross-reactivity.
5. The kinetics of drug binding and unbinding, especially the residence time, play a crucial role in a drug's in vivo efficacy. SPR can rank the kinetic selectivity of drug analogs for the selection of the best drug candidates.
6. Combining kinetic information with affinity and potency data early in the drug discovery process ensures that promising compounds are not being discarded.
7. Promiscuous binders, which appear as false positives in biochemical inhibitor assays, can be identified by SPR when used as the secondary screening technology.
8. SPR-based assays are sensitive high-throughput options to accurately measure plasma protein binding of analytes.

Probable questions:

1. What is UV spectrometer?
2. Describe the working principle of UV spectrometer?
3. State the advantages and disadvantages of UV spectrometer.
4. What is auxochrome?
5. What is chromophore?
6. What is circular dichroism?
7. State the use of Circular dichroism (CD)?
8. What is Surface plasma resonance methods
9. State the application of SPR.

Suggested Reading:

1. Wilson and Walker's Principles and Techniques of Biochemistry and Molecular Biology 2018, ISBN: 9781316677056
2. Biochemistry Laboratory: Modern Theory and Techniques, Latest Edition, Rodne Boyer, Pearson Prentice Hall™ is a trademark of Pearson Education, Inc.ISBN-10: 0-13-604302-X, ISBN-13: 978-0-13-604302-7

References:

1. https://www.uou.ac.in/sites/default/files/slm/MSCCH-509.pdf
2. https://www.technologynetworks.com/analysis/articles/uv-vis-spectroscopy-principle-strengths-and-limitations-and-applications-349865
3. https://www.sciencing.com/advantages-disadvantages-uvvis-spectrometer-6466475/
4. Steven S. Andrews* and James Tretton Physical Principles of Circular Dichroism J. Chem. Educ. 2020, 97, 4370”4376
5. https://www.ntu.edu.sg/docs/librariesprovider12/default-document-library/circular-dichroism.pdf?sfvrsn=ec10f4d4_2
6. chem.libretexts.org/Bookshelves/Analytical_Chemistry/Physical_Methods_in_Chemistry_and_Nano_Science_(Barron)/07%3A_Molecular_and_Solid_State_Structure/7.07%3A_Circular_Dichroism_Spectroscopy_and_its_Application_for_Determination_of_Secondary_Structure_of_Optically_Active_Species
7. Schuck P. Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. Annu. Rev. Biophys. Struct. 1997; 26:541–566. doi: 10.1146/annurev.biophys.26.1.541.
8. https://www.reactionbiology.com/services/biophysical-assays/surface-plasmon-resonance-spr#:~:text=SPR%20Binding%20Assay%20Principle,the%20surface%20plasmon%20resonance%20phenomenon).
9. https://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/S001174BS/P001858/M030458/ET/1526641812P10M32eTextDec2.pdf

Unit XVIII

Blotting Methods: Southern, Northern & Western blotting. RFLP, RAPD and AFLP techniques

Objective: In this unit we will discuss about different Blotting Methods: Southern, Northern & Western blotting. RFLP, RAPD and AFLP techniques will be discussed also.

Introduction

Blotting is used in molecular biology for the identification of proteins and nucleic acids and is widely used for diagnostic purposes. This technique immobilizes the molecule of interest on a support, which is a nitro cellulosic membrane or nylon. It uses hybridization techniques for the identification of the specific nucleic acids and genes. The blotting technique is a tool used in the identification of biomolecules such ad DNA, mRNA and protein during different stages of gene expression. In this technique the molecules are separated by gel electrophoresis.

Gel electrophoresis is a method for separation and analysis of macromolecules (DNA, RNA and proteins) and their fragments, based on their size and charge.

Types of Blotting Techniques

With the discovery of agarose gel electrophoresis, it was possible to separate nucleic acid (DNA and RNA) fragments based on their molecular size. Edwin Southern in year 1975 transferred DNA fragments separated by electrophoresis to a membrane and the molecule of interest was detected by some probe which was hybridized with the membrane. The concept gave the basis for a range of techniques for understanding the organization and expression of genetic materials. After the name of Edwin Southern, the technique is called as Southern Blotting. Thus, Southern blotting refers to transfer of DNA from gel to membrane. Subsequently, procedures were developed for transfer of RNA and protein on membrane and these techniques are called Northern Blotting and Western Blotting respectively. The technique of DNA transfer (Southern Blotting) and RNA transfer (Northern Blotting) are at times grouped as Nucleic acid transfer techniques. Once molecules are transferred to membrane, these are detected by employing some detection method.

Southern blotting

Principle

Southern blot analysis is a laboratory method used to study DNA. Specifically, purified DNA from a biological sample (such as blood or tissue) is digested with a restriction enzyme(s), and the resulting DNA fragments are separated by using an electric current to move them through a sieve-like gel or matrix, which allows smaller fragments move faster than larger fragments. The DNA fragments are transferred out of the gel or matrix onto a solid membrane, which is then exposed to a DNA probe labelled with a radioactive, fluorescent or chemical tag. The tag allows any DNA fragments containing complementary sequences with the DNA probe sequence to be visualized within the Southern blot. The method is named for its creator, British molecular biologist Edwin Southern.

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 DN 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 non-specifically or remain unbound so that only labelled 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 colour on the membrane.

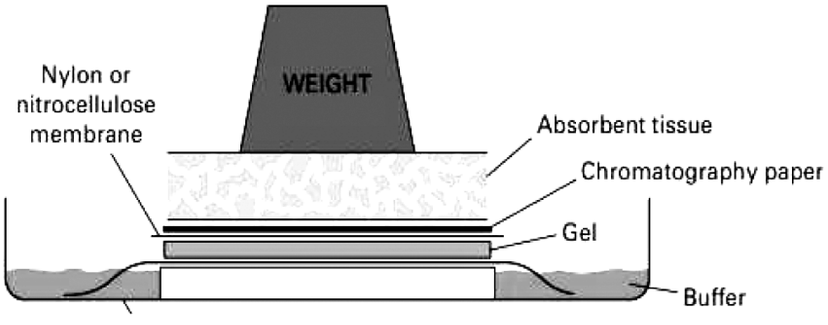


Fig. 18.1: Southern Bolt apparatus
[Source: Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.]

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.

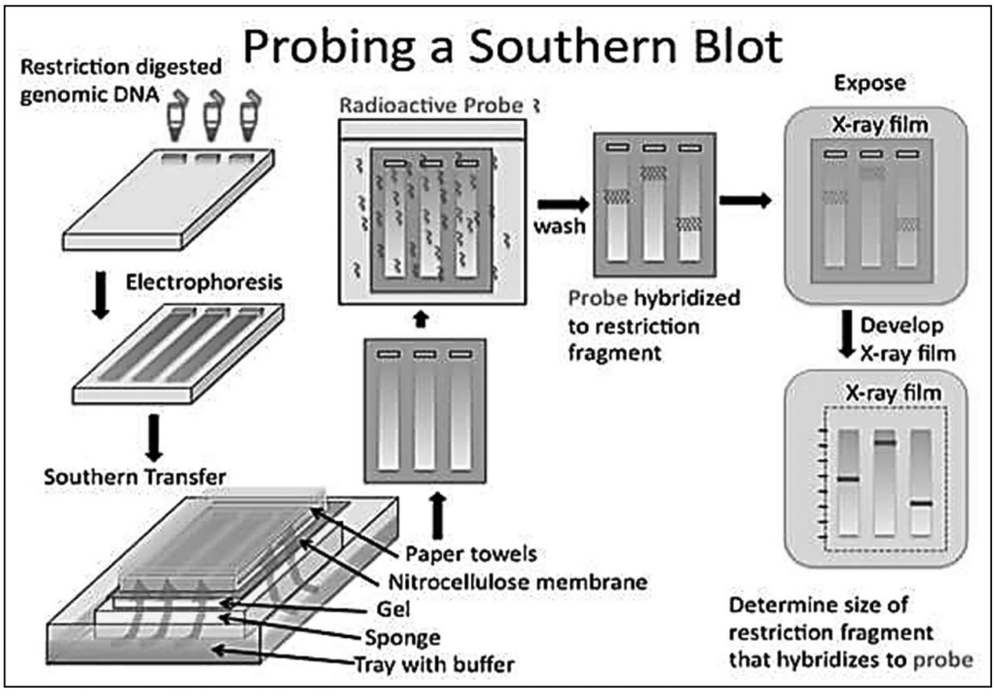


Image Source: Biology LibreTexts.

Fig. 18.2

Applications of Southern Blot

1. Southern blotting has many applications in the field of gene discovery, mapping, evolution, and diagnostic studies.
2. The technique can be used for DNA analysis to detect point mutations and other structural rearrangements in the DNA sequences.
3. The method also allows the determination of molecular weights of the restriction fragments, which helps in the analysis of such fragments.
4. Since the technique enables the detection of a particular DNA segment, it can be used in personal identification via fingerprinting.
5. It can be used in disease diagnosis as well as prenatal diagnosis of genetic diseases.

Limitations of Southern Blot

1. The method is costly as it requires expensive equipment and reagents as compared to other tests.

2. It is a complex process consisting of multiple steps. The process is also labor intensive that requires trained personnel.
3. It is a time-consuming process that can be replaced by other faster processes like Polymerase Chain Reaction.
4. It is a semi-quantitative process that only provides estimated sizing of the DNA fragments.
5. Southern blotting is not a suitable method for detecting mutations at the base-pair level.
6. 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 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.

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.

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.

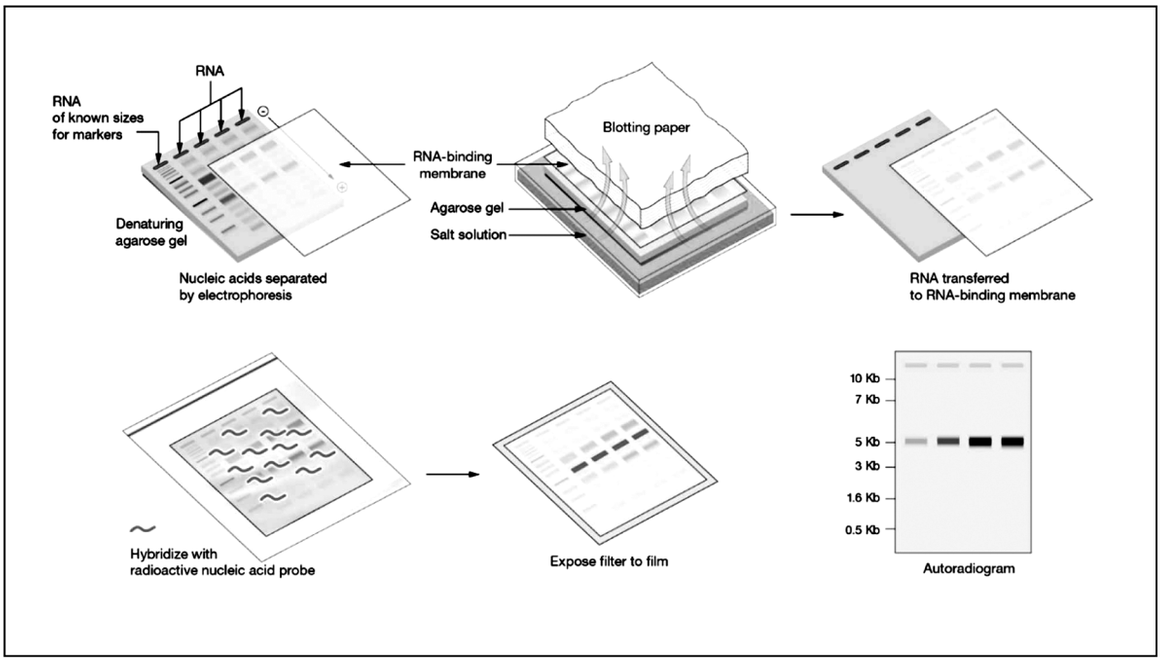


Fig. 18.3: Northern Blotting technique

[Source: https://www.genome.gov/genetics-glossary/Northern-Blot]

- 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 furtherincubated 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

1. The technique can be used for the identification and separation of RNA fragments collected from different biological sources.
2. 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.
3. It also allows the detection and quantification of specific mRNAs from different tissues and different living organisms.
4. Northern blotting is used as a tool for gene expression studies related to overexpression of cancer-causing genes, and gene expression during transplant rejects.
5. Northern blotting has been used as a molecular tool for the diagnosis of diseases like Crohn's disease.
6. The process is used as a method for the detection of viral microRNAs that play important roles in viral infection.

Limitations of Northern Blot

1. Northern blotting has a lower sensitivity as compared to other modern techniques like RT-PCR and nuclease protection assays.
2. The method requires a large amount of sample RNA, and these should be of high quality.
3. The technique is time-consuming and complex, especially in cases where multiple probes are to be added.

Western blotting

Introduction

Developed by Harry Towbin and his colleagues in 1979, the western blot procedure is a typical cell and molecular biology procedure widely used in the analysis of proteins. Often known as protein immunoblotting, the western blot method allows researchers

to determine the presence, size and quantity of particular proteins in a given sample. Three elements comprise the western blotting method:

- separation by size,
- transfer to a solid support or membrane, and
- marking target protein using a proper primary and secondary antibody to visualise.

western blotting procedure is very useful for identifying individual proteins from a complicated mixture of proteins isolated from cells that may have similar characteristics or sizes.

Principle:

- Western blotting technique is used for identification of particular protein from the mixture of protein.
- In this method labelled antibody against particular protein is used identify the desired protein, so it is a specific test. Western blotting is also known as immunoblotting because it uses antibodies to detect protein.

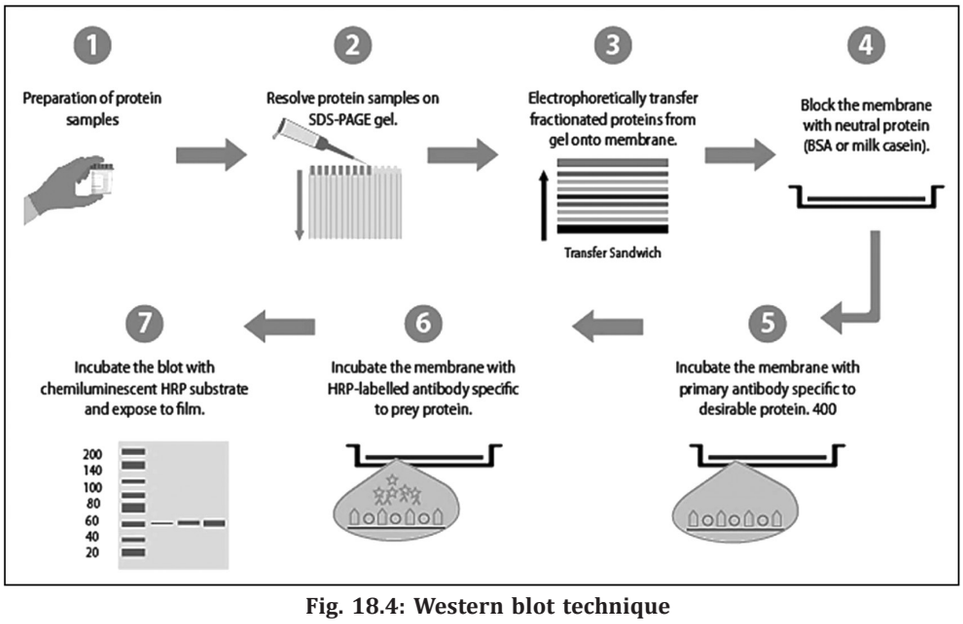


Fig. 18.4: Western blot technique
[Source: https://www.h-h-c.com/western-blot-procedures-analysis-and-purpose/]

Procedure/Steps:

1. Extraction of protein
2. Gel electrophoresis: SDS PAGE

3. Blotting: electrical or capillary blotting
4. Blocking: BSA
5. Treatment with primary antibody
6. Treatment with secondary antibody (enzyme labelled anti Ab)
7. Treatment with specific substrate; if enzyme is alkaline phosphatase, substrate is p- nitro phenyl phosphate which give color.

Step I: Extraction of Protein

- Cell lysate is most common sample for western blotting.
- Protein is extracted from cell by mechanical or chemical lysis of cell. This step is also known as tissue preparation.
- To prevent denaturing of protein protease inhibitor is used.
- The concentration of protein is determined by spectroscopy.
- When sufficient amount of protein sample is obtained, it is diluted in loading buffer containing glycerol which helps to sink the sample in well.
- Tracking dye (bromothymol blue) is also added in sample to monitor the movement of proteins.

Step II: Gel electrophoresis

- The sample is loaded in well of SDS-PAGE Sodium dodecyl sulfate- poly-acrylamide gel electrophoresis.
- The proteins are separated on the basis of electric charge, isoelectric point, molecular weight, or combination of these all.
- The smaller sized protein moves faster than large size protein.
- Proteins are negatively charged, so they move toward positive (anode) pole as electric current is applied.

Step III: Blotting

- The nitrocellulose membrane is placed on the gel. The separated protein from gel gets transferred to nitrocellulose paper by capillary action. This type of blotting is time consuming and may take 1-2 days.

- For fast and more efficient transfer of desired protein from the gel to nitrocellulose paper electro-blotting can be used.
- In electro-blotting the nitrocellulose membrane is sandwiched between gel and cassette of filter paper and then electric current is passed through the gel causing transfer of protein to the membrane.

Step IV: Blocking

- Blocking is very important step in western blotting.
- Antibodies are also protein so they are likely to bind the nitrocellulose paper. So before adding the primary antibody the membrane is non-specifically saturated or masked by using casein or Bovine serum albumin (BSA).

Step V: Treatment with Primary Antibody

- The primary antibody (1°Ab) is specific to desired protein so it forms Ag-Ab complex

Step VI: Treatment with secondary antibody

- The secondary antibody is enzyme labelled. For eg. alkaline phosphatase or Horseradish peroxidase (HRP) is labelled with secondary antibody.
- Secondary antibody (2°Ab) is antibody against primary antibody (anti-antibody) so it can bind with Ag-Ab complex.

Step VII: Treatment with suitable substrate

- To visualize the enzyme action, the reaction mixture is incubated with specific substrate.
- The enzyme converts the substrate to give visible colored product, so band of color can be visualized in the membrane.
- Western blotting is also a quantitative test to determine the amount of protein in sample.

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

1. Western blotting is an excellent method with high sensitivity in order to detect a particular protein even in low quantity.
2. 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.
3. The technique has been used to quantify proteins and other gene products in gene expression studies.
4. 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.
5. Western blotting is also used for the analysis of different biomarkers like growth factors, cytokines, and hormones.

Limitations of Western Blot

1. Since it is a very sensitive process, any imbalance in the process can affect the results of the entire process.
2. In some cases, no bands or erroneous bands might be observed due to the insufficient transfer of the proteins.
3. The test can only be used as a semi-quantitative test as the estimation is not always precise.
4. The process is time-consuming and complex, thus can only be performed by well-trained personnel.

5. Western blotting can only be performed for proteins if the primary antibodies for the proteins are available.
6. Some antibodies might exhibit off-target effects by interacting with more than one protein in the sample.
7. The technique is a costly process with the cost of antibodies and expensive detection methods.
8. Small proteins might not be retained by the membrane, whereas larger proteins are difficult to transfer to the membrane.

RFLP, RAPD and AFLP techniques

RFLP techniques

Introduction

Restriction fragment length polymorphism (RFLP) is a technique invented in 1984 by the English scientist Alec Jeffreys during research into hereditary diseases. It is used for the analysis of unique patterns in DNA fragments in order to genetically differentiate between organisms – these patterns are called Variable Number of Tandem Repeats (VNTRs).

Genetic polymorphism is defined as the inherited genetic differences among individuals in over 1% of normal population. The RFLP technique exploits these differences in DNA sequences to recognize and study both intraspecies and interspecies variation.

Principle

Restriction endonucleases are enzymes that cut lengthy DNA into short pieces. Each restriction endonuclease targets different nucleotide sequences in a DNA strand and therefore cuts at different sites.

The distance between the cleavage sites of a certain restriction endonuclease differs between individuals. Hence, the length of the DNA fragments produced by a restriction endonuclease will differ across both individual organisms and species.

Procedure

RFLP is performed using a series of steps briefly outlined below:

1) DNA Extraction

To begin with, DNA is extracted from blood, saliva or other samples and purified.

2) DNA Fragmentation

The purified DNA is digested using restriction endonucleases. The recognition sites of these enzymes are generally 4 to 6 base pairs in length. The shorter the sequence recognized, the greater the number of fragments generated from digestion.

For example, if there is a short sequence of GAGC that occurs repeatedly in a sample of DNA. The restriction endonuclease that recognizes the GAGC sequence cuts the DNA at every repetition of the GAGC pattern.

If one sample repeats the GAGC sequence 4 times whilst another sample repeats it 2 times, the length of the fragments generated by the enzyme for the two samples will be different.

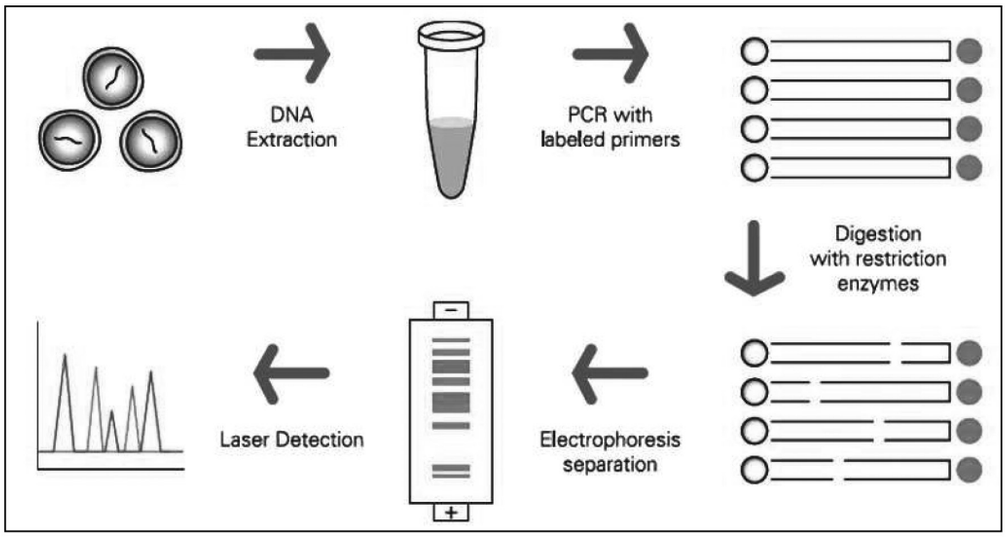


Fig. 18.5

3) Gel Electrophoresis

The restriction fragments produced during DNA fragmentation are analyzed using gel electrophoresis.

The fragments are negatively charged and can be easily separated by electrophoresis, which separates molecules based on their size and charge. The fragmented DNA samples are placed in the chamber containing the electrophoretic gel and two electrodes.

When an electric field is applied, the fragments migrate towards the positive electrode. Smaller fragments move faster through the gel leaving the larger ones behind and thus the DNA samples are separated into distinct bands on the gel.

4) Visualization of Bands

The gel is treated with luminescent dyes in order to make the DNA bands visible.

Applications of RFLP

Key applications of RFLP are listed below:

1. To determine the status of genetic diseases such as Cystic Fibrosis in an individual.
2. To determine or confirm the source of a DNA sample such as in paternity tests or criminal investigations.
3. In genetic mapping to determine recombination rates that show the genetic distance between the loci.
4. To identify a carrier of a disease-causing mutation in a family.

Disadvantages of RFLP

1. RFLP has been a widely used genome analysis techniques employed in forensic science, medicine, and genetic studies. However, it has become almost obsolete with the advent of relatively simple and less expensive DNA profiling technologies such as the polymerase chain reaction (PCR).
2. The RFLP procedure requires numerous steps and takes weeks to yield results, while techniques such as PCR can amplify target DNA sequences in a mere few hours.
3. Additionally, RFLP requires a large DNA sample, the isolation of which can be a laborious and time-consuming process. In contrast, PCR can amplify minute amounts of DNA in a matter of hours.

RAPD techniques

Introduction

Random Amplified Polymorphic DNA is molecular technique which allows evolutionary biologists to determine the genetic characters of a different variety of closely related animals. It is a type of PCR reaction, but the segments of DNA that are amplified are random.The discovery of the RAPD had a major impact on the research of eukaryotic genomes and contributed to the development of various DNA markers. In RAPD, no knowledge of the DNA sequence for the targeted gene is required, as the primers will bind somewhere in the sequence, but it is not certain exactly where. RAPDs are particularly useful to study the genetic structure of populations because they reveal polymorphisms in non-coding regions of the genome (Caccon et al, 1997).

The RAPD markers method has been reported to be an efficient tool to differentiate geographically and genetically isolated population, and has been used to verify the existence of population of species that have arisen either through genetic selection under different environmental conditions or as a result of genetic drift (Rajesh et al. 2017).

Principle of the rapd technique

The principle is that, a single, short oligonucleotide primer, which binds to many different loci, is used to amplify random sequences from a complex DNA template. This means that the amplified fragment generated by PCR depends on the length and size of both the primer and the target genome. The assumption is made that a given DNA sequence (complementary to that of the primer) will occur in the genome, on opposite DNA strands, in opposite orientation within a distance that is readily amplifiable by PCR. These amplified products (of up to 3.0 kb) are usually separated on agarose gels (1.5-2.0%) and visualised by ethidium bromide staining. The use of a single decamer oligonucleotide promotes the generation of several discrete DNA products and these are considered to originate from different genetic loci. Polymorphisms result from mutations or rearrangements either at or between the primer binding sites and are detected as the presence or absence of a particular RAPD band (Fig.). This means that RAPDs are dominant markers and, therefore, cannot be used to identify heterozygotes.

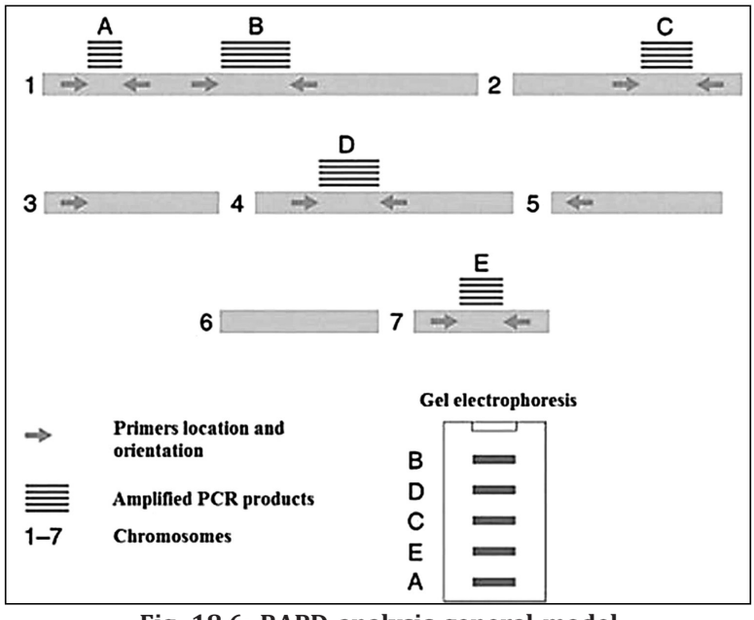


Fig. 18.6: RAPD analysis general model
[www.usask.ca/.../pawlin/resources/rapds.html] (adapted from Griffiths et al.)

Applications:

RAPD has been used for a variety of applications, including:

- **Genetic diversity analysis:** Identifying differences in DNA between individuals or populations.
- **Gene mapping:** Creating genetic maps by identifying the location of RAPD markers on chromosomes.
- **Species identification:** Differentiating between different species or strains.
- **Plant and animal breeding:** Selecting for desirable traits by identifying RAPD markers associated with those traits.

Limitations:

- RAPD markers are often dominant, meaning that it can be difficult to distinguish between heterozygous and homozygous genotypes.
- RAPD markers can be sensitive to PCR conditions, which can lead to reproducibility issues.
- RAPD markers are not always mappable, meaning that a polymorphism detected between one pair of strains may not be useful for another pair of strains.

AFLP techniques

Principle

AFLP is a powerful molecular tool allowing simultaneous screening of a large number of DNA regions that are spread randomly throughout the genome. One of the key features of AFLP analysis is the detection of polymorphisms of genomic DNA fragments obtained through restriction digestion. Further more, AFLP markers are very useful for studying genetic variation between closely related species. “For a wide range of taxa, including plants, fungi, animals and bacteria, AFLP markers have been used to uncover cryptic genetic variation of strains, or closely related species, that had been impossible to resolve with morphological or other molecular systematic characters” (Mueller et al). Besides detecting fragment polymorphisms, “The AFLP technique may be used for fingerprinting of cloned DNA segments like cosmids, P1 clones, bacterial artificial chromosomes or yeast artificial chromosomes (Vos et al). AFLP can also be applied to criminal investigations and paternity testing.

Steps

AFLP is a DNA fingerprinting technique that can be used to answer a variety of research questions. This PCR based tool utilizes restriction enzymes to digest genomic DNA to smaller fragments and consists of three main steps. The first step involves restriction digestion of DNA with two types of enzymes (a rare and a frequent cutter), followed by the ligation of double-stranded oligonucleotide adapters, specific to the ends of the generated fragments. There are several reasons for the use of two types restriction enzymes: generation of short DNA segments with good amplification and separation on the gel, the rare cutter will complement the frequent cutter by decreasing the number of generated fragments, and only one strand of the double stranded PCR products can be labeled. The second step consists of the amplification of certain restriction fragments with PCR primers (the primers are designed to target the adapters fused to the end of the fragments). And the third step is the analysis of the amplified fragments after visualization on a denaturing polyacrylamide gel.

Applications

- To detect various polymorphisms in different genomic regions.
- For identification of **genetic variation** in strains or closely related species of plants, fungi, animals and bacteria.
- AFLP technique has been used in criminal and paternity tests to determine slight differences within populations and in linkage studies to generate maps for **QTL**(Quantitative trait locus) analysis.

Advantages of AFLP marker

- As restriction sites are present across the whole genome of an individual which makes AFLP marker to analyse **multiple locus** at once.
- The sequence information about the organism is not essential as the primers complementary to the adapter sequences are designed.
- In contrast to RFLP which takes longer time for probe hybridization and more skills, AFLP is **comparatively simple** as PCR amplification of fragments is done
- AFLP is possible with **lesser** amount of **genomic template**
- The results are highly reproducible considering have a high quality of DNA as input.

Disadvantages of AFLP

- AFLP cannot be done with **poor quality of DNA** or degraded DNA
- As AFLP are **dominant markers** in nature they cannot detect **homozygous** or **heterozygous** individuals.
- One cannot ascertain which fragment belongs to which dna locus as AFLP are multi-locus in nature.

Probable Questions:

1. What is called blotting?
2. Mention basic principle of Southern blotting?
3. Elastrate different steps of Southern Blotting with a suitable diagram.
4. Why is NaOH used in southern blotting?
5. What is the purpose of southern blotting?
6. What are the applications of Southern Blotting?
7. State the limitations of Southern Blot.
8. Who developed the northern blot technique?
9. Write down the basic principle of Northern blotting?
10. Describe different steps of Northern Blotting with a suitable diagram.
11. Which buffer is used in Northern blotting?
12. What is the main purpose of the northern blot technique in molecular biology research?
13. What are the applications of Northern Blotting?
14. Write down the basic principle of Western blotting?
15. What is the primary detection method used in northern blotting?
16. Describe different steps of Western Blotting with a suitable diagram.
17. What are the applications of Western Blotting?
18. Compare Southern, Northern and Western Blotting techniques.
19. What are the techniques of RFLP?
20. Short notes on RAPD.
21. Write the disadvantages of AFLP.

Suggested Readings:

1. Wilson and Walker: Principle and technique of biochemistry and molecular biology, 7th Edition.
2. Biophysical Chemistry by Upadhyay and Upadhyay.

References:

https://www.genome.gov/genetics-glossary/Southern-Blot

https://www.mybiosource.com/learn/southern-blotting/

Southern Blotting.Green MR, Sambrook J.Cold Spring Harb Protoc. 2021 Jul 1;2021(7). doi: 10.1101/pdb.prot100487.PMID: 34210769

https://kvmwai.edu.in/upload/StudyMaterial/Southern_blot.pdf

https://kvmwai.edu.in/upload/StudyMaterial/Northern_blot.pdf

https://kvmwai.edu.in/upload/StudyMaterial/Western_blot.pdf

https://www.nlm.nih.gov/visibleproofs/education/dna/rflp.pdf

http://www.bio.davidson.edu/courses/genomics/method/rflp.html

Random amplified polymorphic DNA (RAPD) based molecular characterization of evolutionary advancement amongst hierarchal five vertebrate species. Rajesh B, Ramesh B et. al. Int J Anat Res 2017, Vol 5(2.1):3687-92. ISSN 2321-4287

Vos, Pieter. "AFLP: a new technique for DNA fingerprinting." *Nucleic Acids Research* 25 (1995): 4407-414. Print.

https://www.biotechfront.com/2021/04/aflp-marker-principle-steps.html

Unit-XIX
Pesticide Formulation

1. Introduction to Pesticide Formulation

Pesticides play a pivotal role in modern agriculture by preventing, destroying, or controlling pests that damage crops, livestock, and human health. However, pesticides, in their pure (technical) form, are often unsuitable for direct application because they may be chemically unstable, difficult to handle, insoluble in water, or too concentrated. Pesticide formulation refers to the process of blending active ingredients (AIs) with other materials (adjuvants) to improve their performance, ease of application, stability, safety, and effectiveness.

A formulation thus combines the active pesticide chemical with inert substances (carriers, surfactants, solvents) to produce a product that is easy to measure, mix, and apply. *It is the process by which the active ingredient is mixed with inert substances to yield a final product that is stable, effective, easy to apply, and safe to handle.*

2. Objectives of Pesticide Formulation

The major goals of pesticide formulation are:

- Ease of handling: Making pesticides easier and safer to use.
- Improved application: Enhancing the distribution and adherence of the pesticide to the target area.
- Stability enhancement: Preventing decomposition during storage and use.
- Controlled release: Modifying the release rate of active ingredients for longer or targeted action.
- Compatibility: Ensuring compatibility with other agricultural products (e.g., fertilizers).
- Safety: Reducing risks to users, non-target organisms, and the environment (Matthews, 2008).

3. Components of a Pesticide Formulation

A pesticide formulation generally consists of the following:

3.1 Active Ingredient (AI)

The chemical substance that controls the pest. It is responsible for the pesticidal effect.

3.2 Inert Ingredients

These include:

- Carriers: Substances that dilute the AI (e.g., water, clay, talc).
- Surfactants: Surface-active agents that improve wetting, spreading, and penetration.
- Solvents: Liquids used to dissolve the AI.
- Adjuvants: Substances added to enhance the effectiveness or application characteristics.
- Stabilizers: Agents that prevent chemical degradation.
- Safeners: Compounds that protect crops without reducing pesticide activity.

4. Types of Pesticide Formulations

Pesticide formulations are classified broadly into solid and liquid types. Specialized types such as gaseous formulations also exist.

4.1 Solid Formulations

- **Dusts (D):** Finely ground powders containing 0.1–10% active ingredient. No need for mixing.
- **Granules (G):** AI mixed with large inert particles, suitable for soil application.
- **Wettable Powders (WP):** Powders that are mixed with water to form a suspension.
- **Water-Dispersible Granules (WDG) or Dry Flowables (DF):** Granules that disintegrate into fine particles when added to water.

4.2 Liquid Formulations

- **Emulsifiable Concentrates (EC):** AI dissolved in oil with emulsifiers; forms an emulsion when added to water.

- **Suspension Concentrates (SC):** AI particles suspended in a liquid carrier.
- **Soluble Liquids (SL):** AI completely dissolved in a solvent.

4.3 Gaseous Formulations

- **Fumigants:** Volatile substances that release gas to control pests in soil, stored products, or structures (e.g., methyl bromide).

5. Specialized Formulations

With the advancement of technology, specialized pesticide formulations have been developed:

5.1 Microencapsulated Formulations (ME): Active ingredients are enclosed in microscopic capsules, allowing slow and controlled release.

5.2 Ultra-Low Volume (ULV) Formulations: Highly concentrated pesticides intended to be applied in very small amounts without dilution.

5.3 Controlled Release Formulations: Designed to release the AI over an extended period, reducing the need for repeated applications and lowering environmental impacts.

6. Importance of Adjuvants in Pesticide Formulation

Adjuvants improve the performance of the pesticide by modifying physical properties or enhancing biological effectiveness:

- Spreaders: Allow pesticide to spread more evenly on surfaces.
- Stickers: Help pesticide adhere to surfaces longer.
- Penetrants: Facilitate entry of pesticides into plant tissues.
- Buffers: Maintain optimal pH conditions for pesticide stability.

7. Factors Affecting Formulation Selection

Several factors influence the choice of formulation:

- Target pest and crop type
- Application method (e.g., aerial, ground sprayers)
- Environmental conditions (e.g., temperature, rainfall)

- Safety requirements for applicators
- Storage and transport facilities
- Cost considerations

8. Safety and Environmental Considerations

Formulation science aims to minimize environmental contamination and user exposure. Innovations focus on:

- Reduced toxicity
- Better biodegradability
- Reduced volatilization and drift
- Precision application technologies

For example, nano-formulations are being explored for targeted delivery with minimal environmental footprint.

9. Future Trends in Pesticide Formulation

Emerging trends include:

- Nano-pesticides: Use of nanotechnology to enhance delivery and effectiveness.
- Biopesticide formulations: Use of naturally derived substances like microbial or botanical pesticides.
- Eco-friendly formulations: Focused on biodegradable carriers and minimal residual toxicity.

Research continues to develop “smart formulations” that are environmentally sustainable and highly effective against pests.

Conclusion

Pesticide formulation is a crucial field that merges chemistry, environmental science, and agricultural engineering to optimize pest control while ensuring safety and sustainability. Understanding the principles and applications of different formulations helps in responsible pesticide management, contributing to agricultural productivity and environmental protection.

Aspect	Details
Definition	Process of mixing active ingredients (AIs) with inert substances to improve effectiveness, stability, ease of application, and safety.
Objectives	Ease of handling, improved application, enhanced stability, controlled release, compatibility with other inputs, user safety.
Main Components	-Active Ingredient (AI) - Inert Ingredients (carriers, surfactants, solvents, stabilizers, safeners)
Types of Formulations	Solid Formulations: Dusts (D), Granules (G), Wettable Powders (WP), Water-Dispersible Granules (WDG) Liquid Formulations: Emulsificable Concentrates (EC), Suspension Concentrates (SC), Soluble Liquids (SL) Gaseous Formulations: Fumigants
Specialized Formulations	-Microencapsulated Formulations (ME) -Ultra-Low Voluem (ULV) -Controlled Release Formulations
Role of Adjuvants	-Spreaders, stickers, penetrants, buffers -Enhance application efficiency and pesticide performance.
Factors Influencing Formulation Selection	Target pest/crop type, application methods, environment, applicator safety, storage/transport needs, cost.
Safety and Environmental Considerations	Reduced toxicity, biodegradability, minimized volatilization and drift, precision application.
Future Trends	Nano-pesticides, biopesticide formulations, eco-friendly and biodegradable carriers, smart formulations.

Probable Questions

1. Define pesticide formulation. Discuss the necessity and objectives of formulating pesticides instead of using technical-grade active ingredients directly.
2. Explain the various components of a pesticide formulation and elaborate on the roles of each component with examples.
3. Classify the different types of pesticide formulations and compare their properties, applications, and limitations.
4. Describe the formulation and applications of solid pesticide formulations such as dusts, granules, wettable powders, and water-dispersible granules.

5. Differentiate between emulsifiable concentrates (EC), suspension concentrates (SC), and soluble liquids (SL) in terms of formulation chemistry and field application.
6. What are microencapsulated formulations and ultra-low volume (ULV) formulations? Discuss their advantages and significance in modern pest management.
7. Elucidate the importance of adjuvants in pesticide formulations. Classify different adjuvants and explain their functional mechanisms.
8. What are the key factors influencing the selection of a pesticide formulation for specific crop and pest conditions?
9. Discuss the safety and environmental considerations involved in pesticide formulation. How do modern formulations address toxicity and ecological concerns?
10. Outline and critically assess the future trends in pesticide formulation technology, including nano-pesticides and biopesticide formulations.

Suggested Readings

1. Ware, G.W., & Whitacre, D.M. (2004). *The Pesticide Book* (6th ed.). MeisterPro Information Resources, Willoughby, OH, USA.
2. Matthews, G. A. (2008). *Pesticide Application Methods* (3rd ed.). Wiley-Blackwell.
3. Green, J.M., & Foy, C.L. (2000). *Adjuvants: Testing and Mechanisms of Action*. Weed Technology, 14(4), 819–825.
4. Kah, M., Beulke, S., Tiede, K., & Hofmann, T. (2012). *Nanopesticides: State of Knowledge, Environmental Fate, and Exposure Modeling*. Critical Reviews in Environmental Science and Technology, 43(16), 1823–1867.

Unit-XX

Database search tool; Sequence alignment and database searching; Computational tools and biological databases, NCBI, EMBL, PDB, Sequence similarity tools; Blast and FASTA Phylogenetic analysis with the program PHYLIP, DISTANCES, and GROWTREE. Basics of designing a microarray, image analysis and normalization, annotations

1. Database Search Tool

Introduction

In the current era of biological sciences, the amount of biological data has increased exponentially. Database search tools have been developed to help researchers retrieve specific information from vast databases efficiently. Database search tools are software applications or web interfaces that allow users to query and retrieve biological information such as DNA sequences, protein structures, literature, and more.

Key Concepts

- **Database**: An organized collection of data that is easily accessible, managed, and updated.
- **Search Tool**: A software or interface used to locate data within a database using specific queries.

Types of Database Search Tools

1. **Text-based search tools**
    - Retrieve information based on keywords.
    - Example: PubMed search engine for biomedical literature.
2. **Sequence-based search tools**
    - Retrieve biological sequences similar to a query sequence.
    - Example: BLAST (Basic Local Alignment Search Tool).

3. **Structure-based search tools**
    - Retrieve molecular structures based on chemical or protein 3D configurations.
    - Example: RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) search tools.

Important Database Search Tools

Tool	Description	Application
BLAST	Compares nucleotide or protein sequences to sequence databases	Identifying homologous sequences
FASTA	Sequence alignment search tool	Finding similar sequences quickly
HMMER	Searches sequences databases for homologs using probabilistic models (Hidden Markov Models)	Domain detection and functional annotation
Entrez	Search engine for integrated databases at NCBI (National Center for Biotechnology Information)	Search across genomic, proteomic, and literature databases
SRS (Sequence Retrieval System)	Indexes and retrieves biological data	Cross-database searches

Search Strategies

- Use specific keywords for text-based databases.
- Use sequence similarity thresholds (e.g., E-value in BLAST).
- Choose appropriate databases based on research needs.

Conclusion

Understanding the functionality of database search tools is crucial for biological research. They allow efficient mining of biological information, forming the basis for further experimental or computational studies.

2. Sequence Alignment and Database Searching

Introduction

Sequence alignment is a method to arrange sequences of DNA, RNA, or proteins to identify regions of similarity that may indicate functional, structural, or evolutionary

relationships. When aligned sequences are compared to databases, it is called **database searching**.

Key Concepts

- **Sequence Alignment**: Arranging sequences to identify similar regions.
- **Database Searching**: Comparing a query sequence against a database to find similar entries.

Types of Sequence Alignment

1. **Pairwise Alignment**:
    - Comparison between two sequences.
    - Can be global (full length) or local (sub-regions).
2. **Multiple Sequence Alignment (MSA)**:
    - Aligns more than two sequences simultaneously.
    - Useful for identifying conserved motifs or domains.

Common Sequence Alignment Algorithms

Algorith	Description	Application
Needleman-Wunsch Algorithm	Global alignment	Aligns entire sequences end-to-end
Smith-Waterman Algorithm	Local alignment	Finds optimal sub-sequence alignment
ClustalW/Clustal Omega	Multiple sequence alignment	Detects conserved regions among multiple sequences

Database Searching through Sequence Alignment

- **BLAST**: Fast local alignment; provides E-value and bit score.
- **PSI-BLAST (Position-Specific Iterated BLAST)**: Detects distant relatives using profile-based searches.
- **TBLASTN**: Searches translated nucleotide databases using protein queries.

Important Terms

- **E-value**: Expectation value; lower E-value indicates more significant matches.
- **Bit Score**: Indicates the quality of sequence alignment.

Example Workflow

1. Submit query sequence in BLAST.
2. Database searches for local alignments.
3. Result provides hits ranked by E-value.
4. Analyze the top hits for further research.

Conclusion

Sequence alignment and database searching form the backbone of comparative genomics, evolutionary studies, and functional annotations.

3. Computational Tools and Biological Databases

Introduction

Computational tools and biological databases provide the fundamental resources required for modern biological research. Key among these are **NCBI (National Center for Biotechnology Information)**, **EMBL (European Molecular Biology Laboratory)**, and **PDB (Protein Data Bank)**.

National Center for Biotechnology Information (NCBI)

- **Established**: 1988.
- **Function**: Curates genomic information, literature, and bioinformatics tools.
- **Major Resources**:
    - **GenBank**: Repository of nucleotide sequences.
    - **PubMed**: Biomedical literature database.
    - **BLAST**: Sequence comparison tool.
    - **Gene**: Database of gene-specific information.
    - **Protein**: Repository of protein sequences and functional information.

European Molecular Biology Laboratory (EMBL)

- **Established**: 1974.
- **Function**: Europe's flagship life sciences laboratory with focus on molecular biology research and data services.

- **Major Resources**:
    - **EMBL-EBI (European Bioinformatics Institute)**: Hosts databases like ENA (European Nucleotide Archive) and UniProt (protein sequence and annotation database).
    - **EMBL-GenBank-DDBJ**: International nucleotide sequence collaboration.

Protein Data Bank (PDB)

- **Established**: 1971.
- **Function**: Archive for 3D structural data of proteins and nucleic acids.
- **Major Resources**:
    - **RCSB PDB**: Research Collaboratory for Structural Bioinformatics manages PDB in the USA.
    - **PDBe**: PDB Europe.
    - **PDBj**: PDB Japan.
- **Data Types**:
    - Protein structures.
    - Nucleic acid structures.
    - Small molecule structures bound to macromolecules.

Important Computational Tools

Tool	Purpose	Database
BLAST	Sequence similarity search	NCBI
Clustal Omega	Multiple sequence alignment	EMBL-EBI
PyMOL	Molecular visualization	PDB structures
Swiss-PDB Viewer	Protein structure modeling and visualization	PDB

Data Access and Usage

- **NCBI Entrez system** allows integrated searching across multiple datasets.
- **EMBL-EBI search tools** provide access to bioinformatics resources for sequence alignment, protein structure prediction, and literature mining.
- **PDB search** allows retrieval of structure files in formats like PDB, mmCIF, and visualization through integrated viewers.

Conclusion

NCBI, EMBL, and PDB serve as cornerstones for biological data storage, access, and analysis. Their computational tools facilitate crucial bioinformatics operations ranging from basic sequence retrieval to complex structure-function studies.

Summary

Topic	Highlights
Database Search Tool	Helps in retrieving sequence, structure, and literature information efficiently.
Sequence Alignment and Database Searching	Basis for comparative genomics, evolutionary studies, and identification of homologous sequences.
Computational Tools and Biological Databases	Essential resources like NCBI, EMBL, and PDB provide curated biological data and analytical tools.

4. Sequence Similarity Tools: BLAST and FASTA

Introduction

Sequence similarity searching is one of the cornerstones of bioinformatics. It allows researchers to find similar sequences across huge databases, providing crucial insights into biological function, evolutionary relationships, and gene annotation. Two of the most widely used sequence similarity tools are **BLAST (Basic Local Alignment Search Tool)** and **FASTA (Fast-All Alignment Search Tool)**.

Concept of Sequence Similarity

- **Sequence Similarity**: Refers to the percentage of aligned nucleotides or amino acids that are identical or similar between two sequences.
- **Homology**: Inferred from high similarity; two sequences are homologous if they share a common ancestor.
- **Importance**: Identifying conserved domains, predicting protein structure and function, annotating new genomes, and studying evolutionary relationships.

BLAST (Basic Local Alignment Search Tool)

Introduction

BLAST is a heuristic search tool that finds regions of local similarity between a query sequence and database sequences. It is faster than exhaustive search methods because it approximates the best alignments rather than finding the absolute best.

Developed by: Stephen Altschul and colleagues in 1990.

Key Features:

- Focuses on local alignments.
- Extremely fast for large datasets.
- Provides **E-value** (expectation value) for significance testing.
- Multiple versions available: **BLASTn**, **BLASTp**, **BLASTx**, **tBLASTn**, **tBLASTx**.

Basic Working Principle:

1. Break the query sequence into small words (e.g., triplets).
2. Search database for exact matches to these words.
3. Extend matches in both directions to find High Scoring Pairs (HSPs).
4. Report alignments above a certain threshold.

Advantages:

- Rapid searches.
- Handles large datasets efficiently.
- Provides statistical significance for alignments.

Limitations:

- Might miss weaker alignments.
- Less sensitive than some slower exhaustive algorithms.

BLAST Output Elements:

- Score: Indicates the quality of the match.
- E-value: Probability of finding a match by chance.
- Identities: Exact matches.
- Gaps: Insertions or deletions introduced to maximize alignment.

FASTA (Fast-All Alignment Search Tool)

Introduction

FASTA was one of the first widely used programs for sequence similarity searching. Developed by **William Pearson and David Lipman in 1985**, FASTA finds regions of similarity between sequences by performing an initial fast search for short matches.

Key Features:

- Finds regions of local similarity.
- Conducts initial seed word matches.
- Extends alignments and refines them.
- Provides alignment scores and statistical significance.

Basic Working Principle:

1. Identify exact matches (k-tuples) between query and database sequences.
2. Calculate initial similarity scores for k-tuple matches.
3. Perform a banded Smith-Waterman alignment around high-scoring regions.

Advantages:

- More sensitive than BLAST for certain applications.
- Useful for DNA and protein sequences.

Limitations:

- Slightly slower than BLAST.
- May produce more false positives at lower similarity levels.

FASTA Output Elements:

- Initial score and refined alignment score.
- E-value and statistical significance.
- Alignment visualization showing matching regions.

Comparison between BLAST and FASTA

Feature	BLAST	FASTA
Speed	Faster	Slower
Sensitivity	Slightly lower	Higher
Algorithm	Heuristic	Heuristic + Smith-Waterman refinement
Use	DNA and protein	DNA and protein
Popularity	Higher	Moderate

Conclusion

Both BLAST and FASTA are invaluable tools for sequence similarity analysis. The choice between them often depends on the balance between speed and sensitivity required for the research.

5. Phylogenetic Analysis with the Programs PHYLIP, DISTANCES, and GROWTREE

Introduction

Phylogenetic analysis aims to reconstruct evolutionary relationships among species or genes. Computational tools like **PHYLIP (Phylogeny Inference Package)**, **DISTANCES**, and **GROWTREE** enable researchers to create and visualize phylogenetic trees based on sequence data.

Key Concepts in Phylogenetics

- **Phylogenetic Tree**: A diagram that represents evolutionary relationships.
- **Distance Matrix**: A table showing genetic distances between sequences.
- **Clustering Methods**: Algorithms used to create trees based on distance matrices (e.g., Neighbor-Joining, UPGMA).
- **Evolutionary Model**: Mathematical descriptions of sequence evolution (e.g., Jukes-Cantor, Kimura 2-parameter).

PHYLIP (Phylogeny Inference Package)

Introduction

PHYLIP is one of the oldest and most comprehensive software packages for phylogenetic analysis. Developed by **Joseph Felsenstein** at the University of Washington.

Key Features:

- Suite of over 30 programs for different phylogenetic analyses.
- Supports distance matrix, maximum likelihood, parsimony, and bootstrapping.
- Command-line based.

Main Programs in PHYLIP:

- **DNADIST**: Calculates distance matrices for DNA sequences.
- **PROTDIST**: Calculates distance matrices for protein sequences.
- **NEIGHBOR**: Constructs trees using Neighbor-Joining or UPGMA methods.
- **CONSENSE**: Builds consensus trees from multiple runs.
- **SEQBOOT**: Generates bootstrap replicates.

Basic Workflow:

1. Sequence input (usually in PHYLIP format).
2. Calculate distance matrix using **DISTANCES** (e.g., DNADIST or PROTDIST).
3. Build a tree using **GROWTREE** (e.g., NEIGHBOR).
4. Visualize the tree.

DISTANCES Program

Introduction

DISTANCES refers generally to programs within PHYLIP that compute distance matrices, particularly **DNADIST** for nucleotide sequences.

Working:

- Input: Aligned sequences.
- Output: Pairwise distance matrix showing evolutionary distances.

Methods in DISTANCES:

- Jukes-Cantor model (assumes equal substitution rates).
- Kimura 2-parameter model (accounts for transitions and transversions separately).
- F84 model (accounts for base composition biases).

GROWTREE Program

Introduction

GROWTREE is the process or program responsible for constructing trees based on distance matrices.

Programs Involved:

- **NEIGHBOR**: Uses the Neighbor-Joining algorithm (preferred for phylogenetic inference).
- **UPGMA (Unweighted Pair Group Method with Arithmetic Mean)**: Assumes a constant rate of evolution (molecular clock hypothesis).

Working:

- Takes distance matrix as input.
- Constructs tree topology minimizing total branch length (for NJ).
- Outputs tree in Newick format or PHYLIP tree format.

Example Workflow Using PHYLIP Suite

1. Align sequences using external tool (e.g., ClustalW).
2. Input aligned file into **DNADIST** to calculate distances.
3. Input distance matrix into **NEIGHBOR** to build the tree.
4. Use **CONSENSE** to produce a consensus tree if bootstrapping.
5. Visualize with external tools (e.g., FigTree, iTOL).

Applications of PHYLIP, DISTANCES, and GROWTREE

- Molecular systematics.
- Comparative genomics.
- Studying gene family evolution.
- Viral evolutionary analysis (e.g., COVID-19 strains).

Conclusion

PHYLIP and its associated programs, DISTANCES and GROWTREE, remain critical tools for reconstructing phylogenetic relationships. Despite newer graphical tools, PHYLIP's robustness and flexibility ensure its continued relevance.

6. Basics of Designing a Microarray, Image Analysis, and Normalization, Annotations

Introduction

In the era of high-throughput biology, **DNA microarrays** have emerged as a revolutionary technology to monitor gene expression patterns of thousands of genes simultaneously. Designing a microarray experiment, analyzing the resulting images, normalizing the data to eliminate technical biases, and accurately annotating the resulting data are essential steps for obtaining biologically meaningful results.

1. Basics of Designing a Microarray

Definition

- A **microarray** is a collection of microscopic DNA spots attached to a solid surface (e.g., glass slide, silicon chip).
- Each spot contains **a specific DNA probe** that can hybridize with a complementary DNA or RNA target sequence.
- By measuring the hybridization intensity, we can infer gene expression levels or detect polymorphisms.

Full Forms:

- **cDNA**: Complementary DNA
- **RNA**: Ribonucleic Acid
- **DNA**: Deoxyribonucleic Acid

Types of Microarrays

- **cDNA Microarrays**: Use PCR-amplified cDNA clones as probes.
- **Oligonucleotide Microarrays**: Use synthetic short sequences (25-80 mers).
- **SNP Microarrays**: Detect single nucleotide polymorphisms.
- **Tiling Arrays**: Cover genomic regions at high resolution.

Key Steps in Microarray Designing

a) Probe Design

- Design probes that are specific and have minimal cross-hybridization.
- Avoid repetitive regions.
- Optimize GC content (~40-60%) for uniform melting temperatures.

b) Surface Chemistry

- The surface (e.g., poly-lysine coated glass, aldehyde slides) must allow for efficient immobilization of probes.

c) Spotting/Printing

- Robotic printers deposit nanoliter quantities of probes on the array surface.
- Precise spotting ensures reproducibility and minimal spot size variability.

d) Hybridization

- Labelled target molecules are hybridized to the array.
- Conditions such as temperature, ionic strength, and washing stringency are critical.

e) Detection

- Fluorescent dyes (e.g., Cy3, Cy5) are commonly used.
- Laser scanners capture fluorescent signals from each spot.

2. Image Analysis and Normalization

Image Analysis

After hybridization and scanning, the resulting microarray image contains intensity values that must be extracted and processed.

Steps in Image Analysis:

1. **Spot Detection**: Locate each spot on the grid.
2. **Quantification**:
    - Measure intensity at each spot (signal).
    - Subtract local background fluorescence.
3. **Quality Control**:
    - Identify poorly hybridized or damaged spots.
    - Flag unreliable data points.

Software Tools:

- **GenePix Pro**
- **ScanArray**
- **Agilent Feature Extraction Software**

Important Terms

- **Foreground Intensity**: Signal from the hybridized spot.
- **Background Intensity**: Signal from the surrounding area.
- **Signal-to-Noise Ratio (SNR)**: Measure of signal quality.

Normalization

Definition:

Normalization is the process of adjusting data to remove systematic technical variations, allowing meaningful biological comparisons.

Why is Normalization Important?

- Corrects for dye bias (e.g., Cy3 vs. Cy5 efficiency differences).
- Corrects for differences in hybridization conditions across arrays.
- Makes data across experiments comparable.

Types of Normalization

Type	Description	Examples
Global Normalization	Adjusts all data based on overall mean or median intensity.	Median normalization
Lowess (Locally Weighted Scatterplot-Smoothing)	Corrects intensity-dependent dye biases.	Within-array normalization
Quantile Normalization	Forces distribution of proble intensities across arrayw to be the same.	Between-array normalization
Scaling Normalization	Multiples intensities by a constant factor.	Total intensity scaling

Standard Normalization Workflow

1. Raw intensity extraction.
2. Background subtraction.
3. Within-array normalization (e.g., dye-bias correction).
4. Between-array normalization (e.g., quantile normalization).

3. Annotations

What is Annotation?

- **Annotation** in the context of microarrays refers to linking each probe (or spot) to biological information.
- This includes:

- Gene name
- Gene function
- Pathways
- Ontologies (e.g., GO - Gene Ontology terms)

Importance of Accurate Annotation

- Facilitates biological interpretation.
- Essential for downstream functional enrichment and pathway analysis.

Annotation Databases and Tools

Database/Tool	Purpose
NCBI Gene	Gene-centric information
Ensembl	Genomic annotations
UniProt	Protein-centric annotations
DAVID (Database for Annotation, Visualization, and Integrated Discovery)	Functional annotation clustering
GO (Gene Ontology Consortium)	Standardized functional classification

Updating Annotations

- Genome annotations continuously evolve.
- Regular updates ensure accuracy and prevent analysis based on outdated or obsolete probe IDs.

Annotation Pipelines:

- Mapping probe IDs to updated databases.
- Manual curation if automated pipelines fail.

Summary Table

Step	Key Points
Microarry Designing	Careful probe design, surface preparation, spotting, and hybridization
Image Anslysis	Spot detection, intensity extraction, quality control
Normalization	Correcting technical variation (global, Lowess, quantile normalization)
Annotations	Linking proble data to biological knowledge (NCBL, Ensembl, GO)

Conclusion

Microarrays provide an efficient platform for studying gene expression on a genomic scale. However, designing high-quality arrays, conducting rigorous image analysis, applying robust normalization techniques, and performing accurate annotation are crucial for obtaining biologically valid and interpretable results. Mastery of these fundamentals forms the basis of reliable microarray data analysis and interpretation.

Tabular Summary of the Main Concepts

Features	Key Concepts	Important Tools/Resources
Database Search Tool	Efficient retrieval of biological information (sequences, structures, literature) using text-based, sequence-based, and structure-based tools.	PubMed, BLAST, HMMER, Entrez, SRS
Sequence Allignment and Database Searching	Comparising sequences to find region of similarity, using pairwise or multiple sequence alignment, followed by data base searching	Needleman-Wunsch, Smith-Waterman, ClustalW, Clustal Omega
Computational Tools and Biological Databases (NCBI, EMBL, PDB)	Using major databases like NCBI, EMBI, and PDB for sequences, structure data, and associated computational tools (BLAST, Clustal Omega, PyMOI, Swiss-PDB Viewer).	NCBI (GenBank, PubMed, BLAST, Gene, Protein), EMBL-EBI (ENA, UniProt), RCSB PDB
Sequence Similarity Tools: BLAST and FASTA	BLAST: Fast heuristic local alignment; FASTA: More sensitive, uses k-tuple matches and Smith-Waterman refinement.	BLAST variants (BLASTn, BLASTp, BLASTx), FASTA
Phylogenetic Analysis with PHYLIP, DISTANCES, and GROWTREE	Creating evolutionary trees by computing distance matrices (DNADIST, PROTDIST) and building trees using methods like Neighbor-Joing and UPGMA.	PHYLIP suite (DNADIST, PROTDIST, NEIGHBOR, CONSENSE, SEQBOOT)
Basics of Designing a Microarry	Designing microarrays by specific probe design, surface preparation,	Robotic printers, fluorescent dyes (Cy3,

7. Probable Questions

1. Explain the types of biological database search tools and describe their applications with relevant examples.
2. Discuss the principles, advantages, and limitations of BLAST and FASTA tools in sequence similarity searching.
3. Describe the different types of sequence alignment methods and compare the algorithms used for pairwise and multiple sequence alignments.
4. Elaborate on the structure and function of key biological databases such as NCBI, EMBL, and PDB. How are they used in computational biology?
5. Write a detailed account of the workflow of a phylogenetic analysis using PHYLIP. Mention the roles of DNADIST, NEIGHBOR, and CONSENSE.
6. What are the major steps involved in designing a DNA microarray? Discuss the significance of probe design and surface chemistry.
7. Describe the process of image analysis in microarray experiments. What software tools are used and how is data quality controlled?
8. What is normalization in microarray data analysis? Compare the different normalization techniques and justify their importance.
9. Define and explain the process of annotation in microarray experiments. Why is regular updating of annotations necessary?
10. Compare and contrast the Neighbor-Joining and UPGMA methods in the context of phylogenetic tree construction. Discuss their assumptions and applications.

8. Suggested Readings

1. Mount, D. W. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press.
2. NCBI BLAST tutorial: https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html
3. Pearson, W. R. (2013). *An introduction to sequence similarity ("homology") searching*. Current Protocols in Bioinformatics, 3(3.1): 3.1.1–3.1.8. https://doi.org/10.1002/0471250953.bi0301s42
4. NCBI Homepage: https://www.ncbi.nlm.nih.gov/
5. EMBL-EBI Homepage: https://www.ebi.ac.uk/
6. RCSB PDB Homepage: https://www.rcsb.org/

7. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235
8. Altschul, S.F., et al. (1990). *Basic local alignment search tool*. J Mol Biol. 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
9. Pearson, W.R., & Lipman, D.J. (1988). *Improved tools for biological sequence comparison*. Proc Natl Acad Sci USA. 85(8):2444–2448. https://doi.org/10.1073/pnas.85.8.2444
10. NCBI BLAST Tool: https://blast.ncbi.nlm.nih.gov/Blast.cgi
11. FASTA Overview: https://www.ebi.ac.uk/Tools/sss/fasta/
12. Felsenstein, J. (1989). *PHYLIP — Phylogeny Inference Package*. Cladistics. 5(2):164–166. https://evolution.gs.washington.edu/phylip.html
13. Saitou, N., & Nei, M. (1987). *The neighbor-joining method*. Mol Biol Evol. 4(4):406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454
14. Kimura, M. (1980). *A simple method for estimating evolutionary rates of base substitutions*. J Mol Evol. 16(2):111–120. https://doi.org/10.1007/BF01731581
15. Schena, M., Shalon, D., Davis, R.W., & Brown, P.O. (1995). *Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270(5235), 467–470. https://doi.org/10.1126/science.270.5235.467*
16. Quackenbush, J. (2002). *Microarray data normalization and transformation. Nature Genetics, 32(Suppl), 496–501. https://doi.org/10.1038/ng1032*
17. Smyth, G. K., & Speed, T. (2003). *Normalization of cDNA microarray data. Methods, 31(4), 265–273. https://doi.org/10.1016/S1046-2023(03)00155-5*
18. Huang da, W., Sherman, B.T., & Lempicki, R.A. (2009). *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211*

References:

- NCBI BLAST tool: https://blast.ncbi.nlm.nih.gov/Blast.cgi
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). *Basic local alignment search tool*. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

DISCLAIMER : This Self Learning Material (SLM) has been compiled from various authentic books, Journals articles, e-journals and other web sources.