MSCZO-507

M.Sc. II Semester

MOLECULAR BIOLOGY & BIOTECHNOLOGY

DEPARTMENT OF ZOOLOGY
SCHOOL OF SCIENCES
UTTARAKHAND OPEN UNIVERSITY

UNIT EDITOR AND WRITERS

EDITOR

Dr. Manisha Nigam

Professor & Head

Department of Biochemistry

School of Life Sciences

H.N.B.G. (Central) University, Srinagar,

Garhwal, Uttarakhand

UNIT WRITER

Ms. Poornima Nailwal (Unit No. 1 & 5)

Assistant Professor

Department of Zoology,

Uttarakhand Open University

Haldwani, Nainital.

Dr. Jaya Upreti (Unit No.4 & 7)

Assistant Professor

Department of Zoology, School of Sciences,

Uttarakhand Open University

Haldwani, Nainital

Dr. Mukta Joshi (Unit No:2 & 3)

Assistant Professor

Department of Zoology

Uttarakhand Open University

Nainital, Uttarakhand

Dr. Shyam S. Kunjwal (Unit No.6)

Assistant Professor

Department of Zoology,

Uttarakhand Open University

Haldwani, Nainital.

COURSE 1: MOLECULAR BIOLOGY & BIOTECHNOLOGY

Course code: MSCZO- 507   **Credit: 3**

UNIT 1: DNA REPLICATION

CONTENTS

1.1 Objectives

1.2 Introduction

1.3 Prokaryotic and eukaryotic DNA replication

1.4 Mechanics of DNA replication

1.5 DNA damage and repair mechanism

1.6 Enzymes and accessory protein involved in DNA replication

1.7 Summary

1.8 Terminal questions and answers

1.9 Glossary

1.10 Reference

1.1 OBJECTIVES

In this unit you will learn

- ❖ What is DNA replication?

- ❖ How does DNA replication take place in Prokaryotic and Eukaryotic organisms?

- ❖ What is the mechanism of DNA replication?

- ❖ DNA damage and DNA repair mechanism in detail.

- ❖ What are the various enzymes and accessory proteins involved in DNA replication?

- ❖ What is the role of enzymes and proteins in DNA replication?

1.2 INTRODUCTION

DNA replication is the process of synthesis of a new strand of DNA from the parent strand. During the process of DNA replication, the two strands of parent DNA would separate and act as a template for the synthesis of new complementary strands. The model of DNA replication was proposed by Watson and Crick. After completion of DNA synthesis, it was observed that each DNA molecule consists of one parental strand and one newly synthesised strand. For double stranded DNA, each strand of the original DNA molecule serves as template for the production of the complementary strand. This process is called **semi conservative replication** because the parental double helix is half conserved. It is accomplished by DNA dependent DNA synthesis by the enzyme DNA Polymerase. In this process, many other enzymes, such as helicase, primases, topoisomerase, ligases are also utilized.

There are three major types of DNA polymerase present in the nucleus for nuclear DNA replication; these are DNA Pol α, DNA Pol δ, and DNA Pol ε. DNA Pol α has both DNA polymerase and primase activity. DNA Pol α initiates synthesis of DNA in both leading and lagging strands. DNA Pol ε is the most accurate DNA polymerase in the eukaryotic system, having proofreading activity. DNA Pol γ is the only polymerase participating in mitochondrial DNA replication.

1.3 PROKARYOTIC AND EUKARYOTIC DNA REPLICATION

DNA replication employs a large number of proteins and enzymes, each of which plays a critical role during the process. One of the key players is the enzyme DNA polymerase, which adds nucleotides one by one to the growing DNA chain that are complementary to the template strand. The addition of nucleotides requires energy; this energy is obtained from the nucleotides that have three phosphates attached to them, similar to ATP which has three phosphate groups attached. When the bond between the phosphates is broken, the energy released is used to form the phosphodiester bond between the incoming nucleotide and the growing chain. In prokaryotes, three main types of polymerases are known: DNA pol I, DNA pol II, and DNA pol III. DNA pol I is known as Kornberg enzyme. DNA pol III is the enzyme required for chromosomal DNA synthesis. There are five functional types of DNA Polymerase in Prokaryotes. Their functions are compared in table 1.

Table 1. Functional characteristics of DNA Polymerase in E. coli

DNA Polymerase	5' to 3' polymerization	3' to 5' exonuclease	5' to 3' exonuclease	Function
Type I	Yes	Yes	Yes	Removes and replace primers
Type II	Yes	Yes	No	DNA repair; restarts replication after damaged DNA halts synthesis
Type III	Yes	Yes	No	Responsible for most of the chromosomal DNA replication
Type IV	Yes	Yes	No	DNA repair
Type V	Yes	No	No	DNA repair; translesion DNA synthesis

DNA polymerase III holoenzyme is the primary enzyme complex involved in prokaryotic DNA replication. It was discovered by Thomas Kornberg (son of Arthur Kornberg) and Malcolm Gefter in 1970. The complex has high processivity (i.e. the number of nucleotides added per

binding event). It is the main enzyme involved in replication activity. The DNA Pol III holoenzyme also has 5'→3' polymerase activity as well as proofreading capabilities that corrects replication mistakes by means of exonuclease activity reading 3'→5'. DNA Pol III is a component of the replisome, which is located at the replication fork.

The pol III holoenzyme synthesizes DNA at the rate of about 730 nucleotides per second (nt/sec) in vitro, just a little slower than the rate of almost 1000 nt/sec observed in vivo. This enzyme is also highly processive, both in vitro and in vivo.

The Pol III core (**αθε**) does not function processivity by itself, so it can replicate only a short stretch of DNA before falling off the template. By contrast, the core plus the **β**-subunit can replicate DNA processivity at a rate approaching 1000 nt/sec. The **β**-subunit forms a dimer that is ring-shaped. This ring fits around a DNA template and interacts with the **α**-subunit of the core to tether the whole polymerase and template together. This is why the holoenzyme stays on its template so long and is therefore so processive. Proliferating cell nuclear antigen (PCNA) is a DNA clamp that acts as a processivity factor for DNA polymerase δ in eukaryotic cells and is essential for replication. It forms a trimer with a similar ring shape that can encircle DNA and hold DNA polymerase on the template.
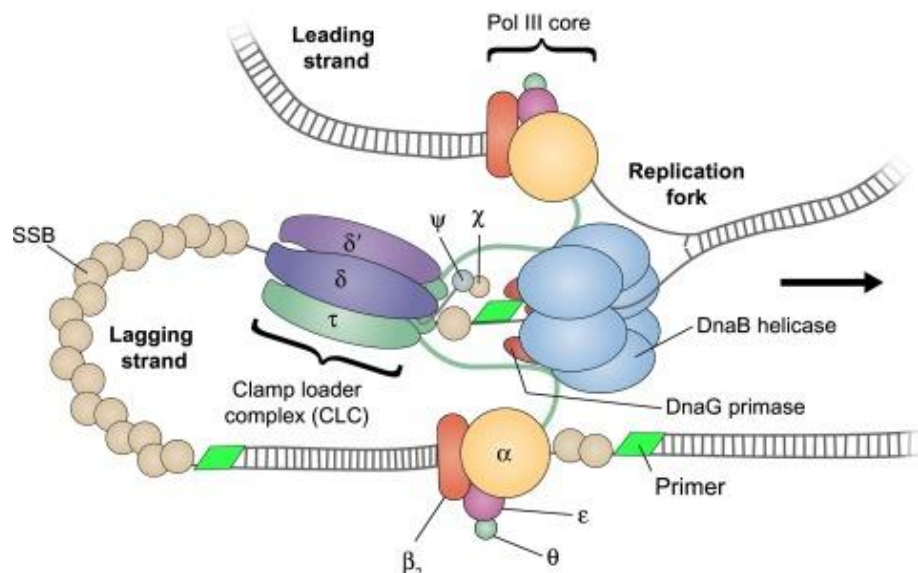
Figure 1.1: E. coli DNA Polymerase holoenzyme (Source: lifescience.org)

The $\boldsymbol{\beta}$-subunit needs help from the $\boldsymbol{\gamma}$ complex ($\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, $\boldsymbol{\delta}$', $\boldsymbol{\chi}$, and Ψ) to load onto the DNA Template. The $\boldsymbol{\gamma}$ complex acts catalytically in forming this processive complex, so it does not remain associated with the complex during processive replication. Clamp loading is an ATP-dependent process. The energy from ATP changes the conformation of the clamp loader such that the $\boldsymbol{\delta}$-subunit can bind to one of the $\boldsymbol{\beta}$-subunits of the clamp. This binding opens the clamp and allows it to encircle DNA.

Table 2: Subunits of DNA Polymerase III holoenzymes and their functions

Subunits	Functions
α	Synthesizes DNA
ε	3' to 5' proofreading
θ	Accessory protein that stimulates the proofreading function
β	Clamp protein, which allows polymerase to slide along the DNA without falling off
τ, γ, δ, δ', χ, Ψ	Clamp loader complex, involved with helping the clamp protein bind to the DNA
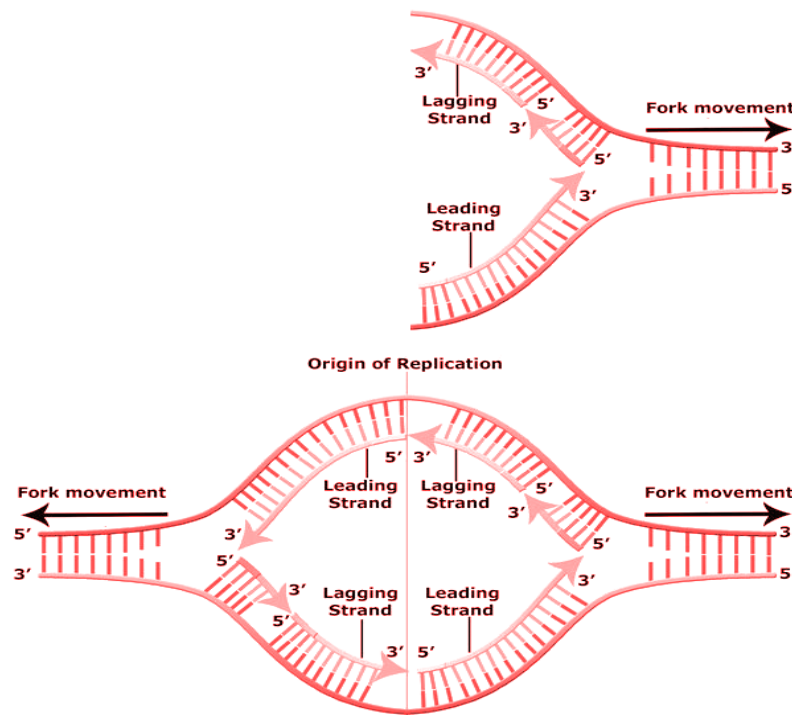
Figure 1.2: Origin of Replication (Source: Sciencedirect.com)

Origins of Replication:

In prokaryotic cells, the circular chromosome contains a unique origin of replication. DNA replication proceeds bidirectionally from the origin to the terminus. Therefore, the whole bacterial genome is a single replicon (monorepliconic). In *E. coli* single origin of the replication present in the chromosome is referred to as oriC. The genome of *E. coli* is replicated bidirectionally from a single origin, oriC. *E. coli* replicon is circular with no free ends. Replication of DNA in *E. coli* is also known as **theta replication** and it occurs in three steps: Initiation, elongation and termination. The origin of replication is recognized by certain proteins that bind to this site. An enzyme called helicase unwinds the DNA by breaking the hydrogen bonds between the nitrogenous base pairs. ATP hydrolysis is required for this process.

Replication fork:

As the DNA opens up, Y-shaped structures called replication forks are formed. Two replication forks at the origin of replication are extended bi-directionally as replication proceeds.
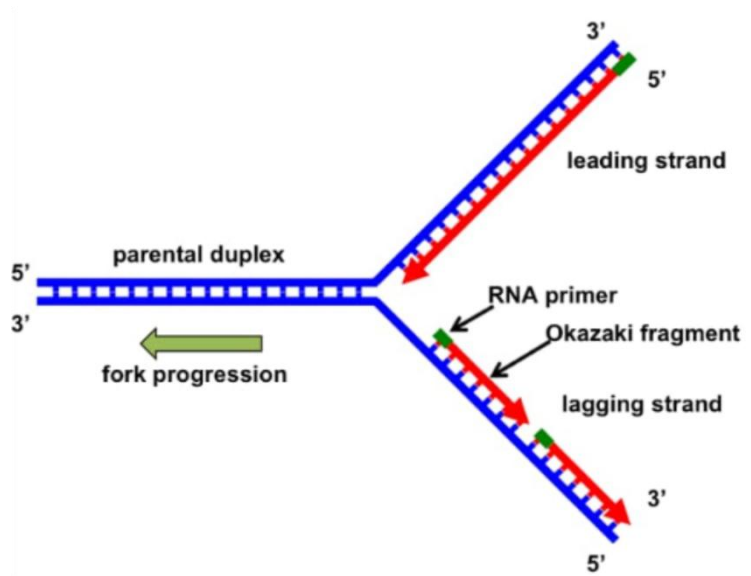
Figure 1.3: Schematic diagram of Replication fork (Source: vedantu.com)

The Leading and Lagging Strands:

DNA polymerase can only synthesize new strands in the 5′ to 3′ direction. Therefore, the two newly-synthesized strands grow in opposite directions because the template strands at each replication fork are antiparallel. The "leading strand" is synthesized continuously toward the replication fork as helicase unwinds the template double-stranded DNA.

The "lagging strand" is synthesized in the direction away from the replication fork and away from the DNA helicase unwinds. This lagging strand is synthesized in pieces because the DNA polymerase can only synthesize in the 5′ to 3′ direction, and so it constantly encounters the previously-synthesized new strand. The pieces are called Okazaki fragments, and each fragment begins with its own RNA primer.

The leading strand can be extended by one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments. The overall direction of the lagging strand will be 3′ to 5′, while that of the leading strand will be 5′ to 3′.
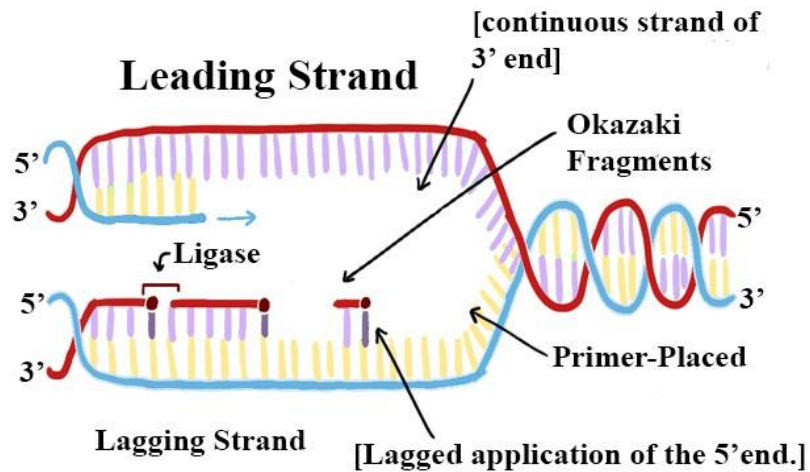
Figure 1.4: Leading and lagging strand synthesis in DNA Replication (Sources: onlinesciences.com)

The sliding clamp (a ring-shaped protein that binds to the DNA) holds the DNA polymerase in place as it continues to add nucleotides. Topoisomerase prevents the over-winding of the DNA double helix ahead of the replication fork as the DNA is opening up; it does so by causing temporary nicks in the DNA helix and then resealing it. As synthesis proceeds, the RNA primers are replaced by DNA. The primers are removed by the exonuclease activity of DNA pol I, while the gaps are filled in by deoxyribonucleotides. The nicks that remain between the newly-synthesized DNA (that replaced the RNA primer) and the previously-synthesized DNA are sealed by the enzyme DNA ligase that catalyzes the formation of phosphodiester linkage between the 3′-OH end of one nucleotide and the 5′ phosphate end of the other fragment.

In eukaryotic cells, 13 different types of DNA polymerase have been identified. Their comparison is given in the following Table 3.

Table 3: Functional characteristics of DNA Polymerase in Eukaryotes

DNA Polymerase	5' to 3' Polymerase activity	3' to 5' exonuclease activity	Cellular function
α (alpha)	Yes	No	Initiation of nuclear DNA synthesis and DNA repair; has primase activity
β(beta)	Yes	No	DNA repair and recombination of nuclear DNA
γ (gamma)	Yes	Yes	Replication of repair of mitochondrial DNA
δ (delta)	Yes	Yes	Leading and lagging strand synthesis of nuclear DNA, DNA repair, and translesion DNA synthesis
ε(epsilon)	Yes	Yes	Unknown; probably repair and replication of nuclear DNA
ζ (zeta)	Yes	No	Translesion DNA synthesis
η (eta)	Yes	No	Translesion DNA synthesis
θ(theta)	Yes	No	DNA repair
ι(iota)	Yes	No	Translesion DNA synthesis
κ(kappa)	Yes	No	Translesion DNA synthesis
λ (lambda)	Yes	No	DNA repair
μ (mu)	Yes	No	DNA repair
σ (sigma)	Yes	No	Nuclear DNA replication (possibly), DNA repair, and sister-chromatid cohesion

Eukaryotes usually have multiple linear chromosomes, each with multiple origins of replication (multirepliconic). Eukaryotic cells contain numerous types of DNA polymerases, which are present in nucleus and organelles such as mitochondria and chloroplasts.

1.4 MECHANICS OF DNA REPLICATION

Bacterial origin of replication and initiation of replication:

The origin of replication is a particular sequence in a genome at which replication is initiated. The specific structure of the origin of replication varies somewhat from species to species, but all share some common characteristics such as high AT content. The origin of replication binds the pre-replication complex, a protein complex that recognizes, unwinds, and begins to copy DNA.

In *E. coli,* the origin of replication- *oriC*- consists of three A-T rich 13-mer repeats and four 9-mer repeats. Ten to 20 monomers of the replication initiator protein Dna A bind to the 9 mer repeats, and the DNA coils around this protein complex forming a protein core. This coiling stimulates the AT rich region in the 13 mer sequence to unwind, allowing the helicase loader Dna C to load the replicative helicase Dna B to each of the two unwound DNA strands. The helicase Dna B forms the basis of the replisome, a complex of enzymes that performs DNA replication.

Further melting of the two DNA strands is mediated by the enzyme helicase. One molecule of Helicase clamps around each single strand in the open complex. The attachment of the helicase yields the prepriming complex. The helicase then separates the two strands and opens up a replication bubble (or Replication eye). Helicases move along and open the DNA duplex utilizing the energy of ATP hydrolysis.

Another protein essential for the unwinding process is the enzyme DNA gyrase reduces torsional strain (torque) that builds up ahead of the replication fork as a result of unwinding.

In *E. coli,* the separated strands are inhibited from subsequently reannealing by single stranded binding proteins (SSB protein), which binds to both separated strands.
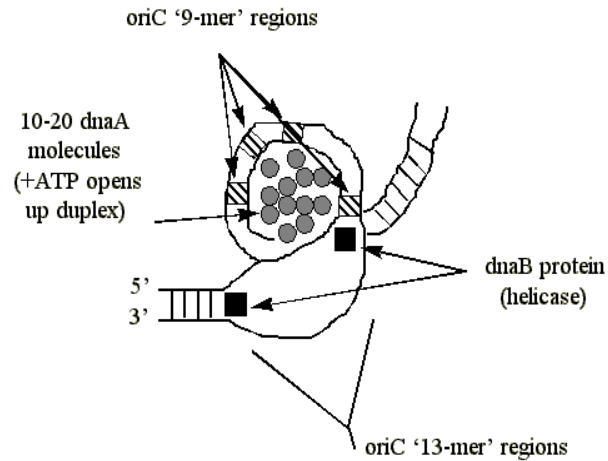
Figure 1.5: Origin of Replication (oriC) in E. coli

Priming: DNA polymerase cannot initiate the synthesis of a polynucleotide; they can only add nucleotides to the 3' end of an already existing chain that is base-paired with the template strand. The Initial nucleotides chain is a short one called a primer. This primer is always a small strand of RNA, synthesized by an enzyme called Primase. The primases synthesize short primer NAs complementary to both strands of duplex DNA, and then they dissociate from the single-stranded template. A complex between primase, helicase, and other accessory proteins is called a primosome.

The first dNTP addition: After Primer formation, DNA Pol III then adds a DNA nucleotide to the 3'end of the RNA primer. Once the first deoxyribonucleotide is added to the Primer, Initiation is said to have been achieved.

Elongation phase of Bacterial DNA replication:
Both strands of the DNA template are copied as the replication bubble enlarges. Each end of the bubble represents a growing fork where both new strands are synthesized. At each growing fork, one strand, called the leading strand, is synthesized continuously from a single primer on the leading-strand template and grows in the 5'-3' direction. Growth of the leading strand proceeds in the same direction as movement of the growing fork.
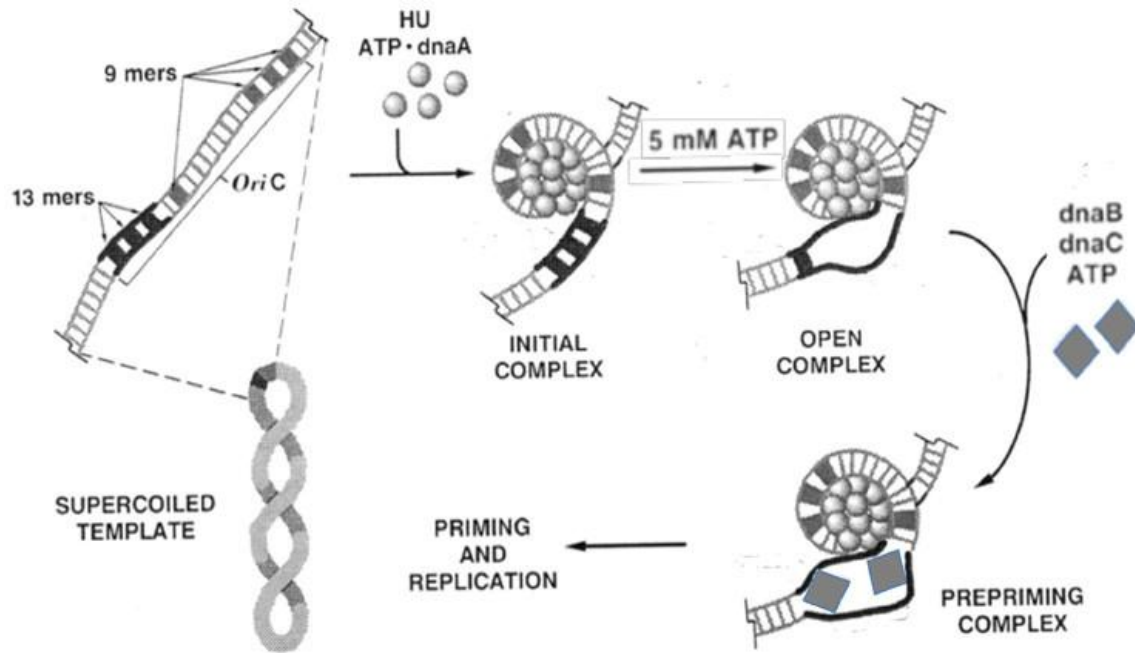
Figure 1.6: Initiation of Replication in E. coli (Source: researchgate.net)

Synthesis of the lagging strand is discontinuous, because DNA polymerase can add nucleotides in the 5' to 3' direction. Movement of the growing fork is however in the opposite direction. In both prokaryotes and eukaryotes this mismatch of direction leads to the discontinuous replication of the lagging strand from multiple primers. As a result on the lagging strand, DNA synthesized in fragments (Okazaki fragments) in a discontinuous fashion. The synthesis of each Okazaki fragment is initiated in an independent manner by the help of a Primase.

DNA Pol I is required repeatedly to remove primers. The same DNA Pol I also carries out the DNA synthesis in the gaps created by primer removal. DNA Ligase finally creates the phosphodiester linkage to join two adjacent Okazaki Fragments.

Elongation in bacteria proceeds at the rate of about 1000 nt/sec (while only at 50-100 nt/sec in eukaryotes, due to histone complexing). Replication proceeds bidirectionally from each replication bubble is called a Replication Fork.

DNA Polymerase action: DNA polymerase are template-directed enzymes that catalyzes the formation of phosphodiester bonds in 5' to 3' direction by the 3' hydroxyl group nucleophilic attack on the innermost phosphorus atom of a dioxyribo-nucleoside 5' triphosphate.

In this exergonic reaction a pyrophosphate group (PPi) is released. PPi further breaks into Pi-Pi. This second reaction provides further free energy to the process. This is why, in DNA replication only deoxy-ribonucleoside triphosphates (dNTP) are used and not dNMPs or dNDPs.

Termination of Bacterial DNA replication:

In some DNA molecules, replication is terminated whenever two replication forks meet. In others, specific termination sequences such as Ter sequences block further replication. A termination protein, called Tus in E. coli binds to these sequences. Tus blocks the movements of helicase, thus stalling the replication fork and preventing further DNA replication.

Often in bacteria due to circular DNA molecules, the daughter molecules become interlocked. This interlocking is resolved by a Type II Topoisomerase that causes a double stranded cut in one of the daughter DNA molecules.
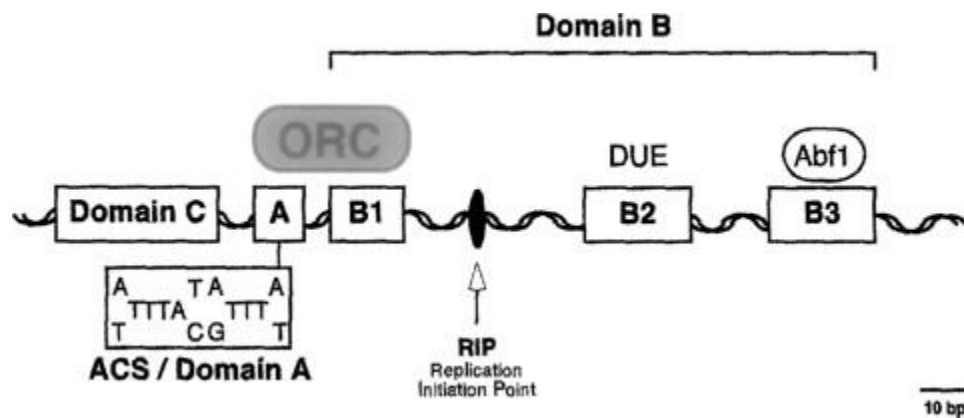
Figure 1.7: Autonomously replicating sequence (oriC) in Yeast (Source: Sciencedirect.com)

DNA replication in Yeast:

The yeast origins of replication are contained within autonomously replicating sequences (ARSs) that are composed of four important regions (A, B1, B2, and B3). Region A is 10-15 bp long and contains an 11-bp consensus sequence that is highly conserved in ARSs and rich in AT base pairs. Region B3 may allow for an important DNA bend within ARS1.

DNA replication in Eukaryotes:
Initiation:

Eukaryotic DNA is bound to proteins known as histones to form structures called nucleosomes. During initiation, the DNA is made accessible to the proteins and enzymes involved in the replication process. There are specific chromosomal locations called origins of replication where replication begins. In some eukaryotes, like yeast, these locations are defined by specific sequence of base pairs to which the replication initiation proteins bind. In other eukaryotes, like humans, there does not appear to be a consensus sequence for their origins of replication. Instead, the replication initiation proteins might identify and bind to specific modifications to the nucleosomes in the origin region.

Certain proteins recognize and bind to the origin of replication and then allow the other proteins necessary for DNA replication to bind the same region. The first proteins to bind the DNA are said to "recruit" the other proteins. Two copies of an enzyme called helicase are among the proteins recruited to the origin. Each helicase unwinds and separates the DNA helix into single-stranded DNA. As the DNA opens up, Y-shaped structures called replication forks are formed. Because two helicases bind, two replication forks are formed at the origin of replication; these are extended in both directions as replication proceeds creating a replication bubble. There are multiple origins of replication on the eukaryotic chromosome which allow replication to occur simultaneously in hundreds to thousands of locations along each chromosome.

Elongation

During elongation, an enzyme called DNA polymerase adds DNA nucleotides to the 3′ end of the newly synthesized polynucleotide strand. The template strand specifies which of the four DNA nucleotides (A, T, C, or G) is added at each position along the new chain. Only the nucleotide complementary to the template nucleotide at that position is added to the new strand.

DNA polymerase contains a groove that allows it to bind to a single-stranded template DNA and travel one nucleotide at a time. For example, when DNA polymerase meets an adenosine nucleotide on the template strand, it adds a thymidine to the 3′ end of the newly synthesized strand, and then moves to the next nucleotide on the template strand. This process will continue until the DNA polymerase reaches the end of the template strand.
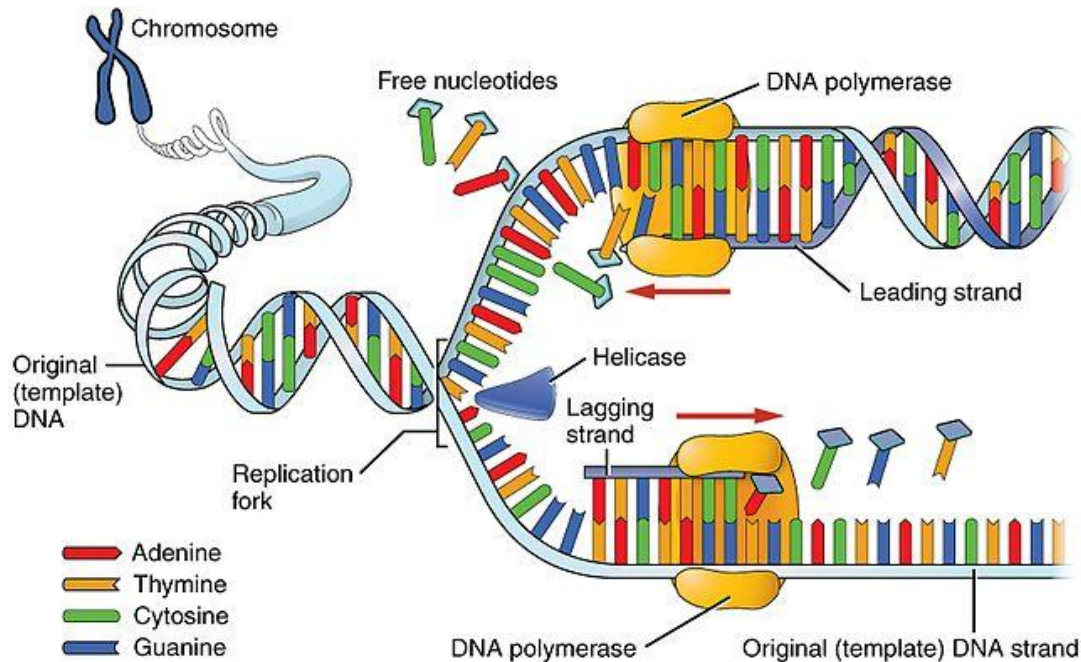
Figure 1.8: DNA replication in Eukaryotes (Source: openoregon.pressbook.in)

DNA polymerase cannot initiate new strand synthesis; it only adds new nucleotides at the 3′ end of an existing strand. All newly synthesized polynucleotide strands must be initiated by a specialized RNA polymerase called primase. Primase initiates polynucleotide synthesis by creating a short RNA polynucleotide strand complementary to the template DNA strand. This short stretch of RNA nucleotides is called the primer. Once RNA primer has been synthesized at the template DNA, primase exits, and DNA polymerase extends the new strand with nucleotides complementary to the template DNA.

Eventually, the RNA nucleotides in the primer are removed and replaced with DNA nucleotides. Once DNA replication is finished, the daughter molecules are made entirely of continuous DNA nucleotides, with no RNA portions.

Termination:

Eukaryotic chromosomes have multiple origins of replication, which initiate replication almost simultaneously. Each origin of replication forms a bubble of duplicated DNA on either side of the origin of replication. Eventually, the leading strand of one replication bubble reaches the

lagging strand of another bubble, and the lagging strand will reach the 5′ end of the previous Okazaki fragment in the same bubble.

DNA polymerase halts when it reaches a section of DNA template that has already been replicated. However, DNA polymerase cannot catalyse the formation of a phosphodiester bond between the two segments of the new DNA strand, and it drops off. These unattached sections of the sugar-phosphate backbone in an otherwise full-replicated DNA strand are called nicks.

Once all the template nucleotides have been replicated, the replication process is not yet over. RNA primers need to be replaced with DNA, and nicks in the sugar-phosphate backbone need to be connected.

The group of cellular enzymes that remove RNA primers include the proteins FEN1 (flap endonuclease 1) and RNase H. The enzymes FEN1 and RNase H remove RNA primers at the start of each leading strand and at the start of each Okazaki fragment, leaving gaps of unreplicated template DNA. Once the primers are removed, a free-floating DNA polymerase lands at the 3′ end of the preceding DNA fragment and extends the DNA over the gap. However, this creates new nicks (unconnected sugar-phosphate backbone).

In the final stage of DNA replication, the enzyme ligase joins the sugar-phosphate backbones at each nick site. After the ligase has connected all nicks, the new strand is one long continuous DNA strand, and the daughter DNA molecule is complete.

Telomere Replication

As DNA polymerase alone cannot replicate the ends of chromosomes, telomerase aids in their replication and prevents chromosome degradation.

The End Problem of Linear DNA Replication

Linear chromosomes have an end problem. After DNA replication, each newly synthesized DNA strand is shorter at its 5′ end than at the parental DNA strand's 5′ end. This produces a 3′ overhang at one end (and one end only) of each daughter DNA strand, such that the two daughter DNAs have their 3′ overhangs at opposite ends.

Every RNA primer synthesized during replication can be removed and replaced with DNA strands except the RNA primer at the 5′ end of the newly synthesized strand. This small section

of RNA can only be removed, not replaced with DNA. Enzymes RNase H and FEN1 remove RNA primers, but DNA Polymerase will add new DNA only if the DNA Polymerase has an existing strand 5′ to it ("behind" it) to extend. However, there is no more DNA in the 5′ direction after the final RNA primer, so DNA polymerase cannot replace the RNA with DNA. Therefore, both daughter DNA strands have an incomplete 5′ strand with 3′ overhang.

In the absence of additional cellular processes, nucleases would digest these single-stranded 3′ overhangs. Each daughter's DNA would become shorter than the parental DNA, and eventually entire DNA would be lost. To prevent this shortening, the ends of linear eukaryotic chromosomes have special structures called telomeres.

Telomere Replication

The ends of the linear chromosomes are known as telomeres: repetitive sequences that code for no particular gene. These telomeres protect the important genes from being deleted as cells divide and as DNA strands shorten during replication.

In humans, a six base pair sequence, TTAGGG, is repeated 100 to 1000 times. After each round of DNA replication, some telomeric sequences are lost at the 5′ end of the newly synthesized strand on each daughter DNA, but because these are noncoding sequences, their loss does not adversely affect the cell. However, even these sequences are not unlimited. After sufficient rounds of replication, all the telomeric repeats are lost, and the DNA risks losing coding sequences with subsequent rounds.

The discovery of the enzyme telomerase helped in the understanding of how chromosome ends are maintained. The telomerase enzyme attaches to the end of a chromosome and contains a catalytic part and a built-in RNA template. Telomerase adds complementary RNA bases to the 3′ end of the DNA strand. Once the 3′ end of the lagging strand template is sufficiently elongated, DNA polymerase adds the complementary nucleotides to the ends of the chromosomes; thus, the ends of the chromosomes are replicated.
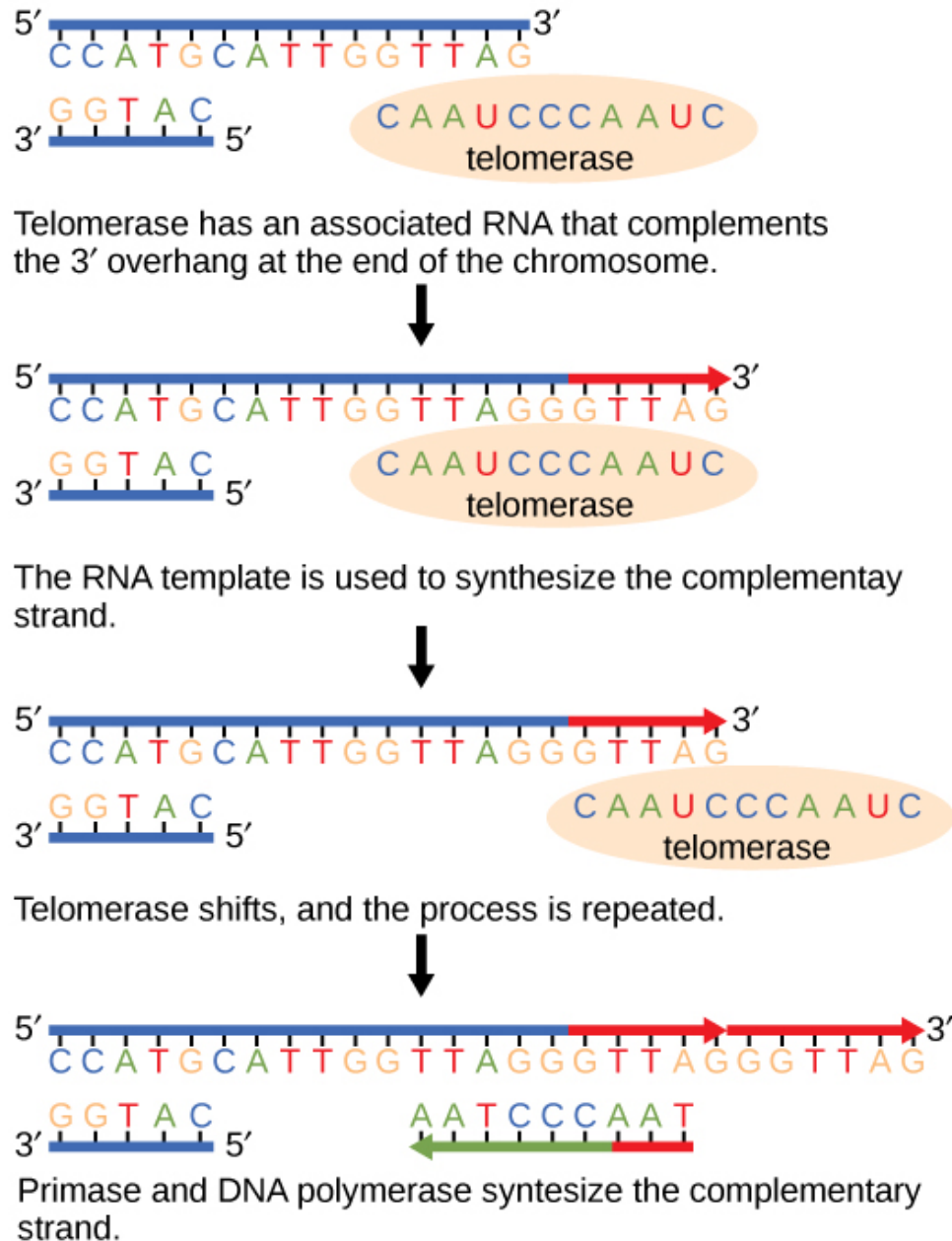
Figure 1.9: Telomere Replication (Source: bio.libretexts.org)

1.5 DNA DAMAGE AND REPAIR MECHANISM

DNA damage can be caused by physical and chemical mutagens. Physical mutagens are primarily radiation such as UV radiation from the sun. UV radiation forms covalent bonds between adjacent pyrimidine bases in the DNA strands. Ionizing radiation (X rays) initiates DNA mutations by generating free radicals within the cell which creates reactive oxygen species (ROS) and results in single or double strand breaks in the DNA helix.

DNA repair mechanism involves a set of processes which are involved in the identification of damaged nucleotides, synthesis of correct nucleotides and ligating the correct nucleotides.

There are three types of repair mechanisms:

(1) Direct reversal of the damage: Direct reversal repair is specific to the damage. For example, in a process called photoreactivation, pyrimidine bases fused by UV light is separated by DNA photolyase (a light-driven enzyme). For direct reversal of alkylation events, a DNA methyltransferase or DNA glycosylase detects and removes the alkyl group.

(2) Excision repair: Excision repair can be specific or nonspecific. In base excision repair, DNA glycosylases specifically identify and remove the mismatched base. In nucleotide excision repair, the repair machinery recognizes a wide array of distortions in the double helix caused by mismatched bases; in this form of repair, the entire distorted region is excised

a. Base excision repair (BER) typically acts on subtle base damage. This process begins with a DNA glycosylase, which extrudes a base in a damaged base pair, then clips out the damaged base, leaving an apurinic or apyrimidinic site that attracts the DNA repair enzymes that remove the remaining deoxyribose phosphate and replace it with a normal nucleotide. In bacteria, DNA polymerase I is the enzyme that fills in the missing nucleotide in BER; in eukaryotes, DNA polymerase β plays this role. However, this enzyme makes mistakes, and has no proofreading activity, so APE1 carries out the necessary proofreading. Repair of 8-oxoguanine sites in DNA is a special case of BER that can happen in two ways. Since oxoG mispairs with A, the A can be removed after DNA replication by a specialized adenine DNA glycosylase. However, if replication has not yet occurred, the oxoG will still be paired with C, and the oxoG can be removed by another DNA glycosylase, the oxoG repair enzyme.

b. Nucleotide excision repair typically handles bulky damage that distorts the DNA double helix. NER in E. coli begins when the damaged DNA is clipped by an endonuclease on either side of the lesion, at sites 12–13 Nt apart. This allows the damaged DNA to be removed as part of the resulting 12–13-base oligonucleotide. DNA polymerase I fill the gap and DNA ligase seals the final nick. Eukaryotic NER follows two pathways. In GG-NER, a complex composed of XPC and hHR23B initiates repair by binding to a lesion anywhere in the genome and causing a limited amount of DNA melting. This protein apparently recruits XPA and RPA. TFIIH then

joins the complex, and two of its subunits (XPB and XPD) use their DNA helicase activities to expand the melted region. RPA binds two exonucleases (XPF and XPG) and positions them for cleavage of the DNA strand on either side of the lesion. This releases the damage on a fragment between 24 and 32 Nt long. TC-NER is very similar to GG-NER, except that RNA polymerase plays the role of XPC in damage sensing and initial DNA melting.

(3) Post-replication repair: Postreplication repair occurs downstream of the lesion, because replication is blocked at the actual site of damage. In order for replication to occur, short segments of DNA called Okazaki fragments are synthesized. The gap left at the damaged site is filled in through recombination repair, which uses the sequence from an undamaged sister chromosome to repair the damaged one, or through error-prone repair, which uses the damaged strand as a sequence template. Error-prone repair tends to be inaccurate and subject to mutation.

DNA can be damaged in many different ways, and this damage, if left unrepaired, can lead to mutations: changes in the base sequence of a DNA. This distinction is worth emphasizing at the outset: DNA damage is not the same as mutation, although it can lead to mutation. DNA damage is simply a chemical alteration to DNA. A mutation is a change in a base pair. For example, the change from a G–C pair to an ethyl-G–C pair is DNA damage; the change from a G–C pair to any other natural base pair (A–T or T–A or C–G) is a mutation. If a particular kind of DNA damage is likely to lead to a mutation, we call it genotoxic. Indeed, we will see in the next section that the ethyl-G in our example is genotoxic because it is likely to mispair with T instead of C during DNA replication. If this happens, then another round of replication will place an A across from the mispaired T, and conversion of the normal G–C pair to an A–T pair (a true mutation) will be complete. Notice that this example illustrates the importance of DNA replication in conversion of DNA damage to mutation.

Different kinds of radiation cause different kinds of damage. Ultraviolet rays have comparatively low energy, and they cause a moderate type of damage i.e. pyrimidine dimers. Gamma and x-rays are much more energetic. They can interact directly with the DNA molecule and form highly reactive free radicals that can attack DNA, altering bases or breaking strands. However, they cause most of their damage by ionizing the molecules, especially water, surrounding the

DNA. This forms free radicals, chemical substances with an unpaired electron. These free radicals, especially those containing oxygen (e.g., OH⁻), are extremely reactive, and they immediately attack neighbouring molecules. When such a free radical attacks a DNA molecule, it can change a base, or it can cause a single- or double-stranded break.
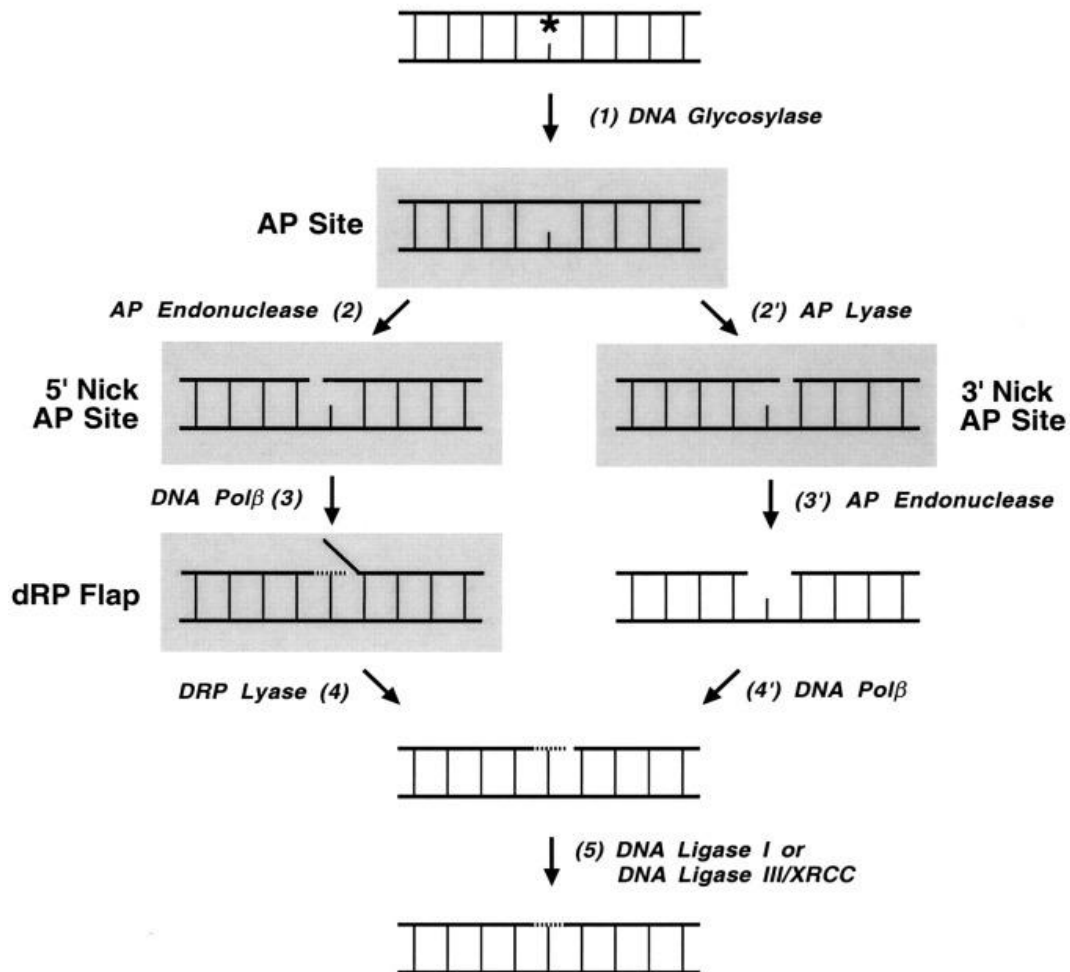

Figure 1.10: Base excision repair mechanism (Source: jbc.org)

1.6 ENZYMES AND ACCESSORY PROTEINS INVOLVED IN DNA REPLICATION

DNA replication is a complex process involving numerous enzymes and proteins. Some of the important enzymes and protein and their role is given below-

DNA helicase: Enzyme which unwinds the DNA helix at the replication fork to provide a single stranded   template.

Single-stranded binding (SSB) proteins: The separated strands are inhibited from subsequent reannealing by SSB protein, which binds to both the separated DNA strands.

Primase: It synthesizes short RNA primers (<15 nucleotides) to prime DNA chain elongation.

DNA polymerase I: RNA primers are removed and replaced with DNA.

DNA polymerase III: It catalyses the synthesis of DNA by adding nucleotides at 3' end. Elongates DNA strand by adding deoxyribonucleotides to the 3′end of the chain. Synthesis can only occur in the 5′to 3′direction because of DNAP III.

DNA ligase: It catalyses the joining of the ends of two DNA chains by making phosphodiester bonds. Forms a 3′-5′phosphodiester bond between adjacent fragments of DNA

Topoisomerase: It works ahead of the replication fork to prevent supercoiling of the DNA.

DNA topoisomerase II (Gyrase): Relieves the strain on the DNA helix during replication by forming supercoils in the helix through the creation of nicks in both strands of DNA. Removes positive supercoiling and introduces negative supercoiling into DNA by utilising ATP.

DNA topoisomerase I: Relaxes the DNA helix during replication through creation of a nick in one of the DNA strands.

DNA clamp: A protein which prevents DNA polymerase III from dissociation from the DNA parent strand.

1.7 SUMMARY

DNA replication is a semi conservative process in which a double stranded DNA helix is replicated; the daughter DNA molecules consist of one strand from the original helix and one newly synthesized. DNA synthesis involves initiation, elongation, and termination. Initiation in E. coli begins with primer synthesis. Primer synthesis in E. coli requires a primosome composed of the DNA helicase, DnaB, and the primase, DnaG. Primosome assemble at the origin of replication, oriC. DnaB binds to the open complex and facilitates binding of the primase to complete the primosome. The primosome remains with the replisome, repeatedly priming

Okazaki fragment synthesis, at least on the lagging strand. DnaB also has a helicase activity that unwinds the DNA as the replisome progresses. Single-strand binding proteins coat the strands of DNA near the replication fork to prevent the single-stranded DNA from winding back into a double helix. DNA polymerase is able to add nucleotides only in the 5′ to 3′ direction. It also requires a free 3′-OH group to which it can add nucleotides by forming a phosphodiester bond between the 3′-OH end and the 5′ phosphate of the next nucleotide. This means that it cannot add nucleotides if a free 3′-OH group is not available. Another enzyme, RNA primase, synthesizes an RNA primer that is about five to ten nucleotides long and complementary to the DNA, priming DNA synthesis. A primer provides the free 3′-OH end to start replication. DNA polymerase then extends this RNA primer, adding nucleotides one by one that are complementary to the template strand. Eukaryotes usually have multiple linear chromosomes, each with multiple origins of replication (multirepliconic). Eukaryotic cells contain numerous types of DNA polymerases, which are present in nucleus and organelles such as mitochondria and chloroplasts. Eukaryotic chromosomes have special structures known as telomeres at their ends. One strand of these telomeres is composed of many tandem repeats of short, G-rich regions whose sequence varies from one species to another. The G-rich telomere strand is made by an enzyme called telomerase, which contains a short RNA that serves as the template for telomere synthesis. The C-rich telomere strand is synthesized by ordinary RNA-primed DNA synthesis, like the lagging strand in conventional DNA replication. This mechanism ensures that chromosome ends can be rebuilt and therefore do not suffer shortening with each round of replication. DNA damage can be caused by physical and chemical mutagens. Physical mutagens are primarily radiation such as UV radiation from the sun. DNA repair mechanism involves a set of processes which are involved in the identification of damaged nucleotides, synthesis of correct nucleotides and ligating the correct nucleotides.

1.8 TERMINAL QUESTION AND ANSWER

Long answer questions:

1. Describe the process of replication in Eukaryotes.

2. Write an account of the enzymology of DNA replication.

3. Write short notes on:

a) Okazaki fragments

b) Events in replication fork.

c) Rolling circle mode of replication

d) Proteins and enzymes involved in initiation of replication in prokaryotes

Multiple choice questions:

1. During replication, Okazaki fragments elongate

(a) leading strand towards the replication fork

(b) lagging strand towards the replication fork

(c) leading strand away from the replication fork

(d) lagging strand away from the replication fork

2. Which of the following enzymes separates the two strands of DNA during replication?

(a) Gyrase

(b) Topoisomerase

(c) Helicase

(d) DNA polymerase

3. DNA replication is

(a) conservative

(b) conservative and discontinuous

(c) semi-conservative and discontinuous

(d) semi-conservative and semi-discontinuous

4. Which of the following is used in DNA replication studies?

(a) *Neurospora crassa*

(b) *Drosophila melanogaster*

(c) *Escherichia coli*

(d) *Pneumococcus*

5. Which of the following helps in opening of the DNA double helix in front of the replication fork?

(a) Topoisomerase

(b) DNA polymerase-I

(c) DNA gyrase

(d) DNA ligase

6. Termination of replication is triggered by

(a) DNA polymerase

(b) Helicase

(c) SSB

(d) Tus protein

7. DNA polymerase synthesizes

(a) DNA in 5'-3' direction

(b) DNA in 3'-5' direction

(c) mRNA in 3'-5' direction

(d) mRNA in 5'-3' direction

8. Association of histones H1 with nucleosome shows

(a) the occurrence of transcription

(b) the occurrence of replication

(c) exposed DNA double helix

(d) the condensation of DNA into chromatin fibre

9. The 3' – 5' phosphodiester linkage joins

(a) two DNA strands

(b) two nucleotides

(c) a nitrogenous base with pentose sugar

(d) two nucleosides

10. The fragments of DNA are joined together by which of the following enzymes?

(a) Endonuclease

(b) DNA polymerase

(c) Primase

(d) Ligase

Answers: 1. (d) 2. (c) 3. (d) 4. (c) 5. (c) 6. (d) 7. (a) 8. (d) 9. (b) 10. (d)

1.9 GLOSSARY

DNA helicase: Enzyme that is involved in opening the DNA helix into its single strands for DNA   replication.

DNA ligase: Enzyme that joins the ends of two strands of DNA together with a covalent bond to make   a   continuous   DNA   strand.

DNA methylation: Addition of methyl groups to DNA.

DNA polymerase: Enzyme that synthesizes DNA by joining nucleotides together using a DNA template   as   a   guide.

DNA primase: Enzyme that synthesizes a short strand of RNA on a DNA template, producing a primer   for   DNA   synthesis.

DNA repair: A set of processes for repairing the many accidental lesions that occur continuously   in   DNA.

DNA replication: Process by which a copy of a DNA molecule is made.

DNA supercoiling: A conformation with loops or coils that DNA adopts in response to super helical tension; conversely, creating various loops or coils in the helix can create such tension.

1.10 REFERENCES

Watson J. D. et al., 2008. *Molecular Biology of the Gene*, 6th ed. Cold Spring Harbor Laboratory Press, 2008, 841 pp

Karp. G., 2001. *Cell and Molecular Biology: Concepts and experiments*, 3rd ed. Wiley Johnand Sons, 2001, 856 pp.

Arthur Kornberg, Tania A. Baker. 2005. DNA replication and related cellular process.

UNIT-2 TRANSCRIPTION

CONTENTS

2.1 OBJECTIVES

After reading this unit you will be able to understand

1. The first step in the process of functioning of a gene in prokaryotes and eukaryotes.

2. Understand and have an overview of Prokaryotic and Eukaryotic transcription

3. Will know about the Regulatory elements and mechanism of transcription regulation

4. Details of Transcriptional polymerase, Capping, Elongation & Termination

2.2 INTRODUCTION

The transfer of genetic information DNA to RNA (mRNA) is known as transcription. The three types of non-genetic RNAs were register, which take an active part in protein synthesis, i) mRNA (messenger RNA), tRNA(transfer RNA), and iii) rRNA (ribosomal RNA).The process of transcription copies a segment of DNA into RNA.The segments of DNA transcribed into RNA molecules that can encode proteins are said to produce messenger RNA (mRNA). Other segments of DNA are copied into RNA molecules called non-coding RNAs (ncRNAs). Averaged over multiple cell types in a given tissue, the quantity of mRNA is more than 10 times the quantity of ncRNA (though in particular single cell type's ncRNAs may exceed mRNAs). The general preponderance of mRNA in cells is valid even though less than 2% of the human genome can be transcribed into mRNA, while at least 80% of mammalian genomic DNA can be actively transcribed (in one or more types of cells), with the majority of this 80% considered to be ncRNA.

2.3 PROKARYOTIC AND EUKARYOTIC TRANSCRIPTION

2.3.1 TRANSCRIPTION IN PROKARYOTES

Prokaryotic transcription also known as bacterial transcription is the process in which a segment of bacterial DNA is copied into a newly synthesized strand of messenger RNA (mRNA) which is later translated to produce proteins with the use of the enzyme RNA polymerase and other

transcription   factors.

The process in which a segment of bacterial DNA is copied into a newly synthesized strand of messenger RNA (mRNA) with use of the enzyme RNA polymerase is known as bacterial transcription. The process of transcription occurs in main three steps: initiation, elongation, and termination; and the end result area strand of mRNA that is complementary to a single strand of DNA. Generally, the transcribed region accounts for more than one gene. Actually, many prokaryotic genes occur in operons, which are a series of genes that work together to code for the same protein or gene product and are controlled by a single promoter. Bacterial RNA polymerase is made up of four subunits and when a fifth subunit attaches, called the σ-factor, the polymerase can recognize specific binding sequences in the DNA, called **promoters.** The binding of the σ-factor to the promoter is the first step in initiation. Once the σ-factor releases from the polymerase, elongation proceeds. The polymerase persists the double stranded DNA, unwind it and synthesizing the new mRNA strand until it reaches a termination site. Termination is essential at specific sites for proper gene expression to occur. Gene expression decides how much gene product, such as protein, is made by the gene. Transcription is performed by RNA polymerase but   particularity   it is   controlled   by   sequence-specific DNA   binding proteins called transcription factors. Transcriptions factors work to recognize specific DNA sequences and based on the cell requirements promote or inhibit additional transcription.

Bacterial transcription is different from eukaryotic transcription in various ways. In bacteria, transcription and translation can occur simultaneously in the cytoplasm of the cell, although in eukaryotes transcription occurs in the nucleus and translation occurs in the cytoplasm. There is only one type of bacterial RNA polymerase whereas eukaryotes have 3 types. Bacteria have a σ-factor that detects and binds to promoter sites but eukaryotes do not need a σ-factor. In place of this, eukaryotes have transcription factors that recognize and bind to promoter sites.

Generally, transcription within bacteria is a highly regulated process that is controlled by the integration of many signals at a particular time. Bacteria heavily rely on transcription and translation to generate proteins that help them respond specifically to their environment.

Bacterial transcription and translation both occur simultaneously in the cytoplasm of the cell. The process occurs in main three steps: Initiation, Elongation and termination.

Initiation – In this step, the double helix DNA strand is unwounded and becomes single-stranded near the initiation site. The RNA polymerase enzyme binds to the promoter sequence and starts the transcription of the single- stranded DNA. The sigma factor is responsible for the binding of the RNA polymerase enzyme to the promoter.

Elongation - After the synthesis of more than 10 base pairs of long RNA, the sigma factor is removed from the promoter. The RNA polymerase enzyme then moves in 5'- 3' direction continuously while synthesizing RNA.

Termination –This is the last stage that stops the process of transcription. It has two types of mechanism-

Rho-independent transcription termination - Transcription termination that is independent of termination factors is known as Rho-independent termination, also called intrinsic termination. Essential elements of a Rho-independent terminator consist of a GC-rich dyad repeat that forms a stem-loop (hairpin) structure followed by a T-rich stretch, generating a U-rich tail in the RNA after termination. Rho-independent termination is achieved by formation of the stem-loop structure, which is facilitated by RNA polymerase pausing during transcription of the T-rich tract. The T-rich stretch is highly conserved among Rho-independent terminators in bacteria, while sequences of the stem-loop seem not to be conserved except for their GC-rich characteristic. This consequence in the dissociation of RNA polymerase from the DNA template.

Rho-dependent transcription termination– Rho, a protein having helicase activity binds with the RNA transcript and moves along the RNA polymerase in 5'-3' direction, encouraging the dissociation of hydrogen bonds between the DNA template and RNA transcript. When Rho factor reaches the transcription bubble, it pulls DNA/RNA hybrid apart and releases the transcript from the transcription bubble. When this occurs, it terminates the transcription.
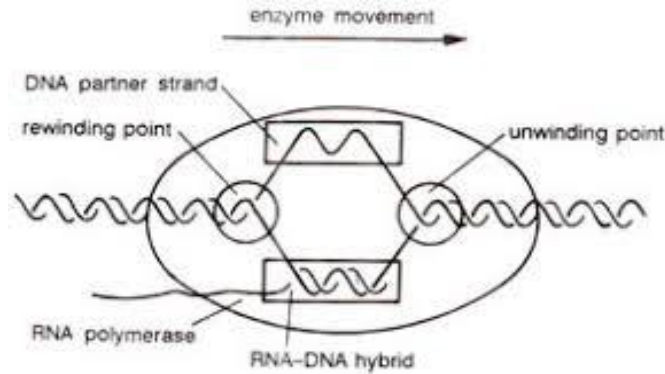
Fig. 2.1 Bacterial RNA polymerase activity

2.3.2 TRANSCRIPTION OF EUKARYOTES

The actual process of transcription in eukaryotes is nearly the same as in the prokaryotes, there is much variability in the factors associated with the process.

The process of transcription, the information in a DNA strand is copied into a new molecule of **RNA**. It is the first step of gene expression, in which a particular portion of DNA is copied into RNA (particularly mRNA) by the enzyme RNA polymerase. It consequences in a complementary, ant parallel RNA strand called a primary transcript.

2.3.2.1 EUKARYOTIC RNA POLYMERASES

In eukaryotes there are at least three different nuclear RNA polymerase.

The three RNA polymerase can be distinguished on the basis of their sensitivity to certain inhibitors of transcription, e.g. α amanitin. Each of these enzymes is a large protein (~ 500,000 daltons), with two large and several (8 to 10) smaller subunits, some of them showing homology to the subunits of *E. coli* RNA polymerase. Eukaryotic RNA polymerase II, also called polymerase II consists of as many as 12 subunits. The largest subunits (~ 200,000 daltons) has homology with β', and the second large subunit ((~ 140,000 daltons) has homology with β subunit of *E. coli* RNA polymerase. One small subunit of RNA polymerase II also has homology with α subunit. The second largest subunit of RNA polymerase II also shares the

catalytic sites of ribonucleases, which may be used for RNA cleavage and may thus help in proof reading.

The different form of the prokaryotes is found to contain three different types of RNA polymerase, called I, II and III RNA-polymerases. RNA polymerase I is believed to be localized exclusively in the nucleolus and involved in the synthesis of ribosomal RNA. Polymerase II and III are found outside the nucleolus and carry out the rest of nuclear transcription. All three enzymes have molecular weight greater than 500,000 and are presumed to be composed of subunits but their molecular structures are not known like the prokaryotic enzymes.

2.3.2.2 EUKARYOTIC PROMOTERS

The three types of eukaryotic RNA polymerases identify different kinds of promoters viz., polymerase I recognizes class I promoter, polymerase II does class II while polymerase III recognize class III promoters. Of this class II promoters are relatively well studied and bear a greater resemblance to those of the prokaryotes.

A) Promoters for RNA polymerase I:

These promoters are accepted by RNA polymerase I and are connected with major rRNA genes. Their series are more unequal compared to those of class II promoters and show sequential difference from species to species. In general, these promoters have two elements

i) A core element situated at the start of transcription between positions -45 and +20, and

ii) An upstream control element (UCE) situated between positions -180 to – 107.

B) Class II promoters:

These promoters can include four different kinds of elements

i) An initiator (I) at the site of transcription initiation.

ii) Downstream elements (D) to the downstream of the transcription initiation site.

iii) A 'TATA box' with a consent of TATAAAA 25 base up stream of the transcription initiation site.

iv) An upstream element farther upstream of TATA box. Many promoters lack at least one of these elements.

There is a major resemblance between the eukaryotic TATA box and prokaryotic -10 box in the sequences. But the major difference is their position (-25vs.-10). The TATA box of yeast may have a more changeable position ranging from -30 to -120. By deleting the TATA box, it was

shown that there is a loss of specificity of initiation relatively than a decrease in the efficiency of transcription initiation. Under natural condition also, such TATA box-less promoters occur in two classes of genes.

i) Housekeeping genes which are constitutively active in all cells of an organism and control basic metabolic pathways essential to sustain cellular life, some of these genes may have GC boxes as substitutes for the TATA boxes.

ii) Homeoticgenes are any of a group of genes that control the pattern of body formation during early embryonic development of organisms. These genes encode proteins called transcription factors that direct cells to form various parts of the body by producing specific types of proteins only in certain types of cells like keratin in skin cells, hemoglobin in RBC.

The upstream elements generally contain GC rich sequences such as GGGCGG and CCGCCC and therefore commonly called GC boxes. These GC boxes are orientations-independent while they are active even after rotation through 108°. But if they are moved added by a few dozen bases away from the TATA box, they lose their ability to effect transcription. Another commonly found sequence in the upstream element is CCAAT(cat box).

C) Class III promoters:

Three classes of promoters have been described for RNA polymerase III; (i) tRNA gene promoters have two elements each 10bp long, spaced 30-120bp apart, and located downstream. These are connected with the genes that encode small RNAs are renowned by RNA polymerase III. Generally, these promoters are internal to genes whose junction they control. Their internal regions are particular into three sequential group-I. Box A (involving bases 50-83 of the transcribed sequence), ii) a comparatively short intermediate element and iii) box C.

Two kinds of class III promoters have been reported:

a) Class III promoters connected with "classical" genes such as 5s rRNA, tRNA and adenovirus VA RNA genes. These promoters have all three elements- box A, box C and intermediate.

b) Class III promoters connected with "non-classical" genes such as Ub snRNA, 7SL RNA, Epstein-Barr virus EBER genes. These promoters have only box A and box C and are similar to the class II promoters as the promoters are not internal to the gene (not including 75-RNA) and contains TATA boxes.

2.3.2.3 ENHANCERS AND SILENCERS

These are cis-acting DNA elements which control the gene activity. They are not a part of the promoter. These are location and direction autonomous elements. The sequences/ element that increase/ stimulate transcription are called enhancers and those which inhibit/ decrease transcription are called silencers. Equally act though proteins which bind to them and these are called transcription factors/ enhancer- binding proteins/ activator proteins.

Commonly the enhancers are upstream of the promoter they control, except some exceptions, e.g mouse immunoglobulin Y_2 b gene.

 Silencer is a sequence-specific element that induces a negative effect on the transcription of its particular gene. There are many positions in which a silencer element can be located in DNA. The most common position is found upstream of the target gene where it can help repress the transcription of the gene.They 'somehow' cause the chromatin to coil up into a condensed unapproachable and consequently inactive form thereby preventing the transcription of neighboring genes.

A short time, the same DNA element can have evenly enhancer and silencer activity depending on the protein bound to it. For example, the thyroid hormone response element binds to it without its ligand (thyroid hormone) but acts as an enhancer when the thyroid hormone receptor binds along with the ligand.

2.4 TRANSCRIPTIONAL ACTIVATORS AND TRANSCRIPTION REGULATION

Transcription activators are modular proteins that promote the transcription of genes from DNA to RNA. In most cases, these proteins include two separate domains - a domain that binds to DNA and a domain for activating transcription. Although in some cases, a single domain is responsible for both binding and activation of transcription.

The binding domains are able to recognize and interact among regulatory sequences on the DNA. These domains are classified into dissimilar families and named according to the structural individuality that enables DNA recognition and binding. Some common types of binding domains include the leucine zipper, zinc finger, and helix-turn-helix motifs. Domains responsible

for gene transcription activation are regularly short, simple sequences and are less complex than the binding domains. They are classified by amino acid composition into categories, such as glutamine-rich, proline-rich, and alanine-rich.

Transcription activators support the recruitment of different proteins required for transcription such as general transcription factors, RNA polymerase, and co-activators. All these proteins jointly known as the pre-initiation complex and depend on transcription activators for their enrollment to the appropriate location. These activators can connect to a site close to the gene's promoter or several thousand base pairs away from the gene to carry out their function. In cases where they are bound to a site away from the gene, they rely on the elasticity of the DNA to bend and bring them in closeness to the gene promoter. Transcription activators are also equired to continue a transcript's elongation or the re-initiation of transcription in cases where the process stops midway. Transcription activators are known to act synergistically. The transcription achieved by the action of numerous activators is higher than what would occur as a sum of individual factors working independently.

Like other proteins, transcription activators are also subjected to post-transcriptional modifications. In several cases these modifications assist in positive regulation of transcription. For example, acetylation of p53, an activator that regulates genes responsible for tumor suppression, increases its capacity for binding to DNA.

In genetics and molecular biology, transcriptional regulation is the means by which a cell regulates the change of DNA to RNA (transcription), thereby orchestrating gene activity. A single gene can be regulated by number of ways, from varying the number of copies of RNA that are transcribed, to the temporal control when the gene is transcribed. These controls allow the cell or organism to react to a diversity of intra- and extracellular signals and therefore mount a response. Some examples of this comprise producing the mRNA that encode enzymes in response to a change in a food source, producing the gene products related to cell cycle and producing the gene products responsible for cellular demarcation in multi cellular eukaryotes in evolutionary developmental biology.

2.5 REGULATORY ELEMENTS AND MECHANISM OF TRANSCRIPTION REGULATION

The main regulatory element in gene expression regulation is the rate of transcription initiation. The rate is controlled by core promoter elements as well as distant-acting regulatory elements such as enhancers. Processes of histone modifications and/or DNA methylation include a crucial regulatory contact on transcription. If a region is not existing for the transcriptional mechanism, e.g. in the case wherever the chromatin structure is compressed due to the occurrence of definite histone modifications, or if the promoter DNA is methylated, transcription could not start at all. Last but not least, gene activity is too controlled post-transcriptionally by ncRNAs such as micro RNAs (miRNAs), as well as by cell signaling, resulting in protein change or altered protein-protein interactions.

The mechanisms regulating gene expression are necessary for all living organisms as they state where and how a gene product should be expressed. This regulation could take place at the pre- and co-transcriptional level by calculating how many transcripts should be formed and or which transcript should be produced by regulating splicing. The same gene could encode for diverse versions of the same protein via splicing regulation. These procedures define which parts of the gene will go into the final mRNA that will code for the protein variant. In adding, gene products can be regulated post-transcriptionally where certain molecules bind to RNA and mark them for degradation even before they can be used in protein manufacture.

Gene regulation drives cellular differentiation; a procedure during which different tissues and cell types are produced. It moreover helps cells maintain differentiated states of cells/tissues. As a result of this procedure, at the final stage of differentiation, different kinds of cells preserve different expression profiles, though they include the same genetic material.

Transcriptional regulation

In genetics and molecular biology, **transcriptional regulation** is is an essential process by which a cell regulates the change of DNA to RNA (transcription), thus orchestrating gene activity. A single gene can be regulated in a variety of activities, from varying the number of copies of RNA that are transcribed; to the sequential control while the gene is transcribed. This control allows the cell or organism to react to a range of intra- and extracellular signals and therefore mount a response.

It is orchestrated by transcription factors and other proteins working in concert to finely tune the amount of RNA being produced during a variety of mechanisms. Bacteria and eukaryotes have very different strategies of accomplishing manage over transcription, but some significant features remain conserved among the two. Most accepted is the idea of combinatorial control, where any given gene is controlled by a definite combination of factors to control transcription. In a hypothetical example, the factors A and B may regulate a different set of genes from the mixture of factors A and C. This combinatorial character extends to complexes of distant more than two proteins, and allows a very small subset (less than 10%) of the genome to control the transcriptional program of the entire cell.

2.6 TRANSCRIPTIONAL POLYMERASE, CAPING, ELONGATION & TERMINATION

RNA polymerases: RNA polymerases are enzymes that read the template strand of the DNA and create an equivalent mRNA strand. They are made up of several subunits.

Prokaryotic cells: prokaryotic cell have only one type of RNA polymerase and require only a single protein, known as sigma factor, to bind to the promoter sequence.

Eukaryotic cells: Eukaryotic cells are three types of RNA polymerase

i) RNA polymerase I synthesizes rRNA.

ii) RNA polymerase II synthesizes mRNA.

iii) RNA polymerase III synthesizes tRNA.

Several transcription factors are essential to bind to the DNA at the promoter sequence (RNA polymerase II cannot bind on its own).

The process of transcription involves three major steps: initiation, elongation, and termination.

1. Initiation

Initiation is the creation of transcription. It occurs while the enzyme **RNA polymerase** binds to a section of a gene called the **promoter.** This signals the DNA to unwind, so the enzyme can ''read'' the bases in one of the DNA strands. The enzyme is now ready to make a strand of mRNA with a complementary sequence of bases. The transcription initiation involves four steps:

1. Development of a closed promoter complex by loose binding of enzyme of RNA polymerase to DNA.

2. Exchange of closed promoter complex to an open promoter complex due to melting of DNA locally.

3. Polymerizing the first few nucleotides (up to ~ 10) while the polymerase remains at the promoter, this is called abortive transcription.

4. Clearing of promoter when the s factor gets detached from the enzyme and polymerase and the polymerase core enzyme moves away from the promoter.

2. Elongation

One DNA strand (the template strand) is read in a 3′ to 5′ direction, and therefore provides the template for the new mRNA molecule. The other DNA strand is referred to as the coding strand. This is because its base sequence is equal to the synthesized mRNA, apart from the substitution of thiamine bases with uracil. RNA polymerase uses inward ribonucleotides to structure the new mRNA strand. It does this by catalyzing the formation of phosphodiester bonds among adjacent ribonucleotides, using corresponding base pairing (A to U, T to A, C to G and G to C). Bases can be added singlyto the 3′ (three-prime) end, so the strand elongate in a 5' to3' direction.
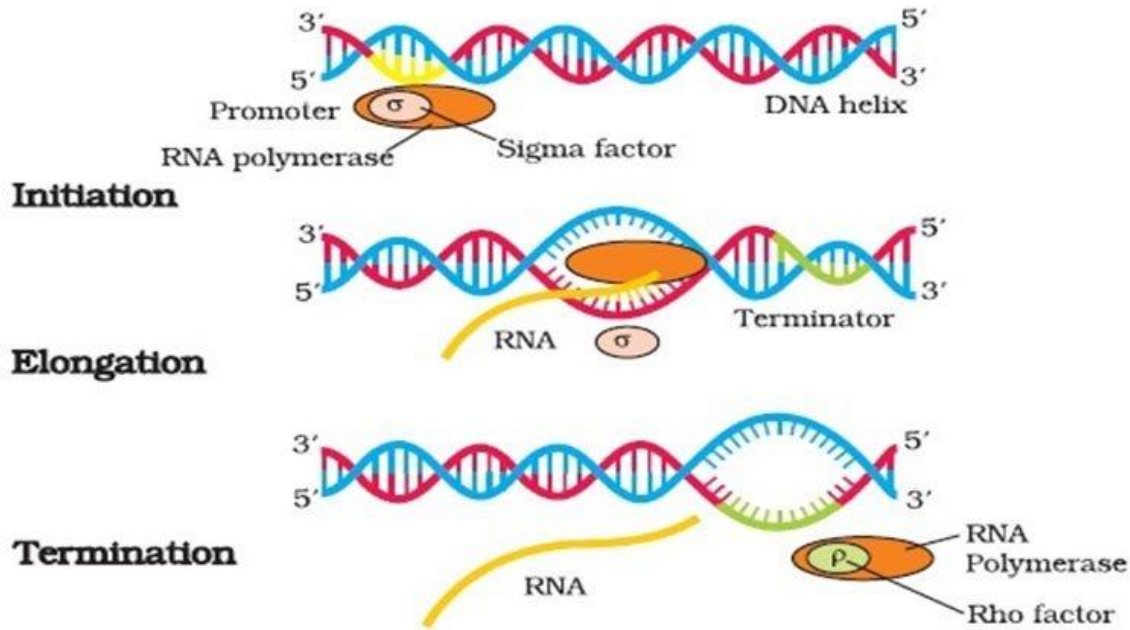
Fig.2.2The three stages of transcription: Initiation, Elongation and Termination

(https://www.google.com/search?q=Bacterial+RNA+polymerase+activity+diagram&source=lnms&tbm)

3. Termination

The ending process of transcription is termination and occurs when RNA polymerase crosses an end (termination) sequence in the gene. The mRNA strand is complete, and it detaches from DNA template.

Pre-RNA and mRNA

After transcription, the eukaryotic **pre-mRNA**s must undergo numerous processing steps before they can be translated. Eukaryotic and prokaryotic rRNAs and tRNAs also go through processing before they can function as components in the protein synthesis machinery.

Processing of mRNA

The eukaryotic pre-mRNA undergoes wide processing before it is ready to be translated. The extra steps involved in eukaryotic mRNA maturation create a molecule having much longer half-life than prokaryotic mRNA. Eukaryotic mRNAs last for numerous hours, whereas the typical *E. coli* mRNA lasts no more than five seconds.

The most three important steps of pre-mRNA processing are the addition of stabilizing and signaling factors at the 5′ and 3′ ends of the molecule, and the deletion of intervening sequences that do not identify the suitable amino acids.

5′ Capping

A cap is additional to the 5′ end of the increasing transcript by a phosphate linkage. This addition protects the mRNA from degradation. Factors involved in protein synthesis identify the cap to initiate translation by ribosomes.

3′ Poly-A Tail Poly-A Tail

Once elongation is complete, an enzyme called poly-A polymerase adds a series of about 200 A residues, called the **poly-A tail** to the pre-mRNA. The poly (A) tail protects the mRNA from degradation, aids in the export of the mature mRNA to the cytoplasm, and is involved in binding proteins involved in initiating translation.

Pre-mRNA Splicing

Eukaryotic genes are composed of exons, which correspond to protein-coding sequences (ex-on signifies that they are expressed), and intervening sequences called introns (int-ron denotes their intervening role), which may be involved in gene regulation, but are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional proteins.

All of a pre-mRNA's introns must be totally and specifically detached before protein synthesis. If the procedure errs by even a single nucleotide, the reading structure of the rejoined exons would shift, and the consequential protein would be dysfunctional. The procedure of removing introns and reconnecting exons is called **splicing**.

2.6.1 STRUCTURE AND FUNCTION OF DIFFERENT TYPE OF RNA

Ribonucleic Acid (RNA)

Some plant viruses (e.g.,TMV,wound tumour viruses etc.), animal viruses (e.g., influenza viruses, foot and mouth viruses, poliomyelitis viruses, reoviruses, etc) and bacteriophages e.g.,

MS_2), contain ribonucleic acid (RNA) as their genetic material, like DNA. RNA is polymeric nucleic acid of four monomeric ribotids or ribonucleotides. Every ribonucleotide contains a pentose sugar (D- ribose), a molecule of phosphoric acid and a nitrogen base. The nitrogen bases of RNA are two purines, adenine and guanine and two pyimidines, cytosine and uracil (fig. 2.3).
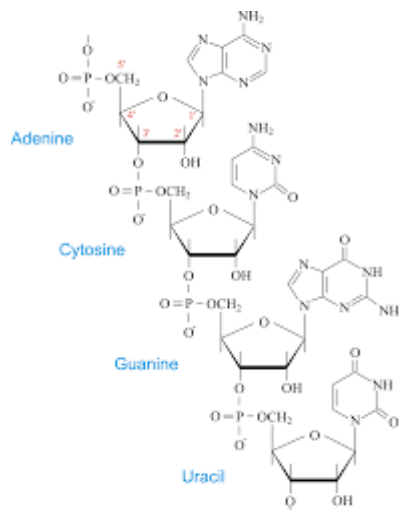
Fig.2.3 Structure of RNA

(Source:https://www.google.com/search?q=structure+of+rna+diagram&tbm=isch&ved)

Molecular structure of RNA

RNA molecule may be either single stranded or double stranded but not helical like DNA molecule. Single stranded RNA occurs as genetic material in plant viruses (e.g TMV, TYM) animal viruses (e.g., influenza viruses, foot and mouth viruses, poliomyelitis viruses, reoviruses, and bacteriophages e.g., MS_2). The non- genetic RNAs except tRNA of prokaryotes and eukaryotes also have single strand RNA molecules. The double stranded but non- helical RNA occurs as the genetic material in some plant viruses (e.g., reoviruses). The transfer or soluble RNA (tRNA or sRNA) which is non-genetic RNA of prokaryotes and eukaryotes is double stranded but non- helical structure.

Each strand of RNA is polynucleotidic, that is made up of many ribonucleotides. In the polynucleotide strand of RNA, the ribose and phosphoric acids of nucleotides remain linked by phophodiester bonds. The organisms which have only RNA, employ RNA in their genetic mechanism and that type of RNA is called **genetic RNA**, while the organism which have DNA along with RNA, they use the RNA forcarrying the informationof DNA,thusRNA has no genetic

role, so called **non- genetic RNA**. The non- genetic RNA is heterogeneous and has DNA-dependent replication of itself, that is, it is not self-replicating like DNA or transcribed by DNA. It includes following three genera.

I. RIBOSOMAL RNA (r RNA)

Ribosomal RNA (r RNA) or insoluble RNA constitutes the largest part (up to 80%) of the total cellular RNA. It is found primarily in the ribosomes although; since it is synthesized in the nucleus it is also detected in that organelle. It contains the four major RNA bases with a slight degree of methylation, and shows differences in the relative proportions of the bases between species. Its molecules appear to be single polynucleotide strands which are unbranched and flexible. At low ionic strength, rRNA behaves as random coil, but with increasing ionic strength the molecule shows helical regions produced by base pairing between adenine-uracil and guanine-cytosine.

Types are, namely 28S rRNA (the sedimentation constant varies between 25S and 30S depending upon the species), 18S rRNA and 5S rRNA. The 28S rRNA and 5S rRNA occur in 60S ribosomal subunit, while 18S rRNA occurs in 40S ribosomal subunit of 80S ribosomes of eukaryotes. The prokaryotic cells also contain three kinds of rRNA molecules, namely 23S rRNA, 16S rRNA and 5S rRNA. The 23S rRNA and 5S rRNA occur in 50S ribosomal subunits of 70S ribosomes of prokaryotes. In eukaryotic cells of plants and animals the 18S and 28S rRNA molecules are transcribed by rDNA of nucleolar organizerregion of chromosomes and nucleolus acts as site of synthesis and maturation of these rRNA molecules. 5S rRNA is transcribed by rDNA, outside the nucleolar organizer region.

The functions of rRNA molecules are still little understood.Although, they are supposed to have some definite role in protein synthetic process, besides being the major constituent molecules of ribosomes.

II. MESSENGER RNA (mRNA)

These RNA molecules are transcribed from large number of genes of the total genome (i.e., 99% genes of the total genome of *E.coli),* and have base sequence complementary to DNA. The sequence of nucleotides in mRNA, like DNA, contains genetic information, which is organized into codons of three ribonucleotides apiece. Except for stop codons, which cease protein synthesis, each codon codes for a distinct amino acid. Two further types of RNA are required for

the translation of codons into amino acids: transfer RNA, which identifies the codon and gives the correct amino acid, and ribosomal RNA (rRNA), which is the core component of the ribosome's protein-making machinery.The name messenger RNA has been proposed by Jacob and Monod (1961).
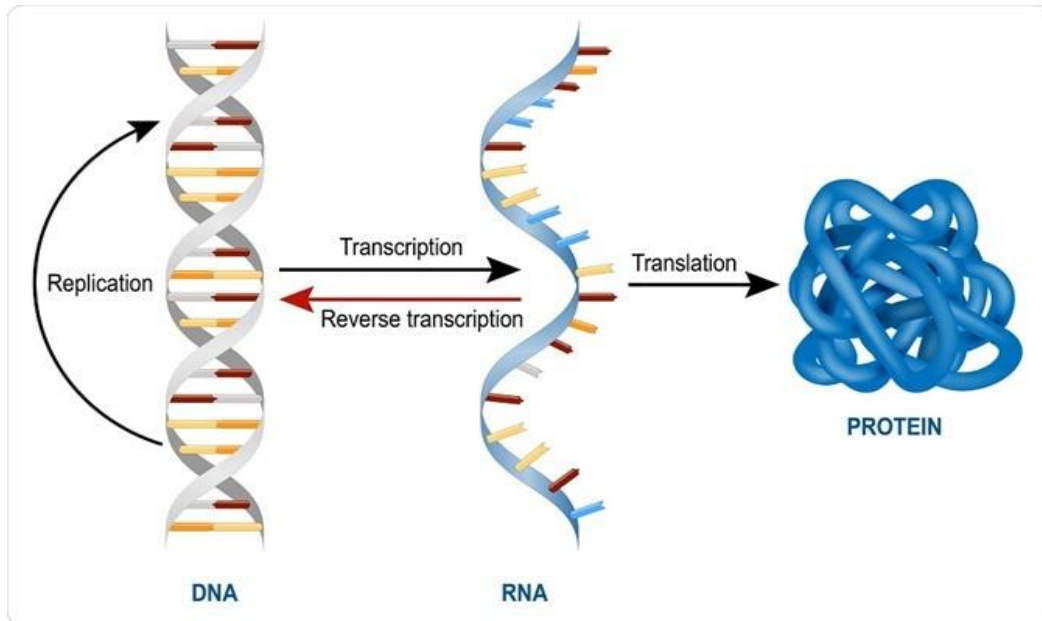
Fig.2.4 Types of RNA; mRNA, rRNA and tRNA(Source: https://www.news-medical.net/life-sciences/-Types-of-RNA-mRNA-rRNA-and-tRNA.aspx)

The molecule of mRNA is single-stranded like the rRNA molecule and it is DNA-like in its base composition so that GC contents of mRNA correspond to its base composition so that GC contents of mRNA correspond to the GC contents of the genome's total DNA.

Synthesis of mRNA- The origin and fate of mRNA in eukaryotic cells is much more complex than in bacteria. The formation of a functionally active mRNA is the consequence of a complex series of steps that comprise

i) the actual transcription of DNA into mRNA precursors

ii) the intra-nuclear processing of these precursors

iii) the transport of the mRNA into the cytoplasm

Life Span of mRNA

In most prokaryotic and eukaryotic cells, mRNA has short life time. For example, the average life of mRNA of *E. coli* is about 2 minutes due to presence of cytoplasmic ribonuclease enzyme. So that, at most times, mRNA makes up only 5% of the total cellular RNA. Besides, in most eukaryotes the average life span of mRNA is one to four hours. Although, in both bacteria and eukaryotes, metabolically stable mRNAs are known that are apparently resistant to nucleases and also survive for long period of time. For example, mRNA with life time of six hours has been detected in the bacterium *Bacillus cerus* at a time when the cells are induced to become long lived spores. Similarly, in differentiated eukaryotic cells mRNA with a life time of days have been detected. For example, in immature red blood cells (reticulocytes) of the mammals the mRNA is synthesized initially by the nucleus in early stages and excluded to the cytoplasm. In later stages the nuclei of maturing reticulocytes degenerate but the mRNA exists up to 2 days for prolong use in the synthesis of globin protein of Haemoglobin. Further, in severe cases, such as in the state of dormancy accepted by many animal eggs and plant seeds, mRNA is maintained in a constant form for months or even years.

III TRANSFER RNA (t RNA)

The RNA which possesses the capability to combine exclusively with only one amino acid in a reaction mediated by a set of amino acid-specific enzymes called **aminoacyl-tRNA synthetases,** which transport that amino acid from the "amino acid pool" to the place of protein synthesis and recognizes the codons of the mRNA is known as the soluble RNA (sRNA) or transfer RNA (tRNA). Therefore,tRNA molecule acts as interpreter of genetic code and has to perform several highly complex functions during protein synthesis. It interacts with a specific synthetase enzyme, possesses a site for binding of an amino acid, a second site for interacting with a ribosome and also contains an anticodon that must be exposed to the codons of mRNA.

Structure and maturation of tRNA

The tRNA molecules that perform all these function are relatively small they contain only 70 to 80 nucleotides so that each has a molecular weight of about 30,000 and a sedimentation coefficient of 4S. Like the rRNA and mRNA, the molecules of tRNA are found to be matured or tailored in the nucleus prior to their movement to the cytoplasm. For example, in *Ecoli* the precursors of tRNA molecules have been isolated, each of which has about 40 extra nucleotides, principally at 5' end but also at the 3' end. These extra nucleotides are consequently cleaved off

by still unidentified enzymes to yield a molecule of final 70 to 80 nucleotide size. Aftersuch large scale tailoring each tRNA molecule is more precisely modified before it become fully active. Three base, 5'-CCA-3' are added to the 3' end of every tRNA molecule, regardless of its amino acid affinity, by an enzyme called tRNA phosphorylase. Byadding, specific nucleotides a tRNA is transformed into what are called minor or unusual bases by specific chemical modification, for example, methyl group are added at certain position to yield 3-methyl-cytosine or l-methylguanosine, certain uracils are reduced to dihydrouracil or rearranged into a form, known as pseudouracil, and adenine is deaminated to yield inosine.

The importance of these unusual bases of tRNA was understood well by molecular biologists during the structure of two-dimensional model from the primary- sequences of nucleotides of known tRNA. Therefore, it was realized that most bases of tRNA pair according to Watson – Crick's pairing rule, but unusual base fail to do so because they carry substitutions or alterations in those positions that usually participate in hydrogen bonding. Therefore, the presence of these bases forces the model builder to construct several non-base-paired loops in the tRNA molecule. By working on these lines, R.Holley (1968) first of all proposed a clover leaf model for yeast tRNA. The clover leaf model of tRNA because accommodated several of the known functions oftRNA, so it gained general acceptance. A typical clover-leaf model of tRNA depicts following structural peculiarities.

(i) All tRNA molecules contain the same terminal sequence of 5'-CCA-3' bases at 3'-end of the polynucleotide chain. The last residue, adenylic acid (A), is the amino acid attachment site.

(ii) All tRNAhas a loop called Tψ arm of seven unpaired bases including pseudouridine. The Tψ arm is involved in the binding of the tRNA molecules to the ribosomes.

(iii)All tRNA molecules contain a site for the recognition of the amino acid activating synthetase enzymes. This is the function of thedihydrouridine loop or DHU arm which is made up of 8 to 12 unpaired bases.

(iv) There is one nucleotide triplet in the tRNA molecule which is different in all tRNAs examined. This is the codon recognition site or anticodon and it is complementary to the corresponding triplet codon of mRNA.

(v) Some tRNA with long chains may form a short, extra arm.

2.6.2 RNA TRANSPORT

In the process of RNA transport, ribonucleic acid (RNA) molecules are actively transported from one location within the cell to another. A small number of RNAs are retained in the nucleus, and may be embattled to sub nuclear domains, but the majority of RNAs are transported to the cytoplasm.
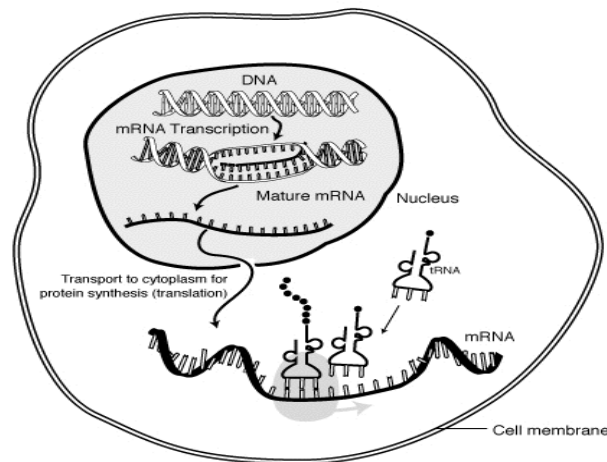
Fig. 2.5MessengerRNA (Source Wikipedia en.wikipedia.org)

RNA molecules synthesized in the nucleus are transported to their sites of function through the eukaryotic cell by specific transport pathways. This analysis focuses on transport of messenger RNA, small nuclear RNA, ribosomal RNA and transfer RNA between the nucleus and the cytoplasm.

Transport of messenger RNA (mRNA) starting the nucleus to the cytoplasm is a necessary step of eukaryotic gene expression. In the cell nucleus, an ancestor mRNA undergoes a sequence of processing steps, with capping at the 5' ends, splicing and cleavage/polyadenylation at the 3' ends. Throughout this procedure, the mRNA associates with a wide variety of proteins, forming a messenger ribonucleoprotein (mRNP) element. Relationship with factors involved in nuclear export too occurs during transcription and processing, therefore nuclear export is fully included into mRNA maturation. The coupling between mRNA maturation and nuclear export is a significant mechanism for providing only fully useful and competent mRNA to the cytoplasmic translational machinery, thus ensuring accuracy and swiftness of gene expression.

2.7 SUMMARY

The transfer of genetic information DNA to RNA (mRNA) is known as transcription. It involves three types of RNA namely: mRNA (messenger RNA), tRNA(transfer RNA), and iii) rRNA (ribosomal RNA). The process of transcription involves copying a segment of DNA into RNA. The segments of DNA transcribed into RNA molecules that can encode proteins are said to produce messenger RNA (mRNA). Other segments of DNA are copied into RNA molecules called non-coding RNAs (ncRNAs).

Prokaryotic transcription also known as bacterial transcription is the process in which a segment of bacterial DNA is copied into a newly synthesized strand of messenger RNA (mRNA) which is later translated to produce proteins with the use of the enzyme RNA polymerase and other transcription factors.

Bacterial transcription is different from eukaryotic transcription in various ways. In bacteria, transcription and translation can occur simultaneously in the cytoplasm of the cell, although in eukaryotes transcription occurs in the nucleus and translation occurs in the cytoplasm. Generally, transcription within bacteria is a highly regulated process that is controlled by the integration of many signals at a particular time. Bacterial transcription occurs in main three steps: Initiation, Elongation and termination.

The actual process of transcription in eukaryotes is nearly the same as in the prokaryotes, there is much variability in the factors associated with the process.

The three types of eukaryotic RNA polymerases identify different kinds of promoters viz., polymerase I recognizes class I promoter, polymerase II does class II while polymerase III recognize class III promoters. Of this class II promoters are relatively well studied and bear a greater resemblance to those of the prokaryotes.

Transcription activators are modular proteins that promote the transcription of genes from DNA to RNA. Transcription activators support in the recruitment of different proteins required for transcription such as general transcription factors, RNA polymerase, and co-activators. All these proteins jointly are known as the pre-initiation complex and depend on transcription activators for their enrollment to the appropriate location. These activators can connect to a site close to the gene's promoter or several thousand base pairs away from the gene to carry out their function.

Transcription activators are also subjected to post-transcriptional modifications that assist in positive regulation of transcription. In genetics and molecular biology, **transcriptional regulation** is the means by which a cell regulates the change of DNA to RNA (transcription), there by orchestrating gene activity.

2.8 TERMINAL QUESTION & ANSWERS

2.8.1 Multiple Choice Questions:

1. The enzyme required for transcription is

a) RNAase

b) RNA polymerase

c) DNA polymerase

d) Restriction enzymes

2. Transcription is the transfer of genetic information from

a) DNA to RNA

b) DNA to mRNA

c) mRNA to tRNA

d) tRNA to mRNA

3. Sigma factor is a component of

a) DNA ligase

b) DNA polymerase

c) Endonuclease

d) RNA polymerase

4. Which of these subunits of RNA polymerase is totally required to initiate transcription?

a) alpha (α)

b) sigma (σ)

c) omega (ω)

d) beta (β)

5. In eukaryotes, in order to initiate transcription

a) RNA strand must be present

b) RNA polymerase must be present

c) Core promoter sequence must be present

d) None of these

Answers: 1 b, 2 b, 3 d, 4 b, 5 c.

2.8.2 Short Answer Question:

1. What is transcription?

2. Write short notes on: i) Enhancers and silencers ii) Initiation iii) Elongation

3. Write about eukaryotic promoters.

4. What is transcriptional regulation?

2.8.3 Long Answer Question:

1. Describe the Structure and Function of different type of RNA.

2. Explain the Prokaryotic transcription.

3. Describe the essential mechanism of transcription.

4. Explain transfer RNA (tRNA).

2.9 GLOSSARY

Adenine: A purine base occurring in DNA and RNA. Pairs normally with thymine in DNA.

Adenosine triphosphate(ATP): The major source of usable energy in cell metabolism composed of adenine, ribose and three phosphate groups. On hydrolysis, ATP loses one phosphate and one hydrogen to become adenosine diphosphate (ADP), releasing energy in the process.

Base: On the DNA molecule, one of the four chemical units that, according to their order, represent the different amino acids. The four base are: adenine (A), cytosine(C), guanine (G) and thymine (T). In RNA, uracil (U) substitutes for thymine.

Base pair: Two nucleotide base on different strands of a nucleic acid molecule that bond together. The bases generally pair in only two combination, adenine with thymine (DNA) or uracil (RNA), and guanine with cytosine.

Deoxyribonucleic acid (DNA): The molecule that carries the genetic information for most living systems. DNA molecule consists of four bases (adenine, cytosine, guanine, and thymine) and a sugar-phosphate backbone, arranged in two connected strands to form a double helix.

Enzyme: A large biological molecule, typically a protein that catalyzes a chemical reaction.

Eukaryote: A cell or organism containing a true nucleus, with a well-defined membrane surrounding the nucleus. All organisms except bacteria, arch bacteria, viruses, and blue-green algae are eukaryotic. *Cf.* Prokaryote.

Exon: Any segment of a gene that is represented in the mature messenger RNA (m RNA) product.

Gene: A segment of chromosome that encodes the necessary regulatory and sequence information to direct the synthesis of a protein or RNA product.

Gene Expression: The process through which a gene is activated at particular time and place so the it's functional product is produced.

Messenger RNA (mRNA): Nucleic acid that carries instructions to a ribosome for the synthesis of a particular protein.

Nucleic acid: Large molecules, generally found in the cell's nucleus and/or cytoplasm that are made up of nucleotide bases. The two kinds of nucleic acid are DNA and RNA.

Nucleotides: The building blocks of nucleic acids. Each nucleotide is composed of sugar, phosphate, and one of four nitrogen bases. If the sugar is ribose, the nucleotide is termed a "ribonucleotide," whereas deoxyribonucleotides have deoxyribose as the sugar component (i. e. adenine, cytosine, guanine and thymine in the case of DNA). The sequence of the nucleotides within the nucleic acid determines, for example, the amino acid sequence of an encoded protein.

Prokaryote: A cellular organism (e.g., bacterium, blue-green algae) whose DNA is not enclosed within a nuclear membrane. *Cf.* Eukaryote.

Promoter: A DNA sequence that is located near or even partially within encoding nucleotide sequences and which controls gene expression. Promoters are required for binding of RNA polymerase to initiate transcription.

Ribonucleic acid (RNA): A single stranded nucleic acid molecule comprising a linear chain made up from four nucleotide subunits (A, C, G, and U). There are three types of RNA: from four nucleotide subunits (A, C, G, and U). There are three types of RNA: messenger, transfer and ribosomal. (Actually there are also ribosomes etc.)

Splicing: The removal of introns and joining of exons to form a continuous coding sequence in RNA.

Template: A molecule that serves as the pattern for synthesizing another molecule.

Transcription: The enzymatic process, involving base pairing, by which the genetic information contained in DNA is used to specify a complementary sequence of base in an RNA molecule.

Transfer RNA (t RNA): RNA molecules that carry amino acids to sites on ribosomes where proteins are synthesized.

Translation: Process by which the information on a messenger RNA molecule is used to direct the synthesis of a protein.

2.10 REFERENCES

1. Genetics by P. S. Verma& V. K.Agarwal

2. Gupta, P. K,.Cell and Molecular Biology 1st Ed. Rastogi Publications., Meerut

3. https://www.jove.com/science-education/11613/eukaryotic-transcription-activators

4. https://en.wikipedia.org/wiki/Transcriptional regulation

5. https://compgenomr.github.io/book/elements-of-gene-regulation.html

6. https://teachmephysiology.com/biochemistry/protein-synthesis/dna-transcription/

7. https://www.google.com/search?q=rna+transport+wikipedia&tbm=isch&source

8. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4488659/

UNIT 3: POST-TRANSCRIPTIONAL MODIFICATIONS IN RNA

CONTENTS

3.1 OBJECTIVES

After reading this unit the readers will be able to:

➢   Understand Post- transcriptional modification.

➢   Discuss 5'- cap formation, end processing and polyadenylation.

➢   Discuss splicing and editing

➢   Explain Nuclear export of mRNA

3.2 INTRODUCTION

PTMs (post-transcriptional modifications) or **co-transcriptional modification** is a set of biological mechanisms that help to produce mature, functional RNA. This fast-acting regulatory mechanism allows various proteins to be produced from a single gene and act as phenotypic and proliferation rate regulators. There are various types of post-transcriptional modifications achieved throughout a diverse class of molecular mechanisms. Some types of cancer and neurological illnesses are linked to these changes. For stability and protection, the pre-messenger RNA (mRNA), also known as heterogeneous nuclear RNA (hnRNA), is changed by adding a 5' 7-methylguanosine cap and a 3' poly-A (polyadenylate) tail. Splicing is also performed on hnRNA that contains introns (noncoding sequences) among the transcribed sequences or exons. Introns are removed during this step, resulting in a mature mRNA that contains the coding sequence for translation. Alternative splicing, on the other hand, excludes the introns but connects exons in different ways, resulting in distinct proteins from the same mRNA. The mRNA sequence is changed in RNA editing, and it differs from the transcribed DNA template. Transfer RNA and ribosomal RNA are made up of longer precursor molecules that have been methylated, trimmed, and nucleotides added to them.

The addition of a 5' cap, the addition of a 3' polyadenylated tail, and RNA splicing are three important stages that drastically alter the chemical structure of the RNA molecule. Because the initial precursor mRNA produced by transcription often contains both exons (coding sequences) and introns (non-coding sequences), splicing removes the introns and directly links the exons, while the cap and tail facilitate the transport of the mRNA to a ribosome and protect it from molecular degradation.

Post-transcriptional modifications can also occur throughout the processing of other transcripts which eventually become transfer RNA, ribosomal RNA, or any of the extra types of RNA used by the cell.

3.3 5'- CAP FORMATION

The **five-prime cap** (**5′ cap**) is a particularly altered nucleotide on the 5′ end of a few primary transcripts such as precursor messenger RNA. This process, known as mRNA capping, is tightly controlled and essential for the production of stable and mature messenger RNA capable of being translated during protein synthesis. Mitochondrial and chloroplastic mRNAs aren't capped.

In eukaryotes, the 5′ cap (cap-0), begin on the 5′ end of an mRNA molecule and consist of a guaninenucleotide connected to mRNA via an abnormal 5′ to 5′ triphosphate linkage. This guanosine is methylated on the 7 situation directly after capping *in vivo* through a methyltransferase.It is referred to as 7-methylguanylate cap, i. e.m^{7}G.

Further alterations exist in multicellular eukaryotes and some viruses, including methylation of the 2′ hydroxy-groups of the first two ribose sugars of the 5′ end of the mRNA. The first ribose sugar in cap-1 has a methylated 2′-hydroxy group, but the first two ribose sugars in cap-2 have methylation 2′-hydroxy groups, The 5′ cap is chemically similar to the 3′ end of an RNA molecule (the 5′ carbon of the cap ribose is bonded, and the 3′ unbonded). This provides significant resistance to 5′ exonucleases.Small nuclear RNAs contain unique 5′-caps. Sm-class snRNAs are found with 5′-trimethylguanosine caps, whereas Lsm-class snRNAs possess 5′-monomethylphosphate caps.
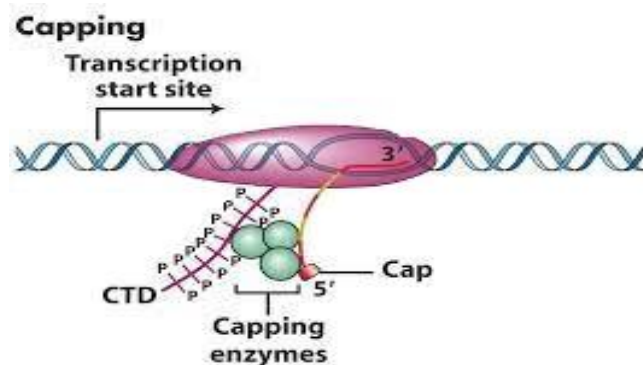
Fig.3.1RNA processing eukaryotic mRNAs (sourcehttp://www.csun.edu/~cmalone/pdf360/Ch13-2RNAprocess.pdf)

In bacteria, and also in higher organisms, some RNAs are capped with NAD^+, NADH, or 3′-dephospho-coenzyme A.

In all organisms, mRNA molecules can be decapped in a process known as messenger RNA decapping.

Capping process:

The initiation point for capping with 7-methylguanylate is the unchanged 5′ end of an RNA molecule, which terminates at a triphosphate group. This features a resultant nucleotide followed by three phosphate groups attached to the 5′ carbon. The capping procedure is initiated previous to the completion of transcription, as the nascent pre-mRNA is being synthesized.

1. One of the terminal phosphate groups is detached by RNA triphosphatase, leaving a bisphosphate group (i.e. $5′(ppN)[pN]_n$);

2. mRNA guanylyltransferase adds GTP to the terminal bisphosphate, losing a pyrophosphate from the GTP substrate in the process. The 5′–5′ triphosphate linkage is formed, yielding 5′(Gp)(ppN)[pN]n.

3. Through mRNA (guanine-N7-)-methyltransferase, the 7-nitrogen of guanine is methylated, and S-adenosyl-L-methionine is demethylated to produce S-adenosyl-L-homocysteine, resulting in 5′(m7Gp)(ppN)[pN]n (cap-0);

4. Cap-adjacent alterations can proceed at the first and second nucleotides, resulting in 5′(m7Gp)(ppN)(pN)[pN]n (cap-1 and cap-2);

5. If the closest cap-adjacent nucleotide is 2′-O-ribose methyl-adenosine (i.e. 5′(m7Gp)(ppAm)[pN]n), it can be methylated at the N6 methyl site to create N6-methyladenosine, yielding 5′(m7Gp)(ppm6Am)[pN]n.

Capping with NAD+, NADH, or 3′-dephospho-coenzyme A has a distinct process. Capping with NAD+, NADH, or 3′-dephospho-coenzyme A occurs through a "ab initio capping mechanism," in which NAD+, NADH, or 3′-dephospho-coenzyme A acts as a "non-canonical initiating nucleotide" (NCIN) for RNA polymerase transcription initiation and is thus directly incorporated

into the RNA product. [9] This "ab initio capping method" can be carried out by both bacterial and eukaryotic RNA polymerase II.

Targeting:

The capping enzyme complex (CEC) attaches to RNA polymerase II before transcription begins to cap with 7-methylguanylate. The CEC completes the capping process as soon as the 5′ end of the new transcript emerges from RNA polymerase II (this kind of mechanism ensures capping, as with polyadenylation). The capping enzymes can only bind to RNA polymerase II, ensuring that only these transcripts, which are essentially completely mRNA, are targeted.

The promoter sequence targets capping with NAD+, NADH, or 3′-dephospho-coenzyme A. NAD+, NADH, or 3′-dephospho-coenzyme A capping occurs only at promoters with specific sequences near and immediately upstream of the transcription start site, and hence only for RNAs generated from those promoters.

Function:

The 5′cap have four main functions:

1. Nuclear export regulation

2. Exonuclease degradation prevention;

3 Translation promotions;

4. 5′ proximal intron excision promotion

The cap binding complex (CBC), which binds solely to 7-methylguanylate-capped RNA, regulates RNA nuclear export. The nuclear pore complex then recognizes the CBC and exports it. After the pioneer round of translation, the CBC is replaced in the cytoplasm by the eIF4F complex's translation factors eIF4E and eIF4G. Other translation initiation machinery, such as the ribosome, identifies this complex.

By two ways, capping with 7-methylguanylate prevents 5′ degradation. First, by functionally resembling a 3′ end, 5′ exonucleases are inhibited from degrading the mRNA. Second, the CBC and eIF4E/eIF4G prevent decapping enzymes from accessing the cap. This prolongs the half-life of mRNA, which is important in eukaryotes because the export and translation processes require a long time.

The decapping complex, which consists of at least Dcp1 and Dcp2, catalyses the decapping of a 7-methylguanylate-capped mRNA and must compete with eIF4E to bind the cap, catalyses the decapping of a 7-methylguanylate-capped mRNA. As a result, the 7-methylguanylate cap serves as a marker for actively translating mRNA, and cells employ it to modify mRNA half-lives in response to novel stimuli. Unwanted mRNAs are transferred to P-bodies for temporary storage or decapping, with the intricacies still being worked out.

Although the mechanism favouring 5′ proximal intron excision is unknown, the 7-methylguanylate cap appears to loop around and interact with the spliceosome during the splicing process, encouraging intron excision.

3.4 END PROCESSING AND POLYADENYLATION

Recent developments in the molecular mechanism of mRNA end processing have revealed the formerly unknown complex system of transcriptional and RNA-processing mechanisms. The relevance of the accuracy of the complex end-processing machinery is reflected in a wide range of human disorders, and gene-specific dysregulation of end processing can come from mutations in RNA sequence regions that bind essential specialized processing components. Interestingly, more widespread dysregulation of end processing can be produced by mutations in these processing components or by a disruption in their well-coordinated equilibrium.

Both loss of function and increase of function can be functionally relevant from a clinical perspective, and an increasing number of disease entities demonstrate that improper end creation of human mRNAs can have a huge impact on health and illness.

The addition of a poly (A) tail to an RNA transcript, usually a messenger RNA, is known as polyadenylation (mRNA). The poly (A) tail is a stretch of RNA that solely comprises adenine bases and is made up of numerous adenosine monophosphates. Polyadenylation is a step in the process of producing mature mRNA for translation in eukaryotes. The poly (A) tail accelerates mRNA breakdown in several bacteria. As a result, it is a part of the wider gene expression process.

When a gene's transcription is finished, the polyadenylation process begins. A collection of proteins cleaves the 3′-most section of the newly generated pre-mRNA, and these proteins

subsequently construct the poly (A) tail at the RNA's 3′ end. These proteins add a poly (A) tail to some genes at one of several potential locations. As a result, polyadenylation, like alternative splicing, can yield many transcripts from a single gene (alternative polyadenylation).

The poly (A) tail is required for nuclear export, translation, and mRNA stability. The tail shortens over time, and the mRNA is enzymatically destroyed when it is short enough. mRNAs with short poly(A) tails are retained in a few cell types for subsequent activation by re-polyadenylation in the cytoplasm. Polyadenylation, on the other hand, accelerates RNA breakdown in bacteria. This is also true for eukaryotic non-coding RNAs on occasionally. Both prokaryotes and eukaryotes have polyadenylated 3′-ends on their mRNA molecules, with prokaryotic poly (A) tails being shorter and fewer mRNA molecules polyadenylated.

3.5 SPLICING AND EDITING

Exons, which correspond to protein-coding regions (ex-on specifies that they are expressed), and introns, which may be involved in gene regulation but are eliminated from pre-mRNA during processing, make up eukaryotic genes. Functional proteins are not encoded by intron regions in mRNA.
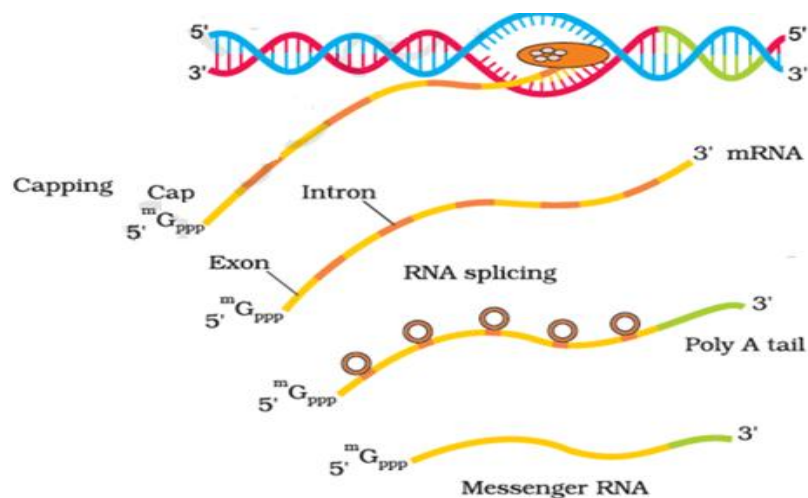

Fig.3.2 Splicing of hnRNA (source-https://goprep.co/a-describe-the-process-of-transcription-in-bacteria-b-i-1nka7w)

Before protein synthesis, all of a pre-introns mRNA's must be eliminated entirely and accurately. The reading frame of the rejoined exons will shift if the process is off by even a single nucleotide, and the resulting protein will be defective. Splicing is the process of eliminating introns and reconnecting exons.

RNA editing was identified by Benne and colleagues, as a sort of RNA processing, in a mitochondrion-encoded mRNA of a kinetoplastid trypanosome. The term RNA editing was first used to describe a process that happens in trypanosomes and involves the insertion and deletion of uridylate (UMP) inside nascent transcripts after transcription. The term was expanded once more examples of such post-transcriptional sequence changes were discovered. The insertion and deletion of nucleotides other than UMP, base deamination, and cotranscriptional insertion of none coded nucleotides are now referred to as RNA editing. RNA editing has been found in mRNAs, tRNAs, and rRNAs, as well as mitochondrial and chloroplast-encoded RNAs and nuclear RNAs. Many metazoa, unicellar eukaryotes including trypanosomes, and plants have been found to have RNA editing. RNA editing has never been seen in a prokaryote before.

RNA editing reactions are usually divided into two categories depending on their response mechanisms i.e. Substitution editing and insertion/deletion editing. Insertion/deletion RNA editing, for example, involves inserting or deleting nucleotides to change the length of the target RNA. RNA editing via base modification, on the other hand, turns one encoded nucleotide into a new nucleotide without modifying the RNA's overall length. Both types of RNA editing result in the formation of transcripts with a different sequence than the genome template. Genetic recoding occurs when the mature transcript and the encoding genome have different RNA sequences. The sequence alterations caused by RNA editing are distinct from those caused by other RNA processing activities such as 5′-capping, 3′-polyadenylation, and splicing that occur during mRNA synthesis in eukaryotes. In eukaryotic organisms and their viruses, RNA editing is common. Editing, like splicing, is a type of processing that can increase genetic diversity and change the function of gene products by altering the information transfer mechanism at the posttranscriptional level.

3.6 NUCLEAR EXPORT OF mRNA

Transport of messenger RNA (mRNA) from the nucleus to the cytoplasm is a necessary step of eukaryotic gene expression. In the cell nucleus, a precursor mRNA undergoes a sequence of

processing steps, together with capping at the 5' ends, splicing and cleavage/polyadenylation at the 3' ends. During this process, the mRNA associates with a broad range of proteins, forming a messenger ribonucleoprotein (mRNP) element. Association with factors involved in nuclear export too occurs through transcription and processing, and thus nuclear export is completely integrated into mRNA maturation. The coupling between mRNA maturation and nuclear export is a significant mechanism for providing only fully functional and competent mRNA to the cytoplasmic translational machinery, thus ensuring exactness and quickness of gene expression. The transport of mRNA transcripts from the nucleus's site of synthesis to the cytoplasm, where they can be translated into proteins, is important for regulating eukaryotic gene expression. The nuclear export of mRNA transcripts can be broken down into distinct stages: first, pre-mRNA is transcribed in the nucleus, where it is processed and packaged into messenger ribonucleoprotein (mRNP) complexes; second, the mRNPs are targeted to and translocate through nuclear pore complexes (NPCs) that are embedded in the nuclear envelope; and third, the mRNPs are directionally released into the cytoplasm for translation. Each of these processes involves a lot of mechanistic interaction. Furthermore, changes in components required for mRNA nuclear export have been linked to a variety of diseases.
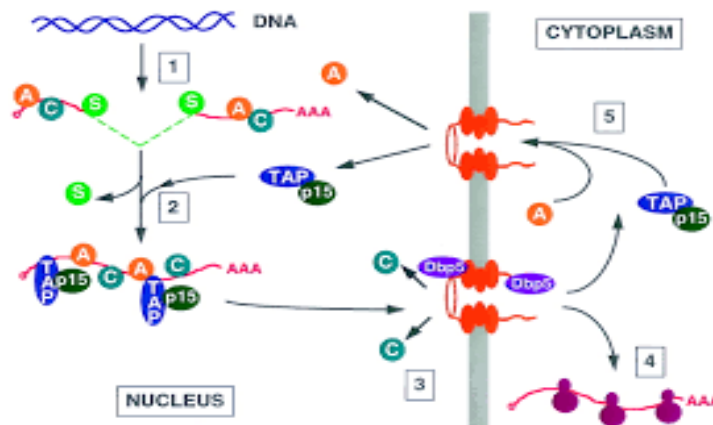

Fig.3.3 Nuclear RNA export pathways

(Source:https://www.google.com/search?q=nuclear+transport+of+mRNA+diagram&tbm)

mRNAs are linked to a wide range of proteins, including adaptor proteins. The presence of a diverse assortment of adaptor proteins within a single mRNP may boost the likelihood of the transport receptor recognizing it. Multiple copies of transport receptors may be beneficial for efficient transport of large mRNPs, as it has been indicated for ribosomal particles. Alternatively, these adaptor proteins may allow a single transport receptor to export multiple mRNPs from the nucleus. It's also possiblethat distinct adaptors work in a sequential manner during mRNP development.

3.7 SUMMARY

Post-transcriptional modifications (PTMs) are process that help the generation of mature, functional **RNA.** These mechanisms permit different proteins to be formed from one gene and proceed as regulators of the phenotype and proliferation rate. The pre-messenger RNA (mRNA), called heterogeneous nuclear RNA (hnRNA), is adapted by adding a 5' 7-methylguanosine cap and a 3' poly-A (polyadenylate) tail for stability and safety. Alternative splicing, on the other hand, moreover excludes the introns, except varying combinations of exons are linked, producing different proteins from the original mRNA. In RNA editing, the mRNA sequence is changed and differs from the transcribed DNA template.

Post-transcriptional modifications can also occur throughout the processing of other transcripts which eventually become transfer RNA, ribosomal RNA, or any of the extra types of RNA used by the cell.

The five-prime cap (5′ cap) is a particularly altered nucleotide on the 5′ end of a few primary transcripts such as precursor messenger RNA. In eukaryotes, the 5′ cap (cap-0), begin on the 5′ end of an mRNA molecule and consist of a guanine nucleotide connected to mRNA via an abnormal 5′ to 5′ triphosphate linkage

Polyadenylation is the adding of a poly (A) tail to an RNA transcript, naturally a messenger RNA (mRNA). The poly (A) tail consists of several adenosine monophosphates. In eukaryotes, polyadenylation is part of the procedure that produces mature mRNA for translation. mRNA molecules in prokaryotes and eukaryotes both have polyadenylated 3′-ends, with the prokaryotic poly(A) tails usually shorter and less mRNA molecules polyadenylated.

Eukaryotic genes are collection of **exons** and *dominant* sequences called **introns** which are detached from the pre-mRNA during processing. Intron sequences in mRNA do not encode useful proteins. All of a pre-mRNA's introns must be totally and specifically detached before protein synthesis. If the procedure errs by even a single nucleotide, the reading structure of the rejoined exons would shift, and the consequential protein would be dysfunctional. The procedure of removing introns and reconnecting exons is called **splicing**.

RNA editing is a kind of RNA processing, which involves the post-transcriptional insertion and deletion of uridylate (UMP) inside nascent transcripts.

Typically, RNA editing reactions are put into two broad categories based on their reaction mechanisms. One type, insertion/deletion RNA editing, involves the insertion or deletion of nucleotides and actually changes the length of the target RNA. The second type, RNA editing through base modification, changes an encoded nucleotide into a dissimilar nucleotide, with no varying the overall length of the RNA. Transport of messenger RNA (mRNA) from the nucleus to the cytoplasm is a necessary step of eukaryotic gene expression.

3.8 TERMINAL QUESTION & ANSWERS

3.8.1. Multiple Choice Questions:

1. Which nucleotide is present in the 5' cap?

a) ADP

b) CDP

c) GDP

d) UDP

2. Which is the first enzyme in capping?

a) Guanylyltransferase

b) Guanul transferase

c) N7G methytransferase

d) RNA 5' triphosphatase

3. mRNA of which of the following organism does not undergo processing?

a) Human

b) Yeast

c) Fungi

d) Bacteria

4. Which of the following is not a type of RNA processing?

a) Polyadenylation at the 3' end

b) Capping of 5' end

c) Removal of exons

d) Splicing

5. The first RNA processing event is _____

a) Capping

b) Tailing

c) Splicing

d) Editing

Answers: 1 c, 2 a, 3 d, 4 c, 5 a.

3.8.2. Short Answer Question:

1. What is Post-transcriptional modification?

2. Write about 5'- cap formation.

3. Write short notes on: i) **splicing ii)** RNA editing

4. What is Polyadenylation?

5. Write the function of 5′cap.

3.8.3. Long Answer Question:

1. Describe the Post- transcriptional modification.

2. Explain Capping process.

3. Describe the nuclear export of mRNA.

4. Explain the 5'-cap formation.

5. Describe the **splicing and** editing.

GLOSSARY

Adenosine triphosphate (ATP): The major source of usable energy in cell metabolism composed of adenine, ribose and three phosphate groups. On hydrolysis, ATP loses one phosphate and one hydrogen to become adenosine diphosphate (ADP), releasing energy in the process.

Deoxyribonucleic acid (DNA): The molecule that carries the genetic information for most living systems. DNA molecule consists of four bases (adenine, cytosine, guanine, and thymine) and a sugar-phosphate backbone, arranged in two connected strands to form a double helix.

Enzyme: A large biological molecule, typically a protein that catalyzes a chemical reaction.

Exon: Any segment of a gene that is represented in the mature messenger RNA (m RNA) product.

Gene: A segment of chromosome that encodes the necessary regulatory and sequence information to direct the synthesis of a protein or RNA product.

Gene Expression: The process through which a gene is activated at particular time and place so the its functional product is produced.

Messenger RNA (mRNA): Nucleic acid that carries instructions to a ribosome for the synthesis of a particular protein.

Nucleic acid: Large molecules, generally found in the cell's nucleus and/or cytoplasm that are made up of nucleotide bases. The two kinds of nucleic acid are DNA and RNA.

Nucleotides: The building blocks of nucleic acids. Each nucleotide is composed of sugar, phosphate, and one of four nitrogen bases. If the sugar is ribose, the nucleotide is termed a "ribonucleotide," whereas deoxyribonucleotides have deoxyribose as the sugar component (i. e. adenine, cytosine, guanine and thymine in the case of DNA). The sequence of the nucleotides within the nucleic acid determines, for example, the amino acid sequence of an encoded protein.

Promoter: A DNA sequence that is located near or even partially within encoding nucleotide sequences and which controls gene expression. Promoters are required for binding of RNA polymerase to initiate transcription.

Promoter site: A region in the operon between the operator and the structural genes, RNA polymerase attaches to the promoter.

Ribonucleic acid (RNA): A single stranded nucleic acid molecule comprising a linear chain made up from four nucleotide subunits (A, C, G, and U). There are three types of RNA: from four nucleotide subunits (A, C, G, and U). There are three types of RNA: messenger, transfer and ribosomal. (Actually there are also ribosomes etc.)

Splicing: The removal of introns and joining of exons to form a continuous coding sequence in RNA.

Template: A molecule that serves as the pattern for synthesizing another molecule.

Transfer RNA (t RNA): RNA molecules that carry amino acids to sites on ribosomes where proteins are synthesized.

3.9 REFERENCES

1. De Robertis, E.D.P and E.M.F. De Robertis. Jr. (1987). Cell Biology 8th Edn. B. I. Waverly Pvt. Ltd., New Delhi

2. Gupta, P. K,.Cell and Molecular Biology 1st Ed. Rastogi Publications., Meerut

3. Robert F.Weaver. Molecular Biology, 1999.WCB-McGraw Hill, Boston.

4. https://en.wikipedia.org/wiki/Post-transcriptional modification

5. https://www.sciencedirect.com/topics/neuroscience/rna-editing

6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4488659/

UNIT 4 TRANSLATION

CONTENTS

4.1 OBJECTIVES

In this unit students will learn:

- Details of the Genetic Code
- The mechanism of Prokaryotic and eukaryotic translation and its regulation.
- About the Post-translational modification of proteins

4.2 INTRODUCTION

Translation is a process of formation of protein from messenger RNA (mRNA).The mRNA consists of a series of codons which provides the information for the amino acid sequence of a protein molecule. Translation is highly regulated and involves several enzymes and molecules to synthesize a protein. Codons of mRNA interact with anticodon of aminoacyl-tRNAs with the help of a ribosome to synthesize a specific sequence of polypeptide chains. Transfer RNA is a RNA molecule (76 to 90 nucleotides) that serve as a link between mRNA and amino acid sequence of protein (Fig 4.1).
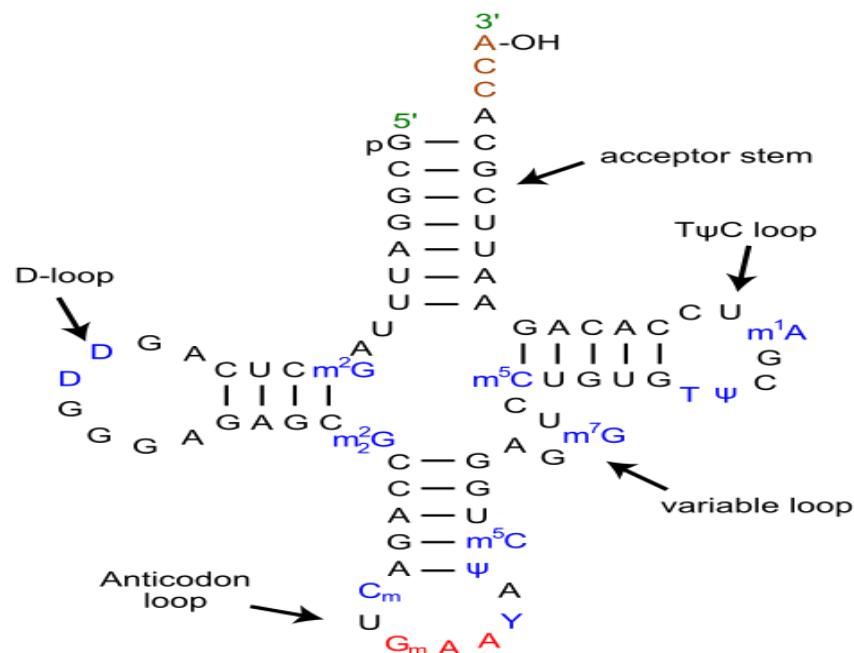

Fig.4.1 tRNA structure (credit: https://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_en.svg)

The ribosome consists of two subunits that play a specific role during translation (Table1). Ribosomes are made up of RNA (rRNA) and protein (known as *r-protein*). Ribosomes are conserved during evolution with little variations in the overall size and proportions of RNA and protein in the ribosomes of bacteria, eukaryotic cytoplasm and organelles. The ribosomal RNA and protein interact with each other to form several active sites which involve protein synthesis.

Ribosome	large Subunit	Small Subunit
Prokaryotes (E.coli)	50S 33 Proteins 23S rRNA 5S rRNA	30S 21 Proteins 16S rRNA
Eukaryotes (S. Cerevisiae)	60S 46 Proteins 5.8S rRNA 25S rRNA 5S rRNA	40S rRNA 33 Proteins 18S rRNA

Table 4.1 Prokaryotic and Eukaryotic Ribosomes

4.3 *GENETIC CODE*

DNA is a genetic material and encodes for the mRNA during the process of transcription. mRNA then encodes the polypeptide chain (protein) during the process of translation. The relationship between a sequence of DNA and the sequence of the corresponding polypeptide is known as genetic code. The genetic code is read in groups of three nucleotides (triplet) called codon. A codon codes for one amino acid in a polypeptide chain. The series of codons that read in a particular direction, starting from one point to terminate at another is called a gene. RNA molecule is made up of four type of nucleotide: Adenine (A), guanine (G), cytosine (C) and uracil (U) and combination of three nucleotide lead to formation of a codon therefore there are

64 possible codons (table). 61 codon codes for a particular amino acid and remaining 3 codon doesn't code for any amino acid but helps in termination of protein synthesis. AUG is a starting codon and code for amino acid methionine. UAA, UAG and UGA are known as stop codons and signal the termination of polypeptide chain. Mutation that changes a normal codon into a stop codon is known as nonsense mutation. Mutation that results in addition or deletion of individual bases causing shift in the triplet sets after the site of mutation is called frameshift mutation. Depending on the starting point, nucleotide sequences have three possible ways for translation and these are called reading frame. A reading frame that consists exclusively of triplets encoding amino acids is called an Open reading frame (ORF).

Table 4.2 Genetic code (Credit: https://commons.wikimedia.org/wiki/File:Genetic_Code.png)

Characteristics of genetic code

1) The genetic code is almost universal

2) The genetic code is non-overlapping means each base is part of one and only one codon.

3) The genetic code is comma less, meaning there are no intervening bases between adjacent codons.

4) The genetic code is highly degenerate, meaning most amino acids are specific by two or more codons except methionine and tryptophan. Example: CCU,CCC,CCA and CCG code for proline.

5) The genetic code is unambiguous, meaning each codon specific only one amino acid.

Various different codons specify one amino acid and the difference between them generally lies at the third base position (at the 3 end). For example, alanine has GCU, GCC, GCA, and GCG codons. The codons for most amino acids can be symbolized by XYA/G or XYU/C. The first two letters of each codon determine the specificity. So, Crick concluded that the third base of most codons pairs rather loosely with the corresponding base of its anticodon. The third base of such codons is called "wobbles." Crick proposed a set of four relationships called the wobble hypothesis:

 1. The first two bases of an mRNA codon always form strong Watson-Crick base pairs with the corresponding bases of the tRNA anticodon and confer most of the coding specificity.

2. The first base of the anticodon (reading in the 5→3 direction; this pairs with the third base of the codon) determines the number of codons recognized by the tRNA. When the first base of the anticodon is C or A, base pairing is specific and only one codon is recognized by that tRNA. When the first base is U or G, binding is less specific and two different codons may be read. When inosine (I) is the first (wobble) nucleotide of an anticodon, three different codons can be recognized—the maximum number for any tRNA.

3. When an amino acid is specified by several different codons, the codons that differ in either of the first two bases require different tRNAs.

4. A minimum of 32 tRNAs are required to translate all 61 codons (31 to encode the amino acids and 1 for initiation).

The wobble (or third) base of the codon confers specificity, but, as it pairs loosely with its corresponding base in the anticodon, it allows rapid dissociation of the tRNA from its codon during protein synthesis. If there is strong base pairing between all three bases of a codon with the three bases of the anticodon, tRNAs would dissociate slowly and this would limit the rate of protein synthesis. Codon-anticodon interactions balance the requirements for accuracy and speed.

4.4 *PROKARYOTIC AND EUKARYOTIC TRANSLATION*

Translation in prokaryotes as well as in eukaryotes is divided into three stages.

Prokaryotic Initiation: The formation of polypeptide starts from a start codon which is AUG (but in bacteria GUG or UUG is also used). mRNA binds to the 30s subunit of ribosome with the help of a special sequence called ribosome binding site (Shine-Dalgarno sequence). This is a small sequence in mRNA (~30 bases long 5`……AGGAGG……3`) which is complementary with the 16s rRNA sequence. The ribosomes in the free pool can dissociate into subunits; this means 70S ribosomes are in dynamic equilibrium with 30S and 50S subunits. The 30S binds with initiation factor IF -3 (protein) stabilizes it and prevents its binding with the 50S subunit. IF-3 also helps 30S subunit to bind with initiation site of mRNA and as 30S-mRNA part it checks the accuracy of recognition of the first aminoacyl-tRNA. IF-1 binds with 30S subunit in the vicinity of A site and prevents aminoacyl-tRNA from entering. The initiator tRNA in bacteria, mitochondria and chloroplast code for formylated methionine called as N-formyl-methionyltRNA (tRNAfMet). The methionine is modified by two stage reactions.

$$Methionine + tRNAfMet + ATP \longrightarrow Met\text{-}tRNAfMet + AMP + PPi$$

$$N10\text{-}Formyltetrahydrofolate + Met\text{-}tRNAfMet \longrightarrow tetrahydrofolate + fMet\text{-}tRNAfMe$$
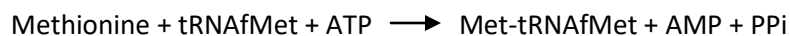
The aminoacyl-tRNAsynthetase is Mg^{++} dependent activating enzyme and the reaction involve use of ATP. tRNA binds with AUG codon by base pairing with it using an anticodon region which is complementary with it. IF-2 binds with the initiator tRNA and controls its entry into the ribosome (partial P site in 30S subunit). IF-2 has a ribosome dependent GTPase activity. The GTP hydrolysis causes change in the conformation leading to association of 50S subunit and

initiator tRNA to 30S form complete ribosome. All the initiation factors are released. The ribosome has three tRNA binding sites. The entry site of aminoacyl-tRNA is A site. The binding site of peptidyl-tRNA is the P site and the exiting site of deacylatedtRNA is the E site. Initiation is a slow step and usually determines the rate of translation.

Eukaryotic Initiation: Eukaryotic initiation also involves the use of start codon and initiator tRNA but methionine is not formylated. Eukaryotic initiation required a large number of initiation factors as compared to prokaryotes. First eIF2, tRNA, eIF3, eIF1 and eIF1A bind with 40S subunit of ribosome to form 43S preinitiation complex. eIF4A, eIF4B, eIF4E and eIF4G bind to 5` end of the mRNA to form the cap-binding complex. This complex binds with 3` end of mRNA by eIF4G which in turn binds with poly(A)-binding protein (PABP). The 43S complex associates with initiation factors at the 5` end mRNA and scans the start codon. Form a 48S complex at initiation codon. eIF2 is a GTP binding protein and eIF2-GTP binds with Met-tRNA called ternary complex. Ternary complex assembly is regulated by the guanine nucleotide exchange factor (GEF) eIF2B (exchange GDP for GTP). eIF5 induces GTP hydrolysis by eIF2 and the eIF2 and eIF3 are released (eIF3 maintain free 40s subunit). eIF5B mediates the association of 60S unit to from complete ribosome.
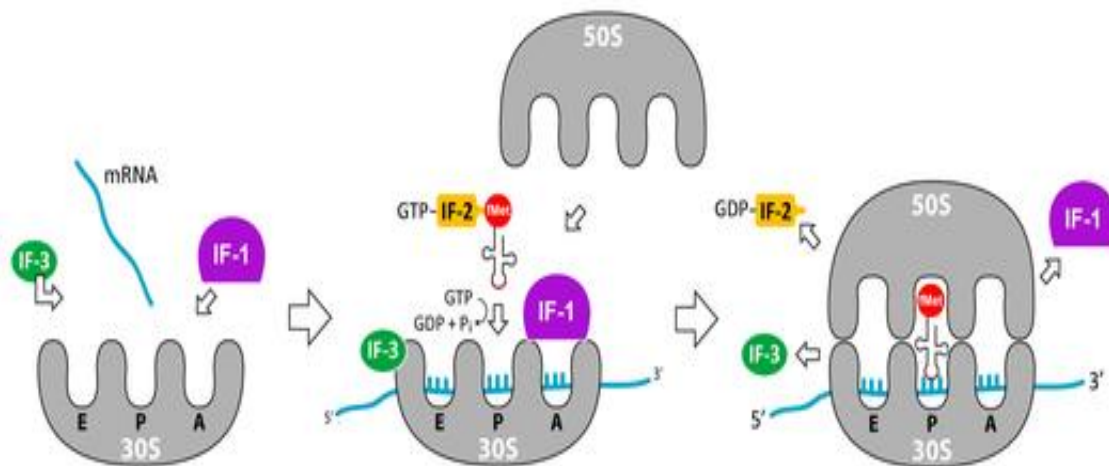

Fig.4.2 Prokaryotic translation initiation (Credit:
https://commons.wikimedia.org/wiki/File:Initiation_der_Translation.png)

Prokaryotic Elongation

Elongation step is assisted by an elongation factor (EF-Tu) in bacteria and its eukaryotic counterpart is also similar as it is highly conserved during evolution. After the formation of a complete ribosome a cyclic process starts in which the aminoacyl-tRNA enters the A site of the ribosome whose P site is already occupied by peptidyl-t RNA. EF-Tu binds with the ribosome only during the process of aminoacyl-tRNA entry. After aminoacyl-tRNA is placed, EF-Tu leaves the ribosome to repeat the cycle with another aminoacyl-tRNA to place it in A site. This association and dissociation cycle is a GTP dependent process. EF-Tu is a monomeric GTP binding protein. EF-Tu-GTP binds aminoacyl-tRNA to form a ternary complex which only binds with the A site of ribosome whose P site is already occupied by peptidyl-tRNA. This step is very important to ensure the correct position of aminoacyl-tRNA and peptidyl-tRNA for the formation of peptide bond.

First the anticodon end of aminoacyl-tRNA binds with the codon in the A site of the 30S subunit. Then codon-anticodon recognition leads to conformational change in ribosomes which stabilize the tRNA binding causing GTP hydrolysis by EF-Tu. tRNA CCA end moves into A site on 50S subunit. EF-Tu-GDP is now released and it cannot bind with aminoacyl-tRNA effectively. Conversion of EF-Tu-GDP into EF-Tu-GTP is mediated by guanine nucleotide exchange factor EF-Ts. The hydrolysis of EF-Tu-GTP is a relatively slow process which allows the incorrect aminoacyl-tRNA to dissociate from the A site.

The large subunit of ribosome has peptidyl transferase activity which mediates the elongation of the polypeptide chain by transferring the polypeptide attached to the tRNA in the P Site to the aminoacyl-tRNA in the A site. The reaction is initiated when EF-Tu releases the aminoacyl end of its tRNA, which then comes closer to the end of the peptidyl-tRNA. The processes of addition of amino acid to the growing polypeptide chain completes by the translocation of the ribosome. During the translocation the ribosome advances three nucleotides along the mRNA and expels the uncharged tRNA from the P site to E site allowing the new peptidyl-tRNA to enter. Now A site is available for another aminoacyl-tRNA corresponding to the next codon. Translocation of the ribosome occurs due to the activity of another GTP binding elongation factor EF-G. GTP hydrolysis triggers the change in EF-G which in turn causes change in

ribosome. EF-Tu and EF-G bind with the ribosome alternately. EF-Tu-GTP binds with ribosome and releases as EF-Tu-GDP after GTP hydrolysis. Only after the release of EF-Tu-GDP from the ribosome, only EF-G-GTP is able to binds with ribosome. Hydrolysis of GTP causes release of EF-G from the ribosome, then again EF-Tu-GTP binds with the ribosome.

Eukaryotic Elongation

In eukaryotes the eEF1α factor (similar to EF-Tu) mediates the transfer of aminoacyl-tRNA to the ribosome. The process is similar to prokaryotes involving GTP hydrolysis. eEF1βγ factor (similar to EF-Ts) regenerates the active form eEF1α-GTP from eEF1α-GDP. Translocation of the ribosome is done by the activity of eEF2 (similar to EF-G).
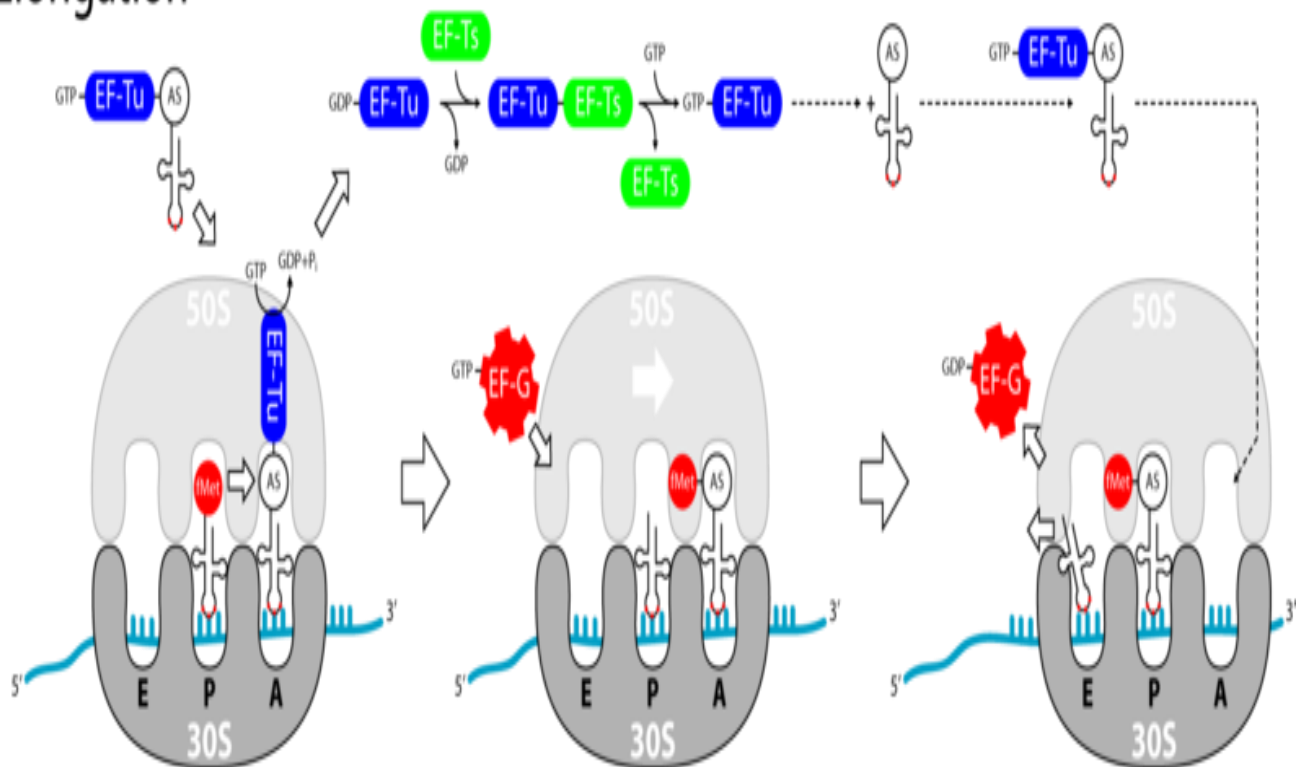
Fig. 4.3 Prokaryotic translation elongation (Credit:
https://commons.wikimedia.org/wiki/File:Elongation_der_Translation.png)

Prokaryotic Termination

Termination of translation required a stop codon and release factors (RF). The stop codon (termination codon) are UAA (ochre), UAG (amber) and UGA (opal). The termination codon is not recognized by tRNA but they recognized by protein factors (RF). Termination codon are recognized by class 1 release factors for e.g. RF1 (recognizes UAA and UAG) and RF2 (recognizes UGA and UAA). Class 2 release factors are not codon specific and they are GTP binding protein. Class 2 release factors structure resemble aminoacyl-tRNA-EF-Tu and EF-G. Class 2 release factors (RF3) help in releasing Class 1 release factors from ribosome. RF1 and RF2 recognize the stop codon which leads to activation of ribosome to hydrolyze the peptidyl tRNA. Hydrolysis of polypeptide is analogous to reaction of peptidyl transfer but in this case, acceptor is water (H_2O) instead of aminoacyl-tRNA. Then RF1 and RF2 are released from ribosome by RF3. Before termination RF3-GDP bind with ribosome and GDP is replaced by GTP which enable it to contact ribosome GTPase centre, where it cause release of RF1 and RF2 when polypeptide chain is terminated. This termination reaction release polypeptide chain but other component (deacylated tRNA, mRNA and ribosome) are still associated. Ribosome recycling factor (RRF) together with EF-G utilize GTP hydrolysis to dissociate remaining component. IF-3 removes deacylated tRNA from 30s subunit and prevents its reassociation with large ribosomal subunit.

Eukaryotic termination

Eukaryotes have a single class 1 release factor (eRF1) which recognizes all the three termination codons. Class 1 release factor (eRF1) can terminate the process of translation without the help of class 2 release factor although yeast use class 2 release factor eRF2.

Termination
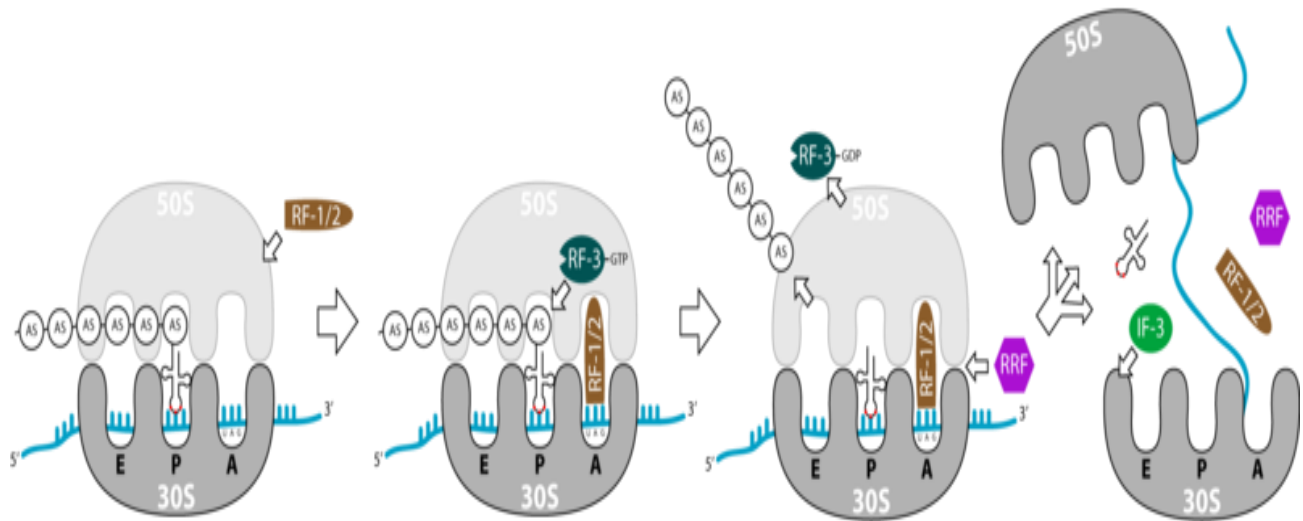
Fig.4.4 Prokaryotic translation termination (Credit:

https://commons.wikimedia.org/wiki/File:Termination_der_Translation.png)

There are various inhibitors of the translation process. Example

1) Puromycin is an antibiotic, synthesized by *Streptomyces alboniger*. Its structure is very similar to the 3 end of an aminoacyl-tRNA so, it binds to the ribosomal A site and participates in peptide bond formation, producing peptidyl-puromycin.

2) Tetracyclines inhibit translation in bacteria by blocking the A site on the ribosome, preventing the binding of aminoacyl-tRNAs.

3) Chloramphenicol blocking peptidyl transfer in bacterial and mitochondrial and chloroplast ribosomes but does not affect cytosolic protein synthesis in eukaryotes.

4) Cycloheximide blocks the peptidyl transferase of only 80S eukaryotic ribosomes

5) Streptomycin leads to misreading of the genetic code (in bacteria) at relatively low concentrations and inhibits initiation at higher concentrations.

6) Diphtheria toxin inactivates eukaryotic elongation factor eEF2 by catalyzing the ADP-ribosylation of a diphthamide (a modified histidine) residue.

7) Ricin toxic protein of the castor bean, it inactivates the 60S subunit of eukaryotic ribosomes by depurinating a specific adenosine in 23S rRNA.

4.5 *REGULATION OF TRANSLATION*

If the untranslated region of mRNA (5` UTR of mRNA) has a sequence that does not allow rapid ribosome binding or movement onto the ORF, then it results in slow translation. Like promoters in transcription mRNA sequences can be good or poor.

Translation process is also controlled by codon usage. There are multiple codons for most of the amino acids and these codons are not decoded equally by tRNAs therefore some have abundant tRNAs and some do not. So, mRNA with codons of abundant tRNA can be rapidly translated compared to mRNA with codons of less abundant tRNAs.

Repressor proteins (for e.g. R17 coat protein, T4 RegA etc) can regulate the process of translation by inhibiting a ribosome from binding to an initiation codon.

The secondary structure of mRNA in a polycistronic mRNA also regulates translation. In the case of RNA phage whose genes are always expressed in a set order. Phage RNA takes up a secondary structure in which only one initiation sequence is accessible. The second sequence cannot be recognized by the ribosome because it base pair with RNA in the other region. Then the translation of the first gene disrupts the secondary structure, enabling binding of the ribosome to the initiation site of the next gene.

4.6 *POST-TRANSLATIONAL MODIFICATIONS OF PROTEIN*

The newly synthesis polypeptide chain passes through a peptide exit tunnel that extends from the peptide transferase centre to the ribosome surface. This exit tunnel allows only α-helical conformation of nascent polypeptide because it is too narrow for more extensive folding. When nascent polypeptide exits from the tunnel, various enzymes interact with it that leads to co-translational modification, its proper folding is assisted by chaperones and its proper transport is facilitated by signal recognition particle. As the nascent polypeptide emerges from the exit tunnel of bacterial enzyme peptide deformylase cleave the N-terminal formyl group and then another

enzyme methionine aminopeptidase removes N-terminal methionine. In eukaryotic polypeptide N-terminal formyl group is absent but it does begin with methionine. So, here also ribosome-bound enzyme methionine aminopeptidases cleave the N-terminal methionine.

In prokaryotes co-translational protein folding involves a chaperone (48 kDa) known as trigger factor. It is present at bacterial ribosomes exit pore. Trigger factor is a dragon shaped structure formed by three domains. The N-terminal domain forms the tail which associates with the 50S subunit of ribosome. The C-terminal domain forms the middle part of the dragon and is involved in trigger chaperon activity. The middle domain forms head which has peptidyl propyl cis/trans isomerase activity. Eukaryotes also have chaperons associated with large subunit of ribosome.

In several cases individual amino acid is also modified like hydroxyl group Ser, Thr and Tyr in some protein is phosphorylated by ATP using enzyme for e.g. milk protein casein. In some protein extra carboxyl group is added to Glu residues for e.g. blood clotting protein prothrombin. Proteins are modified by attaching carbohydrate side chain (glycoprotein), by addition of Isoprenyl group and also by addition of prosthetic group. In some case protein becomes functional only after cleavage ex. Proinsulin, Chymotrypsinogen and Trypsinogen. Some proteins are modified by formation of disulfide cross bond between Cys residues.

The newly synthesized protein is transported to endoplasmic reticulum lumen or integrated into the membrane. Blobel and Sabatini proposed the signal hypothesis to explain this transport process. This transport is very specific and in a mammalian cell is mediated by single sequence, signal recognition particle (SRP), SRP receptor and translocon. Signal sequence is characterized by a positively charged short N-terminal, a center hydrophobic stretch of 7to 13 amino acid and a more polar C-terminal. This signal is recognized by SRP (made up of six proteins) and then it binds with the receptor SR (contains two GTP-binding protein subunit). After binding with the receptor, the signal peptide is transferred to the translocon from SRP.

4.7 *SUMMARY*

Translation is a process of formation of protein from messenger RNA (mRNA). An mRNA consists of a series of codons which provides the information for the amino acid sequence of a protein molecule. Translation is highly regulated and involves several enzymes and molecules to

synthesize a protein. Codons of mRNA interact with anticodon of aminoacyl-tRNAs with the help of a ribosome to synthesize a specific sequence of polypeptide chains. DNA is a genetic material and encodes for the mRNA during the process of transcription. mRNA then encodes the polypeptide chain (protein) during the process of translation. The relationship between a sequence of DNA and the sequence of the corresponding polypeptide is known as genetic code. The genetic code is read in groups of three nucleotides (triplet) called codon. A codon code for one amino acid in a polypeptide chain. Translation in prokaryotes as well as in eukaryotes is divided into three stages. The formation of polypeptide starts from a start codon which is AUG (but in bacteria GUG or UUG is also used). mRNA binds to the 30s subunit of ribosome with the help of a special sequence called ribosome binding site (Shine-Dalgarno sequence). Elongation step is assisted by an elongation factor (EF-Tu) in bacteria and its eukaryotic counterpart is also similar as it is highly conserved during evolution. Termination of translation required a stop codon and release factors (RF). The stop codon (termination codon) are UAA (ochre), UAG (amber) and UGA (opal). The termination codons are not recognized by tRNA but they are recognized by protein factors (RF). The process of translation is highly regulated by various regulatory mechanisms. When nascent polypeptide exits from the tunnel, various enzymes interact with it that leads to co-translational modification, its proper folding is assisted by chaperones and its proper transport is facilitated by signal recognition particles. As the nascent polypeptide emerges from the exit tunnel of bacteria enzyme peptide deformylase cleave the N-terminal formyl group and then another enzyme methionine aminopeptidase removes N-terminal methionine. In eukaryotic polypeptide N-terminal formyl group is absent but it does begin with methionine. So, here also ribosome-bound enzyme methionine aminopeptidases cleave the N-terminal methionine.

4.8 *TERMINAL QUESTION & ANSWERS*

1. Translation is a process of

A) Formation of Protein from DNA

B) Formation of Protein from mRNA

C) Formation of DNA

D) None of the above

2. Genetic Code is

A) Gene in DNA

B) Three nucleotides (triplet) code called codon

C) Protein structure

D) Four nucleotide code

3. Translation takes place in

A) Nucleus

B) vacuoles

C) Cytoplasm

D) lysosome

4. Ribosome are involved in

A) DNA synthesis

B) Protein synthesis

C) RNA synthesis

D) Carbohydrate synthesis

5. Enzyme involved in amino acid activation is

A) Aminoacyl-tRNAsynthetases

B) Aminoacyl-mRNA synthetases

C) Aminoacyl-DNA synthetases

D) None of the above

6. Adaptor molecule during translation is

A) tRNA

B) mRNA

C) DNA

D) All of the above

7. First amino acid coded by prokaryotic

A) N- formylated methionine

B) N- methylated methionine

C) Methionine

D) None of the above

8. In prokaryotes the Ribosomal bind site on RNA is known as

A) Shine-DNA sequence

B) Shine dalgarno sequence

C) Shine-RNA sequence

D) None of the above

9. Eukaryotic initiation required

A) ATP

B) GTP

C) Both A and B

D) None of the above

10. Termination of translation required a

A) Stop codon and release factors

B) Start codon and release factors

C) Stop codon and termination factors

D) Start codon and stop codon

Answers 1(B), 2(B), 3(C), 4(B), 5(A), 6(A), 7(A), 8(B), 9(B), 10(A)

Q11. Describe the process of Prokaryotic translation.

Q12. Describe the process of Eukaryotic translation.

Q13 Write a Short note on

A) Regulation of Translation
B) Post translational modification

4.9 *GLOSSARY*

Aminoacyl-tRNA: An aminoacyl ester of a tRNA.

Aminoacyl-tRNA synthetases: Enzymes that catalyze synthesis of an aminoacyl-tRNA at the expense of ATP energy.

Amino-terminal residue: The only amino acid residue in a polypeptide chain with a free α-amino group; defines the amino terminus of the polypeptide**.**

Antibiotic: One of many different organic compounds that are formed and secreted by various species of microorganisms and plants, are toxic to other species, and presumably have a defensive function.

Binding site: The crevice or pocket on a protein in which a ligand binds.

Codon: A sequence of three adjacent nucleotides in a nucleic acid that codes for a specific amino acid.

Enzyme: A biomolecule, either protein or RNA, that catalyzes a specific chemical reaction. It does not affect the equilibrium of the catalyzed reaction; it enhances the rate of the reaction by providing a reaction path with a lower activation energy.

Frame shift: A mutation caused by insertion or deletion of one or more paired nucleotides, changing the reading frame of codons during protein synthesis; the polypeptide product has garbled amino acid sequence beginning at the mutated codon.

Gene: A chromosomal segment that codes for a single functional polypeptide chain or RNA molecule.

Genetic code: The set of triplet code words in DNA (or mRNA) coding for the amino acids of proteins.

Initiation codon: AUG (sometimes GUG or, even more rarely, UUG in bacteria and archaea); codes for the first amino acid in a polypeptide sequence: N-formylmethionine in bacteria; methionine in archaea and eukaryotes.

Initiation complex: A complex of a ribosome with an mRNA and the initiating Met-tRNAMet or fMet-tRNAfMet, ready for the elongation steps.

Messenger RNA (mRNA): A class of RNA molecules, each of which is complementary to one strand of DNA; carries the genetic message from the chromosome to the ribosomes.

Nonsense codon: A codon that does not specify an amino acid, but signals the termination of a polypeptide chain.

Nonsense mutation: A mutation that results in the premature termination of a polypeptide chain.

Peptide: Two or more amino acids covalently joined by peptide bonds.

Peptide bond: A substituted amide linkage between the α-amino group of one amino acid and the α-carboxyl group of another, with the elimination of the elements of water

Peptidyl transferase: The enzyme activity that synthesizes the peptide bonds of proteins; a ribozyme, part of the rRNA of the large ribosomal subunit.

Posttranslational modification: Enzymatic processing of a polypeptide chain after translation from its mRNA.

Reading frame: A contiguous, nonoverlapping set of three-nucleotide codons in DNA or RNA.

Shine-Dalgarno sequence: A sequence in an mRNA that is required for binding bacterial ribosomes.

Translation: The process in which the genetic information present in an mRNA molecule specifies the sequence of amino acids during protein synthesis.

4.10 REFERENCES

1)   Molecular Biology by Burton E.Tropp, Sudbury, Mass, Jones and Bartlett Publishers

2) Lewin's Gene XI by Jocelyn E. Krebs, Elliott S. Goldstein, Stephen T. Kilpatrick, Jones & Bartlett Learning

3) Lehninger Principles of Biochemistry by David L. Nelson, Michael M. Cox Fifth edition, W. H. Freeman and Company

UNIT 5 RECOMBINANT DNA TECHNOLOGY

CONTENTS

5.1 OBJECTIVES

After studying this Unit, students will know about:

- The gene cloning and steps involved in it.
- Restriction enzymes, ligases, linkers and adaptors.
- Selection of recombinants, hybridisation techniques, gene probes, DNA fingerprinting, molecular markers used in genome analysis, genome library and blotting techniques used in recombinant DNA technology.

5.2 INTRODUCTION

Recombinant DNA Technology is a genetic engineering technology of joining together distinct DNA molecules to produce new genetic combinations as "recombinant DNA" (rDNA). In recombinant DNA technology, molecules of DNA from two different species are inserted into a host organism to produce new genetic combinations that are of value to science, medicine, agriculture and industry. A series of procedures are used to join together (recombine) DNA segments from two or more different molecules. Under certain conditions, a recombinant DNA molecule enters the cell and replicates autonomously inside the host and integrates with the host chromosome. Recombinant DNA technology was first developed in 1977 by Herbert Boyer, Paul Berg and Stanley N. Cohen, when they successfully expressed somatostatin in bacteria.

There are five steps involved in recombinant DNA technology:

The complete process of recombinant DNA technology includes multiple steps, maintained in a specific sequence to generate the desired product.

Step1.Isolation of Genetic Material

The first and the initial step in Recombinant DNA technology is to isolate the desired DNA in its pure form i.e. free from other macromolecules.

Step2.Cutting the gene at the recognition sites

The restriction enzymes play a major role in determining the location at which the desired gene is inserted into the vector genome. These reactions are called 'restriction enzyme digestion'.

Step3.Amplifying the gene copies through Polymerase chain reaction (PCR)

It is a process to amplify a single copy of DNA into thousands to millions of copies once the proper gene of interest has been cut using the restriction enzymes.

Step4.Ligation of DNA Molecules

This step of involves Ligation (joining) of the two pieces –a cut fragment of DNA and the vector together with the help of the enzyme DNA ligase

Step5. Insertion of Recombinant DNA into Host

In this step, the recombinant DNA is introduced into a recipient host cell. This process is termed as Transformation. Once inserted into the host cell, the recombinant DNA gets multiplied and expressed in the form of the manufactured protein under optimal conditions.

5.3 GENE CLONING- THE BASIC STEPS

"DNA cloning, gene cloning or molecular cloning is a technique used to make multiple identical copies of a particular segment of DNA or gene." A clone is a cluster of individual entities or cells that are descended from one progenitor. Clones are genetically identical as the cell simply replicates producing identical daughter cells every time. Scientists are able to generate multiple copies of a single fragment of DNA, a gene which can be used to create identical copies constituting a DNA clone.

Cloning takes place through the insertion of DNA fragments into a tiny DNA molecule. This molecule is made to replicate within the living cell, for instance, a bacterium. The tiny replicating molecule is known as the carrier of the DNA vector. Yeast cells, viruses, Plasmids are the most commonly used vectors. Plasmids are the circular DNA molecules that are introduced from bacteria. They are not part of the main cellular genome. It carries genes, which provide the host cell with beneficial properties such as mating ability, drug resistance. They can be

conveniently manipulated as they are small enough and they are capable of carrying extra DNA which is weaved into them.

Steps involved in gene cloning:

1. A gene segment is cloned and inserted into a circular DNA molecule i.e. plasmid vector to produce an r DNA.

2. The vector carries the gene segment into the host cell, mostly bacterium.

3. Inside the host cell, the vector multiplies, producing numerous identical copies of the DNA of the vector as well as the gene segment inserted.

4. When the host multiplies the r DNA passed to the daughter cells.

5. After many cell divisions, colonies of identical host cells are produced. Each cell contains one or more copies of r DNA molecules.
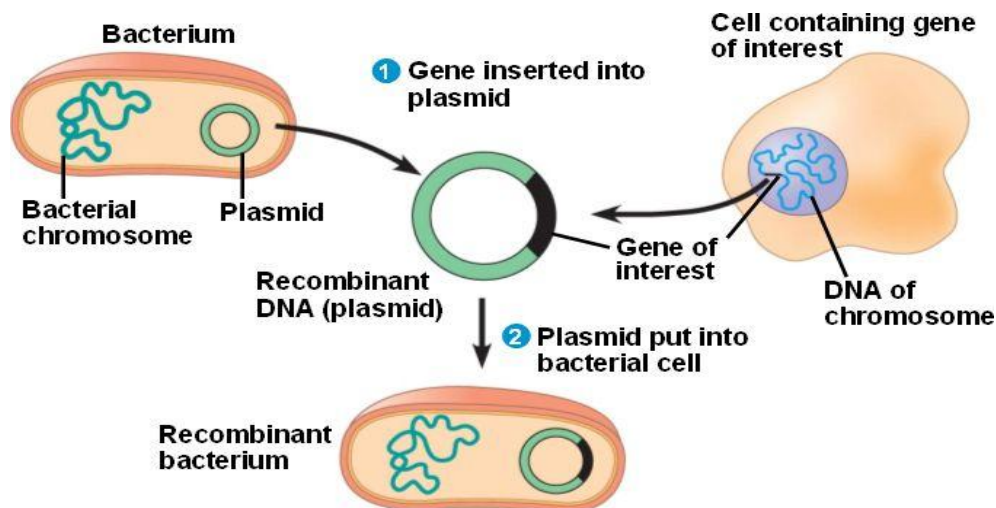

Figure 5.1: Gene cloning (Creative biolabs.com)

5.4   RESTRICTION   ENZYMES   LIGASE   LINKERS   AND ADAPTORS C DNA–TRANSFORMATION

Restriction enzymes are found in bacteria (and other prokaryotes). They recognize and bind to specific sequences of DNA, called **restriction sites**. A restriction enzyme is a protein that recognizes a specific, short nucleotide sequence and cuts the DNA *only* at that specific site, which is known as restriction site or target sequence.

Bacterium restriction-modification systems have two components- a restriction endonuclease and DNA methylase. Each restriction enzyme recognizes just one or a few restriction sites. When it finds its target sequence, a restriction enzyme will make a double-stranded cut in the DNA molecule. Endonuclease is a nuclease enzyme which cleaves the nucleic acid chain within the DNA, rather than at the ends (exonuclease). The first restriction enzyme was isolated from *E. coli K-12*laboratory strains in 1968 by Meselson and Yuan.

A bacterium is immune to its own restriction enzymes, even if it has the target sequences ordinarily targeted by them. This is because the bacterial restriction sites are highly methylated, making them unrecognizable to the restriction enzyme. More than 400 restriction enzymes have been isolated from the bacteria that manufacture them. In live bacteria, restriction enzymes function to defend the cell against invading viral bacteriophages.

Restriction enzymes are named for the organism from which they were first isolated. For example: **1.***Eco*RI is isolated from the *E. coli* strain RY13.*Eco* refers to the genus and species (1st letter of genus; 1st two letters of specific epithet)R is the strain of *E. coli.* I (Roman numeral) indicate it was the first enzyme of that type isolated from *E. coli* RY13.
2. *BamHI* is isolated from *Bacillus amylolique faciens* strain H.

Linkers are short pieces of double stranded (ds) DNA, of known nucleotide sequence that is synthesized in the test tube i.e. synthetic oligonucleotides. These are blunt ends but contain a restriction site (BamHI). **DNA ligase** can attach linkers to the ends of larger blunt-ended DNA molecules. In a blunt end ligation, this particular reaction can be performed very efficiently because linkers can be made in very large amounts and added into the ligation mixture at a high concentration. More than one linker will attach to each end of the DNA molecule, producing the chain structure. However, digestion with BamHI cleaves the chains at the recognition sequences,

producing a large number of cleaved linkers and the original DNA fragment, now carrying BamHI sticky ends. This modified fragment is ready for ligation into a cloning vector restricted with BamHI.

Adaptors, like linkers, are short synthetic oligonucleotides. But unlike linkers, an adaptor is synthesized so that it already has one sticky end. Using adaptors has one more problem. The sticky ends of individual adaptor molecules could base pair with each other to form dimers. This left the new DNA molecule still blunt-ended. The sticky ends could be recreated by digestion with a restriction endonuclease, but that would defeat the purpose of using adaptors in the first place. To solve this problem we remove the phosphate group from the 5′ position of sticky ends of the adaptors by the help of alkaline phosphatase enzyme. Due to this DNA ligase is now unable to form a phosphodiester bridge between 5′-OH and 3′-OH ends. The result of this manipulation is that, although base pairing always occurs between the sticky ends of adaptor molecules, the association is never stabilized between them by ligation. Adaptors can, therefore, be ligated to a blunt-ended DNA molecule but not to themselves. After the adaptors have been attached, the abnormal 5′-OH terminus is converted to the natural 5′-P form by treatment with the enzyme polynucleotide kinase, producing a sticky-ended fragment that can be inserted into an appropriate vector.

Complementary DNA (cDNA)is a double stranded DNA synthesized in the laboratory from messenger RNA. cDNA is not genomic DNA, because the transcript of genomic RNA has been processed (i.e., it lacks promoters and introns). The enzyme reverse transcriptase is used to synthesize double-stranded cDNA that is a complimentary copy of the mRNA. Using an mRNA molecule as a template, reverse transcriptase synthesizes a single-stranded DNA molecule that can then be used as a template for double-stranded DNA synthesis. cDNA does not need to be cut in order to be cloned because it lacks introns. The addition of linker sequences to the end of this DNA, which contain the restriction site, followed by treatment with a restriction enzyme, produces a cDNA preparation with cohesive ends ready for insertion into a vector. A preparation of cDNA represents the genes that were actively being expressed in a cell, an organ, or a whole organism at the time of harvesting and is called a cDNA library.

5.5 SELECTION OF RECOMBINANTS

The need to identify the cells that contain the desired insert at the appropriate and right orientation and isolate these from those not successfully transformed is of utmost importance to researchers. Modern cloning vectors include selectable markers that allow only cells in which the vector, but not necessarily the insert, has been transformed to grow. Additionally, the cloning vectors may contain color selection markers which provide blue/white screening (via α-factor complementation) on X-gal medium. The resulting colonies are required to confirm that cloning was successful. This may be accomplished by means of PCR, restriction fragment analysis and/or DNA sequencing.

1. Colony immunoassay

Colony hybridization was developed for screening recombinants where upon blotting onto nitrocellulose filters and hybridization with a highly radioactive probe, screening of many thousands of colonies per plate for the presence of a DNA sequence carried by a plasmid and complementary to the probe was achieved.

An immunological approach to screen recombinant clones is possible if the gene of interest encodes a polypeptide for which specific antibodies can be prepared. In one approach, DNA complementary to mRNA is inserted in frame with the coding regions of genes present in *E. coli* plasmids. These results in "fused polypeptides" consisting of the N-terminal region of an *E. coli* polypeptide covalently linked to a sequence encoded by the cloned cDNA segment. The identification of cloned genes by colony immunoassays has not been common and one limitation of all previous colony immunoassays is that each fused polypeptide molecule must simultaneously bind to two different antibody molecules. Typically, the first antibody, immobilized on a solid support such as chemically activated paper, is used to entrap the fused polypeptide at the site of the lysed colony, and a second labelled antibody is then bound to the fused polypeptide and detected by autoradiography. A potential disadvantage of all immunological methods is that only one in six sequences inserted at random into the vector would have the orientation and frame consistent with translation into a recognizable fused polypeptide.

A simple immunoassay has been developed by Reggie and Comeron (1986) for isolation of a particular gene from a clone bank of recombinant plasmids. A clone bank of the DNA is constructed with a plasmid vector in *Escherichia coli* and filtered onto a hydrophobic grid membrane and grown up into individual colonies, and a replica was made onto nitrocellulose paper. The bacterial cells upon lysis are immobilized onto the nitrocellulose paper which is reacted with a rabbit antibody preparation made against the particular antigenic product to detect the recombinant clone which carries the corresponding gene. The bound antibodies can be detected easily by a colorimetric assay using goat anti-rabbit antibodies conjugated to horseradish peroxidase.

2. The blue-white screening

The blue white screening is one of the most common molecular techniques that allow detection of the successful ligation of genes of interest in vectors. The α-Complementation plasmids are among the most commonly used vectors for cloning and sequencing the DNA fragments, as they generally have a good multiple cloning site and an efficient blue-white screening system for identification of recombinants in presence of a histochemical dye, 5-bromo-4-chloro-3-indolyl-β-d-galactoside (X-gal), and binding sites for commercially available primers for direct sequencing of cloned fragments.

The molecular mechanism for blue/white screening is based on a genetic engineering of the lac operon in the *E. coli* as a host cell combined with a subunit complementation achieved with the cloning vector. The lacZ product, a polypeptide of 1029 amino acids, gives rise to the functional enzyme after tetramerization and is easily detected by chromogenic substrates either in cell lysates or directly on fixed cells in situ.
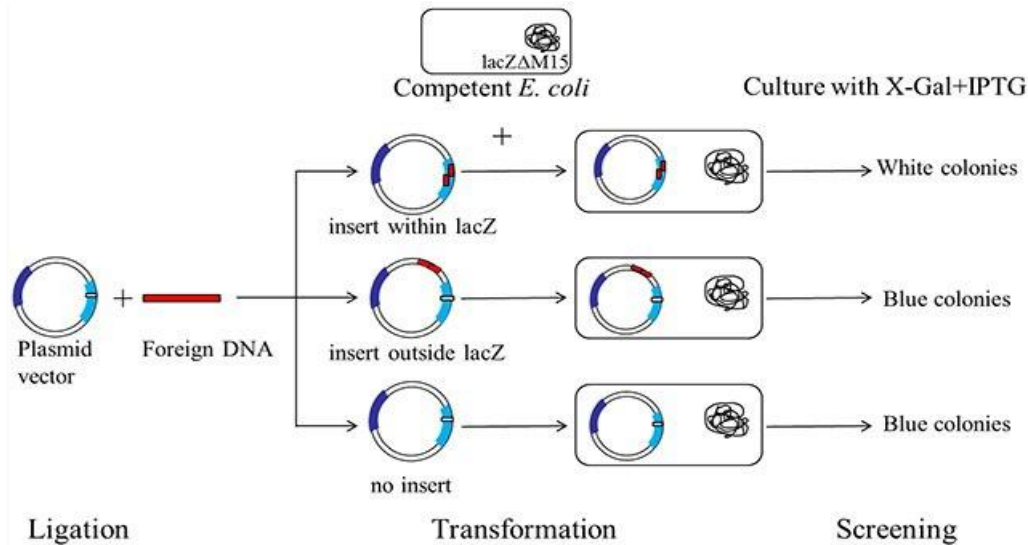
Figure 5.2: Selection of recombinants using blue white screening in Escherichia coli. (Source: proprofs)

The hydrolysis of colorless X-gal by the β-galactosidase causes the characteristic blue colour in the colonies indicating that the colonies contain vectors without insert. White colonies indicate insertion of foreign DNA and loss of the cell's ability to hydrolyse the marker.

Bacterial colonies in general, however, are white, and so bacterial colonies with no vector will also appear white. These are usually suppressed by the presence of an antibiotic in the growth medium. Blue white screening is thus a quick and easy technique that allows for the screening of successful cloning reactions through the colour of the bacterial colony.

3. Reporter gene based screening

Another method for screening and identification of recombinant clones is by using the green fluorescent protein (GFP) obtained from jellyfish *Aequorea victoria*. It is a reporter molecule for monitoring gene expression, protein localization, protein-protein interaction etc. GFP has been expressed in bacteria, yeast, slime mold, plants, drosophila, zebrafish and in mammalian cells. The bacterial cloning vector, pGreenscript A, derived from the mutated Aequorea green fluorescent protein (GFP-S65A) gene, when expressed in *E. coli* produced colonies that showed yellow color under daylight and strong green fluorescence under long-wave ultraviolet light. The vector is used to select for inserted foreign genes based on the loss of the yellow color/green fluorescence of E. coli cells caused by the insertional inactivation of GFP production.

4. Screening clones by positive selection

A variety of positive selection cloning vectors has been developed that allow growth of only those bacterial colonies that carry recombinant plasmids. Typically, these plasmids express a gene product that is lethal for certain bacterial hosts and insertion of any DNA fragment that insertionally inactivates the expression of the toxic gene product, results in growth of colonies.

Positive selection has been a powerful method of screening insert containing transformants. Here the toxic property of the molecule to the host cells is utilized for recombinant selection.

The DNA sequence coding for the toxic product is directly cloned under the promoter elements recognized by the host cells. Positive selection in these vectors is achieved by either inactivation or replacement of toxic genes by the target gene. In general former is much more

convenient than the latter. The advantage of these systems is that no background colonies (non-recombinants) appear on vector alone plates since the relegated vectors carrying toxic intact genes are lethal to the host cells.

5.6 HYBRIDIZATION TECHNIQUES

DNA is usually found in the form of a double-stranded molecule. These two strands bind to one another in a complementary fashion by a process called hybridization. DNA, when replicated naturally, its new strand hybridizes to the old strand. In the laboratory we can take advantage of hybridization by generating nucleic acid probes which we can use to screen for the presence or absence of certain DNA molecules or RNA molecules in the cell.

Hybridization methods represent standard techniques in molecular biology. In general, they are used to detect particular sequences (target) within a complex mixture of DNA or RNA molecules. DNA or RNA are usually transferred and immobilized to nitrocellulose or, more commonly, to nylon membranes. Complementary single-stranded probes are labelled radioactively or non-radioactively. When hybridized to the filter, probes bind to their complementary target sequence via hydrogen bonds. Unhybridized probe is then washed away, and specifically-bound probe is detected by autoradiography or color reaction.

Applications of direct probe hybridization in diagnostic microbiology include identification of bacteria in blood culture bottles using fluorescence *in situ* hybridization (FISH). In diagnostic molecular pathology *in situ* hybridization is used to detect gene duplications and gene mutations (e.g., *in situ* hybridization for HER-2 amplification). Southern blotting, is an another

hybridization-based method used in genetics assays to assess for large alleles, such as those that may be seen in triplet repeat disorders such as Fragile X syndrome or Huntington disease.

5.7 GENE PROBE- MOLECULAR FINGERPRINTING (DNA FINGERPRINTING)

Gene probe (DNA probe) is a single-stranded DNA or RNA fragment used in genetic engineering to search for a particular gene or other DNA sequence. The probe has a base sequence complementary to the target sequence and will thus attach to it by base pairing. Gene probes can be produced in several ways and fall broadly into three types: gene-specific probes, oligonucleotide probes and polymorphic probes.

Gene-specific probes are produced from specific mRNA by the enzyme reverse transcriptase, which synthesizes a complementary DNA copy (cDNA) from mRNA. If radioactive bases are added to the reaction mixture, the cDNA will be labelled and can thus be used as a hybridization probe to look for the complementary sequences.

Probes can be used in dot-blot hybridisation, where serial dilutions of DNA samples are held on DNA-binding membranes and the complementary radioactively labelled probes are hybridized *in vitro*, so that the amount of radioactive signal is proportional to the amount of target DNA present.

The cDNA can also be cloned by synthesizing a second DNA strand from cDNA using a bacterial DNA polymerase, which is incorporated into a plasmid and grown in bacterial cells. These **oligonucleotide probes** recognize short sequences of DNA that correspond to the sequence known to occur in the gene. With a probe of this length, a single mismatched base pair is sufficient to impair hybridization and can be used to detect changes to a single base (point mutations). Similarly, probes can be developed that recognize the various DNA polymorphisms within the non-coding sequences, for example RFLP and VNTR. DNA probes can also be used to identify abnormal genes or gene products at the molecular level within cell cytoplasm or nucleus, by in situ hybridization. This technique utilizes labelled DNA or mRNA probes which hybridize to the expressed genes in the cell in a manner similar to that used

for immunohistochemistry. In situ hybridization can thus establish whether the genomic material of interest is present in the DNA of the cell in vitro.

DNA fingerprinting:

DNA Fingerprinting is a molecular biology technique which was first developed in 1984 by British geneticist Alec Jeffreys. Sir Alec Jeffrey noticed that certain sequences of highly variable DNA (i.e., VNTRs), which do not contribute to the functions of genes, are repeated within genes are known as minisatellites. Jeffreys recognized that each individual has a unique pattern of minisatellites. Therefore, this technique can assist in the identification of individuals or samples by their respective DNA profiles. This was the main idea behind this marvelous invention.

More than 99.1% of the genome is same throughout the human population, the remaining 0.9% of human DNA shows variations between individuals. These variable DNA sequences, termed molecular markers, can be used to both differentiate and correlate individuals.

The procedure for creating a DNA fingerprint involve following steps:

1. Obtain a sample of cells, such as skin, hair, or blood cells, which contain DNA.
2. The DNA is extracted from the cells and purified.
3. The DNA was then cut at specific points along the strand with restriction endonuclease enzymes which are known as molecular scissors.
4. This enzyme produces fragments of varying lengths.
5. These fragments are then separated by gel electrophoresis technique.

6. The shorter the fragment, the more quickly it moved toward the positive pole (anode).
7. The sorted double-stranded DNA fragments were then subjected to a blotting technique in which they were split into single strands and transferred to a nylon sheet.

8. The fragments underwent autoradiography in which they were exposed to DNA probes (i.e., pieces of synthetic DNA) that were made radioactive and that bound to the minisatellites.

9. A piece of X-ray film was then exposed to the fragments, and a dark mark was produced at any point where a radioactive probe had become attached. The resultant pattern of marks could then be analyzed.
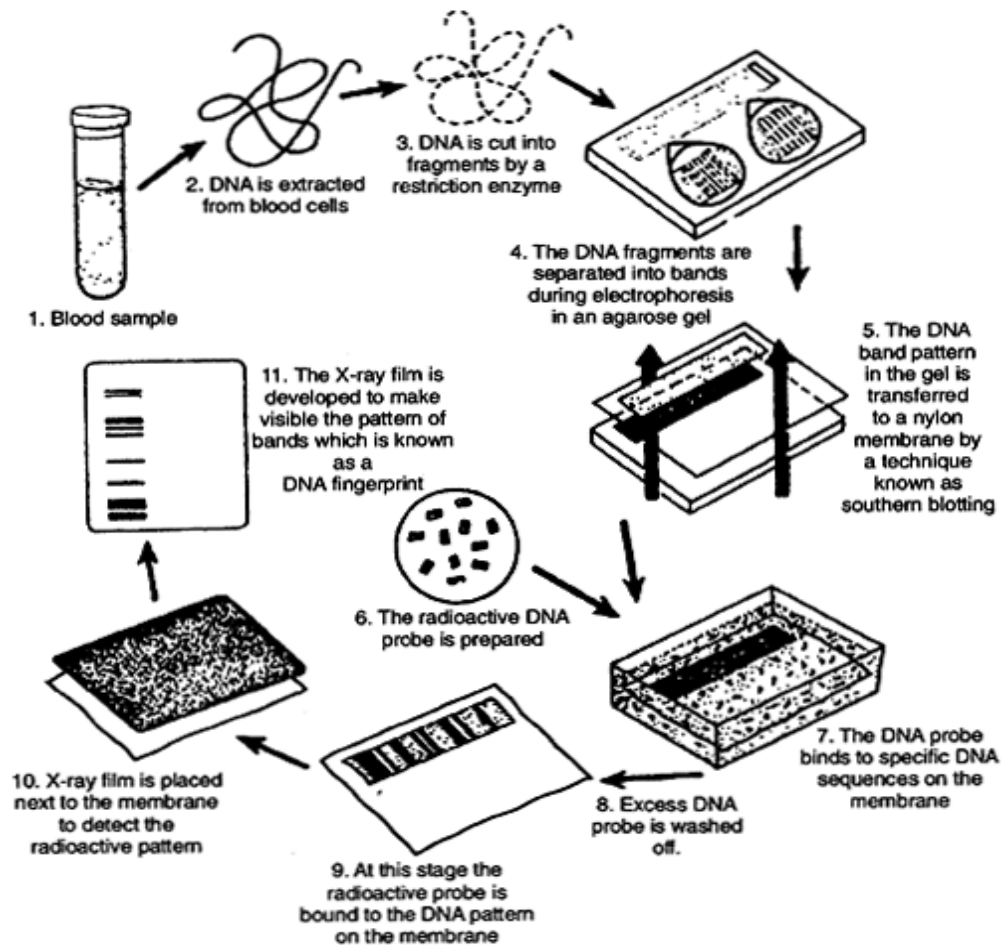

Figure 5.3: Steps involved in DNA fingerprinting (source: pinterest.com)

5.8 MOLECULAR MARKERS IN GENOME ANALYSIS (RFLP, RAPD AND AFLP)

A molecular marker is a molecule contained within a sample taken from an organism (biological markers) or other matter. DNA, for example, is a molecular marker containing information about genetic disorders and the evolutionary history of life.

Molecular markers help in detection of variations or polymorphisms that exist among individuals in the population for specific regions of DNA (e.g. RFLP, AFLP, SNP, etc.). Applications of Molecular markers in gene mapping are:

(1) A marker allows the direct identification of the gene of interest instead of the gene product, and consequently, it serves as a useful tool for screening somatic cell hybrids.

(2) Use in several DNA probes and easy-to-screen techniques, a marker also helps in the physical mapping of the genes using *in situ* hybridization.

(3) Molecular markers provide sufficient markers for construction of genetic maps using linkage analysis.

1. Restriction fragment length polymorphism:
A Restriction fragment length polymorphism (or RFLP) is a variation in the DNA sequence of a genome that can be detected by breaking the DNA into pieces with restriction enzymes and analysing the size of the resulting fragments by gel electrophoresis. It is the sequence that makes DNA from different sources different, and RFLP analysis is a technique that can identify some differences in sequence (when they occur in a restriction site). Though DNA sequencing techniques can characterize DNA very thoroughly, RFLP analysis was developed first and was cheap enough to see wide application. Analysis of RFLP variation was an important tool in genome mapping, localization of genetic disease genes, determination of risk for a disease, genetic fingerprinting, and paternity testing.

2. RAPD (Random Amplification of Polymorphic DNA)

It is a type of PCR reaction, but the segments of DNA that are amplified are random. The scientist performing RAPD creates several arbitrary, short primers (8-12 nucleotides), then proceeds with the PCR using a large template of genomic DNA, hoping that fragments will amplify. By resolving the resulting patterns, a semi-unique profile can be gleaned from a RAPD reaction.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. This makes the method popular for comparing the DNA of biological systems that have not had the attention of the scientific community, or in a system in which relatively few DNA sequences are compared (it is not suitable for forming a DNA databank). Due to the fact that it relies on a large,intact DNA

template sequence, it has some limitations in the use of degraded DNA samples. Its resolving power is much lower than targeted, species specific DNA comparison methods, such as short tandem repeats. In recent years, RAPD is used to characterize, and trace, thephylogeny of diverse plant and animal species.Random Amplified Polymorphic DNA (RAPD) markers are ten mer (10 nucleotide length) DNA fragments from PCR amplification of random segments of genomic DNA with single primer of arbitrary nucleotide sequence and which are able to differentiate between genetically distinct individuals, although not necessarily in a reproducible way.

3. Amplified fragment length polymorphism: Amplified fragment length polymorphism PCR (or AFLPPCR or just AFLP) is a PCR-based tool used in genetics research, DNA fingerprinting, and in the practice of genetic engineering. AFLP uses restriction enzymes to cut genomic DNA, followed by ligation of adaptors to the sticky ends of the restriction fragments. A subset of the restriction fragments are then amplified using primers complementary to the adaptor and part of the restriction site fragments (as described in detail below). The amplified fragments are visualized on denaturing polyacrylamide gels either through autoradiography or fluorescence methodologies. AFLP-PCR is a highly sensitive method for detecting polymorphisms in DNA. The technique was originally described by Vos and Zabeau in 1993.

In detail, the procedure of this technique is divided into three steps: **1**. Digestion of total cellular DNA with one or more restriction enzymes and ligation of restriction half-site specific adaptors to all restriction fragments. **2**. Selective amplification of some of these fragments with two PCR primers that have corresponding adaptor and restriction site specific sequences. **3**. Electrophoretic separation of amplicons on a gel matrix, followed by visualisation of the band pattern.

Flowchart: Dipicting categorisation of molecular markers
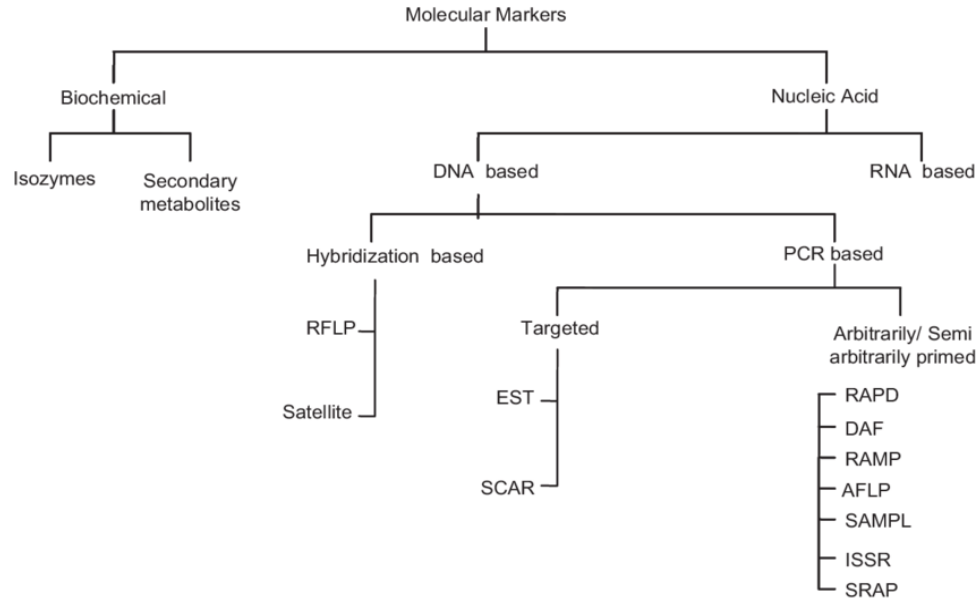
5.9 GENOMIC LIBRARY

A genomic library is a collection of overlapping segments of genomic DNA, cloned into a backbone vector, which statistically includes all regions of the genome of an organism. The resulting cloned DNA is then transformed into a suitable host cell line. Construction of a genomic library is an important initial step in many genetic studies or in the isolation and cloning of genes from an organism. Screening of genomic libraries has been useful in identifying genes of interest to the medical field and the biotechnology industry as well as in finding genes related to particular cellular functions. Additionally, creating a representative genomic library of an organism is a prerequisite for genomic mapping or complete genome sequencing. The success of a study involving genomic libraries is dependent upon the quality and features of the library. These features typically include the vector backbone used, the size of the genomic DNA insert, and the number of recombinant clones contained within the library.

To make a prokaryotic gene library, the complete bacterial chromosomal DNA is cut with a restriction enzyme and each of the fragments is inserted into a vector, usually a simple ColE1-derived plasmid. This mixture of vectors containing a different piece of the bacterial chromosome is transformed into a suitable bacterial host strain and a large number of colonies,

each containing a single vector plus insert, are kept. These must then be screened for the gene of interest. If the gene has an observable phenotype, this may be used. Otherwise, more general methods such as hybridization or immunological screening are necessary.
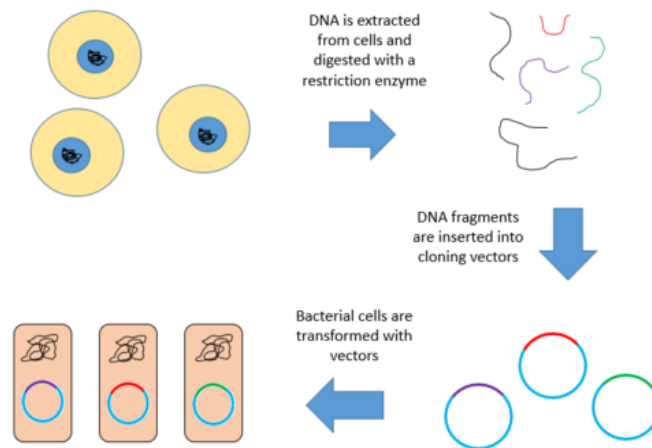
Figure 5.4: Construction of genomic library (source: Wikipedia.org)

5.10 BLOTTING TECHNIQUES: SOUTHERN NORTHERN AND WESTERN BLOTTING

Southern blotting:

Southern blot technique involves detection and immobilization of target DNA onto a membrane. This technique was invented in 1975 by E. M. Southern.

Steps:

1. The DNA fragments are first separated on an agarose gel and then transferred and immobilized onto a membrane.

2. After the gel has been run, the DNA fragments are denatured (i.e. the strands are separated) using strong alkali.

3. As electrophoretic gels are fragile, the DNA in them can diffuse within the gel, it is usual to transfer the denatured DNA fragments by blotting onto a nitrocellulose membrane.

4. The single-stranded nature of the DNA on the membrane is important to allow complementary DNA sequences to be able to bind to the DNA fragments attached to the membrane.

5. The transfer process can be mediated either electrophoretically or through capillary attraction by placing the gel and membrane in contact with each other and allowing the buffer to flow through the gel onto the membrane – the DNA fragments move with the buffer and become trapped on the membrane. Initially, nitrocellulose membranes were used, but these were fragile and easily broken. Nylon membranes are commonly used today.

6. After transfer, the DNA fragments need to be fixed to the membrane so that they cannot detach. A number of methods of fixing are available including baking at 80°C and ultraviolet cross-linking. UV cross-linking is based on the formation of crosslinks between a small fraction of the T residues in the DNA and the positively charged amino groups on the surface of the nylon membrane.

7. Following fixation, the membrane is placed in a solution of labelled (often radioactive) single-stranded nucleic acid – either single-stranded DNA or RNA.

8. The labeled nucleic acid (or **probe**) is allowed to hybridize to its complementary partner sequence on the membrane.

9. The interaction between the single-stranded probe and its complementary sequence will result in the binding of the probe to the membrane through non-covalent hydrogen bonding that normally holds the DNA double helix together.

10. The membrane is then washed extensively to remove non-specifically bound probes, and specific interactions are detected by exposing the membrane to X-ray film. Southern blotting is used to detect DNA sequences that are either identical or similar to the sequence of the probe.

Northern blotting

Northern blotting or Northern blot hybridization is a variant of Southern blotting in which the target nucleic acid is RNA instead of DNA. This method is used to measure the amount and size of RNA. The detection of RNA sequences is performed by the use of probes.
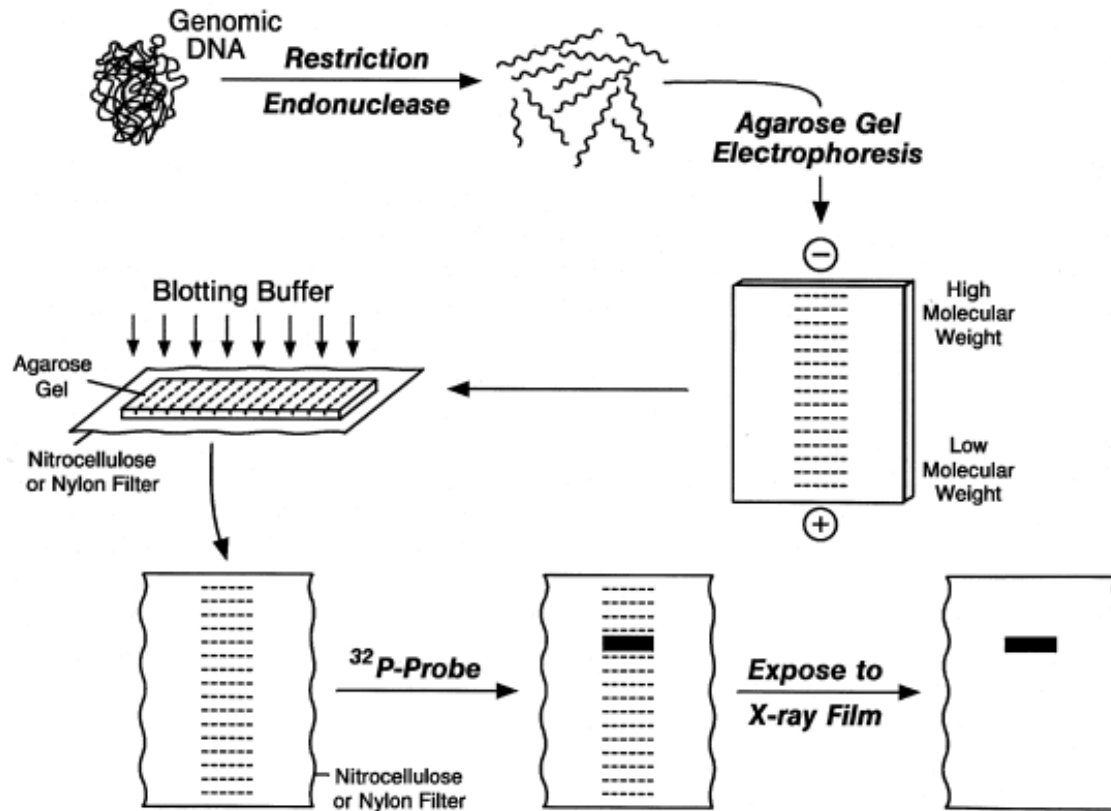
Figure 5.5: Steps involved in the Southern blotting (copyright MyBioSource)

Steps:

1. RNA extract is electrophoresed in an agarose gel, using a denaturing buffer to ensure that the RNA do not form inter or intramolecular base pairs.

2. After separation on an agarose gel, the gel is blotted onto a reactive DBM (diazo benzyloxymethyl) paper, and hybridized with a labelled probe.

3. RNA bands can also be blotted onto nitrocellulose paper under appropriate conditions and suitable nylon membranes.

4. Specific RNA molecules are then detected by hybridization using labelled single-stranded DNA or RNA sequences that are complementary to particular RNAs.

Nylon is a generic name for any long-chain synthetic polymer having recurring polyamide groups. Two types of nylon membranes are: unmodified (or neutral) nylon and charge-modified nylon which carries amine groups and is therefore, also known as positively charged nylon. Both types of nylons are able to bind with single and double stranded nucleic acids.
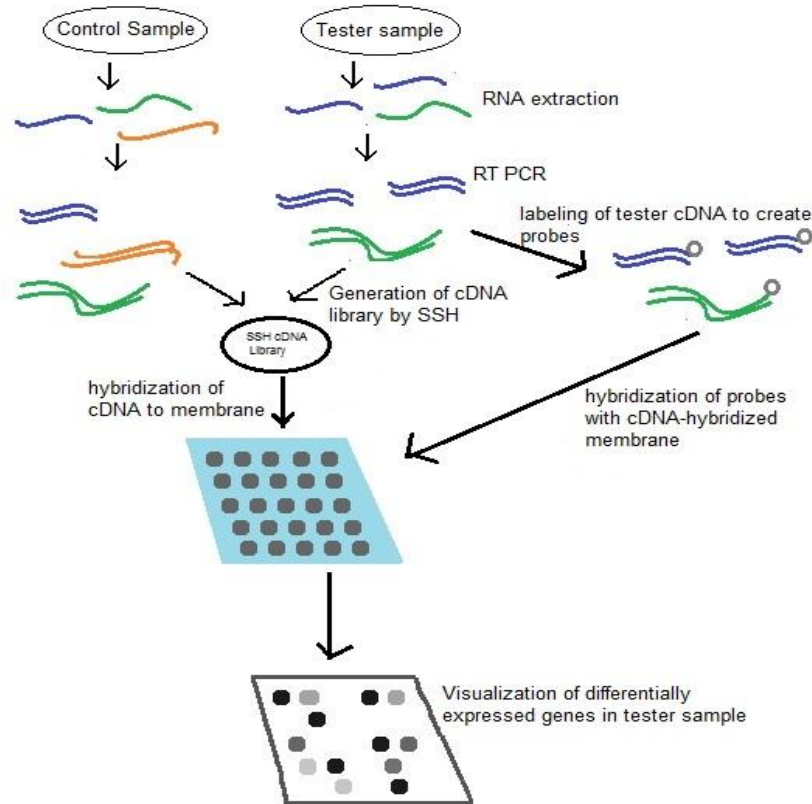
Figure 5.6: Northern Blot Hybridization (Slideshare.net

Western blotting

Western blotting is the method of detection of specific proteins using antibodies or immune blotting techniques. This technique requires antibodies against the test protein. Proteins are separated through a polyacrylamide gel containing the detergent SDS to keep them in an unfolded (**denatured**) state.

Steps:

1. The proteins are transferred from the gel onto a membrane in much the same way as Southern blotting.

2. Particular proteins are then detected using antibodies. The specific interaction between the antibody and its antigen occurs on the membrane, and the position of the bound antibody is detected.

3. Initially radio-labelled antibodies were used, but these have been largely superseded by antibody 'sandwiches'.

4. The sandwiches work through the binding of one unlabelled antibody (the primary antibody) to the antigen on the membrane.

5. A second, labeled, antibody (the secondary antibody) is then used to detect the presence of the first antibody.

This has several advantages; firstly, the multivalent nature of antibody binding means that a substantial increase in sensitivity is achieved, and secondly, a single secondary antibody can be used to detect a number of different primary antibodies.
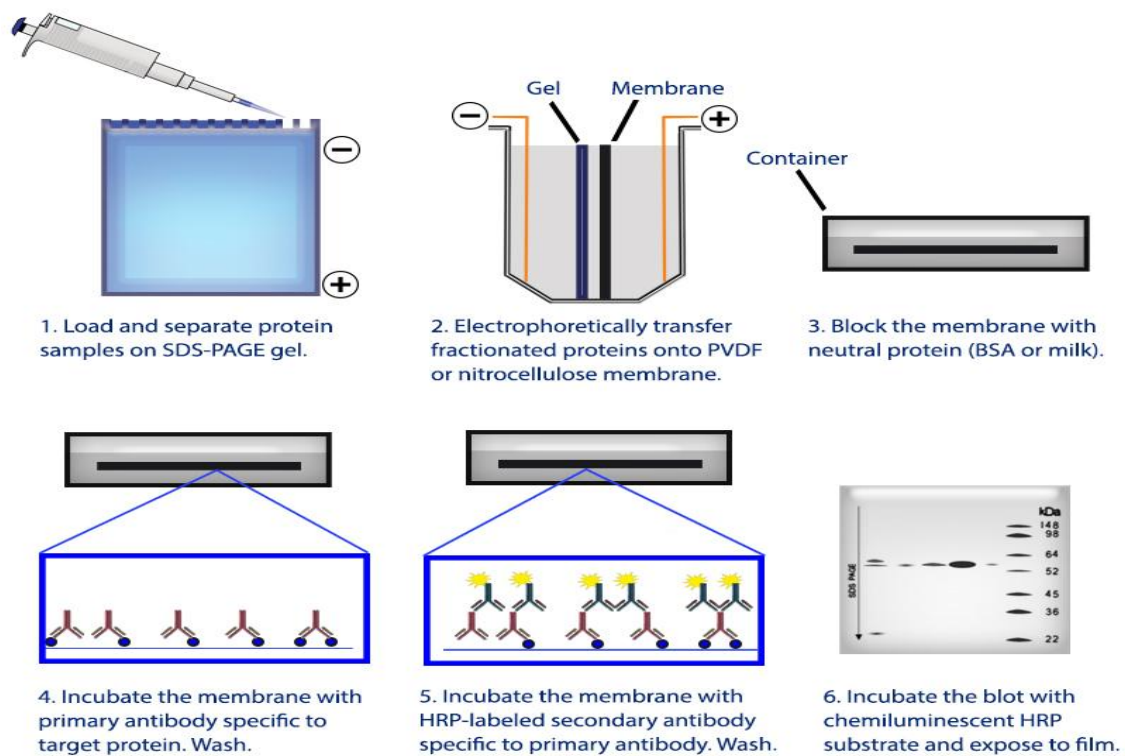
Figure 5.7: Western Blot Hybridization (Source: creative Biolabs)

5.11 SUMMARY

The basic process of recombinant DNA technology revolves around the activity of DNA in the synthesis of protein. By inserting genes into the genome of an organism, the scientist can induce the organism to produce a protein it does not normally produce. Gene cloning is a set of experimental methods in molecular biology that are used to assemble recombinant DNA molecules and to direct their replication within host organisms. Restriction enzyme, also called restriction endonuclease, is a protein produced by bacteria that cleaves DNA at specific sites along the molecule. After introduction of recombinant DNA into the host cell it is essential to identify those cells which received the recombinant DNA molecule, hence selection of recombinant and screening is performed by colony hybridization, blue white screening and by positive selection methods. Hybridization is the process of combining two complementary single-stranded DNA or RNA molecules and allowing them to form a single double-stranded molecule through base pairing. Hybridization is a part of many important laboratory techniques such as polymerase chain reaction and Southern blotting. A gene probe (nucleic acid probe) is a single-stranded nucleic acid fragment that interacts with a complementary sequence of a target nucleic acid. The use of DNA fingerprinting depends upon the presence of repeating base sequences that exist in the human genome. The repeating sequences are called restriction fragment length polymorphisms (RFLPs). As the pattern of RFLPs is unique for every individual, it can be used as a molecular fingerprint. A genomic library is a collection of the total genomic DNA from a single organism. Blots are techniques for transferring DNA, RNA and proteins onto a carrier so they can be separated, and often follows the use of gel electrophoresis. Southern blot is used for transferring DNA and northern blot is for transferring RNA and western blot is used for protein.

5.12 GLOSSARY

Amino acid: The constituent subunits of proteins. Amino acids polymerize to form linear chains linked by peptide bonds; such chains are termed polypeptides (or proteins if large enough). There are twenty commonly occurring amino acids of which all proteins are made.

Annealing: Spontaneous alignment of two complementary single polynucleotide (RNA, or DNA, or RNA and DNA) strands to form a double helix.

Anticodon: Triplet of nucleotide cases (codon) in transfer RNA that pairs with (is complementary to) a triplet in messenger RNA. For example, if the codon is UCG, the anticodon might be AGC.

Antisense RNA: RNA produced by copying and reversing a portion of an RNA-encoding DNA, usually including a protein-specifying region, and placing it next to a transcription-control sequence. This cassette can be delivered to the target cell, resulting in genetic transformation and production of RNA that is complementary to the RNA that is produced from the original, not-reversed, DNA segment. This complementary, or antisense, RNA is able to bind to the complementary sequences of the target RNA, resulting in inhibition of expression of the target gene.

Base: On the DNA molecule, one of the four chemical units that, according to their order, represent the different amino acids. The four bases are: adenine (A), cytosine(C), guanine (G), and thymine (T). In RNA, uracil (U) substitutes for thymine.

Base pair: Two nucleotide bases on different strands of a nucleic acid molecule that bond together. The bases generally pair in only two combinations; adenine with thymine (DNA) or uracil (RNA), and guanine with cytosine.

Bacillus thuringiensis (Bt): A naturally occurring microorganism that produces a toxin protein that only kills organisms with alkalineing stomachs, namely such as insect larvae. As a When delivered as a part of the whole killed organism, this toxin protein has been used for biological control for decades. The genetic information that encodes the toxin protein was identified and moved into plants to make them insect tolerant.

Biotechnology: Development of products by a biological process. Production may be carried out by using intact organisms, such as yeasts and bacteria, or by using natural substances (e.g. enzymes) from organisms.

Deoxyribonucleic acid (DNA): The molecule that carries the genetic information for most living systems. The DNA molecule consists of four bases (adenine, cytosine, guanine, and thymine) and a sugar-phosphate backbone, arranged in two connected strands to form a double helix. A term often used to describe the configuration of the DNA molecule. The helix consists of two spiralling strands of nucleotides (a sugar, phosphate, and base), joined crosswise by specific pairing of the bases. See also Deoxyribonucleic acid; Base; Base pair.

Endonuclease: An enzyme that breaks nucleic acids at specific interior bonding sites; thus producing nucleic acid fragments of various lengths.

Enzyme: A protein catalyst that facilitates specific chemical or metabolic reactions necessary for cell growth and reproduction.

Escherichia coli (E. coli): A bacterium that inhabits the intestinal tract of most vertebrates. Much of the work using recombinant DNA techniques has been carried out with this organism because it has been genetically very well characterized.

Eukaryote: A cell or organism containing a true nucleus, with a well-defined membrane surrounding the nucleus. All organisms except bacteria, archebacteria, viruses, and blue-green algae are eukaryotic. Cf. Prokaryote.

Exonuclease: An enzyme that breaks down nucleic acids only at the ends of polynucleotide chains, thus releasing one nucleotide at a time, in sequential order.

Gene therapy: The replacement of a defective gene in an organism suffering from a genetic disease. Recombinant DNA techniques are used to isolate the functioning gene and insert it into cells. Over three hundred single gene genetic disorders have been identified in humans. A significant   percentage   of   these   may   be   amenable   to   gene   therapy. **Genetic code:** The mechanism by which genetic information is stored in living organisms. The code uses sets of three nucleotide bases (codons) to make the amino acids that, in turn, constitute proteins. Genetic engineering A technology used to alter the genetic material of living cells in order to make them capable of producing new substances or performing new functions. **Genetic Map:** A map showing the positions of genetic markers along the length of a

chromosome relative to each other (genetic map) or in absolute distances from each other

Genetic screening: The use of a specific biological test to screen for inherited diseases or medical conditions. Testing can be conducted prenatally to check for metabolic defects and congenital disorders in the developing foetus as well as postnatal to screen for carriers of heritage diseases.

5.13 TERMINAL QUESTIONS AND ANSWERS

Question1. What is recombinant DNA technology? What are the basic steps involved in recombinant DNA technology?

Question2. Write short notes on the following:

a) Ligase

b) Linkers

c) Adapters

d) cDNA

Question3. Write a note on DNA fingerprinting and the steps involved in it?

Question4. What are molecular markers in recombinant DNA technology?

Question5. Write note on following:

a) RFLP

b) RAPD

c) AFLP

Question6. Elaborate the following techniques:

a) Southern blotting

b) Northern blotting

c) Western blotting

7. Plasmids are used in genetic engineering because they:

a) are easily available

b) can replicate

c) can integrate with host's chromosome

d) are inert

8. Which of the following enzyme is used to join bits of DNA?

a) Ligase

b) Primase

c) DNA polymerase

d) endonuclease

9. The sites of DNA where restriction enzymes act are generally

a) Palindromic

b) Tandem repeats

c) CG rich regions

d) TATA Boxes

10. Disease causing organisms are identified by using

a) MAb's

b) Microprojectiles

c) DNA probes

d) Gene transfer

5.14 REFERENCES

1. T. A. Brown. Gene Cloning and DNA Analysis: AN Introduction, 8th edition.

2. David P. Clark and Nanette J. Pazdernik. Biotechnology: Applying the genetic revolution

3. Sandy B. Primrose, Richard. Principles of gene manipulation and genomics.

4. https://www.ndsu.edu/pubweb/~mcclean/plsc731/mapping/mapping1.htm

5. https://www.nature.com/scitable/topicpage/restriction-enzymes-545/

6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287216/

UNIT 6: CLONING VECTORS

CONTENTS

6.1 OBJECTIVES

After studying this module, you shall be able to learn and understand:

i. Basic concept of cloning

ii. Cloning vectors and their desirable properties

iii. Types of cloning Yeast

iv. *E. coli*

v. pBR 322

6.2 INTRODUCTION

The term cloning in recombinant technology refers to the production of 'clones' or genetically identical organisms. In natural cloning, the asexual reproduction methods are used by bacteria, fungi and plants to produce genetically identical offspring. A number of organisms like fungi, algae, protozoan and plants produce clones naturally through asexual reproductions that are genetically identical to the parent organism. Using processes like binary fission, budding, fragmentation and parthenogenesis organism can produce genetically identical offspring. Formation of identical twins in case of humans is also natural cloning.

Different Types of Cloning:

1. Molecular Cloning/ DNA Cloning/ Gene Cloning: Molecular cloning involves the production of large quantities/ copies of gene or DNA fragments. Molecular cloning involves the construction of recombinant DNA by insertion of gene of interest in vector DNA. Several copies of this constructed recombinant DNA are produced after it is introduced into host cells and replicates as the host cells proliferate. Finally either the gene or expressed protein is extracted, purified and sequenced. Commonly, these vectors belong to bacteria, yeasts, viruses, or plasmids.

2. Reproductive Cloning: Reproductive cloning is used for animals cloning. In reproductive cloning involves a mature somatic cell is taken out from the animal that has to be cloned. After this the somatic cell DNA is transferred into an enucleated egg cell (produced by removing the

nucleus). The cell division in the restructured egg is stimulated using electric current. After few cell divisions the restructured egg develops into an embryo. The embryo is then implanted into the womb of a healthy adult female where it will develop until birth. The genetic makeup of this young animal is similar to the somatic cell donor animal and is therefore known to be the clone of the donor animal. Dolly, the clone sheep was also created using reproductive cloning by Ian Wilmut in 1996. There are enormous potential benefits of this cloning method in medicinal and agricultural fields.

3. Therapeutic Cloning/Embryo Cloning: Therapeutic cloning essentially has the same process as reproductive cloning, however, with distinct outcomes and objectives. The objective of therapeutic cloning is to learn and understand human development and cure disease by production and use of stem cells. Stem cells can be used for the production of different types of specialized cells in the human body. Embryonic stems can be extracted out and used to produce specialized cells. Ethical concerns have been raised as extraction of stem cells cause the embryo destruction. But owing to the potential and crucial role of these stem cells in treating different diseases and also in replacing damaged cells to cure injuries, scientists and researchers find extreme hope in therapeutic cloning or embryo cloning.

6.3 PLASMID BIOLOGY

Plasmids may carry from half dozen to several hundreds of functional genes. The limiting features of plasmids are that they can multiply only within a host cell. Mostly plasmids are carried by bacteria and are essential part of 50% of bacteria found on earth. In some higher organisms also these plasmids are found as extra chromosomal segments like yeast and fungi. One of the best examples of higher organism's plasmid is 2m circle, a well known cloning vector of yeast origin. The number of plasmids found in bacteria per chromosome is known as the copy number of that plasmid, mostly bacteria have one or two copy number, but in some cases they may carry a copy number of 50 to hundreds, these are called high copy number plasmids. Similarly the size of plasmids is also enormously variable and it varies from few hundred base pairs to thousands of base pairs. Mostly plasmids carry genes which are responsible for their own maintenance, but some plasmids attribute characters to the host cells. In molecular biology, we

use a number of genetically engineered plasmids to carry the genes for tailoring of genetic information, and other purposes.

Plasmids are circular, double stranded DNA (dsDNA) molecules, found free in host cells, and these are additional DNA than chromosomal DNA. The occurrence of these extra chromosomal DNAs naturally as parasitic or symbiotic relationship in bacteria and some lower eukaryotic cells like yeast. The distribution of plasmids in each daughter cell is performed just like chromosomal DNA, these plasmids replicate and segregate themselves to distribute equally. Plasmids are naturally occurring extra chromosomal DNA found in various groups of bacteria. These plasmids have capability to self replicated due to presence of origin of replication site in it. The size of plasmid DNA ranges from 1 kb to 250 kilo base pairs. In nature plasmids have very important role in bacterial evolution. Almost all plasmids are double stranded circular DNA, conferring an extra phenotypic character to the bacteria. There are various functions which are served by these independent molecules. These characters may include antibiotics resistance (ampicillin, tetracycline, kanamycin etc.), abiotic stress tolerance (heat, cold, salt, toxic etc.) or may responsible for synthesis of some special polypeptides (toxins, metabolites). The closed-circular DNA of plasmid is cross coiled over its own axis in three-dimensional space to form a Super coil. Plasmids can be classified on the basis of various features found in it.

- Fertility F-plasmids, having a tra gene which enables bacterium to conjugation through sex pili.

- Resistance (R) plasmids, containing antibiotics or poisons resistant genes, conferring antibiotics resistance to the host bacterium also known as R-factors.

- Col plasmids are genes that transcribed and translated for bacteriocins, these proteins are responsible for killing of other bacteria.

- Degradative plasmids, provide the capability to bacterium for degradation of various organic/inorganic substances for harmful for bacteria, e.g. toluene and salicylic acid.

- Virulence plasmids, confers pathogenicity to the bacterium.

Plasmids having resistance and Defense related mechanisms:

- Antibiotic resistance for aminoglycosides, b-lactams, chloramphenicol, sulfonamides, trimethoprim, fusidic acid, tetracyclines, macrolides, fosfomycin

- Heavy metal ions resistance for Ni, Co, Pb, Cd, Cr, Bi, Sb, Zn, Cu and Ag

- Tolerance to mercury and other mercury compounds

- Toxic anions resistance such as chromate, selenate, tellurite, arsenate, arsenite, borate, etc

- Intercalating agent resistance such as acridines and EtBr

- Radiation like UV and X-rays damage protecting plasmids

- Bacteriophage DNA restriction systems

- Resistance towards some bacteriophages

Virulence related plasmids:

- Bacteriocins synthesizing plasmids

- Antibiotics synthesizing plasmids

- Crown gall tumors and hairy root inducing plasmids

- Nodulation in legumes related plasmids

Metabolic Pathways:

- Solbulization of sugars like lactose, raffinose, sucrose related plasmids

- Biodegradation of aliphatic and aromatic hydrocarbons

- Biodegradation of halogenated hydrocarbons like polychlorinated biphenyls

- Bioegradation of proteins

- Hydrogen sulfide Synthesising Plasmids

- Alcaligenes Denitrification

- Pigment synthesising plasmids

Fertility F-plasmids: With the discovery that genetic recombinants can be done by mixing particular Escherichia coli K 1 2 strains together, it was realized that the reason behind getting

recombination is an iniderectional transfer of DNA segments from F- donor strains that contained an infectious "fertility factor," F. The F was found to replicate independently and inducts its DNA crossing the cell envelopes of bacteria which comes in contact. The injected DNA gets recombined with the chromosomal DNA at various locations. The F plasmid is covalently closed circular plasmid having approximately 60 genes with a total length of 100 kilo bases.

Resistance (R) plasmids: The resistance in a bacterium attributed to various mechanisms such as chromosomal mutations, in chromosomal alterations the resistance is most commonly associated with extra chromosomal elements which are commonly acquired from other bacteria. These moving transposing elements may be plasmids, transposons, and/or integrons. The intrinsic mechanisms that evolves efflux pumps to out multiple kinds of antibiotics, are supposed to be major contributors to multidrug resistance. Bacteria can acquire antibiotic resistance by any of the two mechanisms whether intrinsic or acquired mechanisms. The bacteria can have naturally occurring genes such as, AmpC, β-lactamase of gram-negative bacteria may have Intrinsic mechanisms in combination with efflux systems.

Col plasmids: Escherichia coli and some other bacteria produce Colicins, a toxic protein. The production of colicin involved in bacteria to bacteria competition and virulence determination. These colins are different in their activity to kill other bacteria though they share some common features like lethal colicin release, smililar genetic sequences having genes like a colicin, lysis and immunity related genes producing peptides that interacting with a specific locus in the colicin protein seuqence. These have immunity towards colicin but when colicin is produced by a cell, it dies. These colicin related gene clusters are carried by some special plasmids.

Degradative Plasmids: In our daily life microorganisms play a great role in degradation of various products and by products like sewage, oil waste, agricultural byproducts, various pesticides, toxic substances etc. Many microorganisms have capacity to break down complex organic molecules and ability to recycle them by including these in to their meabolic cycles exhibiting a variety of degradative functions. Bacteria like Pseudomonas, Alcaligenes, Chromobacterium, have plasmids which have different genes responsible for degradation of a variey of inorganic and organic substances, these are known as biodegrading bacteria.

Virulence plasmids: The bacteria can be subdivided into many groups on the basis of their pathagenicity, the pathogenicity of a bacterium mainly depends upon the presence or absence of pathogenic DNA sequences which are frequently associated with various pathotypes. In these bacteria the genetic information for pathogenicity have been acquired horizontally through plasmids, bacteriophages and genomic islands. These genomic rearrangements are utilized by bacteria for their evolution though variants efficient in rearrangements, excision and transfer for affinity to additional DNA for creation of new (pathogenic) variants.

Plasmids are stably-inherited extra chromosomal, autonomously replicating; double stranded and covalently linked circular DNA molecules. Their size varies from 5000 to 400000 bp. These are commonly found in bacteria and also archaea and eukaryotes. An ideal cloning vector should have the following three properties:

(i) **Low Molecular weight:** This aid in easy handling and isolation due to resistance to shearing also low molecular weight vectors are present in multiple copies. Transfer efficiency of low molecular weight vector is quite high.

(ii) **Origin of replication:** The vectors should have a compatible an origin of replication with the host. It is important to have a multiple copies of a vector inside the host.

(iii) **Selection markers:** These markers help in the selection process of the chimaeric molecules over self annealed plasmid molecules (like antibiotic resistance).

(iv) **Multiple Cloning Site (MCS)/ Polylinker:** Presence of MCS or polylinker, which contains many unique restriction sites on vector is important requirement as for cloning both the vector and gene of interest has to be digested with the same restriction enzymes and then are ligated using DNA ligase.

Plasmids carry their own origin of replication and therefore can be maintained stably in host cells. These plasmids can be introduced in host cells by transformation. The typical approach is to use a plasmid that carries a gene that is required for the growth of host cell under particular condition (s) like antibiotic resistance gene. Under these conditions only the transformed host cells carrying the recombinant molecule are able to grow in the pressure of that antibiotic. Such genes aids in the selection of cells containing the recombinant plasmid are known as marker/selection genes. A DNA fragment (insert) capacity of these plasmid vectors is up to

15000 bp. pBR322 and its derivatives of like pUC 18/19 are most commonly used plasmid vectors in cloning.

Plasmid vectors

To carry and replicate the DNA fragment for various purposes, a molecular vehicle is needed in the process of molecular cloning. The transfer of DNA fragments with the help of molecular cloning vectors became possible due their ability to self-replication in E. coli or any other host cell, autonomous replication are responsible for this feature. Thus the origin of replication plays an important role in the molecular cloning. Most cloning vectors were engineered originally from extra chromosomal elements found in nature such as bacteriophage and plasmids. Plasmids are the DNA molecules that can replicate itself inside a host cell. These molecular vectors are used for carrying cloned fragments of DNA. The vectors may be a small multi-copy plasmid or a designed and engineered virus. Almost all plasmid vectors are the engineered extra chromosomal naturally found plasmids, isolated from different types of bacteria. Naturally found plasmids have several limitations; for example, some are stringent and not relaxed (pSC101), some has poor marker genes (ColE1), and some are too large (RSF2124). To overcome the limitations of natural vectors, artificial plasmid are designed and engineered by combining different elements. Artificial plasmids vectors are classified into two broad types based on their use:

 1. Cloning vector

2. Expression vector

Apart from these two, there is another class of vectors known as shuttle vector. Shuttle vectors can be propagated in two or more different host species (both in prokaryotes and eukaryotes). Hence, inserted DNA can be manipulated and replicated in two different cellular systems. Cloning vectors are designed for efficient transfer of foreign DNA into the host. Expression vectors have efficient machinery for cloning and expression of foreign gene in the hostsystem. Selection of a vector depends upon various criteria decided by the experimental goal.

Plasmid Expression vector:

Once a recombinant DNA has been made into a vector, the expression of that gene may or may not take place. The proper expression of a gene the structural gene and the corresponding promoter should be cloned on the same segment. In specialized expression vectors these promoter genes are provided by the vectors, only the structural genes have to be inserted on proper places. The best example of it is the vectors having capability of blue white screening have lac promoter, in these the multiple cloning site lies just in the promoter region. The most important and exploit use of genetic engineering is to use recombinant microorganism for production of various biochemicals. These biochemicals are being used widely in different industries for well being of the human used in pharmaceutical industries, cosmetics, bio-plastics etc. These biochemical have different origins, but due to advantage of recombinant DNA technology we are able to produce in bacteria. The basic of phenomenon of this is, the over expression of genes by some modifications or rearrangements in the natural sequences. This technique of over expression of genes is also being used in various structural studies, in determining biological functions etc. As convenient Escherichia coli have been widely used as the prokaryotic host for expression and overproducing various proteins. Its advantages are:

 (1) The genetics and physiology or E. coli is well characterized and arrays of expression vectors are present.

(2) The manipulation in E. coli is easily and cost efficient.

(3) The production of foreign protein can be achieved up to the level of 5%-30% or more of the total protein of the cells. In many instances, however, we cannot use E. coli. Like, for proteins that requires post translational modification. In the case of heterologous expression, it fails to produce polypeptide to assume its native configuration.

6.4 CLONING VECTOR

Cloning vectors are designed for efficient transfer of foreign DNA into the host. Expression vectors have efficient machinery for cloning and expression of foreign gene in the host system. Selection of a vector depends upon various criteria decided by the experimental goal.

One of the most requirements for cloning is a vector. A cloning vector is self replicating small piece of DNA of plasmid, virus, yeast or cell of higher organism into which gene of interest can

be cloned and expressed. The vector and foreign DNA (insert) are cleaved to produce the staggered cut generating similar ends. Using DNA ligase these are then spliced together. This recombinant molecule is then introduced into host cells by transformation or transfection. Since the vectors have origin of replication recognizable by host cells manifold multiple copies of the foreign DNA can be produced. The most commonly used cloning vectors are the following: 1. Plasmids. 2. Bacteriophages 3. Bacterial artificial chromosomes (BAC). 4. Yeast artificial chromosome (YAC). 5. Mammalian artificial chromosome (MAC).

6.4.1 YEAST ARTIFICIAL CHROMOSOMES (YACS)

First described in 1983 by Murray and Szostak, a yeast artificial chromosome has sequences to exist inside E. coli as a circular plasmid and contains sequences to maintain as linear nuclear chromosome in yeast. The number of clones in a genomic library can be greatly reduced. YAC vectors have following elements:

1. coli origin of replication

2. Yeast origin of replication

3. Elements of eukaryotic yeast chromosome (centromere and telomere region)

4. Selection markers for both the hosts (Bacterial as well as Yeast)

5. Maintained as linear DNA-like chromosome.

 6. Introduced into the yeast cells by electroporation.

YAC is a vector used to clone DNA fragments larger than 100 kb and up to 3,000 kb. YACs are useful for the physical mapping of complex genomes and for the cloning of large genes. Yeast artificial chromosomes are created placing a centromere(CEN), telomeres (TEL), and an autonomous replicating sequence (ARS) element together to replicate autonomously and conserve itself in yeast cells. ARS are supposed to function as initiation point for DNA replication in yeasts. A circular plasmid is restriction digested to create a lenear YAC vector. These YAC are cleaved in cented to creat two arms and the foreign DNA are ligated between these two arms forming a single linear molecule. TRP1 and URA3 genes are included in the YAC vector to provide a selection system for identifying transformed yeast cells that include

YAC by complementing recessive alleles trp1 and ura3 in yeast host cell.
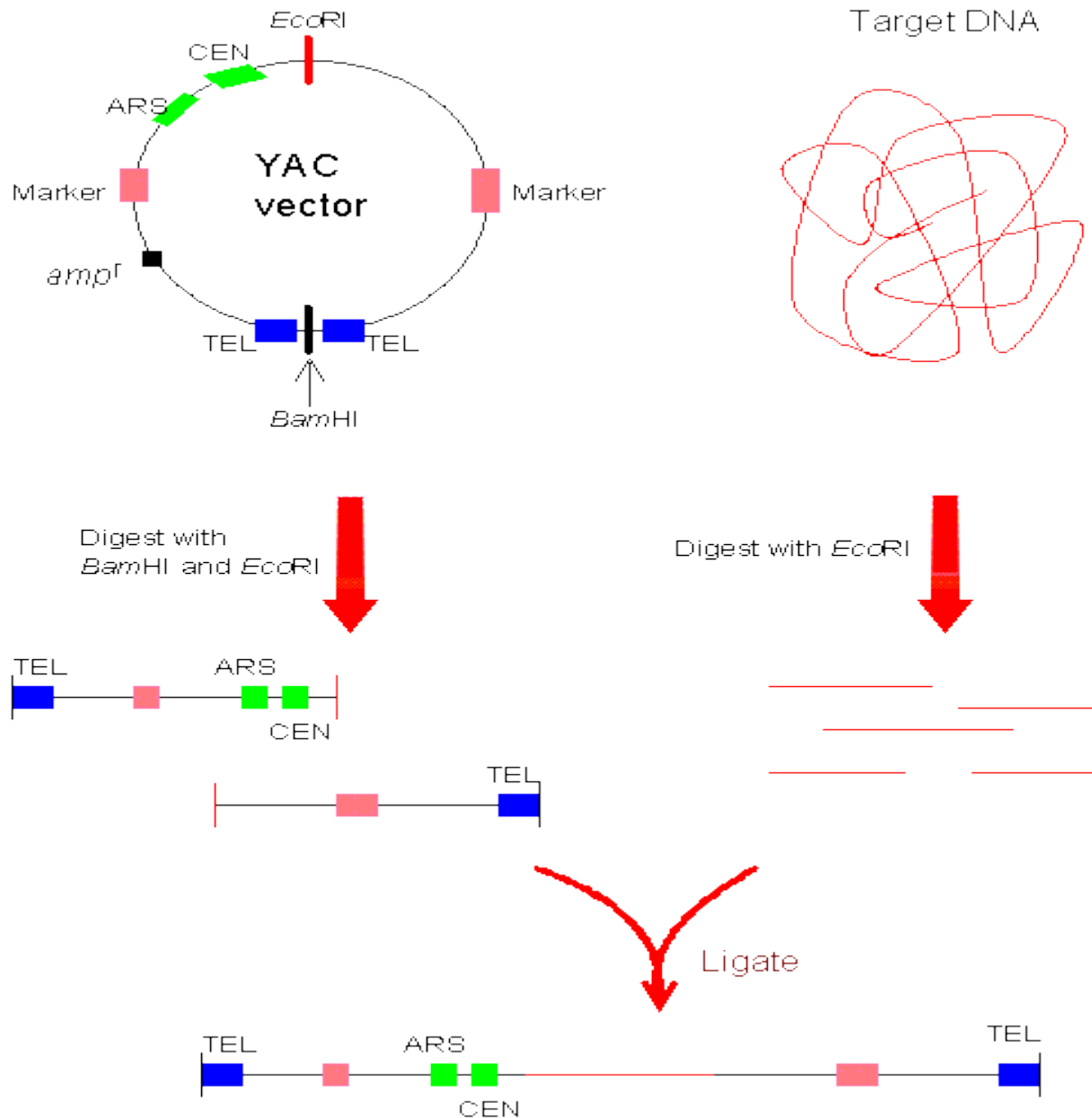
Fig 6.1 Cloning by using YAC vector.

Source: http://www.web-books.com/MoBio/Free/Ch9A4.htm

6.4.2 BACTERIAL ARTIFICIAL CHROMOSOMES (BACS):

Bacterial artificial chromosomes (BACs) cloning vectors are based on the fertility factor (F factor) and developed in early 1990s by Mel Simmons and coworkers. BACs plasmids are

constructed for the cloning very long DNA fragments (typically 100,000 to 300,000 bp). They carry origin of replication and specific marker like chloramphenicol (CmR) resistance gene. Bacterial Artificial Chromosomes were prepared during human genome project for preparation of large insert libraries having a cloning capacity of about 300 kbps. During the human genome project BACs got popularized due their high cloning capacity and easy maintenance. Usually BACs are maintained as single copy plasmids in per host cells and stable as such over hundreds of generations. For partitioning basically three genes are responsible, namely parA, parB and parC. The BACs can be maintained only in bacterial host deficient in homologous recombination (means it should have recA- gene)The total size of a BAC is about 7.4 kb and is introduced through electroporation in to the host cells. A single copy F-plasmid origin of replication (ori) is found in BAC.

Fig.6.2 BACs as a Cloning Vector
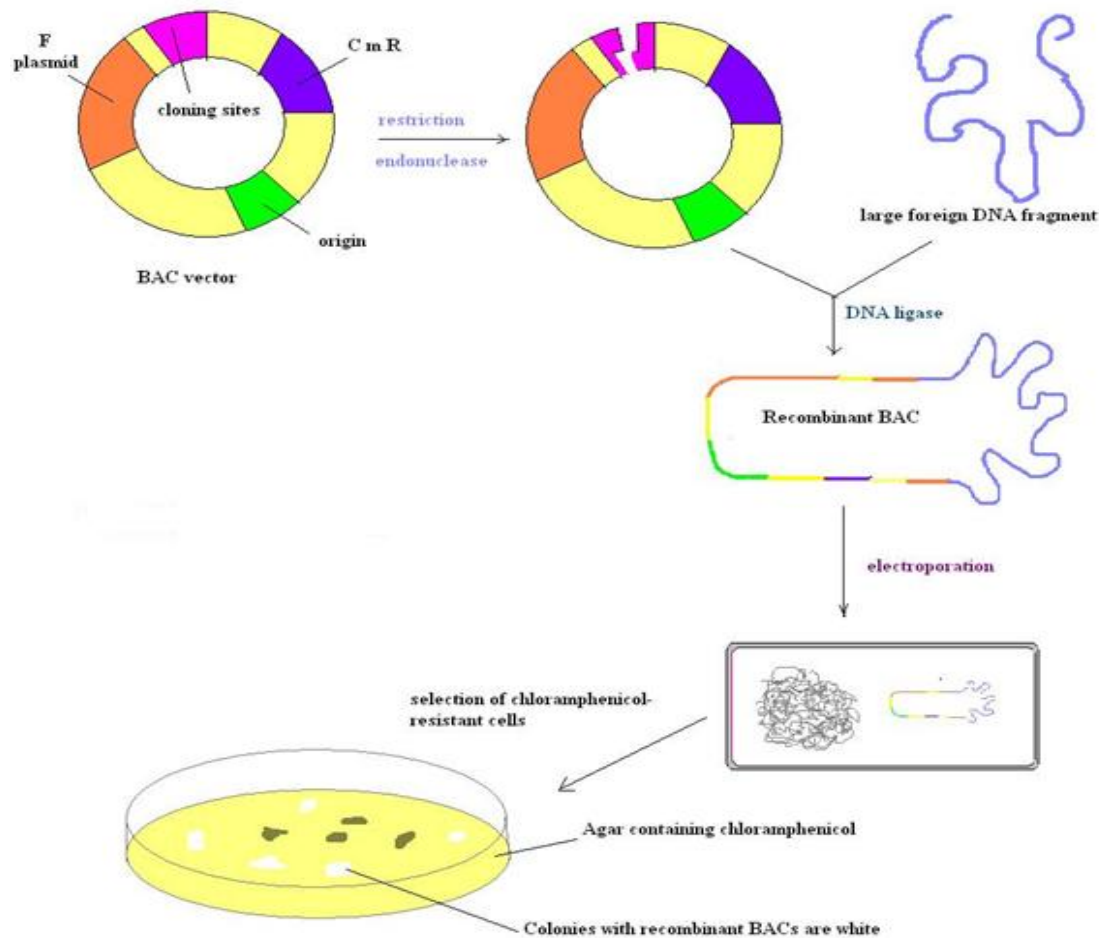
The F (fertility) plasmid is relatively large and vectors derived from it have a higher capacity than normal plasmid vectors. F-plasmid has F (fertility) factor which controls the replication and maintain low copy number. Also conjugation can take place between F+ bacteria (male) and F- bacteria (female) to transfer F-plasmid via pilus. Common gene components of a bacterial artificial chromosome are:

- oriS, repE – F for plasmid replication and regulation of copy number.
- parA and parB for maintaining low copy number and avoiding two F plasmids in a single cell during cell division.
- A selectable marker for antibiotic resistance; some BACs also have lacZ at the cloning site for blue/white selection.
- T7 and Sp6 phage promoters for transcription of inserted genes.

Bacterial Artificial Chromosomes are Bacterial cloning vector (derived from F plasmid) that can accommodate up to 350 kb (most commonly 120–150 kb) DNA sequences, and has a considerably lower error rate than the still larger capacity yeast artificial chromosome (YAC). BACs usually exist in a single copy per cell. Random BACs are selected at random from a genomic library and are then shotgun-sequenced. Most BAC vectors lack selectable markers suitable for mammalian cell selection but can be retrofitted by employing the Cre/loxP site specific recombination system.

For BACs to be used as cloning vehicle, the vector is linearized using restriction end nucleases, treated with phosphatase and then ligated with insert DNA fragment. After the recombinant DNA is produced it is introduced in E. coli host cells.

6.4.3 pBR 322

Plasmids are used as vectors to clone DNA in bacteria. One example of a plasmid used for DNA cloning is called pBR322 Plasmid. The pBR322 plasmid contains a gene that allows the bacteria to be resistant to the antibiotics tetracycline and amipicillin. To use pBR322 plasmid to clone a gene, a restriction endonuclease first cleaves the plasmid at a restriction site. PBR322 plasmid contains three restriction sites: PstI, SalI and ecoRI. The first two restriction sites are located

within the gene that codes for ampicillin and tetracycline resistance, respectively. Cleaving at either restriction site will inactivate their respective genes and antibiotic resistance. The target DNA is cleaved with a restriction endonuclease at the same restriction site. The target DNA is then annealed to the plasmid using DNA ligase. After the target DNA is incorporated into the plasmid, the host cell is grown in a environment containing ampicillin or tetracycline, depending on which gene was left active. Many copies of the target DNA is created once the host is able to replicate.
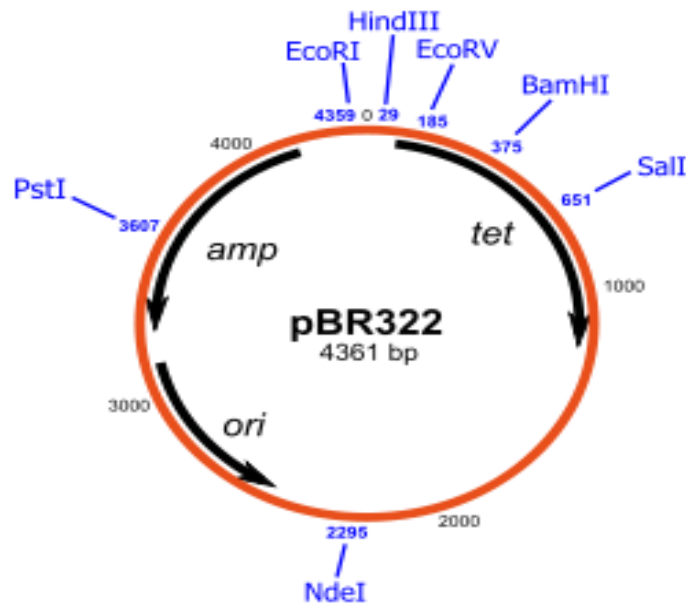
Fig.6.3 pBR322 Plasmid

pBR322 was one of the first plasmids used for the purpose of cloning. It contains genes for the resistance to tetracycline and ampicillin. Insertion of the DNA at specific restriction sites can inactivate the gene for tetracycline (an effect known as an insertional inactivation) or ampicillin resistance.

6.5 SUMMARY

Cloning in recombinant technology refers to the production of 'clones' or genetically identical organisms. Cloning vector is self replicating small piece of DNA of plasmid, virus, yeast or cell of higher organism into which gene of interest can be cloned and expressed. The most commonly

used cloning vectors are: Plasmids, Bacteriophages, Bacterial artificial chromosomes (BAC), Yeast artificial chromosome (YAC), Mammalian artificial chromosome (MAC). Molecular vectors are used for carrying cloned fragments of DNA. The vectors may be a small multi-copy plasmid or a designed and engineered virus. We also learned about various. Key steps for cloning in plasmid vectors, Bacteriophage λ and derived vectors, Insertion vectors, Replacement vectors, Lysogenic pathway, Lytic pathway, In-vitro packaging of λ DNA with the use of helper phage, and about various Bacteriophage λ derived cloning vectors. Plasmids are stably-inherited extra chromosomal, autonomously replicating; double stranded and covalently linked circular DNA molecules. To identify the clones containing the gene of interest screening is carried out. Selection of clones carrying gene of interest is done on the basis of marker genes and or a reporter gene present on the vector used in cloning.

REFERENCES

- Cooke H. 2001. Mammalian artificial chromosomes as vectors: progress and prospects; Cloning Stem Cells, 3(4): 243-249.

- Dente L., Cesareni G., Cortese R. 1983. pEMBL: a new family of single stranded plasmids; Nuc. Acid Res. 11 (6) 1645-1655.

- Hall B.G. 2004. Predicting the evolution of antibiotic resistance genes. Nat Rev Microb 2 (5): 430–435.

- http://bioinfo2010.wordpress.com/2009/07/08/vector-bacteriophage-lambda-and-m13-7thapril/

- Kim et al. 1992. Stable propagation of cosmid-sized human DNA inserts in an F-factor based vector. Nucleic Acids Res. 20 (5): 1083–1085.

- Kim et al. 1995. Construction and utility of a human chromosome 22-specific Fosmid library; Genetic Analysis: Biomol Eng 12 (2): 81–84.

- Brown T.A. 2010. Gene cloning and DNA analysis: an introduction (6th edition); Willey Blackwell Ltd.

- https://epgp.inflibnet.ac.in/Home/ViewSubject?catid=t5vt4STquHRj94mcOBMr5g==

- Clark, David P., and Nanette J. Pazdernik. Molecular biology. Elsevier, 2012.

- Griffiths A.J.F, Miller J.H., Suzuki D.T, Lewontin R.C., Gelbert W.M. 2000. An Introduction to Genetic Analysis (7th edition); New York: W. H. Freeman.
- Robertis E.D.P.De, Robertis E.M.F. De. 2010. Cell and Molecular biology (8th edition); Lippincott Williams and Wilkins.

6.6 TERMINAL QUESTIONS AND ANSWERS

Question No.1 Explain the Plasmid Biology in detail?

Question No.2 what do you understand by the Cloning Vectors?

Question No.3 Write a short note on *pBR322 Plasmid?*

Question No.4 Explain in detail the Yeast Artificial Chromosomes?

Question No.5 Explain in detail the Bacterial Artificial Chromosomes?

Unit 7 ANIMAL BIOTECHNOLOGY AND ITS APPLICATION

CONTENTS

7.1 *OBJECTIVES*

In this unit students will learn:

- Details of the cell, organ and whole embryo culture
- The mechanism of In vitro fertilization.
- About the transgenic animal and gene therapy

7.2 *INTRODUCTION*

Biotechnology is technology based on biology and it utilizes the scientific and engineering principles to produce products by biologic agents to provide goods and services. Biotechnology offers new tools for improving health and productivity of organisms. It helps in detection, treatment and prevention of diseases. Biotechnology helps in various sectors of agriculture by improving overall health, breed and productivity of livestock. Biotechnology helps in improving food quality by introducing desirable traits through new genes into farm livestock and poultry. Biotechnology helps in the improvement of Farm animals and their feeds which results in reduction of animal wastes and minimizing the impact on the environment. Artificial insemination, embryo transfer, in vitro fertilization, genetic mapping and cloning techniques help in breeding programs. As a result of high demand for meat and the degradation of agricultural land, biotechnology is providing new methods to improve productivity in animal agriculture. Genetically engineered poultry, swine, goats, cattle, and other livestock also are starting to be used as 1) producers of pharmaceutical and other products 2) potential sources for replacement organs for humans and 3) models for human disease. Animal-made pharmaceuticals (AMPs) change biotech animals into "factories" to yield therapeutic proteins in their milk, eggs, and blood. Biotechnology can be used to produce human-compatible transplant organs, tissues and cells in pigs that can be vital to improving human health. Various methods are utilized to produce transgenic animals. Reproductive and cloning techniques provide the possibility of preserving the genetics of endangered species. The development in the field of biotechnology has so many advantages but the advancement in the field raises various concerns also. The biological

technologies and other advances are yet to be considered and their use in the field of agricultural and biomedical work causes concerns regarding the safety of end products consumption, its undesirable effects on the environment, and its adverse effects on animal welfare.

7.3 CELL, ORGAN AND WHOLE EMBRYO CULTURE

CELL CULTURE

Cell culture refers to laboratory methods that enable the growth of cells in physiological conditions in vitro. The cell culture technique originated in the 20th century and it was used to 1) study growth and maturation of tissue 2) particular gene role in disease and metabolism 3) for large-scale production of biopharmaceuticals by hybrid cell lines and 4) studies of viruses and vaccine development. There are various experimental applications of cultured cells e.g. creating model systems to study basic cell biology, replicate disease mechanisms, or investigate the toxicity of novel drug compounds. Cell culture helps in manipulating genes and molecular pathways for research purposes. Cell culture leads to production of clone cell populations or specific cell types and well-defined culture systems which help in removing interfering genetic or environmental variables, this help in generation of high reproducible and consistent data that cannot be produced with whole organ systems study.

Safety protocol in Cell Culture Laboratory

Lab coats, gloves, and goggles should be worn to ensure protection of laboratory workers from potentially hazardous sources. When working with cell lines and biohazardous agents, personal protective equipment (PPE) must be worn in the cell culture lab. Culture work should be done inside biosafety cabinets which provide a steady, unidirectional flow of HEPA-filtered air and create an enclosed, ventilated workspace. Researchers should read the Material Safety Data Sheet (MSDS) of a regent before using it. Researchers should be well trained in handling various instruments and procedures.

Equipment used for cell culture

Numbers of equipment are required for the cell culture experiment and they provide desired conditions to carry out the cell culture experiment. The cell culture process required sterile conditions, so a separate space should be dedicated to this work. Various equipment can aid in achieving such a sterile workspace.

1) Biosafety cabinet:- To provide sterile work surface

2) Humid CO_2 incubator:- To create physiological environment for cellular growth

3) Inverted light microscope:- To evaluate cell morphology and count cells

4) Fridge, freezers ($-20°C$, $-80°C$), liquid nitrogen storage :- To store cells, cell material, and culture components

5) Centrifuge :- Help in cell condensation

6) pH meter:- To determine the correct pH of media components

7) Pipettes and pipettors:- To aliquot different volumes

8) Cell media and supplementary components:- To culture cells in desirable components

9) Hemocytometer:- To count cells, determine growth kinetics and prepare suitable plating densities

10) Autoclave:- To sterilize pipettes and other equipment in contact with cells

11) Vacuum pump:- To aspirate cell culture medium

12) Water bath: - To provide desired temperature.

13) Cell culture dishes:- To culture cells in different formats

14) Containers for waste :-To correctly dispose of waste

Aseptic Cell Culture procedure

Any kind of contamination can result in damage to cell culture and abnormal results leading to wrong scientific interpretations. Various precautions are taken to reduce the contamination of cell culture.

1) The air is full with microparticles which have the ability to cause infection, So all work should be done in a biosafety which restricts nonsterile aerosols and airborne components from contaminating cultured cells.

2) The biosafety cabinet should be kept in that part of the laboratory that does not obstruct its airflow by external sources of wind.

3) Biosafety cabinets should be sterile using UV light and the work surface should be decontaminated with an antifungal detergent (e.g., 5% Trigene) followed by 70% ethanol.

4) All equipment used inside the biosafety cabinet should be sprayed and wiped with 70% ethanol.

5) Minimum number of items should be kept inside the biosafety cabinet,

6) All instruments (incubator, centrifuge, microscope, water bath, fridge, and freezer) and equipment should be kept clean and free of dust. All instruments should be properly maintained.

7) Laboratory should be properly clean.

8) Treatment solutions can be added to water baths to prevent the growth of microbes.

9) Disposable gloves should be worn during experiments which are sprayed with 70% ethanol. And gloves should be discarded after the experiment.

10) Clean lab coats should be worn during experiments and should be clean at hot temperatures on a regular basis.

Reagents and Media for Cell Culture

The commercially available media and supplementary cell culture products are generally supplied in sterile condition. In the laboratory the material used for the culture can be autoclaved. Use of antibiotics (e.g., Penicillin/Streptomycin) help in controlling bacterial growth in media bottles after opening and in cell culture vessels.

Contaminations

The contamination cannot be altogether prevented, so it is important to recognize early signs of contamination in order to prevent it. Contaminants are generally due to biological agent bacteria, fungi, viruses, and parasites. The biological contaminants should be avoided because they can change the phenotype and genotype of the cultured cell line by competing for 1) nutrients 2) synthesis of alkaline and acidic or toxic by-products and 3) the potential interference of viral components with the cell culture genome. Other contaminants are chemicals (e.g., plasticizers in cell culture vessels) or other cell types co-cultured in the lab.

Bacterial Contamination

The bacteria have a fast doubling rate and this not only leads to easy detection in cell culture supernatants shortly after infection, but also facilitates quick spread. Cell cultures contaminated by bacteria generally appear turbid in appearance. The high metabolic rates of bacteria can also alter the pH of the culture media and thus change the color of phenol red to yellow. Bacteria can

be detected by using microscopic techniques. Mycoplasma infections are hard to detect so it is advisable to routinely test cultures for their presence using polymerase chain reaction (PCR), enzyme-linked immunosorbent assay (ELISA), or immunostaining .

Fungal Contamination

Yeast's cell can easily be detected at low microscope magnifications. The supernatants of yeast contaminated cell culture appear turbid and have a distinct smell. The pH of the culture remains stable during the initial stages of infection but it increases in high contaminant concentrations.

Viral Contamination

Viruses are small obligatory parasites and they are not visible in generic light microscopy. Detection of viral contamination is hard. Some viruses may cause morphological changes in the cultured cells and other species may integrate into the cellular genome and alter the phenotype of the investigated cell line. Virus's contamination in cell cultures occurs due to the use of animal-derived cell culture products such as trypsin or fetal bovine serum. The presence of viral contaminants can be detected by PCR, ELISA, immunocytochemistry, or electron microscopy.

Removing contaminants

1) To eliminate the contamination regardless of the type of contamination identified, affected cell cultures should be removed from the cell culture room and discarded to prevent the spread of infectious agents to other cultures.

2) The source of contamination should be identified.

3) The culture media and other cell culture components that have been in contact with the contaminated cells should be disposed

4) The surfaces should be clean that come in contact with the contaminated vessel (e.g., incubator, biosafety cabinet, microscope, aspirator).

The Cell Line

Cells cultured can be classified into three different types: primary cells, transformed cells, and self-renewing cells.

Primary cells can be obtained directly from human tissue or fibroblasts can be isolated from skin biopsies and hepatocytes can be obtained from liver explants. These types of cells can be used in Biomedical and translational research because they very well represent their tissue of origin. The

disadvantage associated with them is related to biosafety restrictions associated with their handling and their finite life.

Transformed cells can be produced either naturally or by genetic manipulation. They can be used to provide fast growth rates and stable conditions for maintenance and cloning. But their manipulated genotype may cause karyotypic abnormalities and non physiological phenotypes e.g., Chinese hamster ovary (CHO), HeLa, human umbilical vein endothelial cells (HUVEC)

Self-renewing cells have the ability to differentiate into a diversity of other cells types, while their self-renewing property allows for long-term maintenance in vitro e.g. embryonic stem cells, induced pluripotent stem cells, neural and intestinal stem cells. Cell lines can be obtained commercially.

To grow cells in vitro in a culture medium, it is necessary to provide a proper microenvironment. Cells can either grow in a suspension culture or attach to substratum. Cells grown in suspension generally adopt spherical shapes, while adherent cells display spiked or polygonal morphologies.

The Cell Culture Medium

To grow a cell in vitro a proper cell culture medium is required. Medium contains nutrients (such as carbohydrates, vitamins, amino acids, minerals) growth factors and hormones. Cell culture medium should have a proper pH and cellular osmotic pressure. To provide physiologically relevant microenvironment a solid or semisolid growth substrate is used which allow cell–matrix anchoring and cell–cell interactions respectively. Numbers of cell culture medium compositions have been synthesized according to the requirements of specific cell types. Serum (fetal bovine serum) is added to basal media that already contain a standard formulation based on amino acids, vitamins, carbon sources (e.g., glucose), and inorganic salts. Serum provides 1) growth factors and hormones 2) acts as a carrier for lipids and enzymes and 3) helps in the transportation of micronutrients and trace elements.

Temperature, pH, CO_2, and O_2 Levels

The cell culture depends on the temperature and the desired temperature for culture is the body temperature of the species and the microenvironment from which the cultured cell types were taken. For human and mammalian cell lines the incubated temperature is 36–37°C, but for cell lines of cold-blooded animals a wider temperature is used (ranges between 15°C and 26°C).

Maintaining pH level is also important and for most human and mammalian cell lines pH should be tightly controlled and kept at a physiological pH level of 7.2–7.4. Cell culture temperature is maintained by incubators. Glucose fulfills the cell's energy needs and its metabolism releases pyruvic acid, lactic acid, and CO_2. Since the pH level is dependent on the balance of CO_2 and HCO_3^- (bicarbonate), the addition of bicarbonate-based buffers to cell culture media can equilibrate the CO_2 concentrations. Variations in atmospheric CO_2 concentrations can also change the pH level. Cells should therefore be cultured in incubators which regulate CO_2 tensions to be adjusted to 5–7%.

Subculturing

As the time passes, the number of cells increases in the cell culture medium with depletion of nutrients, and release of the toxic metabolites. So, to expand and maintain a healthy cell culture, it is necessary to do a new culture with a subset of cells from the originating culture, removing toxic by-products, and replenishing nutrients with fresh medium. So, when the available space in the cell culture vessel reaches ~80% confluency, then the cell should be transferred to a new vessel. This process is called "passaging". To subculture cells, cultured cells are enzymatically digested or mechanically dissociated to remove them from their substrate. Then cells are washed with phosphate-buffered saline (PBS) with no Mg^{2+} and Ca^{2+} to remove dead cells and are incubated at 37°C with sufficient digestive enzymes or chelating agent to cover the monolayer (e.g., trypsin, dispase, collagenase, ethylenediaminetetraacetic acid (EDTA)). The dissociated cells are collected in a sterile Falcon tube and the collected cells seeded in new culture vessels at the desired concentrations. The growth of adherent cells depends on available surface area, but in suspension culture the rate-limiting step is the concentration of cells in the medium. So, it is necessary to monitor growth rates in suspension cultures over time.

The subculturing of suspension cultures can be done by aseptically removing one-third of the cell suspension solution and replacing the volume with prewarmed complete medium. The cells need to be concentrated for transfer to new cell culture vessels, freezing, or other experimental assays. For that the cell must be suspension is centrifuged at 300 ×g for 10 minutes. After discarding the supernatant, the cell pellet is resuspended in the desired medium through gently pipetting cells up and down three times. The cells are fragile so centrifuge should not be done at high speed or vigorous pipetting should be avoided.

Quantification of Cells and Determination of Cell Viability

During the process of cell culturing, handling and passing cells can die, but a specific amount of cells are required to start culture and perform assay, so it is important to distinguish between live and dead cells. Cell counting is also useful for assessing growth rates. Cells are generally cultured in the millions; the number of cells are first counted in a small volume and then extrapolated to the full cell volume.

Cryopreservation of Cells

The cell can be preserved by freezing with cryoprotective agents (e.g., glycerol or dimethyl sulfoxide (DMSO)) that can prevent the formation of harmful extra- or intracellular crystals. DMSO is toxic to personnel and cultured cells and therefore cannot be added to cells without prior dilution. Chemically protective gloves should be worn to safeguard personnel from the hazards of DMSO as its solutes can easily penetrate membranes, including the skin. Cells can be preserved in liquid nitrogen (<130°C) for years since all life processes cease at these temperatures.

Thawing Cryopreserved Cells

To recover cells, 10 ml of complete medium is prewarmed in a water bath. Then the frozen vial is immediately placed into a 37°C water bath until two-thirds of the contents are completely thawed. The vial is wiped with 70% ethanol and placed in a biosafety cabinet. Then prewarmed medium (1ml) is added dropwise to the partially thawed vial to minimize the osmotic stress imposed upon the cells when DMSO is diluted. The contents of completely thawed vial is added to the remaining 9 ml of complete medium and centrifuged at $300 \times g$ for 3 minutes. After that the supernatant is discarded and the cell pellet can be washed once in medium to remove residual cryopreservatives. Cells are then resuspended in complete medium and transferred to a cell culture vessel. Cell attachment should occur within 24 hours.

Application of cell culture

1) Model Systems to study Health and Disease: Cell culture technique helps to study the various cellular functions of wild-type cells and diseased cells. The co-cultures can be used to study interaction and route of infection between wild-type cells and pathogenic agents. Cancer cell lines are used to study mechanisms of cancer induction and its

treatment. Human-induced pluripotent stem cells (hIPSCs) help in studying various molecular mechanisms of a disease.

2) Helps in Drug Development and Drug Testing: Cell culture technique is used as a tool that can screen novel chemicals, cosmetics, and drug compounds for their effectiveness and evaluate drug cytotoxicity in specific cell types.

3) Production of biological important components: cell cultures can be used for the production of genetically engineered proteins, antibodies, hormones and biopharmaceuticals that can be isolated and used therapeutically.

4) Virology and Vaccine Production: Cell culture of mammalian cells provide a host for viruses to replicate which allow scientists to study their growth rates, development, and conditions required for their infectious cycle. Also, the attenuated viruses used in vaccines against polio, measles, chicken pox, rabies, and hepatitis B are produced in animal cell cultures.

5) Tissue Regeneration and Transplantation: hIPSCs, embryonic stem cells, and adult stem cells have the ability to regenerate and differentiate into specialized cell types that can help in tissues or organs replacement.

6) Genetic Engineering and Gene Therapy: cell culture helps in introduction of new genetic material into the nucleus of cultured mammalian cells to study expression of specific genes and their impact on cells.

ORGAN CULTURE

Cell culture has few limitations like it doesn't represent the in vivo responses due to lack of an appropriate micro environmental context of the responding cell types. So, organ culture provides an opportunity to better replicate the tissue microenvironment.

Procedure

1) Tissue for human organ culture (HOC) is taken to the laboratory as quickly as possible to minimize deterioration, preferably within minutes of collection.

2) Then the excess blood is removed from the sample by immersing tissue in sterile phosphate-buffered saline prior to transfer into a sterile petri-dish for dissection into <1 mm^3 fragments using two sterilized carbon steel single-edged razor blades stuck

together under a dissecting microscope). Multiple samples can be collected from various regions of a tissue e.g. kidney samples can be taken from the medulla through to the cortex.

3) The dissected fragments are immediately immersed in tissue culture medium in a sterile plate containing culture inserts, treated as indicated, or left untreated and maintained in a 37°C incubator for the desired time periods.

4) Culture conditions may need to be changed depending upon the tissue source.

Application of Organ culture

1) It can be used to study Inflammatory Diseases
2) It can be used to study Systemic Diseases
3) It can be used to reveal altered response and signaling pathways in disease
4) It can be used to study Organ Transplantation
5) It can be used to study Cancer
6) It can be used to study Therapeutics

EMBRYO CULTURE

Embryo culture is a procedure for the cultivating of an embryo under aseptic conditions on a culture medium. The aim of this technique is to improve the quality of embryos developed in the laboratory. The embryos are cultured for 2 to 3 days to reach the four-to-eight cell stage before transfer to the uterus. Premature replacement of the human embryo to the uterus may lead to low implantation rates associated with human In vitro fertilization (IVF). The preimplantation embryo has specific needs depending on the developmental stage and the improvements in culture media are due to better understanding of the environment of the oviduct and uterus. A culture medium is an external environment for the human embryo. The component of cultured media is selected carefully in proper concentration So, as to minimize the stress on cultured embryos. The better understanding of both the physiological changes in oviduct and uterus and the different metabolic requirements of the cleavage-stage and blastocyst-stage embryo led to the concept of stage-specific or "sequential" complex media.

Generally the culture media contain following components

1) Carbohydrates: Zygotes and subsequent cleavage stages favor pyruvate as the primary source of energy, while the eight-cell-stage embryo uses glucose. Glucose is the key anabolic precursor and it is involved in the synthesis of triacylglycerols and phospholipids, and acts as a precursor for complex sugars and glycoproteins. Glucose is also utilized by the pentose phosphate pathway (PPP) to generate ribose moieties required for nucleic acid production.

2) Amino acids: Amino acids are essential for the development and they regulate the mammalian preimplantation. Before the embryonic genome expression, the embryo uses carboxylic acids and AA as energy sources. Certain AA also functions as biosynthetic precursor molecules, osmolytes, buffers of internal pH, antioxidants and chelators, especially for heavy metals. The nitrogen requirement of embryos specifically changes with different requirements. The seven non-essential AA and glutamine stimulate the development of the early cleavage embryo. However, the presences of 13 essential AAs have an inhibitory effect on blastocyst development and viability at an early stage. Amino acids in culture media spontaneously undergo breakdown and produce ammonium into the culture medium. Ammonium is toxic to the embryo and culture media protocols are used to remove the accumulated ammonium.

3) EDTA (Ethylenediaminetetraacetic acid): It is a ligand and chelating agent (ability to "sequester" metal ions). Metal ions after bonding with EDTA remain in solution but exhibit diminished reactivity.

4) Regulation of cell volume-osmolytes: It is necessary to maintain the specific osmotic pressure in the cell. The osmotic pressure of oviduct fluid is >360 mOsmol. Addition of extracellular organic osmolytes, such as glycine, betaine, proline, alanine and hypotaurine protects the preimplantation embryo against hypertonicity and increases embryo development.

5) Impact of pH and buffers: The pH range for embryo culture media is between pH 7.4 and 7.2. Culture media pH is regulated by the balance of CO_2 concentration, and by the concentration of bicarbonate in the media. The intracellular pH in human

cleavage embryonic cells is pH=7.2. pH plays an important role in maintaining intracellular homeostasis. The most commonly used buffer is 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

6) Macromolecules: In culture media sources of macromolecules are proteins e.g. human serum albumin or synthetic serum. Both are added at concentrations of 5 to 20%. Albumin helps in maintaining the stability of cell membranes and chelate trace amounts of toxic components presented in culture water, media components and culture dishes. Other functions are capillary membrane permeability and osmoregulation. Physiological alternative to albumin is the glycosaminoglycan hyaluronate (also called hyaluronic acid or hyaluronan).

7) Vitamins: They work as antioxidants in culture media containing glucose and phosphate and also help in prevention of loss during respiration and metabolic control. Examples of vitamins used in culture media are ascorbic acid, cyanocobalamin, folic acid and tocopherol.

8) Growth factors: Growth factors play an important role in growth and differentiation from the time of morula to blastocyst transition.

9) Antibiotics: Embryo culture media are routinely supplemented with antibiotics to prevent bacterial contamination e.g. penicillin, streptomycin and gentamycin.

7.4 IN VITRO FERTILIZATION (IVF) TECHNOLOGY

The modern technologies provide the possibility of freezing the embryos and semen for their long term preservation. In the laboratory a needle is used to aspirate immature oocytes from the ovaries and the oocytes are placed in a culture containing hormones for one day to mature. When the oocytes reach a point midway through the second division of meiosis, they are fertilized with live sperm. In rare cases, single sperm or sperm head is used for fertilization, which is injected through the tough outer zona pellucida of the oocyte, either below the zona or directly into the cytoplasm (intracytoplasmic injection, or ICSI). The zygotes result after fertilization is cultured until the embryo reaches a more advanced stage of development. In humans, of course, these combined techniques form the basis of in vitro fertilization procedures. In vitro fertilization (IVF) is a complex series of events in which mature eggs are collected (retrieved) from ovaries

and fertilized by sperm in a lab. Then the embryos are transferred to the uterus of a female. IVF is an assisted reproductive technology and done using a couple's own eggs and sperm or from donor. The success of this technique and the birth of a healthy baby depend on many factors, such as your age and the cause of infertility. IVF procedure is required when infertility occurs for any reasons like Fallopian tube damage or blockage, Ovulation disorders etc. The risk factors associated with IVF are Multiple births, Premature delivery and low birth weight, Ovarian hyperstimulation syndrome (due to use of fertility drugs during IVF) , Miscarriage, Egg-retrieval procedure complications, Ectopic pregnancy, Birth defects. This technique also helps in production of embryos for experimental purposes in agricultural research. During 1980 the embryos were bisected in order to provide zygotic twins (genetically identical in terms of both their nuclear and mitochondrial genes). These are then placed in an empty zona before being transferred to different recipient mothers to carry them to term.

7.4.1 DOLLY, EMBRYO TRANSFER IN HUMAN

DOLLY

The female Finn Dorset sheep named Dolly is the first clone (February 1997) of an adult mammal, produced by British developmental biologist Ian Wilmut and colleagues of the Roslin Institute, near Edinburgh, Scotland. The clones had been produced earlier in the laboratory, but they are derived from embryonic cells which are either undifferentiated or only partially differentiated. The clone of lower species like frogs has been produced previously from fully differentiated (adult) cells (e.g., skin or muscle cells).

The clone Dolly is produced from the mammary gland cell taken from an adult Finn Dorset ewe. Wilmut et al produce Dolly by using electrical pulses to fuse the mammary cell with an unfertilized egg cell, the nucleus of which had been removed. The fusion procedure caused the transfer of the mammary cell nucleus into the egg cell, which then began to divide. Derived from mammary cell nuclei of Finn Dorset ewe and host egg cytoplasm from Scottish Blackface ewes, a number of fused couplets successfully formed embryos. Then the embryos were transferred to surrogate Scottish Blackface ewes. Out of 13 recipient ewes, one became pregnant, and after 148 days of gestation period, Dolly was born. Dolly remained alive and well long after her birth and her body organs function properly. The technique which produces Dolly is now known as

somatic cell nuclear transfer (SCNT). This technique is now used to produce a variety of mammalian clones from different types of adult cells. Dolly suffered from progressive lung disease and on February 14, 2003 she was euthanized by veterinarians. Her body was preserved and displayed at the National Museum of Scotland in Edinburgh.

HUMAN EMBRYO TRANSFER

Embryo transfer is a process of transplantation of a mammalian preimplantation embryo into the reproductive tract of a recipient female so that it may implant and continue to develop to birth. Mammalian embryos of many species are able to develop in vitro from fertilization to the blastocyst stage, but at this stage, they must be implanted in the uterus in order for embryogenesis to proceed normally. The first successful embryo transfer was performed in 1890 in the rabbit. Since then various improvements have been done to improve this technique. In 1978, the first birth of a human occurred from a transferred embryo. Embryo transfers are required when natural fertilization is not an option or has difficulty occurring due to ovulation disorders, damage to fallopian tubes, endometriosis, premature ovarian failure, uterine fibroids, genetic disorders, impaired sperm production, etc.

Embryo Transfer is of various types:

1) Fresh embryo transfer: The fertilized eggs are cultured for 1-2 days. The best embryos are chosen to transfer directly to the woman's uterus.

2) Frozen embryo transfer: Any healthy embryos that were not used in the first transfer can be frozen and stored for future use. These can be thawed and transferred to the uterus.

3) Blastocyst embryo transfer: If many healthy embryos develop after fertilization, it is common to wait to see if the embryos develop into blastocysts. And then transfer to the woman's uterus.

4) Assisted hatching (AH) before Transfer: Assisted hatching is done in the laboratory just before embryos are transferred into the woman's uterus. Using a laser, the embryologist makes a tiny, precise opening in the shell of each embryo after which it is transferred.

Process of Embryo Transfer

1) An embryo transfer is the last step of the in vitro fertilization (IVF) process.

2) During IVF, the ovaries are stimulated to release eggs by using fertility medications

3) These eggs are then removed from a woman's ovaries and fertilized in a lab.

4) After the fertilized eggs have multiplied, the best embryo is chosen and transferred to the womb by placing it in a soft catheter and injected in the uterine cavity through the cervix.

5) Pregnancy occurs when the embryo attaches itself to the wall uterus (implantation).

The complication associated with embryo transfer are increased hormonal stimulation (resulting in increased risk of blood clot blocking a blood vessel), bleeding, changes in vaginal discharge, infections, and complications of anesthesia if it is used. The greatest risk of embryo transfer is the chance of multiple pregnancies results due to multiple embryos attachment to the uterus. This may increase the risk of stillbirth and children born with disabilities.

Importance of Embryo Transfer Technique

- Embryo transfer helps in experiments involving removal of the developing embryo from the reproductive tract of one female animal and transfer to another.

- This technique helps in production of transgenic animals, knockout mice, and pathogen-free colonies.

- This technique helps in experiments involving separation of maternal and fetal genetic effects.

7.5 TRANSGENIC ANIMAL

Transgenic animals are created by deliberately inserting a gene into the genome of an animal. Numerous approaches are nowadays used for genetic engineering of various animal species. Various methods are used to access the germline of animals: (1) direct manipulation of the fertilized egg and then its implantation into the uterus; (2) manipulation of the sperm (3) manipulation of early embryonic tissue (4) the use of embryonic stem (ES) cell lines which, after manipulation and selection ex vivo, can then be introduced into early embryos and (5) manipulation of cultured somatic cells, whose nuclei is fused with enucleated oocytes to produce a whole animal (e.g. Dolly).

Generally two methods are used for inserting DNA into vertebrate germline cells, transfection and infection with retrovirus vectors. A third method based on the use of mobile genetic elements, has been commonly used for insects.

Transfection

Transfection process involve: (1) direct microinjection of DNA into the cell nucleus; (2) electroporation— method of DNA introduction through transient pores created by controlled electrical pulses; (3) use of polycations to neutralize charges on DNA and the cell surface that prevent efficient uptake of DNA; (4) lipofection, or enclosure of DNA in lipid vesicles that enter a cell by membrane fusion much in the manner of a virus, and (5) sperm-mediated transfection, possibly in conjunction with intracytoplasmic sperm injection (ICSI) or electroporation. The DNA structure introduced into a cell by any of the above methods is highly variable and uncertain. Generally, only a fragment of the transfected DNA is integrated into the chromosome, frequently in multiple copies, that often are integrated in long tandem arrays.

Retrovirus Vectors

Retroviruses are a class of viruses which replicate by a specific process involving copying of the viral RNA genome into DNA (reverse transcription) followed by its specific and stable introduction into host cell DNA (integration). Then the host transcriptional machinery is used to express the integrated DNA. Retroviruses are used to introduce genes of interest into cells in culture or into somatic tissue in experimental animals. They are also used for germline modification of fish, mollusks, chickens, mice and cattle. Retrovirus vectors are prepared using DNA constructs containing the gene of interest lined by sequences necessary for replication as a virus. These sequences include transcriptional promoters in the long terminal repeats (LTR's), which flank the integrated DNA, or provirus. Other necessary sequences are Signals necessary for packaging of the transcript in virions (virus particles), for reverse transcription, and for integration of the resulting DNA. These DNA constructs are introduced into cells which express viral proteins, but that are unable to make infectious viruses, leading to the creation of infectious virions containing an RNA copy of the gene of interest. When cells are infected with such virions, the RNA is copied into DNA and integrated at random sites in the cell genome. To select cells containing the desired virus construct selectable markers are included in the construct.

Transposons

Transposons are DNA elements that can transfer their information from one site to another in the same cell. A variety of transposons have been found in insects and fish and some are used as vectors for the generation of transgenic insects.

Directed genetic manipulation

Another goal of transgenic technology is to produce engineered animals that lack specific genes (knockout), or have these genes replaced by one that has been engineered in a specific way (knockin). For example, transplantation of organs or tissues from non-primates (pigs) to humans (xenotransplantation) is currently impossible, due to immune response by human recipients to a carbohydrate on the surface of pig cells (galactose-1,3-galactose). Inactivation of the enzyme (galactosyl transferase, GT) in donor pigs could solve this problem.

From the 1990s, attempts have been made to yield transgenic animals that synthesize a variety of human proteins. The coagulation factors VII, VIII and IX are used for lifelong treatment of hereditary diseases and can be produced in the milk of transgenic animals. An immune response to therapeutic agents develops in most patients over time, despite the highly efficient purification of the proteins produced in bacterial or yeast systems. Thus production of them by transgenic animals is very useful. The demand of monoclonal Antibodies (mAb) is high in the pharmaceutical industry and they are produced exclusively in mammalian cell cultures, since proper post-translational modifications are required to ensure therapeutic efficiency. So, transgenic animals used for the production of mAb will be very useful.

7.6 *HUMAN GENE THERAPY, CRYOBIOLOGY*

HUMAN GENE THERAPY

Gene therapy is a procedure in which an effective gene is inserted into a human cell to correct a genetic error or to introduce a new function to the cell. Gene therapy has become a reality due to advancement in the field of bioengineering that helps in manipulating vectors for delivery of extrachromosomal material to target cells. Numerous approaches, including retroviral vectors and non-viral vectors, have been developed for both ex vivo and in vivo transfer of genes into the cells. The main objective of gene therapy is optimization of delivery vehicles (vectors) that are

mostly plasmids, nanostructured or viruses. The viruses are used as a vector due to their quality to invading cells and inserting their genetic material into the host genome. But the problem with viral vectors is that it exacerbates immune responses and genome manipulation, especially in germ line cells. Gene therapy is a very complex process and requires new developments. It required identification and access to specific cells that needed to be treated. The method should efficiently distribute the gene copies to the cells. The diseases and their strict genetic bonds need to be totally known. The gene therapy is of two type 1) gene therapy of the germline and 2) gene therapy of somatic cells.

In germline gene therapy, the stem cells (sperm and egg) are altered by the introduction of functional genes, which are integrated into their genome. The germline modifications are hereditary and pass on to subsequent generations. Somatic cell gene therapy involves integration of therapeutic genes into the patient's somatic cells. Any change and any effects are limited only to that patient and are not inherited by future generations. Gene therapy involves the insertion of a normal gene into the genome to replace an abnormal gene responsible for producing a certain disease. To release the gene into the stem cell the molecular carrier called a "vector" is used. The vectors must be very specific, show efficacy in the release of one or more genes of the sizes required for clinical applications, not be acknowledged by the immune system, and can be purified in large amounts. After the it is inserted into the patient, it cannot induce an immune response; it should increase the normal functions, correct deficiencies, or inhibit deleterious activities. The vector should be safe for the environment, patients and for the professionals who manipulate it. Lastly, the vector should be able to express the gene, for the patient's entire life. There are numerous safety and ethical issues concerning manipulation of the human genome that need to be resolved. Gene therapy has potential for the effective treatment of genetic disorders.

CRYOBIOLOGY

Cryobiology is the branch of science that deals with the effects of freezing and low temperatures on living organisms". Various cryobiological techniques are used today e.g. cryosurgery, embryo and gamete preservation, tissue preservation and transplantation, blood and blood product preservation, and cryotransport. Cryosurgery is used for the treatment of many types of skin lesions (including cancer), for benign and dysplastic mucosal lesions, for uterine cervicitis and

intraepithelial neoplasia, and for cardiac surgery in the treatment of tachyarrhythmias. In this technique selective necrosis of tissues is done through freezing at very low temperatures. The low temperature is achieved by the use of a specific solution called cryogen. Cryosurgical procedures also involve the use of a dipstick apparatus, which is simply a cotton-tipped applicator. The dipstick is dipped into liquid nitrogen and applied to a lesion until adequate necrosis is achieved. Other methods involve the use of open-spray apparatus which help in spraying cryogen on the lesion and solidified carbon dioxide which can be directly applied to the skin with a mix of acetone to treat certain types of acne (slush therapy). Another technique used is cryoprobe in which a precooled metal accessory is applied directly to the lesion.

Cryosurgery has both advantages and disadvantages. The advantages of cryosurgery over other procedures are 1) fit for nursing home, or outpatient facility, 2) no need of general anesthesia, for local anesthesia is optional, 3) no operative suits required, the procedure is simple and safe, 4) no physical restrictions afterward and 5) availability for pregnant patients.

B. Rubinsky, et. al., of the University of California-Berkeley (1990) show negative impact of cryosurgery. They show that cryosurgery done on the liver results in formation of ice in the hepatic sinusoids, expands the sinusoids, dehydrates the hepatocytes, and causes structural damage.

Embryo and gamete cryopreservation

Cryopreservation technique employed since 1972 firstly used in case of mice then for goat (1976) and later for humans (1983). This technique results in successful birth of live young ones from frozen embryos. Cryopreservation techniques involve a number of steps which are specific and straightforward methods of freezing and are exclusive for each organism. The first step involves the exposure of the embryo to cryoprotective solution and then the embryo is cooled to slightly lower than zero degrees Celsius, followed by a seeding procedure. Seeding is a technique utilized for induction of ice formation under controlled conditions. After that, under controlled conditions of cooling the embryo is subjected to intermediate subzero temperature. The ideal temperature for cryopreservation is temperatures as low as -80 to -100 degrees Celsius and temperatures as high as -20 to -40 degrees Celsius. For long-term storage the freezing is done

very rapidly to -196 degrees Celsius. If the cryopreservation is done properly the embryo can survive for thousands of years. Whenever the cryopreserved embryos are used they should thaw properly. The warming and thawing depends on the amount of freezing that had initially been done. After that the cryoprotective solution is removed otherwise improper handling of this solution is toxic to the embryo. The cryoprotective solution is like body fluids of the womb, containing sodium ion- rich saline solution.

Tissue and Blood Cryopreservation

The body tissue and blood can also be preserved using cryopreservation techniques and can be reintroduced into the body whenever required. Previously transplants like born marrow relied on time to be effective, but now cells can be frozen and stored for years and still be effective. Under laboratory conditions the desired cell can be preserved for short-term in a controlled freezer that can cool at a rate of one degree per minute, or for long-term in a large, liquid nitrogen freezer at -180 to -195 degrees Celsius. Likewise blood and blood products can be stored and used whenever required. This technique can also be used for the preservation of placentas and umbilical cords, which can be used as blood and hematopoietic stem cell sources. Later on these preserved cells can be used for stem cell matching for a child, life-threatening diseases treatment, or gene therapy.

Although cryopreservation techniques are very useful, there are various reasons for its opposition. Firstly cryobiological procedure may affect the cells and the surrounding areas leading to damage. Cryosurgery can also cause damage and this damage can result in future problems that will need to be cured, generating a vicious cycle. During cryopreservation embryos or gametes can get damaged resulting in less viable births. The damage is due to poor technique, simply inactive embryos, or damage sustained by cryoprotective solutions or the freezer. The cryopreservation technique is very expensive for the average person to afford. Legal issues are also associated with embryo preservation. For example a couple decided to preserve their embryo but in future if they get separated then the custody issue arises "Who will have the legal right to embryo?"

7.7 SUMMARY

Biotechnology offers new tools for improving health and productivity of organisms. It helps in detection, treatment and prevention of diseases. Biotechnology helps in various sectors of agriculture by improving overall health, breed and productivity of livestock. Cell culture refers to laboratory methods that enable the growth of cells in physiological conditions in vitro. The cell culture technique originated in the 20th century and it was used to study growth and maturation of tissue, particular gene role in disease and metabolism, for large-scale production of biopharmaceuticals by hybrid cell lines, study of viruses and vaccine development. Cells cultured can be classified into three different types: primary cells, transformed cells, and self-renewing cells. Cell culture has few limitations like it doesn't represent the in vivo responses due to lack of an appropriate micro environmental context of the responding cell types. So, organ culture provides an opportunity to better replicate the tissue microenvironment. Embryo culture is a procedure for the cultivating of an embryo under aseptic conditions on a culture medium. The aim of embryo culture technique is to improve the quality of embryos developed in the laboratory condition. The embryos are cultured after 2 to 3 days to reach the four-to-eight cell stage before transfer to the uterus. Embryo transfer is a process of transplantation of a mammalian preimplantation embryo into the reproductive tract of a recipient female so that it may implant and continue to develop to birth. In vitro fertilization (IVF) is a complex series of events in which mature eggs are collected (retrieved) from ovaries and fertilized by sperm in a lab. The female Finn Dorset sheep named Dolly is the first clone (February 1997) of an adult mammal, produced by British developmental biologist Ian Wilmut and colleagues of the Roslin Institute, near Edinburgh, Scotland. Transgenic animals are created by deliberately inserting a gene into the genome of an animal. Gene therapy is a procedure in which an effective gene is inserted into a human cell to correct a genetic error or to introduce a new function to the cell. Gene therapy has become a reality due to advancement in the field of bioengineering that helps in manipulating vectors for delivery of extrachromosomal material to target cells Cryobiology is the branch of science that deals with the effects of freezing and low temperatures on living organisms. Various cryobiological techniques are used today e.g. cryosurgery, embryo and gamete preservation, tissue preservation and transplantation, blood and blood product preservation, and cryotransport.

7.8 TERMINAL QUESTION AND ANSWERS

1. Study of preservation of life at low temperature is

A) Malacology

B) Icebiology

C) Exobiology

D) Cryobiology

2. What does IVF stand For?

A) In vivo fertilization

B) In vitro fermentation

C) In vivo fermentation

D) In vitro fertilization

3. Animals that have had their DNA manipulated to possess and express an extra (foreign) gene are known as _____

A) Transgenic animals

B) Modified animals

C) Infected animals

D) Bt animals

4. The sheep "Dolly" was cloned by using somatic cell from the donor's

A) Mammary gland cell

B) Liver

C) Kidney

D) Eye

5. Disaggregating of cell can be done using

A) Physical disruption

B) Enzymatic digestion

C) Treating with chelating agents

D) All of the above

Answers: 1(D), 2(D), 3(A), 4(A),5(D)

Q 6. Give detail of cell culture technique.

Q 7.Write a short note on

    a) Cryobiology

    b) Embryo culture

    c) Human gene therapy

Q 8. Explain the process of in vitro fertilization.

7.10 *REFERENCES*

1. Cell culture Growing cells as model systems In Vitro Charis, P Segeritz and Ludovic Vallier. Basic Science Methods For Clinical Researchers. 2017:151-172

2. Human Organ Culture: Updating the Approach to Bridge the Gap from In Vitro to In Vivo in Inflammation, Cancer, and Stem Cell Biology, Rafia S. Al-Lamki, John R. Bradley, and Jordan S., Pober Front Med (Lausanne). 2017; 4: 148.

3. Early Embryo Rescue Renee Drewes-Alvarez, in Reference Module in Life Sciences, J Turk, Ger Gynecol Assoc. 2011; 12(2): 110–117.

4. Embryo culture media for human IVF: which possibilities exist?, Irmhild Gruber and Matthias KleinJ Turk, Ger Gynecol Assoc. 2011; 12(2): 110–117

5. https://www.britannica.com/topic/Dolly-cloned-sheep

6. https://thebiologynotes.com/human-embryo-transfer December 30, 2019 by Yashaswi Sharma

7. Animal Biotechnology: Science-Based Concerns Washington (DC): National Academies Press (US); 2002. ISBN-10: 0-309-08439-3

8. Use of Transgenic Animals in Biotechnology: Prospects and Problems, O. G. Maksimenko, A.V. Deykin, Yu. M. Khodarovich, and P. G. Georgiev, Acta Naturae. 2013 Jan-Mar; 5(1): 33–46

9. Gene therapy: advances, challenges and perspectives, Giulliana Augusta Rangel Gonçalves and Raquel de Melo Alves Paiva, Einstein (Sao Paulo). 2017 Jul-Sep; 15(3): 369–375

10. Cryobiology: Low temperature studies of biological systems By Krista Nussbaum https://www.goshen.edu/bio/Biol410/BSSpapers99/krista/Krista.htm

UTTARAKHAND OPEN UNIVERSITY

**Teenpani Bypass Road, Behind Transport Nagar,
Haldwani- 263139, Nainital (Uttarakhand)
Phone: 05946-261122, 261123; Fax No. 05946-264232
Website: www.uou.ac.in; e-mail: info@uou.ac.in
Toll Free No.: 1800 180 4025**