Indira Gandhi
National Open University
School of Sciences

BBYET-141
CELL AND MOLECULAR BIOLOGY

Block

4

GENE EXPRESSION AND REGULATION

UNIT 12
RNA Structure and Transcription 83

UNIT 13
Translation 123

UNIT 14
Regulation of Gene Expression 152

Course Design Committee

Dr. Anuradha Sharma
Hindu College, University of Delhi

Dr. Bela Bhatia
Deshbandhu College, University of Delhi

Dr. Neeru Bhandari
Dyal Singh College, University of Delhi

Dr. K.K. Koul
Hindu College University of Delhi

Late. Dr. D.K. Mallick
Deshbandhu College, University of Delhi

Dr. Swadesh Taneja (Retd.)
Life Sciences, IGNOU

Dr. Eklavya Chauhan
Sr. Consultant,
SOS, IGNOU

Dr. Bhupinder Dhir
Consultant, SOS, IGNOU

School of Sciences, IGNOU

Prof. Sujatha Varma (Director)

Prof. M.S. Nathawat (Ex. Director)

Prof. Jaswant Sokhi (Retd.)

Prof. Bano Saidullah (Retd.)

Prof. Neera Kapoor

Prof. Amrita Nigam

Block Preparation Team

Late Dr. D.K. Mallick
Deshbandhu College, University of Delhi

Dr. Bhupinder Dhir
Consultant, SOS, IGNOU,
New Delhi-110068

Dr. Kumkum Chaturvedi
Consultant, SOS, IGNOU,
New Delhi-110068

Dr. Eklavya Chauhan
Sr. Consultant, SOS,
IGNOU, New Delhi-110068

Prof. Rupam Kapoor (Editor)
Department of Botany,
Delhi University,
Delhi-110007

Prof. Amrita Nigam
SOS, IGNOU,
New Delhi-110068

ignou
THE PEOPLE'S
UNIVERSITY

Course Coordinator : Prof. Amrita Nigam

Production Team

Mr. Rajiv Girdhar
AR (P), MPDD, IGNOU

Mr. Hemant Kumar
SO(P), MPDD, IGNOU

Acknowledgements:

- Dr. A.K. Kavthekar and Dr. Kumkum Chaturvedi for giving useful inputs.
- Mr. Manoj Kumar for CRC Preparation.

November, 2021
©Indira Gandhi National Open University, 2021

ISBN

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the copyright holder.
Further information on the Indira Gandhi National Open University courses may be obtained from the official website of IGNOU at www.ignou.ac.in.
Printed and published on behalf of Indira Gandhi National Open University, New Delhi by Director, SOS, IGNOU

BLOCK 4 : GENE EXPRESSION AND REGULATION

Fourth and last block of this course introduces you to different aspects of gene function and regulation. The study of this block will make you aware of the latest aspects of RNA structure and transcription. This block highlights the principles of genetic code, which forms the basis of polypeptide synthesis. The units in this block have been well supplemented with latest information on the role of different proteinaceous factors involved in protein synthesis, along with a discussion on post-translational modifications of polypeptides. The block ends with a detailed discussion on the major theme in modern molecular genetics-the regulation of gene expression. Various aspects of gene silencing and its potential in solving many congenital diseases and cancer have also been dealt.

The block comprises three units which provide information about transcription, RNA structure, genetic code, translation, and regulation of gene expression.

Unit 12 provides information regarding the structure of RNA, its various types, and their role in the process of genetic message being encoded from DNA to RNA.The process of RNA transcription has been described, enumerating the various steps involved. In addition, different modes of post transcriptional processing and modification of pre mRNA have also been discussed. Salient features of the genetic code have been explained along with a review of some notable exceptions found in bacteria, mitochondria, and some other organisms.

Unit 13 provides a comprehensive account of various steps involved in translation. A comparative study of protein synthesis in prokaryotes and eukaryotes will enable you to appreciate the major differences between the two. Post-translational modifications have also been discussed which highlight the role of various metabolites in the formation of a fully functional protein.

Unit 14 deals with the most modern theme in molecular biology-the regulation of gene expression. This unit includes a detailed insight into the structure and function of *lac* and *tryp* operons in prokaryotes, highlighting various inducible and repressible systems. Gene regulation in eukaryotes is discussed with a view to appreciate the differences with its prokaryotic counterparts. Also included in this unit is a mind-stimulating account of the modern cutting-edge technology of gene silencing.

Objectives

After studying this block, you should be able to:

- study the structure and function of various types of RNA;
- describe the transcription machinery in prokaryotes and eukaryotes;
- appreciate the process of post-transcriptional modifications;
- understand the genetic code and its attributes;
- get an understanding of protein synthesis in prokaryotes and eukaryotes, and
- describe the principles of gene regulation and operon systems in prokaryotes viz. eukaryotic gene regulation.

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 12

RNA STRUCTURE AND TRANSCRIPTION

Structure

12.1 Introduction

Objectives

12.2 Structure of RNA

12.3 Types of RNA

Messenger RNA (mRNA)

Transfer RNA (tRNA)

Ribosomal RNA (rRNA)

Small RNA Molecules

12.4 Transcription

DNA Template and Transcription apparatus

Transcription in Prokaryotes

Transcription in Eukaryotes

12.5 Processing and Modification of RNA

Addition of 5′ Cap

Addition of Poly (A) Tail

RNA Splicing

RNA Editing

12.6 Genetic code

12.7 Summary

12.8 Terminal Questions

12.9 Answers

12.1 INTRODUCTION

In the previous units, you have studied that DNA is the genetic material of organisms. You have also come to understand its structure, chemistry, replication, as well as the way it is packaged into chromosomes and parceled out to daughter cells during mitotic and meiotic cell divisions. In this unit you are going to explore the world of RNA (mRNA, tRNA, rRNA, and some small RNAs), its structure, types and diverse roles in the cells. It is interesting to know that the behaviour of a cell is determined not only by what genes it inherits but also by which of those genes are expressed at a given time. In this unit you will also explore how the genetic information is stored in the DNA by

means of a triplet code called as genetic code that helps in DNA expression and guides the synthesis of RNA and protein molecules. The principle of this directional information flow from DNA to RNA to protein is known as the **Central Dogma of Molecular Biology**, a term coined by Francis Crick soon after the double-helix model of DNA was proposed. We will focus on the initial phases of gene expression by addressing two major questions. First, how is genetic information encoded? Second, how does the transfer of genetic information occur from DNA to RNA? thus explaining the process of transcription. We will study about transcription in bacteria and then compare this prokaryotic model with transcription in eukaryotes. You will study as to how messenger RNA is modified in eukaryotes after transcription that permits a single gene to produce multiple mRNAs coding for different polypeptides. You will also study how various editing mechanisms that allow the alteration in the base sequence of mRNA molecules have opened doors for exploring the cure for different diseases, for sustainable crop productivity and fuel resources. We will conclude this unit by exploring the various characteristics of genetic code - the set of rules that determine as to which nucleotides in mRNA correspond to which amino acids. The cracking of that code, which tells us how DNA can code for proteins, will also help you to become familiar with the terms which you will study in next unit.

Objectives

After studying this unit you would be able to:

- differentiate between various types of RNA and know its functions in the cells;
- distinguish between the process of transcription in prokaryotes and eukaryotes;
- appreciate the elaborate process of modification of pre-mRNA of eukaryotes and know its significance; and
- appreciate the importance of genetic code that helps in expression of DNA and guide the synthesis of RNAs and proteins.

12.2 STRUCTURE OF RNA

RNA, like DNA is a polynucleotide. Ribonucleotides of this single stranded polymer are also joined by phosphodiester bonds. You have already studied the major differences between DNA and RNA in the preceding unit 11. Early life was probably centered around RNA which served as genetic material as well as biological catalyst (see Box 12.1).

RNA is less stable than DNA and gets degraded under alkaline conditions. This is because the ribose sugar of RNA has a free hydroxyl group on the 2'-carbon atom. Such a free hydroxyl group is absent on the deoxyribose sugar of DNA. Moreover, the pyrimidine uracil is present in RNA in place of thymine. (Fig.12.1).

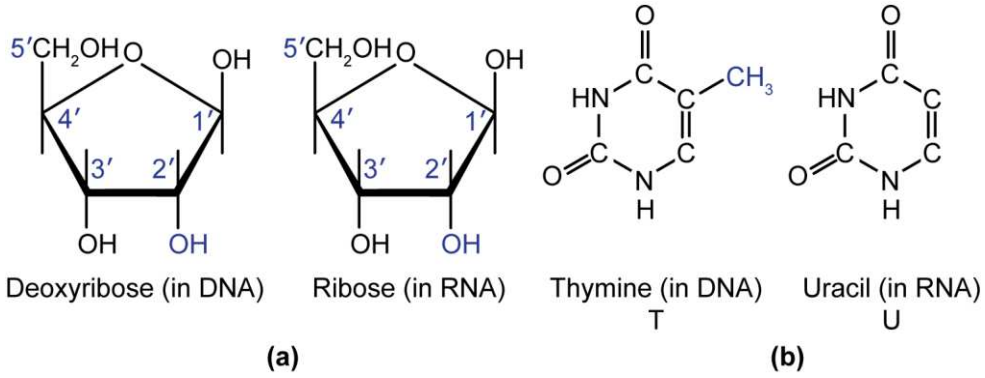


Fig. 12.1: a) Ribonucleotides contain pentose sugar ribose instead of deoxyribose found in deoxyribonucleotide; b) RNA contains the pyrimidine uracil in place of thymine, found in DNA.

RNA is usually single stranded, consisting of single polynucleotide strand, except some viruses which contain double stranded genome. RNA also forms secondary structure when short complementary regions within a nucleotide strand pair, and is called as hairpin–loop or stem-loop structures (Fig.12.2).

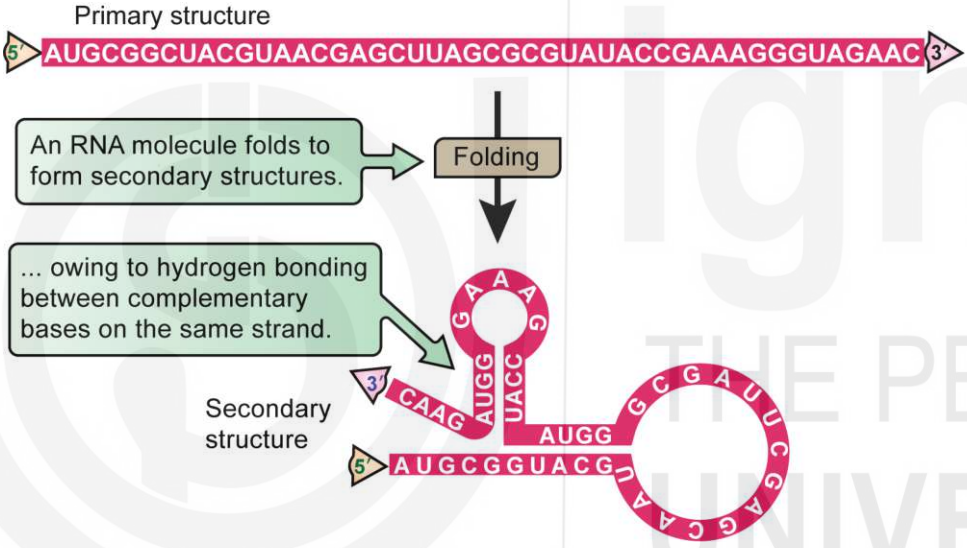


Fig. 12.2: Primary RNA structure and its folding into secondary structure forming a hairpin loop or stem-loop structure.

The strands in the two regions of a single RNA molecule remain antiparallel when they pair up. Cytosine pairs with guanine, while adenine pairs with uracil (although occasionally, guanine pairs with uracil). Interestingly, (Box 12.1) it is believed that the early life was an RNA world and RNA served both as carriers of genetic information and as catalyst that drove the chemical reactions needed to sustain life.

Box 12.1: An Early RNA World

The biochemical paradox- nucleic acid is responsible for information storage and proteins serve as biological catalysts or enzymes, making reactions to take place rapidly within the cell – revealed a dilemma in our understanding of early stages of evolution of life. Which came first: proteins or nucleic acids? If nucleic acids carry the coding instructions for proteins, how could proteins be generated without them? Because nucleic acids are unable to copy themselves, how could they be generated without proteins? If DNA and proteins each require the other, how could life begin?

This apparent paradox disappeared by the discovery of catalytic RNA molecules called **Ribozymes** by Thomas Cech and his colleagues in 1981 from the protozoan

Tetrahymena thermophila. These RNA molecules have the capacity to cut and join other RNA molecules, and replicate them. Also they can catalyze the peptide bonds between amino acids during translation. The discovery of ribozymes leads further support to other evidences, indicating that the original genetic material was RNA, and early life was an **RNA world,** serving both as carriers of genetic information and as catalyst that drove the chemical reactions needed to sustain life. Scientists speculate that these catalytic RNAs may have acquired the ability to synthesize proteinaceous enzymes which proved to be more efficient catalysts than RNA itself. Much later, RNA got replaced by DNA as the primary carrier of genetic information. This seemed to be logical, as DNA is chemically more stable and shows a higher fidelity (faithfulness) in replication.

In modern cells RNA still plays very important role in both DNA replication and protein synthesis.

Functionally RNA is more diverse than DNA

The secondary structures of RNA participate in a variety of cellular functions. Different RNA molecules can assume different structures since their secondary structure is determined by the base sequence of the nucleotide strand. RNA molecules have the potential for tremendous variation in functions because their structure determines their function. In contrast, the DNA has fewer functional roles in the cell because the two complementary strands form a helix which is much more restricted in the range of secondary structure that it can assume.

We will now study the functions of each type of RNA. But before reading further you may like to try SAQ 1.

SAQ 1

Write whether the following statements are true or false.

a) RNA is generally double stranded. [ ]

b) Thymine, one of the pyrimidines present in DNA, is replaced by uracil in RNA. [ ]

c) Functionally RNA is more diverse than DNA. [ ]

d) The presence of free hydroxyl group on the 2'-carbon atom of ribose sugar provides stability to RNA. [ ]

e) The discovery of ribozymes support evidences that RNA was the original genetic material and early life was an RNA world. [ ]

12.3 TYPES OF RNA

RNAs can be grouped into two general classes:

a) Messenger RNA (mRNA) - This RNA acts as a messenger of genetic information for making of polypeptides (proteins) which is encoded onto it by DNA. In other words, mRNA acts as an intermediary that passes information from DNA to protein.

b) Functional RNA -This type of RNA is itself the final product and does not participate in the making of proteins. Because RNA is such a versatile molecule, it is of different types and can serve many functional roles and participate in a variety of cellular processes. We are going to describe each type of RNA and its role now.

12.3.1 Messenger RNA

Messenger RNA carries genetic information from DNA (by transcription) and with the participation of ribosomes passes this genetic information to assemble amino acids in their correct order (translation). In bacteria, mRNA is transcribed directly from DNA but in eukaryotes, a pre-mRNA called primary transcript is first transcribed from DNA and then processed to yield mature mRNA. Both prokaryotic and eukaryotic mRNAs contain three primary regions (Fig. 12.3).

The short sequences of common nucleotides are called "**Consensus Sequences**". Consensus sequence simply means the sequences having considerable similarity. The presence of consensus in a set of nucleotides usually implies that the sequence is associated with an important function. Consensus sequences are conserved, i.e., identical or similar across different species.

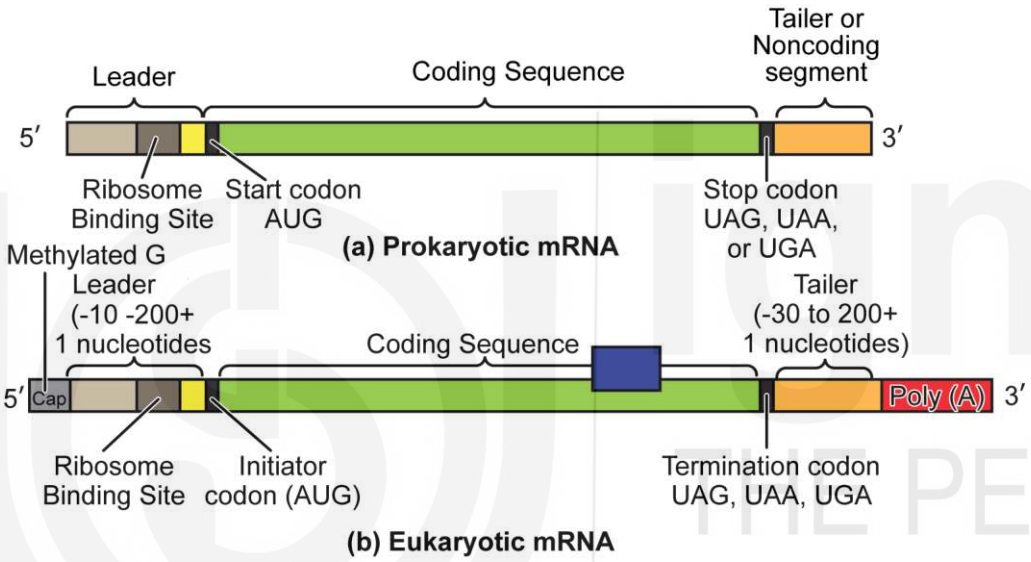


Fig. 12.3: a) Three primary regions of a mature (fully processed) mRNA --- 5´ - untranslated region, the protein- coding region, and the 3´ – untranslated region; b) Eukaryotic mRNA with its characteristic methylated G-cap and a Poly A tail.

- The 5´end of mRNA begins with nucleotides which are not translated into proteins. This 5´ untranslated region (5´UTR) is called as the Leader and has no ribosome binding sites. In bacterial mRNA the end of 5´UTR has a consensus (in all cases) sequence that serves as a ribosome-binding site. This consensus sequence is a hexamer (six nucleotides) called the **Shine-Dalgarno sequence.** This sequence is located seven nucleotides upstream of the first codon (initiation/start codon) where the first amino acid is incorporated. In contrast, the eukaryotic mRNA lacks the equivalent consensus sequence in its 5´ – untranslated regions. In eukaryotic cells the ribosomes bind to a modified 5´ – end of mRNA.

- The next region of mRNA is the protein coding region which begins with the **start codon** and ends with the **stop codon**.

- The last region of the mRNA is the 3´ – untranslated region (3´ – UTR sometimes called **trailer**).The nucleotides at the 3´ – end of the mRNA are not translated into a protein. The 3´ – UTR provides stability to mRNA, and to the protein coding sequences for translation.

12.3.2 Transfer RNAs

These are also known as Functional RNAs. We will first describe the structure followed by function of tRNAs.

Structure: All RNAs have four standard bases (adenine, guanine, cytosine and uracil) but the unique feature of tRNA is the presence of some additional modified nitrogenous bases, viz., methyl adenine, methyl cytosine, methyl guanine, pseudouracil and many others. They arise from the chemical changes made to the four standard bases after transcription (post-transcription modification).These changes are carried out by special tRNA-modifying enzymes. For example, addition of a methyl group to uracil changes it to modified base ribothymine. All the tRNAs have similar structure, a feature critical to tRNA function.

Most tRNAs contain between 74 and 95 nucleotides, some of which are complementary to each other. These may pair and form intramolecular hydrogen bonds resulting in clover leaf structure of each tRNA. The cloverleaf has four arms (Fig.12.4), the acceptor arm, the TΨC, anticodon arm, and DHU arm. The three of the four arms (the TΨC, anticodon and DHU arms) consists of a stem and a loop.

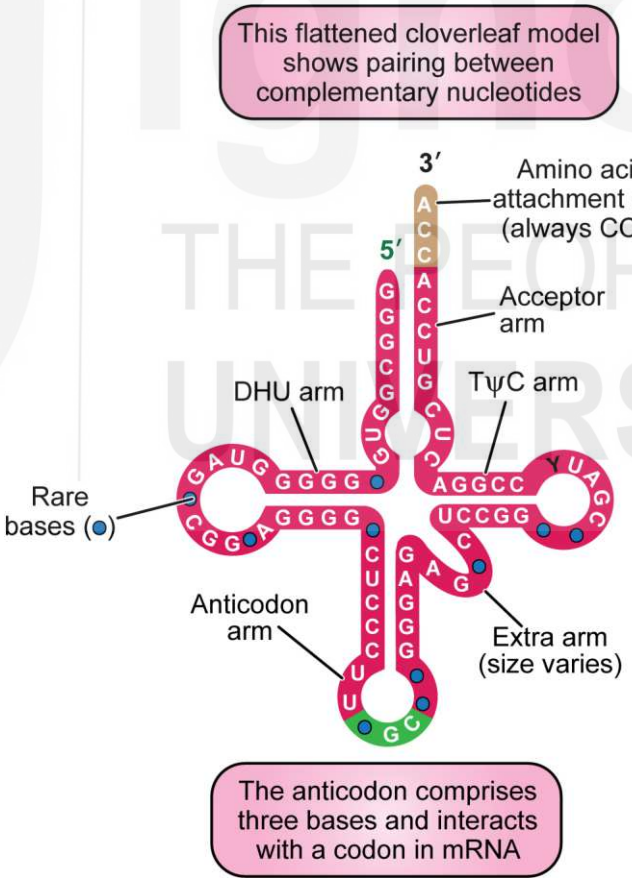


Fig. 12.4: The cloverleaf secondary structure of tRNA showing four arms.

The stem is formed by the pairing of complementary nucleotides. At the terminus of the stem lies the loop, with no nucleotide pairing. The acceptor arm includes the 5´ and 3´ ends of the tRNA molecule and is without a loop. All tRNAs possess same sequence (CCA) at the 3´ end of acceptor arm, where the amino acid attaches to the tRNA clearly showing that this sequence is not responsible for specifying which amino acid will attach to the tRNA.

The TΨC arm is named for the three bases of three nucleotides in the loop of this arm: thymine (T), pseudouridine (Ψ) and cytosine (C). The anticodon arm lies at the bottom of the tRNA. The three nucleotides at the end of this arm make up the anticodon which pairs with corresponding codon on mRNA to ensure that the amino acids link in a correct order. The DHU arm is so named because it often contains the modified base dihydrouridine.

You have studied about structure of tRNA. Now we will study its functions. Transfer RNA (tRNA) is also called adapter RNA molecule as it transports amino acids to the ribosomes and interacts with codons in mRNA, placing amino acids in their correct order in protein synthesis. The tRNA serves as a link between genetic code in mRNA and the amino acids that make up the protein. Each tRNA attaches to a specific amino acid and carries it to the ribosome, where this amino acid is added to the growing polypeptide chain at the position specified by the genetic instructions in the mRNA (genetic code).The complex formed by attaching tRNA with its respective amino acid is designated by an abbreviated form by adding a three- letter superscript that represents the attached amino acid. For example, a tRNA that attaches the amino acid alanine is written as $tRNA_{Ala.}$ There are minimum 20 different types of tRNAs present for 20 different amino acids found in proteins and most of the organisms possess from at least 30 to 40 different types of tRNA, each encoded by a different gene.

12.3.3 Ribosomal RNA

Ribosomes typically contain about 80 % of the total cellular RNA. They are complex structures, each consisting of more than 50 different proteins and RNA molecules. Ribosomal RNA is the major structural component of the ribosome and has sequence complementary to regions of mRNA with which it can interact. It is now clear that the RNA components of the ribosome perform all important catalytic functions associated with translation. Molecular hybridization studies have established the degree of redundancy of the genes coding for rRNA components. The genes for rRNA, like those for tRNA, can be present in multiple copies and their number varies among the species.

Eukaryotic cells possess two types of rRNA genes: a large gene that encodes 18S rRNA, 28S rRNA, and 5.8S rRNA (S values designated on the basis of velocity centrifugation), and a small gene that encodes the 5S rRNA.

12.3.4 Small RNA Molecules

Another class of functional RNAs is **small RNA molecules** specific to eukaryotes. The common perception is that the early RNA world died out billions of years ago, when many of RNA's functions were replaced by more-stable DNA molecule and more efficient protein catalysts. However, within last 20 years, a group of small RNA molecules (most of them 20 – 30 nucleotides long) have been discovered that greatly influence many basic biological processes, including the formation of chromatin structure, transcription and translation. These small RNA molecules are found in all plants and animals and play important roles in gene expression, development, and cancer. The discovery of small RNA molecules has greatly influenced our understanding of how genes are regulated and importance of DNA sequences that do not encode proteins. These findings demonstrate that we still live very much in RNA world.

Classes of small RNA molecules:

i) **Small nuclear RNAs (snRNAs)** – They are small RNA molecules (90 to 400 nucleotides long) that associate with proteins to form RNP (ribonucleoprotein) particles known as snRNPs or *snurps*. This small RNA species also form ribonucleoprotein processing complex (*spliceosome*) that edits and removes introns from eukaryotic mRNAs.

A very large fraction of eukaryotic genome is transcribed into a group of functional RNAs that participate in the regulation of gene expression at various levels.

ii) **MicroRNAs (miRNAs) and Small interfering RNAs (siRNAs)** – These small RNAs found in many eukaryotes are responsible for a variety of different functions, including the regulation of gene expression, defense against viruses, suppression of transposons, and modification of chromatin structure. These two abundant classes of RNA molecules although differ in how they originate, they have a number of features in common and their functions overlap considerably as shown in Table 12.1.

- Both miRNAs and siRNAs are about 22 nucleotides long. Small interfering RNAs arise from cleavage of mRNAs, RNA transposons and RNA viruses. Some miRNAs are cleaved from RNA molecules transcribed from sequences that encode miRNA only, but others are encoded in the introns and exons (a segment that codes for proteins) of mRNAs.
- Each miRNA is cleaved from a single-stranded RNA precursor that forms small hairpins, whereas multiple siRNAs are produced from the cleavage of an RNA duplex consisting of two different RNA molecules.
- Usually siRNA have exact complementarity (a relationship between two molecules following lock and key principle) with their target mRNA or DNA sequences and suppresses gene expression by degrading mRNA or inhibiting transcription whereas miRNAs have limited complementarities with their target mRNAs and often suppresses gene expression by inhibiting translation.
- Finally, miRNAs usually silence the genes that are distinct from those from which the miRNA were transcribed, whereas siRNAs typically silence the gene from which the siRNAs were transcribed.

The mechanism of silencing the genes by these small RNA molecules is called **RNA interference** (Box. 12.2). Both siRNA and miRNA molecule combine with proteins to form an **RNA-induced silencing complex** (RICS). Key to functioning of RICSs is a protein called **Argonaut**. This is achieved by a mechanism involving the pairing of RICS with an mRNA molecule that possesses a sequence complementary to its siRNA or miRNA component. This pairing results in cleavage of mRNA. An analogous group of small RNAs with silencing functions – called prokaryotic silencing RNAs (psiRNAs) have recently been detected in prokaryotes. Gene silencing is brought about by either repression of translation or by mRNA degradation.

Box 12.2: RNA Interference

RNA Interference was discovered by Andrew Fire and Craig Mello in 2006 who were awarded the Nobel Prize.

RNA interference (RNAi) is a precise mechanism operative in eukaryotic cells which helps a cell to check the invasion of foreign genes (from viruses and transposons and also to check the expression of their own genes). RNA interference is triggered by double stranded RNA molecules, which may arise in several ways: i) by the transcription of inverted repeats into an RNA molecules that then base pairs with itself to form double- stranded RNA, ii) by simultaneous transcription of two different RNA molecules that are complementary to one another that pair, forming double-stranded RNA, or iii) by infection of viruses that make double stranded RNA. These double-stranded RNA molecules are chopped up by an enzyme appropriately called DICER, resulting in tiny RNA molecules that are unwound to produce siRNAs and miRNAs.

Geneticists speculate that RNA interference evolved as a defense against RNA viruses and transposable elements that move through RNA intermediate. However, RNA interference is also responsible for number of key genetic and developmental processes including changes in chromatin structure, translation, cell fate and proliferation, and cell death. Geneticists also use RNAi machinery as an effective tool for blocking the expression of specific genes and hence in gene silencing.

Table 12.1: Location and functions of different classes of RNA molecules.

Class of RNA	Cell Type	Location in Eukaryotic Cells*	Function
Ribosomal RNA (rRNA)	Bacterial and eukaryotic	Cytoplasm	Structural and functional components of the ribosome
Messenger RNA (mRNA)	Bacterial and eukaryotic	Nucleus and cytoplasm	Carries genetic code for proteins
Transfer RNA (tRNA)	Bacterial and eukaryotic	Cytoplasm	Helps incorporate amino acids into polypeptide chain
Small nuclear RNA (snRNA)	Eukaryotic	Nucleus	Processing of pre-mrRNA
Small nucleolar RNA (snoRNA)	Eukaryotic	Nucleus	Processing and assembly of rRNA
Small cytoplasmic RNA (scRNA)	Eukaryotic	Cytoplasm	Variable
MicroRNA (miRNA)	Eukaryotic	Cytoplasm	Inhibits translation of mRNA
Small interfering RNA (siRNA)	Eukaryotic	Cytoplasm	Triggers degradation of other RNA molecules
Piwi-interacting RNA (piRNA)	Eukaryotic	Cytoplasm	Thought to regulate gametogenesis, but function poorly defined

*All eukaryotic RNAs are transcribed in the nucleus

We have described above the structure, types and functions of different RNAs, but before moving further you may like to work out an SAQ 2 based on these.

SAQ 2

Fill in the blanks with the suitable word from the words given below.

a) The tRNA is also called ……. molecule because it serves as a link between genetic code in mRNA and the amino acids that make up the protein.

b) Ribosomal RNA is the major structural component of the ……. .

c) ……. is a sequence of nucleotides at the 5´ end of mRNA that does not encode any of the amino acids of a protein.

d) The ribonucleoprotein processing complex that removes introns from eukaryotic mRNAs is called …… .

e) The mechanism of silencing the genes by the small RNA molecules is called …….. .

(Ribosome, 5´ UTR, adaptor, RNA interference, spliceosome)

12.4 TRANSCRIPTION

12.4.1 DNA Template and Transcription Apparatus

The process by which RNA molecules are synthesized on DNA template is called transcription. An mRNA molecule is synthesized which is complementary to the gene sequence of one of the two strands of DNA double helix. Soon after the double-helical model of DNA was first proposed, Francis Crick proposed a principle called **Central Dogma of Molecular Biology,** which explains the directional flow of genetic information from DNA to RNA to protein (Fig. 12.5)

However, since its formulation, the central dogma has been refined in various ways. For example, many viruses with RNA genomes have been found to synthesize RNA molecules using RNA as template. Other RNA viruses, such as HIV, carry out reverse transcription, whereby the viral RNA is used as a template for DNA synthesis – a “backward” flow of genetic information. But despite these variations on the original model, the principle that information flows from DNA to RNA to protein remains the main operating principle by which all cells use their genetic information.

The idea that RNA is involved as an intermediate molecule in the process of information flow between DNA and protein is suggested by following observations:

1. DNA for the most part remains associated with chromosomes in the nucleus of the eukaryotic cells. However, protein synthesis occurs in association with ribosomes located outside the nucleus, in the cytoplasm, Therefore, DNA does not appear to participate directly in protein synthesis.

2. RNA is synthesized in the nucleus of eukaryotic cells, in which DNA is found and is chemically similar to DNA.

3. Following its synthesis, most RNA migrates to the cytoplasm where protein synthesis (translation) occurs.

4. The amount of RNA is generally proportional to the amount of proteins in a cell.

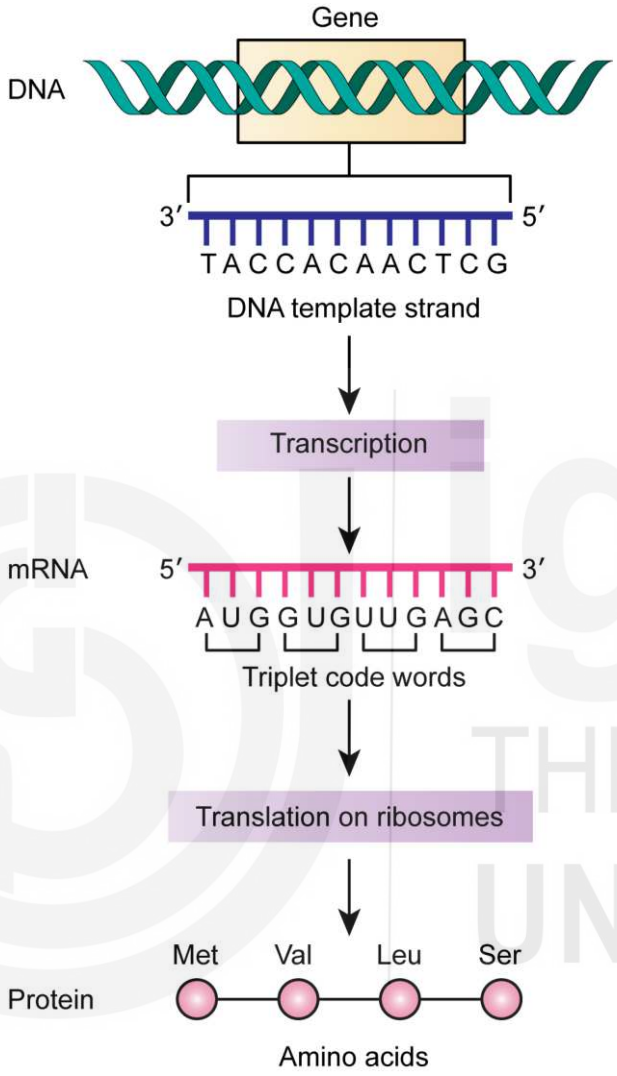


Fig. 12.5: Flow chart showing how genetic information encoded in DNA produces a protein.

At this point, it would be interesting to compare transcription with DNA replication, which you have studied earlier.

Transcription is in many ways similar to replication, but one fundamental difference relates to the length of the template used. During replication, all the nucleotides in the DNA template are copied, but, during transcription, only specific small fragments of the DNA molecule- usually a single gene or, at the most, a few genes – get transcribed into RNA. This seems to be logical since all the gene products are not needed at the same time or in the same cell. And it would be highly uneconomical for a cell to constantly transcribe all the genes regardless of their utilization. So transcription is, in fact, a highly selective and a customized process where only those individual genes are transcribed whose products are required.

Like replication, transcription requires three major components:

1) A DNA template;

2) The raw materials required to build a new RNA molecule,

3) The transcription apparatus, consisting of enzymes, proteinaceous factors, cofactors, and activators necessary to catalyze the synthesis of RNA.

DNA template

The process of transcription also uses a single strand of DNA double helix as template. However, unlike replication, where both the strands are replicated individually, the transcription of a gene takes place on only one of the two DNA strands. The nucleotide strand used for transcription is called as the template strand. The other strand, called the non template strand, is not ordinarily transcribed. The template strand contains a transcription unit (Box 12.3), which is a stretch of DNA that encodes an RNA molecule and the sequences necessary for its transcription. During transcription, an RNA molecule is synthesized that is complementary and antiparallel to the DNA template strand. The RNA transcript has the same polarity and base sequence as that of the non-template strand, with the exception that RNA contains U in place of T. In most organisms, each gene is transcribed from a single strand, but different genes may also be transcribed from different strands as shown in Figure 12.6.

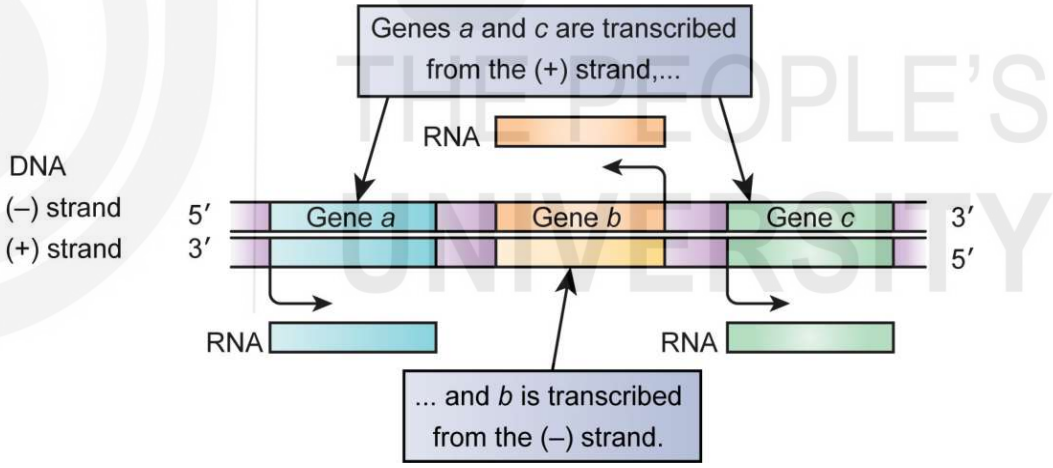


Fig. 12.6: RNA transcription from one DNA strand. In some cases, different genes may be transcribed from one or the other of the two DNA strands.

The terms **upstream** and **downstream** are generally used to designate the direction of the transcription and the location of nucleotide sequences surrounding the RNA-coding sequence on the DNA template. The transcription apparatus that includes *RNA polymerase* is supported by a number of accessory proteins that join and leave the *RNA polymerase* at different stages of the process. *RNA polymerase* binds to the promoter (located usually upstream of the start/initiation site) and moves towards the terminator (positioned downstream of the start site).

Generally only one (nontemplate strand) is written out for DNA sequences because it will be the same as RNA transcribed from the template (with the

exception that U in RNA replaces T in DNA). By convention, the sequence on the nontemplate strand is written with the 5′ end on the left whereas 3′ end is written on the right. The first nucleotide transcribed (transcription start site) is designated as +1. Nucleotides downstream of the start site are assigned positive number, and the nucleotides upstream of the start site are assigned negative numbers. For example, the nucleotide +34 would be 34 nucleotides downstream of the start site, whereas nucleotide -72 would be 72 nucleotides upstream of the start site. There is no nucleotide numbered as zero.

Box 12.3: The Transcription Unit

A transcription unit is a stretch of DNA that encodes an RNA molecule and the sequences necessary for its transcription. It includes three critical regions: a promoter, an RNA-coding sequence and a terminator (Fig. 12.7). The transcription apparatus recognizes and binds to the promoter DNA sequence. The binding decides which of the two DNA strands is to be read as template and what will be the direction of transcription. In addition, the promoter also determines the transcription start site, the first nucleotide that will be transcribed into RNA. Although in most transcription units, the promoter is located next to transcription start site, it does not itself get transcribed.

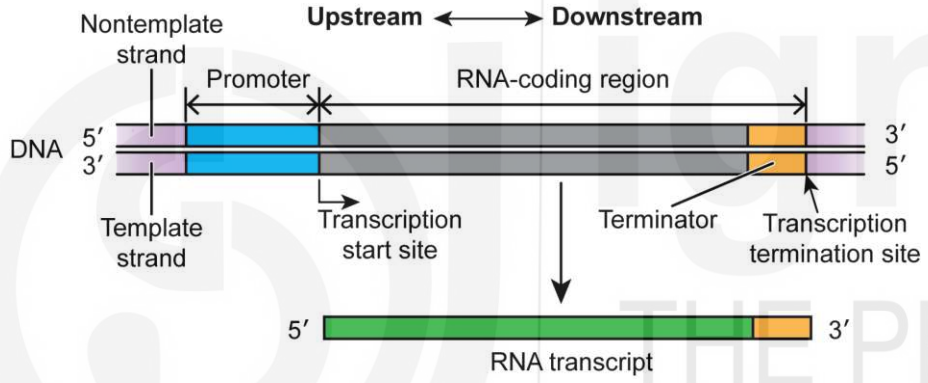


Fig. 12.7: A transcription unit showing a promoter, a RNA-coding region, and a terminator.

The second critical region of the transcription unit is a sequence of DNA nucleotides that is copied into a RNA molecule. This is called as the RNA- coding region. The third component of a transcription unit is the terminator, a sequence of nucleotides that signals where transcription is to end. The terminator portion is usually the part of the RNA-coding sequence. It is only after the terminator has been copied into RNA that the process of transcription stops.

The raw materials (Substrates)

During RNA synthesis ribonucleoside triphosphates (rNTPs) are added one at a time to 3′-OH group of the growing RNA molecule. Two phosphate groups from the incoming rNTPs are cleaved, whereas the remaining phosphate group participates in the formation of a phosphodiester bond that connects the nucleotides to the growing RNA molecule. The overall chemical reaction summarizing the synthesis of RNA on a DNA template can be expressed as:

$$n(NTP) \rightarrow (NMP)_n + {}_n(PP_i)$$

where ribonucleoside triphosphates (rNTPs) serve as substrate for the enzyme *RNA polymerase* as shown in Fig. 12.8, which catalyzes the polymerization of nucleoside monophosphate (NMPs) or nucleotides into a polynucleotide chain $(NMP)_n$. The energy released by cleaving the

triphosphate precursor into the monophosphate form drives the reaction, and inorganic phosphate (PP_i) is produced. Nucleotides are always added to the 3´ – end of RNA molecule and the direction of transcription is therefore, 5´ → 3´.

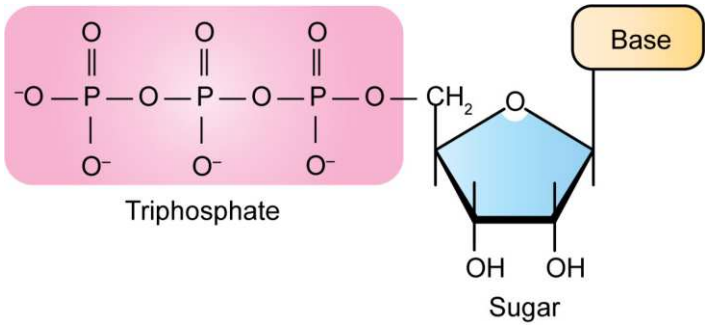


Fig 12.8: Structure of ribonucleoside triphosphate.

Transcription apparatus (*RNA polymerase* and accessory proteins)

Bacterial *RNA polymerase*: The enzyme capable of synthesizing RNA on DNA template was discovered by several investigators including Samuel Weiss (1959) who isolated it from rat liver and called it *RNA polymerase*. *RNA polymerase* from *E. coli* has been extensively investigated. Bacterial *RNA polymerase* is a multimeric enzyme (i.e., it is made up of several polypeptide chains) consisting of five subunits that make up the core enzyme: two copies of subunits called alpha (α) and single copies each of subunits beta (β), beta prime (β'), and omega (ω) (Fig.12.9).

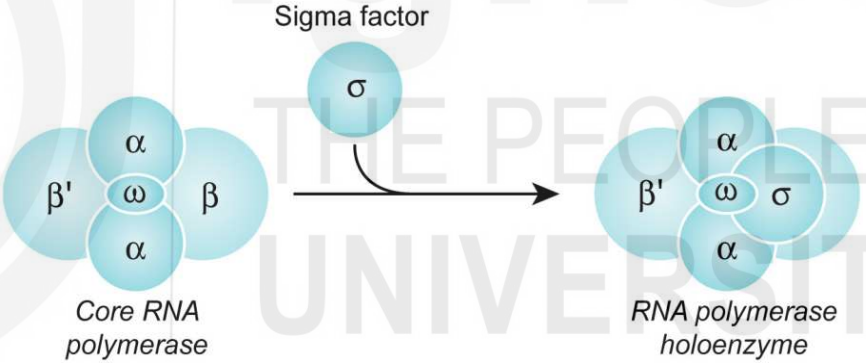


Fig 12.9: The core enzyme of bacterial *RNA polymerase* consisting of five subunits. The core enzyme catalyzes the elongation of RNA molecule and adds new RNA nucleotides.

Beta (β) and Beta prime (β') polypeptides provide the catalytic mechanism and the active site for transcription. Although the ω subunit does not directly participate in transcription, it helps in stabilizing the enzyme. Other functional subunits join and leave the core enzyme at particular stages of transcription process. The sigma (σ) factor plays a regulatory role in initiation of RNA transcription. It controls the binding of *RNA polymerase* to the promoter thereby ensuring that initiation takes place at the precise site. Without sigma factor, *RNA polymerase* will initiate transcription at random point along the DNA. The complex active form of the enzyme, i.e., the holoenzyme is formed only when sigma factor binds with the core enzyme. It is only after this binding that the *RNA polymerase* acquires the capability to stably bind to the promoter region, and not anywhere else, and initiate transcription at proper start site. After a few RNA nucleotides have joined together, the sigma factor usually detaches from the core enzyme.

Eukaryotic *RNA polymerases*: Eukaryotic cells contain three distinct nuclear *RNA polymerases* viz., *RNA polymerase* I, II and III. These enzymes transcribe different classes of genes forming different RNAs (see Table 12.2). Protein-coding genes are transcribed by *RNA polymerase II* to yield mRNAs. *RNA polymerase II* also transcribes microRNAs (miRNAs), snRNAs, snoRNAs and long noncoding RNAs (lncRNAs), which are critical regulators of gene expression in eukaryotic cells. Ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are transcribed by *RNA polymerase I* and III. *RNA polymerase I* is specifically devoted to transcription of three largest species of rRNAs, which are designated as 28S, 18S, and 5.8S. *RNA polymerase III* transcribes the genes for tRNAs and for the smallest species of ribosomal RNAs (5.8S rRNAs). Some of the small RNAs involved in splicing and protein transport (snRNAs and scRNAs) are also transcribed by *RNA polymerase III*. Interestingly, chloroplasts and mitochondria also possess separate *RNA polymerases* which are similar to bacterial *RNA polymerases*. These enzymes specifically transcribe the DNAs for the organelles.

All the three *RNA polymerases* are complex enzymes consisting of 12 to 17 different subunits each. Although they recognize and transcribe distinct classes of genes, they share several features common with each other as well as with bacterial *RNA polymerase*. In particular, all three eukaryotic *RNA polymerases* contain nine conserved subunits, five of which are related to the, β, β', and ω subunits of bacterial *RNA polymerase*.

Table: 12.2: Eukaryotic *RNA polymerases.*

Type	Present in	Transcribes
RNA polymerase I	All eukaryotes	large RNAs
RNA polymerase II snRNAs, snoRNAs, some miRNAs	All eukaryotes	pre-mRNAs
RNA polymerase III, some snRNAs and miRNAs	All eukaryotes	tRNAs, some rRNAs,
RNA polymerase IV	Plants and some siRNAs	
RNA polymerase V	Plants RNA molecules taking part in heterochromatin formation	

12.4.2 Transcription in Prokaryotes

Extensive studies on *E. coli* have provided the model for subsequent investigations of transcription in eukaryotes. *E. coli* was also the first organism in which mRNA was discovered, *RNA polymerase* was purified and studied and basic mechanisms by which transcription is regulated was elucidated by pioneering experiments in which regulated gene expression allows the cell to respond to variations in the environment such as changes in the availability of nutrients. *E. coli* has thus, provided the foundation for studies of the far more

complex mechanisms that regulate gene expression in eukaryotic cells. Transcription in eukaryotes is similar to bacterial transcription in that the basic principles of transcription known in prokaryotes also apply to eukaryotes.

Transcription involves four stages:

1. Binding of *RNA polymerase* to a promoter sequence
2. Initiation of RNA synthesis
3. Elongation of the RNA chain
4. Termination of RNA synthesis

An overview of all the four stages of transcription has been depicted (Fig.12.10).

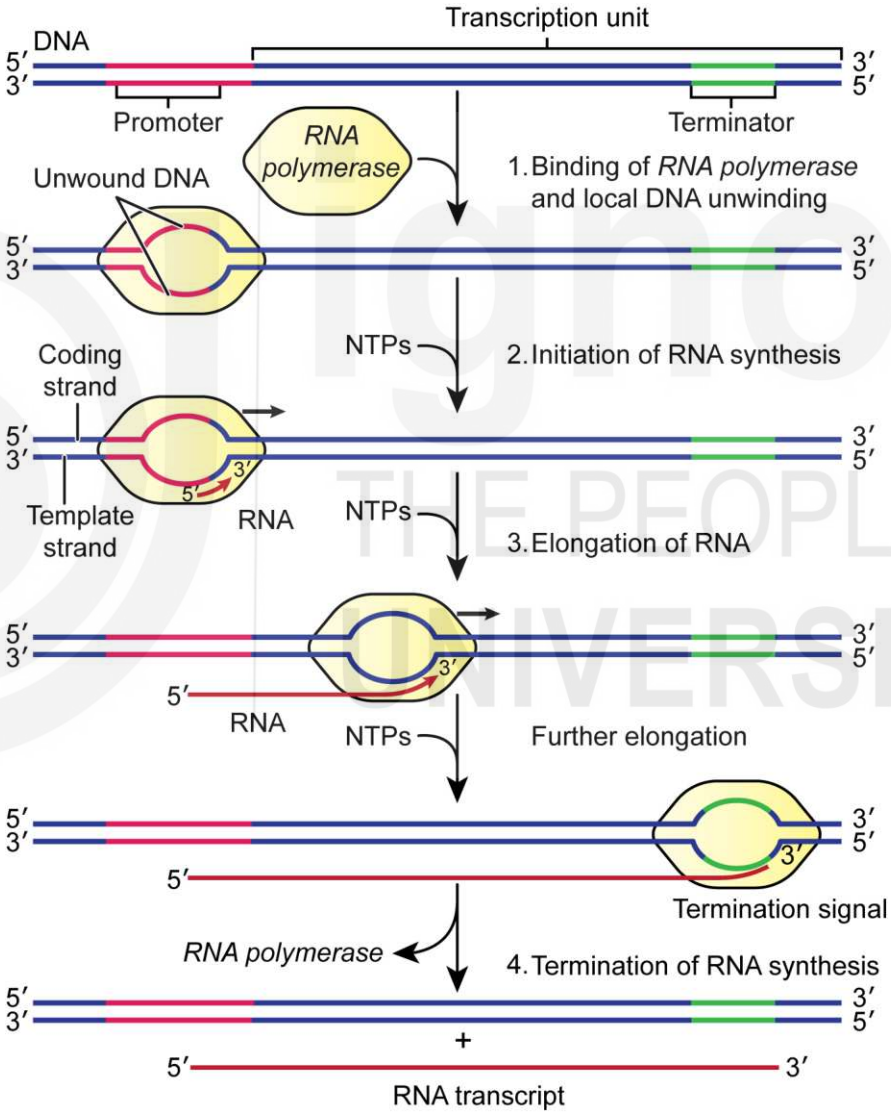


Fig 12.10: An overview of transcription showing the four main stages. Note that *RNA polymerase* moves along the template strand of the DNA in the 3′ →5′ direction and the newly formed RNA molecule grows in 5′ →3′ direction. NTPs- Nucleoside triphosphates

Binding of *RNA Polymerase* to a Promoter

The first step of RNA synthesis is the binding of *RNA polymerase* to a DNA promoter site- a specific sequence made up of several dozen base pairs. This

binding determines which DNA strand will serve as the template strand, and the point from where RNA synthesis would begin.

Bacterial Promoters

In bacterial cells promoters are usually positioned adjacent to an RNA-coding sequence. Although the nucleotide sequence of many promoters of *E. coli* and other bacteria vary in sequence, short stretches of nucleotides are common to many. Furthermore, the spacing and location of the sequences relative to the transcription start site are similar in most promoters (Fig. 12.11). The most common consensus sequence found in almost all bacterial promoters is centered about 10 bp upstream of the start site, called the **-10 consensus sequence**, and popularly called as **Pribnow box.** It is simply written as TATAAT.

5′ –T A T A A T – 3′

3′ –A T A T T A – 5′

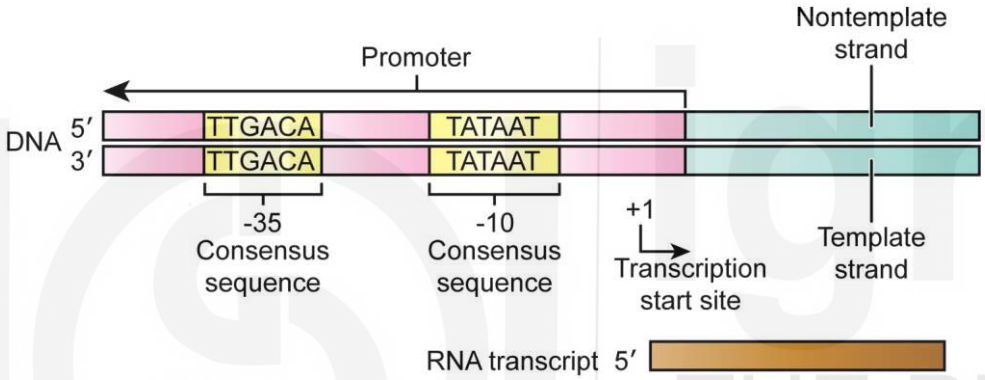


Fig. 12.11: Organization of a bacterial promoter sequence showing the transcription start point (designated + 1 or usually A) and two consensus sequences (the six nucleotides -10 sequence and the six nucleotides -35 sequence).

Another consensus sequence common to most bacterial promoters is TTGACA. It lies approximately 35 nucleotides upstream of the start site and is termed as -35 consensus sequence. The nucleotides on either side of the -10 and -35 consensus sequence or those between them vary considerably from promoter to promoter, suggesting that these nucleotides are not very important in promoter recognition.

During binding of *RNA polymerase* to promoter, the sigma factor associates with the core enzyme to form a holoenzyme which binds to the -35 and -10 consensus sequence in the DNA promoter. Although it binds only to the nucleotides of consensus sequences, the enzyme extends from -50 to +20 when bound to the promoter. The holoenzyme initially binds weakly to the promoter but then undergoes a change in structure that allows it to bind more tightly and unwind the double-stranded DNA. DNA unwinding begins within -10 consensus sequence and extends downstream for about 14 nucleotides, including the start site (from nucleotides -12 to +2).

Initiation of RNA Synthesis

After positioning itself over the start site for transcription (at position + 1), *RNA polymerase* unwinds the DNA to produce single - stranded template. The

orientation and spacing of consensus sequence on a DNA strand determines which strand will be the template for transcription thereby also determining the direction of transcription. The RNA molecule is synthesized when *RNA polymerase* pairs the base on a ribonucleoside triphosphate with its complementary base at the start site on the DNA template strand as the nucleotide is added to the 3´ end of the growing RNA molecule. Unlike DNA synthesis, primer is not required to initiate the synthesis of the 5´ end of the RNA molecule. A phosphodiester bond is incorporated at the 3´end of the growing RNA molecule by cleaving two of the three phosphate groups from the ribonucleoside triphosphate. However, there is no cleavage of phosphate groups in the first ribonucleoside. This is because 5´ end of the first ribonucleoside does not take part in the formation of a phosphodiester bond.

After several abortive attempts of generating short RNA transcripts, from 2 to 6 nucleotides in length, *RNA polymerase* eventually synthesizes an RNA molecule from 9 to 12 nucleotides in length. With this initiation, the *RNA polymerase* now enters the elongation phase.

Elongation

Once the initiation of transcription is completed, there is conformational (shape) change in the *RNA polymerase*. This change helps the enzyme to escape from the promoter and begin transcription downstream. The sigma factor subunit is usually released after initiation. *RNA polymerase* progressively unwinds the DNA at the leading (downstream) edge of the transcription bubble. New nucleotides are joined to the RNA molecule according to the sequence on the template. Thereafter, the enzyme rewinds the DNA at the trailing (upstream) edge of the bubble (see Fig. 12.12). In bacterial cell at 37°C, about 40 nucleotides are added per second and this rate of RNA synthesis is much lower than the 1000 to 2000 nucleotides added per second in DNA synthesis.

The transcription bubble

The transcription bubble is a short stretch of 18 nucleotides unwound DNA within which transcription takes place. RNA is continuously synthesized within this region, with single stranded DNA being used as a template.

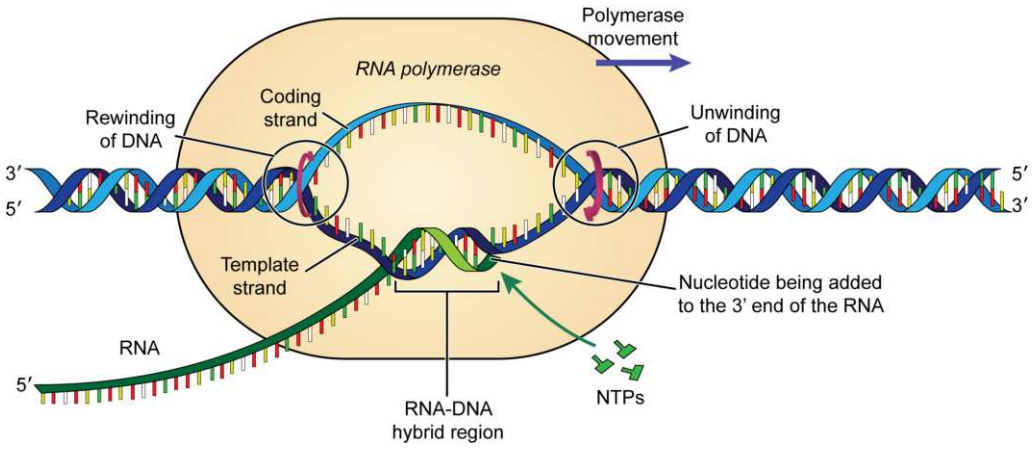


Fig. 12.12: A bacterial elongation complex showing the movement of the enzyme along the template DNA strand from the 3´ towards the 5´ end and RNA strand is elongated in the 5´ →3´ direction. NTPs- Nucleoside triphosphates

Termination of RNA Synthesis

There exist specific sequences of DNA and RNA, both in prokaryotes and eukaryotes which act as transitory transcription pauses. Such RNA pauses help in coordinating RNA processing.

Elongation of the growing RNA chain continues until *RNA polymerase* copies a special sequence, called a termination signal. This triggers the end of transcription. Since most terminators are found upstream of the site at which termination actually takes place, it is believed that transcription does not stop abruptly when *polymerase* reaches a terminator. Instead, transcription stops only after the terminator has been transcribed. There are several overlapping events that take place at the terminator site which bring an end to transcription.*RNA polymerase* stops synthesizing RNA; the RNA molecule is released from *RNA polymerase*; the newly synthesized RNA molecule dissociate fully from the DNA; and *RNA polymerase* detaches itself from the DNA template.

Bacterial cell exhibits two major types of terminators based on their requirement of the participation of a protein called rho (*p*) factor. An account of both rho-independent and rho-dependent terminators is given in Box 12.4

Box 12.4: Terminators of Transcription

Rho-independent terminators

Rho-independent terminators (also known as intrinsic terminators) which make up about 50 percent of all terminators in prokaryotes have two common features.

i) They contain inverted repeats (sequence of nucleotides on one strand that are inverted and complementary. Transcription of inverted repeats into RNA results in the formation of a hairpin secondary structure (see Fig. 12.13).

ii) In rho-independent terminators, a string of seven to nine adenine nucleotides are located following the second inverted repeats in the template strand. The transcription of these adenine nucleotides produces a string of uracil nucleotides after the hairpin in the transcribed RNA. The string containing many Uracils in RNA molecule causes the RNA molecule to pause, thus allowing time for the inverted repeats to form a hairpin structure. The formation of hairpin structure eventually leads to destabilization of the RNA-RNA pairing, causing the RNA molecule to separate from its DNA template. This separation is facilitated by the weaker adenine-uracil base pairing as compared to other types of base pairings.

Rho-dependent terminators

These terminators require a rho factor and exhibit two important features while carrying out termination:

i) DNA sequences that produce a pause in transcription and

ii) a DNA sequence that encodes a stretch of RNA upstream of the terminator that is without any secondary structure. This unstructured RNA serves as a binding site for the rho protein, which binds to the RNA and moves towards its 3´ end, following the *RNA polymerase* (see Fig. 12.14). When *RNA polymerase* encounters the terminator, it pauses, allowing rho to catch up. The

helicase property of the rho protein unwinds the RNA-DNA hybrid, bringing transcription to an end.

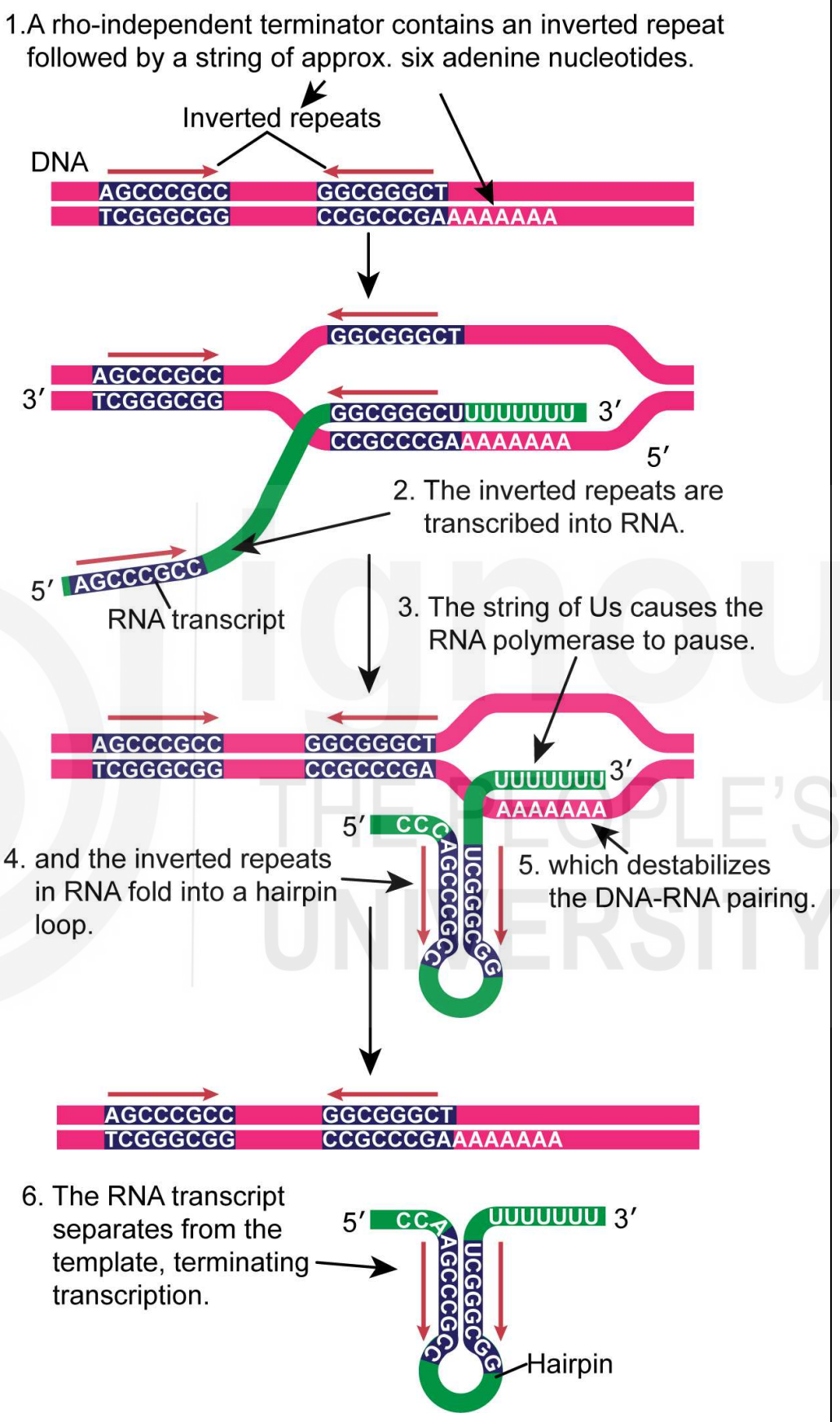


Fig. 12.13: Rho-independent termination in bacteria showing transcription of inverted repeats. The inverted repeats fold into hairpin loop resulting in separation of RNA transcript from the template, terminating transcription.

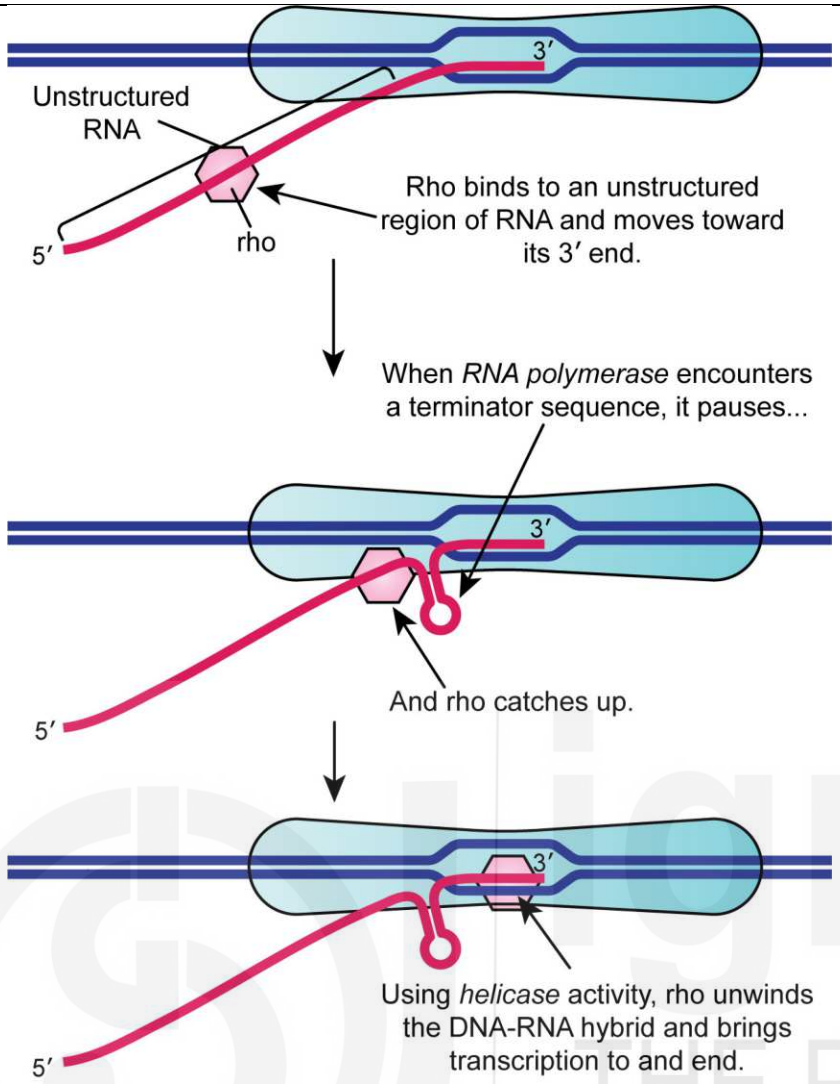


Fig. 12.14: Rho protein dependent termination of transcription in bacterial cells.

12.4.3 Transcription in Eukaryotes

Though much of our knowledge of transcription has been derived from studies in prokaryotes, most of the general aspects of the mechanisms like initiation, elongation and termination are similar in eukaryotic transcription too. However, there are some basic differences (see Table 12.5) .We shall discuss them at the end of this subsection.

Various steps of transcription in eukaryotes are discussed below**.**

Binding of *RNA polymerase* to the promoter and initiation of transcription

There are three different kinds of *RNA polymerases* (*RNA polymerase I, RNA polymerase II* and *RNA polymerase III*). These enzymes recognize different types of promoters. Promoter recognition is in fact, carried out by accessory proteins that bind to the promoter and then assign a specific *RNA polymerase* to the promoter. These accessory proteins comprise general transcription factors, which along with *RNA polymerase*, form basal transcription apparatus, which is a group of proteins that assemble near the start site and are sufficient to initiate minimum level of transcription.

Another group of accessory proteins consists of transcriptional activator proteins which bind to specific DNA sequences and bring about higher level of transcription by activating the basal transcription apparatus at the start site.

As *RNA polymerase II* is responsible for the synthesis of mRNA from protein-coding genes, it has been the focus of most studies of transcription in eukaryotes. The promoter of genes transcribed by RNA polymerase II contains several different sequence elements surrounding their transcription sites. These sequences are categorized into two groups:

1) Core promoter

There are at least four types of DNA sequences that are involved in core promoter function (see Fig.12.15). These four elements are:

i) a short initiator (Inr) sequence surrounding the transcription start point;

ii) a Goldberg-Hogness, or TATA box which consists of consensus sequence of TATA followed by two or three more A's, usually located 25 to 30 nucleotide upstream (-30) of the start site. The sequence and function of TATA boxes are analogous to those found in the – 10 promoter region of prokaryotic genes. However, unlike prokaryotes, *RNA polymerase II* cannot bind directly to this promoter region;

iii) TFIIB recognition element (BRE), which is located upstream of the transcription start site: and

iv) the downstream promoter element (DPE) located about 30 nucleotides downstream of the start point.

These four elements are organized into two general types of core promoters; TATA- driven promoter which includes an Inr sequence and TATA box with or without associated BRE, and DPE-driven promoters, which contain DPE and Inr sequences, but no TATA box or BRE.

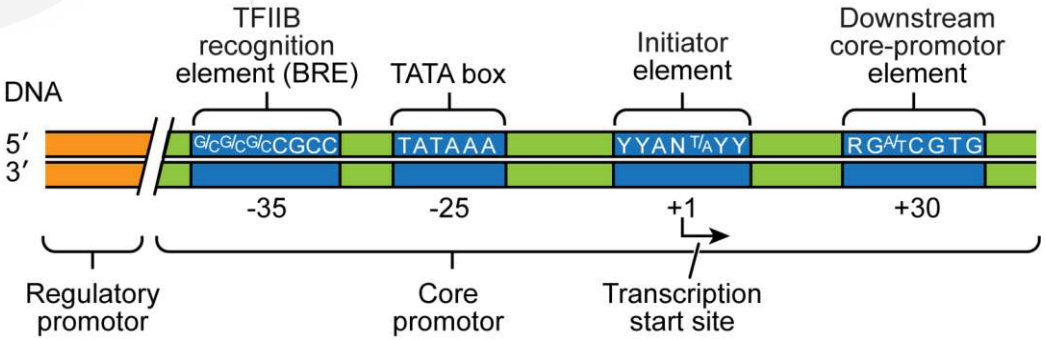


Fig.12.15: The promoters of genes transcribed by *RNA polymerase II* with core promoter and regulatory sequences.

2) Regulatory sequences

Regulatory DNA sequences include **upstream elements**, proximal promoter elements, enhancers and silencers. A core promoter (TATA-driven or DPE-driven) is capable of supporting only a basal (low level) of transcription. However, most protein–coding genes have additional short sequences further upstream called as regulatory sequences that improve promoter's efficiency.

Some of the upstream elements are common to many different genes; examples include CAAT box (consensus sequence GCCCAATCT in animals and yeasts) and the GC box (consensus sequence GGGCGG). The locations of these elements relative to a gene's start point vary from gene to gene. The elements within 100 – 200 nucleotides of the start point are often called **proximal control elements**. Other elements known as **enhancers** and **silencers** have more dramatic effects on eukaryotic gene transcription. As their name suggests, enhancers increase transcription levels and silencers decrease them. They change transcription in response to a cell's requirement for a particular gene product, or at a particular time during the development or place within the organism. Each eukaryotic gene has its own unique arrangement of proximal promoter, enhancer and silencer. The location of these elements can vary from immediately upstream from the promoter to downstream, within, or kilobases away from a gene. Thus, they can modulate transcription even from a distance.

Transcription is initiated through the assembly of transcriptional machinery on the promoter. This machinery consists of *RNA polymerase II* and a series of transcriptional factors that together form a giant complex consisting of 50 or more polypeptides. Assembly of transcription machinery begins when regulatory proteins bind DNA near the promoter and modify the chromatin structure so that transcription can take place. These proteins and other regulatory proteins then recruit basal transcriptional apparatus to the core promoter. The basal transcriptional apparatus consists of *RNA polymerase*, a series of general transcriptional factors, and a complex of proteins known as mediator. The general transcription factor (GTF) includes TFIIA, TFIIB, TFIIE, and TFIIH (TF stands for transcription factor for *RNA polymerase II* and the final letter designates the individual factor). *RNA polymerase II* and transcription factors assemble at the core promoter, forming a **pre-initiation complex** that is analogous to closed complex seen in bacterial initiation. The role of sigma factor that recognizes and binds to the promoter in prokaryotes is replaced by general transcription factor in eukaryotes. Initiation begins with binding of general transcription factor to the promoter in a defined way. TFIID binds directly to the TATA- box sequence. TFIID consists of at least nine polypeptides. One of them is TATA-binding protein (TBP), which recognizes and binds to the TATA consensus sequence. The other transcription factors interact primarily with each other. Hence protein to protein interaction plays a crucial role in the binding stage. After the *RNA polymerase* and transcription factors have assembled on the core promoter, conformational changes take place both in DNA and the *polymerase*. These changes cause 11 to 15 bp of DNA surrounding the transcription start site to separate, thus producing single stranded DNA that will serve as template for transcription. The single-stranded DNA template now positions itself within the active site of RNA polymerase, creating structure called the **open complex**. Once the open complex is formed, the synthesis of RNA begins as phosphate groups are cleaved off nucleoside triphosphate and nucleotides are joined together to form an RNA molecule (see Fig. 12.16). As in bacterial transcription, there are several abortive transcriptions before *polymerase* initiates the synthesis of a full-length RNA molecule.

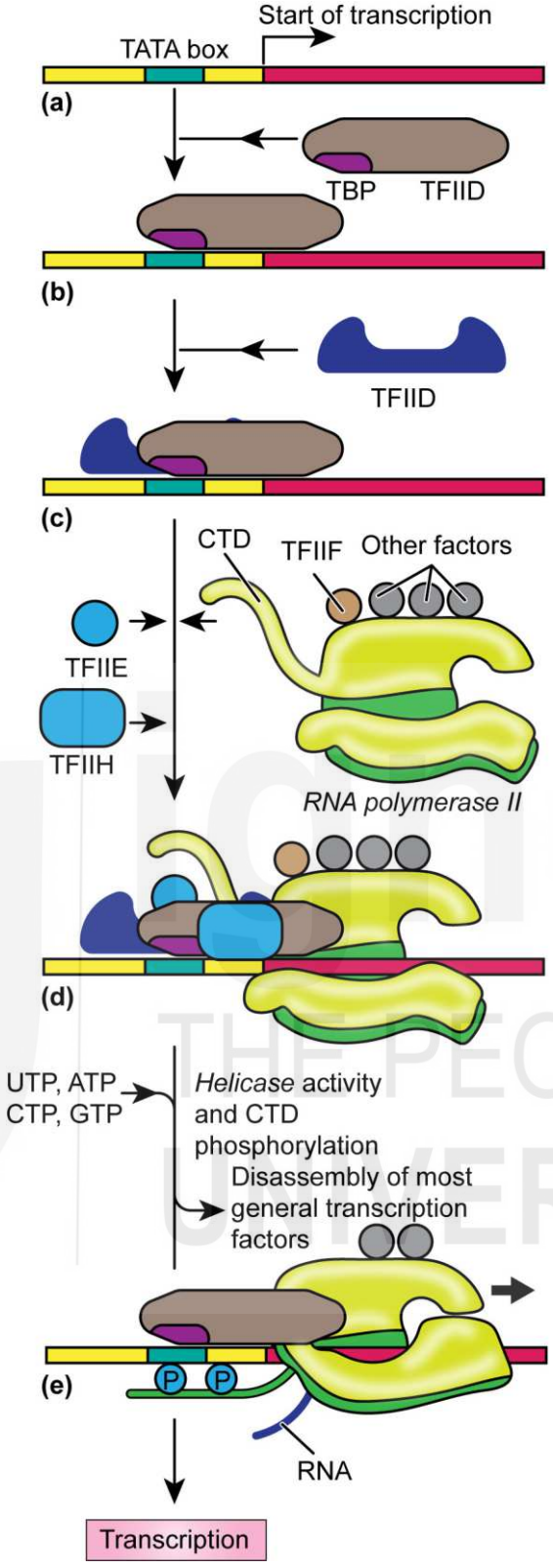


Fig. 12.16: Different steps of Initiation of transcription of a eukaryotic gene by *RNA polymerase II* a) TATA box on the promoter; b) Subunit TBP of TFIID recognizes TATA box and bind, which enables the adjacent binding of TFIID; c) The rest of the general transcription factors as well as RNA polymerase II itself assemble at the promoter; d) TFIIH utilizes ATP to pull apart DNA double helix at the transcription start point, locally exposing the template strand; e) TFIIH also phosphorylates *RNA polymerase II*, changing its conformation so that *polymerase* is released from the general factor and can begin the elongation phase of transcription. At the site of phosphorylation is a long C – terminal domain (CTD) that extends from the *polymerase* molecule.

The general transcription factors and its roles in initiation of eukaryotic transcription have been summarized in Table 12.3.

Table 12.3: General transcription factors needed for transcription initiation by *polymerase II*.

NAME	NUMBER OF SUBUNITS	ROLES IN INITIATION OF TRANSCRIPTION
TFIID		
TBP subunits	1	recognizes TATA box
TAF subunits	~11	recognizes other DNA sequences near the transcription start point; regulates DNA-binding by TBP
TFIIB	1	recognizes BRE element in promoters; accurately positions *RNA polymerase* at the start site of transcription
TFIIF	3	stabilizes *RNA polymerase* interaction with TBP and TFIIB; helps attract TFIIE and TFIIH
TFIIE	2	attracts and regulates TFIIH
TFIIH	9	unwinds DNA at the transcription start point, phosphorylates Ser5 of the *RNA polymerase* CTD; releases *RNA polymerase* from the promoter

(TFIID is composed of TBP and - 11 additional subunits called TAFs (TBP associated factors); CTD (C-terminal domain)

Elongation and Termination of transcription

Once 30 bp of RNA have been synthesized, *RNA polymerase* leaves the promoter and enters into elongation phase. Most of the transcription factors are detached at the promoter and can quickly assist other *RNA polymerases* to continue transcription. Roger Kornberg (Nobel Prize in Chemistry in 2006) and his colleagues worked out the molecular structure and mode of functions of *RNA polymerase* II during elongation. Transcription bubble is maintained by *RNA polymerase* during elongation. The DNA double helix enters a cleft in the p*olymerase* which is gripped by jaw like extension of the enzyme. The two strands of the DNA are unwound and RNA nucleotides that are complementary to the template strand are added to the growing 3´ –end of the RNA molecule. DNA-RNA hybrid hits a wall of amino acids and bends at almost right angle while passing through the funnels of the *polymerase*. This bend positions the end of the DNA-RNA hybrid at the active site of the *polymerase*, and new nucleotides are added to the three end of the growing RNA molecule. The newly synthesized RNA molecule is separated from the DNA and runs through another groove before exiting from the *polymerase* (Fig. 12.17).

The termination of transcription by *RNA polymerase* II does not take place at any specific sequences. *RNA polymerase* II often continues to synthesize additional RNA, hundreds or even thousands of nucleotides past the coding sequences necessary to produce the mRNA. While the end of pre-mRNA is

cleaved at specific site, designated by consensus sequence, the transcription still continues to take place at the 3′ – end of the molecule. The pre-mRNA is cut into two pieces by cleavage: (i) the mRNA that will eventually encode the protein and (ii) and portion of RNA that has its 5′ end trailing out of the RNA polymerase. An enzyme called *Rat* I attaches to the 5′ end of this RNA and moves towards the 3′- end where *RNA polymerase* continues the transcription of RNA. *Rat* I is a *5′→3′ exonuclease* enzyme capable of degrading RNA in 5′→3′ direction. *Rat* I comes near the *polymerase*, and starts breaking up the RNA as it moves. Transcription stops when *Rat* I reaches the transcriptional machinery.

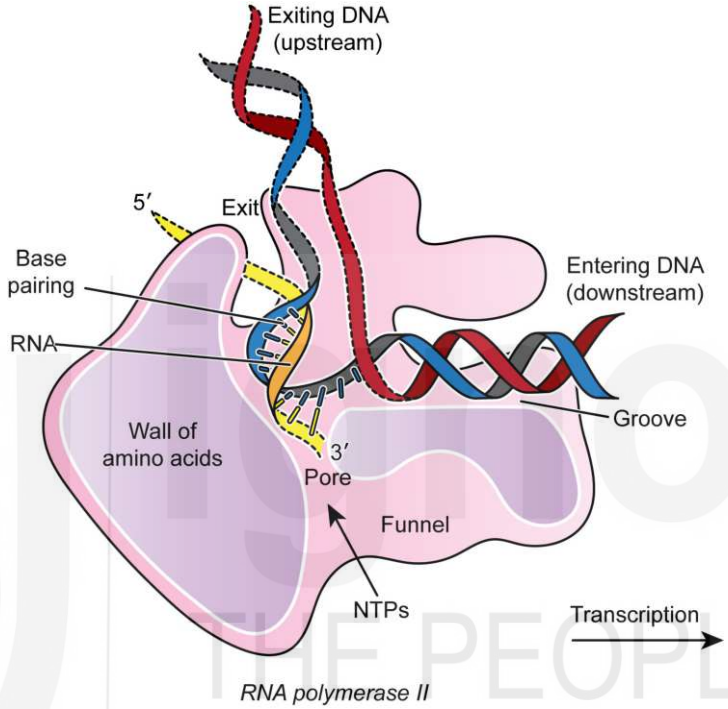


Fig. 12.17: A 3-D model of *RNA polymerase*.

The basic differences between prokaryotic and eukaryotic transcription has been summarized in Table 12.4.

Table 12.4: Basic differences between prokaryotic and eukaryotic transcription.

Transcription in prokaryotes	**Transcription in eukaryotes**
1. There is no spatial separation of transcription and translation since prokaryotes lack nucleus. Transcription is carried out by single *RNA polymerase*.	1. Transcription in eukaryotes occurs in nucleus under the direction of three separate form of *RNA polymerase*. For the mRNA to be translated, it must move out from the nucleus into the cytoplasm.
2. For initiation of transcription of prokaryote genes, DNA remains accessible for *RNA polymerase* and other accessory proteins since DNA is not complexed with histone proteins in highly compressed chromatin.	2. Initiation of transcription in eukaryotic genes requires compact nucleosomes coiling to be uncoiled and DNA to be made accessible to *RNA polymerase* and other accessary proteins. This transition refers to as histone modification and chromatin remodeling

3.	3. For promoter recognition in prokaryotic cells the holoenzyme (*RNA polymerase* plus the sigma factor) recognizes and binds directly to the sequences in the promoter	3.	In eukaryotic cells the promoter recognition is carried out by accessary proteins called as general transcription factors that bind to the promoter and then recruit a specific *RNA polymerase* (I, II or III) to the promoter. In addition to promoter other control units in eukaryotes are *enhancer* and *silencer.*
4.	In prokaryotes the primary RNA transcripts does not require being processed (removal of non-coding sequences- introns) before being translated because prokaryotic genes mostly lack introns.	4.	In eukaryotes the primary transcripts are generally much larger than those that are eventually translated into protein known as pre-mRNA or hnRNA (heterogeneous nuclear RNA) that need to be processed before being transported to the cytoplasm for translation.

So far, you have studied in detail about the mechanism of bacterial and eukaryotic transcription. Let us now turn our attention to the processing and modification of eukaryotic mRNA, the immediate product of transcription. But before we do that, how about trying some SAQ first.

SAQ 3

a) Put the following in correct order for over all transcription process in both prokaryotes and eukaryotes. Indicate this order by putting 1, 2 …… in the boxes given against each statement.

i) Elongation of the RNA chain [ ]

ii) Initiation of RNA synthesis [ ]

iii) Termination of RNA synthesis [ ]

iv) Release of RNA transcript [ ]

v) Binding of *RNA polymerase* to a promoter sequence [ ]

b) Fill in the blanks with the suitable word from the words given below.

i) The three critical regions of a transcription unit are: a promoter, an …… region and a terminator.

ii) The sigma (σ) factor of bacterial *RNA polymerase* controls the binding of *RNA polymerase* to the ……. .

iii) The termination of transcription in bacteria with the formation of hairpin loop structure is called…….. termination.

iv) The type of *RNA polymerase* responsible for the synthesis of mRNA from the protein-coding genes in eukaryotes is called …….. .

v) The type of RNA polymerase that synthesize precursor of tRNA is …….. .

(*RNA polymerase* I, promoter, mRNA-coding, rho-independent, *RNA polymerase* II)

12.5 PROCESSING AND MODIFICATION OF RNA

Although transcription is the first and most highly regulated step in gene expression, it is usually only the beginning of the series of events required to produce functional RNA. Most of the newly synthesized RNAs must be modified in various ways to be converted to their functional forms. You have studied that the transcription and translation in bacterial cells take place simultaneously and hence they are coupled. Therefore, bacterial mRNA has little opportunity to be modified before protein synthesis. In contrast, transcription and translation are separated in both time and space in eukaryotic cells. Transcription takes place in nucleus and translation in cytoplasm. Substantial processing is required in the nucleus to convert primary transcripts into mature mRNA molecules that are ready to be transported to the cytoplasm and translated. Primary transcripts are often very long, ranging from 2000 to 20,000 nucleotides. This heterogeneity in size is reflected in the term heterogeneous nuclear RNA (hnRNA). hnRNA consists of a mixture of mRNA molecules and their precursors i.e. pre-mRNA. The processing of mRNA includes modification of both ends of the initial transcripts as well as alteration of protein-coding section of RNA molecule. The C-terminal domain (CTD) of *RNA polymerase* II plays a key role in coordinating these processes by serving as binding site for the enzyme complexes involved in mRNA processing. The association of these processing enzymes with CTD of *polymerase* II accounts for their specificity in processing mRNAs. *RNA polymerase* I and III lack CTD, so their transcripts are not processed by the same enzyme complexes.

The mRNA processing in eukaryotes involves (i) addition of addition of 5´ –cap at its 5´ – end (ii) addition of poly (A) tail at its 3´ – end, (iii) RNA splicing, and (iv) RNA editing.

12.5.1 Addition of the 5´ Cap

The first step in mRNA processing is the modification of the 5´ end of the transcript by the addition of a structure called a 7-methylguanosine cap.

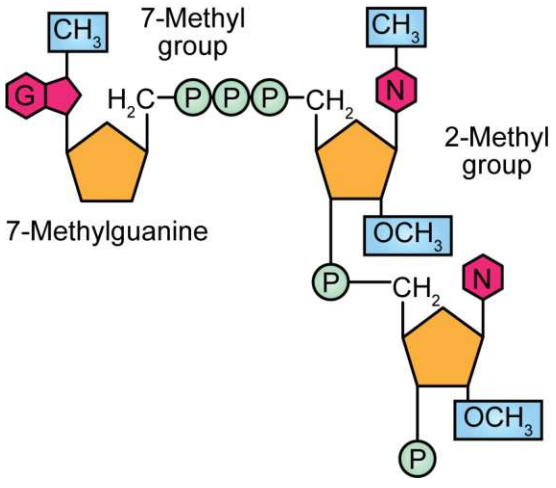


Fig.12.18: Eukaryotic mRNA 5´ cap consisting of nucleotide with 7 – methylguanine attached to the pre-mRNA by a unique 5´ – 5´ bond.

The enzyme responsible for capping are recruited to the phosphorylated CTD following initiation of transcription, and the cap is added after transcription of

the 20 to 30 nucleotides of the RNA. Capping is initiated by the addition of a guanosine triphosphate (GTP) in reverse orientation to the 5´ nucleotides of the RNA. Then methyl groups are added to the G residues and to the ribose moieties of one of the 5´ nucleotides of the RNA chain (see Fig. 12.18). The 5´ cap contributes to mRNA stability by protecting the molecule from degradation by nucleases that attack RNA at the 5´ end. The 5´ cap also plays an important role in positioning and stabilizing mRNA on the ribosome for initiation of translation.

12.5.2 Addition of Poly (A) Tail at 3´ End

In addition to the 5´ cap, a poly (A) tail ranging from 50 – 250 nucleotides in length is present at the 3´ end of the most of the eukaryotic mRNA molecules. The mRNAs coding for the major histones in animal cells are among the few mRNAs known to lack such a poly (A) tail. The poly (A) tail is always added after transcription as genes do not contain long stretches of thymine (T) nucleotides that could serve as a template for the addition of poly (A) tail. This process of addition of poly (A) tail is known as adenylation (see Fig. 12.19).The isolation of the enzyme *poly (A) polymerase*, which catalyzes the addition of poly (A) sequences to RNA without requiring a DNA template has given direct support to the above conclusion. The addition of poly (A) is a part of the process that creates the 3´ end of the most eukaryotic mRNA molecules. Processing of the 3´ end of pre-mRNA requires sequences both upstream and downstream of the cleavage site. Unlike bacteria, where specific termination sequences halt transcription at 3´ end of newly forming mRNAs, the transcription of eukaryotic pre-mRNAs often proceeds hundreds or even thousands of nucleotides beyond the site destined to become 3´ end of the final mRNA molecule. A special signaling (consensus) sequence AAUAAA is usually located slightly upstream of the cleavage site and a sequence rich in uracil (U) nucleotides or GU-rich nucleotides downstream of the cleavage site determines where the poly (A) tail should be added. This signaling element triggers cleavage of the primary transcript 10 -35 nucleotides downstream from the AAUAAA sequence and *poly (A) polymerase* catalyses the formation of poly (A) tail.

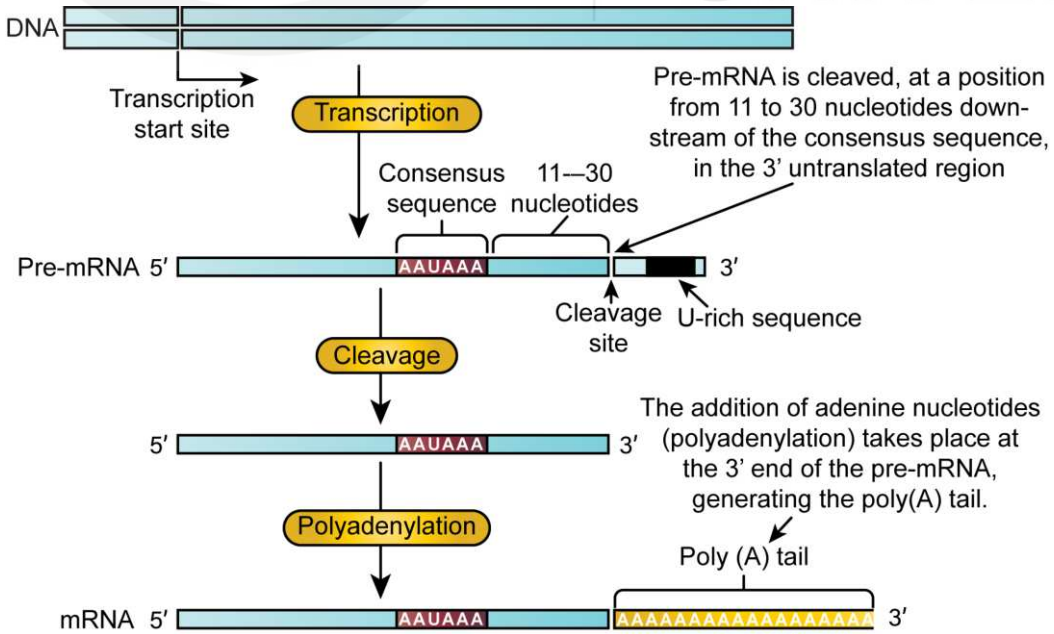


Fig.12.19: Eukaryotic mRNA with 3´ poly (A) tail. In pre-mRNA processing poly (A) tail is added through cleavage and polyadenylation.

Like the 5´ cap, poly (A) tail protects mRNA from *nuclease* attack and, as a result, the length of the poly (A) influences mRNA stability (the longer the tail, the longer the life span of the mRNA in the cytoplasm). In addition, poly (A) is recognized by specific proteins involved in exporting mRNA from the nucleus to the cytoplasm, and it may also help ribosomes recognize mRNA as molecule to be translated. Figure 12.20 shows a mRNA with both 5´ cap and poly(A) tail.

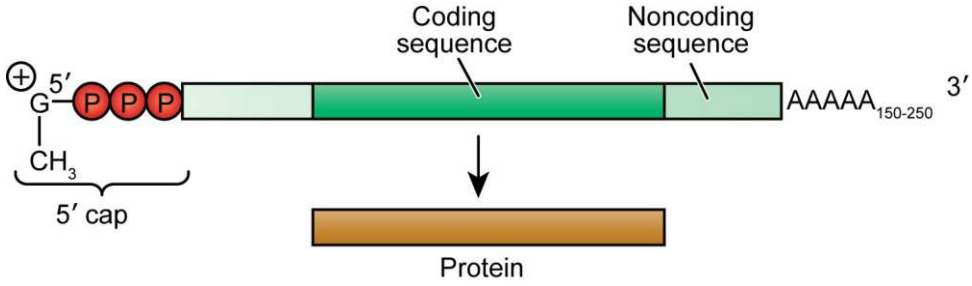


Fig.12.20: Formation of eukaryotic mRNAs by addition of a 5´ cap, cleavage of pre-mRNA transcript, and the addition of poly (A) tail.

12.5.3 RNA Splicing

One of the most exciting breakthroughs in the history of molecular genetics occurred in 1977, when internal DNA sequences in eukaryotes were discovered that are not expressed in the amino acid sequence of the protein they code. These internal DNA sequences are represented in initial RNA transcripts, but they are removed before the mature mRNA is translated. Such nucleotides segments are called **intervening sequence** and the genes that contain them are called **split genes**. DNA sequences that are not represented in the final mRNA product are also called **introns** and those retained and expressed are called **exons** (see Fig.12.21). Although the vast majority of eukaryotic genes examined thus far contain introns, there are several exceptions. Interestingly, the genes coding for histones and interferon lack introns. It is not clear why or how the genes encoding these molecules have been maintained throughout evolution without acquiring introns which is the characteristics of all other genes.

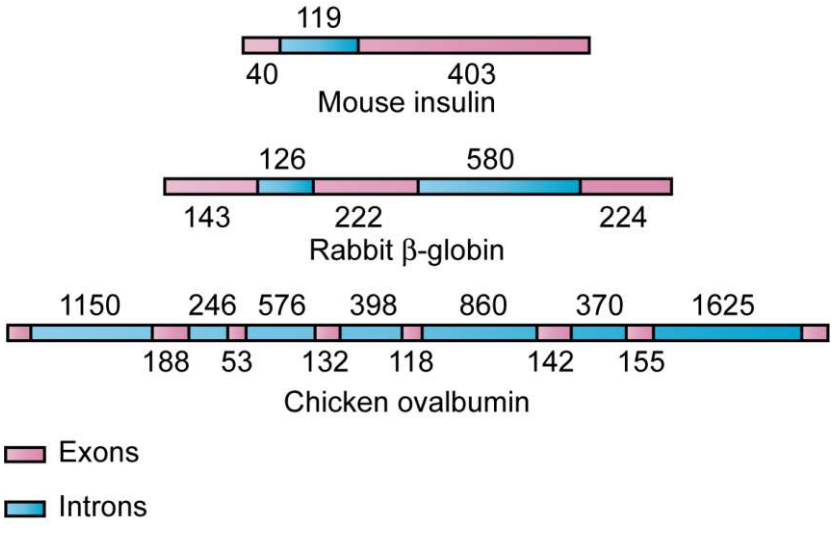


Fig.12.21: Split genes in eukaryotes. The intervening sequences contain various exons and introns. The numbers indicate the number of nucleotides present in various introns and exons regions.

Why do majority of the eukaryotic genes contain introns that need to be spliced out to form the final mRNA product? Do these introns serve any useful purpose to the cell? You can find answers to these questions in the Box 12.4.

The most striking modification of pre-mRNAs therefore, is the removal of introns through splicing and rejoining together the remaining RNA segments (exons).

Splicing requires the presence of three **consensus sequences** in the intron**.** One end of the intron is referred to as the **5´ splice site**, and the other end is the **3´ splice site.** These splice sites possess short consensus sequences (see Fig. 12.22). Most introns in pre-mRNA begin with GU and end with AG. This is indicative of the fact that these sequences play crucial role in splicing. The third important requisite for splicing is at the branch point, which is an adenine nucleotide that lies from 18 to 40 nucleotides upstream of the 3´ splicing site. The sequences surrounding the branch point does not have strong consensus. The deletion or mutation of the adenine nucleotide at the branch point will prevent splicing. Splicing takes place within a large structure called **spliceosome**, which is one of the largest and most complex of all molecular complexes. The spliceosome consists of five RNA molecules and almost 300 proteins. The RNA components are small nuclear RNAs (snRNAs) ranging in length from 107 to 210 nucleotides. These snRNAs associate with proteins to form **snRNPs** (small ribonucleoproteins). Each snRNPs (pronounced “snurp”) contains a single RNA molecule and multiple proteins. The spliceosome is composed of five snRNPs (U1, U2, U4, U5 and U6), and some proteins not associated with an snRNA.

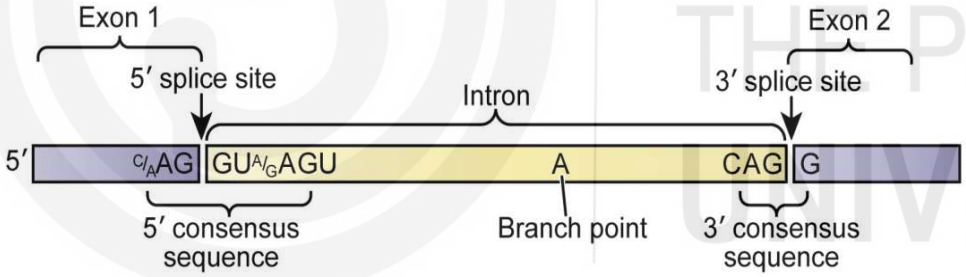


Fig. 12.22: Pre-mRNA showing consensus sequences required for splicing.

Many human genetic diseases arise from mutations that affect pre-mRNA splicing. Approximately, 15% of the single-base substitutions that result in human genetic diseases alter pre-mRNA splicing. Some of these mutations interfere with the recognition of the normal 5´ and 3´ splicing sites. Others create new splice sites.

Self-splicing introns

Although the participation of spliceosomes is almost always required for intron removal, there are a few exceptions. Few types of genes possess self-splicing RNA introns. The RNA transcript of such a gene is capable of carrying out the entire process of RNA splicing in the absence of any protein. The intron RNA itself catalyses the process. As you have studied earlier in this unit, such RNA molecules that function as catalysts in the absence of protein are called ribozymes.

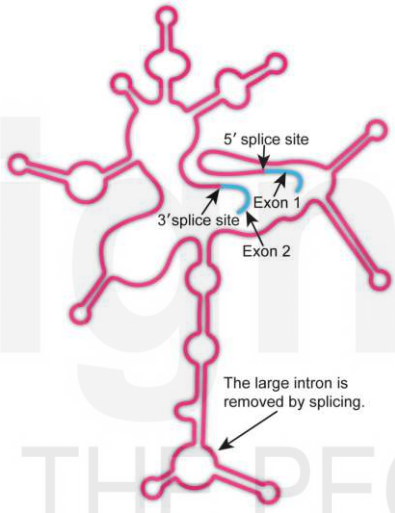
Box 12.5: More about Ribozymes

There are two classes of introns where the intron RNA can function as ribozyme and catalyzes its own removal:

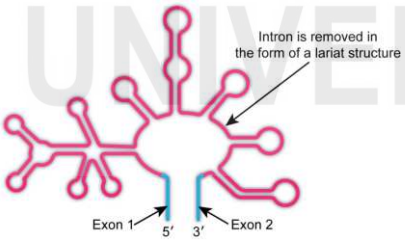
Group I introns: The first class of introns is group 1 introns which is present in the mitochondrial genome of fungi, in rRNA genes and in the genes coding for its electron transport chain. Group 1 introns also occur in plant mitochondrial genes, in some

rRNA and tRNA genes of chloroplasts, in nuclear rRNA genes of some unicellular eukaryotes, in also some tRNA genes in bacteria and in few bacteriophage genes coding for mRNAs. Although the lengths of group 1 introns vary, all of them fold into a common secondary structure with nine looped structure (see Fig. 12.23 a), and are necessary for splicing.

Group II introns: In contrast to group I introns, the splicing of group II introns is accomplished by a mechanism that has some similarities to the spliseosomal-mediated splicing of nuclear genes, and splicing generates lariats where an adenine within the intron forms the branch point, just as in spliceosome mechanism as shown in Figure 12.23 b. Group II introns are found in some mitiochondrial and chloroplasts genes of plants and unicellular eukaryotes and in the genomes of some archaea and bacteria. Because of these similarities, group II introns and nuclear pre-mRNA introns have been suggested to be evolutionary related. Biologists think that today's prevailing splicing mechanism, based on spliceosomes, evolved from group II introns, with the intron RNA's catalytic role being taken over by the snRNAs molecules of the spliceosomes.



(a)



(b)

Fig. 12.23: a) Group I introns; b) Group II introns.

Group I and Group II introns fold into characteristic secondary structure

12.5.4 RNA Editing

Now you will be introduced to another very interesting aspect. It has been believed since long that all information about the amino acid sequence of a protein resides in DNA. This may not always be true. The process of RNA editing weakens this assumption. In RNA editing, the coding sequence of an mRNA molecule is altered after transcription, and so the protein has an amino acid sequence that differs from that encoded by the gene. RNA editing was first detected in 1986 when the coding sequences of mRNAs were compared with the coding sequences of the DNAs from which they had been transcribed.

In some nuclear genes in mammalian cells and some mitochondrial genes in plant cells, there had been substitutions in some of the nucleotides of mRNAs. The mitochondrial gene in trypanosome parasites (which causes African sleeping sickness) undergoes extensive RNA editing. Different types of RNA editing have now been observed in mRNAs, tRNAs, and rRNAs from the wide range of organisms that include insertion and deletion of the nucleotides and conversion of one base to another. A variety of mechanisms can bring about changes in RNA sequences. In some cases, molecules called **guide RNAs** (gRNAs) play a crucial role. A gRNA contains sequences that are partly complementary to segments of the pre-edited RNA, and the two molecules undergo base pairing in these sequences (Fig. 12.24). After the mRNA is anchored to the gRNA, the mRNA undergoes cleavage and nucleotides are added, deleted, or altered to the template provided by gRNA. A recently discovered RNA editing technique that has revolutionized the field of genome engineering is CRISPER/cas9.

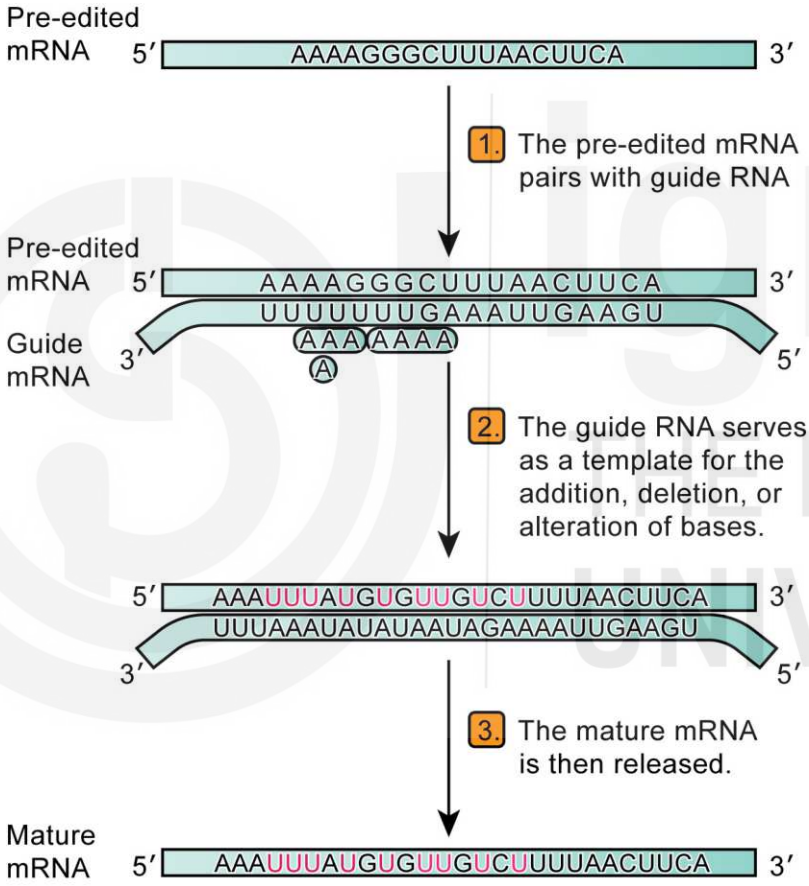


Fig. 12.24: Process of RNA editing by guide RNAs.

We have described above the various mechanisms of RNA modification, processing and its significance. You may like to work out an SAQ based on these before we describe the genetic code.

SAQ 4

Fill in the blanks with the suitable word from the words given below.

a) The first step in mRNA processing is the modification of the 5′ end of the transcript by the addition of a structure called a …… .

b) **……** play a crucial role in RNA editing.

c) In the split genes the sequences that do not contribute to the amino acid sequence of the protein are called …….. .

d) The classes of introns in which the intron RNA can function as ribozyme to catalyse its own removal is called ……. Introns.

(guide RNAs, intron, self-splicing, 7-methylguanosine cap)

12.6 GENETIC CODE

The essence of gene expression lies in the relationship between the nucleotide base sequence of DNA molecule and the linear order of amino acids in protein molecules. This relationship is based on a set of rules known as the genetic code. How information residing in a DNA base sequence could be passed to mRNA through the mechanism of complementary base pairing and how does a base sequence in mRNA use its "message" to guide the synthesis of a protein molecule, which consists of a sequence of amino acids? What is needed, of course, is the knowledge of the appropriate code – the set of rules that determines which nucleotides in mRNA correspond to which amino acids. The cracking of that code, which tells us how DNA can code for proteins, is one of the major landmarks of twentieth-century biology. The awareness of the exact coding relationship between the base sequence of a DNA molecule and the amino acid sequence of a protein arose from the discovery that mutations in DNA can lead to changes in proteins.

Features of genetic code

The genetic code is a triplet code: One of the first questions about the genetic code to be addressed was how many nucleotides are necessary to specify a single amino acid? The basic unit of genetic code – the set of bases that encode a single amino acid is a codon. The investigations made by many early scientists recognized that codons must contain a minimum of three nucleotides. Each nucleotide's position in mRNA can be occupied by one of the four bases; A, G, C, or U. If a codon consisted of a single nucleotide, only four different codons (A,G, C, and U) would be possible, which is not enough to encode 20 different amino acids commonly found in proteins. If codons were made up of two nucleotides each (i.e., GU, AC and so on), there would be 4 x 4 = 16 possible codons – still not enough to encode all 20 amino acids. With three nucleotides per codon, there are 4 x 4 x 4 = 64 possible codons, which is more than enough to specify 20 different amino acids and that would be far less cumbersome than a four nucleotides per codon, 4 x 4 x 4 x 4 = 256 possible codons. Therefore, a triplet code requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids. Using mutations in bacteriophage, Francis Crick and his colleagues in 1961 confirmed that genetic code is indeed a triplet code.

The genetic code is degenerate and unambiguous

One amino acid is encoded by three consecutive nucleotides in mRNA, and each nucleotide can have one of the four possible bases (A, G, C, and U) at

each nucleotide position, thus permitting $4^3 = 64$ possible codons. Three of these codons are stop codons (UAA, UAG, and UGA), specifying the end of translation. Thus, 61 codons, called sense codons, encode amino acids. Because there are 61 sense codons and only 20 different amino acids commonly found in proteins, the code contains more information than is needed to specify the amino acids and is said to be degenerate code. Degenerate is the term that Francis Crick borrowed from quantum physics, where it describes multiple physical states that have equivalent meaning. The degeneracy of the genetic code means that the amino acids may be specified by more than one codon. Tryptophan and methionine are exceptions as they are encoded by a single codon. All other amino acids are encoded by two codons, and some, such as leucine are specified by six different codons (see Fig. 12.25). Codons that specify the same amino acid are said to be synonymous. It is also clear from the Figure 12.25 below that the genetic code is unambiguous: every codon has one and only one meaning. Base pairing between codon and anticodon in which there is nonstandard pairing, usually at the third (3´) position of the codon and is called as **wobble hypothesis.** This nonstandard pairing allows more than one codon to pair with the same anticodon of the tRNA.

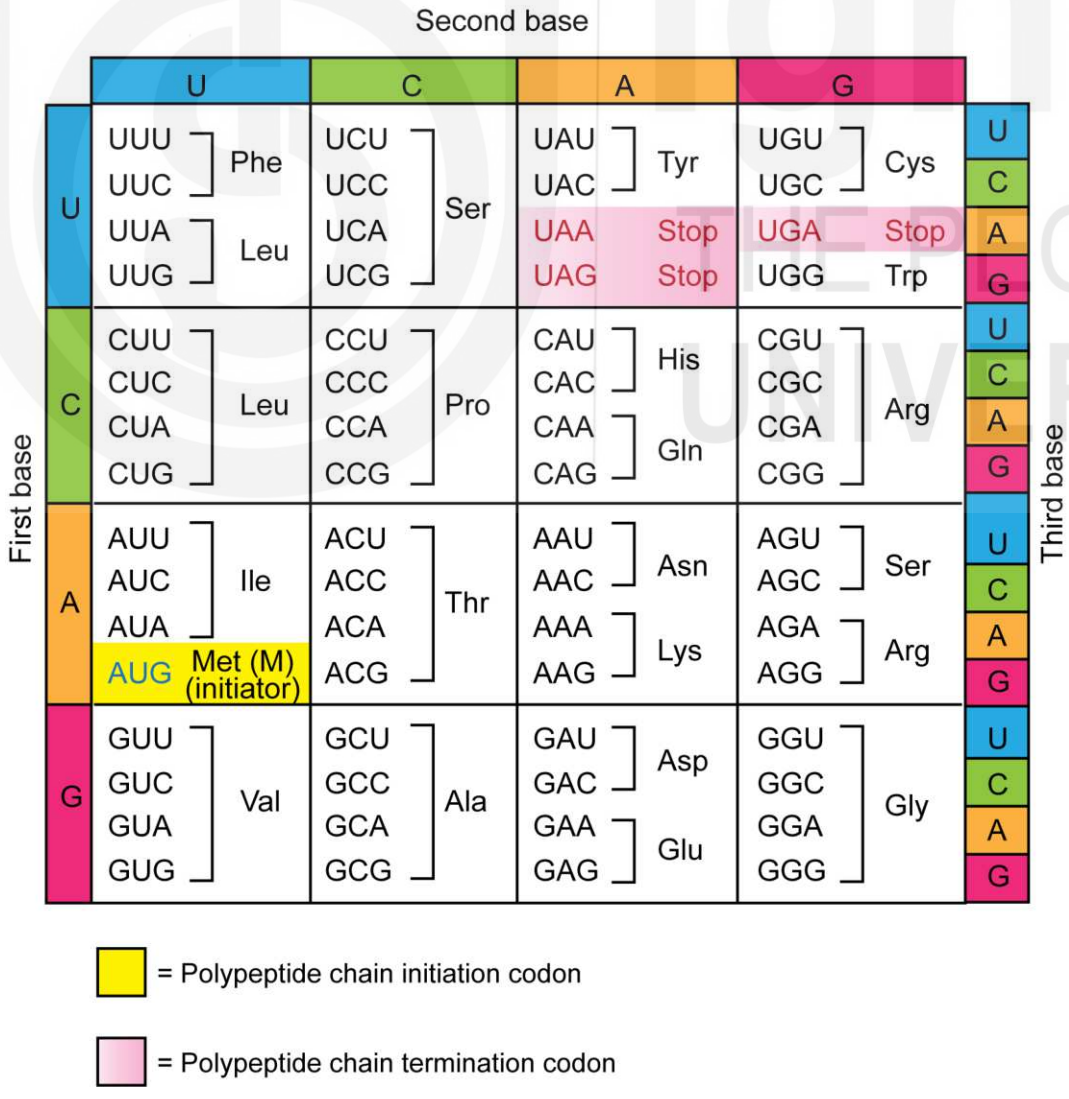


Fig. 12.25: Genetic code consisting of 64 codons. The codons are written in 5´ → 3´ as they appear in mRNA.

Box 12.6: Wobble Hypothesis

The cells of most organisms possess from about 30 to 50 different tRNAs which serve as adaptor molecules, binding particular amino acids and delivering them to a ribosome, where amino acids are assembled to form polypeptide chain. Since, there are only 20 different amino acids in proteins, thus some amino acids are carried by more than one tRNA. Different tRNAs that accept the same amino acid but have different anticodons are called **isoaccepting tRNAs**. Even though some amino acids have multiple (isoaccepting) tRNAs, there are still more codons than anticodons, because different codons can sometimes pair with the same anticodon through flexibility in base pairing at the third position of the codon. Figure 12.26 shows reveals that many synonymous codons differ only in the third position. For example, alanine is encoded by codons GCU, GCC, GCA, and GCG, all of which begin with GC. When the codon on the mRNA and the anticodon of the tRNA join, the first (5′) base of the codon pairs with the third (3′) base of the anticodon strictly according to Watson and Crick base pair ruling: A with U ; C with G. Next the middle bases of codon and anticodon also strictly follow Watson and Crick ruling. After these bases have hydrogen bonded, the third base pairs weakly, there may be flexibility, or wobble, in their pairing. This is called **wobble hypothesis** developed by Francis Crick which proposed that some nonstranded pairings of bases could take place at the third position of a codon (see Fig.12.26). For example, a G in the anticodon may pair with either C or a U in third position of the codon.

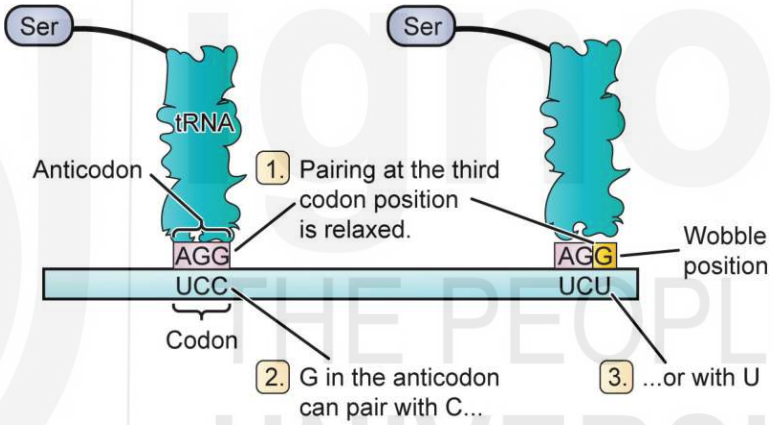


Fig. 12.26: Existence of wobble in pairing of a codon and anticodon.

Genetic code is nonoverlapping and commaless

The findings of Francis Crick, Sydney Brenner, and their colleagues indicated that the genetic code is generally nonoverlapping. An overlapping code is one in which a single nucleotide may be included in more than one codon frame, as for example:

Nucleotide sequence: A U A C G A G U C

Nonoverlapping code: A U A C G A G U C

Ile Arg Val

Overlapping code: A U A U A C A C G……..

Ile Tyr Thr

Generally, each nucleotide is part of a single codon. However, a few overlapping genes are found in viruses, but codons within the same gene do not overlap, and the genetic code is considered nonoverlapping. The different way of reading the sequence of nucleotides is called a reading frame. The

reading frame is set by the initiation codon, which is the first codon of mRNA to specify an amino acid. After the initiation codon, the other codons are read as successive groups of three nucleotides. No bases are skipped between the codons: so there are no punctuation marks to separate the codons. Therefore, genetic code is **comma less/comma free**. If the DNA duplex is mutated by insertion of single base pair, the mRNA will have an additional nucleotide. This insertion alters the reading frame beyond that point, so that remainder of the mRNA is read incorrectly and all the amino acids are wrong.

The genetic code is nearly universal

The final property of the genetic code is its near universality. Except for a few cases, all organisms studied so far - prokaryotes, eukaryotes, viruses use the same basic genetic code. In other words, the 64 codons almost always stand for the same amino acid suggesting that this coding system was established early in the history of life on earth and has remained largely unchanged over billions of year of evolution. However, several exceptions to the standard genetic code do exist, most notably in mitochondria and few bacteria and other unicellular organisms (see Table 12.5). For example one difference involves the codon UGA, which is stop codon in the standard case but is translated as tryptophan in mammalian and yeast mitochondria.

Table 12.5: Some exceptions to the univarsal genetic code.

Genome	Codon	Universal Code	Altered Code
Bacterial DNA			
Mycoplasma capricolum	UGA	Stop	Trp
Mitochondrial DNA			
Human	UGA	Stop	Trp
Human	AUA	Ile	Met
Human	AGA, AGG	Arg	Stop
Yeast	UGA	Stop	Trp
Trypanosomes	UGA	Stop	Trp
Plants	CGG	Arg	Trp
Nuclear DNA			
Tetrahymena	UAA	Stop	Gln
Paramecium	UAG	Stop	Gln

SAQ 5

Choose the correct answers from the following:

a) What does the term ‘degenerate’ mean with regard to the genetic code?

i) The genetic code is universal

ii) The code is nonoverlapping

iii) Some amino acids are specified by more than one codon

iv) The genetic code is a triplet codon

b) Which of these combinations are the termination codons in the standard genetic code?

i) UAA, UAG, UGA

ii) UGA, UGC, UGG

iii) UAG, UGA, UGG

iv) UAA, UAC, UAG

c) The amino acids which are coded by single codon are:

i) Tryptophan and Methionine

ii) Tryptophan and Leucine

iii) Methionine and Serine

iv) Methionine and Arginine

12.7 SUMMARY

In this unit you have learnt that:

- Like DNA, RNA is a polymer consisting of nucleotides joined together by phosphodiester bonds. RNA molecule consists of a ribose sugar, a phosphate backbone and a nitrogenous base. In place of thymine, RNA contains uracil and is usually single stranded, which allows it to form secondary structures. The secondary structure helps RNA to participate in a variety of cellular functions.
- An early world was full of RNA and used RNA as both carrier of genetic information and as biological catalysts.
- Cells possess a number of different classes of RNA.There is directional flow of genetic information. Instructions stored in DNA are transcribed and processed into molecules of mRNA, rRNA and tRNA for use in protein synthesis.
- Transcription is the process by which a DNA template strand is copied by *RNA polymerase* to produce a complementary molecule of RNA. The transcription unit consists of a promoter, an RNA-coding region, and a terminator.
- The molecular mechanism of transcription in both prokaryotes and eukaryotes involves four stages: binding of *RNA polymerase* to the promoter sequence in the DNA template, initiation of RNA synthesis, elongation of RNA chain, and termination.
- The start site of the transcription is determined by the consensus sequence of the promoter. The single strand of DNA functions as a template strand and RNA synthesis stops after a terminator sequence has been transcribed.

- Bacterial cell possess two types of terminators: rho-dependent and rho-independent terminators. The transcription in eukaryotes involve the same general principles as in bacteria but is more complex. The three nuclear *RNA polymerases* recognize different families of DNA promoter sequences and are specialized for synthesizing different types of RNA.
- The presence of introns (non-coding sequences) interrupts the coding sequences (exons) of genes. The pre-mRNA molecules generally undergo some type of processing, both during and after transcription, before RNA can perform its normal function. The major exception is bacterial mRNA, whose translation often begins before transcription is completed.
- The processing of eukaryotic pre-mRNA is very elaborate involving the addition of a 5′ cap and a 3′ poly (A) tail, as well as removal of introns by spliceosome. The presence of introns and exons allows pre-mRNA molecules to be spliced in more than one way, there by permitting a single gene to produce multiple mRNAs coding for different polypeptides.
- There exist various editing mechanisms by which mRNAs may be altered by the addition, deletion, or modification of nucleotiodes in the coding sequences. These editing mechanisms create mRNAs that code for polypeptides whose amino acid sequences could not have been predicted by knowing the base sequence of the corresponding genes.
- The mRNA is read in units of three bases called codon. The three nucleotides specify a single amino acid therefore, the genetic code is a triplet code. The genetic code is unambiguous, nonoverlapping, degenerate, and nearly universal. Different codons may pair with the same anticodon through wobble, which can exist at the third position of the codon and allows some nonstandard pairing of bases in this position.

12.8 TERMINAL QUESTIONS

1. The following diagram represents DNA that is part of the RNA-coding sequence of a transcription unit. The bottom strand is the template strand. Give the sequence found on the mRNA molecule transcribed from this DNA and identify the 5′ and 3′ ends of RNA.

   5′ –ATAGGCGATGCCA- 3′

   3′ –TATCCGCTACGGT -5′ ← Template strand

2. Write four characteristics of genetic code.
3. Outline the functions of various types of non-coding RNA found in eukaryotic cells.
4. Differentiate between prokaryotic and eukaryotic transcription.
5. Describe the structure of RNA and its different types.

12.9 ANSWERS

Self-Assessment Questions

1. a) False; b) True; c) True; d) False; e) True
2. a) Adapter
   b) Ribosome
   c) 5′ UTR
   d) Spliceosome
   e) RNA interference
3. a) i) Initiation of RNA synthesis
      ii) Binding of *RNA polymerase* to a promoter sequence
      iii) Elongation of the RNA chain
      iv) Termination of RNA synthesis
      v) Release of RNA transcript
   b) i) mRNA-coding
      ii) promoter
      iii) rho-independent
      iv) *RNA polymerase* II
      v) *RNA polymerase* I
4. a) 7-methylguanosine cap
   b) guide RNAs
   c) introns
   d) self-splicing
5. a) Some amino acids are specified by more than one codon
   b) UAA, UAG, UGA
   c) Tryptophan and Methionine

Terminal Questions

1. 5′ –AUAGGCGAUGCCA – 3′.
2. Refer to Section 12.6.
3. Refer to Subsection 12.4.3.
4. Refer to Table 12.5.
5. Refer to Sections 12.2 and 12.3.

UNIT 13

TRANSLATION

Structure

13.1 Introduction

Objectives

13.2 Translation in Prokaryotes

Cast of Characters

Initiation of Translation

Chain Elongation

Termination of Polypeptide Synthesis

13.3 Translation in Eukaryotes

The Cast of Characters

Initiation of Translation

Chain Elongation

Termination of Polypeptide Synthesis

13.4 Post Translational Modification of Proteins

Polypeptide Folding

Protein Cleavage

Glycosylation

Attachment of Lipids

Protein Modification by Small Molecules

13.5 Summary

13.6 Terminal Questions

13.7 Answers

13.1 INTRODUCTION

In the previous unit (Unit 12) you studied the world of RNA (mRNA, tRNA, rRNA and some small RNAs), its structure, types, and diverse role in the cells. You have also come to understand the initial phases of gene expression by exploring how the genetic information stored in the DNA helps in the expression of genes and guides the synthesis of RNA known as transcription. In that unit you also learnt various mechanisms of processing, modification and editing in the base sequences of mRNA molecules that opened door for exploring the cure for different diseases, for sustainable crop productivity and fuel resources. This unit describes how genetic information contained within the order of nucleotides in messenger RNA (mRNA) is interpreted to generate linear sequences of amino acids in the proteins. This process is known as

translation. Thus, translation is the key step in the production of protein molecules, involving decoding of language from the nucleotide base sequence of an mRNA molecule and translating it to the amino acid sequence of a polypeptide chain. During this process, there are three major players, 1) a sequence of mRNA nucleotides, read as triplet codons, which specifies the order in which the amino acids are added to a growing polypeptide chain, 2) Ribosomes which serve as intracellular sites for translation, and 3) tRNA molecules which are the adaptor agents that ensure insertion of correct amino acids at each position in the polypeptide. Our discussion in this unit will mainly focus on translation in bacterial cell, where the mechanisms are especially well understood. The comparable events in the eukaryotic cells are rather similar and the differences are confined mostly in the initiation stage which will be discussed. Lastly, we will discuss how the translated proteins get altered before they can become fully functional (post translational modifications).

Objectives

After studying this unit, you should be able to:

- describe the various stages of translation in prokaryotes;
- understand how the various stages of translation in eukaryotes differ from prokaryotes; and
- appreciate the process of posttranslational modifications of proteins and realize its importance.

13.2 TRANSLATION IN PROKARYOTES

In the previous unit you studied about genetic code that stores information in the form of triplet codons in DNA, and that this information is initially expressed through the process of transcription. However, the final product of gene expression, in most of the instances, is a polypeptide chain which consists of a linear series of amino acids whose sequence has been specified by the genetic code. We will first start surveying the cell's cast of characters for performing translation, and then we will examine the mechanism of each step in detail.

13.2.1 The Cast of Characters

The cellular machinery involved in translating mRNAs into polypeptides includes five major components: 1) Ribosomes, 2) tRNA molecules, 3) *aminoacyl tRNA synthetases*, 4) mRNA molecule, and 5) Protein factors.

1) Ribosomes

- Ribosomes are macromolecular machines that direct the synthesis of proteins. Ribosomes, therefore, serve as protein factories and play crucial role in orienting the mRNA and amino acid- carrying tRNAs in such a manner that the genetic code can be read accurately. They also

help in catalyzing peptide bond formation that links the amino acids into a polypeptide.

- Ribosomes are particles made of ribosomal RNA (rRNA) and ribosomal protein and are present in the cytoplasm. The ribosomes of prokaryotes (archaea and bacteria) are smaller than those of eukaryotes, although many of the ribosomal proteins, translation factors, and tRNAs used by archaea resemble their eukaryotic counterparts more closely than do the comparable components of bacteria.

- Ribosomes are made up of two dissociable subunits called the large and small subunits. The bacterial ribosome has a sedimentation coefficient of about 70S and consists of a 30S small subunit and a 50S large subunit. The larger ribosomal subunit of prokaryotes consists of 23S rRNA, and a 5S rRNA molecule along with 34 proteins. On the other hand, the smaller subunit has a 16S rRNA component and 21 proteins (see Figure 13.1).

- The rRNA component of ribosome performs all the important catalytic functions associated with translation while ribosomal proteins are thought to promote binding of various molecules involved in translation and, in general, to fine-tune the process.

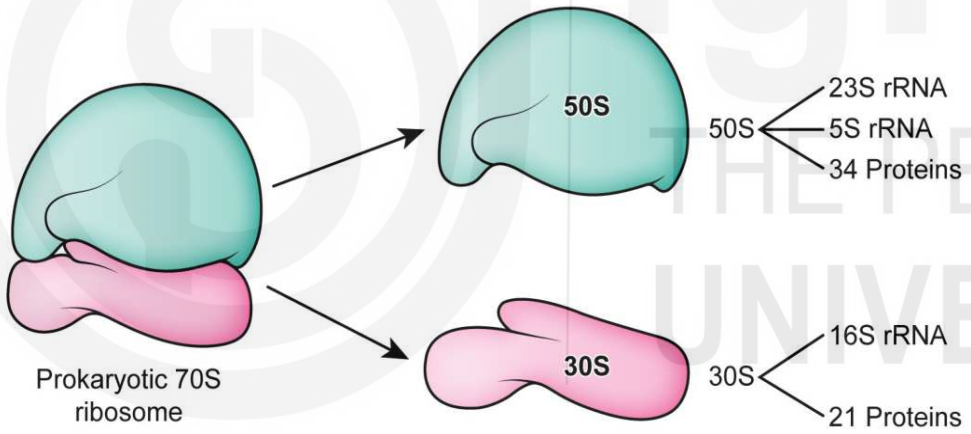


Fig. 13.1: Components of prokaryotic ribosome showing large and small subunits which contain both ribosomal proteins and rRNAs.

- Molecular hybridization studies have revealed that there exists some degree of redundancy of the genes coding for rRNA component. The *E. coli* genome contains seven copies of a single sequence that encode all the three components, viz., 23S, 16S and 5S rRNAs.

- Ribosomes have four important sites which are particularly important for protein synthesis (see Figure 13.2). In addition to an **mRNA binding site** there are three sites where tRNA can bind. The three tRNA binding sites are: an **A (aminoacyl) site,** that binds each newly arriving tRNA with its attached amino acid, **a P (peptidyl) site,** where the tRNA carrying the growing/elongating polypeptide chain resides, and an **E (exit) site**, from where tRNA , after discharging its amino acids, exits the ribosomes.

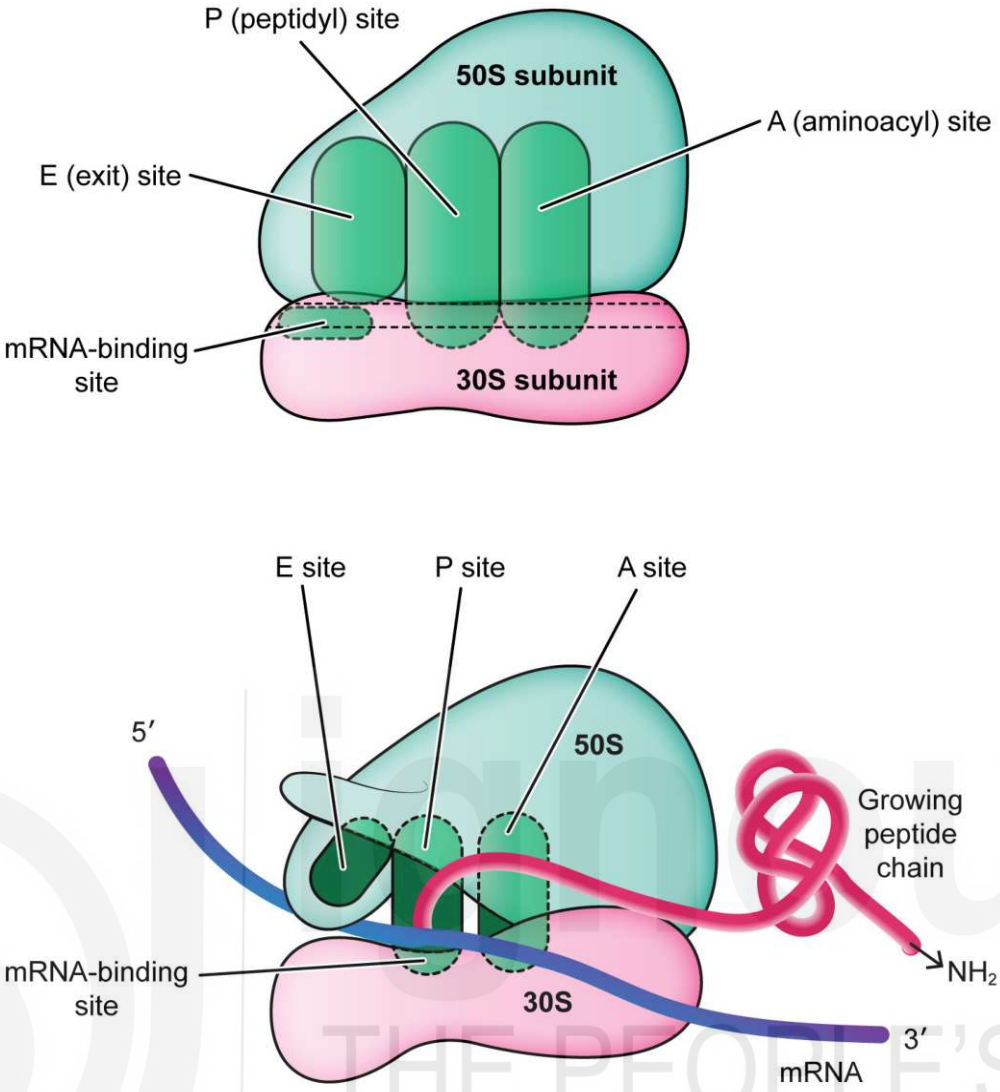


Fig. 13.2: Binding sites on a ribosome showing A (aminoacyl) site, P (peptidyl) site and an E (exit) site. In addition, there is an mRNA- binding site which binds a specific nucleotide sequence near the 5´ end of mRNA.

2. tRNA molecules

You have already studied the structure and function of tRNAs in previous unit 12. Here you are going to learn how tRNA molecules serve as adaptors, enabling sequence of codons in mRNA to ultimately determine the amino acid sequence of polypeptide chain.

- Francis Crick, postulated in early 1957 that as the amino acids cannot directly recognize nucleotide base sequences, a hypothetical ‘’adapter’’ must be present that can mediate between amino acids and mRNA. In the year following Crick’s adaptor hypothesis, Mahlon Hogland, while investigating protein synthesis in cell free system, discovered a family of adaptor molecules that were named as transfer RNAs (tRNAs).

- Transfer RNA plays a vital role as an intermediary between mRNA and amino acids. Each tRNA has two levels of specificity. On one hand, and as specified by genetic code, each tRNA binds to one specific amino acid. On the other hand, each tRNA recognizes one or more mRNA codons specifying that amino acid.

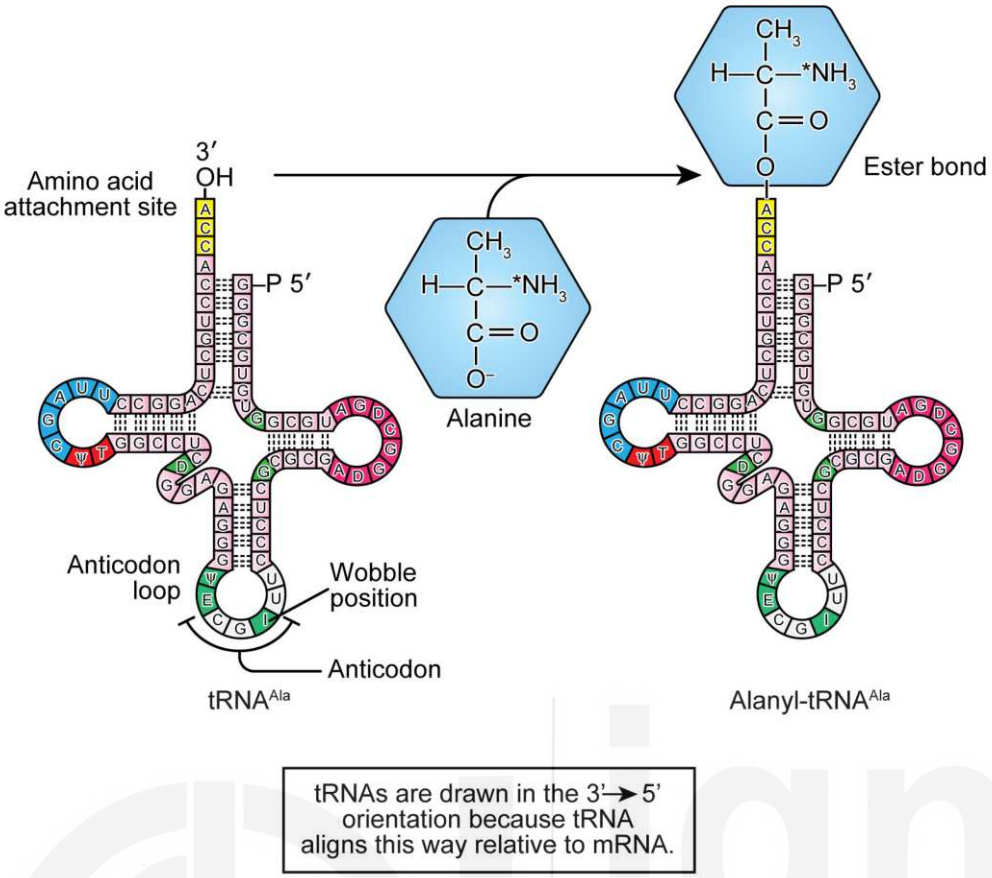


Fig. 13.3: Secondary structure of tRNA, before and after amino acid attachment.

- The tRNAs are linked to their corresponding amino acids by an ester bond. This bond joins an amino acid to the 2' - or 3´ – hydroxyl group of the adenine (A) nucleotide located at 3´ end of all tRNA molecules (see Figure 13.3). The enzymes that catalyze the formation of ester bond also assist in choosing the correct amino acid for each tRNA.
- The name of the amino acid that attaches to given tRNA is indicated by superscript. For example, tRNA molecules specific for amino acid alanine are designated as $tRNA^{Ala}$. After attachment, the tRNA is now called an **aminoacyl tRNA** (e.g., alanyl $tRNA^{Ala}$). The tRNA is now said to be in its charged form, and the amino acid is said to be activated.
- Each tRNA possesses an **anticodon**, which is a special trinucleotide sequence located within one of the loops of the tRNA molecule. The tRNA molecules can recognize codons in mRNA .The anticodon of each tRNA are complementary to one or more mRNA codons that specify the amino acid being carried by that tRNA. The codons in mRNA are represented in the 5´ → 3´ direction, whereas anticodons in tRNA are usually written in the 3´ → 5´ orientation. Thus, if one of the codons for alanine is 5´ –GCC -3´, the corresponding anticodon in tRNA is 3´ – CGG – 5´.

5´ G G C 3´ **mRNA** Codon

tRNA Anticodon 3´ C C G 5´

In the previous unit you have studied in detail about the characteristics of genetic code. You know the number of different tRNAs is significantly

less than the 61 codons employed by genetic code to specify amino acids. This is possible simply because mRNA and tRNA line up on the ribosome in such a manner so as to permit flexibility between the third base of the codon and the corresponding base in the anticodon (wobble hypothesis). It is precisely due to this wobble that fewer tRNA molecules are required for some amino acids than the number of codons that specify those amino acids.

3. *Aminoacyl-tRNA Synthetases*

- *Aminoacyl-tRNA synthetases* are the enzymes responsible for linking amino acids to their corresponding tRNAs. Cells normally have 20 different *aminoacyl-tRNA synthetases*, one for each 20 amino acids commonly used in protein synthesis.
- Number of *aminoacyl-tRNA synthetases* in some cells may be less than 20. In such cases, the same *aminoacyl-tRNA synthetase* may catalyze the attachment of two different amino acids to their corresponding tRNAs, or it may attach an incorrect amino acid to a tRNA molecule. These latter 'errors' are corrected by a second enzyme that alters the incorrect amino acid after it has been attached to the tRNA.
- *Aminoacyl-tRNA synthetases* catalyze the attachment of amino acids to their corresponding tRNA through an ester bond. This is accompanied by the hydrolysis of ATP to AMP and pyrophosphate (Figure 13.4).

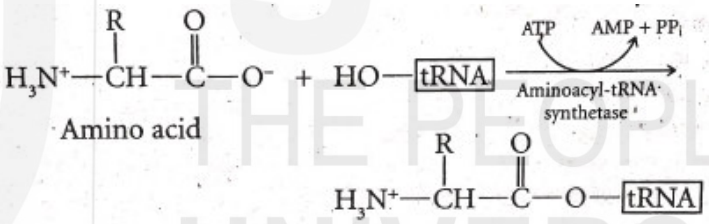


Fig. 13.4: Attachment of amino acid to tRNA via ester bond catalyzed by *Aminoacyl-tRNA synthetase*.

This enzyme catalyzes the formation of an ester bond between the carboxyl group of an amino acid and 3´ OH of the appropriate tRNA, thereby generating an aminoacyl tRNA. This ester bond is thought to be a '' high energy bond'' as the hydrolysis of this bond releases sufficient energy to drive the formation of the peptide bond that will eventually join the amino acids to a growing polypeptide chain. This process of aminoacylation of tRNA is called amino acid activation (see Figure 13.5).

The *aminoacyl-tRNA synthetases* recognize nucleotides located at least two different regions of tRNA molecules where they identify and pick up the correct tRNA that is to become linked to a particular amino acid. Changes in the base sequence of either the anticodon template or the 3´ end of a tRNA molecule will result in a change in the amino acid that will get attached to a tRNA. The *aminoacyl-tRNA synthetases* also perform proofreading function to ensure that the correct amino acid has been incorporated. A site on the *aminoacyl-tRNA synthetase* molecule recognizes incorrect amino acids and releases them by hydrolyzing the bond that links the amino acid to the tRNA. The appropriate codon in mRNA is recognized by tRNA itself once the correct amino acid has

been joined to its tRNA. The specificity of the *aminoacyl-tRNA synthetase* reaction is crucial to the accuracy of gene expression because it ensures that the proper amino acid is linked to each tRNA.

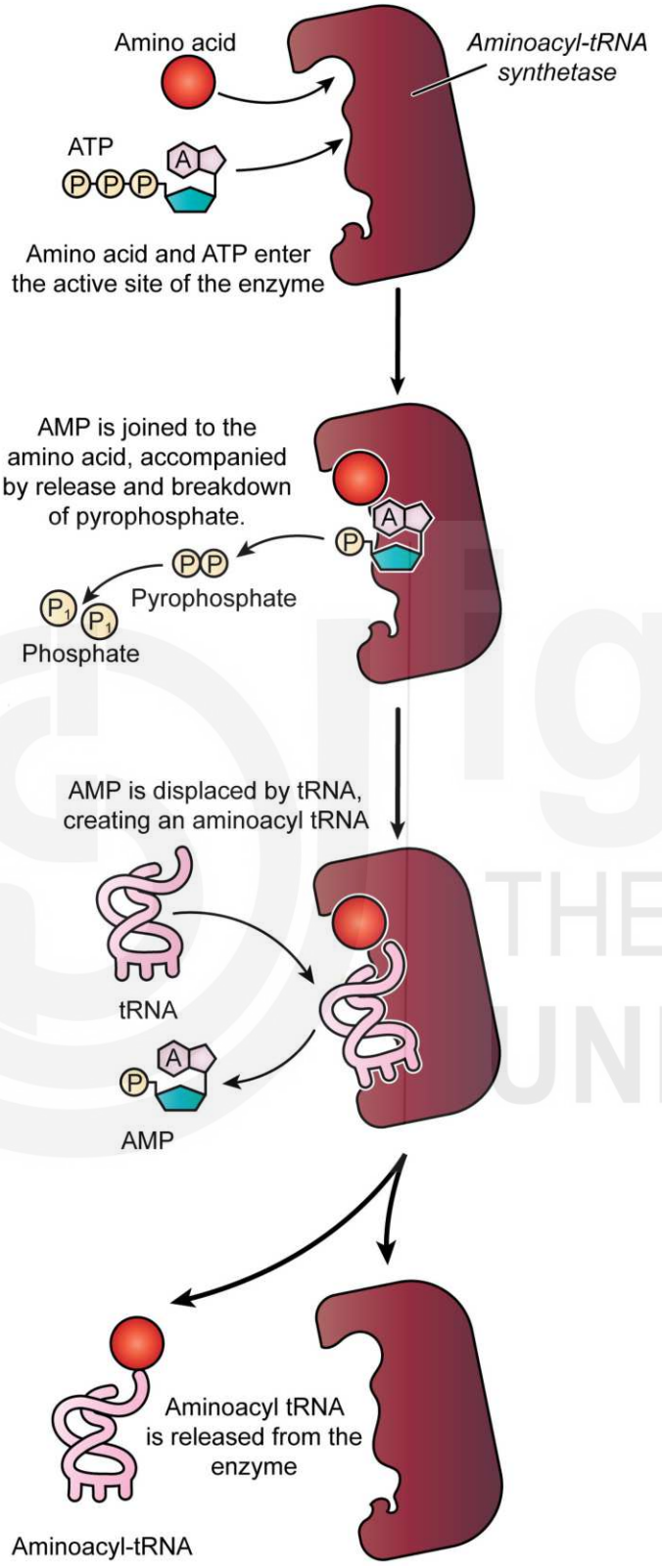


Fig. 13.5: Amino acid activation by *Aminoacyl-tRNA synthetase.*

4. mRNA template

The genetic information is encoded onto mRNA which acts as a template for polypeptide synthesis**.** Prokaryotes have polycistronic mRNAs with multiple translation start sites (Fig.13.6).

A

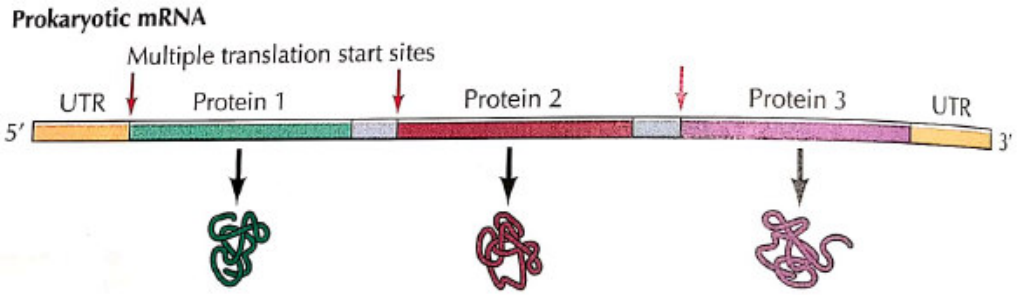


Fig. 13.6: A portion of prokaryotic polycistronic mRNAs encodes multiple proteins, each of which is translated from an independent start point.

5. Protein factors

In addition to *aminoacyl-tRNA synthetases* and protein components of ribosome, the process of translation requires participation of several other kinds of protein molecules. These protein factors participate in the initiation, elongation, and termination of polypeptide chain. The precise role played by these factors will now be discussed in the mechanism of each of the events occurring during translation.

Process of Translation

Translation of mRNA leads to the synthesis of polypeptides in the N-terminal to C-terminal direction. The complete translation process is subdivided into three stages; i) **Initiation stage**, where mRNA gets bound to the ribosome and positions itself for proper translation, ii) **Elongation stage**, in which amino acids are sequentially joined together through peptide bonds according to the arrangement of codons in mRNA, and iii) **Termination stage**, where both mRNA and the newly formed polypeptide chains are released from the ribosome.

An overview of translation has been shown in Figure 13.7.

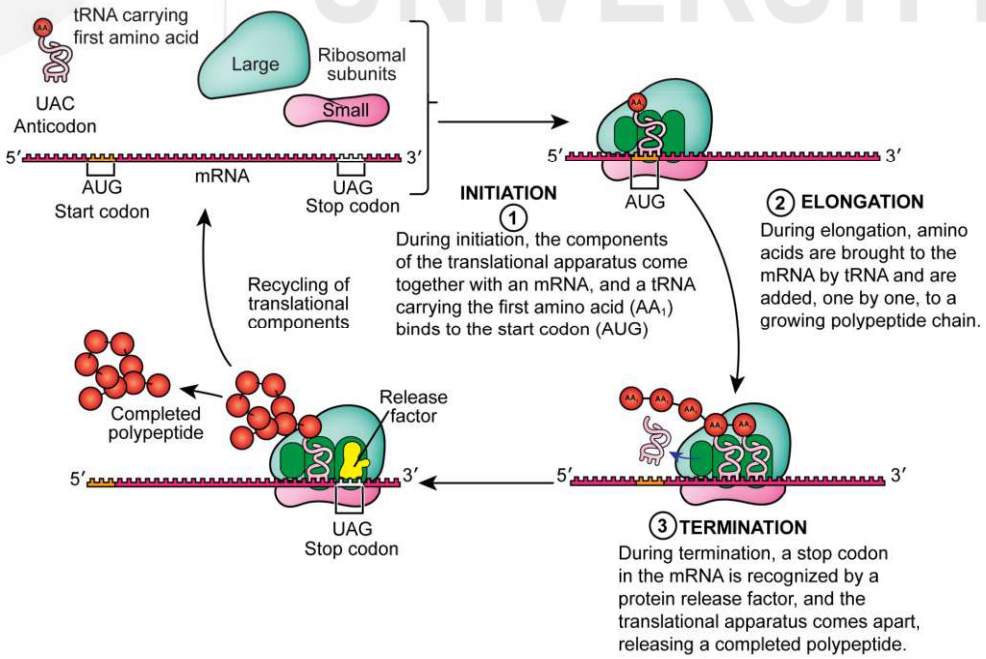


Fig. 13.7: An overview of translation showing three stages: Initiation, Elongation, and Termination.

13.2.2 Initiation of Translation

The functional ribosome of bacteria exists as two subunits, the small 30S subunit and large 50S subunit. An mRNA molecule can bind to the small ribosome subunit only when the subunits are separate. The initiation process of translation in bacteria can be subdivided into three distinct steps:

Step 1: The three initiation factors *IF1, IF2* and *IF3* first bind to the small (30S) ribosomal subunit, with GTP attaching to *IF2*. The presence of IF3 at this early stage prevents the 30S subunit from prematurely associating with the 50S subunit. The sequences on mRNA necessary for ribosome binding have been identified. The sequence covered by ribosome during initiation is from 30 to 40 nucleotides long and includes the AUG initiation codon. Within the ribosome binding site, there exists a special nucleotide sequence called the **Shine–Dalgarno consensus sequence** (Fig.13.8) .This consensus sequence AGGAGG is located seven nucleotides upstream to the initiation codon AUG and is complementary to a sequence of nucleotides UCCUCC located at the 3′ end of 16S rRNA (part of 30S subunit of ribosome).The nucleotides of Shine-Dalgarno sequence pair with their complementary nucleotides in the 16S rRNA during initiation.

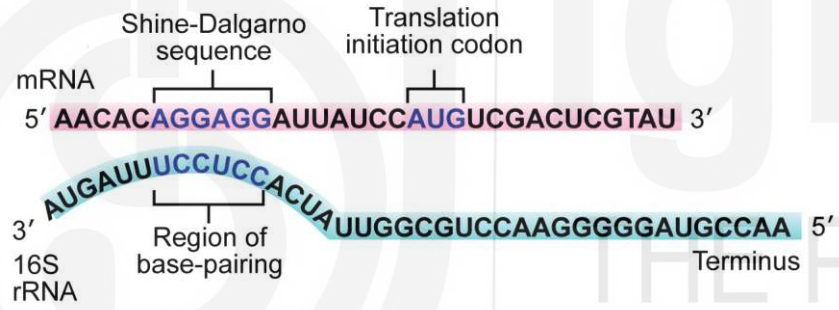


Fig. 13.8: Shine-Dalgarno consensus sequence in mRNA required for attachment of small subunit of ribosome.

Step 2: The binding of mRNA to the mRNA binding site of the small ribosomal subunit places the mRNA's AUG start codon at the ribosome's P site, where it can bind to the anticodon of the appropriate tRNA carrying the first amino acid It was discovered that methionine is the first amino acid, and the bacterial cells contain two different methionine-specific tRNAs. One, designated $tRNA^{Met}$ carries a normal methionine destined for insertion into the internal regions of polypeptide chain. The other, called $tRNA^{fMet}$, carries a methionine that is converted to the derivative *N-formyl methionine* (*fMet*) after linkage to the tRNA. In *N-formyl methionine*, the amino group of methionine is blocked by addition of a formyl group and so cannot form a peptide bond with another amino acid and only the carboxyl group is available for bonding to another amino acid. Hence N-formyl methionine can be situated only at the N-terminal end of polypeptide chain- suggesting that $tRNA^{fMet}$ functions as an **initiator tRNA** which starts the process of translation.

During initiation, the initiator tRNA with its attached N-formyl methionine is bound to the P site of the 30S ribosomal subunit by the action of initiation factor IF2 which forms a complex with GTP, which can distinguish initiator

tRNA^{fMet} from other kind of tRNA. This quality of IF2 helps us to explain as to why AUG start codons bind to the initiator tRNA^{fMet}, whereas AUG codons located elsewhere in mRNA bind to the non-initiation tRNA^{fMet}. Once the tRNA^{fMet} enters the P site, its anticodon becomes base-paired with the AUG start codon in the mRNA, and IF3 is released. At this point the 30S subunit with its associated IF1, IF2-GTP, mRNA and N-formyl methionyl tRNA^{fMet} is referred to as the **30S initiation complex**.

Step 3: The 30S initiation complex can now bind to a free 50S ribosomal subunit once IF3 has been released, generating the **70S initiation complex**. The 50S subunit then promotes hydrolysis of the IF-bound GTP, leading to the release of IF2 and IF1. At this stage all the three initiation factors have been released (Fig. 13.9).

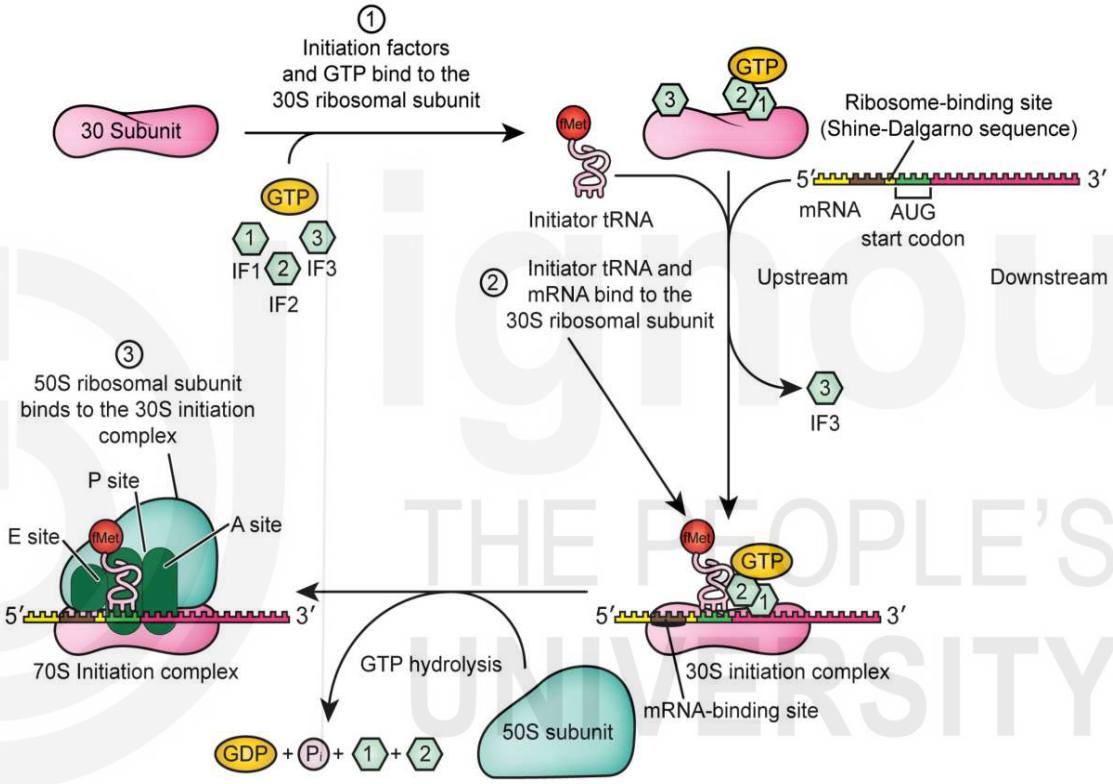


Fig.13.9: Steps involved in Initiation of translation in bacteria. Assembly of 70S initiation complex occurs in three steps. (1) Three initiation factors (IF1, IF2 and IF3) plus GTP bind to the small ribosomal subunit. (2) The initiator aminoacyl tRNA and mRNA are attached. (3) The large ribosomal subunit joins the complex. The resulting 70S initiation complex has fMet-tRNA^{fMet} residing in the ribosome's P site.

13.2.3 Chain Elongation

At the onset of elongation stage, the AUG start codon in the mRNA is located at the ribosomal P site and the second codon(the codon immediately downstream of the start codon) is located at the A site. Elongation takes place in three steps. In the first step an aminoacyl tRNA (a charged tRNA) whose anticodon is complementary to the second codon binds to the ribosomal A site. The binding of this new aminoacyl tRNA to the codon in the A site requires two protein **elongation factors**, EF-Tu and EF-Ts, and is driven by hydrolysis of GTP. From now on, every incoming aminoacyl tRNA will bind first to the A (aminoacyl) site-hence the site's name.

The EF-Tu-GTP complex promotes the binding of all aminoacyl tRNAs to the ribosome except the initiator tRNA, thus ensuring that AUG codons located downstream from the start codon do not mistakenly bind an initiator tRNA to the ribosome. As the aminoacyl tRNA gets transferred to the ribosome, the GTP is hydrolyzed and the EF-Tu-GTP complex is released. The role of EF-Ts is to generate EF-Tu-GTP from EF-Tu-GDP for the next round of elongation cycle.

After the appropriate amino acyl tRNA has been bound to the ribosomal A site, the second step of elongation is formation of peptide bond between the amino group of amino acid bound at A site and the carboxyl group that links the initiating amino acid (or growing polypeptide chain) to the tRNA at P site. The source of energy during peptide bond formation is provided by cleavage of high energy bond (Box 13.1). The formation of this peptide bond causes the growing polypeptide chain to be transferred from the tRNA located at P site to the tRNA located at A site (Fig. 13.10).

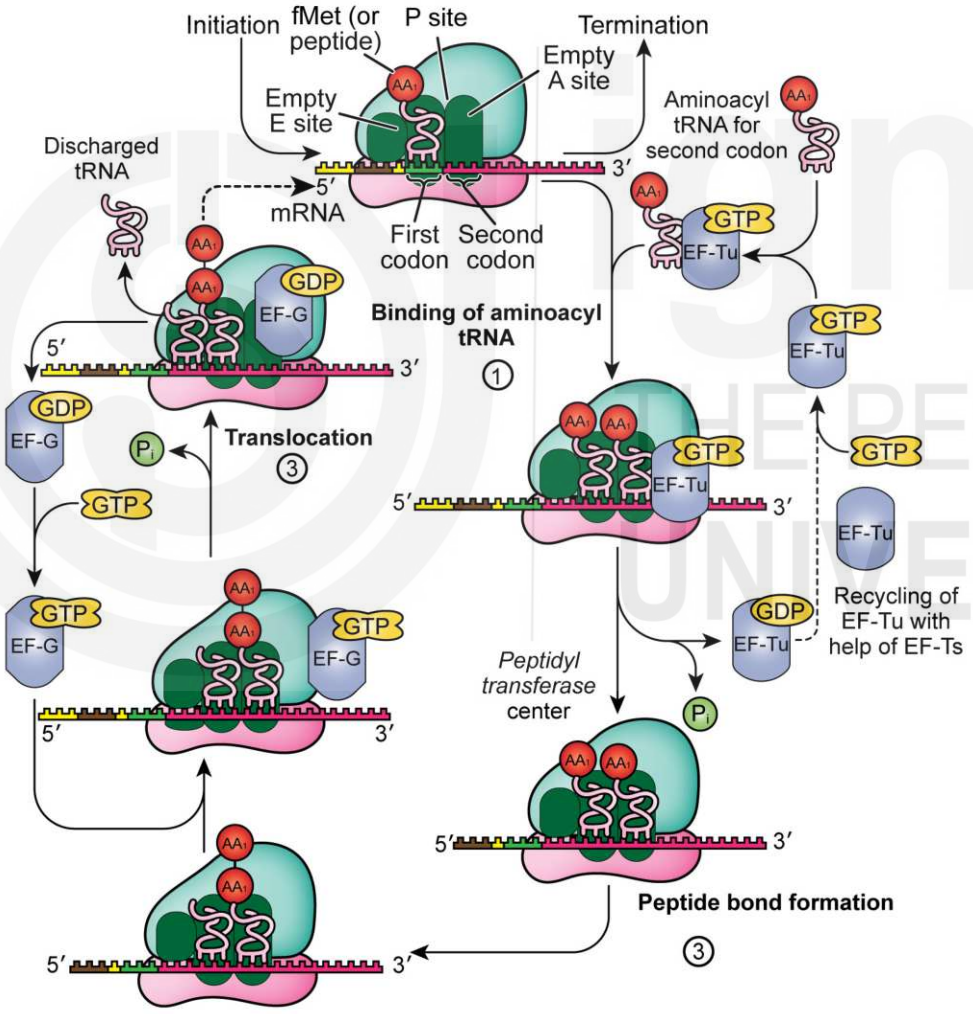


Fig. 13.10: Three stages of Polypeptide Chain Elongation in Bacteria: 1. An aminoacyl tRNA binds to the A site with the help of GTP-bound EF-Tu. During binding of tRNA, GTP gets hydrolyzed and EF-Tu is released. The recycling is supported by EF-Tu. 2. A peptide bond is formed at the P site between the –COOH group of fMet (and COOH of terminal amino acid in later cycles) and the newly arrived amino acid at the A-site. 3. The mRNA advances by three nucleotides. The peptidyl tRNA moves from the A site to the P site. Also the empty tRNA moves from the P site to the E site. During the process, GTP bound to EF-G gets hydrolysed.

Box 13.1: Source of Energy During Peptide Bond Synthesis

The peptide bond formation is the only step in protein synthesis that requires neither non ribosomal protein factors nor an outside source of energy such as GTP or ATP. The necessary energy is provided by cleavage of the high-energy bond that joins the amino acid or peptide chain to the tRNA located at the P site.

For many years, peptide bond formation was thought to be catalyzed by a hypothetical ribosomal protein that was given the name ***peptidyl transferase*.** However, in 1992 Harry Noller and his colleagues showed that the large subunit of bacterial ribosomes retains *peptidyl transferase* activity after all ribosomal proteins have been removed. In contrast, *peptidyl transferase* activity is quickly destroyed when rRNA is degraded by exposing ribosomes to *ribonuclease*. These observations suggested rRNA rather than a ribosomal protein is responsible for catalyzing peptide bond formation. In bacterial ribosomes, *peptidyl transferase* activity has been localized to the 23S rRNA of large ribosomal subunit, and high-resolution X-ray data have pinpointed the catalytic site to a specific region of the RNA chain. Hence 23S rRNA is an example of a ***ribozyme***, an enzyme made entirely of RNA.

The third step in elongation is translocation. After a peptide bond has been formed, the P site is occupied by an empty tRNA while the A site contains a peptidyl tRNA (the tRNA to which the growing polypeptide chain is attached). The movement of the ribosome down the mRNA in the 5′ → 3′ direction is called translocation. This step positions the ribosome over the next codon and requires elongation factor G (EF-G) and the hydrolysis of GTP to GDP. Because the tRNAs in the P and A sites are still attached to the mRNA through codon-anticodon pairing, they do not move with the ribosome as it translocates further. Consequently, the ribosome shifts so that the tRNA that previously occupied the P site now occupies the E site (exit site), from which it moves into the cytoplasm where it can be recharged with another amino acid. Translocation also causes the tRNA that occupied the A site (which is attached to the growing polypeptide chain) now to be relocated in the P site, leaving the A site open. Thus, the progress of each tRNA through the ribosome during elongation can be summarized as : cytoplasm → A site → P site → E site → cytoplasm.

The net effect of translocation is to bring the next mRNA codon into the S site, so the ribosome is now set to receive the next aminoacyl tRNA and repeat the elongation cycle. The only difference between the succeeding elongation cycles and the first elongation cycle is that an initiator tRNA occupies the P site at the beginning of the first elongation cycle, and the peptidyl tRNA occupies the P site at the beginning of all subsequent cycles. As each successive amino acid is added, the mRNA is progressively read in the 5′ → 3′ direction. The amino terminal of the growing polypeptide passes out of the ribosome through an exit tunnel in the 50S subunit. Immediately after its exit the polypeptide chain is folded and modified into its proper three-dimensional shape. Polypeptide synthesis is a very rapid process and a polypeptide of 400 amino acids can be made within 10 seconds in a growing *E. coli* cell.

13.2.4 Termination of Polypeptide Synthesis

The elongation process continues, reading one codon after another and adding successive amino acids in polypeptide chain, until any one of the three chain-termination (stop) codons (UAG, UAA, or UGA) in the mRNA arrives and

enters the ribosome's A site. Since there are no tRNAs with anticodons complementary to the termination codon, no tRNA enters the A site of the ribosome when a termination codon is encountered. Instead, the stop codons are recognized by proteins called **release factors** which possess special regions ('peptide anticodons') that bind to mRNA stop codons present at ribosomal A site. *E. coli* has three release factors – RF_1, RF_2 and RF_3. Release factor1 recognizes and binds to the termination codons UAA and UAG, while RF_2 binds to UGA and UAA. The binding of RF_1 or RF_2 to the A site of the ribosome promotes the cleavage of the tRNA in the P site from the polypeptide chain and the release of polypeptide. The release factor 3 binds to the ribosome and forms a complex with GTP. This binding brings about conformational change in the ribosome, releasing RF1 or RF2 from the A site and causing the tRNA in the P site to move to the E site. In this process GTP is hydrolyzed to GDP (Figure 13.11). Additional factors help bring about the release of the tRNA from the P site, the release of the mRNA from the ribosome, and the dissociation of the ribosome

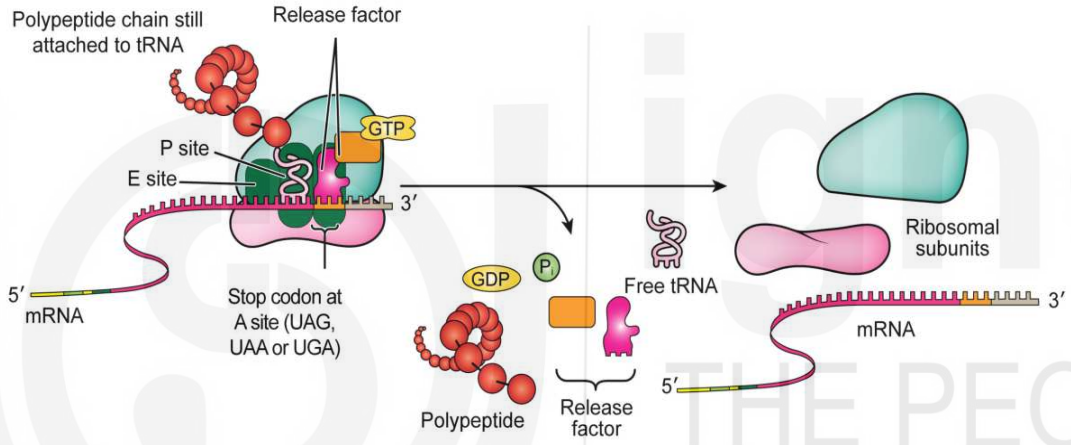


Fig. 13.11: A schematic diagram showing the mechanism of termination of translation: when a stop codon – UAA, UAG or UGA arrives at A site, it is recognized and bound by protein release factors associated with GTP. Hydrolysis of GTP is accompanied by release of the completed polypeptide, followed by dissociation of the tRNA, mRNA, ribosomal subunits, and release factors.

We have described above the machinery involved in translation and the mechanism of translation in prokaryotes. But before moving further you may like to work out some SAQs based on these.

SAQ 1

Fill in the blanks with the suitable word from the words given below.

a) The bacterial ribosome has a sedimentation coefficient of about…….. and is built from a ……. small subunit and …….. large subunit.

b) The enzymes responsible for linking amino acids to tRNAs are called …… .

c) The messenger RNAs that encode multiple polypeptides in bacteria are called…… .

d) The mRNA is bound to the 30S subunits in its proper orientation by means of a special nucleotide sequence called the …… .

e) The aminoacyl tRNA is escorted to the ribosome by an elongation factor ……. in prokaryotes, which is complexed with GTP.

(*Aminoacyl-tRNA Synthetases*, EF-Tu, Shine- Dalgarno sequence, polycistronic, 70S, 30S, 50S)

13.3 TRANSLATION IN EUKARYOTES

Like prokaryotes, the cellular machinery and mechanisms involved for translating mRNAs into polypeptides are similar in most respects in eukaryotic translation, although some significant differences do exist that will be discussed here.

13.3.1 The Cast of Characters

1) Ribosomes

Ribosomes are site of protein synthesis in both prokaryotes and eukaryotes. The fact that cells typically contain many ribosomes reflects the central importance of protein synthesis in cell metabolism. The general structures of prokaryotic and eukaryotic ribosomes are similar, although they differ in some details. Based on sedimentation coefficient, ribosomes of eukaryotes are 80S, somewhat larger than prokaryotic 70S. Like bacterial ribosome, eukaryotic ribosome is also built from two dissociable subunits called the large and small subunits. The large subunit contains the *peptidyl transferase* center, which is responsible for the formation of peptide bonds. The small subunit contains the decoding center where charged tRNAs read or "decode" the codon units of mRNA. The 80S ribosomes of eukaryotes consist of 40S and 60S subunits. The subunits of eukaryotic ribosomes are also larger and contain more proteins than their prokaryotic counterparts. The large subunit (60S) of eukaryotic ribosomes is composed of the 28S, 5.8S, and 5S rRNAs along with 46 proteins while the small subunit (40S) contains 18S rRNA and 33 proteins (Fig. 13.12). Also, the four important sites on the ribosome are like those found in eukaryotic ribosomes.

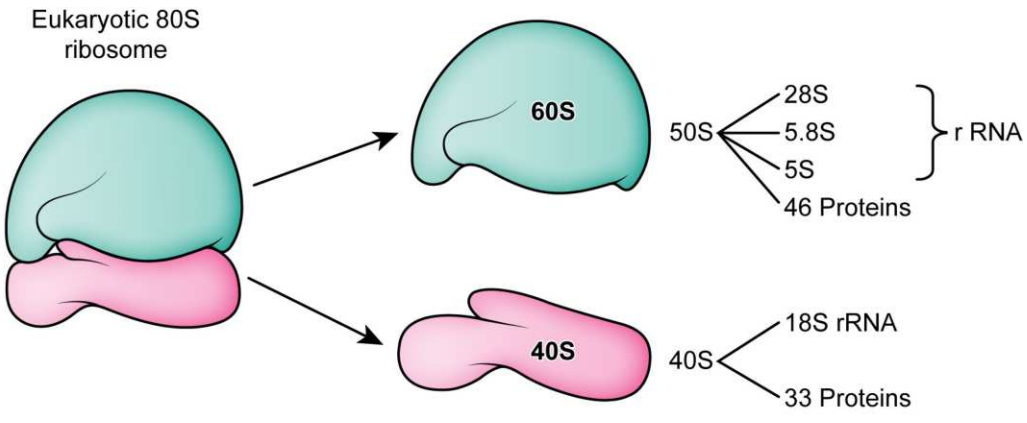


Fig.13.12: Components of eukaryotic ribosomes.

Each time a protein is synthesized, the translation components undergo a specific series of events in which the small and large subunits of the ribosome associate with each other and the mRNA, translate the target mRNA, and then dissociate after completing the synthesis of protein. This sequence of association and dissociation is known as the **ribosome cycle**. Although a ribosome can synthesize only one polypeptide at a time, each mRNA can be translated simultaneously by multiple ribosomes. An mRNA bearing multiple ribosomes is known as **polyribosome** or a **polysome** (Fig.13.13).

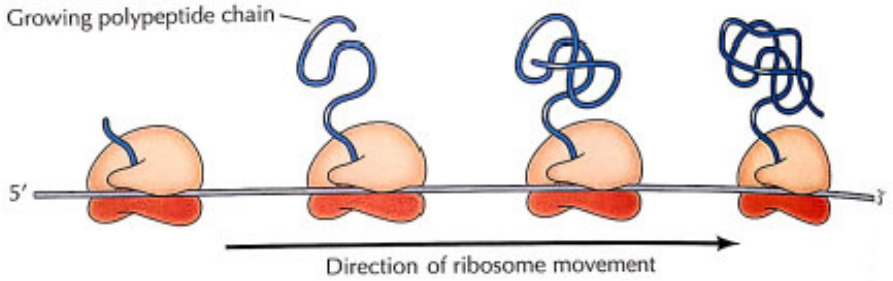


Fig. 13.13: Messenger rRNAs are translated by a series of ribosomes (a polysome).

2) mRNA template

The basic structure of the mRNA of prokaryotes and eukaryotes is similar. The protein- coding region of each mRNA is composed of a contiguous, nonoverlapping string of codons called an **open reading frame** (commonly known as ORF). Translation starts at the 5′ end of the ORF and proceeds one codon at a time to 3′ end. The first and last codons of an ORF are known as the start and stop codons respectively. The eukaryotic cells always use 5′-AUG-3′ as the start codon whereas for the bacterial cell also the start codon is usually 5′-AUG-3′ , in some cases 5′-GUG-3′ and sometimes even 5′-UUG-3′ are also used. The number of ORFs per mRNA is different between eukaryotes and prokaryotes. Eukaryotic mRNAs almost always contain a single ORF. In contrast prokaryotic mRNAs frequently contain two or more ORFs and hence can encode multiple polypeptide chains. mRNAs containing multiple ORFs are known as polycistronic mRNA (Fig. 13.6) and those encoding a single ORF are known as monocistronic mRNAs (Fig. 13.14). Unlike prokaryotic counterparts, eukaryotic mRNAs recruit ribosomes using a specific chemical modification called 5′ cap which is located at the extreme 5′ end of the mRNA and a poly-A tail at the extreme 3′ end of the mRNA (see unit 12).

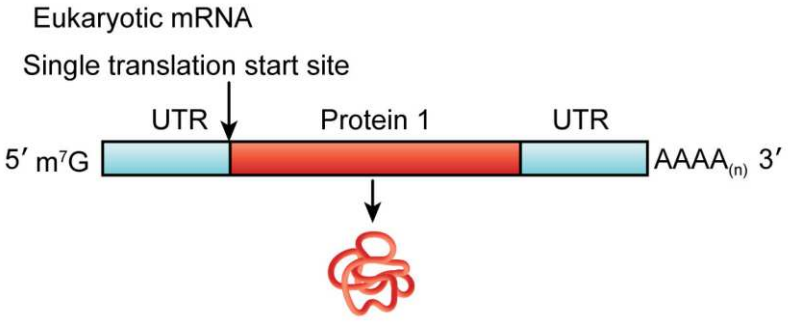


Fig. 13.14: Monocistronic eukaryotic mRNA showing 5′ cap and 3′ poly (A) tail.

13.3.2 Initiation of Translation

The initiation of translation in eukaryotes is more complicated than prokaryotes. Although the events of the initiation in eukaryotic cells are like the prokaryotes, there are some important differences. As you have studied in the section 13.2.2, that in prokaryotes the initiation sequences in 16S rRNA of the small subunit of the ribosome bind to the Shine-Dalgarno sequence in mRNA. No such analogous consensus sequence exists in eukaryotes. Instead, the cap at the 5´ end of the mRNA plays a critical role in the initiation of translation.

Unlike the situation in bacteria, the AUG start codon in eukaryotes specifies the amino acid methionine rather than N-formyl methionine. Eukaryotes possess a different set of initiation factors known as eIFs. There are about a dozen proteins (each consisting of multiple polypeptide chains) viz. eIF1, eIF2 and many more along with a special initiator $tRNA^{Met}$ that carries methionine.

Unlike bacteria this tRNA does not become formylated. The factors eIF1, eIF1A, and eIF3 binds to the 40S ribosomal subunit, while eIF2 (in a complex with GTP) binds to the initiator methionyl tRNA. Thereafter, the factor eIF5 associates with 40S subunit and the initiator tRNA to form **preinitiation complex**.

The eIF5 group of factors helps in recognizing and bringing mRNA to the ribosome. The factor eIF4E forms a complex with eIF4A and eIF4G to recognize the 5´ cap of the mRNA. The factor eIF4G also binds to the poly-A binding protein (PABP), which is associated with poly-A tail of the 3´ end of the mRNA.

Thus, the eukaryotic initiation factors recognize 5´ end and 3´ end of mRNAs, resulting in stimulation of polyadenylation on translation. The initiation factors eIF4E and eIF4G, in association with eIF4A and eIF4B then bring the mRNA to the 40S ribosomal subunit, with eIF4G interacting with eIF3.

After binding to the mRNA, the small 40S ribosomal subunit with initiator methionyl tRNA and eIFs, scans along the mRNA and usually begins translation at the first AUG triplet it encounters. The nucleotides on either side of the eukaryotic start codon appear to be involved in its recognition.

A common consensus start sequence is 5´ A(G) CC**AUG**G 3´ (also called **Kozak sequence**), where the underlined triplet AUG is the actual start codon (named after M. Kozak). After the initiator $tRNA^{Met}$ becomes base-paired with the start codon, initiation factor eIF5B triggers the hydrolysis of GTP bound to eIF2. Initiation factors including eIF2 bound to GDP are then released and the large 60S ribosomal subunit joins the complex in a reaction facilitated by the hydrolysis of GTP bound to initiation factor eIF5B to form 80S initiation complex of eukaryotic cells (Fig. 13.15).

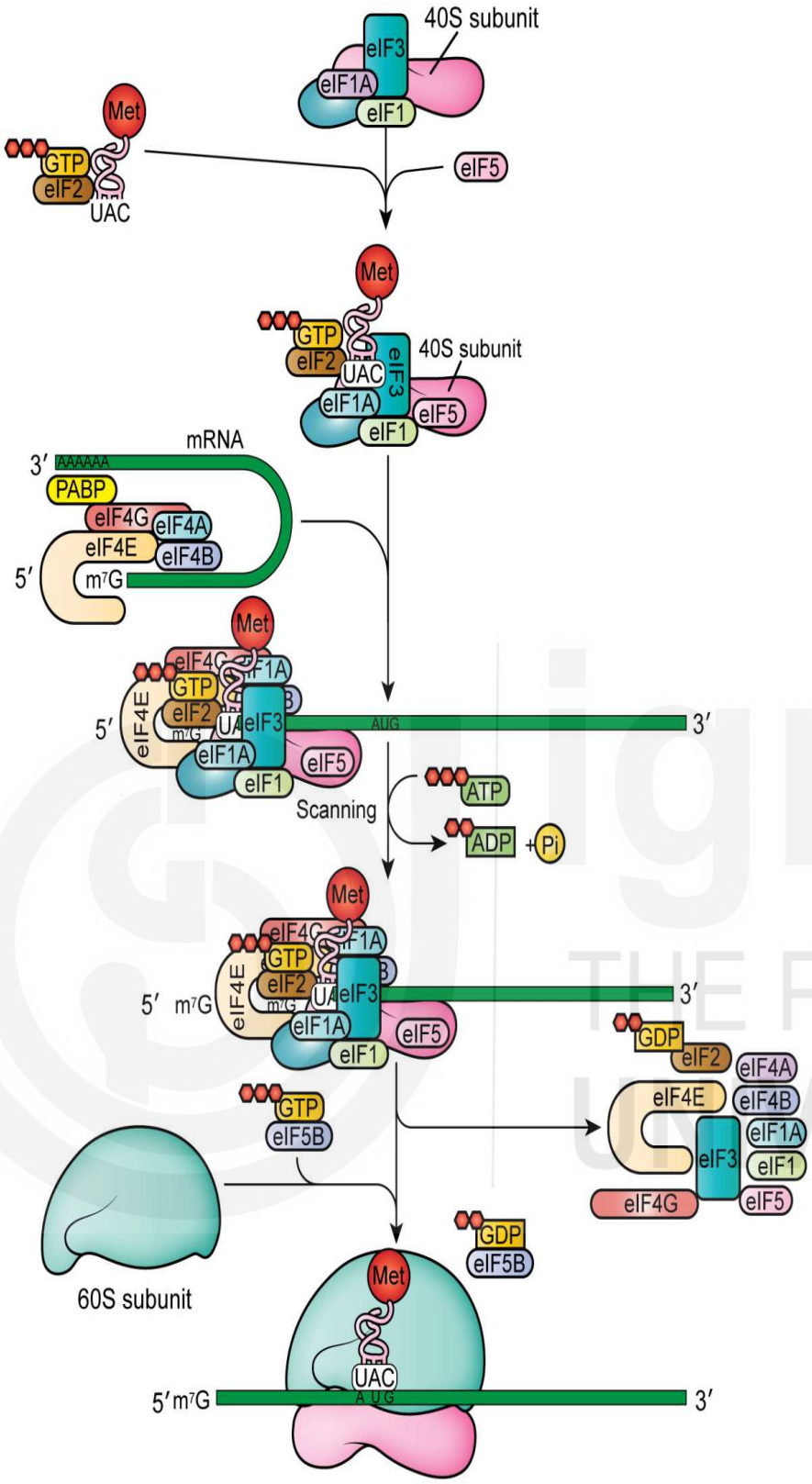


Fig. 13.15: Different steps involved in initiation of translation in eukaryotic cells: i) Initiation factors eIF1, eIF1A, and eIF3 bind to the 40S ribosomal subunit, ii) The initiator methionyl tRNA is bound by eIF2 (complexed to GTP) and forms a complex with the 40S subunit and eIF5, iii) The mRNA is brought to the 40S subunit by eIF4E (which binds to the 5′ cap), eIF4G (which binds to both eIF4E at the 5′ cap and PABP at the 3′ poly-A tail), eIF4A, and eIF4B. The ribosome then scans down the mRNA to identify the first AUG initiation codon. The energy required for scanning is provided by ATP hydrolysis. When the initiating AUG codon is identified, eIF5 triggers hydrolysis of GTP bound to eIF2, followed by the release of eIF2 (complexed to GTP) and other initiation factors. The 60S ribosomal subunit then joins the 40S complex, facilitated by eIF5B.

13.3.3 Chain Elongation

After the formation of initiation complex, translation proceeds by elongation of the polypeptide chain. The mechanism of elongation in eukaryotic cells is very similar to prokaryotic cells except for the elongation factors involved in this process. Eukaryotes possess at least three elongation factors, one of which also acts in initiation and termination. Another of the elongation factors used in eukaryotes, called elongation factor 2 (EF-2) is the target of a toxin produced by bacteria that causes diphtheria that until recently was a leading killer of children. The diphtheria toxin inhibits EF-2, preventing translocation of ribosome along the mRNA, and protein synthesis ceases.

13.3.4 Termination of Polypeptide Synthesis

Translation in eukaryotic cells terminates in a similar way as takes place in bacteria, except that there are two release factors. The release factor eRF1, which recognizes all the three termination codons (UAA, UAG, and UGA) and eRF2, which binds GTP and stimulates the release of the polypeptide from the ribosome.

We shall now turn our attention to posttranslational modification of protein. But before reading further you may like to try an SAQ.

SAQ 2

a) State whether the following statements about eukaryotic translation are true (T) or false (F).

i) The initiation of translation in eukaryotes is more complicated than prokaryotes and requires at least 12 initiation factors.

ii) In eukaryotic cells, the initiation codon AUG encodes a modified type of methionine, N-formyl-methionine.

iii) A common start sequence 5´ A (G) CC**AUG**G3´ (also called Kozak sequence) is the sequence on the eukaryotic mRNA where the underlined triplet is the start codon.

iv) The transcription and translation takes place simultaneously in eukaryotic cells.

v) In eukaryotic cells a single release factor (eRF1) recognizes all three termination codons.

b) Put the following in correct order for overall translation process in both prokaryotes and eukaryotes. Indicate this order by putting 1, 2 …… in the boxes given against each statement.

i) Chain elongation of amino acid [ ]

ii) Initiation of amino acid synthesis [ ]

iii) Termination of protein synthesis [ ]

iv) Posttranslational modifications [ ]

v) Assembly of the cellular machinery involved in translation [ ]

13.4 POST TRANSLATIONAL MODIFICATION OF PROTEINS

The newly synthesized polypeptides formed by translation must fold into correct three-dimensional shapes and often must be chemically modified before they can perform normal functions. Such alterations are called posttranslational modifications. As mentioned earlier, in bacteria, the N-formyl group located at the N-terminus of polypeptide chains is always removed. The methionine to which it was attached also gets removed. This is also the case with the methionine that starts eukaryotic polypeptides. As a result, relatively few mature polypeptides have methionine at their N-terminus, even though they all started out that way. The common modification events include the chemical alteration of individual amino acids groups by methylation, phosphorylation, or acetylation reactions. In addition, a polypeptide may also undergo glycosylation (addition of carbohydrate side chains or binding to prosthetic groups**)**. In proteins composed of multiple subunits, individual polypeptide chain must bind to one another to form appropriate multisubunit proteins or multi-protein complexes. A common posttranslational modification in eukaryotes is the attachment of a protein called ubiquitin, which targets the protein for degradation. Another modification in some proteins is the removal of 15 to 30 amino acids at the amino end of the protein called the **signal sequence**. The sequence helps direct a protein to a specific location within the cell, after which the sequence is removed by specific enzymes. Some common post translational modifications of proteins will be discussed.

13.4.1 Polypeptide Folding

As mentioned above, the synthesis of a polypeptide, is not equivalent to functional protein as the newly synthesized polypeptides must fold into correct three-dimensional shapes to function properly. The three-dimensional conformations of proteins result from interactions between the side chains of their constituent amino acids. Special proteins called **molecular chaperones** facilitate the folding of polypeptides. The term ‘’chaperones’’ was first used by Ron Laskey et al., to describe a protein (nucleoplasmin) that is required for the assembly of nucleosomes from histones and DNA. Like the nucleoplasmin which binds to histones and helps in their assembly into nucleosomes but itself is not a component of the final nucleosome structure , the chaperones also act as catalysts to facilitate protein folding by assisting the self-assembly process without being part of assembled complex. Chaperones appear to function by binding to and stabilizing unfolded or partially folded polypeptides that are intermediates along the pathway leading to the final correctly folded state. Absence of chaperones results in instability in the unfolded or partially folded polypeptide chains. Chaperones also function during translation and are seen to bind to nascent polypeptide chain that are still being translated on ribosomes (Fig. 13.16). Protein folds into domains of approximately 100 to 300 amino acids. It is, therefore, necessary to protect the nascent chain from aberrant folding or aggregation with other proteins until synthesis of entire domain is completed and the protein could fold into its correct conformation. Chaperone binding stabilizes the amino- terminal portion in an unfolded conformation until the rest of the polypeptide chain is synthesized and the completed protein can fold correctly. Chaperone also helps in stabilizing unfolded protein chain during their transport into subcellular organelles. For

example, during transportation of proteins into mitochondria from cytosol, proteins are transported across the mitochondrial membrane in partially unfolded conformations that are stabilized by chaperones in the cytosol. Chaperones are found throughout the living world, from archaea and bacteria to the various compartments of eukaryotic cells. The two most widely occurring chaperone families are *Hsp70* and *Hsp60*. The ''Hsp'' comes from the original designation of these proteins ''Heat-Shock proteins'' because cells produce them in response to stressful conditions, such as exposure to high temperatures.

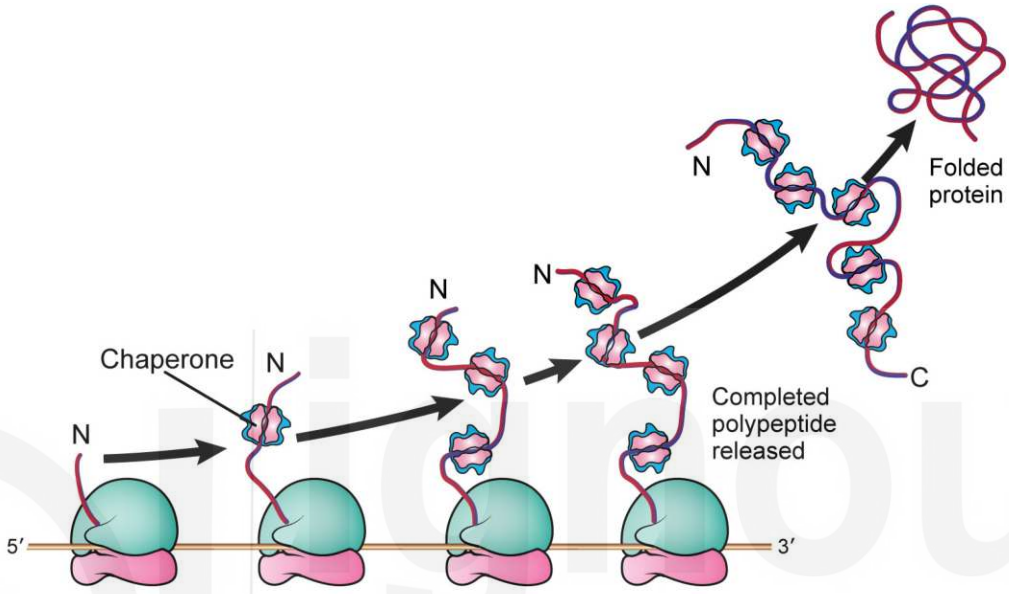


Fig. 13.16: Chaperone binding to the amino (N) terminal portion of the nascent polypeptide chain during translation.

In addition to chaperones that help in protein folding, cells contain at least two types of enzymes that act as chaperone by catalyzing protein folding. ***Protein disulfide isomerase*** **(PDI)** catalyzes disulfide bond formation between cysteine residues that is important in stabilizing the folded structure of many proteins. The second enzyme involved in protein folding is ***peptidyl prolyl isomerase*** that catalyzes the isomerization of peptide bond that involves proline residues. The isomerization between the *cis* and *trans* configurations of prolyl-peptide bond is a rate- limiting step in protein folding. This enzyme is widely distributed in both prokaryotic and eukaryotic cells and plays important role in folding of some proteins.

13.4.2 Protein Cleavage

The cleavage of polypeptide chain is known as **proteolysis.** It is an important step in the maturation of many proteins. The removal of formyl group or the entire initiator amino acid methionine from the amino terminus of many polypeptides is a good example of proteolysis. Proteolytic modification of amino terminus facilitates the translocation of many proteins across the plasma membrane as well as proteins destined to cellular organelles in eukaryotic cells.

13.4.3 Glycosylation

The modification of many proteins by addition of carbohydrates, particularly in eukaryotes is known as **glycosylation**. The proteins to which the

carbohydrate chains have been added are called **glycoproteins**. These are usually secreted and localized on cell surface. However, several cytosolic and nuclear proteins are also glycosylated. The carbohydrate moieties of glycoprotein play three important roles- i) protein folding in the ER, ii) in targeting the proteins for delivery to the appropriate intracellular compartments, and iii) as recognition sites in the cell-cell interaction. Depending on the site of attachment of carbohydrate side chain, most of the glycoproteins are classified as either N-linked or O-linked (Fig.13.17).

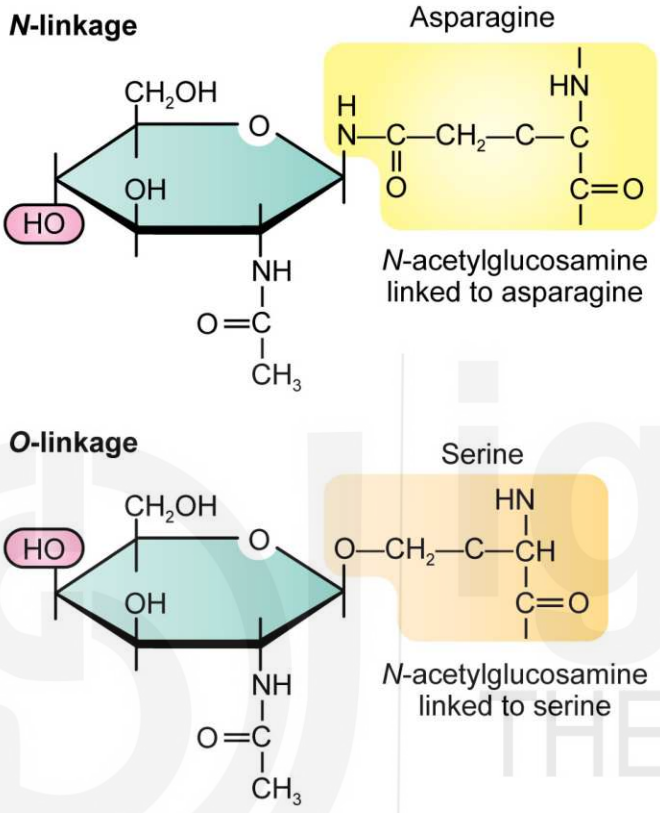


Fig. 13.17: Linkage of carbohydrate side chain to glycoprotein.

In N-linked glycoprotein, the carbohydrate is attached to the nitrogen atom in the side chain of amino acid asparagine and in the O-linked, the oxygen atom in the side chain of amino acids serine or threonine is the site of carbohydrate attachment. The sugars that get directly attached to these positions are usually either N-acetylglucosamine or N-acetylgalactosamine. In addition, mannose residues can be attached to tryptophan residues in some proteins by carbon-carbon bonds. Most of the glycoproteins in eukaryotic cells are destined either for secretion or for incorporation into the plasma membrane. These proteins get usually transferred to endoplasmic reticulum while their translation is still in progress and the glycosylation is initiated there before translation is complete.

13.4.4 Attachment of Lipids

Some proteins in the eukaryotic cells are modified by the attachment of lipids to the polypeptide chain. The addition of lipids helps to target and anchor these proteins to plasma membrane, with which the hydrophobic lipid can interact. There are three general types of lipid additions and are common in eukaryotic proteins associated with the cytosolic face of plasma membrane – i) **N-myristoylation**, ii) **prenylation**, and iii) **palmitoylation** and a fourth type

of modification, **glycosylphosphatidylinositol (or GPI) anchors** where glycolipids are added and these play an important role in anchoring some cell surface proteins to the extracellular face of plasma membrane.

Protein Modification by Lipids

i) **N-myristoylation**: In this process, myristic acid (a 14- carbon fatty acid) is attached to an N-terminus glycine residue (Fig. 13.18). The glycine is usually the second amino acid incorporated into the polypeptide chain. The initiator amino acid methionine is removed by proteolysis before the addition of glycine residue on which the fatty acid myristic acid is added. Many proteins that are modified by N-myristoylation are associated with the inner face of the plasma membrane.

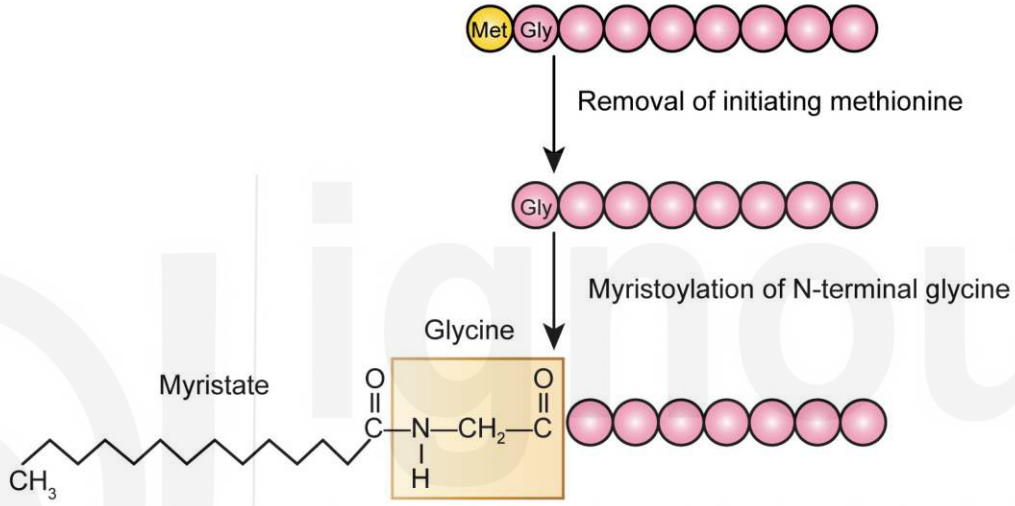


Fig. 13.18: Posttranslational modifications of proteins through addition of fatty acid by N-myristoylation.

ii) **Prenylation**: Lipids can also be attached to the side chains of cysteine, serine, and threonine residues. Prenylation is an important example of this type of modification in which specific types of lipids (prenyl groups) are added to the sulfur atoms in the side chains of cysteine residues located near the C terminus of the polypeptide chain. Many plasma membrane-associated proteins involved in cell growth and differentiation are modified in this way. One good example of prenylation is the **Ras oncogene** proteins, which are responsible for the uncontrolled growth of many human cancers. These proteins terminate with a cysteine residue (Cys) followed by two aliphatic amino acids (A) and any other amino acid (X) at the C -terminus.

The prenylation of these proteins proceeds in three steps: In the first step prenyl group is added to a cysteine residue located three amino acids from the carboxyl terminus of the polypeptide chain. The prenyl groups added in this reaction are either farnesyl (15 carbons) or geranylgeranyl (20 carbons) called **farnesylation**. In the second step there is proteolytic removal of three C- terminal amino acids following cysteine residue (proteolysis). In the last step methyl group is added to the carboxyl group of the C- terminal cysteine residue (methylation) (Fig. 13.19).

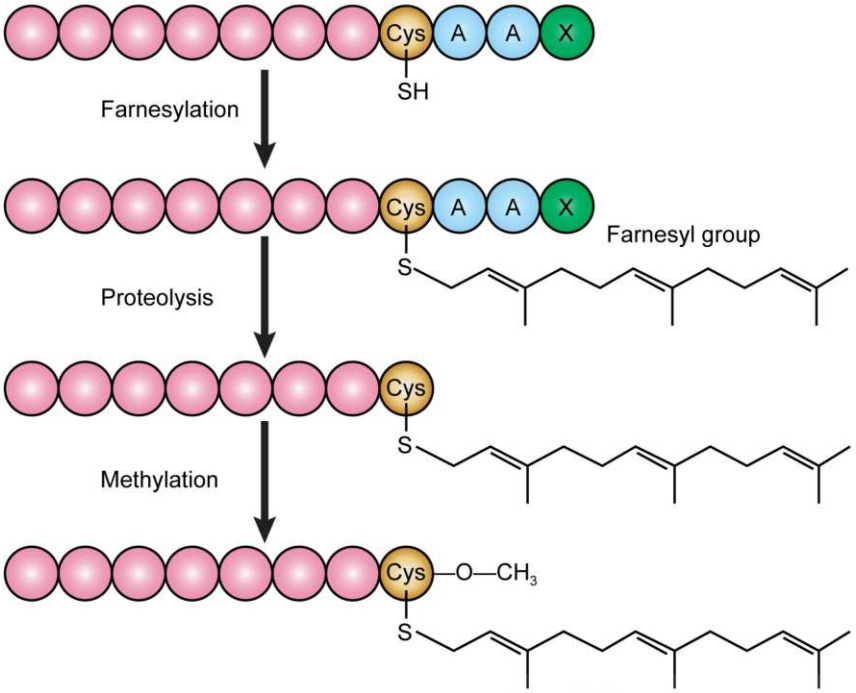


Fig. 13.19: Posttranslational modifications of protein by prenylation of cysteine residue.

iii) **Palmitoylation:** In this type of fatty acid modification, palmitic acids (a 16-carbon fatty acid) are added to sulfur atoms of the side chains of internal cysteine residues (Fig.13.20). Like N-myristoylation and prenylation, palmitoylation plays an important role in the association of some proteins with the cytosolic face of the plasma membrane.

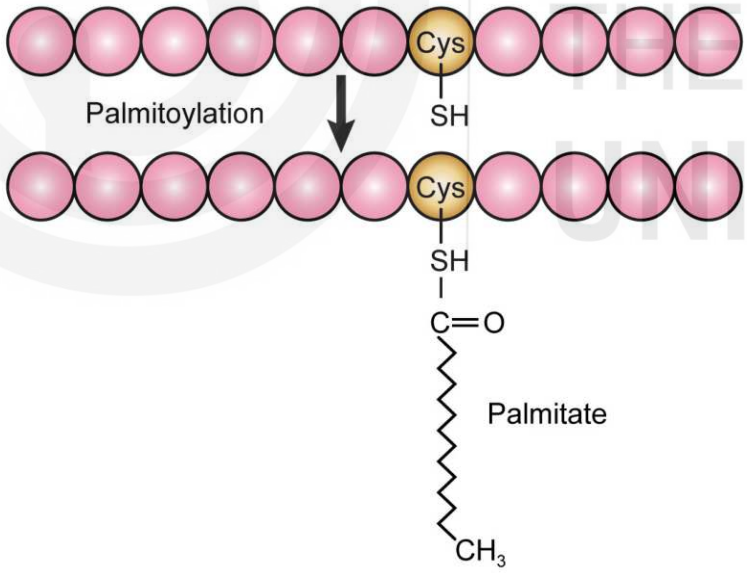


Fig. 13.20: Posttranslational modifications of proteins by palmitoylation.

iv) In many cases, lipids linked to oligosaccharides (**glycolipids**) are added to the C-terminus) carboxyl group of some proteins where they serve as anchors that attach the proteins to the external face of the plasma membrane. Since the glycolipids attached to these proteins contain phosphatidylinositol, they are generally called **glycosylphosphatidylinositol (or GPI) anchors.** In this process the oligosaccharide portions of GPI anchors are attached to the terminal carboxyl groups of polypeptide chains. The inositol head group of

phosphatidylinositol is in turn attached to the oligosaccharide consisting of mannose, N-acetylgalactosamine, and glucosamine residues; therefore, the carbohydrate serves as a bridge between the protein and the fatty acid chains of phospholipid (Fig.13.21). The GPI anchors are synthesized and added to proteins as a preassembled unit within the endoplasmic reticulum. The addition of GPI anchors is accompanied by cleavage of peptide consisting of about 20 amino acids from the C-terminus of the polypeptide chain. The modified protein is then transported to the cell surface, where the fatty acid chains of the GPI anchor mediate its attachment to the plasma membrane.

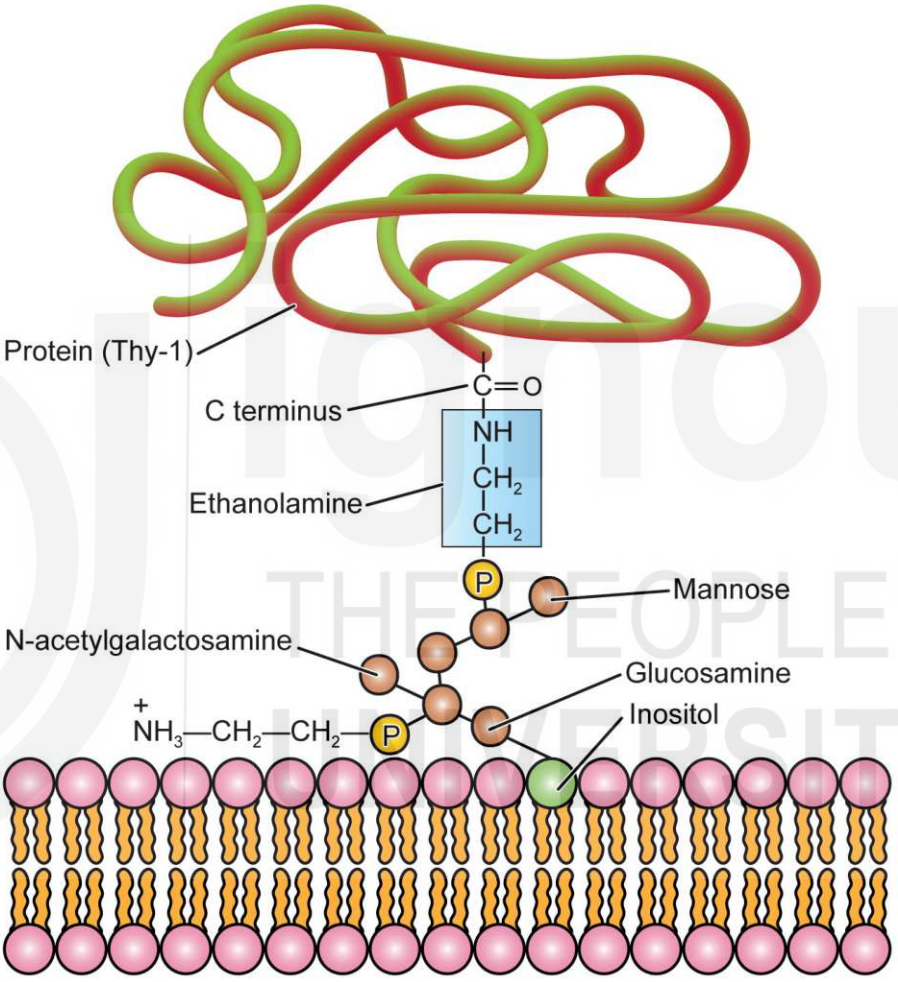


Fig. 13.21: Structure of a GPI anchor.

Various other post translational mopdofication of proteins have been discussed in Box 13.2

Box 13.2: Protein Phosphorylation and Other Modifications

Protein phosphorylation is catalyzed by ***protein kinases***. These enzymes are one of the largest protein families in eukaryotes accounting for approximately 2% of the eukaryotic genes. Most of the protein kinases transfer phosphate groups from ATP to the hydroxyl groups of the side chains of serine, threonine, or tyrosine residues. These enzymes are also called ***serine/threonine kinases or tyrosine kinases***. Protein phosphorylation is reversed (dephosphorylation) by *protein phosphatases*, which catalyzes the hydrolysis of phosphorylated amino acid residues. The combined action of *protein kinases* and *protein phosphatases* mediates the reversible phosphorylation of many cellular proteins (Fig.13.22).

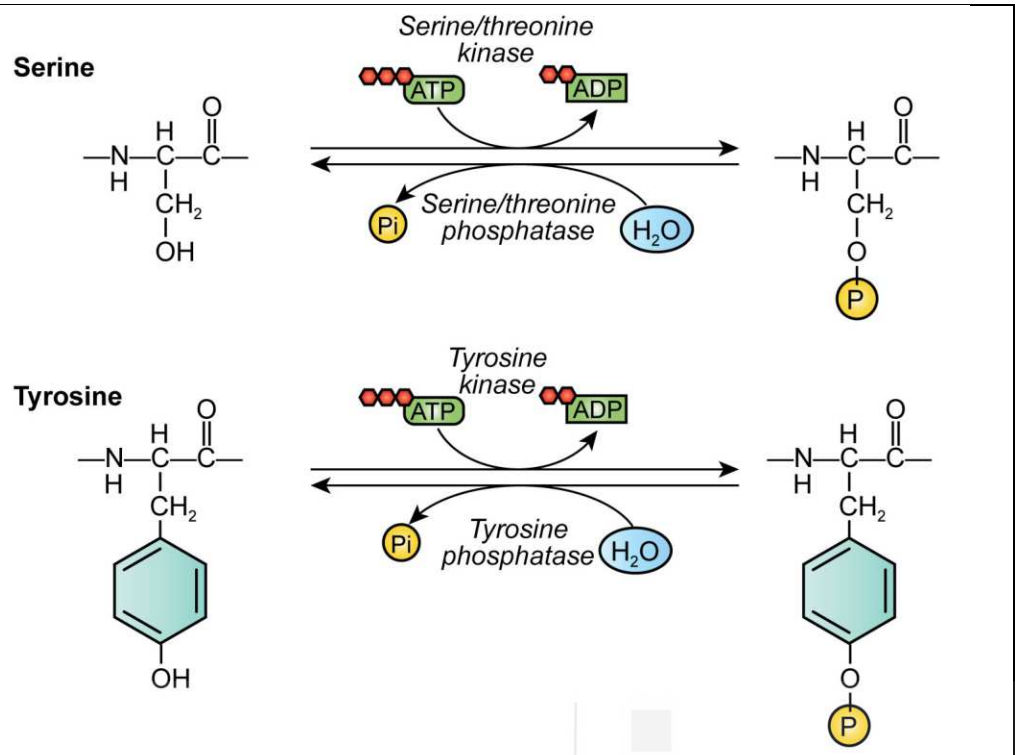


Fig. 13.22: Post translational modifications of proteins by phosphorylation and dephosphorylation by *protein kinases* and *phosphatases*.

13.4.5 Protein Modification by Small Molecules

Protein phosphorylation is the most common and best studied type of covalent modification that regulates protein activity. However, there are several other types of protein modifications by small molecules (Fig.13.23) that play important roles. Some examples are given below.

Acetylation of lysine residues

Acetylation is the process in which an acetyl (CH_3CO) functional group is transferred from one molecule (**Acetyl-Coenzyme A** for example) to another molecule. Deacetylation is simply the reverse reaction where an acetyl group is removed from a molecule. Histone acetylation and deacetylation are the processes by which the lysine residues within the N-terminal tail protruding from the histone core of the nucleosome are acetylated and deacetylated as part of gene regulation. These reactions are typically catalyzed by the enzymes with ‘‘*histone acetyl transferase”* (HAT) or ‘*’histone deacetylase”* (HADC) activity. Recent proteomic studies have shown that lysine acetylation is particularly common posttranslational modification that affects thousands of proteins in both prokaryotic and eukaryotic cells, including many metabolic enzymes as well as transcriptional regulatory proteins.

Methylation of lysine and arginine residues

Protein methylation has emerged as an important and widespread post-translational modification affecting almost all basic cellular processes in prokaryotes and eukaryotes. Methylation can affect the side chain of several residues as well as the amino and carboxyl terminal of proteins. In eukaryotes, methylation is predominantly found on lysine (Lys) and arginine (Arg) residues. Protein methylation has been most studied in histones and in this process,

methyl groups (-CH_3) are transferred to amino acids lysine and arginine of histone proteins that make up nucleosomes, which the DNA double helix wraps around to form chromosomes. The amino acid lysine can be methylated from one to three methyl groups with the help of *lysine methyltransferase* whereas arginine from one to two methyl groups catalyzed by three different classes of *arginine methyltransferases* (Fig. 13.23). The lysine methylation on the tails of histone proteins, in conjunction of acetylation and phosphorylation, controls their interaction with other proteins, affects chromatin compaction and up- or down-regulation of gene expression. The arginine methylation is predominantly known to be associated with RNA regulation and processing.

Nitrosylation to cysteine residue

Protein S-nitrosylation is a rapidly reversible and precisely targeted ubiquitous posttranslational modification wherein nitric oxide (NO) is covalently attached to a thiol group of a cysteine residue (S – NO) (Fig. 13.23). Cysteine thiol groups undergo a variety of posttranslational oxidation modifications, ranging from physiologic and reversible S-nitrosylation (S – NO), hydroxylation (S – OH), and disulfide bond formation (S – S). S-nitrosylation exerts profound effects on cellular signaling and overall response to the microenvironment.

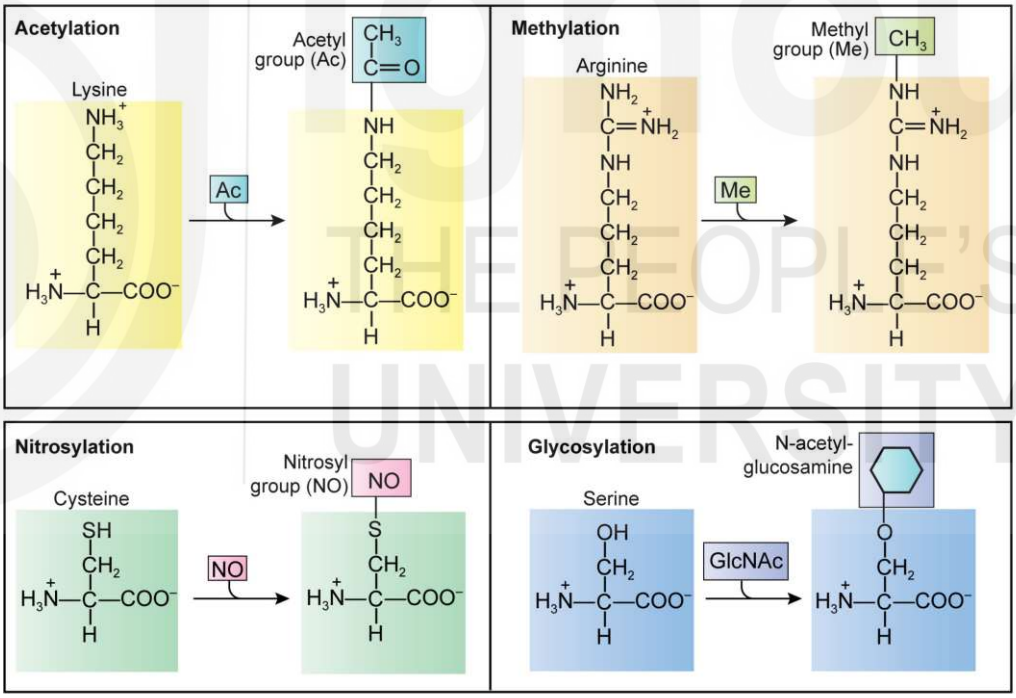


Fig. 13.23: Posttranslational modifications of proteins by small molecules showing acetylation of lysine residues, methylation of lysine and arginine residues, nitrosylation of cysteine residues and glycosylation of serine residues (see section 13.4.3).

In addition to modification by small molecules mentioned above, some proteins are regulated by the covalent attachment of polypeptides. The first polypeptides found to act in this way was **ubiquitin**, a 76 -amino acid polypeptide that is highly conserved in all eukaryotes. Initially the polypeptide ubiquitin was found to target proteins for degradation. In the modification of proteins, ubiquitin is attached to the amino groups of the side chains of lysine residues. The selective addition of ubiquitin to target proteins (ubiquitylation) is a multistep process. Importantly, ubiquitin can also be removed from target

protein by deubiquitylation enzymes, so ubiquitylation is a reversible protein modification. It is now recognized that the modification of proteins by addition of ubiquitin and other ubiquitin-like proteins, such as SUMO (small ubiquitin-related modifier) affects a variety of functions. The histone modification by these polypeptides is one mechanism for regulating the transcriptional activity of chromatin. Ubiquitylation is also important in the regulation of protein kinases, proteins involved in DNA repair, and in control of endocytosis and vesicle trafficking.

SAQ 3

Fill in the blanks with the suitable word from the words given below.

a) The protein folding inside cells is usually facilitated by proteins called …… .

b) The disulfide bond formation between cysteine residues that is important in stabilizing the folded structure of many proteins is catalyzed by **…… .**

c) The modification of many proteins by addition of carbohydrates particularly in eukaryotes is known as …….. .

d) The protein phosphorylation is catalyzed by …….. .

e) The cleavage of polypeptide chain which is an important step in the maturation of many proteins is known as …… .

(Proteolysis, *protein kinases*, glycosylation, Molecular chaperones, *Protein disulfide isomerase* (PDI))

13.5 SUMMARY

In this unit you have learnt that:

- Translation is the key step in the production of protein molecules, which involve reading of information from the nucleotide base sequence of an mRNA molecule and converting it into the amino acid sequence of a polypeptide chain.
- The cellular machinery that is involved for translating mRNAs into polypeptides includes ribosomes, mRNA molecule, tRNA molecules, aminoacyl *tRNA synthetases*, and Protein factors.
- Ribosomes are site of protein synthesis in both prokaryotes and eukaryotes. The general structures of prokaryotic and eukaryotic ribosomes are similar, although they differ in some details.
- Based on sedimentation coefficient, ribosomes of eukaryotes are 80S (60S and 40S), somewhat larger than prokaryotic 70S (50S and 30S) ones. The sequence of mRNA nucleotides specifies the order in which the amino acids are added to the growing polypeptide chain.

- The rRNA component of the ribosome helps position the mRNA and catalyzes peptide bond formation. The *aminoacyl-tRNA synthetases* link amino acids to the tRNA molecules that bring amino acid to the ribosome. The various protein factors trigger specific events associated with the translation factors.
- Translation involves initiation, elongation, and termination stages. Translation of both prokaryotic and eukaryotic mRNAs initiate with a methionine residue. The initiation codon AUG encodes a modified type N-formyl methionine in prokaryotes whereas in eukaryotic cells the initiation codon AUG encodes unformylated methionine.
- In prokaryotic cells initiation codon is preceded by a sequence called Shine-Dalgarno sequence that aligns the mRNA on the ribosome by base pairing with 16S rRNA. In eukaryotes, most initiation codons are identified by scanning from the 5´ end of the mRNA, which is recognized by its 7-methylguanosine cap.
- Translation is initiated by the binding of methionyl tRNA and mRNA to the small ribosomal subunit and the large ribosomal subunit then joins the complex.
- The polypeptide chain elongation involves sequential cycles of aminoacyl tRNA binding, peptide bond formation, and translocation, with each cycle driven by the action of elongation factors. The chain termination occurs when a stop codon in mRNA is recognized by release factors, which causes the mRNA and newly formed polypeptide to be released from the ribosome.
- The GTP binding and hydrolysis are required for the action of several initiation, elongation, and release factors.
- After translation, proteins in both prokaryotic and eukaryotic cells may undergo alterations termed as posttranslational modifications. The newly synthesized polypeptides must fold into correct three-dimensional shapes .The polypeptides must be chemically modified before they can perform normal functions. The molecular chaperones (HSPs) facilitate protein folding by binding to and stabilizing unfolded or partially folded polypeptide chains.
- The common modification events include the chemical modifications of individual amino acids groups by methylation, phosphorylation, or acetylation reactions.
- In addition, a polypeptide may undergo glycosylation (addition of carbohydrate side chains or binding to prosthetic groups. In the case of proteins composed of multiple subunits, individual polypeptide chain must bind to one another to form appropriate multisubunit proteins or multi-protein complexes. A common posttranslational modification in eukaryotes is the attachment of a protein called ubiquitin, which targets the protein for degradation.

13.6 TERMINAL QUESTIONS

1. Write the role of initiation factors in bacteria.
2. Write the three repetitive steps performed during chain elongation of both prokaryotes and eukaryotes.
3. Discuss the role of small molecules in protein modification.

13.7 ANSWERS

Self-Assessment Questions

1. a) 70S, 30S, 50S
   b) *aminoacyl-tRNA synthetase*
   c) polycistronic
   d) Shine-Delgarno sequence,
   e) EF-Tu
2. a) i) True; ii) False; iii) True; iv) False; v) True
   b) i) 5; ii) 2; iii) 1; iv) 3; v) 4
3. a) molecular chaperones
   b) *protein disulfide isomerase*(PDI)
   c) glycosylation
   d) *protein kinases*
   e) proteolysis

Terminal Questions

1. Refer to Subsection 13.2.2.
2. i) binding of an aminoacyl tRNA to the ribosome brings a new amino acid into position to be joined to the polypeptide chain, ii) peptide bond formation links this amino acid to the growing polypeptide, and iii) the mRNA is advanced a distance of three nucleotides by the process of translocation to bring the next codon into position for translation.
3. Refer to Subsection 13.4.5.

UNIT 14

REGULATION OF GENE EXPRESSION

Structure

14.1 Introduction

Objectives

14.2 Principles of Transcriptional regulation

14.3 Gene Regulation in Prokaryotes

Catabolic Pathways and Substrate Induction

Anabolic Pathways and End-Product Repression

Mechanism of Gene Regulation at Transcriptional Level

Model Systems for Gene Regulation in Prokaryotes

Transcription Regulation Through Attenuation

14.4 Gene Regulation in Eukaryotes

Gene Regulation by Chromatin Modification

Gene Regulation by Transcription Factors and Transcriptional Regulator Proteins

Coordinated Gene Regulation

14.5 Gene Silencing

Transcriptional Gene Silencing

Posttranscriptional Gene Silencing

14.6 Summary

14.7 Terminal questions

14.8 Answers

14.1 INTRODUCTION

A major theme of molecular genetics is the central dogma, which states that genetic information flows from DNA to RNA to proteins. In the previous units (Unit 12 and 13) you have been made familiar as to how the biological information flows from the DNA main repository to proteins via RNA through the processes of transcription and translation. Although the central dogma provided a molecular basis for the connection between genotype and phenotype, it failed to address a critical question: How is the flow of information along the molecular pathways regulated? In this unit, we will study as to how are the various steps in the gene expression regulated so that each gene product is made at the right time and in proper amounts for each cell's needs. You will appreciate that regulation is an important aspect of every process in nature. And this is especially true for gene expression. Most genes are not expressed all the time. In some bacteria, selective gene expression

enables a cell to be metabolically economical, synthesizing only those gene products that are of immediate use under the prevailing environmental conditions. In other cases, such as in multicellular organisms, selective gene expression allows cells to fulfill specialized roles. Our initial knowledge about gene regulation came from bacteria. The subsequent advances in DNA technology helped in understanding the gene regulation process in eukaryotes. Gene silencing is the process of regulation by which a certain gene can be prevented from expressing itself. Genes can be silenced both at transcriptional and translational levels. Modern gene silencing techniques have been extensively used to study genes associated with disorders. Our discussion in this unit will mainly focus on gene regulation in bacterial cell and thereafter, in eukaryotes. You will also study as to how gene silencing regulates the gene expression.

Objectives

After studying this unit, you should be able to:

- describe regulation of gene expression and principles of transcriptional regulation;
- understand the transcriptional regulation in prokaryotes and how does it differ from eukaryotes; and
- know the process and importance of gene silencing.

14.2 PRINCIPLES OF TRANSCRIPTIONAL REGULATION

The biological properties of each cell type are largely dependent on the proteins expressed within it. This array of expressed proteins determines much of the cell's architecture, its enzymatic activities, interaction with its environment and many other physiological properties. As the final product of the gene expression is a protein, regulation could be achieved by controlling the transcription of DNA into RNA or the translation of RNA into protein. However, most of the regulation is thought to take place at the level of gene transcription. First, the molecular signals from inside or outside the cell initiate the binding of the regulatory proteins to specific DNA sites outside of protein-encoding regions. Binding of these proteins modulates the rate of transcription. These proteins may directly or indirectly assist *RNA polymerase* in binding to its transcription initiation sites- the promoter; or they may repress transcription by preventing the binding of *RNA polymerase*. The regulation of transcription is a vital process in all living organisms. Most of our information regarding the mechanism of regulation of gene expression has been obtained from *E. coli.* Although bacteria and eukaryotes have much of the logistics of gene regulation in common, there are some fundamental differences in the underlying mechanisms and machinery. Both use sequence-specific DNA-binding proteins to modulate the level of transcription. However, eukaryotic genomes are larger and have a broader range of properties. So naturally their regulation is more complex requiring many more types of regulatory proteins and more extensive interactions with the adjacent regulatory regions in the DNA.

14.3 GENE REGULATION IN PROKARYOTES

Most bacteria are free- living organisms that grow by increasing mass and then divide by binary fission. The growth and division of these bacteria are controlled by genes, the expression of which must be regulated appropriately. *Escherichia coli*, a bacterium that resides in our large intestine, is solely dependent on the nutrients available by us. It must survive under different nutritional environments. *E. coli* makes up for its inability to alter external environment by being internally flexible. For example, if glucose is present, this bacterium utilizes it to generate ATP. In the absence of glucose, it is capable of utilizing lactose, arabinose, maltose, xylose, or any of several other sugars. Similar is the case with amino acids. When amino acids are available, *E. coli* uses them to synthesize proteins. If a particular amino acid is absent, *E. coli* produces enzymes needed to synthesize that amino acid. Thus, *E. coli* has the capability to respond to environmental changes by rapidly regulating gene expression and altering its enzymes. There are two types of genes. Genes whose activity is controlled in response to the needs of a cell or organism are called **regulated genes** and genes whose products are essential to the normal functioning and dividing cells, irrespective of the external conditions are known as **constitutive genes** or **housekeeping genes**. These genes must continue to function under all conditions, e.g., genes that code for enzymes needed for protein synthesis and glucose metabolism.

Bacteria use two different strategies for regulating enzyme synthesis, depending on whether a given enzyme is involved in a catabolic (degradative) or anabolic (synthetic) pathway. The enzymes that catalyze such pathways are often regulated synchronously, synthesis of all the enzymes involved in a particular pathway is turned on and off together. Now we will study two well-understood pathways, one catabolic and one anabolic.

14.3.1 Catabolic Pathways and Substrate Induction

The main function of catabolic enzymes is to degrade specific substrates, often as a means of obtaining energy. The steps involved in catabolic pathway that degrades the disaccharide lactose into simple sugars that are subsequently metabolized by glycolysis are depicted in Figure 14.1. The hydrolysis of lactose by the enzyme *β-galactosidase* into monosaccharide glucose and galactose is the central step of this catabolic pathway. A protein called *galactoside permease* facilitates the transport of lactose into the cell before its hydrolysis. The synthesis of lactose is regulated coordinately with *β-galactosidase.* Since the function of a catabolic enzyme is to degrade a specific substrate, such enzymes are needed only when the cell is confronted by the relevant substrate. For example, the enzyme β-*galactosidase* is required only when cell has excess of substrate lactose; in the absence of lactose the enzyme is superfluous and inactive. Accordingly, for the cellular economy, the synthesis of β-*galactosidase* is to be turned on or induced in the presence of lactose, and to be turned off in its absence. This turning on of enzyme synthesis is called **substrate induction**. Enzymes whose synthesis is regulated by induction in response to substrate are known as **inducible enzymes**.

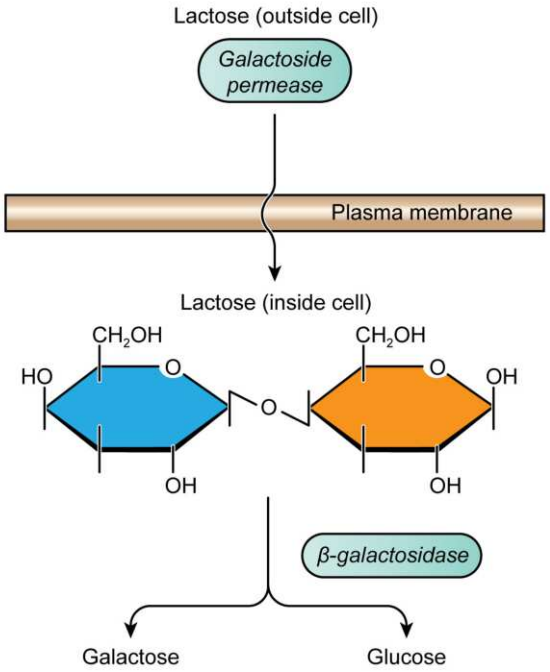


Fig. 14.1: A typical catabolic pathway showing the coordinated regulation of the synthesis of the enzyme *β galactosidase* (for the breakdown of the diasaccharide lactose into glucose and galactose).

14.3.2 Anabolic Pathways and End-Product Repression

The regulation of anabolic pathways is in a sense just the opposite that for catabolic pathways. The enzymes that catalyze the five steps of the anabolic pathways for synthesizing amino acid tryptophan from the precursor chorismate have been shown in the Figure 14.2.

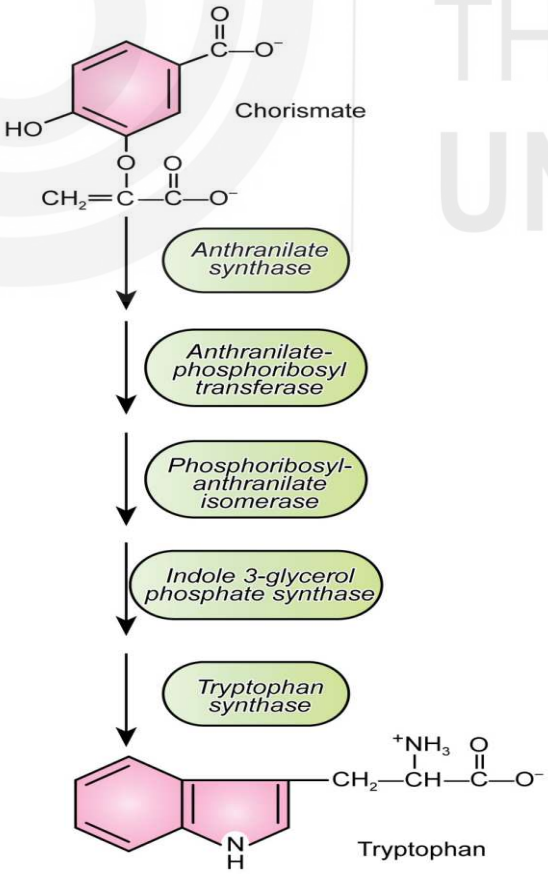


Fig. 14.2: A typical anabolic pathway showing coordinated regulation of enzymes for the synthesis of tryptophan from chorismate.

The first two steps of this pathway, like lactose metabolism is regulated coordinately at the gene level. In the anabolic pathway the concentration of enzyme produced is inversely proportional to the intracellular concentration of the end- product of the pathway. In other words, as the concentration of tryptophan rises, the production of the enzymes involved in synthesizing tryptophan is proportionately reduced. This is advantageous for the cell which economizes on its metabolic resources by reducing the production of enzymes involved in synthesizing tryptophan. Such reduction in the expression of the enzyme-coding genes is called **end-product repression**. Most of the biosynthetic pathways in bacterial cells are regulated in this manner.

14.3.3 Mechanisms of Gene Regulation at Transcriptional Level

Operon Control

Much of what we know about the control of gene expression in bacteria, including the vocabulary used to express that knowledge is based on the pioneering studies carried out by French molecular geneticists Francois Jacob and Jacques Monod. The “operon model” described by them in 1961 probably influenced our understanding of gene regulation more than any other work. In prokaryotes the functionally related genes are clustered together, and their DNA sequences are controlled by a single promoter. A significant difference between bacterial and eukaryotic transcriptional control lies in the organization of functionally related genes that are clustered together with DNA sequences under the control of a single promoter that allows the genes to be turned on (forming single mRNA) and off simultaneously as a single unit. Such a group of bacterial structural genes that are transcribed together is called an **operon**. The organization of functionally related genes into operons is commonly observed in prokaryotes but not in eukaryotes. Nonetheless, the operon model established several basic principles that have shaped our understanding of transcriptional regulation in both prokaryotic and eukaryotic systems.

Operon Structure

An operon is a group of structural genes and sequences that control transcription. The organization of a typical operon has been illustrated in Fig. 14.3. A set of **structural genes** such as gene *a*, gene *b* and gene *c* is present at the end of an operon. These structural genes are transcribed into a single mRNA, which is subsequently translated to produce enzymes A, B, and C. These enzymes carry out a series of biochemical reactions that convert the precursor molecule X into product Y. All the three structural genes are transcribed together under the control of a promoter which lies upstream of the first structural gene. The *RNA polymerase* first binds to the promoter and then moves downstream, transcribing the structural genes.

A **regulator gene** helps to control the transcription of the structural genes of the operon. Even though this gene is not considered as part of operon, it significantly affects operon function. The regulator gene possesses its own promoter and is transcribed into a short mRNA segment which is later translated into a small protein (**regulator protein**)**.**This protein binds to a region of the operon called the **operator** and affects transcription. The operator usually overlaps the 3´ end of the **promoter** and in some cases even the 5´end of the first structural gene.

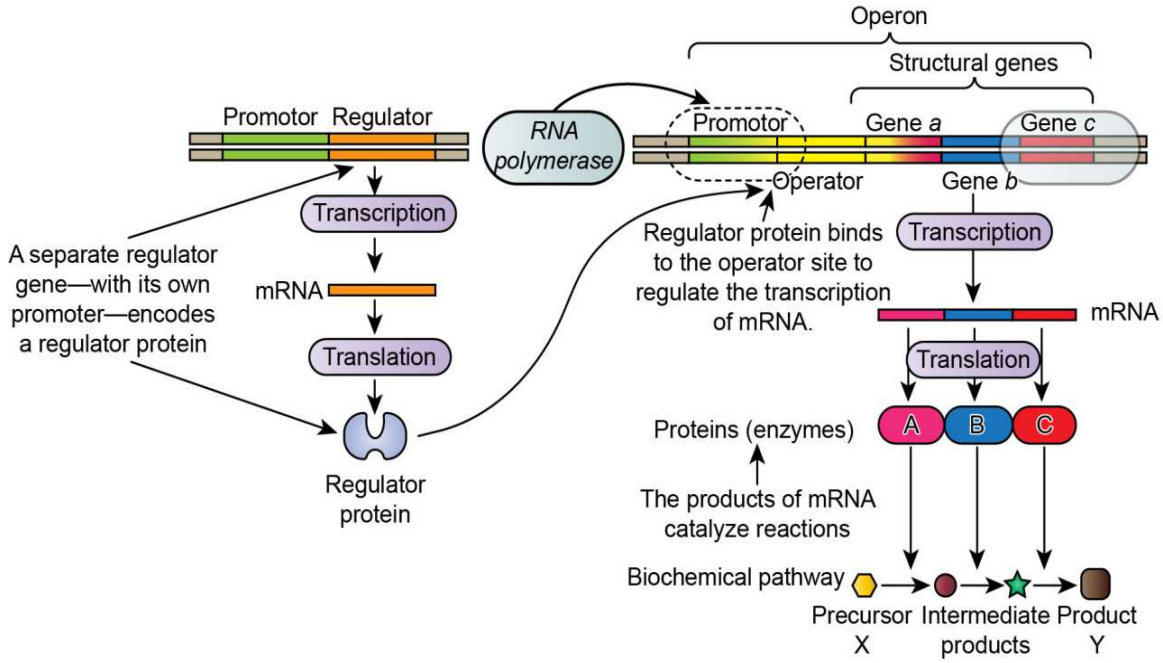


Fig. 14.3: An operon as a single transcriptional unit that includes a series of structural genes, a promoter, and an operator.

Strategies of Gene Regulation

In bacterial cells there are two types of transcriptional regulation: **negative regulation** and **positive regulation** (Fig.14.4)**.**

Negative Regulation: In this type of regulation the regulatory protein present is called an **inhibitor or repressor**. The repressor binds to DNA and inhibiting or suppressing transcription of relevant gene. A signal molecule or inducer is required to react with or remove repressor to allow the initiation of transcription. Therefore, the negative regulation of gene expression is also called repression.

Positive Regulation: In positive regulation an **activator or inducer** molecule activates the promoter which in turn promotes transcription of mRNA. The positive regulation of gene regulation is also called induction and the substance that induces gene action is called inducer.

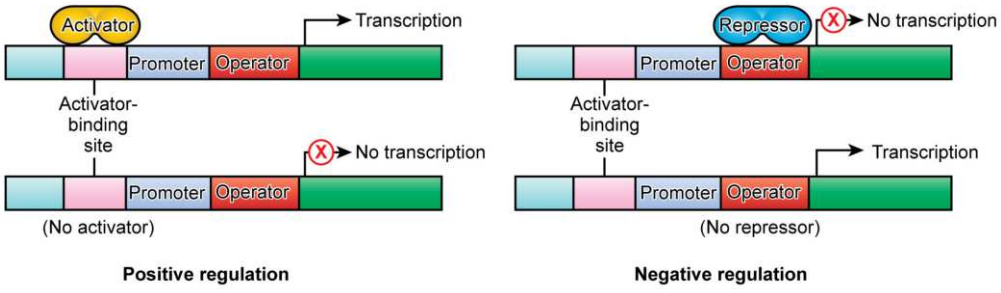


Fig. 14.4: Negative and positive regulation of bacterial gene action by regulatory protein control.

Operons can also be either **inducible** or **repressible**:

Inducible operons: In such systems gene or genes normally remain switched off. They must be induced or 'switched on' to produce mRNA needed for the synthesis of required enzyme or enzymes. The substance which induces the gene for protein synthesis or enzyme production is known as an **inducer**.

Repressible operons: In this system gene or genes are normally switched on. The activity of these genes is suppressed, and the synthesis of specific protein is stopped or reduced. The substance which stops or suppresses the protein synthesis is known as **repressor**.

There are several varieties of basic control mechanisms for regulation of transcription in bacteria:

Negative inducible operons: The **regulator gene** encodes an active repressor that readily binds to the **operator**. In the absence of an inducer, the repressor binds to the operator and physically blocks the binding of *RNA polymerase* to the promoter. This is because the operator sites overlap the promoter site. As a result, the structural genes *d, e*, and *f* (which metabolize precursor V) cannot synthesize (Fig.14.5 a). Hence, there is no transcription.

Transcription can only be turned on when a small molecule, an **inducer**, binds to the repressor (Fig.14.5 b).The regulatory proteins frequently have two binding sites: one that binds to DNA and another that binds to a small molecule called **inducer**. Binding of the inducer (precursor V in Fig. 14.5 b) alters the shape of the repressor, preventing it from binding to DNA. The DNA molecule is now free to transcribe.

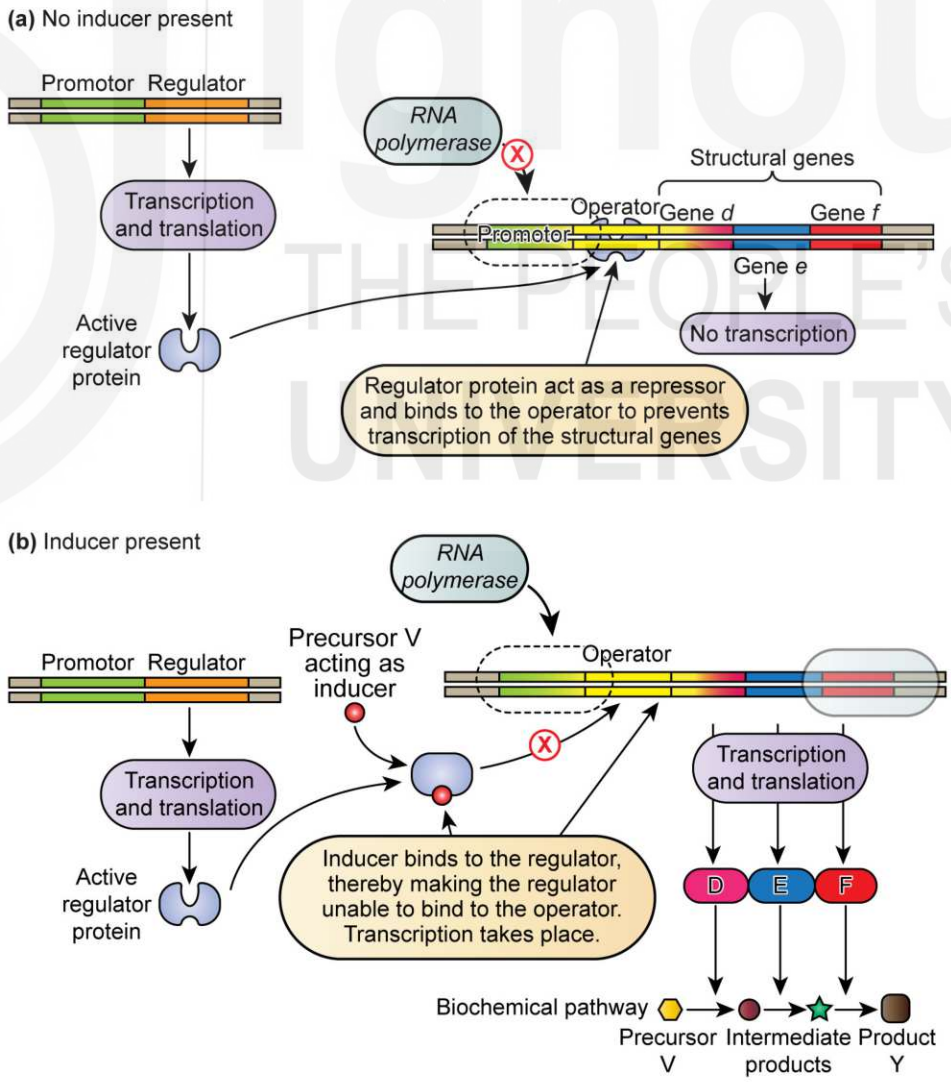


Fig. 14.5: Negative inducible operons. a) Structural genes *d*, *e*, and *f* become switched off when the inducer is absent; b) Genes *d*, *e*, and f and switched on when the inducer is present.

Proteins, which can change their shape on binding to another molecule, are called **allosteric proteins**. This is a type of inducible system as transcription is normally off (inhibited) and must be turned on (induced). This mechanism is based on economy and adaptation. When no precursor V is available, it would be meaningless to have enzyme synthesis as there would be no substrate to metabolize. As soon as precursor V becomes available, some of it binds to the repressor, rendering the repressor inactive. Now the repressor is unable to bind to the operator site. So, an operon with negative inducible control regulates the synthesis of enzyme with utmost economy: the enzyme is synthesized only when their substrate (V) is available.

Negative repressible operons: In such type of repressible system, the activity of gene or genes is normally kept suppressed or turned off. This is because the regulator protein in this type of operon is also a repressor. But it is synthesized in an inactive form and cannot bind to the operator by itself. Since there is no repressor bound to the operator, *RNA polymerase* readily binds to the promoter and transcription of structural genes takes place. The activity of these genes can be turned on only if the repressor is activated. This job is done by a small molecule that binds to the repressor and makes it capable of binding to the operator and is called **corepressor**. As illustrated in Figure 14.6 a, the product U of the metabolic reaction is a corepressor. So, the negative repressible operon regulates the gene expression depending on the availability of the product U of the metabolic reaction.

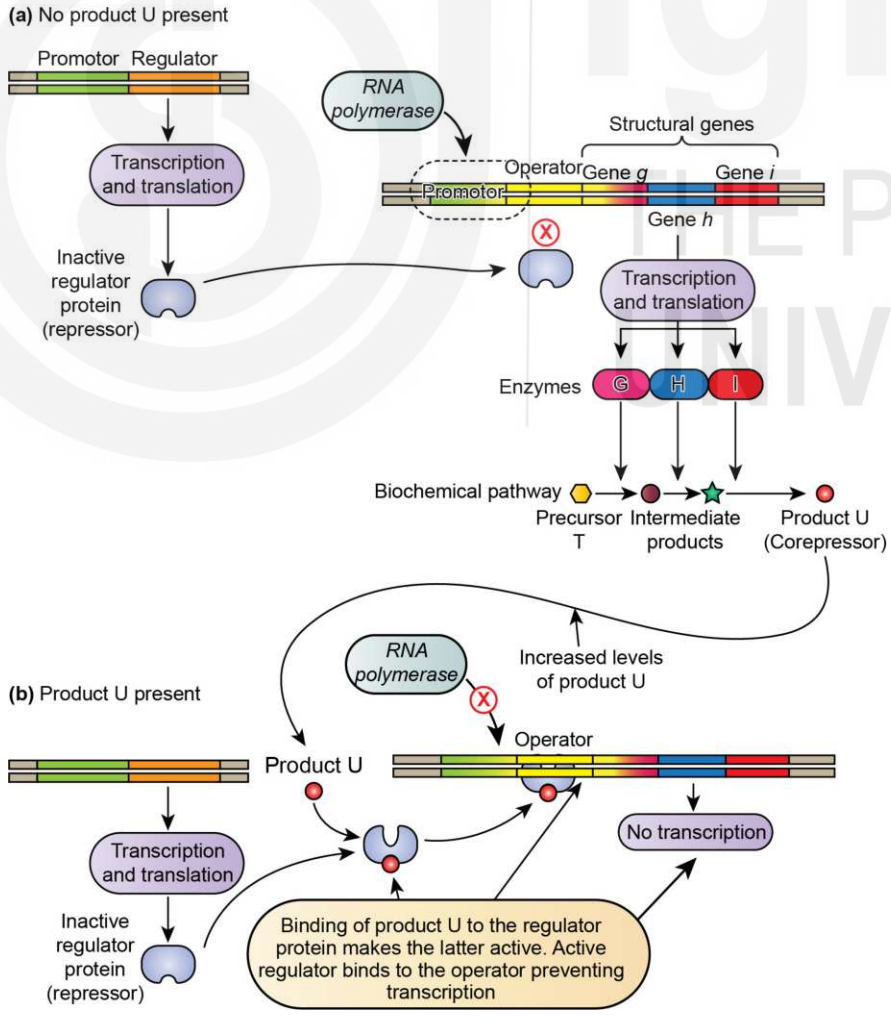


Fig.14.6: Negative repressible operons. a) Structural genes *g, h,* and *i* become switched on when the metabolic product U is absent; b) Structural genes *g, h,* and *i* become switched off when the product U is present.

When the level of the product U is high: The repressor gets activated after binding with the product U, preventing transcription. With the operon repressed, enzymes *g, h,* and *i* are not synthesized, and no more U is produced from precursor T (Fig. 14.6 b).

When the product U is absent: When the entire product U is used up by the cell, the repressor no longer gets activated in the absence of U. Consequently, the repressor cannot bind to the operator. This inactivation of repressor allows the transcription of structural genes and the synthesis of enzymes *g, h* and *i*. This result in conversion of precursor T into product U (Fig.14.6 a). Like inducible operons, repressible operons are also exhibit economy as the enzymes are synthesized only when they are needed.

As mentioned above, both the inducible and the repressible systems represent forms of negative control, where the regulatory protein functions as a repressor. You will now study the positive control in which the regulator protein itself stimulates transcription.

Positive control: In this mode, a regulatory protein is an activator with a positive control. The activator binds to the DNA usually at a site other than the operator and stimulates transcription. Positive control can be inducible or repressible.

Inducible positive operon: In this system transcription normally remains turned off because the regulator protein (an activator) is produced in an inactive form. Transcription can take place only when the regulatory protein is activated through attachment of an inducer. It is, therefore, logical to assume that the inducer should necessarily be the precursor of the reaction which is controlled by the operon so that the necessary enzymes would be produced only when the substrate for their reaction was present (Fig.14.7c).

Repressible positive operon: In this operon, the regulatory protein is produced in an active form that readily binds to DNA to switch on the transcription. Transcription can be stopped/ inhibited only when a substance becomes attached to the activator and makes it unable to bind to the DNA. It means that the product of the reaction controlled by the operon would logically be the repressing substance, as shown in the Fig.14.7d. It is economical for the cell to prevent the transcription of genes that allow the synthesis of the product when these products are already available in excess. The characteristics of positive and negative control in inducible and repressible operons have been summarized in Fig.14.7 (a, b).

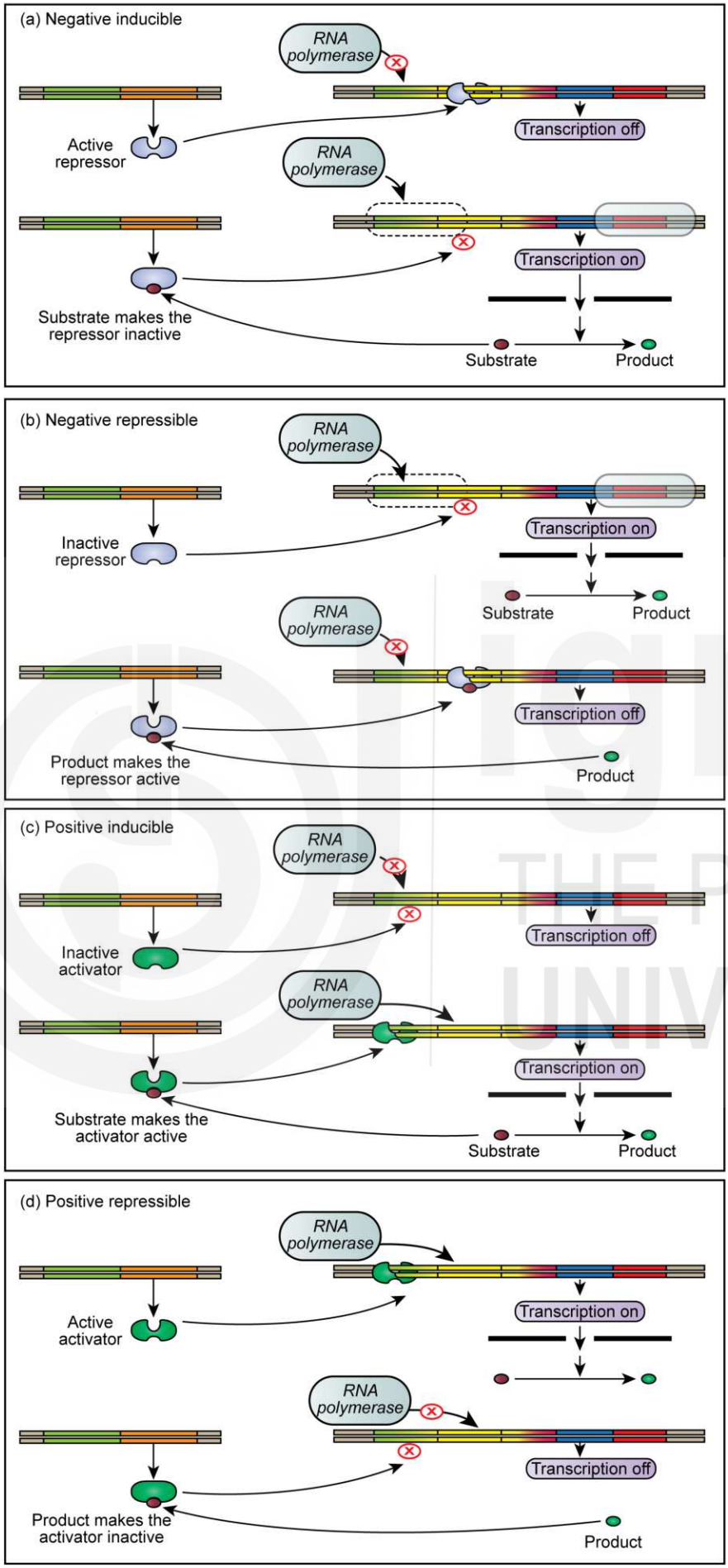


Fig. 14.7: Summary of the characteristics of negative and positive controls in inducible and repressible systems.

14.3.4 Model Systems for Gene Regulation in Prokaryotes

Lac operon and *trp* operon are model systems for understanding of gene regulation in prokaryotic organisms.

The *lac* operon of *E. coli*: An example of a negative inducible operon

The operon model proposed by Jacob and Monod proved to be a basic model of gene regulation with far-reaching implications. A key feature of the operon model is the idea that genes with metabolically related functions are clustered together so their transcription can be regulated as a single unit. But how is this regulation accomplished? Jacob and Monod addressed this question by studying genetic control of lactose metabolism in *E. coli*.

Lactose metabolism and regulation of *lac* operon: Lactose is a major carbohydrate found in milk. It is metabolized by *E. coli* residing in mammalian gut. Lactose does not easily diffuse across the *E. coli* cell membrane and be actively transported into the cell by the protein *permease*. Since lactose cannot be used directly as an energy source. Transport of lactose into the cell is an active process and is mediated by the enzyme *permease*. *E. coli* first breaks it into glucose and galactose. This reaction is catalyzed by the enzyme *β −galactosidase*. In addition, this enzyme also converts lactose into allolactose, a compound that plays important role in regulating lactose metabolism itself. A third enzyme, *thiogalactoside transacetylase* is also produced by *lac* operon, but its function in lactose metabolism is not yet fully understood (Fig. 14.8).

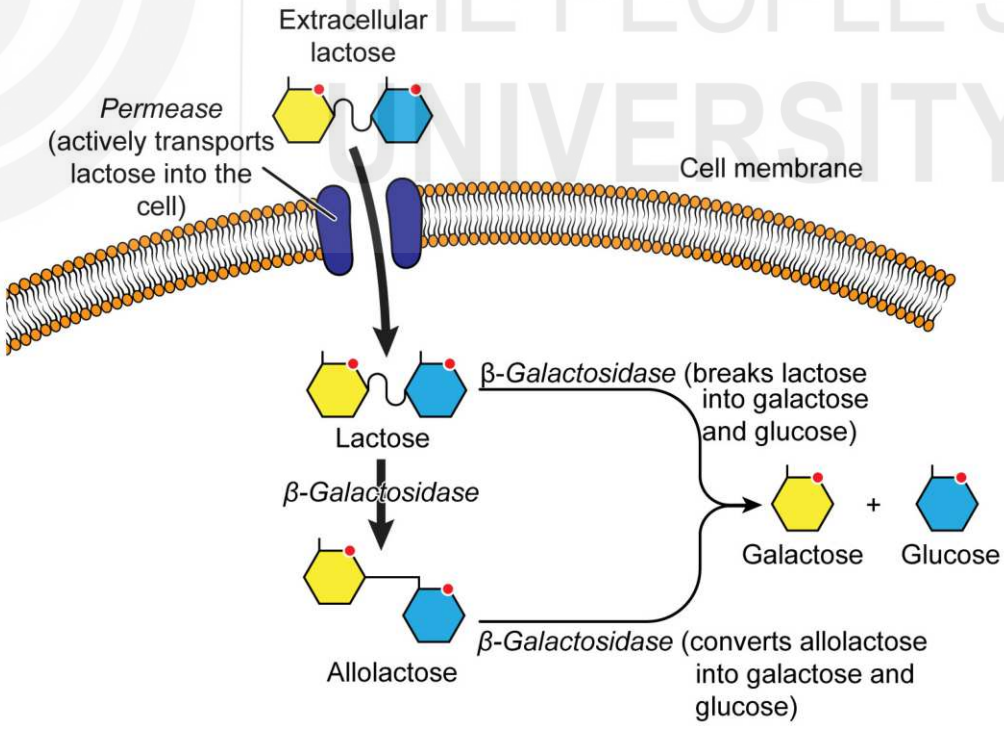


Fig.14.8: Enzymes involved in lactose metabolism in *E. coli* cells.

The *lac* operon is an example of an inducible operon. The adjacently placed structural genes in *lac* operon in *E. coli*, along with a common promoter *lac P* encode the enzymes *β −galactosidase*, *permease*, and *transacetylase*

respectively (Fig. 14.9). *β −galactosidase* is encoded by *lac Z* gene, *permease* by *lacY* gene and *thiogalactoside transacetylase* by the *lac A* gene. As you have already studied about the negative inducible operon earlier, lactose seems to be inducer here and allolactose is responsible for induction. There are two situations in which lactose is metabolized by *lac* operon.

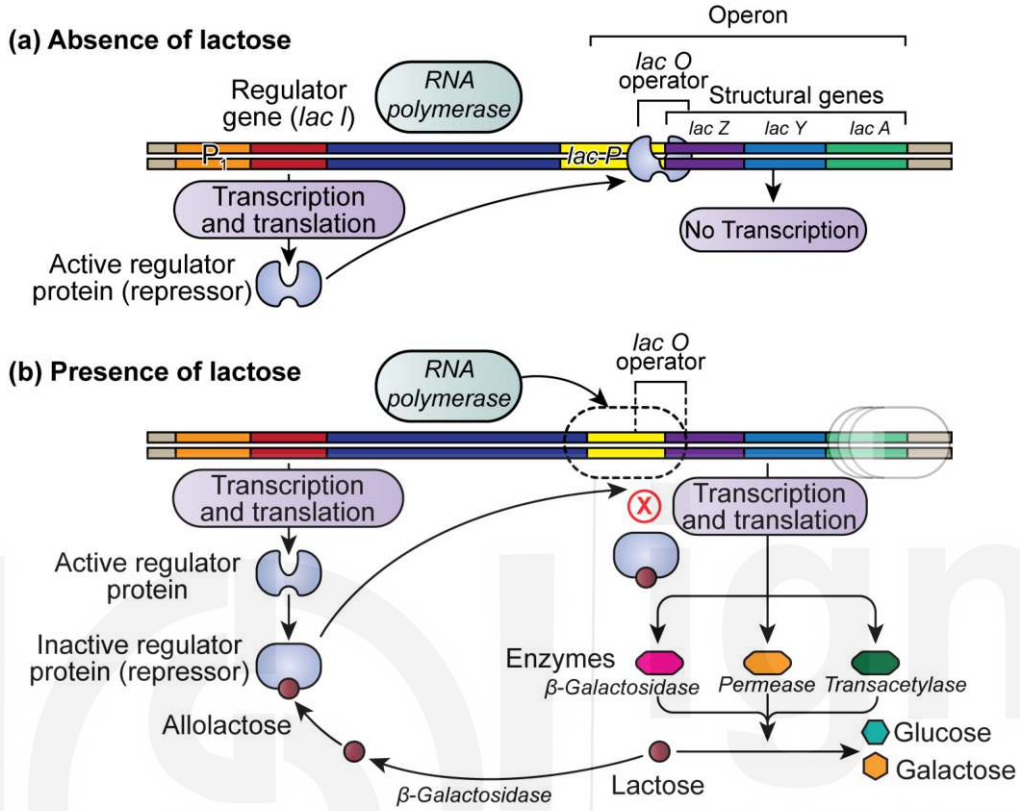


Fig. 14.9: The *lac* operon regulates lactose metabolism.

a) In the absence of lactose, transcription is inhibited because the regulator protein (repressor) binds to the operator which blocks the binding of *RNA polymerase*.

b) When lactose is present, the structural genes are transcribed and translated. This is because some of the lactose gets converted into allolactose which then binds to the regulator protein, making the protein inactive. Since the regulator protein cannot bind to the operator the structural genes get freely transcribed and translated.

The *trp* operon model of *E. coli*: An example of a negative repressible operon

Although the operon concept was formulated by the lac operon of *E. coli*, many other bacterial regulatory systems are now known to follow the same general pattern. As discussed earlier, operons coding for the enzymes involved in catabolic pathway resemble the *lac* operon in being inducible. Here the operons are turned on by a specific allosteric effector, which is usually substrate itself for the pathway involved. As against the catabolic pathway, operons that regulate enzymes involved in anabolic (biosynthetic) pathways resemble tryptophan (*trp*) operon in being repressible operons. They can turn off allosterically by an effector that is itself the product of the reaction. The tryptophan (*trp*) operon contains genes coding for enzymes involved in tryptophan biosynthesis, along with DNA sequences that regulate the production of these enzymes. The effector molecule in this case is the end-

product of the biosynthetic pathway, the amino acid tryptophan. The *trp* operon contains five structural genes *trp*E, *trp*D, *trp*C, *trpB* and *trpA* that produce the component of three enzymes (two of the enzymes consists of two polypeptide chains). These enzymes convert chorismate (the key intermediate of the tryptophan biosynthesis) into tryptophan (Fig. 14.10).

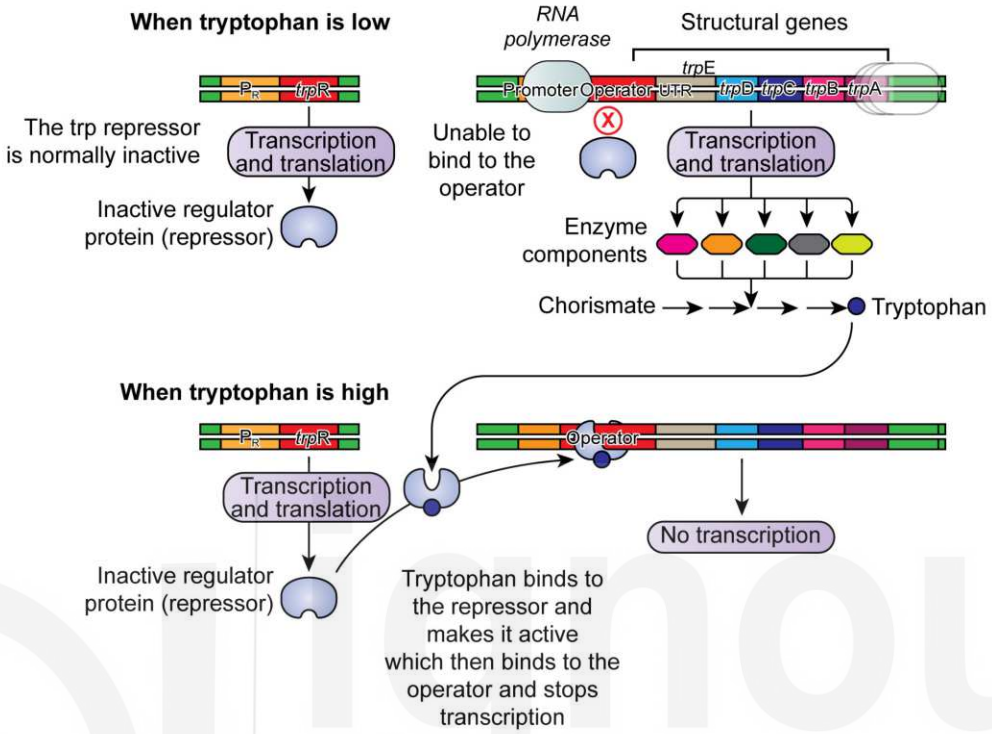


Fig. 14.10: The *trp* operon in *E. coli* controlling the biosynthesis of the amino acid tryptophan.

The first structural gene, *trp e*, contains a long 5´untranslated region (5´UTR) that is transcribed but does not encode any of these enzymes. Instead, this 5´UTR plays an important role in other regulatory mechanisms. The *trp* promoter is located upstream of the 5´ UTR.Tryptophan is synthesized by this operon in two situations.

a) **Low tryptophan levels**: *RNA polymerase* binds to the promoter and transcribes the five structural genes into a single mRNA, which is then translated into enzymes that convert chorismate into tryptophan.

b) **High tryptophan levels:** Some distance from the *trp* operon is a regulator gene, *trpR* which encodes a repressor that alone cannot bind DNA. Like the *lac* repressor, the tryptophan repressor has two binding sites, one that binds to DNA at the operator site and another which binds the tryptophan (the activator). Binding with tryptophan results in a conformational change in the repressor. This change makes the repressor capable of binding with DNA at the operator site. This site overlaps the promoter. When the operator is occupied by the tryptophan repressor, *RNA polymerase* cannot bind to the promoter and the structural genes cannot be transcribed.

Thus, when the cellular levels of tryptophan are low, transcription of *trp* operon takes place and more tryptophan is synthesized but when the cellular levels of tryptophan are high, transcription of the *trp* operon is inhibited and the synthesis of more tryptophan does not take place (Fig.14.10).

14.3.5 Transcription Regulation through Attenuation

You have studied how in different ways a cell can regulate the initiation of transcription in an operon. Some operons have additional level of control that affects the continuation of transcription. However, it does not influence initiation of transcription. This is referred to as **attenuation**. Here, transcription begins at the start site, but termination takes place prematurely (i.e., before *RNA polymerase* even reaches the structural genes). Attenuation takes place in several operons that encode enzymes participating in the biosynthesis of amino acids. One of the best studied examples of the process of attenuation is the *trp* operon of *E. coli*.

Attenuation in the *trp* Operon in *E. coli*:

The *trp* operon is regulated both by repression and by attenuation. This is quite unusual as most operons are regulated by one of these mechanisms but not by both. Charles Yanofsky et al. made several observations in 1970 that indicated that repression at the operator site is not the only method of regulation in the *trp* operon. These scientists isolated several mutants that exhibited high level of transcription even if the control at the operator site was unaffected. Their observations indicated that some mechanism other than repression alone at the operator site was controlling the transcription. Furthermore, they observed that two mRNAs of different sizes were transcribed from the *trp* operon: a long mRNA containing sequences for the structural genes and a much shorter mRNA of only 140 nucleotides (Fig. 14.11 a).

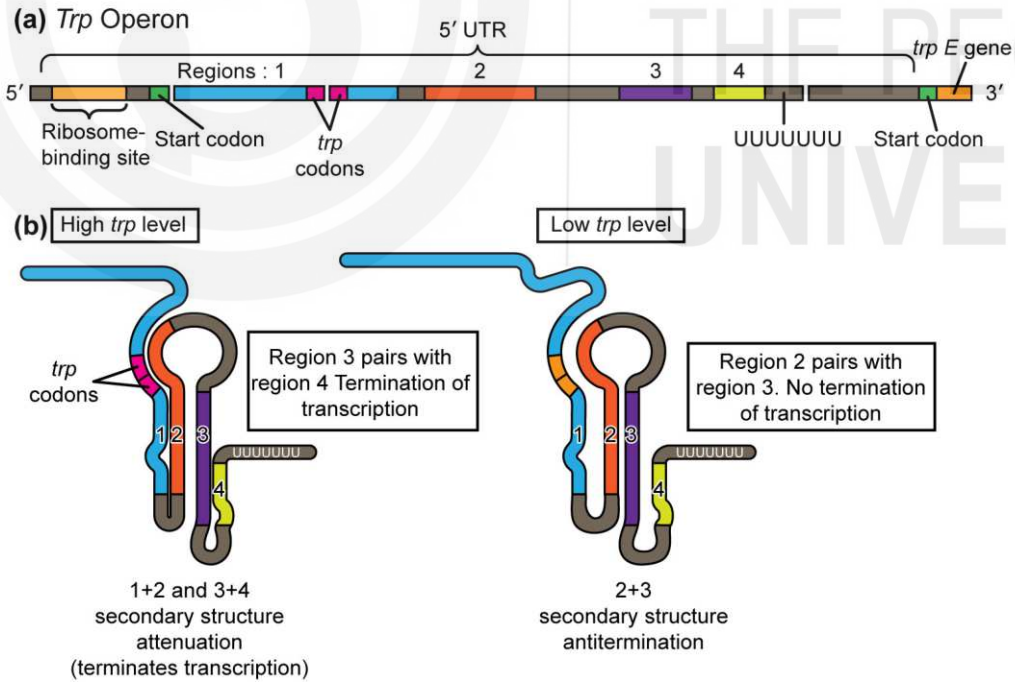


Fig. 14.11: a) The *trp* operon showing different regions of transcription; b) Mechanism of attenuation by two different secondary structures formed by the 5´ UTR of the mRNA transcript of the *trp* operon.

These observations led to the proposal that the premature termination of transcription also regulates transcription in tryptophan operon. A region of 162 nucleotides corresponds to the long 5´ UTR of the mRNA transcribed from the *trp* operon. The 5´ UTR (also called a leader) contains four regions: region 1 is

complementary to region 2, region 2 is complementary to region 3, and region 3 is complementary to region 4. These complementarities allow the 5´ UTR to fold into two different secondary structures (Fig.14.11 b). Only one of these secondary structures causes attenuation when the tryptophan levels are either low or high.

High tryptophan levels:

One of the secondary structures contain one hairpin produced by the base pairing of regions 1 and 2 and another hairpin produced by base pairing of regions 3 and 4. There is a noticeable string of uracil nucleotides that follows the 3 + 4 hairpin. Since a bacterial intrinsic terminator also includes a hairpin followed by a string of uracil nucleotides, this secondary structure in the 5´ UTR of the *trp* operon acts a terminator called **attenuator**.

Low tryptophan levels:

Under low tryptophan levels, the alternative secondary structure of the 5´ UTR is produced by the base pairing of the regions 2 and 3 (Fig.14.11b). This base paring also produces a hairpin, but this hairpin is not followed by a string of uracil nucleotides. Thus, this structure does not function as a terminator. As a result, *polymerase* continues past the 5´ UTR into the coding section of the structural genes and the enzymes for the synthesis of tryptophan are made. Since this 2+3 secondary structure prevents the termination of transcription, it is called an **antiterminator**.

We have described above as to how the expression of genes is regulated in prokaryotes and the details of the mechanisms involved in these processes which are now well understood in prokaryotes. These mechanisms are in principle also applicable to eukaryotes, but with some significant differences, which we are going to understand now. Before describing the process in eukaryotes, you may like to work out the following SAQs.

SAQ 1

Fill in the blanks with the suitable word from the words given below.

a) A group of bacterial structural genes that are transcribed together along with their promoter and additional sequences which control transcription are called …….. .

b) The gene that controls the transcription of structural genes of the operon is called ……

c) The region of the operon where regulator protein can bind and affect whether transcription can take place or not is called…….. .

d) The *lac* repressor cannot bind operator in the presence of …….. .

e) In negative repressible operon, the regulator protein is synthesized as an inactive …….. .

(Regulator gene, Repressor, Operator gene, Operon, Allolactose)

14.4 GENE REGULATION IN EUKARYOTES

In the early days of molecular biology, there was a popular saying that "what is true of *E. coli* is also true of elephants." This maxim expressed the initial conviction that almost everything learnt about bacterial cell function at the molecular level would also apply to eukaryotes. Such predictions have turned out to be only partly true. We now know that there are several similarities in DNA structure, protein synthesis, and enzyme function between the bacterial and eukaryotic worlds and findings from studies with bacteria can often even be extrapolated to eukaryotes. Yet, when it comes to regulation of gene expression, the comparison needs to be scrutinized more carefully. Although eukaryotic cells and bacteria have many features of gene regulation in common, they differ in the following important ways:

i) Most of the eukaryotic genes are not organized into operon and are rarely transcribed together into a single mRNA molecule. Each structural gene possesses its own promoter and is therefore, transcribed individually.

ii) In eukaryotic cells, gene expression is significantly influenced by the chromatin structure. DNA must first unwind from the histone proteins before transcription can take place.

iii) The presence of nuclear membrane in eukaryotic cells separates transcription and translation in a temporal (time) and spatial (space) manner.

Therefore, the regulation of gene expression in eukaryotic cells is characterized by a greater diversity of mechanisms that act at different points in the transfer of information from DNA to protein. Let us now understand some important mechanisms through which eukaryotic genes are regulated.

14.4.1 Gene Regulation by Chromatin Modification

The modification of chromatin structure constitutes an important type of gene control in the eukaryotic cells. Three different processes affect gene regulation by altering chromatin structure:

i) The modification of histone proteins

Histones in the octamer core of nucleosome have two domains: (a) a globular domain that associates with other histones and the DNA and (b) a positively charged tail domain that interacts with negatively charged phosphate groups on the backbone of the DNA. Addition or removal of phosphate groups, methyl groups, or acetyl groups often modify the tails of these histone proteins. These modifications have been called the **histone code** as they encode information affecting genes expression.

Methylation of histones

One type of histone modification is the addition of methyl groups to the tails of histone proteins. Such modifications can bring about either the activation or the repression of transcription, depending on which amino acids in the histone tail are methylated. A common modification is the addition of three methyl groups to lysine 4 in the tail of the H3 histone protein, abbreviated as H3K4me3 (K is the abbreviation for lysine). The H3K4me3 modification is

commonly found in promoters of transcriptionally active genes among eukaryotes.

Acetylation of histones

Histone modification by acetylation involves the addition of acetyl groups (CH_3CO) to histone proteins (Fig. 14.12). Acetylation of histones usually stimulates transcription. For example, the addition of a single acetyl group to lysine 16 in the tail of the H4 histone prevents the formation of the 30 nm chromatin fiber, thereby causing the chromatin to be in an open configuration and available for transcription. Thus, acetyl groups destabilize chromatin structure, allowing transcription to take place.

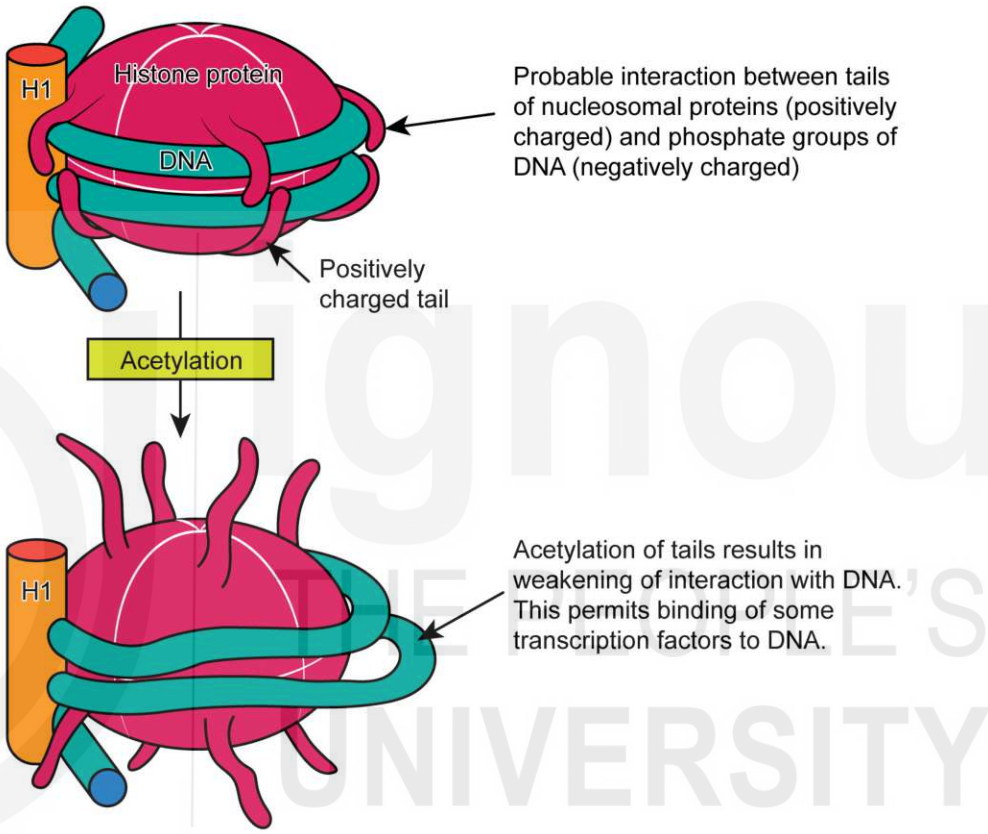


Fig. 14.12: Histone modification by acetylation that alters chromatin structure allowing some transcription factors to bind to bind to DNA.

Addition of acetyl groups to histone proteins is catalyzed by *acetyl transferase* enzymes. On the other hand, *deacetylases* remove acetyl groups from histones and restore chromatin structure, which represses transcription. Certain transcription factors and other proteins that regulate transcription either have *acetyl transferase* activity or attract *acetyl transferases* to DNA.

ii) Chromatin Remodelling

Some transcription factors and other regulatory proteins also alter chromatin structure without altering the chemical structure of the histones directly. These proteins are called **chromatin-remodelling complexes**. They bind directly to sites on DNA and reposition the nucleosomes, allowing transcription factors to bind to promoters and initiate transcription. This complex utilizes energy derived from the hydrolysis of ATP to reposition nucleosomes, exposing promoters in the DNA to the action of other regulatory proteins and *RNA polymerase.*

The remodelling complexes reposition nucleosomes in such a manner to make the DNA more accessible for gene expression by inducing conformational changes.

iii) DNA Methylation

Another change in chromatin structure associated with transcriptional regulation is the methylation of cytosine bases, which yields 5-methylcytosine. The methylation of cytosine in DNA is distinct from the methylation of histone proteins. Heavily methylated DNA is associated with the repression of transcription in vertebrates and plants, whereas transcriptionally active DNA is usually unmethylated in these organisms. Abnormal patterns of methylation are also associated with some types of cancer. DNA methylation is most common on cytosine bases adjacent to guanine nucleotides (CpG, where p represents the phosphate group in the DNA backbone); so, two methylated cytosines sit diagonally across from each other on opposing strands DNA regions with many CpG sequences are called **CpG islands** and are commonly found near transcription start sites.

$$5'-\overset{m}{C}G-3'$$
$$3'-G\underset{m}{C}-5'$$

While genes are not being transcribed, these CpG islands are often methylated, but the methyl groups are removed before the initiation of transcription. Silencing by methylation may influence a single gene, a group of genes or even a whole of chromosome. CpG methylation is also associated with long-term gene repression, such as on the inactivated X chromosome of female mammals.

14.4.2 Gene Regulation by Transcription Factors and Transcriptional Regulator Proteins

We have studied in Unit 12 (Transcription) that during eukaryotic transcription, general transcription factors and *RNA polymerase* assemble into a *basal* **transcription apparatus**, which binds to a **core promoter** located immediately upstream of a gene. The basal transcription apparatus is capable of minimal levels of transcription; **transcriptional regulator proteins** are required to bring about normal levels of transcription. These proteins bind to a regulatory promoter, which is located upstream of the core promoter (Fig. 14.13), and to **enhancers**, which may be located some distance from the gene. Some transcriptional regulator proteins are activators, stimulating transcription. Others are repressors, which inhibit transcription.

Transcriptional Activators and Coactivators

Transcriptional activator proteins stimulate and stabilize the basal transcription apparatus at the core promoter. The activators may interact either directly with the basal transcription apparatus or indirectly through protein coactivators. Some activators and coactivators, as well as the general transcription factors, also show *acetyl transferase* activity thereby further stimulating transcription by altering chromatin structure. Transcriptional activator proteins have two distinct functions (Fig.14.13):

i) They are capable of binding DNA at a specific base sequence, usually a consensus sequence in a regulatory promoter or enhancer; most transcriptional activator proteins contain one or more DNA-binding motifs, such as the helix-turn-helix, zinc finger, and leucine zipper.

ii) They have ability to interact with other components of the transcriptional apparatus and influence the rate of transcription. Within the regulatory promoter are several different consensus sequences to which different transcriptional activators can bind. Among different promoters, activator-binding sites are mixed and matched in different combinations and so each promoter is regulated by a unique combination of transcriptional activator proteins.

Transcriptional activator proteins bind to the consensus sequences in the regulatory promoter and affect the assembly or stability of the basal transcription apparatus at the core promoter. One of the components of the basal transcription apparatus is a complex of proteins called the **mediator** (Fig. 14.13). Transcriptional activator proteins binding to sequences in the regulatory promoter or enhancer come in contact of the mediator and affect the rate of transcription. Some regulatory promoters also contain sequences that are bound by proteins that lower the rate of transcription through inhibitory interactions with the mediator.

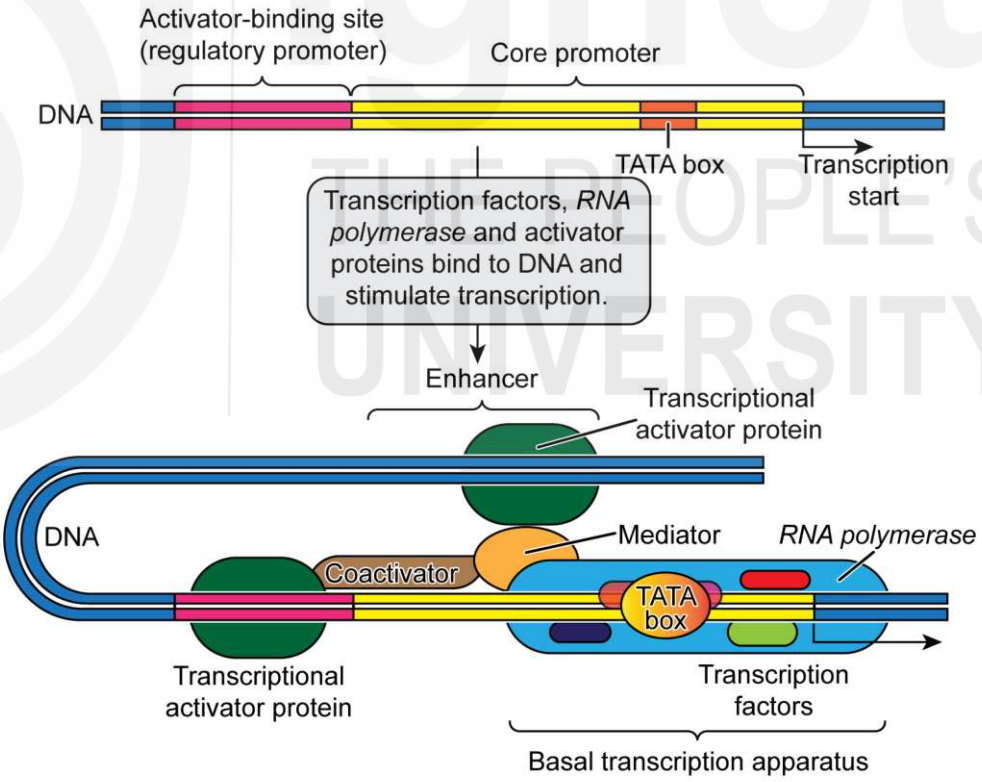


Fig. 14.13: Regulation of transcription by activator proteins.

14.4.3 Coordinated Gene Regulation

As discussed earlier, most eukaryotic cells do not possess operons, still several of the eukaryotic genes may be activated by the same stimulus. As you have studied in bacterial gene regulation, groups of bacterial genes are often coordinately expressed (turned on and off together) because they are physically clustered as an operon and have the same promoter, but

coordinately expressed genes in eukaryotic cells are not clustered. How, then, is the transcription of eukaryotic genes coordinately controlled if they are not organized into an operon? Genes that are coordinately expressed in eukaryotic cells can respond to the same stimulus as their promoters or enhancers have common short regulatory sequences. For example, different eukaryotic heat-shock genes possess a common regulatory element upstream of their start sites. Such short DNA regulatory sequences are called **response elements** and contain consensus sequences at varying distances from the gene being regulated (see Table 14.1).

Table 14.1: Few response elements found in eukaryotes with their consensus sequences.

Response Element	Responds to	Consensus Sequence
Heat-shock element	Heat and other stress	CNNGAANNTCCNNG
Glucocorticoid response element	Glucocorticoids	TGGTACAAATGTTCT
Phorbol ester response element	Phorbol esters	TGACTCA
Serum response element	Serum	CCATATTAGG
Source: After B. Lewin, G*enes IV* (Oxford University Press, 1994), p.880		

The response elements are binding sites for transcriptional activators. A transcriptional activator protein binds to the response element and enhances the rate of transcription. The same response element may be present in different genes, thus allowing multiple genes to be activated by the same stimulus. Also, a single eukaryotic gene may be regulated by several different response elements. For example, the metallothionein gene protects cells from the toxicity of heavy metals by encoding a protein that binds to heavy metals and removes them from cells. The basal transcription apparatus assembles around the TATA box, just upstream of the transcription start site for the metallothionein gene, but the apparatus alone is capable of only low rates of transcription (Fig. 14.14).

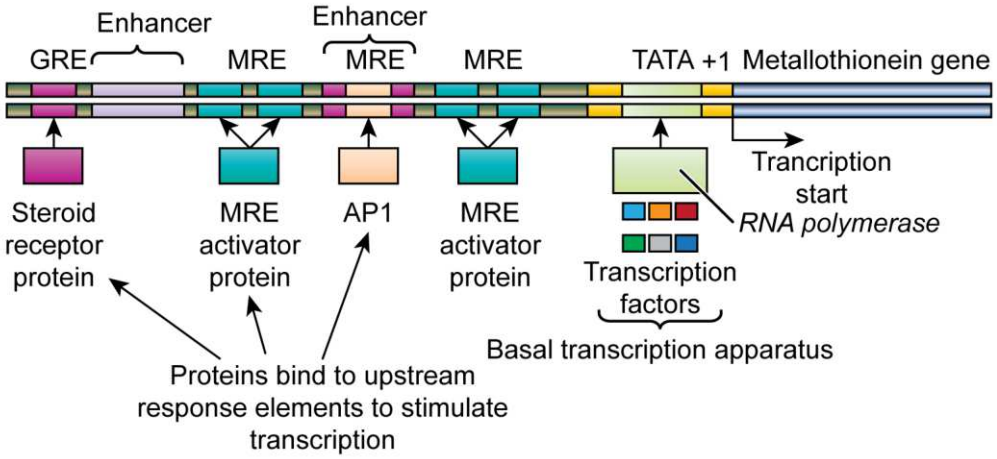


Fig. 14.14: Coordinated gene regulation in eukaryotes with the help of common regulatory sequences in their promoters and enhancers called as response elements. A single metallothionein gene is activated by several different response elements.

We have discussed the gene regulation in eukaryotes in detail and understood how it differs from prokaryotes. Now before understanding how the different genes may be silenced, you may like to try an SAQ.

SAQ 2

State in the space given, whether the following statements about eukaryotic translation are true (T) or false (F).

a) One of the mechanisms of epigenetic changes is methylation of histones. [ ]

b) Most transcriptional activator proteins affect transcription by interacting with basal transcription apparatus. [ ]

c) The tails of histone proteins are often modified by the addition or removal of phosphate groups, methyl groups, or acetyl groups. These modifications have sometimes been called the histone code. [ ]

d) Chromatin-remodelling complexes are proteins that bind directly to sites on DNA and reposition the nucleosomes, allowing transcription factors to bind to promoters and initiate transcription. [ ]

e) Like prokaryotes, most of the eukaryotic genes are organized into operon and are transcribed together into a single mRNA molecule. [ ]

Andrew Z. Fire and Craig C. Mello jointly received Nobel Prize in Physiology or Medicine in 2006 “for their discovery of RNA interference-gene silencing by double stranded DNA. Their discovery revolutionized scientists’ thinking about the regulation of gene expression Small RNA particularly siRNA, can also participate in silencing of genes at the level of transcription. RNAi has now been extensively used to silence genes in *Drosophila melanogaster* and in many plants. This cutting edge technology holds a very promising future in the possible down grading of HIV and many oncogenes.

14.5 GENE SILENCING

We have so far in this Unit discussed the regulation of transcription at the individual gene level, i.e., how do the specific activators and repressors regulate gene transcription and how chromatin structure plays a role in regulating the initiation of transcription. G**ene silencing** is the regulation of gene expression in a cell that prevents the expression of certain genes. Gene silencing is also often considered as **gene knockdown** wherein the expression of one or more gene is reduced. Gene silencing is possible either during transcription or translation.

14.5.1 Transcriptional Gene Silencing

Gene silencing is the phenomenon generally defined as an epigenetic modification of gene expression (heritable changes in genome function that are not associated with DNA sequence alteration) leading to inactivation of previously active individual genes or larger chromosome region. Repression of genes may involve changes in level of DNA methylation and alterations in covalent modifications of histone proteins, which result chromatin condensation thereby making genes inaccessible to transcription machinery. Another example of DNA methylation affecting gene expression is **genomic imprinting** (Fig.14.15).

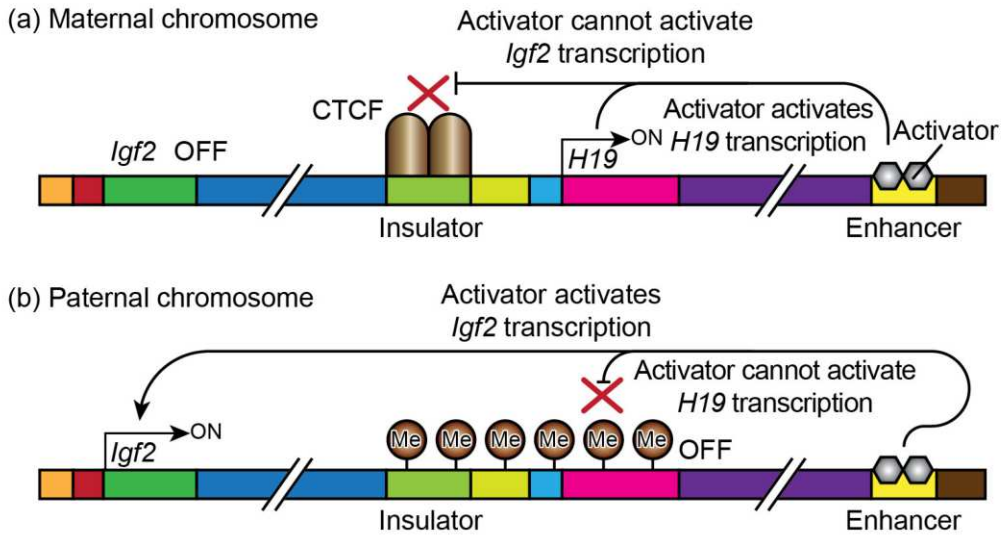


Fig. 14.15: Model for the imprinting of the *Igf2* and *H19* genes for a) maternal chromosome; b) paternal chromosome.

Genomic Imprinting: It is an epigenetic phenomenon in which the expression of certain genes is determined by whether the gene is inherited from the female or male parent. This example involves the first imprinted locus identified. This locus contains a pair of genes – *Igf 2* (which encode insulin like growth factor 2) and *H19* (which encodes an untranslated mRNA of unknown function) – located about 80 kb apart in humans and mice. The imprinting process of *Igf 2* and *H19* is as follows: A single enhancer located downstream of the *Hl9* gene controls the expression of both genes. When activators bind to the enhancer, the transcription machinery could be recruited to both genes. However, another regulatory element located between the genes affects this activation. This regulatory element is an **insulator**. When the insulator is functional it blocks the activation of a promoter to one side of (left side in the Figure) by activators bound to an enhancer on the other side of it (right side in the Figure).

On the maternal chromosome: The genes and the regulatory sequences are not methylated on the maternal chromosome, allowing the protein CCCTC-binding factor (CTCF) to bind to the insulator (Fig.14.15 a). The bound CTCF functions as a transcriptional repressor of *Igf 2*, preventing transcription activation by the activator at the enhancer. It does so by recruiting a complex containing HDAC, which causes chromosome remodeling as described earlier. But the enhancer- bound activator can activate H19 transcription. In other words, on the maternal chromosome, *Igf 2* is inactive and *H19* is active.

On the paternal chromosome: On the paternal chromosome the DNA is methylated for a segment of chromosome encompassing the promoter of *H19* gene and the insulator (Fig. 14.15b). Therefore, the CTCF cannot bind to the insulator, and the activator is then able to activate the transcription of *Igf2*. Because of the methylation of *H19* promoter, and the resulting chromosome remodeling of the transcription of *H19* is silenced. In other words, on the paternal chromosome, *Igf2* is active and *H19* is inactive. Key to imprinting is the methylation of specific DNA sequences and the inheritance of those methylated sequences. Some human genetic diseases such as the Prader-Willi and Angelman syndrome, result from effects involving imprinting.

14.5.2 Posttranscriptional Gene Silencing

Gene silencing can also occur post transcriptionally due to mRNA degradation and/or repression of its translation. These effects are often mediated by small RNA regulators such a small interfering RNAs (siRNAs) and micro RNAs (miRNAs) which are generated from different forms of double- stranded RNA (dsRNA) accumulating in cells.

RNA Silencing

The discovery that small RNA molecules control gene expression has given rise to a new field of research. First discovered in plants, short RNA molecules, ~21 nucleotides long, are now known to regulate gene expression in the cytoplasm of plants, animals, and fungi by repressing translation and triggering the degradation of mRNAs. This form of sequence-specific posttranscriptional regulation is known as **RNA interference (RNAi)**. As mentioned above, RNAi was first discovered in plants, where it was called posttranscriptional gene silencing (PTGS), although the involvement of RNA was not demonstrated till that time. In 1998 Andrew Fire and Craig Mello, working on gene regulation in the nematode worm, *Caenorhabditis elegans* showed the involvement of RNA molecule in this process. They injected mRNA molecules encoding a muscle protein into the worm and saw no difference in behavior. However, when both sense and antisense RNA were injected together ("sense RNA" is the mRNA with the coding information for the gene-encoded polypeptide, while "antisense RNA" is the complementary RNA to mRNA), the worms began moving in a manner like they contained a defective gene for the same muscle protein. Fire and Mello hypothesized that double-stranded RNA formed by pairing of the sense and antisense RNAs was responsible for silencing expression of the gene encoding the same RNA. They deduced that dsRNA could silence genes and named the phenomenon as RNA interference.

Mechanism of Gene Silencing by RNA Interference

Small interfering RNAs and micro RNAs regulate gene expression through four distinct mechanisms:1) cleavage of mRNA, 2) inhibition of translation, 3) transcriptional silencing, or 4) Slicer-independent degradation of mRNA.

1) RNA cleavage

RNA-induced silencing complex (RISCs) that contains a siRNA (and some that contain a miRNA) pairs with mRNA molecules and cleaves the mRNA near the middle of the bound siRNA (Fig. 14.16a). This cleavage is sometimes referred to as "**Slicer activity**." After cleavage, the mRNA is further degraded. Thus, the presence of siRNAs and miRNAs increase the rate of mRNA breakdown thereby decreasing the amount of protein produced.

2) Inhibition of translation

Some micro RNAs (miRNAs) regulate genes by inhibiting the translation of their complementary mRNAs (Fig. 14.16b). For example, an important gene in flower development in *Arabidopsis thaliana* is *APETALA2.* The expression of this gene is regulated by a miRNA that base pairs with nucleotides in the coding region of *APETALA2* mRNA and inhibits its translation.

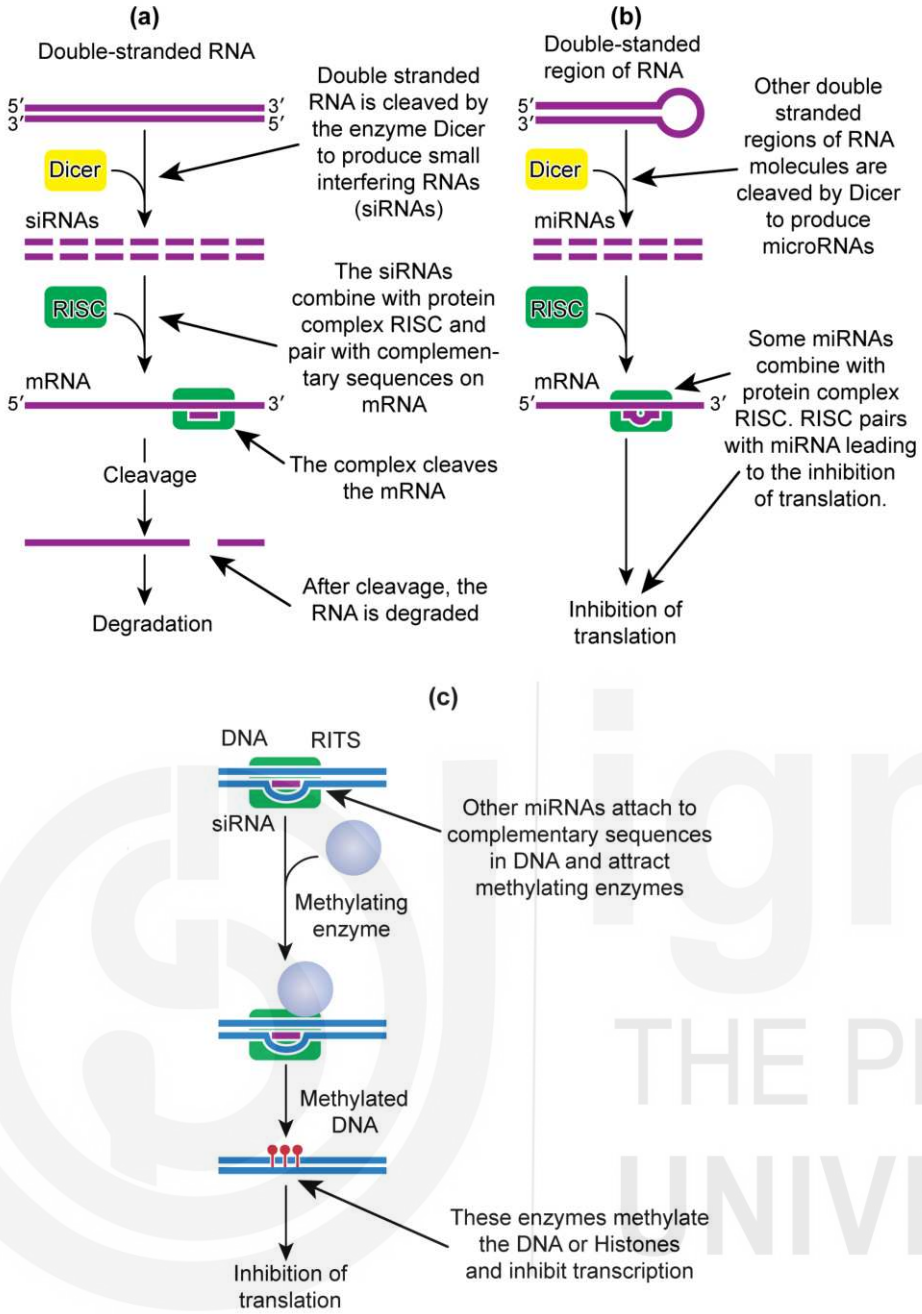


Fig. 14.16: RNA silencing by a) small interfering RNA (siRNAs); b) Micro RNAs (miRNAs);and c) small interfering RNA (siRNAs) that leads to degradation of mRNA or the inhibition of transcription or translation.

3) Transcriptional silencing

Some small interfering RNAs (siRNAs) silence transcription by altering chromatin structure. These siRNAs combine with proteins to form a complex called RNA-induced transcriptional silencing (RITS), which is analogous to RNA-induced transcriptional silencing complex (RISC). The siRNA component of RITS then binds to its complementary sequence in DNA or an RNA molecule in the process of being transcribed and represses transcription by attracting enzymes that methylate the tails of histone proteins. The addition of methyl groups to the histones causes them to bind DNA more tightly, restricting the access of proteins and enzymes necessary to carry out transcription. Some miRNAs bind to complementary sequences in DNA and attract enzymes that methylate the DNA directly, which also leads to the suppression of transcription as discussed in DNA methylation.

4) Slicer-independent degradation of mRNA

Another mechanism by which miRNAs regulate gene expression is by triggering the decay of mRNA in a process that does not requires slicer activity. A short-lived mRNA with an AU-rich element in its 3′ UTR is degraded by an RNA-silencing mechanism. A miRNA with a sequence that is complementary to the consensus sequence in the AU-rich element has also been identified. This miRNA binds to the AU-rich element, and brings about the degradation of the mRNA in a process that requires Dicer and RISC.

To conclude, in this unit we have tried to give you a brief overview of all the aspects that control the regulation of gene expression in both prokaryotes and eukaryotes and how gene silencing occurs in. You have also studied how gene silencing can regulate the expression of selected genes.

SAQ 3

Fill in the blanks with the suitable word from the words given below.

a) An epigenetic phenomenon in which the expression of certain genes is determined by whether the gene is inherited from the female or male parent is called …… .

b) Posttranscriptional gene silencing by small RNA molecules is called…… .

c) …….. helps in transcriptional silencing by altering chromatin structure with the formation of RNA-induced transcriptional silencing complex (RISC).

d) …… of nitrogenous bases in DNA silences genes and is often used to control gene expression during development in higher organism.

e) In RNA silencing, siRNA and miRNA usually binds to …… part of mRNA molecule that they control.

(Methylation, RNA interference, small interfering RNA, 3´ UTR, Genomic imprinting)

14.6 SUMMARY

- Gene expression can be controlled at different levels, including the alteration of DNA or chromatin structure, transcription, mRNA processing, RNA stability, translation, and posttranslational modification. Much of gene regulation is carried out through the action of regulatory proteins binding to specific sequences in DNA.
- Genes in bacterial cells are typically clustered into operons—groups of functionally related structural genes and the sequences that control their transcription. Structural genes in an operon are transcribed together as a single mRNA molecule.

- In negative control, a repressor protein binds to DNA and inhibits transcription. In positive control, an activator protein binds to DNA and stimulates transcription. In inducible operons, transcription is normally off and must be turned on; in repressible operons, transcription is normally on and must be turned off.
- Positive control in the *lac* and other operons is exercised through catabolite repression. When complexed with cyclic AMP, the catabolite activator protein binds to a site in or near the promoter and stimulates the transcription of the structural genes. Levels of cAMP are inversely correlated with glucose; so low levels of glucose stimulate transcription and high levels inhibit transcription.
- The *trp* operon of *E. coli* is a negative repressible operon that controls the biosynthesis of tryptophan.
- Attenuation causes premature termination of transcription. It takes place through the close coupling of transcription and translation, and depends on the secondary structure of the 5′UTR sequence. Antisense RNAs are complementary to sequences in mRNA and may inhibit translation by binding to these sequences, thereby preventing the attachment or progress of the ribosome.
- Eukaryotic cells differ in gene regulation from bacteria in several ways. A characteristic feature of eukaryotes is the presence of chromatin and nuclear membrane. However, operon is absent.
- In eukaryotic cells, chromatin structure represses gene expression. In transcription, chromatin structure may be altered by the modification of histone proteins, including acetylation, phosphorylation, and methylation. The repositioning of nucleosomes and the methylation of DNA also affect transcription.
- The initiation of eukaryotic transcription is controlled by general transcription factors that assemble into the basal transcription apparatus and by transcriptional activator proteins that stimulate normal levels of transcription by binding to regulatory promoters and enhancers.
- Enhancers affect the transcription of distant genes. Regulatory proteins bind to enhancers and interact with the basal transcription apparatus by causing the intervening DNA to loop out.
- DNA sequences called insulators limit the action of enhancers by blocking their action in a position-dependent manner.
- Coordinately controlled genes in eukaryotic cells respond to the same factors because they have common response elements that are stimulated by the same transcriptional activator.
- RNA molecules (antisense RNA) may act as regulators of gene expression in bacteria. Regulation by siRNAs and miRNAs, which is extensive in eukaryotes, is absent from bacterial cells.

- RNA silencing plays an important role in eukaryotic gene regulation. Small RNA molecules (siRNAs and miRNAs) cleaved from double-stranded DNA combine with proteins and bind to sequences on mRNA or DNA. These complexes cleave RNA, inhibit translation, affect RNA degradation, and silence transcription.

14.7 TERMINAL QUESTIONS

1. Explain the operon and steps involved in controlling the expression of structural genes.
2. Describe the three different processes of eukaryotic gene regulation by chromatin modifications.
3. Write the mechanisms involved in gene silencing by RNA interference.

14.8 ANSWERS

Self-Assessment Questions

1. a) Operon
   b) Regulator gene
   c) Operator gene
   d) Allolactose
   e) Repressor
2. a) True; b) True; c) True; d) True; e) False
3. a) Genomic imprinting
   b) RNA interference
   c) Small interfering RNA
   d) Methylation
   e) 3´ UTR

Terminal Questions

1. Refer to Subsection 14.3.3.
2. Refer to Subsection 14.4.1.
3. Refer to Subsection 14.5.2.

GLOSSARY

Acceptor arm : The arm in tRNA to which an amino acid attaches.

Acetylation of protein : Process in which an acetyl (CH_3CO) functional group is transferred from one molecule (Acetyl- coenzyme A for example) to another molecule. Histone acetylation is a common posttranslational modification.

Aminoacyl tRNA : A covalently linked combination of an amino acid and a tRNA molecule. Also referred to as charged tRNA.

Anticodon : Sequence of three nucleotides in transfer RNA that pairs with corresponding codon in messenger RNA in translation.

Antisense RNA : Small RNA molecule that base pairs with a complementary DNA or RNA sequence and affects its functioning.

Attenuator : A nucleotide sequence between the promoter and the structural gene of some bacterial operons that regulates the transit of *RNA polymerase*, reducing transcription of related structural genes.

Branch point : Adenine nucleotide in nuclear pre-mRNA introns that lies from 18 to 20 nucleotides upstream of the 3' splice site.

CAP : Catabolite activator proteins: proteins that bind cAMP and regulate the activation of inducible operons.

Catabolite repression : The selective inactivation of an operon by a metabolic product of an enzyme encoded by the operon.

Central dogma : Concept that genetic information passes from DNA to RNA to protein in a one way information pathway.

CFTR : The protein (cystic fibrosis transmembrane conductance regulator) encoded by CFTR gene. The protein regulate the movement of chloride ions across the plasma membrane. Muations in CFTR leads to cystic fibrosis.

Chaperone : a protein that regulates the folding of a polypeptide into a functional three dimenstional shape.

Chromatin remodeling : A process in which the structure of chromatin is altered by a protein complex, resulting in changes in transcriptional state of genes in the altered region.

Cloverleaf structure : Secondary structure common to all tRNAs.

Codon : Sequence of three nucleotides that encodes one amino acid in a protein.

Consensus sequence : Comprises the most commonly encountered nucleotide sequence found at a specific location in DNA or RNA.

-10 consensus sequence (Pribnow box) : Consensus sequence (TATAAT) found in most bacterial promoters approximately 10 base pairs upstream of the transcription start site.

-35 consensus sequence : Consensus sequence (TTGACA) found in many bacterial promoters approximately 35 base pairs up stream of the start site.

Core enzyme : Part of bacterial *RNA polymerase* that, during transcription, catalyzes the elongation of RNA molecule by the addition of RNA nucleotides; consists of four subunits: two copies of alpha (α), a single copy of beta (β) , and a single copy of beta prime (β').

Core promoter : Located immediately upstream of the eukaryotic promoter DNA sequence to which the basal transcription apparatus binds.

Constitutive genes : Genes whose products are essential to the normal functioning and for dividing cells irrespective of external conditions.

CpG Island : A short region of regulatory DNA found upstream of genes that contain unmethylated stretches of sequence with a high frequency of C and G nucleotides.

C terminus : In a polypeptide, the end that carries a free carboxyl group on the last amino acid.

Dicer : An enzyme (a *ribonuclease*) that cleaves double stranded RNA (dsRNA) and pre-micro RNA (miRNA) to form small interfering RNA (siRNA) molecules about 20-25 nucleotides long that serves as guide molecules for the degradation of mRNA molecule with sequences complementary to the siRNA.

Dolichol phosphate : A lipid carrier in the ER membrane in which N-linked glycosylation takes place.

Enhancer : A DNA sequence that enhances transcription and expression of structural genes. Enhancers can act at a distance of thousands of base pairs and can be located upstream, or internal to the gene they affect. They can stimulate almost any promoter in its vicinity and may be upstream or downstream of the gene.

Exon : Coding region of a split gene (a gene that is interrupted by introns). After processing introns remain in the mRNA.

Exon shuffling : Process by which exons of different genes are exchanged and mixed into new combinations; important in the evolution of eukaryotic genes.

5´ cap : Modified 5´ end of eukaryotic mRNA, consisting of an extra nucleotide (methylated) and methylation of the 2´ position of the ribose sugar in one or more subsequent nucleotide; plays a role in the binding of the ribosome to mRNA and affects mRNA stability and removal of introns.

Gene knockdown : The regulation of gene expression in which expression of one or more genes is reduced.

Gene silencing : The regulation of gene expression in a cell to prevent the expression of certain genes.

Group I intron : Belongs to a class of introns in some ribosomal RNA genes that are capable of self- splicing.

Group II intron : Belongs to a class of introns in some protein-encoding genes that are capable of self-splicing and are found in mitochondria, chloroplasts, and a few eubacteria.

Guide RNA (gRNA) : RNA molecule that serves as a template for an alteration made in mRNA during RNA editing

Glycosylation of proteins : The modification of many proteins by addition of carbohydrates particularly in eukaryotes.

Histone code : Various chemical modifications applied to histone tails (the free ends of histone molecules), projecting from nucleosomes. These modifications influence DNA-histone interactions and promote or repress transcription.

Inducer : An effector molecule that activates transcription.

Inducible enzyme system : An enzyme system under the control of an inducer, a regulatory molecule that acts to block a repressor and allow transcription.

Initiation codon : The nucleotide triplet AUG that in an mRNA molecule codes for incorporation of the amino acid methionine as the first amino acid in the polypeptide chain.Looplike structure created in the splicing of nuclear pre-mRNA in which the 5´ end of an intron is attaced to the branch point in pre-mRNA.

Isoschizomers : *Restriction endonucleases* that cleave the DNA at the same nucleotide sequence.

Kozak sequence : A short nucleotide sequence adjacent to the initiation codon that is recognized as the translational start site in eukaryotic mRNA. Molecule that carries genetic information for the amino acid sequence of a protein.

***lac* repressor protein** : A protein that binds to the operator in the *lac* operon and blocks transcription.

Lariat : Looplike structure created in the splicing of nuclear pre-mRNA in which the 5´ end of an intron is attached to the branch point in pre-mRNA.

Linking number : The number of times that one strand of a covalently closed circular double stranded DNA winds around the other. It cannot be changed without breaking the covalent bond.

Messenger RNA (mRNA) : RNA molecule that carries genetic information for the amino acid sequence of a protein.

Methylation : Enzymatic transfer of methyl groups from S-adenosylmethionine to biological molecules, including phospholipids, proteins, RNA, and DNA. Methylation of DNA is associated with regulation of gene expression and with epigenetic phenomena such as imprinting.

Methylation of proteins : Protein methylation is an important and widespread posttranslational modification affecting almost all basic cellular process in prokaryotes and eukaryotes.

microRNA (miRNA) : Small RNA molecule, typically 21 or 22 bp length, produced by cleavage of double-stranded RNA arising from small hairpins within RNA that is mostly single stranded. The miRNAs combine with proteins to form a complex that binds to mRNA molecules and inhibits their translation.

Monocistronic mRNAs : Messenger RNAs that encode single polypeptide chain.

Nitrosylation of proteins : Protein S-nitrosylation is a rapidly reversible and precisely targeted ubiquitous posttranslational modification in which nitric oxide (NO) is covalently attached to a thiol group of a cysteine residue (S-NO).

Nonsense codon : Codon in mRNA that signals the end of translation; also called a stop codon or termination codon. The three common nonsense codons are UAA, UAG, UGA.

Operon : A genetic unit consisting of one or more structural genes encoding polypeptides, and an adjacent operator gene that regulates the transcriptional activity of the structural gene or genes.

Operator region : A region in bacterial DNA that interacts with a specific repressor protein to regulate the expression of an adjacent genes or gene set.

Palindrome : A nucleotide sequence that reads the same forward or backward.

Palmitoylation : Fatty acid modification in which palmitic acid (a 16-carbon fatty acid) is added to the sulfur atom of the side chains of internal cysteine residues.

Peptide bond : The covalent bond between the amino group of one amino acid and and carboxyl group of another amino acid.

Polycistronic mRNAs : Messenger RNAs that encode multiple polypeptides.

Polysome : A structure composed of two or more ribosomes associated with an mRNA engaged in translation. Also called a polyribosome.

Posttranslational modification : The processing or modification of translated polypeptide chain by enzymatic cleavage, or the addition of phosphate groups, carbohydrate chains or lipids.

Prenylation : An important example of protein modification by lipids in which prenyl groups are added to the sulfur atoms in the side chains of cysteine residue.

Protein disulfide isomerase **(PDI)** : Catalyzes disufide bond formation between cysteine residues that is important in stabilizing the folded structure of many proteins.

Protein kinases : Enzyme that regulates the biological activity of proteins by phosphorylation of specfic amino acids by using ATP.

Protein phosphorylation : Transfer of phosphate group from ATP to hydroxyl groups of the side chains of serine, threonine, and tyrosine residues by protein kinases.

Proteolysis : Cleavage of polypeptide chain which is an important step in maturation of many proteins.

Reading frame : Particular way in which nucleotide sequence is read in a groups of three nucleotides (codon) in translation. Each reading frame begins with start codon and ends with stop codon.

Repressor : A protein that binds to a regulatory sequence adjacent to a gene and blocks transcription of a gene.

Repressible enzyme system : An enzyme or groups of enzymes whose synthesis is regulated by the intracellular concentration of certain metabolites.

Response elements : Short regulatory sequences that typically contain consensus sequences at varying distances from the gene being regulated. e.g. Heat-shock element.

Rho factor : Subunit of bacterial *RNA polymerase* that facilitates the termination of transcription of some genes.

rho-independent terminator : Sequence in bacterial DNA that does not require the presence of rho subunit of *RNA polymerase* to terminate transcription.

Ribonucleoside triphosphate (rNTP) : Substrate of RNA synthesis consists of ribose, a nitrogenous base, and three phosphates linked to 5´-carbon atom of the ribose. In transcription, two of the phosphates are cleaved, producing an RNA nucleotide.

Ribonucleotide : Ribose containing nucleotide present in RNA.

Ribosomal RNA (rRNA) : RNA molecule that is a structural component of the ribosome.

Ribozyme : RNA molecule that can act as a biological catalyst.

RNA-coding region : Sequence of DNA nucleotides that encodes an RNA molecule.

RNA editing : Process in which the protein-coding sequence of an messenger RNA is altered after transcription.

RNA-induced silencing complex(RISC) : Combination of a small interfering RNA (siRNA) molecule or a micro RNA (miRNA) molecule and proteins that can cleave messenger RNA, leading to the degradation of the messenger RNA, or affects transcription or repress translation of the messenger RNA.

RNA interference (RNAi) : Process in which cleavage of double-stranded RNA produces small interfering RNA(siRNA) that bind to mRNAs containing complementary sequences and bring about their cleavage and degradation.

RNA polymerase : Enzyme that synthesizes RNA from a DNA template during transcription.

RNA polymerase I : Eukaryotic *RNA polymerase* that transcribes large ribosomal RNA molecules (18 S rRNA and 28 S rRNA).

RNA polymerase II : Eukaryotic *RNA polymerase* that transcribes pre-messenger RNA, some small nuclear RNAs, and some microRNAs.

***RNA polymerase* III** : Eukaryotic *RNA polymerase* that transcribes transfer RNA, small ribosomal RNAs (5 S rRNA), some small nuclear RNAs, and some micro RNAs.

***RNA polymerase* IV** : Transcribes small interfering RNAs in plants.

***RNA polymerase* V** : Transcribes RNA that has role in heterochromatin formation in plants.

RNA silencing : Mechanism by which double stranded RNA is cleaved and processed to yield small single-stranded interfering RNA (siRNAs), which bind to complementary sequences in mRNA and bring about cleavage and degradation of mRNA; also known as RNA interference and posttranscriptional RNA gene silencing.

Shine-Delgarno sequence : Consensus sequence found in the bacterial 5´ untranslated region of mRNA; contains the ribosome binding site.

Sigma factor : Subunit of bacterial *RNA polymerase* that allows the *RNA polymerase* to recognize a promoter and initiate transcription.

Signal sequence : The modification of some proteins in which 15 – 30 amino acids (signal sequence) are removed at the amino end of the protein and helps direct a protein to a specific location in the cell.

Small cytoplasmic RNA (scRNA) : Small RNA molecule found in the cytoplasm of eukaryotic cells.

Small interfering RNA (siRNA) : Single stranded RNA molecule (usually from 21 to 25 nucleotides in length) produced by the cleavage and processing of double stranded RNA; binds to complementary sequence in mRNA and brings about cleavage and degradation of the mRNA.

Small nuclear ribonucleoprotein (snRNP) : Structure found in the nuclei of eukaryotic cells that consists of small nuclear RNA (snRNA) and protein; functions in the processing of pre-mRNA.

Small nuclear RNA (snRNA) : Small RNA molecule found in the nuclei of eukaryotic cells; functions in the processing of pre-mRNA.

Spliceosome : Large complex consisting of several RNAs and many proteins that splices protein-encoding pre-mRNA; contains five small ribonucleoprotein particles (U1, U2, U4, U5, and U6).

SR proteins : Group of serine- and argenine-rich proteins that regulate alternative splicing of pre-mRNA

Tm (melting temperature) : The midpoint temperature of the melting curve for the thermal denaturation of DNA/ macromolecule.

TFIIB recognition element (BRE) Consensus sequence [(G or C)(G or C)(G or C)CGCC] found in some *RNA polymerase II* promoters; usually located 32 to 38 nucleotides upstream of the transcription start site.

Twist (Tw): Number of complete revolutions that one strand of a covalently closed circular double stranded DNA makes around the duplex axis. It is positive for right handed helix and negative for left handed super helical coils.

3’ splice site : The 3’ end of an intron where cleavage takes place in RNA splicing.

3’ untranslated region (3’UTR) : Sequence of nucleotides at the 3’ end of mRNA; does not encode the amino acid of a protein but affects the stability of the mRNA and its translation.

Transcription	:	Process by which RNA synthesized from DNA template.
Transcription Bubble	:	Region of a DNA molecule that has unwound to expose a single- stranded template, which is being transcribed into RNA.
Transcription factor	:	Protein that binds to DNA sequences in eukaryotic cells and affect transcription.
Trascription start site	:	The first DNA nucleotide that is transcribed into an RNA molecule.
Transcription unit	:	Sequence of nucleotides in DNA that encodes a single RNA molecule, along with the sequence necessary for its transcription; normally contains a promoter, an RNA-coding sequence, and a terminator.
Transfer RNA (tRNA)	:	RNA molecule that carries an amino acid to the ribosome and transfer it to a growing polypeptide chain in translation.
Trans-splicing	:	The process of splicing together exons from two or more pre-mRNAs.
Triplet code	:	Refers to three nucleotides encode each amino acid in a protein.
Universal genetic code	:	Refers to the fact that particular codons specify the same amino acid in almost all organisms.
5´ untranslated (5´ UTR) region	:	Sequence of nucleotides at the 5´ end of mRNA; does not encode amino acid of a protein.
Upstream element	:	Consensus sequence found in some bacterial promoters that contains a number of A-T pairs and is found about 40 to 60 bp upstream of the transcription start site.
Wobble	:	Base pairing between codon and anticodon in which there is nonstandard pairing, usually at the third (3') position of the codon; allows more than one codon to pair with the same anticodon.