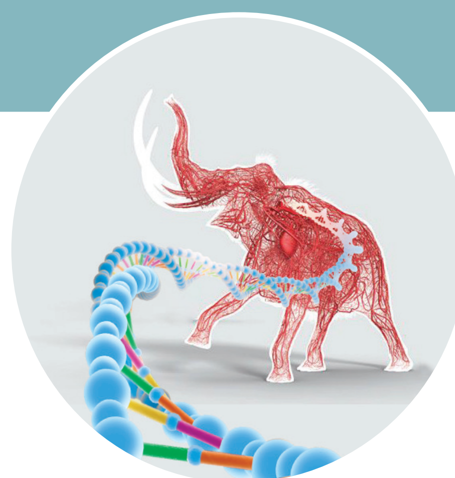
UNIT - II

5 CHAPTER

Molecular Genetics



New life for Woolly Mammoth DNA-researchers can now re-create the genes of mammoth and study the proteins they encoded

Chapter Outline

5.1 Gene as the functional unit of inheritance
5.2 In search of the genetic material
5.3 DNA is the genetic material
5.4 Chemistry of nucleic acids
5.5 RNA world
5.6 Properties of genetic material
5.7 Packaging of DNA helix
5.8 DNA Replication
5.9 Transcription
5.10 Genetic code
5.11 tRNA – the adapter molecule
5.12 Translation
5.13 Regulation of Gene expression
5.14 Human Genome Project (HGP)
5.15 DNA finger printing technique

Learning Objectives

- *Identifies DNA as the genetic material.*
- *Understands the organization of prokaryotic and eukaryotic genome.*
- *Learns to differentiate the nucleotides of DNA and RNA.*
- *Understands gene expression - Replication, Transcription and Translation.*
- *Learns about codons and the salient features of genetic code.*
- *Understands the gene regulation through Lac operon model.*
- *Realizes the importance of Human Genome Project.*
- *Illustrates the applications of DNA finger printing technique.*

Mendel's theory dispelled the mystery of why traits seemed to appear and disappear magically from one generation to the next. Mendel's work reveals the patterns of heredity and reflect the transmission of evolved information from parents to offspring. This information is located on the chromosomes. One of the most advanced realizations of human knowledge was that our unique characteristics are encoded within molecules of DNA. The discovery that DNA is the genetic material left several questions unanswered. How is the information in DNA used? Scientists now know that DNA directs the construction of proteins. Proteins determine the shapes of cells and the rate of chemical reactions, such as those that occur during metabolism and photosynthesis. The hereditary nature of every living organism is defined by its genome, which consists of a long sequence of nucleic acids that provide the information needed to construct the organism. The genome contains the complete set of hereditary information for any organism. The genome may be divided into a number of different nucleic acid molecules. Each of the nucleic acid molecule may contain large number of genes. Each gene is a sequence within the nucleic acid that represents a single protein. In this chapter we will discuss the

structure of DNA, its replication, the process of making RNA from DNA (transcription), the genetic code that determines the sequence of amino acid in protein synthesis (translation), regulation of gene expression and the essentials of human genome sequencing.

5.1 Gene as the functional unit of inheritance

A gene is a basic physical and functional unit of heredity. The concept of the gene was first explained by **Gregor Mendel in 1860's**. He never used the term 'gene'. He called it 'factor'. In 1909, the Danish biologist Wilhelm Johannsen, coined the term 'gene', that was referred to discrete determiners of inherited characteristics.

According to the classical concept of gene introduced by Sutton in 1902, genes have been defined as discrete particles that follow Mendelian rules of inheritance, occupy a definite locus in the chromosome and are responsible for the expression of specific phenotypic character. They show the following properties:

- Number of genes in each organism is more than the number of chromosomes; hence several genes are located on the same chromosome.
- The genes are arranged in a single linear order like beads on a string.
- Each gene occupies a specific position called locus.
- Genes may exist in several alternate forms called alleles.
- Genes may undergo sudden change in positions and composition called mutations.
- Genes are capable of self-duplication producing their own copies.

One gene-one enzyme hypothesis

The experiments of **George Beadle and Edward Tatum** in the early 1940's on *Neurospora crassa* (the red bread mould) led them to propose one gene-one enzyme hypothesis, which states that one gene controls the production of one enzyme.

One gene-one polypeptide hypothesis

It was observed that an enzyme may be composed of more than one polypeptide chain and a gene can code for only one polypeptide chain. Thus one gene-one polypeptide hypothesis states that one gene controls the production of only one polypeptide chain of an enzyme molecule.

5.2 In search of the genetic material

As early as 1848, Wilhelm Hofmeister, a German botanist, had observed that cell nuclei organize themselves into small, rod like bodies during mitosis called **chromosomes**. In 1869, Friedrich Miescher, a Swiss physician, isolated a substance from the cell nuclei and called it as **nuclein**. It was renamed as nucleic acid by Altman (1889), and is now known as DNA. By 1920, it became clear that chromosomes are made up of proteins and DNA. Many experiments were carried out to study the actual carriers of genetic information. Griffith's experiment proved that DNA is the genetic material which has been dealt in class XI. Bacterial transformation experiments provided the first proof that DNA is the genetic material. However, he could not understand the cause of bacterial transformation, and the biochemical nature of genetic material was not defined from his experiments.

Later, Oswald Avery, Colin Macleod and Maclyn McCarty in 1944 repeated Griffith's experiments in an '*in vitro*' system in order

TN_GOVT_XII_Zoology_chapter5.indd 69 25-02-2019 18:23:04

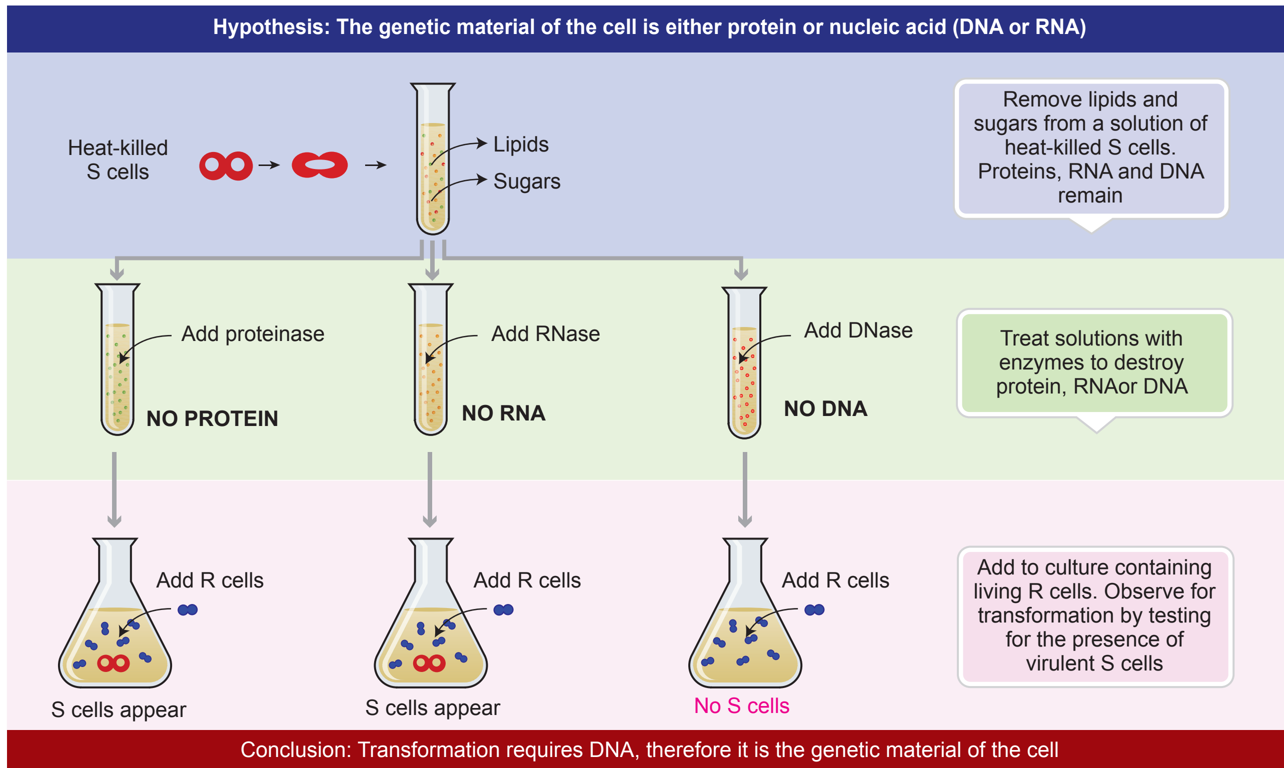
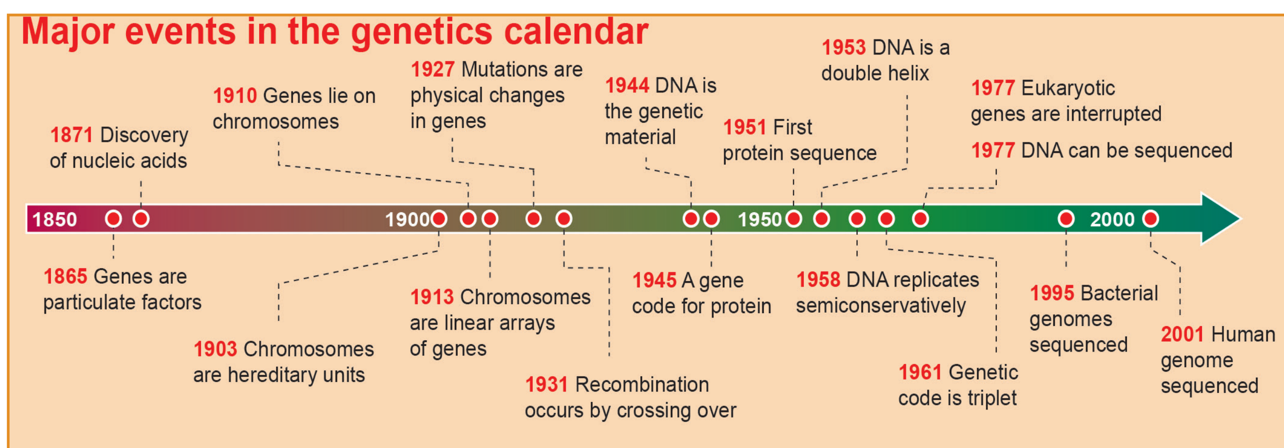


Fig. 5.1 Transformation experiment of Avery *et. al.,* (1944)

to identify the nature of the transforming substance responsible for converting a non-virulent strain into virulent strain. They observed that the DNA, RNA and proteins isolated from the heat-killed S-strain when added to R-strain changed their surface character from rough to smooth and also made them pathogenic **(Fig. 5.1)**. But when the extract was treated with **DNase** (an enzyme which destroys DNA) the transforming ability was lost. **RNase** (an enzyme which destroys RNA) and **proteases** (an enzyme which destroys protein) did not affect the transformation. Digestion with DNase inhibited transformation suggesting that the DNA caused the transformation. These experiments suggested that DNA and not proteins is the genetic material. The phenomenon, by which DNA isolated from one type of cell (R – strain), when introduced into another type (S-strain), is able to retain some of the properties of the R - strain is referred to as **transformation**.



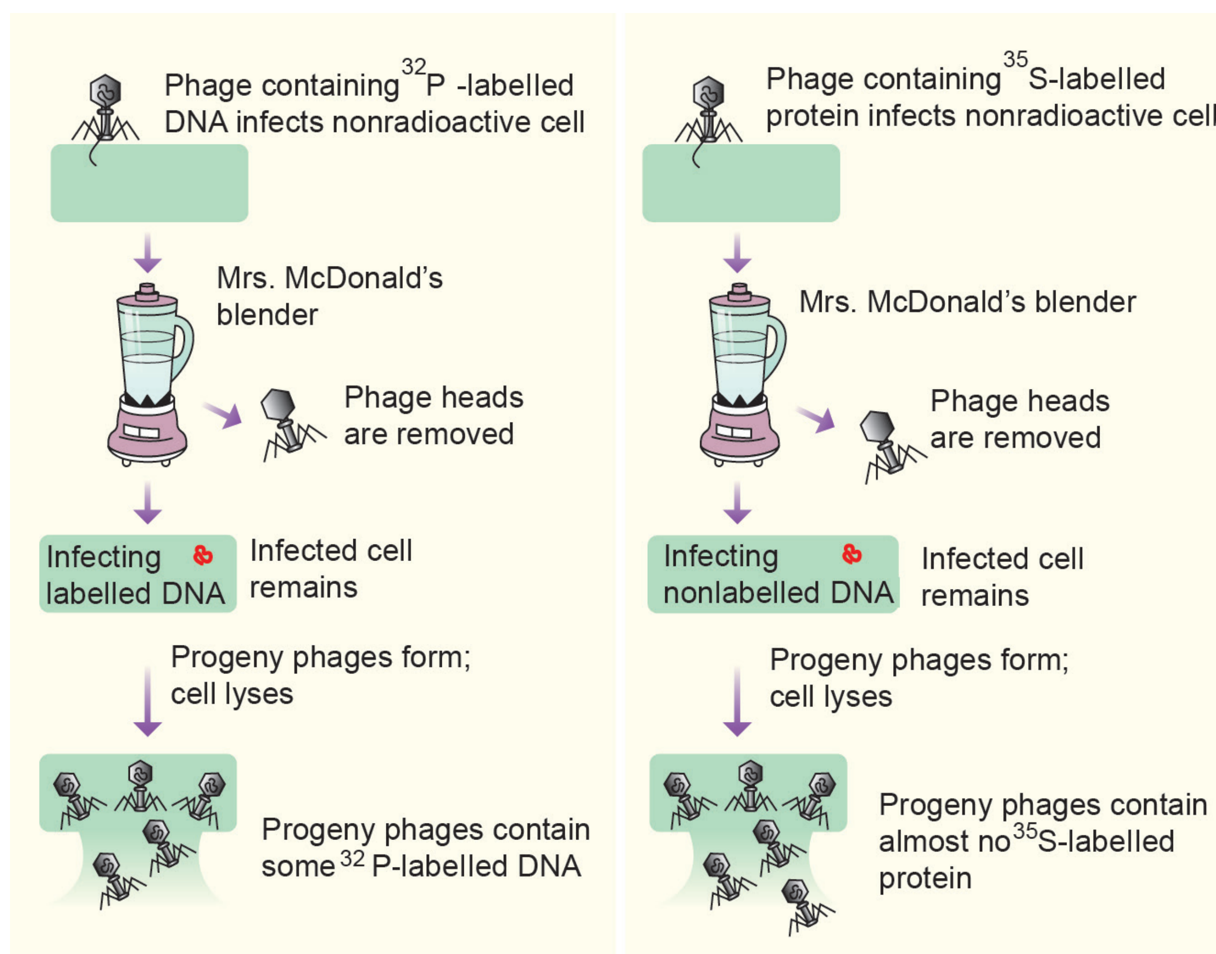


Fig. 5.2 The Hershey-Chase (blender) experiment

5.3 DNA is the genetic material

Many biologists despite the earlier experiments of Griffith, Avery and others, still believed that protein, not DNA, was the hereditary material in a cell. As eukaryotic chromosomes consist of roughly equal amounts of protein and DNA, it was said that only a protein had sufficient chemical diversity and complexity to encode the information required for genetic material. In 1952, however, the results of the Hershey-Chase experiment finally provided convincing evidence that DNA is the genetic material.

5.3.1 Hershey and Chase experiment on T_2 bacteriophage

Alfred Hershey and **Martha Chase** (1952) conducted experiments on bacteriophages that infect bacteria. Phage T_2 is a virus that infects the bacterium *Escherichia coli*. When phages (virus) are added to bacteria, they adsorb to the outer surface, some material enters the bacterium, and then later each bacterium lyses to release a large number of progeny phage. Hershey and Chase wanted to observe whether it was DNA or protein that entered the bacteria. All nucleic acids contain phosphorus, and contain sulphur (in the amino acid cysteine and methionine). Hershey and Chase designed an experiment using radioactive isotopes of Sulphur (^{35}S) and phosphorus (^{32}P) to keep separate track of the viral protein and nucleic acids during the infection process. The phages were allowed to infect bacteria in culture medium which containing the radioactive isotopes ^{35}S or ^{32}P. The bacteriophage that grew in the presence of ^{35}S had labelled proteins and bacteriophages grown in the presence of ^{32}P had labelled DNA.

TN_GOVT_XII_Zoology_chapter5.indd 71 25-02-2019 18:23:04

The differential labelling thus enabled them to identify DNA and proteins of the phage.

Hershey and Chalse mixed the labelled phages with unlabeled *E. coli* and allowed bacteriophages to attack and inject their genetic material. Soon after infection (before lysis of bacteria), the bacterial cells were gently agitated in a blender to loosen the adhering phase particles. It was observed that only ^{32}P was found associated with bacterial cells and ^{35}S was in the surrounding medium and not in the bacterial cells. When phage progeny was studied for radioactivity, it was found that it carried only ^{32}P and not ^{35}S **(Fig. 5.2)**. These results clearly indicate that only DNA and not protein coat entered the bacterial cells. Hershey and Chase thus conclusively proved that it was DNA, not protein, which carries the hereditary information from virus to bacteria.

5.4 Chemistry of Nucleic Acids

Having identified the genetic material as the nucleic acid DNA (or RNA), we proceed to examine the chemical structure of these molecules. Generally nucleic acids are a long chain or polymer of repeating subunits called nucleotides. Each nucleotide subunit is composed of three parts: a nitrogenous base, a five carbon sugar (pentose) and a phosphate group.

Pentose sugar

There are two types of nucleic acids depending on the type of pentose sugar. Those containing deoxyribose sugar are called **Deoxyribo Nucleic Acid** (DNA) and those with ribose sugar are known as **Ribonucleic Acid** (RNA). DNA is found in the nucleus of eukaryotes and nucleoid of prokaryotes. The only difference between these two sugars is that there is one oxygen atom less in deoxyribose.

Nitrogenous bases

The bases are nitrogen containing molecules having the chemical properties of a base (a substance that accepts H^+ ion or proton in solution). DNA and RNA both have four bases (two purines and two pyrimidines) in their nucleotide chain. Two of the bases, Adenine (A) and Guanine (G) have double carbon–nitrogen ring structures and are called purines. The bases, Thymine (T), Cytosine (C) and Uracil (U) have single ring structure and these are called pyrimidines. Thymine is unique for DNA, while Uracil is unique for RNA.

The phosphate functional group

It is derived from phosphoric acid (H_3PO_4), has three active OH^- groups of which two are involved in strand formation. The phosphate functional group (PO_4) gives DNA and RNA the property of an acid (a substance that releases an H^+ ion or proton in solution) at physiological pH, hence the name **nucleic acid**. The bonds that are formed from phosphates are esters. The oxygen atom of the phosphate group is negatively charged after the formation of the phosphodiester bonds. This negatively charged phosphate ensures the retention of nucleic acid within the cell or nuclear membrane.

Nucleoside and nucleotide

The nitrogenous base is chemically linked to one molecule of sugar (at the 1-carbon of the sugar) forming a **nucleoside**. When a phosphate group is attached to the 5' carbon of the same sugar, the nucleoside becomes a **nucleotide**. The nucleotides are joined (polymerized) by condensation reaction to form a polynucleotide chain. The hydroxyl group on the 3' carbon of a sugar of one nucleotide forms an ester with the phosphate of another nucleotide. The chemical bonds that link the sugar

components of adjacent nucleotides are called **phosphodiester bond** (5'⟶3'), indicating the polarity of the strand.

The ends of the DNA or RNA are distinct. The two ends are designated by the symbols 5' and 3'. The symbol 5' refers to carbon in the sugar to which a phosphate (PO_4) functional group is attached. The symbol 3' refers to carbon in the sugar to which hydroxyl (OH) functional group is attached. In RNA, every nucleotide residue has an additional $^{-}$OH group at 2' position in the ribose. Understanding the 5'⟶3' direction of a nucleic acid is critical for understanding the aspects of replication and transcription.

Based on the X - ray diffraction analysis of **Maurice Wilkins** and **Rosalind Franklin**, the double helix model for DNA was proposed by **James Watson** and **Francis Crick in 1953**. The highlight was the base pairing between the two strands of the polynucleotide chain. This proposition was based on the observations of Erwin Chargaff that Adenine pairs with Thymine (A = T) with two hydrogen bonds and Guanine pairs with Cytosine (G ≡ C) with three hydrogen bonds. The ratios between Adenine with Thymine and Guanine with Cytosine are constant and equal. The base pairing confers a unique property to the polynucleotide chain. They are said to be complementary to each other, that is, if the sequence of bases in one strand (template) is known, then the sequence in the other strand can be predicted. The salient features of DNA structure has already been dealt in class XI.

5.5 RNA world

A typical cell contains about ten times as much RNA as DNA. The high RNA content is mainly due to the variety of roles played by RNA in the cell. **Fraenkel-Conrat** and **Singer** (1957) first demonstrated that **RNA** is the genetic material in RNA containing viruses like TMV (Tobacco Mosaic Virus) and they separated RNA from the protein of TMV viruses. Three molecular biologists in the early 1980's (Leslie Orgel, Francis Brick and Carl Woese) independently proposed the '**RNA world**' as the first stage in the evolution of life, a stage when RNA catalysed all molecules necessary for survival and replication. The term 'RNA world' first used by **Walter Gilbert** in 1986, hypothesizes RNA as the first genetic material on earth. There is now enough evidence to suggest that essential life processes (such as metabolism, translation, splicing etc.,) evolved around RNA. RNA has the ability to act as both genetic material and catalyst. There are several biochemical reactions in living systems that are catalysed by RNA. This catalytic RNA is known as **ribozyme.** But, RNA being a catalyst was reactive and hence unstable. This led to evolution of a more stable form of DNA, with certain chemical modifications. Since DNA is a double stranded molecule having complementary strand, it has resisted changes by evolving a process of repair. Some RNA molecules function as gene regulators by binding to DNA and affect gene expression. Some viruses use RNA as the genetic material. Andrew Fire and Craig Mellow (recipients of Nobel Prize in 2006) were of the opinion that RNA is an active ingredient in the chemistry of life. The types of RNA and their role have been discussed in class XI.

5.6 Properties of genetic material (DNA versus RNA)

The experiment by Hershey and Chase clearly indicates that it is DNA that acts as a genetic material. However, in some viruses like Tobacco mosaic virus (TMV), bacteriophage θB, RNA acts as the genetic material. A molecule that can act as a genetic material should have the following properties:

- **Self Replication:** It should be able to replicate. According to the rule of base pairing and complementarity, both nucleic

acids (DNA and RNA) have the ability to direct duplications. Proteins fail to fulfill this criteria.

- **Stability:** It should be stable structurally and chemically. The genetic material should be stable enough not to change with different stages of life cycle, age or with change in physiology of the organism. Stability as one of property of genetic material was clearly evident in Griffith's transforming principle. Heat which killed the bacteria did not destroy some of the properties of genetic material. In DNA the two strands being complementary, if separated (denatured) by heating can come together (renaturation) when appropriate condition is provided. Further 2' OH group present at every nucleotide in RNA is a reactive group that makes RNA liable and easily degradable. RNA is also known to be catalytic and reactive. Hence, DNA is chemically more stable and chemically less reactive when compared to RNA. Presence of thymine instead of uracil in DNA confers additional stability to DNA.
- **Information storage:** It should be able to express itself in the form of '**Mendelian characters**'. RNA can directly code for protein synthesis and can easily express the characters. DNA, however depends on RNA for synthesis of proteins. Both DNA and RNA can act as a genetic material, but DNA being more stable stores the genetic information and RNA transfers the genetic information.
- **Variation through mutation:** It should be able to mutate. Both DNA and RNA are able to mutate. RNA being unstable, mutates at a faster rate. Thus viruses having RNA genome with shorter life span can mutate and evolve faster.

The above discussion indicates that both RNA and DNA can function as a genetic material. DNA is more stable, and is preferred for storage of genetic information.

5.7 Packaging of DNA helix

The distance between two consecutive base pairs is 0.34nm (0.34×10^{-9}m) of the DNA double helix in a typical mammalian cell. When the total number of base pairs is multiplied with the distance between two consecutive base pairs ($6.6 \times 10^{-9} \times 0.34 \times 10^{-9}$ m/bp), the length of DNA double helix is

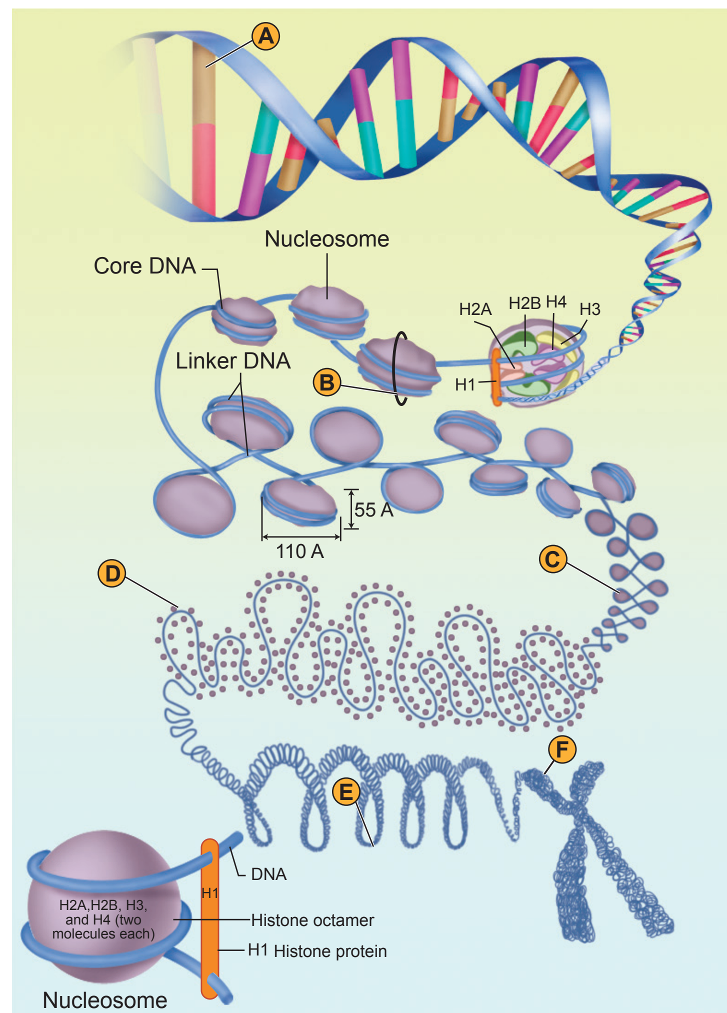


Fig. 5.3 Condensation of DNA - A - DNA, B-Nucleosomes and Histones, C- Chromatin fiber, D- Coiled chromatin fiber, E- Coiled coil, F- metaphase chromatid

TN_GOVT_XII_Zoology_chapter5.indd 74 25-02-2019 18:23:05

approximately 2.2 m. (The total length of the double helical DNA = total number of base pairs × distance between two consecutive base pairs). If the length of *E. coli* DNA is 1.36 mm, the number of base pairs in *E. coli* is 4×10^6m (1.36×10^3 m/0.34×10^{-9}). The length of the DNA double helix is far greater than the dimension of a typical mammalian nucleus (approximately 10^{-6} m). How is such a long DNA polymer packaged in a cell?

Chromosomes are carriers of genes which are responsible for various characters from generation to generation. Du Praw (1965) proposed a single stranded model (unineme), as a long coiled molecule which is associated with histone proteins in eukaryotes. Plants and animals have more DNA than bacteria and must fold this DNA to fit into the cell nucleus. In prokaryotes such as *E. coli* though they do not have defined nucleus, the DNA is not scattered throughout the cell. DNA (being negatively charged) is held with some proteins (that have positive charges) in a region called the nucleoid. The DNA as a nucleoid is organized into large loops held by protein. DNA of prokaryotes is almost circular and lacks chromatin organization, hence termed **genophore**.

In eukaryotes, this organization is much more complex. Chromatin is formed by a series of repeating units called **nucleosomes**. Kornberg proposed a model for the nucleosome, in which 2 molecules of the four histone proteins H2A, H2B, H3 and H4 are organized to form a unit of eight molecules called **histone octamere**. The negatively charged DNA is wrapped around the positively charged histone octamere to form a structure called **nucleosome**. A typical nucleosome contains 200 bp of DNA helix. The histone octameres are in close contact and DNA is coiled on the outside of nucleosome. Neighbouring nucleosomes are connected by linker DNA (H1) that is exposed to enzymes. The DNA makes two complete turns around the histone octameres and the two turns are sealed off by an H1 molecule. Chromatin lacking H1 has a **beads-on-a-string** appearance in which DNA enters and leaves the nucleosomes at random places. H1 of one nucleosome can interact with H1 of the neighbouring nucleosomes resulting in the further folding of the fibre. The chromatin fiber in interphase nuclei and mitotic chromosomes have a diameter that vary between 200-300 nm and represents inactive chromatin. 30 nm fibre arises from the folding of nucleosome, chains into a **solenoid** structure having six nucleosomes per turn. This structure is stabilized by interaction between different H1 molecules. DNA is a solenoid and packed about 40 folds. The hierarchical nature of chromosome structure is illustrated in **(Fig. 5.3)**. Additional set of proteins are required for packing of chromatin at higher level and are referred to as non-histone chromosomal proteins (NHC). In a typical nucleus, some regions of chromatin are loosely packed (lightly stained) and are referred to as euchromatin. The chromatin that is tightly packed (stained darkly) is called heterochromatin. Euchromatin is transcriptionally active and heterochromatin is transcriptionally inactive.

5.8 DNA Replication

Replication of DNA takes place during the S phase of cell cycle. During replication, each DNA molecule gives rise to two DNA strands, identical to each other as well as to the parent strand. Three hypotheses of DNA replication have been proposed. They are conservative replication, dispersive replication, and semi-conservative replication.

In conservative replication, the original double helix serves as a template. The original molecule is preserved intact and an entirely new double stranded molecule is synthesized. In dispersive replication, the original molecule is broken into fragments and each fragment

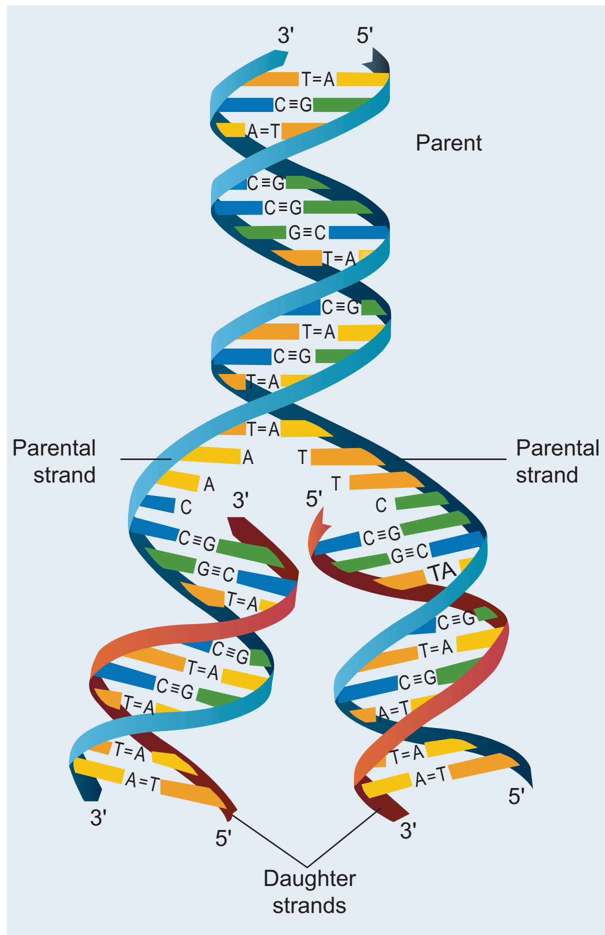


Fig. 5.4 Semiconservative DNA replication

serves as a template for the synthesis of complementary fragments. Finally two new molecules are formed which consist of both old and new fragments.

Semi-conservative replication was proposed by Watson and Crick in 1953. This mechanism of replication is based on the DNA model. They suggested that the two polynucleotide strands of DNA molecule unwind and start separating at one end. During this process, covalent hydrogen bonds are broken. The separated single strand then acts as template for the synthesis of a new strand. Subsequently, each daughter double helix carries one polynucleotide strand from the parent molecule that acts as a template and the other strand is newly synthesised and complementary to the parent strand **(Fig. 5.4)**.

5.8.1 Experimental proof of DNA replication

The mode of DNA replication was determined in 1958 by Meselson and Stahl. They designed an experiment to distinguish between semi conservative, conservative and dispersive replications. In their experiment, they grew two cultures of *E.coli* for many generations in separate media. The 'heavy' culture was grown in a medium in which the nitrogen source (NH_4Cl) contained the heavy isotope ^{15}N and the 'light' culture was grown in a medium in which the nitrogen

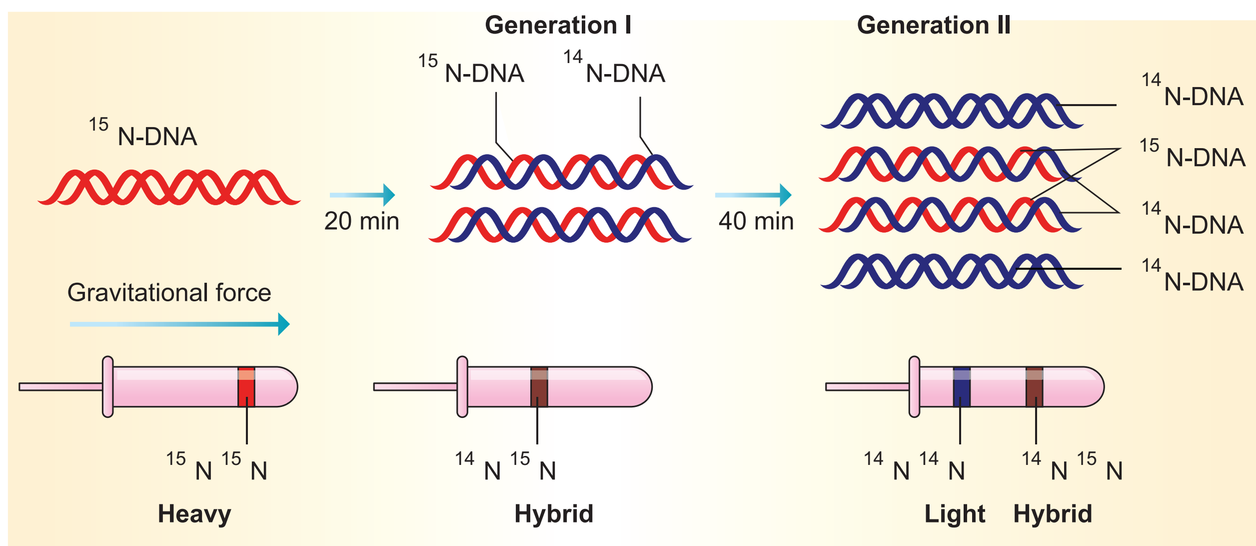


Fig. 5.5 Meselson and Stahl experiment to support semiconservative mode of DNAreplication

source contained light isotope ^{14}N for many generations. At the end of growth, they observed that the bacterial DNA in the heavy culture contained only ^{15}N and in the light culture only ^{14}N. The heavy DNA could be distinguished from light DNA (^{15}N from ^{14}N) with a technique called **Cesium Chloride (CsCl) density gradient centrifugation**. In this process, heavy and light DNA extracted from cells in the two cultures settled into two distinct and separate bands (hybrid DNA) (**Fig. 5.5**).

The heavy culture (^{15}N) was then transferred into a medium that had only NH_4Cl, and took samples at various definite time intervals (20 minutes duration). After the first replication, they extracted DNA and subjected it to density gradient centrifugation. The DNA settled into a band that was intermediate in position between the previously determined heavy and light bands. After the second replication (40 minutes duration), they again extracted DNA samples, and this time found the DNA settling into two bands, one at the light band position and one at intermediate position. These results confirm Watson and Crick's semi conservative replication hypothesis.

5.8.2 Enzymes and mechanism of replication

In prokaryotes, replication process requires three types of DNA polymerases (DNA polymerase I, II, and III). DNA polymerase III is the main enzyme involved in DNA replication. DNA polymerase I (also known as **Kornberg enzyme**) and DNA polymerase II are involved in DNA repair mechanism. Eukaryotes have five types of DNA polymerases that catalyses the polymerization of nucleotides at the 3' OH of the new strand within a short period of time. *E.coli* that has 4.6 X 10^6 bp completes its replication process within 38 minutes. Replication takes place faster at the same time accurately. Any error will lead to mutation. However replication errors are corrected by repair enzymes such as nucleases. Deoxy nucleotide triphosphate acts as substrate and also provides energy for polymerization reaction.

Replication begins at the initiation site called the site of '**origin of replication' (ori)**. In prokaryotes, there is only one origin of replication, whereas in eukaryotes with giant DNA molecules, there can be several origins of replication (replicons). Since the two strands of DNA cannot be separated throughout at a time (due to large requirement of energy) the replication occurs within a small opening of the DNA helix called as replication fork. Unwinding of the DNA strand is carried out by DNA helicase. Thus, in one strand (template strand with polarity 3'⟶5') the replication is continuous and is known as the **leading strand** while in the other strand (coding strand with polarity 5'⟶3') replication is discontinuous, known as the **lagging strand (Fig. 5.6)**. The discontinuously synthesized fragments of the lagging strand (called the **Okazaki fragments**) are joined by the enzyme DNA ligase.

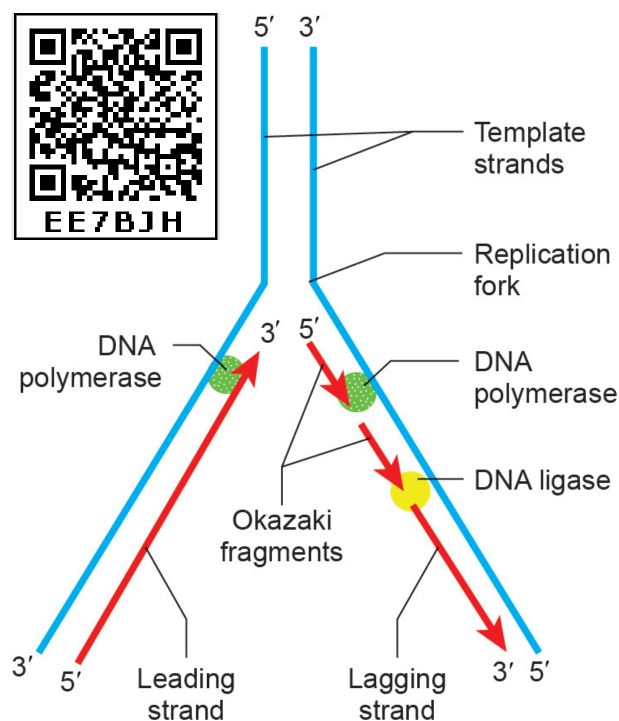


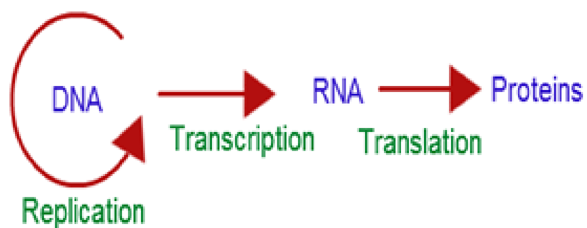
Fig 5.6 Mechanism of replication showing a replication fork

As they move away in both directions, newly synthesized complementary nucleotides are paired with the existing nucleotides on the parent strand and covalently bonded together by **DNA polymerase**. Formation of new strand requires a primer (a short stretch of RNA)for initiation. The primer produces a 3'-OH end on the sequence of ribonucleotides, to which deoxy ribonucleotides are added. The RNA primer is ultimately removed leaving a gap in the newly synthesized DNA strand. It is removed from 5' end one by one by the exonuclease activity of DNA polymerase. Finally, when all the nucleotides are in position, gaps are sealed by the enzyme **DNA ligase**.

At the point of origin of repliction, the helicases and topoisomerases (DNA gyrase) unwind and pull apart the strands, forming a Y-Shaped structure called the **replication fork**. There are two replication forks at each origin. The two strands of a DNA helix have an antiparallel orientation. The enzyme DNA polymerase can only catalyse the addition of a nucleotide to the new strands in the 5'⟶3' direction, as it can only add nucleotides to the 3' carbon position.

5.9 Transcription

Francis Crick proposed the **Central dogma** in molecular biology which states that genetic information flows as follows:



The process of copying genetic information from one strand of DNA into RNA is termed **transcription**. This process takes place in presence of DNA dependent RNA polymerase.

In some retroviruses that contain RNA as the genetic material (e.g, HIV), the flow of information is reversed. RNA synthesizes DNA by reverse transcription, then transcribed into mRNA by transcription and then into proteins by translation.

For a cell to operate, its genes must be expressed. This means that the gene products, whether proteins or RNA molecules must be made. The RNA that carries genetic information encoding a protein from genes into the cell is known as messenger RNA (mRNA). For a gene to be transcribed, the DNA which is a double helix must be pulled apart temporarily, and RNA is synthesized by RNA polymerase. This enzyme binds to DNA at the start of a gene and opens the double helix. Finally, RNA molecule is synthesized. The nucleotide sequence in the RNA is complementary to the DNA template strand from which it is synthesized.

Both the strands of DNA are not copied during transcription for two reasons. 1. If both the strands act as a template, they would code for RNA with different sequences. This in turn would code for proteins with different amino acid sequences. This would result in one segment of DNA coding for two different proteins, hence complicate the genetic information transfer machinery. 2. If two RNA molecules were produced simultaneously, double stranded RNA complementary to each other would be formed. This would prevent RNA from being translated into proteins.

5.9.1 Transcription unit and gene

A transcriptional unit in DNA is defined by three regions, a **promoter**, the **structural gene** and a **terminator**. The promoter is located towards the 5' end. It is a DNA sequence that provides binding site for RNA polymerase. The presence of promoter in a transcription unit, defines the template and coding strands. The terminator region located towards the 3' end of the coding strand contains a DNA sequence that causes the RNA polymerase to stop transcribing. In eukaryotes the promoter has AT rich regions called **TATA box (Goldberg-**

TN_GOVT_XII_Zoology_chapter5.indd 78

25-02-2019 18:23:05

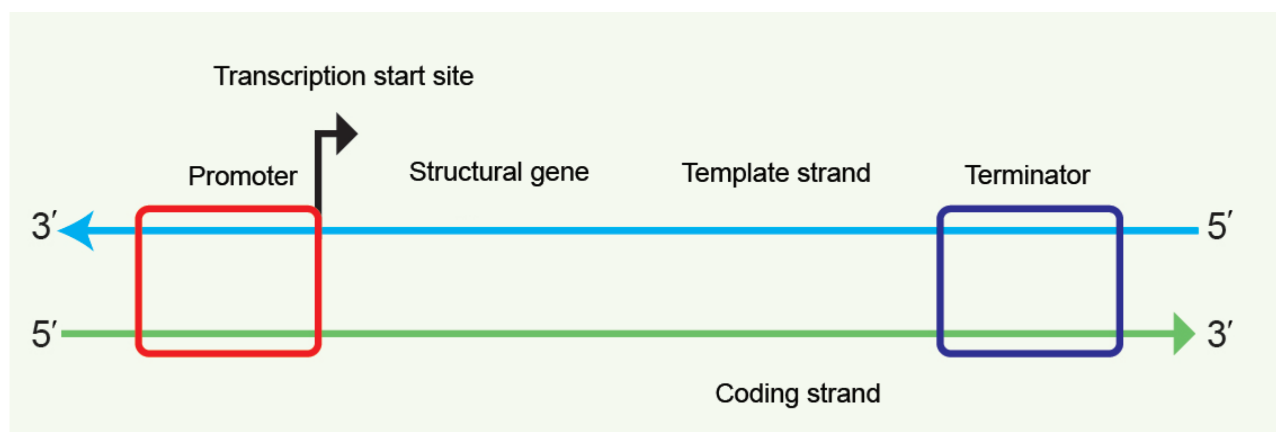


Fig. 5. 7 Schematic structure of a transcription unit

Hogness box) and in prokaryotes this region is called **Pribnow box**. Besides promoter, eukaryotes also require an enhancer.

The two strands of the DNA in the structural gene of a transcription unit have opposite polarity. DNA dependent RNA polymerase catalyses the polymerization in only one direction, the strand that has the polarity 3'⟶5' acts as a template, and is called the **template strand**. The other strand which has the polarity 5'⟶3' has a sequence same as RNA (except thymine instead of uracil) and is displaced during transcription. This strand is called **coding strand** (**Fig. 5.7**).

The structural gene may be **monocistronic** (eukaryotes) or **polycistronic** (prokaryotes). In eukaryotes, each mRNA carries only a single gene and encodes information for only a single protein and is called monocistronic mRNA. In prokaryotes, clusters of related genes, known as operon, often found next to each other on the chromosome are transcribed together to give a single mRNA and hence are polycistronic.

Before starting transcription, RNA polymerase binds to the promoter, a recognition sequence in front of the gene. Bacterial (prokaryotic) RNA polymerase consists of two major components, the core enzyme and the sigma subunit. The core enzyme (β^1, β, and α) is responsible for RNA synthesis whereas a sigma subunit is responsible for recognition of the promoter. Promoter sequences vary in different organisms. RNA polymerase opens up the DNA to form the transcription bubble. The core enzyme moves ahead, manufacturing RNA leaving the sigma subunit behind at the promoter region. The end of a gene is marked by a terminator sequence that forms a hair pin structure in the RNA. The sub-class of terminators require a recognition protein, known as rho (ρ), to function.

5.9.2 Process of transcription

In prokaryotes, there are three major types of RNAs: mRNA, tRNA, and rRNA. All three RNAs are needed to synthesize a protein in a cell. The mRNA provides the template, tRNA brings amino acids and reads the genetic code, and rRNAs play structural and catalytic role during translation. There is a single DNA-dependent RNA polymerase that catalyses transcription of all types of RNA. It binds to the promoter and initiates transcription (Initiation). The polymerases binding sites are called promoters. It uses nucleoside triphosphate as substrate and polymerases in a template depended fashion following the rule of complementarity. After the initiation of transcription, the polymerase continues to elongate the RNA, adding one nucleotide after another to the growing RNA chain. Only a short stretch of RNA remains bound to the enzyme, when the polymerase reaches a terminator at the end of a gene, the nascent RNA falls off, so also the RNA polymerase.

The question is, how the RNA polymerases are able to catalyse the three steps initiation, elongation and termination? The RNA

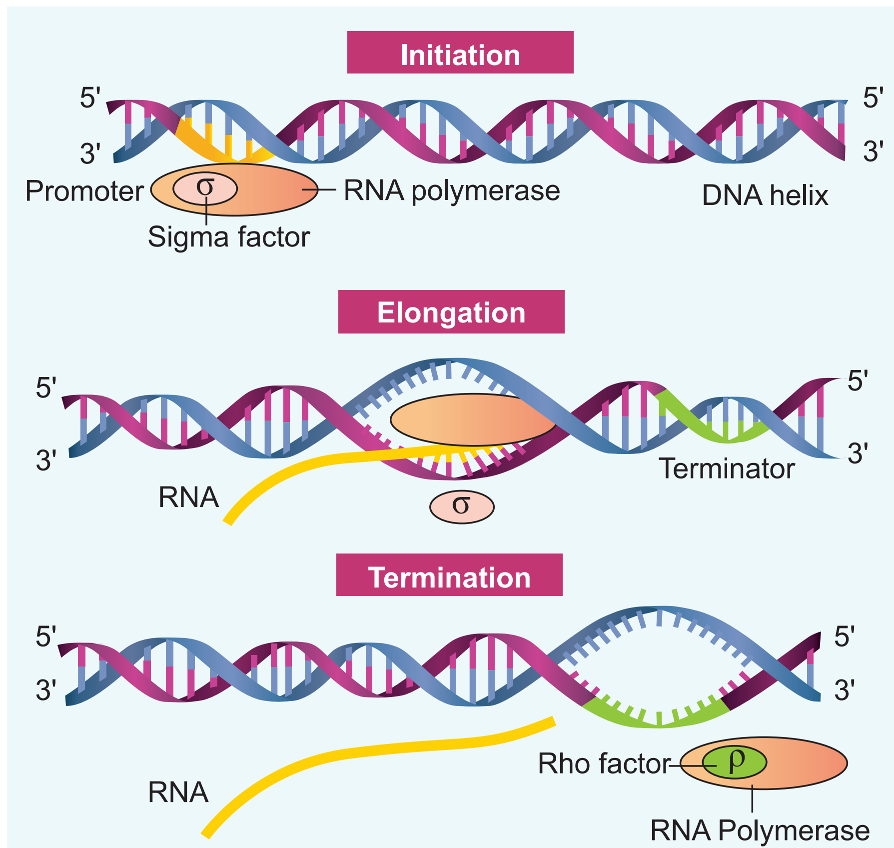


Fig. 5. 8 Process of transcription in prokaryotes

polymerase is only capable of catalyzing the process of elongation. The RNA polymerase associates transiently with initiation factor sigma (σ) and termination factor rho (ρ) to initiate and terminate the transcription, respectively. Association of RNA with these factors instructs the RNA polymerase either to initiate or terminate the process of transcription (**Fig. 5.8**).

In bacteria, since the mRNA does not require any processing to become active and also since transcription and translation take place simultaneously in the same compartment (since there is no separation of cytosol and nucleus in bacteria), many times the translation can begin much before the mRNA is fully transcribed. This is because the genetic material is not separated from other cell organelles by a nuclear membrane consequently; transcription and translation can be coupled in bacteria.

In Eukaryotes, there are at least three RNA polymerases in the nucleus (in addition to RNA polymerase found in the organelles). There is a clear division of labour. The RNA polymerase I transcribes rRNAs (28S, 18S and 5.8S), whereas the RNA polymerase III is responsible for transcription of tRNA, 5srRNA and snRNA. The RNA polymerase II transcribes precursor of mRNA, the hnRNA (heterogenous nuclear RNA). In eukaryotes, the monocistronic structural genes have interrupted coding sequences known as **exons** (expressed sequences) and non-coding sequences called **introns** (intervening sequences). The introns are removed by a process called **splicing**. hnRNA undergoes additional processing called as **capping** and **tailing**. In capping an unusual nucleotide, methyl guanosine triphosphate is added at the 5' end, whereas adenylate residues (200-300) (Poly A) are added at the 3' end in tailing (**Fig. 5.9**). Thereafter, this processed hnRNA, now called mRNA is transported out of the nucleus for translation.

The split gene feature of eukaryotic genes is almost entirely absent in prokaryotes. Originally each exon may have coded for a single polypeptide chain with a specific function. Since exon arrangement and intron removal are flexible, the exon coding for these polypeptide subunits act as domains combining in various ways to form new genes. Single genes can produce different functional proteins by arranging their exons in several different ways through alternate splicing

TN_GOVT_XII_Zoology_chapter5.indd 80 25-02-2019 18:23:05

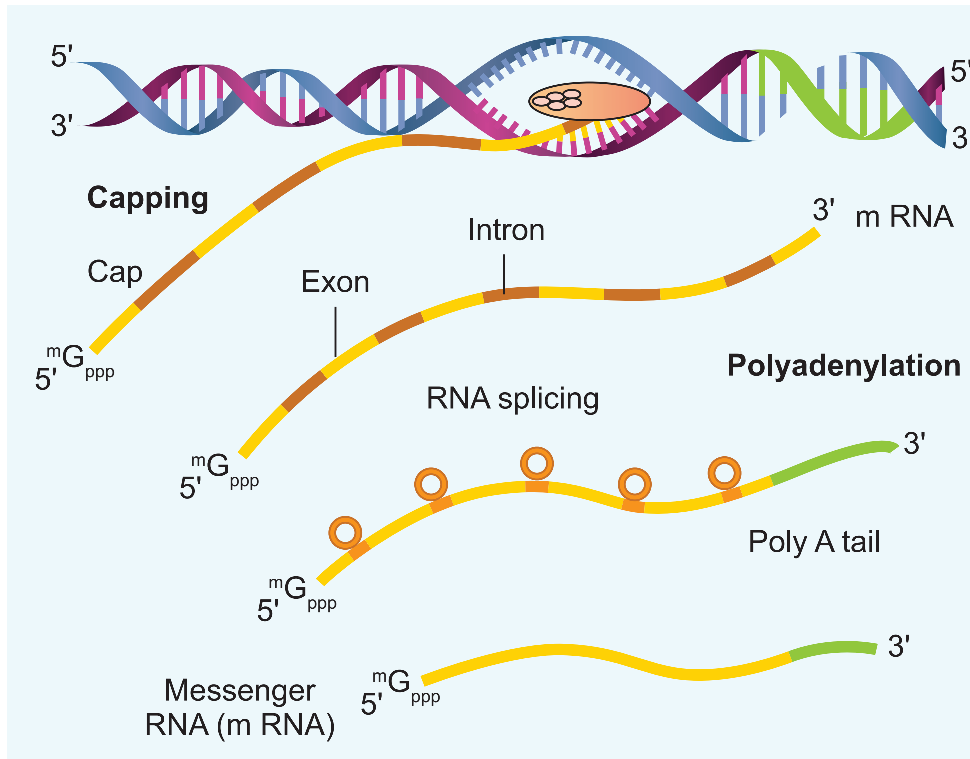


Fig. 5.9 Process of transcription in eukaryotes

patterns, a mechanism known to play an important role in generating both protein and functional diversity in animals. Introns would have arosen before or after the evolution of eukaryotic gene. If introns arose late how did they enter eukaryotic gene? Introns are mobile DNA sequences that can splice themselves out of, as well as into, specific 'target sites' acting like mobile transposon-like elements (that mediate transfer of genes between organisms – Horizontal Gene Transfer - HGT). HGT occurs between lineages of prokaryotic cells, or from prokaryotic to eukaryotic cells and between eukaryotic cells. HGT is now hypothesized to have played a major role in the evolution of life on earth.

5.10 Genetic Code

DNA is the genetic material that carries genetic information in a cell and from generation to generation. At this stage, an attempt will be made to determine in what manner the genetic information exists in DNA molecule? Are they written in coded language on a DNA molecule? If they occur in the language of codes what is the nature of genetic code? The translation of proteins follows the triplet rule; a sequence of three mRNA base (a codon) designates one of the 20 different kinds of amino acids used in protein synthesis.

Genetic code is the sequence relationship between nucleotide in genes (or mRNA) and the amino acids in the proteins they encode. There are 64 possible triplets, and 61 of them are used to represent amino acids. The remaining three triplet codons are termination signals for polypeptide chains. Since there are only 20 amino acids involved in protein synthesis, most of them are encoded by more than one triplet. Two things make this multiple (degenerate) coding possible. First, there is more than one tRNA for most amino acids. Each tRNA has a different anticodon. Second, this pairing is highly specific for the first two portions on the codon, permitting Watson and Crick base pairs (A – U and G - C) to be formed. But at the third position there is a great deal of flexibility as to which base pairs are acceptable. Most part of the genetic code is universal, being the same in prokaryotes and eukaryotes.

The order of base pairs along DNA molecule controls the kind and order of amino acids found in the proteins of an organism. This specific order of base pairs is called genetic code, the blue print establishing the kinds of proteins to be synthesized which makes an organism unique.

Marshall Nirenberg, **Severo Ochoa** (enzyme polynucleotide phosphorylase called Ochoa's enzyme), **Hargobind Khorana**, **Francis Crick** and many others have contributed significantly to decipher the genetic code. The order in which bases are arranged in mRNA decides the order in which amino acids are arranged in proteins. Finally a checker board for genetic code was prepared (**table 5.1**).

The salient features of genetic code are as follows:

- The genetic codon is a **triplet code** and 61 codons code for amino acids and 3 codons do not code for any amino acid and function as **stop codon** (Termination).
- The genetic code is universal. It means that all known living systems use nucleic acids and the same three base codons (triplet codon) direct the synthesis of protein from amino acids. For example, the mRNA (UUU) codon codes for phenylalanine in all cells of all organisms. Some exceptions are reported in prokaryotic, mitochondrial and chloroplast genomes. However similarities are more common than differences.
- A non-overlapping codon means that the same letter is not used for two different codons. For instance, the nucleotide sequence GUU GUC represents only two codons.
- It is comma less, which means that the message would be read directly from one end to the other i.e., no punctuation are needed between two codes.
- A degenerate code means that more than one triplet codon could code for a specific amino acid. For example, codons GUU, GUC, GUA and GUG code for valine.
- Non-ambiguous code means that one codon will code for one amino acid.
- The code is always read in a fixed direction i.e. from 5'→3' direction called polarity.
- AUG has dual functions. It acts as a initiator codon and also codes for the amino acid methionine.
- UAA, UAG (tyrosine) and UGA (tryptophan) codons are designated as termination (stop) codons and also are known as "non-sense" codons.

Table 5.1 Genetic code dictionary

Second Nucleotide in Codon

First nucleotide in codon (5' end)	U	C	A	G	Third nucleotide in codon (3' end)
U	UUU Phe F Phenylalanine	UCU Ser S Serine	UAU Tyr Y Tyrosine	UGU Cys C Cysteine	U
	UUC Phe F Phenylalanine	UCC Ser S Serine	UAC Tyr Y Tyrosine	UGC Cys C Cysteine	C
	UUA Leu L Leucine	UCA Ser S Serine	UAA Termination	UGA Termination	A
	UUG Leu L Leucine	UGC Ser S Serine	UAG Termination	UGG Trp W Tryptophan	G
C	CUU Leu L Leucine	CCU Pro P Proline	CAU His H Histidine	CGU Arg R Arginine	U
	CUC Leu L Leucine	CCC Pro P Proline	CAC His H Histidine	CGC Arg R Arginine	C
	CUA Leu L Leucine	CCA Pro P Proline	CAA Gln Q Glutamine	CGA Arg R Arginine	A
	CUG Leu L Leucine	CCG Pro P Proline	CAG Gln Q Glutamine	CGG Arg R Arginine	G
A	AUU Ile I Isoleucine	ACU Thr T Threonine	AAU Asn N Asparagine	AGU Ser S Serine	U
	AUC Ile I Isoleucine	ACC Thr T Threonine	AAC Asn N Asparagine	AGC Ser S Serine	C
	AUA Ile I Isoleucine	ACA Thr T Threonine	AAA Lys K Lysine	AGA Arg R Arginine	A
	AUG Met M Methionine	ACG Thr T Threonine	AAG Lys K Lysine	AGG Arg R Arginine	G
G	GUU Val V Valine	GCU Ala A Alanine	GAU Asp D Aspartic acid	GGU Gly G Glycine	U
	GUC Val V Valine	GCC Ala A Alanine	GAC Asp D Aspartic acid	GGC Gly G Glycine	C
	GUA Val V Valine	GCA Ala A Alanine	GAA Glu E Glutamic acid	GGA Gly G Glycine	A
	GUG Val V Valine	GCG Ala A Alanine	GAG Glu E Glutamic acid	GGG Gly G Glycine	G

Codon

Three-letter and single-letter abbreviations

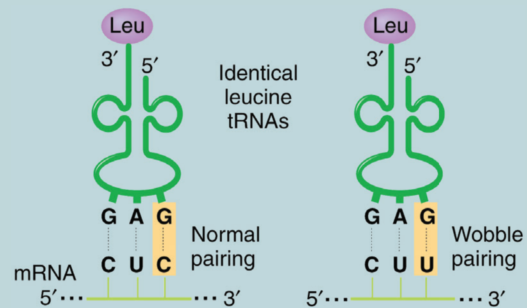
TN_GOVT_XII_Zoology_chapter5.indd 82 25-02-2019 18:23:06

5.10.1 Mutation and genetic code

Comparative studies of mutations (sudden change in a gene) and corresponding alteration in amino acid sequence of specific protein have confirmed the validity of the genetic code. The relationship between genes and DNA are best understood by mutation studies. The simplest type of mutation at the molecular level is a change in nucleotide that substitutes one base for another. Such changes are known as base substitutions which may occur spontaneously or due to the action of mutagens. A well studied example is sickle cell anaemia in humans which results from a point mutation of an allele of β-haemoglobin gene (βHb). A haemoglobin molecule consists of four polypeptide chains of two types, two α chains and two β-chains. Each chain has a heme group on its surface. The heme groups are involved in the binding of oxygen. The human blood disease, sickle cell anaemia is due to abnormal haemoglobin. This abnormality in haemoglobin is due to a single base substitution at the sixth codon of the beta globin gene from GAG to GTG in β -chain of haemoglobin. It results in a change of amino acid glutamic acid to valine at the 6th position of the β -chain. This is the classical example of point mutation that results in the change of amino acid residue glutamic acid to valine (**Fig. 5.10**). The mutant haemoglobin

Wobble Hypothesis

It is a hypothesis proposed by Crick (1966) which states that tRNA anticodon has the ability to wobble at its 5' end by pairing with even non-complementary base of mRNA codon. According to this hypothesis, in codon-anticodon pairing the third base may not be complementary. The third base of the codon is called wobble base and this position is called wobble position. The actual base pairing occurs at first two positions only. The importance of Wobbling hypothesis is that it reduces the number of tRNAs required for polypeptide synthesis and it overcomes the effect of code degeneracy.



In the above example though the codon and the anti codon do not match perfectly, yet the required amino acid is brought perfectly. This enables the economy of tRNA, GUU, GUC, GUA and GUG code for the amino acid - Valine.

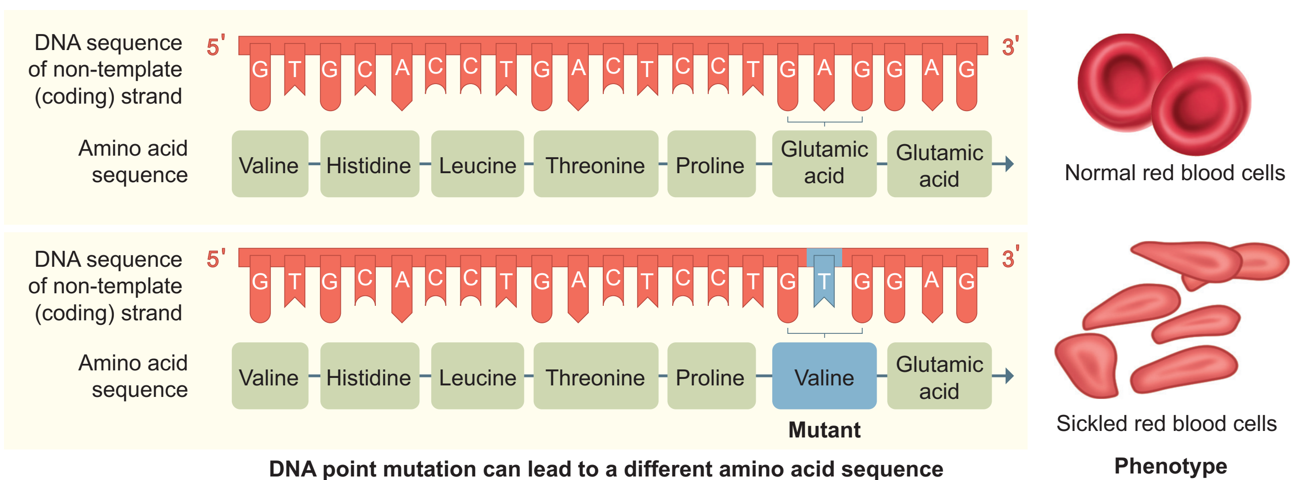


Fig. 5. 10 DNA point mutation

TN_GOVT_XII_Zoology_chapter5.indd 83 25-02-2019 18:23:06

undergoes polymerisation under oxygen tension causing the change in the shape of the RBC from biconcave to a sickle shaped structure.

The effect of point mutation can be understood by the following example.

ABC DEF GHI JKL

If we insert a letter **O** between **DEF** and **GHI** the arrangement would be

ABC DEF OGH IJK L

If we insert **OQ** at the same place the arrangement would be

ABC DEF OQG HIJ KL

The above information shows that insertion or deletion of one or two bases, changes the reading frame from the point of insertions or deletions. Such mutations are referred to as frame shift insertion or deletion mutations. This forms the genetic basis of proof that codon is a triplet and is read in a continuous manner

5.11. tRNA – the adapter molecule

The transfer RNA, (tRNA) molecule of a cell acts as a vehicle that picks up the amino acids scattered through the cytoplasm and also reads specific codes of mRNA molecules. Hence it is called an **adapter molecule.** This term was postulated by Francis Crick.

The two dimensional clover leaf model of tRNA was proposed by Robert Holley. The secondary structure of tRNA depicted in **Fig. 5.11** looks like a **clover leaf**. In actual structure, the tRNA is a compact molecule which looks like an **inverted L**. The clover leaf model of tRNA shows the presence of three arms namely DHU arm, middle arm and TΨC arm. These arms have loops such as amino acyl binding loop, anticodon loop and ribosomal binding loop at their ends. In addition it also shows a small lump called variable loop or extra arm. The amino acid is attached to one end (amino acid acceptor end) and the other end consists of three anticodon nucleotides. The anticodon pairs with a codon in mRNA ensuring that the correct amino acid is incorporated into the growing polypeptide chain. Four different regions of double-stranded RNA are formed during the folding process. Modified bases are especially common in tRNA. Wobbling between anticodon and codon allows some tRNA molecules to read more than one codon.

The process of addition of amino acid to tRNA is known as **aminoacylation** or **charging** and the resultant product is called aminoacyl- tRNA (charged tRNA). Without aminoacylation tRNA is known as uncharged tRNA **(Fig. 5.12).** If two such tRNAs are brought together peptide bond formation is favoured energetically. Numbers of amino acids are joined by peptide bonds to form a polypeptide chain. This aminoacylation is catalyzed by an enzyme **aminoacyl – tRNA synthetase**. This is an endothermic reaction and is associated with ATP hydrolysis. 20

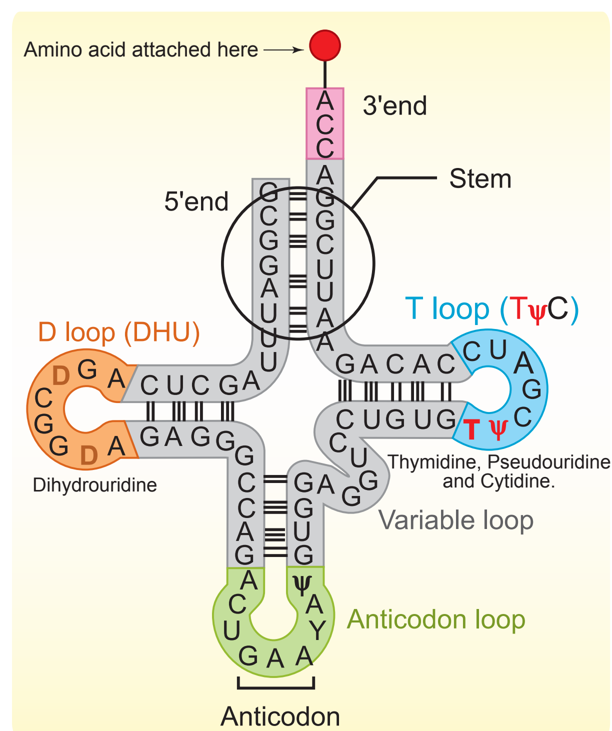


Fig. 5.11 Holley's two-dimensional clover leaf model of transfer RNA

different aminoacyl – tRNA synthetases are known. The power to recognize codon on the mRNA lies in the tRNA and not in the attached amino acid molecule.

The tRNA charged with amino acid serves as an adapter molecule to decode the information on mRNA. This is achieved by the interaction of tRNA with mRNA. The tRNA molecule has a region that contains complementary bases (anticodon) to the codon on the mRNA. For initiation, there is another specific tRNA that is referred to as initiator tRNA. There are no tRNAs for stop codons.

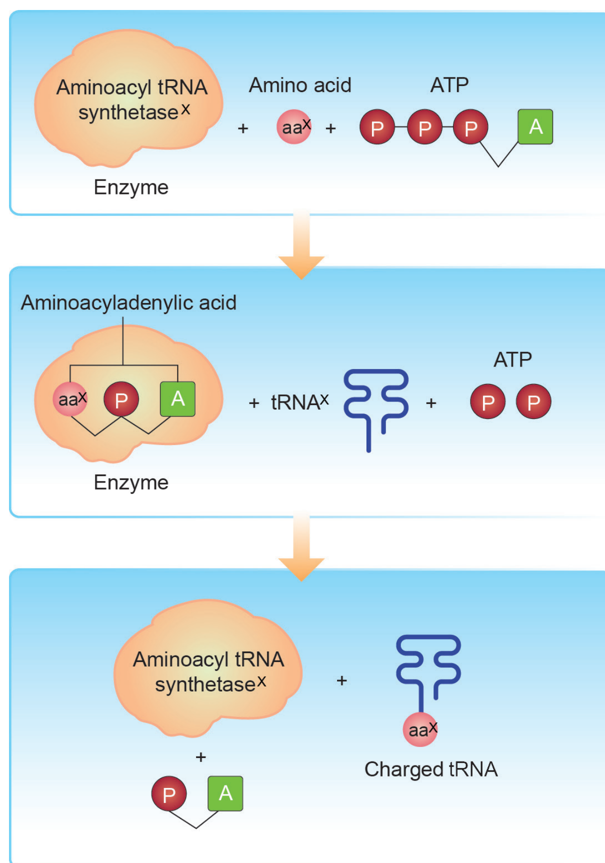


Fig. 5.12 Steps involved in charging tRNA. The 'X' denotes that for each amino acid only the corresponding specific tRNA and specific aminoacyl tRNA synthetase enzyme are involved in the charging process.

5.12 Translation

Translation refers to the process of polymerization of amino acids to form poly peptide chain. The decoding process is carried out by ribosomes that bind mRNA and charged tRNA molecules. The mRNA is translated, starting at the 5' end. After binding to mRNA, the ribosomes move along it, adding new amino acids to the growing polypeptide chain each time it reads a codon. Each codon is read by an anticodon on the corresponding tRNA. Hence the order and sequence of amino acids are defined by the sequence of bases in the mRNA.

5.12.1 Mechanism of Translation

The cellular factory responsible for synthesizing protein is the ribosome. The ribosome consists of structural RNAs and about 80 different proteins. In inactive state, it exists as two subunits; large subunit and small subunit. When the subunit encounters an mRNA, the process of translation of the mRNA to protein begins. The prokaryotic ribosome (70 S) consists of two subunits, the larger subunit (50 S) and smaller subunit (30 S). The ribosomes of eukaryotes (80 S) are larger, consisting of 60 S and 40 S sub units. 'S' denotes the sedimentation efficient which is expressed as Svedberg unit (S). The 30 S subunit of bacterial ribosome contains 16Sr RNA and 50 S subunit contains 5Sr RNA molecules and 23 S RNA and 31 ribosomal proteins. The larger subunit in eukaryotes consist of a 23 S RNA and 5Sr RNA molecule and 31 ribosomal proteins. The smaller eukaryotic subunit consist of 18Sr RNA component and about 33 proteins.

One of the alternative ways of dividing up a sequence of bases in DNA or RNA into codons is called **reading frame**. Any sequence of DNA or RNA, beginning with a start codon and which can be translated into a protein is known as an **Open Reading Frame (ORF)**. A translational unit in mRNA is the sequence of RNA that is flanked by

TN_GOVT_XII_Zoology_chapter5.indd 85 25-02-2019 18:23:06

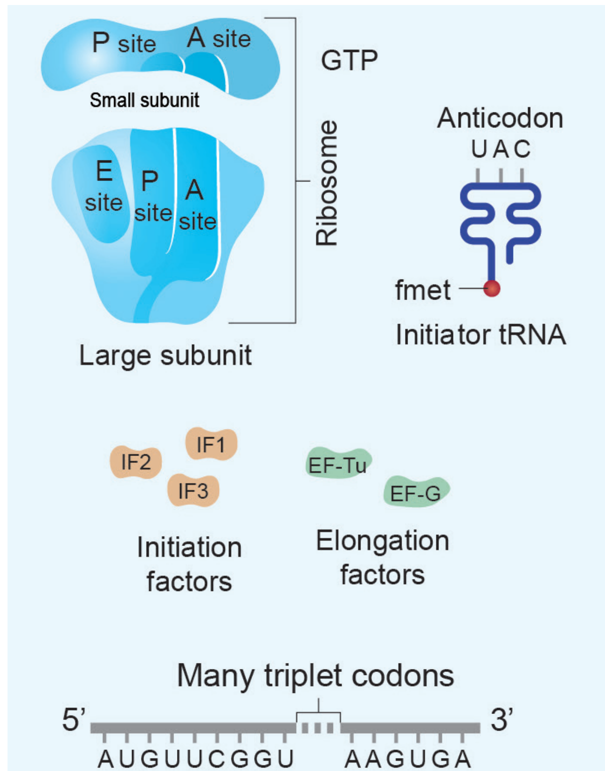


Fig. 5.13 a-Translation components

the start codon (AUG) and the stop codon and codes for polypeptides. mRNA also have some additional sequences that are not translated and are referred to as **Untranslated Regions (UTR).** UTRs are present at both 5' end (before start codon) and at 3' end (after stop codon). The start codon **(AUG)** begins the coding sequence and is read by a special tRNA that carries methionine (met). The initiator tRNA charged with methionine binds to the AUG start codon. In prokaryotes, N - formyl methionine (f^{met}) is attached to the initiator tRNA whereas in eukaryotes unmodified methionine is used. The 5' end of the mRNA of prokaryotes has a special sequence which precedes the initial AUG start codon of mRNA. This ribosome binding site is called the **Shine – Dalgarno sequence** or **S-D sequence**. This sequences base-pairs with a region of the 16Sr RNA of the small ribosomal subunit facilitating initiation. The subunits of the ribosomes (30 S and 50 S) are usually dissociated from each other when not involved in translation **(Fig. 5.13a)**.

Initiation of translation in *E. coli* begins with the formation of an initiation complex, consisting of the 30S subunits of the ribosome, a messenger RNA and the charged N-formyl methionine tRNA (f^{met} – t RNA f^{met}), three proteinaceous initiation factors (IF1, IF2, IF3), GTP(Guanine Tri Phosphate) and Mg $^{2+}$.

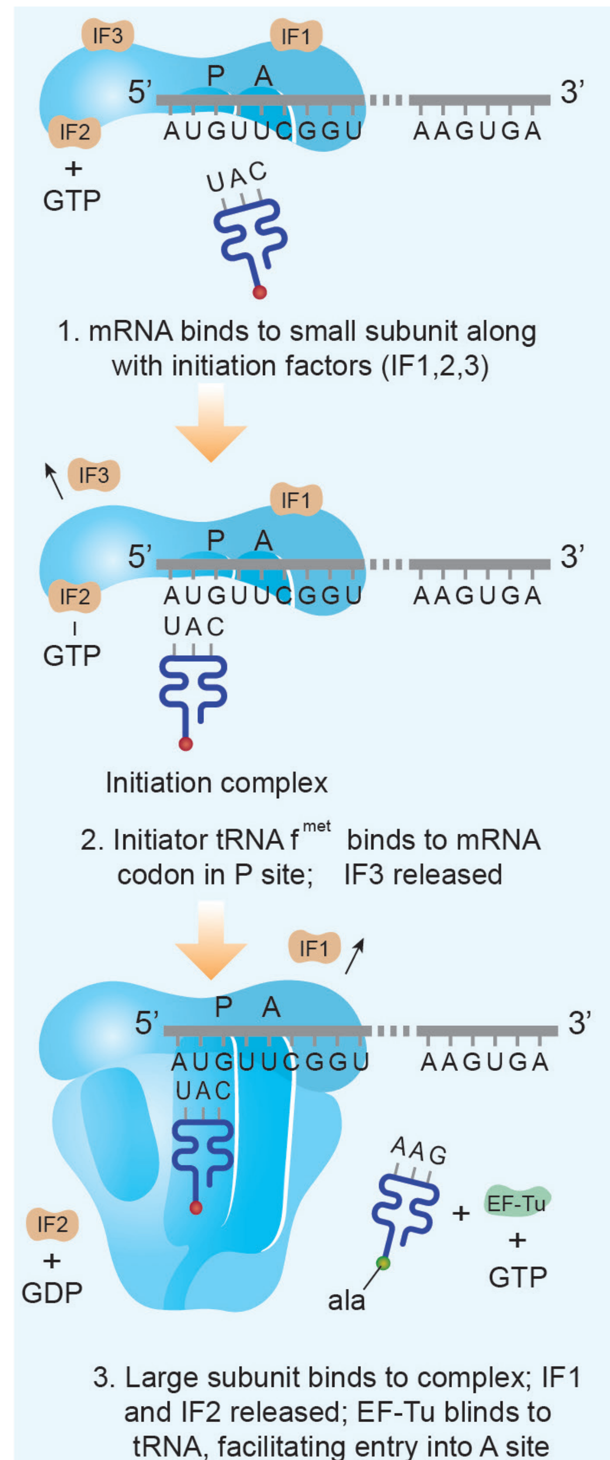


Fig. 5.13 b- Initiation

The components that form the initiation complex interact in a series of steps. IF3 binds to the 30S and allows the 30S subunit to bind to mRNA. Another initiation protein (IF2) then enhances the binding of charged formyl methionine tRNA to the small subunit in response to the AUG triplet. This step 'sets' the reading frame so that all subsequent groups of three ribonucleotides are translated accurately.

The assembly of ribosomal subunits, mRNA and tRNA represent the initiation complex. Once **initiation complex** has been assembled, IF3 is released and allows the initiation complex to combine with the 50S ribosomal subunit to form the complete ribosome (70S). In this process a molecule of GTP is hydrolyzed providing the required energy and the initiation factors (IF2 and IF2 and GDP) are released (**Fig. 5.13 b**).

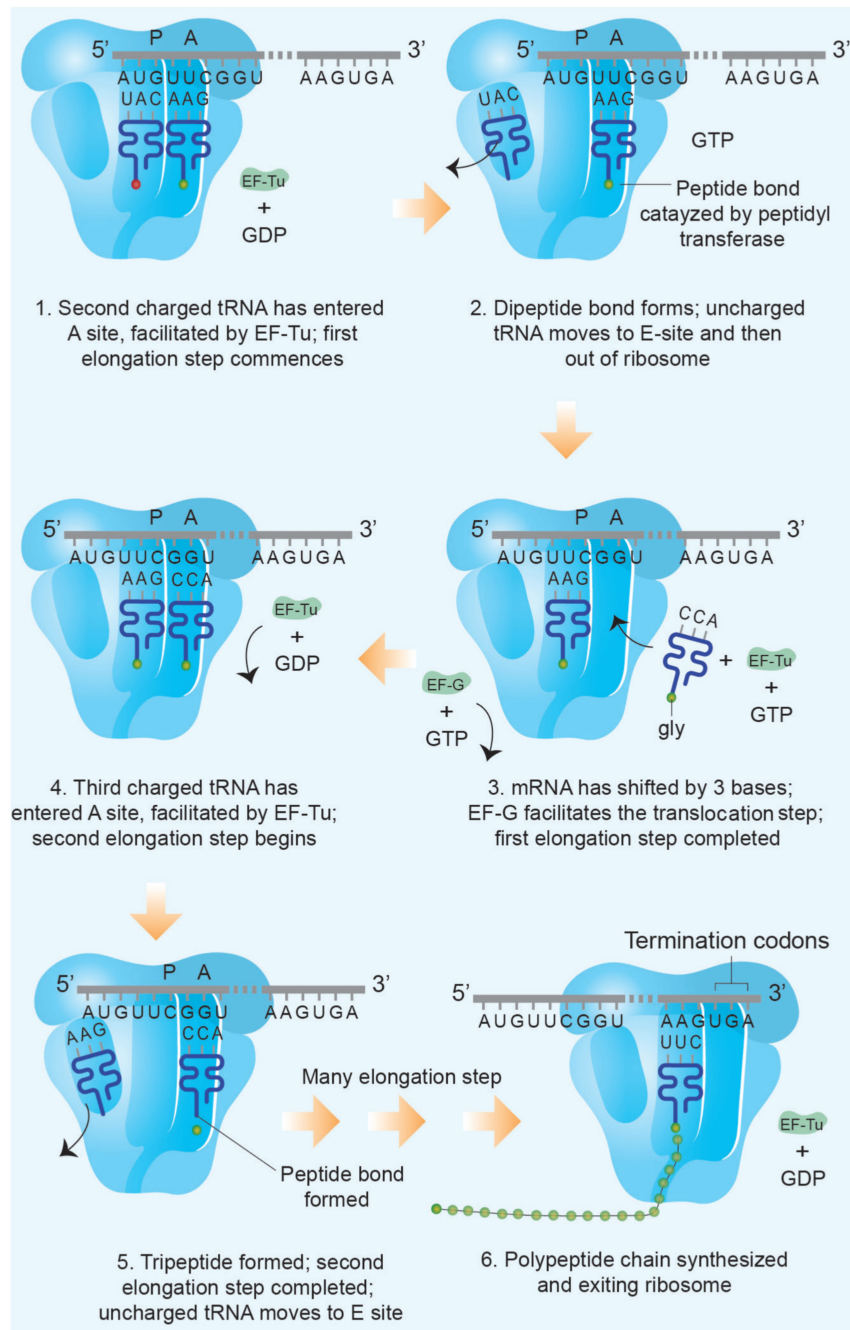


Fig. 5.13 c- Elongation of the growing polypeptide chain during translation

Elongation is the second phase of translation. Once both subunits of the ribosomes are assembled with the mRNA, binding sites for two charged tRNA molecules are formed. The sites in the ribosome are referred to as the aminoacyl site (A site), the peptidyl site (P site) and the exit site (E site). The charged initiator tRNA binds to the P site. The next step in prokaryotic translation is to position the second tRNA at the 'A' site of the ribosome to form hydrogen bonds between its anticodon and the second codon on the mRNA (step1). This step requires the correct transfer RNA, another GTP and two proteins called elongation factors (EF-Ts and EF-Tu).

Once the charged tRNA molecule is positioned at the A site, the enzyme peptidyl transferase catalyses the formation of peptide bonds that link the two amino acids together (step 2). At the same time, the covalent bond between the amino acid and tRNA occupying the P site is hydrolyzed (broken). The product of this reaction is a

TN_GOVT_XII_Zoology_chapter5.indd 87 25-02-2019 18:23:07

dipeptide which is attached to the 3' end of tRNA still residing in the A site. For elongation to be repeated, the tRNA attached to the P site, which is now uncharged is released from the large subunit. The uncharged tRNA moves through the 'E' **site** on the ribosome.

The entire mRNA-tRNA-aa1-aa2 complex shifts in the direction of the **'P' site** by a distance of three nucleotides (step 3). This step requires several elongation factors (EFs) and the energy derived from hydrolysis of GTP. This results in the third triplet of mRNA to accept another charged tRNA into the A site (step 4). The sequence of elongation is repeated over and over (step 5 and step 6). An additional amino acid is added to the growing polypeptide, each time mRNA advances through the ribosome. Once a polypeptide chain is assembled, it emerges out from the base of the large subunit (**Fig. 5.13 c**).

Termination is the third phase of translation. Termination of protein synthesis occurs when one of the three stop codons appears in the 'A' site of the ribosome. The terminal codon signals the action of **GTP – dependent release factor,** which cleaves the polypeptide chain from the terminal tRNA releasing it from the translational complex (step 1). The tRNA is then released from the ribosome, which then dissociates into its subunits (step 2) (**Fig. 5.13 d**).

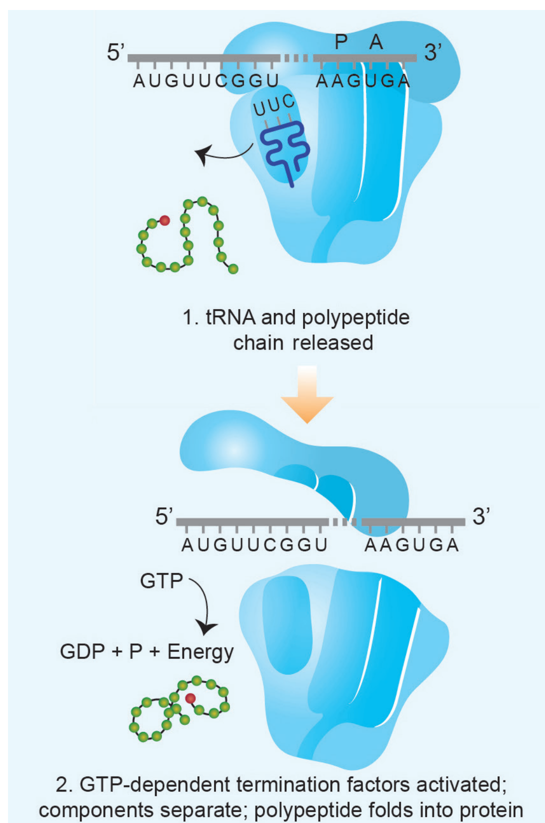


Fig. 5.13 d- Termination of the process of translation

Many antibiotics do not allow pathogenic bacteria to flourish in animal host because they inhibit one or the other stage of bacterial protein synthesis. The antibiotic tetracycline inhibits binding between aminoacyl tRNA and mRNA. Neomycin inhibits the interaction between tRNA and mRNA. Erythromycin inhibits the translocation of mRNA along the ribosome. Streptomycin inhibits the initiation of translation and causes misreading. Chloramphenicol inhibits peptidyl transferase and formation of peptide bonds.

5.13 Regulation of gene expression

We have previously established how DNA is organized into genes, how genes store genetic information, and how this information is expressed. We now consider the most fundamental issues in molecular genetics. How is genetic expression regulated? Evidence in support of the idea that genes can be turned on and off is very convincing. Regulation of gene expression has been extensively studied in prokaryotes, especially in *E. coli*. Gene expression can be controlled or regulated at transcriptional or post transcriptional or translational level. Here, we are going to discuss regulation of gene expression at transcriptional level. Usually, small extracellular or intracellular metabolites trigger initiation or inhibition of gene expression. The clusters of gene with related functions are called **operons**. They usually transcribe single mRNA molecules.

TN_GOVT_XII_Zoology_chapter5.indd 88 25-02-2019 18:23:07

In *E.coli*, nearly 260 genes are grouped into 75 different operons.

Structure of the operon: Each operon is a unit of gene expression and regulation and consists of one or more **structural genes** and an adjacent operator gene that controls transcriptional activity of the structural gene.

i) The structural gene codes for proteins, rRNA and tRNA required by the cell.

ii) Promoters are the signal sequences in DNA that initiate RNA synthesis. RNA polymerase binds to the promoter prior to the initiation of transcription.

iii) The operators are present between the promoters and structural genes. The repressor protein binds to the operator region of the operon.

The *Lac* (Lactose) operon: The metabolism of lactose in *E.coli* requires three enzymes – permease, β-galactosidase (β-gal) and transacetylase. The enzyme permease is needed for entry of lactose into the cell, β-galactosidase brings about hydrolysis of lactose to glucose and galactose, while transacetylase transfers acetyl group from acetyl Co A to β-galactosidase.

The *lac operon* consists of one regulator gene ('i' gene refers to inhibitor) promoter sites (p), and operator site (o). Besides these, it has three structural genes namely *lac* z,y and *lac* a. The *lac* '**z**' gene codes for β-galactosidase, *lac* '**y**' gene codes for permease and '**a**' gene codes for transacetylase.

Jacob and **Monod** proposed the classical model of *Lac* operon to explain gene expression and regulation in *E.coli*. In *lac* operon, a polycistronic structural gene is regulated by a common promoter and regulatory gene. When the cell is using its normal energy source as glucose, the '**i**' gene transcribes a repressor mRNA and after its translation, a **repressor protein** is produced. It binds to the operator region of the operon and prevents translation, as a result, β-galactosidase is not produced. In the absence of preferred carbon source such as glucose, if lactose is available as an energy source for the bacteria then lactose enters the cell as a result of permease enzyme. Lactose acts as an inducer and interacts with the repressor to inactivate it.

The repressor protein binds to the operator of the operon and prevents RNA polymerase from transcribing the operon. In the presence of inducer, such as lactose or allolactose, the repressor is inactivated by interaction with the inducer. This allows RNA polymerase to bind to the promotor site and transcribe the operon to produce *lac* mRNA which enables formation of all the required enzymes needed for lactose metabolism (**Fig. 5.14**). This regulation of *lac* operon by the repressor is an example of negative control of transcription initiation. *Lac*

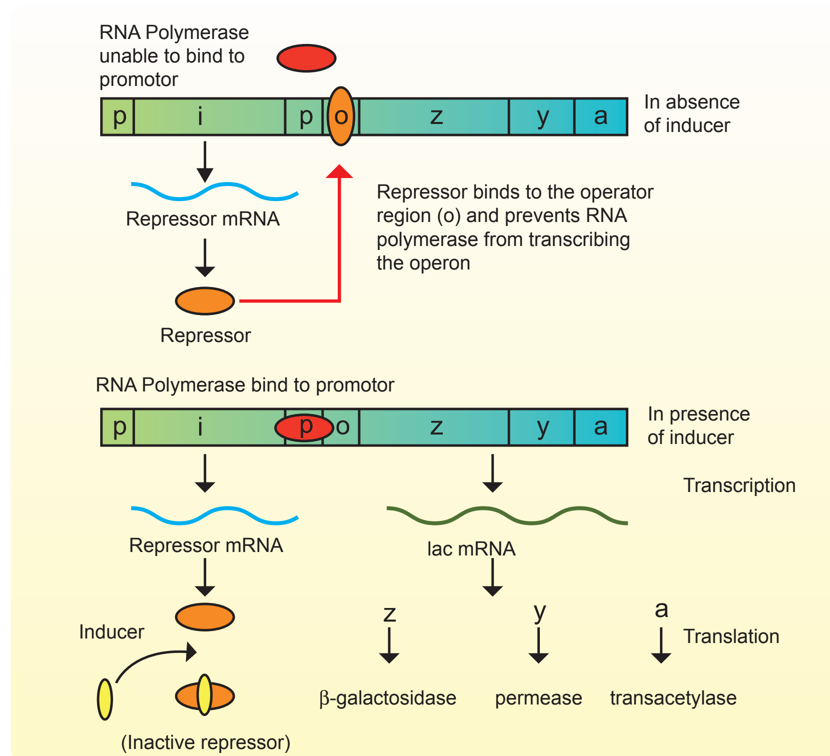


Fig. 5.14 *Lac* Operon model

operon is also under the control of positive regulation as well.

5.14 Human Genome Project (HGP)

The international human genome project was launched in the year 1990. It was a mega project and took 13 years to complete. The human genome is about 25 times larger than the genome of any organism sequenced to date and is the first vertebrate genome to be completed. Human genome is said to have approximately 3×10^9 bp. HGP was closely associated with the rapid development of a new area in biology called bioinformatics.

5.14.1 Goals and methodologies of Human Genome Project

The main goals of Human Genome Project are as follows

- Identify all the genes (approximately 30000) in human DNA.
- Determine the sequence of the three billion chemical base pairs that makeup the human DNA.
- To store this information in databases.
- Improve tools for data analysis.
- Transfer related technologies to other sectors, such as industries.
- Address the ethical, legal and social issues (ELSI) that may arise from the project.

The methodologies of the Human Genome Project involved two major approaches. One approach was focused on identifying all the genes that are expressed as RNA (**ETS_S – Expressed Sequence Tags**). The other approach was sequence annotation. Here, sequencing the whole set of genome was taken, that contains all the coding and non-coding sequences and later assigning different regions in the sequences with functions. For sequencing, the total DNA from a cell is isolated and converted into random fragments of relatively smaller sizes and cloned in suitable hosts using specialized vectors. This cloning results in amplification of pieces of DNA fragments so that it could subsequently be sequenced with ease. Bacteria and yeast are two commonly used hosts and these vectors are called as **BAC** (**Bacterial Artificial Chromosomes**) and **YAC** (**Yeast Artificial Chromosomes**). The fragments are sequenced using automated **DNA sequencers** (developed by Frederick Sanger). The sequences are then arranged based on few overlapping regions, using specialized computer based programs. These sequences were subsequently annotated and are assigned to each chromosome. The genetic and physical maps on the genome are assigned using information on polymorphism of restriction endonuclease recognition sites and some repetitive DNA sequences, called **microsatellites**. The latest method of sequencing even longer fragments is by a method called **Shotgun sequencing** using super computers, which has replaced the traditional sequencing methods.

5.14.2 Salient features of Human Genome Project:

- Although human genome contains 3 billion nucleotide bases, the DNA sequences that encode proteins make up only about 5% of the genome.
- An average gene consists of 3000 bases, the largest known human gene being **dystrophin** with 2.4 million bases.
- The function of 50% of the genome is derived from transposable elements such as LINE and ALU sequence.
- Genes are distributed over 24 chromosomes. Chromosome 19 has the highest gene density. Chromosome 13 and Y chromosome have lowest gene densities.

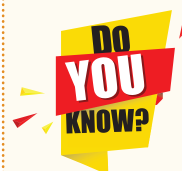
TN_GOVT_XII_Zoology_chapter5.indd 90 25-02-2019 18:23:07

- The chromosomal organization of human genes shows diversity.
- There may be 35000-40000 genes in the genome and almost 99.9 nucleotide bases are exactly the same in all people.
- Functions for over 50 percent of the discovered genes are unknown.
- Less than 2 percent of the genome codes for proteins.
- Repeated sequences make up very large portion of the human genome. Repetitive sequences have no direct coding functions but they shed light on chromosome structure, dynamics and evolution (genetic diversity).
- Chromosome 1 has 2968 genes whereas chromosome 'Y' has 231 genes.
- Scientists have identified about 1.4 million locations where single base DNA differences (**SNPs – Single nucleotidepolymorphism** – pronounce as '**snips**') occur in humans. Identification of '**SNIPS**' is helpful in finding chromosomal locations for disease associated sequences and tracing human history.

5.14.3 Applications and future challenges

The mapping of human chromosomes is possible to examine a person's DNA and to identify genetic abnormalities. This is extremely useful in diagnosing diseases and to provide genetic counselling to those planning to have children. This kind of information would also create possibilities for new gene therapies. Besides providing clues to understand human biology, learning about non-human organisms, DNA sequences can lead to an understanding of their natural capabilities that can be applied towards solving challenges in healthcare, agriculture, energy production and environmental remediation. A new era of molecular medicine, characterized by looking into the most fundamental causes of disease than treating the symptoms will be an important advantage.

- Once genetic sequence becomes easier to determine, some people may attempt to use this information for profit or for political power.
- Insurance companies may refuse to insure people at 'genetic risk' and this would save the companies the expense of future medical bills incurred by 'less than perfect' people.
- Another fear is that attempts are being made to "breed out" certain genes of people from the human population in order to create a 'perfect race'.



Pharmacogenomics is the study of how genes affect a person's response to drugs. This relatively new field combines pharmacology (the science of drugs) and genomics (the study of genes and their functions) to develop effective, safe medications and doses that will be tailored to a person's genetic makeup.

5.15 DNA fingerprinting technique

The DNA fingerprinting technique was first developed by Alec Jeffreys in 1985 (Recipient of the Royal Society's Copley Medal in 2014). Each of us have the same chemical structure of DNA. But there are millions of differences in the DNA sequence of base pairs. This makes the uniqueness among us so that each of us except identical twins is different from each other genetically. The DNA of a person and finger prints are unique. There are 23 pairs of human chromosomes with 1.5 million pairs of genes. It is a well known fact that genes are segments of DNA which differ in the sequence

TN_GOVT_XII_Zoology_chapter5.indd 91 25-02-2019 18:23:07

of their nucleotides. Not all segments of DNA code for proteins, some DNA segments have a regulatory function, while others are intervening sequences (introns) and still others are repeated DNA sequences. In DNA fingerprinting, short repetitive nucleotide sequences are specific for a person. These nucleotide sequences are called as **variable number tandem repeats (VNTR)**.The VNTRs of two persons generally show variations and are useful as genetic markers.

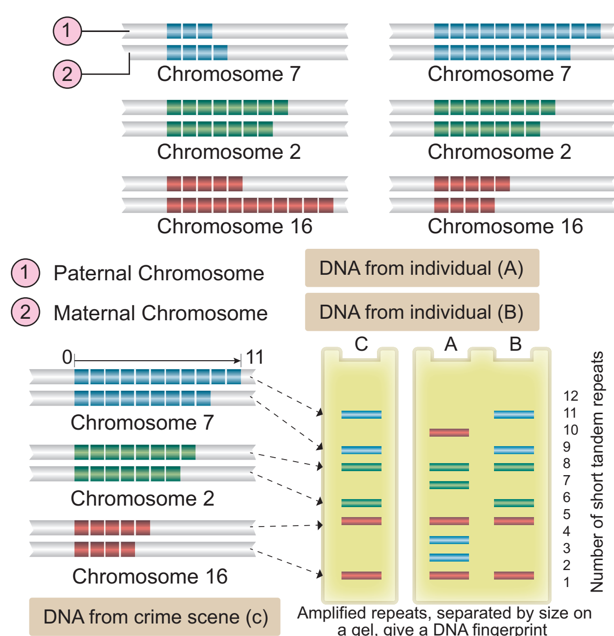


Fig. 5.15 Schematic representation of DNA fingerprinting : Few representative chromosomes have been shown to contain different copy number of VNTR

DNA finger printing involves identifying differences in some specific regions in DNA sequence called **repetitive DNA**, because in these sequences, a small stretch of DNA is repeated many times. These repetitive DNA are separated from bulk genomic DNA as different peaks during density gradient centrifugation. The bulk DNA forms a major peak and the other small peaks are referred to as **satellite DNA**. Depending on base composition (A : T rich or G : C rich), length of segment and number of repetitive units, the satellite DNA is classified into many sub categories such as micro-satellites, mini-satellites, etc., These sequences do not code for any proteins, but they form a large portion of human genome. These sequences show high degree of polymorphism and form the basis of DNA fingerprinting **(Fig. 5.15)**. DNA isolated from blood, hair, skin cells, or other genetic evidences left at the scene of a crime can be compared through VNTR patterns, with the DNA of a criminal suspect to determine guilt or innocence. VNTR patterns are also useful in establishing the identity of a homicide victim, either from DNA found as evidence or from the body itself.

The Steps in DNA Fingerprinting technique is depicted in Fig. 5.16.

1. **Extraction of DNA**

   The process of DNA fingerprinting starts with obtaining a sample of DNA from blood, semen, vaginal fluids, hair roots, teeth, bones, etc.,

2. **Polymerase chain reaction (PCR)**

   In many situations, there is only a small amount of DNA available for DNA fingerprinting. If needed many copies of the DNA can be produced by PCR (DNA amplification).

3. **Fragmenting DNA**

   DNA is treated with restriction enzymes which cut the DNA into smaller fragments at specific sites.

4. **Separation of DNA by electrophoresis**
   During electrophoresis in an agarose gel, the DNA fragments are separated into bands of different sizes. The bands of separated DNA are sieved out of the gel using a nylon membrane (treated with chemicals that allow for it to break the hydrogen bonds of DNA so there are single strands).

5. **Denaturing DNA**

   The DNA on gels is denatured by using alkaline chemicals or by heating.

TN_GOVT_XII_Zoology_chapter5.indd 92 25-02-2019 18:23:08

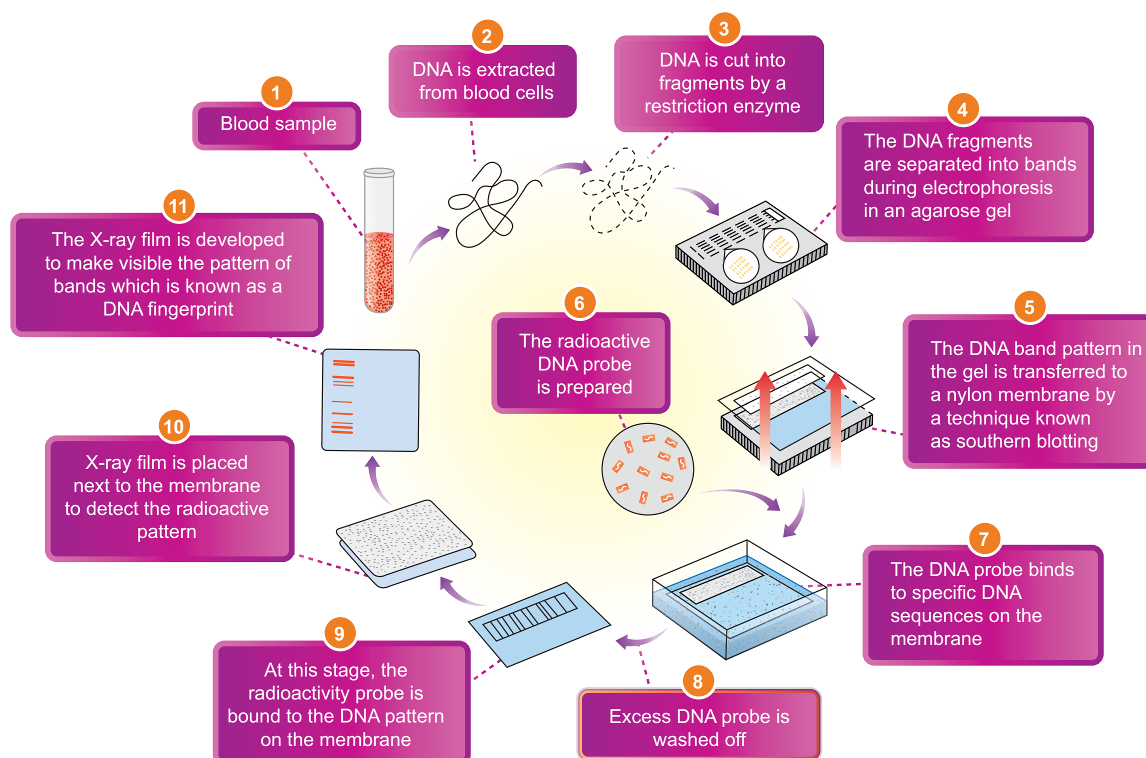


Fig. 5.16 Steps in DNA finger printing

6. **Blotting**

   The DNA band pattern in the gel is transferred to a thin nylon membrane placed over the 'size fractionated DNA strand' by **Southern blotting**.

7. **Using probes to identify specific DNA**

   A radioactive probe (DNA labeled with a radioactive substance) is added to the DNA bands. The probe attaches by base pairing to those restriction fragments that are complementary to its sequence. The probes can also be prepared by using either 'fluorescent substance' or 'radioactive isotopes'.

8. **Hybridization with probe**

   After the probe hybridizes and the excess probe washed off, a photographic film is placed on the membrane containing 'DNA hybrids'.

9. **Exposure on film to make a genetic/ DNA Fingerprint**

   The radioactive label exposes the film to form an image (image of bands) corresponding to specific DNA bands. The thick and thin dark bands form a pattern of **bars** which constitutes a genetic fingerprint.

Application of DNA finger printing

- **Forensic analysis** - It can be used in the identification of a person involved in criminal activities, for settling paternity or maternity disputes, and in determining relationships for immigration purposes.
- **Pedigree analysis** – inheritance pattern of genes through generations and for detecting inherited diseases.
- **Conservation of wild life** – protection of endangered species. By maintaining DNA records for identification of tissues of the dead endangered organisms.
- **Anthropological studies**–It is useful in determining the origin and migration of human populations and genetic diversities.

Summary

In the twentieth century, one of the landmark discovery in biology was the identification of DNA, as genetic material of living organisms. Gene may be defined as a segment of DNA which is responsible for inheritance and expression of a particular character.

In 1953, James Watson and Francis Crick proposed DNA structure based on X-ray crystallographic studies provided by Maurice Wilkins and Rosalind Franklin. Nucleotides are the structural units of nucleic acids. Each nucleotide has three components, i) pentose sugar ii) nitrogenous base and iii) phosphate. DNA and RNA are polynucleotides. DNA has double stranded helical structure while RNA is a single stranded structure. DNA acts as genetic material of almost all the living organism except few viruses.

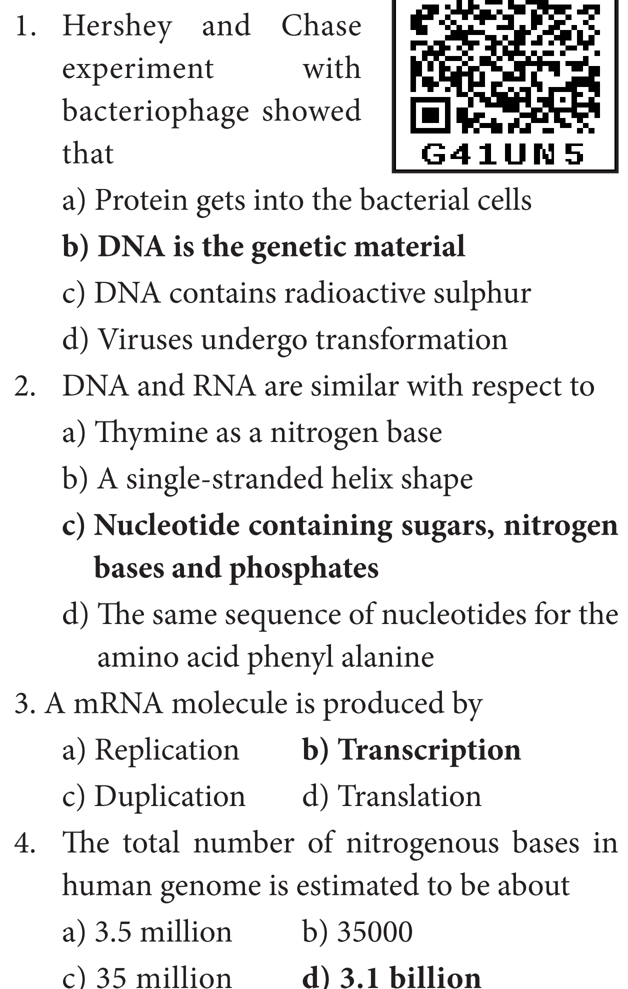
The non genetic RNAs are of three types; m-RNA, r-RNA and t-RNA. They help in protein synthesis. DNA has capacity of replication, while the three types of RNA are transcribed on DNA. Meselson and Stahl (1958) proved experimentally the semi-conservative nature of DNA replication using heavy isotope of nitrogen N^{15} in *E.coli*.

In 1958 Crick proposed that DNA determines the sequence of amino acids in a polypeptide (protein) through mRNA, and proposed the central dogma of protein synthesis which involves transcription and translation. The process of copying genetic information from one strand of DNA into RNA is termed transcription. The DNA transcribed RNA molecules serve as a template for the synthesis of polypeptides by a process termed translation. Each amino acid in a polypeptide chain is represented by a sequence of three nucleotides in the RNA known as the genetic code. RNA transfers genetic message from nucleus to the cytoplasm. DNA is always present in the nucleus and synthesis is also confined to the nucleus

Jacob and Monod proposed the classical model of *Lac* operon to explain gene expression and regulation in *E. coli*. In *lac* operon a polycistronic structural gene is regulated by a common promoter and regulator. It is an example of negative control of transcription initiation.

Human genome project, a mega project was aimed to sequence every gene in the human genome. Polymerase chain reaction is an *in vitro* method of synthesis of nucleic acids wherein, a specific DNA segment is amplified rapidly without concomitant replication of the rest of the DNA molecule. DNA fingerprinting is a technique to identify variations in individuals of a population at the DNA level. It has immense applications in the field of forensic analysis, pedigree analysis, anthropological studies, and conservation of wild life.

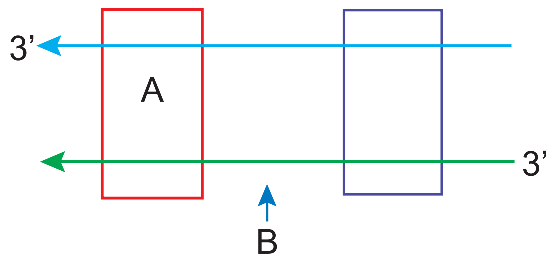
Evaluation



1. Hershey and Chase experiment with bacteriophage showed that
   a) Protein gets into the bacterial cells
   b) DNA is the genetic material
   c) DNA contains radioactive sulphur
   d) Viruses undergo transformation
2. DNA and RNA are similar with respect to
   a) Thymine as a nitrogen base
   b) A single-stranded helix shape
   c) Nucleotide containing sugars, nitrogen bases and phosphates
   d) The same sequence of nucleotides for the amino acid phenyl alanine
3. A mRNA molecule is produced by
   a) Replication **b) Transcription**
   c) Duplication d) Translation
4. The total number of nitrogenous bases in human genome is estimated to be about
   a) 3.5 million b) 35000
   c) 35 million **d) 3.1 billion**

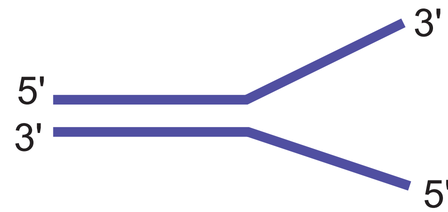
5. *E. coli* cell grown on ^{15}N medium are transferred to ^{14}N medium and allowed to grow for two generations. DNA extracted from these cells is ultracentrifuged in a cesium chloride density gradient. What density distribution of DNA would you expect in this experiment?
   (a) One high and one low density band.
   (b) One intermediate density band.
   (c) One high and one intermediate density band.
   (d) One low and one intermediate density band.
6. What is the basis for the difference in the synthesis of the leading and lagging strand of DNA molecules?
   (a) Origin of replication occurs only at the 5' end of the molecules.
   (b) DNA ligase works only in the 3' → 5' direction.
   (c) DNA polymerase can join new nucleotides only to the 3' end of the growing stand.
   (d) Helicases and single-strand binding proteins that work at the 5' end.
7. Which of the following is the correct sequence of event with reference to the central dogma?
   (a) Transcription, Translation, Replication
   (b) Transcription, Replication, Translation
   (c) Duplication, Translation, Transcription
   (d) Replication, Transcription, Translation
8. Which of the following statements about DNA replication is not correct?
   (a) Unwinding of DNA molecule occurs as hydrogen bonds break.
   (b) Replication occurs as each base is paired with another exactly like it.
   (c) Process is known as semi conservative replication because one old strand is conserved in the new molecule.
   (d) Complementary base pairs are held together with hydrogen bonds.
9. Which of the following statements is not true about DNA replication in eukaryotes?
   (a) Replication begins at a single origin of replication.
   (b) Replication is bidirectional from the origins.
   (c) Replication occurs at about 1 million base pairs per minute.
   (d) There are numerous different bacterial chromosomes, with replication ocurring in each at the same time.
10. The first codon to be deciphered was __________ which codes for ________.
   (a) AAA, proline (b) GGG, alanine
   (c) UUU, Phenylalanine (d)TTT, arginine
11. Meselson and Stahl's experiment proved
   (a)Transduction
   (b) Transformation
   (c) DNA is the genetic material
   (d) Semi-conservative nature of DNA replication
12. Ribosomes are composed of two subunits; the smaller subunit of a ribosome has a binding site for _________ and the larger subunit has two binding sites for two __________.**Ans (mRNA, tRNA)**
13. An operon is a:
   (a) Protein that suppresses gene expression
   (b) Protein that accelerates gene expression
   (c) Cluster of structural genes with related function
   (d) Gene that switched other genes on or off
14. When lactose is present in the culture medium:
   (a) Transcription of *lac y, lac z, lac a* genes occurs.
   (b) Repressor is unable to bind to the operator.
   (c) Repressor is able to bind to the operator.
   (d) Both (a) and (b) are correct.

15. Give reasons: Genetic code is 'universal'.
16. Name the parts marked 'A' and 'B' in the given transcription unit:



17. Differentiate - Leading stand and lagging strand
18. Differentiate - Template strand and coding strand.
19. Mention any two ways in which single nucleotide polymorphism (SNPs) identified in human genome can bring revolutionary change in biological and medical science.
20. State any three goals of the human genome project.
21. In *E.coli*, three enzymes β- galactosidase, permease and transacetylase are produced in the presence of lactose. Explain why the enzymes are not synthesized in the absence of lactose.
22. Distinguish between structural gene, regulatory gene and operator gene.
23. A low level of expression of *lac* operon occurs at all the windows for treatment of various genetic disorders. Justify the statement.
24. Why the human genome project is called a mega project?
25. From their examination of the structure of DNA, What did Watson and Crick infer about the probable mechanism of DNA replication, coding capability and mutation?
26. Why tRNA is called an adapter molecule?
27. What are the three structural differences between RNA and DNA?
28. Name the anticodon required to recognize the following codons: AAU, CGA, UAU, and GCA.
29. a) Identify the figure given below

    b) Redraw the structure as a replicating fork and label the parts



    c) Write the source of energy for this replication and name the enzyme involved in this process.

    d) Mention the differences in the synthesis of protein, based on the polarity of the two template strands.
30. If the coding sequence in a transcription unit is written as follows:

    5' TGCATGCATGCATGCATGCATGCATGC 3'

    Write down the sequence of mRNA.
31. How is the two stage process of protein synthesis advantageous?
32. Why did Hershey and Chase use radioactively labelled phosphorous and sulphur only? Would they have got the same result if they use radiolabelled carbon and nitrogen?
33. Explain the formation of a nucleosome.
34. It is established that RNA is the first genetic material. Justify giving reasons.

TN_GOVT_XII_Zoology_chapter5.indd 96

25-02-2019 18:23:09

Concept Map

TN_GOVT_XII_Zoology_chapter5.indd 97 25-02-2019 18:23:10

ICT CORNER

Molecular Genetics

Lets us explore the gene expression

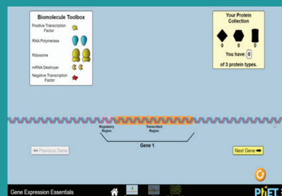
Procedure :

Step 1: Use the URL or scan the QR Code to launch the "Gene Expression Essentials" activity page.

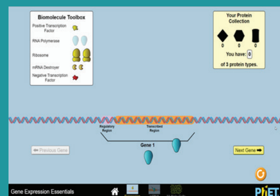
Step 2: Click "Expression" pick the genetic material from the Biomolecule Toolbox, understand the changes for the three different genes.

Step 3: Click "mRNA" and slide through the slider in Positive Transcription factors and Negative Transcription factors such as Concentration, Affinity. Also Slide through "Affinity" in RNA Polymerase.

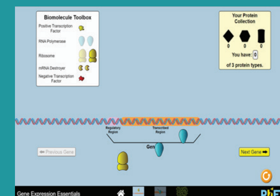
Step 4: Click "Multiple Cells" and find the average protein level vs Time in the graph indicated above.



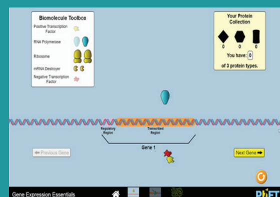
Step 1



Step 2



Step 3



Step 4

Molecular Genetics URL:

https://phet.colorado.edu/sims/html/gene-expression-essentials/latest/gene-expression-essentials_en.html

B267_12_Z00_EM

*Pictures are indicative only

*Allow flash player

TN_GOVT_XII_Zoology_chapter5.indd 98 25-02-2019 18:23:12