[TOPIC 1] The DNA and RNA World

1.1 The DNA World

Over the years after **Mendel,** the nature of the genetic material was investigated, resulting in the realisation that **DNA** is the genetic material in majority of organisms.

Deoxyribonucleic acid (DNA) and **Ribonucleic acid** (RNA) are the two types of nucleic acid found in living systems. Nucleic acids are polymers of nucleotides. Bacteriophage $\phi \times 174$ has 5386 nucleotides, bacteriophage λ has 48502 bp. *E. coli* has 4.6×10^6 bp and haploid content of human DNA is 3.3×10^9 bp.

DNA acts as a **genetic material** in most organisms, whereas RNA acts as a genetic material in some viruses.

Structure of Polynucleotide Chain

(*i*) A nucleotide has three parts, i.e. a nitrogenous base, a pentose sugar (deoxyribose in DNA and ribose in RNA) and a phosphate group.

(*ii*) **Nitrogenous bases** can be purines, i.e. adenine and guanine or pyrimidines, i.e. cytosine, uracil and thymine.

(*iii*) Cytosine is common for both DNA and RNA and thymine is present in DNA. Uracil is present in RNA at the place of thymine.

(*iv*) A nitrogenous base is linked to the first carbon of pentose sugar through **N-glycosidic linkage** to form a nucleoside, e.g. adenosine, guanosine, etc.

(*v*) When a phosphate group is linked to 5′—OH of a nucleoside on the fifth carbon of pentose sugar through phosphodiester linkage, a corresponding nucleotide is formed.

(*vi*) Two nucleotides are linked through **3′→ 5′ phosphodiester linkage** to form a dinucleotide.

(*vii*) Several nucleotides can be joined to form a polynucleotide chain. In this polynucleotide chain at one end free phosphate group is present which is known as 5′ end, while at one end free OH group is present, which is known as 3′ end.

(*viii*) The backbone in a polynucleotide chain is formed of sugar and phosphates.

(*ix*) The nitrogenous bases linked to sugar moiety project from the backbone.

In case of RNA, every nucleotide residue has an additional N—OH group present at 2′-position in the ribose. Also, the uracil is found at the place of thymine (5-methyl uracil).

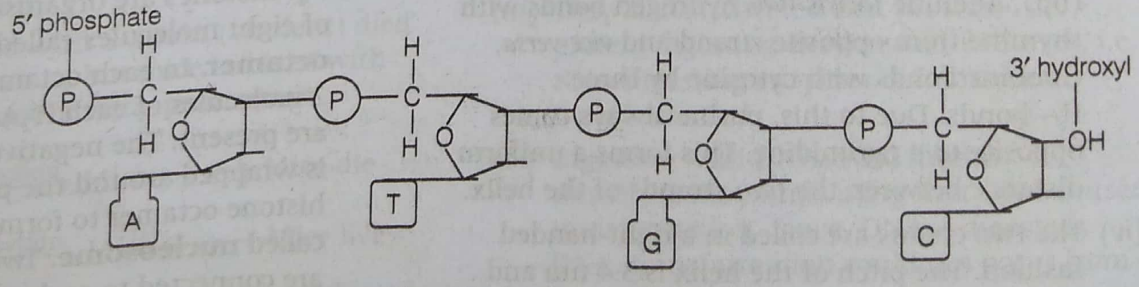

Figure 6.1 A polynucleotide chain

Discoveries Related to Structure of DNA

(*i*) **Friedrich Meischer** (1869), first identified DNA as an acidic substance present in the nucleus and named it as 'nuclein'.

(*ii*) **Levene** (1910) found DNA to contain phosphoric acid as well as deoxyribose sugar. He characterised four type of nucleotides present in DNA.

(*iii*) **Erwin Chargaff** proposed that for a double-stranded DNA, the ratios between adenine (A), thymine (T) and guanine (G), cytosine (C) are constant and equals to one.

$$\frac{A+T}{C+G}=1$$

(*iv*) **James Watson** and **Francis Crick**, in 1953 proposed a very simple **double helix** model for the structure of DNA based on X-ray diffraction data, produced by **Maurice Wilkins** and **Rosalind Franklin** in the same year.

Salient Features of Double Helix Structure of DNA

(*i*) DNA is a long polymer of deoxyribonucleotides. It is made up of two polynucleotide chains, where the backbone is constituted by sugar-phosphate and the bases project inside. So, nitrogenous bases can be considered as steps of the coiled helix.

(*ii*) The two chains have anti-parallel polarity, i.e. $5'\rightarrow3'$ for one, $3'\rightarrow5'$ for another.

(*iii*) The bases in two strands are paired through hydrogen bonds (H—bonds) forming base pairs (bp). Adenine forms two hydrogen bonds with thymine from opposite strand and *vice-versa*. Guanine bonds with cytosine by three H—bonds. Due to this, purine always comes opposite to a pyrimidine. This forms a uniform distance between the two strands of the helix.

(*iv*) The two chains are coiled in a right-handed fashion. The pitch of the helix is 3.4 nm and there are roughly 10 bp in each turn. Due to this, the distance between a base pair in a helix is about 0.34 nm.

(*v*) The plane of one base pair stacks over the other in double helix. This confers stability to the helical structure in addition to H—bonds.

(*vi*) Diameter of DNA double helix is 20 Å.

(*vii*) The length of a DNA double helix is about 2.2 metres $(6.6\times10^{9}\text{bp}\times0.34\times10^{-9}\text{m/bp})$. Therefore, it needs special **packaging in a cell.**

Packaging of DNA Helix

(*i*) **In prokaryotic cells** (which do not have a defined nucleus), such as *E. coli*, DNA being negatively charged is held with some proteins that have positive charge in a region called as **nucleoid.** The DNA in nucleoid is organised in large loops held by proteins.

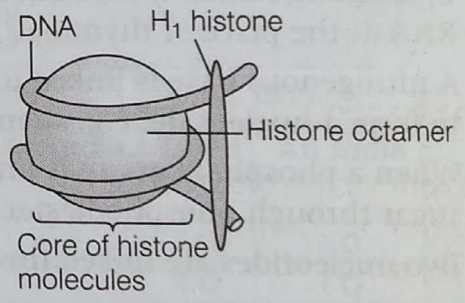

Figure 6.2 A nucleosome

(*ii*) **In eukaryotes,** there is a set of positively charged proteins called **histones** that are rich in basic amino acid residues, lysines and arginines (both positive). There are 5 types of histone proteins known as H_1, H_2A, H_2B, H_3 and H_4. Histones are organised to form a unit of eight molecules called **histone octamer.** In each octamer 2 molecules of each H_2A, H_2B, H_3 and H_4 are present. The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called **nucleosome.** Two nucleosomes are connected to each other with the help of linker DNA, on which H_1 histone is present, so H_1 is also known as **linker histone**.

(*iii*) A typical nucleosome contains 200 bp of DNA helix. Nucleosomes constitute the repeating unit of a structure in nucleus called **chromatin**, i.e. thread-like stained structure. Under electron microscope, the nucleosomes in chromatin can be seen as **beads-on-string.**

This structure in chromatin is packaged to form chromatin fibres that further coils and condense to form **chromosomes** at metaphase stage.

(*iv*) The packaging of chromatin at higher level requires additional set of proteins, which are collectively called **Non-Histone Chromosomal** (NHC) **proteins.**

(*v*) In a nucleus, some regions of chromatin are loosely packed (stains light) called **euchromatin** (transcriptionally active chromatin). While in some regions, chromatin is densely packed (stains dark) called **heterochromatin** (inactive chromatin).

Transforming Principle

(*i*) **Frederick Griffith** (1928) carried out a series of experiments with *Streptococcus pneumoniae* (bacterium causing pneumonia).

(*ii*) According to him, when the bacteria are grown on a culture plate, some produce smooth shiny colonies (S), while others produce rough (R) colonies.

(*iii*) This is because the S-strain bacteria have a mucous (polysaccharide) coat, while R-strain does not. So, S-strain was infectious, while R-strain was non-infectious.

(*iv*) Mice infected with S-strain (virulent) died from pneumonia, but mice infected with R-strain do not develop pneumonia.

S-strain $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice die

R-strain $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice live

(*v*) Griffith killed bacteria by heating and observed that heat-killed S-strain bacteria injected into mice did not kill them. On injecting mixture of heat-killed S and live R bacteria, the mice died. He recovered both living S-bacteria and R-bacteria from dead mice.

S-strain (Heat-killed) $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice live

S-strain (Heat-killed) + R-strain (live) $\xrightarrow{\text{Injection}}$ Mice $\longrightarrow$ Mice die

(*vi*) From this experiment, he concluded that the 'R-strain bacteria' had been transformed by the heat-killed S-strain bacteria. Some transforming principle transferred from heat-killed S-strain, had enabled the R-strain to synthesise a smooth polysaccharide coat and become virulent. This must be due to transfer of the genetic material. However biochemical nature of genetic material was not defined from his experiments.

Biochemical Nature of Transforming Principle

(*i*) **Oswald Avery, Colin MacLeod and Maclyn McCarty** (1933-44), worked to determine the biochemical nature of transforming principle in Griffith's experiment.

(*ii*) They purified biochemicals such as proteins, RNA and DNA, etc. from heat-killed S-cells and discovered that DNA alone from S-bacteria caused R-bacteria to be transformed.

(*iii*) They also discovered that protease, i.e. protein digesting enzyme and RNAases, i.e. RNA-digesting enzymes did not affect transformation.

(*iv*) Digestion with DNAse did inhibit transformation, indicating that DNA caused transformation. Since, DNAse digested DNA, transformation could not occur from R-strain to S-strain.

(*v*) They concluded that DNA is the hereditary material. But, still all the biologists were not convinced.

DNA is the Genetic Material

(*i*) **Alfred Hershey** and **Martha Chase** (1952) gave unequivocal proof that DNA is the genetic material.

(*ii*) In their experiments, bacteriophages, i.e. viruses that infect bacteria, were used. Head of bacteriophage is made up of protein and inside this protein coat DNA is present as its genetic material.

(*iii*) They grew some viruses on a medium that contained radioactive phosphorus and some others on sulphur containing radioactive medium. ^{32}P was used as radioactive phosphorus, while ^{35}S was used as radioactive sulphur.

(*iv*) Viruses grown in the presence of **radioactive phosphorus** contained **radioactive DNA**, but not radioactive protein because DNA contains **phosphorus**, but protein does not. In the same way, viruses grown on radioactive sulphur contained radioactive protein, but not radioactive DNA because DNA does not contain sulphur while protein does.

(*v*) Radioactive phages were allowed to attach to *E. coli* bacteria. As the infection proceeded, viral coats were removed from the bacteria by agitating them in a blender. The virus particles were separated from the bacteria by spinning them in a centrifuge.

(*vi*) Bacteria which were infected with viruses that had radioactive DNA were radioactive, as ^{32}P was detected in the cells indicating that DNA was the material that passed from the virus to the bacteria.

(*vii*) Bacteria that were infected with viruses that had radioactive proteins were not radioactive. As ^{35}S was not detected in the cells, while ^{35}S was present in supernatant. This indicated that the proteins did not enter the bacteria from viruses.

It proved that DNA is a genetic material that is passed from virus to bacteria.

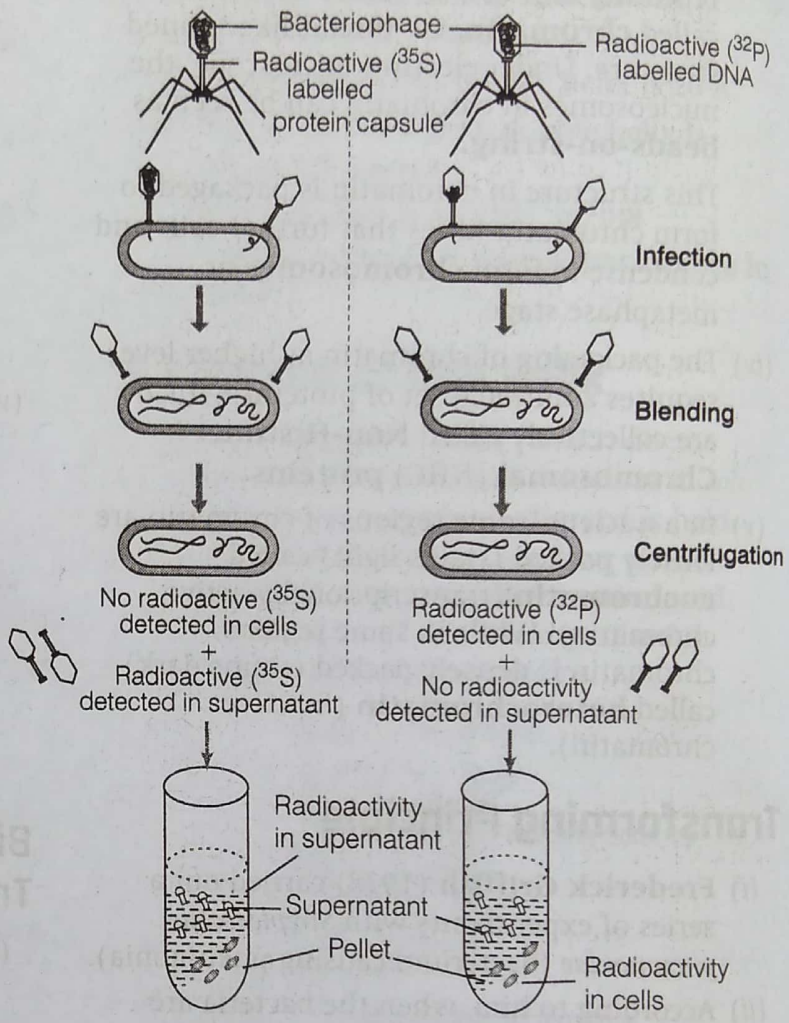

Figure 6.3 Hershey-Chase experiment

Properties of Genetic Material

(*i*) It became establised that DNA is the genetic material from the Hershey-Chase experiment.

(*ii*) In some viruses, RNA was also reported as genetic material, e.g. Tobacco mosaic viruses, QB bacteriophage, etc.

(*iii*) **Characteristics of a Genetic Material**

(a) It should be able to replicate.

(b) It should be chemically and structurally stable.

(c) It should provide scope for slow changes (mutation) that are required for evolution.

(d) It should be able to express itself in the form of 'Mendelian characters'.

(*iv*) According to the above mentioned rules, both the nucleic acids (DNA and RNA) have the ability to direct duplications, because of the rule of base pairing and complementarity. Stability can be explained in DNA as the two strands being complementary if separated by heating come together in appropriate conditions.

(*v*) The 2′—OH group present at every nucleotide in RNA is a reactive group and makes RNA labile and easily degradable, hence it is reactive. RNA is also catalytic in nature.

(*vi*) DNA is chemically less reactive and structurally more stable as compared to RNA. Presence of thymine also confers additional stability to DNA. So, among the two nucleic acids, the DNA is a predominant genetic material.

(*vii*) Both RNA and DNA are able to mutate. Viruses having RNA genome and having shorter life span mutate and evolve faster. So, faster mutation makes RNA more unstable.

(*viii*) DNA is dependent on RNA for protein synthesis, while RNA can directly code for it. The protein synthesising machinery has evolved around RNA. This concluded that the DNA being more stable is suitable for storage of genetic information, while for the transmission of genetic information, RNA is suitable.

Francis Crick proposed the **central dogma** in molecular biology, which states that the flow of genetic information is as follows:

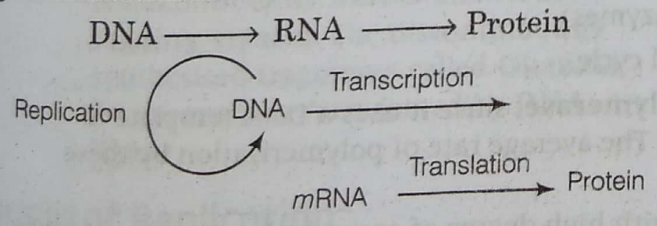

Figure 6.4 Central dogma

In some viruses, the flow of information is in reverse direction, i.e. from RNA to DNA. This process is known as **reverse transcription**. An enzyme is very important for this process known as **reverse transcriptase**.

1.2 RNA World

RNA was the first genetic material to be discovered. There are evidences to prove that essential life processes, such as metabolism, translation, splicing, etc. have evolved around RNA.

(*i*) There are some important biochemical reactions in living systems that are catalysed by RNA catalysts and not by protein enzymes.

(*ii*) DNA has evolved from RNA with chemical modifications that made it more stable, because RNA being a catalyst was reactive and hence, unstable.

There are following three types of RNAs:

(*i*) *m*RNA (messenger RNA) provides the template for transcription.

(*ii*) *t*RNA (transfer RNA) brings amino acids and reads the genetic code.

(*iii*) *r*RNA (ribosomal RNA) plays structural and catalytic role during translation.

All the three RNAs are needed to synthesise a protein in a cell.

1.3 Replication

The scheme for replication of DNA is termed as **semi-conservative DNA replication**. It was proposed by **Watson** and **Crick** (1953). According to it,

(*i*) The two strands would separate and act as a template for the synthesis of new complementary strands.

(*ii*) After replication, each DNA molecule would have one parental and one newly synthesised strand.

(*iii*) The basis of semi-conservative method of replication is complementarity between nitrogenous bases.

Experimental proof that DNA replicates semi-conservatively, comes first from *E. coli* and later from higher organisms, such as plants and human cells.

Matthew Meselson and Franklin Stahl performed the following experiments to prove this in 1958.

(*i*) *E. coli* was grown in a medium containing $^{15}NH_4Cl$ as the only nitrogen source for many generations. ^{15}N (heavy isotope of nitrogen) got incorporated into newly synthesised DNA (and other nitrogen containing compounds). This heavy DNA molecule could be distinguished from the normal DNA by centrifugation in a cesium chloride (CsCl) density gradient.

(*ii*) They then transferred the cells into a medium with normal $^{14}NH_4Cl$ and took samples at various definite intervals as the cells multiplied and extracted the DNA that remained as double-stranded helices. DNA samples were separated independently on CsCl gradients to measure DNA densities.

(*iii*) The DNA that was extracted from the culture, first generation (after 20 min) after the transfer from ^{15}N to ^{14}N medium had a hybrid or intermediate density. Means one strand of DNA had ^{15}N, while another strand had ^{14}N. DNA extracted from the culture after second generation (after 40 min) was composed of equal amounts of this hybrid DNA and of light DNA, i.e. one DNA molecule was of $^{15}N^{14}N$, while another DNA molecule was of $^{14}N^{14}N$. So, it explains semi-conservative method of DNA replication.

(*iv*) Very similar experiments were carried out by **Taylor** and **Colleagues** on *Vicia faba* (faba beans) using radioactive thymidine and the same results were obtained as in earlier experiments, i.e. DNA replicates semi-conservatively.

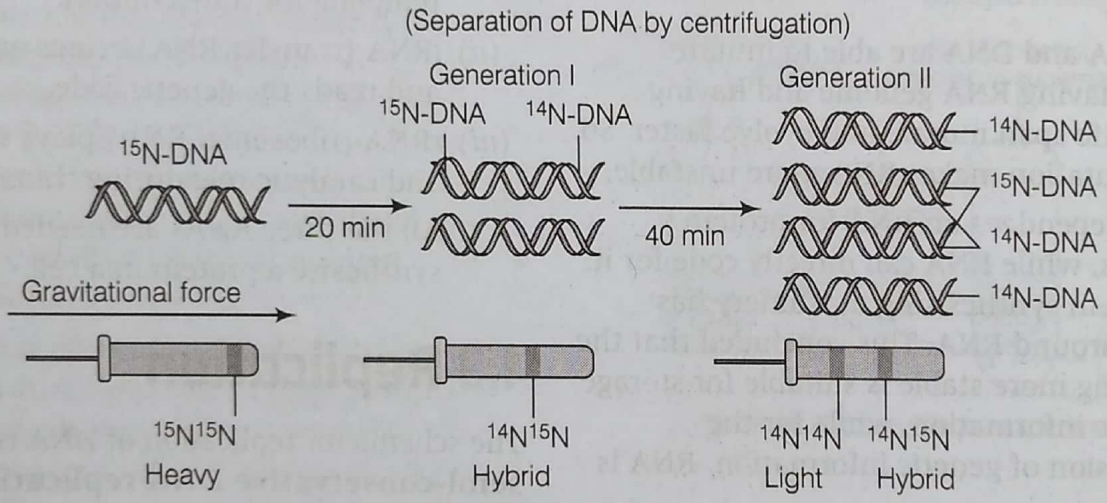

Figure 6.5 Meselson and Stahl's experiment

DNA Replication Machinery and Enzymes

The process of replication requires a set of catalysts (enzymes).

(*i*) DNA replication takes place in the S-phase of cell cycle.

(*ii*) The main enzyme is **DNA-dependent DNA polymerase**, since it uses a DNA template to catalyse the polymerisation of deoxynucleotides. The average rate of polymerisation by these enzymes is approximately 2000 bp/second.

(*iii*) These polymerases has to catalyse the reaction with high degree of accuracy because any mistake during replication would result into mutations.

(*iv*) DNA polymerisation is an energy demanding process, so deoxyribonucleoside triphosphates serve dual purposes, i.e. act as substrates and provide energy for polymerisation reaction.

(*v*) Many additional enzymes are also required in addition to DNA-dependent DNA polymerase such as helicase, topoisomerase, SSB, ligase, etc.

(*vi*) Both the strands of DNA get separated by the activity of enzyme **helicase**. So, it causes unwinding of DNA.

(*vii*) **Topoisomerase** enzyme removes tension from uncoiled DNA strand.

(*viii*) **SSB protein** stabilise uncoiled DNA strand so that it cannot coil again.

(*ix*) This uncoiling of DNA strands form a Y-shape structure is known as **replication fork.**

(*x*) (a) **Replication in DNA strand** occurs within a small opening of the DNA helix, known as **replication fork.**

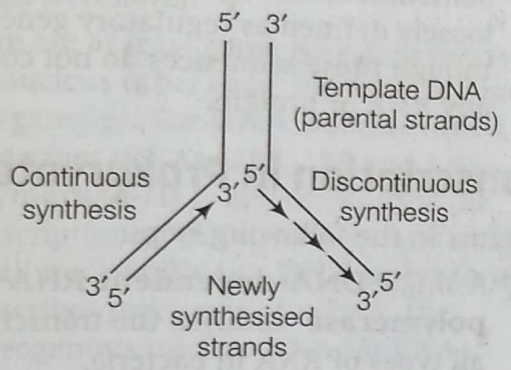

Figure 6.6 Replicating fork of DNA

(b) DNA polymerase cannot initiate the process of replication on their own, so it need a primer to start the process of DNA replication.

(c) DNA-dependent DNA polymerases catalyse polymerisation only in one direction, i.e. 5′→ 3′. It creates additional complications at the replicating fork. Consequently, on one strand (template 3′→ 5′), the replication is continuous. This is known as **leading strand**, while on the other strand (template 5′→ 3′), it is discontinuous. This is known as **lagging strand**. The discontinuously synthesised fragments called **Okazaki fragments** are later joined by **DNA ligase.**

Origin of Replication

(*i*) DNA polymerases cannot initiate the process of replication on their own. Also, replication does not initiate randomly at any place in DNA. So, there is a definite region in *E. coli* DNA, where the replication originates. This region is termed as **origin of replication.**

(*ii*) Due to this requirement, a piece of DNA, if needed to be propagated during recombinant DNA procedures, requires a vector. The vectors provide the origin of replication.

1.4 Transcription

It is the process of copying genetic information from one strand of the DNA into RNA. The principle of complementarity governs the process of transcription, except the adenosine now forms base pair with uracil instead of thymine.

(*i*) In transcription, only a segment of DNA is duplicated and only one of the strands is copied into RNA. Both the strands are not copied because

(a) If both the strands code for RNA, two different RNA molecules and two different proteins would be formed, hence complicating the genetic information transfer machinery.

(b) Since, two RNA produced would be complementary to each other, they would form a double-stranded RNA without translation, making the process of transcription futile.

(*ii*) A transcription unit in DNA is defined by three regions in the DNA, which are as follows:

(a) A promoter

(b) The structural gene

(c) A terminator

(*iii*) The two strands of DNA have opposite polarity and the **DNA-dependent RNA** polymerase also catalyse the polymerisation in only one direction that is 5′→ 3′.

(*iv*) The strand that has the polarity (3′→ 5′) acts as a template and is referred to as **template strand.** The other strand, which has the polarity (5′→ 3′) and the sequence same as RNA (T at the place of U) is displaced during transcription. This strand is called as **coding strand.** All the reference point while defining a transcription unit is made with coding strand.

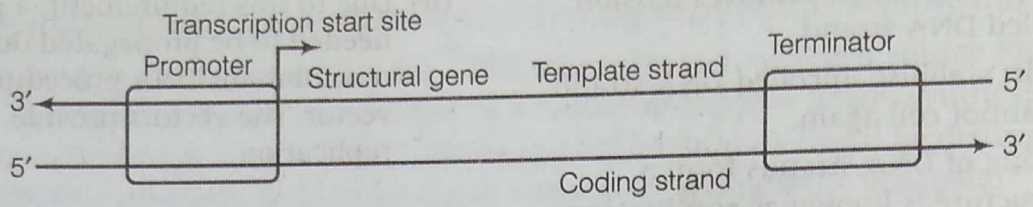

Figure 6.7 Schematic structure of a transcription unit

(*v*) The **promoter** and **terminator** flank the **structural gene** in a transcription unit.

(*vi*) The **promoter** is located towards 5′ end (upstream) of the structural gene.

(*vii*) It is the DNA sequence that provides binding site for RNA polymerase and the presence of promoter defines the template and coding strands. By switching its position with terminator, the definition of coding and template strands could be reversed.

(*viii*) The terminator is located towards 3′ end (downstream) of the coding strand and it usually defines the end of the process of transcription.

(*ix*) There are additional regulatory sequences that may be present further upstream or downstream to the promoter.

Transcription Unit and the Gene

(*i*) A gene can be defined as the functional **unit of inheritance.** Genes are located on the DNA.

(*ii*) A **cistron** is a segment of DNA coding for a polypeptide.

(*iii*) The structural gene in a transcription unit could be said as **monocistronic** (mostly in eukaryotes) or **polycistronic** (mostly in bacteria or prokaryotes). In monocistronic gene only one cistron is present in one transcription unit, while in polycistronic gene many cistrones are present. In eukaryotes monocistronic gene is splitted.

(*iv*) The coding sequences or expressed sequences are defined as **exons.** Exons appear in mature or processed RNA. The exons are interrupted by **introns**, i.e. the non-coding sequences.

(*v*) Introns or intervening sequences do not appear in mature or processed RNA.

(*vi*) Sometimes, the regulatory sequences are loosely defined as regulatory genes, even though these sequences do not code for any RNA or protein.

Transcription in Prokaryotes

It occurs in the following steps:

(*i*) A single **DNA-dependent RNA polymerase** catalyse the transcription of all types of RNA in bacteria.

(*ii*) RNA polymerase binds to promotor and **initiates transcription** (initiation).

(*iii*) It uses nucleoside triphosphates as substrate and polymerises in a template depended fashion following the rule of complementarity.

(*iv*) It also facilitates opening of the helix and continues elongation.

(*v*) Once the polymerase reaches the terminator region, the nascent RNA and RNA polymerase falls off. This results in **termination of transcription.**

(*vi*) RNA polymerase is only capable of catalysing the process of **elongation.** It associates transiently with **initiation-factor** (σ) and **terminator factor** (ρ), to initiate and terminate the process of transcription, respectively. In other words, RNA polymerase can identify promoter only when σ factor is attached on the promoter. Similarly it terminates the process of transcription when a ρ factor is attached on the terminator. Thus, catalysing all the three steps.

(*vii*) In bacteria, since the *m*RNA does not require any processing to become active and also since transcription and translation take place in the same compartment, many times the translation can begin much before the *m*RNA is fully transcribed. As a result, transcription and translation can be coupled in bacteria.

Transcription in Eukaryotes

Eukaryotes have additional complexities than prokaryotes as follows:

(*i*) There are at least three RNA polymerases in the nucleus other than the RNA polymerase in organelles. The **RNA polymerase-I** transcribes *r*RNAs (28S, 18S and 5.8S). **RNA polymerase-III** is responsible for the transcription of *t*RNA, 5*sr*RNA and *sn*RNAs (small nuclear RNAs). **RNA polymerase-II** transcribes precursor of *m*RNA, the heterogenous nuclear RNA (*hn*RNA).

(*ii*) Another complexity is that, the primary transcripts contain both the exons and the introns and are non-functional. Hence, it is subjected to a process called **splicing.** In this process, introns are removed and exons are joined in a definite order.

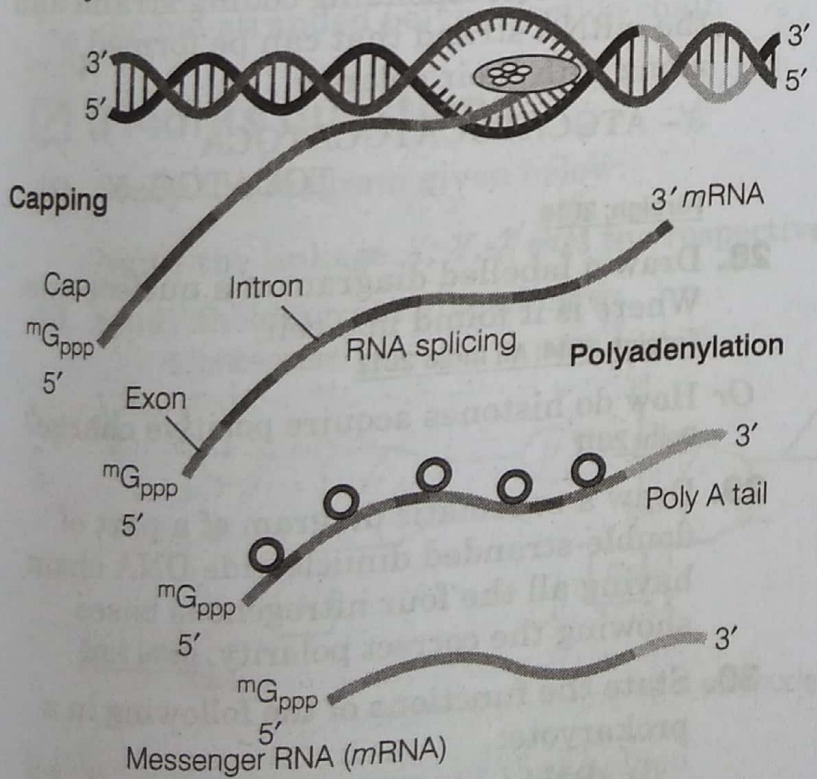

Figure 6.8 Process of transcription in eukaryotes

[TOPIC 2] Genetic Code, Human Genome Project and DNA Fingerprinting

2.1 Genetic Code

It is the relationship between the sequences of nucleotides on *m*RNA and the sequences of amino acids in the polypeptide chain. Genetic code is a sequence of three nucleotides present on *m*RNA, which codes for a specific amino acid during translation.

Deciphering the Code

(*i*) **George Gamow** a physicist suggested that the genetic code should be made up of three nucleotides. He stated that since there are only four bases and if they have to code for 20 amino acids, the code should constitute a combination of bases. But, a permutation combination of $4^3(4 \times 4 \times 4)$ would generate 64 codons, generating many more codons than required.

(*ii*) **Har Gobind Khorana** could synthesise RNA molecules with defined combinations of bases (homopolymers and copolymers).

(*iii*) **Marshall Nirenberg** made cell-free system for protein synthesis and finally the code was deciphered.

(*iv*) **Severo Ochoa enzyme** (polynucleotide phosphorylase) was also helpful in polymerising RNA with defined sequences in a template independent manner (enzyme synthesis of RNA).

(*v*) All of these investigations, finally helped to make a checker-board for genetic code as given below:

First position	Second position U	Second position C	Second position A	Second position G	Third position
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Figure 6.9 Codons for the various amino acids

The Salient Features of Genetic Code

These are as follows:

(*i*) It is a triplet, out of 64 codons 61 codons codes for amino acids and 3 codons do not code for any amino acids. Hence, they function as **stop codons**. [UAA, UAG and UGA]

(*ii*) One codon codes for only one amino acid, hence, it is **unambiguous** and **specific**. For example, CCG codes for proline, it will not code for any other amino acid.

(*iii*) Some amino acids are coded by more than one codon, hence the **code is degenerate.**

(*iv*) The codon is read in *m*RNA in a contiguous fashion. There are no punctuations.

(*v*) The code is nearly **universal**. For example, from bacteria to human, UUU would code for phenylalanine (Phe). Some exceptions are found in mitochondrial codons and in some protozoans.

(*vi*) **AUG** has dual functions. It codes for **methionine** (met) and also acts as **initiator codon**.

Mutations and Genetic Code

The sudden inheritable change in the genetic material is defined as mutation.

(*i*) **Point mutation** occurs due to change in single base pair. Its example is a change of single base pair in the gene for β-globin chain of haemoglobin that results in the change of amino acid residue glutamate to valine. It results in **sickle-cell anaemia**.

(*ii*) **Frameshift mutation** occurs where addition/insertion or deletion of one or two bases changes the reading frame from the site of mutation, resulting in a protein with a different set of amino acids. Insertion or deletion of three of its multiples of bases do not alter the reading frame, but one/more amino acids are coded in the protein gets translated.

(*iii*) **Silent mutation** occurs when a base change in a codon does not alter the amino acid coded. This forms the genetic basis of proof that codon is a triplet and it is read in a contiguous manner.

Francis Crick postulated the presence of an adapter molecule that would on one hand read the code and on other hand would bind to specific amino acids. This molecule is called ***t*RNA**. It may also be called as ***s*RNA**, i.e. soluble RNA as it is soluble in water. ***t*RNA** was known before the genetic code was postulated but its role as an adapter molecule was assigned much later.

(*i*) Clover leaf model of *t*RNA has an **anticodon loop** that has bases complementary to the code and also has an **amino acid acceptor end,** which it binds to amino acids. It also possesses an enzyme site and a ribosome site.

(*ii*) *t*RNAs are specific for each amino acid, they are **clover-shaped**.

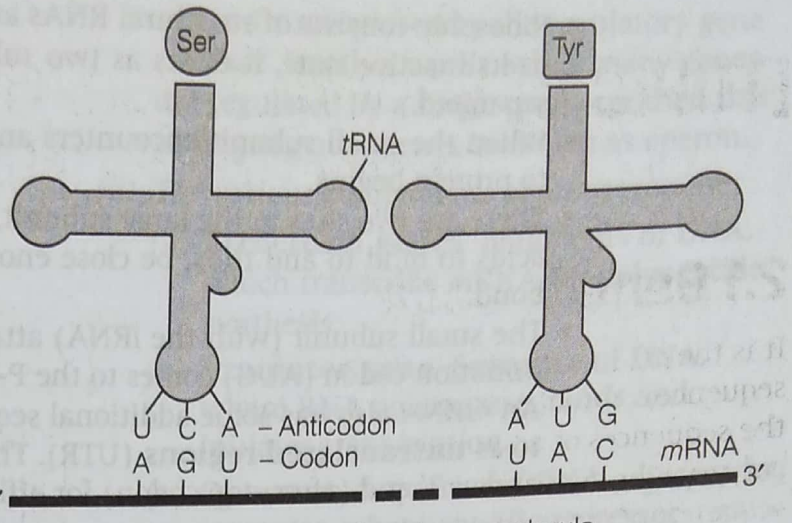

Figure 6.10 *t*RNA-the adapter molecule

(*iii*) For initiation, there is specific *t*RNA (initiator *t*RNA). There are no *t*RNAs for stop codons.

2.2 Translation

It is the process of polymerisation of amino acids to form a polypeptide.

(*i*) The order and sequence of amino acids are defined by the sequence of bases in the *m*RNA.

(*ii*) The amino acids are joined by a bond, which is known as a **peptide bond.** This process requires energy.

(*iii*) The process of translation requires the following components– *m*RNA, ribosome, amino acids as raw material, ATP, adaptor molecule *t*RNA and aminoacyl *t*RNA synthetase enzyme for charging of *t*RNA.

Phases of Translation

(*i*) **Activation of amino acids** occurs in the presence of ATP and link to their cognate *t*RNA, in the presence of aminoacyl *t*RNA synthatase enzyme, i.e. **charging of *t*RNA** or **aminoacylation of *t*RNA**. If two such charged *t*RNAs are brought closer, the formation of peptide bond between them would occur energetically in the presence of a catalyst.

(*ii*) **Initiation of polypeptide synthesis** occurs in ribosomes which is known as the cellular factory for protein synthesis.

- Ribosome consists of structural RNAs and about 80 different proteins.
- In its inactive state, it exists as two subunits, i.e. a large and a small subunit, both are separated.
- When the small subunit encounters an *m*RNA, the process of translation of the *m*RNA to protein begins.
- There are two sites in the large subunit, i.e. the **P-site** and **A-site** for subsequent amino acids to bind to and thus, be close enough to each other for the formation of a peptide bond.
- The small subunit (with the *t*RNA) attaches to the large subunit in such a way that the initiation codon (AUG) comes to the P-site.
- An *m*RNA also has some additional sequences that are not translated. These are referred to as **untranslated regions** (UTR). They are present at both 5′ end (before start codon) and at 3′ end (after stop codon) for efficient translation process.

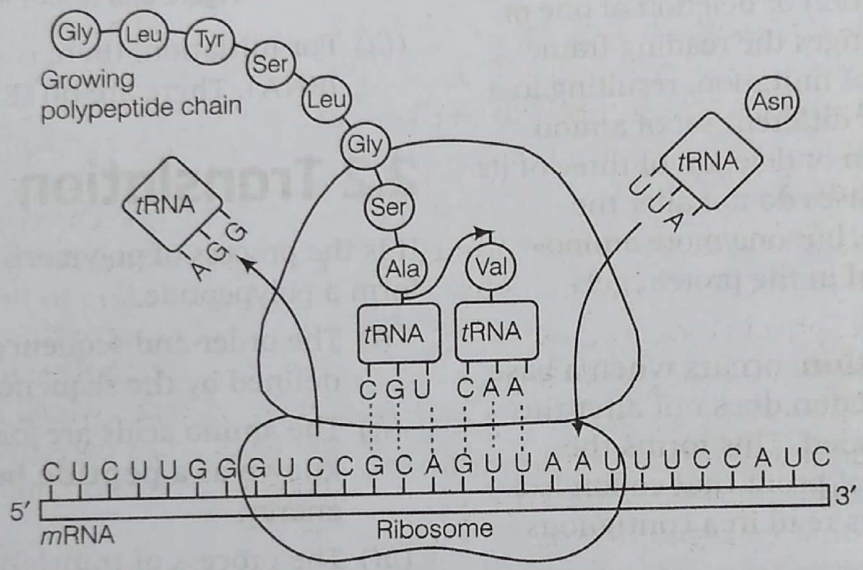

Figure 6.11 Translation

(*iii*) **Elongation of polypeptide chain** occurs when a second *t*RNA charged with an appropriate amino acid binds to the A-site of the ribosome.

- A peptide bond (CO—NH) forms between the carboxyl group of methionine and the amino group of the second amino acid. The reaction is catalysed by the enzyme **peptidyl transferase**.
- The complexes composed of an amino acid linked to *t*RNA, sequentially bind to the appropriate codon in *m*RNA by forming complementary base pairs with the *t*RNA anticodon.
- The ribosome moves from codon to codon along the *m*RNA in 5′ → 3′ direction. Amino acids are added one by one, translated into polypeptide sequences dictated by DNA and represented by *m*RNA.

(*iv*) **Termination of polypeptide synthesis** occur when a **release factor** binds to the stop codon. As a result, the polypeptide synthesis or elongation process stops, releasing the complete polypeptide from the ribosome.

2.3 Regulation of Gene Expression

It occurs at various levels. It results in the formation of a polypeptide.

(*i*) **In prokaryotes,** gene expression is regulated by the rate of initiation of transcription.

(*ii*) **In eukaryotes,** regulation is achieved at four levels:

(a) Transcriptional level (formation of primary transcript).

(b) Processing levels (regulation of splicing).

(c) Transport of *m*RNA from nucleus to the cytoplasm.

(d) Translational level.

(*iii*) Genes in a cell are expressed to perform a particular function or a set of functions.

(*iv*) The metabolic, physiological or environmental conditions regulate expression of genes.

(*v*) The development and differentiation of embryo into adult organisms are also a result of coordinated regulation of expression of several sets of genes.

(*vi*) **In a transcriptional unit,** the activity of RNA polymerase at a given promoter is in turn regulated by the interaction with accessory proteins.

(*vii*) The accessibility of promoter regions of prokaryotic DNA in many cases is regulated by the interaction of proteins with sequences termed as **operators**.

(*viii*) The sequences of the operator bind a repressor protein. Each operon has its specific operator and specific repressor. For example, *lac* operon interacts with *lac* repressor only.

F Jacob and **J Monod** were the first to describe a transcriptionally regulated system.

(*i*) An **operon** is a unit of **prokaryotic gene expression**, which includes sequentially regulated (structural) genes and control elements recognised by the regulatory gene product. In other words, when many genes are regulated by a single promoter then this arrangement of genes is known as operon.

(*ii*) The various components of an operon are:

(a) **Structural genes** Fragments of DNA, which transcribe *m*RNA for polypeptide synthesis.

(b) **Promoter gene** Sequence of DNA, where RNA polymerase, binds and initiates transcription.

(c) **Operator** Sequence of DNA adjacent to promoter, where specific repressor protein binds.

(d) **Regulator gene** Codes for the repressor protein that binds to the operator and suppresses its activity, hence transcription does not occur. Also, represented as 'i' gene.

(e) **Inducer** Prevents the repressor from binding to the operator. Due to this, transcription is switched on. It may be a metabolite, hormone, etc.

In ***lac* operon,** which is related to lactose metabolism, a polycistronic structural gene is regulated by a common promoter and regulatory genes.

(*i*) The *lac* operon consists of one regulatory gene (i gene) and three structural genes (z, y and a).

(a) i—for repressor of *lac* operon

z—for beta-galactosidase (β-gal) that catalyses the hydrolysis of lactose into galactose and glucose.

y—for permease, increases the permeability of the cell.

a—for transacetylase.

(b) All the three gene products in *lac* operon are required for the metabolism of lactose.

(c) For bacteria, when glucose is present in the nutrition media *lac* operon is closed. But when glucose is replaced by lactose, then it needs β-galactosidase enzyme to break lactose into glucose and galactose, hence *lac* operon is activated.

(*ii*) **Lactose** is a substrate for enzyme β-galactosidase and it regulates switching on and off of the operon, hence termed as **inducer**.

(*iii*) The lactose induces operon in the following ways:

(a) When glucose is present in the medium, **repressor** of the operon is synthesised from the i gene.

(b) Repressor protein binds to the operator region of the operon, so RNA polymerase cannot bind on promoter and transcription does not occur. Hence, β-galactosidase is not formed.

(c) In presence of an inducer, such as lactose or allolactose, the repressor is inactivated by interaction with the inducer. So, now inactivated repressor cannot bind on operator. This allows RNA polymerase access to the promoter and transcription proceeds. Means all the genes are transcribed and all three enzymes are formed.

(d) **Regulation of *lac* operon** by a repressor is referred to as **negative regulation.**

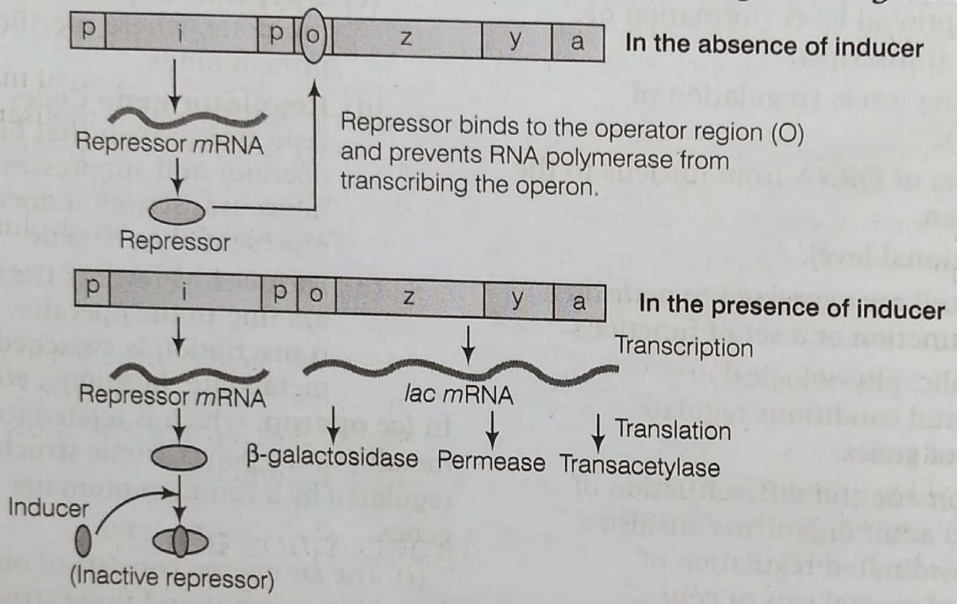

Figure 6.12 *Lac* operon

2.4 Human Genome Project (HGP)

It is a 13 year project coordinated by the US Department of Energy and the National Institute of Health. It was launched in the year 1990 and completed in 2003. It was closely associated with rapid development of a new area in biology called **Bioinformatics**. HGP was called as a megaproject because:

(*i*) total cost of the project was approximately 9 billion US dollars.

(*ii*) to store the information, it will require 3300 books of 1000 pages each, if 1000 letters are written on each page.

(*iii*) it needs high speed computational devices to store, retrieve and analyse the information.

The important **goals of HGP** are to:

(*i*) identify all the approximately 20000-25000 genes in human DNA.

(*ii*) determine the sequences of the 3 billion chemical base pairs that make up human DNA.

(*iii*) store this information in databases.

(*iv*) improve tools for data analysis.

(*v*) transfer related technologies to other sectors, such as industries.

(*vi*) address the Ethical, Legal and Social Issues (ELSI) that may arise from the project.

Methodologies of HGP

These are focussed on two main lines, i.e. expressed sequence tags and sequence annotation.

(*i*) **Expressed Sequence Tags** (ESTs) method is based on identifying all the genes that are expressed as RNA.

(*ii*) **Sequence annotation** is the approach of simply sequencing the whole set of genome that contains all the coding and non-coding sequences and later assigning different regions in the sequence with functions.

(a) For sequencing, the total DNA from cell is isolated and converted into relatively smaller size as fragments.

(b) DNA fragments are cloned in suitable host using specialised vectors, such as Bacterial Artificial Chromosome (BAC) and Yeast Artificial Chromosome (YAC).

(c) **Fragments of DNA** are then sequenced by automated DNA sequences, which work on principle developed by **F Sanger**.

(d) These sequences are arranged accordingly on the basis of overlapping regions on DNA fragments.

(e) The alignments of these sequences based on computer programmes are developed.

(f) At last, the genetic and physical maps of the genome are constructed by collecting information about certain repetitive DNA sequences and DNA polymorphism.

Salient Features of Human Genome

These are as follows:

(*i*) The human genome contains 3164.7 million nucleotide bases.

(*ii*) The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin of 2.4 million bases.

(*iii*) The total number of genes is estimated at 30000, much lower than previous estimates of 80000-140000 genes. Almost all, (99.9%) nucleotide bases are exactly the same in all people.

(*iv*) The functions are unknown for over 50% of the discovered genes.

(*v*) Less than 2% of the genome codes for proteins.

(*vi*) Repeated sequences make up very large portion of the human genome.

(*vii*) Repetitive sequences are stretches of DNA sequences that are repeated many times, sometimes hundred to thousand times. They are thought to have no direct coding functions, but shed light on chromosome structure, dynamics and evolution.

(*viii*) Chromosome 1 has most genes (2968) and the Y has the fewest (231).

(*ix*) Scientists have identified about 1.4 million locations, where single base DNA differences, **Single Nucleotide Polymorphisms** (SNPs) occur in humans.

Applications of HGP

(*i*) Knowledge from DNA sequences will define research leading to our knowledge of biological systems.

(*ii*) It will enable a radically new approach to biological research.

(*iii*) It can be used to study genes in a genome.

(*iv*) It can create new ways to diagnose, treat and prevent the disorders that affect humans.

(*v*) Many non-human model organisms such as *C. elegans*, *Drosophila*, rice, *Arabidopsis* have also been sequenced.

2.5 DNA Fingerprinting

It is a quick way to compare the DNA sequences of any two individuals.

(*i*) **Dr. Alec Jeffreys** developed the technique of DNA fingerprinting in an attempt to identify DNA marker for the inherited diseases.

(*ii*) DNA fingerprinting involves identifying differences in some specific regions in DNA sequence called as **repetitive DNA**, i.e. a small stretch of DNA is repeated many time.

(*iii*) The repetitive DNAs 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**.

(*iv*) Satellite DNA can be classified as microsatellites, minisatellites, etc. depending on base composition, the length of segment and number of repetitive units.

(*v*) These sequences show high degree of polymorphism and form the basis of DNA fingerprinting.

(*vi*) Since DNA from every tissue (as blood, hair, etc.) from an individual differs, they become very useful tool in forensic applications.

(*vii*) As the polymorphisms are inheritable, DNA fingerprinting is the basis of paternity testing as well.

(*viii*) Polymorphism (variation at genetic level) arises due to mutations. If an inheritable mutation is observed in a population of high frequency, it is called **DNA polymorphism.**

(*ix*) The mutations keep on accumulating generation after generation and form one of the basis of variability/polymorphism.

(*x*) There is a variety of different types of polymorphism ranging from single nucleotide change to very large scale changes.

Methodology of DNA Fingerprinting

This technique involves Southern blot hybridisation, using radiolabelled VNTR as a probe.

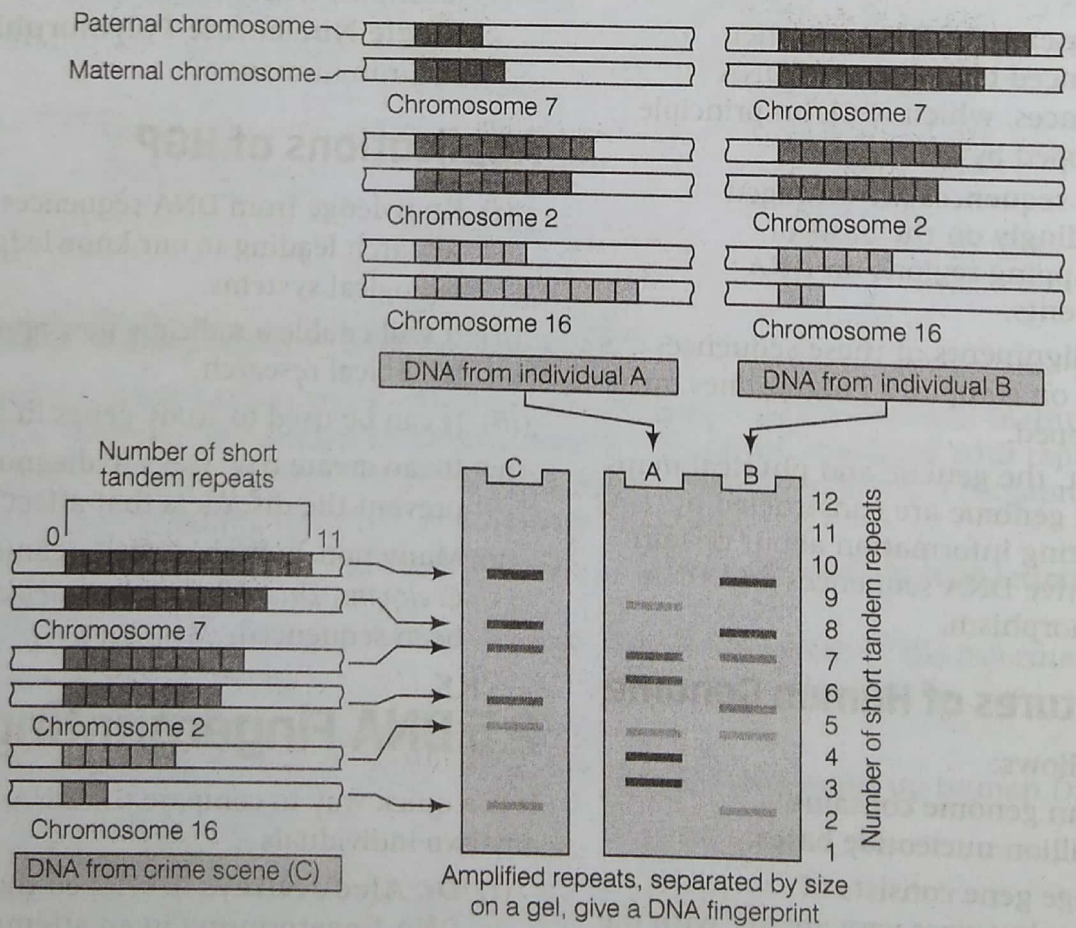

Figure 6.13 Schematic representation of DNA fingerprinting

The methodology includes:

(*i*) DNA is isolated and digested by the **restriction endonucleases**.

(*ii*) DNA fragments are separated by **electrophoresis**.

(*iii*) Separated DNA fragments are transferred to synthetic membranes like **nitrocellulose** or **nylon.**

(*iv*) Hybridisation using labelled VNTR probe.

(*v*) Hybridised DNA fragments are detected by **autoradiography**.

The sensitivity has been increased by the use of Polymerase Chain Reaction (PCR).

Variable Number of Tandem Repeats (VNTRs) belongs to a class of satellite DNA called **minisatellite.**

(*i*) The chromosome number repeats show very high degree of polymorphism. Due to this, the size of VNTR varies in size from 0.1-20 kb.

(*ii*) This DNA differs in individual to individual except in case of monozygotic (identical) twins.

(*iii*) Consequently, after hybridisation with VNTR probe, the autoradiogram gives many bands of different sizes. These bands give characteristic pattern of an individual DNA.

Applications of DNA Fingerprinting

(*i*) It is useful as identification tool in forensic applications.

(*ii*) It is the basis of paternity testing in case of disputes.

(*iii*) It is used in determining population and genetic diversities and also in evolutionary biology.