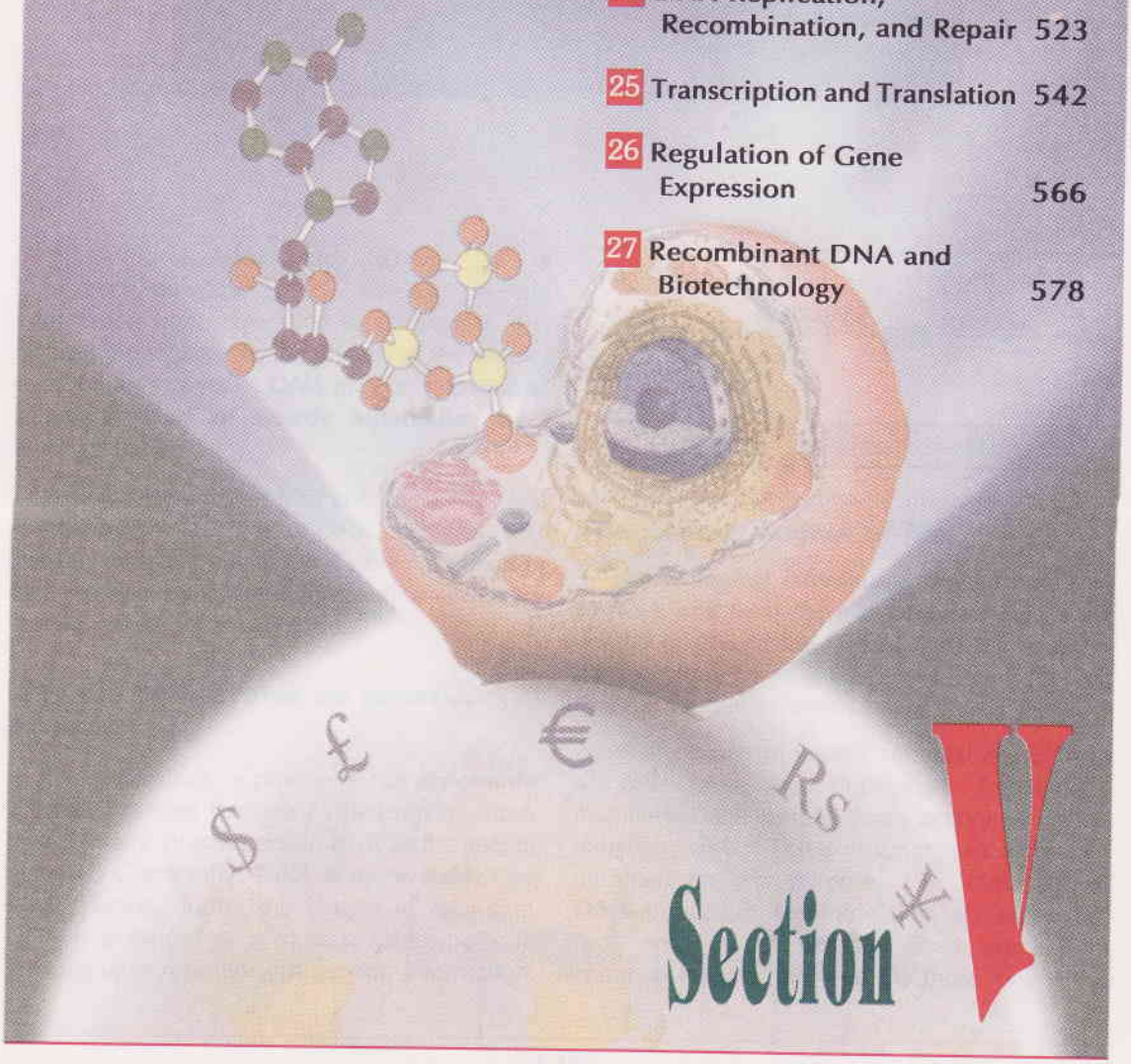
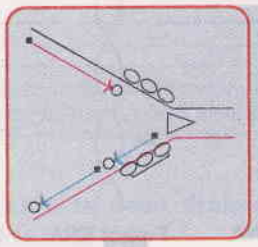
MOLECULAR BIOLOGY AND BIOTECHNOLOGY



Section V

24 DNA–Replication, Recombination, and Repair



The hereditary molecule, DNA speaks :

"I replicate and recombinate,
To permit the cells to proliferate,
Environmental insults try to damage me,
But I protect myself with adequate repairs"

Deoxyribonucleic acid (DNA) is a macromolecule that carries genetic information from generation to generation. It is responsible to preserve the identity of the species over millions of years. ***DNA*** may be regarded as a ***reserve bank of genetic information*** or a memory bank.

A single mammalian fetal cell contains only a few picograms (10^{-12} g) of DNA. It is surprising that this little quantity of DNA stores information that will determine the differentiation and every function of an adult animal.

Why did DNA evolve as genetic material?

RNA molecules, in principle, can perform the cellular functions that are carried out by DNA. In fact, many viruses contain RNA as the genetic material. Chemically, ***DNA is more stable*** than RNA. Hence, during the course of evolution, DNA is preferred as a more suitable molecule for long-term repository of genetic information.

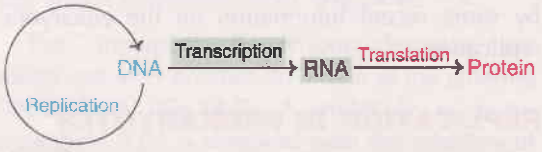


Fig. 24.1 : The central dogma of life.

The central dogma of life

The ***biological information flows from DNA to RNA, and from there to proteins.*** This is the central dogma of life (***Fig.24.1***). It is ultimately the DNA that controls every function of the cell through protein synthesis.

As the carrier of genetic information, DNA in a cell must be duplicated (replicated), maintained and passed down accurately to the daughter cells. Three distinct processes are designed for this purpose. The '***three Rs' of DNA-r***eplication, ***r***ecombination, and ***r***epair, are dealt with in this chapter. There are certain common features between the three Rs.

- They act on the same substrate (DNA).
- They are primarily concerned with the making and breaking of phosphodiester bonds (the backbone of DNA structure).
- Enzymes used in the three processes are mostly similar/comparable.

REPLICATION OF DNA

DNA is the genetic material. When the cell divides, the daughter cells receive an identical copy of genetic information from the parent cell.

Replication is a ***process in which DNA copies itself to produce identical daughter molecules of DNA***. Replication is carried out with high fidelity which is essential for the survival of the species. Synthesis of a new DNA molecule is a complex process involving a series of steps.

The salient features of replication in prokaryotes are described first. This is followed by some recent information on the eukaryotic replication.

REPLICATION IN PROKARYOTES

Replication is semiconservative

The parent DNA has two strands complementary to each other. Both the strands undergo simultaneous replication to produce two daughter molecules. Each one of the newly synthesized DNA has one-half of the parental DNA (one strand from original) and one-half of new DNA (***Fig.24.2***). This type of replication is known as semiconservative since ***half of the original DNA is conserved in the daughter DNA.*** The first experimental evidence for the semiconservative DNA replication was provided by Meselson and Stahl (1958).

Initiation of replication

The initiation of DNA synthesis occurs at a site called ***origin of replication***. In case of prokaryotes, there is a single site whereas in eukaryotes, there are multiple sites of origin.

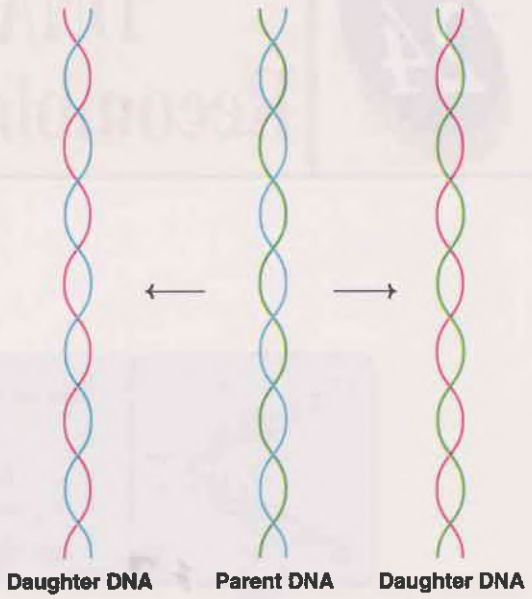


Fig. 24.2 : DNA replication—semiconservative.

These sites mostly consist of a short sequence of A-T base pairs. A specific protein called ***dna A*** (20-50 monomers) binds with the site of origin for replication. This causes the double-stranded DNA to separate.

Replication bubbles

The two complementary strands of DNA separate at the site of replication to form a bubble. Multiple replication bubbles are formed in eukaryotic DNA molecules, which is essential for a rapid replication process (***Fig.24.3***).

RNA primer

For the synthesis of new DNA, a short fragment of RNA (about 5-50 nucleotides, variable with species) is required as a primer. The enzyme ***primase*** (a specific RNA polymerase) in association with single-stranded binding proteins forms a complex called ***primosome***, and produces RNA primers. A constant synthesis and supply of RNA primers should occur on the lagging strand of DNA. This is in contrast to the leading strand which has almost a single RNA primer.

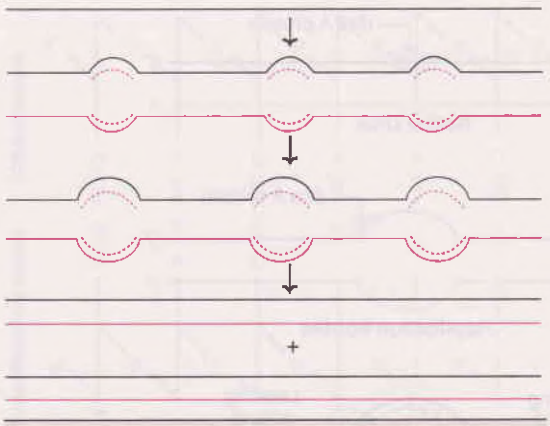


Fig. 24.3 : Schematic representation of multiple replication bubbles in DNA replication.

DNA synthesis is semidiscontinuous and bidirectional

The replication of DNA occurs in 5′ to 3′ direction, simultaneously, on both the strands of DNA. On one strand, the ***leading (continuous or forward) strand—the DNA synthesis is continuous***. On the other strand, the ***lagging (discontinuous or retrograde) strand—the synthesis of DNA is discontinuous***. Short pieces of DNA (15-250 nucleotides) are produced on the lagging strand.

In the replication bubble, the DNA synthesis occurs in both the directions (bidirectional) from the point of origin.

Replication fork and DNA synthesis

The separation of the two strands of parent DNA results in the formation of a replication fork. The active synthesis of DNA occurs in this region. The replication fork moves along the parent DNA as the daughter DNA molecules are synthesized.

DNA helicases : These enzymes bind to both the DNA strands at the replication fork. Helicases move along the DNA helix and separate the strands. Their function is comparable with a ***zip opener***. Helicases are dependent on ATP for energy supply.

Single-stranded DNA binding (SSB) proteins : These are also known as DNA helix-destabilizing proteins. They possess no enzyme activity. SSB proteins bind only to single-stranded DNA (separated by helicases), keep the two strands separate and provide the template for new DNA synthesis. It is believed that SSB proteins also protect the single-stranded DNA degradation by nucleases.

DNA synthesis catalysed by DNA polymerase III

The synthesis of a new DNA strand, catalysed by DNA polymerase III, occurs in 5′→3′ direction. This is antiparallel to the parent template DNA strand. The presence of all the four deoxyribonucleoside triphosphates (dATP, dGTP, dCTP and dTTP) is an essential prerequisite for replication to take place.

The synthesis of two new DNA strands, simultaneously, takes place in the opposite direction—one is in a direction (5′→3′) towards the replication fork which is continuous, the other in a direction (5′→3′) away from the replication fork which is discontinuous (***Fig.24.4***).

The incoming deoxyribonucleotides are added one after another, to 3′ end of the growing DNA chain (***Fig.24.5***). A molecule of pyrophosphate (PPi) is removed with the addition of each nucleotide. The template DNA strand (the parent) determines the base sequence of the newly synthesized complementary DNA.

Polarity problem

The DNA strand (leading strand) with its 3′-end (3′-OH) oriented towards the fork can be elongated by sequential addition of new nucleotides. The other DNA strand (lagging strand) with 5′-end presents some problem, as there is no DNA polymerase enzyme (in any organism) that can catalyse the addition of nucleotides to the 5′ end (i.e. 3′→5′ direction) of the growing chain. This problem however is solved by synthesizing this strand as a series of small fragments. These pieces are made in the normal 5′→3′ direction, and later joined together.

Okazaki pieces : The small ***fragments of the discontinuously synthesized DNA*** are called

Okazaki pieces. These are produced on the lagging strand of the parent DNA. Okazaki pieces are later joined to form a continuous strand of DNA. DNA polymerase I and DNA ligase are responsible for this process (details given later).

Proof-reading function of DNA polymerase III

Fidelity of replication is the most important for the very existence of an organism. Besides its 5′→3′ directed catalytic function, DNA polymerase III also has a ***proof-reading activity.*** It checks the incoming nucleotides and allows only the correctly matched bases (i.e. complementary bases) to be added to the growing DNA strand. Further, DNA polymerase edits its mistakes (if any) and removes the wrongly placed nucleotide bases.

Replacement of RNA primer by DNA

The synthesis of new DNA strand continues till it is in close proximity to RNA primer. Now the DNA polymerase I comes into picture. It removes the RNA primer and takes its position. DNA polymerase I catalyses the synthesis (5′→3′ direction) of a fragment of DNA that replaces RNA primer (***Fig.24.6***).

The enzyme ***DNA ligase*** catalyses the formation of a phosphodiester linkage between the DNA synthesized by DNA polymerase III and the small fragments of DNA produced by DNA polymerase I. This process—nick sealing-requires energy, provided by the breakdown of ATP to AMP and PPi.

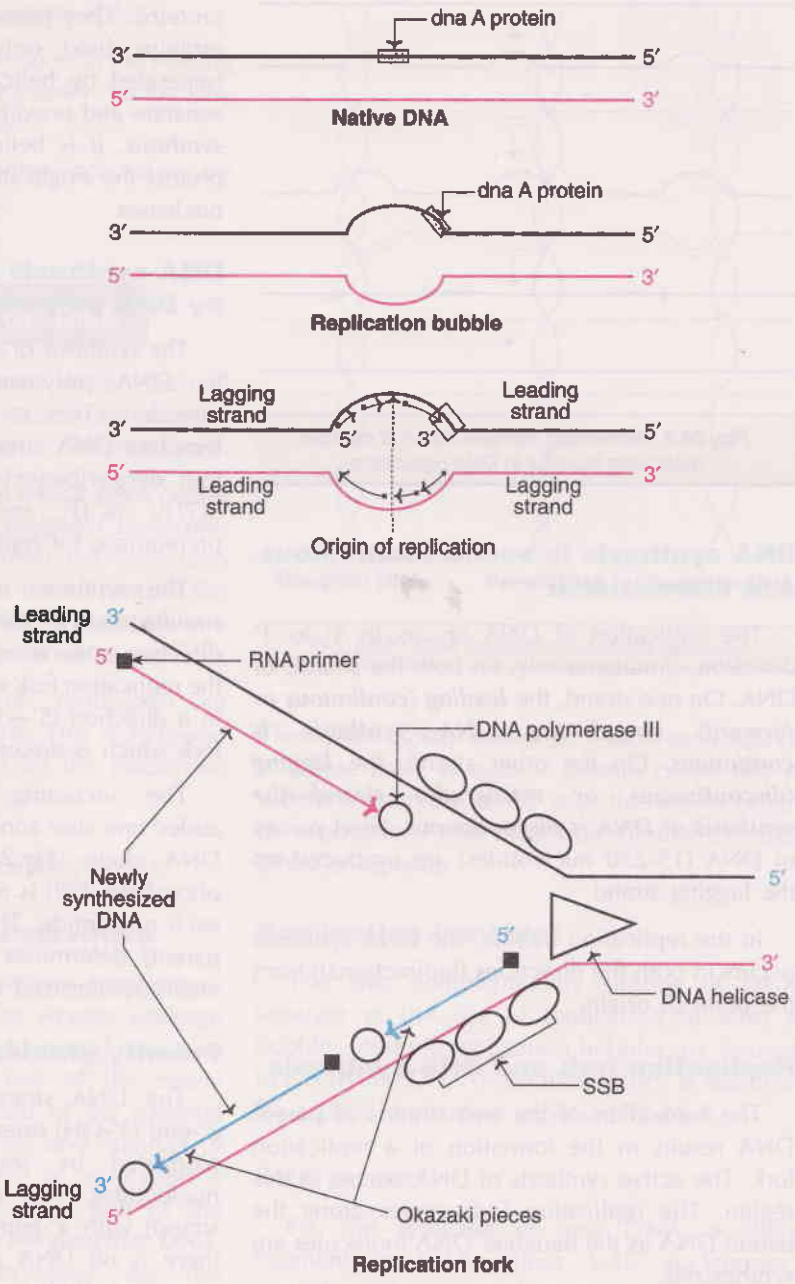


Fig. 24.4 : Overview of DNA replication process (SSB–Single-stranded binding protein).

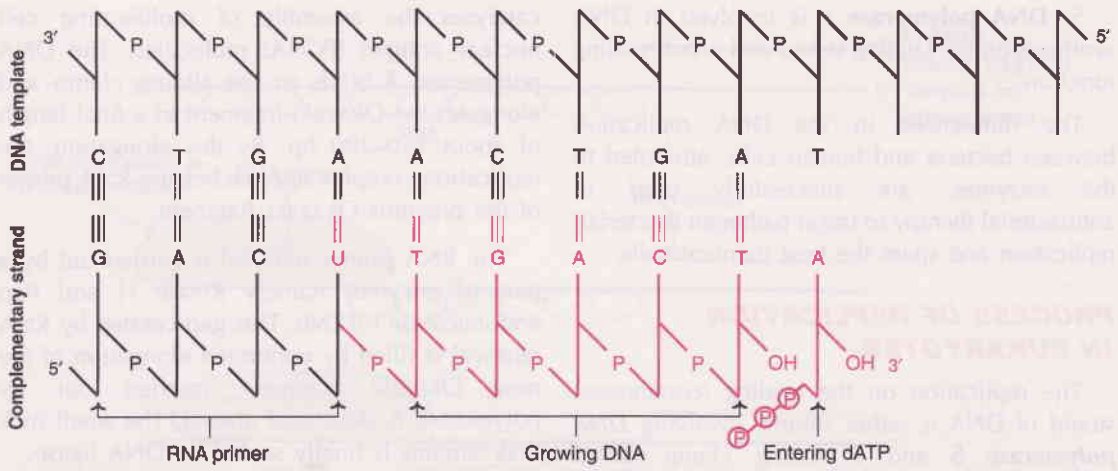


Fig. 24.5 : DNA replication with a growing complementary strand.

Another enzyme—***DNA polymerase II***—has been isolated. It participates in the ***DNA repair process.***

Supercoils and DNA topoisomerases

As the double helix of DNA separates from one side and replication proceeds, supercoils are formed at the other side. The formation of supercoils can be better understood by comparing DNA helix with two twisted ropes tied at one end. Hold the ropes at the tied end in a fixed position. And let your friend pull the ropes apart from the other side. The formation of supercoils is clearly observed.

The problem of supercoils that comes in the way of DNA replication is solved by a group of enzymes called DNA topoisomerases. Type I DNA topoisomerase cuts the single DNA strand (nuclease activity) to overcome the problem of supercoils and then reseals the strand (ligase activity). Type II DNA topoisomerase (also known as DNA gyrase) cuts both strands and reseals them to overcome the problem of supercoils.

REPLICATION IN EUKARYOTES

Replication of DNA in eukaryotes closely resembles that of prokaryotes. Certain differences, however, exist. ***Multiple origins of replication*** is a characteristic feature of eukaryotic cell. Further, at least ***five distinct DNA polymerases*** are known in eukaryotes. Greek letters are used to number these enzymes.

1. **DNA polymerase** α is responsible for the synthesis of RNA primer for both the leading and lagging strands of DNA.

2. **DNA polymerase** β is involved in the repair of DNA. Its function is comparable with DNA polymerase I found in prokaryotes.

3. **DNA polymerase** γ participates in the replication of mitochondrial DNA.

4. **DNA polymerase** δ is responsible for the replication on the leading strand of DNA. It also possesses proof-reading activity.

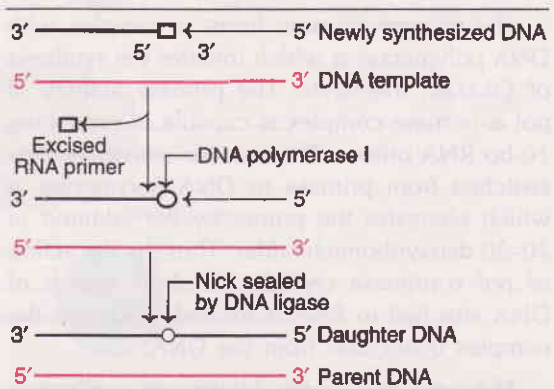


Fig. 24.6 : Overview of the action of DNA polymerase I and DNA ligase.

5. **DNA polymerase** ε is involved in DNA synthesis on the lagging strand and proof-reading function.

The differences in the DNA replication between bacteria and human cells, attributed to the enzymes, are successfully used in antibacterial therapy to target pathogen (bacterial) replication and spare the host (human) cells.

PROCESS OF REPLICATION IN EUKARYOTES

The replication on the leading (continuous) strand of DNA is rather simple, involving ***DNA polymerase*** δ and a sliding clamp called ***proliferating cell nuclear antigen (PCNA)***. PCNA is so named as it was first detected as an antigen in the nuclei of replicating cells. PCNA forms a ring around DNA to which DNA polymerase δ binds. Formation of this ring also requires another factor namely ***replication factor C*** (***RFC***).

The ***replication on the lagging*** (discontinuous) ***strand in eukaryotes is more complex*** when compared to prokaryotes or even the leading strand of eukaryotes. This is depicted in ***Fig.24.7***, and briefly described hereunder.

The parental strands of DNA are separated by the enzyme helicase. A single-stranded DNA binding protein called ***replication protein A*** (***RPA***) binds to the exposed single-stranded template. This strand has been opened up by the replication fork (a previously formed Okazaki fragment with an RNA primer is also shown in ***Fig.24.4***).

The enzyme primase forms a complex with DNA polymerase α which initiates the synthesis of Okazaki fragments. The primase activity of pol α-primase complex is capable of producing 10-bp RNA primer. The enzyme activity is then switched from primase to DNA polymerase α which elongates the primer by the addition of 20–30 deoxyribonucleotides. Thus, by the action of pol α-primase complex, a short stretch of DNA attached to RNA is formed. And now the complex dissociates from the DNA.

The next step is the binding of replication factor C (RFC) to the elongated primer (short RNA-DNA). RFC serves as a clamp loader, and catalyses the assembly of proliferating cell nuclear antigen (PCNA) molecules. The DNA polymerase δ binds to the sliding clamp and elongates the Okazaki fragment to a final length of about 150–200 bp. By this elongation, the replication complex approaches the RNA primer of the previous Okazaki fragment.

The RNA primer removal is carried out by a pair of enzymes namely RNase H and flap endonuclease I (FENI). This gap created by RNA removal is filled by continued elongation of the new Okazaki fragment (carried out by polymerase δ, described above). The small nick that remains is finally sealed by DNA ligase.

Eukaryotic DNA is tightly bound to histones (basic proteins) to form nucleosomes which, in turn, organize into chromosomes. During the course of replication, the chromosomes are relaxed and the nucleosomes get loosened. The DNA strands separate for replication, and the parental histones associate with one of the parental strands. As the synthesis of new DNA strand proceeds, histones are also produced simultaneously, on the parent strand. At the end of replication, of the two daughter chromosomal DNAs formed, one contains the parental histones while the other has the newly synthesized histones.

INHIBITORS OF DNA REPLICATION

Bacteria contain a specific type II topoisomerase namely ***gyrase***. This enzyme cuts and reseals the circular DNA (of bacteria), and thus overcomes the problem of supercoils. Bacterial gyrase is inhibited by the antibiotics ciprofloxacin, novobiocin and nalidixic acid. These are widely used as antibacterial agents since they can effectively block the replication of DNA and multiplication of cells. These antibacterial agents have almost no effect on human enzymes.

Certain compounds that ***inhibit human topoisomerases*** are used as anticancer agents e.g. adriamycin, etoposide, doxorubicin. The nucleotide analogs that inhibit DNA replication are also used as anticancer drugs e.g. 6-mercaptopurine, 5-fluorouracil.

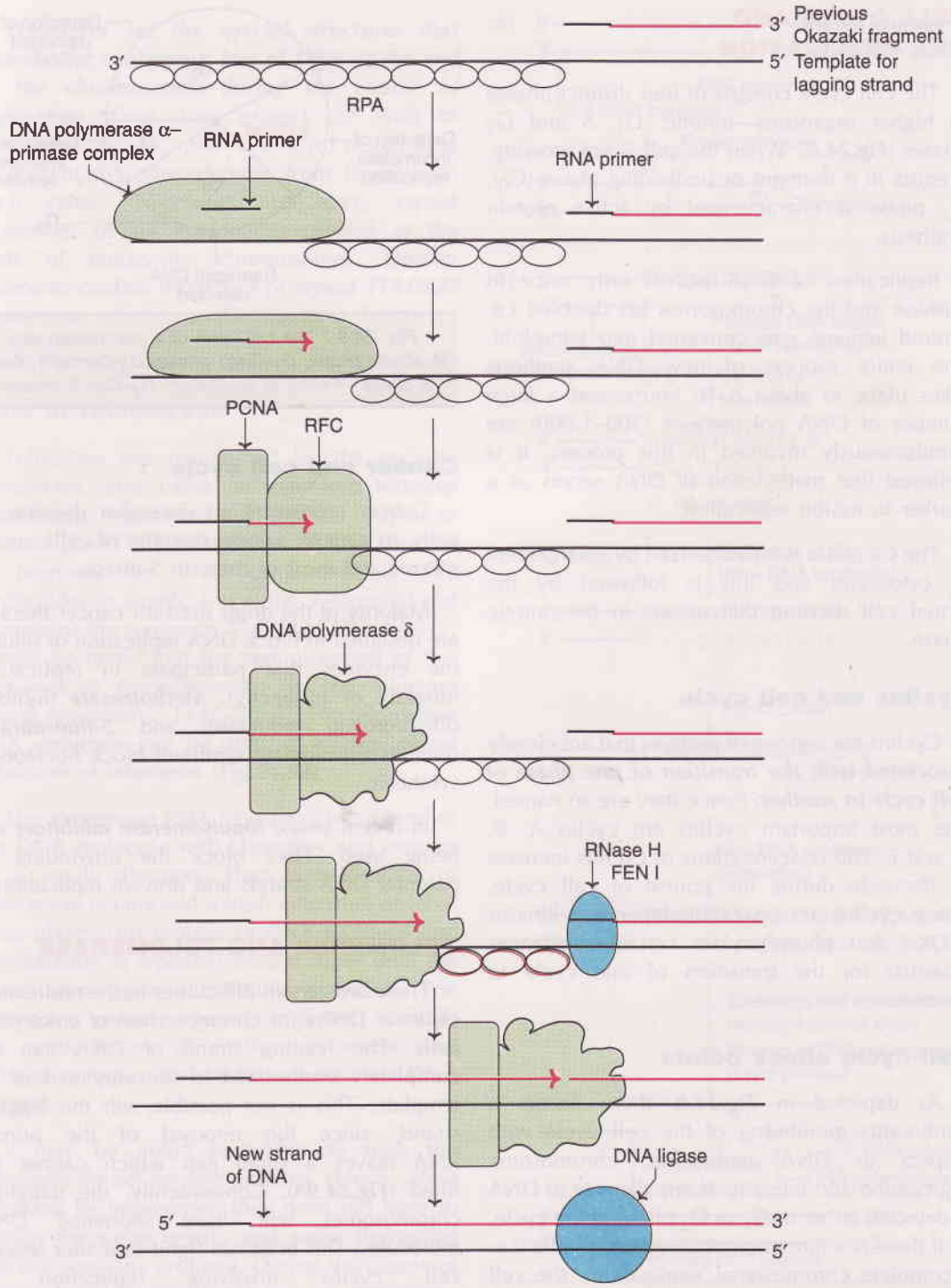


Fig. 24.7 : *An outline of DNA replication on the lagging strand in eukaryotes (RPA–Replication protein A; PCNA–Proliferating cell nuclear antigen; RFC–Replication factor C; RNase H–Ribonuclease H; FENI–Flap endonuclease I;* ***Note :*** *Leading strand not shown).*

CELL CYCLE AND DNA REPLICATION

The cell cycle consists of four distinct phases in higher organisms—mitotic, G_1, S and G_2 phases (***Fig.24.8***). When the cell is not growing, it exists in a dormant or undividing phase (G_0). G_1 phase is characterized by active protein synthesis.

Replication of DNA occurs only once in S-phase and the chromosomes get doubled i.e. diploid genome gets converted into tetraploid. The entire process of new DNA synthesis takes place in about 8–10 hours, and a large number of DNA polymerases (500–1,000) are simultaneously involved in this process. It is believed that methylation of DNA serves as a marker to inhibit replication.

The G_2 phase is characterized by enlargement of cytoplasm and this is followed by the actual cell division that occurs in the mitotic phase.

Cyclins and cell cycle

Cyclins are a group of proteins that are closely ***associated with the transition of one phase of cell cycle to another***, hence they are so named. The most important cyclins are cyclin A, B, D and E. The concentrations of cyclins increase or decrease during the course of cell cycle. These cyclins act on cyclin-dependent kinases (CDKs) that phosphorylate certain substances essential for the transition of one cycle to another.

Cell cycle check points

As depicted in ***Fig.24.8***, there occurs a continuous monitoring of the cell cycle with respect to DNA replication, chromosome segregation and integrity. If any damage to DNA is detected either in G_1 or G_2 phase of the cycle, or if there is a formation of defective spindle (i.e. incomplete chromosomal segregation), the cell cycle will not progress until appropriately corrected. If it is not possible to repair the damage done, the cells undergo ***apoptosis*** (programmed cell death).

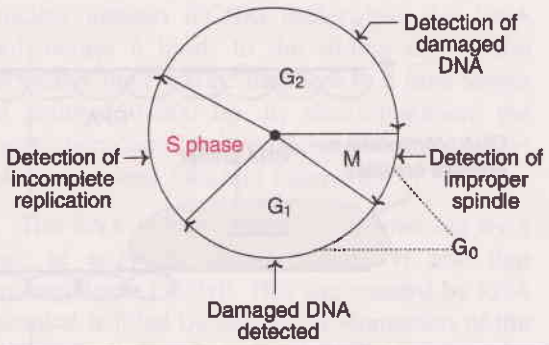


Fig. 24.8 : The cell cycle of a mammalian cell (M–Mitotic phase; G_1–Gap1 phase; G_0–Dormant phase; S phase – Period of replication; G_2–Gap 2 phase).

Cancer and cell cycle

Cancer represents an excessive division of cells. In cancer, a large quantity of cells are in mitosis and most of them in S-phase.

Majority of the drugs used for cancer therapy are designed to block DNA replication or inhibit the enzymes that participate in replication (directly or indirectly). ***Methotrexate*** (inhibits dihydrofolate reductase) and ***5-fluorouracil*** (inhibits thymidylate synthase) block nucleotide synthesis.

In recent years, ***topoisomerase inhibitors*** are being used. They block the unwinding of parental DNA strands and prevent replication.

TELOMERES AND TELOMERASE

There are certain difficulties in the replication of linear DNAs (or chromosomes) of eukaryotic cells. The leading strand of DNA can be completely synthesized to the very end of its template. This is not possible wih the lagging strand, since the removal of the primer RNA leaves a small gap which cannot be filled (***Fig.24.9A***). Consequently, the daughter chromosomes will have shortened DNA molecules. This becomes significant after several cell cycles involving replication of chromosomes. The result is that over a period of time, the chromosomes may lose certain essential genes and the cell dies. This is however, avoided to a large extent.

Telomeres are the special structures that prevent the continuous loss of DNA at the end of the chromosomes during the course of replication. Thus, they protect the ends of the chromosomes, and are also responsible to prevent the chromosomes from fusing with each other. Telomeres are many repeat sequences of six nucleotides present at the ends of eukaryotic chromosomes. Human telomeres contain thousands of ***repeat TTAGGG sequences***, which can be up to a length of 1500 bp.

Role of telomerase

Telomeres are maintained by the enzyme telomerase, also called as ***telomere terminal transferase***. Telomerase is an unusual enzyme as it is composed of both protein and RNA. In case of humans, the RNA component is 450 nucleotides in length, and at the 5′-terminal and it contains the sequence 5′-CUAACCCUAAC-3′. It may be noted that the central region of this sequence is complementary to the telomere repeat sequence 5′-TTAGGG-3′. The telomerase RNA sequence can be used as a template for extension of telomeres (***Fig.24.9B***).

The telomerase RNA base pairs to the end of the DNA molecule with telomeres and extends to a small distance. Then translocation of telomerase occurs and a fresh extension of DNA takes place. This process of DNA synthesis and translocation is repeated several times until the chromosome gets sufficiently extended. The extension process gets completed through the participation of DNA polymerase and primase complex and sealing of the new DNA formed.

It may be noted here that as such the telomeres do not encode proteins. Hence, when extended by telomerase, they need not have to remain the same length, and some shortening will not pose any problem. During the course of repeated cell cycles, there occurs ***progressive shortening of telomeres***, and this has to be ***prevented***, which is appropriately carried out ***by telomerase***.

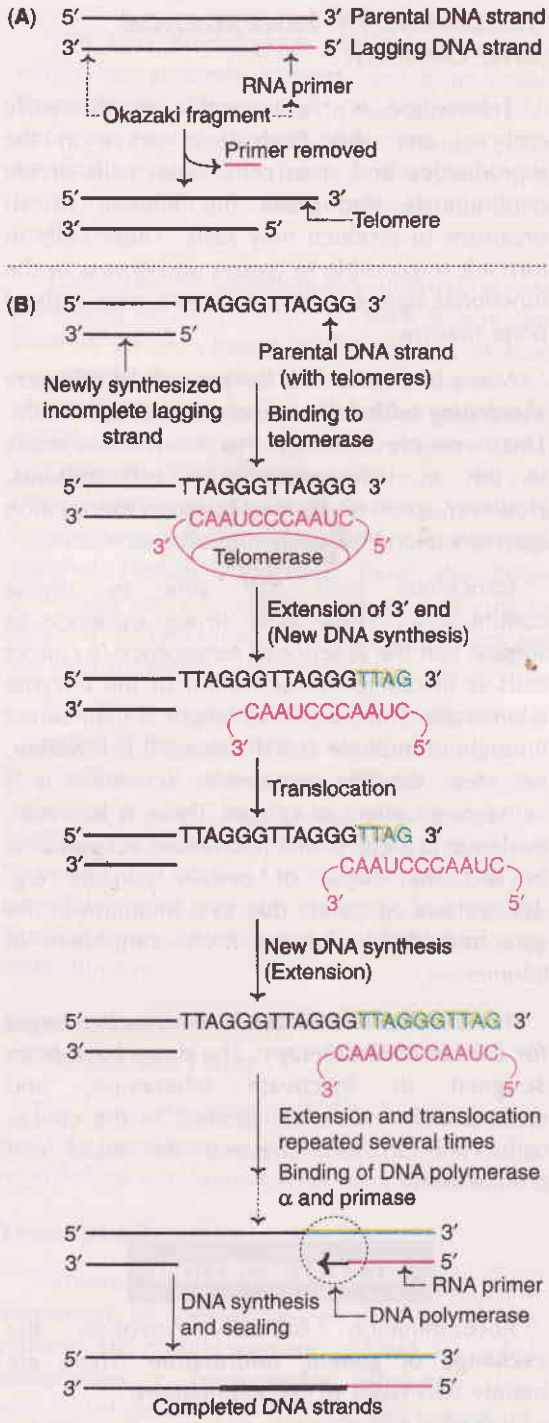


Fig. 24.9 : *Replication of DNA with telomeres (**A**) Formation of telomere (**B**) Role of telomerase in the replication of telomere (**Note :** Only 2/3 repeat sequences of telomere shown for clarity).*

TELOMERE IN SENESCENCE AND CANCER

Telomerase is highly active in the early embryo, and after birth it is active in the reproductive and stem cells. Stem cells divide continuously throughout the lifetime of an organism to produce new cells. These cells in turn are responsible to tissues and organs in the functional state e.g. hematopoietic stem cells of bone marrow.

Many biologists ***link the process of telomere shortening with cell senescence*** (i.e. cell death). This is mainly based on the observations made in the *in vitro* mammalian cell cultures. However, some researchers question this relation between telomere shortening and senescence.

Cancerous cells are able to divide continuously. There is a strong evidence to suggest that the absence of senescence in cancer cells is linked to the activation of the enzyme telomerase. Thus, telomere length is maintained throughout multiple cell divisions. It is however, not clear whether telomerase activation is a cause or an effect of cancer. There is however, evidence to suggest that telomerase activation is in fact the cause of certain cancers e.g. dyskeratosis congenita due to a mutation in the gene responsible for the RNA component of telomerase.

The enzyme ***telomerase is an attractive target for cancer chemotherapy***. The drugs have been designed to inactivate telomerase, and consequently induce senescence in the cancer cells. This in turn prevents the rapid cell proliferation.

RECOMBINATION

Recombination basically involves the ***exchange of genetic information***. There are mainly two types of recombinations.

1. **Homologous recombination :** This is also called as ***general recombination***, and occurs between identical or nearly identical chromosomes (DNA sequences). The best example is the recombination between the paternal and maternal chromosomal pairs (***Fig.24.10***).

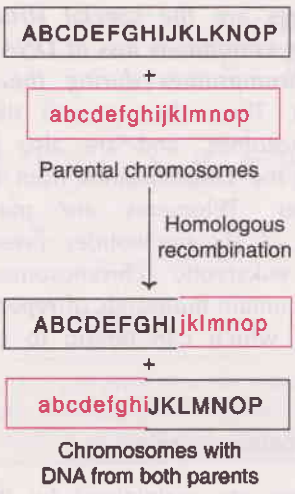


Fig. 24.10 : A diagrammatic representation of homologous recombination.

2. **Non-homologous recombination :** This is regarded as ***illegitimate recombination*** and does not require any special homologous sequences. ***Transposition*** is a good example of non-homologous recombination. Random integration of outside genes into mammalian chromosomes is another example.

HOMOLOGOUS RECOMBINATION

It is a known fact that the chromosomes are not passed on intact from generation to generation. Instead, they are inherited from both the parents. This is possible due to homologous recombination. Three models have been put forth to explain homologous recombinations.

- Holliday model
- Meselson-Radding model
- Double-strand break model.

Holliday model

Holliday model (proposed by Holliday in 1964) is the simplest among the homologous recombination models. It is depicted in ***Fig.24.11***, and briefly explained in the next page.

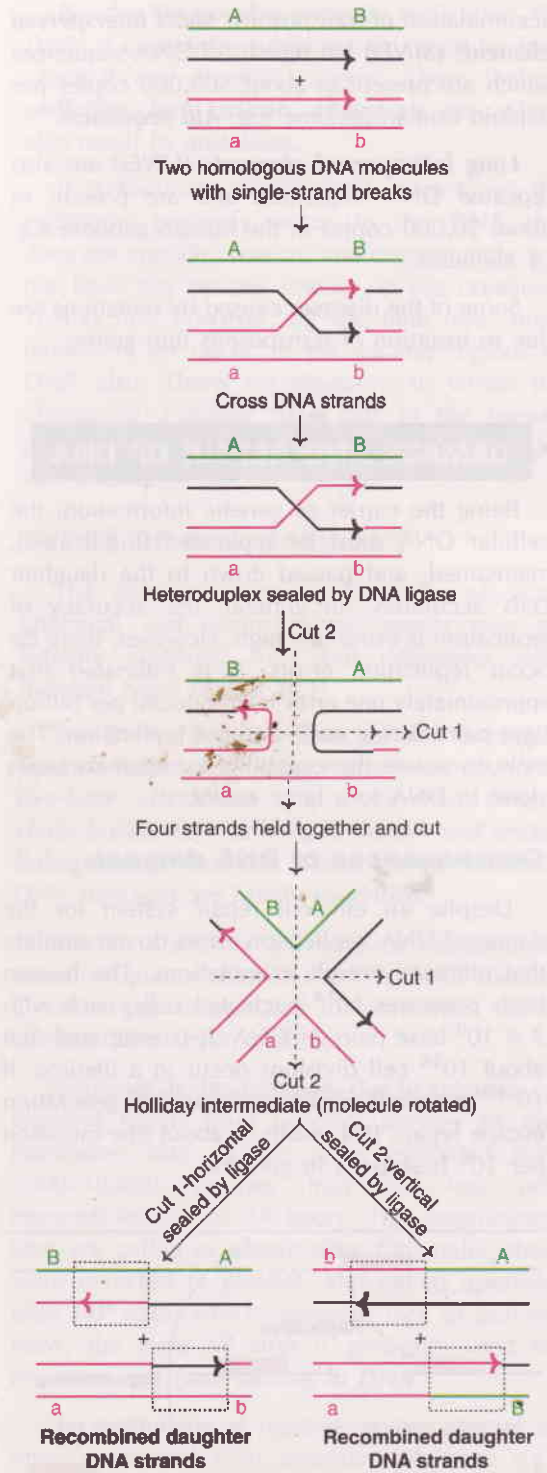


Fig. 24.11 : *Holliday model for homologous recombination (**Note :** Heteroduplex regions are shown in dotted boxes).*

The two homologous chromosomes come closer, get properly aligned, and form single-strand breaks. This results in two aligned DNA duplexes. Now the strands of each duplex partly unwind and invade in the opposite direction to form a two strands cross between the DNA molecules.

There occurs simultaneous unwinding and rewinding of the duplexes in such a way that there is no net change in the amount of base pairing, but the position of crossover moves. This phenomenon referred to as ***branch migration***, results in the formation of ***heteroduplex DNA***. The enzyme DNA ligase seals the nick. The two DNA duplexes (4 strands of DNA), joined by a single crossover point can rotate to create a ***four-standed Holliday junction***. Now the DNA molecules are subjected to symmetrical cuts in either of the two directions, and the cut ends are resealed by ligase.

The DNA exchange is determined by the direction of the cuts, which could be horizontal or vertical. If the corss strands are cut horizontally (cut 1), the flanking genes (or markers, i.e. AB/ab) remain intact, and no recombination occurs. On the other hand, if the parental strands are cut vertically (cut 2), the flanking genes get exchanged (i.e. Ab/aB) due to recombination.

NON-HOMOLOGOUS RECOMBINATION

The recombination process without any special homologous sequences of DNA is regarded as non-homologous recombination.

Transposition

Transposition primarily involves the ***movement of specific pieces of DNA in the genome***. The mobile segments of DNA are called ***transposons*** or ***transposable elements***. They were first discovered by Barbara McClintock (in 1950) in maize, and their significance was ignored for about two decades by other workers.

Transposons are mobile and can move almost to any place in the target chromosome. There are two modes of transposition. One that

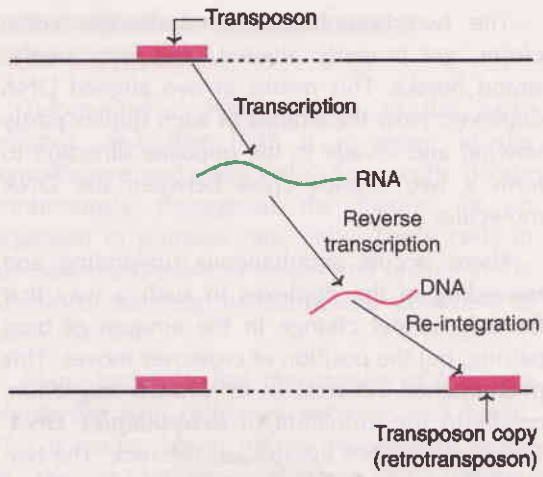


Fig. 24.12 : A diagrammatic representation of retrotransposition.

involves an RNA intermediate, and the other which does not involve RNA intermediate.

Retrotransposition : Transposition involving RNA intermediate represents retrotransposition (***Fig.24.12***). By the normal process of transcription, a copy of RNA formed from a transposon (also called as retrotransposon). Then by the enzyme reverse transcriptase, DNA is copied from the RNA. The newly formed DNA which is a copy of the transposon gets integrated into the genome. This integration may occur randomly on the same chromosome or, on a different chromosome. As a result of the retro-transposition, there are now two copies of the transposon, at different points on the genome.

DNA transposition : Some transposons are capable of direct transposition of DNA to DNA. This may occur either by replicative transposition or conservative transposition (***Fig.24.13***). Both the mechanisms require enzymes that are mostly coded by the genes within the transposons.

DNA transposition is less common than retrotransposition in case of eukaryotes. However, in case of prokaryotes, DNA transposons are more important than RNA transposons.

Significance of transposition

It is now widely accepted that a large fraction of the human genome has resulted due to the accumulation of transposons. ***Short interspersed elements (SINEs)*** are repeats of DNA sequences which are present in about 500,000 copies per haploid human genome e.g. ***Alu sequences.***

Long interspersed elements (LINEs) are also repeated DNA sequences and are present in about 50,000 copies in the human genome e.g. L1 elements.

Some of the diseases caused by mutations are due to insertion of transposons into genes.

DAMAGE AND REPAIR OF DNA

Being the carrier of genetic information, the cellular DNA must be replicated (duplicated), maintained, and passed down to the daughter cells accurately. In general, the accuracy of replication is extremely high. However, there do occur replication errors. It is estimated that approximately one error is introduced per billion base pairs during each cycle of replication. The cells do posses the capability to repair damages done to DNA to a large extent.

Consequences of DNA damage

Despite an efficient repair system for the damaged DNA, replication errors do accumulate that ultimately result in mutations. The human body possesses 10^{14} nucleated cells, each with 3×10^9 base pairs of DNA. It is estimated that about 10^{16} cell divisions occur in a lifetime. If 10^{-10} mutations per base pair per cell generation escape repair, this results in about one mutation per 10^6 base pairs in genome.

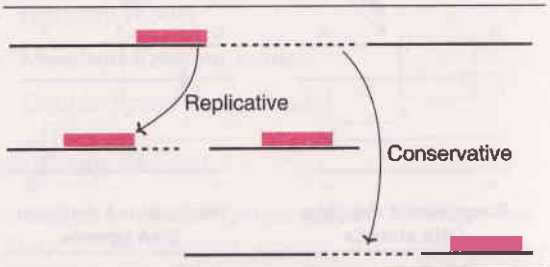


Fig. 24.13 : A diagrammatic representation of DNA transposition (coloured blocks represent transposons).

Besides the possible errors in replication, the DNA is constantly subjected to attack by both physical and chemical agents. These include ***radiation, free radicals, chemicals*** etc., which also result in ***mutations***.

It is fortunate that a great majority of the mutations probably occur in the DNA that does not encode proteins, and consequently will not have any serious impact on the organism. This is not, however, all the time true, since mutations do occur in the coding regions of DNA also. There are situations in which the change in a single base pair in the human genome can cause a serious disease e.g. sickle-cell anemia.

TYPES OF DNA DAMAGES

The damages done to DNA by physical, chemical and environmental agents may be broadly classified into four categories with different types (***Table 24.1***).

The DNA damage may occur due to ***single-base alterations*** (e.g. depurination, deamination), ***two-base alterations*** (e.g. pyrimidine dimer) ***chain breaks*** (e.g. ionizing radiation) and ***cross-linkages*** (e.g. between bases). Some selected DNA damages are briefly described.

The occurrence of spontaneous deamination bases in aqueous solution at 37°C is well known. Cytosine gets deaminated to form uracil while adenine forms hypoxanthine.

Spontaneous depurination, due to cleavage of glycosyl bonds (that connect purines to the backbone) also occurs. It is estimated that 2000–10,000 purines may be lost per mammalian cell in 24 hours. The depurinated sites are called as ***abasic sites***. Originally, they were detected in purines, and called ***apurinic sites (AP sites)*** which represent lack of purine. Now, the term AP sites is generally used to represent ***any base lacking in DNA***.

The production of reactive oxygen species is often associated with alteration of bases e.g. formation of 8-hydroxy guanine. Free radical formation and oxidative damage to DNA increases with advancement of age.

Table 24.1 Major types of DNA damages

Category	*Types*
Single-base alteration	Deamination (C→U; A→hypoxanthine)
	Depurination
	Base alkylation
	Insertion or deletion of nucleotides
	Incorporation of base analogue
Two-base alteration	UV light induced pyrimidine dimer alteration (T–T)
Chain breaks	Oxidative free radical formation
	Ionizing radiation
Cross-linkage	Between bases in the same or opposite strands
	Between the DNA and protein molecules

Ultraviolet radiations result in the formation of covalent links between adjacent pyrimidines along the DNA strand to form ***pyrimidine dimers***. DNA chain breaks can be caused by ionizing radiations (e.g. X-rays).

MUTATIONS

The genetic macromolecule DNA is highly stable with regard to its base composition and sequence. However, DNA is not totally exempt from gradual change.

Mutation refers to ***a change in the DNA structure of a gene.*** The substances (chemicals) which can induce mutations are collectively known as ***mutagens***.

The changes that occur in DNA on mutation are reflected in replication, transcription and translation.

Types of mutations

Mutations are mainly of two major types—point mutations, frameshift mutations (***Fig.24.14***).

1. **Point mutations :** The replacement of one base pair by another results in point mutation. They are of two sub-types.

(a) **Transitions :** In this case, a purine (or a pyrimidine) is replaced by another.

(b) **Transversions :** These are characterized by replacement of a purine by a pyrimidine or vice versa.

2. **Frameshift mutations :** These occur when one or more base pairs are inserted in or deleted from the DNA, respectively, causing ***insertion or deletion mutations.***

Consequences of point mutations

The change in a single base sequence in point mutation may cause one of the following (***Fig.24.15***).

1. **Silent mutation :** The codon (of mRNA) containing the changed base may code for the same amino acid. For instance, *UCA* codes for serine and change in the third base *(UCU)* still codes for serine. This is due to degeneracy of the genetic code. Therefore, there are ***no detectable effects*** in silent mutation.

2. **Missense mutation :** In this case, the changed base may code for a different amino acid. For example, *UCA* codes for serine while *ACA* codes for threonine. The mistaken (or missense) amino acid may be ***acceptable***, ***partially acceptable*** or ***unacceptable*** with regard to the function of protein molecule. ***Sickle-cell anemia*** is a classical example of missense mutation.

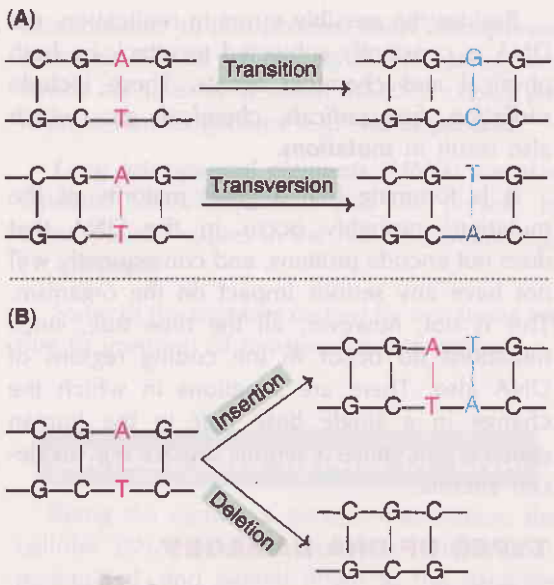


Fig. 24.14 : *An illustration of mutations (**A**)-Point mutations; (**B**)-Frameshift mutations.*

BIOMEDICAL / CLINICAL CONCEPTS

- *A few micrograms (10^{-12} g) of DNA in a fetal cell stores the genetic information that will determine the differentiation and every function of an adult animal. This is the marvel of molecular biology.*
- *Topoisomerase inhibitors (e.g. adriamycin, etoposide) are useful to prevent DNA replication, and thus uncontrolled cell proliferation in cancer. These compounds block the unwinding of DNA strands.*
- *The progressive shortening of telomeres (DNA sequences at chromosomal ends) is prevented by telomerase. This enzyme is an attractive target for cancer therapy.*
- *Mutations may sometimes result in serious diseases. e.g. sickle-cell anemia, cancer.*
- *Xeroderma pigmentosum is a rare disease characterized by photosensitivity and risk for skin cancer. This is due to a defect in the nucleotide excision repair of the damaged DNA (caused by UV rays).*
- *Hereditary nonpolyposis colon cancer is a common inherited cancer, and is due to a faulty mismatch repair of defective DNA.*

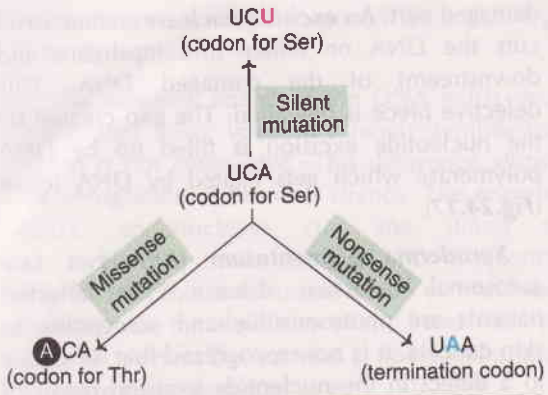


Fig. 24.15 : An illustration of point mutations (represented by a codon of mRNA).

3. **Nonsense mutation** : Sometimes, the codon with the altered base may become a ***termination*** (or nonsense) ***codon***. For instance, change in the second base of serine codon *(UCA)* may result in *UAA*. The altered codon acts as a stop signal and causes termination of protein synthesis, at that point.

Consequences of frameshift mutations

The insertion or deletion of a base in a gene results in an ***altered reading frame of the mRNA*** (hence the name frameshift). The machinery of mRNA (containing codons) does not recognize that a base was missing or a new base was added. Since there are no punctuations in the reading of codons, translation continues. The result is that the protein synthesized will have several altered amino acids and/or prematurely terminated protein.

Mutations and cancer

Mutations are permanent alterations in DNA structure, which have been implicated in the etiopathogenesis of cancer.

REPAIR OF DNA

As already stated, damage to DNA caused by replication errors or mutations may have serious consequences. The cell possesses an inbuilt system to repair the damaged DNA. This may be achieved by four distinct mechanisms (***Table 24.2***).

1. Base excision-repair
2. Nucleotide excision-repair
3. Mismatch repair
4. Double-strand break repair.

Base excision-repair

The bases cytosine, adenine and guanine can undergo spontaneous depurination to respectively form uracil, hypoxanthine and xanthine. These altered bases do not exist in the normal DNA, and therefore need to be removed. This is carried out by base excision repair (***Fig.24.16***).

Table 24.2 Major mechanisms of DNA repair

Mechanism	*Damage to DNA*	*DNA repair*
Base excision-repair	Damage to a single base due to spontaneous alteration or by chemical or radiation means.	Removal of the base by N–glycosylase; abasic sugar removal, replacement.
Nucleotide excision-repair	Damage to a segment of DNA by spontaneous, chemical or radiation means.	Removal of the DNA fragment (~ 30 nt length) and replacement.
Mismatch repair	Damage due to copying errors (1-5 base unpaired loops).	Removal of the strand (by exonuclease digestion) and replacement.
Double-strand break repair	Damage caused by ionizing radiations, free radicals, chemotherapy etc.	Unwinding, alignment and ligation.

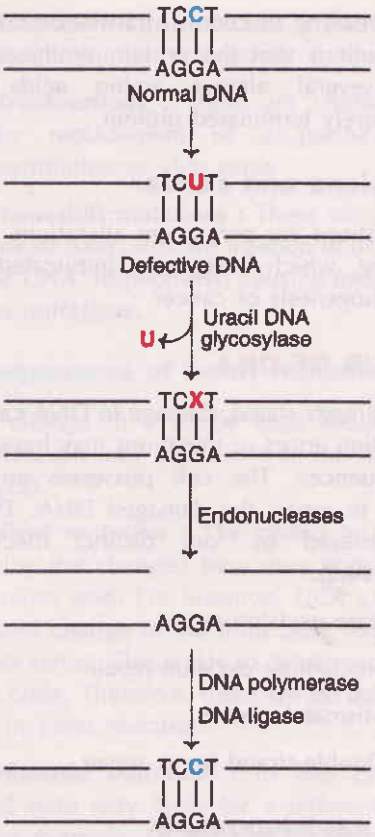


Fig. 24.16 : A diagrammatic representation of base excision-repair of DNA.

A defective DNA in which cytosine is deaminated to uracil is acted upon by the enzyme uracil DNA glycosylase. This results in the removal of the defective base uracil. An endonuclease cuts the backbone of DNA strand near the defect and removes a few bases. The gap so created is filled up by the action of repair DNA polymerase and DNA ligase.

Nucleotide excision-repair

The DNA damage due to ultraviolet light, ionizing radiation and other environmental factors often results in the modification of certain bases, strand breaks, cross-linkages etc. Nucleotide excision-repair is ideally suited for such large-scale defects in DNA. After the identification of the defective piece of the DNA, the DNA double helix is unwound to expose the damaged part. An ***excision nuclease*** (exinuclease) cuts the DNA on either side (upstream and downstream) of the damaged DNA. This defective piece is degraded. The gap created by the nucleotide excision is filled up by DNA polymerase which gets ligated by DNA ligase (***Fig.24.17***).

Xeroderma pigmentosum (XP) is a rare autosomal recessive disease. The affected patients are photosensitive and susceptible to skin cancers. It is now recognized that XP is due to a defect in the nucleotide excision repair of the damaged DNA.

Mismatch repair

Despite high accuracy in replication, defects do occur when the DNA is copied. For instance, cytosine (instead of thymine) could be incorporated opposite to adenine. Mismatch

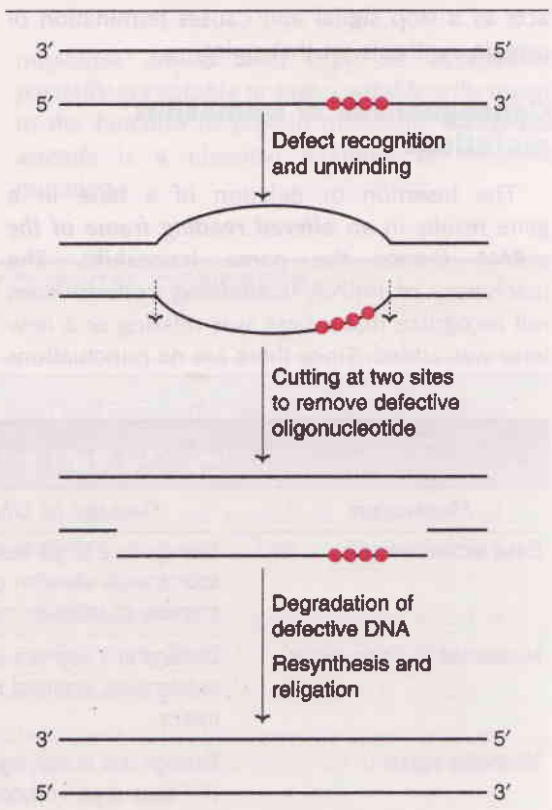


Fig. 24.17 : A diagrammatic representation of nucleotide excision-repair of DNA.

repair corrects a single mismatch base pair e.g. C to A, instead of T to A.

The template strand of the DNA exists in a methylated form, while the newly synthesized strand is not methylated. This difference allows the recognition of the new strands. The enzyme GATC endonuclease cuts the strand at an adjacent methylated GATC sequence (***Fig.24.18***). This is followed by an exonuclease digestion of the defective strand, and thus its removal. A new DNA strand is now synthesized to replace the damaged one.

Hereditary nonpolyposis colon cancer (HNPCC) is one of the most common inherited cancers. This cancer is now linked with ***faulty mismatch repair*** of defective DNA.

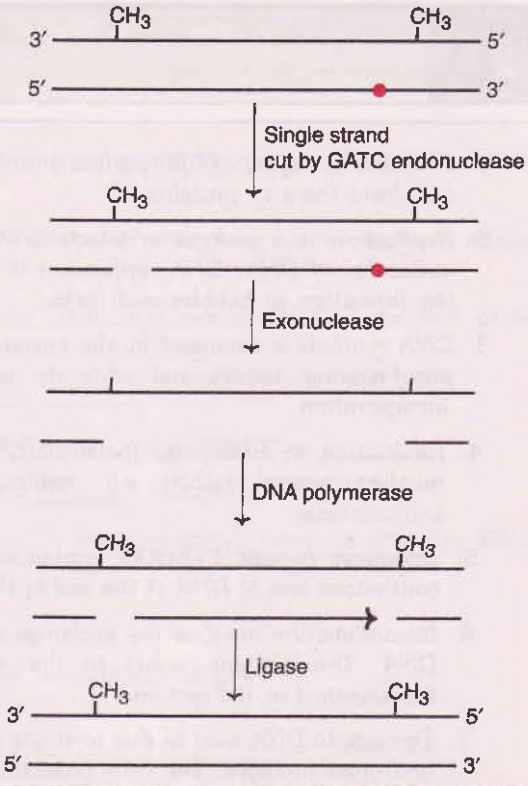


Fig. 24.18 : *A diagrammatic representation of mismatch repair of DNA.*

Double-strand break repair

Double-strand breaks (DSBs) in DNA are dangerous. They result in genetic recombination which may lead to chromosomal translocation, broken chromosomes, and finally cell death. DSBs can be repaired by homologous recombination or non-homologous end joining. Homologous recombination occurs in yeasts while in mammals, non-homologous and joining dominates.

DEFECTS IN DNA REPAIR AND CANCER

Cancer develops when certain genes that regulate normal cell division fail or are altered. Defects in the genes encoding proteins involved in nucleotide-excision repair, mismatch repair and recombinational repair are linked to human cancers. For instance, as already referred above, HNPCC is due to a defect in mismatch repair.

SUMMARY

1. *The central dogma of life revolves around the flow of information from DNA to RNA, and from there to proteins.*
2. *Replication is a process in which DNA copies itself to produce identical daughter molecules of DNA. DNA replication is semiconservative, bidirectional and occurs by the formation of bubbles and forks.*
3. *DNA syntheis is catalysed by the enzyme DNA polymerase III. This enzyme possesses proof-reading activity and edits the mistakes that might occur during nucleotide incorporation.*
4. *Replication in eukaryotes (particularly on the lagging strand) is more complex and involves several factors e.g. replication protein A, replication factor C, flap endonuclease.*
5. *Telomeres (repeat TTAGGG sequences) are the special structures that prevent the continuous loss of DNA at the end of the chromosome during the course of replication.*
6. *Recombination involves the exchange of genetic information through the exchange of DNA. Transposition refers to the movement of specific pieces of DNA (called transposons) in the genome.*
7. *Damage to DNA may be due to single base alteration, two-base alteration, chain breaks and cross linkages. The cells possess an inbuilt system to repair the damaged DNA*

SELF-ASSESSMENT EXERCISES

I. Essay questions

1. Describe the replication of DNA.
2. Give an account of recombination of DNA.
3. Discuss different types of DNA damages, and the repair mechanisms.
4. What are mutations? Describe different types, and consequences of mutations.
5. Give an account of telomeres and their role in senescence and cancer.

II. Short notes

(a) Replication fork, (b) Okazaki pieces, (c) RNA primer, (d) DNA topoisomerases, (e) Inhibitors of DNA replication, (f) Telomerase, (g) Holliday model of DNA recombination, (h) Transposition, (i) Frameshift mutations, (j) Missense mutation, (k) Mismatch repair, (l) Xeroderma pigmentosum.

III. Fill in the blanks

1. DNA strands for replication process are separated by the enzyme _____.
2. The small fragments of DNA produced during replication are called _____.
3. During the course of DNA replication, the proof-reading function is carried out by the enzyme _____.

SELF-ASSESSMENT EXERCISES

4. The problem of supercoils in DNA replication is overcome by a group of enzymes, namely _____.
5. The proteins that are associated with the transition of one phase of cell cycle to another _____.
6. Name the DNA sequence that prevents the continuous loss of DNA at the end of the chromosome during the course of replication _____.
7. *The mobile segments of DNA are called* _____.
8. Any change in the DNA sequence of a gene is commonly referred to as _____.
9. Sickle-cell anemia is a good example of _____ mutation.
10. *One common example of inherited cancer with faulty mismatch repair of defective DNA* _____.

IV. Multiple choice questions

11. *The chemical nature of the primer required for the synthesis of DNA*

    (a) DNA (b) Histone (c) RNA (d) hnRNA.

12. *The enzyme responsible for the synthesis of RNA primer in eukaryotes*

    (a) DNA polymerase α (b) DNA polymerase β (c) DNA polymerase γ (d) Topisomerases.

13. The repeat sequence of nucleotides in telomeres

    (a) TTGGGA (b) TTAGGG (c) GGGATT (d) TTGAGG.

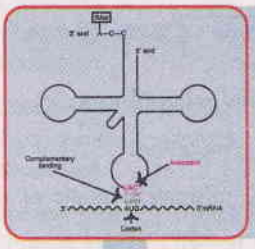
14. *The DNA damage caused by deamination is an example of*

    (a) Single-base alteration (b) Two-base alteration (c) Chain breaks (d) Cross linkage.

15. *The mutation involving the replacement of one purine by another*

    (a) Frameshift mutation (b) Transition (c) Transversion (d) None of the above.

25 Transcription and Translation



The genetic code speaks :

"Triplet base sequence of messenger RNA, I am;
Universal, specific, non-overlapping, degenerate, in character;
Faithfully work under the dictates of DNA :
To execute my master's orders for protein synthesis."

The conventional concept of central dogma of life which in essence is "***DNA makes RNA makes protein***" is an oversimplification of molecular biology. With the advances in cell biology and rapid developments in bio-informatics, the terms genome, transcriptome and proteome are in current use to represent the central dogma of molecular biology (***Fig.25.1***). Some information on the new concepts and terminology is given hereunder.

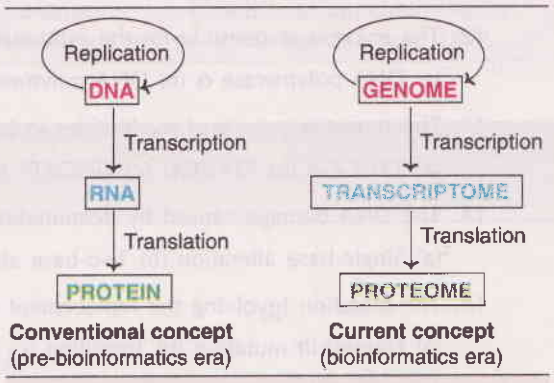


Fig. 25.1 : The central dogma of life (or molecular biology) represented in the form of conventional and current concepts.

GENOME

The ***total DNA*** (genetic information) contained ***in an organism*** or ***a cell*** is regarded as the genome. Thus, the genome is the storehouse of biological information. It includes the chromosomes in the nucleus and the DNA in mitochondria, and chloroplasts.

Genomics : The study of the structure and function of genome is genomics. The term ***functional genomics*** is used to represent the gene expression and relationship of genes with gene products. ***Structural genomics*** refers to the structural motifs and complete protein structures. ***Comparative genomics*** involves the study of comparative gene function and phylogeny.

TRANSCRIPTOME

The ***RNA copies of*** the active protein coding ***genes*** represent transcriptome. Thus, transcriptome is the initial product of gene

expression which directs the synthesis of proteins.

Transcriptomics : The study of transcriptome that involves all the RNA molecules made by a cell, tissue or an organism is transcriptomics.

PROTEOME

The ***cell's repertoire*** (repository/storehouse) of ***proteins*** with their nature ***and biological functions*** is regarded as proteome. Thus, proteome represents the entire range of proteins and their biological functions in a cell.

Proteomics : The study of the proteome.

Metabolomics : The use of genome sequence analysis for determining the capability of a cell, tissue or an organism to synthesize small molecules (metabolites) is metabolomics.

Whether the central dogma of life is represented in the conventional or more recent form, replication, transcription and translation are the key or core processes that ultimately control life. Replication of DNA has been described in ***Chapter 24***, while transcription and translation are discussed in this chapter.

TRANSCRIPTION

Transcription is a process in which ribonucleic acid ***(RNA) is synthesized from DNA.*** The word ***gene*** refers to the ***functional unit of the DNA*** that can be transcribed. Thus, the genetic information stored in DNA is expressed through RNA. For this purpose, one of the two strands of DNA serves as a ***template*** (non-coding strand or ***sense strand***) and produces ***working copies of RNA molecules.*** The other DNA strand which does not participate in transcription is referred to as coding strand or antisense strand (frequently referred to as coding strand since with the exception of T for U, primary mRNA contains codons with the same base sequence).

Transcription is selective

The entire molecule of DNA is not expressed in transcription. RNAs are synthesized only for some selected regions of DNA. For certain other regions of DNA, there may not be any transcription at all. The exact reason for the selective transcription is not known. This may be due to some inbuilt signals in the DNA molecule.

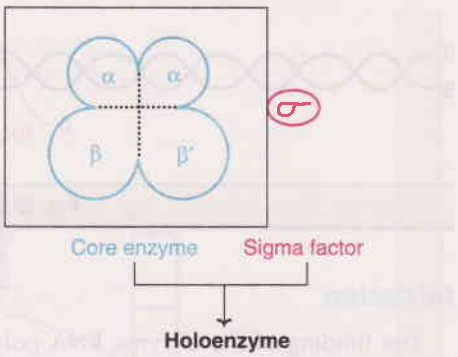


Fig. 25.2 : RNA polymerase of E. coli.

The product formed in transcription is referred to as ***primary transcript.*** Most often, the primary RNA transcripts are inactive. They undergo certain alterations (splicing, terminal additions, base modifications etc.) commonly known as ***post-transcriptional modifications,*** to produce functionally active RNA molecules.

There exist certain differences in the transcription between prokaryotes and eukaryotes. The RNA synthesis in prokaryotes is given in some detail. This is followed by a brief discussion on eukaryotic transcription.

TRANSCRIPTION IN PROKARYOTES

A single enzyme—DNA dependent RNA polymerase or simply ***RNA polymerase***—synthesizes all the RNAs in prokaryotes. RNA polymerase of *E. coli* is a complex holoenzyme (mol wt. 465 kDa) with five polypeptide subunits—2α, 1β and 1β′ and one sigma(s) factor (***Fig.25.2***). The enzyme without sigma factor is referred to as core enzyme ($\alpha_2\beta\beta'$).

An overview of RNA synthesis is depicted in ***Fig.25.3***. Transcription involves three different stages—initiation, elongation and termination (***Fig.25.4***).

SELF-ASSESSMENT EXERCISES

I. Essay questions

1. Describe the role of second messengers in hormonal action.
2. Write an account of the anterior pituitary hormones.
3. Discuss in detail the synthesis and biochemical functions of thyroid hormones.
4. Describe the hormones of adrenal cortex with special reference to glucocorticoids.
5. Write briefly on the synthesis and biochemical functions of sex hormones.

II. Short notes

(a) 'G'-Proteins, (b) Inositol triphosphate, (c) Hypothalamic hormones, (d) ACTH, (e) Goiter, (f) Epinephrine, (g) Cortisol, (h) Gastrin, (i) ADH, (j) Aldosterone.

III. Fill in the blanks

1. The enzyme that catalyses the formation of cAMP from ATP is _____.
2. The inorganic ion that can act as a second messenger for certain hormones is _____.
3. The endocrine organ responsible for the synthesis of trophic hormones is _____.
4. The compounds that produce opiate-like effects on the central nervous system are _____.
5. The enzyme that converts iodide (I^-) to active iodine (I^+) _____.
6. The most predominant mineralocorticoid synthesized by adrenal cortex _____.
7. The major urinary excretory product of catecholamines _____.
8. The male sex hormone, testosterone, is converted to a more active form, namely _____.
9. The precursor for the synthesis of steroid hormones _____.
10. The gastrointestinal hormone that increases the flow of bile from the gall bladder _____.

IV. Multiple choice questions

11. Impairment in the synthesis of dopamine by the brain is a major causative factor for the disorder
    (a) Parkinson's disease (b) Addison's disease (c) Cushing's syndrome (d) Goiter.
12. One of the following hormones is an amino acid derivative
    (a) Epinephrine (b) Norepinephrine (c) Thyroxine (d) All of them.
13. The most active mineralocorticoid hormone is
    (a) Cortisol (b) Aldosterone (c) 11-Deoxycorticosterone (d) Corticosterone.
14. Name the hormone, predominantly produced in response to fight, fright and flight
    (a) Thyroxine (b) Aldosterone (c) Epinephrine (d) ADH.
15. The hormone essentially required for the implantation of fertilized ovum and maintenance of pregnancy
    (a) Progesterone (b) Estrogen (c) Cortisol (d) Prolactin.

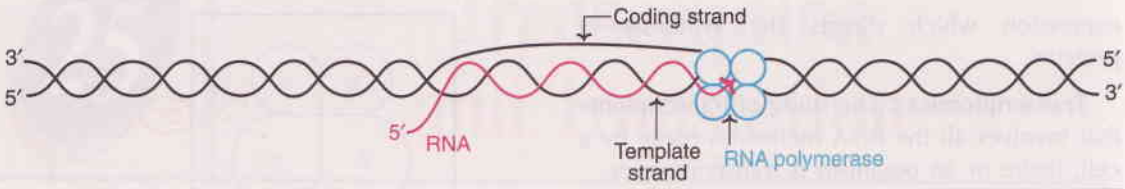


Fig. 25.3 : An overview of transcription.

Initiation

The binding of the enzyme RNA polymerase to DNA is the prerequisite for the transcription to start. The specific region on the DNA where the enzyme binds is known as ***promoter region.*** There are two base sequences on the ***coding DNA strand*** which the sigma factor of RNA polymerase can recognize for initiation of transcription (***Fig.25.5***).

1. **Pribnow box (TATA box) :** This consists of 6 nucleotide bases (TATAAT), located on the left side about 10 bases away (upstream) from the starting point of transcription.

2. **The '–35' sequence :** This is the second recognition site in the promoter region of DNA. It contains a base sequence TTGACA, which is located about 35 bases (upstream, hence –35) away on the left side from the site of transcription start.

Elongation

As the holoenzyme, RNA polymerase recognizes the promoter region, the sigma factor is released and transcription proceeds. RNA is synthesized from 5′ end to 3′ end (5′→3′) antiparallel to the DNA template. RNA polymerase utilizes ribonucleotide triphosphates (ATP, GTP, CTP and UTP) for the formation of RNA. For the addition of each nucleotide to the growing chain, a pyrophosphate moiety is released.

The sequence of nucleotide bases in the mRNA is complementary to the template DNA strand. It is however, identical to that of coding strand except that RNA contains U in place of T in DNA (***Fig.25.6***).

RNA polymerase differs from DNA polymerase in two aspects. No primer is required for RNA polymerase and, further, this enzyme does not possess endo- or exonuclease activity. Due to lack of the latter function (proof-reading activity), RNA polymerase has no ability to repair the mistakes in the RNA synthesized. This is in contrast to DNA replication which is carried out with high fidelity. It is, however, fortunate that mistakes in RNA synthesis are less dangerous, since they are not transmitted to the daughter cells.

The double helical structure of DNA unwinds as the transcription goes on, resulting in supercoils. The problem of supercoils is overcome by topoisomerases (more details in ***Chapter 24***).

Termination

The process of transcription stops by termination signals. Two types of termination are identified.

1. **Rho (ρ) dependent termination :** A specific protein, named ρ factor, binds to the growing RNA (and not to RNA polymerase) or weakly to DNA, and in the bound state it acts as ATPase and terminates transcription and releases RNA. The ρ factor is also responsible for the dissociation of RNA polymerase from DNA.

2. **Rho (ρ) independent termination :** The termination in this case is brought about by the formation of ***hairpins*** of newly synthesized RNA. This occurs due to the presence of ***palindromes.*** A palindrome is a word that reads alike forward and backward e.g. madam, rotor. The presence of palindromes in the base sequence of DNA template (same when read in opposite direction) in the termination region is known. As a result of this, the newly synthesized RNA folds to form hairpins (due to complementary base pairing) that cause termination of transcription.

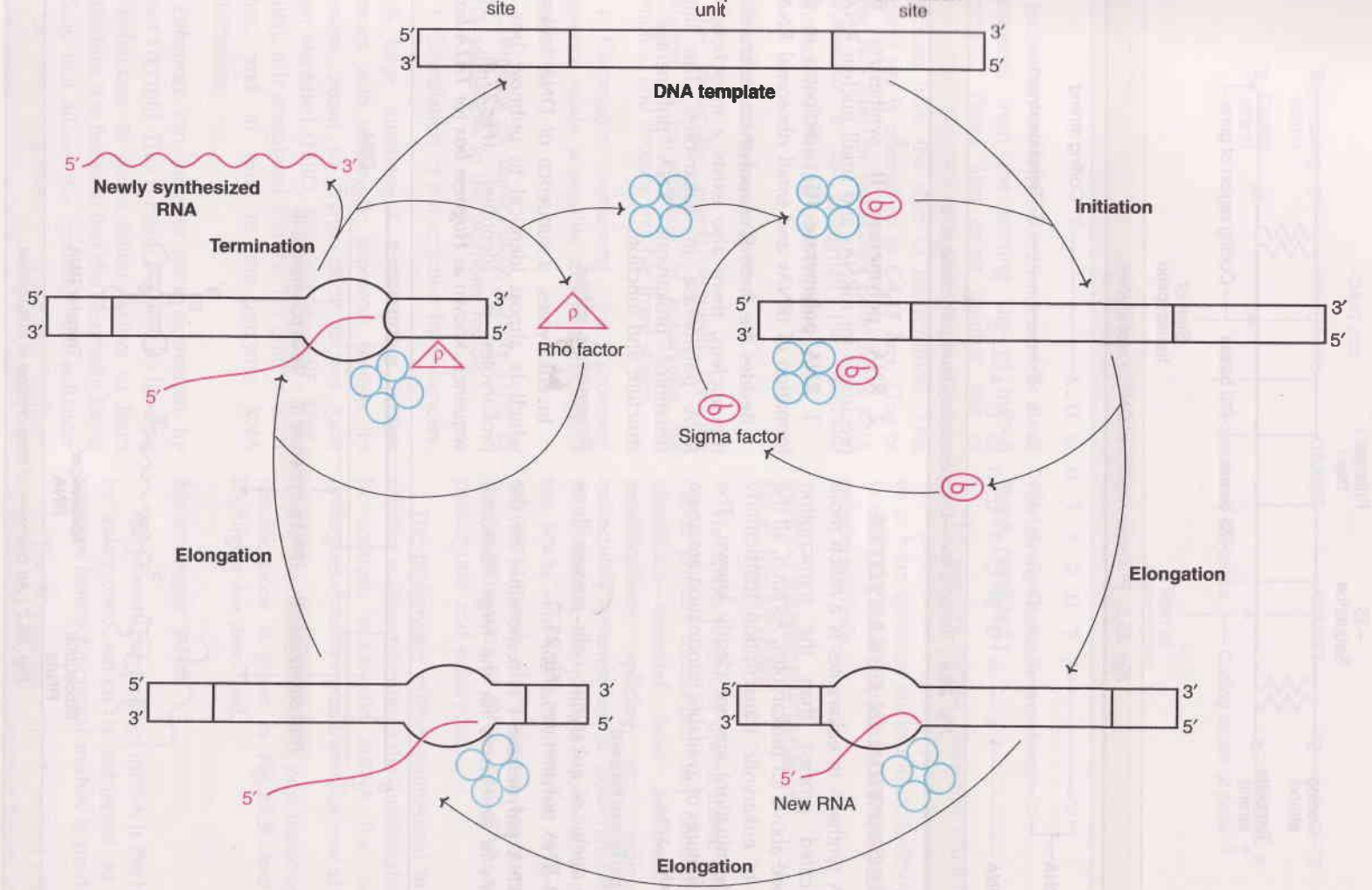


Fig. 25.4 : Synthesis of RNA from DNA template (transcription).

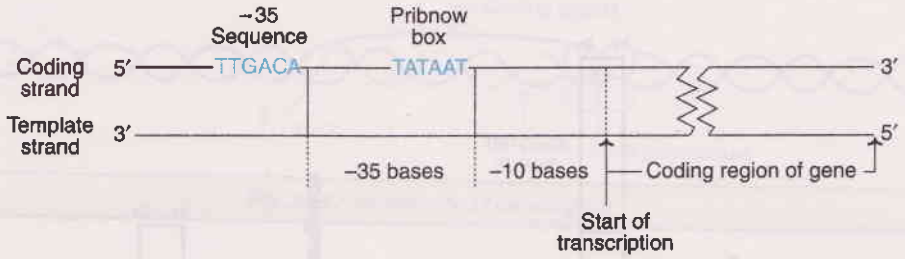


Fig. 25.5 : Promoter regions of DNA in prokaryotes.

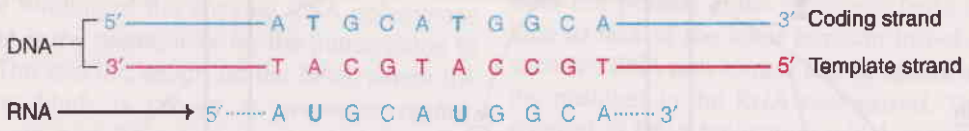


Fig. 25.6 : Transcription—Complementary base pair relationship.

TRANSCRIPTION IN EUKARYOTES

RNA synthesis in eukaryotes is a much more complicated process than the transcription described above for prokaryotes. As such, all the details of eukaryotic transcription (particularly about termination) are not clearly known. The salient features of available information are given hereunder.

RNA polymerases

The nuclei of eukaryotic cells possess three distinct RNA polymerases (***Fig.25.7***).

1. **RNA polymerase I** is responsible for the synthesis of precursors for the large ribosomal RNAs.

2. **RNA polymerase II** synthesizes the precursors for mRNAs and small nuclear RNAs.

3. **RNA polymerase III** participates in the formation of tRNAs and small ribosomal RNAs.

Besides the three RNA polymerases found in the nucleus, there also exists a mitochondrial RNA polymerase in eukaryotes. The latter resembles prokaryotic RNA polymerase in structure and function.

Promoter sites

In eukaryotes, a sequence of DNA bases—which is almost identical to pribnow box of prokaryotes—is identified (***Fig.25.8***). This sequence, known as ***Hogness box*** (or ***TATA box***),

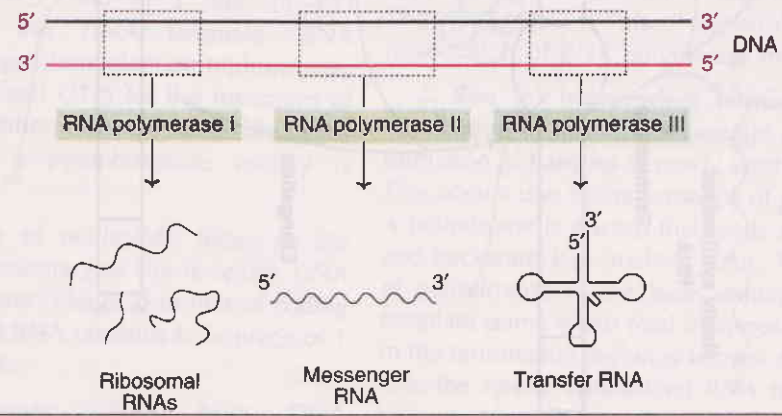


Fig. 25.7 : An overview of transcription in eukaryotes.

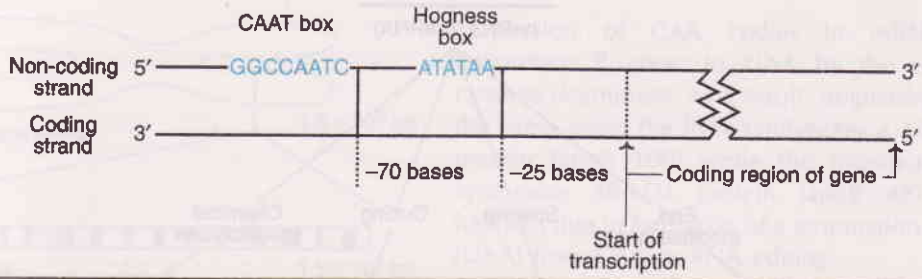


Fig. 25.8 : Promoter regions of DNA in eukaryotes.

is located on the left about 25 nucleotides away (upstream) from the starting site of mRNA synthesis. There also exists another site of recognition between 70 and 80 nucleotides upstream from the start of transcription. This second site is referred to as ***CAAT box***. One of these two sites (or sometimes both) helps RNA polymerase II to recognize the requisite sequence on DNA for transcription.

Initiation of transcription

The molecular events required for the initiation of transcription in eukaryotes are complex, and broadly involve three stages.

1. Chromatin containing the promoter sequence made accessible to the transcription machinery.

2. Binding of transcription factors (TFs) to DNA sequences in the promoter region.

3. Stimulation of transcription by enhancers.

A large number of ***transcription factors*** interact with eukaryotic promoter regions. In humans, about six transcription factors have been identified (TFIID, TFIIA, TFIIB, TFIIF, TFIIE, TFIIH). It is postulated that the TFs bind to each other, and in turn to the enzyme RNA polymerase.

Enhancer can increase gene expression by about 100 fold. This is made possible by binding of enhancers to transcription factors to form ***activators***. It is believed that the chromatin forms a loop that allows the promoter and enhancer to be close together in space to facilitate transcription.

Heterogeneous nuclear RNA (hnRNA)

The ***primary mRNA transcript*** produced by RNA polymerase II in eukaryotes is often referred to as heterogeneous nuclear RNA (hnRNA). This is then processed to produce mRNA needed for protein synthesis.

POST-TRANSCRIPTIONAL MODIFICATIONS

The RNAs produced during transcription are called primary transcripts. They undergo many alterations—***terminal base additions, base modifications, splicing*** etc., which are collectively referred to as post-transcriptional modifications. This process is required to convert the RNAs into the active forms. A group of enzymes, namely ribonucleases, are responsible for the processing of tRNAs and rRNAs of both prokaryotes and eukaryotes.

The prokaryotic mRNA synthesized in transcription is almost similar to the functional mRNA. In contrast, eukaryotic mRNA (i.e. hnRNA) undergoes extensive post-transcriptional changes.

An outline of the post-transcriptional modifications is given in ***Fig.25.9***, and some highlights are described.

Messenger RNA

The primary transcript of mRNA is the hnRNA in eukaryotes, which is subjected to many changes before functional mRNA is produced.

1. **The 5′ capping :** The 5′ end of mRNA is capped with 7-methylguanosine by an unusual

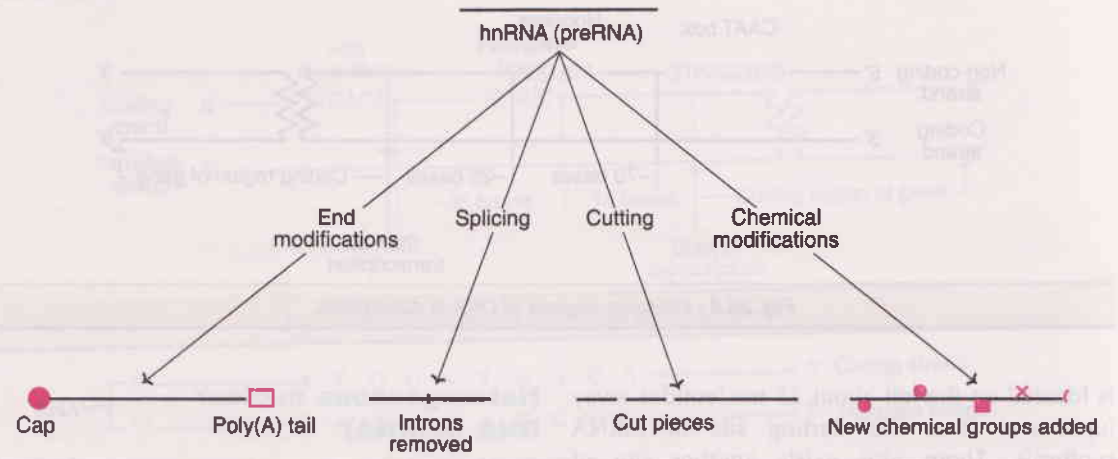


Fig. 25.9 : An outline of post-transcriptional modifications of RNA (hnRNA-Heterogeneous nuclear RNA).

5′→5′ triphosphate linkage. S-Adenosyl-methionine is the donor of methyl group. This cap is required for translation, besides stabilizing the structure of mRNA.

2. **Poly-A tail :** A large number of eukaryotic mRNAs possess an adenine nucleotide chain at the 3′-end. This poly-A tail, as such, is not produced during transcription. It is later added to stabilize mRNA. However, poly-A chain gets reduced as the mRNA enters cytosol.

3. **Introns and their removal :** Introns are the intervening nucleotide sequences in mRNA which do not code for proteins. On the other hand, ***exons of mRNA possess genetic code and are responsible for protein synthesis***. The splicing and excision of introns is illustrated in ***Fig.25.10***. The removal of introns is promoted by small nuclear ribonucleo-protein particles (snRNPs). snRNPs (pronounced as snurps) in turn, are formed by the association of small nuclear RNA (snRNA) with proteins.

The term ***spliceosome*** is used to represent the snRNP association with hnRNA at the exon-intron junction.

Post-transcriptional modifications of mRNA occur in the nucleus. The mature RNA then enters the cytosol to perform its function (translation).

A diagrammatic representation of the relationship between eukaryotic chromosomal DNA and mRNA is depicted in ***Fig.25.11***.

Different mRNAs produced by alternate splicing

Alternate patterns of hnRNA splicing result in different mRNA molecules which can produce

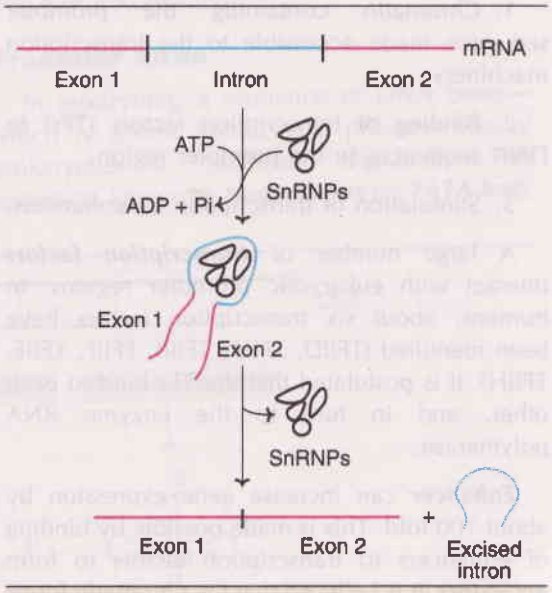


Fig. 25.10 : Formation of mature RNA from eukaryotic mRNA (SnRNPs–Small nuclear ribonucleoprotein particles).

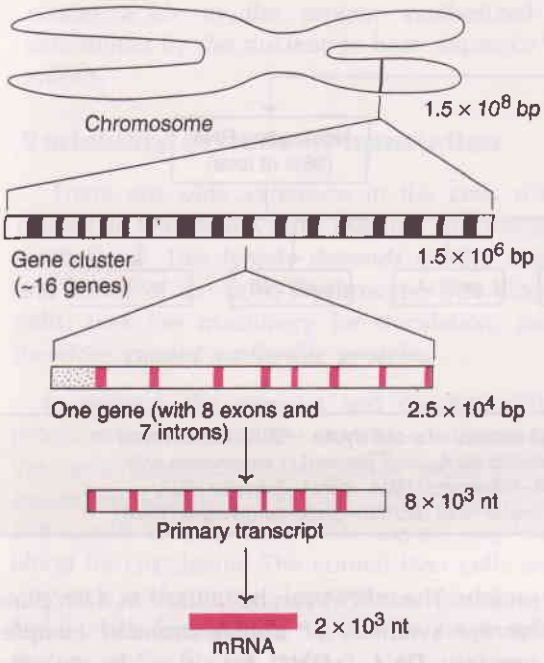


Fig. 25.11 : A diagrammatic representation of the relationship between eukaryotic chromosomal DNA and mRNA (bp–Base pair; nt–Nucleotides).

different proteins. Alternate splicing results in mRNA heterogeneity. In fact, the processing of hnRNA molecules becomes a site for the regulation of gene expression.

Faulty splicing can cause diseases : Splicing of hnRNA has to be performed with precision to produce functional mRNA. Faulty splicing may result in diseases. A good example is one type of ***β-thalassemia*** in humans. This is due to a mutation that results in a nucleotide change at an exon-intron junction. The result is a diminished or lack of synthesis of β-chain of hemoglobin, and consequently the disease β-thalassemia.

mRNA editing

The sequence in the DNA determines the coding sequence in mRNA, and finally the amino acid sequence in the protein. However, in recent years, changes in the coding information by editing of mRNA have been reported. It is estimated that about 0.01% of the mRNAs undergoes editing. One example is the conversion of CAA codon in mRNA (of apoprotein B gene) to UAA by the enzyme cytidine deaminase. As a result, originating from the same gene, the liver synthesizes a 100-kDa protein (apoB 100) while the intestinal cells synthesize 48-kDa protein (apoB 48). This happens due to formation of a termination codon (UAA) from CAA in RNA editing.

Transfer RNA

All the tRNAs of prokaryotes and eukaryotes undergo post-transcriptional modification. These include trimming, converting the existing bases into unusual ones, and addition of CCA nucleotides to 3′ terminal end of tRNAs.

Ribosomal RNA

The preribosomal RNAs originally synthesized are converted to ribosomal RNAs by a series of post-transcriptional changes.

Inhibitors of transcription

The synthesis of RNA is inhibited by certain antibiotics and toxins.

Actinomycin D : This is also known as dactinomycin. It is synthesized by *Streptomyces*. Actinomycin D binds with DNA template strand and blocks the movement of RNA polymerase. This was the very first antibiotic used for the treatment of tumors.

Rifampin : It is an antibiotic widely used for the treatment of tuberculosis and leprosy. Rifampin binds with the β-subunit of prokaryotic RNA polymerase and inhibits its activity.

α-Amanitin : It is a toxin produced by mushroom, *Amanita phalloides*. This mushroom is delicious in taste but poisonous due to the toxin α-amanitin which tightly binds with RNA polymerase II of eukaryotes and inhibits transcription.

CELLULAR RNA CONTENTS

A typical bacterium normally contains 0.05-0.10 pg of RNA which contributes to about 6% of the total weight. A mammalian cell, being larger in size, contains 20–30 pg RNA, and this

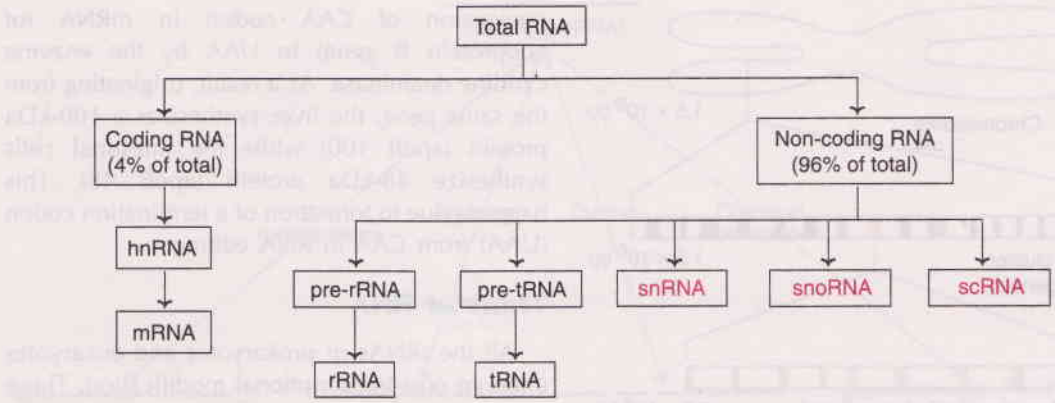


*Fig. 25.12 : A diagrammatic representation of RNA content of a cell (**Note** : RNAs represented in black are found in all organisms; RNAs in colour and exclusively present in eukaryotes only; hnRNA–Heterogeneous nuclear RNA; rRNA–Ribosomal RNA : tRNA–Transfer RNA; snRNA–Small nuclear RNA; snoRNA–Small nucleolar RNA; scRNA–Small cytoplasmic RNA).*

represents only 1% of the cell weight. Transcriptome, representing the RNA derived from protein coding genes actually constitutes only 4%, while the remaining 96% is the non-coding RNA (***Fig.25.12***). The different non-coding RNAs are ribosomal RNA, transfer RNA, small nuclear RNA, small nucleolar RNA and small cytoplasmic RNA. The functions of different RNAs are described in ***Chapter 2*** (***Refer Table 2.3***).

REVERSE TRANSCRIPTION

Some of the viruses—known as ***retroviruses***—possess RNA as the genetic material. These viruses cause cancers in animals, hence known as ***oncogenic***. They are actually found in the transformed cells of the tumors.

The enzyme RNA dependent DNA polymerase —or simply ***reverse transcriptase***—is responsible for the formation of ***DNA from RNA*** (***Fig.25.13***). This DNA is complementary (cDNA) to viral RNA and can be transmitted into host DNA.

Synthesis of cDNA from mRNA : As already described, the DNA expresses the genetic information in the form of RNA. And the mRNA determines the amino acid sequence in a protein. The mRNA can be utilized as a template for the synthesis of ***double-stranded complementary DNA*** (***cDNA***) by using the enzyme reverse transcriptase. This cDNA can be used as a probe to identify the sequence of DNA in genes.

TRANSLATION

The genetic information stored in DNA is passed on to RNA (through transcription), and ultimately expressed in the language of proteins. The ***biosynthesis of a protein or a polypeptide in a living cell*** is referred to as translation. The term translation is used to represent the biochemical translation of four-letter language information from nucleic acids (DNA and then RNA) to 20 letter language of proteins. The sequence of

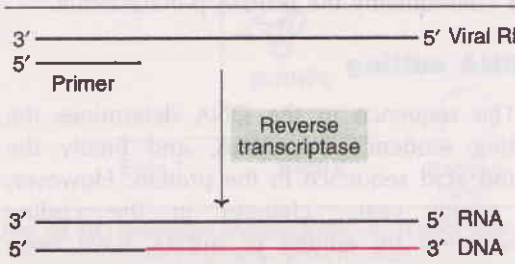


Fig. 25.13 : Reverse transcription of RNA virus.

amino acids in the protein synthesized is determined by the nucleotide base sequence of mRNA.

Variability of cells in translation

There are wide variations in the cells with respect to the quality and quantity of proteins synthesized. This largely depends on the need and ability of the cells. ***Erythrocytes*** (red blood cells) lack the machinery for translation, and therefore ***cannot synthesize proteins.***

In general, the growing and dividing cells produce larger quantities of proteins. Some of the cells continuously synthesize proteins for export. For instance, liver cells produce albumin and blood clotting factors for export into the blood for circulation. The normal liver cells are very rich in the protein biosynthetic machinery, and thus the ***liver*** may be regarded as the ***protein factory in the human body.***

GENETIC CODE

The ***three nucleotide*** (triplet) ***base sequences in mRNA that act as code words for amino acids*** in protein constitute the genetic code or simply ***codons.*** The genetic code may be regarded as a dictionary of nucleotide bases (A, G, C and U) that determines the sequence of amino acids in proteins.

The codons are composed of the four nucleotide bases, namely the purines—adenine (A) and guanine (G), and the pyrimidines—cytosine (C) and uracil (U). These four bases produce 64 different combinations (4^3) of three base codons, as depicted in ***Table 25.1.*** The nucleotide sequence of the codon on mRNA is written from the 5′-end to 3′ end. Sixty one codons code for the 20 amino acids found in protein.

The three codons ***UAA, UAG*** and ***UGA*** do not code for amino acids. They act as ***stop signals*** in protein synthesis. These three codons are collectively known as ***termination codons*** or non-sense codons. The codons UAG, UAA and UGA are often referred to, respectively, as ***amber***, ***ochre*** and ***opal*** codons.

The codons ***AUG***—and, sometimes, ***GUG***—are the chain ***initiating codons.***

Other characteristics of genetic code

The genetic code is universal, specific, non-overlapping and degenerate.

1. **Universality :** The same codons are used to code for the same amino acids in all the living organisms. Thus, the genetic code has been conserved during the course of evolution. Hence genetic code is appropriately regarded as universal. There are, ***however, a few exceptions.*** For instance, AUA is the codon for methionine in mitochondria. The same codon (AUA) codes for isoleucine in cytoplasm. With some exceptions noted, the genetic code is universal.

2. **Specificity :** A particular codon always codes for the same amino acid, hence the genetic code is highly specific or unambiguous e.g. UGG is the codon for tryptophan.

3. **Non-overlapping :** The genetic code is read from a fixed point as a continuous base sequence. It is non-overlapping, commaless and without any punctuations. For instance, UUUCUUAGAGGG is read as UUU/CUU/AGA/GGG. Addition or deletion of one or two bases will radically change the message sequence in mRNA. And the protein synthesized from such mRNA will be totally different. This is encountered in ***frameshift mutations*** which cause an alteration in the reading frame of mRNA.

4. **Degenerate :** Most of the amino acids have more than one codon. The codon is degenerate or redundant, since there are ***61 codons*** available to code ***for*** only ***20 amino acids.*** For instance, glycine has four codons. The codons that designate the same amino acid are called ***synonyms.*** Most of the synonyms differ only in the third (3′ end) base of the codon.

The Wobble hypothesis explains codon degeneracy (described later).

Codon-anticodon recognition

The codon of the mRNA is recognized by the anticodon of tRNA (***Fig.25.14***). They pair with

each other in antiparallel direction (5′ → 3′ of mRNA with 3′ → 5′ of tRNA). The usual conventional complementary base pairing (A=U, C≡G) occurs between the first two bases of codon and the last two bases of anticodon. The third base of the codon is rather lenient or flexible with regard to the complementary base. The ***anticodon region*** of tRNA consists of seven nucleotides and it recognizes the three letter codon in mRNA.

Wobble hypothesis

Wobble hypothesis, put forth by Crick, is the phenomenon in which a ***single tRNA can recognize more than one codon.*** This is due to the fact that the third base (3′-base) in the codon often fails to recognize the specific complementary base in the anticodon (5′-base). Wobbling is attributed to the difference in the spatial arrangement of the 5′-end of the anticodon. The possible pairing of 5′-end base of anticodon (of tRNA) with the 3′-end base of codon (mRNA) is given

Anticodon		Codon		
C	—	G	}	Conventional base pairing
A	—	U		
U	—	G or A	}	Non-conventional base (coloured) pairing
G	—	U or C		

Wobble hypothesis explains the degeneracy of the genetic code, i.e. existence of multiple codons for a single amino acid. Although there are 61 codons for amino acids, the number of tRNAs is far less (around 40) which is due to wobbling.

Table 25.1 The genetic code along with respective amino acids

First base 5′end	*Second base (middle one)*				*Third base 3′end*
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Try	UGU Cys	U
	UUC Phe	UCC Ser	UAC Try	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

**AUG serves as initiating codon, besides coding for methionine residue in protein synthesis; UAA, UAG and UGA called as nonsense codons, are responsible for termination of protein synthesis.*

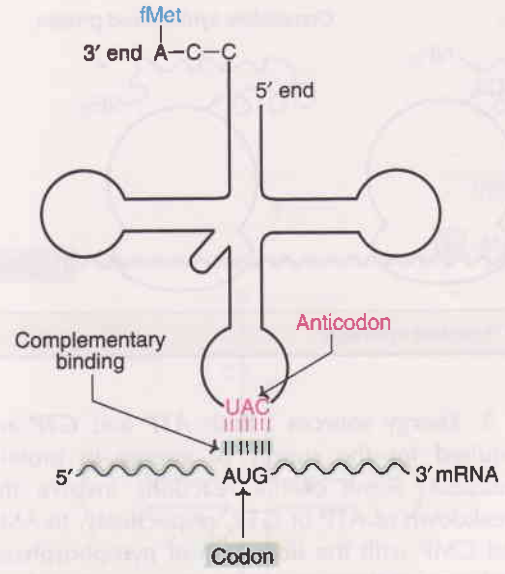


Fig. 25.14 : Complementary binding of codon (of mRNA) and anticodon (of tRNA).

Mutations and genetic code

Mutations result in the change of nucleotide sequences in the DNA, and consequently in the RNA. The different types of mutations are described in ***Chapter 24***. The ultimate effect of mutations is on the translation through the alterations in codons. Some of the mutations are harmful.

The occurrence of the disease ***sickle-cell anemia*** due to a single base alteration (CTC → CAC in DNA, and GAG → GUG in RNA) is a classical example of the seriousness of mutations. The result is that glutamate at the 6th position of β-chain of hemoglobin is replaced by valine. This happens since the altered codon GUG of mRNA codes for valine instead of glutamate (coded by GAG in normal people).

Frameshift mutations are caused by deletion or insertion of nucleotides in the DNA that generate altered mRNAs. As the reading frame of mRNA is continuous, the codons are read in continuation, and amino acids are added. This results in proteins that may contain several altered amino acids, or sometimes the protein synthesis may be terminated prematurely.

PROTEIN BIOSYNTHESIS

The ***protein synthesis*** which involves the translation of nucleotide base sequence of mRNA into the language of amino acid sequence may be divided into the ***following stages*** for the convenience of understanding.

I. Requirement of the components

II. Activation of amino acids

III. Protein synthesis proper

IV. Chaperones and protein folding

V. Post-translational modifications.

I. REQUIREMENT OF THE COMPONENTS

The protein synthesis may be considered as a biochemical factory operating on the ribosomes. As a factory is dependent on the supply of raw materials to give a final product, the protein synthesis also requires many components.

1. **Amino acids :** Proteins are polymers of amino acids. Of the ***20 amino acids*** found ***in protein structure***, half of them (10) can be synthesized by man. About ***10 essential amino acids*** have to be provided through the diet. Protein synthesis can occur only when all the amino acids needed for a particular protein are available. If there is a deficiency in the dietary supply of any one of the essential amino acids, the translation stops. It is, therefore, necessary that a regular dietary supply of essential amino acids, in sufficient quantities, is maintained, as it is a prerequisite for protein synthesis.

As regards prokaryotes, there is no requirement of amino acids, since all the 20 are synthesized from the inorganic components.

2. **Ribosomes :** The functionally active ribosomes are the ***centres*** or ***factories for protein synthesis***. Ribosomes may also be considered as workbenches of translation. Ribosomes are huge complex structures (70S for prokaryotes and 80S for eukaryotes) of proteins and ribosomal RNAs. Each ribosome consists of two subunits—one big and one small. The functional ribosome has two

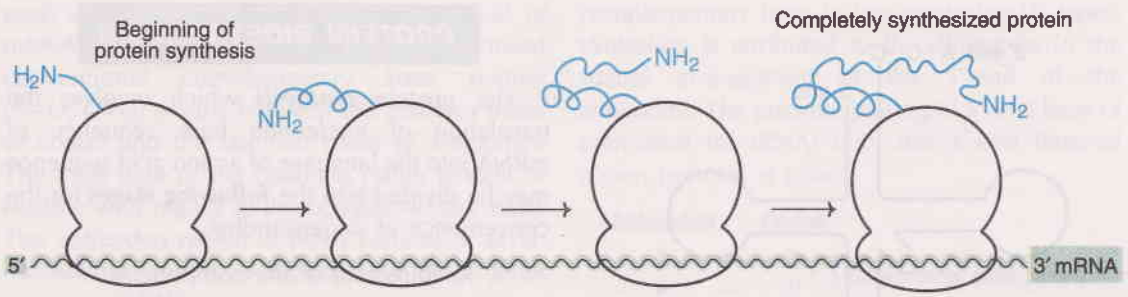


Fig. 25.15 : A polyribosome in protein synthesis.

sites—A site and P site. Each site covers both the subunits. ***A site*** is for binding of aminoacyl tRNA and ***P site*** is for binding peptidyl tRNA, during the course of translation. Some authors consider A site as acceptor site, and P site as donor site. In case of eukaryotes, there is another site called ***exit* site** or ***E site***. Thus, eukaryotes contain three sites (A, P and E) on the ribosomes.

The ribosomes are located in the cytosomal fraction of the cell. They are found in association with rough endoplasmic reticulum (RER) to form clusters RER—ribosomes, where the protein synthesis occurs. The term ***polyribosome*** (polysome) is used when several ribosomes simultaneously translate on a single mRNA (***Fig.25.15***).

3. **Messenger RNA (mRNA) :** The specific information required for the synthesis of a given protein is present on the mRNA. The DNA has passed on the genetic information in the form of ***codons*** to mRNA to translate into a protein sequence.

4. **Transfer RNAs (tRNAs) :** They carry the amino acids, and hand them over to the growing peptide chain. The amino acid is covalently bound to tRNA at the 3′-end. Each tRNA has a three nucleotide base sequence—the ***anticodon***, which is responsible to recognize the codon (complementary bases) of mRNA for protein synthesis.

In man, there are about 50 different tRNAs whereas in bacteria around 40 tRNAs are found. Some amino acids (particularly those with multiple codons) have more than one tRNA.

5. **Energy sources :** Both ***ATP*** and ***GTP*** are required for the supply of energy in protein synthesis. Some of the reactions involve the breakdown of ATP or GTP, respectively, to AMP and GMP with the liberation of pyrophosphate. Each one of these reactions consumes two high energy phosphates (equivalent to 2 ATP).

6. **Protein factors :** The process of translation involves a number of protein factors. These are needed for initiation, elongation and termination of protein synthesis. The protein factors are more complex in eukaryotes compared to prokaryotes.

II. ACTIVATION OF AMINO ACIDS

Amino acids are activated and attached to tRNAs in a two step reaction. A group of enzymes—namely aminoacyl tRNA synthetases—are required for this process. These enzymes are highly specific for the amino acid and the corresponding tRNA.

The amino acid is first attached to the enzyme utilizing ATP to form enzyme-AMP-amino acid complex. The amino acid is then transferred to the 3′ end of the tRNA to form aminoacyl tRNA (***Fig.25.16***).

III. PROTEIN SYNTHESIS PROPER

The protein or polypeptide synthesis occurs on the ribosomes (rather polyribosomes). The ***mRNA is read in the 5′→3′ direction and the polypeptide synthesis proceeds from N-terminal end to C-terminal end.*** Translation is directional and collinear with mRNA.

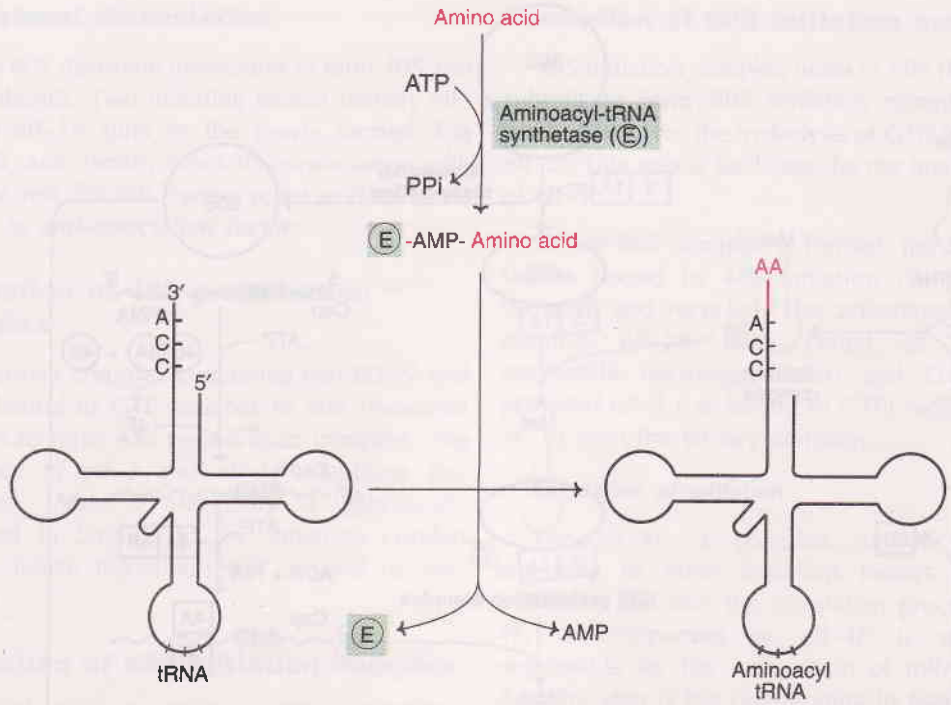


Fig. 25.16 : Formation of aminocacyl tRNA (AA–Amino acid; E–Enzyme).

The prokaryotic mRNAs are ***polycistronic,*** since a single mRNA has many coding regions that code for different polypeptides. In contrast, eukaryotic mRNA is ***monocistronic,*** since it codes for a single polypeptide.

In case of prokaryotes, translation commences before the transcription of the gene is completed. Thus, simultaneous transcription and translation are possible. This is not so in case of eukaryotic organisms since transcription occurs in the nucleus whereas translation takes place in the cytosol. Further, the primary transcript (hnRNA) formed from DNA has to undergo several modifications to generate functional mRNA.

Protein synthesis is comparatively simple in case of prokaryotes compared to eukaryotes. Further, many steps in eukaryotic translation were not understood for quite sometime. For these reasons, majority of the textbooks earlier used to describe translation in prokaryotes in detail, and give most important and relevant information for eukaryotic translation. With the advances in molecular biology, the process of protein biosynthesis in eukaryotes is better understood now.

Translation in eukaryotes is briefly ***described here***, along with some relevant features of prokaryotic protein biosynthesis. Translation proper is divided into three stages—initiation, elongation and termination (as it is done for transcription).

INITIATION OF TRANSLATION

The initiation of translation in eukaryotes is complex, involving at least ***ten eukaryotic initiation factors (eIFs)***. Some of the eIFs contain multiple (3-8) subunits. The process of translation initiation can be divided into four steps (***Fig.25.17***).

1. Ribosomal dissociation.
2. Formation of 43S preinitiation complex.
3. Formation of 48S initiation complex.
4. Formation of 80S initiation complex.

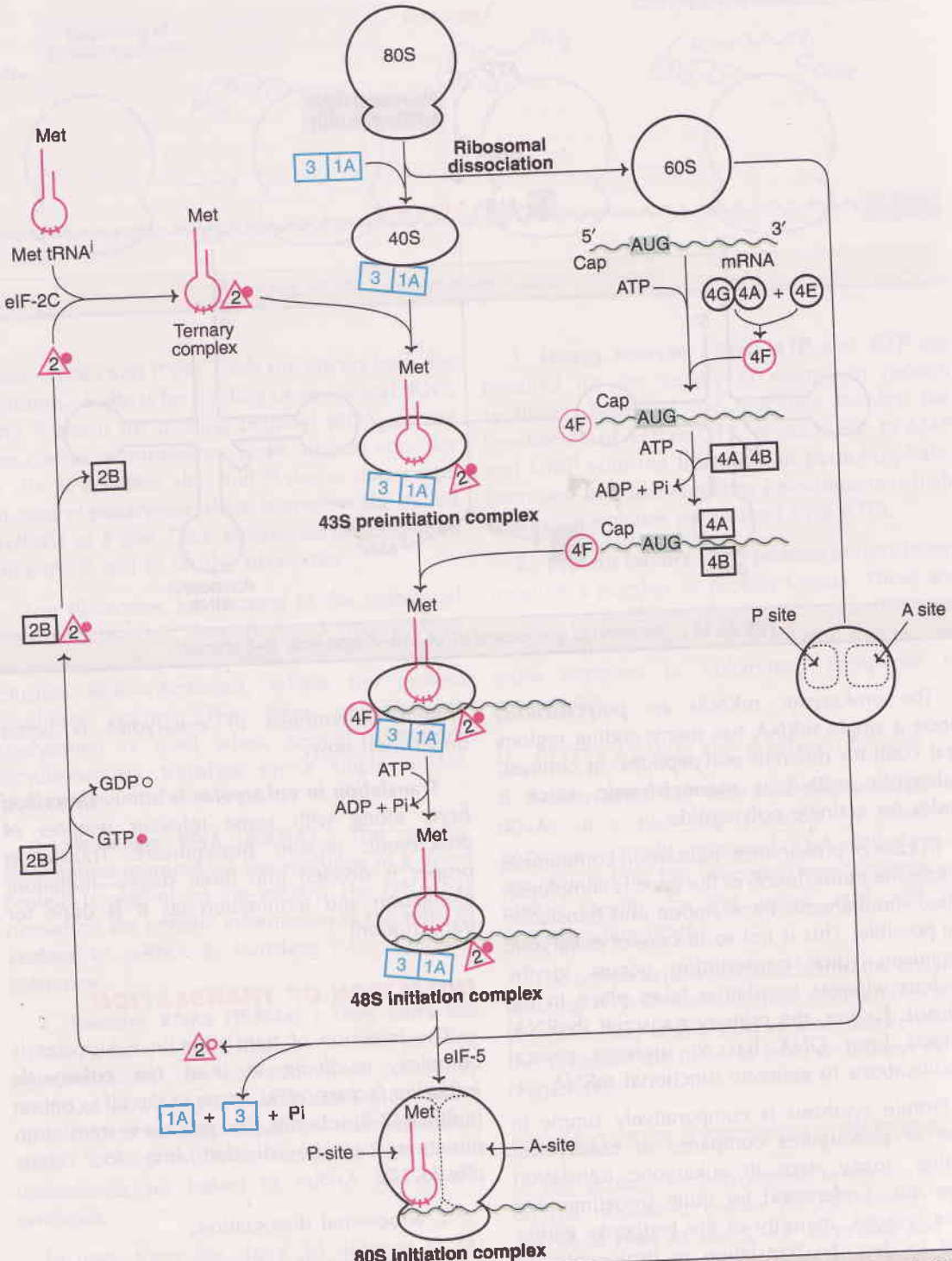


Fig. 25.17 : A diagrammatic representation of initiation of protein biosynthesis (translation) in eukaryotic cells (The eukaryotic initiation factors are represented by symbols □, △, and ○. By prefixing with eIF, the full names of the factors are obtained e.g. [3] represents eIF–3).

Ribosomal dissociation

The 80S ribosome dissociates to form 40S and 60S subunits. Two initiating factors namely eIF-3 and eIF-1A bind to the newly formed 40S subunit, and thereby block its reassociation with 60S subunit. For this reason, some workers name eIF-3 as ***anti-association factor***.

Formation of 43S preinitiation complex

A ternary complex containing met-tRNA$_i$ and eIF-2 bound to GTP attaches to 40S ribosomal subunit to form 43S preinitiation complex. The presence of eIF-3 and eIF-1A stabilizes this complex (***Note :*** Met-tRNA is specifically involved in binding to the ***i***nitiation condon AUGs; hence the superscript^i is used in met-tRNA$_i$).

Formation of 48S initiation complex

The binding of mRNA to 43S preinitiation complex results in the formation of 48S initiation complex through the intermediate 43S initiation complex. This, however, involves certain interactions between some of the eIFs and activation of mRNA.

eIF-4F complex is formed by the association of eIF-4G, eIF-4A with eIF-4E. The so formed eIF-4F (referred to as cap binding protein) binds to the cap of mRNA. Then eIF-4A and eIF-4B bind to mRNA and reduce its complex structure. This mRNA is then transferred to 43S complex. For the appropriate association of 43S preinitiation complex with mRNA, energy has to be supplied by ATP.

Recognition of initiation codon : The ribosomal initiation complex scans the mRNA for the identification of appropriate initiation codon. 5′-AUG is the initiation codon and its recognition is facilitated by a specific sequence of nucleotides surrounding it. This marker sequence for the identification of AUG is called as ***Kozak consensus sequences***. In case of ***prokaryotes*** the recognition sequence of initiation codon is referred to as ***Shine- Dalgarno sequence***.

Formation of 80S initiation complex

48S initiation complex binds to 60S ribosomal subunit to form 80S initiation complex. The binding involves the hydrolysis of GTP (bound to eIF-2). This step is facilitated by the involvement of eIF-5.

As the 80S complex is formed, the initiation factors bound to 48S initiation complex are released, and recycled. The activation of eIF-2 requires eIF-2B (also called as guanine nucleotide exchange factor) and GTP. The activated eIF-2 (i.e. bound to GTP) requires eIF-2C to form the ternary complex.

Regulation of initiation

The eIF-4F, a complex formed by the assembly of three initiation factors controls initiation, and thus the translation process. eIF-4E, a component of eIF-4F is primarily responsible for the recognition of mRNA cap. And this step is the rate-limiting in translation.

eIF-2 which is involved in the formation of 43S preinitiation complex also controls protein biosynthesis to some extent.

Initiation of translation in prokaryotes

The formation of translation initiation complex in prokaryotes is less complicated compared to eukaryotes. The 30S ribosomal subunit is bound to initiation factor 3 (IF-3) and attached to ternary complex of IF-2, formyl met-tRNA and GTP. Another initiation factor namely IF-I also participates in the formation of preinitiation complex. The recognition of initiation codon AUG is done through Shine-Dalgarno sequence. A 50S ribosome unit is now bound with the 30S unit to produce 70S initiation complex in prokaryotes.

ELONGATION OF TRANSLATION

Ribosomes elongate the polypeptide chain by a sequential addition of amino acids. The amino acid sequence is determined by the order of the codons in the specific mRNA. Elongation, a cyclic process involving certain elongation

factors (EFs), may be divided into three steps (***Fig.25.18***).

1. Binding of aminoacyl t-RNA to A-site.
2. Peptide bond formation.
3. Translocation.

Binding of aminoacyl—tRNA to A-site

The 80S initiation complex contains met-$tRNA^{i}$ in the P-site, and the A-site is free. Another aminoacyl-tRNA is placed in the A-site. This requires proper codon recognition on the mRNA and the involvement of elongation factor 1a (EF-Ia) and supply of energy by GTP. As the aminoacyl-tRNA is placed in the A-site, EF-1α and GDP are recycled to bring another aminoacyl-tRNA.

Peptide bond formation

The enzyme ***peptidyltransferase*** catalyses the formation of peptide bond (***Fig.25.19***). The activity of this enzyme lies on 28S RNA of 60S ribosomal subunit. It is therefore the ***rRNA*** (and not protein) referred to as ***ribozyme*** that catalyses the peptide bond formation. As the amino acid in the aminoacyl-tRNA is already activated, no additional energy is required for peptide bond formation.

The net result of peptide bond formation is the attachment of the growing peptide chain to the tRNA in the A-site.

Translocation

As the peptide bond formation occurs, the ribosome moves to the next codon of the mRNA (towards 3'-end). This process called translocation, basically involves the movement of growing peptide chain from A-site to P-site. Translocation requires EF-2 and GTP. GTP gets hydrolysed and supplies energy to move mRNA. EF-2 and GTP complex recycles for translocation.

In recent years, another site namely ***exit site*** (E-site) has been identified in eukaryotes. The deacylated tRNA moves into the E-site, from where it leaves the ribosome.

In case of prokaryotes, the elongation factors are different, and they are EF-Tu, EF-Ts (in place of of EF-1a) and EF-G (instead of EF-2).

Incorporation of amino acids

It is estimated that about ***six amino acids per second*** are incorporated during the course of elongation of translation ***in eukaryotes.*** In case of prokaryotes, as many as 20 amino acids can be incorporated per second. Thus the process of protein/polypeptide synthesis in translation occurs with ***great speed*** and ***accuracy.***

TERMINATION OF TRANSLATION

Termination is a simple process when compared to initiation and elongation. After several cycles of elongation, incorporating amino acids and the formation of the specific protein/polypeptide molecule, one of the ***stop or termination signals (UAA, UAG and UCA)*** terminates the growing polypeptide. The termination codons which act as stop signals do not have specific tRNAs to bind. As the termination codon occupies the ribosomal A-site, the release factor namely eRF recognizes the stop signal. eRF-GTP complex, in association with the enzyme peptidyltransferase, cleaves the peptide bond between the polypeptide and the tRNA occupying P-site. In this reaction, a water molecule, instead of an amino acid is added. This hydrolysis releases the protein and tRNA from the P-site. The 80S ribosome dissociates to form 40S and 60S subunits which are recycled. The mRNA is also released.

INHIBITORS OF PROTEIN SYNTHESIS

Translation is a complex process and it has become a favourite target for inhibition by antibiotics. Antibiotics are the substances produced by bacteria or fungi which inhibit the growth of other organisms. Majority of the antibiotics interfere with the bacterial protein synthesis and are harmless to higher organisms. This is due to the fact that the process of translation sufficiently differs between prokaryotes and eukaryotes. The action of a few important antibiotics on translation is described next.

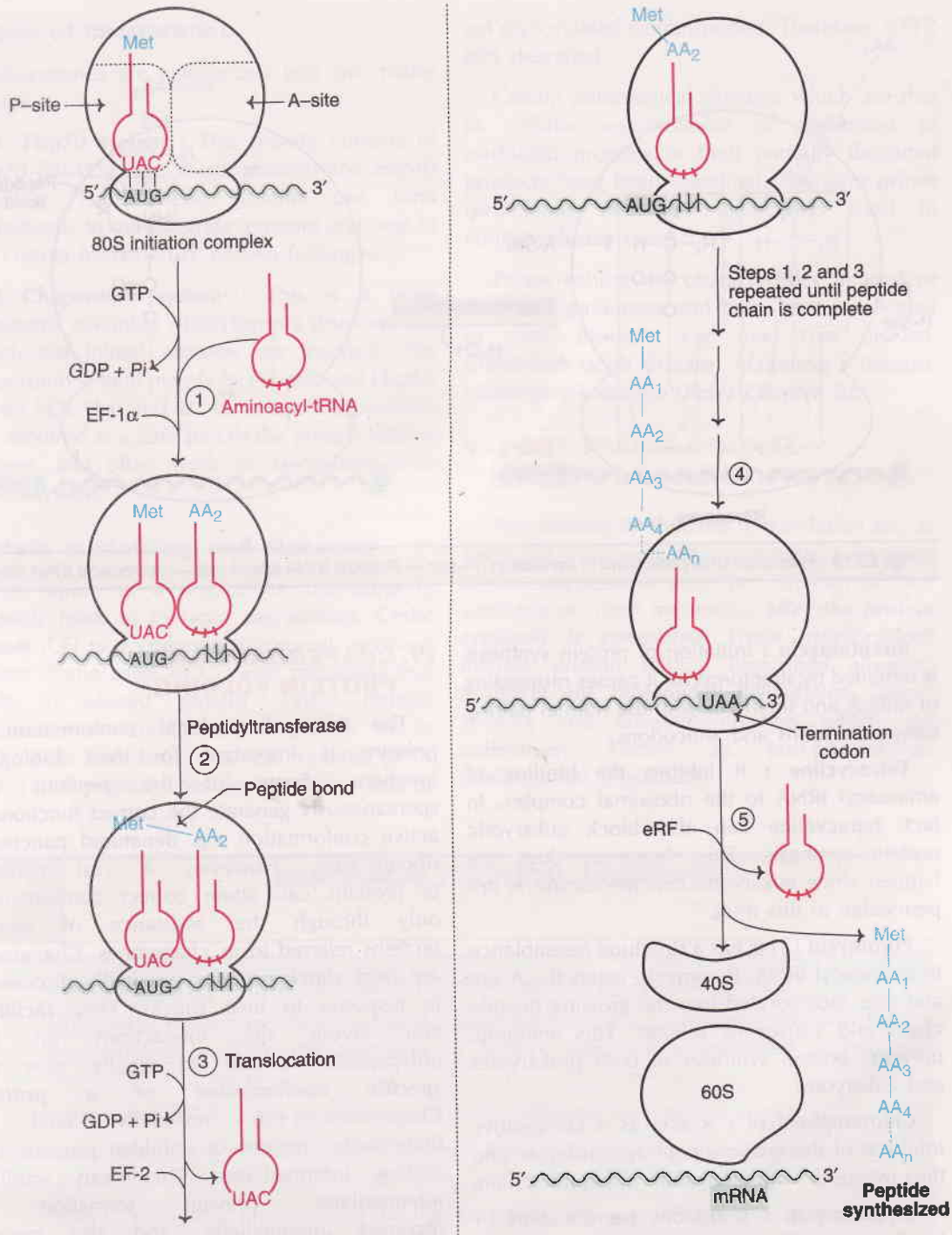


Fig. 25.18 contd. next column

*Fig. 25.18 : Protein biosynthesis — Elongation and termination (for initiation **See Fig. 25.17**). Met-Methionine; P-site — Peptidyl tRNA binding site; A-site — Aminoacyl tRNA binding site. AA-Amino acid; EF-Elongation factor; RF-Releasing factor.*

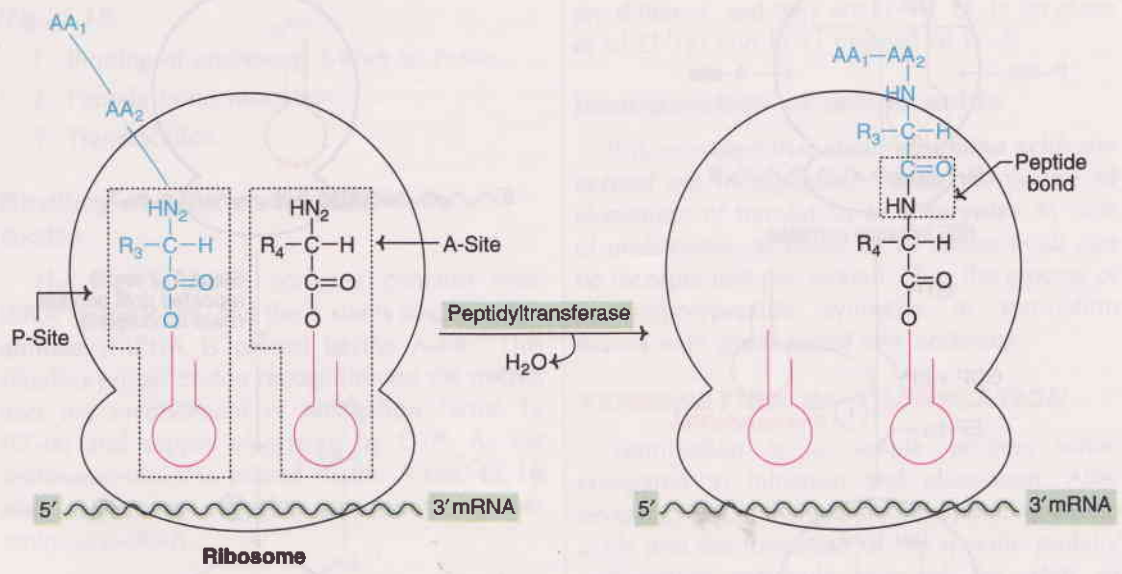


Fig. 25.19 : Formation of peptide bond in translation (P–site — Peptidyl tRNA site; A–site — Aminoacyl tRNA site).

Streptomycin : Initiation of protein synthesis is inhibited by streptomycin. It causes misreading of mRNA and interferes with the normal pairing between codons and anticodons.

Tetracycline : It inhibits the binding of aminoacyl tRNA to the ribosomal complex. In fact, tetracycline can also block eukaryotic protein synthesis. This, however, does not happen since eukaryotic cell membrane is not permeable to this drug.

Puromycin : This has a structural resemblance to aminoacyl tRNA. Puromycin enters the A site and gets incorporated into the growing peptide chain and causes its release. This antibiotic prevents protein synthesis in both prokaryotes and eukaryotes.

Chloramphenicol : It acts as a competitive inhibitor of the enzyme peptidyltransferase and thus interferes with elongation of peptide chain.

Erythromycin : It inhibits translocation by binding with 50S subunit of bacterial ribosome.

Diphtheria toxin : It prevents translocation in eukaryotic protein synthesis by inactivating elongation factor eEF_2.

IV. CHAPERONES AND PROTEIN FOLDING

The three dimensional conformation of proteins is important for their biological functions. Some of the proteins can spontaneously generate the correct functionally active conformation e.g. denatured pancreatic ribonuclease. However, a vast majority of proteins can attain correct conformation, only through the assistance of certain proteins referred to as chaperones. Chaperones are ***heat shock proteins*** (originally discovered in response to heat shock). They ***facilitate*** and favour the interactions on the polypeptide surfaces to finally give the ***specific conformation of a protein.*** Chaperones can reversibly bind to hydrophobic regions of unfolded proteins and folding intermediates. They can stabilize intermediates, prevent formation of incorrect intermediates, and also prevent undesirable interactions with other proteins. All these activities of chaperones help the protein to attain compact and biologically active conformation.

Types of chaperones

Chaperones are categorized into two major groups

1. **Hsp70 system :** This mainly consists of Hsp70 (70-kDa ***h***eat ***s***hock ***p***rotein) and Hsp40 (40-kDa Hsp). These proteins can bind individually to the substrate (protein) and help in the correct formation of protein folding.

2. **Chaperonin system :** This is a large oligomeric assembly which forms a structure into which the folded proteins are inserted. The chaperonin system mainly has Hsp60 and Hsp10 i.e. 60 kDa Hsp and 10 kDa Hsp. Chaperonins are required at a later part of the protein folding process, and often work in association with Hsp70 system.

Protein misfolding and diseases

The failure of a protein to fold properly generally leads to its rapid degradation. ***Cystic fibrosis*** (CF) is a common autosomal recessive disease. Some cases of CF with mutations that result in altered protein (cystic fibrosis transmembrane conductance regulator or in short CFTR) have been reported. Mutated CFTR cannot fold properly, besides not being able to get glycosylated or transported. Therefore, CFTR gets degraded.

Certain neurological diseases which are due to cellular accumulation of aggregates of misfolded proteins or their partially degraded products have been identified. The term ***prions*** (***pr***oteinous ***in***fectious agents) is used to collectively represent them.

Prions exhibit the characteristics of viral or microbial pathogens and have been implicated in many diseases. e.g. mad cow disease, Creutzfeldt-Jacob disease, Alzheimer's disease, Huntington's disease (***Refer Chapter 22***).

V. POST-TRANSLATIONAL MODIFICATIONS OF PROTEINS

The proteins synthesized in translation are, as such, not functional. Many changes take place in the polypeptides after the initiation of their synthesis or, most frequently, ***after the protein synthesis is completed***. These ***modifications*** include protein folding (described already), trimming by proteolytic degradation, intein splicing and covalent changes which are collectively known as post-translational modifications (***Fig.25.20***).

BIOMEDICAL / CLINICAL CONCEPTS

- *Faulty splicing of hnRNA may result in certain diseaes e.g. β-thalassemia.*
- *Inhibitors of transcription are used as therapeutic agents. Thus, actinomycin D was the first antibiotic used in the treatment of tumors. Rifampin is employed to treat tuberculosis and leprosy.*
- *Retroviruses (RNA is the genetic material) are oncogenic i.e. cause cancers.*
- *Several antibiotics selectively block bacterial translation, and thus inhibit their growth e.g. streptomycin, tetracycline, puromycin.*
- *Protein misfolding often results in the formation of prions (proteinous infectious agents) which have been implicated in many diseases e.g. mad cow disese, Alzheimer's disease.*
- *Lebers' hereditary optic neuropathy is caused by mutation in mtDNA in males. The victims become blind due to loss of central vision as a result of neuroretinal degeneration.*

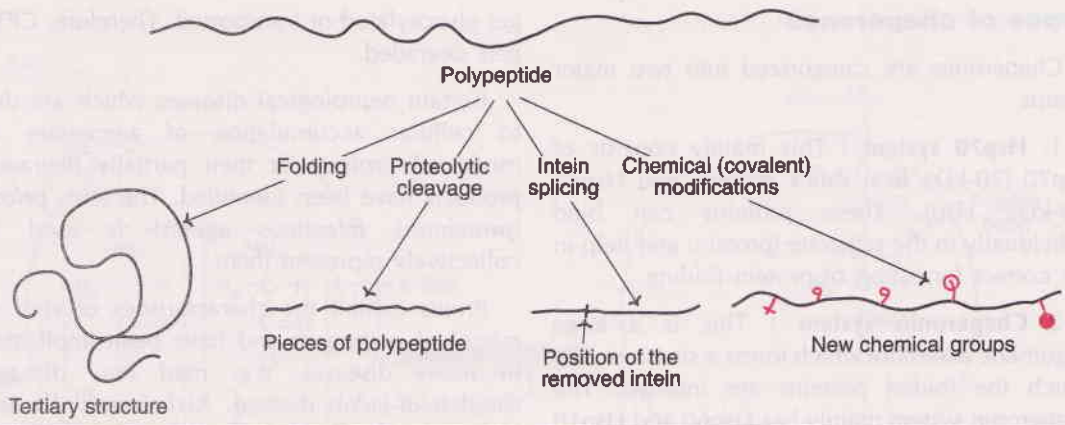


Fig. 25.20 : An outline of post-translational modifications of proteins.

Proteolytic degradation

Many proteins are synthesized as the precursors which are much bigger in size than the functional proteins. Some portions of precursor molecules are removed by proteolysis to liberate active proteins. This process is commonly referred to as trimming. The formation of insulin from preproinsulin, conversion of ***zymogens*** (inactive digestive enzymes e.g. trypsinogen) to the active enzymes are some examples of trimming.

Intein splicing

Inteins are ***intervening sequences in certain proteins.*** These are comparable to introns in mRNAs. ***Inteins*** have to be ***removed, and exteins ligated*** in the appropriate order for the protein to become active.

Covalent modifications

The proteins synthesized in translation are subjected to many covalent changes. By these modifications in the amino acids, the proteins may be converted to active form or inactive form. Selected examples of covalent modifications are described below.

1. **Phosphorylation :** The hydroxyl group containing amino acids of proteins, namely serine, threonine and tyrosine are subjected to phosphorylation. The phosphorylation may either increase or decrease the activity of the proteins. A group of enzymes called ***protein kinases*** catalyse phosphorylation while protein phosphatases are responsible for dephosphorylation (removal of phosphate group). Many enzymes that undergo phosphorylation or dephosphorylation are known in metabolisms (e.g. glycogen synthase).

2. **Hydroxylation :** During the formation of collagen, the amino acids proline and lysine are respectively converted to hydroxyproline and hydroxylysine. This hydroxylation occurs in the endoplasmic reticulum and requires ***vitamin C.***

3. **Glycosylation :** The attachment of carbohydrate moiety is essential for some proteins to perform their functions. The complex carbohydrate moiety is attached to the amino acids, serine and threonine (O-linked) or to asparagine (N-linked), leading to the synthesis of glycoproteins.

Vitamin K dependent ***carboxylation*** of glutamic acid residues in certain clotting factors is also a post-translational modification.

In the ***Table 25.2***, selected examples of post-translational modification of proteins through their amino acids are given.

PROTEIN TARGETING

The eukaryotic proteins (tens of thousands) are distributed between the cytosol, plasma membrane and a number of cellular organelles

Table 25.2 Selected examples of post-translational modifications of proteins through their amino acids

Amino acid	*Post-translational modification(s)*
Amino-terminal amino acid	Glycosylation, acetylation, myristoylation, formylation.
Carboxy terminal amino acid	Methylation, ADP-ribosylation
Arginine	Methylation
Aspartic acid	Phosphorylation, hydroxylation
Cysteine (—SH)	Cystine (—S—S—) formation, selenocysteine formation, glycosylation.
Glutamic acid	Methylation, γ-carboxylation.
Histidine	Methylation, phosphorylation.
Lysine	Acetylation, methylation, hydroxylation, biotinylation.
Methionine	Sulfoxide formation.
Phenylalanine	Glycosylation, hydroxylation.
Proline	Hydroxylation, glycosylation.
Serine	Phosphorylation, glycosylation.
Threonine	Phosphorylation, methylation glycosylation.
Tryptophan	Hydroxylation.
Tyrosine	Hydroxylation, phosphorylation, sulfonylation, iodination.

(nucleus, mitochondria, endoplasmic reticulum etc.). At the appropriate places, they perform their functions.

The proteins, synthesized in translation, have to ***reach*** their ***destination*** to exhibit their *biological activities. This is carried out by a* process called protein targeting or ***protein sorting*** or ***protein localization***. The proteins *move from one compartment to another by multiple mechanisms.*

The protein transport from the endoplasmic reticulum through the Golgi apparatus, and beyond uses carrier vesicles. It may be, however, noted that only the ***correctly folded proteins*** are recognized as the ***cargo for transport***. Protein targeting and post-translational modifications occur in a well coordinated manner.

Certain glycoproteins are targeted to reach lysosomes, as the lysosomal proteins can recognize the glycosidic compounds e.g. N-acetylglucosamine phosphate.

For the transport of secretory proteins, a special mechanism is operative. A ***signal peptide*** containing 15–35 amino acids, located at the amino terminal end of the secretory proteins facilitates the transport.

Protein targeting to mitochondria

Most of the proteins of mitochondria are synthesized in the cytosol, and their transport to mitochondria is a complex process. Majority of the proteins are synthesized as larger preproteins with N-terminal presequences for the entry of these proteins into mitochondria. The ***transport of unfolded proteins*** is often ***facilitated by chaperones***.

One protein namely ***mitochondrial matrix targeting signal***, involved in protein targeting has been identified. This protein can recognize mitochondrial receptor and transport certain proteins from cytosol to mitochondria. This is an energy-dependent process.

Protein targeting to other organelles

Specific signals for the transport of proteins to organelles such as nuclei and peroxisomes have been identified.

The smaller proteins can easily pass through nuclear pores. However, for larger proteins, ***nuclear localization signals*** are needed to *facilitate their entry into nucleus.*

MITOCHONDRIAL DNA, TRANSCRIPTION AND TRANSLATION

The mitochondrial DNA (***mtDNA***) has structural and functional resemblances with prokaryotic DNA. This fact supports the view that mitochondria are derivatives of prokaryotes. mtDNA is circular in nature and contains about 16,000 nucleotide bases.

A vast majority of structural and functional proteins of the mitochondria are synthesized in the cytosol, under the influence of nuclear DNA. However, certain proteins (around 13), most of them being the components of electron transport chain, are synthesized in the mitochondria (e.g. cytochrome b of complex III, two subunits of ATP synthase). Transcription takes place in the mitochondria leading to the synthesis of mRNAs, tRNAs and rRNAs. Two types of rRNA and about 22 species of tRNA have been so far identified. Transcription is followed by translation resulting in protein synthesis.

The mitochondria of the sperm cell do not enter the ovum during fertilization, therefore, ***mtDNA is inherited from the mother.*** Mitochondrial DNA is subjected to high rate of mutations (about 10 times more than nuclear DNA) that causes inherited defects in oxidative phosphorylation. The best known among them are certain mitochondrial myopathies and Leber's hereditary optic neuropathy. The latter is mostly found in males and is characterized by blindness due to loss of central vision as a result of neuroretinal degeneration. ***Leber's hereditary optic neuropathy*** is a consequence of single base mutation in mtDNA. Due to this, the amino acid histidine, in place of arginine, is incorporated into the enzyme NADH coenzyme Q reductase.

SUMMARY

1. *Transcription is the process in which RNA is synthesized from DNA, which is carried out in 3 stages–initiation, elongation and termination.*
2. *In case of prokaryotes, a single enzyme synthesizes all the RNAs. In eukaryotes, RNA polymerase I, II and III respectively catalyse the formation of rRNAs, mRNAs and tRNAs.*
3. *The primary mRNA transcript (i.e. hnRNA) undergoes post-transcriptional modifications e.g. base modifications, splicing etc.*
4. *Reverse transcription is the process of synthesizing DNA from RNA by the enzyme reverse transcriptase.*
5. *Biosynthesis of a protein or a polypeptide is known as translation. The amino acid sequence of a protein is determined by the triplet nucleoside base sequences of mRNA, arranged as codons.*
6. *The genetic code (codons)-composed of A, G, C and U–is universal, specific, non-overlapping and degenerate. Of the 64 codons, three (UAA, UAG, UGA) are termination codons while the rest code for amino acids.*
7. *Ribosomes are the factories of protein biosynthesis. Translation involves activation of amino acids, protein synthesis proper (initiation, elongation and termination), protein folding and post-translational modifications.*
8. *The post-translational modifications include proteolytic degradation, intein splicing and covalent modifications (phosphorylation, hydroxylation, glycosylation etc.). These modifications are required to make the proteins biologically active.*
9. *The proteins synthesized in translation reach the destination to exhibit their biological activity. This is carried out by a process called protein targeting or protein sorting.*
10. *The mitochondria possess independent DNA with the machinery for transcription and translation. However, only a few proteins (around 13) are actually synthesized in the mitochondria.*

SELF-ASSESSMENT EXERCISES

I. Essay questions

1. Give an account of transcription. Compare the RNA synthesis between prokaryotes and eukaryotes.
2. Describe protein biosynthesis (translation).
3. Discuss the inhibitors of transcription and translation.
4. Give an account of post-transcriptional and post-translational modifications.
5. What is genetic code? Describe the characteristics of genetic code. Add a note on the effects of mutations on genetic code.

II. Short notes

(a) Genome, (b) Heterogeneous nuclear RNA (hnRNA), (c) Eukaryotic RNA polymerases, (d) Introns and exons, (e) Reverse transcription, (f) Wobble hypothesis, (g) Anticodon, (h) Shine-Dalgarno sequence, (i) Peptidyltransferase, (j) Chaperones, (k) Protein targeting.

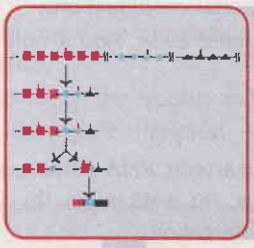
III. Fill in the blanks

1. The total DNA (genetic information) contained in an organism (or a cell) is referred to as _____.
2. The primary transcript produced by RNA polymerase II is eukaryotes _____.
3. The intervening nucleotide sequences in mRNA that do not code for proteins _____.
4. The synthesis of complementary DNA (cDNA) from mRNA is catalysed by the enzyme _____.
5. A single tRNA is capable of recognizing more than one codon, and this phenomenon is referred to as _____.
6. The factories for protein biosynthesis are _____.
7. The enzyme peptidyltransferase calalyses the formation of peptide bond during translation. The chemical nature of this enzyme is _____.
8. The proteins that facilitate the formation of specific conformation of proteins are _____.
9. The common term used for the diseases due to misfolding of proteins _____.
10. The process of delivery of proteins in a cell to the site their biological activity is _____.

IV. Multiple choice questions

11. The codon(s) that terminate(s) protein biosynthesis
    (a) UAA (b) UAG (c) UGA (d) All of them.
12. The nitrogenous base that is never found in the genetic code
    (a) Adenine (b) Guanine (c) Thymine (d) Cytosine.
13. The total DNA (genetic information) contained in a living cell (or organism) is regarded as
    (a) Genome (b) Transciptome (c) Proteome (d) Gene.
14. The enzyme responsible for the synthesis of mRNAs in eukaryotic cells
    (a) RNA polymerase I (b) RNA polymerase II (c) RNA polymerase III (d) RNA polymerase α.
15. Mitochondrial DNA is inherited from
    (a) Mother only (b) Father only (c) Both of them (d) Either mother or father.

26 Regulation of Gene Expression



The genes speak :

"Functional units of DNA, we are;
Ultimate for all cellular activities;
Tailored to express as per tissue demands;
Mystery of our molecular actions await unfolding."

DNA, the chemical vehicle of heredity, is composed of ***functional units***, namely ***genes***. The term ***genome*** refers to the total genetic information contained in a cell. The bacterium *Escherichia coli* contains about 4,400 genes present on a single chromosome. The ***genome of humans*** is more complex, with 23 pairs of (diploid) chromosomes containing 6 billion (6×10^9) base pairs of DNA, with an estimated ***30,000–40,000 genes***. At any given time, only a fraction of the genome is expressed.

The living cells possess a remarkable property to adapt to changes in the environment by regulating the gene expression. For instance, insulin is synthesized by specialized cells of pancreas and not by cells of other organs (say kidney, liver), although the nuclei of all the cells of the body contain the insulin genes. Molecular regulatory mechanisms facilitate the expression of insulin gene in pancreas, while preventing its expression in other cells.

GENE REGULATION—GENERAL

The ***regulation of the expression of genes*** is absolutely ***essential for the growth, development, differentiation*** and the very existence of an organism. There are two types of gene regulation-positive and negative.

1. **Positive regulation :** The gene regulation is said to be positive when its expression is increased by a regulatory element (positive regulator).

2. **Negative regulation :** A decrease in the gene expression due to the presence of a regulatory element (negative regulator) is referred to as negative regulation.

It may be noted here that double negative effect on gene regulation results in a positive phenomenon.

Constitutive and inducible genes

The genes are generally considered under two categories.

1. **Constitutive genes :** The products (proteins) of these genes are ***required all the time*** in a cell. Therefore, the constitutive genes (or ***housekeeping genes***) are expressed at more or less constant rate in almost all the cells and, further, they are not subjected to regulation e.g. the enzymes of citric acid cycle.

2. **Inducible genes :** The concentration of the proteins synthesized by inducible genes is ***regulated by*** various ***molecular signals***. An inducer increases the expression of these genes while a repressor decreases, e.g. tryptophan pyrrolase of liver is induced by tryptophan.

One cistron-one subunit concept

The chemical product of a gene expression is a protein which may be an enzyme. It was originally believed that each gene codes for a specific enzyme, leading to the popular concept, one gene-one enzyme. This however, is not necessarily valid due to the fact that several enzymes (or proteins) are composed of two or more nonidentical subunits (polypeptide chains).

The cistron is the smallest unit of genetic expression. It is the fragment of DNA coding for the subunit of a protein molecule. The original concept of ***one gene-one enzyme is replaced by one cistron-one subunit.***

Models for the study of gene expression

Elucidation of the regulation of gene expression in prokaryotes has largely helped to understand the principles of the flow of information from genes to mRNA to synthesize specific proteins. Some important features of prokaryotic gene expression are described first. This is followed by a brief account of eukaryotic gene expression.

THE OPERON CONCEPT

The operon is the ***coordinated unit of genetic expression in bacteria***. The concept of operon was introduced by Jacob and Monod in 1961 (Nobel Prize 1965), based on their observations on the regulation of lactose metabolism in *E. coli*. This is popularly known as ***lac operon***.

LACTOSE (LAC) OPERON

Structure of lac operon

The lac operon (***Fig.26.1***) consists of a regulatory gene (I; I for inhibition), operator gene (O) and three structural genes (Z, Y, A). Besides these genes, there is a promoter site (P), next to the operator gene, where the enzyme RNA polymerase binds. The structural genes Z, Y and A respectively, code for the enzymes β-galactosidase, galactoside permease and galactoside acetylase. β-Galactosidase hydrolyses lactose (β-galactoside) to galactose and glucose while permease is responsible for the transport of lactose into the cell. The function of acetylase (coded by A gene) remains a mystery.

The structural genes Z, Y and A transcribe into a single large mRNA with 3 independent translation units for the synthesis of 3 distinct enzymes. An mRNA coding for more than one protein is known as ***polycistronic mRNA***. Prokaryotic organisms contain a large number of polycistronic mRNAs.

Repression of lac operon

The regulatory gene (I) is constitutive. It is expressed at a constant rate leading to the synthesis of lac repressor. Lac repressor is a tetrameric (4 subunits) regulatory protein (total mol. wt. 150,000) which specifically binds to the operator gene (O). This prevents the binding of the enzyme RNA polymerase to the promoter site (P), thereby blocking the transcription of structural genes (Z, Y and A). This is what happens in the absence of lactose in *E. coli*. The ***repressor molecule acts as a negative regulator of gene expression.***

Derepression of lac operon

In the presence of lactose (inducer) in the medium, a small amount of it can enter the *E. coli* cells. The repressor molecules have a high affinity for lactose. The lactose molecules bind and induce a conformational change in the repressor. The result is that the repressor gets

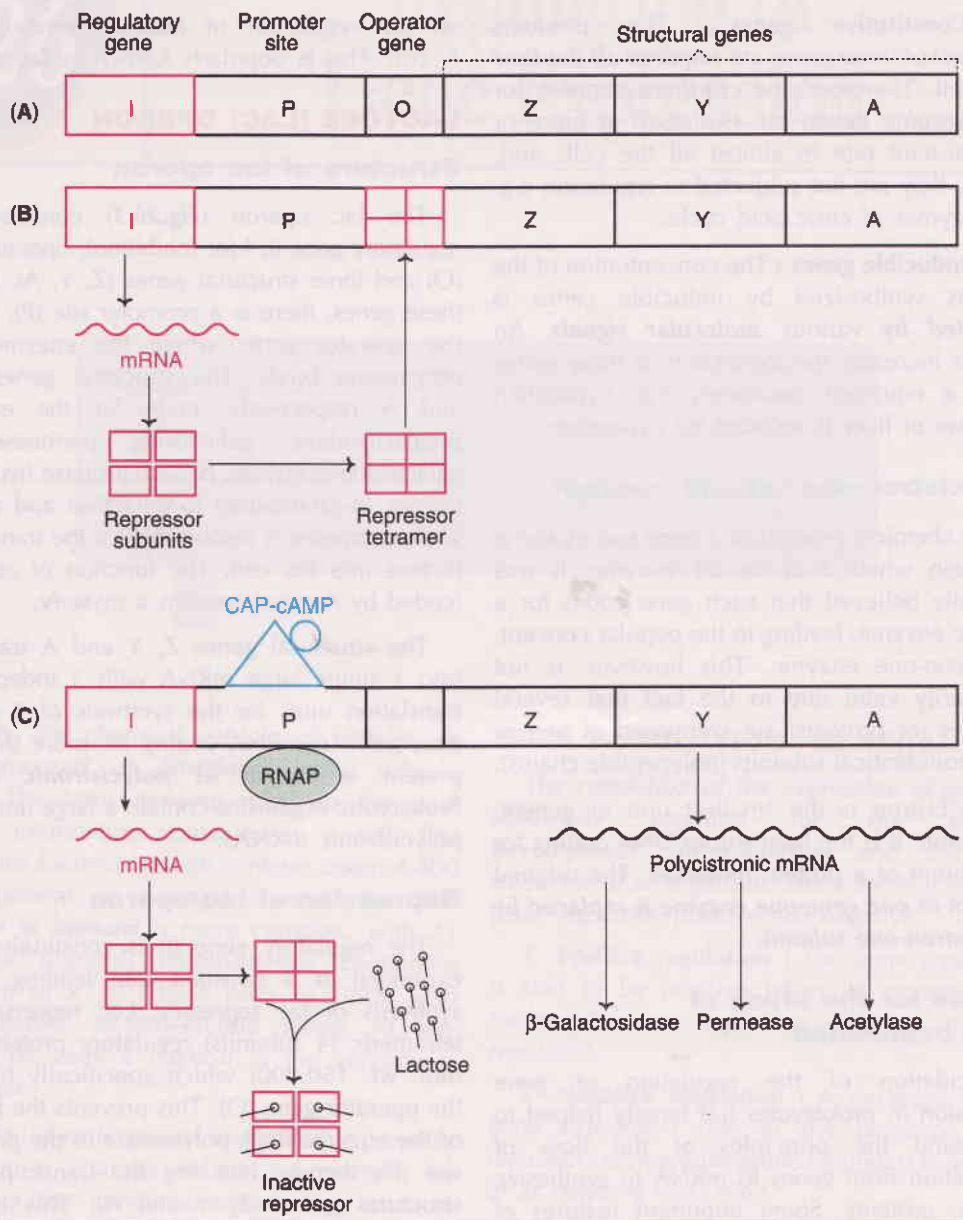


Fig. 26.1 : Model of lactose operon in E.coli *(**A**) Structure of lac operon (**B**) Repression of lac operon (**C**) Derepression of lac operon. (CAP—cAMP–catabolite gene activator protein bound to cAMP; RNAP–RNA polymerase).*

inactivated and, therefore, cannot bind to the operator gene (O). The RNA polymerase attaches to the DNA at the promoter site and transcription proceeds, leading to the formation of polycistronic mRNA (for genes Z, Y and A) and, finally, the 3 enzymes. Thus, lactose induces the synthesis of the three enzymes β-galactosidase, galactoside permease and galactoside acetylase. ***Lactose acts by inactivating the repressor molecules***, hence this process is known as derepression of lac operon.

Gratuitous inducers : There are certain structural analogs of lactose which can induce the lac operon but are not the substrates for the

enzyme β-galactosidase. Such substances are known as gratuitous inducers. Isopropyl-thiogalactoside (IPTG) is a gratuitous inducer, extensively used for the study of lac operon.

The catabolite gene activator protein : The cells of *E. coli* utilize glucose in preference to lactose; when both of them are present in the medium. After the depletion of glucose in the medium, utilization of lactose starts. This indicates that glucose somehow interferes with the induction of lac operon. This is explained as follows.

The attachment of RNA polymerase to the promoter site requires the presence of a ***catabolite gene activator protein (CAP)*** bound to cyclic AMP (***Fig.26.2***). The presence of glucose lowers the intracellular concentration of cAMP by inactivating the enzyme adenylyl cyclase responsible for the synthesis of cAMP. Due to the diminished levels of cAMP, the formation of CAP-cAMP is low. Therefore, the binding of RNA polymerase to DNA (due to the absence of CAP-cAMP) and the transcription are almost negligible in the presence of glucose. Thus, glucose interferes with the expression of lac operon by depleting cAMP levels. Addition of exogenous cAMP is found to initiate the transcription of many inducible operons, including lac operon.

It is now clear that the presence of CAP-cAMP is essential for the transcription of structural genes of lac operon. Thus, CAP-cAMP acts as a positive regulator for the gene expression. It is, therefore, evident that lac operon is subjected to both positive (by repressor, described above) and negative regulation.

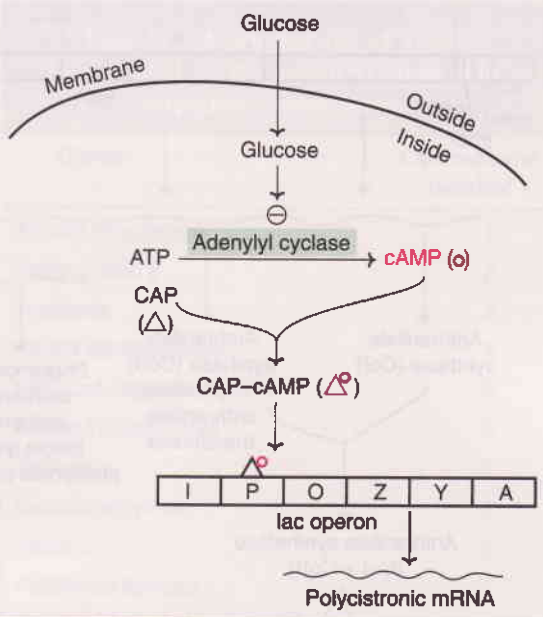


Fig. 26.2 : Control of lac operon by catabolite gene activator protein (CAP) and the role of glucose.

TRYPTOPHAN OPERON

Tryptophan is an aromatic amino acid, and is required for the synthesis of all proteins that contain tryptophan. If tryptophan is not present in the medium in adequate quantity, the bacterial cell has to make it, as it is required for the growth of the bacteria.

The tryptophan operon of *E. coli* is depicted in ***Fig.26.3***. This operon contains five structural genes (*trpE, trpD, trpC, trpB, trpA*), and the regulatory elements—primary promoter (*trpP*), operator (*trpO*), attenuator (*trpa*), secondary internal promoter ($TrpP_2$), and terminator (*trpt*).

The five structural genes of tryptophan operon code for three enzymes (two enzymes contain two different subunits) required for the synthesis of tryptophan from chlorismate.

The tryptophan repressor is always turned on, unless it is repressed by a specific molecule called corepressor. Thus lactose operon (described already) is inducible, whereas ***tryptophan operon is repressible***. The tryptophan operon is said to be derepressed when it is actively transcribed.

Tryptophan operon regulation by a repressor

Tryptophan acts as a corepressor to shut down the synthesis of enzymes from tryptophan operon. This is brought out in association with a specific protein, namely ***tryptophan repressor***.

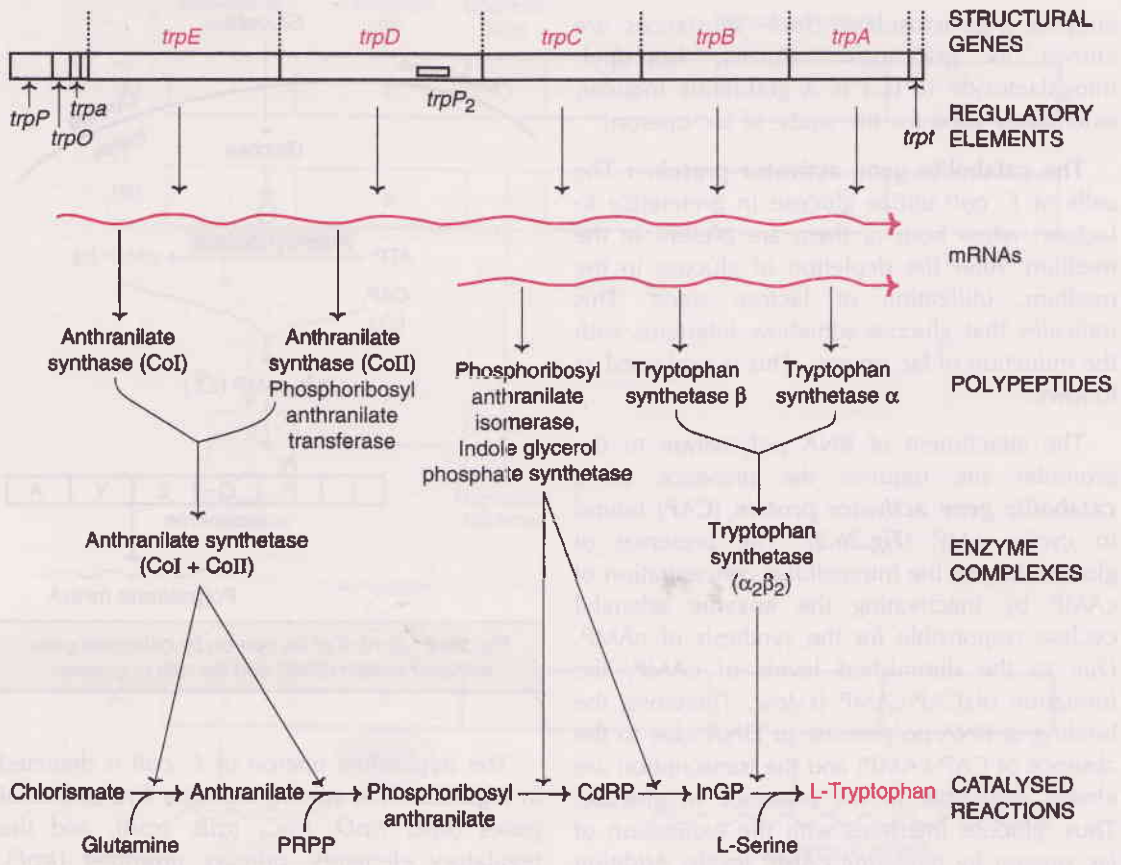


Fig. 26.3 : *Tryptophan operon in* E.coli *[regulatory elements are promoter (trpP), operator (trpO), attenuator (trpa), secondary internal promoter ($trpP_2$) and terminator (trpt); CoI, CoII–Component I and component II; PRPP – 5-Phosphoribosyl 1-pyrophosphate; CdRP–Carboxyl-phenylamino 1 deoxyribulose 5-phosphate; InGP–Indole 3-glycerol phosphate].*

Tryptophan repressor, a homodimer (contains two identical subunits) binds with two molecules of tryptophan, and then binds to the *trp* operator to turn off the transcription. It is of interest to note that tryptophan repressor also regulates the transcription of the gene (*trpR*) responsible for its own synthesis.

Two polycistronic mRNAs are produced from tryptophan operon—one derived from all the five structural genes, and the other obtained from the last three genes.

Besides acting as a corepressor to regulate tryptophan operon, tryptophan can inhibit the activity of the enzyme anthranilate synthetase. This is referred to as ***feedback inhibition***, and is brought out by binding of tryptophan at an allosteric site on anthranilate synthetase.

Attenuator as the second control site for tryptophan operon

Attenuator gene (*trpa*) of tryptophan operon lies upstream of *trpE* gene. Attenuation is the second level of regulation of tryptophan operon. The attenuator region provides RNA polymerase which regulates transcription. In the presence of tryptophan, transcription is prematurely terminated at the end of attenuator region. However, in the absence of tryptophan, the attenuator region has no effect on transcription. Therefore, the polycistronic mRNA of the five structural genes can be synthesized.

GENE EXPRESSION IN EUKARYOTES

Each cell of the higher organism contains the entire genome. As in prokaryotes, gene expression in eukaryotes is regulated to provide the appropriate response to biological needs. This may occur in the following ways

- Expression of certain genes (housekeeping genes) in most of the cells.
- Activation of selected genes upon demand.
- Permanent inactivation of several genes in all but a few types.

In case of prokaryotic cells, most of the DNA is organized into genes which can be transcribed. In contrast, in mammals, ***very little of the total DNA is organized into genes*** and their associated regulatory sequences. The function of the bulk of the extra DNA is not known.

Eukaryotic gene expression and its regulation are ***highly complex***. Some of the important aspects are briefly described.

CHROMATIN SRUCTURE AND GENE EXPRESSION

The DNA in higher organisms is extensively folded and packed to form ***protein-DNA complex*** called chromatin. The structural organization of DNA in the form of chromatin plays an important role in eukaryotic gene expression. In fact, chromatin structure provides an additional level of control of gene expression.

A selected list of genes (represented by the products) along with the respective chromosomes on which they are located is given in ***Table 26.1***.

In general, the genes that are transcribed within a particular cell are less condensed and more open in structure. This is in contrast to genes that are not transcribed which form highly condensed chromatin.

Histone acetylation and deacetylation

Eukaryotic DNA segments are wrapped around histone proteins to form nucleosome.

Table 26.1 A selected list of genes (represented by the products) along with respective chromosomes

Genes	*Chromosome number*
Alkaline phosphatase	1
Apolipoprotein B	2
Transferrin	3
Alcohol dehydrogenase	4
HMG CoA reductase	5
Steroid 21-hydroxylase	6
Arginase	7
Carbonic anhydrase	8
Interferon	9
Parathyroid hormone	11
Glyceraldehyde 3-phosphate dehydrogenase	12
Adenosine deaminase	13
α_1-Antitrypsin	14
Cytochrome P_{450}	15
Hemoglobin α-chain	16
Growth hormone	17
Prealbumin	18
Creatine phosphokinase (M chain)	19
Adenosine deaminase	20
Superoxide dismutase	21
Immunoglobulin (λ chain)	22
Glucose 6-phosphate dehydrogenase	X
Steroid sulfatase	Y

Acetylation or deacetylation of histones is an ***important factor in determining the gene expression***. In general, acetylation of histones leads to activation of gene expression while deacetylation reverses the effect.

Acetylation predominantly occurs on the lysine residues in the amino terminal ends of histones. This modification in histones reduces the positive charges of terminal ends (tails), and decreases their binding affinity to negatively charged DNA. Consequently, nucleosome structure is disrupted to allow transcription.

Methylation of DNA and inactivation of genes

Cytosine in the sequence CG of DNA gets methylated to form 5'-methylcytosine. A major portion of CG sequences (about 20%) in human DNA exists in methylated form. In general, ***methylation leads to*** loss of transcriptional activity, and thus ***inactivation of genes.*** This occurs due to binding of methylcytosine binding proteins to methylated DNA. As a result, methylated DNA is not exposed and bound to transcription factors. It is interesting to note that methylation of DNA correlates with deacetylation of histones. This provides a double means for repression of genes.

The activation and normal expression of genes, and gene inactivation by DNA methylation are depicted in ***Fig.26.4.***

ENHANCERS AND TISSUE-SPECIFIC GENE EXPRESSION

Enhancers (or activators) are DNA elements that facilitate or ***enhance gene expression.*** The enhancers provide binding sites for specific proteins that regulate transcription. They facilitate binding of the transcription complex to promoter regions.

Some of the enhancers possess the ability to promote transcription in a tissue-specific manner. For instance, gene expression in lymphoid cells for the production immunoglobulins (Ig) is promoted by the enhancer associated with Ig genes between J and C regions.

Transgenic animals are frequently used for the study of tissue-specific expression. The available evidence from various studies indicates that the ***tissue-specific gene expression is largely mediated through the involvement of enhancers.***

COMBINATION OF DNA ELEMENTS AND PROTEINS IN GENE EXPRESSION

Gene expression in mammals is a complicated process with several environmental stimuli on a single gene. The ultimate response of the gene which may be positive or negative is brought out by the association of DNA elements and proteins.

In the illustration given in the ***Fig.26.5***, gene I is activated by a combination of activators 1, 2 and 3. Gene II is more effectively activated by the combined action of 1, 3 and 4. Activator 4 is not in direct contact with DNA, but it forms a bridge between activators 1 and 3, and activates gene II. As regards gene III, it gets inactivated by a combination of 1, 5 and 3. In this case, protein 5 interferes with the binding of protein 2 with the DNA, and inactivates the gene.

MOTIFS IN PROTEINS AND GENE EXPRESSION

A motif literally means a dominant element. Certain motifs in proteins mediate the binding of regulatory proteins (transcription factors) to

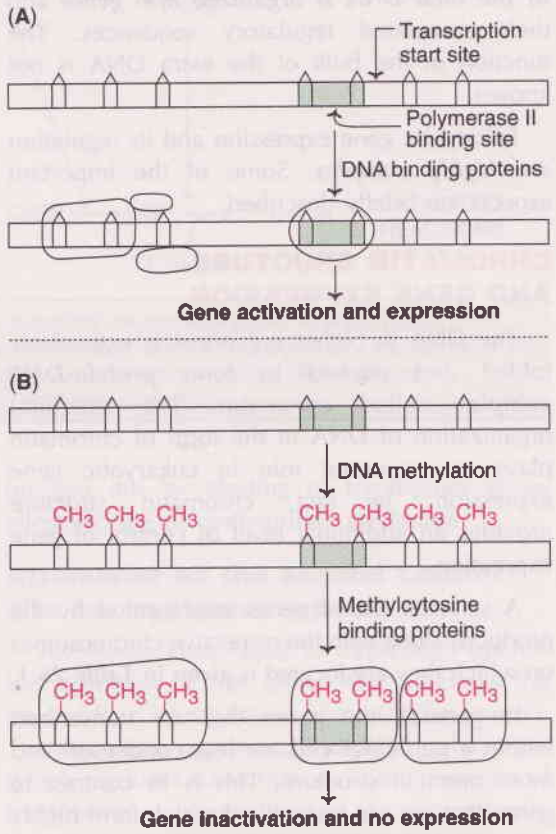


Fig. 26.4 : Methylation of DNA and inactivation of genes (A) Gene activation in the absence of DNA methylation (B) Gene inactivation due to methylation (⬠ represent CG sequences).

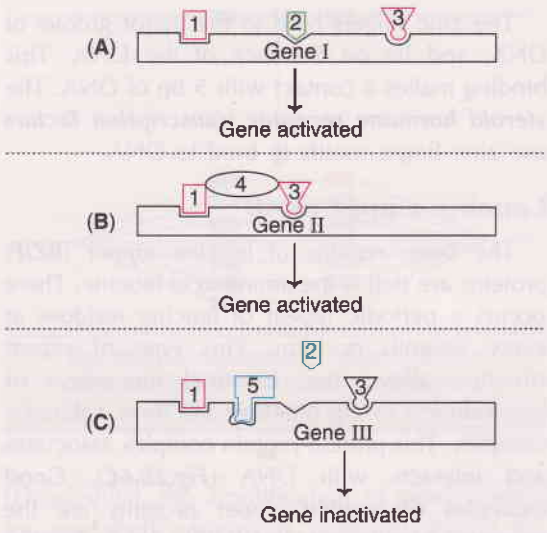


Fig. 26.5 : A diagrammatic representation of the association of DNA elements and proteins in gene regulation. **A**, **B** *and* **C** *represent genes I, II and III (1...5 represent proteins).*

DNA. The specific control of transcription occurs by the binding of regulatory proteins with high affinity to the correct regions of DNA.

A great majority of specific ***protein-DNA interactions*** are brought out by four unique motifs—helix-turn-helix (HTH), zinc finger, leucine zipper, helix-loop-helix (HLH).

These ***amino acid motifs*** bind with high affinity to the specific site and low affinity to other parts of DNA. The motif-DNA interactions are maintained by hydrogen bonds and van der Waals forces.

Helix-turn-helix motif

The helix-turn-helix (***HTH***) motif is about 20 amino acids which represents a small part of a large protein. HTH is the domain part of the protein which specifically interacts with the DNA (***Fig.26.6A***). Examples of helix-turn-helix motif proteins include lactose repressor, and cyclic AMP catabolite activator protein (CAP) of *E. coli*, and several developmentally important transcription factors in mammals.

Zinc finger motif

Sometime ago, it was recognized that the transcription factor TFIIIA requires zinc for its

BIOMEDICAL / CLINICAL CONCEPTS

- *Regulation of gene expression to adapt to the changes in the environment is a remarkable property of living cells e.g. synthesis of insulin by β-cells of pancreas and nowhere else.*
- *The growth, development and differentiation of an organism involves complex mechanisms which ultimately depend on gene regulation.*
- *The house-keeping genes or constitutive genes are expressed at almost a constant rate in the cells, and they are not usually subjected to regulations e.g. enzymes of Krebs cycle.*
- *The malignant cells develop drug resistance to long term administration of methotrexate. This occurs by amplification of the genes coding for dihydrofolate reductase.*
- *The human body has the capability to produce around 10 billion antigen-specific immunoglobulins. This is achieved by a process called gene rearrangement.*
- *Knowledge on the gene expression and its regulation helps in the understanding and control of several diseases, including cancer.*

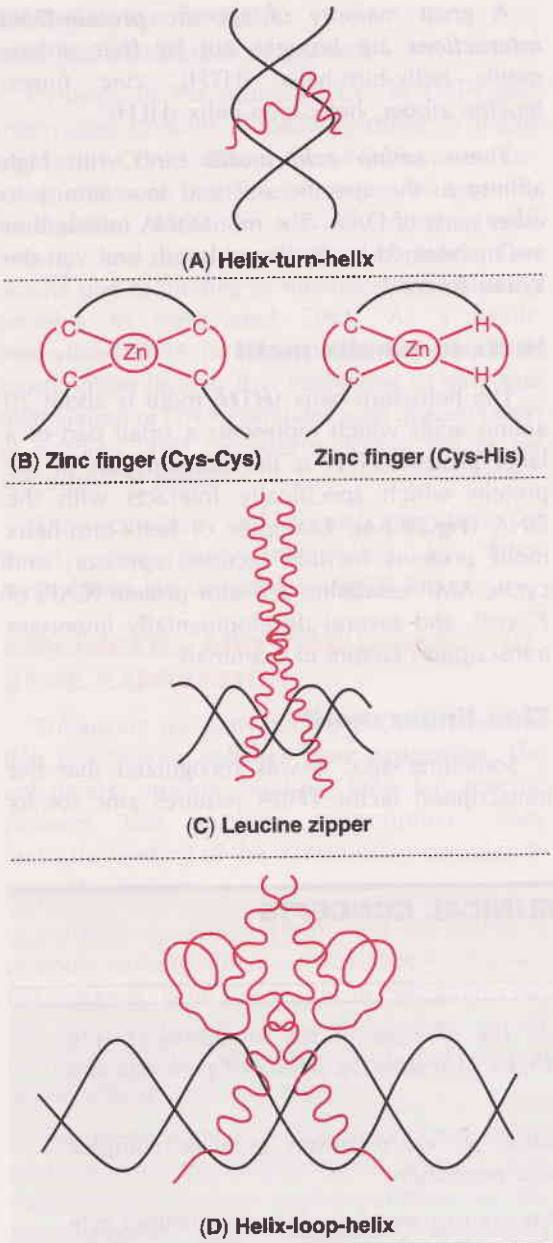


Fig. 26.6 : A diagrammatic representation of common motifs in proteins interacting with DNA.

activity. On analysis, it was revealed that each TFIIIA contains zinc ions as a repeating coordinated complex. This complex is formed by the closely spaced amino acids cysteine and cysteine, followed by a histidine—histidine pair. In some instances, His-His is replaced by a second Cys-Cys pair (***Fig.26.6B***).

The zinc fingers bind to the major groove of DNA, and lie on the face of the DNA. This binding makes a contact with 5 bp of DNA. The ***steroid hormone receptor transcription factors*** use zinc finger motifs to bind to DNA.

Leucine zipper motif

The ***b***asic regions of leucine ***zip***per (***bZIP***) proteins are rich is the amino acid leucine. There occurs a periodic repeat of leucine residues at every seventh position. This type of repeat structure allows two identical monomers or heterodimers to ***zip together*** and form a dimeric complex. This protein-protein complex associates and interacts with DNA (***Fig.26.6C***). Good examples of leucine zipper proteins are the enhancer binding proteins (EBP)—fos and jun.

Helix-loop-helix motif

Two amphipathic (literally means a feeling of closeness) α-helical segments of proteins can form helix-loop-helix motif and bind to DNA. The dimeric form of the protein actually binds to DNA (***Fig.26.6D***).

GENE REGULATION IN EUKARYOTES

The important features of eukaryotic gene expression along with the regulatory aspects are described in the preceeding pages. Besides transcription, eukaryotic cells also employ variety of other mechanisms to regulate gene expression. The most important ones are listed below, and briefly described next.

1. Gene amplification
2. Gene rearrangement
3. Processing of RNA
4. Alternate mRNA splicing
5. Transport of mRNA from nucleus to cytoplasm
6. Degradation of mRNA.

Gene amplification

In this mechanism, the expression of a gene is increased several fold. This is commonly observed during the developmental stages of eukaryotic organisms. For instance, in fruit fly

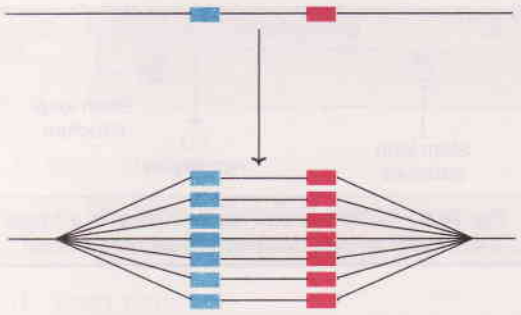


Fig. 26.7 : A diagrammatic representation of gene amplification (the genes are depicted in colour shade and colour).

(*Drosophila*), the amplification of genes coding for egg shell proteins is observed during the course of oogenesis. The amplification of the gene (DNA) can be observed under electron microscope (***Fig.26.7***).

The occurrence of gene amplification has also been reported in humans. Methotrexate is an anticancer drug which inhibits the enzyme ***dihydrofolate reductase***. The malignant cells develop drug resistance to long term administration of methotrexate by amplifying the genes coding for dihydrofolate reductase.

Gene rearrangment

The body possesses an enormous capacity to synthesize a wide range of antibodies. It is estimated that the human body can produce about 10 billion (10^{10}) antibodies in response to antigen stimulations. The molecular mechanism of this antibody diversity was not understood for long. It is now explained on the basis of gene rearrangement or ***transposition of genes*** or ***somatic recombination of DNA***.

The structure of a typical immunoglobulin molecule consists of two light (L) and two heavy (H) chains. Each one of these chains (L or H) contains an N-terminal variable (V) and C-terminal constant (C) regions (***Refer Fig.9.3***). The V regions of immunoglobulins are responsible for the recognition of antigens. The phenomenon of gene rearrangement can be understood from the mechanism of the synthesis of light chains of immunoglobulins (***Fig.26.8***).

Each light chain can be synthesized by three distinct DNA segments, namely the variable (V_L), the joining (J_L) and the constant (C_L). The mammalian genome contains about 500 V_L segments, 6 J_L segments and 20 C_L segments. During the course of differentiation of B-lymphocytes, one V_L segment (out of the 500) is brought closer to J_L and C_L segments. This occurs on the same chromosome. For the sake of illustration, 100^{th} V_L, 3^{rd} J_L and 10^{th} C_L segments are rearranged in ***Fig.26.8***. The rearranged DNA (with V_L, J_L and C_L fragments) is then transcribed to produce a single mRNA for the synthesis of a specific light chain of the antibody. By

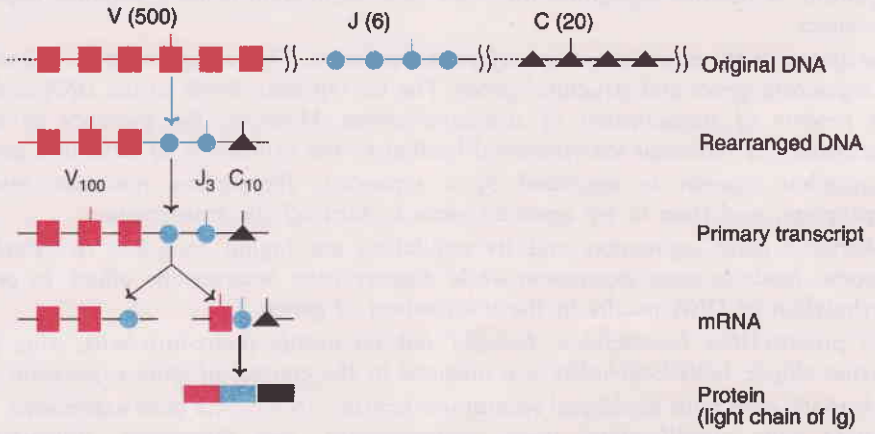


Fig. 26.8 : A diagrammatic representation of gene rearrangement for the synthesis of light chain of immunoglobulin.

innumerable combinations of V_L, J_L and C_L segments, the body's immune system can generate millions of antigen specific immunoglobulin molecules.

The formation of heavy (H) chains of immunoglobulins also occurs by rearrangement of 4 distinct genes—variable (V_H), diversity (D), joining (J_H) and constant (C_H).

Processing of RNA

The RNA synthesized in transcription undergoes modifications resulting in a functional RNA. The changes include intron-exon splicing, polyadenylation etc. (***Chapter 25***).

Alternate mRNA splicing

Eukaryotic cells are capable of carrying out alternate mRNA processing to control gene expression. Different mRNAs can be produced by alternate splicing which code for different proteins (for more details, ***Refer Chapter 25***).

Degradation of mRNA

The expression of genes is indirectly influenced by the stability of mRNA. Certain hormones regulate the synthesis and degradation of some mRNAs. For instance, estradiol prolongs the half-life of vitellogenin mRNA from a few hours to about 200 hours.

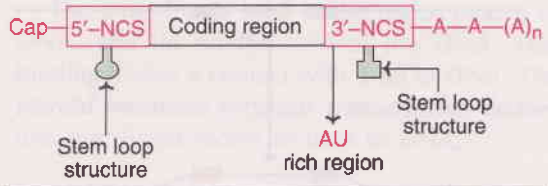


Fig. 26.9 : A diagrammatic representation of a typical eukaryotic mRNA. (NCS-Non-coding sequences)

It appears that the ends of mRNA molecules determine the stability of mRNA. A typical eukaryotic mRNA has 5'-non-coding sequences (5'-NCS), a coding region and a 3'-NCS. All the mRNAs are capped at the 5' end, and most of them have a polyadenylate sequence at the 3' end (***Fig.26.9***). The 5' cap and poly (A) tail protect the mRNA against the attack by exonuclease. Further, stem-loop structures in NCS regions, and AU rich regions in the 3' NCS also provide stability to mRNA.

SUMMARY

1. *DNA, the chemical vehicle of heredity, is composed of genes. The regulation of gene expression is absolutely essential for the growth, development and differentiation of an organism. A positive regulation increases gene expression while a negative regulation decreases.*
2. *The operon is the coordinated unit of gene expression. The lac operon of* E. coli *consists of regulatory genes and structural genes. The lac repressor binds to the DNA and halts the process of transcription of structural genes. However, the presence of lactose inactivates the repressor (derepression) leading to the expression of structural genes.*
3. *Tryptophan operon is regulated by a repressor. Tryptophan repressor binds to tryptophan, and then to trp operator gene to turn off the transcription.*
4. *Eukaryotic gene expression and its regulation are highly complex. Acetylation of histones leads to gene expression while deacetylation reverses the effect. In general, methylation of DNA results in the inactivation of genes.*
5. *The protein-DNA interactions, brought out by motifs (helix-turn-helix, zinc finger, leucine zipper, helix-loop-helix), are involved in the control of gene expression.*
6. *Eukaryotic cells have developed several mechanisms to regulate gene expression. These include gene amplification, gene rearrangement, and processing, transport and degradation of DNA.*

SELF-ASSESSMENT EXERCISES

I. Essay questions

1. Describe lactose (lac) operon.
2. Write briefly on the gene expression and its regulation in eukaryotes.

II. Short notes

(a) One cistron-one subunit concept, (b) Catabolite gene activator protein, (c) Gene inactivation by DNA methylation, (d) Zinc finger motif, (e) Gene amplification.

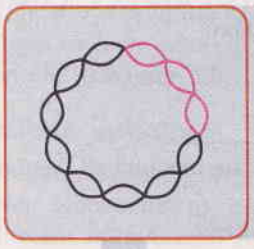
III. Fill in the blanks

1. The number of genes found in human genome _____.
2. The genes responsible for the production of proteins that are required all the time in a cell are regarded as _____.
3. The earlier concept of one gene-one enzyme is replaced by _____.
4. The chromatin in higher organisms is chemically composed of _____.

IV. Multiple choice questions

5. The structural 'Z' gene of lactose (lac) operon is responsible for the synthesis of the enzyme(s)
   (a) β-Galactosidase (b) Permease (c) Acetylase (d) All of them.
6. Methylation of DNA results in
   (a) Activation of genes (b) Inactivation of genes (c) No effect on genes (d) Inactivation of protein motifs.
7. The production of a wide range of immunoglobulins is explained on the basis of
   (a) Gene amplification (b) Gene rearrangement (c) Alternate mRNA splicing (d) mRNA degradation.
8. The specific control of transcription involves the following motif(s)
   (a) Helix-turn-helix (b) Zinc finger (c) Leucine zipper (d) All of them.

27 Recombinant DNA and Biotechnology



The recombinant DNA speaks :

"I am the hybridized DNA molecule;
Created by cutting and sealing;
When introduced into host cells;
I multiply and code for desired proteins."

The term biotechnology represents a fusion or an alliance between ***bio***logy and ***technology***. Frankly speaking, biotechnology is a newly discovered discipline for age-old practices e.g. preparation of wine, beer, curd, bread. These natural processes are regarded as ***old*** or ***traditional biotechnology***.

The ***new*** or ***modern biotechnology*** embraces all the ***genetic manipulations***, cell fusion techniques, and improvements made in the old biotechnological processes. The biotechnology with particular reference to recombinant DNA in human health and disease is briefly described in this chapter.

Genetic engineering primarily involves the ***manipulation of genetic material (DNA)*** to achieve the desired goal in a pre-determined way. Some ***other terms*** are also in common use to describe genetic engineering.

- ***Gene manipulation***
- ***Recombinant DNA (rDNA) technology***
- ***Gene cloning (molecular cloning)***
- ***Genetic modifications***
- ***New genetics.***

Brief history of recombinant DNA technology

The present day DNA technology has its roots in the experiments performed by ***Boyer and Cohen in 1973***. In their experiments, they successfully recombined two plasmids (pSC 101 and pSC 102) and cloned the new plasmid in *E.coli*. In the later experiments the genes of a frog could be successfully transplanted, and expressed in *E.coli*. This made the real beginning of modern rDNA technology and laid foundations for the present day ***molecular biotechnology***.

Some biotechnologists who admire Boyer-Cohen experiments divide the subject into two chronological categories.

1. ***BBC***-biotechnology ***B***efore ***B***oyer and ***C***ohen.

2. ***ABC***-biotechnology ***A***fter ***B***oyer and ***C***ohen.

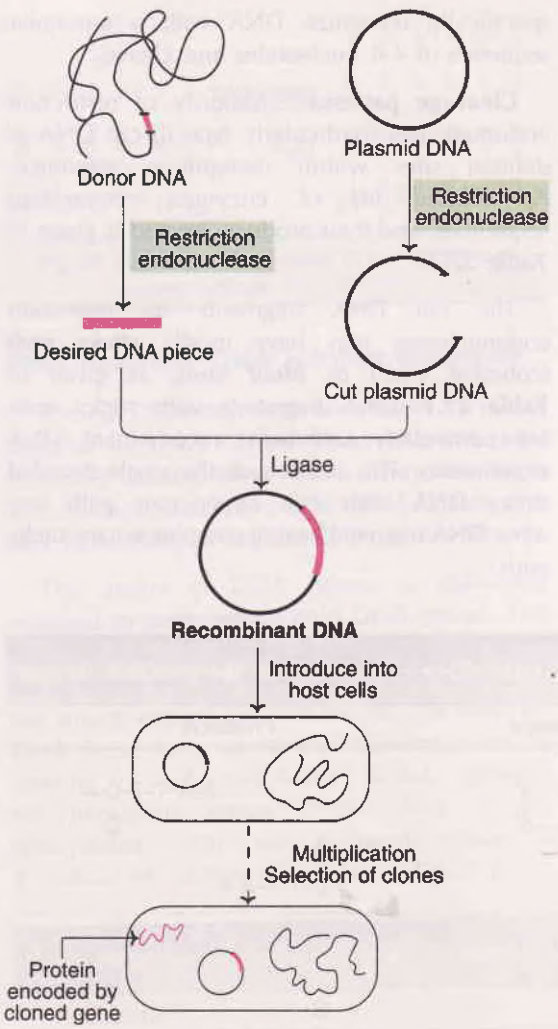


***Fig. 27.1** : The basic principle of recombinant DNA technology.*

Recombinant DNA technology is a vast field. The basic principles and techniques of rDNA technology along with the most important applications are briefly described in this chapter.

BASIC PRINCIPLES OF rDNA TECHNOLOGY

There are many diverse and complex techniques involved in gene manipulation. However, the basic principles of recombinant DNA technology are reasonably simple, and broadly involve the following stages (***Fig.27.1***).

1. Generation of DNA fragments and selection of the desired piece of DNA (e.g. a human gene).

2. Insertion of the selected DNA into a cloning vector (e.g. a plasmid) to create a ***recombinant DNA*** or ***chimeric DNA*** (Chimera is a monster in Greek mythology that has a lion's head, a goat's body and a serpent's tail. This may be comparable to Narasimha in Indian mythology).

3. Introduction of the recombinant vectors into host cells (e.g. bacteria).

4. Multiplication and selection of clones containing the recombinant molecules.

5. Expression of the gene to produce the desired product.

Recombinant DNA technology with special reference to the following aspects is described

- Molecular tools of genetic engineering.
- Host cells-the factories of cloning.
- Vectors-the cloning vehicles.
- Methods of gene transfer.
- Gene cloning strategies.

MOLECULAR TOOLS OF GENETIC ENGINEERING

The term genetic engineer may be appropriate for an individual who is involved in genetic manipulations. The ***genetic engineer's toolkit*** or molecular tools namely the ***enzymes*** most commonly used in recombinant DNA experiments are briefly described.

Restriction endonucleases– DNA cutting enzymes

Restriction endonucleases are one of the most important groups of enzymes for the manipulation of DNA. These are the ***bacterial enzymes*** that can ***cut/split DNA*** (from any source) at specific sites. They were first discovered in *E.coli* restricting the replication of bacteriophages, by cutting the viral DNA (The host *E.coli* DNA is protected from cleavage by addition of methyl groups). Thus, the enzymes that restrict the viral replication are known as ***restriction enzymes*** or restriction endonucleases.

Nomenclature : Restriction endonucleases are named by a standard procedure, with particular reference to the bacteria from which they are isolated. The first letter (in italics) of the enzymes indicates the genus name, followed by the first two letters (also in italics) of the species, then comes the strain of the organism and finally a Roman numeral indicating the order of discovery. A couple of examples are given below.

EcoRI is from *Escherichia* (***E***) *coli* (***co***), strain Ry13 (***R***), and first endonuclease (***I***) to be discovered. ***HindIII*** is from *Haemophilus* (***H***) *influenzae* (***in***), strain Rd (***d***) and, the third endonucleases (***III***) to be discovered.

Recognition sequences : Recognition sequence is the ***site where the DNA is cut*** by a restriction endonuclease. Restriction endonucleases can specifically recognize DNA with a particular sequence of 4-8 nucleotides and cleave.

Cleavage patterns : Majority of restriction endonucleases (particularly type II) cut DNA at defined sites within recognition sequence. A selected list of enzymes, recognition sequences, and their products formed is given in ***Table 27.1.***

The cut DNA fragments by restriction endonucleases may have mostly ***sticky ends*** (cohesive ends) or ***blunt ends***, as given in ***Table 27.1***. DNA fragments with sticky ends are particularly useful for recombinant DNA experiments. This is because the single-stranded sticky DNA ends can easily pair with any other DNA fragment having complementary sticky ends.

TABLE 27.1 Some restriction enzymes with sources, recognition sequences and the products formed

Enzyme (source)	*Recognition sequence*	*Products*
EcoRI (*Escherichia coli*)	5′.....G✂A–A–T–T–C.....3′ 3′.....C–T–T–A–A✂G.....5′	A–A–T–T–C..... G..... .....G C–T–T–A–A
BamHI (*Bacillus amyloliquefaciens*)	5′.....G✂G–A–T–C–C.....3′ 3′.....C–C–T–A–G✂G.....5′	G–A–T–C–C..... G..... .....G .....C–C–T–A–G
HaeIII (*Haemophilus aegyptius*)	5′.....G–G✂C–C.....3′ 3′.....C–C✂G–G.....5′	*C–C..... G–G..... .....*G–G ..... C–C
HindIII (*Haemophilus influenzae*)	5′.....A✂A–G–C–T–T.....3′ 3′.....T–T–C–G–A✂A.....5′	A–G–C–T–T..... A..... .....A .....T–T–C–G–A
NotI (*Nocardia otitidis*)	5′.....G–C✂G–G–C–C–G–C.....3′ 3′.....C–G–C–C–G–G✂C–G.....5′	G–G–C–C–G–C..... C–G..... .....G–C .....C–G–C–C–G–G.....

(***Note :*** *Scissors indicate the sites of cleavage.* * *The products are with blunt ends while for the rest, the products are with sticky ends).*

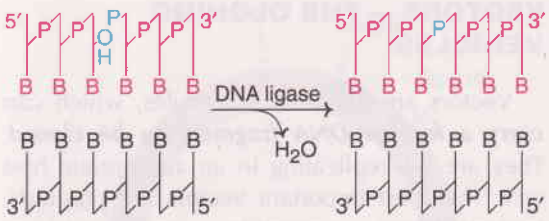


Fig. 27.2 : Action of DNA ligase in the formation of phosphodiester bond (B-base).

DNA ligases—DNA joining enzymes

The cut DNA fragments are covalently joined together by DNA ligases. These enzymes were originally isolated from viruses. They also occur in *E.coli* and eukaryotic cells. DNA ligases actively participate in cellular DNA repair process.

The action of DNA ligases is absolutely required to permanently hold DNA pieces. This is so since the hydrogen bonds formed between the complementary bases (of DNA strands) are not strong enough to hold the strands together. DNA ligase joins (seals) the DNA fragments by forming a phosphodiester bond between the phosphate group of 5′-carbon of one deoxyribose with the hydroxyl group of 3′-carbon of another deoxyribose (***Fig.27.2***).

Many enzymes are used in the recombinant DNA technology/genetic engineering. A selected list of these enzymes and the reactions catalysed by them is given in ***Table 27.2***.

HOST CELLS — THE FACTORIES OF CLONING

The hosts are the ***living systems or cells*** in which the carrier of recombinant DNA molecule or ***vector can be propagated***. There are different types of host cells-prokaryotic (bacteria) and eukaryotic (fungi, animals and plants). Some examples of host cells used in genetic engineering are given in ***Table 27.3***.

Host cells, besides effectively incorporating the vector's genetic material, must be conveniently cultivated in the laboratory to collect the products. In general, ***microorganisms*** are ***preferred*** as host cells, since they ***multiply faster*** compared to cells of higher organisms (plants or animals).

Prokaryotic hosts

Escherichia coli : The bacterium, *Escherichia coli* was the first organism used in the DNA technology experiments and continues to be ***the host of choice*** by many workers.

Table 27.2 The most commonly used enzymes in recombinant DNA technology/genetic engineering

Enzyme	*Use/reaction*
Alkaline phosphatase	Removes phosphate groups from 5′-ends of double/single-stranded DNA (or RNA).
Bal 31 nuclease	For the progressive shortening of DNA.
DNA ligase	Joins DNA molecules by forming phosphodiester linkages between DNA segments.
DNA polymerase I	Synthesizes DNA complementary to a DNA template.
DNase I	Produces single-stranded nicks in DNA.
Exonuclease III	Removes nucleotides from 3′-end of DNA.
λ exonuclease	Removes nucleotides from 5′-end of DNA.
Polynucleotide kinase	Transfers phosphate from ATP to 5′-OH ends of DNA or RNA.
Restriction enzymes	Cut double-stranded DNA with a specific recognition site.
Reverse transcriptase	Synthesizes DNA from RNA.
RNase A	Cleaves and digests RNA (and not DNA).
RNase H	Cleaves and digests the RNA strand of RNA-DNA heteroduplex.
Taq DNA polymerase	Used in polymerase chain reaction
SI nuclease	Degrades single-stranded DNA and RNA.
Terminal transferase	Adds nucleotides to the 3′-ends of DNA or RNA. Useful in homopolymer tailing.

Table 27.3 Some examples of host cells used in genetic engineering

Group	Examples
Prokaryotic	
Bacteria	*Escherichia coli*
	Bacillus subtilis
	Streptomyces sp
Eukaryotic	
Fungi	*Saccharomyces cerevisiae*
	Aspergillus nidulans
Animals	Insect cells
	Oocytes
	Mammalian cells
	Whole organisms
Plants	Protoplasts
	Intact cells
	Whole plants

The ***major drawback*** however, is that *E.coli* (or even other prokaryotic organisms) ***cannot perform post-translational modifications.***

Bacillus subtilis : *Bacillus subtilis* is a rod shaped non-pathogenic bacterium. It has been used as a host in industry for the production of enzymes, antibiotics, insecticides etc. Some workers consider *B.subtilis* as an ***alternative to E.coli.***

Eukaryotic hosts

Eukaryotic organisms are preferred to produce human proteins since these hosts with complex structure (with distinct organelles) are more suitable to synthesize complex proteins. The ***most commonly used*** eukaryotic organism is the ***yeast, Saccharomyces cerevisiae.***

Mammalian cells : Despite the practical difficulties to work with and high cost factor, mammalian cells (such as mouse cells) are also employed as hosts. The advantage is that certain complex proteins which cannot be synthesized by bacteria can be produced by mammalian cells e.g. tissue plasminogen activator. This is mainly because the mammalian cells ***possess the machinery to modify the protein to the active form*** (post-translational modifications).

VECTORS — THE CLONING VEHICLES

Vectors are the DNA molecules, which can ***carry a foreign DNA fragment to be cloned.*** They are self-replicating in an appropriate host cell. The most important vectors are plasmids, bacteriophages, cosmids and artificial chromosome vectors.

Plasmid

Plasmids are ***extrachromosomal,*** double-stranded, circular, ***self-replicating DNA*** molecules. Almost all the bacteria have plasmids containing a low copy number (1-4 per cell) or a high copy number (10-100 per cell). The size of the plasmids varies from 1 to 500 kb. Usually, plasmids contribute to about 0.5 to 5.0% of the total DNA of bacteria (**Note :** A few bacteria contain linear plasmids e.g. *Streptomyces* sp, *Borella burgdorferi*).

Nomenclature of plasmids : It is a common practice to designate ***p***lasmid by a lower case ***p,*** followed by the first letter(s) of researcher(s) names and the numerical number given by the workers. Thus, ***pBR322*** is a ***p***lasmid discovered by ***B***olivar and ***R***odriguez who designated it as ***322.*** Some plasmids are given names of the places where they are discovered e.g. ***pUC*** is ***p***lasmid from ***U***niversity of ***C***alifornia.

pBR322 – the most common plasmid vector : pBR322 of *E.coli* is the most popular and widely used plasmid vector, and is appropriately regarded as the parent or grand parent of several other vectors.

pBR322 has a DNA sequence of 4,361 bp. It carries genes resistance for ampicillin (Amp^r) and tetracycline (Tel^r) that serve as markers for the identification of clones carrying plasmids. The plasmid has unique recognition sites for the action of restriction endonucleases such as *EcoRI, HindIII, BamHI, SalI* and *PstI* (***Fig.27.3***).

Other plasmid cloning vectors : The other plasmids employed as cloning vectors include pUC19 (2,686 bp, with ampicillin resistance gene), and derivatives of pBR322–pBR325, pBR328 and pBR329.

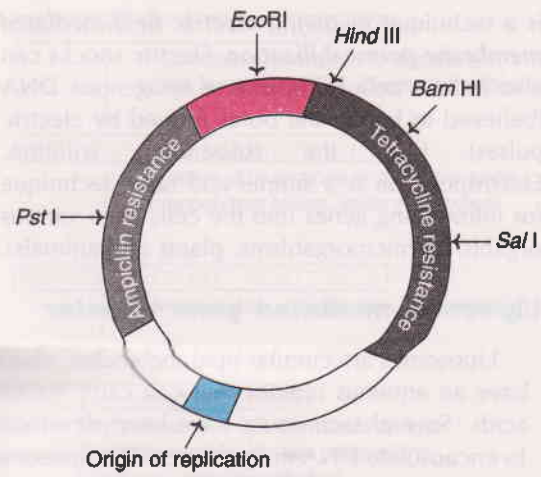


Fig. 27.3 : Genetic map of plasmid cloning vector pBR322.

Bacteriophages

Bacteriophages or simply ***phages*** are the ***viruses*** that ***replicate within the bacteria***. In case of certain phages, their DNA gets incorporated into the bacterial chromosome and remains there permanently. Phage vectors can accept short fragments of foreign DNA into their genomes. The advantage with phages is that they ***can take up larger DNA segments than plasmids***. Hence phage vectors are preferred for working with genomes of human cells. The most commonly used phages are bacteriophage λ (phage λ) and bacteriophage (phage M_{13}).

Cosmids

Cosmids are the vectors possessing the characteristics of both plasmid and bacteriophage λ. ***Cosmids*** can be constructed by adding a fragment of phage λ DNA including ***cos*** site, to plas***mids***. A foreign DNA (about 40 kb) can be inserted into cosmid DNA. The recombinant DNA so formed can be packed as phages and injected into *E.coli*. Once inside the host cell, cosmids behave just like plasmids and replicate. The advantage with cosmids is that they can carry larger fragments of foreign DNA compared to plasmids.

Artificial chromosome vectors

Human artificial chromosome (HAC) : Developed in 1997 (by H. Willard), human artificial chromosome is a ***synthetically produced vector DNA, possessing the characteristics of human chromosome***. HAC may be considered as a self-replicating microchromosome with a size ranging from 1/10th to $^1/_5$th of a human chromosome. The advantage with HAC is that it can carry human genes that are too long. Further, HAC can carry genes to be introduced into the cells in gene therapy.

Yeast artificial chromosomes (YACs) : Introduced in 1987 (by M. Olson), yeast artificial chromosome (YAC) is a synthetic DNA that can accept large fragments of foreign DNA (particularly human DNA). It is thus possible to clone large DNA pieces by using YAC.

Bacterial artificial chromosomes (BACs) : The construction of BACs is based on one F-plasmid which is larger than the other plasmids used as cloning vectors. BACs can accept DNA inserts of around 300 kb.

Choice of vector

Among the several factors, the size of the foreign DNA is very important in the choice of vectors. The efficiency of this process is often crucial for determining the success of cloning. The sizes of DNA insert that can be accepted by different vectors is shown in ***Table 27.4***.

METHODS OF GENE TRANSFER

Introducing a foreign DNA (i.e. the gene) into the cells is an important task in biotechnology. The efficiency of this process is often crucial for determining the success of cloning. The most commonly employed gene transfer methods, namely transformation, conjugation, electroporation and lipofection, and direct transfer of DNA are briefly described.

Transformation

Transformation is the method of introducing foreign DNA into bacterial cells (e.g. *E.coli*). The

Table 27.4 The different cloning vectors with the corresponding hosts and the sizes of foreign insert DNAs

Vector	*Host*	*Foreign insert DNA size*
Phage λ	*E. coli*	5–25 kb
Cosmid λ	*E. coli*	35–45 kb
Plasmid artifical chromosome (PAC)	*E. coli*	100–300 kb
Bacterial artificial chromosome (BAC)	*E. coli*	100–300 kb
Yeast chromosome	*S. cerevisiae*	200–2000 kb

uptake of plasmid DNA by *E.coli* is carried out in ice-cold $CaCl_2$ (0-5°C), and a subsequent heat shock (37–45°C for about 90 sec). By this technique, the ***transformation frequency***, which refers to ***the fraction of cell population that can be transferred***, is reasonably good e.g. approximately one cell per 1000 (10^{-3}) cells.

Conjugation

Conjugation is ***a natural microbial recombination process.*** During conjugation, two live bacteria (a donor and a recipient) come together, join by cytoplasmic bridges and transfer single-stranded DNA (from donor to recipient). Inside the recipient cell, the new DNA may integrate with the chromosome (rather rare) or may remain free (as is the case with plasmids).

The natural phenomenon of conjugation is exploited for gene transfer. This is achieved by transferring plasmid-insert DNA from one cell to another. In general, the plasmids lack conjugative functions and therefore, they are not as such capable of transferring DNA to the recipient cells. However, some plasmids with conjugative properties can be prepared and used.

Electroporation

Electroporation is based on the principle that high voltage electric pulses can induce cell plasma membranes to fuse. Thus, electroporation is a technique involving ***electric field-mediated membrane permeabilization.*** Electric shocks can also induce cellular uptake of exogenous DNA (believed to be via the pores formed by electric pulses) from the suspending solution. Electroporation is a simple and rapid technique for introducing genes into the cells from various organisms (microorganisms, plants and animals).

Liposome-mediated gene transfer

Liposomes are circular lipid molecules, which have an aqueous interior that can carry nucleic acids. Several techniques have been developed to encapsulate DNA in liposomes. The liposome-mediated gene transfer is referred to as ***lipofection.***

On treatment of DNA fragment with liposomes, the DNA pieces get encapsulated inside liposomes. These liposomes can adhere to cell membranes and fuse with them to transfer DNA fragments. Thus, the DNA enters the cell and then to the nucleus. The positively charged liposomes very efficiently complex with DNA, bind to cells and transfer DNA rapidly.

Direct transfer of DNA

It is possible to directly transfer the DNA into the cell nucleus. ***Microinjection*** and ***particle bombardment*** are the two techniques commonly used for this purpose.

GENE CLONING STRATEGIES

A clone refers to ***a group of*** organisms, cells, ***molecules*** or other objects, ***arising from a single individual.*** Clone and colony are almost synonymous.

Gene cloning strategies in relation to recombinant DNA technology broadly involve the following aspects (***Fig.27.4***).

- Generation of desired DNA fragments.
- Insertion of these fragments into a cloning vector.
- Introduction of the vectors into host cells.
- Selection or screening of the recipient cells for the recombinant DNA molecules.

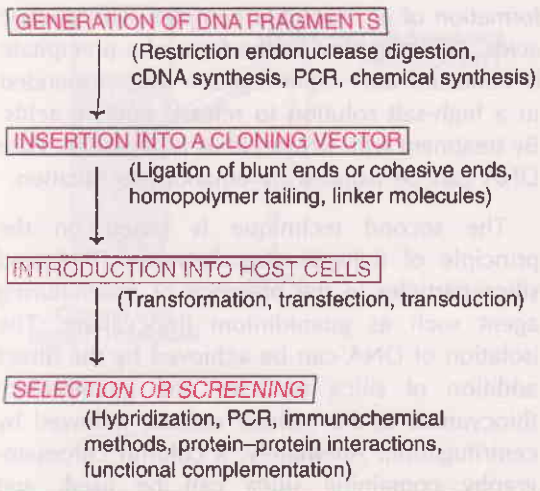


Fig. 27.4 : An overview of cloning strategies in recombinant DNA technology.

CLONING FROM GENOMIC DNA OR mRNA?

DNA represents the complete genetic material of an organism which is referred to as genome. *Theoretically speaking, cloning from genomic* DNA is supposed to be ideal. But the DNA contains non-coding sequences (introns), control regions and repetitive sequences. This complicates the *cloning strategies, hence DNA as a source* material is not preferred, by many workers. However, if the objective of cloning is to elucidate the control of gene expression, then genomic DNA has to be invariably used in cloning.

The use of mRNA in cloning is preferred for the following reasons.

- mRNA represents the actual genetic information being expressed.
- *Selection and isolation mRNA are easy.*
- As introns are removed during processing, mRNA reflects the coding sequence of the gene.
- The synthesis of recombinant protein is much easier with mRNA cloning.

Besides the direct use of genomic DNA or *mRNA, it is possible to synthesize DNA in the* laboratory (by polymerase chain reaction), and use it in cloning experiments. This approach is useful if the gene sequence is short and the complete sequence of amino acids is known.

BASIC TECHNIQUES IN GENETIC ENGINEERING

There are several techniques used in recombinant DNA technology or gene manipulation. The most frequently used methods are listed.

- *Isolation and purification of nucleic acids.*
- Nucleic acid blotting techniques.
- DNA sequencing.
- Methods of gene transfer (described already).
- Polymerase chain reaction.
- Production of monoclonal antibodies (***Chapter 41***).
- Construction of gene library.
- Site-directed mutagenesis and protein engineering.

ISOLATION AND PURIFICATION OF NUCLEIC ACIDS

Almost all the experiments dealing with gene *manipulations require pure forms of either DNA* or RNA, or sometimes even both. Hence there is a need for the reliable isolation of nucleic acids from the cells. The purification of nucleic acids broadly involves three stages.

1. Breaking or opening of the cells to expose nucleic acids.

2. Separation of nucleic acids from other cellular components.

3. Recovery of nucleic acids in a pure form.

The basic principles and procedures for nucleic acid purification are briefly described.

PURIFICATION OF CELLULAR DNA

The first step for DNA purification is to open the cells and release DNA. The method should *be gentle to preserve the native DNA. Due to* variability in cell structure, the approaches to break the cells are also different.

Lysis of cells

Bacterial cells : The bacterial cells (e.g. *E. coli*) can be lysed by a combination of enzymatic and chemical treatments. The enzyme lysozyme and the chemical ethylenediamine tetraacetate (EDTA) are used for this purpose. This is followed by the addition of detergents such as sodium dodecyl sulfate (SDS).

Animal cells : Animal cells, particularly cultured animal cells, can be easily opened by direct treatment of cells with detergents (SDS).

Plant cells : Plant cells with strong cell walls require harsh treatment to break open. The cells are frozen and then ground in a morter and pestle. This is an effective way of breaking the cellulose walls.

Methods to purify DNA

There are two different approaches to purify DNA from the cellular extracts.

1. **Purification of DNA by removing cellular components :** This involves the degradation or complete removal of all the cellular components other than DNA. This approach is suitable if the cells do not contain large quantities of lipids and carbohydrates.

The cellular extract is centrifuged at a low speed to remove the debris (e.g. pieces of cell wall) that forms a pellet at the bottom of the tube. The supernatant is collected and treated with phenol to precipitate proteins at the interface between the organic and aqueous layers. The aqueous layer, containing the dissolved nucleic acids, is collected and treated with the enzyme ribonuclease (RNase). The RNA is degraded while the DNA remains intact. This DNA can be precipitated by adding ethanol and isolated after centrifugation, and suspended in an appropriate buffer.

2. **Direct purification of DNA :** In this approach, the DNA itself is selectively removed from the cellular extract and isolated. There are two ways for direct purification of DNA.

In one method, the addition of a detergent ***cetyltrimethyl ammonium*** (CTAB) results in the formation of an insoluble complex with nucleic acids. This complex, in the form of a precipitate is collected after centrifugation and suspended in a high-salt solution to release nucleic acids. By treatment with RNase, RNA is degraded. Pure DNA can be isolated by ethanol precipitation.

The second technique is based on the principle of tight binding between DNA and silica particles in the presence of a denaturing agent such as guanidinium thiocyanate. The isolation of DNA can be achieved by the direct addition of silica particles and guanidinium thiocyanate to the cellular extract, followed by centrifugation. Alternately, a column chromatography containing silica can be used, and through this the extract and guanidinium thiocyanate are passed. The DNA binds to the silica particles in the column which can be recovered.

PURIFICATION OF mRNA

Among the RNAs, mRNA is frequently required in a pure form for genetic experiments.

After the cells are disrupted on lysis by different techniques (described above), the cellular extract is deproteinised by treatment with phenol or phenol/chloroform mixtures. On centrifugation, the nucleic acids get concentrated in the upper aqueous phase which may then be precipitated by using isopropanol or ethanol.

The purification of mRNA can be achieved by ***affinity chromatography*** using oligo (dT)-cellulose (***Fig.27.5***). This is based on the principle that oligo (dT)-cellulose can specifically bind to the poly (A) tails of eukaryotic mRNA. Thus, by this approach, it is possible to isolate mRNA from DNA, rRNA and tRNA.

As the nucleic acid solution is passed through an affinity chromotographic column, the oligo(dT) binds to poly(A) tails of mRNA. By washing the column with high-salt buffer, DNA, rRNA and tRNA can be eluted, while the mRNA is tightly bound. This mRNA can be then eluted by washing with low-salt buffer. The mRNA is precipitated with ethanol and collected by centrifugation (***Fig.27.5***).

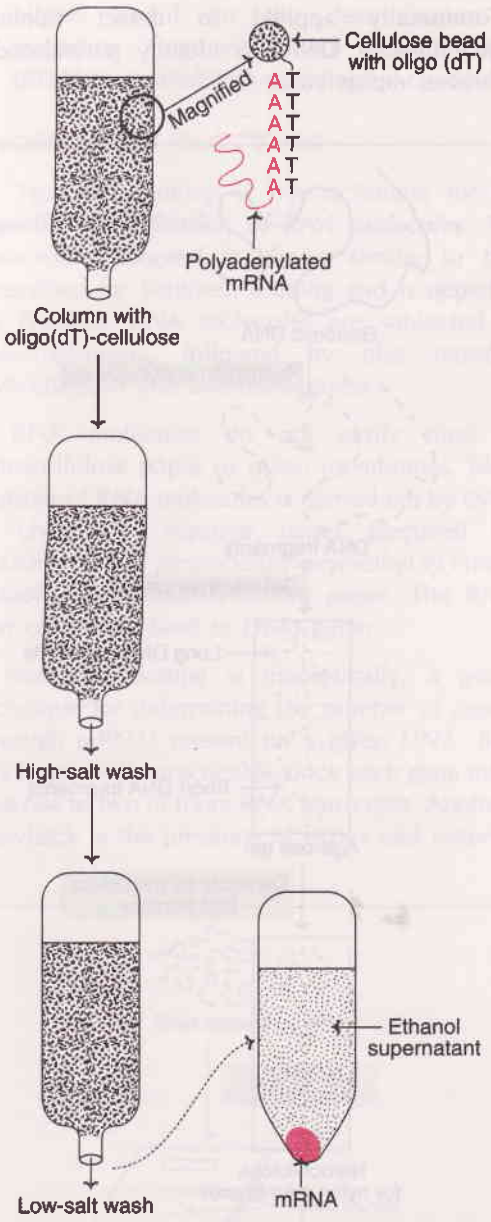


Fig. 27.5 : Purification of mRNA by affinity chromatography with oligo(dT)-cellulose.

NUCLEIC ACID BLOTTING TECHNIQUES

Blotting techniques are very widely used analytical tools for the specific ***identification of desired DNA or RNA fragments*** from thousands of molecules. Blotting refers to the ***process of immobilization of sample nucleic acids on solid support*** (nitrocellulose or nylon membranes). The blotted nucleic acids are then used as targets in the hybridization experiments for their specific detection. An outline of the nucleic acid blotting technique is depicted in ***Fig.27.6.***

Types of blotting techniques

The most comonly used blotting techniques are listed below

- Southern blotting (for DNA)
- Northern blotting (for RNA)
- Dot blotting (DNA/RNA)

The Southern blotting is named after the scientist Ed Southern (1975) who developed it. The other names Northern blotting and Western blotting are laboratory jargons which are now accepted. ***Western blotting involves the transfer of protein blots and their identification by using specific antibodies.***

A diagrammatic representation of a typical blotting apparatus is depicted in ***Fig.27.7.***

SOUTHERN BLOTTING

Southern blotting technique is the first nucleic acid blotting procedure developed in 1975 by

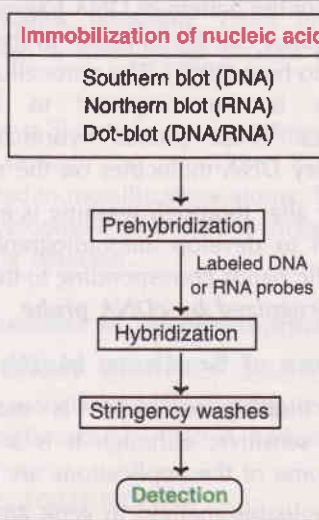


Fig. 27.6 : An outline of the nucleic acid blotting techniques.

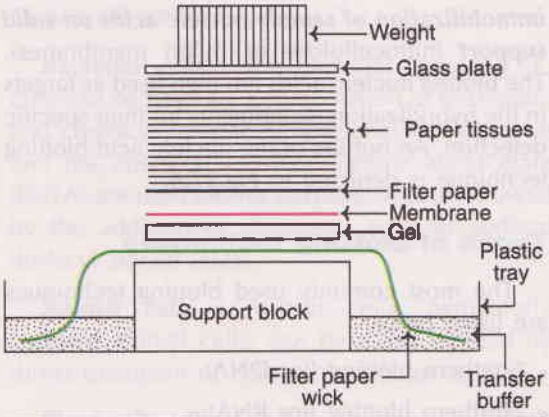


Fig. 27.7 : Diagrammatic representation of a typical blotting apparatus.

Southern. It is depicted in ***Fig.27.8***, and briefly described.

The genomic DNA isolated from cells/tissues is digested with one or more restriction enzymes. This mixture is loaded into a well in an agarose or polyacrylamide gel and then subjected to electrophoresis. DNA, being negatively charged migrates towards the anode (positively charged electrode); the smaller DNA fragments move faster.

The separated DNA molecules are denatured by exposure to a mild alkali and transferred to nitrocellulose or nylon paper. This results in an exact replica of the pattern of DNA fragments on the gel. The DNA can be annealed to the paper on exposure to heat (80°C). The nitrocellulose or nylon paper is then exposed to labeled cDNA probes. These probes hybridize with complementary DNA molecules on the paper.

The paper after thorough washing is exposed to X-ray film to develop autoradiograph. This reveals specific bands corresponding to the ***DNA fragments recognized by cDNA probe***.

Applications of Southern blotting

Southern blotting technique is extremely specific and sensitive, although it is a simple technique. Some of the applications are listed.

- It is an invaluable method in gene analysis.
- Important for the confirmation of DNA cloning results.
- Forensically applied to detect minute quantities of DNA (to identify parenthood, thieves, rapists etc.).

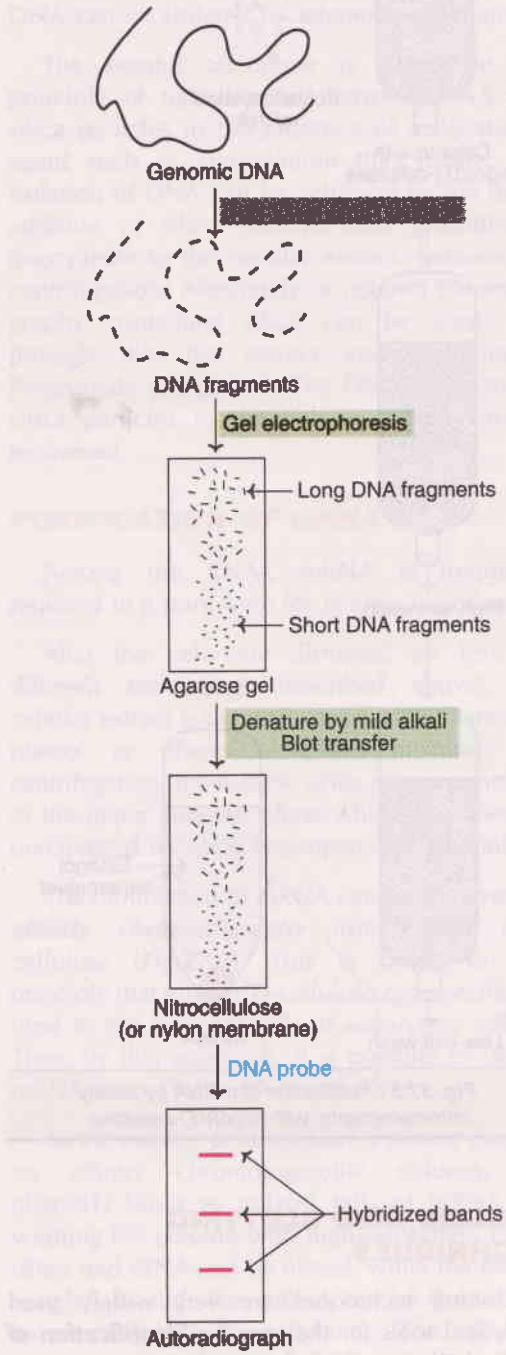


Fig. 27.8 : Southern blotting technique.

- Highly useful for the determination of restriction fragment length polymorphism (RFLP) associated with pathological conditions.

NORTHERN BLOTTING

Northern blotting is the technique for the ***specific identification of RNA molecules.*** The procedure adopted is almost similar to that described for Southern blotting and is depicted in ***Fig.27.9***. RNA molecules are subjected to electrophoresis, followed by blot transfer, hybridization and autoradiography.

RNA molecules do not easily bind to nitrocellulose paper or nylon membranes. Blot-transfer of RNA molecules is carried out by using a chemically reactive paper prepared by diazotization of aminobenzyloxymethyl to create diazobenzyloxymethyl (DBM) paper. The RNA can covalently bind to DBM paper.

Northern blotting is theoretically, a good technique for determining the number of genes (through mRNA) present on a given DNA. But this is not really practicable since each gene may give rise to two or more RNA transcripts. Another drawback is the presence of exons and introns.

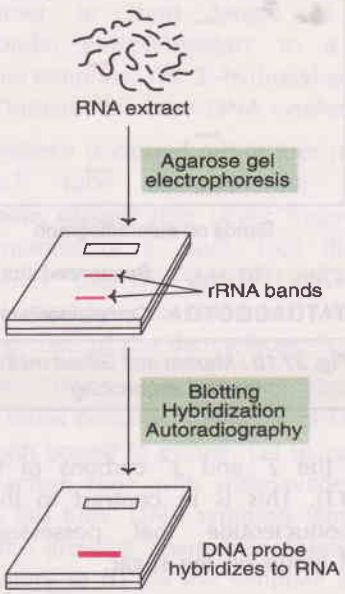


Fig. 27.9 : An outline of Northern blotting.

DOT-BLOTTING

Dot-blotting is a modification of Southern and Northern blotting techniques described above. In this approach, the ***nucleic acids (DNA or RNA) are directly spotted onto the filters***, and not subjected to electrophoresis. The hybridization procedure is the same as in original blotting techniques.

Dot-blotting technique is particularly useful in obtaining quantitative data for the evaluation of gene expression.

Western blotting

Western blotting involves the ***identification of proteins***. It is very useful to understand the nucleic acid functions, particularly during the course of gene manipulations.

The technique of Western blotting involves the transfer of electrophoresed protein bands from polyacrylamide gel to nylon or nitrocellulose membrane. These proteins can be detected by specific protein-ligand interactions. Antibodies or lectins are commonly used for this purpose.

Autoradiography

Autoradiography is the process of ***localization and recording of a radiolabel*** within a solid specimen, with the production of an ***image in a photographic emulsion***. These emulsions are composed of silver halide crystals suspended in gelatin.

When a β-particle or a γ-ray from a radiolabel passes through the emulsions, silver ions are converted to metallic silver atoms. This results in the development of a visible image which can be easily detected.

Applications of autoradiography

As already described, autoradiography is closely associated with blotting techniques for the detection of DNA, RNA and proteins.

DNA SEQUENCING

Determination of nucleotide sequence in a DNA molecule is the basic and fundamental

Table 27.6 A selected list of rDNA-derived therapeutic agents (approved by FDA) with trade names and their applications in humans

rDNA product	*Trade name(s)*	*Applications/uses*
Insulin	Humulin	Diabetes
Growth hormone	Protropin/Humatrope	Pituitary dwarfism
α-Interferon	Intron A	Hairy cell leukemia
Hepatitis B vaccine	Recombinax HB/Engerix B	Hepatitis B
Tissue plasminogen activator	Activase	Myocardial infarction
Factor VIII	Kogenate/Recombinate	Hemophilia
DNase	Pulmozyme	Cystic fibrosis
Erythropoietin	Epogen/Procrit	Severe anemia with kidney damage

medical and commercial importance. An ***approval***, by the concerned authorities, for using recombinant insulin ***for the treatment of diabetes mellitus*** was ***given in 1982***. And in 1986, Eli Lilly company received approval ***to market human insulin*** under the trade name ***Humulin***.

Technique for production of recombinant insulin : The orginal technique (described briefly above) of insulin synthesis in *E. coli* has undergone several changes, for improving the yield. e.g. addition of signal peptide, synthesis of A and B chains separately etc.

The procedure employed for the synthesis of two insulin chains A and B is illustrated in ***Fig.27.28***. The genes for insulin A chain and B chain are separately inserted to the plasmids of two different *E. coli* cultures. The *lac* operon system (consisting of inducer gene, promoter gene, operator gene and structural gene Z for β-galactosidase) is used for expression of both the genes. The presence of lactose in the culture medium induces the synthesis of insulin A and B chains in separate cultures. The so formed insulin chains can be isolated, purified and joined together to give a full-fledged human insulin.

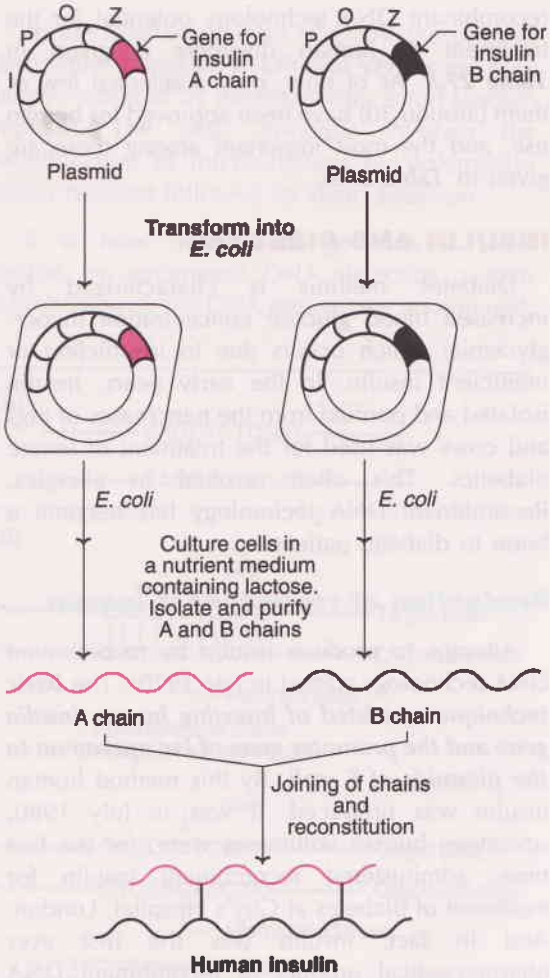


Fig. 27.28 : The production of recombinant insulin in E. coli *(I-Inducer gene, P-Promoter gene, O-Operator gene, Z-β-Galactosidase gene; all these genes are of lac operon system).*

RECOMBINANT VACCINES

Recombinant DNA technology in recent years, has become a boon to produce ***new generation vaccines***. By this approach, some of

the limitations (low yield, high cost, side effects) of traditional vaccine production could be overcome. In addition, several new strategies, involving gene manipulation are being tried to create novel recombinant vaccines.

The list of diseases for which recombinant vaccines are developed or being developed is given in ***Table 22.7***. It may be stated here that due to very stringent regulatory requirements to use in humans, the new generation vaccines are first tried in animals, and it may ***take some more years before most of them are approved for use in humans***.

Hepatitis B vaccine —the first synthetic vaccine

In 1987, the recombinant vaccine for hepatitis B (i.e. ***HBsAg***) became the first synthetic vaccine for public use. It was marketed by trade names ***Recombivax*** and ***Engerix-B***. Hepatitis B vaccine is safe to use, very effective and produces no allergic reactions. For these reasons, this recombinant vaccine has been in use since 1987. The individuals must be administered three doses over a period of six months. Immunization against hepatitis B is strongly recommended to anyone coming in contact with blood or body secretions. All the health professionals—physicians, surgeons, medical laboratory technicians, nurses, dentists, besides police officers, firefighters etc., must get vaccinated against hepatitis B.

Hepatitis B vaccine in India

India is the ***fourth country*** (after USA, France and Belgium) in the world to develop an indigenous hepatitis B vaccine. It was launched in 1997, and is now being used.

Table 27.7 A selected list of diseases along with the pathogenic organisms for which recombinant vaccines are developed or being developed

Disease	*Pathogenic organism*
Viral diseases	
Acute infantile gastroenteritis	Rotavirus
Acute respiratory diseases	Influenza A and B viruses
AIDS	Human immunodeficiency virus
Chicken pox	Varicella–zoster virus
Encephalitis	Japanese encephalitis virus
Genital ulcers	Herpes simplex virus type–2
Hemorrhagic fever	Dengue virus
Liver damage	Hepatitis A virus
Liver damage	Hepatitis B virus
Upper and lower respiratory tract lesions	Yellow fever virus
Bacterial diseases	
Cholera	*Vibrio cholerae*
Diarrhea	*E. coli*
Dysentery	*Shigella* strain
Gonorrhea	*Niesseria gonorroheae*
Leprosy	*Mycobacterium leprae*
Meningitis	*Neisseria meningitidis*
Pneumonia	*Streptococcus pneumoniae*
Rheumatic fever	*Streptococcus* group A
Tetanus	*Clostridium tetani*
Tuberculosis	*Mycobacterium tuberculosis*
Typhoid	*Salmonella typhi*
Urogenital tract infection	*Streptococcus* group B
Parasitic diseases	
Filariasis	*Wuchereria bancrofti*
Malaria	*Plasmodium* sp
River blindness	*Onchocerca volvulus*
Schistosomiasis	*Schistosoma mansoni*
Sleeping sickness	*Trypanosoma* sp

DNA VACCINES (GENETIC IMMUNIZATION)

Genetic immunization by using DNA vaccines is a novel approach that came into being in 1990. The ***immune response of the body is stimulated by a DNA molecule***. A DNA

Applications of DNA fingerprinting

The amount of DNA required for DNA fingerprint is remarkably small. The ***minute quantities of DNA*** from blood strains, body fluids, hair fiber or skin fragments ***are enough***. ***Polymerase chain reaction is used to amplify*** this DNA for use in fingerprinting. DNA profiling has wide range of applications—most of them related to medical forensics. Some important ones are listed below.

- Identification of criminals, rapists, thieves etc.
- Settlement of paternity disputes.
- Use in immigration test cases and disputes.

In general, the fingerprinting technique is carried out by collecting the DNA from a suspect (or a person in a paternity or immigration dispute) and matching it with that of a reference sample (from the victim of a crime, or a close relative in a civil case).

DNA MARKERS IN DISEASE DIAGNOSIS AND FINGERPRINTING

The ***DNA markers are highly useful for genetic mapping of genomes***. There are four types of DNA sequences which can be used as markers.

1. Restriction fragment length polymorphisms (***RFLPs***, pronounced as rif-lips).

2. Minisatellites or variable number tandem repeats (***VNTRs***, pronounced as vinters).

3. Microsatellites or simple tandem repeats (***STRs***).

4. Single nucleotide polymorphisms (***SNPs***, pronounced as snips).

The general aspects of the above DNA markers are described along with their utility in disease diagnosis and DNA fingerprinting.

RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLPs)

A RFLP represents a stretch of DNA that serves as a marker for mapping a specified gene. RFLPs are located randomly throughout a person's chromosomes and have no apparent function.

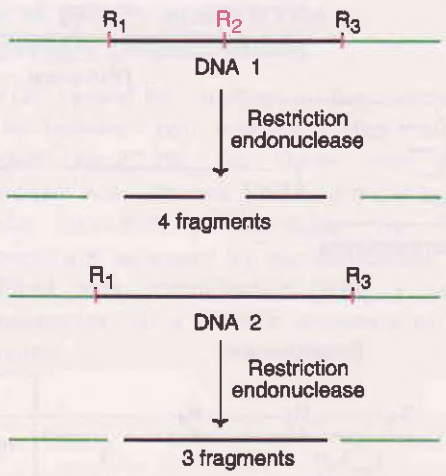


Fig. 27.22 : *An outline of the restriction fragment length polymorphism (RFLP) (R_1, R_2, R_3 represent the sites for the action of restriction endonucleases).*

A DNA molecule can be cut into different fragments by a group of enzymes called restriction endonucleases (***See Table 27.1***). These fragments are called ***polymorphisms*** (literally means ***many forms***).

An outline of RFLP is depicted in ***Fig.27.22***. The DNA molecule 1 has three restriction sites (R_1, R_2, R_3), and when cleaved by restriction endonucleases forms 4 fragments. Let us now consider DNA 2 with an inherited mutation (or a genetic change) that has altered some base pairs. As a result, the site (R_2) for the recognition by restriction endonuclease is lost. This DNA molecule 2 when cut by restriction endonuclease forms only 3 fragments (instead of 4 in DNA 1).

As is evident from the above description, a stretch of DNA exists in ***fragments*** of various ***lengths*** (***polymorphisms***), derived by the action of ***restriction*** enzymes, hence the name restriction fragment length polymorphisms.

RFLPs in the diagnosis of diseases

If the RFLP lies within or even close to the locus of a gene that causes a particular disease, it is possible to trace the defective gene by the

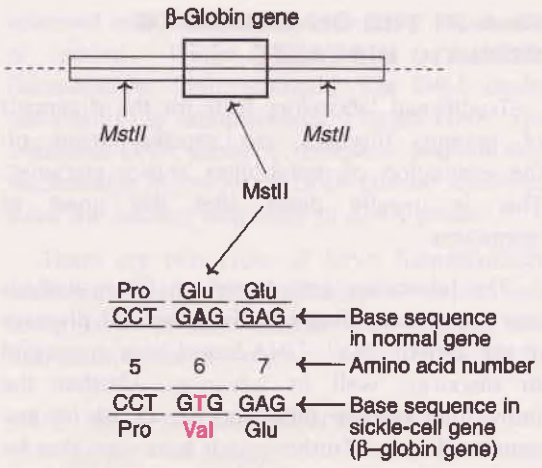


*Fig. 27.21 : Single-base change resulting in sickle-cell anemia. (**Note :** A small and relevant DNA fragment of β-globin gene magnified and shown with encoded amino acids).*

DNA fragments in and around β-globin gene, followed by the electrophoretic pattern of the DNA fragments formed. The change in the base from A to T in the β-globin gene destroys the recognition site (CCTGAGG) for *MstII* (***Fig.27.21***). Consequently, the DNA fragments formed from a sickle-cell anemia patient for β-globin gene differ from that of a normal person. Thus, sickle-cell anemia can be detected by digesting mutant and normal β-globin genes by restriction enzyme and performing a ***hybridization with a cloned β-globin DNA probe.***

GENE BANKS — A NOVEL CONCEPT

As the search continues by scientists for the identification of more and more genes responsible for various diseases, the enlightened public (particularly in the developed countries), is very keen to enjoy the fruits of this research outcome. As of now, DNA probes are available for the detection a limited number of diseases. Researchers continue to develop DNA probes for a large number of genetically predisposed disorders.

Gene banks are the centres for the storage of individual's DNAs for future use to diagnose diseases. For this purpose, the DNA isolated from a person's cells (usually white blood cells) is stored. As and when a DNA probe for the detection of a specific disease is available, the stored DNA can be used for the diagnosis or risk assessment of the said genetic disease.

In fact, some institutions have established gene banks. They store the DNA samples of the interested customers at a fee (one firm was charging $200) for a specified period (say around 20–25 years). For the risk assessment of any disease, it is advisable to have the DNAs from close relatives of at least 2-3 generations.

DNA FINGERPRINTING OR DNA PROFILING

DNA fingerprinting is the present day ***genetic detective*** in the practice ***of modern medical forensics.*** The underlying principles of DNA fingerprinting are briefly described.

The structure of each person's genome is unique. The only exception being monozygotic identical twins (twins developed from a single fertilized ovum). The unique nature of genome structure provides a good opportunity for the specific identification of an individual.

It may be remembered here that in the traditional fingerprint technique, the individual is identified by preparing an ink impression of the skin folds at the tip of the person's finger. This is based on the fact that the nature of these skin folds is genetically determined, and thus the fingerprint is unique for an individual. In contrast, the ***DNA fingerprint is an analysis of the nitrogenous base sequence in the DNA of an individual.***

History and terminology

The original DNA fingerprinting technique was developed by Alec Jaffreys in 1985. Although the DNA fingerprinting is commonly used, a more general term ***DNA profiling*** is preferred. This is due to the fact that a wide range of tests can be carried out by DNA sequencing with improved technology.

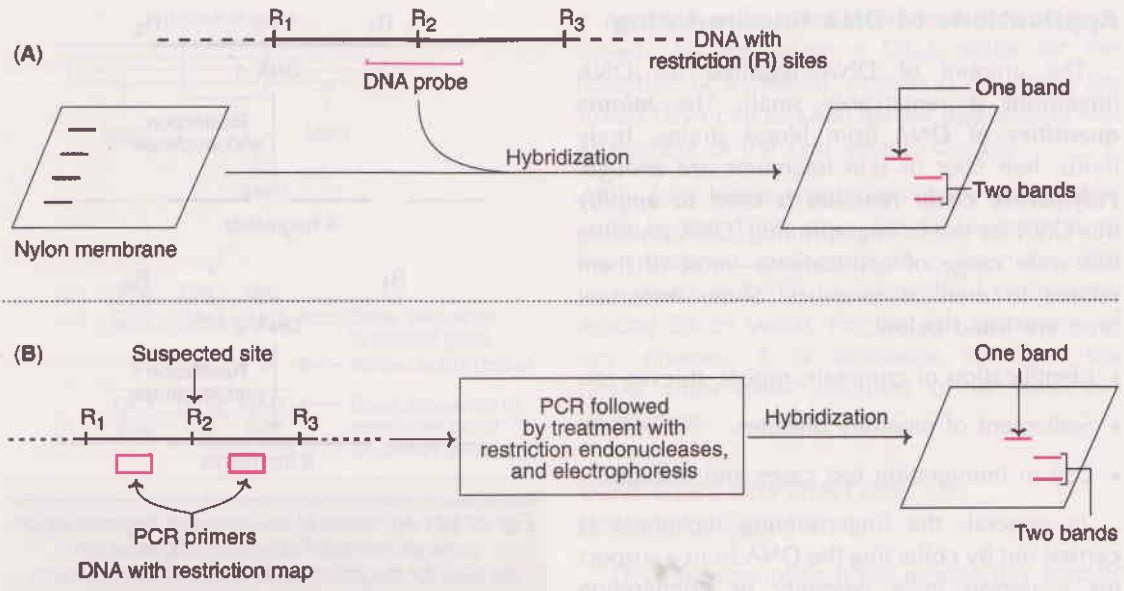


Fig. 27.23 : Two common methods used for scoring restriction fragment length polymorphism (RFLP) (A) RFLP by Southern hybridization (B) RFLP by polymerase chain reaction (PCR).

analysis of RFLP in DNA. The person's cellular DNA is isolated and treated with restriction enzymes. The DNA fragments so obtained are separated by electrophoresis. The RFLP patterns of the disease suspected individuals can be compared with that of normal people (preferably with the relatives in the same family). By this approach, it is possible to determine whether the individual has the marker RFLP and the disease gene. With 95% certainity, RFLPs can ***detect single gene-based diseases.***

Methods of RFLP scoring : Two methods are in common use for the detection of RFLPs (***Fig.27.23***).

1. **Southern hybridization :** The DNA is digested with appropriate restriction enzyme, and separated by agarose gel electrophoresis. The so obtained DNA fragments are transferred to a nylon membrane. A DNA probe that spans the suspected restriction site is now added, and the hybridized bands are detected by autoradiograph. If the restriction site is absent, then only a single restriction fragment is detected. If the site is present, then two fragments are detected (***Fig.27.23A***).

2. **Polymerase chain reaction :** RFLPs can also be scored by PCR. For this purpose, PCR primers that can anneal on either side of the suspected restriction site are used. After amplification by PCR, the DNA molecules are treated with restriction enzyme and then analysed by agarose gel electrophoresis. If the restriction site is absent only one band is seen, while two bands are found if the site is found (***Fig.27.23B***).

Applications of RFLPs : The approach by RFLP is very powerful and has helped many genes to be mapped on the chromosomes. e.g. sickle-cell anemia (chromosome 11), cystic fibrosis (chromosome 7), Huntington's desease (chromosome 4), retinoblastoma (chromosome 13), Alzheimer's disease (chromosome 21).

VARIABLE NUMBER TANDEM REPEATS (VNTRs)

VNTRs, also known as minisatellites, like RFLPs, are DNA fragments of different length. The main difference is that RFLPs develop from random mutations at the site of restriction enzyme activity while VNTRs are formed due to

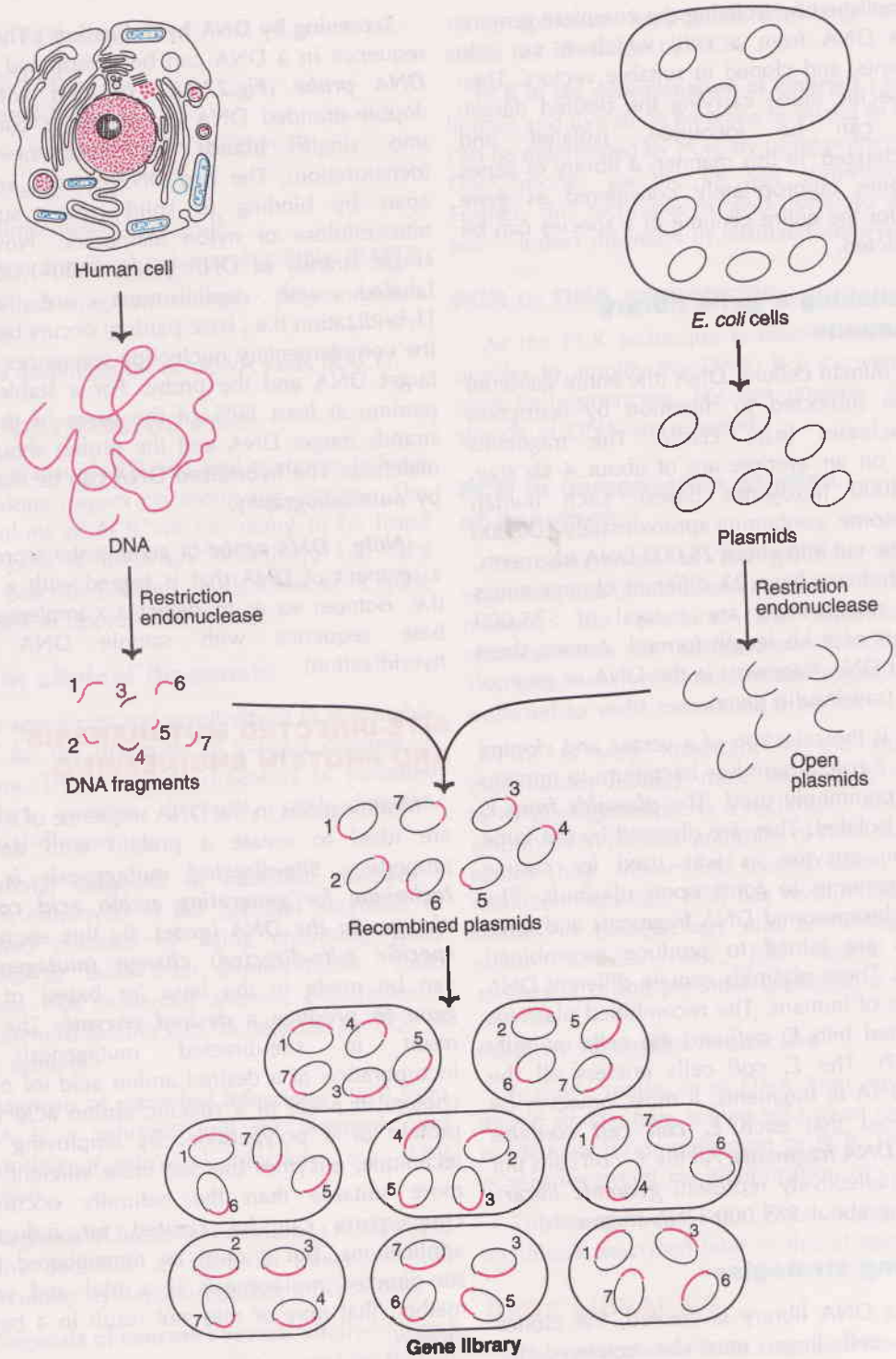


*Fig. 27.17 : Creation of a genomic library for humans. (**Note :** Double-stranded DNA is represented by single lines or circles for clarity; human DNA fragments are coloured).*

accomplished by isolating the complete genome (entire DNA from a cell) which is cut into fragments, and cloned in suitable vectors. Then the specific clone carrying the desired (target) DNA can be identified, isolated and characterized. In this manner, a library of genes or clones (appropriately considered as ***gene bank***) for the entire genome of a species can be constructed.

Establishing a gene library for humans

The human cellular DNA (the entire genome) may be subjected to digestion by restriction endonucleases (e.g., *EcoRI*). The fragments formed on an average are of about 4 kb size. (i.e., 4000 nitrogenes bases). Each human chromosome, containing approximately 100,000 kb can be cut into about 25,000 DNA fragments. As the humans have 23 different chromosomes (24 in man), there are a total of 575,000 fragments of 4 kb length formed. Among these 575,000 DNA fragments is the DNA or gene of interest (say insulin gene).

Now is the selection of a vector and cloning process. *E.coli*, a harmless bacterium to humans is most commonly used. The ***plasmids from E. coli*** are isolated. They are digested by the same restriction enzyme as was used for cutting human genome to form open plasmids. The human chromosomal DNA fragments and open plasmids are joined to produce recombined plasmids. These plasmids contain different DNA fragments of humans. The recombined plasmids are inserted into *E. coli* and the cells multiply (***Fig.27.17***). The *E. coli* cells possess all the human DNA in fragments. It must, however be remembered that each ***E. coli cell contains different DNA fragments***. All the *E. coli* cells put together collectively represent ***genomic library*** (containing about 575,000 DNA fragments).

Screening strategies

Once a DNA library is created, the clones (i.e., the cell lines) must be screened for identification of specific clones. The screening techniques are mostly based on the sequence of the clone or the structure/function of its product.

Screening by DNA hybridization : The target sequence in a DNA can be determined with a ***DNA probe*** (***Fig.27.18***). To start with, the double-stranded DNA of interest is converted into single strands by heat or alkali (denaturation). The two DNA strands are kept apart by binding to solid matrix such as nitrocellulose or nylon membrane. Now, the single strands of DNA probe (100–1,000 bp) labeled with radioisotope are added. Hybridization (i.e., base pairing) occurs between the complementary nucleotide sequences of the target DNA and the probe. For a stable base pairing, at least 80% of the bases in the two strands (target DNA and the probe) should be matching. The hybridized DNA can be detected by ***autoradiography***.

(***Note : DNA probe*** or gene probe represents a segment of DNA that is tagged with a label (i.e. isotope) so as to detect a complementary base sequence with sample DNA after hybridization)

SITE-DIRECTED MUTAGENESIS AND PROTEIN ENGINEERING

Modifications in the DNA sequence of a gene are ideal to create a protein with desired properties. ***Site-directed mutagenesis is the technique for generating amino acid coding changes in the DNA (gene)***. By this approach ***specific (site-directed) change (mutagenesis)*** can be made in the base (or bases) of the gene ***to produce a desired enzyme***. The net result in site-directed mutagenesis is incorporation of a desired amino acid (of one's choice) in place of a specific amino acid in a protein or a polypeptide. By employing this technique, enzymes that are more efficient and more suitable than the naturally occurring counterparts can be created for industrial applications. But it must be remembered that site-directed mutagenesis is a trial and error method that may or may not result in a better protein.

A couple of proteins developed by site-directed mutagenesis and protein engineering are given next.

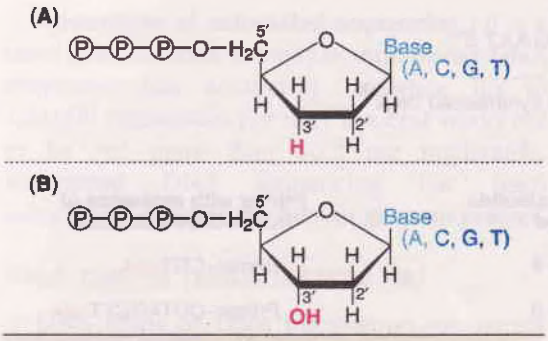


Fig. 27.11 : Structure of (A) dideoxynucleotide triphosphate and (B) deoxynucleotide triphosphate (Note the difference at 3'-carbon).

incoming nucleoside triphosphate is attached by its 5'-phosphate group to the 3'-hydroxyl group of the last nucleotide of the growing chain (***Refer Chapter 24***) when a dideoxynucleotide is incorporated to the growing chain, no further replication occurs. This is because dideoxynucleotide, lacking a 3'-hydroxyl group, cannot form a phosphodiester bond and thus the DNA synthesis terminates.

Sequencing method : The process of sequencing DNA by dideoxynucleotide method is briefly described. A single-stranded DNA to be sequenced is chosen as a template. It is attached to a primer (a short length of DNA oligonucleotide) complementary to a small section of the template. The 3'-hydroxyl group of the primer initiates the new DNA synthesis.

DNA synthesis is carried out in four reaction tubes. Each tube contains the primed DNA, ***Klenow subunit*** (the larger fragment of DNA polymerase of *E. coli*), four dideoxy-ribonucleotides (ddATP, ddCTP, ddGTP or ddTTP). It is necessary to radiolabel (with ^{32}P) the primer or one of the deoxyribonucleotides.

As the new DNA synthesis is completed, each one of the tubes contains fragments of DNA of varying length bound to primer. Let us consider the first reaction tube with dideoxyadenosine (ddATP). In this tube, DNA synthesis terminates whenever the growing chain incorporates ddA (complementary to dT on the template strand). Therefore, this tube will contain a series of different length DNA fragments, each ending with ddA. In a similar fashion, for the other 3 reaction tubes, DNA synthesis stops as the respective dideoxynucleotides are incorporated.

The synthesis of new DNA fragments in the four tubes is depicted in ***Fig.27.12.***

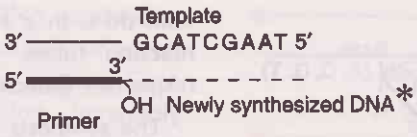
The DNA pieces are denatured to yield free strands with radiolabel. The samples from each tube are separated by polyacrylamide gel electrophoresis. This separation technique resolves DNA pieces, different in size even by a single nucleotide. The shortest DNA will be the fastest moving on the electrophoresis.

The sequence of bases in a DNA fragment is determined by identifying the electrophoretic (radiolabeled) bands by autoradiography. In the ***Fig.27.13***, the sequence of the newly synthesized DNA fragment that is complementary to the original DNA piece is shown. It is conventional to read the bands from bottom to top in 5' to 3' direction. By noting the order of the bands first C, second G, third T and so on, the sequence of the DNA can be determined accurately. As many as 350 base sequences of a DNA fragment can be clearly identified by using autoradiographs.

Modifications of dideoxynucleotide method : Replacement of ^{32}P-radiolabel by ^{33}P or ^{35}S improves the sharpness of autoradiographic images. DNA polymerase of the thermophilic bacterium, *Thermus aquaticus* (in place of Klenow fragment of *E. coli* DNA polymerase I) or a modified form of phage T7 DNA polymerase (sequenase) improves the technique.

AUTOMATED DNA SEQUENCING

DNA sequencing in the recent years is carried out by an ***automated DNA sequencer***. In this technique, flourescent tags are attached to chain-terminating nucleotides (dideoxynucleotides). This tag gets incorporated into the DNA molecules, while terminating new strand synthesis. Four different fluorescent dyes are used to identify chain-terminating reactions in a sequencing gel. The DNA bands are separated by electrophoresis and detected by their fluorescence. Recently, four dyes that exhibit strong absorption in laser are in use for automated sequencing.



Reaction tube with dideoxynucleotide	Primer with nucleotide extended	Primer with sequence of nucleotides extended
ddATP	Primer + 4	Primer–CGTddA
	Primer + 9	Primer–CGTAGCTTddA
ddCTP	Primer + 1	Primer–ddC
	Primer + 6	Primer–CGTAGddC
ddGTP	Primer + 2	Primer–C**ddG**
	Primer + 5	Primer–CGTA**ddG**
ddTTP	Primer + 3	Primer–CGddT
	Primer + 7	Primer–CGTAGCddT
	Primer + 8	Primer–CGTAGCTddT

*Fig. 27.12 : Synthesis of new DNA fragments in the presence of dideoxynucleotides (*the size of thenew DNA is variable, depending on the chain termination).*

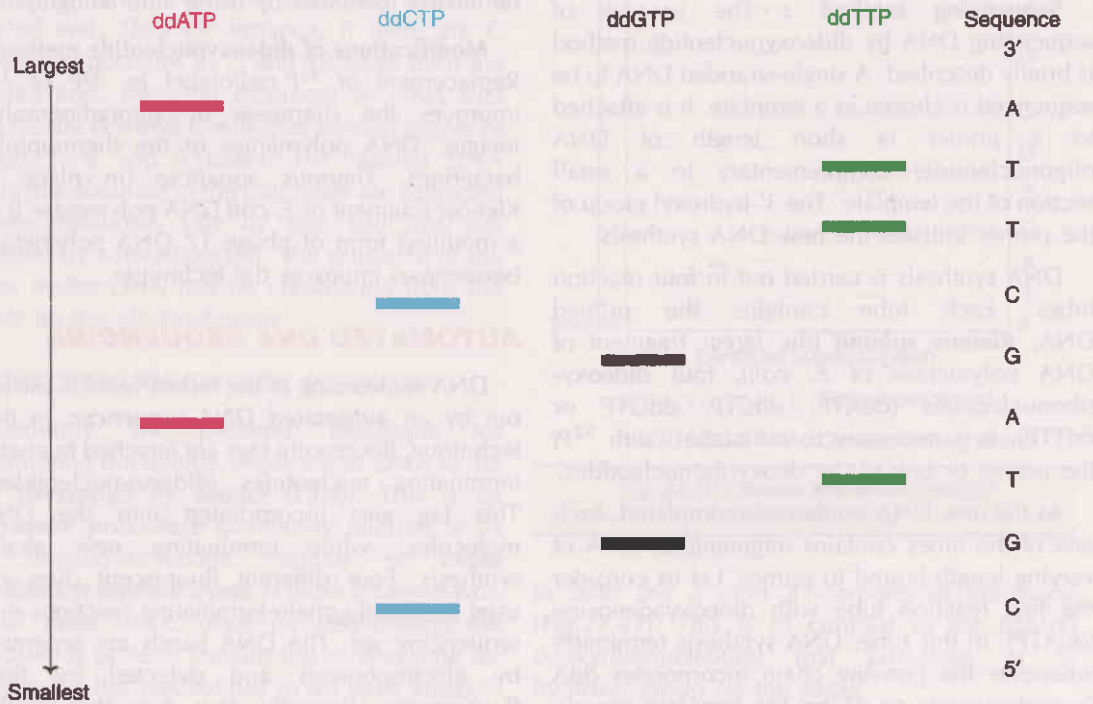


Fig. 27.13 : Sequence of the newly synthesized DNA fragment (complementary to original strand).

Advantages of automated sequencing : It is a rapid and accurate technique. Automated DNA sequencer can accurately sequence up to 100,000 nucleotides per day. The cost works out to be not more than $0.2 per nucleotide. Automated DNA sequencing has been successfully ***used in the human genome project.***

DNA CHIPS (MICROARRAYS)

DNA chips or DNA microarrays are recent developments for DNA sequencing as result of advances made in automation and miniarization. A large number of DNA probes, each one with different sequence, are immobilized at defined positions on the solid surface, made up of either nylon or glass. The ***probes*** can be short DNA molecules such as ***cDNAs or synthetic oligonucleotides.***

For the preparation of high density arrays, oligonucleotides are synthesized *in situ* on the surface of glass or silicon. This results in an ***oligonucleotide chip*** rather than a DNA chip.

Technique of DNA sequencing

A DNA chip carrying an array of different oligonucleotides can be used for DNA sequencing. For this purpose, a fluorescently labeled DNA test molecule, whose sequence is to be determined, is applied to the chip. Hybridization occurs between the complementary sequences of the test DNA molecule and oligonucleotides of the chip. The positions of these hybridizing oligonucleotides can be determined by confocal microscopy. Each hybridizing oligonucleotide represents an 8-nucleotide sequence that is present in the DNA probe. The sequence of the test DNA molecule can be deduced from the overlaps between the sequences of the hybridizing oligonucleotides (***Fig.27.14***).

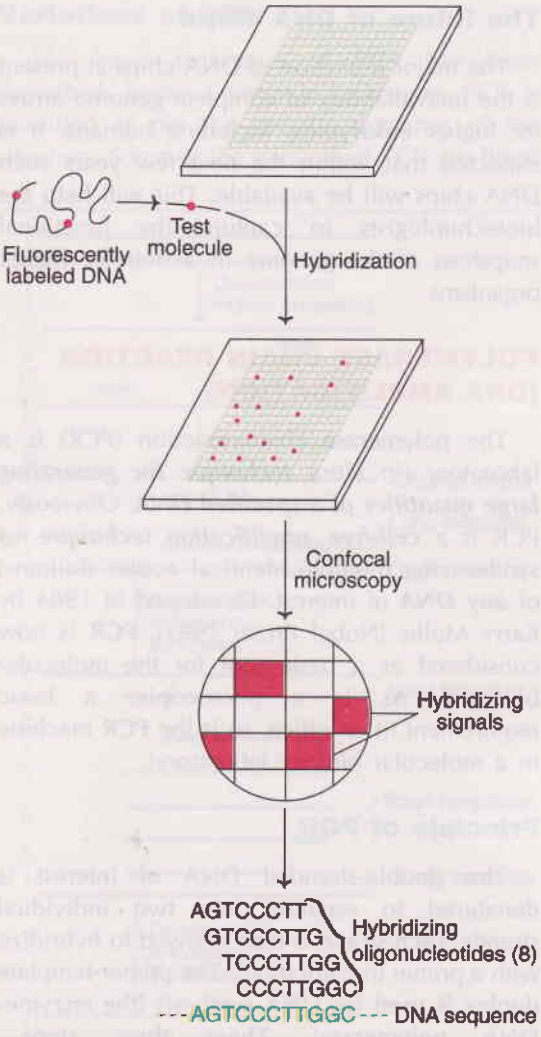


Fig. 27.14 : Microarray (or chip) technology in DNA sequencing.

Applications of DNA chips

There have been many successes with this relatively new technology of DNA chips. Some of them are listed.

- Identification of genes responsible for the development of nervous systems.
- Detection of genes responsible for inflammatory diseases.
- Construction of microarrays for every gene in the genome of *E. coli,* and almost all the genes of the yeast *Saccharomyces cerevisiae.*
- Expression of several genes in prokaryotes has been identified.
- Detection and screening of single nucleotide polymorphisms (SNPs).
- Rapid detection of microorganisms for environmental monitoring.

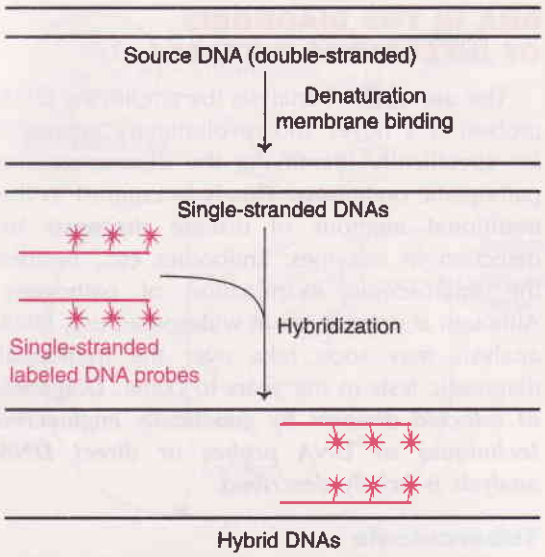


Fig. 27.18 : Screening by DNA hybridization (✳ indicates radioisotope label in the DNA probe)

Tissue plasminogen activator (tPA)

Tissue plasminogen activator is therapeutically used to lyse the blood clots that cause myocardial infarction. Due to its shorter half-life (around 5 minutes), tPA has to be repeatedly administered. By ***replacing asparagine residue*** (at position 120) ***with glutamine, the half-life of tPA can be substantially increased***. This is due to the fact that glutamine is less glycosylated than asparagine and this makes a difference in the half-life of tPA.

Hirudin

Hirudin is a protein secreted by leech salivary gland, and is a strong thrombin inhibitor (i.e., acts as an anticoagulant). By ***replacing asparagine*** (at 47 position) ***with lysine, the potency of hirudin can be increased*** several-fold.

DNA IN DISEASE DIAGNOSIS AND MEDICAL FORENSICS

DNA, being the genetic material of the living organisms, contains the information that contributes to various characteristic features of the specific organism. Thus, ***the presence of a disease-causing pathogen can be detected by identifing a gene or a set of genes of the organism.*** Likewise an ***inherited genetic defect can be diagnosed by identifying the alterations in the gene.*** In the modern laboratory diagnostics, DNA analysis is a very useful and a sensitive tool.

The basic principles underlying the DNA diagnostic systems, and their use in the diagnosis of certain pathogenic and genetic diseases are described. Besides these, the various approaches for DNA fingerprinting (or DNA profiling) are also discussed.

METHODS OF DNA ASSAY

The specific ***identification of the DNA sequence*** is absolutely ***essential in the laboratory diagnostics***. This can be achieved by employing the following principles/tools.

Nucleic acid hybridization

Hybridization of nucleic acids (particularly DNA) is the basis for reliable DNA analysis. Hybridization is based on the principle that a single-stranded DNA molecule recognizes and specifically binds to a complementary DNA strand amid a mixture of other DNA strands. This is comparable to a specific key and lock relationship. The general procedure adopted for nucleic acid hybridization has been described (See p. 597 and ***Fig.27.18***). Some more information is given below (***Fig.27.19***).

The single-stranded target DNA is bound to a membrane support. Now the DNA probe (single-

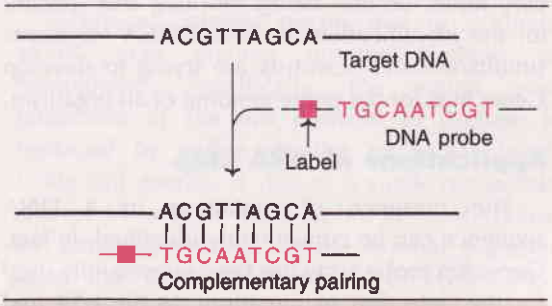


Fig. 27.19 : Hybridization of target DNA with DNA probe (with radioactive isotope label).

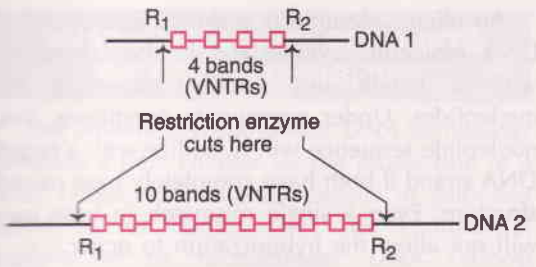


Fig. 27.24 : A diagrammatic representation of variable number tandem repeats (VNTRs). Each band (or copy) represents a repeating sequence in the DNA (e.g. 100 base pairs each). R_1 and R_2 indicate the sites cut by a restriction enzyme.

different number of base sequences between two points of a DNA molecule. In general, VNTRs are made up of tandem repeats of short base sequences (10–100 base pairs). The number of elements in a given region may vary, hence they are known as variable number tandem repeats.

An individual's genome has many different VNTRs and RFLPs which are unique to the individual. The ***pattern of VNTRs and RFLPs forms the basis of DNA fingerprinting*** or DNA profiling.

In the ***Fig.27.24***, two different DNA molecules with different number of copies (bands) of VNTRs are shown. When these molecules are subjected to restriction endonuclease action (at two sites R_1 and R_2), the VNTR sequences are released, and they can be detected due to variability in repeat sequence copies. These can be used in mapping of genomes, besides their utility in DNA fingerprinting.

VNTRs are useful for the detection of certain genetic diseases associated with alterations in the degree of repetition of microsatellites e.g. Huntington's chorea is a disorder which is found when the VNTRs exceed 40 repeat units.

Limitations of VNTRs : The major drawback of VNTRs is that they are not evenly distributed throughout the genome. VNTRs tend to be localized in the telomeric regions, at the ends of the chromosomes.

Use of RFLPs and VNTRs in genetic fingerprinting

RFLPs caused by variations in the number of VNTRs between two restriction sites can be detected (***Fig.27.25***). The DNAs from three individuals with different VNTRs are cut by the specific restriction endonuclease. The DNA fragments are separated by electrophoresis, and identified after hybridization with a probe complementary to a specific sequence on the fragments.

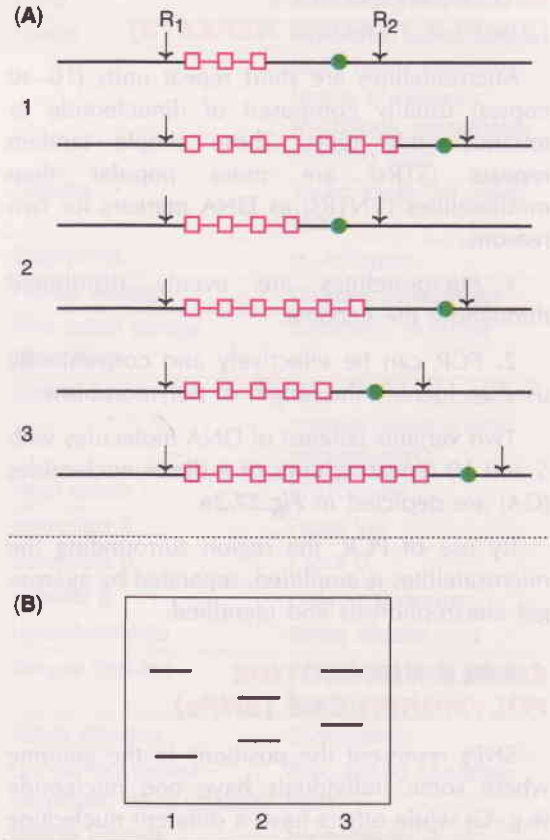


Fig. 27.25 : *Use of restriction fragment length polymorphisms (RFLPs) caused by variable number tandem repeats (VNTRs) in genetic fingerprinting (**A**) An illustration of DNA structure from three individuals (**B**) Hybridized pattern of DNA fragment with a probe complementary to the sequence shown in green circles (1, 2 and 3 represent the individuals; R_1 and R_2 indicate restriction sites; coloured squares are the number of VNTRs)*

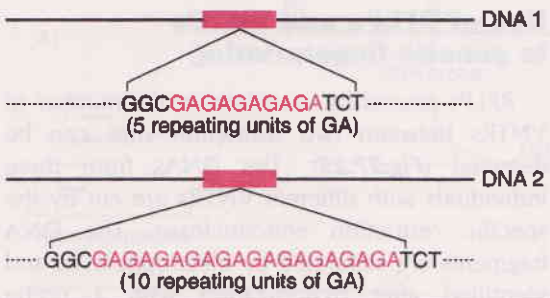


Fig. 27.26 : Two alleles of DNA molecules representing 5 and 10 dimer repeating units.

MICROSATELLITES (SIMPLE TANDEM REPEATS)

Microsatellites are short repeat units (10–30 copies) usually composed of dinucleotide or tetranucleotide units. These simple tandem repeats (STRs) are more popular than minisatellites (VNTRs) as DNA markers for two reasons.

1. Microsatellites are evenly distributed throughout the genome.

2. PCR can be effectively and conveniently used to identify the length of polymorphism.

Two variants (alleles) of DNA molecules with 5 and 10 repeating units of a dimer nucleotides (GA) are depicted in ***Fig.27.26***.

By use of PCR, the region surrounding the microsatellites is amplified, separated by agarose gel electrophoresis and identified.

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

SNPs represent the positions in the genome where some individuals have one nucleotide (e.g. G) while others have a different nucleotide (e.g. C). There are large numbers of SNPs in genomes. It is estimated that the human genome contains at least 3 million SNPs. Some of these SNPs may give rise to RFLPs.

SNPs are highly useful as DNA markers since there is no need for gel electrophoresis and this saves a lot of time and labour. The detection of SNPs is based on the oligonucleotide hybridization analysis (***Fig.27.27***).

An oligonucleotide is a short single-stranded DNA molecule, synthesized in the laboratory with a length not usually exceeding 50 nucleotides. Under appropriate conditions, this nucleotide sequence will hybridize with a target DNA strand if both have completely base paired structure. Even a single mismatch in base pair will not allow the hybridization to occur.

DNA chip technology is most commonly used to screen SNPs hybridization with oligonucleotide (*See* p. 593).

CURRENT TECHNOLOGY OF DNA FINGERPRINTING

In the forensic analysis of DNA, the original techniques based on RFLPs and VNTRs are now largely replaced by microsatellites (short tandem repeats). The basic principle involves the amplification of microsatellites by polymerase chain reaction followed by their detection.

It is now possible to generate a DNA profile by automated DNA detection system (comparable to the DNA sequencing equipment).

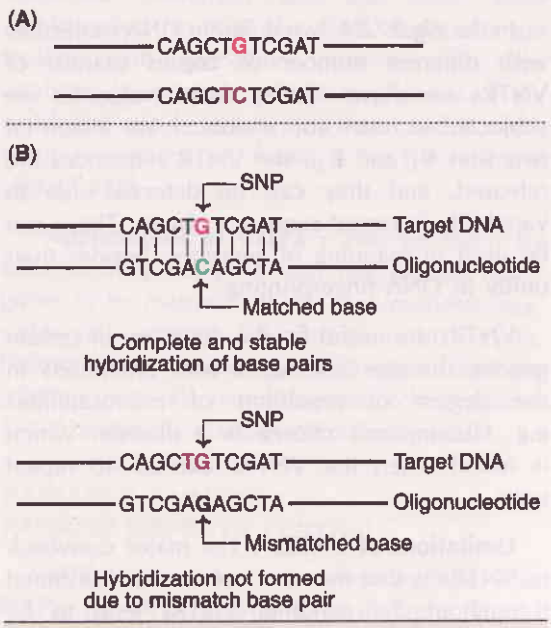


Fig. 27.27 : (A) An illustration of single nucleotide polymorphism (SNP) (B) Oligonucleotide hybridization to detect SNP.

stranded and labeled with a detector substance) is added. Under appropriate conditions (temperature, ionic strength), the DNA probe pairs with the complementary target DNA. The unbound DNA probe is removed. Sequence of nucleotides in the target DNA can be identified from the known sequence of DNA probe.

There are two types of DNA hybridization-radioactive and non-radioactive respectively using DNA probes labeled with isotopes and non-isotopes as detectors.

THE DNA CHIP-MICROARRAY OF GENE PROBES

The DNA chip or ***Genechip contains thousands of DNA probes*** (4000,000 or even more) arranged on a small glass slide of the size of a postage stamp. By this recent and advanced approach, thousands of target DNA molecules can be scanned simultaneously.

Technique for use of DNA chip

The unknown DNA molecules are cut into fragments by restriction endonucleases. Fluorescent markers are attached to these DNA fragments. They are allowed to react with the probes of the DNA chip. Target DNA fragments with complementary sequences bind to DNA probes. The remaining DNA fragments are washed away. The target DNA pieces can be identified by their fluorescence emission by passing a laser beam. A computer is used to record the pattern of fluoresence emission and DNA identification.

The technique of employing DNA chips is very rapid, besides being sensitive and specific for the identification of several DNA fragments simultaneously. Scientists are trying to develop Genechips for the entire genome of an organism.

Applications of DNA chip

The presence of mutations in a DNA sequence can be conveniently identified. In fact, Genechip probe array has been successfully used for the detection of mutations in the p53 and BRCA I genes. Both these genes are involved in cancer (***See*** p. 593 also).

DNA IN THE DIAGNOSIS OF INFECTIOUS DISEASES

The use of DNA analysis (by employing DNA probes) is a novel and revolutionary approach for specifically identifying the disease-causing pathogenic organisms. This is in contrast to the traditional methods of disease diagnosis by detection of enzymes, antibodies etc., besides the microscopic examination of pathogens. Although at present not in widespread use, DNA analysis may soon take over the traditional diagnostic tests in the years to come. ***Diagnosis of selected diseases by genetically engineered techniques or DNA probes or direct DNA analysis is briefly described.***

Tuberculosis

Tuberculosis is caused by the bacterium *Mycobacterium tuberculosis.* The commonly used diagnostic tests for this disease are very slow, and sometimes may take several weeks. This is because *M. tuberculosis* multiplies very slowly (takes about 24 hrs. to double; *E. coli* takes just 20 minutes to double).

A novel diagnostic test for tuberculosis was developed by genetic engineering, and is illustrated in ***Fig.27.20.*** A gene from ***firefly,*** encoding the enzyme ***luciferase*** is introduced into the bacteriophage specific for *M. tuberculosis.* The bacteriophage is a bacterial virus, frequently referred to as ***luciferase reporter phage or mycophage.*** The genetically engineered phage is added to the culture of *M. tuberculosis.* The phage attaches to the bacterial cell wall, penetrates inside, and inserts its gene (along with luciferase gene) into the *M. tuberculosis* chromosome. The enzyme luciferase is produced by the bacterium. When luciferin and ATP are added to the culture medium, luciferase cleaves luciferin. This reaction is accompanied by a flash of light which can be detected by a luminometer. This diagnostic test is quite sensitive for the confirmation of tuberculosis.

The ***flash of light is specific for the identification of M. tuberculosis*** in the culture. For other bacteria, the genetically engineered phage cannot attach and enter in, hence no flash of light would be detected.

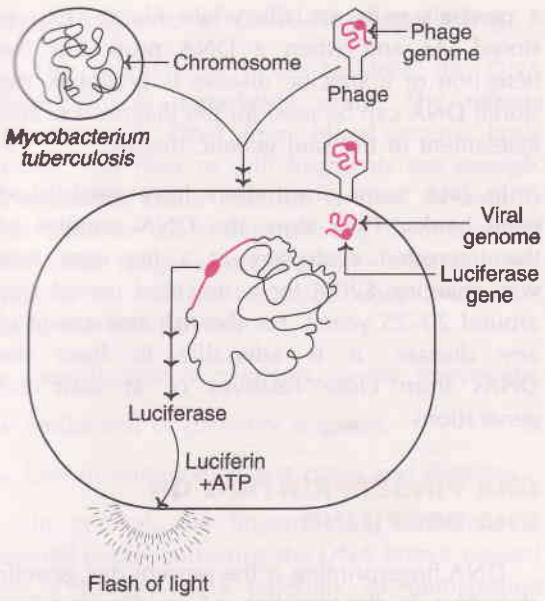


Fig. 27.20 : Diagnosis of tuberculosis by using a genetically engineered bacteriophage (phage).

Malaria

Malaria, mainly caused by *Plasmodium falciparum*, and *P. vivax*, affects about one-third of the world's population. The commonly used laboratory tests for the diagnosis of malaria include microscopic examination of blood smears, and detection of antibodies in the circulation. While the former is time consuming and frequently gives false-negative tests, the latter cannot distinguish between the past and present infections.

A ***specific DNA diagnostic test*** for identification of the current infection of *P. falciparum* has been developed. This is carried out by ***using a DNA probe that can bind and hybridize with a DNA fragment of P. falciparum genome***, and not with other species of *Plasmodium*. It is reported that this DNA probe can detect as little as 1 ng of *P. falciparum* in blood or 10 pg of its purified DNA.

Acquired immunodeficiency syndrome (AIDS)

DNA probes, with radioisotope label, for HIV DNA are now available. By ***using PCR and DNA probes, AIDS can be specifically diagnosed in the laboratory***.

DNA IN THE DIAGNOSIS OF GENETIC DISEASES

Traditional laboratory tests for the diagnosis of genetic diseases are mostly based on the estimation of metabolites and/or enzymes. This is usually done after the onset of symptoms.

The laboratory tests based on DNA analysis can specifically diagnose the inherited diseases at the genetic level. DNA-based tests are useful to discover, well in advance, whether the individuals or their offsprings are at risk for any genetic disease. Further, such tests can also be employed for the prenatal diagnosis of hereditary disorders, besides identifying the carriers of genetic diseases.

Although not in routine use in the laboratory service, methods have been developed or being developed for the analysis of DNA in the diagnosis of several genetic diseases. These include sickle-cell anemia, cystic fibrosis, *Duchenne's muscular dystrophy*, *Huntington's* disease, fragile X syndrome, Alzheimer's disease, certain cancers (e.g. breast cancer, colon cancer), type II diabetes, obesity, Parkinson's disease and baldness.

Sickle-cell anemia

Sickle-cell anemia is a genetic disease characterized by the irregular sickle (crescent like) shape of the erythrocytes. Biochemically, this disease results in severe anemia and progressive damage to major organs in the body (heart, brain, lungs, joints).

Sickle-cell anemia occurs due to a single amino acid change in the β-chain of hemoglobin. Specifically, the amino acid ***glutamate at the 6th position of β-chain is replaced by valine***. At the molecular level, sickle-cell anemia is due to a single-nucleotide change (***A → T***) in the β-globin gene of coding (or antisense) strand. In the normal β-globin gene the DNA sequence is CCTGAGGAG, while in sickle-cell anemia, the sequence is CCTGTGGAG. This single-base mutation can be detected by using restriction enzyme *MstII* to cut

PHARMACEUTICAL PRODUCTS OF DNA TECHNOLOGY

The advent of recombinant DNA technology heralded a new chapter for the production of a wide range of therapeutic agents in sufficient quantities for human use. The commercial exploitation of ***recombinant DNA (rDNA) technology*** began in late 1970s by a few biotechnological companies to produce proteins. There are at least 400 different proteins being produced (by DNA technology) which may serve as therapeutic agents for humans. A selected list of some important human proteins produced by recombinant DNA technology potential for the treatment of human disorders is given in ***Table 27.5***. As of now, only a selected few of them (around 30) have been approved for human use, and the most important among these are given in ***Table 27.6***.

INSULIN AND DIABETES

Diabetes mellitus is characterized by increased blood glucose concentration (hyperglycemia) which occurs due to insufficient or inefficient insulin. In the early years, insulin isolated and purified from the pancreases of pigs and cows was used for the treatment of severe diabetics. This often resulted in allergies. Recombinant DNA technology has become a boon to diabetic patients.

Production of recombinant insulin

Attemps to produce insulin by recombinant DNA technology started in late 1970s. The ***basic technique consisted of inserting human insulin gene and the promoter gene of lac operon on to the plasmids of* E. coli**. By this method human insulin was produced. It was in July 1980, seventeen human volunteers were, for the first time, administered recombinant insulin for treatment of diabetes at Guy's Hospital, London. And in fact, insulin was the first ever pharmaceutical product of recombinant DNA technology administered to humans. Recombinant insulin worked well, and this gave hope to scientists that DNA technology could be successfully employed to produce substances of

TABLE 27.5 A selected list of human proteins produced by recombinant DNA technology for treatment of human disorders

Disorder	*Recombinant protein(s)*
Anemia	Hemoglobin, erythropoietin
Asthma	Interleukin-I receptor
Atherosclerosis	Platelet-derived growth factor
Delivery	Relaxin
Blood clots	Tissue plasminogen activator, urokinase
Burns	Epidermal growth factor
Cancer	Interferons, tumor necrosis factor, colony stimulating factors, interleukins, lymphotoxin, macrophage-activating factor
Diabetes	Insulin, insulin-like growth factor
Emphysema	α_1-*Antitrypsin*
Female infertility	Chorionic gonadotropin
Free radical damage (minimizing)	Superoxide dismutase
Growth defects	Growth hormone, growth hormone-releasing factor, somatomedin-C
Heart attacks	Prourokinase
Hemophilia A	Factor VIII
Hemophilia B	Factor IX
Hepatitis B	Hepatitis B vaccine
Hypoalbuminemia	Serum albumin
Immune disorders	Interleukins, β-cell growth factors
Kidney disorders	Erythropoietin
Lou Gehrig's disease (amytrophic lateral sclerosis)	Brain-derived neurotropic factor
Multiple sclerosis	Interferons (α, β, γ)
Nerve damage	Nerve growth factor
Osteomalacia	Calcitonin
Pain	Endorphins and enkephalins
Rheumatic disease	Adrenocorticotropic hormone
Ulcers	Urogastrone
Viral infections	Interferons (α, β, γ)

- Nested PCR
- Inverse PCR
- Anchored PCR
- Reverse transcription PCR (RT-PCR)
- Asymmetric PCR
- Real-time quantitative PCR
- Random amplified polymorphic DNA (RAPD)
- Amplified fragment length polymorphism (AFLP)
- Rapid amplification of cDNA ends (RACE).

APPLICATIONS OF PCR

The advent of PCR had, and continues to have tremendous impact on molecular biology. The applications of PCR are too many to be listed here. Some of them are selectively and very briefly described. Other applications of PCR are discussed at appropriate places.

PCR in clinical diagnosis

The specificity and sensitivity of PCR is highly useful for the diagnosis of various diseases in humans. These include diagnosis of inherited disorders (genetic diseases), viral diseases, bacterial diseases etc.

Prenatal diagnosis of inherited diseases : PCR is employed in the prenatal diagnosis of inherited diseases by using chorionic villus samples or cells from amniocentesis. Thus, diseases like sickle-cell anemia, β-thalassemia and phenylketonuria can be detected by PCR in these samples.

Diagnosis of retroviral infections : PCR from cDNA is a valuable tool for diagnosis and monitoring of retroviral infections, e.g., HIV infection.

Diagnosis of bacterial infections : PCR is used for the detection of bacterial infections e.g., tuberculosis by *Mycobacterium tuberculosis.*

Diagnosis of cancers : Several virally-induced cancers (e.g., cervical cancer caused by human papilloma virus) can be detected by PCR. Further, some cancers which occur due to chromosomal translocation (chromosome 14 and 18 in follicular lymphoma) involving known genes are identified by PCR.

PCR in sex determination of embryos : Sex of human and live stock embryos fertilized *in vitro,* can be determined by PCR, by using primers and DNA probes specific for sex chromosomes. Further, this technique is also useful to detect sex — linked disorders in fertilized embryos.

PCR in DNA sequencing

As the PCR technique is much simpler and quicker to amplify the DNA, it is conveniently used for sequencing. For this purpose, single-strands of DNA are required.

PCR in comparative studies of genomes

The differences in the genomes of two organisms can be measured by PCR with random primers. The products are separated by electrophoresis for comparative identification. Two genomes from closely related organisms are expected to yield more similar bands.

PCR is very important in the study of ***evolutionary biology***, more specifically referred to as ***phylogenetics.*** As a technique which can amplify even minute quantities of DNA from any source (hair, mummified tissues, bone, or any fossilized material), PCR has revolutionized the studies in palaentology and archaelogy. The movie 'Jurassic Park', has created public awareness of the potential applications of PCR!

PCR in forensic medicine

A single molecule of DNA from any source (blood strains, hair, semen etc.) of an individual is adequate for amplification by PCR. Thus, PCR is very important for identification of criminals.

The reader may refer DNA finger printing technique described later in this chapter.

GENE LIBRARIES

The ***collection of DNA fragments*** (specifically genes) from a particular species represents gene libraries. The creation or construction of gene libraries (broadly ***genomic libraries***) is

The future of DNA chips

The major limitation of DNA chips at present is the unavailability of complete genome arrays for higher eukaryotes, including humans. It is expected that within the next few years such DNA chips will be available. This will help the biotechnologists to capture the functional snapshots of the genome in action for higher organisms.

POLYMERASE CHAIN REACTION (DNA AMPLIFICATION)

The polymerase chain reaction (PCR) is a laboratory (*in vitro*) ***technique for generating large quantities of a specified DNA***. Obviously, PCR is a ***cell-free amplification technique*** for synthesizing multiple identical copies (billions) of ***any DNA*** of interest. Developed in 1984 by Karry Mullis (Nobel Prize, 1993), PCR is now considered as a basic tool for the molecular biologist. As is a photocopier a basic requirement in an office, so is the PCR machine in a molecular biology laboratory!

Principle of PCR

The double-stranded DNA of interest is denatured to separate into two individual strands. Each strand is then allowed to hybridize with a primer (renaturation). The primer-template duplex is used for DNA synthesis (the enzyme-DNA polymerase). These three steps—***denaturation, renaturation*** and ***synthesis*** are repeated again and again to generate multiple forms of target DNA.

Technique of PCR

The essential requirements for PCR are listed below

1. A target DNA (100–35,000 bp in length).

2. Two primers (synthetic oligonucleotides of 17–30 nucleotides length) that are complementary to regions flanking the target DNA.

3. Four deoxyribonucleotides (dATP, dCTP, dGTP, dTTP).

4. A DNA polymerase that can withstand at a temperature up to 95°C (i.e., thermostable).

The actual technique of PCR involves repeated cycles for amplification of target DNA. Each cycle has three stages.

1. **Denaturation :** On raising the temperature to about 95°C for about one minute, the DNA gets denatured and the two strands separate.

2. **Renaturation or annealing :** As the temperature of the mixutre is slowly cooled to about 55°C, the primers base pair with the complementary regions flanking target DNA strands. This process is called renaturation or annealing. High concentration of primer ensures annealing between each DNA strand and the primer rather than the two strands of DNA.

3. **Synthesis :** The initiation of DNA synthesis occurs at 3′-hydroxyl end of each primer. The primers are extended by joining the bases complementary to DNA strands. The synthetic process in PCR is quite comparable to the DNA replication of the leading strand (***Refer Chapter 24***). However, the temperature has to be kept optimal as required by the enzyme DNA polymerase. For *Taq* DNA polymerase, the optimum temperature is around 75°C (for *E. coli* DNA polymerase, it is around 37°C). The reaction can be stopped by raising the temperature (to about 95°C).

The 3 stages of PCR in relation to temperature and time are depicted in ***Fig.27.15***. Each cycle of PCR takes about 3-5 minutes. In the normal practice, the PCR is carried out in an automated machine.

As is evident from the ***Fig.27.16*** (cycle I), the new DNA strand joined to each primer is beyond the sequence that is complementary to the second primer. These new strands are referred to as ***long templates***, and they will be used in the second cycle.

For the second cycle of PCR, the DNA strands (original + newly synthesized long template) are denatured, annealed with primers and subjected to DNA synthesis. At the end of second round, long templates, and short templates (DNA

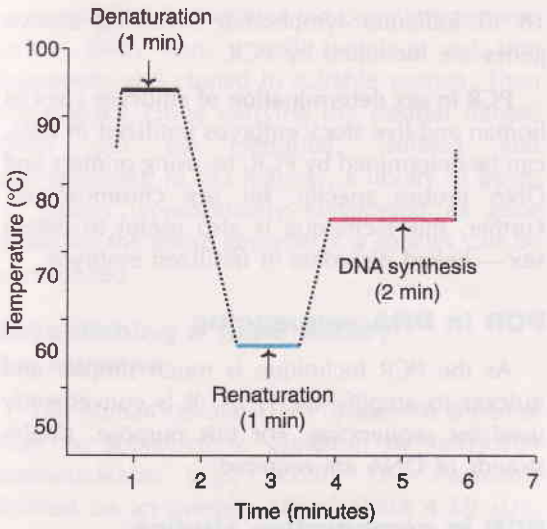


Fig. 27.15 : *The three stages in each cycle of PCR in relation to temperature and time (Each cycle takes approximately 3-5 minutes).*

strands with primer sequence at one end, and sequence complementary to the other end primer) are formed.

In the third cycle of PCR, the original DNA strands along with long and short templates are the starting materials. The technique of denaturation, renaturation and synthesis are repeated. This procedure is repeated again and again for each cycle. It is estimated that at the end of 32nd cycle of PCR, about a million-fold target DNA is synthesized. The short templates possessing precisely the target DNA as double-stranded molecules accumulate.

Sources of DNA polymerase

In the original technique of PCR, Klenow fragment of *E. coli* DNA polymerase was used. This enzyme, gets denatured at higher temperature, therefore, fresh enzyme had to be added for each cycle. A breakthrough occurred (Lawyer 1989) with the introduction of *Taq* DNA polymerase from thermophilic bacterium, *Thermus aquaticus*. The ***Taq DNA polymerase is heat resistant, hence it is not necessary to freshly add this enzyme for each cycle of PCR.***

Variations of PCR

The basic technique of PCR has been described. Being a versatile technique, PCR is modified as per the specific demands of the situation. Some of the variants of PCR are listed

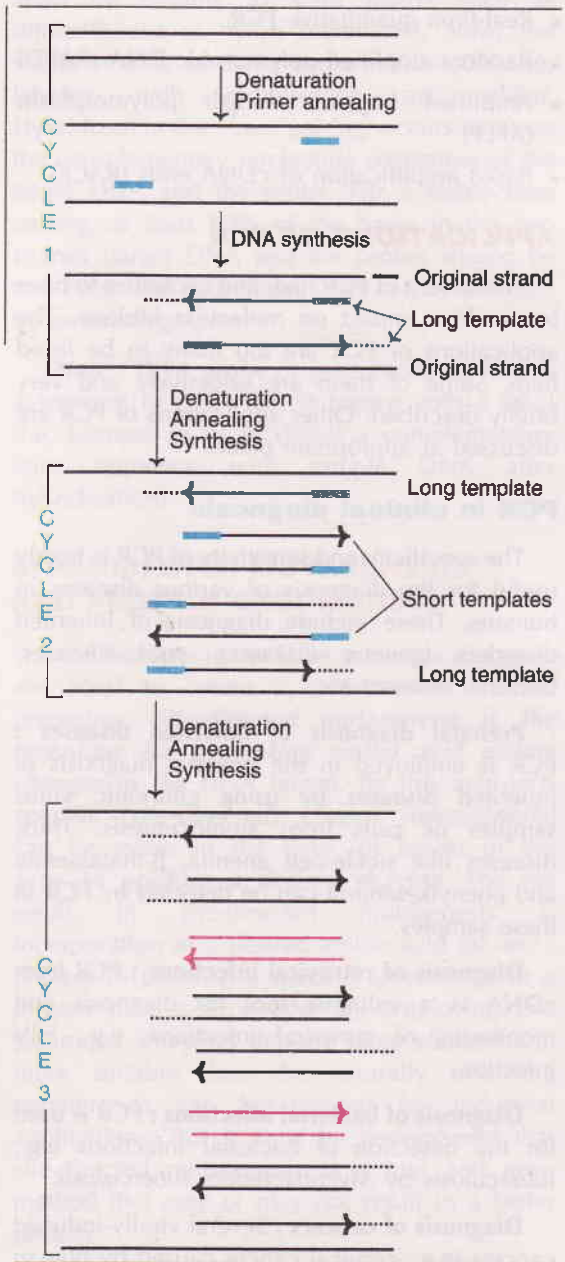


Fig. 27.16 : *The polymerase chain reaction (PCR) representing the initial three cycles (—— indicate primers).*

requirement in biotechnology. DNA sequencing is important to understand the functions of genes, and basis of inherited disorders. Further, DNA cloning and gene manipulation invariably require knowledge of accurate nucleotide sequence.

MAXAM AND GILBERT TECHNIQUE

The first DNA sequencing technique, ***using chemical reagents***, was developed by Maxam and Gilbert (1977). This method is briefly described below (**Fig.27.10**).

A strand of source DNA is labeled at one end with ^{32}P. The two strands of DNA are then separated. The labeled DNA is distributed into four samples (in separate tubes). Each sample is subjected to treatment with a chemical that specifically destroys one (G, C) or two bases (A + G, T + C) in the DNA. Thus, the DNA strands are partially digested in four samples at sites G, A + G, T + C and C. This results in the formation of a series of labeled fragments of varying lengths. The actual length of the fragment depends on the site at which the base is destroyed from the labeled end. Thus for instance, if there are C residues at positions 4, 7, and 10 away from the labeled end, then the treatment of DNA that specifically destroys C will give labeled pieces of length 3, 6 and 9 bases. The labeled DNA fragments obtained in the four tubes are subjected to electrophoresis side by side and they are detected by autoradiograph. The sequence of the bases in the DNA can be constructed from the bands on the electrophoresis.

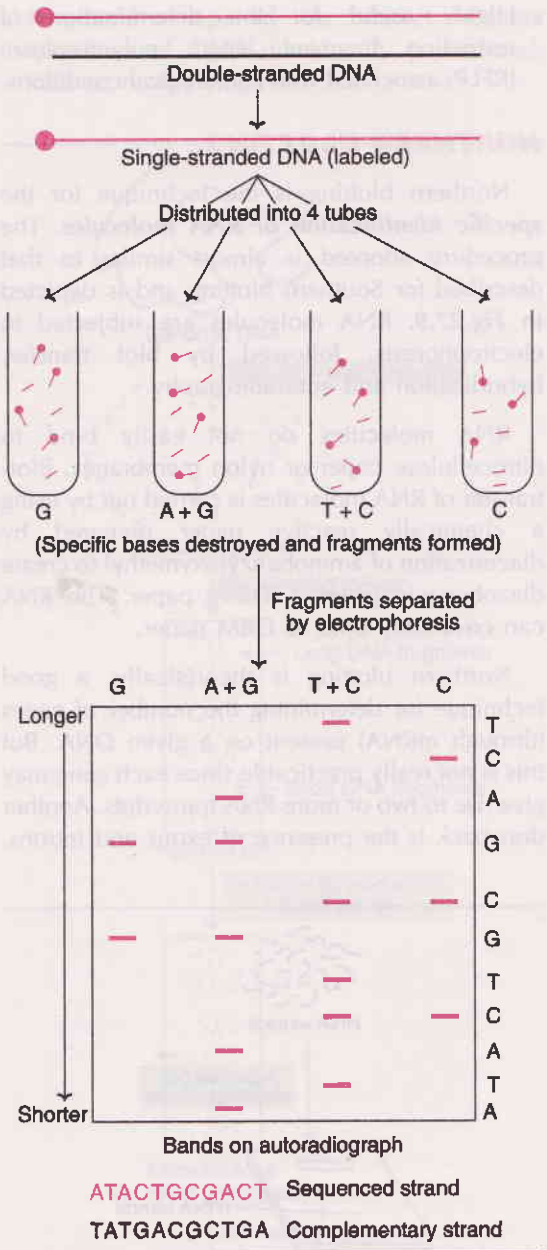


Fig. 27.10 : Maxam and Gilbert method for DNA sequencing.

DIDEOXYNUCLEOTIDE METHOD

Currently, the preferred technique for determining nucleotide sequence in DNA is the one ***developed by Sanger*** (1980). This is an ***enzymatic procedure*** commonly referred to as the dideoxynucleotide method or ***chain termination method*** (**Note :** Fredrick Sanger won Nobel prize twice, once for determining the structure of protein, insulin; the second time for sequencing the nucleotides in an RNA virus).

A dideoxynucleotide is a laboratory-made chemical molecule that lacks a hydroxyl group at both the 2′ and 3′ carbons of the sugar (***Fig.27.11***). This is in contrast to the natural deoxyribonucleotide that possesses at 3′ hydroxyl group on the sugar.

Termination role of dideoxynucleotide : In the normal process of DNA replication, an

vaccine consists of a gene encoding an antigenic protein, inserted onto a plasmid, and then incorported into the cells in a target animal. The plasmid carrying DNA vaccine normally contains a promoter site, cloning site for the DNA vaccine gene, origin of replication, a selectable marker sequence (e.g. a gene for ampicillin resistance) and a terminator sequence (a poly—A tail).

DNA vaccine—plasmids can be administered to the animals by one of the following delivery methods.

- Nasal spray
- Intramuscular injection
- Intravenous injection
- Intradermal injection
- Gene gun or biolistic delivery (involves pressure delivery of DNA-coated gold beads).

DNA VACCINE AND IMMUNITY

An illustration of a DNA vaccine and the mechanism of its action in developing

BIOMEDICAL / CLINICAL CONCEPTS

- *Biotechnology is a newly discovered discipline for age-old practices (e.g. preparation of curd, wine, beer), with special emphasis on genetic manipulations.*
- *Human artificial chromosome (HAC) is a synthetic vector, possessing the characteristics of human chromosome. HAC is capable of carrying large-sized human genes that may be useful in gene therapy.*
- *Southern blotting technique (that specifically detects DNA) is employed for the identification of thieves, rapists, and settlement of parenthood.*
- *Polymerase chain reaction is useful for the diagnosis of inherited diseases, in DNA sequencing, and in forensic medicine.*
- *By employing site-directed mutagenesis, it is possible to produce more efficient and more suitable enzymes for therapeutic and industrial purposes.*
- *The analysis of genetic material DNA (gene/genes) is employed for the diagnosis of certain diseases, and in medical forensics e.g. AIDS, sickle-cell anemia, certain cancers, DNA fingerprinting.*
- *The pharmaceutical products of rDNA technology have revolutionized the treatment of certain diseases e.g. diabetes, asthma, atherosclerosis, heart attacks, hemophilia.*
- *Recombinant vaccine for hepatitis B is the first synthetic vaccine. It is effective, safe and produces no allergic reactions.*
- *Genetic immunization by using DNA vaccines is a novel concept. It has been shown that the immune response (humoral and cellular) of the body can be stimulated by a DNA molecule.*
- *Transgenic mice that serve as animal models for human diseases have been developed. These include human mouse (model for immune system), Alzheimer's mouse, oncomouse (model for cancer), prostate mouse, knockout mice (for allergy, transplantation etc.).*
- *Transgenic animals serve as bioreactors for the production of therapeutically important proteins e.g. interferon, lactoferrin, urokinase.*
- *Certain pet animals (cats, dogs) are being cloned by some companies.*

immunity is given in ***Fig.27.29***. The plasmid vaccine carrying the DNA (gene) for antigenic protein enters the nucleus of the inoculated target cell of the host. This DNA produces RNA *and in turn the specific antigenic* protein. The antigen can act directly for developing humoral immunity or as fragments in association with major histocompatability class (MHC) molecules for developing cellular immunity.

Humoral immunity

As the antigen molecules bind to B-lymphocytes, they trigger the production of antibodies which can destroy the pathogens. Some of the B-lymphocytes become memory cells that can protect the host against future infections.

Cellular immunity

The protein fragments of the antigen bound to MHC molecules can activate the cytotoxic T-lymphocytes. They are capable of destroying the infected pathogenic cells. Some of the activated T-lymphocytes become memory cells which can kill the future infecting pathogens.

TRANSGENIC ANIMALS

With the advent of modern biotechnology, it is now possible to carry out manipulations at the genetic level to get the desired characteristics in animals. ***Transgenesis refers to the phenomenon of introduction of exogeneous DNA into the genome to create and maintain a stable heritable character.*** The foreign DNA that is introduced is called ***transgene.*** And the animal whose genome is altered by adding one or more transgenes is said to be transgenic. The transgenes behave like other genes present in the animals' genome, and are passed on to the offsprings. Thus, transgenic animals are ***genetically engineered*** or ***genetically modified organisms (GMOs)*** with a new heritable character.

Importance of transgenic animals—general

Transgenesis has now become a powerful tool for studying the gene expression and developmental processes in higher organisms, besides the improvement in their genetic characteristics. Transgenic animals serve as good models for understanding the human diseases. Further, several proteins produced by transgenic animals are important for medical and pharmaceutical applications. Thus, the transgenic farm animals are a part of the lucrative world-wide biotechnology industry, with great benefits to mankind.

TRANSGENIC MICE AND THEIR APPLICATIONS

Mouse, although not close to humans in its biology, has been and continues to be the most exploited animal model in transgenesis experiments. The common feature between man and mouse is that both are mammals. Transgenic mice are extensively used as animal models for understanding human diseases, and for the production of therapeutic agents. Adequate care, however, must be exercised before extrapolating data of transgenic mice to humans.

Mouse models for several human diseases (cancers, muscular dystrophy, arthritis, Alzheimer's disease, hypertension, allergy, coronary heart disease, endocrine diseases, neurodegenerative disorders etc.) have been developed. A selected few of them are listed.

- ***The human mouse***, the transgenic mouse that displays human immune system.
- ***The Alzheimer's mouse*** to understand the pathological basis of Alzheimer's disease.
- ***The oncomouse***, the animal model for cancer.
- ***The prostate mouse***, the transgenic mouse to understand prostate cancer.
- ***The knockout mice***, (developed by eliminating specific genes) for certain diseases e.g. SCID mouse, knockout mouse for trans-plantation.

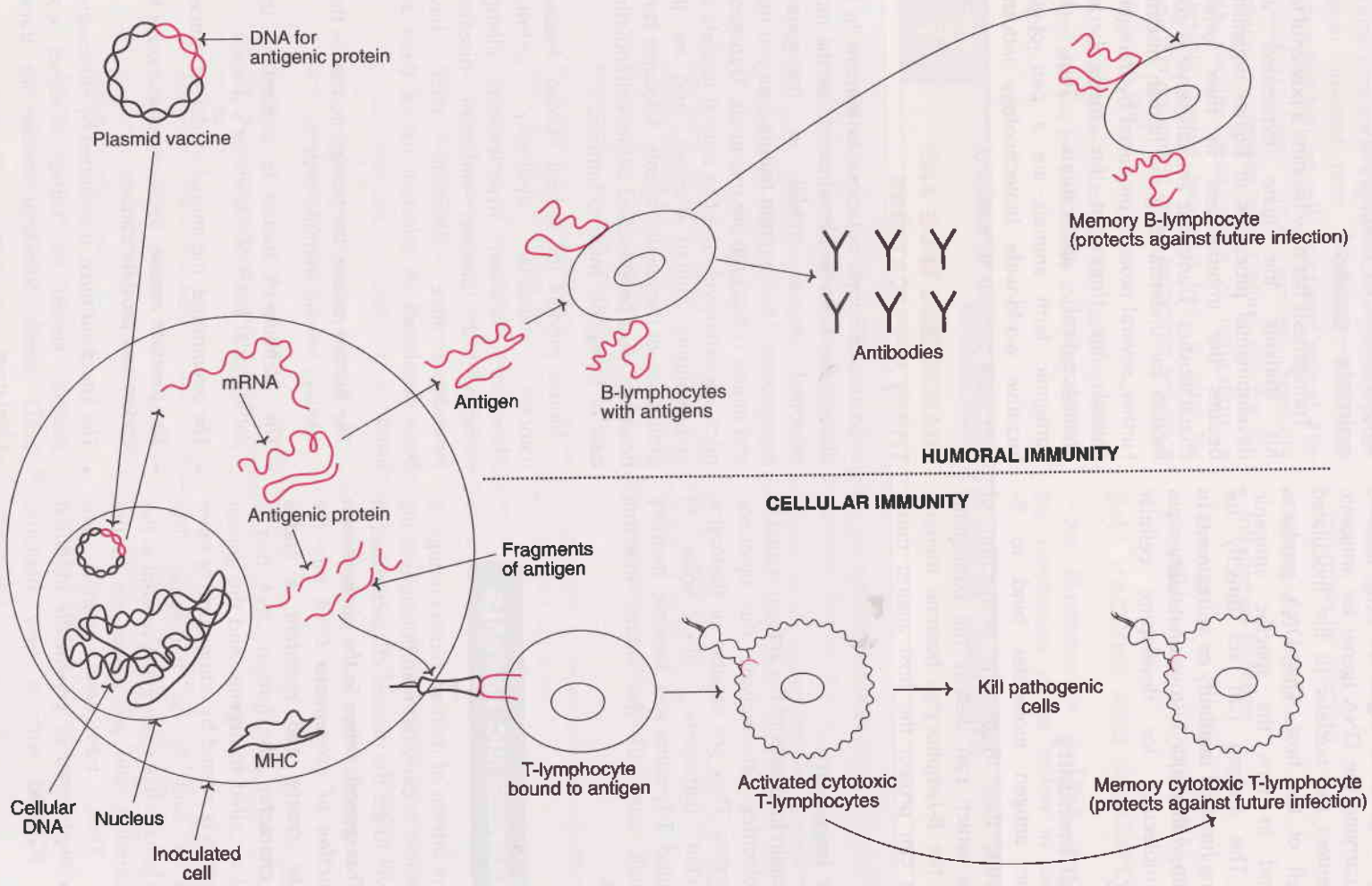


Fig. 27.29 : DNA vaccine and mechanism of its action in developing immunity (MHC–Major histocompatability complex molecule)

ANIMAL BIOREACTORS

Transgenesis is wonderfully utilized for production of proteins for pharmaceutical and medical use. In fact, any protein synthesized in the human body can be made in the transgenic animals, provided that the genes are correctly programmed. The advantage with transgenic animals is to produce scarce human proteins in huge quantities. Thus, the ***animals serving as factories for production of biologically important products*** are referred to as ***animal bioreactors or*** sometimes ***pharm animals***. Some transgenic animals that serve as bioreactors are listed

- Transgenic cow for the production of lactoferrin and interferons.
- Transgenic goat to synthesize tissue plasminogen activator, and antithrombin III.
- Transgenic mouse for the production of immunoglobulins, and urokinase.
- Transgenic pig to produce hemoglobin.

DOLLY – THE TRANSGENIC CLONE

Dolly, the first ever mammal clone was developed by Wilmut and Campbell in 1997. It is a ***sheep*** (female lamb) ***with a mother and no father.***

The ***technique primarily involves nuclear transfer and the phenomenon of totipotency.*** The character of a cell to develop into different cells, tissues, organs, and finally an organism is referred to as ***totipotency*** or ***pluripotency.*** Totipotency is the basic character of embryonic cells. As the embryo develops, the cells specialize to finally give the whole organism. As such, the cells of an adult lack totipotency. Totipotency was induced into the adult cells for developing Dolly.

The cloning of sheep for producing Dolly, illustrated in ***Fig.27.30***, is briefly described here. The mammary gland cells from a donor ewe were isolated. They were subjected to total nutrient deprivation (starvation) for five days. By this process, the mammary cells abandon their

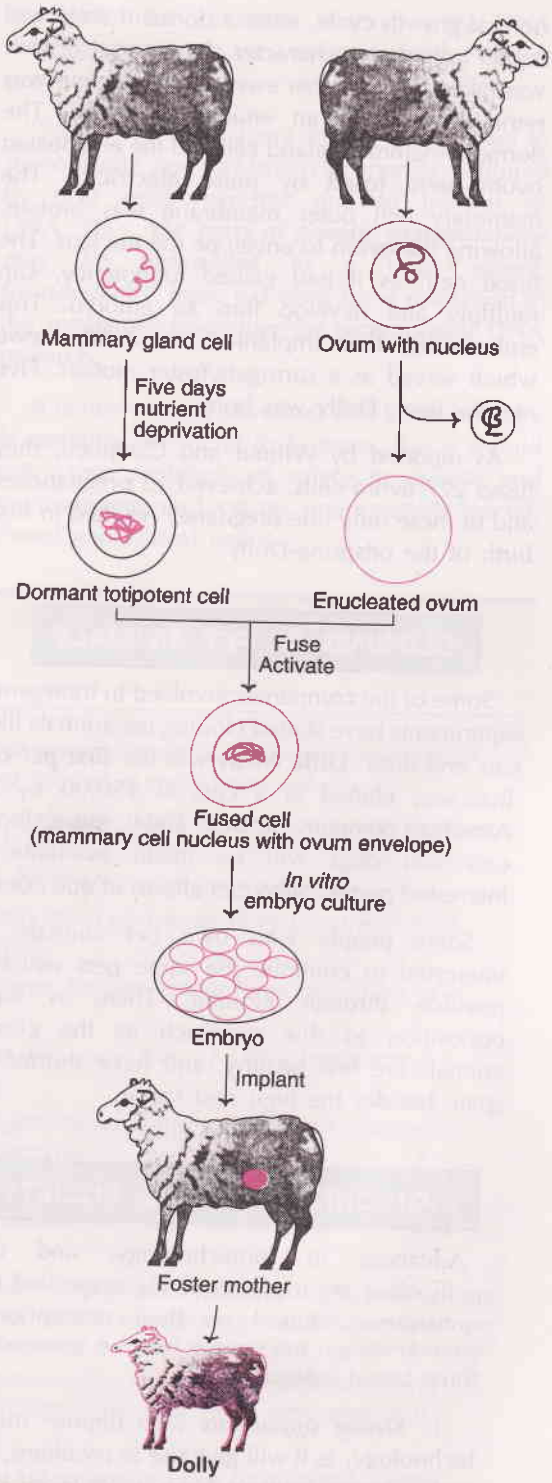


Fig. 27.30 : The cloning of sheep for developing Dolly.

normal growth cycle, enter a dormant stage and regain totipotency character. An ovum (egg cell) was taken from another ewe, and its nucleus was removed to form an enucleated ovum. The dormant mammary gland cell and the enucleated ovum were fused by pulse electricity. The mammary cell outer membrane was broken, allowing the ovum to envelope the nucleus. The fused cell, as it had gained totipotency, can multiply and develop into an embryo. This embryo was then implanted into another ewe which served as a surrogate/foster mother. Five months later, Dolly was born.

As reported by Wilmut and Campbell, they fused 277 ovum cells, achieved 13 pregnancies, and of these only one pregnancy resulted in live birth of the offspring-Dolly.

CLONING OF PET ANIMALS

Some of the companies involved in transgenic experiments have started cloning pet animals like cats and dogs. ***Little Nicky*** was the ***first pet cat*** that was ***cloned*** at a cost of $50,00 by an American company (in Dec. 2004). More cloned cats and dogs will be made available to interested parties (who can afford) in due course.

Some people who own pet animals are interested to continue the same pets which is possible through cloning. There is some opposition to this approach as the cloned animals are less healthy, and have shorter life span, besides the high cost factor.

BIOTECHNOLOGY AND SOCIETY

Advances in biotechnology, and their applications are most frequently associated with controversies. Based on their perception to biotechnology, the people may be grouped into three broad categories.

1. ***Strong opponents*** who oppose the new technology, as it will give rise to problems, issues and concerns humans have never faced before. They consider biotechnology as an unnatural manipulative technology.

2. ***Strong proponents*** who consider that the biotechnology will provide untold benefits to society. They argue that for centuries the society has safely used the products and processes of biotechnology.

3. A ***neutral group*** of people who have a ***balanced approach*** to biotechnology. This group believes that research on biotechnology (with regulatory systems), and extending its fruits to the society should be pursued with a cautious approach.

BENEFITS OF BIOTECHNOLOGY

The fruits of biotechnology are beneficial to the fields of ***healthcare***, ***agriculture***, ***food production***, ***manufacture of industrial enzymes*** and appropriate environmental management.

It is a fact that modern technology in various forms is woven tightly into the fabric of our lives. Our day-to-day life is inseparable from technology. Imagine life about 1-2 centuries ago where there was no electricity, no running water, sewage in the streets, unpredictable food supply and an expected life span of less than 40 years. Undoubtedly, technology has largely contributed to the present day world we live in. Many pepople consider biotechnology as a technology that will improve the quality of life in every country, besides maintaining living standards at a reasonably higher level.

ELSI OF BIOTECHNOLOGY

Why so much uproar and negativity to biotechnology? This is mainly because the major part of the modern biotechnology deals with genetic manipulations. These unnatural genetic manipulations, as many people fear, may lead to unknown consequences.

ELSI is the short form to represent the ***ethical, legal and social implications of biotechnology.*** ELSI broadly covers the relationship between biotechnology and society with particular reference to ethical and legal aspects.

Risks and ethics of biotechnology

The modern biotechnology deals with genetic manipulations of viruses, bacteria, plants,

animals, fish and birds. Introduction of foreign genes into various organisms raises concerns about the safety, ethics and unforeseen consequences. Some of the popular phrases used in the media while referring to experiments on recombinant DNA technology are listed.

- Manipulation of life
- Playing God
- Man-made evolution

The major apprehension of genetic engineering is that through recombinant DNA experiments, unique microorganisms or viruses (either inadvertently, or sometimes deliberately for the purpose of war) may be developed that would cause epidemics and environmental catastrophes. Due to these fears, the regulatory guidelines for research dealing with DNA manipulation were very stringent in the earlier years.

So far, risk assessment studies have failed to demonstrate any hazardous properties acquired by host cells/organisms due to transfer of DNA. Thus, ***the fears of genetic manipulations may be unfounded*** to a large extent. Consequently, there has been some relaxation in the regulatory guidelines for recombinant DNA research.

It is now widely accepted that biotechnology is certainly beneficial to humans. But it should not cause problems of safety to people and environment, and create unacceptable social, moral and ethical issues.

SUMMARY

1. *Recombinant DNA (rDNA) technology is primarily concerned with the manipulation of genetic material (DNA) to achieve the desired goal in a pre-determined way.*
2. *The procedure for rDNA technology involves molecular tools (enzymes e.g. restriction endonucleases), host cells (E. coli, S. cerevisiae), vectors (plasmids, bacteriophages), gene transfer (transformation, electroporation) and the strategies of gene cloning.*
3. *Blotting techniques are employed for the identification of desired DNA (Southern blot), RNA (Northern blot), and protein (Western blot).*
4. *Polymerase chain reaction is an in vitro technique for generating large quantities of a specified DNA i.e. cell-free amplification.*
5. *Gene libraries or genomic libraries represents the collection of DNA fragments (i.e. genes) from a genome of a particular species.*
6. *Site-directed mutagenesis is the technique for generating amino acid coding changes in the DNA (gene) to produce a desired protein/enzyme.*
7. *Analysis of DNA (i.e. detection of gene/genes) can be used as a diagnostic system for the detection of many pathogenic and genetic diseases e.g. tuberculosis, malaria, AIDS, sickle-cell anemia, certain cancers.*
8. *DNA fingerprinting or DNA profiling is the present day genetic detective in the practice of modern medical forensics. Four types of DNA markers are used in DNA fingerprinting–RFLFs, VNTRs, STRs, and SNPs.*
9. *Many pharmaceutical compounds of health importance (for disease treatment) are being produced by rDNA technology e.g. insulin, growth hormone, interferons, erythropoietin, hepatitis B vaccine.*
10. *Transgenic animals can be developed by introducing a foreign DNA (transgene). These animals are genetically modified or engineered with new heritable characters e.g. oncomouse, knockout mouse, prostate mouse.*

SELF-ASSESSMENT EXERCISES

I. Essay questions

1. Describe the basic principles underlying the recombinant DNA technology.
2. Give an account of the nucleic acid blotting techniques. Add a note on their importance.
3. Describe the polymerase chain reaction along with its applications.
4. Write briefly on the utility of DNA in disease diagnosis and medical forensics.
5. Give an account of the pharmaceutical products of DNA technology.

III. Short notes

(a) Restriction endonucleases, (b) Plasmids, (c) Methods of gene transfer, (d) Purification of nucleic acids, (e) Western blotting, (f) DNA sequencing, (g) DNA chips (h) Gene libraries, (i) Restriction fragment length polymorphisms, (j) Recombinant vaccines.

III. Fill in the blanks

1. The most commonly used prokaryotic host in rDNA technology is _____.
2. Northern blotting technique is used for the detection of _____.
3. Name the blotting technique in which nucleic acids (DNA or RNA) are directly blotted onto the filters without electrophoresis_____.
4. The bacterial source of the enzyme *Taq* DNA polymerase, that is widely used in polymerase chain reaction _____.
5. The collection of DNA fragments from the genome of a particular species represents _____.
6. The technique for generating amino acid coding changes in the DNA (gene) is regarded as _____.
7. The trade name for insulin produced by rDNA technology _____.
8. The first synthetic veccine developed by rDNA technology _____.
9. The most commonly used animal model in transgenesis to represent humans _____.
10. Name the first ever mammmal that has been cloned _____.

IV. Multiple choice questions

11. One of the following enzyme produces single-stranded nicks in DNA

    (a) DNA ligase (b) DNA polymerase (c) DNase I (d) SI nuclease.
12. Western blotting is the technique for the identification of

    (a) DNA (b) RNA (c) Carbohydrates (d) Proteins.
13. The DNA markers used in the diagnosis of diseases and DNA fingerprinting

    (a) Restriction fragment length polymorphisms, (b) Minisatellites and microsatellites, (c) Single nucleotide polymorphisms, (d) Any one of the above.
14. The first pharmaceutical product of recombinant DNA technology approved for human use

    (a) Insulin (b) Growth hormone (c) Interferon (d) Hypatitis B vaccine.
15. Genetic immunization involves the administration of

    (a) Antigens (b) Antibodies (c) DNA (d) RNA.