MICROBIAL GENETICS AND MOLECULAR BIOLOGY (DMB23) (MSC MICROBIOLOGY)

ACHARYA NAGARJUNA UNIVERSITY

CENTRE FOR DISTANCE EDUCATION

NAGARJUNA NAGAR,

GUNTUR

ANDHRA PRADESH

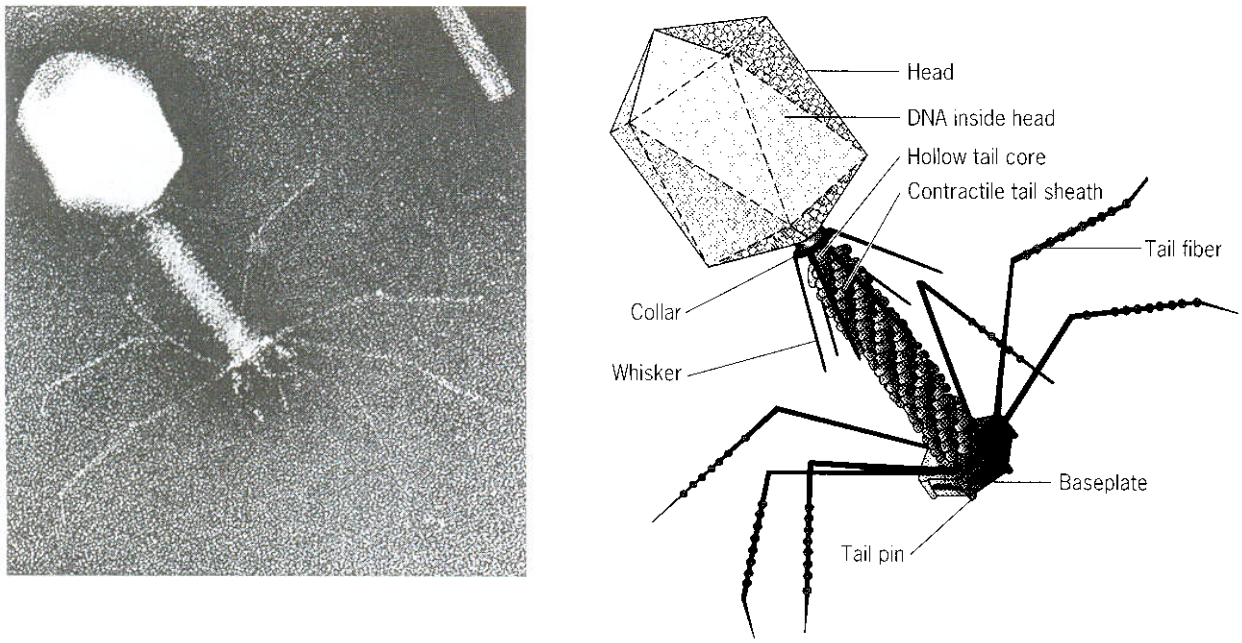
Lesson No. 4

Genetic Recombination in Phage - Fine structure of gene and map of T4 Phage

4.1 INTRODUCTION

Viruses are ultramicroscopic intracellular parasites, capable of autonomous replication inside a host cell. Viruses were established as distinct from bacteria by the work of Frederick Twort and Felix D'Herelle during 1910-1917. D'He'relle called viruses, destroying bacteria as bacteriophages. Researchers began using bacteriophages as experimental organisms in 1930s to understand more about the genetic fine structure.

T phages, TI through T7 of *E. coli* are among the most intensely studied bacterial viruses. Though some structural variation occurs between T-phages, they are all tadpole-shaped (Fig. 4.1). The head is composed of several proteins, a polyhedron. Inside the head is a single molecule of double-stranded DNA. Attached to the head is a tail, composed of hollow tubes. The tail fibres uses to find a bacteria to infect. The tail pins anchor to virus to the host cell during infection.



(a) (b)

Fig. 4.1 Electron micrograph (a) and diagram (b) showin the structure of bacteriophage T4.

The structure of bacteriophage T4 is given (Fig. 4.1). When phage has absorbed to the bacterial cell surface, phage genetic material enters the bacterial cell. During the infection, the cells genetic material is destroyed, while the genetic material of the virus is replicated many times. New virus particles are assembled within the host cell, which bursts open releasing several hundred viral particles.

Certain phages are capable of replicating through different life-cycle stages lytic and lysogenic cycles.

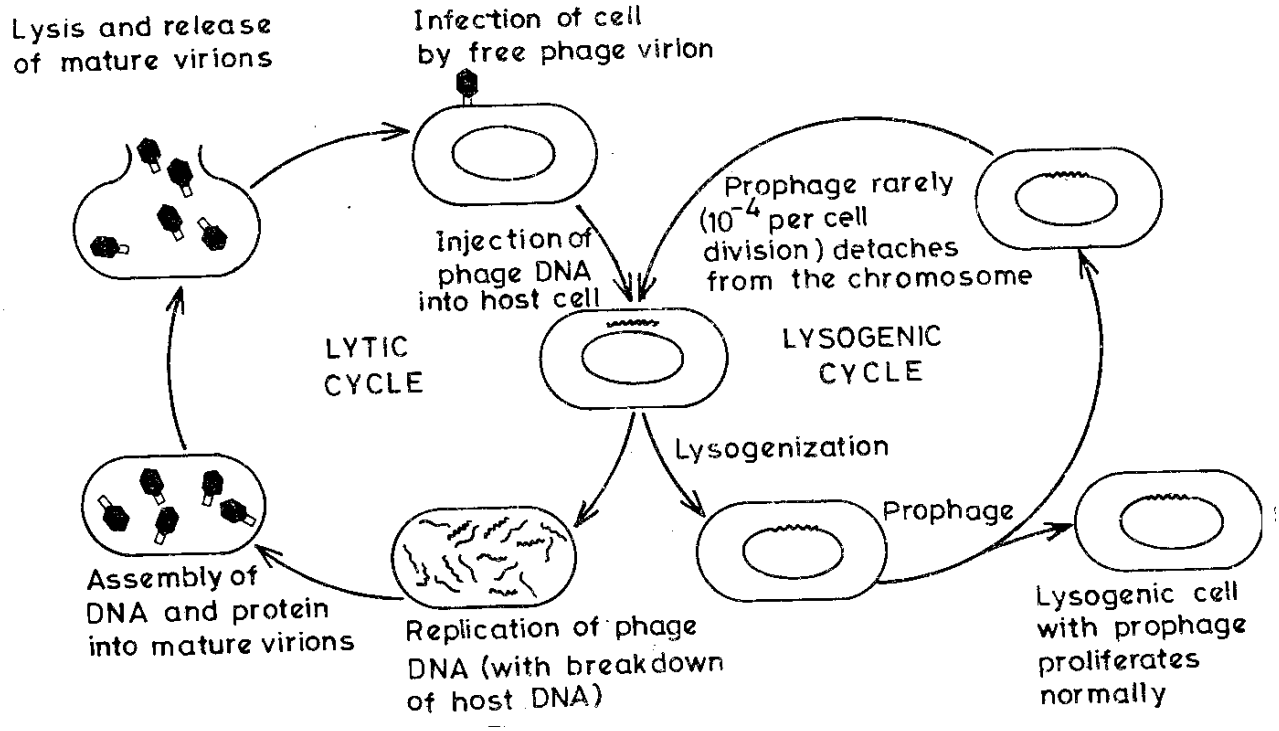
4.2 PHAGE LIFE CYCLES

4.2.1 Lytic cycle

They replicate in the host cytoplasm and cause destruction of the host cell and liberate phage particles (Fig. 4.2).

4.2.2 Lysogenic cycle

These phages are capable of integrating into the host chromosome. Some temperate bacteriophages such as phage lambda, when infects *E. coli*, it may either cause lysis or become lysogenic depending on the prevailing conditions. The integrated phage is termed a prophage (Fig. 4.2). Majority of research on lysogeny has been done on phage λ



4.4 PHAGE PHENOTYPES

Plaque morphology: It is a phenotypic trait used in phage genetics research. A plaque is a clear area in a lawn of bacterial cells that result from killing of contiguous cells by several cycles of bacteriophage growth. Plaques may be large or small with sharp or fuzzy edges (Fig. 4.3).

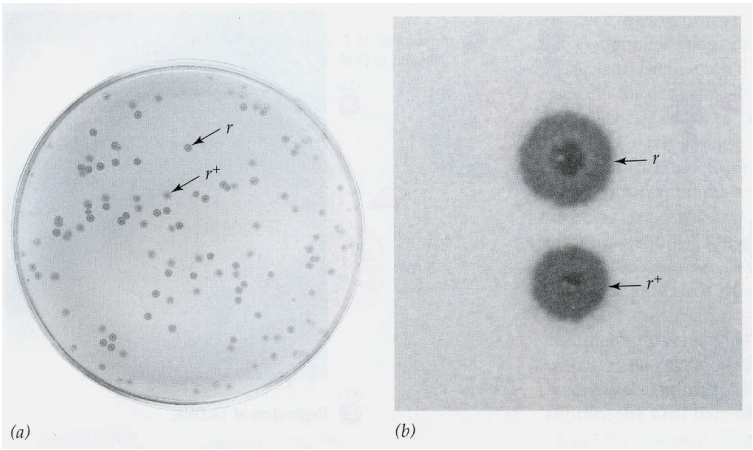


Fig. 4.3 (a) Wild-type (r^+) and rapid-lysis (R) mutant plaques of bacteriophage. T4 growing on a confluent lawn of *E. coli* strain B cells. (b) Higher magnification of individual r^+ and *r* plaques.

Host range: It is another frequently studied phenotypic trait in bacteriophages. Host range mutants are able to infect cells of some bacterial strains but not those of other strains. All wild type e.g. *E. coli* strain B cells can be infected by T-even bacteriophages. However, *E. coli*, B/2, a mutant *E. coli* is resistant to infection by T2 phage.

Genetic recombination in phage

In 1946, Max Delbrick and Alfred Hershey independently announced, the discovery of genetic recombination in phages. They simultaneously infected *E. coli* cells with two different strains of bacteriophage (Fig. 4.4) T2 – (1) h^+r (2) $h\ r^+$ i.e.

h^+ wild type phages
h - host range mutants
r^+ - wild type producing small plaques
r - rapid lysis mutants producing large plaques

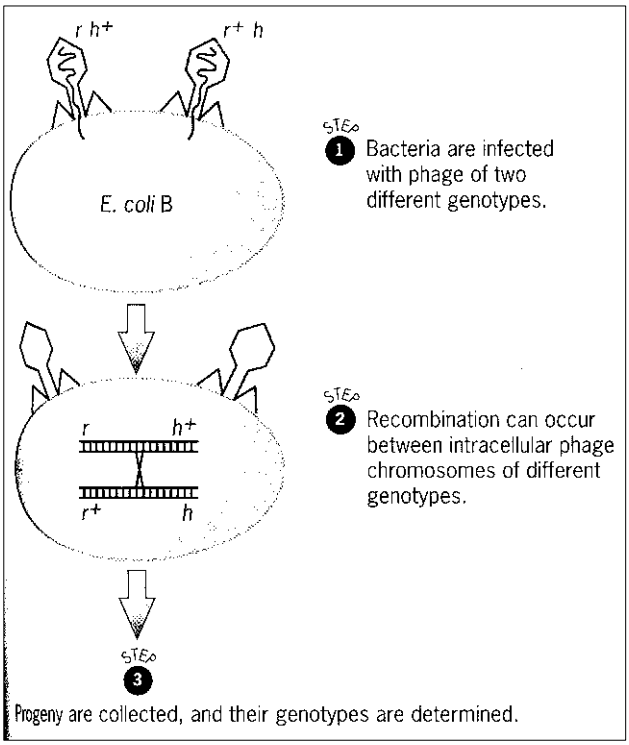


Fig. 4.4 A cross between *r* h^+ and r^+ *h* bacteriophage. Phage crosses are performed by simultaneously infecting bacteria with viruses of two different genotypes.

(1) They simultaneously infected *E. coli* cells with different strains of bacteriophage i.e. h^+r, hr^+.
(2) They collected progeny phage from this mixed infection.
(3) They produced a distinct plaque morphology when grown on these mixed lawns (Fig. 4.5).

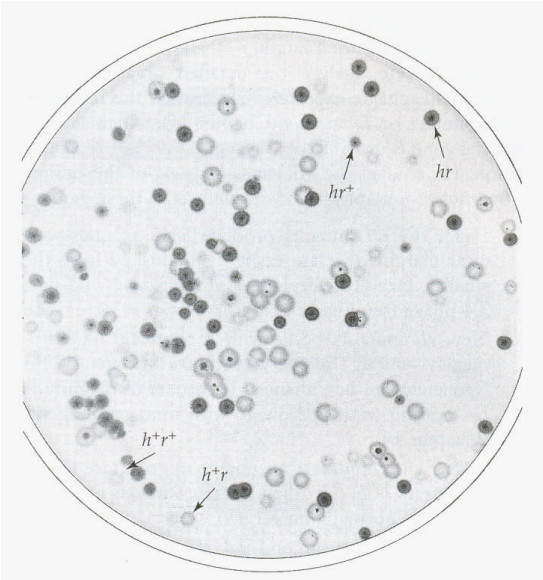


Fig. 4.5 Photograph showing the types of plaques formed by T2 *h r*, h^+ r^+, *h* r^+ , and h^+ *r* phage on a mixed lawn of *E. coli* B and B/2 cells

(4) h^+r plaques were large and turbid; and h r^+ plaques were small and clear. These are parental genotypes

(5) Wild type (h^+ r^+) plaques were small and turbid, nr plaques were large and clear. They are recom genotypes.

(6) Hershey and Rotman found that 2 per cent of the progeny were recombinant and 98 per cent of the progeny were parental.

Parental types

h + r

h r^+

Recombinants

h^+r^+

h r

A single cross over in this region will produce the recombinants. The percentage of recombination is the map distance, which is a relative index of distance between loci. The greater the physical distance, the greater the amount of recombination. The discovery of recombination allowed them to estimate that the distance between h and r genes was about 2 map units.

Two important points emerged from the early phage mapping studies.

1. Recombination appeared to occur by a reciprocal process as occurred during meiosis in eukaryotic organisms. If one recombinant class was present among the progeny produced by a population of doubly infected bacteria, the other recombinant class was present at approximately the same frequency.

2. Recombination in phages is not the same as meiotic recombination. In meiosis, chromosomes synapse and cross over at a specific stage. In viruses there is no process as complicated as meiosis. Recombination can occur at any time during the phage life cycle as long as DNA has not been packaged into protein coats. Multiple rounds of genetic exchange can occur.

4.5 GENE FINE STRUCTURE

Gene mapping took an entirely new dimension in mid 1950s, when Seymour Benzer mapped 2400 r II mutants to 304 sites separable by recombination in two contiguous genes in a small region of the bacteriophage T4 genome. His experiments extended the results of earlier studies on *Drosophila*, which showed that the gene is divisible by both mutation and recombination. Benzer studied r II genes because of the unique properties of mutant phage with defects in these genes.

1. Phage T4 r II mutants produce large plaques with clear margins, easily distinguished from the smaller , fuzzy-edged plaques, produced by wild type T4 phage.

2. New r II mutants occur with a relatively high frequency – about in every $100{,}000^{r+}$ progeny.

3. Phage T4 r II mutants are unable to grow on a strain of *E. coli* called K12 (λ), whereas wild type T4 grows normally on K12 (λ).

4. Wild type T4 phage and T4 r II mutants are both able to grow on *E. coli* strain B cells.

Benzer utilized the system in the analysis of r II region by crossing two independently occurring r II mutants on *E. coli* B cells (Fig. 4.6), then collecting the progeny. If recombination occurred between the two mutants, a wild type chromosome would result, along with a double mutant (Fig. 4.6). A major advantage of r II system is, the recombination frequencies as low as 10^{-6} could have been detected in Benzer's experiments. The r II mutants are conditional lethal mutants i.e. wild type T4 phage can grow on both *E. coli* B and K12 (λ), r II mutants can grow only on *E. coli* strain B.

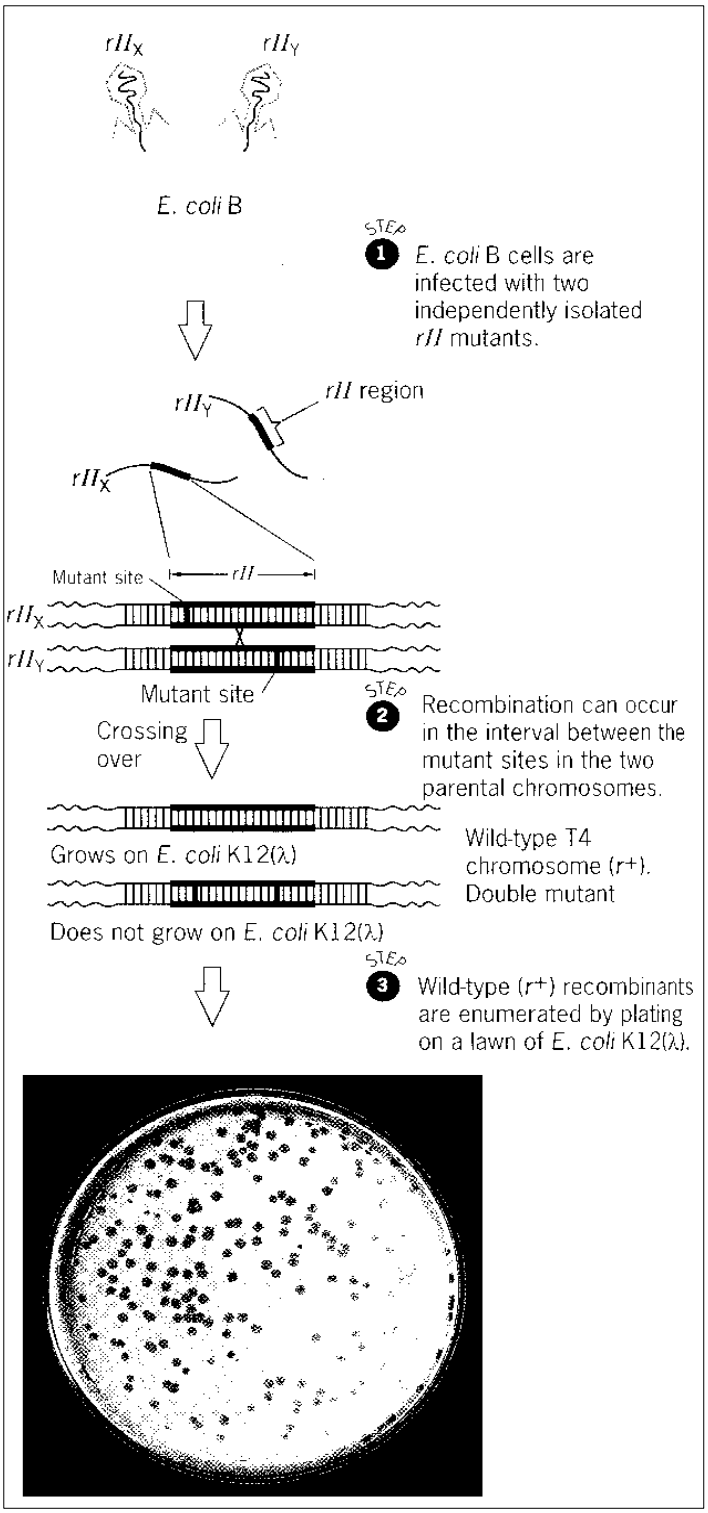


Fig. 4.6 A genetic cross in which **E. coli** B cells are infected with two independently isolated phage Tr r II mutants. The recombinant progeny are wild-type (*r*+) and double mutants.

If progeny phage produced from a mixed infection of *E. coli* B cells are plated on laws of K12 (λ) cells, only wild type T_4 phage can form plaques. If 20 plaques form on K12 (λ) lawn, 20 wild type recombinant phages are present. However, if recombination is a reciprocal process, one double mutant recombinant will be present for each wild type produced by recombination. The 20 wild types are produced by recombination. The 20 wild type plaques indicate that 40 recombinant phages are present. Thus recombination frequencies and map distances are calculated a follows:

$$\text{Recombination frequency} = \frac{2 \times \text{number of wild-type recombinants}}{\text{total number of progeny}}$$

and

$$\text{map distance (\% recombinants)} = \frac{2 \times \text{number of plaques on K12 } (\lambda)}{\text{number of plaques on B}} \times 100$$

To estimate the frequency of recombination, the total number of progeny phage produced per unit volume of mixed infection lysate must be determined. All the progeny phage produced in a cross between two r II mutants can grow on *E. coli* B cells. If 100 plaques are counted in a sample diluted 10^6 fold, the original sample contained 10^8 phage per unit volume (100×10^6). Thus, if 40 recombinant phage (wild type and double mutant) are present among 10^8 total progeny, the recombination frequency between the two mutant sites will be $40/10^8 = 0.0000004$

These calculations overlook only one fact, if an r II mutant may undergo spontaneous mutation back to wild type (r^+). However, such reversion rate of mutations are usually negligible and therefore, the above calculations are considered.

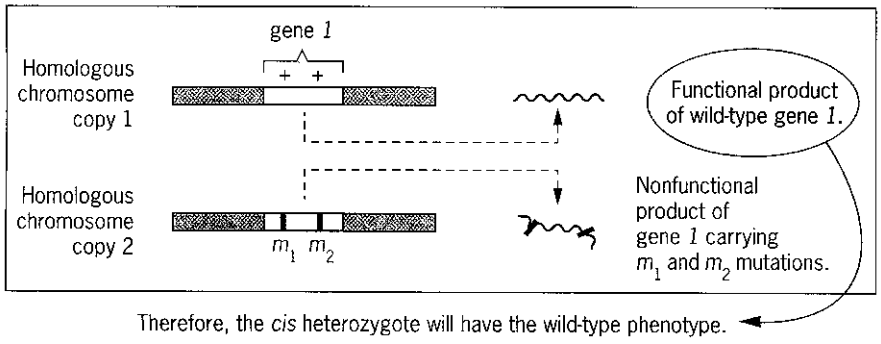
All of Benzer's r II mutants contained mutations in one of two contiguous genes, designated r II A and r II B. He determined this by performing complementation test.

3.5.1 Complementation test – Operational definition of the gene

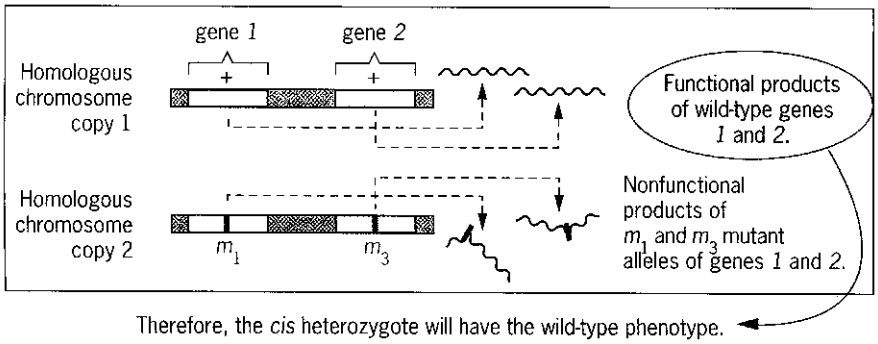
A double heterozygote, carries two mutations and their wild type alleles i.e., **m_1** and **m^+ plus m_2 and m_2^+.** They can exist in two arrangements. When two mutations are on the same chromosome, the arrangement is called the **coupling** or **cis configuration**; a heterozygote with this genotype is called **cis heterozygote**. When two mutations are on different chromosomes, the arrangement is called the repulsion or **trans configuration**. An organism with this genotype is a **trans-heterozygote**.

The complementation test allows geneticists to determine whether mutations that produce the same or similar phenotypes are in the same gene or in different genes, complementation test is referred as the **cis-trans test**. Each **cis heterozygote** which contains one wild type chromosome, should have wild type phenotype, whether the mutations are in the same gene or in two different genes (Fig. 4.7). If the ***trans* heterozygote** has the mutant phenotype then the two mutations are in

the same unit of function, the same gene (Fig. 4.8a). When a ***trans*** **heterozygote** has the wild type phenotype, the two mutations are said to exhibit complementation or to complement each other and are located in two different units of function, two different genes (Fig. 4.8b).



(a) *cis* heterozygote: mutations in one gene.

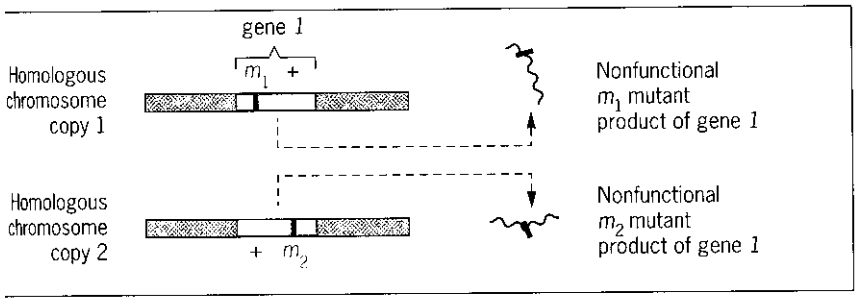


(b) *cis* heterozygote: mutations in two different genes.

Fig. 4.7 The *cis* test. The *cis* heterozygote should have the wild-type phenotype whether the mutations are in the same gene (a) or in two different genes (b).

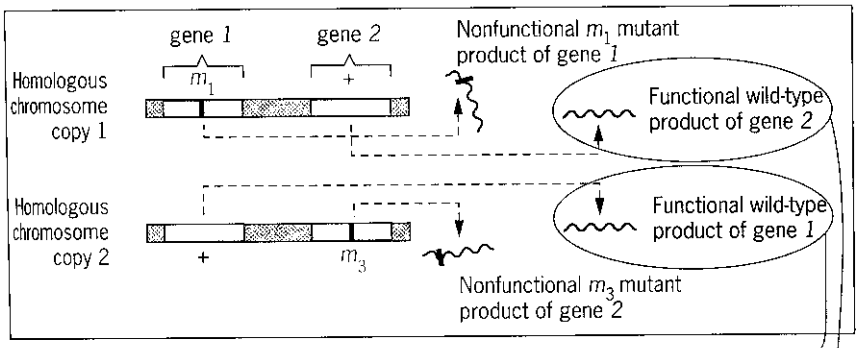
All of Benzer's r II mutants contained mutations in one of two contiguous genes, designated r II A and r II B. He determined this fact by performing complementation test.

Benzer infected *E. coli* K12 (λ) cells with a pair of r II mutants. If the two r II mutants contained mutations in the same gene, they failed to grow on *E. coli* K 12 (λ), if they were mutant in different genes (one was r II A and other was r II B) they grew on *E. coli* K12 (λ). Benzer characterized a number of r II mutants by the following procedure.



No functional gene *1* product is synthesized in the *trans* heterozygote; therefore, it will have a mutant phenotype.

(a) trans heterozygote: mutations in one gene.



Functional products of both genes are synthesized in the *trans* heterozygote; therefore, it will have the wild-type phenotype.

(b) trans heterozygote: mutations in two different genes.

Fig. 4.8 The *trans* test. The *trans* heterozygote should have (a) the mutant phenotype if the two mutations are in the same gene, and (b) the wild-type phenotype if the mutations are in two different genes.

4.5.2 Deletion mapping

After mapping about 60 mutations by two-factor crosses, Benzer devised a new, more efficient procedure, to map the rest of 2400 r II mutants. Benzer's short cut procedure is called **deletion mapping**. He demonstrated that some of his r II mutants contained deletions of all or part of r II region. A phage that carries a deletion cannot mutate back to wild, nor it recombine with another phage. Hence, a deletion mutant can't produce wild phenotype. Therefore, any r II mutant that produce wild type recombinants carry a mutation that maps outside the deleted segment, and any mutant that fails to generate wild-type recombinants, must carry a mutation that maps within the area defined by the deletion.

Benzer accomplished this by crossing different deletion mutants (Fig. 4.9). The appearance of wild-type recombinant progeny indicate that the deletions did not overlap using deletion mapping, Benzer estimated the extent of the deletions.

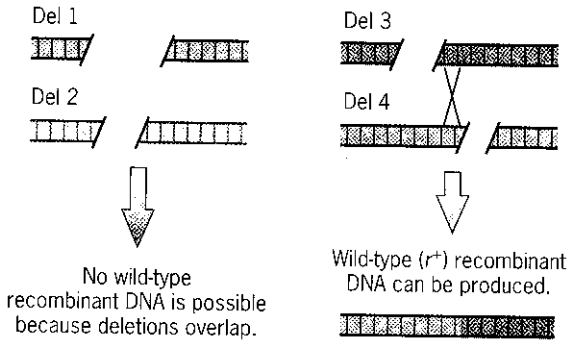


Fig. 4.9 Crosses between pairs of deletion mutants. If the deletions overlap, no wild-type recombinants are produced, if they do not overlap, *r*+ progeny are produced.

Benzer identified seven large deletions, that were missing overlapping segments of the r II locus containing r II A and r II B genes.

Benzer also isolated and characterized many smaller deletions that defined 47 short intervals within the r II region (Fig. 4.10). Benzer used this procedure to map 2400 r II mutants within r II A and r II B genes. Four important conclusions emerged from Benzer's analysis of the r II region.

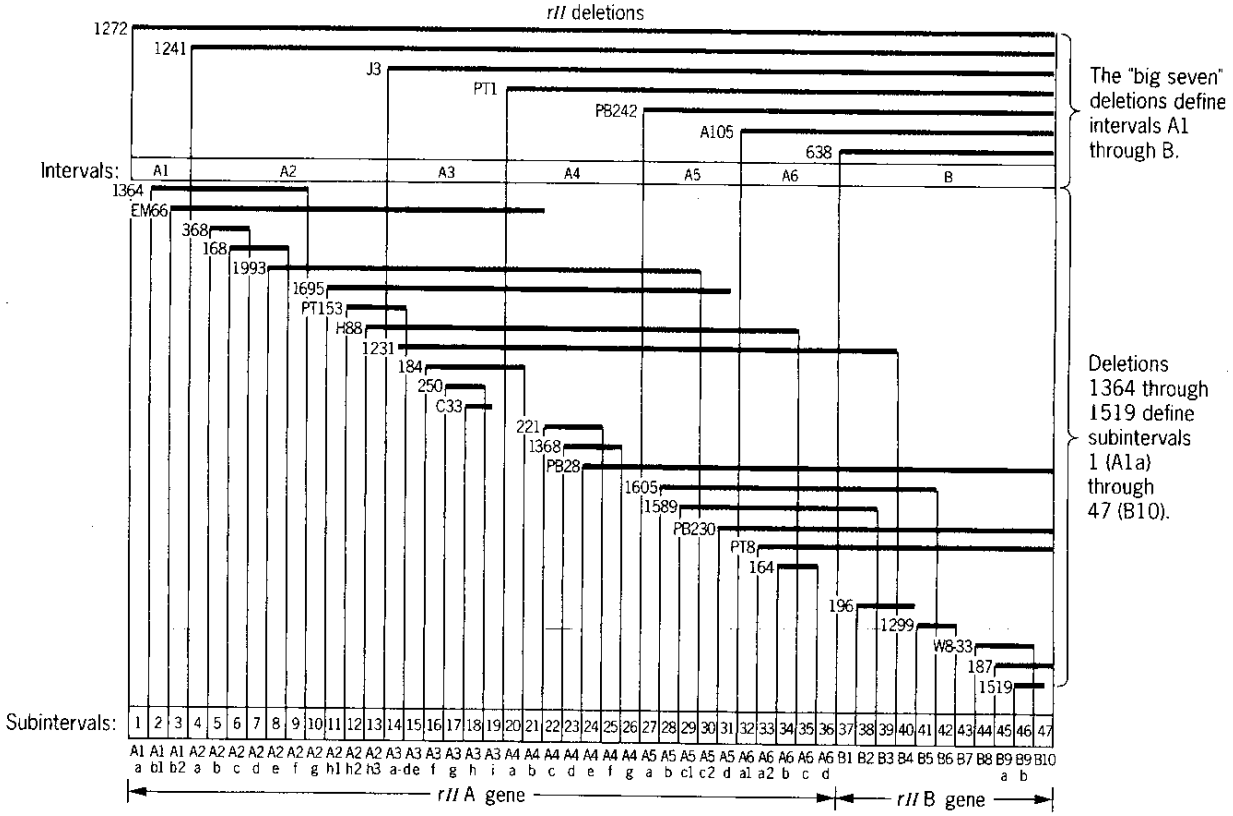


Fig. 4.10 The extent of the deletions carried by the deletion mutants used by Benzer to construct his fine structure map of the *r II* locus of bacteriophage T4.

(1) Since 2400 *r II* mutants mapped to 304 different sites are separable by recombination, the mutants were not randomly distributed over the 304 sites. Some r II sites are called mutation **hot spots** since 500 spontaneous mutations occurred at one hot spot in r II B gene (Fig. 4.11).

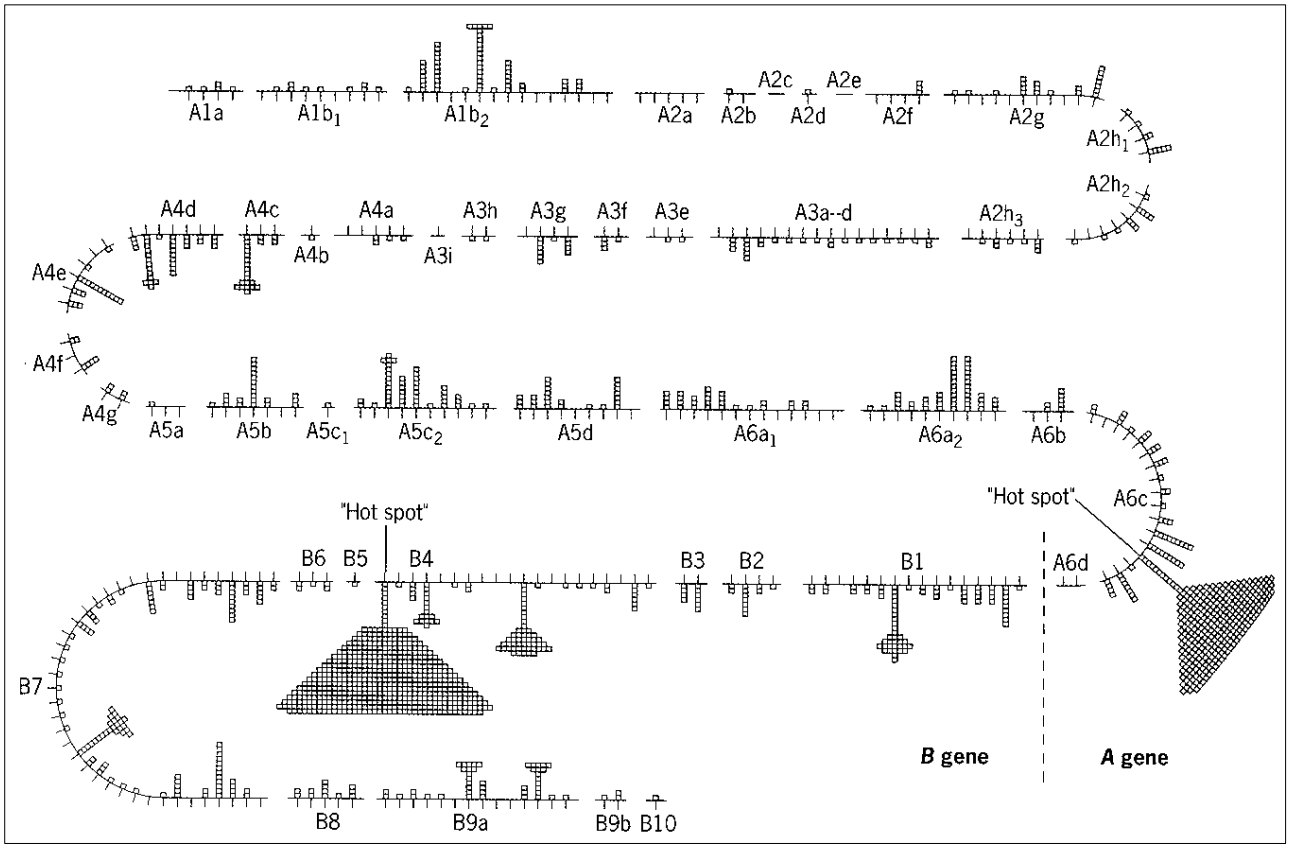


Fig. 4.11 The genetic map of the *r II* locus of phage T4. The *r II* locus contains two genes, *r II A* and *r II B*. Each square represents the independent occurrence of a spontaneous mutation at the indicated site. Vertical lines at which no squares occur represent positions defined by mutations induced by mutagenic agents. Some sites, called hot spots, mutate at very high frequencies.

(2) All 2400 r II mutants were mapped to either r II A gene or r II B gene or both genes.

(3) The smallest recombination frequency that Benzer observed was 0.02 per cent. The complete T_4 map is about 1500 map units length, this frequency represents 0.02/1500 = 0.00133 per cent of the map given that the T_4 genome is 173,000 base pairs in length, the smallest interval in which recombination can occur is (0.0000133) x 173,000 or 2.3 base pairs. Thus Benzer's results suggested that recombination can occur between any pair of adjacent nucleotides.

(4) Benzer's work extended the results of earlier experiments with *Drosophila* demonstrating that the gene is divisible by mutation and recombination.

The existence of a detailed map of *r II* locus of phage T_4 provided researchers, more opportunities to understand the fine structure of gene, insights about recombination mechanisms. Indeed Benzer's analysis of T_4 *r II* locus stands today as one of the great achievements in understanding modern gene concept.

4.6 GENETIC MAP OF T4 PHAGE

Bacteriophage T4 has been used for many important genetic studies. The results of early genetic mapping experiments indicated that the genes of phage. T4 might be distributed over as many as seven different linkage groups. However, physical studies showed that the T4 virion contained a single linear molecule of DNA. Thus, all of the genes of T4 had to be located on the same DNA molecule.

The results of genetic crosses with T4 mutants and their wild-type alleles revealed some intriguing inconsistencies. For example, the results of mapping experiments established the following linear gene order: *h*42-*ac*41-*r*67. But when these three genes were used in three-point crosses, the rarest recombinant classes—presumably the double crossovers—indicated that *h*42, not *ac*41, was located in the middle. These inconsistencies between the results of different crosses could not be resolved by repeating the experiments.

However, the dilemma could be resolved by making the T4 genetic map circular (Fig. 4.12). With a circular map, *h*42 maps to the left and to the right of both *ac*41 and *r*67, depending on the direction taken around the circle. Other crosses confirmed this circular map. But the next problem was that physical studies showed that the T4 DNA molecule is linear, not circular.

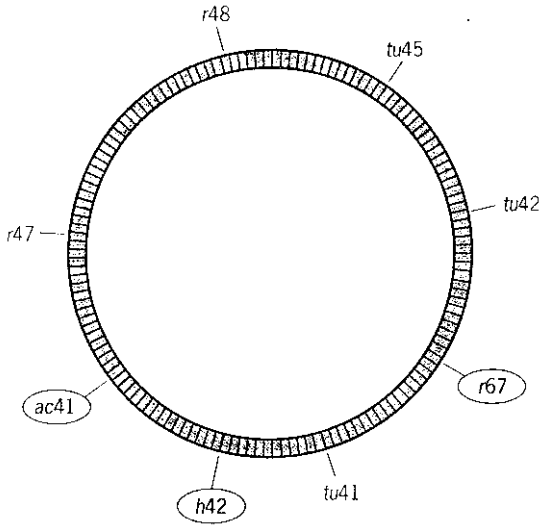
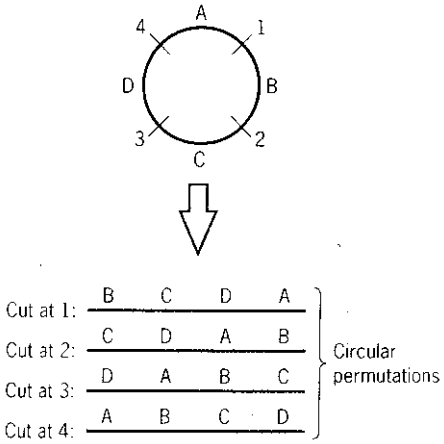


Fig. 4.12 The circular genetic map of bacteriophage T4

How can a circular genetic map be generated form a linear chromosome? In the mid-1960s, George Streisinger proposed that the T4 chromosome was both **terminally redundant** and **circularly permuted**. Initially, both geneticists and virologists were skeptical of this proposal. But both components of the hypothesis were soon proven correct.

Terminally redundant DNA molecules contain the same nucleotide sequence at both ends of a linear molecule, as in the sequence AAGGCCTTGACTA……TACGTAAGGCCTT. This DNA strand is terminally redundant: the sequence AAGGCCTT is present at both ends.

Circularly permuted sequences are obtained by breaking a circular structure containing a linear sequence (with the ends joined) of markers at random in order to obtain a collection of linear structures, as shown below:



The ends of a T4 chromosome contain the same genes in the same order; that is, they are terminally redundant (*abcdefg…..wxyzabc*). Note that the long, terminally redundant nucleotide sequences of the phage chromosome are denoted *abc* here for brevity. However, the endpoints for each chromosome are different; that is, one chromosome may have the sequence *abcdef….xyzabc*, whereas another will have the sequence *defghi…..xyzabcdef*, and a third will have the sequence *fghijk…..xyzabcdefgh*. Thus, a population of T4 chromosomes consists of a set of circularly permutated DNA molecules that are also terminally redundant. Sometimes a particular segment is within the terminally redundant ends of the molecule, and sometimes it is in the middle of the DNA molecule.

The combination of these two features—circular permutations and terminal redundancy—results in a circular genetic map from a population of linear chromosomes. But how can a population of circularly permitted, terminally redundant progeny phage chromosomes be generated in a cell infected with a single T4 virus particle? That is, how can a parental phage that is redundant for *abc* at its ends produce progeny that are redundant for *cde, def, efg,* and so forth? The answer is in the way that T4 DNA replicates and is packaged into phage heads. During replication of T4 DNA molecules, recombination occurs between the terminally redundant ends of

these molecules and generates long DNA molecules called **concatamers** (Fig. 4.13). Each concatamer contains many copies of the phage T4 genome joined end to end.

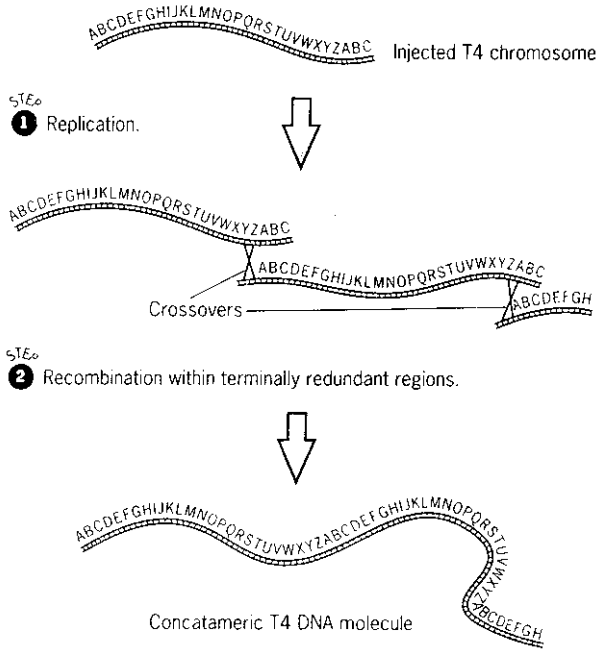


Fig. 4.13 The formation of concatameric replicative DNA molecules by "head-to-tail" recombination within the terminally redundant regions of T4 chromosomes.

During the formation of progeny viruses, the T4 head proteins condense around these concatameric DNA molecules until the head is full. The amount of DNA required to fill a head is **slightly more** than one complete set of T4 genes, usually symbolized *abc….xyz*. Because there is still room left in the head after one complete set of genes (*abc….xyz*) has entered, a few additional genes are added, yielding a phage head containing the gene sequence *abc….xyzabc*. The head is now full and the DNA is cut. This particular virus is redundant for genes *a*, *b*, and *c*. The net virus particle begins packaging DNA at *d*, gets to *c*, and then keeps going to *d*, *e*, and *f*. The next virus begins packaging at *g* and is redundant for *g*, *h* and *i*. This mode of DNA packaging, called the **headful mechanism**, produces T4 progeny chromosomes that are both terminally redundant and circularly permuted (Fig. 4.14). All the progeny viruses contain terminally redundant DNA molecules, and the molecules are redundant for different gens.

The circularly permuted chromosomes of phage T4 will yield a circular genetic map, even though each phage contains a linear molecule of DNA. Map distances are based on average recombination frequencies, which are population parameters. Although two closely linked genes may be located at opposite (terminally redundant) ends of one phage chromosome, they will be present in contiguous segments of DNA within the non-redundant regions of the majority of chromosomes in the T4 progeny population.

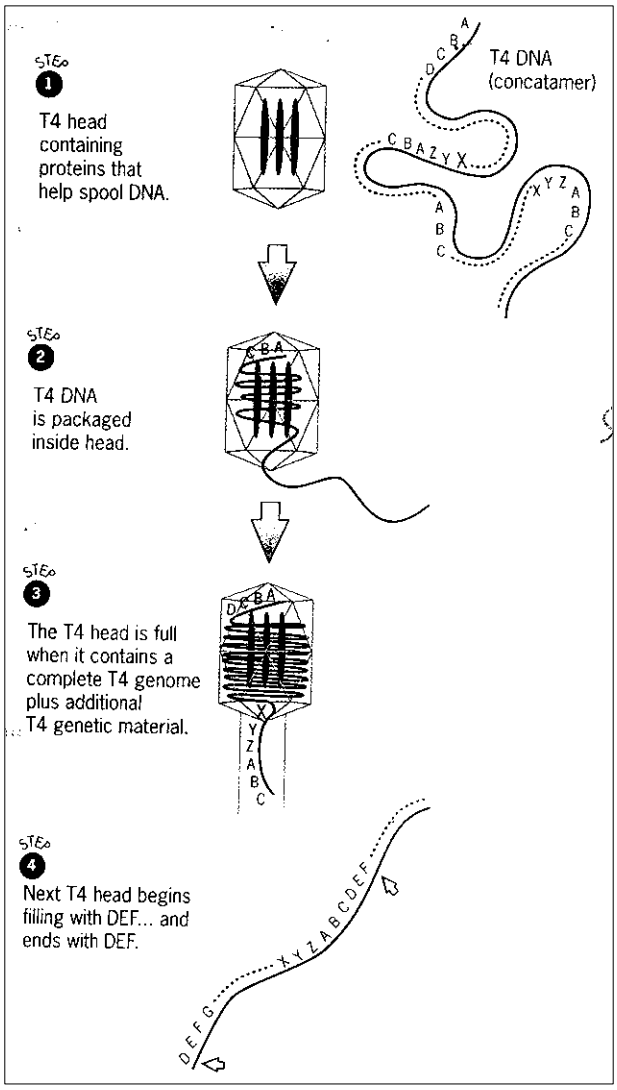


Fig. 4.14 Generation of circularly permuted, terminally redundant T4 chromosomes by the headful mechanism.

4.7 SUMMARY

Genetic recombination in phages is understood through Benzers analysis of *r II* mutants. The analysis of mutants by deletion mapping and the frequency of recombination between mutant sites led to the understanding of fine structure of the gene. Hence, gene as a divisible unit of function-**cistron**, unit of recombination-**recon** and unit of mutation-**muton** is described. Phage T4

has a circular genetic map but a linear chromosome. This apparent paradox is explained by the results of experiments showing that T4 chromosomes, which are terminally redundant and circularly permuted are described in this lesson.

4.8 MODEL QUESTIONS

1. Give a brief account of process of Genetic recombination in phages.
2. Discuss Benzer's work that have led to the understanding of fine structure of a gene.
3. Described the genetic map of T4 phage.
4. Write notes on:
   a) Deletion mapping
   b) Lysogenic cycle
   c) Cis-trans test
   d) Concatamer

4.9 REFERENCE BOOKS

1. P.J. Russell, 1987. **Essential Genetics**, Blackwell Publishers.
2. F.J. Ayala & J.A. Kiger, Jr. 1980. **Modern Genetics**. The Benjamin/Cummings Publ. Com.
3. M.W. Strickberger. **Genetics** – Prentice Hall of India, New Delhi.
4. Tamarin. **Principles of Genetics**.
5. Snustad, D.P., M.J. **Principles of Genetics.** Simmons, 2003. John Wiley & Sons.

*****

Prof. T.N. MARY

Lesson - 5

DNA REPLICATION

5.0 Objective

This lesson plan is aimed to understand the mechanism of replication of DNA and the enzymology of replication process.

5.1 Introduction

Genetic information is transferred from parent to progeny organisms by a faithful replication of the parental DNA molecules. In prokaryotes, the replication of DNA occurs during Log phase of growth and conjugation process. In eukaryotes, replication of DNA occurs during the 'S' sub-phase of interphase in both mitosis and meiosis. Replication of double-stranded DNA is a complicated process and involves various enzymes. The complexity is from the facts that – (1) a supply of energy is required to unwind the helix, (2) the single strands resulting from the unwinding tend to form intrastrand base pairs, (3) a single enzyme can catalyze only a limited

number of physical and chemical reactions and many reactions are needed in replication, (4) several safeguards have evolved that are designed both to prevent replication errors and to eliminate the rare errors that do occur and (5) both circularity and the enormous size of DNA molecules impose geometric constraints on the replicative system.

There are three hypothetically possible modes of DNA replication viz., (1) Dispersive, (2) Conservative and (3) Semi-conservative. But the most practical mode of replication that normally occurs in vivo is the semi-conservative replication. In dispersive replication, the old DNA molecule would break into several pieces, each fragment would replicate, and the old and new segments would recombine randomly to yield progeny of DNA molecules with a combination of old and new segments along its length. According to the conservative replication, the two newly synthesized strands would associate to form one double helix, while the two old strands would remain together as one double helix. In contrast, in the semi-conservative mode of DNA replication, each newly synthesized strand of DNA would remain associated with the old strand against which it was synthesized. Thus each progeny DNA molecule would consist of one 'old' and one 'newly synthesized' strand.

5.2 Semi-conservative replication

The semi-conservative mode of DNA replication was postulated by Watson and Crick along with the double-helix model of DNA. The main features of this mode of DNA replication are –

(1) a progressive separation of the two strands of a DNA molecule

(2) complementary base-pairing of the bases located in the single-stranded regions thus produced with the appropriate free deoxyribonucleotides

(3) formation of phosphodiester linkages between the neighbouring deoxyribonucleotides that have base-paired with the single-stranded regions, thereby producing the regions of new strand

(4) this ensures that the base sequences of the new strands are strictly complementary to those of the old strands

(5) as a result, each DNA molecule produced by replication has one 'old' and one 'new' strand

(6) the base sequence of a newly synthesized strand is dictated by the base sequence of the old strand, since the old strand serves as a template for the synthesis of the new strand

5.3 Evidence for semi-conservative replication

The evidence for semi-conservative replication of DNA was first presented by Meselson and Stahl in 1958. They grew *E. coli* on 15N (a heavy isotope of ^{14}N) for 14 cell generations so that the nitrogen present in DNA bases of these cells was ^{15}N. DNA having ^{15}N has a detectably

higher density than that having ^{14}N, therefore they are called heavy and light DNA, respectively. The heavy and light DNAs can be readily separated through equilibrium density gradient centrifugation as they form distinct bands in the centrifuge tube. In density gradient centrifugation, a heavy salt solution, $CsCl_2$ is centrifuged at 30,000 – 50,000 rpm for 48-72 hrs which leads to the formation of a linear gradient of increasing density from the top of centrifuge tube to the bottom of tube. When DNA is centrifuged in such a solution, it will move to a position where the density of salt solution is the same as that of DNA.

Meselson and Stahl transferred the *E. coli* cells grown on ^{15}N medium to a medium containing normal ^{14}N. they withdrew samples from these *E. coli* cells after approximately one, two and three cell generations. DNA from these cell samples was separated and subjected to density gradient centrifugation. After one cell generation, the DNA formed a single band intermediate between the heavy and light DNAs. The DNA obtained after two cell generations formed two bands of comparable intensity, one of the bands was intermediate and the other was light in density. The same two bands were recovered in the DNA isolated after three cell generations, although the intermediate band was relatively lower in intensity than the light band.

These findings can be readily explained on the basis of semi-conservative replication of DNA. The DNA from *E. coli* cells grown on ^{15}N had ^{15}N in both the strands; therefore it was heavier than the normal DNA having ^{14}N. When these *E. coli* cells were allowed one semi-conservative replication of their DNA on the ^{14}N medium, each of the resultant DNA molecules would have one heavy (^{15}N) and one light (^{14}N) strand. Therefore, these DNA molecules would have intermediate density. One more semi-conservative replication of these DNA molecules in the ^{14}N medium would generate two types of DNA molecules – (i) half of the molecules would have one heavy and one light strand (intermediate density) and (ii) the remaining half would have both light strands (light density). These molecules would obviously form one intermediate and one light band. On the third round of DNA replication on ^{14}N, the intermediate density DNA molecules would yield half intermediate and half light molecules, while all the molecules obtained from light DNA molecules would be light. This is the reason for the lower intensity of the intermediate band after three cell generations. All the expectations based on the semi-conservative mode of DNA replication are satisfied by these findings to the full extent.

5.4 Enzymes or proteins involved in DNA replication

DNA replication involves several proteins and enzymes. In *E.coli*, at least two dozen gene products are involved in DNA replication. Many of these gene products have been purified and their roles in DNA replication have been studies in vitro. Many of these proteins were first identified through the studies of mutants of *E. coli*, while some enzymes eg., ligase, DNA polymerase I etc., were first discovered biochemically.

5.4.1 DNA polymerase : DNA polymerase is the chief enzyme of DNA replication. The DNA polymerase activity was first demonstrated by Kornberg in 1956. It catalyzes the covalent addition of deoxyribonucleotides to the 3′-OH of a preexisting polynucleotide called as primer. Clearly this

enzyme cannot initiate the synthesis of a polynucleotide, it can only add nucleotides to a primer polynucleotide. All known polymerases have an absolute requirement of a free 3′-OH on the primer and , therefore, the direction of synthesis of the new chain is always 5′ to 3′. They also require a template DNA strand which specifies the base sequence of the new DNA chain. These enzymes utilize the 5′ triphosphates of the four deoxyribonucleotides viz., dATP, dGTP, dCTP and dTTP.

The complementary base pairing during DNA replication, purely on the basis of chemical and physical forces involved in the process, would give rise to error at a frequency of approximately 10^3 per base pair replicated. The actual rate of errors in bacteria seems to be about 10^{-8} - 10^{-10}. This phenomenal decline in the error rate may be due to the following two processes: (1) DNA ploymerase may scrutinize the bases entering into pairing with the bases of DNA strand being replicated and allow only the correct pairing, and (2) it does scrutinize the base pair after the new base has been added to the chain and deletes the wrong ones (the process is called 'proof reading'). DNA polymerase definitely proof-reads new chains. DNA polymerases isolated from prokaryotes and eukaryotes differ from each other in several respects.

Prokaryotic DNA Polymerases

Three different types of DNA polymerases, designated as DNA polymerase I, II, and III, have been isolated from prokaryotes.

DNA polymerase I or Kornberg enzyme was the first to be isolated from *E. coli* by Arthur Kornberg and his associates and was used for in vitro DNA synthesis in 1957. For quite some time, it was believed that this enzyme catalyses DNA replication in vivo. But it is now known that this enzyme is chiefly a DNA repair enzyme. DNA polymerase I possesses the activities like (1) 5′ → 3′ polymerase, (2) 3′ → 5′ exonuclease and (3) 5′ → 3′ exonuclease. The 5′ → 3′ and 3′ → 5′ exonuclease activities appear to be located in different sites of the enzyme. An exonuclease is an enzyme that degrades nucleic acids from one end and is unable to make internal cuts in a polynucleotide.

The 3′ → 5′ exonuclease activity enables the enzyme to remove deoxyribonucleotides, one by one, from the 3' end of a DNA strand. This activity is crucial for 'proof reading' or 'editing' of the polynucleotide chain being newly synthesized. Whenever the DNA chain being synthesized has a wrong base insertion , the 3′ → 5′ exonuclease activity of the enzyme removes the wrong base, and the 5′ → 3′ polymerase activity reinitiates the synthesis of the chain. This activity reduces the chances of errors in DNA replication. On the other hand, the 5′ → 3′ exonuclease activity functions in the removal of the DNA segments damaged by UV light, irradiation and other agents. This activity excises a small groups of nucleotides, up to 10. An endonuclease must cleave the DNA strand close to the site of damage before the 5′ → 3′ exonuclease action of the DNA polymerase I may take place. This activity of DNA polymerase is also responsible for the excision of RNA primers during DNA replication, and the ribonucleotides are immediately replaced by appropriate deoxyribonucleotides due to the 5′ → 3′ polymerase activity of the enzyme.

The DNA polymerase I enzyme is coded by the gene polA. It is a single chain polypeptide of 1,09,000 daltons can be cleaved by proteolytic treatments into a large and a small segments. The larger segment is of 68,000 daltons and is known as 'Klenow fragment' which exhibits the 5′ → 3′ polymerase and 3′ → 5′ exonuclease activities. The small fragment is of 65,000 daltons and possess the 5′ → 3′ exonuclease activity.

The other two DNA polymerases, DNA polymerase II and III, were discovered in pol A mutants of *E. coli* which are deficient in DNA polymerase I. The pol A mutants replicate their DNA at normal rates but are deficient in DNA repair activity. The DNA polymerase II has 5′ → 3′ polymerase activity and 3′ → 5′ exonuclease activity. The exact in vivo function of this enzyme is not known, although it most likely functions in DNA repair, particularly, in the absence of DNA polymerases I and III.

DNA polymerase III possesses 5′ → 3′ polymerase and 3′ → 5′ exonuclease activities and is responsible for DNA replication in vivo. This enzyme has a higher affinity for nucleotide triphosphates than DNA polymerases I and II, and catalyzes the synthesis of DNA chains at very high rates. This enzyme has several active subunits (Fig. 5.1) ; the α subunit of 1,30,000 daltons which possess polymerase activity, while another subunit, ε is of 25,000 daltons exhibits the proof reading ability. Generally, the α and ε subunits function in conjuction with each other. Three other subunits, called δ, γ and τ, confer onto the enzyme the ability to remain on a sing template.

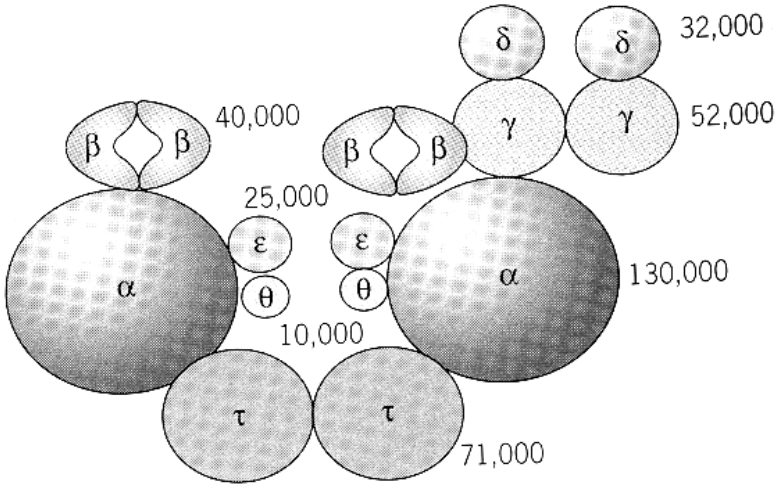


Fig. 5.1 Structure of Prokaryotic DNA Polymerase

The major properties of the three DNA polymerases are summarized in the table 5.1 below.

Characteristic	**DNA Pol I**	**DNA Pol II**	**DNA Pol III**
Molecular weight (Daltons)	1,09,000	1,20,000	>2,50,000
Constitution	Monomer	Not known	Heteromultimer
Molecules per cell	400	Not known	10-20
Nucleotides polymerized at 37^{o}C/min/molecule	Up to 1,000	Up to 50	Up to 15,000
Affinity for 5′ triphosphates of deoxyribonucleosides	Low	Low	High
Activities:			
5′ → 3′ polymerase	Present	Present	Present
3′ → 5′ exonuclease	Present	Present	Present
5′ → 3′ exonuclease	Present	Absent	Absent
Functions in	DNA repair, excision of RNA primers	DNA repair	DNA replication, it is the real replicase

Table 5.1 – Properties of DNA polymerases I, II and III of *E. coli*

Eukaryotic DNA polymerases

There are four different types of eukaryotic DNA polymerases called α, β, γ and δ. DNA polymerases α and β are confined to the nuclei of cells. DNA polymerase α most likely catalyzes chromosome replication in eukaryotes. DNA polymerase γ is found in mitochondria and chloroplasts, and is believed to be responsible for the replication of chromosomes of these organelles. DNA polymerase δ has been isolated from calf thymus and rabbit bone marrow, but the functions of this enzyme are not yet known. Of the four polymerases only DNA polymerase δ possesses the 3′ → 5′ exonuclease activity. And none of the eukaryotic polymerases has the 5′ → 3′ exonuclease activity required for the removal of RNA primers during DNA replication. This function is carried out by some other enzyme in eukaryotes.

5.4.2 Unwinding Proteins

The unwinding of DNA double helix to generate single-stranded regions of the DNA duplex for DNA replication begins at the origin site, and is the first event in replication initiation. In *E. coli*, it requires six proteins, viz., DnaA, DnaB, DnaC, HU, gyrase and SSB, for the formation of a prepriming complex in vitro. The OriC of *E. coli* has four 9 bp consensus repeats in its right side, while three 13 bp consensus repeats occur in its left side. The first step in the formation of prepriming complex is the binding of Dna A protein to the 9 bp repeats, one molecule to each repeat. More molecules of Dna A go on binding cooperatively till 20-40 molecules form a core and the OriC DNA folds around this core. The Dna A molecules act on the three 13 bp repeats in the left side of the OriC. As a result, each of these sites 'melts' to become opened up and get separated to yield single-stranded regions.

Six molecules of DnaB form a hexamer to which six monomers of DnaC become associated and this large aggregate binds to the single-stranded structure generated in the 13 bp repeats of OriC. DnaB provides the helicase activity which actually unwinds DNA. It actually replaces the DnaA from the 13 bp repeats and begins the unwinding of DNA. Gyrase provides the swivel that allows one strand to unwind around the other. Without this action unwinding would generate torsion in the DNA molecule. The SSB protein bind to the single-stranded segment of DNA so generated and stabilizes it and prevents duplex formation. The HU protein is a histone-like protein of *E. coli* is not absolutely necessary for replication initiation in vitro but it enhances the reaction. The initiation of replication generally requires about 60 bp to be unwound or single-stranded. ATP is required for activities of helicase, gyrase, primase and DNA polymerase III.

5.4.3 Primase

These enzymes catalyze the synthesis of RNA primers which are a prerequisite for the initiation of DNA replication in the vast majority of organisms. Primase is quite distinct from the RNA polymerase which is normally used for the transcription of DNA. The primase of *E. coli* and certain viruses are also called DnaG protein since it is produced by the Dna G gene.

5.4.4 Polynucleotide ligase

Ligase enzyme is an important enzyme both in DNA replication and in DNA repair. DNA ligase catalyzes the formation of phosphodiester linkage between the 5′-phosphoryl group of one nucleotide and the 3′-OH group of the immediate neighbour nucleotide at the side of a nick in a DNA strand. The property of formation of phosphodiester linkage by DNA ligase is similar to that of DNA polymerase. But the DNA ligase cannot fill in the gaps in a DNA strand, while DNA polymerases do this precisely. In addition ligase seals the nicks left behind by DNA polymerase I during DNA repair and among the Okazaki fragments generated during discontinous DNA replication. DNA ligases isolated from different organisms differ in their properties from each other.

5.5 Mechanism of replication

DNA replication is a complex process and involves various enzymes (Fig. 5.2). The replication of DNA is catalyzed by the enzyme DNA polymerase. Replication of DNA begins at certain unique and fixed point called 'origin site'. Two enzymes, DNA gyrase and DNA helicases, bind to the origin points and induce the unwinding of complementary strands of DNA double helix. Certain proteins, called single-stranded binding proteins, bind to the single-stranded regions thus produced to keep them in single-stranded condition. An enzyme called primase initiates transcription of the strand whose 3′-end is single-stranded (3′→5′ strand) which generates a 10-60 nucleotide-long primer RNA transcribed in 5′→3′ direction. The free 3′-OH of this primer RNA provides the initiation point for DNA polymerase for the sequential addition of deoxyribonucleotides. DNA polymerase has an absolute requirement for a free 3′-OH of a pre-existing polynucleotide for the initiation of DNA replication. DNA polymerase III progressively adds deoxyribonucleuotides to the free 3′-OH of this growing polynucleotide chain so that the replication of the 3'→5′ strand of the DNA molecule is continuous i.e., growth of the new strand in 5′→3′ direction.

The replication of the second strand (5′→3′ strand) of the DNA molecule is discontinuous. The replication of this 5′→3′ strand begins somewhat later than that of the 3′→5′ strand. Consequently, a segment of the 5′→3′ strand of a DNA molecule always replicates later than the homologous segment of the 3′→5′ strand. Therefore, the 3′→5′ strand of a DNA molecule is known as the 'leading strand', while the 5′→3′ strand is termed as the 'lagging strand'. When the replication of the 3′→5′ strand has progressed for some time, primase initiates the synthesis of RNA primer on the 5′→3′ strand close to the replication fork and away from the origin. The RNA synthesis progresses towards the origin in the 3′→5′ direction of the lagging strand. The 3′-OH of this primer RNA provides the initiation point for DNA polymerase to catalyze the replication of the 'lagging strand'. The replication of lagging strand proceeds in the direction which is opposite to that of the leading strand. However, on both leading and lagging strands, the new strand is synthesized from the 5′-end to the 3′-end. The replication of lagging strand generates small polynucleotide fragments called 'Okazaki fragments' named after the scientist R.Okazaki, who first identified them. The replication of 5′→3′ strand is discontinuous in that it has to be initiated several times, and every time one Okazaki fragment is produced. These Okazaki fragments are about 1000-2000 nucleotides long in *E. coli*, while they are only 100-200 nucleotides long in eukaryotes.

The RNA primer associated with the newly synthesized DNA strands / Okazaki fragments are most likely digested by the DNA polymerase I in prokaryotes. This enzyme also catalyzes the filling of the gaps so generated in the new strands through semiconservative replication of the old strands. The Okazaki fragments, after the gap-filling by DNA polymerase I, are joined together by the enzyme polynucleotide ligase which catalyzes the formation of phosphodiester bonds between the immediate neighbour nucleotides of the adjacent fragments. The replication fork generated at the origin may produce an 'eye' in a linear DNA molecule, while in a circular DNA it generates a 'θ-structure'. The eye and θ-structures shall be produced irrespective of whether the DNA replication is unidirectional or bi-directional.

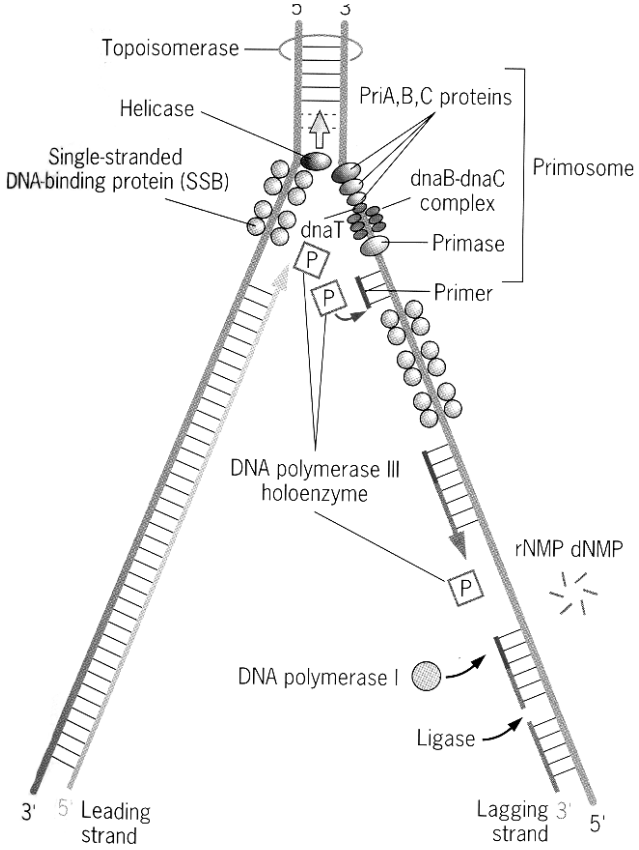


Fig. 5.2 Mechanism of DNA replication

5.6 Unidirectional and Bidirectional Replication

If the replication of the DNA proceeds only in one direction from the origin point, it is referred as unidirectional replication. In bi-directional replication, the process proceeds in both the directions. In unidirectional replication, one branch point remains at a fixed position with respect to the replication bubbles and this position defines the replication origin. In bi-directional replication both branch points move with respect to the bubbles, so each branch point is a replication fork. Bi-directional replication has been widely observed with phage, bacterial, and plasmid DNA. A small number of phages and plasmids use the unidirectional mode exclusively. Bi-directional replication from a fixed origin has also been demonstrated for several organisms with chromosomes that replicate as linear structures. Replication of the chromosome of phage T7, begins at a unique site near one end to form a so-called 'eye' structure and then proceeds bidirectionally until one fork reaches the nearest end. Replication of DNA molecules in the chromosomes of eukaryotes is also bi-directional. However, bidirection replication is not universal. The chromosome of coliphage P2, which like the lambda chromosome is circular during replication, replicates unidirectionally from a unique origin.

5.7 Rolling circle replication of DNA

The rolling circle model of DNA replication applies to the replication of several viral DNAs such as ΦX174 and λ and to the replication of the *E. coli* F factor during conjugation and

transfer of donor DNA to a recipient. The unique aspect of rolling-circle replication is that one parental circular DNA strand remains intact and rolls while serving as a template for the synthesis of a new complementary strand (Fig. 5.3). The first step is the generation of a specific cut or nick in one of the two strands at the origin of replication by sequence-specific endonuclease producing 3′ -OH and 5′ -phosphate termini. The 5′ end of the cut strand is then displaced from the circular molecule. This creates a replication fork structure and leaves a single stranded stretch of DNA that serves as a template for the addition of deoxyribonucleotides to the free 3′ end by DNA polymerase III using the intact circular DNA as a template. This new DNA synthesis occurs continuously as the 5′ cut end continues to be displaced from the circular molecule; thus, the intact circular DNA is acting as the leading strand template.

The 5′ end of the cut DNA strand is rolled out as a free "tongue" of increasing length as replication proceeds. This single-stranded DNA tongue becomes covered by single stranded binding proteins. New DNA synthesis on the displaced DNA occurs in the 5′ -to-3′ direction i.e., from the circle out toward the end of the displaced DNA. With further displacement, new DNA synthesis again must begin at the circle and move outward along the displaced DNA strand. Thus, synthesis on this strand occurs discontinuously as the displaced strand acts as lagging strand template. In this case, primase synthesizes short RNA primers that are extended as DNA in the form of Okazaki fragments by DNA polymerase III. The RNA primers are ultimately removed and adjacent Okazaki fragments are joined through the action of DNA ligase. The so formed linear molecule will become circularized to form into a circular molecule.

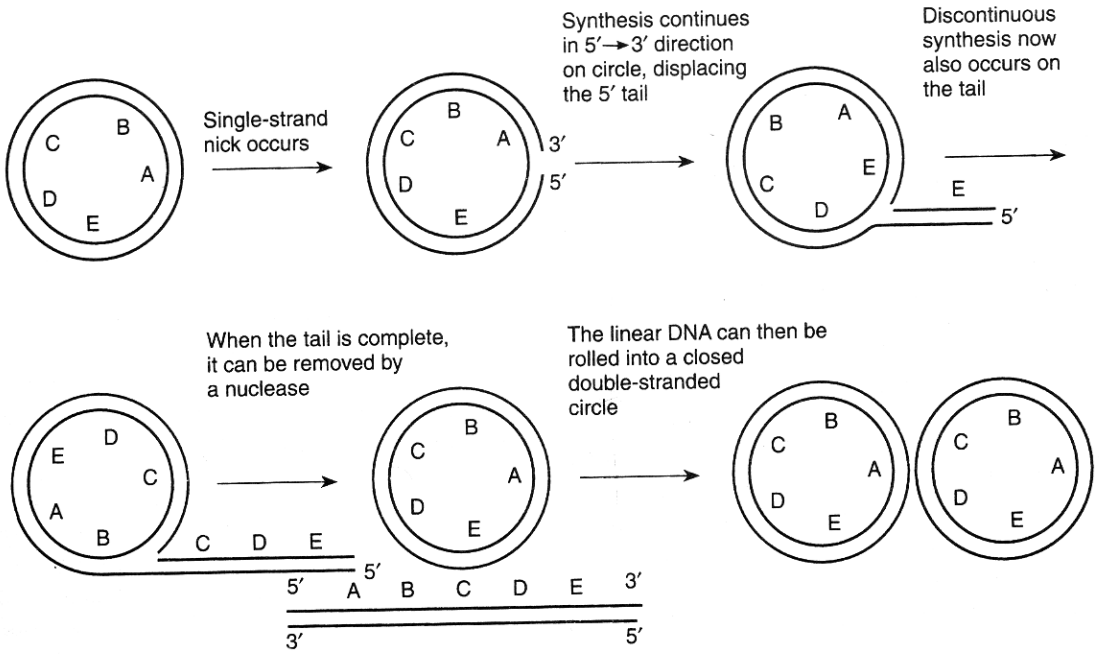


Fig. 5.3 Rolling circle mechanism of Replication

5.8 Denaturation and Renaturation of DNA

Denaturation: During the replication of DNA and copying of RNA from DNA, the two strands of the DNA molecule should be separated. And as suggested by Watson and Crick, the two strands of a DNA molecule are held together by weak, non-covalent bonds. When DNA is dissolved in saline solution and the solution is slowly warmed, a temperature is reached when strand separation begins. Within a few degrees, the process is generally complete and the solution contains single-stranded molecules that are completely separated from their original partners. The progress of thermal denaturation or DNA melting is usually monitored by following the increase in absorbance of the dissolved DNA. The nitrogenous bases of a nucleic acid molecule absorb ultraviolet radiation with an absorbance maximum near 260 nm. Once the two DNA strands have separated, the hydrophobic interactions that result from base stacking are greatly decreased, which changes the electronic nature of the bases and increases their absorbance of UV radiation. The temperature at which the shift in absorbance is half completed is termed the melting temperature (T_m). The higher the GC content of the DNA, the higher the T_m. This increased stability of GC-containing DNA reflects the presence of the extra hydrogen bond between the bases as compared with AT pairs.

Renaturation: When the bacterial DNA solution that is thermally denatured is slowly cooled, the DNA regained the properties of the double helix and it absorbed less UV light and behaved like genetic material. Similar results were obtained by heating the DNA to 100^oC to denature it, rapidly dropping the temperature of the solution to approximately 25^oC below the T_m, and allowing the DNA to incubate at this temperature for a period of time. This phenomenon of reassociation of complementary single-stranded DNA molecules is termed as 'Renaturation or Reannealing'. On one hand, reannealing has served as the basis for an investigation into the complexity of the genome and on the other hand it led to the development of a methodology called 'nucleic acid hybridization' in which complementary strands of nucleic acids from different sources can be mixed to form double-stranded molecules. This technique serves as an important tool for the assessment of the relatedness between two organisms.

5.9 Summary

DNA replication is semi-conservative, which indicates that one half of the parent duplex is transmitted to each of the daughter cells during cell division. The mechanism of replication was first suggested by Watson and Crick as their part of DNA model structure. Replication begins at a single origin on the circular bacterial chromosome and proceeds outward in both directions as a pair of replication forks. The initiation of strand separation at the origin site involves a number of proteins. DNA synthesis and editing or proof-reading are catalyzed by a family of DNA polymerases. To catalyze the polymerization reaction, the enzyme requires all four deoxyribonucleoside triphosphates, a template strand to copy, and a primer containing a free 3′OH to which nucleotides can be added.

The synthesis of new strand always carried out in 5′→3′ direction. One of the newly synthesized strands (the leading strand) grows toward the replication fork and is synthesized

continuously. The other newly synthesized strand (the lagging strand) grows away from the fork and is synthesized discontinuously. Events at the replication fork require a variety of different types of proteins having specialized functions. These proteins include DNA gyrase, DNA helicase, SSBs, Primase etc. The DNA polymerase is the prime replicase responsible for replication and the DNA polymerase is for editing or proof-reading function and for the removal of RNA primers after the DNA strand synthesis. And the DNA ligase seals the fragments of the lagging strand into a continuous polynucleotide.

5.10 Model Questions

1. Describe the mechanism of DNA replication.
2. Write an essay on the enzymology of DNA replication.
3. Short questions
   a. DNA polymerase III
   b. Continuous and discontinous replication
   c. Rolling circle mechanism
   d. Denaturation and Renaturation of DNA

5.11 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 6

DNA DAMAGE AND REPAIR

6.0 Objective

To make the learner to understand about the types of DNA damages and the mechanisms for the repair of the damage.

6.1 Introduction

Spontaneous and induced mutations constitute damage to the DNA of a cell or an organism. Especially with high doses of mutagens, the mutational damage can be considerable. Both prokaryotic and eukaryotic cells have a number of repair systems to deal with damage to DNA. All of the systems use enzymes to make the correction. Some of the systems directly correct the lesion while others first excise the lesion creating a single-stranded gap and then synthesize new DNA for the resulting gap. If the repair systems are unable to correct all of the lesions, the result is a mutant cell or, if too many mutations remain, death of the cell. Clearly, DNA repair systems are very important for the survival of the cell. The fact that the repair systems are not 100

percent efficient makes it possible to isolate mutants for study. At high doses of mutagens, repair systems are unable to correct all of the damage, and cell death can result.

6.2 Types of DNA damages

6.2.1 Alkylation

The transfer of an alkyl group, either a methyl or ethyl group, to reactive sites on the bases or phosphates is referred to as alkylation. This reaction is catalyzed by some alkylating agents namely nitrogen mustard, sulphur mustard, methylmethane sulphonate, ethylethane sulphonate etc. This addition results in the base modification and thereby base pair potential. A particular reactive site is the oxygen of carbon 6 in the guanine. The alkylated guanine pairs with thymine and produces a GC-to-AT transition when the DNA is replicated. Alkylation damage like this can be removed by specific DNA repair enzymes. In this repair system the modified base is not removed from the DNA. In this case, an enzyme encoded by the ada gene, called O^6-methylguanine methyltransferase, recognizes the O^6-methylguanine in the DNA and removes the methyl group, thereby changing it back to its original form.

6.2.2. Deamination

Deamination is the removal of an amino group from a base. This is one of the most common chemical event that occurs to produce a spontaneous mutation. This deamination of cytosine to uracil (Fig. 6.1) can also be catalyzed by Nitrous acid (HNO_2), which is a very potent mutagen. For example, the deamination of cytosine produces uracil. Uracil is not a normal base in DNA, although it is a normal base in RNA. The deamination is a disastrous change because the deamination product, uracil, pairs with adenine rather than with guanine. This has two effects: (1) an incorrect base will appear in mRNA and (2) an adenine instead of a guanine will occur in newly replicated DNA strands. An intracellular repair system exists to remove most of the uracils produced by deamination of cytosine, thereby minimizing the mutational consequences of this event. However, if the uracil is not repaired, an adenine will be incorporated in the new DNA strand opposite it during replication. Ultimately this will result in the conversion of a CG base pair to a TA base pair i.e., a transition mutation.

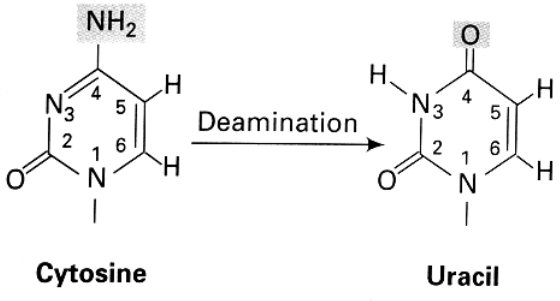


Fig. 6.1 Deamination of Cytosine

In this case, an uracil-dependent DNA glycosylase enzyme detect an individual unnatural base and catalyze its removal from the deoxyribose sugar to which it is attached. This catalytic activity leaves a gap in the DNA where the base was removed. This hole is called an AP site (apurinic site or apyrimidinic site). The enzyme AP endonuclease recognizes the hole and cuts the DNA backbone on the 3′ side beside the missing base. This leaves a primer end from which DNA polymerase I initiates repair synthesis and subsequently the gap is filled followed by sealing of nick by DNA ligase.

6.2.3 Pyrimidine Dimers

Ultraviolet light (UV) rays are non-ionizing and they have insufficient energy to induce ionizations. However, UV is a useful mutagen and at high enough doses it can kill cells. In some applications, UV light is used as a sterilizing agent. Ultraviolet light causes mutations because the purine and pyrimidines in DNA absorb light very strongly in the UV range (254 to 260 nm). At this wavelength UV light induces gene mutations primarily by causing photochemical changes in the DNA. One of the effects of UV radiation is the formation of abnormal chemical bonds between adjacent pyrimidine molecules in the same strand. This bonding is induced mostly between adjacent thymines, forming the thymine dimers (Fig. 6.2), usually designated as T ^ T.

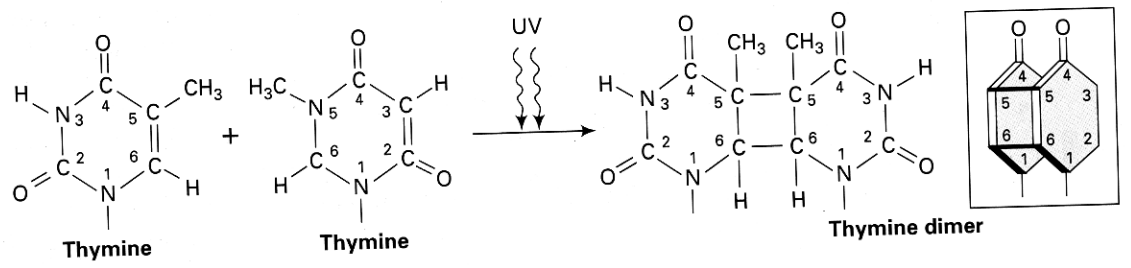


Fig. 6.2 Formation of thymine dimer

Formation of a thymine dimer involves the disruption of double bonds between 4 and 5 carbon residues of the two thymine molecules as well as of the hydrogen bonds between these thymines and the corresponding adenine residues of the complementary DNA strand, and the formation of covalent bonds between the two 4 and the two 5 carbons of the two adjacent thymines. This unusual pairing produces a bulge in the DNA strand and disrupts the normal pairing of the Ts with the corresponding As on the opposite strand. So, the significant effects of the presence of thymine dimers are (1) the DNA helix becomes distorted as the thymines, which are in the same strand, are pulled toward one another and (2) as a result of distortion, hydrogen bonding to adenines in the opposing strand, though possible is significantly weakenend. This structural distortion causes inhibition of advance of the replication fork as the DNA polymerase is unable to catalyze replication of a DNA molecule in the region distorted by a pyrimidine dimer formation. This UV damage of DNA can be repaired by different mechanisms viz., photoreactivation, excision repair and post-replication recombination repair.

6.3 DNA damage Repair Mechanisms

6.3.1 Light-Dependent Repair

This light-dependent repair mechanism for DNA repair is also be called as 'Light repair' or 'Photoreactivation'(Fig. 6.3). This involves in the direct correction and repair of UV-light induced thymine dimmers. It is an enzymatic cleavage of thymine dimers. The dimers are reverted directly to the original form by exposure to visible light in the wavelength range of 320 to 370 nm. In *E. coli*, it is catalyzed by a light-activated enzyme called 'DNA photolyase' or 'Photolyase' or 'PR enzyme' which is encoded by the phr gene. The photolyase enzyme searches for and binds to the thymine dimers on the DNA, and uses light energy to cleave the covalent cross-links between the thymine residues of the dimer. Photolyase will bind to thymine dimers in DNA in the dark, but it cannot catalyze cleavage of the bonds joining the thymine moieties without energy derived from visible light, specifically light within the blue region of the spectrum. Photolyase also splits cytosine dimers and cytosine-thymine dimers. Thus, when ultraviolet light is used to induce mutations in bacteria, the irradiated cells are grown in the dark for a few generations to maximize the mutation frequency.

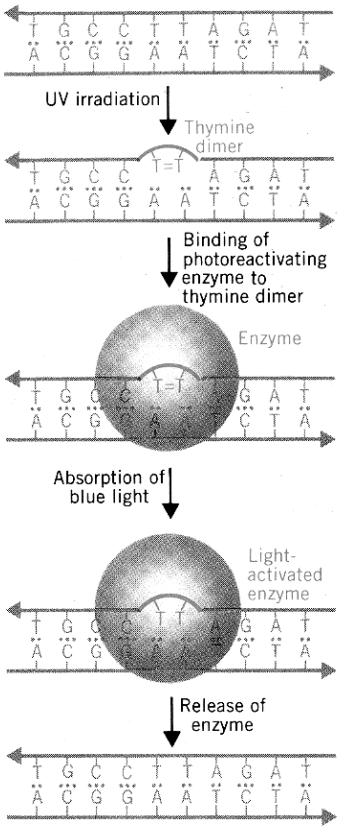


Fig. 6.3 Photoreactivation mechanism

6.3.2 Mismatch Repair

The DNA polymerases proofread the DNA strands during their synthesis, removing any mismatched nucleotides at the 3′ termini of growing strands. The mismatch repair pathway provides a backup to this replicative proofreading by correcting mismatched nucleotides remaining in DNA after replication. Many mismatched base pairs left after DNA replication may be corrected by this mechanism. Mismatches often involve the normal four bases in DNA. For example a T may be mispaired with a G. because both T and G are normal components of DNA, mismatch repair systems need some way to determine whether the T or the G is the correct base at a given site. The repair system makes this distinction by identifying the template strand, which contains the original nucleotide sequence, and the newly synthesized strand, which contains the misincorporated base. This distinction can be made based on the pattern of methylation in newly replicated DNA. In *E. coli*, the A in palindromic GATC is usually methylated by the action of dam-methylase, an enzyme encoded by the dam gene. Thus, both A nucleotides in the DNA segment are methylated. However, after replication, the parental DNA strand has a methylated A nucleotide in the GATC sequence, while the A nucleotide in the GATC of the newly replicated DNA strand is not methylated until a short time after its synthesis. Therefore, for a short while after replication, the parental strand has methylated GATC sequence, while the new strand has an unmethylated GATC sequence, called hemimethylation. The mismatch repair system uses this difference in methylation state to excise the mismatched nucleotide in the nascent strand and replace it with the correct nucleotide by using the methylated parental strand of DNA as template (Fig. 6.4).

In *E. coli*, the products of three genes namely mutS, mutL, and mutH are involved in the initial stages of mismatch repair. The first step in the repair process is the binding of the mutS-encoded protein, MutS, to the mismatch, which may be a single base pair mismatch. MutH and MutL proteins then join the complex of MutS and DNA. This results in the DNA looping which brings the unmethylated GATC sequence located about 1 to 2 kb away from the replication site close the mismatch. MutH contains a GATC-specific endonuclease activity that cleaves the unmethylated strand at hemimethylated GATC sites either 5′ or 3′ to the mismatch. The incision sites may be 1000 nucleotide pairs or more from the mismatch. The subsequent excision process requires MutS, MutL, DNA helicase II (Mut U or Uvr D), and an appropriate exonulcease. If the incision occurs at a GATC sequence 5′ to the mismatch, a 5′ → 3′ exonuclease like *E. coli* exonuclease VII is required. If the incision occurs 3′ to the mismatch, a 3′ → 5′ nuclease activity like that of *E. coli* exonuclease I is needed. After the excision process has removed the mismatched nucleotide from the unmethylated strand, DNA polymerase fills in the gap, and DNA ligase seals the nick. The homologs of the *E. coli* MutS and MutL proteins have been identified in both *Saccharomyces cerevisiae* and humans which indicates the presence of similar mismatch repair pathways occur in eukaryotes. In fact, mismatch excision has been demonstrated in vitro with nuclear extracts prepared from human cells. Thus mismatch repair is probably a universal or nearly universal mechanism for safeguarding the integrity of genetic information stored in double stranded DNA.

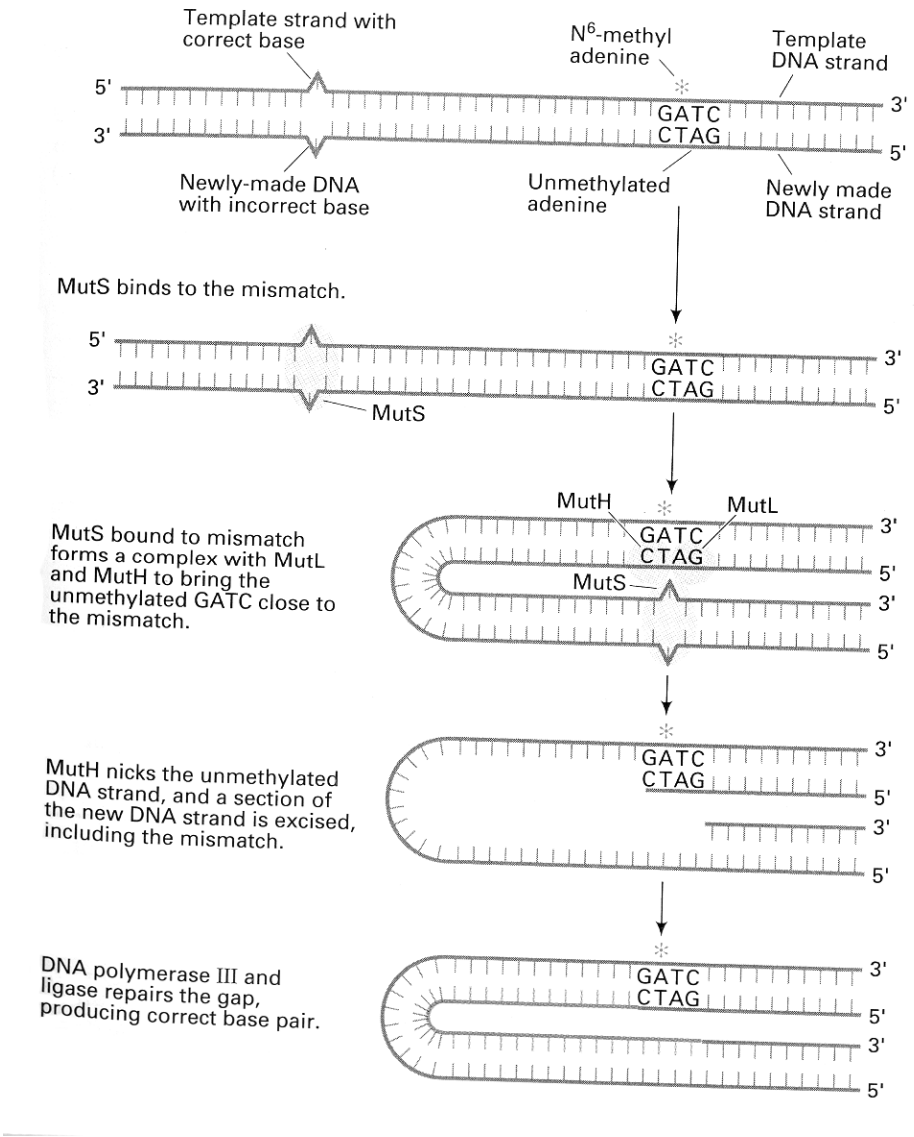


Fig. 6.4 Mismatch repair mechanism

6.3.3 Nucleotide Excision Repair

The nucleotide excision repair pathways remove larger defects like thymine dimmers and bases with bulky side-groups from DNA. This pathway is operative in the dark, and occur by very similar mechanisms in *E. coli* and humans. In nucleotide excision repair, a unique excision nuclease activity produces cuts on either side of the damaged nucleotides and excises an oligonucleotide containing the damaged bases. This nuclease is called as 'excinuclease' to distinguish it from the endonucleases and exonucleases that play other roles in DNA metabolism.

In *E. coli*, excinuclease activity requires the products of three genes viz., uvrA, uvrB, and uvrC. A trimeric protein containing two UvrA polypeptides and one UvrB polypeptide recognizes

the defect in DNA, binds to it, and uses energy from ATP to bend the DNA at the damaged site. The UvrA dimer is then released, and UvrC protein binds to the UvrB-DNA complex. The UvrB protein cleaves the fifth phosphodiester bond from the damaged nucleotides on the 3′ side, and the UvrC protein hydrolyzes the eight phosphodiester linkage from the damage on the 5′ side. The uvrD gene product, DNA helicase II, releases the excised oligomer. In the last steps of the pathway, DNA polymerase I replaces UvrB protein and fill in the gap using the complementary strand as template followed by the sealing of the nick left by polymerase by DNA ligase.

Nucleotide excision repair mechanism in humans occurs through a pathway similar to the one in *E. coli*, but it involves about four times as many proteins. In humans, the excinuclease activity requires at least 17 polypeptides. Protein XPA (for xeroderma pigmentosum protein A) recognizes and binds to the damaged nucleotides in DNA. It then recruits the other proteins required for excinuclease activity. In humans, the excised oligomer is 29 nucleotides long rather than the 12-mer removed in E. coli. The gap is filled by either DNA polymerase δ or ε in humans, and DNA ligase completes the job.

6.3.4. Post-replication Recombination repair

In *E. coli*, light-dependent repair, excision repair, and mismatch repair can be eliminated by mutations in the phr gene, uvr gene and mut gene, respectively. In multiple mutants deficient in these repair mechanisms, still another DNA repair system called Post-replication repair is operative. When a DNA ploymerase III encounters a thymine dimer in a template strand, its progress is blocked. DNA polymerase restarts DNA synthesis at some position past the dimer, leaving a gap in the nascent strand opposite the dimer in the template strand. At this point, the original nucleotide sequence has been lost from both strands of this progeny double helix. The damaged DNA molecule is repaired by a recombination-dependent repair process mediated by the *E. coli* recA gene product. The RecA protein, which is required for homologous recombination, stimulates the exchange of single strands between homologous double helices. During post-replication repair, the RecA protein binds to the single strand of DNA at the gap and mediates pairing with the homologous segment of the sister double helix. The gap opposite the dimer is filled with the homologous DNA strand from the sister DNA molecule. The resulting gap in the sister double helix is filled in by DNA polymerase, and the nick is sealed by DNA ligase. The thymine dimer remains in the template strand of the original progeny DNA molecule, but the complementary strand is now intact. If the thymine dimer is not removed by the nucleotide excision repair system, this post-replication repair must be repeated after each round of DNA replication. The mechanism of this repair is given in figure 6.5 below.

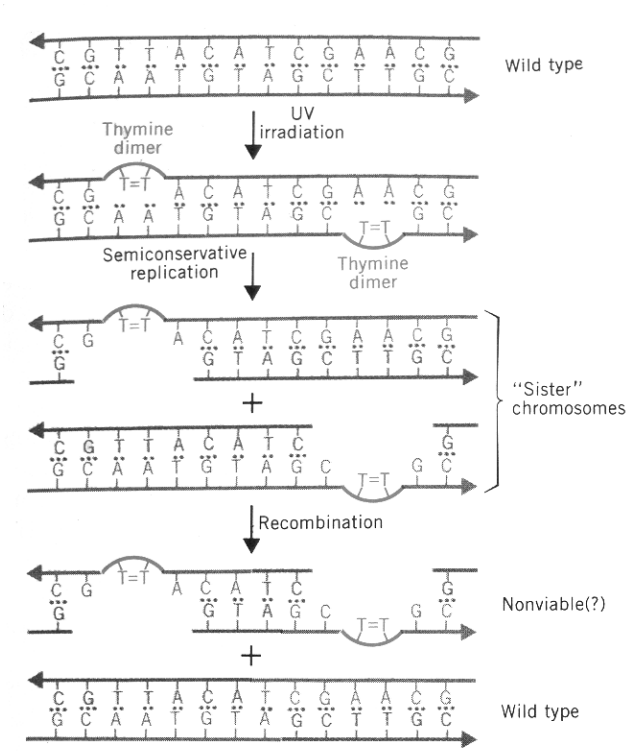


Fig. 6.5 Post-replication Recombination repair

6.3.5 SOS Repair

When the DNA of *E. coli* cells is heavily damaged by mutagenic agents such as UV light, the cells take some drastic steps in their attempt to survive. They go through a so-called SOS response, during which a whole battery of DNA repair, recombination, and replication proteins are synthesized. SOS repair includes a bypass system that allows DNA chain growth across damaged segments at the cost of fidelity of replication. It is an error-prone process; even though intact DNA strands are formed, the strands contain incorrect bases. The principle involved is that survival with mutations is better than no survival at all. The SOS response appears to be a somewhat desperate and risky attempt to escape the lethal effects of heavily damaged DNA. When the error-prone repair system is operative, mutation rates increase sharply. SOS repair is thought to invoke a relaxation of the editing system in order to allow polymerization to proceed across a dimer despite the distortion of the helix.

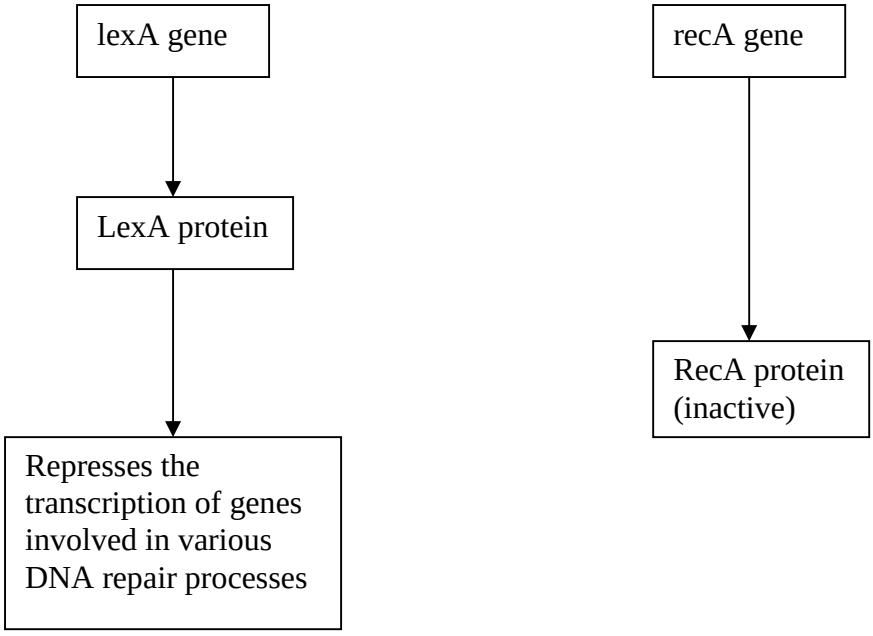
The SOS response has been studied best in *E. coli*, where the synthesis of many of the enzymes involved in the repair of DNA damage is regulated by the SOS system. Two genes are key to controlling the SOS system: lex A and recA. The *E. coli* cells with mutant recA and lexA genes have their SOS response permanently turned on. In the uninduced state, when there is no DNA damage, the lexA encoded protein, LexA functions as a repressor to prevent transcription of about 17 genes whose protein products are involved in the repair of various kinds of DNA damages, including excision repair of gaps. All of the genes involved have in common a 20-nucleotide regulatory sequence called the SOS box. The recA encoded protein, RecA, is a

regulator of the induction of the SOS response. When there is sufficient damage of DNA, somehow the RecA protein becomes activated, perhaps by binding to single-stranded DNA. The activated RecA stimulates the LexA protein to cleave itself and this relieves the repression of the DNA repair genes. As a result, the DNA repair genes are transcribed, and DNA repair proceeds. After the DNA damage is dealt with, RecA again becomes inactivated, and newly synthesized LexA protein again acts to repress the DNA repair genes.

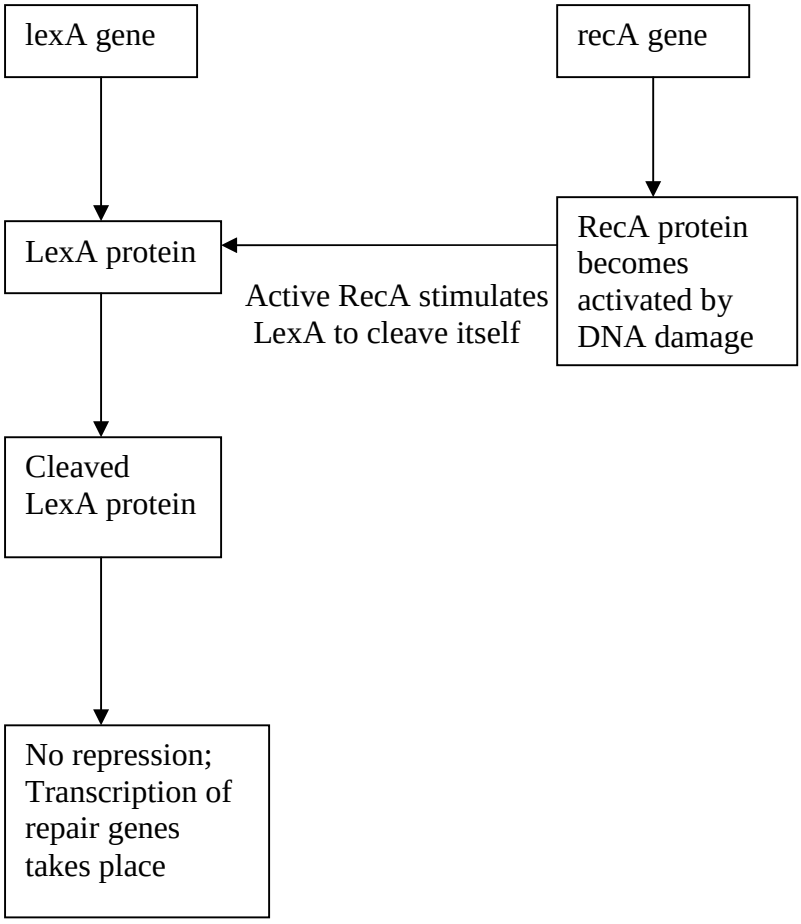
The role of the RecA product is twofold. One of these is its regulatory function and the other is a direct effect on editing. The RecA protein binds tightly to single-stranded DNA but only very weakly to double-stranded DNA. The distortion resulting from a pyrimidine dimer produces a short stable single stranded region to which RecA binds. When DNA polymerase III encounters a dimer site to which RecA is bound, RecA interacts with the ε subunit of the polymerase, which is responsible for editing, and inhibits the editing function. As a result, the replication fork advances. The presence of RecA at the dimer site inhibits editing and causes the mispaired base to remain in the daughter strand as a mutation.

SOS repair also requires two other genes, umuC and umuD. The role of the gene products is not clearly known but three hypotheses have been suggested: (1) they facilitate tight binding of RecA at the small distortion, (2) they facilitate binding of pol III to be distorted region, or (3) they enable the pol III to leave the damaged site either by releasing pol III from RecA or RecA from the DNA.

Un-induced State



Induced State



6.4 Summary

Genetic damage can occur to the DNA spontaneously, through replication errors, or through treatment with radiation or chemical mutagens. If the genetic damage is nor repaired mutations will result, and if there has been too much damage, cell death may result. Damage through alkylation, deamination and UV irradiation are quite common. Cells possess a number of repair mechanisms that function to correct at least some damage to DNA. These repair mechanisms include: (1) repair by DNA polymerase proofreading, in which a base pair mismatch in DNA being synthesized is immediately repaired by $3' \rightarrow 5'$ excision; (2) photoreactivation of pyrimidine dimers induced by UV light; (3) excision repair, in which pyrimidine dimers and other DNA damage that distorts the DNA helix are excised and replaced with new DNA; (4) repair of damaged bases by glycolysaes and AP endonuclease; and (5) repair by mismatch correction in which the methylation state of a DNA sequence signals which DNA strand is newly synthesized so that the mismatched base on that strand is corrected. Any DNA damage that is not repaired may result in a mutation and may have the potential to be lethal to the cell. The collective array of

repair enzymes, then, serves to reduce mutation rates for spontaneous errors by several orders of magnitude.

6.5 Model Questions

1. What are the thymine dimers. Explain the mechanisms for the removal of thymine dimers.
2. Describe the various mechanisms of DNA damage repair systems.
3. Short questions
   a. Pyrimidine dimers
   b. SOS response
   c. Nucleotide excision repair
   d. Photoreactivation

6.6 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. Uma Maheswara Rao

Lesson No. 7

MUTATIONS

7.0 Objective

The main purpose of this lesson plan is to make reader to know about the various types of mutations that occur in organisms and types of chemical mutagens.

7.1 Introduction

Mutation is the process by which a DNA base pair change or a chromosome change is produced. Thus, a mutation may be the result of any detectable change that affects DNA's chemical or physical constitution, its replication, its phenotypic function, or the sequence of one or more DNA base pairs. A mutation can be transmitted to daughter cells and even to succeeding generations, thereby giving rise to mutant cells or mutant individuals. If a mutant cell gives rise only to somatic cells in multicellular organisms, a mutant spot or area is produced, but the mutant characteristic is to passed on to the succeeding generation. This type of mutation is called a somatic mutation. However, mutations in the germ-line of sexually reproducing organisms may be transmitted by the gametes to the next generation, producing an individual with the mutation in both its somatic and germ line cells. Such mutations are called germ-line mutations. A somatic mutation affects the individual in which it happens, while a germ-line mutation affects individuals of the subsequent generations.

Mutations can occur spontaneously, but they can also be induced experimentally by the application of a mutagen, any physical or chemical agent that significantly increases the frequency of mutational events above the spontaneous mutation rate. Mutations that result from treatment with mutagens are called induced mutations and naturally occurring mutations are spontaneous mutations. There are no qualitative differences between spontaneous and induced mutations. The manifestation of a mutant phenotype is typically the result of a change in DNA that results in the altered function or production of a protein.

7.2 Types of Mutations

Mutations can be categorized in several ways. One distinction is based on the nature of the change specifically, on the number of bases changed. Mutations can also be defined according to their effects on amino acid sequences in proteins and these may be caused by base pair substitutions or additions or deletions.

7.2.1 Point mutation : the gene mutation in which there is only a single changed base pair of DNA is called point mutation. A point mutation may be a base substitution, a base insertion or a base deletion, but the term most frequently refers to a base substitution.

7.2.2 Transition mutation : it is a specific type of base pair substitution mutation involving a change from one purine-pyrimidine base pair to the other purine-pyrimidine base pair. The four types of transition mutations are AT to GC, GC to AT, TA to CG, and CG to TA (Fig. 7.1).

7.2.3 Transversion mutation : this is also a specific type of base pair substitution mutation which involves a change from a purine-pyrimidine base pair to a pyrimidine-purine base pair. The four types of transversion mutations are AT to TA, GC to CG, AT to CG, and GC to TA (Fig. 7.1).

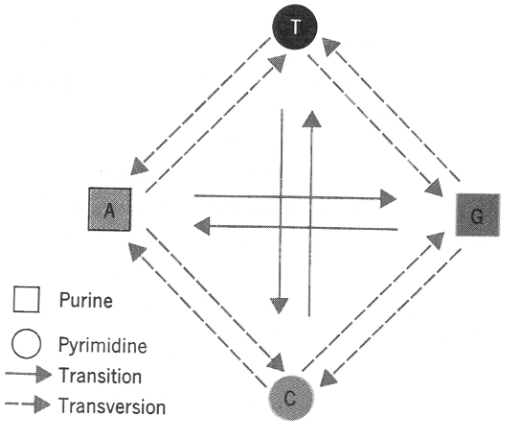


Fig. 7.1 Transition and Transversion of bases

7.2.4 Missense mutation : a gene mutation in which a base pair change in the DNA causes a change in an mRNA codon so that a different amino acid is inserted into the polypeptide in place of the one specified by the wild-type codon, resulting in an altered phenotype.

7.2.5 Nonsense mutation: it is a base pair change in the DNA that results in the change of an mRNA codon from one that specifies an amino acid to a chain terminating or nonsense codon i.e., UAG, UAA, and UGA. For example, an AT to TA transversion mutation changes the lysine codon 5′-AAA-3′ on the mRNA to 5′-UAA-3′ which is a nonsense codon. Because a nonsense mutation gives rise to chain termination at an incorrect place in a polypeptide, the mutation prematurely ends the polypeptide. As a result, instead of complete polypeptide, polypeptide fragments are released from the ribosome which are usually nonfunctional.

7.2.6 Neutral mutation : it is a base pair change in a gene that changes a codon in the mRNA such that the resulting amino acid substitution produces no detectable change in the function of the protein translated from that message. A neutral mutation is a subset of missense mutations and is the case where the new codon codes for a different amino acid that is chemically equivalent to the original and hence does not affect the protein's function. As a simple example, the AT to GC transition mutation changes the codon from 5′-AAA-3′ to 5′-AGA-3′ which substitutes arginine for lysine. Both arginine and lysine are basic amino acids and are sufficiently similar in properties so that the protein's function may well not be altered significantly.

7.2.7 Silent mutation : it is also a subset of missense mutations and is the case where a base pair change in a gene that alters a codon in the mRNA such that the same amino acid is inserted in protein. The protein in this case obviously has wild-type function. For example, a silent mutation

results from an AT to GC transition mutation which changes the codon from 5′-AAA-3′ to 5′-AAG-3′ , both of which specify lysine.

7.2.8 Frameshift mutation : this mutation results from the addition or deletion of one or more base pairs in a gene. An addition of deletion of one base pair shifts the mRNA's reading frame by one base so that incorrect amino acids are added to the polypeptide chain after the mutation site. Often, frameshift mutations generate new codons resulting in a shortened protein, or they result in read-through of the normal stop codon, resulting in longer than normal proteins. In any case, a frameshift mutation usually results in a nonfunctional protein.

7.3 Reverse and suppressor mutations

Point mutations generally fall into two classes in terms of their effects on the phenotype in comparison to the wild type. *Forward mutations* are mutations that cause the genotype to change from wild type to mutant, and *reverse mutations* or *reversions* or *back mutations* are the mutations that cause the genotype to change from mutant to wild type. A gene reversion is a mutational event that causes a change from a mutant phenotype to wil-dype or partially wild-type function. Reversion of a nonsense mutation, for instance, occurs when a base pair change results in a change of the mRNA nonsense codon to a codon for an amino acid, the mutation is a *true reversion*. If the reversion is to some other amino acid, the mutation is a *partial reversion*, and complete function may be restored.

The effects of a mutation may be diminished or suppressed by another mutation that occur at a different site from the original mutation and this is referred as *suppressor mutation.* This suppressor mutation is also called second-site mutation. A suppressor mutation does not result in a reversal of the original mutation; instead, it makes or compensates for the effects of the initial mutation. There are two major classes of suppressor mutations: those which occur within the same gene as the original mutations but at a different site called intragenic suppressors, and those occurring in a different gene called intergenic suppressors. Both intragenic and intergenic suppressors operate to allow the production of functional or partially functional copies of the protein which were initially rendered inactive by the original deleterious mutation. Thus, function can be restored only when both the original mutation and the suppressor mutation are present together in the same cell.

Intragenic suppressors act in one of two ways: by altering a different nucleotide in the same codon in which the original mutation occurred, or by altering a nucleotide in a different codon. As an example, a DNA sequence of three base pairs in the wild type specifies the mRNA codon 5′-CGU-3′, which is read as arginine. The original or first mutation is a GC to AT transversion at the first base pair, resulting in the mRNA codon 5′-AGU-3′, which specifies serine. The suppressor or second mutation is a TA to AT transversion at the third position, giving the mRNA codon 5′-AGA-3′, which is an arginine codon. Thus, with both mutations, the protein will be completely functional in cells.

Intergenic suppression is the suppression of a mutational defect by a second mutation in another gene. Genes that cause suppression of mutations in other genes are called suppressor genes. Many intergenic suppressors work by changing the way the mRNA encoded by the mutant gene is read. Each suppressor gene can suppress the effects of only one type of nonsense, missense, or frameshift mutation; hence, suppressor genes can suppress only a small proportion of the point mutations that theoretically can occur within a gene. On the other hand, a given suppressor gene will suppress all mutations for which it is specific, whatever gene the mutation is in.

Suppressors of nonsense mutations have been well characterized, particularly in *E. coli* and yeast. The suppressor genes in this case often are mutant tRNA genes. That is, particular tRNA genes can mutate so that their anticodons recognize a chain terminating codon and put an amino acid into the chain. Thus, instead of polypeptide chain synthesis being stopped prematurely as a result of a nonsense mutation, the altered or suppressor tRNA inserts an amino acid at that position, and full or partial function of the polypeptide may be restored. There are three classes of nonsense suppressors, one for each of the nonsense codons UAG, UAA, and UGA.

For example, a gene for a tyrosine tRNA which has the anticodon 3′-AUG-5′ is mutated so that the tRNA has the anticodon 3′-AUC-5′, the mutated suppressor tRNA which will still carry tyrosine will read the nonsense codon 5′-UAG-3′. So, instead of chain termination occurring, tyrosine is inserted at that point in the polypeptide. How functional the complete protein will be depend on the effects of the inserted tyrosine in the protein. If it is an important part of the protein, then the incorrect amino acid may not restore function to a significant degree. If it is in a less crucial area, the protein may have some or complete function.

7.4 Spontaneous Mutations

Spontaneous mutations occur naturally without any apparent or known cause, i.e., they are not induced by an exposure to a mutagenic agent. There are two possible sources of origin of spontaneous mutations – (1) errors during DNA replication, and (2) mutagenic effects of the natural environment of the organisms. The DNA polymerase incorporates wrong bases during DNA replication with a frequency of 10^{-5}. But DNA polymerase also has $3' \rightarrow 5'$ exonuclease activity which enables it to proofread the newly synthesized DNA strands and, thereby, correct the errors made during DNA replication. The DNA polymerase proofreads every nucleotide newly added to the growing end of the polynucleotide chains. It detects a mis-matched nucleotide, incorporates the correct nucleotide in its place and then proceeds with the further synthesis of the chain.

The contribution of environment to spontaneous mutations is not clearly understood. Some of the solar radiations are surely mutagenic and are likely contributors to spontaneous mutations. For example, the UV rays of sunlight are highly mutagenic but they are unable to penetrate more than skin-deep in animals and higher plants. The mutagenic action of UV rays becomes dramatically apparent in human beings suffering from a genetic disease called xeroderma pigmentosum. These individuals lack an effective mechanism for the repair of damage to DNA caused by UV rays. As a result, they develop mild to severe skin cancer in the areas of body

exposed to sunlight. On the other hand, both eukaryotic and prokaryotic genomes contain some mobile DNA elements, e.g., insertion sequences, which integrate at specific sites in their genomes and produce gene mutations. The rate of spontaneous mutations is very low and generally ranges between 10^{-8} and 10^{-10} per nucleotide per generation for forward mutations in different genes of bacteria and viruses. For eukaryotes the estimates range from 10^{-7} to 10^{-9} per nucleotide per generation. The rates of reverse mutations have been estimated in prokaryotes and are generally much lower than those for forward mutations.

7.5 Induced Mutations

Mutations produced due to the treatment with either a chemical or a physical agent are called induced mutations. The agents capable of inducing mutations are known as mutagens and their capacity for inducing mutations is termed as mutagenic property. Induced mutations are useful in two different ways – (1) in genetic and biochemical studies, and (2) in crop improvement. The process of inducing mutations through treatment with a mutagen is known as mutagenesis. Since the rate of spontaneous mutation is so low, geneticists use mutagens to increase mutation frequency so that a significant number of organisms have mutations in the gene being studied. Treatments with various mutagens increase the mutation rate by several orders of magnitude. Some potent chemical mutagens may produce mutations at the rate of more than 1% per gene/generation. Generally two classes of mutagens, radiation and chemical, are used and both involve specific mechanisms of action.

7.5.1 Radiation

Both X rays and UV rays are used to induce mutations. X rays are an example of ionizing radiation. Ionizing radiation can penetrate tissues, hence the use of X rays as a diagnostic tool. Collision of ionizing radiation with atoms in its path gives rise to ions and reactive chemical radicals that can break chemical bonds, including those in DNA. That is, the products of ionizing radiation can induce chromosome breakages, chromosome rearrangements, and damage to DNA, e.g., point mutations. In fact, ionizing radiation is the leading cause of gross chromosomal mutations in humans. At certain low levels of ionizing radiation, point mutations are commonly produced and at these levels, there is a linear relationship between the rate of point mutations and radiation dosage.

UV rays are non-ionizing and they have insufficient energy to induce ionizations. However, UV light is a useful mutagen, and at high enough doses it can kill cells. UV light is used as a sterilizing agent in some applications. Ultraviolet light causes mutations because the purine and pyrimidine bases in DNA absorb light very strongly in the UV range (254 to 260 nm). At this wavelength UV light induces gene mutations primarily by causing photochemical changes in the DNA. One of the effects of UV radiation on DNA is the formation of abnormal chemical bonds between adjacent pyrimidine molecules in the same strand. This bonding is induced mostly between adjacent thymines, forming the structures called thymine dimers, usually designated T ^ T, C ^ C, C ^ T and T ^ C pairs are also produced by UV radiation, but in much lower amounts. This unusual pairing produces a bulge in the DNA strand and disrupts the normal pairing of the Ts with

the corresponding As on the opposite strand. Many of the thymine dimers are repaired leaving no trace of the original damage.

7.5.2 Chemical Mutagens

Many chemicals are reported to be mutagenic as they induce mutations. Basing on their mechanism of action, these chemicals may be grouped into different classes. Some of the most important chemical mutagens are Base analogues, Nitrous acid, Acridines, and alkylating and hydroxylating agents. Chemical mutagens may be classified into two broad groups indicating on the state of DNA they affect. The first group consists of those mutagens which affect both replicating as well as non-replicating DNA, e.g., alkylating agents, nitrous acid, hydroxylamine and acridine dyes. The second group includes those chemicals which affect only replicating DNA, e.g., base analogues.

7.5.2.1 Base analogues

Base analogues, in their molecular structures, are very similar to the bases normally found in DNA. Base analogue mutagens cause mutations because they can exist in alternate states or tautomeric states, a normal state and a rare state. In each of the two states the base analogue pairs with a different base in DNA so that base-pair substitution mutations can be produced. The most common base analogues are 5-bromouracil (5BU) and 2-aminopurine (2-AP).

Structurally 5-bromouracil has a bromine residue instead of the methyl group of thymine (Fig.7.2). In the normal state 5BU resembles thymine and will pair only with adenine in DNA. In its rare state it pairs only with guanine. 5BU induces mutations by switching between the two forms once the base analogue has been incorporated into the DNA. Since the base analogues exist at a reasonably high frequency, the frequency of mutations induced by base analogues is much higher than the frequency of spontaneous mutations.

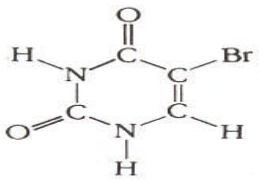


Fig. 7.2 Chemical structure of 5BU

If 5BU is incorporated in its normal state, it pairs with adenine. If it changes into its rare state during replication, then guanine will be incorporated as the complementary base pair instead of adenine. In the next round of replication the G-5BU base pair will be resolved into a G-C base pair instead of the A-T base pair. By this process a transition mutation is produced, in this case from AT to GC. The 5BU will induce a mutation from GC to AT if it is first incorporated into

DNA in its rare state and then switches to the normal state during replication (Fig.7.3). Thus 5BU-induced mutations can be reverted by a second treatment of 5BU.

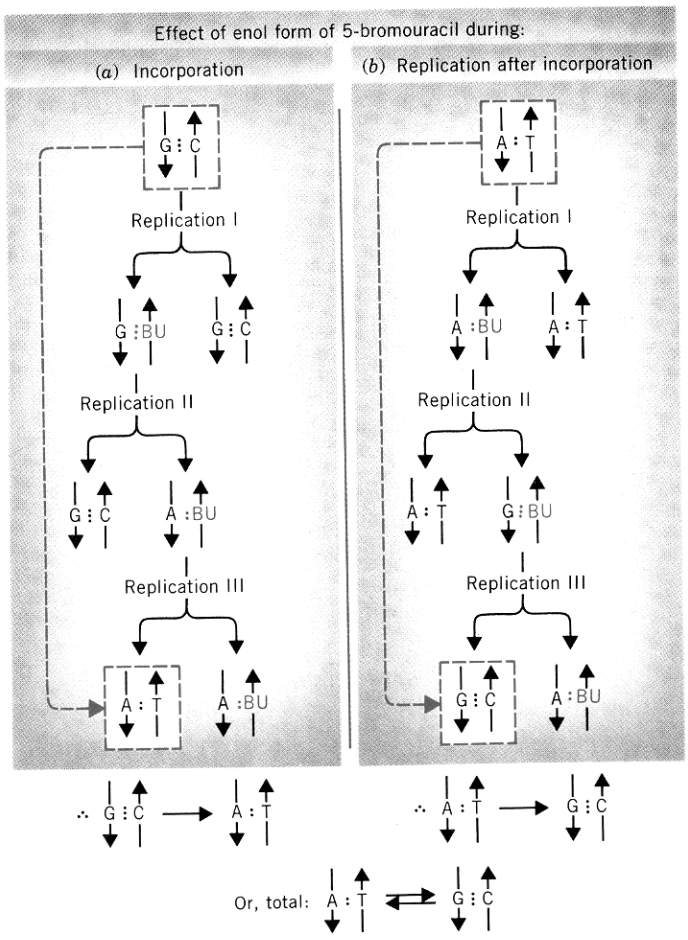


Fig. 7.3 Transition mutation by 5BU

The base analogue 2-aminopurine acts as a mutagen in essentially the same way as 5BU does, and like 5BU, 2AP exists in normal and rare states. In its normal state 2AP resembles adenine but has an amino group at a different position on the purine ring (Fig.7.4) than does adenine. In its normal state it base pairs with thymine. In its rare state 2AP resembles guanine and base pairs with cytosine. As with 5BU, 2AP induces transition mutations, which can be reverted by a second application of 2AP. The mutation will be either AT to GC or GC to AT, depending on its state during its initial incorporation into the DNA and on its state during replication.

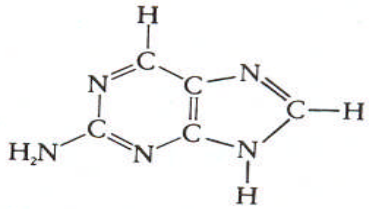


Fig. 7.4 Chemical structure of 2AP

7.5.2.2 Nitrous acid

Nitrous acid (HNO_2) is a potent mutagen which acts both on replicating and non-replicating DNA. Nitrous acid is a deaminating agent that removes amino groups from the bases. It leads to oxidative deamination of cytosine, adenine and guanine in which the amino groups are replaced by keto (=O) groups yielding uracil, hypoxanthine and xanthine, respectively. Treatment of guanine with nitrous acid produces xanthine, but since this purine base has the same pairing properties as guanine, no mutation results. However, when cytosine is treated with nitrous acid, uracil is produced which pairs with adenine. The deamination of cytosine by nitrous acid, then, produces a CG-to-TA transition mutation during replication. Likewise, nitrous acid modifies adenine to produce hypoxanthine, a base that pairs with cytosine rather than thymine, thus resulting in an AT-to-GC transition mutation. Thus nitrous acid produces both AT to GC as well as GC to AT transitions and these transitions are produced only when the DNA treated with HNO_2 undergoes replication. A nitrous acid-induced mutation can be reverted by a second treatment with nitrous acid.

7.5.2.3 Acridines

The acridine group includes acridine orange, proflavine and acriflavine. These compounds are intercalating substances. The acridines are planar, three-ringed molecules (Fig. 7.5) whose dimensions are roughly same as those of a purine-pyrimidine pair. In aqueous solution, these substances form stacked arrays and are also able to stack with a base pair. This insertion occurs between bases in adjacent pairs, a process called intercalation (Fig. 7.6). Since the thickness of the acridine molecules is approximately that of a base pair and because the two bases of a pair are normally in contact, the intercalation of one acridine molecule causes adjacent base pairs to move apart by a distance equal to that of the thickness of one base pair. This has bizarre effects on the outcome of DNA replication, though the mechanism of action of the mutagen is not known. When DNA containing intercalated acridines is replicated, additional bases appear in the sequence. The usual addition is a single base, though occasionally two bases are added. Deletion of a single base also occurs but this is far less common than base addition. Mutations of this sort are called frameshift mutations. This is because the base sequence is read in groups of three bases when it is being translated into an amino acid sequence and the addition of a base changes the reading frame.

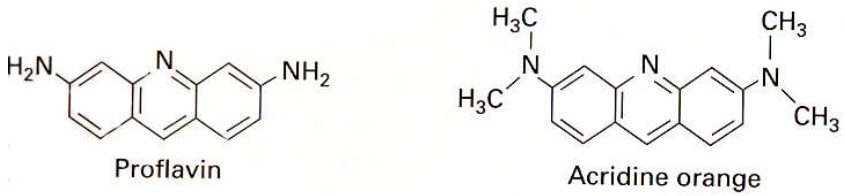


Fig. 7.5 Chemical structures of Proflavin and Acridine orange

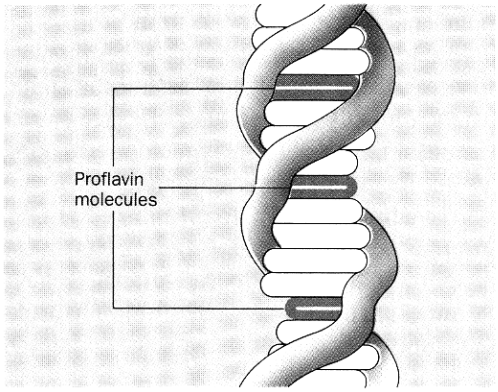


Fig. 7.6 Intercalation of Proflavin molecule between bases

7.5.2.4 Alkylating and Hydroxylating Agents

Several chemicals have reactive alkyl groups (methyl and ethyl) which they transfer to DNA and to phosphate groups of DNA. These chemicals are called alkylating agents and the process is known as alkylation. Alkylating agents such as nitrogen and sulfur mustards, methyl methanesulfonate (MMS), ethyl methanesulfonate (EMS) and nitrosoguanidine (NTG) and many others have several effects on DNA. One major mechanism of mutagenesis by alkylating agents involves the transfer of methyl or ethyl groups to the bases such that their base pairing potentials are altered and transitions result.

The transfer of alkyl groups to the phosphate residues of DNA produces phosphate trimesters. Most of these alkyl groups are removed from DNA through hydrolysis. But some alkyl groups may not be removed, which may interfere with DNA replication as well as cause breakage in the sugar-phosphate backbone. Ethylation at the 7-N position and at the 6-O position are believed to be two effects of EMS. 7-ethylguanine is then believed to base pair with thymine. 7-ethyl guanine often dissociates from deoxyribose and is removed from the DNA strand resulting in depurination. The gap in the DNA strand thus produced may lead to a break in the strand, may interfere with DNA replication or may lead to the incorporation of a wrong base during repair or replication producing a mutation. Other base alkylation products are believed to somehow activate repair processes in much the same way as thymine dimers do. The occasional errors occurring in these repair processes may lead to transversions and frameshift mutations in addition to transitions. Some alkylating agents, particularly difunctional alkylating agents, cross-link DNA strands and/or molecules and induce chromosome breaks and the various kinds of chromosomal

aberrations found correlated with breaks. The alkylating agents exhibit less specificity in their mutagenicity than base analogues, nitrous acid, or acridines. They induce all types of mutations, including transitions, transversions, frameshifts, and even chromosome aberrations, with various relative frequencies, depending on the specific alkylating agent employed.

Nitrosoguanidine (NTG), one of the most potent chemical mutagens known, has been found to induce clusters of closely linked mutations in the segment of the chromosome that is replicating during the mutagenic treatment. Mutants isolated after NTG treatment often carry multiple, closely linked mutations, making them less useful for most genetic studies.

The hydroxylating agent, hydroxylamine (NH_2OH) has a very specific mutagenic effect in contrast to many of the alkylating agents. It induces only GC to AT transitions. Although the exact mechanism of hydroxylamine mutagenesis is still somewhat uncertain, it apparently acts by hydroxylating the amino group of cytosine. The resulting hydroxylaminocytosine can base pair with adenine to produce the observed GC to AT transitions. Because of its speicificity, hydroxylamine has been very useful in classifying transition mutations.

7.6 Summary

Mutation is the process by which a change in DNA base pairs or a change in the chromosomes is produced. Mutations at the level of chromosomes are called chromosomal mutations. Mutations in the sequences of genes at the level of the base pair are called gene mutations. Gene mutations may occur by a substitution of one base pair for another or by the addition or deletion of one or more base pairs. The consequences of a gene mutation to an organism depend upon a number of factors, especially the extent to which the amino acid coding information is changed. A missense mutation cause the substitution of one amino acid for another, an nonsense mutation cause premature termination of synthesis of the polypeptide.

The effects of a gene mutation can be reversed either by reversion of the gene sequence to its original state, or by a mutation at a site distinct from that of the original mutation. The latter is called a suppressor mutation. Suppressor mutations that occur within the same gene as the original mutation are intragenic suppressors. They act either by altering a different nucleotide in the same codon affected by the original mutation, or by altering a nucleotide in a different codon. Suppressor mutations that occur in a different gene from the original mutation are called intergenic suppressors. Often, suppression by this class of suppressors involves a tRNA with an altered anticodon.

Mutations occur spontaneously at a low rate. The mutation rate can be increased through the use of mutagens like irradiation and certain chemicals. Typically, mutagens are used by researchers so that a mutant of interest is more likely to be found in a population of cells. Chemical mutagens work in a number of different ways, such as by acting as base analogues, by modifying bases, or by intercalating into the DNA. The later results in frameshift mutations, while the others result in base pair substitution mutations.

7.7 Model Questions

1. Define mutation? Explain the different types of mutations.
2. What is a mutagen? Explain different types of chemical mutagens and their mode of actions.
3. Short Questions
   a. Transition and Transversion mutations
   b. Base analogues
   c. Induced mutations
   d. Spontaneous mutations
   e. Suppressor mutations

7.8 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 8

GENETIC CODE

8.0 Objective

The aim of this lesson to acquaint the reader with the nature and features of the genetic language used during translation for the synthesis of a polypeptide chain.

8.1 Introduction

The number and the sequence of bases in mRNA specifying an amino acid is known as codon, while the set of bases in a tRNA that base pair with a codon of an mRNA is known as anticodon. The sequence of bases in an anticodon is exactly run antiparallel to each other when they base pair. The set of all the codons that specify the 20 amino acids is termed as the genetic code or genetic language or coding dictionary. The genetic language has only four alphabets, A, T, G and C, the four bases making up polynucleotides. If it were a one-letter code, only four amino acids could be encoded. If it were a two-letter code, then only 16 amino acids could be encoded. A three-letter code generates more than enough to code for the 20 amino acids. So, the genetic code uses combinations of three bases to specify a particular amino acid. A group of three bases is termed as triplet codon. The linear order of sense codons in the mRNA dictates the linear order of

amino acids in the polypeptide. Special codons also direct initiation and termination of the polypeptide chain. Codons are read by the protein synthesis machinery from initiation to termination in a non-overlapping and unpunctuated manner, i.e., each triplet of nucleotides is independent of any other and there are no gaps between codons. This allows any nucleotide sequence to be divided up into codons in three ways, i.e., there are three reading frames. The reading frame used for protein synthesis is determined by the position of the initiation codon.

8.2 Nature and Characteristics of the Genetic Code

The salient features of the genetic code are

- The code is triplet code
- The code is comma-free; that is, it is continuous
- The code is non-overlapping
- The code is almost universal
- The code is degenerate
- The code has start and stop signals
- Wobble occurs in the anticodon

8.3 The genetic code is a triplet code

Twenty different amino acids are incorporated during translation. Thus, at least 20 different codons must be formed using the four symbols (bases-A,U,G,C) available in the message (mRNA). Two bases per codon would yield only 4^2 or 16 possible codons that clearly not enough. Three bases per codon yield 4^3 or 64 possible codons (Fig. 8.1)- and apparent excess.

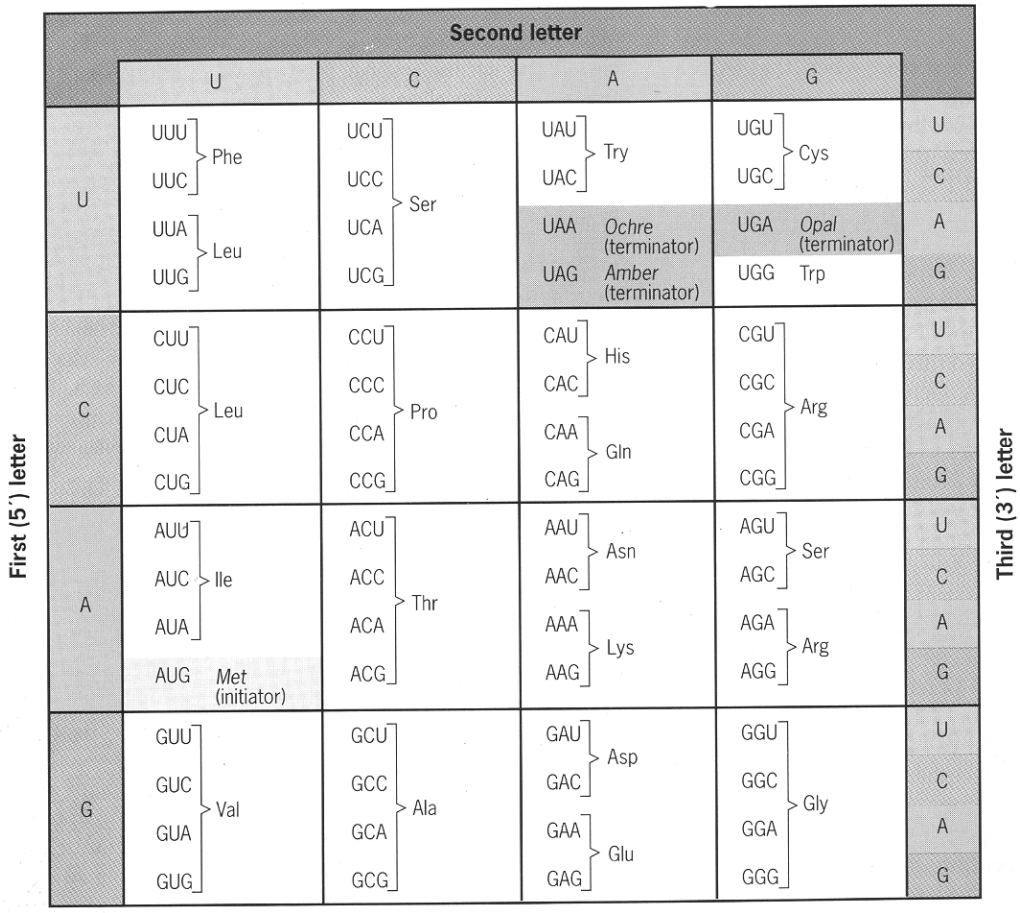


Fig. 8.1 Chart representing the 64 codons

The first strong evidence that the genetic code was in fact a triplet code (three nucleotides per codon) resulted from a genetic analysis of proflavin-induced mutations in the rII locus of phage T4 carried out by F. H. C. Crick and colleagues in 1961. Crick and colleagues isolated proflavin-induced revertants of a proflavin-induced mutation. These revertants were shown to result form the occurrence of suppressor mutations rather that from back-mutation at the original site of mutation. Crick and colleagues reasoned that if the original mutation was a single base pair addition or deletion, then the suppressor mutations must be single base pair deletions or additions, respectively, occurring at a site or sites near the original mutation. A single base pair addition or deletion will alter the reading frame of the gene and mRNA for that portion of the gene distal to the mutation. When the suppressor mutations were isolated as single mutants by screening progeny of backcrosses to wild type, they were found to produce mutant phenotypes, just like the original mutation.

Crick and colleagues next isolated proflavin-induced suppressor mutations of the original suppressor mutations, and so on. All the isolated mutations were then classified into two groups, plus (+) and minus (-) for additions and deletions. They reasoned that a (+) mutation would

suppress a (-) mutation, but not another (+) mutation, and vice versa. Next, Crick et al. constructed recombinants that carried various combinations of the (+) and the (-) mutations. Recombinants with two (+) mutations or two (-) mutations always had mutant phenotypes, just like the single mutants. Recombinants carrying three (+) mutations or three (-) mutations, however, often had wild-type phenotypes. This indicated that the additions of three base pairs or the deletion of three base pairs left the distal portion of the gene with the correct wild-type reading frame, a result that would be expected only if each codon contained three nucleotides. The further confirmation that the coding ratio is indeed three has come from many other studies using *in vitro* translation systems.

8.4 Deciphering the genetic code

The exact relationship of the 64 codons to the 20 amino acids was determined by experiments done mostly in the laboratories of Marshall Nirenberg and Ghobind Khorana. Essential to these experiments was the use of cell-free, protein-synthesizing systems, which contained ribosomes, tRNAs with amino acids attached, and all the necessary protein factors for polypeptide synthesis. These cell-free, synthesizing systems were assembled from components isolated and purified from *E. coli*. Protein synthesis in these systems is inefficient, in that very little protein is made. Therefore, radioactively labeled amino acids have to be used to measure the incorporation of amino acid into new proteins.

In one approach to establish which codons specify which amino acids, synthetic mRNAs containing one, two, or three different types of bases were made and added to cell-free, protein-synthesizing systems. The polypeptides made in these systems were then analyzed. When the synthetic mRNA contained only one type of base, the results were unambiguous. Synthetic poly(U) mRNA, for example, directed the synthesis of a polypeptide consisting of phenylalanines, a polyphenylalanine chain. Since the genetic code is a triplet code, this result indicated that UUU is a codon for phenylalanine. Similarly, a synthetic poly(A) mRNA directed the synthesis of a polylysine, and poly(C) directed the synthesis of polyproline, indicating that AAA is a codon for lysine and CCC is a codon for proline. The results from poly(G) were inconclusive since the poly(G) folds up upon itself, so it cannot be translated in vitro.

Synthetic mRNAs made by the random incorporation of two different bases called random copolymers were also analyzed in the cell-free, protein-synthesizing systems. When mixed copolymers are made, the bases are incorporated into the synthetic molecule in a random way. Thus poly(AC) molecules can contain eight different codons (CCC, CCA, CAC, ACC, CAA, ACA, AAC, and AAA), and the proportions of each depend on the A-to-C ration used to make the polymer. In the cell-free, protein-synthesizing system, poly(AC) synthetic mRNAs caused the incorporation of asparagine, glutamine, histidine, and threonine into polypeptides, in addition to the lysine expected from AAA codons and the proline expected from CCC codons. The proportions of asparagine, glutamine, histidine, and threonine incorporated into the polypeptides produced depend on the A/C ratio used to make the mRNA, and these observations were used to deduce information about the codons that specify the amino acids. For example, since an AC random copolymer containing much more A than C resulted in the incorporation of many more asparagines than histidines, researches concluded that asparagine is coded by two As and one C

and histidine by two Cs and one A. With experiments of this kind, the base composition of the codons for a number of amino acids was determined.

Another experimental approach also used copolymers, but these copolymers had been synthesized so that they had a known sequence, not a random one. For example, a repeating copolymer of U and C gives a synthetic mRNA of **UCUCUCUCUC**. When this copolymer is tested in a cell-free, protein-synthesizing system, the resulting polypeptide had a repeating amino acid pattern of leucine-serine-leucine-serine. From this result, researches conclude that UCU and CUC specify leucine and serine, although they could not determine from this information alone which coded for which.

Another approach developed by Nirenberg and Philip Leder in 1964 used a ribosome-binding assay. This assay depended on the fact that, in the absence of protein synthesis, specific tRNA molecules will bind to complexes formed between ribosomes and mRNAs. For example, when synthetic mRNA poly(U) is mixed with ribosomes, it forms a poly(U)-ribosome complex, and only tRNA.Phe i.e., the tRNA that will bring phenylalanine to an mRNA and that has the appropriate anticodon, AAA, for the UUU codon) will bind to the UUU codon. Importantly, the specific binding of the appropriate tRNA to the mRNA-ribosome complex does not require the presence of long mRNA molecules; the binding of a trinucleotide will suffice. The discovery of the trinucleotide-binding property made it possible to determine many codons and the amino acids for which they code. Using the ribosome-binding assay, many of the ambiguities that had arisen from other approaches were resolved. For example, UCU was found to promote the binding of a tRNA.Ser, that is, a tRNA that brings the amino acid serine to the message. Thus UCU must code for serine. Similarly, CUC causes tRNA.Leu binding, so CUC codes for leucine. All in all, about 50 codons were clearly identified by using this approach.

8.5 Genetic code is non-overlapping

The mRNA is read in successive groups of three nucleotides. A message of AAGAAGAAG… in the cell would be read as lysine-lysine-lysine….., which is what the AAG specifies. Theoretically, three readings are possible from this message, depending on where the reading is begun namely, the repeating AAG, the repeating AGA, and the repeating GAA.

8.6 Ambiguity in the Genetic Code

There is no evidence that the genetic code is ambiguous *in vivo*. Ambiguity denotes that a single codon may code for more than one amino acid. The only exception appears to be the AUG codon in prokaryotes. It codes for formylmethionine at the initiation site, while at other positions it specifies methionine. It is not clear how this precise distinction is always made, but it is believed that a secondary structure of the mRNA at the initiation site may be involved in coding for formylmethionine. However, the genetic code is sometimes ambiguous in vitro. For example, poly-U incorporates small amounts of leucine in addition to phenylalanine. If the ribosomes are treated with streptomycin, codon UUU codes for, in addition to phenylalanine, isoleucine, serine and leucine, although less frequently. Low temperature, ethyl alcohol and high Mg^{++} concentration

also cause ambiguity. These agents as well as streptomycin most likely produce this effect by interacting with a protein, S 10 of the 30 S subunit, of ribosomes.

8.7 Universality of genetic code

The genetic code is nearly universal i.e., the codons have the same meaning and codes for same amino acids in all the organisms, with minor exceptions. The evidence for universality comes from a variety of experiments. Firstly, haemoglobin mRNA from man is precisely translated in in vitro systems having tRNA from *E. coli* and rabbit ribosomes, and rabbit haemoglobin mRNA is translated in α and β haemoglobin chains in *Xenopus* eggs. Secondly, synthetic polyribonucleotides e.g., poly-U, poly-C etc., direct the incorporation of the same amino acids in the cell-free extracts from *E. coli, Chlamydomonas*, rat and mice.

The most important exceptions to the universality of the code occur in mitochondria of mammals, yeast, and several other species. Mitochondria have their own chromosomes and protein synthesizing machinery. Although the mitochondrial and cytoplasmic systems are similar, there are some differences. In the mitochondria of humans and other mammals, (1) UGA specifies tryptophan rather than chain termination (2) AUA is a methionine codon, not an isoleucine codon, and (3) AGA and AGG are chain-termination codons rather than arginine codons. The other 60 codons have the same meaning in mammalian mitochondria as in nuclear mRNAs. There are also rare differences in codon meaning in the mitochondria of other species and in nuclear transcripts of some protozoa. However, since these exceptions are rare, the genetic code should be considered nearly universal.

8.8 Degenerate and Ordered code

All the amino acids except methionine and tryptophan are specified by more than one codon. Three amino acids – leucine, serine, and arginine—are each specified by six different codons. Isoleucine has three codons. The other amino acids each have either two or four codons. The occurrence of more than one codon per amino acid is called degeneracy. The degeneracy in the genetic code is not at random; instead, it is highly ordered. In most cases, the multiple codons specifying a given amino acid differ by only one base i.e., the third or 3′ base of the codon. The degeneracy is primarily of two types. (1) Partial degeneracy occurs when the third base my be either of the two pyrimidines (U or C) or, alternatively, either of the two purines (A or G). With partial degeneracy, changing the third base from a purine to a pyrimidine, or vice versa, will change the amino acid specified by the codon. (2) In the case of complete degeneracy, any of the four bases may be present at the third position in the codon, and the codon will still specify the same amino acid. For example, valine is encoded by GUU, GUC, GUA, and GUG.

8.9 Wobble hypothesis

Since 61 sense codons specify amino acids in mRNA, a total of 61 tRNA molecules could have the appropriate anticodons. Theoretically, though, the complete set of 61 sense codons can be read by fewer than 61 distinct tRNAs because of the wobble in the anticodon. The hydrogen bonding between the bases in the anticodon of rRNAs and the codons of mRNAs appears to

follow strict base pairing rules only for the first two bases of the codon. The base pairing involving the third base of the codon is apparently less stringent, and called wobble at this site. In 1966, Crick postulated the wobble hypothesis to account for the degeneracy of genetic code. According to this hypothesis, the base pairing at the third position is wobble i.e., not precise (Fig 8.2). The base U in the third position of anticodon my pair with A or G in the codon, G may pair with C or U, and I (inosinic acid having the base hypoxanthine) may pair with U, C or A. For example, phenylalanine tRNA ($tRNA^{phe}$) from yeast having the anticodon AAG translates equally well the codons UUU and UUC. Similarly, a tRNA specific for alanine ($tRNA^{ala}$), isolated again from yeast, has the anticodon CGI, and is able to pair with the codons GCU, GCC and GCA, all the codons coding for alanine. The occurrence of three tRNAs for the six serine codons have also been predicted because of wobble hypothesis. Three serine tRNAs have been characterized: (1) $tRNA^{Ser1}$ (anticodon AGG) binds to codons UCU and UCC, (2) $tRNA^{Ser2}$ (anticodon AGU) binds to codons UCA and UCG, and (3) $tRNA^{Ser3}$ (anticodon UCG) binds to codons AGU and AGC. These specificities were verified by the trinucleotide stimulated binding of purified aminoacyl-tRNAs to ribosomes *in vitro*.

Several tRNAs contain the base inosine, which is a ribonucleoside made from the purine hypoxanthine. Inosine is produced by a post-transcriptional modification of adenosine. Crick's wobble hypothesis predicted that when inosine is present at the 5′ end of an anticodon (the wobble position), it would base pair with adenine, uracil, or cytosine in the codon. In fact, purified alanyl-tRNA containing inosine (I) at the 5′ position of the anticodon binds to ribosomes activated with GCU, GCC, or GCA trinucleotides. The same results has been obtained with other purified tRNAs with inosine at the 5′ position of the anticodon.

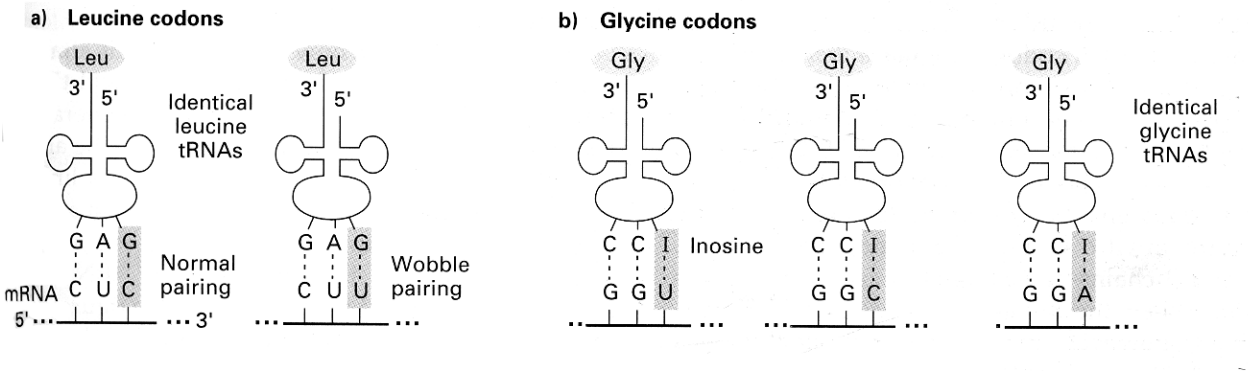


Fig. 8.2 Wobbling of base pair

8.10 Initiation and Termination Codons

The genetic code also provides for punctuation of genetic information at the level of translation. In both prokaryotes and eukaryotes, the codon AUG is used to initiate polypeptide chains. In rare instances, GUG is used as an initiation codon. In both cases, the initiation codon is recognized by an initiator tRNA, $tRNA_f^{Met}$ in prokaryotes and $tRNA_i^{Met}$ in eukaryotes. In prokaryotes, an AUG codon must follow an appropriate nucleotide sequence in the 5′

nontranslated segment of the mRNA molecule in order to serve as translation initiation codon. In eukaryotes, the codon must be the first AUG encountered by the ribosome as it scans from the 5′ end of the mRNA molecule. At internal positions, AUG is recognized by $tRNA^{Met}$, and GUG is recognized by a valine tRNA.

Three codons namely UAA (Ochre codon), UAG (Amber codon) and UGA (Opal codon) specify polypeptide chain termination. These three codons are the stop codons, also called nonsense codons, or chain terminating codons. These codons do not specify an amino acid, and no tRNAs in normal cells carry the appropriate anticodons. These stop codons are used singly or in tandem groups to specify the end of the translation process of a polypeptide chain. These codons are recognized by protein release factors, rather than by tRNAs. Prokaryotes contain two release factors, RF-1 and RF-2. RF-1 terminates polypeptides in response to codons UAA and UAG, whereas RF-2 causes termination at UAA and UGA codons. Eukaryotes contain a single release factor that recognizes all the three termination codons.

8.11 Summary

The genetic code is a triplet code, a sequence of three nucleotides making up each codon. The code is degenerate in the sense that all 64 possible triplet nucleotide sequences are used and several different codons frequently code for the same amino acid. Each of the 61 amino acid-specifying codons in mRNA is capable of base pairing with a three-nucleotide sequence present at the anticodon site of one or more tRNAs. The base pairing between the third (3′) base of the codon and the 5′ base of the anticodon does not follow the normal strict base pairing rules; instead, there is ‘wobble’ at this site, permitting base pairs to form other than the usual for base pairs. Thus, the anticodon of a single tRNA may recognize one, two or three different codons. The genetic code is nearly universal and the codons appear to have almost the same meaning. Protein synthesis is typically initiated by codon AUG, and it is terminated by terminated by three codons, UAG, UAA, UGA, either singly or in combination.

8.12 Model Questions

1. Write an account on salient features of Genetic code.
2. Short questions
   a. Deciphering the triplet code
   b. Wobble hypothesis
   c. Universality and non-overlapping of genetic code.

8.13 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 9

GENE EXPRESSION

9.0 Objective

The purpose of this lesson plan is to make the reader very clear about the nature of gene expression in prokaryotes.

9.1 Introduction

Crick, in 1958, proposed that the information present in DNA in the form of base sequence is transferred to RNA and then from RNA it is transferred to protein in the form of amino acid sequence, and that this information does not flow in the reverse direction, that is, from protein to RNA to DNA. The DNA molecules provide the information for their own replication. This relationship between DNA, RNA and protein molecules is known as the **Central Dogma**. The structure, function, development and reproduction of an organism depends on the properties of the proteins present in each cell and tissue. A protein consists of one or more chains of amino acids. Each chain of amino acids is called a polypeptide. The sequence of amino acids in a polypeptide chain is coded for by a gene that is a specific base pair sequence in DNA. When a protein is needed in the cell, the genetic code for that protein's amino acid sequence must be read from the

DNA and processed into the finished protein. Two major steps occur in the process of protein synthesis – **transcription** and **translation**.

Transcription is the transfer of information from a double-stranded, template DNA molecule to a single-stranded RNA molecule. Translation or protein synthesis is the conversion in the cell of the messenger RNA (mRNA) base sequence information into the amino acid sequence of a polypeptide. Unlike the DNA replication, transcription and translation generally occur throughout the cell cycle.

9.2 Bacterial RNA Polymerase

The RNA polymerase is similar in most bacteria but has a different structure in cyanobacteria and archaebacteria. In most bacteria, a single type of enzyme is responsible for the transcription of all the various types of RNA, e.g., mRNA, rRNA and tRNA. The *E. coli* complete RNA polymerase enzyme is capable of recognizing the appropriate promoter site for initiating the transcription and to continue the synthesis of RNA using DNA as template. The complete enzyme is known as the 'holoenzyme'. The holoenzyme consists of two components namely the 'core enzyme' and the 'sigma factor'. The holoenzyme is composed of different subunits plus sigma factor and is symbolized as $\alpha_2\beta\beta'\sigma$. The core enzyme can not initiate transcription at the proper sites, but it can synthesize RNA using DNA as a template. The core RNA polymerase consists of four polypeptides which are of three types –

- The α subunit- this polypeptide is coded by rpoA gene. It is 40,000 daltons in molecular weight and present in two copies / core enzyme molecule. It functions in binding with promoter of DNA.
- The β subunit – this subunit is coded by rpoB gene. It is 1,55,000 daltons in molecular weight and present in a single copy/enzyme molecule. It is involved in binding with the incoming nucleotides for RNA synthesis.
- The β′ subunit – the β′ polypeptide is coded by rpoC and 1,60,000 daltons in molecular weight. It is involved in binding with the template DNA i.e., the single stranded stretch of DNA generated to allow the transcription to proceed.

Each subunit contributes to the function of the core enzyme as a whole, which should have at least the following four distinct functional sites viz., a DNA unwinding site, the site binding to the antisense strand, a site binds to the sense strand and a DNA rewinding site.

The sigma factor is involved in binding of the RNA polymerase to the promoter DNA and possibly with the transcription initiation. It is not involved in the RNA synthesis as it is released when the RNA chain reaches 8-9 bases, the core enzyme then continues transcription. The *E. coli* sigma factor (σ^{70}) is a single polypeptide of 85,000 daltons and is coded by the rpoD gene. The chief function of the sigma factor is to ensure the stable binding of RNA polymerase to only at the promoters. The binding of RNA polymerase to any DNA segment is called loose binding. The association of σ with the core enzyme drastically reduces its ability for loose binding with DNA.

The association of core enzyme with σ factor is a critical factor in making the holoenzyme capable of recognizing the promoter binding sites for transcription initiation. When the holoenzyme forms a tight binding complex with DNA, σ factor contacts the DNA upstream of the

start point. In the case of tight binding, a short stretch of the DNA associated with the holoenzyme becomes single-stranded and such complexes are very stable. The loose binding of the enzyme yields a closed binary complex in which the DNA is present as normal duplex in association with the enzyme. In contrast, the tight binding of the enzyme gives rise to the open binary complex in which a part of the DNA associated with the enzyme is single stranded. When RNA synthesis begins, the open binary complex is converted into a ternary complex since the enzyme is now complexed with a DNA-RNA hybrid segment as well.

9.3 Transcription

Transcription begins with the attachment of RNA polymerase holoenzyme with the promoter region of transcription unit and comes to an end when the core enzyme reaches the terminator site and dissociates from the DNA. The entire process may be divided into three steps viz., Initiation, Elongation and Termination. Mechanism of transcription is given in the figure 9.1 below.

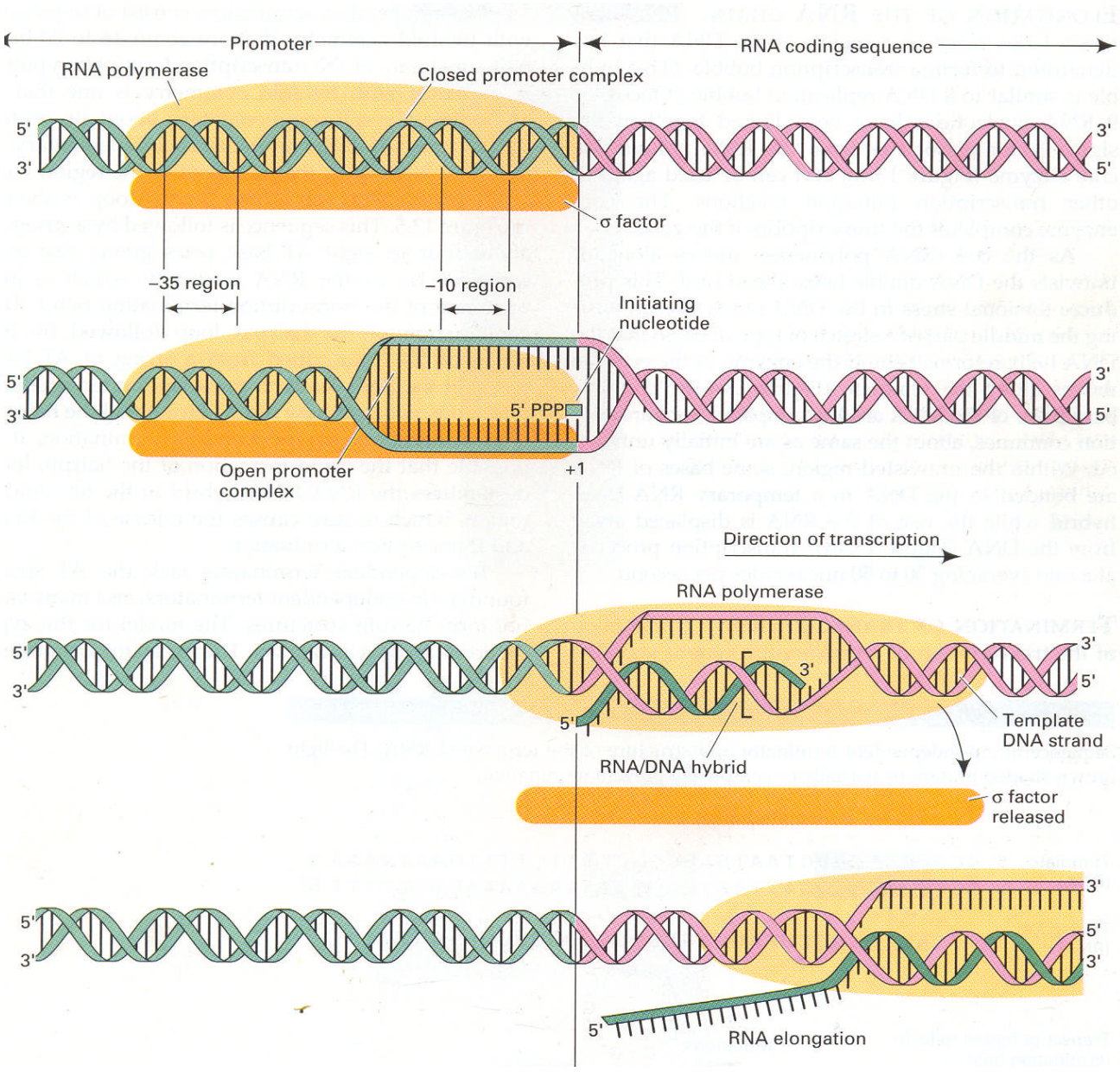


Fig. 9.1 Mechanism of Transcription

9.3.1 Initiation of RNA chain

Initiation of RNA chain involves three steps – (1) binding of the RNA polymerase holoenzyme to a promoter region in DNA, (2) the localized unwinding of the two strands of DNA by RNA polymerase, providing a template strand free to base pair with incoming ribonucleotides, and (3) the formation of phosphodiester bonds between the first few ribonucleotides in the nascent RNA chain. Initially the enzyme binds to the duplex DNA of the promoter region forming the closed binary complex. The sigma subunit of RNA polymerase mediates its binding to promoter in

DNA. Two DNA sequences in most promoters of *E. coli* genes are considered to be critical in specifying the initiation of transcription. These consensus sequences are generally called as –35 sequence and –10 sequence as they are centered at 35 and 10 base pairs upstream from the base pair at which transcription starts. The –10 consensus sequence in the non-template strand is **TATAAT** and the –35 consensus sequence is **TTGACA**. The sigma subunit initially recognizes and binds to the –35 sequence and this sequence is sometimes called the recognition sequence. The AT-rich –10 sequence facilitates the localized unwinding of DNA, which is an essential prerequisite to the synthesis of a new RNA chain. The formation of closed binary complex induces strand separation in the stretch of DNA duplex bound to the enzyme. The melted region is < 17 bp long and begins from the midpoint of the –10 hexamer consensus sequence of the promoter, and extends just beyond the startpoint so that the latter becomes available for transcription initiation. The complex of the melted DNA with the enzyme is known as the open binary complex. Once the template strand becomes available, the enzyme begins to incorporate the RNA nucleotides beginning at the startpoint. As in the case of DNA replication, the RNA synthesis also proceeds in the $5' \rightarrow 3'$ direction and the template DNA strand must be oriented as $3' \rightarrow 5'$.

9.3.2 Elongation of the RNA Chain

RNA synthesis takes place in a region of the DNA that has denatured to form a transcription bubble. Once 8 or 9 RNA nucleotides have been linked together, the sigma factor dissociates from the RNA polymerase core enzyme and can be used again in other transcription initiation reactions. The core enzyme completes the transcription of the gene. The covalent extension of RNA chains takes place within the transcription bubble, a locally unwound segment of DNA. The RNA polymerase molecule contains both DNA unwinding and DNA rewinding activities. RNA polymerase continuously unwinds the DNA double helix ahead of the polymerization site and rewinds the complementary DNA strands behind the polymerization site as it moves along the double helix. In *E. coli*, the average length of a transcription bubble is 18 nucleotide pairs, and about 40 ribonucleotides moves along the DNA molecule. Within the untwisted region, some bases of RNA are bound to the DNA in a temporary RNA-DNA hybrid while the rest of the RNA is displaced away from the DNA.

9.3.3 Termination of RNA Chain

Termination of the transcription of a prokaryotic gene is signaled by controlling elements called terminators. When the core enzyme reaches the terminator site, there is no further addition of ribonucleotides to the RNA chain. The RNA molecule dissociates from the template DNA strand, and the core enzyme frees itself from the template DNA molecule. The DNA strands in the melted region reassociate to form normal double helix and the process of transcription catalyzed by this polymerase molecules comes to an end. The released core enzyme will soon bind to a segment of DNA irrespective of the base sequence, and would begin the search for a promoter once the sigma factor binds to it. Basing on the requirement of a polypeptide or protein namely ‘rho-factor’, the terminators are divided into two groups – (1) simple or rho-independent terminators, and (2) rho-dependent terminators or type II terminators.

Rho-independent terminators consist of sequences with twofold symmetry that are bout 16 to 20 base pairs upstream of the transcription termination point. A sequence with twofold symmetry is one that is approximately self-complementary about its center, that is, one half of the sequence is complementary to the other half. Thus, the transcript of the region with twofold symmetry can form a hairpin loop. This sequence is followed by a string of about four to eight AT base pairs giving rise to a series of Us in the RNA transcript which is just upstream of the transcription termination point. The combination of the hairpin loop followed by the string of Us in the RNA leads to transcription termination. the rapid formation of the hairpin loop destabilizes the RNA-DNA hybrid in the terminator region, which in turn causes the release of the RNA and transcription termination.

Rho-dependent terminators lack the AT string found in rho-independent terminators, and many cannot form hairpin structures. The rho-factor is a protein with two domains – one domain binds to RNA, and the other domain binds to ATP. For transcription termination, rho-factor first binds to ATP, and is activated by it. The activated rho then binds to the RNA transcript at some recognition site located upstream of the termination site. It then moves along the RNA till it catches the RNA polymerase. When the rho-factor catches the RNA polymerase at the terminator site, it leads to unwinding of the RNA-DNA hybrid and interacts with the RNA polymerase itself which leads to the release of the RNA transcript, the RNA polymerase and the rho-factor from the DNA template bringing the termination of transcription.

9.4 Translation

The translation process or the protein synthesis can be divided into three stages – (1) polypeptide chain initiation, (2) chain elongation, and (3) chain termination. the main features of the initiation step are binding of mRNA to the ribosome, selection of the initiation codon, and binding of the charged tRNA bearing the first amino acid (fMet or Met). In the elongation stage there are two processes – joining together two amino acids by peptide formation and moving the mRNA and the ribosome with respect to one another in order that each codon can be translated successively. This movement is called translocation. In the termination stage the completed protein is dissociated from the synthetic machinery and the ribosomes are released to begin another cycle of synthesis.

9.4.1 Initiation of polypeptide chain

Initiation comprises all the events that precede the formation of the first peptide bond. In prokaryotes, the initiation complex is formed when the initiation codon, AUG, of the mRNA binds to the P site of the 30S subunit of a ribosome, and a $tRNA_f^{met}$ molecule carrying formylmethionine base pairs with the codon AUG. In *E. coli*, the formation of initiation complex requires a number of initiation factors, viz., GTP, and the factors IF1, IF2, and IF3. In addition to the AUG initiation codon, the initiation of protein synthesis in prokaryotes requires other information coded in the base sequence of an mRNA molecule upstream to the initiation codon. These sequences serve to align the ribosome on the message in the proper reading frame so that polypeptide synthesis can proceed correctly. Most of these binding sequences are purine-rich and are about 8 to 12

nucleotides upstream from the initiation codon. Evidence from the work of John Shine and Lynn Dalgarno indicates that this purine-rich sequence, and other nucleotides in this region, are complementary to a pyrimidine rich region at the 3′ end of 16S rRNA. The mRNA region that binds in this way is known as *Shine-Dalgarno* sequence. The formation of complementary base pairs between the mRNA and 16S rRNA allows the ribosome to locate the true sequence in the mRNA for the initiation of protein synthesis.

In *E. coli*, translation commences with the formation of a 30S initiation complex which involves the recognition of the AUG initiation codon by the small ribosome subunit. At the beginning of this process, three protein initiation factors, IF1, IF2, and IF3, are bound to the 30S ribosomal subunit along with a molecule of GTP. The fMet-tRNA.fMet and the mRNA then attach to the 30S-IF-GTP complex to form the 30S initiation complex. IF3 is released as a result of this process. Next, the 50S subunit binds, leading to GTP hydrolysis and the release of IF1 and IF2. The final complex is called the 70S initiation complex (Fig. 9.2). All three IF molecules are recycled for use in other initiation reactions. the 70S ribosome has two binding sites for aminoacyl-tRNA, the peptidyl (P), and aminoacyl (A) sites and the fMet-tRNA.fMet is bound to the mRNA in the P site.

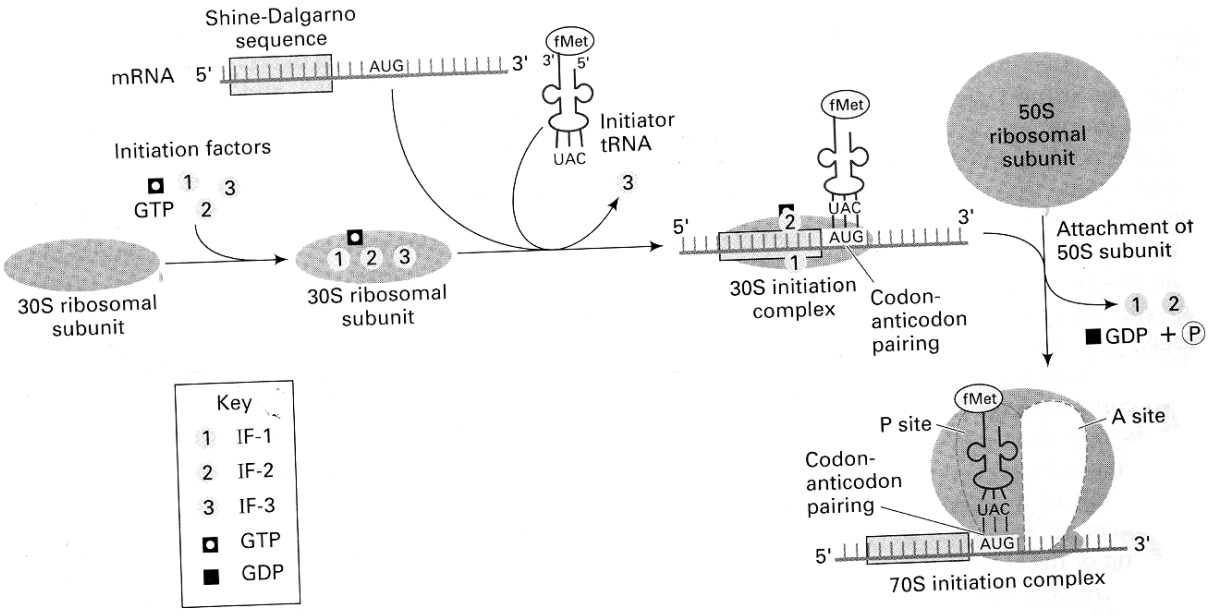


Fig. 9.2 Formation of final 70S initiation complex

9.4.2 Elongation of polypeptide chain

After initiation, the elongation phase of translation (Fig. 9.3) begins. This phase has three steps –

(1) the binding of aminoacyl-tRNA (charged tRNA) to the ribosome

(2) the formation of a peptide bond

(3) the movement (translocation) of the ribosome along the mRNA, one codon at a time

At the start of the elongation phase in prokaryotes, the fMet-tRNA.fMet is hydrogen-bonded to the AUG initiation codon in the peptidyl (P) site of the ribosome in the 70S initiation complex. FMet-tRNA.fMet is the only tRNA to bind before the complete ribosome is formed. The orientation of this tRNA-codon complex exposes the next codon in the mRNA in the aminoacyl (A) site. Next, the appropriate aminoacyl-tRNA binds to the exposed mRNA codon in the A site. This aminoacyl-tRNA is brought to the ribosome complexed with the protein elongation factor EF-Tu and a molecule of GTP. When the aminoacyl-tRNA binds to the codon in the A site, the GTP is hydrolyzed and EF-Tu is released bound to the GDP produced. This EF-Tu is recycled. First, a second elongation factor, EF-Ts, binds to EF-Tu and displaces the GDP. Next, GTP binds to the EF-Tu—EF-Ts complex to produce an Ef-Tu-GTP complex simultaneously with the release of EF-Ts. The aminoacyl-tRNA binds to the EF-Tu-GTP, and that complex can then bind to the A site in the ribosome when the appropriate codon is exposed.

When ribosome maintains the two aminoacyl-tRNAs in the correct positions (one in P site and another in A site), a peptide bond formation occurs between the two amino acids. The peptide bond formation occurs in two steps. The first step is the breakage of the bond between the carboxyl group of the amino acid and the tRNA in the P site. The second step is the formation of the peptide bond between the now-freed amino acid and the amino acid attached to the tRNA in the A site. This reaction is catalyzed by peptidyl transferase. Once the peptide bond has formed, a tRNA without an attached amino acid (an uncharged tRNA) is left in the P site. The tRNA in the A site, now called peptidyl-tRNA, has the first two amino acids of the polypeptide chain attached to it.

The last step in the elongation cycle is ‘translocation’. Once the peptide bond is formed and the growing polypeptide chain is on the tRNA in the A site, the ribosome moves one codon along the mRNA toward the 3′ end. In prokaryotes, the translocation requires the activity of another protein elongation factor, EF-G. An EF-G—GTP complex binds to the ribosome, and translocation then takes place along with movement of the uncharged tRNA away from the P site. The uncharged tRNA moves from the P site and then binds to the Exit site or E site. After translocation, the EF-G is then released in a reaction requiring GTP hydrolysis; EF-G can then be reused. During the translocation step the peptidyl-tRNA remains attached to its codon on the mRNA. And since the ribosome has moved, the peptidyl-tRNA is now located in the P site. The exact mechanism for the physical translocation of the ribosome is not known. After the translocation is completed, the A site is vacant. An aminoacyl-tRNA with the correct anticodon binds to the newly exposed codon on the mRNA in the A site. The whole process is repeated until translation terminates at a stop codon.

In both prokaryotes and eukaryotes, once the ribosome moves away from the initiation site on the mRNA, the initiation site is open for another initiation event to occur. Thus many ribosomes may simultaneously be translating each mRNA. The complex between an mRNA molecule and all the ribosomes that are translating it simultaneously is called a polyribosome or polysome. An average mRNA may have eight to ten ribosomes synthesizing protein from it. Simultaneous translation enables a large amount of protein to be produced from each mRNA molecule.

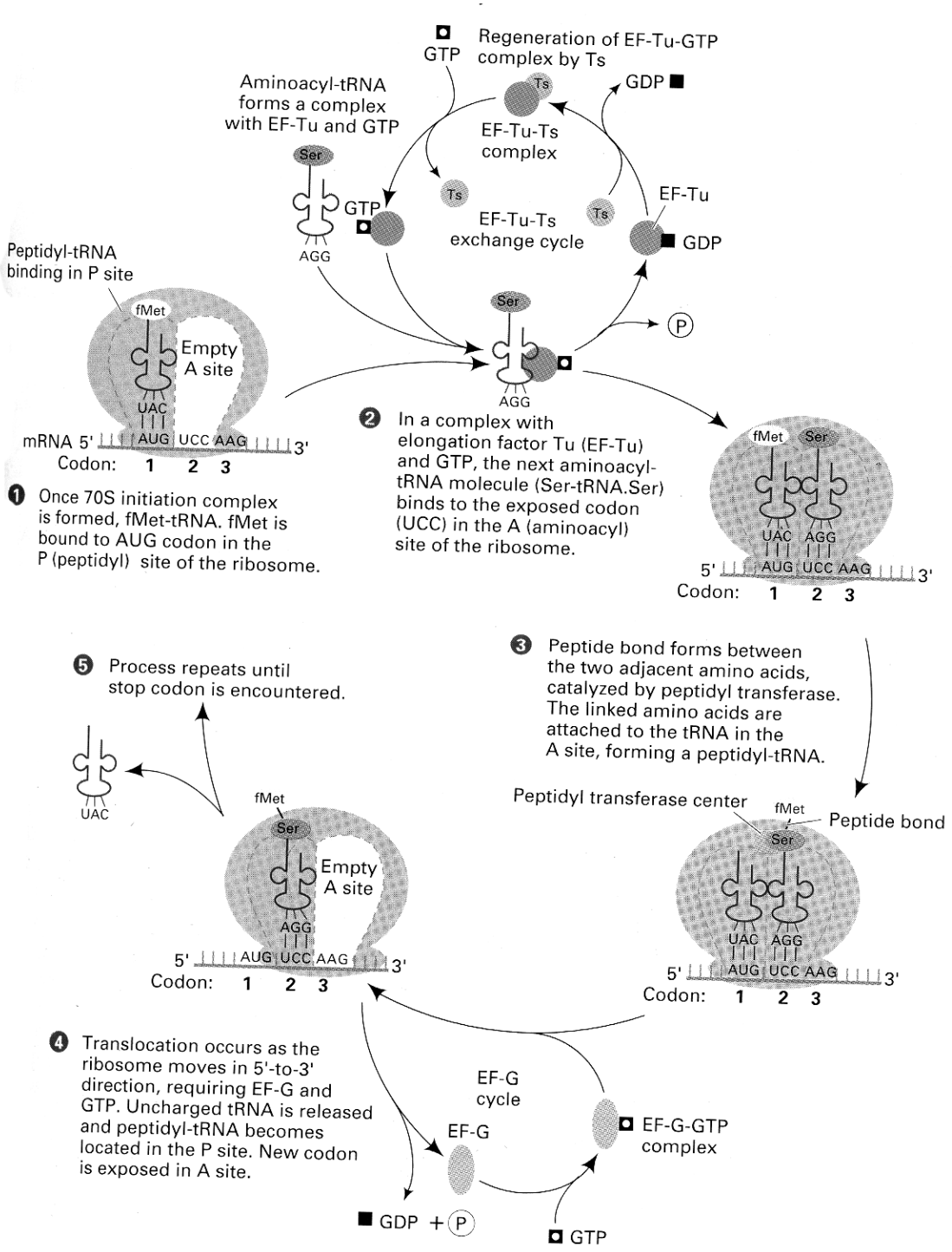


Fig. 9.3 Elongation and Translocation steps during Translation

9.4.3 Termination of Translation

Elongation continues until the polypeptide coded for in the mRNA is completed. The end of a polypeptide chain is signaled by one of the three stop codons, UAG, UAA, and UGA, which are the same in both prokaryotes and eukaryotes. The stop codons do not code for any amino acid, and so no tRNAs in the cell have anticodons for them. The ribosome recognizes a chain termination codon only with the help of proteins called termination factors or release factors (RF), which read the chain termination codons, and then initiate a series of specific termination events.

The *E. coli* has three RFs – RF1, RF2, and RF3 – and each is a single polypeptide. Factor RF1 recognizes UAA and UAG, while RF2 recognizes UAA and UGA. Thus these two release factors have overlapping specificity in codon recognition. Factor RF3, which does not recognize any of the stop codons, stimulates the termination events. The specific termination events triggered by the release factors are : (1) release of the polypeptide from the tRNA in the P site of the ribosome in a reaction catalyzed by peptidyl transferase, and then (2) release of the tRNA from the ribosome and dissociation of the two ribosomal subunits from the mRNA. The initiating amino acid, formylmethionine is usually cleaved from the completed polypeptide.

9.5 Comparison of protein synthesis between prokaryotes and eukaryotes

Characteristic	Prokaryotes	Eukaryotes
RNA polymerase		
Types	One in each species	Three; RNA polymerase I,II and III
Molecular weight	4,80,000 daltons	5,00,000 daltons
Subunits	$\alpha_2\beta\beta'\sigma$	2 large subunits + < 10 smaller subunits
Transcription factors	Not known	Many different types of transcription factors
Promoter	A simpler and relatively smaller sequence	A relatively larger sequence and has variable modules
Enhancer	A part of the promoter may act in a manner similar to enhancer	Present at variable distances from the promoter
Transcription initiation	The holoenzyme binds to promoter and initiates transcription	Transcription factors first bind to promoter, then RNA polymerase associates with them and initiates
Transcription complex		
Composition	Core RNA polymerase + σ factor	RNA polymerase + transcription factors
Separation of	σ factor dissociates from the	Transcription factors dissociate

components	core enzyme after initiation	only when transcription is complete
MRNA	Polycistronic RNA transcripts In general, there is no processing after transcription. Very short life of 2-5 minutes; degradation may begin even before transcription is complete. Poly –(A) tail not found. 5′ cap is not present. Shine-Dalgarno consensus present at the 5′ end.	Monocistronic RNA transcripts Mature mRNA molecules are processed from longer precursor hnRNA molecues. Much longer life of ~ 6hr or more; mRNA in seeds may survive more than 100 years Most mRNA's have 100-200 adenine nucleotides at the 3′ -end; producing a poly-(A) tail Partially methylated 5′ cap present. Shine-Dalgarno consensus is absent.
rRNA and ribosomes	Ribosomes of 70S; dissociate into 30S and 50S subunits. Ribosomes free in the cytoplasm. rRNA of 16S (30S subunit), 23S and 5S (50S subunit)	Ribosomes of 80S; dissociates into 40S and 60S subunits. Ribosomes attached to endoplasmic reticulum as well as free in the cytoplasm. rRNA is 18 S (40S subunit), 28S, 7.8 S and 5 S (60S subunit)
Translation	Translation often simultaneous with transcription. Translation occurs while mRNA is attached to the chromosome. Translation initiation involves base-pairing between 16S rRNA and mRNA in the Shine-Dalgarno consensus sequence. Formylmethionine is incorporated by the initiation codon AUG. Initiation site of mRNA first binds with the 30S subunit of ribosome which then accepts formylmethionly-$tRNA_f^{met}$ complex.	Translation is not simultaneous with transcription. mRNA has to be processed and transported out of the nucleus as translation occurs mostly in the cytoplasm. Translation initiation is based on the recognition of the 5′ cap of mRNA by the 40S subunit of ribosomes and by some proteins. Methionine is incorporated at the initiation point by the codon AUG. Methionyl-tRNAfmet first binds to the 40S subunit, to which an mRNA molecule is then attached to form the initiation complex.

9.6 Summary

The gene expression refers to the transcription and translation of mRNA which leads to the synthesis of protein. When a gene is expressed, the DNA base pair sequence is transcribed into the base sequence of an RNA molecule. When a gene is transcribed, only one of the two DNA strands is copied. The direction of RNA synthesis is 5′ to 3′ and the reaction is catalyzed by RNA polymerase. The promoter sequence of the gene specify where the transcription is to begin, and terminator sequence specify where transcription is to stop. In bacteria, there is only one type of RNA polymerase. The promoter is recognized by a complex between the RNA polymerase core enzyme and a protein factor called sigma. Once transcription is initiated correctly, the sigma factor dissociates from the enzyme and is reused in other transcription initiation events. Termination occurs in one of the two ways, rho-dependant and rho-independent termination processes.

Protein synthesis occurs on ribosomes, where the genetic message encoded in mRNA is translated. Amino acids are brought to the ribosome on charged tRNA molecules. Each tRNA has an anticodon, which binds specifically to a codon in the mRNA. As a result, the correct amino acid sequence is achieved by (1) the specific binding of each amino acid to its own specific tRNA, and (2) the specific binding between the codon of the mRNA and the complementary anticodon in the tRNA. The initiator codon for the start of the protein is AUG. In prokaryotes, the sequence upstream to the AUG codon required for the small ribosomal subunit binding is the Shine-Dalgarno sequence. The initiation of protein synthesis requires the initiation factors. During elongation phase, the polypeptide chain is elongated one amino acid at a time. This occurs simultaneously with the movement of the ribosome towards the 3′ end of the mRNA one codon at a time. Protein factors called elongation factors play important catalytic roles. The signal for polypeptide chain growth to stop is the presence of a chain terminating codon (UAG, UAA, or UGA) in the mRNA. Specific protein factors called release factors read the stop codon and initiate the events characteristic of protein synthesis termination.

9.7 Model questions

1. Write an essay on the protein synthesis in prokaryotes.
2. Explain the transcription process in prokaryotes.
3. Give an account on the translation process in prokaryotes.
4. Short questions
    a. Bacterial RNA polymerase
    b. Comparison of protein synthesis between prokaryotes and eukaryotes
    c. 30S and 70S initiation complexes
    d. Types of transcription termination

9.8 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 10

REGULATION OF GENE EXPRESSION - Lac OPERON

10.0 Objective

The main intention of this lesson plan is to enlighten the student regarding the regulation of *lac* genes at the transcription level that encode the enzymes required for the lactose utilization in *E. coli*.

10.1 Introduction

Microorganisms exhibit remarkable capacities to adapt to diverse environmental conditions. This adaptability depends in part on their ability to turn on and turn off the expression of specific sets of genes in response to changes in the environment. Most microorganisms, exhibit a striking ability to regulate the expression of specific genes in response to environmental signals. The expression of particular genes is turned on when the products of these genes are needed for growth. Their expression is turned off when the gene products are no longer needed. Clearly, the ability of an organism to regulate gene expression in this way increases its ability to reproduce under a variety of environmental conditions. The synthesis of gene transcripts and translation products requires the expenditure of considerable energy. By turning off the expression of genes when their products are not needed, an organism can avoid wasting energy and can utilize the conserved energy to synthesize products that maximize growth rate. Gene expression in prokaryotes is regulated at several different levels: transcription, mRNA processing, mRNA turnover, translation, and enzyme function. However, the regulatory mechanisms with the largest effects on phenotype act at the level of transcription.

Certain gene products, such as tRNA molecules, rRNA molecules, ribosomal proteins, RNA polymerase subunits, and enzymes catalyzing metabolic processes that are frequently referred to as cellular 'housekeeping' functions are essential components of almost all living cells. Genes that specify products of this type are continually being expressed in most cells. Such genes are said to be expressed constitutively and are referred to as 'constitutive genes' or 'housekeeping genes'. Other gene products are needed for cell growth only under certain environmental conditions. Constitutive synthesis of such gene products would be wasteful, using energy that could otherwise be utilized for more rapid growth and reproduction. A change in the set of gene products synthesized in a prokaryotic cell involves regulatory mechanisms that control gene expression. Genes whose activity is controlled in response to the needs of a cell or organism are called ' regulated genes'. In bacterial systems in which several enzymes act in sequence in a single metabolic pathway, it is often the case that either all of these enzymes are present or all are absent. This phenomenon is called coordinate regulation. This results from control of the synthesis of a single polycistronic mRNA that encodes all of the gene products.

10.2 Operon concept

Induction and repression of gene expression can be accomplished by the same mechanism, with one relatively minor modification. Based on the results of their studies of genetic mutations affecting the regulation of the synthesis of the lactose utilization enzymes, Francois Jacob and Jacques Monod proposed a classical model to explain the regulation of genes encoding the enzymes required for lactose utilization in *E. coli*. This model of mechanism is called 'operon model'. For their proposal, Jacob and Monod received the Nobel Prize in 1965. By definition, an operon is a cluster of genes, the expressions of which are regulated together by operator-regulator protein interactions, plus the operator region itself and the promoter. According to the proposal, the transcription of one or a set of contiguous structural genes is regulated by two controlling elements. One of these elements, the 'regulator or repressor gene', encodes a protein called the 'repressor'. Under the appropriate conditions, the repressor binds to the second element, the operator.

When the repressor binds to operator, transcription of the structural genes cannot occur. Transcription is initiated at RNA polymerase-binding sites called promoters located just upstream from the structural genes. Operator regions are usually located between the promoters and the structural genes that they regulate. Whether the repressor will bind to the operator and turn off the transcription of the structural genes in an operon is determined by the presence or absence of effector molecules in the environment. These effector molecules are usually small molecules such as amino acids, sugars, and similar metabolites. In the case of inducible operons, the effector molecules are inducers. Those active on repressible operons are called corepressors. These effector molecules, inducers and corepressors, are bound by the repressors and cause changes in the three-dimensional structures of the repressors. Conformations changes in protein structure resulting from the binding of small molecules are called allosteric transitions. In the case of repressors, the allosteric changes caused by the binding of effector molecules usually alter their ability to bind to operator regions. Inducible operons and repressible operons can be distinguished from one another

by determining whether the naked repressor or the repressor-effector molecule complex is active in binding to the operator.

In the case of inducible operon, the free repressor binds to the operator, turning off transcription. When the effector molecule (the inducer) is present, it is bound by the repressor, causing the repressor to be released from the operator, that is the repressor-inducer complex cannot bind to the operator. Thus the addition of inducer turns on or induces the transcription of the structural genes in the operon. In the case of a repressible operon, the situation is just reversed. The free repressor cannot bind to the operator. Only the repressor-effector molecule (co-repressor) complex is active in binding to the operator. Thus transcription of the structural genes in a repressible operon is turned on in the absence of and turned off in the presence of the effector molecule. Except for this difference in the operator-binding behaviour of the free repressor and the repressor-effector molecule complex, inducible and repressible operons are identical. They operate by exactly the same mechanism.

10.3 Types of Regulation of Transcription

There are several common patterns of regulation of transcription. These depend on the type of metabolic activity of the system being regulated. In a catabolic or degradative system the concentration of the substrate of an enzyme early in the pathway often determines whether the enzymes in the pathway are synthesized. Enzymes that are involved in the catabolic or degradative pathways such as in lactose, galactose, or arabinose utilization, are characteristically inducible. This induction occurs at the level of transcription. Induction alters the rate of synthesis of enzymes, not the activity of existing enzyme molecules. Induction is different from the enzyme activation, which occurs when the binding of a small molecule to an enzyme increases the activity of the enzyme, but does not affect its rate of synthesis.

In contrast, in an anabolic or biosynthetic pathway the final product is the regulatory substance. Bacteria can synthesize most of the organic molecules, such as amino acids, purines, pyrimidines, and vitamins, required for growth. For example, the *E. coli* genome contains five genes encoding enzymes that catalyze steps in the biosynthesis of tryptophan. These five genes must be expressed in *E. coli* cells growing in an environment devoid of tryptophan in order to provide adequate amounts of this amino acid for ongoing protein synthesis. And a regulatory mechanism has evolved in E. coli that turns off the synthesis of the tryptophan biosynthetic enzymes when external tryptophan is available. This process of turning off the expression of genes in response to an environmental signal is called repression. A gene whose expression has been turned off in this way is said to repressed; when its expression is turned on, the gene is said to be derepressed and the response is called derepression. Enzymes that are components of anabolic or biosynthetic pathways often are repressible. Like induction, repression also occurs at the level of transcription. Repression should not be confused with feed-back inhibition, which occurs when the product of a biosynthetic pathway binds to and inhibits the activity of the first enzyme in the pathway, but does not effect the synthesis of the enzyme.

The molecular mechanisms for each of the regulatory patterns vary quite widely but usually fall in one of the two major categories – negative regulation and positive regulation. In

negative regulation, an inhibitor is present in the cell that keeps the transcription turned off. In this mechanism, the association of a specific protein called repressor with the operator DNA prevents the transcription of the operon. The transcription is believed to be prevented due to a change in the configuration of the operator DNA so that RNA polymerase is unable to perform its function. Since the regulation of gene action in such a system is achieved by the prevention of transcription by a repressor, it is known as negative control. In positive regulation, an effector molecule which may be a protein, a small molecule, or a molecular complex activates a promoter. In positive control of transcription, association of a specific protein, termed as activator, to a segment of DNA in the promoter gene or to RNA polymerase enables the transcription of the operon. The promontory effect of the activator is believed to be due to its effect on DNA configuration in the transcription initiator region, which then becomes more favourable for the action of RNA polymerase. Negative and positive regulation are not mutually exclusive, and some systems are both positively and negatively regulated and need two regulators.

10.4 Lactose utilization as a carbon source in *E. coli*

E. coli is able to grow in a simple medium containing salts including a nitrogen source and a carbon source such as glucose. These chemicals provide molecules that can be manipulated by the enzymatic machinery of the cell to produce everything the cell needs to grow and reproduce, such as nucleic acids, proteins, and lipids. The energy for these biochemical reactions in the cell comes from the metabolism of glucose, a process that is of central importance to the functioning of a bacterial cell and of cells of all organisms. The enzymes required for glucose metabolism are coded for by constitutive genes.

If lactose, or one of several other sugars, is provided to *E. coli* as a carbon source instead of glucose, a number of enzymes are rapidly synthesized; these enzymes are needed for the metabolism of this particular sugar. The enzymes are synthesized because the genes that code for them become actively transcribed in the presence of the sugar; these same genes are inactive if the sugar is absent. In other words, the genes are regulated genes whose products are needed only at certain times. Biochemical analysis had shown that when the lactose, a disaccharide, is the sole carbon source in the growth medium, three proteins are synthesized –

1. β-galactosidase: This enzyme has two functions (a) It catalyzes the breakdown of lactose into its two component monosaccharides, glucose and galactose; and (b) It catalyzes the isomerization of lactose to ‘allolactose’ a compound important in the regulation of expression of the lactose utilization genes.

2. β-galactoside permease: This is also called as M protein. This protein is found in the *E. coli* cytoplasmic membrane and is needed for the active transport of lactose from the growth medium into the cell.

3. β-galactoside transacetylase: The function of this enzyme is poorly understood.

In a wild-type *E. coli* that is growing in a medium containing glucose but no lactose, only a few molecules of each of the above three enzymes are produced, indicating a low level of

expression of the three genes that code for the proteins. For example, only an average of three molecules of β-galactosidase is present in the cell under these conditions. If lactose but no glucose is present in the growth medium, the number of molecules of each of the three enzymes increases coordinately about a thousand fold. This occurs because the three essentially inactive genes are now being actively transcribed and the resulting mRNA translated. This process is called coordinate induction. The inducer molecule is directly responsible for the increased production of the three enzymes is actually allolactose, not lactose. Allolactose is produced from lactose by β-galactosidase. Further, the mRNAs for the enzymes have a relatively short half-life, so the transcripts must be made continually in order for the enzymes to be produced. So, when lactose is no longer present, transcription of the three genes is stopped and any mRNAs already present are broken down thereby resulting in a drastic reduction in the amounts of the three proteins in the cell.

10.5 The lactose operon in *E. coli*: Induction and Catabolite Repression

Jacob and Monod proposed the operon model largely as a result of their studies of the lactose (lac) operon in *E. coli*. More is known about this operon than any other operon. The lac operon (Fig. 10.1) contains a promoter (P), an operator (O), and three structural gens, *lac*Z, *lac*Y, and *lac*A, encoding the enzymes β-galactosidase, β-galactoside permease, and β-galactoside transacetylase, respectively. The β-galactoside permease pumps lactose into the cell, where β-galactosidase cleaves it into glucose and galactose. The exact biological role of the transacetylase is not clearly understood.

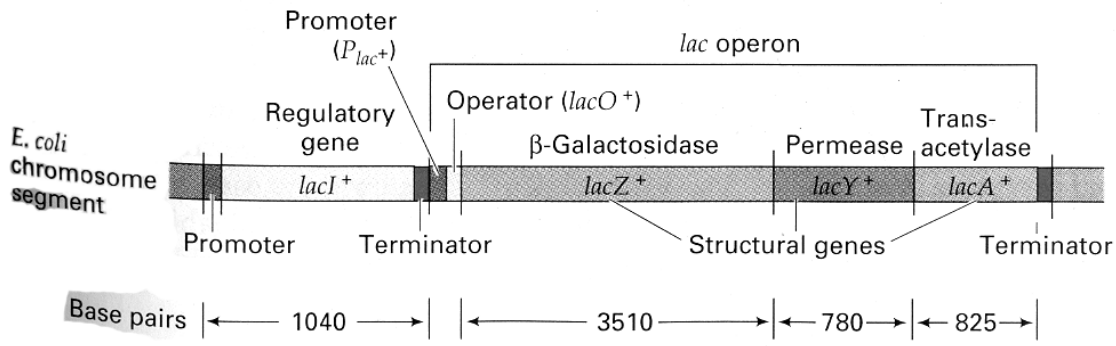


Fig. 10.1 Structure of lac operon

10.5.1 Induction

The lac operon is an inducible operon; the *lac*Z, *lac*Y, and *lac*A genes are expressed only in the presence of lactose. The *lac* regulator gene, designated the *I* gene, encodes a repressor that is 360 amino acids long. The active form of the *lac* repressor is a tetramer containing four copies of the *I* gene product. In the absence of inducer, the repressor binds to the *lac* operator sequence, which, in turn prevents RNA polymerase from catalyzing the transcription of the three structural genes (Fig. 10.2). A few molecules of the *lac*Z, *lac*Y, and *lac*A gene products are synthesized in

the uninduced state, providing a low background level of enzyme activity. This background activity is essential for induction of the *lac* operon, because the inducer of the operon, 'allolactose', is derived from lactose in a reaction catalyzed by β-galactosidase. Once formed, allolactose is bound by the repressor, causing the release of the repressor from the operator (Fig. 10.2). In this way, allolactose induces the transcription of the *lac*Z, *lac*Y, and *lac*A structural genes. The *lac*I gene, the *lac* operator, and the *lac* promoter were all initially identified genetically by the isolation of mutant strains that exhibited altered expression of the *lac* operon genes. Mutations in the *I* gene and the operator frequently result in the constitutive synthesis of the *lac* gene products. These mutations are designated I^- and O^c, respectively.

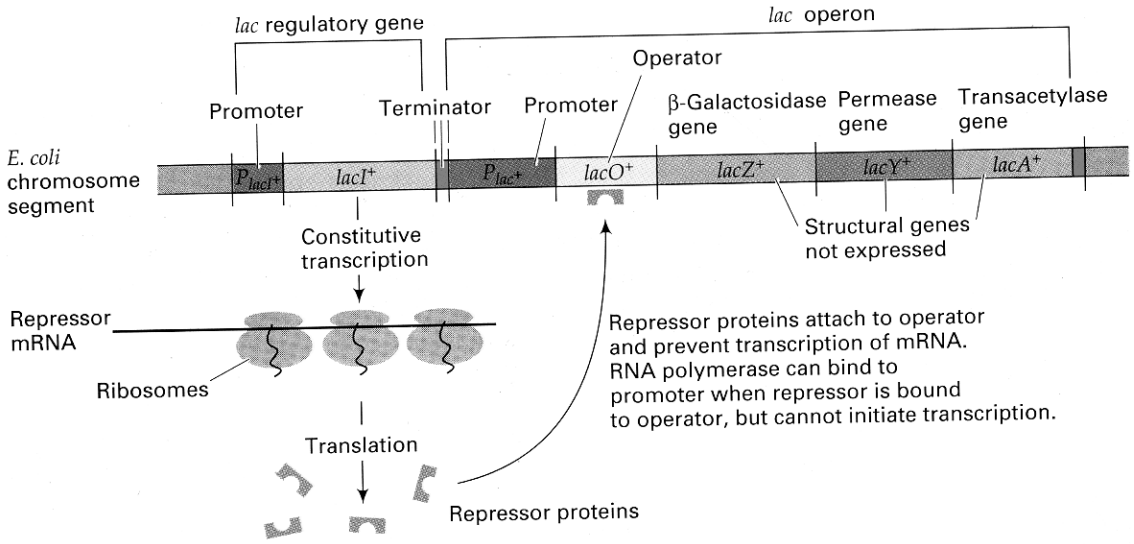


Fig. 10.2 Regulation of lac operon in the absence of lactose

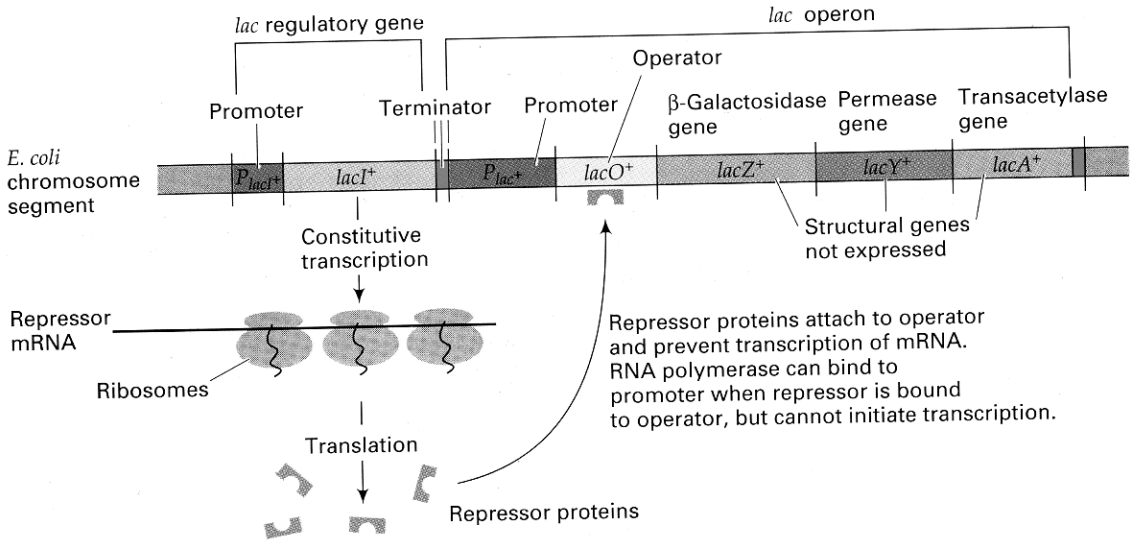


Fig. 10.2 Regulation of lac operon in the presence of inducer

10.5.2 Catabolite Repression

Jacob and Monod proposed the operon model to explain the induction of the biosynthesis of the enzymes involved in lactose utilization when this sugar is added to the medium in which *E. coli* cells are growing. However, the presence of glucose has long been known to prevent the induction of the *lac* operon, as well as other operons controlling enzymes involved in carbohydrate catabolism. This phenomenon, called 'catabolite repression' or the 'glucose effect'. This assures that glucose is metabolized when present, in preference to other, less efficient, energy sources.

The catabolite repression of the *lac* operon and several other operons is mediated by a regulatory protein called CAP (Catabolite Activator Protein) and a small effector molecule called cyclic AMP (adenosine-3′,5′ -phosphate; abbreviated cAMP). Because CAP binds cAMP when this mononucleotide is present at sufficient concentrations, it is sometimes called the cyclic AMP receptor protein. Infact, the *lac* promoter contains two separate binding sites – one for RNA polymerase and one for the CAP-cAMP complex. The CAP-cAMP complex must be present in its binding site in the lac promoter in order for the operon to be induced. The CAP-cAMP complex thus exerts positive control over the transcription of the *lac* operon. It has an effect exactly opposite to that of repressor binding to an operator. Although the precise mechanism by which CAP-cAMP stimulates RNA polymerase binding to the promoter is still uncertain, its positive control of the *lac* operon transcription is firmly established by the results of both in vivo and in vitro experiments. CAP is known to function as a dimer, thus, like the *lac* repressor, it is multimeric in its functional state.

Only the CAP-cAMP complex binds to the *lac* promoter; in the absence of cAMP, CAP does not bind. Thus cAMP acts as the effector molecule, determining the effect of CAP on *lac*

operon transcription. The intracellular cAMP concentration is sensitive to the presence or absence of glucose. High concentrations of glucose cause sharp decrease in the intracellular concentration of cAMP. How glucose controls the cAMP concentration is not clear. Perhaps glucose, or some metabolite that forms in the presence of sufficient concentrations of glucose, inhibits the activity of adenylcyclase, the enzyme that catalyzes the formation of cAMP from ATP. The presence of glucose results in a decrease in the intracellular concentrations of cAMP. In the presence of a low concentration of cAMP, CAP cannot bind to the *lac* operon promoter. In turn, RNA ploymerase cannot bind efficiently to the *lac* promoter in the absence of bound CAP-cAMP. The overall result of the positive control of transcription of the *lac* operon by the CAP-cAMP complex is that in the presence of glucose *lac* operon transcription never exceeds 2 percent of the induced rate observed in the absence of glucose. By similar mechanisms, CAP and cAMP keep the arabinose and galactose operons of *E. coli* from being induced in the presence of glucose.

10.6 Summary

From the studies of the regulation of expression of the three lactose-utilizing genes in *E. coli* bacteria, Jacob and Monod developed a model that is the basis for the regulation of gene expression in a large number of bacterial and bacteriophage systems. The genes for the enzymes are contiguous in the chromosome and are adjacent to a controlling site, an operator, and a single promoter. This complex constitutes a transcriptional regulatory unit called an operon. A regulator gene, which may or may not be nearby, is associated with the operon. The addition of an appropriate substrate to the cell results in the coordinate induction of the operon's structural genes. With respect to the *lac* operon, induction occurs as lactose binds with a repressor protein that is encoded by the regulator gene, inactivating it and preventing it from binding to the operator. As a result, RNA polymerase can transcribe the three genes onto a single polygenic mRNA. As long as lactose is present, mRNA continues to be produced and the enzymes are made. When lactose is no longer present, the repressor protein is no longer inactivated, and it binds to the operator, thereby preventing RNA polymerase from transcribing the *lac* genes.

If both glucose and lactose are present in the medium, the lactose operon is not induced because glucose is the preferred energy source. This phenomenon is called catabolite repression and involves cellular levels of cyclic AMP. In the presence of lactose and absence of glucose, cAMP complexes with CAP to form a positive regulator needed for RNA polymerase to bind to the promoter. Addition of glucose results in a lowering of cAMP level so no CAP-cAMP complex is produced and, therefore, RNA polymerase cannot bind to the promoter and transcribe the *lac* genes.

10.7 Model Questions

1. Describe the mechanism of regulation of *lac* operon.
2. a) Negative control

   b) Glucose effect

   c) Lac operon

10.8 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 11

REGULATION OF GENE EXPRESSION – Trp OPERON

11.0 Objective

This lesson plan is aimed to give a clear idea to the student about the repressive mechanism of gene regulation in the case of tryptophan operon.

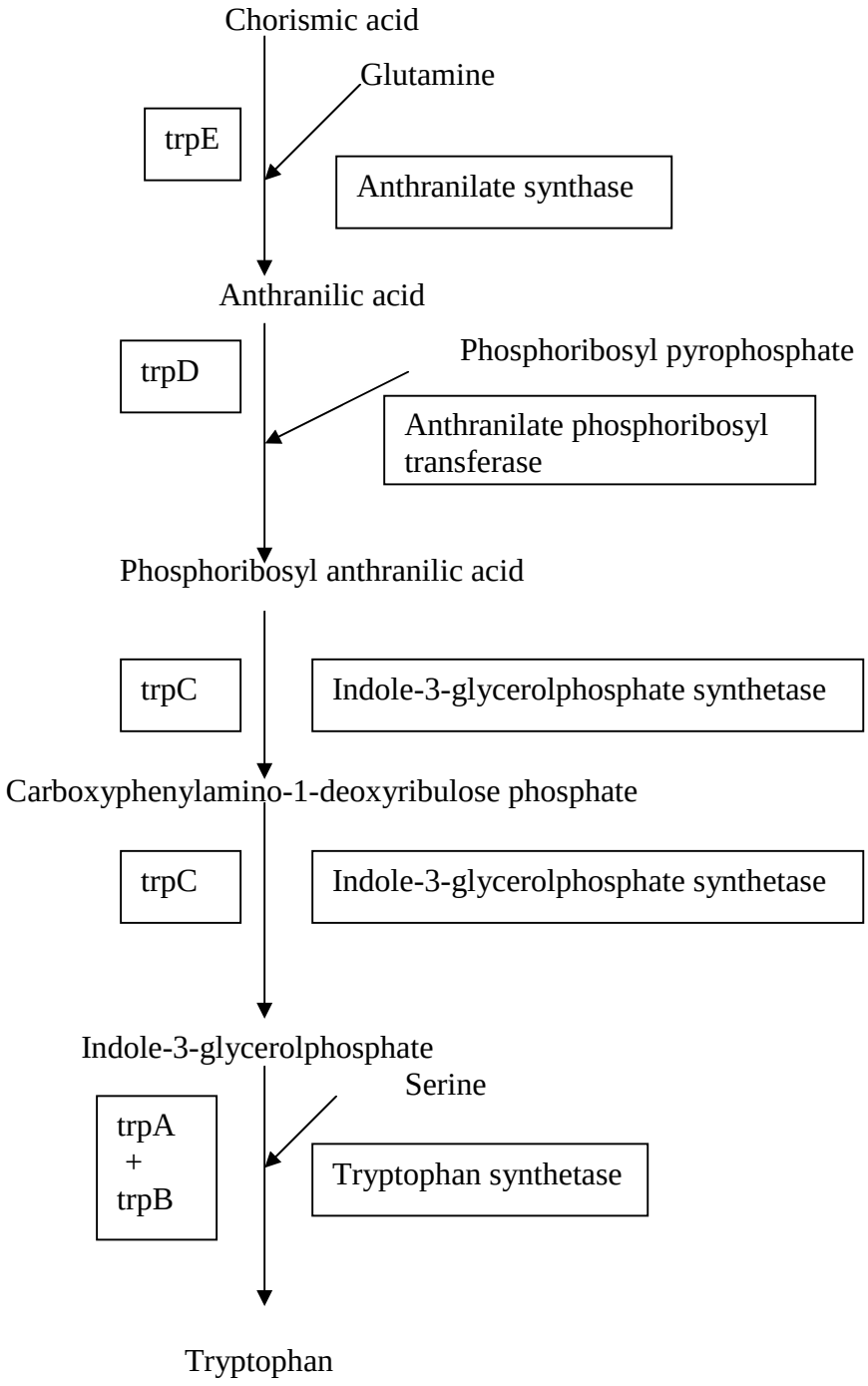
11.1 Introduction

Organisms live and reproduce in changing environments. One way that an organism can adjust to a new environment is to alter its gene activity so that gene products appropriate to the new conditions are synthesized. As a result, the organism is optimally adjusted to grow and reproduce in that environment. In a growth medium, all the necessary amino acids many not be present to enable bacteria to assemble proteins. If an amino acid is missing, a bacterium has certain operons and other gene systems that enable the bacterium to manufacture that amino acid so that it may grow and reproduce. Each step in the biosynthetic pathway through which amino acids are assembled is catalyzed by a specific enzyme coded by a specific gene or genes.

When an amino acid is present in the medium, the genes encoding the enzymes for that amino acid's biosynthetic pathway are tuned off. If an amino acid is not present in the medium, however, those must be turned on in order for the biosynthetic enzymes to be made. Unlike the *lac* operon, where gene activity is induced when lactose is added to the medium, in this case there is a repression of gene activity when an amino acid is added. So, the amino acid biosynthesis operons controlled in this way are the repressible operons. In general, repressible operons function to turn off anabolic or biosynthetic pathways when the end product is readily available. One repressible

operon in *E. coli* that has been extensively studied is the operon for the biosynthesis of the amino acid tryptophan (Trp). Although the regulation of the *trp* operon shows some basic similarities to the regulation of the classical *lac* operon, some intriguing differences also appear to be common among similar, repressible, amino acid biosynthesis operons in bacteria.

11.2 Biosynthetic Pathway of Tryptophan



The biosynthesis of typtophan occurs from chorismic acid. The entire biosynthetic process occurs in five enzymatic steps involving the enzymes encoded by the five structural genes of the *trp* operon. During the first step, the glutamine combines with chorismic acid to produce anthranilic acid. This step is catalyzed by the enzyme, anthranilate synthetase coded by *trp*E gene. In the second step, the anthranilic acid combines with phosphoribosyl pyrophosphate to yield phosphoribosyl anthranilic acid under the mediation of enzyme, anthranilate phosphoribosyl transferase coded by *trp*D. In the next two steps, phosphoribosyl anthranilic acid is first converted to carboxyphenylamino-1-deoxyribulose phosphate and then to indole-3-gycerolphosphate. Both the steps are catalyzed by the same enzyme, indole-3-glycerolphosphate synthetase coded by *trp*C gene. During the last step, indole-3-glycerolphosphate combines with serine to yield Tryptophan. This last step is mediated by the enzyme, tryptophan synthetase which is jointly coded by *trp*A and *trp*B genes.

11.3 Tryptophan operon in *E. coli*: Repression and Attenuation

The trp operon in *E. coli* controls the synthesis of the enzymes that catalyze the biosynthesis of the amino acid typtophan. The functions of the five structural genes and the adjacent regulatory sequences of the *trp* operon (11.1) have been analyzed in detail by Charles Yanofsky and colleagues. The five structural genes (A-E) encode five enzymes that convert chorismic acid to tryptophan in five steps. The genes encoding these enzymes in *E. coli* are adjacent to one another in the same order as their use in the biosynthetic pathway; they are translated from a single polycistronic mRNA and called *trp*E, *trp*D, *trp*C, *trp*B, and *trp*A. The *trp*E gene is the first one translated. The promoter and operator regions are closely integrated in the DNA and are upstream from the *trp*E gene. Between the promoter-operator region and *trp*E is a 162-base pair region called *trp*L, or the leader region. Within *trp*L, relatively close to *trp*E, is an attenuator site (att) that plays an important role in the regulation of the tryptophan operon. The repressor gene *trp*R is located very far form this gene cluster, like the repressor in the *gal* operon. The entire tryptophan operon is approximately 7,000 base pairs long. Transcription of the operon results in the production of a polygenic mRNA for the five structural genes. Each of these transcripts is translated to an equal extent. The expression of the *trp* operon is regulated at two levels – repression, which controls the initiation of transcription, and attenuation, which governs the frequency of premature transcript termination.

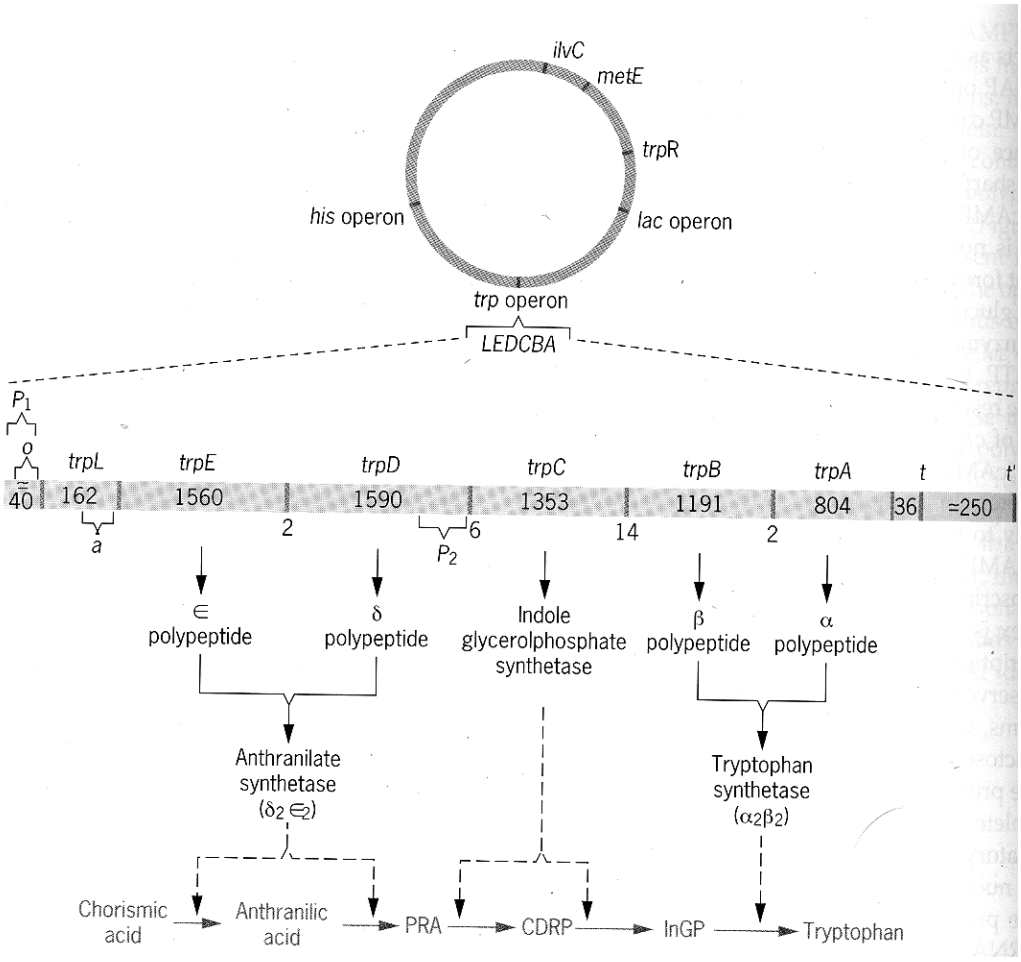


Fig 11.1 Structure of trp operon

11.3.1 Repression

The *trp* operon of *E. coli* is probably the best known repressible operon. The operator (O) region of the *trp* operon lies within the primary promoter (P_1) region. There is also a weak promoter (P_2) at the operator-distal end of the *trp*D gene. The P_2 promoter results in a somewhat increased basal level of transcrtiption of the *trp*C, *trp*B, and *trp*A genes. Two transcription termination sequences (t and t′) are located downstream from *trp*A. The *trp*L region specifies a 162-nucleotide long mRNA leader sequence. The product of the regulatory gene, *trp*R, is an aporepressor protein, which alone cannot bind to the operator. When tryptophan is abundant in the growth medium, it binds to the aporepressor and converts it to an active repressor. The active repressor binds to the operator and prevents the initiation of transcription of the *trp* operon protein-coding genes by RNA polymerase. As a result, the tryptophan biosynthesis enzymes are not produced. In the absence of tryptophan, RNA polymerase binds to the promoter region and transcribes the structural genes of the operon.

The rate of transcription of the *trp* operon in the derepressed state i.e. the absence of typtophan, is 70 times the rate that occurs in the repressed state i.e. in the presence of tryptophan.

In *trp*R mutants that lack functional repressor, the rate of synthesis of the tryptophan biosynthetic enzymes is till reduced about tenfold by the addition of tryptophan to the medium. This additional reduction in *trp* operon expression is caused by attenuation.

11.3.2 Attenuation

Deletions that remove part of the *trp*L region result in increased rates of expression of the *trp* operon. However, these deletions have no effect on the repressibility of the *trp* operon; that is, repression and derepression occur just as in *trp*L^+ strains. These results indicate that the synthesis of the trptophan biosynthetic enzymes is regulated at a second level by a mechanism that is independent of the state of repression or derepression of the *trp* operon and requires nucleotide sequences present in the *trp*L region of the *trp* operon. This second level of regulation of the *trp* operon is called attenuation, and the sequence within *trp*L that controls this phenomenon is called the attenuator. Attenuation occurs by control of the termination of transcription at a site near the end of the mRNA leader sequence. This premature termination of *trp* operon transcription occurs only in the presence of tryptophan-charged $tRNA^{Trp}$. When this premature termination or attenuation occurs, a truncated *trp* transcript of 140 nucleotides is produced.

The second regulatory mechanism is involved in the expression of the *trp* operon under conditions of tryptophan starvation or tryptophan limitation. Under severe tryptophan starvation, the *trp* genes are expressed maximally, while under less severe starvation conditions, the *trp* genes are expressed at less than maximal levels. This is accomplished by a mechanism that controls the ratio of the transcripts that include the five *trp* structural genes to those that are terminated before the structural genes.

The attenuator region has a nucleotide-pair sequence essentially identical to the transcription-termination signals found at the ends of most bacterial operons. These termination signals contain a G:C rich palindrome followed by several A:T base pairs. Transcription of these termination signals yields a nascent RNA with the potential to form a hydrogen-bonded hairpin structure followed by several Us. When a nascent transcript forms this hairpin structure, it is believed to cause a conformational change in the associated RNA polymerase, resulting in termination of transcription within the following, more weakly hydrogen-bonded region of DNA-RNA base pairing. The nucleotide sequence of the attenuator therefore explains its ability to terminate trp operon transcription prematurely.

In prokaryotes, the transcription and translation mechanisms are coupled, that is, ribosomes begin translating mRNAs while they are still being produced by transcription. Thus events that occur during translation may also affect transcription. The 162 nucleotide-long leader sequence of the *trp* operon mRNA contains sequences that can base pair to form alternate stem-and-loop or hairpin structures (Fig. 11.2). The four leader regions that can base pair to form these structures are – (1) nucleotides 59-67, (2) nucleotides 71-79, (3) nucleotides 110-121, and (4) nucleotides 126-134. The nucleotide sequences of these four regions are such that region 1 can base pair with region 2, region 2 can base pair with region 3, and region 3 can base pair with region 4. Region 2 can base pair with either region 1 or region 3, but, obviously, it can pair with only one of these regions at any given time. Thus there are two possible secondary structures for the trp leader sequence – (1) region 1 paired with region 2 and region 3 paired with region 4 or (2) region 2 paired with region 3, leaving regions 1 and 4 unpaired. The pairing of regions 3 and 4 produces the

transcription-termination hairpin. If region 3 is base paired with region2, it cannot pair with region 4, and the transcription-termination hairpin cannot form. The presence or absence of tryptophan determines the formation of these alternative structures.

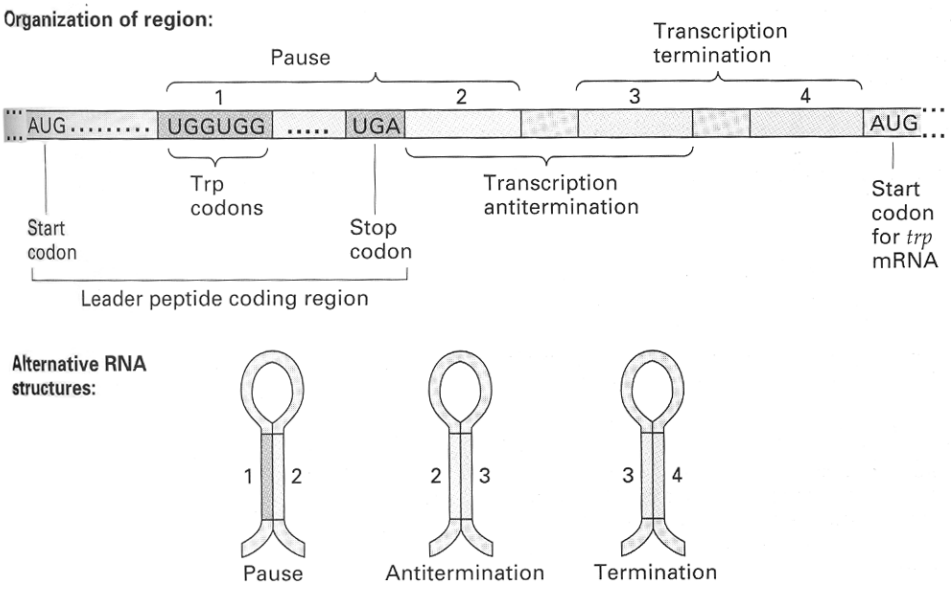


Fig. 11.2 Leader region of trp operon and stem and loop structures

The leader sequence contains an AUG translation-initiation codon, followed by 13 condons for amino acids, followed, in turn, by a UGA translation-termination codon. In addition, the trp leader sequence contains an efficient ribosome-binding site located in the appropriate position for the initiation of translation at the leader AUG initiation codon. All the available evidence indicates that a 14 amino acid 'leader peptide' is synthesized. But the synthesis of this putative leader peptide in vivo has not been verified directly, may be because of their unstable nature. The leader peptide contains two contiguous tryptophan residues. The two Trp codons are positioned such that in the absence of tryptophan, the ribosome will stall before it encounters the base paired structure formed by leader regions 2 and 3 (11.3). Because the pairing of regions 2 and 3 precludes the formation of the transcription-termination hairpin by the base pairing of regions 3 and 4, transcription will continue past the attenuator into the trpE gene in the absence of tryptophan.

In the presence of tryptophan, the ribosome can translate past the Trp codons to the leader-peptide termination codon. In the process, it will disrupt the base pairing between leader regions 2 and 3. This disruption leaves region 3 free to pair with region 4, forming the transcription-termination hairpin. Thus, in the presence of typtophan, transcription frequently terminates at the attenuator, reducing the amount of mRNA for the trp structural genes. The transcription of the *trp* operon can be regulated over a range of almost 700-fold by the combined effects of repression (up to 70-fold) and attenuation (up to 10-fold).

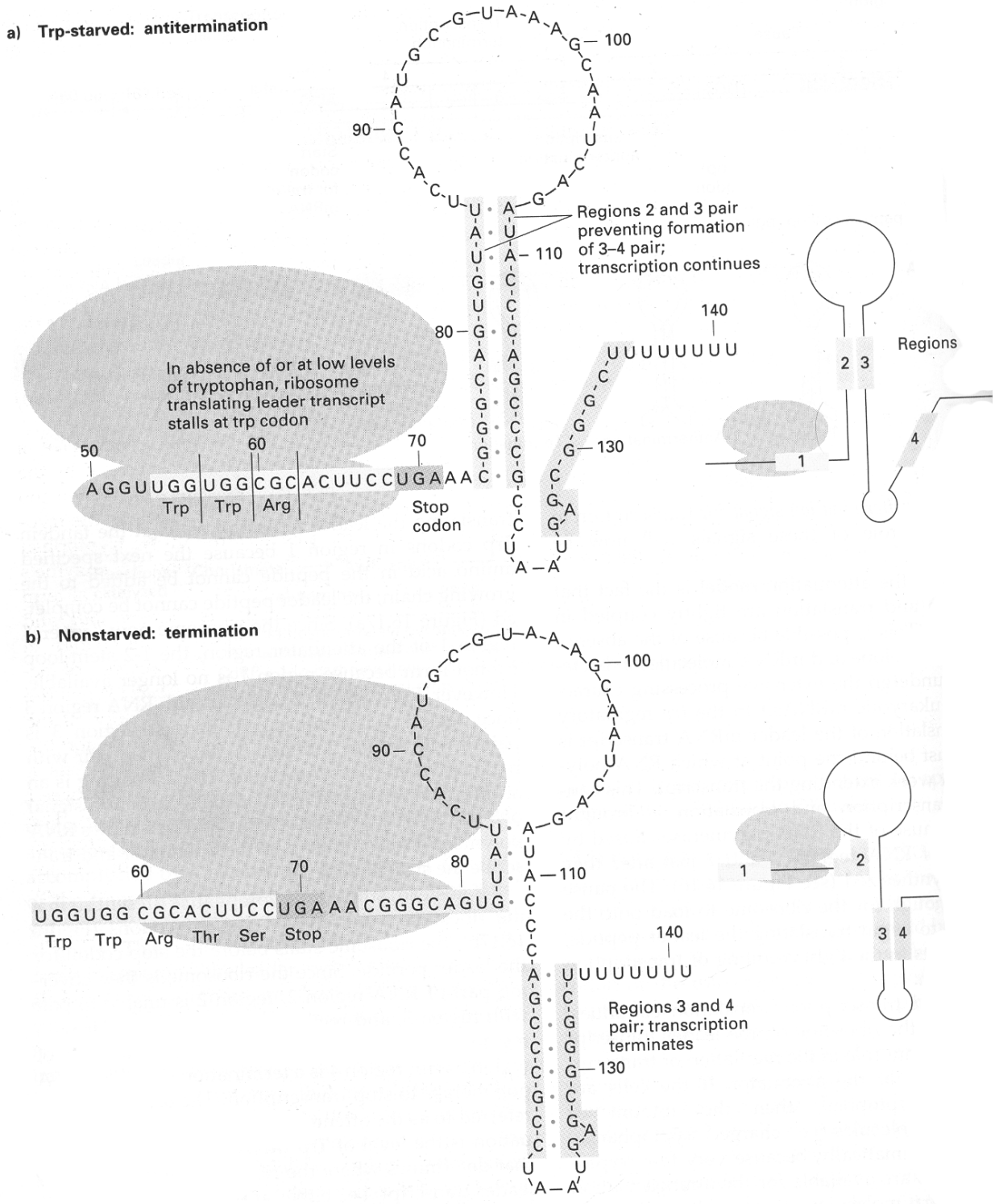


Fig. 11.3 Mechanism of attenuation mode of regulation

11.4 Fine-scale regulation of the trp operon

In *E. coli*, in addition to the regulation at the gene expression level in the trp operon, another mechanism also operates to give an overall regulatory system for gene expression. This mechanism is the degradation of the polygenic mRNA transcript of the protein-coding genes. As with the case of most other prokaryotic genes, the transcripts of trp genes have a relatively short half-life, so mRNA must continually be made in order for the enzymes to be produced. The short half-life allows the cell to change its physiological state when the environment changes. When tryptophan is added to the medium, transcription termination at the attenuator and repression at the operator block further transcription of the structural genes, and the structural gene mRNAs already made are rapidly degraded. In this way the cell does not make unnecessary materials. Another part of the overall regulatory system is a fine-tuning system. Gene regulation involves control processes that function in response to long-term changes in the environment. Cells must also respond to short-term changes that occur. For example, if tryptophan being made by the tryptophan biosynthetic pathway accumulates at a rate faster than it is being used, a mechanism called feedback inhibition or end-product inhibition is triggered.

In feedback inhibition the end product of a biosynthetic pathway can often be recognized by the first enzyme in the biosynthetic pathway. When too much tryptophan is produced, for example, the tryptophan binds to the first enzyme in the tryptophan biosynthetic pathway, thereby altering the enzyme's three-dimensional conformation. When the shape of the enzyme is changed, the function of the enzyme is impaired, because the substrate for the enzyme can no longer bind to the enzyme. As a result, the tryptophan pathway is temporarily turned off at the first step in the biosynthetic pathway. The enzyme becomes nonfunctional until the tryptophan level in the cell drops and the tryptophan dissociates from the enzyme. This step reverses the feedback inhibition of the enzyme's function, and the pathway is restarted.

11.5 Summary

The genes for a number of bacterial amino acid biosynthesis pathways are also arranged in operons. Expression of these operons is accomplished by a repressor-operator system or, in some cases, through attenuation at a second controlling site called an attenuator, or both. The tryptophan operon is an example of an operon with both types of transcription regulation systems. The repressor-operator system functions essentially like that of the *lac* operon, except that the addition of tryptophan to the cell activates the repressor, thereby turning the operon off. The attenuator is located downstream from the operator in a leader region that is translated. The attenuator is a transcription termination site that allows only a fraction of RNA polymerases to transcribe the rest of the operon. Attenuation involves a coupling between transcription and translation, and the formation of particular RNA secondary structures that signal whether or not transcription can continue.

Key to the attenuation phenomenon is the presence of multiple copies of codons for the amino acid synthesized by the enzymes encoded by the operon. When enough of the amino acid is present in the cell, enough charged tRNAs are produced so that the ribosome can translate the key codons in the leader region, and this causes the RNA that is being made by the RNA polymerase

ahead of the ribosome to assume a secondary structure which signals transcription to stop. However, when the cell is starved for that amino acid, there are insufficient charged tRNAs to be used at the key codons so that the ribosome stalls at that point. As a result the RNA ahead of that point assumes a secondary structure that permits continued transcription into the structural genes. The combination of repressor-operator regulation and attenuation control permits a fine degree of control of transcription of the operon.

11.6 Model questions

1. Give an account on the mechanism of *trp* operon regulation
2. a) Attenuation

   b) Repressible system

   c) Biosynthetic pathway of tryptophan

11.7 Reference Books

1. Freifelder, D. Molecular Biology (1990) – Narosa Publishing House, New Delhi
2. Gardner, E.J., Simmons, M.J. and Snustad, D.P. Principles of Genetics (2001)- John Wiley & Sons, Inc., New York
3. Snustad, D.P., Simmons, M.J. and Jenkins, J.B. Principles of Genetics (1997) - John Wiley & Sons, Inc., New York
4. Russel, P.J. Genetics (1998)- The Benjamin/Cummings Publishing Company, Inc., California
5. Singh, B.D. Fundamentals of Genetics (2001) – Kalyani Publishers, New Delhi

Dr. V. UmaMaheswara Rao

Lesson No. 12

GENETICS OF NITROGEN FIXATION

12.0 Objective

This lesson plan is intended mainly to emphasize the genes involve in nitrogen fixation and nodule formation.

12.1 Introduction

Nitrogen, one of the important elements, in its gaseous form occupy 80% of the atmosphere. But it is inaccessible for utilization by many organisms unless it is turned into organic form. Nitrogen in its organic form is a major component of all living beings as proteins, nucleic acids, vitamins and numerous other vital molecules contain it as a part of them. The utilization of nitrogen gas as a source of nitrogen is called nitrogen fixation. This inert nitrogen gas is made available for utilization, in nature, by only certain prokaryotes. A variety of prokaryotes, both aerobic and anaerobic, reduces or fix the atmospheric nitrogen to an utilizable nitrogen source namely ammonia, which can be used by plants and other microorganisms as building block for the synthesis of amino acids and other nitrogenous compounds. So far, no eukaryotic organism has been reported that can fix the nitrogen. The phenomenon of making or converting the gaseous nitrogen to utilizable nitrogen form is called as ‘nitrogen fixation’. As it is carried out by living prokaryotic organisms, it is called as ‘biological nitrogen fixation’. Some prokaryotes fix the nitrogen in free-living state and the phenomenon is called as ‘non-symbiotic nitrogen fixation’. Some other species, fix the nitrogen in association with specific host plants by forming specialized structures called ‘nodules’ on plant roots. These species are called as ‘symbiotic nitrogen fixers’ and the phenomenon is called as ‘symbiotic nitrogen fixation’.

12.2 Nitrogenase Enzyme complex

In the fixation process, N_2 is reduced to ammonium and the ammonium converted into organic form. The reduction process is catalyzed by the enzyme complex nitrogenase., which consists of two separate proteins called dinitrogenase and dinitrogenase reductase. Both components contain iron, and dinitrogenase contains molybdenum as well. The molecular weight of the dinitrogenase range from 2,20,000 to 2,70,000 daltons. The Fe containing protein or the dinitrogenase reductase possess a molecular weight within the range of 55,000 – 66,800 daltons. The molybdenum and iron in dinitrogenase are contributed by a cofactor known as FeMoCo, and the actual reduction of N_2 involves participation of this iron-molybdenum center. Active nitrogenase can be reconstituted by the addition of purified Mo-Fe and Fe proteins of different microorganisms. For examples, proteins of *Klebsiella pneumoniae* and *Bacillus polymyxa* and those of blue-green algae and photosynthetic bacteria have been combined to reconstitute active nitrogenases.

The Mo-Fe protein has been designated as dinitrogenase because nitrogen binds to the protein moiety whereas the Fe protein has been referred to as the dinitrogen reductase since the second moiety serves the specific function of reducing the Mo-Fe protein. Some nitrogen-fixing bacteria can synthesize more than one nitrogenase under certain growth conditions, and these so-called alternative nitrogenases do not contain molybdenum but instead contain either vanadium (and iron) or iron only. Alternative nitrogenases are not synthesized when sufficient molybdenum is present, as the molybdenum nitrogenase is generally the main nitrogenase in the cell. Alternative nitrogenases presumably serve as a back-up mechanism to ensure that N_2 fixation can still occur when molybdenum is limiting in the habitat. The nitrogenase has been isolated from different genera of free-living nitrogen-fixing microorganisms – *Clostridium, Bacillus, Klebsiella, Chloropseudomonas, Chromatium, Rhodospirillum, Anabaena, Gloeocapsa, Plectonema, Azotobacter* and *Mycobacterium.*

Owing to the stability of the N≡N triple bond, N_2 is extremely inert and its activation is a very energy demanding process. Six electrons must be transferred to reduce N_2 to $2NH_3$, and several intermediate steps might be visualized; but since no intermediates have ever been isolated, it is now assumed that the three successive reduction steps occur with the intermediates firmly bound to nitrogenase. Nitrogen fixation is highly reductive in nature and the process is inhibited by oxygen since dinitrogenase reductase is rapidly and irreversibly inactivated by O_2. in aerobic bacteria, N_2 fixation occurs in the presence of O_2 in whole cells, but not in purified enzyme preparations, and nitrogenase in such organisms is protected from O_2 inactivation either by removal of O_2 by respiration, the production of O_2-retarding slime layers, or by compartmentalizing nitrogenase in a special type of cell.

12.3 Genetics of N_2-fixation in *Klebsiella pneumoniae*

The genetics of N_2-fixation is best studied in *Klebsiella pneumoniae*, a non-symbiotic nitrogen fixer. Genes that involve in nitrogen fixation are called '*nif* genes' and form a complex regulon called '*nif* regulon' or '*nif* complex' or '*nif* gene cluster'. The regulon refers to a large stretch of genetic material with multiple operons. The *nif* regulon includes nitrogenase structural

genes, genes for FeMoCo factor, genes controlling the electron transport proteins and a number of regulatory genes. In *K. pneumoniae* the *nif* region constitutes a cluster of chromosomal genes next to the genes regulating the biosynthesis of histidine. By genetic procedures involving the isolation of mutants lacking *nif* genes, complementation analysis, cloning of *nif* genes, identification of *nif*-coded polypeptides and DNA sequencing, the genes involved in nitrogen fixation and their organization in the chromosome have been deciphered in *K. pneumoniae*. Barring the *nif* J, the entire *nif* cluster of *K. pneumoniae* has now been sequenced. The complete *nif* cluster constitutes 21 genes *nif* JCHDKTYENXUSVWZMFLABQ, of which T,W, and Z are the three potential new genes. These 21 genes are arranged in 7 or 8 transcriptional units. The *nif* gene cluster of *K. pneumoniae* possess a molecular weight of 18 x 10^6 daltons. Thus the physical map of *nif* cluster in *K. pneumoniae* is well understood and all the genes have been cloned into various vectors, which has facilitated the screening for *nif* homology in other nitrogen-fixing microorganisms. This has resulted in the analysis of *nif* genes in *Azotobacter, Azospirillum, Rhizobium, Enterobacter, cyanobacteria, Frankia* and other species.

The properties and functions of some of the *nif* gene products of *K. pneumoniae* have been fairly well understood. These relate to the nitrogenase enzyme, the electron transfer system and the regulatory functions. There are many other gene products where properties and functions are not clear. The Mo-Fe protein moiety of nitrogenase is encoded by *nif* H. The nucleotide sequence of *nif* H and *nif* D is not well established. Five genes (*nif* BNEVQ) are involved in the formation of functional Mo-Fe protein. The genes, *nif* BNE are involved either in synthesis or insertion of FeMoCO, a catalytic cofactor of nitrogenase. The gene *nif* V plays a role in substrate specificity while *nif* Q operates under conditions of Mo deficiency. The genes *nif* M is believed to be necessary in the processing of the Fe protein. The roles of six genes *nif* TYXUSW are presently unknown. Components of the specific electron transfer are the products of *nif* F and *nif* J which contain acid labile sulphur. In the electron chain to nitrogenase, the physiological donor is pyruvate. Electrons are carried from the *nif* J protein having pyruvate flavodoxin oxydoreductase activity to the *nif* F protein which is the substrate of the Fe protein. The genes *nif* H involves in the synthesis of Fe-protein. The *nif* genes D and K involves in the synthesis of α and β subunits of MoFe-protein, respectively. The product of *nif* A gene acts as a positive regulator and *nif* L gene product acts as negative regulator.

Ammonia (NH_4) totally represses *nif* gene product biosynthesis. The genes involved in glutamine synthetase, an enzyme which regulates NH_4 assimilation are referred to as *Gln* while *ntr* denotes genes whose products regulate nitrogen assimilation. The genes which determine uptake hydrogenase activity are known as *hup* genes. In *K. pneumoniae, ntr*BC are linked to *gln*A, the structural gene for glutamine sythetase whereas *ntr*A is unlinked. *Gln*A and *ntr*BC are organized in one or two operons transcribed from two promoters in the order P1glyAP2ntr BC. P1 promotes transcription under conditions of nitrogen limitation where as P2 acts under nitrogen enriched situations. In this way, when nitrogen is the limiting factor, the biosynthesis of glutamine synthetase is derepressed including those operons under the control of *nif* which includes the *nif* regulator also. The product of *ntr*C acts as a general activator of all these operons which is again dependent on *ntr*A product. The organization of the *nif* cluster of *K. pneumoniae* is given in Fig. 12.1.

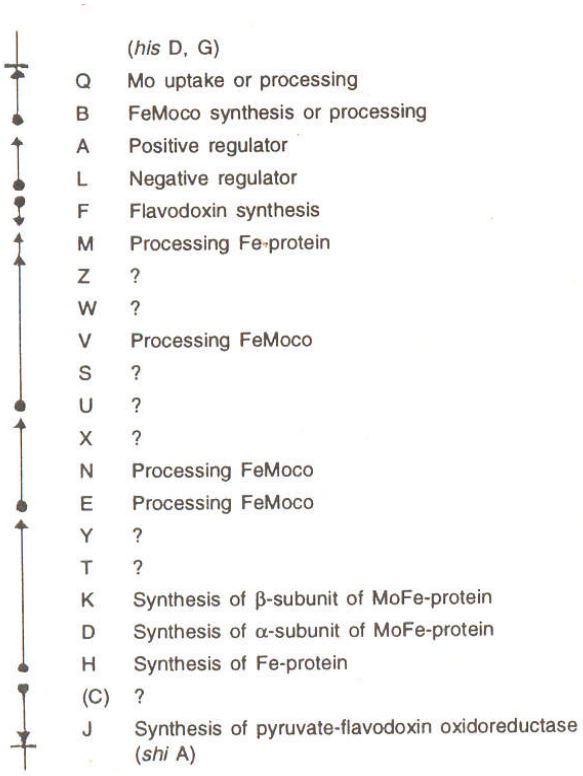


Fig. 12.1 nif gene cluster of *K. pneumoniae*

From the current information on the organization and functioning of the *nif* gene in some N_2 fixing speices, it can be reasonably concluded that a basic group of *nif* genes is perhaps common to all diazotrophs including *nif HDK* genes which are highly conserved. These are the genes concerned with the processing of the metal clusters of the nitrogenase plus the regulatory genes. Genetic analysis of *Azotobacter vinelandii* and *Azotobacter chroococcum* have revealed that a major *nif* cluster comparable to that of one in *K. pneumoniae* and comprising of genes *nif* HDKTY, *nif* ENX, *nif* USV, *nif* WZM, *nif* F, also occur in *Azotobacter*. The *nif* ABQ genes are also linked to each in a separate cluster and have been sequenced in *Azotobacter*. *A. chroococcum* is capable of producing two nitrogenases – a Mo containing one in N-free molybdenum added medium, a Va containing one in the presence of vanadium and absence of molybdenum. *A. vinelandii* can produce three enzymes – a Mo nitrogenase, a Va nitrogenase and one without either of these two metals in a medium devoid of those metals. Even though *nif* regulation in *Azotobacter* appears to be considerably similar to *K. pneumoniae*, the situation is complicated because of the unknown mechanisms through which the three types of nitrogenases are synthesized by the bacteria.

12.4 Regulation of *nif* genes

Nitrogen fixation is regulated by the expression or repression of transcription of *nif* structural genes. A positive regulatory protein encoded by *nif* A gene involves in positive

regulation (expression of transcription) and *nif* L protein involves in negative regulation (repression of transcription). In the presence of a suitable source of fixed nitrogen, such as ammonia, transcription of the *nif* genes is shut down or repressed. As a result of positive regulation *nif* genes are transcripted and N_2 is reduced to ammonia. However, this ammonia produced by nitrogenase does not repress enzyme synthesis as it is incorporated into organic form and used in biosynthesis as soon as it is produced. In certain nitrogen-fixing bacteria, especially phototrophic bacteria, nitrogenase activity is also regulated by ammonia, a phenomenon called the 'ammonia-switch off effect'. In this case, excess of ammonia catalyzes a covalent modification of nitrogenase reductase, which results in a loss of enzyme activity. Again when ammonia becomes limiting, modified protein is converted back to the active form and N_2 fixation resumes.

The *nif* gene regulation is very complex, involving both local control by genes within the *nif* complex and more global control by regulatory genes located elsewhere in the genome. Like the other genes, *nif* genes for its expression requires an RNA polymerase enzyme to transcribe the DNA into messenger RNA. The regulatory genes that involve in global control are *ntr*A, *ntr*B, *ntr*C and glnA genes. During initiation of transcription, RNA polymerase binds to the promoters that present at the beginning of the genes. The nucleotide sequences of the *nif* gene promoters are very different from those of other genes, such that of *E. coli*. So, the RNA polymerase that recognizes the *nif* gene promoters is somewhat different from the enzymes that binds to the promoters of other genes. The *ntr*A gene codes for a protein called as 'sigma factor' which confers on RNA polymerase the ability to recognize the *nif* gene promoter. Without this protein, the *nif* genes including the regulatory *nif* A and *nif* L could not be transcribed and nitrogen fixation would not occur. So, mutations in *ntr*A gene completely abolish nitrogen fixation.

The *nif* A gene, from within the *nif* gene complex, makes a protein that is necessary for inducing *nif* gene transcription. However, in presence of oxygen and ammonia, this induction by *nif* A gene is prevented or inhibited by *nif* L gene product. In the absence of oxygen and other fixed nitrogen sources, the product of *ntr*C gene in concert with *ntr*A product activates the *nif* A,L promoter. But in high ammonia, *ntr*B product coded by *ntr*B gene blocks the activation of *nif* A,L promoters by *ntr*C. The *ntr*B and *ntr*C genes are transcribed by either of two promoters that present, of which one (P1) also involves in the transcription of gln A gene that codes for glutamine synthetase. Thus, the expression of *nif* genes that involve in N_2-fixation is regulated or controlled locally by *nif* A and *nif* L genes that present in the *nif* gene cluster and globally by *ntr*A, *ntr*B and *ntr*C that occur as a part of global system, but located elsewhere in the genome.

12.5 Nodule formation in Symbiotic N_2-fixation

Nitrogen can also be fixed or reduced to ammonia in a symbiotic association between certain bacteria and higher plants. The best known and well studied example is Legume-*Rhizobium* symbiosis. The reduction of N_2 to ammonia in this process is called as symbiotic nitrogen fixation and the organisms involved are known as symbiotic nitrogen fixers. The bacteria that involve in legume-*Rhizobium* symbiosis include species of genera, *Rhizobium* and *Bradyrhizobium*. Both of them are Gram-negative motile rods. *Rhizobium* consists fast growing species and the genus *Bradyrhizobium* includes slow growing species. Both of them are able to form nodules on roots of legume plants as a result of interaction or association.

Initially, a specific adhesion protein called Ricadhesin which is present on surfaces all species of *Rhizobium* and *Bradyrhizobium* binds to the calcium complexes on root hair surface. Lectins are the other substances identified on the surfaces of both plant roots and bacteria and so involve in interaction. In this first step, the bacterium infects the growing root hair at its tip. As a response the root hair curls and an infection thread is produced by plant. Then bacterium invade into infection thread, multiply and induces the formation of incipient nodule. As the nodule develops, the bacterial cells transform into swollen, mis-shapen cells called 'bacteroids'. These bacteroids, either singly or in small groups, surrounded by plant produced cell membrane called 'peribacteroid membrane'. Nitrogenase is localized within the bacteroids themselves and is not released into plant cytosol. The bacteria in this form of bacteroids turn on the nif genes and carryout the nitrogen fixation. *Rhizobium* needs O_2 to generate energy for N_2 –fixation. This needed O_2 is supplied to bacteroids in nodule by the O_2-binding protein called 'leghemoglobin'. This leghemoglobin is a red, iron containing protein which gives the characteristic pink coloration to the healthy nodules. The heme portion of the leghemoglobin is synthesized by bacterium and globin portion is by plant. This leghemoglobin not only supply the necessary O_2 to bacteroids, but acts as oxygen buffer and involves in the maintenance of low but constant levels of oxygen within the nodule to avoid the inactivation of nitrogenase enzyme.

12.6 Genetics of nodule formation and regulation

Genes directing specific steps in nodulation of a legume by a strain of *Rhizobium* have been called '*nod* genes'. Many *nod* genes from different *Rhizobium* species are highly conserved and are borne on large plasmids called 'Sym plasmids'. In addition to *nod* genes which direct specific nodulation events, Sym plasmids contain specificity gens, which restrict a strain of *Rhizobium* to a particular host plant. In the Sym plasmid of *Rhizobium leguminosarum*, *nod* genes are located between two clusters of genes for nitrogen fixation, the *nif* genes. The arrangement of *nod* gens in the *R. leguminosarum* Sym plasmid is shown in figure 12.2. In this species, ten nod genes have been identified. The entire *nod* region has been sequenced and the function of many nod proteins is known, the genes *nod*ABC direct root hair curling by production of specific polysaccharide signal molecules called NOD factors. The *nod*D gene encodes a regulatory protein that controls expression of other nod genes. In the presence of specific nodulation signals, the *nod*D gene product controls transcription of other nod genes.

In most rhizobial species the NodD protein is constitutive and regulates transcription of other nod functions by a simple induction type of control. The NodF protein is involved in the synthesis of a lipid or polysaccharide needed for effective nodulation and is thought to be involved in the transport of substances needed to begin nodulation. The *nod*E and *nod*L genes function in determining host range. The *nod*M gene encodes for glucosamine synthase involves in nod factor synthesis. The two other genes, *nod*I and *nod*J encode membrane proteins that involves in the exportation of Nod factors from bacterial cells. Cloning and sequencing of nod genes in *R. leguminosarum* suggests that they are arranged in at least three operons.

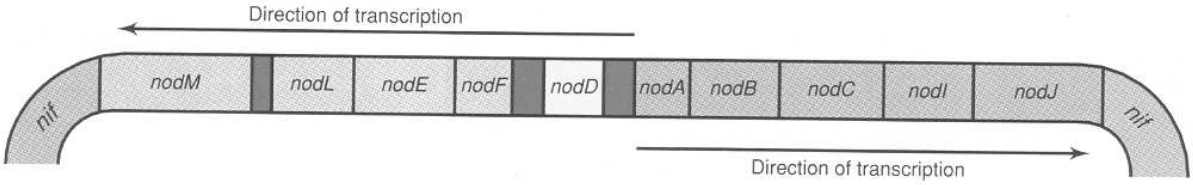


Fig. 12.2 nod gene cluster in *R. leguminosarum*

As a regulatory protein, the NodD protein binds to a highly conserved 35 base pair DNA sequence which preceeds each operon. These regions, referred to as nod boxes, are presumably operator genes to which the NodD protein can bind. Following interaction with inducer molecules, the conformation of the NodD protein changes, causing it to fall off the nod box DNA. This event presumably exposes a promoter site for RNA polymerase binding which initiates transcription of nod structural genes. Several inducer molecules have been identified. In most cases, these are plant flavonoids, complex organic molecules that are widespread plant products. These flavonoids have many functions in plants, including growth regulation and attraction of pollinating animals. But in the case of leguminous plants, their roots secrete large amounts of flavonoids, presumably to trigger nod gene expression in nearby rhizobial cells in the soil. Interestingly, some flavonoids that are structurally very closely related to *nod*D inducers strongly inhibit induction of nod genes in certain rhizobial species, suggesting that part of the specificity observed between plant and bacterium in the *Rhizobium*-legume symbiosis could lie in the chemical nature of the flavonoids excreted by a particular plant.

The *nod* D gene regulates the expression of other nod genes by encoding a regulatory protein which binds to the promoters of other nod genes. The *nod*D gene can also regulate its own transcription. The expression of other nod genes by *nod*D protein increase to several folds in the presence of inducers such as flavones or flavanones that present in the plant exudates. The *nod*D gene, at least in *R. leguminosarum*, is subject to negative regulation by its product, as the high concentrations of the product repress transcription of the *nod*D gene. The *nod*D gene product activates the other genes for expression. Mutations in *nod*D,A,B,C genes completely abolish the nodule formation. However, mutations in other genes delay the onset of nodulation and reduce the number of nodules per plant. Despite the differences in the host range specificities of various *Rhizobium* species, their corresponding nod genes are similar sequence, location and function. The regulatory circuitry that affects *nif* gene expression in *Rhizobium* species is clearly very similar to that in *K. pneumoniae*. The sequences of regulatory nif A gene and *nif* promoters of both *Rhizobium* and *Klebsiella* are also found to be similar. The regulatory mechanism of nod genes by nod D is given in the figure 12.3 below.

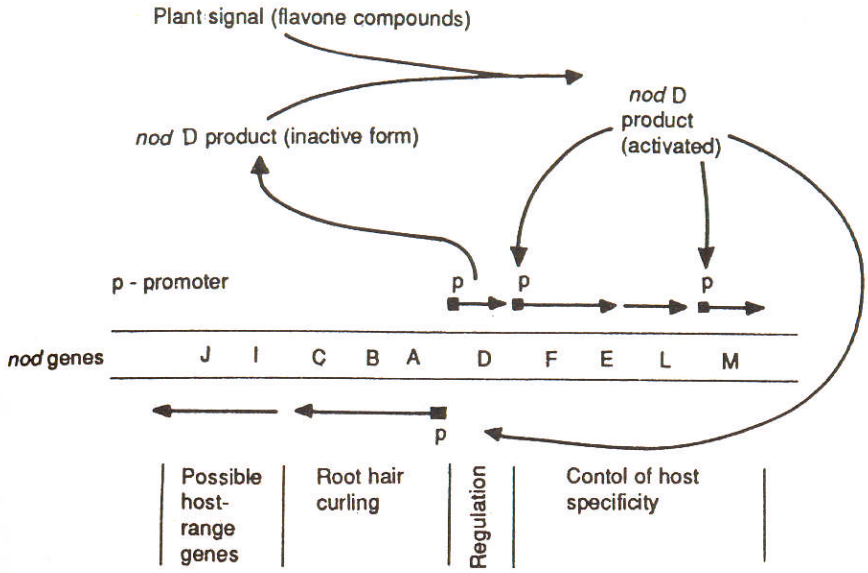


Fig. 12.3 Regulation of nod genes in *R. leguminosarum*

12.7 Summary

The conversion of gaseous nitrogen to utilizable nitrogen source is referred as nitrogen fixation. This event can be carried out only by certain prokaryotes, either free living and symbiotic nitrogen fixers. A number of genes involve in the nitrogen fixation and are called as nif genes. The nitrogen fixation is catalyzed by the enzyme complex called nitrogenase. The transcription of these nif genes is regulated locally and globally by the regulatory genes. In the case of symbiotic nitrogen fixation, nodules are formed by the bacterium on the roots of plants. The bacteria are present in the root nodules in the form of bacteroids which contain the nitrogenase enzyme and act as the sites for the nitrogen fixation. The nodule formation is catalyzed by a number of protein products coded by the genes called nod genes.

12.8 Model Questions

1. What is nif regulon? Write an account on the role of nif genes and their regulation.
2. Write an essay on nod genes and their role in nodulation process in *R. leguminosarum*.

12.9 Reference Books

1. Brock, T.D., Madigan, M.T., Martinko, J.M. and Parker, J. Biology of Microorganisms (1994) – Prentice Hall, New Jersey.
2. Subba Rao, N.S. Soil Microbiology (2001)- Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi.
3. Atlas, R.M., and Bartha, R. Microbial Ecology- Fundamentals and Applications (2000)- Benjamin/Cummings Science Publishing, California.

Dr. V. UmaMaheswara Rao

Lesson 13

TRANSPOSABLE ELEMENTS IN BACTERIA

13.0 **OBJECTIVE:**

The structure, types and mechanism of different transposable elements and their evolutionary significane is described here.

13.1 INTRODUCTION

A long standing assumption about gene organization and chromosome structure was that each gene has a definite position on a particular chromosome. However, movement of genetic elements from one place to another across the genome was not documented. It was only in 1950's, for the first time in eukaryotic system (Maize), during genetic analysis of 'mottling' of the maize Kernels, Barbara McClintock (1952) [Fig. 1] came out with clear observation that there are certain genetic elements in the maize genome that are endowed with capacity to move from one place to another resulting phenotypic effects. Recently genetic instabilities have been found in bacteria, that have led to the identification of transposons which were the first to be studied at the molecular level and were categorized as IS (Insertion Sequences).

Fig.13.1 Barbara McClintock

Transposons: Transposons are mobile genetic elements with DNA repeat sequences of the termini (Fig. 13.2). They are segments of DNA that can insert at random into the genome, plasmids, bacterial chromosomes independent of host cell recombination system. The process of integration and excision of transposable elements is called transposition. All transposable elements share two basic properties:

1. The ability to move from place to place in the genome.
2. The ability to amplify their copy number within the genome via this transposition.

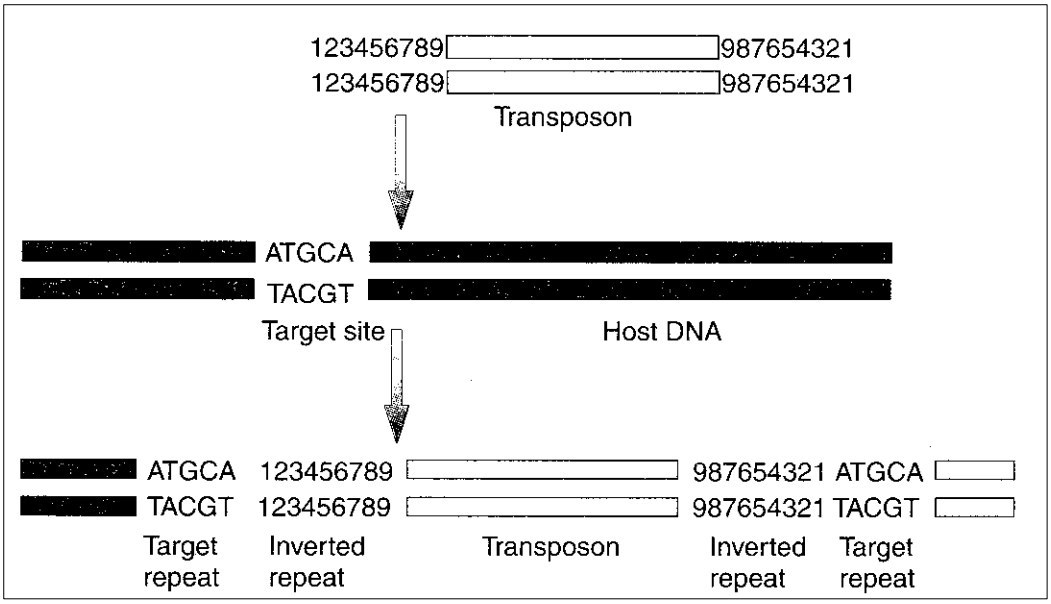


Fig. 13.2 Transposons with terminal inverted repeats generate direct repeats of flanking DNA at the target site. The ends of the transposon consist of inverted repeats of 9 base pairs (1 to 9).

13.2 BARBARA McCLINTOCK'S WORK

Barbara McClintock (Fig. 13.1) discovered jumping genes during 1951 in maize i.e. Ac and Ds resulting color variation among kernels of maize (Fig. 13.3). They can be classified into two classes: 1. Autonomous elements, they have the ability to excise and transpose autonomously (Ac). 2. Non-autonomous elements which are deleted defective elements do not transpose autonomously but require helper functions from an autonomous elements. Ds affects the phenotype by blocking the expression of the genes it transposes e.g. Callele that controls purple colour of Corn Kernel, when Ds enters into Callele it disrupts the activity (C^I) thus Kernel becomes white (Fig. 13.4). Ds element that entered into the genome can't move out of C until Ac element enters the genome by another genetic cross. In the presence of Ac, Ds cleaves its site restoring C-allele activity producing colored spots (Fig. 13.4). Recently Ds and Ac elements have been cloned and sequenced. They are found to be typical transposons.

Fig. 13.3 Color variation among kernels of maize. Studies of the genetic basis of this variation led to the discovery of transposable elements.

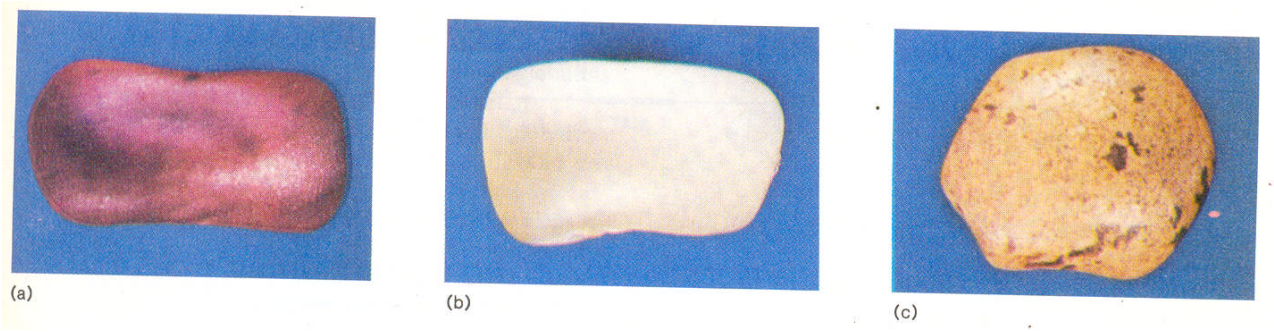


Fig. 13.4 The *Ac-Ds* mutability system in corn. (a) A corn kernel that is purple because it has the dominant C allele. (b) A corn kernel in which the dominant C allele has been inactivated by a previous transposition of the Ds element into the C allele, disrupting it. In this case, the Ac gene is not present; thus, the Ds element cannot leave its position within the C gene. (c) When *Ac* is again present (e.g. brought in by a genetic cross), *Ds* can leave its position and the C allele in those cells is then restored. The result is the spotted corn kernel.

Although each kind of transposable element has its own special characteristics, most of them can be classified into one of three categories based on how they transpose (Table 1.1). In the first category, transposition is accomplished by excising an element from its position in a chromosome and inserting it into another position. Excision and insertion events are catalyzed by an enzyme, transposase. Geneticists refer this mechanism as **cut-and-paste transposition** and the elements in this category as cut and paste transposons. e.g. IS elements, Ac/Ds elements and P elements etc.

Table 13.1 Categorization of Transposable Elements by Transposition Mechanism

Category	**Examples**	**Organism**
I. Cut-and-paste transposons	ISA elements (e.g. IS*50*)	Bacteria
	Composite transposons (e.g. Tn*5*)	Bacteria
	*Acl*Ds elements	Maize
	P elements	*Drosophila*
	mariner elements	*Drosophila*
	bobo elements	*Drosophila*
	Tcl elements	Nematodes
II. Replicative transposons	Tn*3* elements	Bacteria
III. Retrotransposons		
A. Retroviruslike elements (also called long terminal repeat, or LTR, retrotransposons)	*Tyl*	Yeast
	copia	Drosophila
	gypsy	Drosophila
B. Retroposons	*E,G*, and *I* elements	Drosophila
	Telomere-specific retroposons (*HeT-A, TART*)	Drosophila
	LINEs (e.g., *Ll*)	Humans
	SINEs (e.g., *Alu*)	Humans

In the second category, transposition is accomplished through a process that involves replication of transposable elements DNA. During the transposition, the element is replicated and one copy of it is inserted at the new site; one copy remains at the original site. This mechanism is referred by geneticists as replicative transposition. The elements in this category are **replicative transposons** e.g. Tn3 elements of bacteria**.**

In the third category, transposition is accomplished through a process that involves insertion of copies of an element synthesized from the element's RNA. Reverse transcriptase enzyme uses RNA as a template to synthesize DNA molecules to be inserted into new chromosomal sites. Since the genetic information flows in cells from RNA to DNA, the geneticists refer to it as retrotransposition. The elements under this category are referred as **retrotransposons**. Some of the elements that transpose in this way are related to retroviruses; consequently they are called *retroviruslike* elements. Other elements that engage in retrotransposition are called retroposons in e.g. *Tyl* in yeast, Retroposons in Drosophila and Humans.

13.3 STRUCTURE

Transposons contain two functional regions.

1. DNA repeat sequences at the termini of the transposon, required for genomic integration and excision.

2. An internal segment encoding a transposase enzyme required for transposition.

Transposons have been identified in many organisms both prokaryotes and eukaryotes. Bacterial transposons were the first to be characterized at the molecular level. In maze, a variety of transposons have been identified which are classified into families. Transposons have also been identified in other plant species such as snapdragon, *Petunia Arabidopsis* etc. In many plants with complex genomes, transposable elements make up over 50% of the nuclear DNA.

13.4 CLASSIFICATION

On the basis of structure:

(i) With long terminal direct repeats (copia in *Drosophila*, Ty in yeast, IAP in mice).
(ii) With long terminal inverted repeats (FB, TE in *Drosophila*).
(iii) With short terminal inverted repeats (P and I in *Drosophila,* Ac/Ds in maize, Tam I in snapdragon).
(iv) Without terminal repeats (Alu in mammals) they have been grouped into two major superfamilies on the basis of mechanism of transposition.

Class I elements: They are RNA transposable elements. They transpose by reverse transcription of an RNA intermediate. They make DNA copies of their RNA transcripts, that integrate into the genome. They do not excise when they transpose, they make a copy that inserts elsewhere in the genome.

Class II elements: They transpose directly as sequences of DNA coding for proteins that are able to manipulate DNA directly to propagate themselves within the genome.

The categories of transposon family are variable based on their presence in a particular organism, length and number of inverted repeat sequences etc. Thus they are grouped under four types of transposons as mentioned above.

13.5 TRANSPOSABLE ELEMENTS IN BACTERIA

Genetic instabilities have been found in bacteria that have led to the identification of transposable elements (Insertion Sequences), which were first studied at molecular level during 1967. They vary in size characteristic features of IS elements are given in Table 13.2.

Table 13.2 Features of some IS elements

Name	Element size (bp)	TIR left/right (bp)	TSDs (bp)	Target site preference
IS1	768	20/23	9	AT rich terminal G or C
IS2	1327	32/41	5	Hot spots in P_2 genome
IS3	1258	29/40	3	?
IS4	1426	16/18	11-13	$AAAN_{20}TTT$
IS5	1195	16	4	Hot spots
IS10	1329	17/22	9	NGCTNAGCN
IS50R	1531	9	9	Hot spots
IS91	-1800	8/9	0	?

Note: ITRs – Inverted terminal repeats; TSD – Target site duplication; bp – base pairs

13.5.1 IS elements

The simplest bacterial transposons are IS elements. An *E. coli* may contain several < 10 copies of any of the IS elements. The IS elements are autonomous units, coding for proteins needed to sponsor its own transposition. Each IS element is different in sequence but there are some common features in organization.

The terminal sequences are always in inverted orientation and they are called terminal inverted repeats (TIRs), their length range from 9 to 40 nucleotide pairs. The structure of an IS element before and after insertion at a target site is shown in Fig. 13.5.

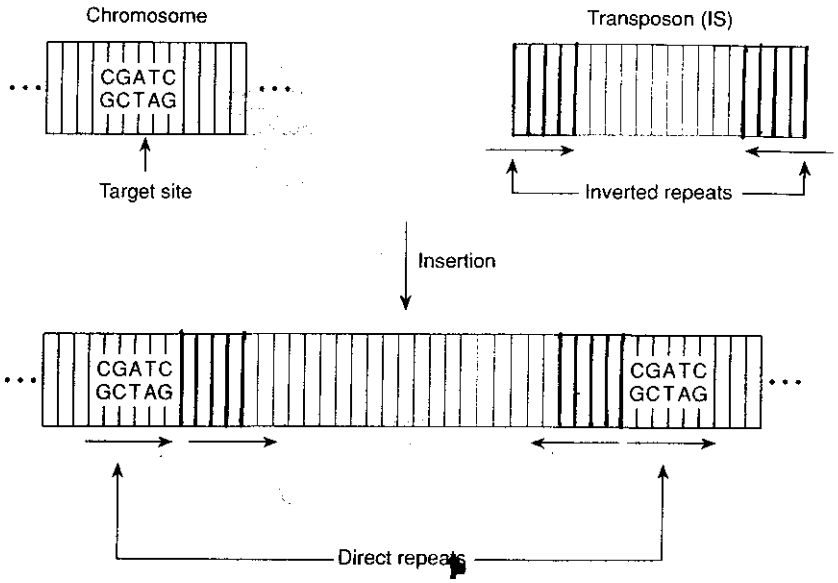


Fig. 13.5 An IS element inserted into a target site in a bacteria chromosome creates a direct repeat on either side of the IS element.

When IS elements insert into a chromosome or plasmid they create a duplication of the DNA sequence at the site of the insertion (Fig. 13.5). One copy of the duplication is located at each side of the element. These short 3-12 direct repeat sequences are called target site duplications and arise from staggered breaks in double stranded DNA (Fig. 13.6).

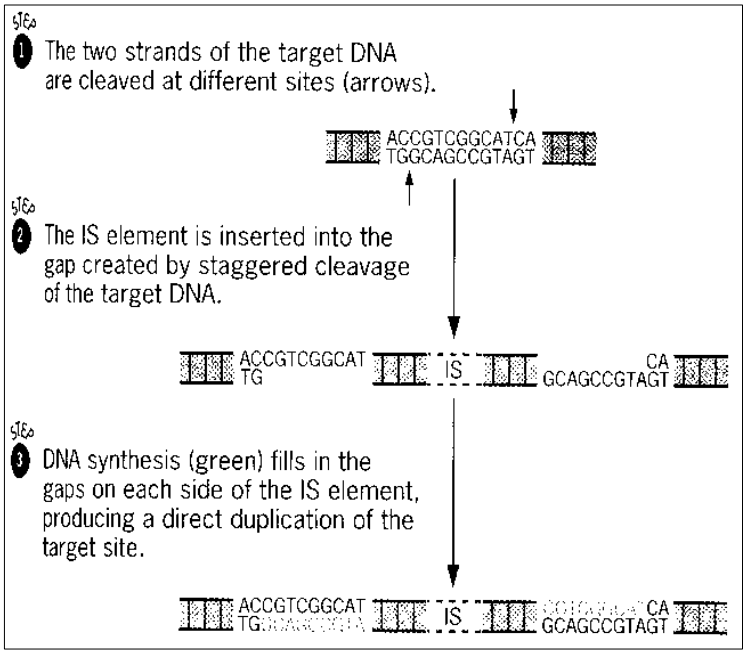


Fig. 13.6 Product of target site duplications by the insertion of an IS element

13.5.2 Composite transposons

Two homologous IS elements combine with other genes to form a composite transposon. Thus, a composite transposon has a central region carrying the drug markers flanked on either side

by arms that consist of IS elements. The arms may be either the same or inverted orientation (Fig. 13.7). In all three transposons, Tn9, Tn5 and Tn10, the genes between the flanking IS elements confer resistance to antibiotics.

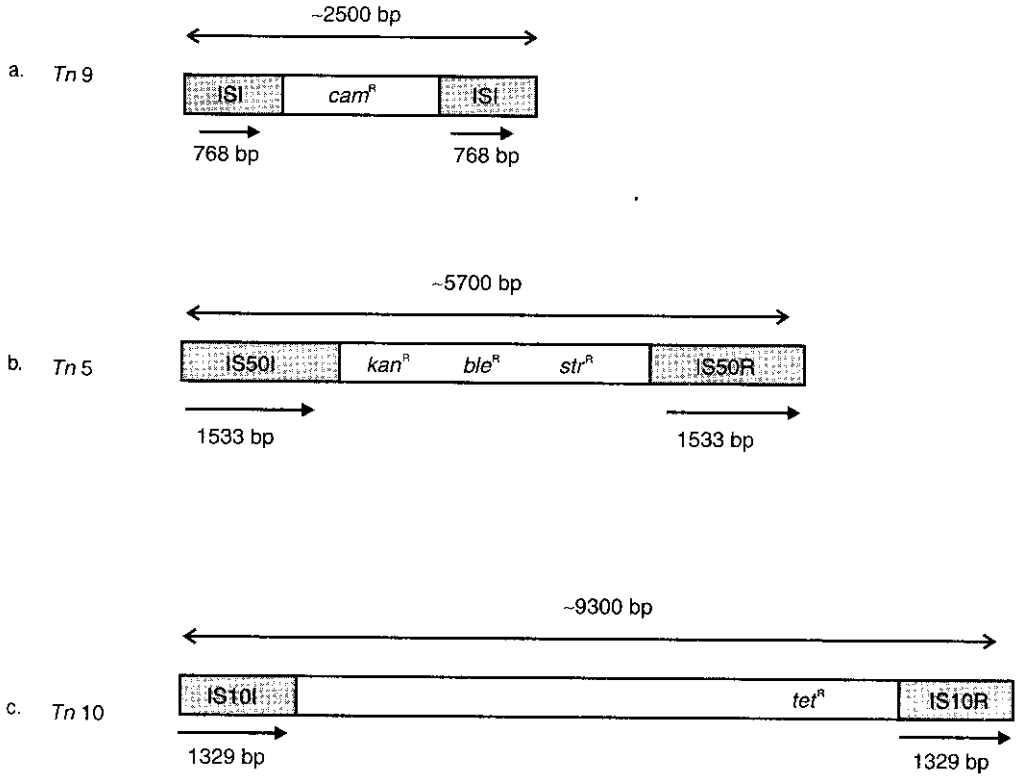


Fig. 13.7 Genetic organization of composite transposons. *Cam*-chloramphenicol; *kan*-kanamycin; *ble*-bieomycin; *str*-streptomycin; *tet*-tetracycline.

Tn5 movement is regulated, when a bacterial cell is infected with a non-lytic bacteriophage that carries Tn5 on its chromosome. The frequency of Tn5 transposition is dramatically reduced, if the infected cell carries a copy of Tn5. Analyses by Michael Syvanen, William Reznikoff and their colleagues have shown that this hypothesis is correct. IS50R element of Tn5 produces two proteins. The transposase catalyzes transposition, whereas other truncated transposase created by translation from a start codon within the transposase gene, prevents transposition. Because the shorter protein is more abundant, Tn5 transposition tends to be repressed (Fig. 13.8).

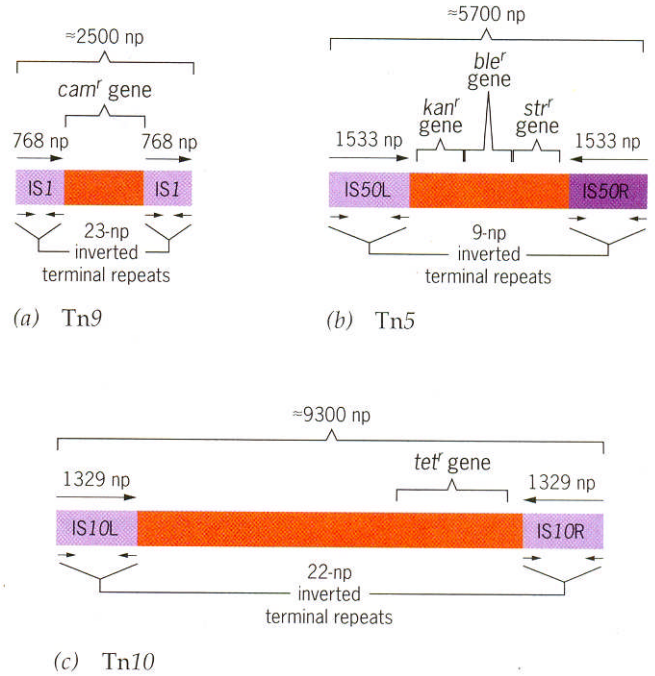


Fig. 13.8 Genetic organization f composite transposons. The orientation and length (in nucleotide pairs, np) of the constituent sequences are indicated. (a) Tn*9* cosnists of two *IS*1 elements flanking a gene for chloramphenicol resistance. (b) Tn*5* consists of two IS*50* elements flanking genes for kanamycin, bleomycin, and streptomycin resistance. (c) Tn*10* consists of two IS*10* elements flnaking a gene for tetracycline resistance.

13.5.3 Tn3 elements

The elements in this group of transposons are larger than the IS elements and contain genes not necessary for transposition. Tn3 elements (Fig. 13.9) have simple inverted repeats of 38-40 nucleotide pairs long and produce target site duplications when they insert into DNA. The transposition of Tn3 occurs in two stages (Fig. 13.10) by resulting a structure called a cointegrate. This event occurs at a sequence in Tn3, the resolution site, generating two molecules, each with a copy of the transposon.

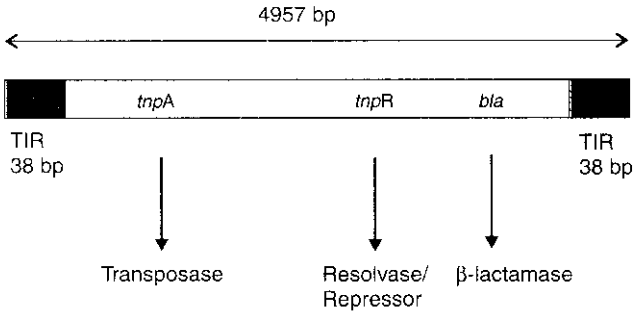


Fig. 13.9 Genetic organization of Tn3 complex transposon

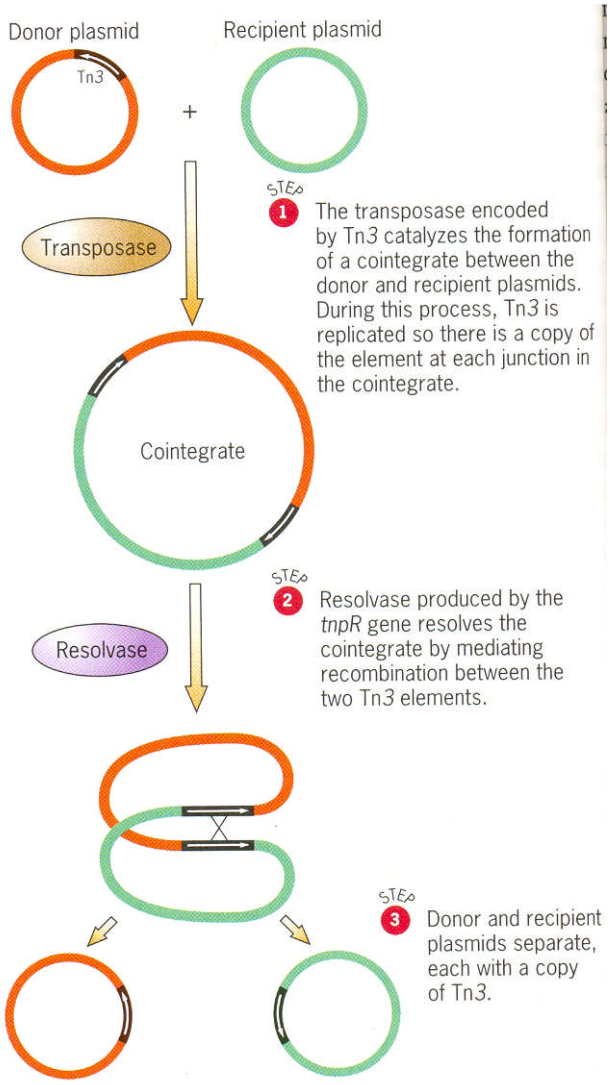


Fig. 13.10 Transposition of Tn3 via the formation of a cointegrate

13.5.4 Complex transposons

Tn3 family complex transposons do not contain IS elements. They have inverted terminal repeats of 38-40 nucleotide pairs long and produce target site duplications of 5 nucleotide pairs on insertion. They are larger than IS elements and contain accessory genes as well as genes needed for transposition. The genetic organization of Tn3 complex transposon is shown in Fig. 13.9. Three genes tnpA, tnpR and bla encode for transposase a resolvase and β-lactamase respectively. The β-lactamase confers resistance to ampicillin while the other two proteins play important role in transposition.

1.6 MECHANISM OF TRANSPOSITION

The transposition does not use the normal recombination machinery of the cell. Shapiro explains the widely recognized model that many tarnsposons in the process of transposition go through a cointegrate stale (Fig. 13.11), a state in which there is a fusion of two elements. During the process of transposition, an intermediate cointegrate stage is formed, made up of both plasmids and two copies of the transposon. Then through a process called **resolution**, the cointegrate is reduced back to the two original plasmids, each containing a copy of the transposon.

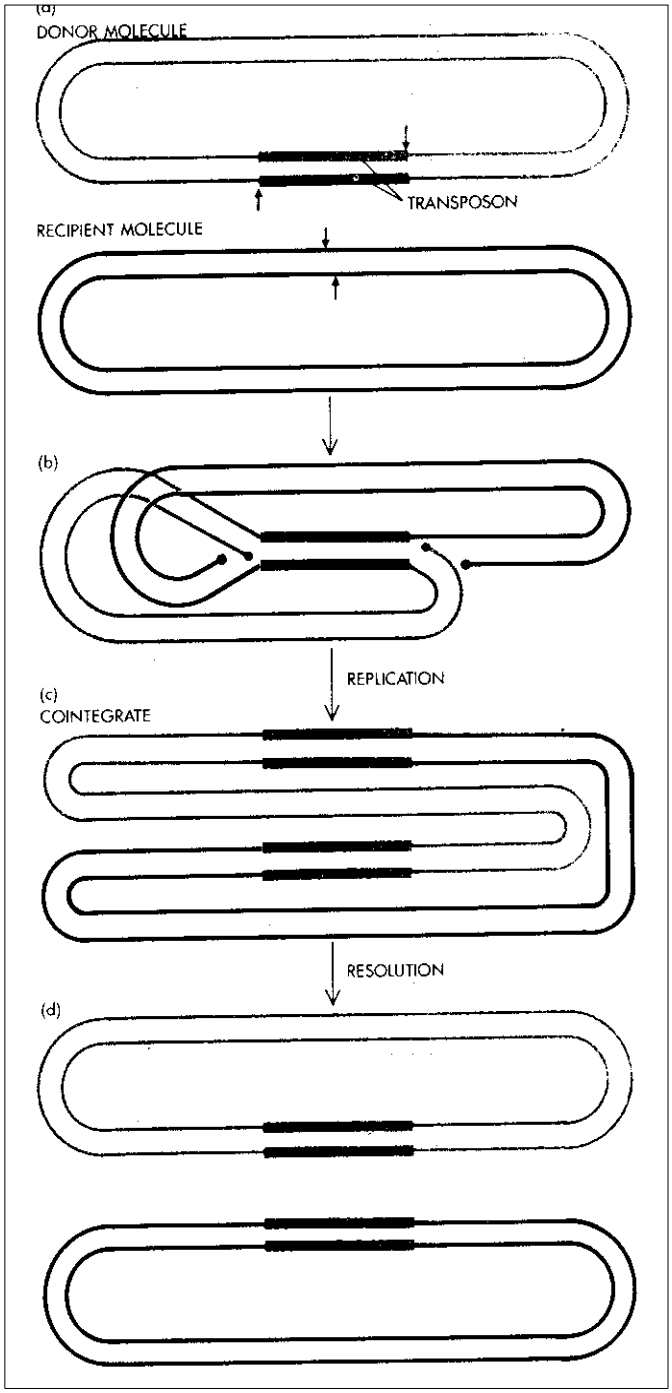


Fig. 13.11 A model for transposition.
(a) Single stranded cuts are made at the ends of the transposon on he donor molecule and staggered cuts 5 to 7 bases apart in the target DNA (recipient molecule). (b) Each transposon end is attached to a protruding target DNA end. Replication from the two forks copies the transposon to produce structure. (c) The resulting "cointegrate" consists of two copies of the transposon in a direct-repeat orientation. (d) A site-specific crossover at the internal resolution or "res" site produces the starting donor molecule and the recipient molecule, now containing a copy of the transposon. →

13.7 EFFECTS OF TRANSPOSITION (phenotypic and genotypic)

A well known case of transposon orientation controlling a phenotype in bacteria occurs in *Salmonella typhimurium*. The flagella of this bacterium occur in two types, determined by the

protein of the flagella. Phase 1 flagella are determined by the H1 gene and Phase 2 flagella are determined by the H_2 gene. The change from one phase to another occurs at a rate of about 10^{-4} per cell division. The H1 and H2 genes are at separate locations on the bacterial chromosome (Fig. 13.8). When the promoter is in proper orientation, the H2 operon is expressed resulting in phase 2 flagella. The rH1 gene product represses the H1 gene. If the inverted repeat ends of the transposon undergo recombination, the transposon will be inverted causing the promoter to the in an incorrect orientation for the transcription of the H2 operon. No H1 repressor will be made, with the result that the H1 gene will be expressed resulting phase flagella.

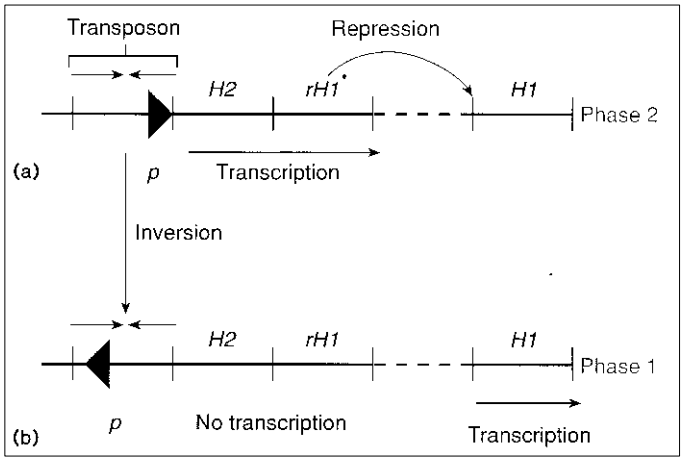


Fig. 13.12 Arrangement of flagellin genes on the Salmonella chromosome. The promoter (p) is within a transposon. In one orientation (a), the *H2* operon is transcribed, which results in *H2* flagella and *rH1* protein, the repressor of the *H1* gene. In the second orientation (b), the *H2* operon is not transcribed, resulting in uninhibited transcription of the *H1* gene.

13.8 THE MEDICAL SIGNIFICANCE OF BACTERIAL TRANSPOSONS

Many bacterial transposons carry genes for antibiotic resistance. They can move from one DNA molecule to another, from a chromosome to a plasmid. This genetic flux has a profound medical significance. Many DNA molecules that acquire resistance genes can be passed on to other cells. Resistance to a particular antibiotic can be spread horizontally between individuals or vertically from one generation to the next in a bacterial population. Hence, all or nearly all bacterial cells become resistant.

This process has occurred in several species pathogenic to humans, including *Staphylococcus, Enterococcus, Shigella* and *Salmonella*. Today many bacterial infections causing diseases such as dysentery, tuberculosis, gonorrhea are difficult to trout; since pathogen has acquired resistance to several different antibiotics.

The spread of multiple drug resistance in bacterial populations has been accelerated by the evolution of conjugative R plasmids that carry the resistance genes. The plasmids have two components one is **resistance transfer factor** (RTF) contains the genes needed for conjugative

transfer between cells. The other is **R-determinant**, contains the genes for antibiotic resistance. They resistance genes carried by transposon can be inserted into the plasmid (Fig. 13.9). Conjugative R plasmids can be transferred rapidly between cells in abacterial population between dissimilar cell types e.g. between a coccus and a bacillus. Thus multiple drug resistance has evolved in a part of the microbial kingdom.

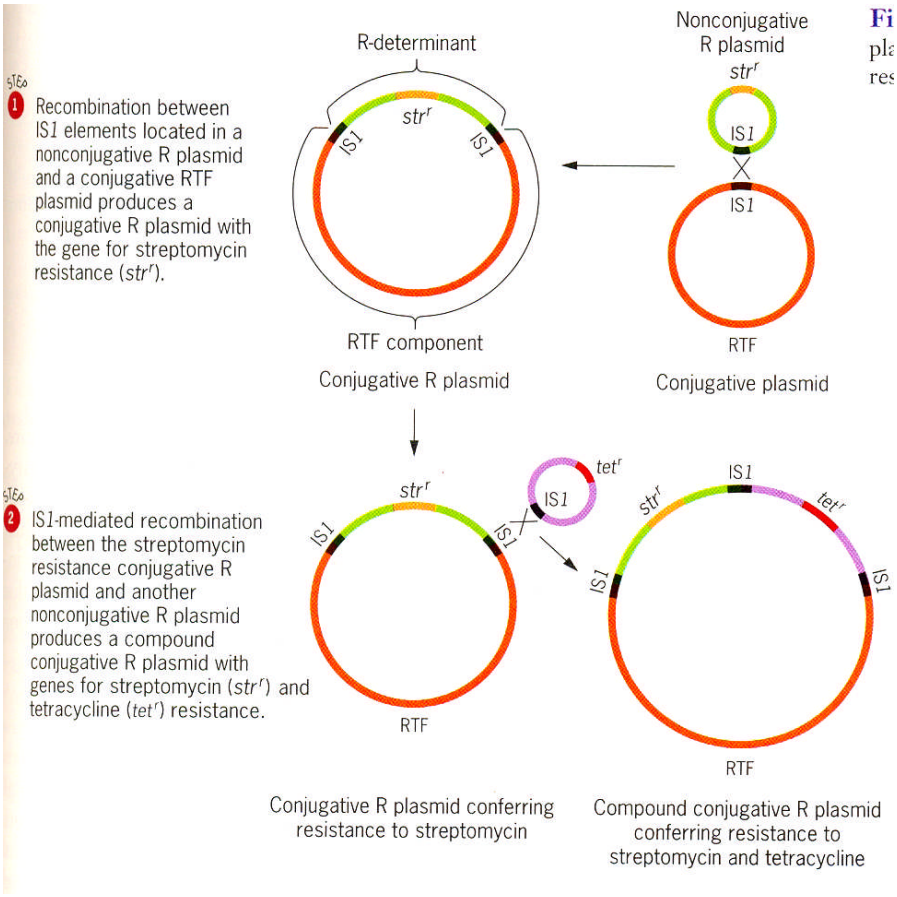


Fig. 13.9 Evolution of conjugative plasmids carrying genes for antibiotic resistance

Thus, transposons can have marked effects on the phenotype by their actions in transposition. However, they can also exist without noticeable consequences. Transposons once created are self-maintaining. Since they may exist without a noticeable benefit to the host's phenotype, transposons have been referred to as **selfish DNA**. Some scientists have suggested that transposons improve the evolutionary fitness of the bacteria that have them.

Applications of Transposons:

1. Transposable elements have been utilized for isolation of genes, when the gene product is unknown.
2. Transposon works as a mutagen causing different phenotypic effects.

3. DNA sequence alterations caused by transposons range from few base pairs to large chromosome rearrangements via deletion, inversion and duplication.
4. Transposons are transposed to a gene of interest, to produce unstable allele.
5. The sequences introduced by insertion have the ability to alter promoters of genes, thus also effect gene regulation pattern.
6. They can be directed towards tagging a specific gene of interest.
7. Gene delivery vectors.

13.9 SUMMARY

Transposons are mobile genetic elements first discovered in maize by Barbara McClintock. They re Ac Ds elements, autonomous and non-autonomous in their function. Mottling of maize kernels were identified and she observed that certain genetic elements move from one place to another in the genome.

During 1970s, the such elements were found in bacteria and named as IS elements. The structure and features of some IS elements are described. The length of a transposons and its repeats at the ends varies among them. Phenotypic effects and medical significance of transposons aredescribed here.

1.10 MODEL QUESTIONS

1. Explain IS elements in bacteria. Describe their significance in medicine.
2. Classify transposons in prokaryotes? Describe the mechanism of transcription.
3. Write short notes on:
   a) Ac-Ds element
   b) Tn3 element
   c) Structure of a transposon
   d) Effect of transposon

1.11 REFERENCE BOOKS

1. P.J. Russell, 1987. **Essential Genetics**, Blackwell Publishers.
2. F.J. Ayala & J.A. Kiger, Jr. 1980. **Modern Genetics**. The Benjamin/Cummings Publ. Com.
3. M.W. Strickberger. **Genetics** – Prentice Hall of India, New Delhi.
4. Tamarin. **Principles of Genetics**.

*****

PROF. T.N. MARY

Lesson 14

TOOLS AND TECHNIQUES IN MOLECULAR BIOLOGY

(Southern blotting, Northern blotting, Western blotting, Polymerase chain reaction and DNA finger printing)

14.1 INTRODUCTION

The development of recombinant DNA techniques has resulted many new approaches to the analysis of genes and gene products. Geneticists can isolate and characterize any gene from any organisms. During the course of several molecular biology experiments (characterization of transgenic organisms) it becomes necessary to confirm the integration, presence and expression of a specific gene, which is known through blotting techniques.

On the other hand, fingerprints have played central role in human identity cases for decades. The use of DNA fingerprints in forensic cases is based on the fact that no two individuals have identical prints. Recent evidence has demonstrated that the human genome contains large families of DNA polymorphisms. These polymorphisms can be used to produce **DNA fingerprints**.

DNA fingerprinting provides a **powerful forensic tool**. Another tool **polymerase chain reaction** is extremely powerful procedure that allows the amplification of a selected DNA sequence in a genome. The procedures of these techniques and their applications in the field of molecular biology are described here.

14.2 SOUTHERN BLOTTING (Analysis of DNAs)

Gel electrophoresis is a powerful tool for the separation of macromolecules with different sizes and charges. DNA molecules separate in agarose and acrylamide gels on the basis of size (Fig. 14.1).

In 1975, EM southern published an important procedure that allowed investigators to identify the location of genes and other DNA sequences on restriction fragments separated by gel electrophoresis. It acquired the name **Southern blotting** after it's author. Later on, other scientists modified the procedure to use with RNA fragments. They named the procedure for transferring RNA from a gel to a membrane **Northern blotting**. Later on other researchers figured out how to transfer proteins from gels membranes, to test them. The name of this procedure is **western blotting**. No eastern blotting has yet been invented.

An outline of a procedure for southern hybridization is shown in Fig. 14.2.

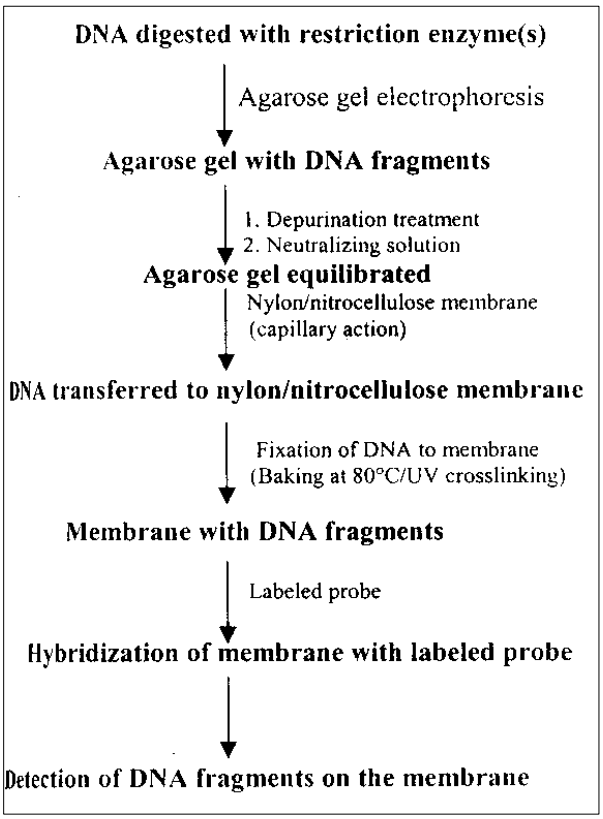


Fig. 14.2 A general scheme for Southern hybridization

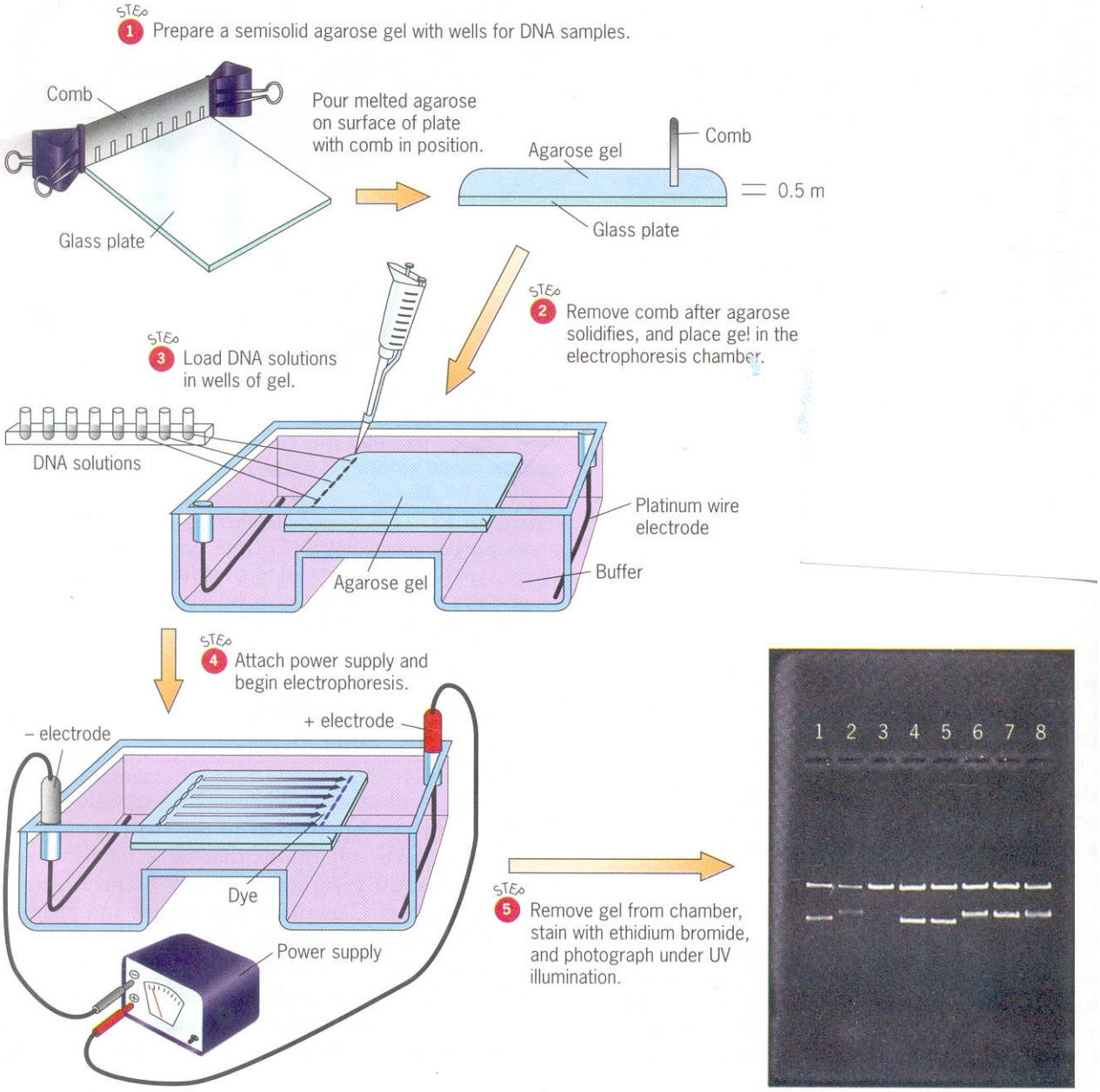


Fig. 14.1. The separation of DNA molecules by agarose gel electrophoresis. The DNAs are dissolved in loading buffer with density greater than that of the electrophoresis buffer so that DNA samples settle to the bottoms of the wells, rather than diffusing into the electrophoresis buffer. The loading buffer also contains a dye to monitor the rate of migration of molecules through the gel. Ethidium bromide binds to DNA and fluoresces when illuminated with ultraviolet light. In the photograph shown, lane 3 contained *Eco*RI-cut pUC119 DNA; the other lanes contained *Eco*RI-cut pUC119 DNAs carrying maize glutamine synthetase cDNA inserts.

1. Genomic DNA is digested with one or more restriction enzymes and resulting fragments are separated by electrophoresis on an agarose gel.

2. A piece of nylon membrane placed over it followed by 2-4 filter papers. Thus the membrane is sandwiched between the gel, stack of filter papers and paper towels.

3. Dry filter paper towels are placed in top of pyramid and a glass plate and weight around 5 kg is placed on top.

4. DNA in agarose gel is denatured *in situ* for efficient blotting, gel is pretreated with alkali that denatured the fragments.

5. DNA fragments from the gel get transferred to a membrane by capillary movement of buffer the transfer is allowed for 12 hrs.

6. The paper towels and membrane are removed. The membrane is dried on the filter paper followed by backing in an oven at 80^{o}C for 2 hr.

7. Following the fixation step, the membrane can be used for hybridization labeling reaction using labeled probe (The composition of hybridization buffer can affect the speed of reaction).

8. After hybridization, the membrane is washed to remove unbound probe and regions of hybridization are detected by autoradiographic method.

Southern blotting hybridization refers to a technique of blotting electrophoresed DNA fragments from a gel onto a nitrocellulose filter (membrane) and subjecting it to a hybridization between a specific DNA fragment and a DNA probe and detection of DNA fragments on the membrane (Fig. 14.3).

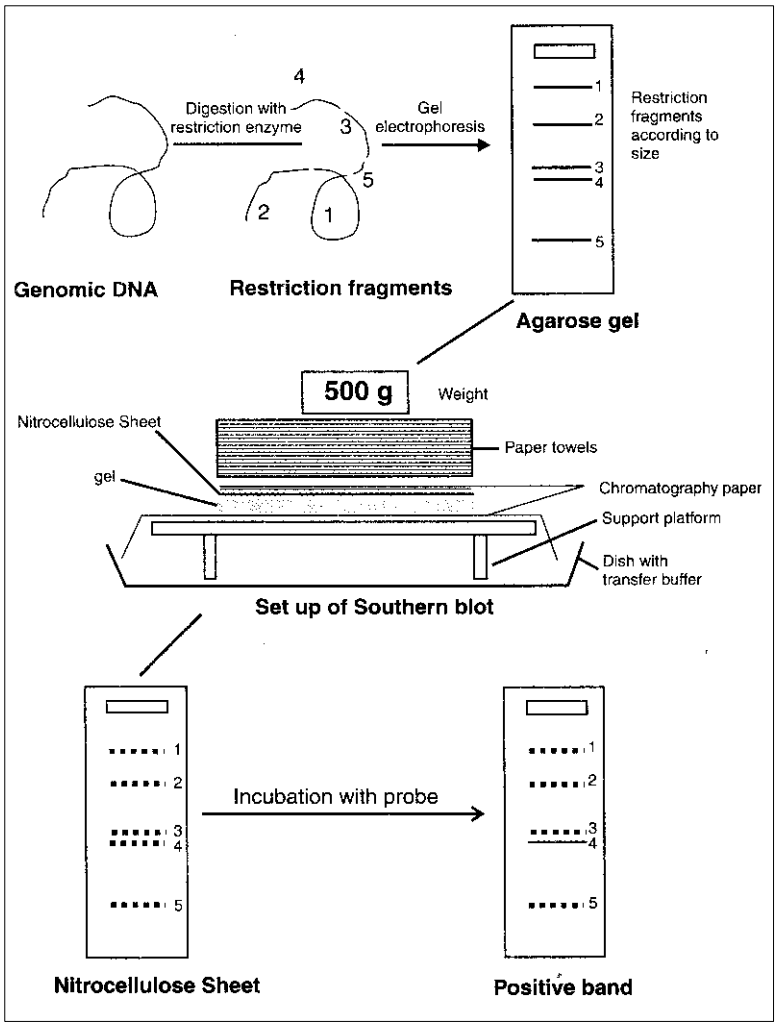


Fig. 14.3. General scheme for southern blotting

14.3 NORTHERN BLOTTING (Analysis of RNAs)

If DNA molecules are transferred from agarose gels to nylon membranes, we expect that RNA molecules separated by agarose gel electrophoresis could be similarly transferred and analyzed. RNA blots are called **northern blots**.

The northern blot procedure is identical to that used for southern blot transfers. However, RNA molecules are very sensitive to degradation by RNases. Thus, care must be taken to prevent contamination of materials. They must be kept denatured during electrophoresis in order to separate them on the basis of size.

After transfer to an appropriate membrane, the RNA blot is hybridized to either RNA or DNA probes just as with a southern blot (Fig. 14.4). The hybridization bands are detected by autoradiography. The nylon membranes or Diazo Benzyloxy Methyl (DBM) papers are used.

Northern blot hybridizations are extremely helpful in studies of gene expression. They can be used to determine when and where a particular gene is expressed since it is mainly used for detection of transcription of a DNA sequences, indicating the gene expression.

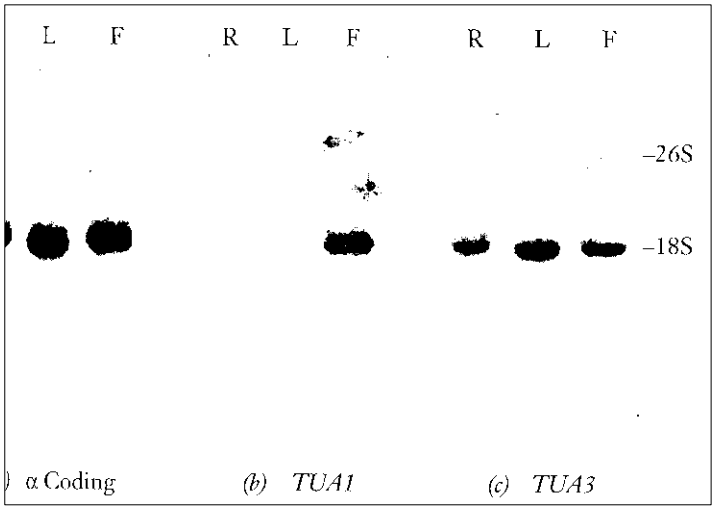


Fig. 14.4 Typical northern blot hybridization data. Total RNAs were isolated from roots (R), leaves (L), and flowers (F) of *A. thaliana* plants, separated by agarose gel electrophoresis, and then transferred to nylon membranes. The autoradiogram shown in (a) is of a blot that was hybridized to a radioactive probe containing an α-tubulin coding coding sequence. This probe hybridizes to the transcripts of all six α-tubulin genes in A. thaliana. The autoradiograms shown in (b) and (c) are of RNA blots that were hybridized to DNA probes specific for the α1- and α3- tubulin genes (*TUA1* and *TUA3*, respectively). The results show that the α3-tubulin transcript is present in all organs analyzed, whereas the α1-tubuln transcript is present only in flowers. The 18S and 26S ribosomal RNAs provide size markers.

14.4 WESTERN BLOTTING (Analysis of Proteins)

A typical western blotting apparatus is shown in Fig. 14.5. Polyacrylamide gel electrophoresis is an important tool for the separation and characterization of proteins. After electrophoresis, the proteins are detected by stainying with coomassie blue. However, the separated polypeptides also can be transferred from the gel to a nitrocellulose membrane, and individual proteins can be detected with antibodies. This transfer of proteins from acrylamide gels to nitrocellulsoe membrane is called western blotting (Fig. 14.5). The application of western blots to identify the chloroplastic form of glutamine synthetase protein is shown in Fig. 14.6.

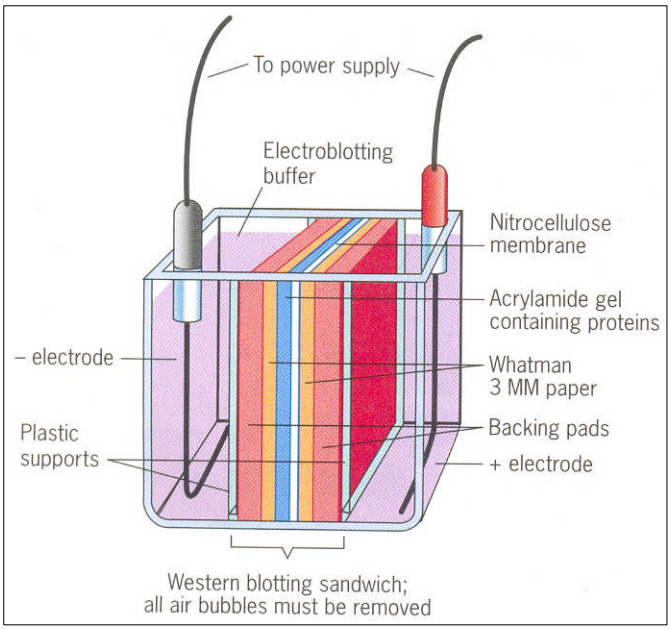


Fig 14.5. A typical western blotting or electroblotting apparatus. An electric current is used to transfer the proteins from a gel to a nitrocellulose membrane placed next to it in the blotting sandwich. All other components of the sandwich function to provide gentle but firm support; tight contact between the gel and the membrane is essential for good transfer.

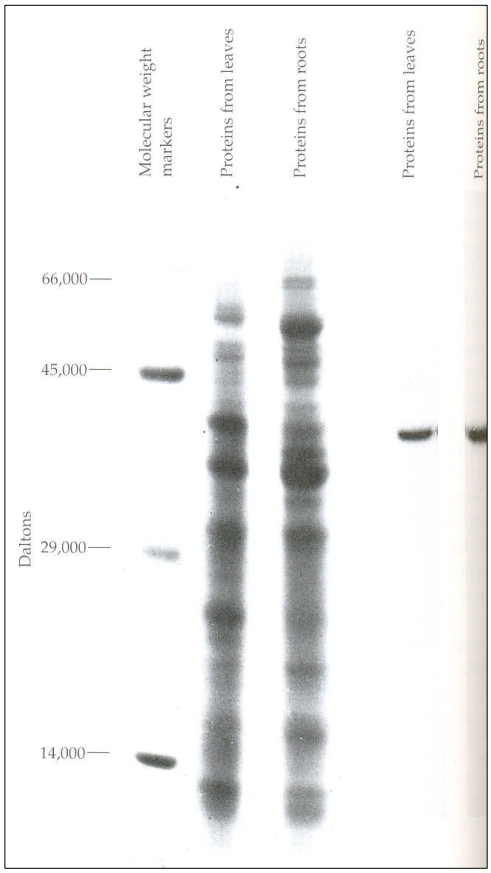


Fig. 14.6. The use of western blots to identify individual proteins separated by polyacrylamide gel electrophoresis, (a) Proteins isolated from roots or leaves of maize were separated by polyacrylamide gel electrophoresis and stained with Coomassie blue. (b) The chloroplastic form of glutamine synthetase was identified by western blot analysis of the gel shown in (a).

The technique is as follows:

(i) Protein bands are separated on polyacrylamide gel by electrophoresis.

(ii) The protein bands on the gel are transferred onto a nitrocellulose membrane by capillary movement of the buffer similar to southern blotting.

(iii) Specific protein bands are identified by use of antibodies for detecting specific antigens, e.g. Lectins are used as probes for identification of glycoproteins.

(iv) Unreacted antibody is washed away and those protein bands that have found the antibody become visible and are identified.

(v) A second antibody (labelled with radioactive method) is used to find with the first antibody that is bound to a protein. This identification involving the binding of the first antibody to the protein band followed by the binding of a second antibody (labelled) to the first antibody that is found to the protein band is called as **Sandwich reaction**.

14.5 PROBES

Probes are short, single stranded nucleotide sequences (15-30 bases long) of DNA or RNA labelled radioactively or non-radioactively to aid detection after their hybridization with complementary sequences in the target gene.

Probes can be prepared in various ways. Highly purified mRNA molecules can be used as probes or cDNA molecule prepared from mRNA by reverse transcription can be used. Probes can also be prepared by PCR by the generation of single-stranded copies of the DNA sequence. Synthetic oligonucleotides can be prepared for use as probes, if the nucleotide sequence of the target gene is known.

14.6 POLYMERASE CHAIN REACTION (PCR)

Molecular biology relies on the techniques that enable the detection and capture of minute quantities of Nucleic acids. Such an alternative, which is now routinely used in molecular cloning to amplify specific DNA or RNA segments within large populations of polynucleotide chains is called PCR (Polymerase Chain Reaction).

- The PCR technique developed into an entirely *in vitro* enzymatic amplification technique in 1985 by Kary Mullis.

- With the introduction of the PCR, more sensitive levels of detection and higher levels of amplification of specific sequences are achieved and in less time.

- PCR is a relatively simple technique by which a DNA or cDNA template is amplified many thousand or million folds quickly and reliably.

A typical amplification reaction includes:

1. The sample of target DNA
2. A thermostable DNA polymerase
3. Two oligonucleotide primers
4. Deoxy nucleotide triphosphates (dNTPs)
5. Reaction buffer
6. Magnesium ions

The components of the reaction are mixed and the reaction is placed in a thermal cycles, which is an automated instrument that takes the reaction through a series of different temperatures for ranging amounts of time. The series of temperature and time adjustments is referred to as one cycle of amplification.

Each PCR cycle theoretically doubles the amount of targeted template sequence in the reaction. Ten cycles theoretically multiply the amplicon by a factor of about one thousand, 20 cycles, by a factor of more than a million in a matter of hours. In principle and virtually in practice, 2^n copies of the duplex segment bordered by primer are produced, where 'n' is the number of cycles. Each cycle of PCR amplification consists of a number of steps which produce two oligo nucleotide primed single stranded DNA template set up the polymerization reaction and synthesize a copy of each strand of the template being targeted. These steps should be optimized for each template and primer pair combination.

Step-1 (Denaturation): The initial step in a cycle denatures the target DNA by heating it to 95^oC or higher for 15 seconds to 2 minutes. In the denaturation process, the two intertwined strands of DNA separate from one another, producing the necessary single-stranded DNA template for the thermostable polymerase.

Step-2 (Annealing): The next step of a cycle reduces the temperature to approximately 40^oC-60^oC. At this temperature, the oligo nucleotide primers can form stable associations (anneal) with the separated target DNA strands and serve as primers.

Step-3 (Extension): Here, the synthesis of new DNA begins when the reaction temperature is raised to the optimum for the thermostable DNA polymerase. For most thermostable DNA polymerases, the optimum temperature is 72^oC.

Extension of the primer by the thermostable polymerase lasts approximately 1-2 minutes. This steps completes one cycle.

The next cycle begins with a return to 95^oC for denaturation. After 20-40 cycles the amplified nucleic acid may then be analysed for size, quantity, sequence etc., or used in further experimental procedures, e.g. cloning.

Thus after completion of step 3, the targeted sequences on both strands are copies and four strands are produced (Fig. 14.7). The three step cycle is repeated which yields 8 copies from 4 strands. The cycle is repeated to 20-40 cycles. If PCR is 100% efficient, when suitable temperature and primers are used, one target molecule would become 2^n after 'n' cycles. To achieve a perfect **PCR** the reaction conditions must be varied accordingly. The target sequence can be a gene from genomic DNA or from a cDNA prepared from mRNA by reverse transcription. The latter process where reverse transcription is followed by PCR is called as RT-PCR. After PCR cycles, the amplified DNA segment is purified by gel electrophoresis and used as desired.

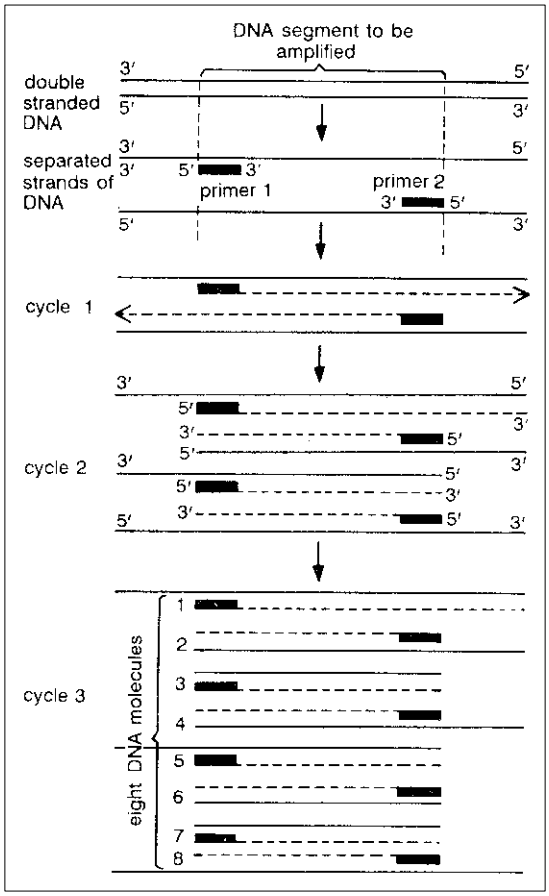


Fig. 14.7 Basic reaction of the PCR (only three cycles of PCR are shown; in each cycle primers are shown by solid boxes, template strands are shown by continuous lines and newly synthesized strands are shown by broken lines).

The PCR technique has been modified in a variety of ways to suit specific situations and its applications. The variations are inverse PCR, anchored PCR, nested PCR etc.

14.6.1 Variations of P.C.R. and Applications

The PCR technique has been modified in a variety of ways to suit specific situations and applications. Some of these variations are **inverse PCR**, anchored PCR, nested PCR etc.

PCR has many exciting applications.

(i) PCR can be used to amplify a specific gene present in different individuals of a species or from a crime spot.

(ii) PCR has been used to study DNA polymorphism in the genomes or related genotypes of a species using known random primers. Such an application of PCR generates random amplified polymorphic DNA (RAPD) which is detected as bands after elecrtrophoresis. The bands of different species or strains can be compared and polymorphism analyzed.

(iii) PCR can be used to detect the presence of a gene transferred into an organism (transgene) for the amplification of DNA from the transgenic organism.

(iv) PCR is used for diagnosis of diseases and causal micro-organisms and also for detection of genetic diseases such as sickle cell anaemia, phenyl ketonuria and muscular dystrophy.

(v) PCR can be used to determine the sex of embryos that have been fertilized *in vitro* (in cattle) using Y-chromosome specific primers before their implantation in the uterus.

14.7 D.N.A. FINGER PRINTING

The use of fingerprints in forensic cases is based on the fact that no two individuals have identical prints. The human genome contains 3×10^9 nucleotide pairs. Recent evidence has demonstrated that the human genome contribute to the evolutionary divergence of genomes. The human genome contains large families of DNA polymorphisms of many different types. These polymorphisms can be used to produce **DNA fingerprint**. DNA fingerprints can be prepared from minute amounts of blood, semen, hair or other cells. DNA is extracted from these cells, amplified by PCR and analyzed with chosen DNA probes by Southern blot procedure.

Cloning and genomic sequencing projects have identified many repetitive sequences in the human genome. Some of these are simple nucleotide repeats that vary in number between individuals but are inherited (VNTRs). There variable number tandem repeats (VNTRs) are important components of DNA finger prints.

Genetic polymorphism: When the same region or locus of a chromosome has two slightly different DNA sequences in different chromosomes or individuals of the same species, these are described as polymorphic and the locus is said to show **genetic polymorphism** within a diploid individual, two alleles of a particular single copy of gene could be different by just one nucleotide and gene can be described as polymorphic. Similarly in a population of individuals, if other alternative DNA sequences exist for the same gene, there would be many different alleles in the whole population. Mutation is a change in the nitrogen bases of a gene. Thus mutational events that cause genetic poymorphism create the single nucleotide polymorphisms (SNPs), but they also create length variations in arrays of repeated sequences. Each different length of repeat is a

different form and each is an example of simple sequence length polymorphism. DNA fingerprinting is based on the analysis of feature of single nucleotide polymorphism.

Human genome contains numerous short DNA sequences that are present as tandem repeats of varied lengths at several chromosomal locations. These VNTRs are important components of DNA fingerprints and are useful in preparing DNA fingerprints.

Thus these polymorphisms can be used to produce DNA fingerprints, specific banding patterns on southern blots of genomic DNA, cleaved with a specific restriction enzyme and hybridized to appropriate DNA probes.

2.7.1 Applications

Paternity tests: In the past, uncertain paternity have been decided by comparing the blood types of the child, the mother and possible father. These blood type comparisons contribute little towards a positive identification of the father. In contrast when DNA fingerprints are compared, all the bands in the child's DNA print should be present in the combined DNA prints of the parents. Approximately half of the bands in the child's DNA print will result from DNA sequences inherited from the mother and the other half from DNA sequences inherited from the father (Fig. 14.8). Thus they help in the positive identification of the true father in cases of disputed parentage.

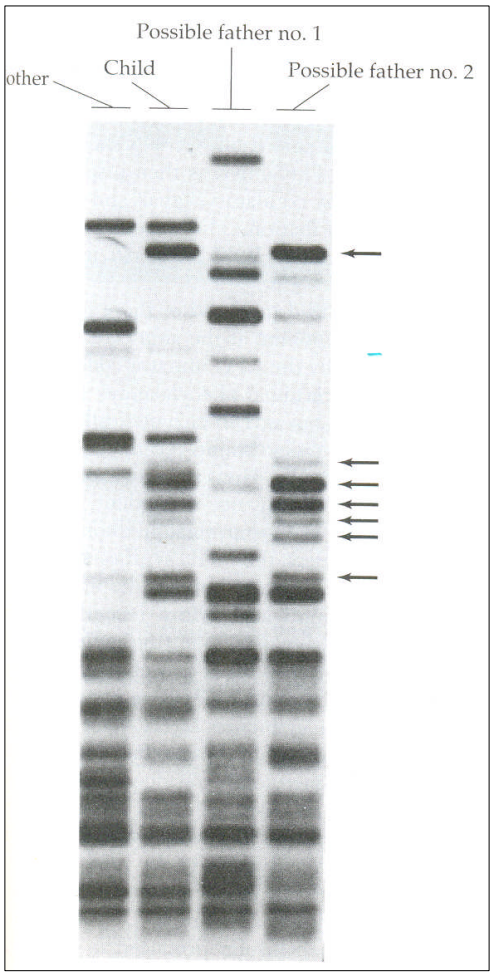


Fig. 14.8 DNA fingerprints of a mother, her child, and two men, each of whom claimed to be the child's father. Arrows mark bands that identify male no. 2 as the biological father.

Forensic cases

1. DNA fingerprints were first used as evidence in a criminal case in 1988.
2. One type of DNA fingerprints, VNTR are used to identify individuals in suspects of committing the crime. Patterns from different samples are comparedand individual bands are assigned to particular identity.

3. They are used for determining pedigree and parentage of an individual.
4. DNA fingerprints provide a tool by which the differences in the genome can be detected and recorded which are useful in the studies of evolutionary biology.

14.8 SUMMARY

The molecular analysis of DNA, RNA and proteins is described by the application of blotting techniques. DNA molecules separated by agarose gel electrophoresis are transferred to nylon membranes to produce DNA gel blots called southern blotting. The same procedure is applied to RNA molecules. When proteins are transferred from gels to membranes and detected with antibodies, the products are called western blots. The amplification of DNA by polymerase chain reaction (PCR) is described. DNA finger prints that detect and record polymorphisms in the genomes of individuals is explained. Applications of PCR and DNA finger printing in paternity tests and forensic cases are discussed.

14.9 MODEL QUESTIONS

1. Describe the procedure involved in different blotting techniques you have studied.
2. Explain the method of polymerase chain reaction and its uses.
3. Write short notes on:
   a) DNA finger printing
   b) Western blotting
   c) Probe
   d) Agarosa gel Electrophoresis

14.10 REFERENCE BOOKS

1. P.J. Russell, 1987. **Essential Genetics**, Blackwell Publishers.
2. F.J. Ayala & J.A. Kiger, Jr. 1980. **Modern Genetics**. The Benjamin/Cummings Publ. Com.
3. M.W. Strickberger. **Genetics** – Prentice Hall of India, New Delhi.
4. Tamarin. **Principles of Genetics**.
5. H.S. Chawla, 2002. Introduction to Plant Biotechnology. Second Edition, Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi.

*****

PROF. T.N. MARY

Lesson 15

GENETIC ENGINEERING -I

CONCEPT OF RECOMBINANT DNA TECHNOLOGY

15.0. Objectives : This lesson deals with concept of r DNA technology, important enzymes and vectors used in gene transfer; and different methods of transformation and screening of recombinants are described in detail.

15.1. Introduction

Recombinant DNA technology, also called gene cloning or molecular cloning, mainly aimed at the transfer of genetic information [DNA] from one organism to another. The major steps in this technology are - The DNA from a donor organism is extracted, enzymatically cleaved (cut, digested), and joined (ligated) to a vector DNA to form a new, recombined DNA molecule (cloning vector-insert DNA construct, DNA construct). This cloning vector-insert DNA construct is transferred into a expression host system. The introduction of DNA into a bacterial host cell is called transformation. Those host cells that take up the DNA construct (transformed cells) are identified and selected (separated, isolated) from those that do not. Different cloning strategies have to be followed to ensure effective production of protein in prokaryotic and eukaryotic organisms.

15.2. Enzymes and Vectors used in gene transfer

For molecular cloning, both the source DNA that contains the target sequence and the cloning vector must be consistently cut into discrete and reproducible fragments. Subjecting

isolated chromosomal DNA either to passage through a small-bore needle or to sonication produces double stranded pieces of DNA that may range from 0.3 to 5 kilobase pairs (kb) in length. Unfortunately, these simple procedures induce breaks randomly, so each time a DNA sample is treated, a different set of fragments is generated. After the discovery of bacterial enzymes that cut DNA molecules internally at specific base pair sequences the molecular cloning became more feasible. These enzymes are called type II restriction endonucleases.

15.2.1. Restriction endonucleases

One of the first characterized restriction endonucleases was from the bacterium *Escherichia coli,* and it was designated as *EcoRI.* In addition to *EcoRI*, hundreds of other type II restriction endonucleases have been isolated from various bacteria. The naming protocol for these enzymes is the same as that for *EcoRI;* the genus is the capitalized letter and the first two letters of the species name are in lower case letters. The strain designation is often omitted from the name, and roman numerals are used to designate the order of characterization of different restriction endonucleases from the same organism. For example, *HpaI* and *HpaII* are the first and second type II restriction endonucleases that were isolated from *Haemophilus parainfluenzae.*

Recognition sites are the palindromic sequences where the type II restriction endonucleases bind and cut a DNA molecule. The length of the recognition site for different enzymes can be four, five, six, eight, or more nucleotide pairs. Because of the frequency with which their recognition sites occur in DNA, restriction endonucleases that cleave within sites of four (four-cutters) and six (six-cutters) nucleotide pairs are used for most of the common molecular cloning protocols. After cleavage by the endonucleases, three types of DNA molecules will result, they are either with 5'- phosphate extensions (protruding ends, sticky ends) or with 3' – hydroxyl extensions, or with blunt ends.

EcoRI enzyme binds to a DNA region with a specific palindromic sequence (the two strands are identical in this region when either is read in the same polarity, i.e., 5' to 3') of 6 base pairs (bp) and cuts between the guanine and adenine residues on each strand.

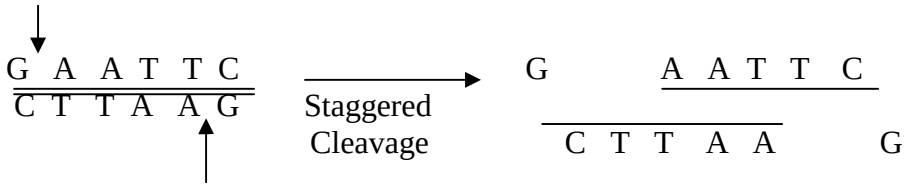


Fig.15.1. Staggered cleavage of a SHORT fragment of DNA by *EcoRI*

It specifically cleaves the inter nucleotide bond between the oxygen-of the 3' carbon of the sugar of one nucleotide and the phosphate group attached to the 5' carbon of the sugar of the adjacent nucleotide. The symmetrical staggered cleavage of DNA by *EcoRI* produces two single-

stranded, complementary cut ends, each with extensions of four nucleotides. In this case, each single-stranded extension ends in a 5' -phosphate group, and the 3' -hydroxyl group from the opposite strand is recessed.

The importance of the type II restriction endonucleases for gene cloning cannot be overstated. When a DNA sample is treated with one of these enzymes, the same set of fragments is always produced, assuming that all of the recognition sites are cleaved. Moreover, physical maps that designate the linear order of restriction endonuclease sites on a specific piece of DNA can be constructed by treating the DNA molecule singly with different restriction endonucleases and then with combinations of restriction endonucleases. The positions of the cleavage sites can be deduced from an analysis of fragment sizes, which are determined by agarose gel electrophoresis. It can be deduced that this piece of DNA has two BamHI sites and two EcoRI sites in a specific order with a specified number of base pairs separating the sites.

15.2.2. Ligases : Enzymes next to endonucleases essential for genetic engineering are ligases. These enzymes are usually isolated from bacteriophase T_4. When the extended ends that are created by restriction enzyme (e.g., BamHI) cleavage are aligned, the hydrogen bonds of the four bases that pair are not strong enough to keep two DNA molecules together. Enzyme ligase catalyzes the formation of phospho diester bonds at the ends of DNA strands. It also joins blunt ends that come in contact when they both bind to the enzyme.

15.2.3. Alkaline phosphatases : The extent of re-circularization of plasmid DNA in the ligation reaction is diminished by treating the cut plasmid DNA with alkaline phosphatases, which removes phosphate groups form 5’ ends of the linearized DNA so that these ends cannot be joined to each other by T4 DNA ligase.

15.2.4. Plasmids (Cloning Vectors)

Plasmids are self-replicating, double-stranded, circular DNA molecules that are maintained in bacteria as independent extra chromosomal entities. Virtually all bacterial genera have plasmids. These are of different types - some plasmids carry information for their own transfer from one cell to another (**F Plasmids**), others encode resistance to antibiotics (**R Plasmids**), others carry specific sets of genes for the utilization of unusual metabolites (**degradative plasmids**), and some have no apparent functional coding genes (**cryptic plasmids**). The size of Plasmids usually ranges from less than 1 to more than 500 kb. Every plasmid must have a sequence that functions as an origin of DNA replication; without which it cannot replicate in a host cell. They also differ in their number of copies per host cell. Some plasmids are represented by 10 to 100 copies per host cell; called high-copy-number plasmids, others maintain as 1 to 4 copies per cell are called low-copy-number plasmids.

In a cell plasmids represent approximately 0.1 to 5.0% of the total DNA. Plasmids are designated as single incompatibility group if two or more types cannot coexist in the same host cell. But plasmids from different incompatibility groups can be maintained together in the same cell. This coexistence is independent of the copy numbers of the individual plasmids. Some microorganisms have been found to contain as many as 8 to 10 different plasmids. In these cases, each plasmid can carry out different functions and have its own unique copy number, and each

belongs to a different incompatibility group. The plasmids can be called as narrow- and broad-host-range plasmids, respectively, if replicate in only one species (because of specific origin of replication) of host cell and can replicate in a number of bacterial species (due to the presence of less specific origin of replication).

The important features required for a plasmid to act as a high-quality cloning vector are (1) small size is preferable, because the efficiency of transfer of exogenous (foreign) DNA into E. *coli* decreases significantly in case of plasmids with more than 15 kb length; (2) contain unique (single) restriction endonuclease recognition sites into which the insert DNA can be cloned; and (3) one or more selectable genetic markers are necessary for identifying recipient cells that carry the cloning vector-insert DNA construct. Therefore, to meet these needs, plasmid cloning vectors have to be genetically engineered.

15.2.4.1. Plasmid pBR322 :

In the 1980s, one of the best-studied and most often used "general-purpose" plasmid cloning vector was pBR322. In general, plasmid cloning vectors are designated by a lowercase p, which stands for plasmid, and some abbreviation that may be descriptive or, as is the case with pBR322, anecdotal. The BR of pBR322 recognizes the work of the researchers F. Bolivar and R. Rodriguez, who created the plasmid; and 322 is a numerical designation that has relevance to these workers. Plasmid pBR322 contains 4,361 bp. As shown in Fig. 15.2. pBR322 carry two antibiotic resistance genes. One confers resistance to ampicillin (Ampr), and the other confers resistance to tetracycline (Tetr). This plasmid also has unique *BamHI*, HindIII, and *SalI* recognition sites within the Tetr gene; a unique *PstI* site in the Ampr gene; a unique EcoRI site that is not within any coding DNA; and an origin of replication that functions only in *E. coli.* It is maintained at a high copy number in *E. coli* and cannot be readily transferred to other bacteria.

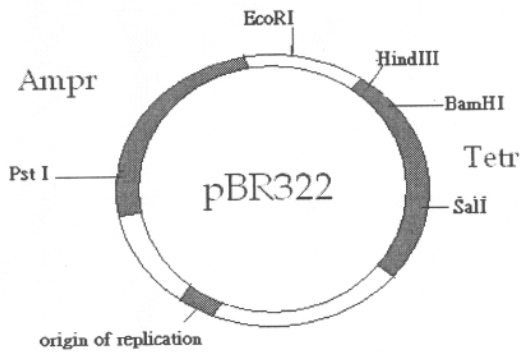


Fig. 15. 2. Genetic map of the plasmid cloning vector pBR322.

Working of pBR322 : Purified, closed circular pBR322 molecules are cut with a restriction enzyme that lies within either of the antibiotic resistance genes and cleaves the plasmid DNA only once to create single, linear, sticky-ended DNA molecules. These linear molecules are combined with prepared target DNA from a source organism. This DNA has been cut with the same restriction enzyme, which generates the same sticky ends as those on the plasmid DNA. The DNA mixture is then treated with T_4 DNA ligase in the presence of ATP. Under these conditions, a number of different ligated combinations are produced, including the original closed circular plasmid DNA. To reduce the amount of this particular unwanted ligation product, the cleaved plasmid DNA preparation is treated with the enzyme alkaline phosphatase to remove the 5'-phosphate groups from the lenearized plasmid DNA. As a consequence, T_4 DNA ligase cannot join the ends of the dephosphorylated linear plasmid DNA. However, the two phosphodiester bonds that are formed by T_4 DNA ligase after the ligation and circularization of alkaline phosphatase-treated plasmid DNA with restriction endonuclease-digested source DNA, which provides the phosphate groups, are sufficient to hold both molecules together, despite the presence of two nicks. After transformation, these nicks are sealed by the host cell DNA ligase system. In addition, although unwanted, fragments from the source DNA are also joined to each other by T_4 DNA ligase.

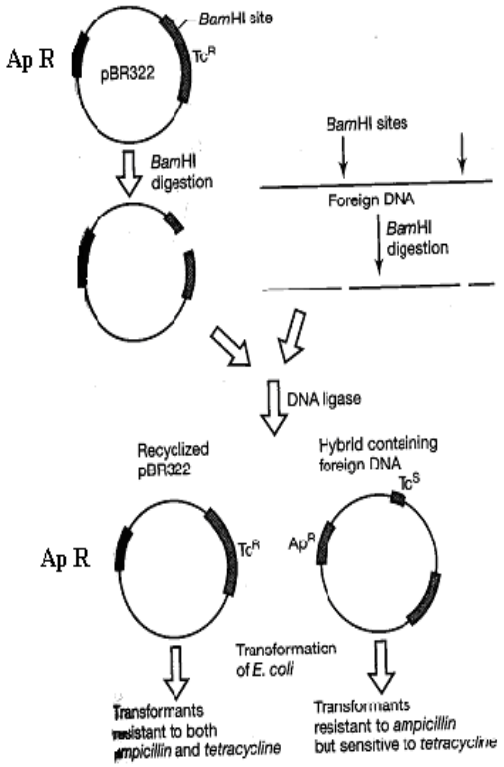


Fig.15.3. The use of plasmid pBR322 as a cloning vector, showing the working nature.

15.2.4.2. Plasmid pUC19

The plasmid pBR322 was a well-conceived cloning vector. It has only a few unique cloning sites, however, and its use requires a time-consuming selection procedure. Thus, it was inevitable that other systems would be developed. For example, the plasmid pUC19 is 2,686 bp long and contains an ampicillin resistance gene, a regulatable segment of the ,B-galactosidase gene *(IaeZ')* of the lactose operon of E. *coli*, a *lacI* gene that produces a repressor protein that regulates the expression of the *IaeZ'* gene, a short sequence with multiple unique cloning sites (i.e., EcoRI, *SacI, KpnI, XmaI, SmaI,* BamHI, *XbaI, SalI, HindI, Aeel, BspMI, PstI, SphI,* and *HindIII),* and the origin of replication from pBR322 (Fig. 15.3.)

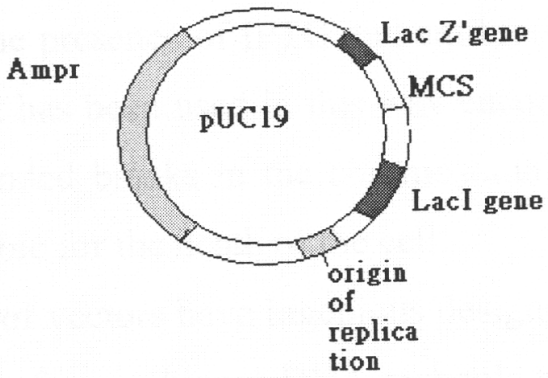


Fig. 15.4. Genetic map of the plasmid cloning vector pUC19.

For a pUC19 cloning experiment, DNA from a source organism is cut with one of the restriction endonucleases for which there is a recognition site in the multiple cloning sequence. This source DNA is mixed with pUC19 plasmids that have been treated initially with the same restriction endonuclease and subsequently with alkaline phosphatase. After ligation with T_4 DNA ligase, the reaction mixture is introduced by transfore1ation into a host cell which can synthesize that part of β -galactosidase (*Lac Z* α) that combines with the product of the *lacZ'* gene to form a functional enzyme. The treated host cells are plated onto medium that contains ampicillin, IPTG, and X-Gal. Non transformed cells cannot grow in the presence of ampicillin. Cells with recircularized plasmids can grow with ampicillin in the medium; because they can form functional β-galactosidase, they produce blue colonies. In contrast, host cells that carry a plasmid-cloned DNA construct produce white colonies on the same medium.

In addition to pBR322 and the pUC-series plasmids, many other cloning vectors have been devised. Some vectors have inventive selection systems to identify clones with insert-vector

constructs. For example, a vector that is derived from the pUC series carries a gene that, when expressed, encodes a protein that kills the cell. This cell-killing gene (suicide gene) is fused in the correct reading frame to the *lacZ'* gene. A cell with an intact plasmid and no IPTG in the medium does not synthesize the suicide protein. With an insert and IPTG, a nonfunctional suicide protein is produced because the insert, in all likelihood, disrupts the reading frame of the suicide gene. By contrast, cells with a plasmid and no insert in the presence of IPTG synthesize the suicide protein and are killed. Non transformed cells are sensitive to an antibiotic, whereas recombinant clones have as part of the vector a gene that confers resistance to the antibiotic. In other words, the only surviving cells in the presence of IPTG and antibiotic are those that carry insert DNA. One suicide gene that has been used in this way encodes an enzyme that prevents the rejoining of double-stranded breaks in the chromosomal DNA. These DNA breaks accumulate and are responsible for the death of the cell.

Although a number of vectors have ingenious designs, in principle they all retain the two basic requirements of recombinant DNA technology: a choice of cloning sites and an easy way to identify cells that have plasmid cloned DNA constructs. It should be noted that unique restriction endonuclease sites have a dual function in recombinant DNA research: they are essential for inserting DNA into a cloning vector, and they allow an inserted DNA sequence to be recovered from the vector. In other words, after a piece of DNA has been cloned into a site, it can be retrieved by cutting a purified plasmid-cloned DNA construct with the original restriction endonuclease because the insertion event creates two recognition sites at the ends of the cloned DNA sequence. Occasionally, the initial recognition site is destroyed. Then the cloned piece of DNA is difficult to retrieve. A recovered DNA fragment can be cloned into cloning vectors for DNA sequencing or vectors that have been specially designed to achieve high levels of expression (transcription and translation) of the cloned gene.

15.2.4.3. Bacteriophage λ Vectors

The plasmid-based vectors used for cloning DNA molecules generally carry up to 10 kb of inserted DNA. However, for the formation of library, it is often helpful to be able to maintain larger pieces of DNA. To meet this, bacteriophage lambda (λ) has been developed as a cloning vector.

In its life-cycle, bacteriophage λ infects *E.coli*, and, after injection of the viral DNA, two possibilities exist. Bacteriophage λ can enter a lytic cycle which, after 20 minutes, leads to the lysis of the host cell and the release of about 100 phage particles. Alternatively, the injected bacteriophage λ DNA can be integrated into the *E. coli* chromosome as a prophage and be maintained more or less indefinitely in the cell (lysogeny). However, under conditions of nutritional or environmental stress, the integrated bacteriophage λ DNA can be excised and enter a lytic cycle. The bacteriophage λ DNA is about 50 kb in length, of which approximately 20 kb is essential for the integration-excision (I/E) events. For forming genomic libraries, it was reasoned that this 20 kb of DNA could be replaced with 20 kb of cloned DNA. This DNA molecule could then be perpetuated as a "recombinant" bacteriophage λ through compulsory lytic cycles.

One of the many bacteriophage λ cloning vectors that have been devised has two B*amHI* sites that flank the I/E region. When purified DNA from this bacteriophage is cut with *BamHI*, three

segments are created. The so-called left arm (L region) contains the genetic information for the production of heads and tails, the right arm (R region) carries the genes for DNA replication and cell lysis, and the middle fragment (I/E) has the genes for the integration-excision processes. The objective of this genetic engineering protocol is to replace the middle segment of the λ DNA with cloned DNA that is approximately 20 kb in length. The source DNA is cut with *BamHI,* and DNA pieces that are 15 to 20 kb in length are isolated. The two DNA samples are combined and incubated with T_4 DNA ligase. Then empty bacteriophage heads and tail assemblies are added. Under these conditions, 50-kb units of DNA are packaged into the heads, and infective bacteriophage particles are formed. Other products from the ligation reaction cannot be packaged, because they are either too large (>52 kb) or too small «38 kb). Recombinant bacteriophage λ can undergo lytic cycles only in an E. *coli* strain that does not allow reconstituted bacteriophage λ with intact I/E regions to grow. Recombinant bacteriophage λ are perpetuated by continual growth on fresh E. *coli.*

15.2.4.4. Cosmids

The cloning vectors, called cosmids, can carry 40 kb of cloned DNA and can be maintained as plasmids in E. *coli.* Cosmids combine the properties of plasmids and bacteriophage λ vectors. For example, the commonly used cosmid pLFR-5 (approximately 6 kb) has two cos sites (cos ends) from bacteriophage λ separated by a *ScaI* restriction endonuclease site, a multiple cloning sequence with six unique sites (HindIII, *Pstl, SalI, BamHI, SmaI,* and *EcoRI),* an origin of DNA replication *(ori),* and a tetracycline resistance (Tetr) gene. This cosmid can carry about 40 kb of cloned DNA. For this vector, pieces of DNA that are approximately 40 kb in length are purified by sucrose density gradient centrifugation from a partial restriction endonuclease digestion of the source DNA with *BamHI* . The pLFR-5 DNA is cleaved initially with *Scal* and then with *BamHI.* The two DNA samples are mixed and ligated. Some of the ligated products will have a 40-kb DNA piece inserted between the two fragments that are derived from the digestions of the pLFR-5 DNA.

Once inside the host cell, the cos ends, which were cleaved during the in vitro packaging, base pair and enable the linear DNA to circularize. This circular form is stable, so the cloned DNA can be maintained as a plasmid-insert DNA construct because the vector contains a complete set of plasmid functions. Moreover, the tetracycline resistance gene allows colonies that carry the cosmid to grow in the presence of tetracycline; non transformed cells are sensitive to tetracycline and die. The advantages of using cosmids are two fold. First, because the cloning capacity of cosmid is greater than that of plasmids, gene clusters and large genes are easier to clone. Second, for creating a library, a larger insert in the cloning vehicle means that fewer clones have to be screened.

15.2.4.5. Very-large-Insert Bacterial Vector Systems

A vector system that carries large inserts (>100 kb) is helpful for the analysis of complex eukaryotic genomes. For example, very-large-insert vector systems are indispensable for the mapping of the human genome and the discovery of human genes. In contrast to a small-insert library, a large insert genomic library is more likely to include all of the genetic material of the

organism, there are fewer clones to maintain, and there is a good chance that most or all of a particular gene is present in a single clone. A low-copy number E. *coli* plasmid vector based on the PI bacteriophage cloning system has been devised for cloning inserts that are from 100 to 300 kb long. These insert vector constructs are called bacteriophage PI-derived artificial chromosomes. Similarly, the F plasmid (F-factor replicon, sex plasmid, fertility plasmid) of E. *coli,* which is present at 1 to 2 copies per cell, has been used along with the *lacZ'* selection system of the pUC vectors to construct an extremely stable cloning vector that carries DNA inserts from 150 to 300 kb. These F-plasmid-based insert vector constructs are called bacterial artificial chromosomes (BACs).

15.3. Expression systems : E. *coli* is the system of first choice for the expression of many heterologous proteins. It is the best choice because, genetic manipulations are straightforward, culturing is easy and inexpensive, and many foreign proteins are well tolerated and may be expressed at high levels.

However, *E. coli* is not always the host of choice. For example, proteins whose full biological activity requires post translation modifications (like glycosylation or cleavage at specific sites) may be best expressed in a eukaryotic host. Therefore, proper expression system to be selected depending on the type of protein. Four expression systems have been developed so far. They include *E. coli, Bacillus subtilis, Saccharomyces cerevisiae,* and cultured insect cells.

15.3.1. Expression in E. coli : Small cytosolic proteins and polypeptides (<100 residues in length) are best expressed as fusion proteins composed of carrier sequences linked by a standard peptide bond to the target protein. The advantage being – the carrier stabilizes the protein of interest against intra cellular degradation and provides a ligand-binding site that can be used for affinity purification of the protein at the time of isolation and purification of the protein from the host. In E. coli, the cytosolic proteins of >100 residues in length are often unstable or form insoluble inclusion bodies.

15.3.2. Expression in *Sacchromyces cerevisiae* : Yeasts are exploited in two ways. First they may be used as hosts for the isolation of mammalian cDNAs by complementation, because many mutants of *S. cerevisiae* can be complemented by expression of the homologous mammalian proteins. Second, for the isolation of genes that encode pairs of interacting proteins, by developing selective systems.

15.3.3. Expression in *Bacillus subtilis* : *B. subtilis* has a well–developed secretary system, and recombinant proteins can often be delivered in to the medium in high yield in a soluble and active form. In spite of the advantage in extracting the proteins in to the medium this system has a disadvantage that *B. subtilis* also secretes proteases with high degradability. However, some proteins exclusively expressed in *B. subtilis* system but not in *E. coli.*

15.3.4. Expression in cultured insect cells : The baculovirus expression system is used to generate large quantities of recombinant proteins in cultured insect cells. The main advantage of the system is that proteins secreted from insect cells undergo post translational change, glycosylation. However, a major disadvantage is that expression of the foreign protein occurs

during acute lytic infection of cells with a recombinant baculovirus stock. The baculovirus most commonly used is *Autographa californica* nuclear polyhedrosis (AcNPV or AcMNPV).

15.4. Methods of genetic transformation in prokaryotes

After construction of a vector-DNA construct, the next important step is transferring the recombinant DNA molecule in to a suitable expression vector. The procedures differs for prokaryotes and eukaryotes. In prokaryotes, it is mainly done by $CaCl_2$ chemical treatment method, electroporation or by conjugation.

15.4.1. Transferring DNA into *E. coli* by chemical method:

Transformation is the process of introducing free DNA into a bacterial cell. For E. *coli,* which is the main host cell for recombinant DNA research, the uptake of plasmid DNA is usually achieved by treating cells with ice-cold $CaCl_2$ and then exposing them for approximately 90 s to a high temperature (42°C). This method has a maximum transformation frequency, which is defined as the fraction of the cell population that can be transformed, of about 1 transformed cell per 1,000 cells (10^{-3}). For this procedure, the transformation efficiency, which is defined as the number of transformants per microgram of added DNA, is approximately 10^7 to 10^8 transformed colonies per microgram of intact plasmid DNA. Although a 100% transformation frequency would be ideal, selection schemes that enable plasmid-transformed cells to be readily identified overcome the drawback of a low transformation frequency. .

The mechanism of action of the $CaCl_2$-heat transformation procedure in E. *coli* is not known. It is assumed that the bacterial cell wall is broken down in localized regions, a condition that allows the plasmid DNA in solution to be taken into the interior of the bacterial cell. Bacterial cells that are able to take up DNA are said to be competent. For E. *coli,* competence must be induced. In some other bacteria, it occurs naturally and sometimes can be enhanced by the use of specific growth media or growth conditions.

15.4.2. Electroporation

Subjecting bacteria to a high strength electric field can induce the uptake of free DNA. This procedure is called electroporation. The experimental protocols for electroporation are different for various bacterial species. For E. *coli,* the cells (~ 50 μL) and DNA are placed in a chamber fitted with electrodes, and a single pulse of approximately 25 microfarads, 2.5 kilovolts, and 200 ohms is administered for about 4.6 milliseconds (ms). This treatment yields transformation efficiencies of 109 transformants per microgram of DNA for small plasmids (about 3 kb) and 106 for large plasmids (about 136 kb). Similar conditions are used to introduce BAC vector DNA into E. *coli.* Thus, electroporation is an effective way to transform E. *coli* with plasmids containing inserts that are longer than 100 kb. Because it is likely that an appropriate set of electroporation conditions can be found for all bacterial species, this procedure will become the standard way for transforming bacteria.

Very little is known about the mechanism of DNA uptake during electroporation. 1t has been deduced, along the lines of the explanation of chemically induced transformation, that transient pores are formed in the cell wall as a result of the electroshock and that, after contact with the lipid bi layer of the cell membrane, the DNA is taken into the cell.

15.4.3. Conjugation

For some bacteria, the natural transmission of plasmid DNA from a donor cell to a recipient cell has been used to transfer plasmid-insert DNA constructs to a host cell that is not readily transformed. Some plasmids are genetically equipped to form cell-to-cell junctions through which plasmid DNA is transferred from one cell to another. Effective contact between a donor cell and a recipient cell is due to conjugative functions; the mechanical transfer of the DNA is the consequence of mobilizing functions. Most of the plasmids that are used for recombinant DNA research lack conjugative functions, and therefore the DNA of these plasmids cannot be passed to recipient cells by conjugation. However, some plasmid cloning vectors can be mobilized and transferred if the conjugative functions are supplied by a second plasmid in the same cell. Thus, by introducing a plasmid with conjugative functions into a bacterial cell that carries a mobilizable plasmid cloning vector, it is possible to transfer the plasmid cloning vector to a recipient cell that is difficult to transform by other means.

The standard experimental protocol for this procedure entails mixing three strains together. When the cells are close to each other, the conjugative plasmid, which in this case is also mobilizable, can be self-transferred to the cell with the mobilizable plasmid cloning vector. Then, with the help of the conjugative plasmid, the plasmid cloning vector is transferred to a targeted recipient cell. All possible combinations of plasmid transfer occur among the cells, but the genetic features of the strains and plasmids are designed to select for the targeted recipient cells that receive the cloning vector.

15.5. Screening for recombinants

After the transformation step, it is necessary to identify, as easily as possible, those cells that contain plasmid-cloned DNA constructs. Depending on the type of cloning vector used and the site of insertion of the foreign gene in the plasmid vector the screening procedure differ. In a pBR322 system in which the target DNA was inserted into the *BamHI* site, this specific identification is accomplished in two steps.

I STEP : The cells from the transformation mixture are plated onto medium that contains the antibiotic ampicillin. Only those cells that have either intact pBR322 or pBR322-cloned DNA constructs, both of which have an intact Ampr gene, can grow under these conditions. The non transformed cells are sensitive to ampicillin. The BamHI site of pBR322 is within the tetracycline resistance gene , so the insertion of DNA into this gene disrupts the coding sequence and tetracycline resistance is lost. Therefore, cells with these plasmid cloned DNA constructs are resistant to ampicillin and sensitive to tetracycline. Cells with re-circularized pBR322 DNA, however, have an intact Tetr gene and are resistant to both ampicillin and tetracycline.

II STEP : The second step in the selection procedure distinguishes between these two possibilities. Cells that grow on the ampicillin-containing medium are transferred to a tetracycline-containing medium. Each location of inoculating cells on a tetracycline-agar plate corresponds to the site of a colony on an original ampicillin-agar plate. Cells that form colonies on the tetracycline agar plates carry re-circularized pBR322, because, as noted above, these cells are resistant to both ampicillin and tetracycline. Those cells that do not grow on the tetracycline-agar plates, however, are sensitive to tetracycline and carry pBR322-cloned DNA constructs.

Individual cultures that are sensitive to tetracycline are established from each of the colonies on the ampicillin-agar plates; or, more often, cells from colonies that are ampicillin resistant and tetracycline sensitive are pooled and sub-cultured together. Later, additional screening procedures can be conducted to identify those cells that carry a specific pBR322-cloned DNA construct. The *HindIII* and *Sall* sites in the tetracycline resistance gene and the *PstI* site in the ampicillin resistance gene of pBR322 provide alternative cloning locations. When the *PstI* recognition site is used for cloning, the principle of the selection scheme is the same but the antibiotic sensitivities are reversed; thus, the first set of plates contains tetracycline and the second set contains ampicillin.

In case of pUC19, the multiple cloning sequence is incorporated into the *IacZ'* gene in the plasmid without interfering with the production of the functional hybrid , β -galactosidase. Under these conditions, colonies containing unmodified pUC19 appear blue, if the substrate 5-bromo-4-chloro-3-indolyl-, β -o-galactopyranoside (X-Gal) is added to the medium, because it is hydrolyzed by this hybrid β -galactosidase to a blue product. Therefore transformed cells appear white in colour. The reason for this is that, usually, DNA inserted into a restriction endonuclease site within the multiple cloning sequence disrupts the correct sequence of DNA codons (reading frame) of the *lacZ'* gene and prevents the production of a functional LacZ′ protein, so no active hybrid β -galactosidase is produced. In the absence of β -galactosidase activity, the X-Gal in the medium is not converted to the blue compound by these colonies, so they remain white.

15.6. SUMMARY

Different steps in genetic engineering include : (1) DNA that is isolated from an organism that contains the target gene is cut with a restriction endonuclease. (2) A DNA cloning vector that can be perpetuated in a host cell-usually a plasmid-is cut with the same restriction endonuclease used to digest the source DNA. (3) The two DNA samples are mixed, and the molecules are joined by the action of T_4 DNA ligase, In the ligation reaction, some of the joined DNA molecules contain plasmid DNA and a DNA fragment from the source organism. (4) The DNA molecules that were joined by T_4 DNA ligase are used to transform host cells. This transformation produces some cells that carry plasmid-insert DNA constructs. Because the plasmid has a DNA sequence that enables it to be replicated (origin of replication) in the host cell, the entire construct is perpetuated.

Strategies have been devised to select for cells with plasmid-insert DNA constructs. After transformation, cells that have taken up the introduced plasmid are identified (1) by testing for

resistance to specific antibiotics or by looking for a specific colorimetric response in growing cells; (2) by growing the colonies onto a medium that contains ampicillin, IPTG, and X-Gal.

15.7. Model Questions

1. Write detail account on different steps in recombinant DNA technology.

2. Give an account on enzymes and vectors used in genetic engineering .

3. Write short notes on the following

a) enzymes used in genetic engineering
b) plasmid vectors
c) λ phage and cosmid vectors.
d) methods of transformation of DNA into prokaryotes
e) Screening methods of transformed cells

15.8. Reference books

Ratledge & Kristinsen 2001 Basic Biotechnology, University press

Hunt & Liveey (eds) 2000 Functional Genomics, Oxford University press

Krenzer & Massey – Recombinant DNA and Biotechnology : A guide for teachers
2nd ed. ASM press

Walker, J. M. and Rapley, R. 2002.Molicular biology and biotechnology. Fourth edition Panima publ. Corp., New Delhi., 563 pp.

Glick B.R. and Pasternak, J.J. 2003. Molecular Biotechnology- principles and applications of recombinant DNA. Panima Publ. Corp., New Delhi., 499pp.

Old and primrose. 1994. Principles of Gene Manipulation : An introduction to genetic engineering. 5 th ed. Blackwell Scientific publ.

Sambrook, J., E.F. Fritsch, T. Maniatis, 2002. Molecular Cloning : A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York.

Dr. M. RaghuRam

Lesson No. 16

GENETIC ENGINEERING – II

TRANSGENIC PLANTS AND ANIMALS

16.0. Objectives : In this lesson, the application of genetic engineering in developing plants and animals with desirable characters was discussed.

16.1. Introduction

Recombinant DNA technology, which has been used extensively with microbial systems, is also an important tool for the direct genetic manipulation of plants and animals. Considerable effort has gone into developing varieties of plants that produce increased yields and have enhanced

nutritional value. Although much of this work was carried out in three major grains-corn, wheat, and rice-successful breeding programs for other food plants and horticultural species have also been established. Similarly, genetic modification of animals by recombinant DNA technology (transgenesis) includes the introduction of a cloned gene(s) into the genome of a cell, so that subsequently it may be possible to establish true breeding lineages.

16.2. Transgenic plants :

There are three major reasons for developing transgenic plants. First, the addition of a gene(s) often improves the agricultural, horticultural, or ornamental value of a crop plant. Second, transgenic plants can act as living bioreactors for the inexpensive production of economically important proteins or metabolites. Third, plant genetic transformation (transgenesis) provides a powerful means for studying the action of genes during development and other biological processes.

There are a number of effective DNA-delivery systems and expression vectors that work with a range of plant cells. Furthermore, most plant cells are totipotent-meaning that an entire plant can be regenerated from a single plant cell-so fertile plants that carry an introduced gene(s) in all cells (i.e., transgenic plants) can often be produced from genetically engineered cells. The transgenic plant flowers and produces viable seed, the desired trait is passed on to successive generations.

Some of the genetically determined traits that can be introduced into plants by a single gene or, possibly, a small cluster of genes include insecticidal activity, protection against viral infection, resistance to herbicides, delay of senescence, tolerance of environmental stresses, altered flower pigmentation, improved nutritional quality of seed proteins, and self-incompatibility. To date, numerous transgenic plants have been generated, including many crop and forest species. In the future, plant biotechnology will have an enormous impact on traditional plant breeding programs because it can significantly decrease the 10 to 15 years that it currently takes to develop a new variety using traditional plant breeding techniques; further, it will also be used to create plants with novel characteristics.

16.2.1. Methodology : Plant Transformation with the Ti Plasmid of *Agrobacterium tumefaciens*

The gram-negative soil bacterium *Agrobacterium tumefaciens* is a phytopathogen, genetically transforms the plant cells and results in the production of crown gall tumors in the infected plant. This agronomically important disease affects only dicotyledonous plants (dicots), including grapes, stone-fruit trees, and roses.

Crown gall formation is the consequence of the transfer, integration, and expression of genes of a specific segment of bacterial plasmid DNA called the T-DNA (transferred DNA)-into the plant cell genome. The T-DNA is actually part of the "tumor-inducing" (Ti) plasmid that is carried by most strains of *A. tumefaciens.* Depending on which bacterial strain the Ti plasmid

comes from, the length of the T-DNA region can vary from approximately 12 to 24 kilobase pairs (kb). Strains of *A. tumefaciens* that do not possess a Ti plasmid cannot induce crown gall tumors.

The initial step in the infection process is the attachment of *A. tumefaciens* to a plant cell at the site of an open wound, often at the base of the stem (i.e., the crown) of the plant. Originally, it was thought that *A. tumefaciens* infected wounded plants because the physical barrier of the cell wall had been breached by injury, thereby facilitating entry of the bacterium. However, it is now believed that these bacteria respond to certain plant phenolic compounds such as acetosyringone and hydroxyacetosyringone which are excreted by susceptible wounded plants. These wound-response compounds resemble some of the products of phenylpropanoid metabolism, which is the major plant pathway for the synthesis of plant secondary metabolites such as lignins and flavonoids. These small molecules (i.e., acetosyringone, hydroxysyringone) act to induce the activity of the virulence *(vir)* genes that are encoded on the Ti plasmid.

The *vir* genes are located on a 35-kb region of the Ti plasmid that lies outside of the T-DNA region. The products of the *vir* genes are essential for the transfer and integration of the T-DNA region into the genome of a plant cell. There are at least seven, and possibly eight, different *vir* genes.

After a Ti plasmid-carrying cell of *A. tumefaciens* attaches to a host plant cell and the *vir* genes are induced, the T-DNA is transferred by a process that is thought to be similar to plasmid transfer from donor to recipient cells during bacterial conjugation. In this model, the T-DNA is transferred as a linear, single-stranded molecule from the Ti plasmid, passes into the plant cell, and eventually becomes integrated in the plant chromosomal DNA. The 5' end of the single-stranded T-DNA carries the right border sequence, and the left border sequence is at the 3' end.

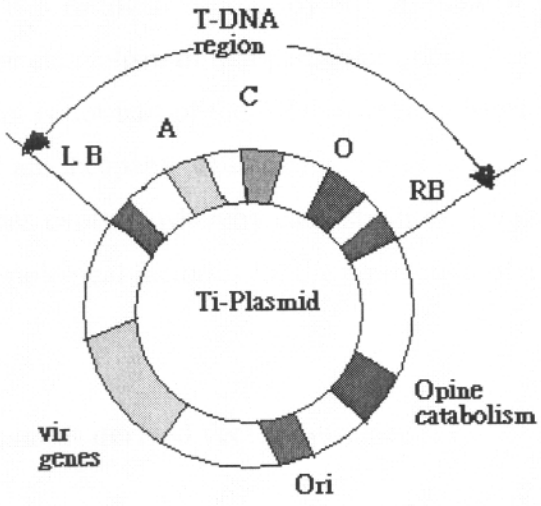


Fig. 16.1. Schematic representation of a Ti Plasmid ; LB – Left border; RB – right border; A – Auxin; C – Cytokinin; O- opine

The formation of the single-stranded form of T-DNA is initiated by strand-specific cutting at both borders of the intact T-DNA region. The integration of the T-DNA into the plant genome is thought to depend on specific sequences that are located at the right border of the T-DNA. This border contains a repeating unit that consists of 25 base pairs (bp). Although the left border contains a similar 25-bp repeat, deletion studies have shown that this region is not involved in the integration process.

Most of the genes that are located within the T-DNA region are activated only after the T-DNA is inserted into the plant genome. The products of these genes are responsible for crown gall formation.

In addition to auxin and cytokinin biosynthesis genes, the T-DNA region from each specific Ti plasmid carries a gene for the synthesis of a molecule called an opine. Opines are unique and unusual condensation products of an amino acid and a keto acid, or an amino acid and a sugar. For example, the condensation product of arginine and pyruvic acid is called octopine; arginine with α-ketoglutaraldehyde is nopaline; and agropine is a bicyc1ic sugar derivative of glutamic acid. The opines are synthesized within the crown gall and then secreted. They can be used as a carbon source (and sometimes also as a nitrogen source) by any *A. tumefaciens* that carries a Ti plasmid-borne gene for the catabolism of that particular opine. The opine catabolism gene(s) is on the Ti plasmid but is not part of the T-DNA region. Most other soil microorganisms that have been tested are incapable of utilizing opines as a carbon source. Thus, a unique set of mechanisms has evolved whereby each strain of *A. tumefaciens* genetically engineers plant cells to be biological factories for the production of a carbon compound that it alone is able to use.

16.2.2. Ti Plasmid- derived vector systems

The ability of the Ti plasmid to genetically transform plants, was exploited to insert a desired DNA sequence into the T-DNA region and then use the Ti plasmid and *A. tumefaciens* to deliver and insert this gene(s) into the genome of a susceptible plant cell. Even though the Ti plasmids are effective as natural vectors, they have several serious limitations as routine cloning vectors, suggesting some modifications before using as a cloning vector:

1. The production of phytohormones by transformed cells growing in culture prevents them from being regenerated into mature plants. Therefore, the auxin and cytokinin genes must be removed from any Ti plasmid-derived cloning vector.

2. A gene encoding opine synthesis is not useful to a transgenic plant and may lower the final plant yield by diverting plant resources into opine production. Therefore, the opine-synthesis gene should be removed.

3. Ti plasmids are large (approximately 200 to 800 kb). For recombinant DNA experiments, however, a much smaller version is preferred, so large segments of DNA that are not essential for a cloning vector must be removed.

4. Because the Ti plasmid does not replicate in *Escherichia coli*, the convenience of perpetuating and manipulating Ti plasmids carrying inserted DNA sequences in this bacterium does not exist. Therefore, in developing Ti plasmid-based vectors, an origin of replication that can be used in *E. coli* must be added.

After considering the above said limitations, a number of plant cell vectors have been constructed with the following components :

1. A selectable marker gene, such as neomycin phosphotransferase, that confers kanamycin resistance to transformed plant cells. Because the neomycin phosphotransferase gene (as well as many other marker genes used in plant transformation) is prokaryotic in origin, it is necessary to put it under the control of plant (eukaryotic) transcriptional regulation signals, including both a promoter and a termination-polyadenylation sequence, to ensure that it is efficiently expressed in transformed plant cells.

2. An origin of DNA replication that allows the plasmid to replicate in *E. coli.* In some vectors, an origin of replication that functions in *A. tumefaciens* has also been added.

3. The right border sequence of the T-DNA region. This region is absolutely required for T-DNA integration into plant cell DNA, although most cloning vectors include both a right and a left border sequence.

4. A polylinker (multiple cloning site) to facilitate insertion of the cloned gene into the region between T-DNA border sequences.

Because these cloning vectors lack *vir* genes, they cannot by themselves effect the transfer and integration of the T-DNA region into recipient plant cells. Two different approaches have been used to meet these ends. In one approach, a binary vector system is used. The binary cloning vector contains both E. *coli* and *A. tumefaciens* origins of DNA replication but no *vir* genes; it is actually an E. *coli-A. tumefaciens* shuttle vector. All the cloning steps are carried out in E. *coli* before the vector is introduced into *A. tumefaciens.* The recipient *A. tumefaciens* strain carries a modified (defective, disarmed) Ti plasmid that contains a complete set of *vir* genes but lacks portions, or all, of the T-DNA region (so that this T-DNA cannot be transferred). In this system, the defective Ti plasmid synthesizes the *vir* gene products that mobilize the T-DNA region of the binary cloning vector plasmid. By providing the proteins encoded by the *vir* genes, the defective Ti plasmid is acting as a helper plasmid, enabling the T-DNA from the binary cloning vector to be inserted into the plant chromosomal DNA.

In the second approach, called the cointegrate vector system, the cloning vector recombines with a modified (disarmed) Ti plasmid which lacks the tumor-producing genes of the T-DNA within *A. tumefaciens,* and the entire cloning vector becomes integrated into the disarmed Ti plasmid. The cointegrate cloning vehicle and the disarmed helper Ti plasmid carry homologous DNA sequences that provide a shared site for in vivo homologous recombination; normally these sequences lie inside the T-DNA region. Following recombination, the cloning vector becomes part

of the disarmed Ti plasmid, which provides the *vir* genes necessary for the transfer of the T-DNA to the host plant cells. The only way that this cloning vector can be maintained in *A. tumefaciens* is as part of a cointegrate structure. In this cointegrated configuration the genetically engineered T-DNA region can be transferred to plant cells.

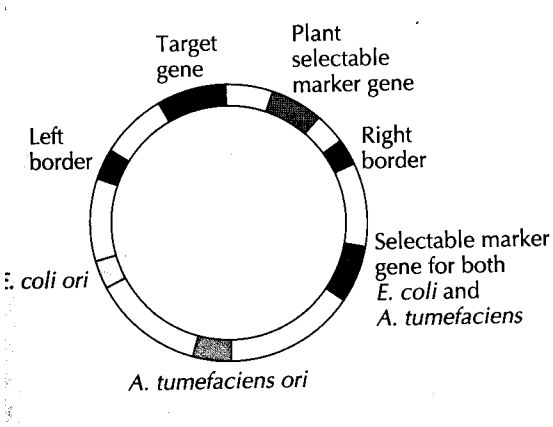


Fig.16.2. Binary cloning vector has origins of replication (ori) for both *E. coli* and *A. tumefaciens*., a selectable marker that can be used in either *E. coli* or *A. tumefaciens.*

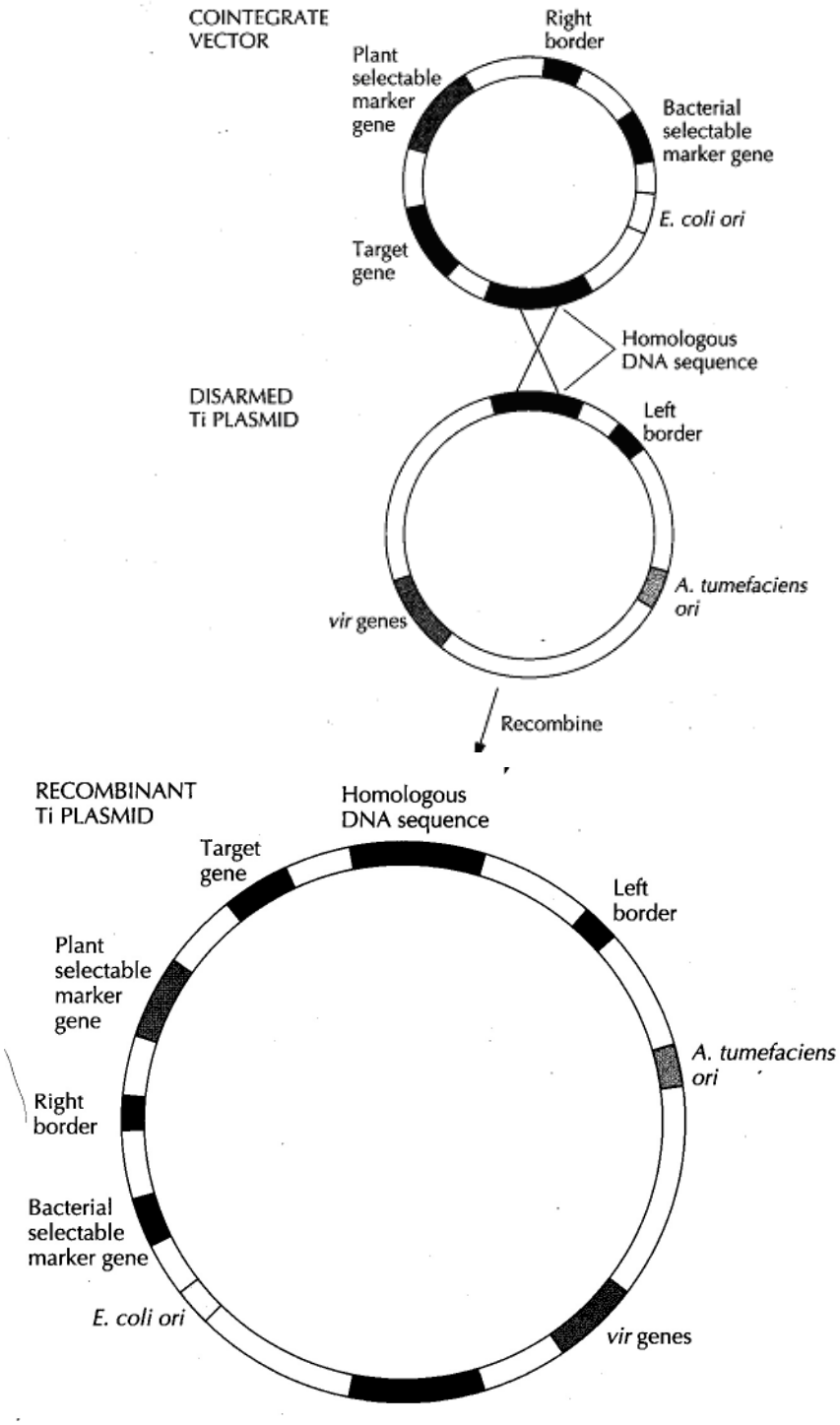


Fig. 16.3. The cointegrate cloning bector (top) carry only an *E.coli* origin of replication and cannot exist autonomously within *A. tumefaciens*.. The disarmed Ti plasmid (middle) contains the T-DNA left border, the *vir* gene cluster, and an *A. tumefacien*s origin of DNA replication. The lower one is the recombinant Ti plasmid that has T-DNA left and right borders.

16.2.3. Methods of transferring genes in to plants

Although *Agrobacterium* based biological systems, as described above, are effective in several species, not all plants can be transformed in this way. In particular, monocotyledonous plants (monocots), including the major cereal crops such as rice, wheat, and corn. However, refined protocols have been devised for the transformation of maize and rice, both monocots, by *A. tumefaciens* carrying Ti plasmid vectors. For example, immature maize embryos were immersed in an *A. tumefaciens* cell suspension for a few minutes and then incubated for several days at room temperature in the absence of selective pressure. The embryos were then transferred to a medium that contained a selective antibiotic that only allowed transformed plant cells to grow. These cells were maintained in the dark for a few weeks. Finally, the mass of transformed plant cells was transferred to a different growth medium and incubated in the light, which permitted regeneration of whole transgenic plants.

Plant protoplasts can be maintained in culture as independently growing cells, or, with a specific culture medium, cell walls can be formed and whole plants can be regenerated from these cells. In addition, transformation methods have been developed that introduce cloned genes into a small number of cells of a plant tissue from which whole plants can be formed, bypassing the need for regeneration from a protoplast. At present, most researchers favor the use of either Ti plasmid-based vectors or microprojectile bombardment (physical method) to deliver DNA into plant cells. Using these various techniques, more than 50 different plants have been genetically transformed.

16.2.3.1. Microprojectile Bombardment

Microprojectile bombardment, also called biolistics, is the most promising physical method of DNA delivery system for plants. Gold or tungsten spherical particles (approximately 0.4 to 1.2 pm in diameter) are coated with DNA that has been precipitated with $CaCl_2$, spermidine, or polyethylene glycol. The coated, particles are accelerated to high speed (300 to 600 meters/second) with a special apparatus called a particle gun, which uses gunpowder, compressed air, or helium to provide the propelling force. At these speeds, the projectiles can penetrate plant cell walls and membranes. The particle density used does not significantly damage the cells.

Once, inside a cell, the DNA, by some unknown process, integrates into the plant DNA. With the microprojectile bombardment system it is possible to transform a large number of different plant species, including monocots and conifers, plants that are not susceptible to *Agrobacterium*-mediated DNA transfer.

Microprojectile bombardment can be used to introduce foreign DNA into plant cell suspensions, callus cultures, meristematic tissues, immature embryos, protocorms, coleoptiles, and pollen in a wide range of different plants. Furthermore, this method has also been used to deliver genes into chloroplasts and mitochondria, thereby opening up the possibility of introducing exogenous (foreign) genes into these organelles.

The configuration of the vector that is used to deliver foreign genes to the plant may have an influence on both the integration and expression in the plant of those genes. For example, transformation is more efficient when linear rather than circular DNA is used. Moreover, large plasmids (>10 kb) may become fragmented during microprojectile bombardment and therefore often result in lower levels of foreign gene expression than when smaller plasmids are used.

16.2.4. Screening for transferred genes

It is essential to be able to detect the foreign DNA that has been integrated into plant genomic DNA so that those cells that have been transformed can be identified. Reporter genes from bacteria with plant-specific regulatory sequences for expression in plant cells were generally used as markers. Dominant marker selection provides a direct means of obtaining only transformed cells in culture. For example, in the presence of the antibiotic kanamycin, only plant cells with a functional neomycin phosphotransferase gene can grow.

The type of reporter gene to be used is based on the needs of the particular experiment. If the expression of a reporter gene interferes with normal plant functions, it cannot be used. Similarly, if the presence of reporter genes alter the quality of the commercial product, it can not be used. As an example, it is better not to introduce antibiotic resistance genes into any crop plants.

Some reporter gene products (e.g., β-D-glucuronidase and both firefly and bacterial luciferases) can be detected in intact plant tissues. The most popular of these systems is the E. *coli* β-D-glucuronidase (GUS) gene. It encodes a stable enzyme that is not normally present in plants and that catalyzes the cleavage of a range of β-D-glucuronides. The GUS activity in transformed plant tissues can be localized by observing the blue color that is formed after the hydrolysis of the uncolored substrate 5-bromo-4-chloro3-indolyl β-D-glucuronic acid. Alternatively, GUS activity in plant extracts can be more quantitatively and sensitively assayed by a fluorometric analysis that involves the hydrolysis of the substrate 4-methylumbelliferyl β-D-glucuronide to form a fluorescent product.

16.2.5. Manipulation of gene expression in plants

The final step in genetic engineering is – recovery, that is, the transformed plants were assayed for the production of the foreign protein and studied physiologically to assess how the presence of an additional protein affected the whole plant. For better recovery of the foreign gene product, the genes must be properly manipulated for over expression. Previously, in most of the experiments promoters (strong constitutive 35S promoter from cauliflower mosaic virus, which is expressed in all plant tissues and throughout the life of the plant) that were expressed constitutively in a range of plant cells were used for better expression . More recently, the promoters that express in specific cell at only certain time of the growth were isolated, characterized and used. For example, instead of the strong constitutive 35S promoter from cauliflower mosaic virus, researchers have used the promoter for the small subunit of the

photosynthetic enzyme ribulose bisphosphate carboxylase, which is active only in photosynthetic tissues such as leaves. Similarly, plant promoters active only in specific tissues such as roots or flowers, or only during periods of environmental stress- e.g., the pathogenesis-related (PR) promoters have been used to control the expression of some foreign genes.

16.2.6. Production of Transgenic Plants

By using the recombinant DNA technology, many transgenic plants with desires traits have been developed. Genes that confer resistance to insects, viruses, herbicides, environmental stress, and senescence have been incorporated into various plants, leading to the development of insect-resistant, virus resistant and herbicide resistant plants. In addition, some work has gone into manipulating the biochemical pathway of flower pigmentation, modifying plant products, and using plants as bioreactors.

16.2.6.1. Insect-Resistant Plants

If crop plants could be genetically engineered to produce functional insecticides, then it might be possible to develop crops that would be intrinsically tolerant of insect predators and would not need to be sprayed (often six to eight times during a growing season) with costly and hazardous chemical pesticides. However, biological insecticides are usually highly specific for a limited number of insect species, and they are generally considered to be non hazardous to humans and other higher animals.

Several different strategies have been used to confer resistance against insect predators. One approach involves a gene for an insecticidal protoxin produced by one of several subspecies of the bacterium *Bacillus thuringiensis.* Another common strategy uses genes for plant proteins such as amylase inhibitors or protease inhibitors that have been shown to be effective against a wide variety of insects. After an insect ingests one of these inhibitors, it is not able to digest food (i.e., plants) because the inhibitor interferes with the hydrolysis of starch or plant proteins. Thus, the insect will feed less.

B. *thuringiensis* protoxin does not persist in the environment, nor is it hazardous to mammals. Thus, it is a safe means of protecting plants. Unfortunately, a number of the economically significant pests of crop plants feed on internal plant tissues and are thus unlikely to be inhibited by B. *thuringiensis* preparations that has been sprayed onto the surface of the plant. To avoid this problem, genes for B. *thuringiensis* toxins can be expressed in plants. With these transgenic plants, there is no need to spray the insecticidal toxin. This limits the environmental distribution of the toxin and avoids problems associated with its limited environmental stability and the timing of the toxin application.

16.2.6.2. Virus-Resistant Plants

Plant viruses often cause considerable crop damage and significantly reduce yields. Therefore, in the absence of effective chemical treatments, plant breeders have attempted to

transfer naturally occurring virus resistance genes from one plant strain (cultivar) to another. The newly developed transgenic plants express the gene for the coat protein (which often is the most abundant protein of a virus particle) of a virus that normally infects the plants. Then, the ability of the virus to subsequently infect these plants and spread systemically is often greatly diminished. Although the precise mechanism by which the presence of coat protein genes inhibits viral proliferation is not known, it is clear that the antiviral effect occurs early in the viral replication cycle and, as a result, prevents any significant amount of viral synthesis. This feature is an advantage because it decreases the probability of selecting for spontaneous viral mutants that can overcome this resistance and replicate in the presence of virus coat protein. The viral coat protein gene approach has been used to confer tolerance to a number of different plant viruses. Using this approach, researchers have developed virus-resistant transgenic plants for a number of different crops.

16.2.6.3. Herbicide-Resistant Plants

Approximately 10% of global crop production is lost through weed infestation every year, despite the expense of $10 billion spent on more than 100 different chemical herbicides. The creation of herbicide resistant crop plants is one way to overcome some of this problem.

Resistant plants have been developed for the herbicide glyphosate. It acts as an inhibitor of 5-enolpyruvylshikimate –3- phosphate synthase (EPSPS), which is the enzyme in the shikmate pathway. This enzyme plays as important role in the synthesis of aromatic amino acids in both bacteria and plants. An EPSPS- encoding gene from glyphosate resistant strain of *E. coli* was isolated and transferred in to plant cells. Transgenic tobacco, petunia, tomato, potato and cotton with this gene are resistant to glyphosate treatment whereas, the weeds are not.

16.3. Transgenic animals

Genetic modification of animals by recombinant DNA technology (transgenesis) includes the introduction of a cloned gene(s) into the genome of a cell that may subsequently establish a true breeding lineage.

Transgenic mice have also played a role in examining the feasibility of the industrial production of human therapeutic drugs by domesticated animals and in the creation of transgenic strains that act as biomedical models for various human genetic diseases.

16.3.1. Methods of gene transfer

For transgenesis, DNA can be introduced into mice by (1) retroviral vectors that infect the cells of an early stage embryo prior to implantation into a receptive female, (2) microinjection into the enlarged sperm nucleus (the male pronucleus) of a fertilized egg, or (3) introduction of genetically engineered embryonic stem cells into an early-stage developing embryo before implantation into a receptive female.

16.3.1.1. Retroviral Vector Method

Of the various gene transfer methods, the use of retroviral vectors has the advantage of being an effective means of integrating the transgene into the genome of a recipient cell. However, these vectors can transfer only small pieces (-8 kilobases [kb]) of DNA, which, because of the size constraint, may lack essential adjacent sequences for regulating the expression of the transgene.

There is a further major drawback to the use of retroviral vectors. Although these vectors are designed to be replication defective, the genome of the retroviral strain (helper virus) that is needed to create large quantities of the vector DNA can be integrated into the same nucleus as the transgene. Despite special precautions, helper strain retroviruses could be produced by the transgenic organism. Consequently, for applications in which either a commercial product is to be synthesized by the transgenic organism or the transgenic organism is to be used as food, it is absolutely necessary that there be no retroviral contamination. Therefore, because alternative methods are available, retroviral vectors are rarely used for creating transgenic animals that have a commercial end use.

16.3.1.2. DNA Microinjection Method

Because of the disadvantages of the retroviral vector method, microinjection of DNA is currently the preferred method for producing transgenic mice. This procedure is performed via the following steps.

1. The number of available fertilized eggs that are to be inoculated by microinjection is increased by stimulating donor females to superovulate. They are given an initial injection of pregnant mare's serum and another injection, about 48 hours later, of human chorionic gonadotropin. A superovulated mouse produces about 35 eggs instead of the normal 5 to 10.

2. The superovulated females are mated and then sacrificed. The fertilized eggs are flushed from their oviducts.

3. Microinjection of the fertilized eggs usually occurs immediately after their collection. The microinjected transgene construct is often in a linear form and free of prokaryotic vector DNA sequences.

In mammals, after entry of the sperm into the egg, both the sperm nucleus (male pronucleus) and female nucleus exist separately. After the female nucleus completes its meiotic division to become a female pronucleus, nuclear fusion (karyogamy) can occur. The male pronucleus, which tends to be larger than the female pronucleus, can be located by using a dissecting microscope. The egg can then be maneuvered, oriented, and held in place while the DNA is microinjected. On a good day, several hundred male pronuclei can be inoculated.

After inoculation, 25 to 40 eggs are implanted micro surgically into a foster mother, which has been made pseudopregnant by being mated to a vasectomized male. In mice, copulation is the only known way to prepare the uterus for implantation. In this case, because the male mate lacks sperm, none of the eggs of the foster mother are fertilized. The foster mother will deliver pups from the inoculated eggs about 3 weeks after implantation.

16.3.1.3. Embryo stem technology

An alternative and powerful strategy for transgenesis involves the introduction of foreign DNA into embryonic stem (ES) cells to establish an ES cell line, cells are removed from the inner cell mass of the developing blastocyst and passaged either on feeder layers, or in the presence of differentiation – inhibiting activity (DIA) to maintain their undifferentiated state. Foreign DNA can be introduced in to the ES cells by a number of means : electroporation, transfection or micro-injection. Selected cells are then reintroduced in to a blastocyst and are reimplanted into a peudopregnant female and allowed to develop to term. An important distinction between pups obtained in this way, and those resulting from micro-injection of the one-cell embryo, is that they will, by definition, be chimearas, since cells harboring the transgene only constitute a proportion of the inner cell mass of the blatocyst. However, providing transgen-containing cells have contributed to the germline, then a suitable breeding strategy will allow the establishment of a transgenic line.

16.3.2. Screening for transferred genes.:

To identify transgenic animals, DNA from a small piece of the tail can be assayed by either Southern blot hybridization or the polymerase chain reaction (PCR) for the presence of the transgene. A transgenic mouse can be mated to another mouse to determine if the trans gene is in the germ line of the founder animal. Subsequently, progeny can be bred to generate pure (homozygous) transgenic lines.

The procedure, although apparently simple, requires the coordination of a number of experimental steps. Even a highly trained practitioner can expect, at best, only 5% of the inoculated eggs to develop into live transgenic animals. None of the steps in the procedure is 100% efficient; consequently, large numbers of microinjected fertilized eggs must be used. With the mouse system, for example, about 66% of the fertilized eggs survive the injection procedure; about 25% of the implanted eggs develop into pups; and about 25% of the pups are transgenic. Thus, from 1,000 inoculated fertilized eggs, 30 to 50 transgenic pups are produced. Furthermore, with this method, the injected DNA integrates at random sites within the genome, and often multiple copies of the injected DNA are incorporated at one site. Not all of the transgenic pups will have the appropriate characteristic. In some individuals, the transgene may not be expressed because of the site of integration, and in others, the copy number may be excessive and may lead to over expression, which disrupts the normal physiology of the animal. Notwithstanding the overall inefficiency, it has become routine to use DNA microinjection to generate lines of mice carrying functional transgenes.

16.3.3. Production and applications of transgenic animals

Though transgenic mice were produced almost 20 years ago, the mouse is still considered as experimental animal for transformation studies. During the recent years, transgenic techniques have been extended to other animals like cow, goat, the pig, and the sheep.

16.3.3.1. Transgenic mice

Transgenic mice have been produced by microinjection of the pronucleus *of* the fertilized egg or transfection of ES cells with YACs that carry either a number of related genes or a single large gene. For example, transgenic mice that carry the human β-globin gene cluster, which has five different functional globin genes and covers about 250 kb, express the genes of the β-globin ensemble in a tissue-specific and time-dependent manner that corresponds to the site-specificity and temporal pattern in humans. This concurrence is attributed to the flanking DNA that contains the promoter region and other important regulatory elements.

The production of mice that synthesize only human antibodies is a noteworthy example of YAC transgenesis. Monoclonal antibodies have the potential to be used as human therapeutic agents, However, it is impossible to generate human monoclonal antibodies routinely. Also, unfortunately, rodent monoclonal antibodies are immunogenic to humans. Elaborate recombinant DNA strategies have been devised to "humanize" existing rodent monoclonal antibodies. These laborintensive procedures produce Fv and Fab antibodies, which sometimes have a low affinity for a particular antigen. A more accessible method for generating complete human antibodies by using hybridomas would be a welcome technological advance.

Transgenic mice have also been used as models for expression systems that are designed for secretion of the product of a transgene into milk. For example, large quantities of authentic cystic fibrosis transmembrane regulator (CFTR) protein are needed to study its function and to formulate potential therapies for treating cystic fibrosis (CF). This prevalent genetic disease affects about 1 in 2,500 live-born persons of European origin. The primary effect of a faulty CF gene is an alteration in the function of CFTR, which normally acts as a chloride channel. As a consequence of the disruption of the proper flow of chloride ions into and out of cells, mucus accumulates in the ducts of several organs, especially the lungs and pancreas. This mucus becomes the site of a bacterial infection that is difficult to control with antibiotics; as the bacteria die, their released DNA makes the mucus very thick. This thickened mucus prevents normal organ function by blocking the ducts and thereby exacerbates the effects of CF. Currently, the life expectancy for persons with CF is about 25 to 30 years.

A reliable supply of CFTR is required to learn more about its mode of action. With various in vitro cell expression systems, the yields of CFTR have been low, possibly as a result of the biological consequences of the accumulation in the cell membrane of transfected cells of CFTR, which is a transmembrane protein. This limitation would be remedied if plasma membranes were frequently shed from the host cell. With such a system, not only would a heterologous transmembrane protein be associated with the released fragments of plasma membrane but also concentrating and purifying the protein would be relatively easy. In fact, during lactation, the

mammary gland cells use this procedure for the production of fat globules. In this instance, fat from within the mammary gland cell is encapsulated by plasma membrane and together they are secreted into milk as a globule.

To test the feasibility of this system, a full-length CFTR cDNA sequence was cloned into the middle of a defective goat β-casein gene from which the end of exon 2 to the beginning of exon 7 had been deleted. The construct kept the promoter and termination sequences of the goat β-casein gene. The CFTR cDNA was cloned into a structural gene to provide introns for enhancing transcription of the transgene. The β -casein gene is actively expressed in mammary glands during lactation, and β -casein is a major milk protein.

Transgenic mouse lines carrying the CFTR sequence under the control of the β -casein gene regulatory sequences were established. As predicted, the milk of transgenic females contained the CFTR protein bound to the membrane of fat globules. There were no negative effects to either CFTR transgenic lactating mothers or pups that were fed milk that contained CFTR. The CFTR protein was glycosylated and readily extracted from the fat-rich fraction of the milk. It remains to be determined whether the CFTR is authentic. However, the possibility of producing other membrane-bound proteins in milk has been established. The mammary gland cells of lactating transgenic mice have also synthesized many other proteins that are potentially therapeutic for humans. Eventually, to obtain large quantities of CFTR, additional transmembrane proteins, and various human therapeutic proteins, the transgenic constructs will have to be incorporated into the genome of a much larger mammal such as a cow, sheep, or goat.

16.3.3.2. Transgenic Cattle

If the mammary gland is to be used as a bioreactor, dairy cattle, which each annually produce approximately 10,000 liters of milk, containing about 35 grams of protein per liter, are likely candidates for transgenesis. More specifically, if a recombinant protein was present at 1 gram per liter of milk and it could be purified with 50% efficiency, the yield from 20 transgenic cows would be about 100 kg per annum. Coincidentally, the annual global requirement for protein C, which is used for the prevention of blood clots, is about 100 kg. On the other hand, one transgenic cow would be more than sufficient for the production of the annual world supply of factor IX (plasma thromboplastin component), which is used by hemophiliacs to facilitate blood clotting.

One of the goals of transgenesis of dairy cattle is to change the constituents of milk. The amount of cheese produced from milk is directly proportional to the κ-casein content. Increasing κ-casein production with an over-expressed κ -casein transgene is a reasonable likelihood. For a different end use, expression of a lactase trans gene in the mammary gland could result in milk that is free of lactose. Such a milk would be welcomed by the many people who are lactose intolerant and experience severe indigestion after the consumption of milk or milk-containing foods. Although transgenesis of cattle holds promise, progress toward producing large numbers of genetically engineered animals will be slow because it takes about 2 years to produce a calf from a fertilized egg.

For livestock, in general, attempts will be made to create animals with inherited resistance to bacterial, viral, and parasitic diseases. For example, genetic resistance to bacterial diseases- such as mastitis (mammary gland abscesses) in dairy cattle, neonatal scours (dysentery) in swine, and fowl cholera-occurs in some breeds. If the basis of each of these resistances is a single gene, it may be possible to create transgenic animals that are resistant to these bacterial infections.

16.4. Industrial applications of genetic engineering

Genetic engineering has many commercial and practical applications. The very important areas in which it was proved to be an effective tool includes- microbial fermentations, virus vaccines, mammalian proteins, transgenic plants and animals, environmental biotechnology and in production of commercially important molecules. From the beginning, microbial systems are used for genetic transformations, because it is easy to culture and extract the products at lowest cost. Presently, plant systems were proved to be effective in the production of some gene products, and probably replace the microbial systems in future.

16.4.1. Use of genetically engineered microorganisms

16.4.1.1. Microbial fermentations :

A number of important products are made industrially using microorganisms, of which antibiotics are the most important. Gentic engineering procedures can be used to manipulate the antibiotic producing organism to increase the yields or to produce modified antibiotics. *Streptomyces* is the major organism used for the production of antibiotics, by cloning the antibiotic biosynthesis genes. Manipulating the genes involved in the biosynthesis of existing antibiotics may produce new antibiotics with unique properties. Some of the antibiotics produced by various strains of S*treptomyces* are

Strain / Plasmid	Antibiotic
S. aoelicolor	Actinorhodine
S. violaceoruber B1140	Granaticin
Streptomyces sp.	Medermycin

16.4.1.2. Virus vaccines

Traditionally, vaccines are either inactivated or attenuated infectious agents (bacteria or viruses) that are injected in to an antibody –producing organism to produce immunity. Recombinant DNA technology can be used in various ways to create reliable vaccines. It is possible to clone the gene(s) that encodes the major antigenic determinant (S) from a pathogenic organism into the genome of a benign carrier organism, (usually a virus), which can be used as a vaccine without concern that pathogenic particles are present in the vaccine. Finally, the genes or segments of genes that encode the major antigenic determinants of pathogenic organisms can be cloned into expression vectors, and large amounts of the product can be harvested, purified and used as a vaccine. With the last strategy, complete genes produce subunit vaccines and cloned

domains of the major antigenic determinants produce peptide vaccines. Sub unit vaccines contain only the subunit of a protein from the pathogenic organism. The first recombinant subunit vaccine for use in humans was made using yeast. The gene encoding the surface protein from hepatitis B virus was cloned and expressed in yeast. The protein from yeast was isolated and purified and used as a vaccine. Similarly vaccines were made with cloned influenza virus hemaglutinin genes.

16.4.1.3. Mammalian proteins

A number of mammalian proteins are of great medical and commercial interest. In the case of human proteins, commercial production by direct isolation from tissues or fluids is complicated and expensive, or even impossible. By cloning the gene for a human protein in an appropriate microorganism or a culturable cell, its commercial production is possible. E.g. Cloning of human insulin gene into bacteria has been carried out.

16.4.1.4. Environmental biotechnology

Because of the enormous metabolic diversity of bacteria, a large gene pool exists in bacteria form natural habitats. In some cases these genes code for protein that degrade environmental pollutants. Genes for the biodegradation of many toxic wastes and waste water pollutants have been shown to exist in natural isolates of bacteria. Genetic engineering is beginning to tap these resources for the purpose of environmental cleanup. For example, many members of the bacterial genus *Pseudomonas* carries plasmids that encode enzymes capable of degrading aromatic and halogenated organic compounds. By combining plasmids from different *Pseudomonas* strains within a single host, a new organism can be created with multiple degradation capabilities, which is widely used in environmental cleanup.

16.4.1.5. Pharmaceutical products

A large number of proteins that have potential as pharmaceutical agents have been synthesized from cloned genes in bacteria. Because most of these proteins are from eukaryotic organisms, the strategy for the isolation of a target gene often involves isolating mRNA enriched in the messenger of interest. In some instances, enzymes may be used as therapeutic agents. Some therapeutic products made by recombinant DNA techniques are -

Blood proteins : Erytroprotein, Factors VII, VIII, IX, Tissue plasminogen activator and Urokinase

Human hormones : Human growth hormone, Insulin, Nerve growth factor, Relaxin

Immune modulators : α interferon, β interferon, Colony stimulating factor, Lysozyme, Tumor necrosis factor

Vaccines : Cytomegalovirus, Hepatisis B virus, Measles , Rabbis

16.4.1.6. Other commercially important molecules

With recombinant DNA technology, it is possible to modify metabolic pathways of organisms either by introducing new genes or by altering existing ones. The goal is to create an

organism with a novel enzymatic activity that can convert an existing substrate in to a commercial compound that with current technology can be produced only by a combination of chemical treatments and fermentation steps. Eg. Synthesis of L- Ascorbic acid using cloned species of C*orynebaterium, Brevibacterium*; microbial synthesis of indigo using modified species of *Pseudomonas*.

16.4.2. Use of genetically engineered plants (Plants as bioreactors)

Plants are easy to grow and can generate considerable biomass. With these features in mind, research has been carried out to determine whether transgenic plants can be used for the production of commercial proteins and chemicals, Unlike recombinant bacteria, which are grown in large bioreactors (a process that requires highly trained personnel and expensive equipment), crops can be produced relatively inexpensively by less-skilled workers, Ultimately, the biggest hurdles to overcome in an endeavor of this sort will be the purification of the product of a transgene from the mass of plant tissue and the overall cost of producing a target protein in a transgenic plant compared with the cost of production in a microorganism. In small-scale trials, plants have been used to produce monoclonal antibodies, functional antibody fragments, and the polymer polyhydroxybutyrate, which can be used to make a biodegradable plastic-like material.

16.4.2.1. Antibodies

The production of antibodies and antibody fragments in transgenic plants has several potential advantages over their synthesis in recombinant microbial cells, For example, transformation of plants generally results in the stable integration of the foreign DNA into the plant genome, while most microorganisms are transformed with plasmids that can be lost during a prolonged or large-scale fermentation. In addition, the processing and assembly of foreign proteins in plants are similar to those in animal cells, whereas bacteria do not readily process, assemble, or post translationally modify eukaryotic proteins. Moreover, plants are inexpensive to grow on a large scale, and their production is not limited by fermentation capability, Finally, since most harvested plant tissues cannot usually be stored for long periods, foreign proteins might be produced in seeds where they will be stable for long periods under ambient conditions.

16.4.2.2. Polymers

It is costly to produce the polymer, poly (3-hydroxybutyric acid), which is used in the synthesis of biodegradable plastics, by bacterial fermentation. Consequently, research has been conducted to determine if this polymer could be produced at less cost in plants. In bacteria such as *Alcaligenes eutrophus,* poly(3-hydroxybutyric acid) is synthesized from acetyl coenzyme A in three steps catalyzed by three enzymes whose genes are organized on a single operon. Since plants are unable to process a transcript of an operon with more than one gene, each of the three genes was cloned separately and targeted to the chloroplast of the plant *Arabidopsis thaliana.* The chloroplast was chosen as the site of synthesis since previous experiments had demonstrated that cytoplasmic synthesis of this polymer resulted in only low levels of polymer formation and

produced highly stunted plants. Moreover, chloroplasts can accumulate high levels of starch, another biological polymer.

Each of the three poly(3-hydroxybutyric acid) biosynthesis genes was fused to a DNA fragment that encodes the chloroplast transit peptide of the small subunit of pea ribulose bisphosphate carboxylase and was placed under the transcriptional control of the cauliflower mosaic virus 35S promoter. Mature leaves of some of the transgenic plants expressing all three of the bacterial poly(3-hydroxybutyric acid) biosynthesis genes produced more than 1 mg of poly(3-hydroxybutyric acid) per gram (fresh weight) of leaf. This work is an important first step in the development of crop plants that produce large amounts of poly(3-hydroxybutyric acid).

16.4.2.3. Foreign Proteins in Seeds

Oleosins, or oil body proteins, are found in the seeds of a wide range of plants. These proteins are quite hydrophobic and are mostly embedded within plant oil bodies, thus stabilizing the oil bodies as discrete organelles. However, the N- and C-terminal regions of oleosins are more hydrophilic than the rest of the protein and are exposed to the aqueous environment. It is therefore possible to engineer at the DNA level fusions between oleosins and water-soluble proteins with the expectation that the fusion protein will be targeted to plant oil bodies, making it relatively easy to purify the fusion protein. In this case, the water soluble target protein will not be embedded in the oil body but rather will be exposed to the aqueous environment. Depending upon the final use of the fusion protein, it is also possible to cleave the target protein from the fusion protein if necessary. Since it is relatively easy to purify oil bodies from plant seeds, the purification of the recombinant protein is simplified. Then the recombinant protein can be recovered by cleavage of the fusion protein. This system has the potential to significantly lower the costs of purifying target proteins produced in plants.

16.5. Summary

Transgenesis, is the introduction of exogenous DNA of desired characters into the genome, such that it is stably maintained in a heritable manner. It becomes more feasible in plants with the advent of identifying the genetically induced tumor gall producing *Agrobacterium tumefaciens.* The main aim of developing these transgenic plants is to improve the agricultural, horticultural, or ornamental value of a crop plants; to use transgenic plants as living bioreactors for the inexpensive production of economically important proteins or metabolites; and to study the action of genes during development and other biological processes. Transformation of genetic material is either by A*grobacterium* based biological systems of physical methods like microprojectile methods. The transferred cells were identified by screening for selective markers which were also transferred along with the desirable genes. For better expression of the transferred genes were manipulated by adding proper promoter sequences. Similarly in animals also the genes with desirable characters can be incorporated either by retroviral vectors or microinjection of DNA or in embryonic stem cells. Transgenic mice were developed for synthesis of human antibodies; while transgenic cattle was aimed at high quality milk yields. Bacteria can also be exploited by genetic manipulation, to act as biological factories for the production of commercially important products like

pharmaceutical proteins, enzymes, antibiotics, vaccines; and can also be used in environmental cleanup.

16.5. Model questions

1. Write an essay on development and uses of transgenic plants.
2. Give an account on development of transgenic animals with desirable genes.
3. Discuss in detail about various industrial applications of genetic engineering
4. Write short notes on the following
    b) Ti – plasmid
    c) *Agrobacterium* based gene transfer
    d) Physical methods of gene transfer in plants and animals
    e) Methods of screening the transformed cells in plants and animals
    e) Plants as bioreactors

16.6. Reference books

Ratledge & Kristinsen 2001 Basic Biotechnology, University press

Hunt & Liveey (eds) 2000 Functional Genomics, Oxford University press

Krenzer & Massey – Recombinant DNA and Biotechnology : A guide for teachers
2nd ed. ASM press

Walker, J. M. and Rapley, R. 2002.Molicular biology and biotechnology. Fourth edition Panima publ. Corp., New Delhi., 563 pp.

Glick B.R. and Pasternak, J.J. 2003. Molecular Biotechnology- principles and applications of recombinant DNA. Panima Publ. Corp., New Delhi., 499pp.

Old and primrose. 1994. Principles of Gene Manipulation : An introduction to genetic engineering. 5 th ed. Blackwell Scientific publ.

Sambrook, J., E.F. Fritsch, T. Maniatis, 2002. Molecular Cloning : A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York.

Dr. M. Raghu Ram