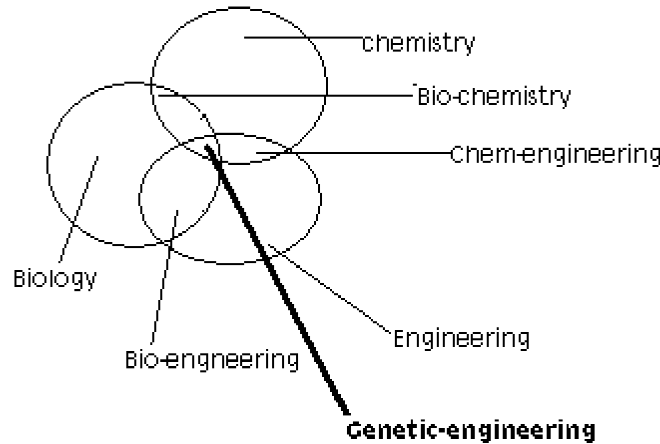
An Introduction to Recombinant DNA Technology



Biotechnology: The term biotechnology is derived from a fusion of biology and technology, which means the exploitation of biological agents or their components for generating useful products/services. In other words, every biological process used for mass production of something to be recurrently consumed by human, plants, or animals come under biotechnology.

The two main features of biotech are- First, Utilization of biological agents (micro-organism, cell of higher organism- either living or dead), their components or constituents (e.g. enzymes) in such a way that, Second, Some product or service is generated. This product or service should of human welfare.

The four classical examples of biotech are-

1. Use of yeast fermentation to make wine or beer and for making leavened bread.
2. Cheese production from milk.
3. Alcohol and vinegar production from molasses.
4. Antibiotic production from moulds, e.g. penicillin.

So far, as recent biotech is concerned it is based on technique of genetic engineering, molecular cloning and recombinant DNA.

Genetic engineering or Recombinant DNA technology- means alteration of genetic material of an organism under highly controlled laboratory condition so that an organism can produced different chemicals or perform completely new functions.

Research on genetic engineering is centered on *in vitro* joining of DNA fragments of different origin, mediated by some highly specific enzymes to produce recombinant DNA. (**Gene splicing**) which is then introduced into an appropriate host wherein its multiply and perform its normal function. (**Gene cloning**)

Recombinant DNA molecules are produced with one or more of the following **objectives**-

1. To obtain a large number of copies of specific DNA fragments
2. To recover large quantities of protein produced by concerned genes.
3. To integrate the gene in question into the chromosome of a target organism where it express it self.

A recombinant DNA molecule is produced by joining together two or more DNA segments usually originating from different organism. Most specifically a recombinant DNA molecule is a vector (e.g. a plasmid, phage or viruses) into which a desired DNA fragments has been inserted to enable its cloning in an appropriate host. This is achieved by using specific enzyme for cutting DNA (Restriction enzyme) into suitable fragment and then joining together the appropriate fragments (ligation). In this manner, a recombinant DNA molecule produced contains a gene from one organism joined together to the regulatory sequence of other organism, such a gene is called **chimaric gene**.

To achieve this the DNA segment are integrated into a self replicating DNA molecule called vector or carrier, most commonly used vector are either bacterial plasmid or DNA viruses. All these steps concerned with piecing DNA segment of diverse origin and placing them in suitable vector together constitutes recombinant DNA technology.

The vectors containing DNA segments to be cloned (DNA insert) is called chimaric vector, are then introduced into a suitable organism, usually a bacterium; this organism is called host while the process is called transformation. The transformed host cells are selected and cloned. The vector present in such host would replicate either in synchrony with or independent of host cell. The steps concerned with transformation of suitable host with a chimaric vector, and cloning of transformed cell is called DNA cloning or gene cloning. The gene cloning produces a large number of copies of gene/ DNA being cloned.

Often, the term recombinant DNA technology and gene cloning are used as synonym to each other. The other popular term, including both of these activities, is Genetic engineering.

The two important tools of recombinant DNA technology are

(1) Restriction enzymes and

(2) Vectors.

Restriction Endonuclease Enzymes: these are enzymes that produce internal cut, called cleavage in DNA molecule. Many endonuclease cut DNA at random site but a class of endonuclease cut DNA only within or near to those site which have specific base sequences; such endonuclease are known as Restriction Endonuclease and site recognized by them are called recognition site or recognition sequences. The recognition sequences are different and specific for different restriction enzymes.

The presence of restriction enzymes was first postulated by W. Arber during 1960, while the first true restriction enzyme, *HindII* was isolated by Simth and Nathen in 1970 for which Simth, Nathan and Arber were awarded the Nobel Prize for physiology and Medicine in 1978.

Restriction endonucleases are essential for DNA cloning and sequencing. They serve as tool for cutting DNA molecule at pre-determined sites, which is basic requirement for recombinant DNA technology.

Types of Restriction Endonucleases: there are types of restriction endonucleases

(i) **Type-I restriction endonuclease:** are complex endonucleases, and have recognition sequence of about 15bp; they cleave the DNA about 1000bp away from 5'- end of the sequence 'TCA' located within recognition site e.g., Eco_K, Eco_B etc.

(ii) **Type-II restriction endonuclease:** are remarkably stable and induce cleavage either within their recognition site or very close to them. They require Mg^{++} ions for the cleavage. The first type II isolated was *Hind* II in 1970 and at present more than 350 deifferent Type II endonulease with more than 100 recognition sequence has been identified. Only Type II restriction endonucleases are used for restriction mapping and gene cloning, since they cleave only at specific site. There target sites are 4-8 base pair long with palindrome with rotational symmetry nature and are predominantly GC rich.

(iii) **Type III restriction endonuclease**: are intermediate between Type I and Type II; they cleave DNA in the immediate vicinity of their recognition site e.g., *EcoP*$_{I}$, *EcoP*$_{15}$.etc. Type I and Type III enzymes are not used for gene cloning.

Recognition Sequence: The recognition sequences for Type II endonucleases form *palindromes with rotational symmetry.*

In **Palindrome**, the base sequences in the second half of a DNA strand is the mirror image of the sequences in its first half; consequently the complimentary DNA strand of a DNA double helix also show the same situation.

5'GAAAAG3' — DNA SINGLE STRAND

5'G A A A A G3'
3'C T T T T C5' — DNA DOUBLE STRAND

But in a **Palindrome with rotational symmetry,** the base sequences in the first half of one strand of DNA double helix is the mirror image of the second half of its complimentary strand. Thus in such palindrome, the base sequence in both strands of the double helix read the same when read from the same end (either 5' or 3' end) of the both strands.

5' G A A T T C 3'
3' C T T A AG 5'

Cleavage pattern: Most of the Type II restriction enzyme cleaves the DNA molecule within their specific recognition sites, but some cut immediate outside the target sequence. Thus there are two types of cuts-

Staggered cuts: Most of the Type II enzymes produces staggered cut in which the two strand of DNA double helix are cleaved at different location, this generate protruding (5' or 3') ends, i.e., one strand of double helix extend some bases beyond the other. Due to palindromic symmetrical nature of target site, the two protruding ends generated by such cleavage by a given enzyme have a complimentary base sequences. As a result they readily pair with each other. Such protruding ends are called cohesive or sticky ends.

An important consequence of this fact is that when a fragments generated by a single restriction enzyme from a DNA of different origin are mixed, they join together due to their sticy ends. Therefore, this property of restriction enzyme is of great value for construction of recombinant DNA.

3'Protruding end
5'--------G AATTC------3'
3'--------CTTAA G------5'
5'Protruding end

DNA of "A" Origin

\+

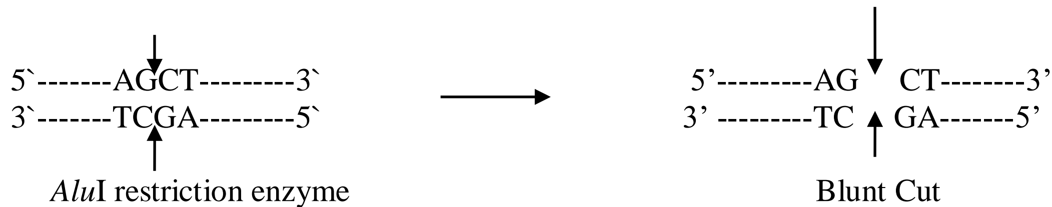
3' Protruding end
5'--------G AATTC------3'
3'--------CTTAA G------5'
5'Protruding end

DNA of "B" Origin

5'--------GAATTC------3'
3'--------CT TAAG-----5'
"A" type recombinant DNA

5'--------GAATTC------3'
3'--------CTTAAG------5'
"B" type recombinant DNA

Blunt or Flush Cuts: Some Type II restriction enzyme cut both strand of DNA molecule at same site so that the resulting ends have blunt or flush ends, in which the two strands end at same point. The blunt cut end can also be effectively utilized for construction of recombinant DNA with help of several strategies.



Steps in Gene Cloning: The entire procedure for recombinant DNA technology can be grouped into following five steps

1. Identification and isolation of desired gene or DNA fragments to be cloned.
2. Insertion of isolated gene in suitable vector.
3. Introduction of these vectors into a suitable organism (host) called host transformation.
4. Selection of the transformed host.
5. Multiplication/expression/integration followed by expression of the introduced gene in the host.

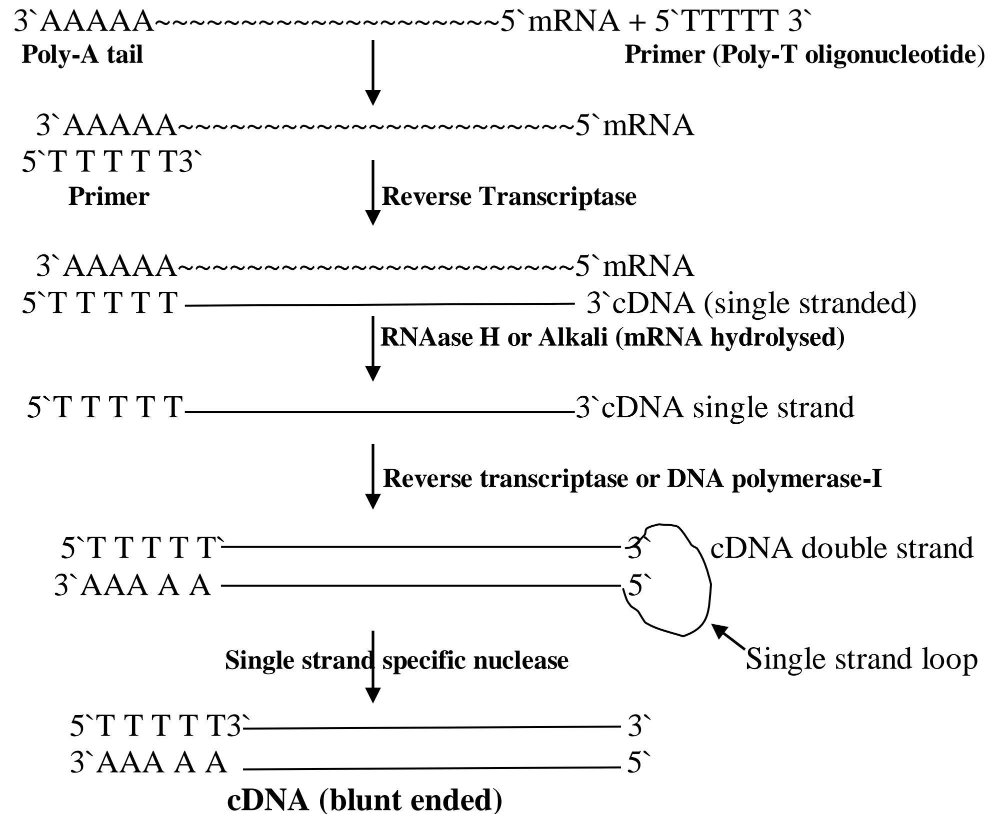
Isolation of the desired gene: The identification and isolation of desired gene or DNA fragments, called DNA insert can be done from the following

A). **c DNA Libraries:** A cDNA library is a population of bacterial transfornants or phage lysates in which each mRNA isolated from an organism or tissue is represented as its cDNA insertion in plasmid or a phage vectors.

Preparation of cDNA: cDNA is the copy or complimentary DNA produced by using mRNA as template. DNA copies of a mRNA molecule is produced by the reverse transcriptase enzyme. This enzyme performs similar reaction as DNA polymerase and has an absolute requirement for primer with a free 3`OH end.

When eukaryotic mRNA is used as a template, a Poly-T oligonucleotides is used as primer since these mRNA have Poly-A tail at their 3`ends, but for prokaryotes Poly-A tail may be added at the 3`ens of mRNA with the help of enzyme Poly-A polymerase.

Use of cDNA library is absolutely essential when the expression of eukaryotic gene is required in prokaryotes. This is because eukaryotic genes have *introns* which must be removed from their transcript to yield mature mRNA.



cDNA (blunt ended)

Diagrammatic representation of cDNA production from mRNA using reverse transcriptase

Problem in cDNA preparation: The ideal situation as mentioned above is rare. In fact, starting with even a pure mRNA , the end product consist of a mixture of cDNAs most of which is shorter than the complete mRNA. This is due to either of following reasons-

1. Incomplete copying of mRNA by reverse transcriptase.
2. Incomplete copying of the cDNA single strand and
3. The nuclease used for cleaving of hair pin loop may also cut away some ends of cDNA duplex.

Thus the double stranded cDNA preparation is always a mixture of different kinds of cDNA molecules. Physical and chemical methods are incapable of resolving these mixtures. Therefore, the cDNA mixture itself is used for cloning and the desired cDNA is identified in pure form from the appropriate bacterial clone.

Genomic Library: A genomic library is a collection of plasmid clone or phage lysates containing recombinant DNA molecule so that the sum total of DNA insert in this collection, represent the entire genome of the concerned organism.

However, in spite of all care taken in the production of genomic libraries, certain DNA fragments should expected to under or over represented or even missing. The possible reasons for this may be that certain fragment code for toxic product, or might replicate slowly or might have been altered by recombinational event during cloning. In addition, endonulease cleavage site are often not recognized equally well.

Construction of Genomic Library: For preparation of genomic library, the total genomic DNA of an organism is extracted. The DNA is broken into a fragment of appropriate size either by mechanical shearing, or sonication, or by using suitable restriction enzyme for partial digestion of the DNA; complete digestion is avoided since its generates fragment that are too heterogeneous in size. For partial digestion, restriction enzyme having 4-base recognition sequences are preferred to those having 6-base recognition sequences. This is because a given 4-base recognition sequences is expected to occur every 4^4 (=256) base pairs in DNA molecules, while a 6-base target sites would occur only after every 4^6 (=4096) base pair.

The partial digest of genomic DNA are subjected to agarose gel electrophoresis or sucrose gradient centrifugation for separation from mixture of fragment of appropriate size. These fragments are then inserted into a suitable vector for cloning. This constitutes the shotgun approach of gene cloning. In principle, any vector can be used, but λ vector and cosmids have been most commonly used since DNA inserts of up to 23-25 kb can be cloned in these vectors. The vectors containing the insert are cloned in a suitable bacterial host.

The minimum size (number of clones or bacterial colonies) of a genomic library depends on following two factors

(a) The complexity of genome (the more complex the genome the larger the size) and

(b) The length of fragment or insert used for cloning (the smaller the fragments, the larger the number of clones for the same size genome)

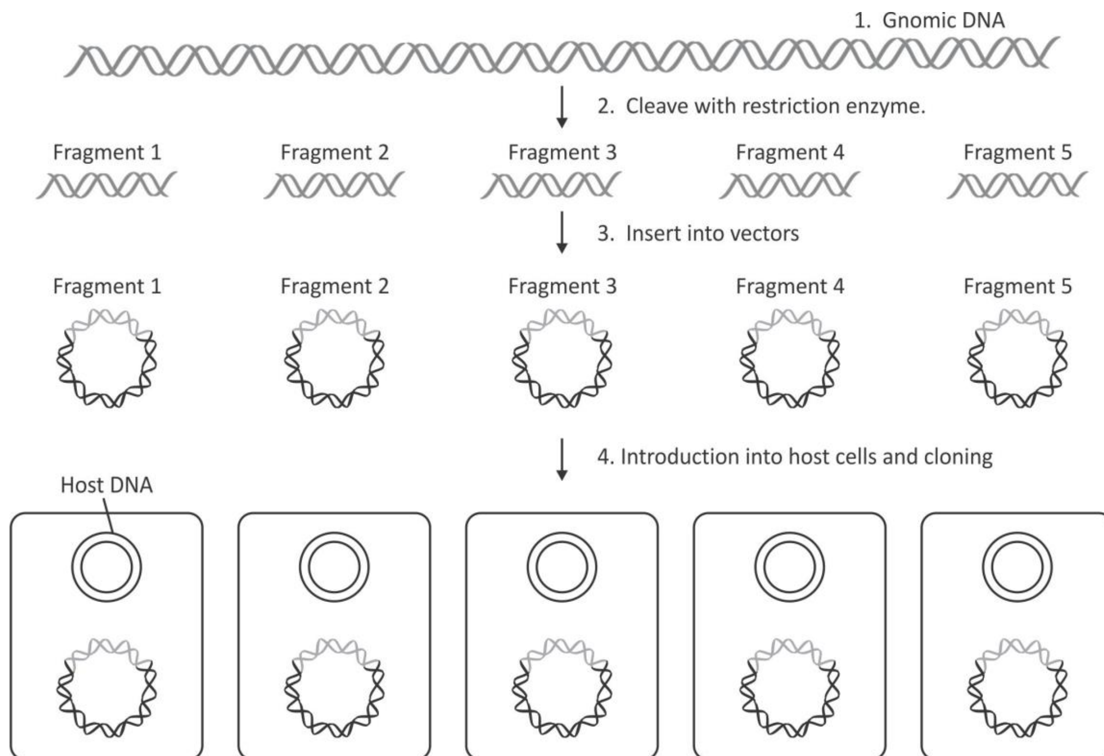


Fig. : Construction of genomic library

Difference between

Genomic Library .	cDNA Library
It includes all possible fragments of DNA from a given cell or organism.	Carries only expressed gene sequences.
It is larger .	It is smaller.
Represents the entire genome of an organism having both coding and non coding regions.	Represents only the expressed part of the genome and contain only coding sequences called ESTs.
Expression of genes taken from genomic library is difficult in prokaryotic system like bacteria due to absence of splicing mechanism.	cDNA has only coding sequences therefore can be directly expressed in prokaryotic system.
Vectors used in genomic library include plasmid, cosmid, lambda phage, BAC and YAC in order to accommodate large fragments.	Vectors used in cDNA library include plasmid, phagemid, lambda phage etc. to accommodate small fragments as cDNA has no introns.

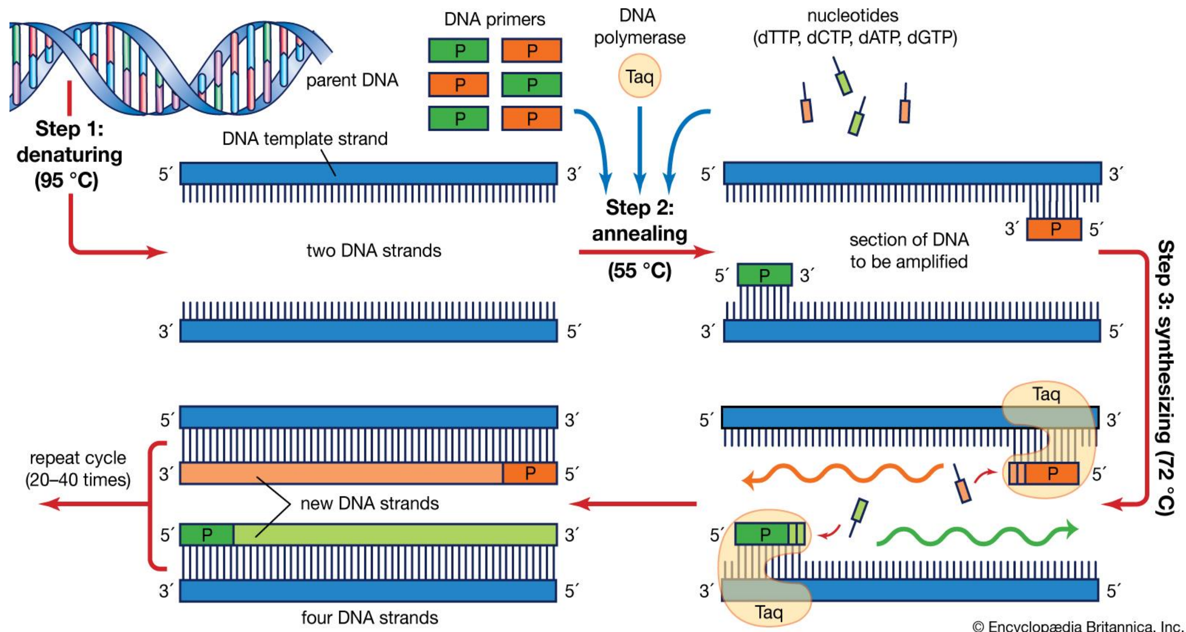
Chemical Synthesis of Gene: The amino acid sequence of the protein (or base sequence of mRNA) produced by gene enables deduction of base sequence of the gene on the basis of codons for the various amino acids. However, the degeneracy of genetic code may present some problem but the functional sequence of the gene nonetheless be worked out. Once the base sequence of gene is worked out, the polynucleotide of the same base sequence can be synthesized either chemically or even enzymatically.

The chemical synthesis of gene utilizes chemical reagents for the various steps of the process. There are three distinct methods, differing mainly in the strategies of protection of OH groups of phosphate residues (1) Phosphodiester approach (2) Phosphotriester approach and (3) Phosphite triester approach.

Gene Amplification through Polymerase Chain Reaction: The polymerase chain reaction (PCR) technique, developed by Kary Mullis in 1985, is extremely powerful. It generates upto billion copies of desired DNA or RNA segment, present even as single copy of initial preparation, in a matter of few hours. The PCR process has been completely automated and compact thermal cycler are available in the markets.

The PCR is carried out *in vitro*. It utilizes (1) DNA preparation containing the desired segment to be amplified, (2) two nucleotide primer (about 20 bases long) specific, i.e., complementary, to the two 3’ borders (the sequences present at the 3’ ends of the two strands) of desired segment, (3) the four deoxynucleoside triphosphate, viz, TTP, dCTP, dATP and dGTP and (4) a heat stable DNA polymerarase, e.g., *Taq* (isolated from bacterium *Thermus aquaticus*), *Pfu* (from P*yrococcus furiosus*) and *vent* (from *Thermococcus litoralis*) polymerase.

Procedure of PCR: At start of PCR, the DNA from which a segment is to be amplified, an excess of two primer molecules, the four deoxyribose triphosphate and heat stable DNA polymerase are mixed together in the reaction mixture chamber of thermal cycler. The following operations are performed sequentially.



A schematic representation of the three steps performed during PCR

The completion of step 3 completes the first cycle of amplification; each cycle may take few (ordinarily 1-3) minutes. Thus, at each cycle, both new and old strands anneal to the primer and serve as templates for DNA synthesis. As a result, at the end of each cycle, the number of copies of desired segment becomes twice the number present at the end of the previous cycle. Thus, at the end of n cycle 2^n copies of segment are expected. After PCR cycles, the amplified DNA segment is purified by gel electrophoresis and can be used for the desired purpose.

VECTORS: A vector is DNA molecule that has an ability to replicate in appropriate host cell and into which the fragment to be cloned (DNA insert) is inserted for cloning. Therefore, a vector must have an origin of DNA replication (represented as *ori*) that functions in the host cell. Any extra chromosomal small genome, e.g., plasmid, phage and virus may be used as a vector.

Properties of Good Vectors:

1. It should replicate autonomously. When objective is cloning of large number of copies of DNA insert, the vector replication must be under relaxed replication control so that it can generate multiple copies of itself in a single host cell.
2. It should be easy to isolate and purify.
3. It should be easily introduced into the host cell.
4. The vector should have suitable marker gene that allow easy detection of and/or isolation of the transformed cell.
5. When the objective is gene transfer, it should have an ability to integrate either itself or the DNA insert it carries into the genome of the host cell.
6. A vector should contain unique target site for as many restriction enzyme as possible into which DNA insert can be integrated without disturbing its essential function.
7. When expression of DNA insert is desired, the vector should contain at suitable control elements, e.g., promoter, operator and ribosome binding site.

It should be kept in mind that (1) the DNA molecule used as vector has coevolved with their specific host. Therefore, choice of vector depends largely on the host species into which DNA insert of gene is to be cloned. (2) Most naturally vector do not have all required functions; therefore, vectors have been created by joining together segments performing specific function (called modules) from two or more natural entities.

Cloning and Expression vectors: All vector used for propagation of DNA insert in a suitable host are called **cloning vectors**. It is desirable that all cloning vectors have relaxed replication control so that they can produce multiple copies per host cell. But when a vector is designed for expression of, i.e., production of specific protein specified by, DNA insert, it is termed as **expression vector**. As a rule such expression vector must contain at least the regulatory sequences.

When an eukaryotic gene to be expressed in a prokaryote, the eukaryotic coding sequence has to placed after prokaryotic promoter and ribosomal binding site since the eukaryotic regulatory sequence are not recognized in prokaryotes. In addition, eukaryotic gene must be intron free since prokaryote lack the mechanism needed for their removal from the RNA transcripts. Thus eukaryotic gene isolated as cDNA are intron free and hence, suitable for expression in prokaryotes.

Several strategies have been attempted for construction of expression vectors. These can be broadly grouped into

i) Translational fusion- Construction of vectors allowing the synthesis of fusion protein comprising amino acid coded by a sequence in the vector and those encoded by the DNA insert.

ii) Transcriptional fusion- Development of vector permitting the synthesis of pure protein encoded exclusively by the DNA inserts.

Host preferred to DNA cloning: Bacteria are the host of choice for DNA cloning. Among them, *E. coli*. Occupies a prominent position since cloning and isolation DNA insert for analysis is the easiest in this host. The *E. coli.* Strain K12 is the most commonly used.

Properties of a Good Host: A good host should have following features:

1. is easy to transform
2. supports the replication of recombinant DNA
3. is free from elements that interfere with replication of recombinant DNA
4. lacks active restriction enzyme
5. does not have methylases since these enzyme would methylate the replicated recombinant DNA which as a result, would become resistant to useful restriction enzyme, and
6. is deficient in normal recombination function so that DNA insert is not altered by recombination events.

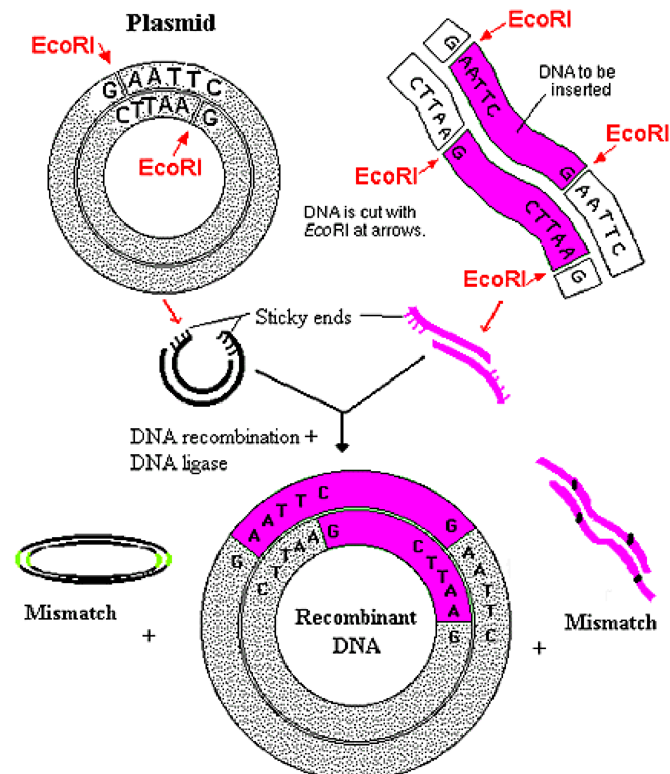
***E. coli* Vectors:** *E. coli.* Supports several types of vectors, some natural, some constructed, which can be grouped as: (1) Plasmids, (2) Bacteriophage (both natural), (3) Cosmids, (4) Phasmid and (5) Shuttle vector

Plasmid: A plasmid is a DNA molecule, other than the bacterial chromosome that is capable of independent replication and transmission. Plasmids are circular and may exist either independent of or may become integrated into the bacterial chromosome; generally they are not essential for the host cell except under specific environment. There are several types of bacterial plasmids, but the three widely studied types are: (1) **F plasmids** (responsible for conjugation). (2) **R plasmids** (carry genes for resistance to antibiotics) and (3) **Col plasmids** (code for colicins, the proteins that kill sensitive E coli cells; they also carry genes that provide immunity to the particular colicin).

The plasmids may either be **conjugative or transmissible** (mediate DNA transfer via conjugation, and as a result spread rapidly among the bacterial cells of population), e.g., F plasmids, many R plasmids and some Col plasmids, or **nonconjugative** (do not mediate DNA transfer through conjugation), e.g., many R plasmids and most Col plasmids.

Stringent and Relaxed Replication: Each plasmid is maintained in the bacterial cell at a characteristic copy number mainly due to its replication control system. In this respect the plasmids are of two types: (1) single copy and (2) multi copy plasmids. The replication control of **single copy plasmids** is the same as that of their bacterial host cells so that they replicate and segregate with the bacterial chromosome; this is called **stringent replication**. In contrast, the replication control of **multicopy plasmids** is different from that of their bacterial host genome so that they undergo more than one replication for each replication of their host genome; this is referred to as **relaxed replication**.

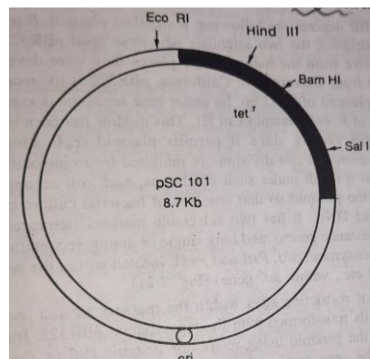
Plasmid Vectors: Many different *E. coli*. Plasmids are used as vector. The natural plasmid have been modified, shortened, reconstructed and recombined both in *in vitro* and *in vivo* to create plasmids of enhances utility and also with specific functions.



Inserting a DNA Sample into a Plasmid

Selection of Recombinant Vectors: When an experiment is performed to insert a DNA fragment into a vector, two type of vectors molecule are obtained (1) many vector molecules will contain the DNA insert (**recombinant** or **chimaeric vector**) and (2) many other will contain the vector sequence (**unaltered vector**). This mixture of vector molecule is used for transformation of host cell. (1) some host will receive the recombinant vector, (2) some other will contain the normal unaltered vector, while (3) the majority of them will contain no vector, i.e., will not be transformed. In a cloning experiment it is critical to effectively select for the low frequency of cells transformed by the recombinant vector from among the cells containing the unaltered vector and the non transformed cells.
Selection of host cells transformed by the recombinant vector is easily achieved by placing two selectable markers, e.g., antibiotic resistance genes, such as, ampicillin resistance (amp^r) and tetracycline resistance (tet^r) in the vector. The DNA insert is integrated within one of the two selectable markers. If the DNA insert is integrated within the ampicillin resistance (amp^r) gene, the cells containing the recombinant vector will be resistant to tetracycline but sensitive to ampicillin. In contrast, no transformed cells will be sensitive to both the antibiotics, while the cell containing the unaltered vector will be resistant to both. Therefore, following transformation with the above recombinant vector, cells are plated on a tetracycline supplemented medium; this eliminates the nontransformed cells. The remaining colonies are now replica - plated on ampicillin containing medium to identify sensitive colonies; these colonies contain the recombinant vector, and are isolated from the master plate.

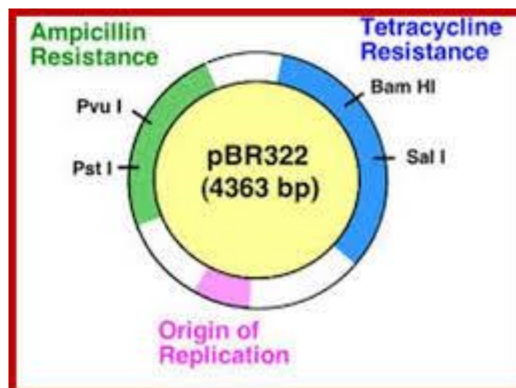
pSC101: Isolated in 1973 by Herbert Boyer and Stanley Norman Cohen as first natural plasmid from *Salmonella typhimurium.* The "SC" stands for Stanley Cohen. This plasmid vector contains the replication module (*ori*) for replication in *E. coli.*, tet^r gene for resistance to tetracycline and single recognition sites for restriction endonuclease *Eco*RI (outside tet^r gene), *Hind*III, *Bam*HI and *Sal*I (within tet^r gene).
Insertion of DNA insert into the *Eco*RI site leaves the tet^r gene intact and functional; as a result, *E. coli*. cells transformed by *p*SC101 become tetracycline resistant but such cells may either have a non recombinant *p*SC101 (without the DNA insert) or a recombinant one (having the DNA insert). On the other hand, insertion of the DNA fragment into the *Hind*III, *Bam*HI or *Sal*I sites disrupts the tet^r gene and makes it nonfunctional, therefore, cells transformed by such a recombinant pSC101 are sensitive to tetracycline and hence can be easily distinguished from those containing the nonrecombinant plasmid which are resistant to the antibiotic. But nontransformed cells too are tetracycline sensitive hence they cannot be separated from those having the recombinant *p*SC101. Clearly *p*SC101 does not permit a direct selection of cells containing the recombinant vector. In addition, it contains unnecessary DNA, and has stringent regulation of replication. Subsequently, several novel plasmid vectors were designed to overcome these deficiencies.



An ideal plasmid vector must have the following functions: (I) minimum amount of DNA, (2) relaxed replication control, (3) at least two selectable markers, (4) only one (unique) recognition site for at least one restriction endonuclease, and (5) for easy selection of the recombinant vector, this unique restriction site must be located within one of the two selectable markers.
Almost all such vectors have been constructed from naturally occurring DNA sequences using both classical genetic (recombination) and recombinant DNA techniques.

pBR322: The name pBR denotes the following p signifies plasmid, B is from Boliver, and R is from Rodrigdez, the two scientists who developed **p**BR322. Some other plasmid names derive from the names of the places they were developed at, e.g., pUC gets its name from University of California. pBR322 is the most popular and most widely used plasmid of 4362 bp. Its entire base sequence is known. It has the replication module of *E coli* plasmid Col El. This module has been included in many other plasmid vectors since it permits plasmid replication even when chromosome replication and cell division are inhibited by amino acid starvation or chloramphemicol; As a result under such conditions, each cell accumulates several thousand copies of the plasmid so that a liter of bacterial culture easily yields a milligram of plasmid DNA. It has two selectable markers (tetracycline, *tet*^{r}, and ampicillin, *amp*^{r}, resistance genes). and only single or unique recognition sites for 12 different restriction enzymes (two. *Pst*I and *Pvu*I, located within the *amp*^{r} gene, and 4, e.g., *Bam*HI, *Sal*I etc., within *tet*^{r} gene).

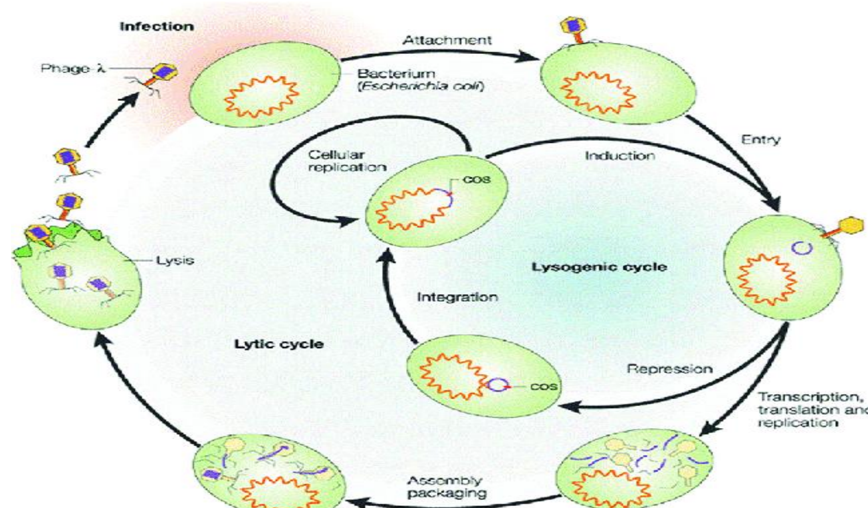
pBR322 has *nic-bom* region which is responsible for its mobilization. **Mobilization** refers to the cell to cell transfer of otherwise nonconjugative plasmid in the presence of a conjugative plasmid provided that both are present in the same cell. The *nic-bom* region also interferes with replication efficiency of extrachromosomal DNA in monkey cell; such prokaryotic sequences are called poison sequences. The poison sequence of pBR322 has been deleted (deletion between bases 1442 and2502) to yield pBR327 and pAT153 (deletion between bases 1644 and 2345)



pUC7: It is a derivative of pBR322 and is much smaller (2.7kb); it has all essential parts of pBR322, e.g., ampicillin resistance gene and Col E1 origin. The second marker is scorable marker, i.e., *E. coli*. Gene *lac*Zα encoding the α fragment of β-galactosidase, the enzyme that hydrolyses galactose. The specific *E.coli*. strains are used as host, e.g., JM103 and JM109 etc. for pUC series vectors in which *lac*Zα is deleted from their *lac* operon. When pUC enters such an *E.coli*. cell, the host genome and the plasmid encode for different parts of the β-galactosidase enzyme which interact with each other to produce the active enzyme enabling these cells to hydrolyse galactose. β-galactosidase also hydrolyses X-gal (5-Brom-4-chloro-3-indoly-β-D-

galactoside) to yield a blue dye. Therefore, appropriate lacZ$^-$ *E.coli*. cells transformed bu pUC vectors behave as lacZ$^+$ and produce blue colored colonies on X al containing medium. A polylinker sequence located within the lacZα provides several unique restriction sites for DNA insertion. The polylinker sequences by itself does not interfere with lacZα expression, but when DNA is placed within it lacZα expression is prevented
The unique restriction sites used for integration of DNA inserts into pUC vectors interrupt the lacZα fragment so that appropriate E.coli cell possessing recombinant pUC vectors are β-galatosidase deficient and as a result, produce white colonies on X-gal medium. Therefore appropriate *E.coli*. cells transformed with pUC vectors are first grown in ampicillin containing medium to eliminate nontransformed cell. The colonies so obtained are replica plated on X -gal containing medium the white colonies are selected as they contain the recombinant vector (in contrast, blue colonies will contain the unaltered vector).

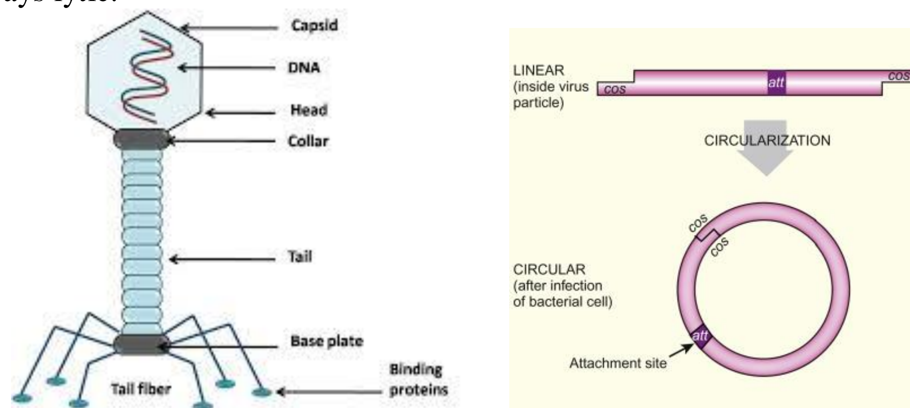
BACTERIOPHAGE VECTORS: Bacteriophages are viruses that attack bacteria. Most phages lyse the bacterial cells they infect (lytic plaque). But many others can choose to follow either a lytic or a lysogenic cycle; in the latter situation, the phage chromosome integrates into the bacterial chromosome and multiplies with the latter as **prophage (temperate** or **lysogenic phages**). The prophage may dissociate from the bacterial chromosome and follow the lytic cycle.



LYTIC AND LYSOGENIC CYCLE OF BACTERIOPHAGE

Several bacteriophage are used as cloning vectors, most commonly used *E.coli.* phages being λ (lambda) and M13 phages. Plasmtid vectors have to be introduced into bacterial cells which are then cloned and selected for the recovery of recombinant vectors. In contrast, the phage vectors are directly tested on appropriate **bacterial lawn** (a continuous bacterial growth on an agar plate) where each phage particle forms a **plaque** (a clear bacteria-free zone in the bacterial lawn). Phage vectors present two advantages over plasmid vectors. (1) They are more efficient than plasmids for cloning of large DNA fragments; the largest cloned insert size in λ vector is just over 24 kb, while that for plasmid vectors it is less than 15 kb. In addition, (2) it is easier to screen a large number of phage plaques than bacterial colonies for the identification of recombinant vectors.

λ Phage Vectors: The λ genome (total 48,502 bp contains an origin of replication: genes for head and tail proteins and enzymes for DNA replication, lysis and lysogeny, and single- stranded protruding cohesive ends of 12 bases (5'GGGCGGCGACCT3'; (the other end is complementary to it ,i.e., 3'CCCGCCGCTGGA5'). The genome remains linear in the phage head, but within *E.coli.* cells the two cohesive ends anneal to form a circular molecule necessary for replication. The sealed cohesive ends are called *cos* sites which are the sites of cleavage during and are necessary for packaging of the mature phage DNA into phage heads. *The λ DNA must be larger than 38 kb and smaller than 52 kb to be packaged into phage particles.* The genes for lysogeny are located in the segment between 20 and 38 kb; the whole or a part of this segment is deleted to create λ vector to (1) accommodate larger DNA insert and (2) to ensure that the recombinant phage is always lytic.

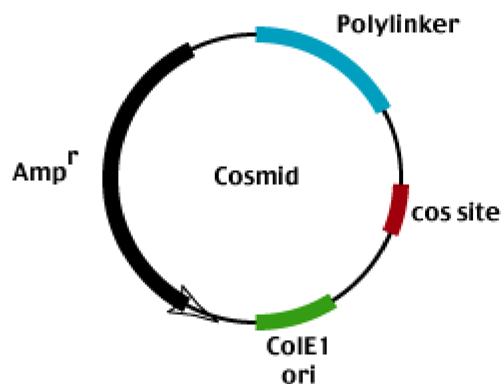


LAMBDA BACTERIOPHAGE AND ITS COS SITE

Phage M 13 Vectors: These vectors are used for obtaining single - strand copies of cloned DNA which are especially suited for DNA sequencing. They are derived from the 6.4 kb genome of the *E.coli.* filamentous bacteriophage M13. This phage has a single - stranded linear DNA genome in phage particles which converts into a double-stranded circular replicative intermediate within the host cells. M13 infects only F^+ cells; it does not kill the cells, but forms turbid plaques due to growth retardation of infected cells. Ordinarily, the double - stranded form is used to produce recombinant molecules since single - stranded DNAs are not cleaved by typeII restriction endonucleases; This form is readily isolated from M 13 - infected *E. coli.* cells. However, the single - stranded form of M 13 is used to recover single - stranded copies of the DNA inserts; This form of vector is available from the phage particles abundant in the growth medium.

The desirable features **of M 13** vectors are as follows: (1) very large inserts can be cloned since packaging does not depend on genome size (as is the case with i vectors) (2) Pure single strand copies of double - strand DNA inserts are obtained in abundance (3) Since DNA inserts are accepted in either of the orientations (which is also the case for plasmid and λ vectors), some recombinant ones will produce single strand copies of one strand of the DNA double strand, while others would produce copies of the complementary strand of DNA insert. The phage particles in a single plaque, as a rule, will yield copies of the same single - strand. This property is very useful for a precise DNA sequencing (using both the strands of a DNA molecule) and for the synthesis of specific radio-labeled DNA probes. (4) Bacterial cells infected by these vectors remain viable as in the case of plasmid vectors this allows easy maintenance of the vector finally, (5) they form plaques like a phage vectors making selection of the recombinant vectors rather easy.

COSMID VECTORS: Cosmids are essentially plasmids that contain a minimum of 250 bp of DNA which includes (1) the *cos* site (the sequence yielding cohesive ends) and (2) sequences needed for binding of and cleavage by terminase so that under appropriate conditions they are packaged *in vitro* into empty a phage particles. A typical cosmid has (1) replication origin, (2) unique restriction sites and (3) selectable markers from the plasmid; therefore selection strategy for obtaining the recombinant vectors is based on that for the contributing plasmid. Cosmid vectors are constructed using recombinant DNA techniques. The cosmid vectors are opened by the appropriate restriction enzyme at a unique site, are then mixed with DNA inserts prepared by using the same enzyme and annealed. Cosmids can accommodate upto 45 bp long DNA inserts. Packaged cosmids infect host cells Iike λ particles, but once inside the host they replicate and propogate like plasmids



The typical features of cosmids are as follows: (1) they can be used to clone DNA inserts of upto 45 kb. (2) They can be packaged into a λ particles which infect host cells, which is many - fold me efficient than plasmid transformation (3) Selection for recombinant vector is based on the procedure applicable to the plasmid making up the cosmid (4) Finally, these vectors are amplified and maintained in the same manner as the argued plasmid.

PHASMID VECTORS: These vectors are shortened linear λ genomes containing DNA replication and lytic function plus the cohesive ends of the plage, their middle nonessential segment is replaced by a linearized plasmid with intact replication module. In practice, a phasmid vector contains several tandem copies of the plasmid to make it longer than 38 kb, the minimum size needed for packaging in a λ particles. During construction of the recombinant DNA, one or more copies of the plasmid are deleted from and the DNA insert is integrated into the vector, but generally one copy of the plasmid is retained in the recombinant vector. Phasmids, both recombinant and unaltered, are packaged in λ particles *in vitro* and used for infection of appropriate *E.coli* cells.

SHUTTLE VECTORS: These vectors have been designed to replicate in cells of two different species; therefore, they contain two origins of replication, one specific for each host species, as well as those genes necessary for their replication and not provided by the host cells. These vectors are created by recombinant techniques. Some of them can be grown in two different prokaryotic species, while others can propagate in a prokaryotic species, usually *E.coli*, and a eukaryotic one, e.g., yeast, plants, animals. Since these vectors can be grown in one host and then moved into another without any extra manipulation, they are called shuttle vectors.

INTRODUCTION OF THE VECTOR INTO A SUITABLE HOST: The recombinant vector is constructed *in vitro*; it is then generally introduced into *E.coli*. to (1) select the recombinant from the unchanged vector, (2) to obtain many copies of the recombinant vector or the DNA insert or (3) to express the insert in *E.coli*. itself. Purified recombinant vector may subsequently be introduced into another bacterium, e.g., *Bacillus subtilis, streptomyces* etc., yeast, higher plants or animals

Cells are generally poorly accessible to DNA molecules. But treatment with $CaCl_2$, makes them permeable to DNA; the process involved is poorly understood. Growing *E.coli*. cells are isolated and suspended in 50 mM $CaCl_2$ at a concentration of $10^8 – 10^{10}$ cells / ml; the cells may be incubated for 12-24 hr to increase the frequency of transformation. The recombinant vector is then added; efficient transformation takes only a few minutes and the cells are plated on a suitable medium for the selection of transformed clones. The frequency of transformed cells is 10^6-10^7 per µg of plasmid DNA; this is about one transformation per 10,000 plasmid molecules.

SELECTION OF RECOMBINANT CLONES: When recombinant vector is constructed and used for transformation of Ecoli, the following types of bacterial cells are obtained (1) majority of the cells nontransformed. (2) a majority of the transformed cells contain nontransformed vector while (3) the remaining cells have recombinant vector. The first objective of cloning experiments is to identify and isolate those small numbers of cells that contain the recombinant vector from among a very large number of transformed cells. Since the DNA insert are generally mixtures, particularly when cDNA preparation and genomic fragments are used, the various transformed clones contain a variety of different DNA insert. The next step, therefore, is to identify the clone having the desired DNA insert from among the large number of clones containing the recombinant vectors. Suitable selection strategies have been devised to achieve these two critical objectives and is the most important step in DNA cloning. This is generally achieved by inserting a selectable marker gene or reporter gene into the vector.

Reporter Genes: A marker gene or reporter gene produces a phenotype which permits either an easy selection or quick identification of the cells in which it is present. Thus marker genes are either **selectable or scorable**. A **selectable marker** governs a feature which enables only such cells that possess it to survive under the selective conditions. For example, genes conferring resistance to antimetabolites, e.g., an antibiotic like kanamycin, are good selectable markers. When a population of bacterial cells is plated on a kanamycin containing medium, only those cells that have the kanamycin resistance gene (kan^r) survive and form colonies. On the other hand, **scorable markers** produce distinct phenotypes which allow easy identification of the cells having them from those which do not contain them. Examples of such genes are gus (β-galactourinidase; produces blue color in the presence of appropriate substrate), *ocs* (octopine synthase), *nos* (nopaline synthase); both *ocs* and *nos* produce specific nitrogen compound called opines) etc. It should be noted that scorable markers do not allow selective multiplication of the cells having them; they only enable their easy identification.

MULTIPLICATION, EXPRESSION AND INTEGRATION OF THE DNA INSERT IN HOST GENOME: Once the clone containing the desired DNA insert is identified, it is multiplied in *E.coli*. to obtain sufficient number of copies to be used in one or more of the following ways. (1) It can be used for a structural analysis of the insert, e.g., DNA sequencing, chromosome walking etc. (2) It may be introduced into a bacterium like *B.subtilis* for production of the protein encoded by the insert since this host secretes proteins into the medium which allows easy purification. (3) It can be introduced into a eukaryotic host, eg, yeast, animal cells plants etc., either to explore the function of the insert or (4) to integrate it into the host genome.