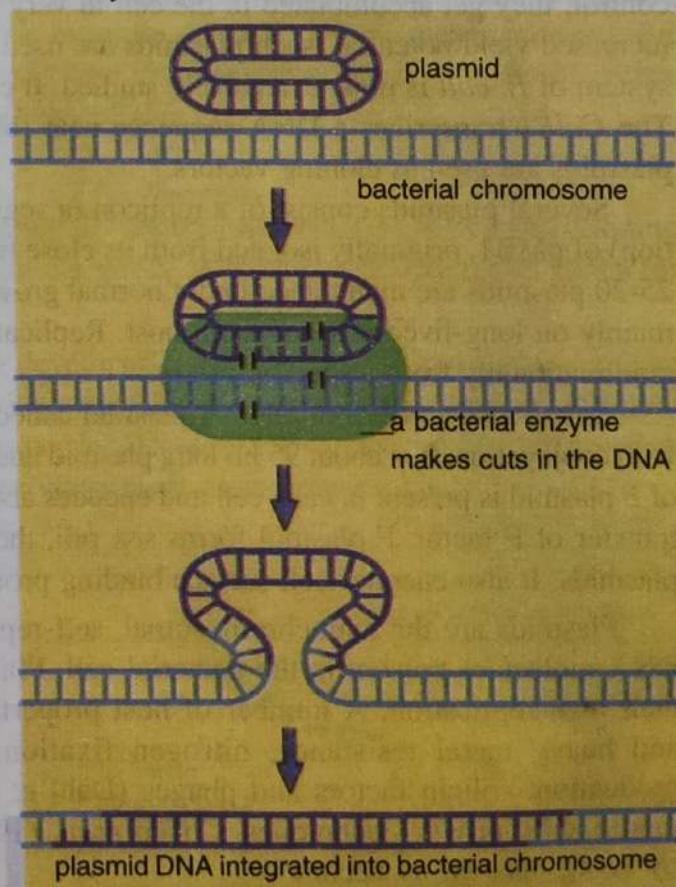




## Tools of Genetic Engineering : III- Cloning Vectors

In recent years a remarkable development has been made in manipulation of prokaryotic and eukaryotic DNA. You can isolate DNA from different sources and cut enzymatically at desired place and re-join the DNA of two organisms at desired site. Similar to DNA enzymes, the other most important requirements for recombinant DNA technology is the cloning and expression vectors. The recombinant DNA is produced by cloning a foreign DNA isolated either from the genome or synthesised chemically or as cDNA using mRNA molecule. However, cloning of this DNA can be done only when 'another DNA molecule' is available that may replicate in the transformed host cell. This 'other DNA molecule' used for joining the foreign DNA is called **vectors**. The vectors are the DNA molecules that can carry a foreign DNA segment and replicate inside the host cell. Vectors may be plasmids, a bacteriophage, cosmids, phagemids, transposons, a virus YAC or BAC. Moreover, a vector must possess the following characteristics:

- It must replicate (through *ori* gene) in host cells after its introduction.
- It must contain marker genes such as *tet*<sup>R</sup> (for tetracycline resistance), *kan*<sup>R</sup> (for kanamycin resistance), *amp*<sup>R</sup> (for ampicillin resistance). These help in selection of transformed cells from untransformed cells.
- A unique cleavage site must be present in one of the marker genes.



Plasmid integration into a bacterial chromosome.

By inactivating it, through restriction enzyme one can detect the recombinant molecules. However, many vectors used commonly contain several recognition and cleavage sites for several restriction enzymes in a small region. These are called *multiple cloning sites* (MCS) or *polylinkers*. Presence of MCS facilitates the use of restriction enzymes of choice.

- (iv) It should contain specific control systems like promoters, terminators, ribosome binding sites, etc. so that the cloned DNA should express properly.

The best known of these vectors is the plasmid vector. The largest number exists for *E. coli*. During the early 1970s, Stanley Cohen and Herb Boyer selected *E. coli* strain K12 for conducting research work on molecular biology. Since then *E. coli* K12 has been used for various experiments in recombinant DNA technology and production of products. One of the unique features of *E. coli* is that it is found in human intestine and cannot thrive outside the laboratory environment. *E. coli* strain K12 mutant in restriction-modification system is used in gene cloning. *E. coli* strain K12 mutant in *endH* gene (encodes DNA specific endonuclease) increases the yield of plasmid DNA and improves DNA quality. The *recA* gene is associated with expression of recombination proteins. Therefore, *E. coli* strain K12 mutant in *recA* gene ensures about the biosafety of *E. coli* strain K12 because they can easily be killed by ultra violet light.

## A. BACTERIAL PLASMID VECTOR

The size of plasmids varies from 1 to 200 kb. It depends on host proteins for replication, maintenance and other functions. They are present in a characteristic number of copies in a bacterial cell, yeast cell, in eukaryotic organelles such as mitochondria. The plasmids are maintained as single copy per cell and they are called **single copy plasmid**. There are **multicopy plasmid** also which are maintained as 10–20 copies per cell. Apart from this when the plasmids are under the relaxed replication control, they get accumulated in the cell in very large copies i.e. over 1000 copies per cell. Due to increased yield potential, such plasmids are used as cloning vectors. The *ColE1* plasmid replication system of *E. coli* is most extensively studied. It expresses colicins which is an antibacterial protein. The *ColE1* transcribes a DNA sequence near the *ori* gene. Due to increased yield potential these plasmids are used as cloning vectors.

Several plasmids consist of a replicon (a segment of genome that contains an origin of replication) of pMB1, originally isolated from its close relative, *ColE1* replicon. In each bacterial cell about 25–20 plasmids are maintained under normal growth conditions. The *ColE1* or pMB1 replicon relies mainly on long-lived enzyme from host. Replication is primed by an RNA primer and takes place unidirectionally from a specific origin.

*E. coli* also consists of another plasmid called **fertility factor** or **F plasmid** which is associated with conjugation. It is about 95 kb long plasmid and transfer F factor to F<sup>-</sup> (F-deficient) cells. One copy of F plasmid is present in each cell and encodes about 100 genes. Some of them are needed during the transfer of F factor. F plasmid forms sex pili, the number of which is governed by the number of plasmids. It also encodes cell surface binding proteins for the filamentous bacteriophage M13.

Plasmids are the extrachromosomal, self-replicating, and double stranded closed and circular DNA molecules present in the bacterial cell. Plasmids contain sufficient genetic informations for their own replication. A number of host properties are specified by plasmids, such as antibiotic and heavy metal resistance, nitrogen fixation, pollutant degradation, bacteriocin and toxin production, colicin factors and phages (Dahl *et al.*, 1981). Naturally occurring plasmids can be modified by *in vitro* techniques. Cohen *et al.* (1973) for the first time reported the cloning DNA by using plasmid as vector.

- A plasmid can be considered a suitable cloning vehicle if it possesses the following features
- (i) It can be really isolated from the cells,
  - (ii) It possesses a single restriction site for one or more restriction enzyme(s),

- (iii) Insertion of a linear molecule at one of these sites does not alter its replication properties,
- (iv) It can be reintroduced into a bacterial cell and cells carrying the plasmid with or without the insert, can be selected or identified (Bernard and Helinski, 1980),
- (v) They do not occur free in nature but are found in bacterial cells.

In many cases the principal objective of cloning experiment is the insertion of a particular restriction fragment into a suitable plasmid vector and its amplification.

### Amplification of plasmids

Amplification is a process of increasing the number of a plasmid in a bacterial cell. In this process, a cell containing a relaxed plasmid is treated with a drug to inhibit protein synthesis. Consequently, cells stop replicating. The relaxed plasmid pBR322 continues to replicate despite drug treatment. Replication of relaxed plasmid neither depends on cell replication nor requires protein synthesis. For example, addition of chloramphenicol causes pBR322 to get increased about 3000 per cell. Finally the ratio of plasmid DNA to chromosomal DNA is increased which makes easy to isolate the plasmid DNA. In order to replicate and clone a fragment of foreign DNA, it is necessary to possess a sequence of nucleotides which are recognised by the host bacterium as an origin of replication. The origin of replication occurs naturally in plasmid and transferred to onward progenies. During cell division these genes are transferred resulting in gradual decrease in number of such genes. However, due to continuous exchange of genetic materials between plasmid and chromosomal DNA, new genes originate with respect to environmental conditions. Therefore, plasmids have been artificially developed from naturally occurring form and only special features that help in cloning have been preserved.

There are several plasmid cloning vectors such as pBR320, pSC102, CoIE1, pUC, pRP4, pRK2, pRSFIO1O, pEY, pWWO, Ti- and R-DNA plasmids, etc. The plasmid cloning vectors are designated by 'p', some abbreviations and a few numbers. For example, in pBR320, 'p' means *plasmid*, BR refers to the researchers F. Bolivar and R. Rodriguez who discovered the plasmid, and 320 is the numerical designation. Some of the plasmids are given in Table 5.1.

### 1. pBR322

The plasmids that occur naturally do not possess all the characteristics to be used as cloning vector. Therefore, they are constructed by inserting the genes of relaxed replication and genes for antibiotic resistance.

The pBR322 is the first artificial cloning vector developed in 1977 by Boliver and Rodriguez from *E. coli* plasmid ColE1. It is 4.362 kb long and most widely used cloning vector. In the transformed *E. coli* cells its copy number remains about 15-20 molecules per cell. However, its number may be amplified by incubating the transformed *E. coli* cells on a medium containing chloramphenicol.

> Table 5.1. Some cloning vectors.

Cloning vectors	Natural occurrence	Size (Kb)*	Selective marker**
<b>Plasmids</b>			
pACYC 177	<i>Escherichia coli</i>	3.7	Amp <sup>r</sup> , Kan <sup>r</sup>
pBR322	<i>E. coli</i>	4.0	Amp <sup>r</sup> , tet <sup>r</sup>
pBR324	<i>E. coli</i>	8.3	Amp <sup>r</sup> , tet <sup>r</sup> , El imm.
pMB9	<i>R. coli</i>	5.8	Tet <sup>r</sup>
pRK646	<i>E. coli</i>	3.4	Amp <sup>r</sup>
pC194	<i>Staphylococcus aureus</i>	3.6	Ery <sup>r</sup>
pSA0501	<i>S. aureus</i>	4.2	Str <sup>r</sup>
pBS161-1	<i>Bacillus subtilis</i>	3.65	Tet <sup>r</sup>

pWWO	<i>Pseudomonas putida</i>		
Cosmids			
pJC74	Derived plasmid from Col EL	16	Amp <sup>r</sup> , El imm.
pJC720	do	24	El imm., Rif <sup>r</sup>
pHC79	Derivative of pBR322	6	Amp <sup>r</sup> , Tet <sup>r</sup>
Viruses			
SV40	Mammalian cells	5.2	—
Phage M13 <sup>+</sup>	<i>E. coli</i>	6.4	—
Phage λ	<i>E. coli</i>	4.9	—

\* 1 Kb (Kilobase pairs) = 1,000 base pairs = 0.66 mega dalton;

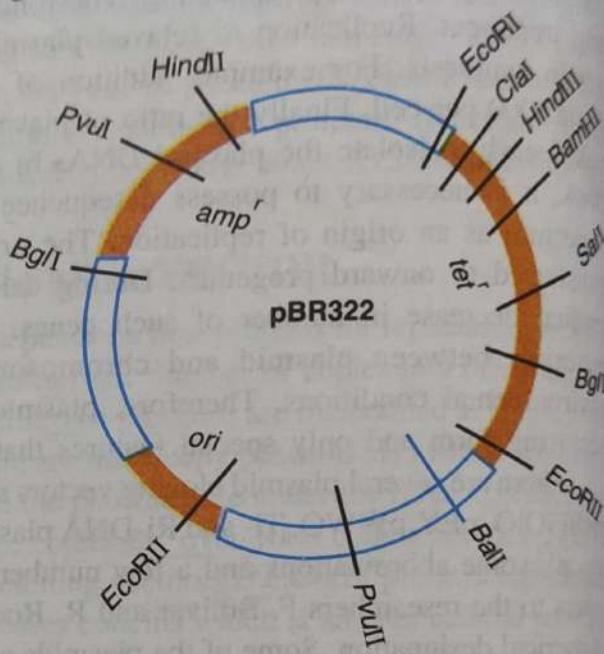
\*\*, Resistance to ampicillin (Amp<sup>r</sup>), tetracycline (Tet<sup>r</sup>), erythromycin (Ery<sup>r</sup>), streptomycin (Str<sup>r</sup>), Kanamycin (Kan<sup>r</sup>), rifampicin (Rif<sup>r</sup>), and colicin production (EL imm.)

A physical map of plasmid pBR322 is shown in (Fig. 5.1). The pBR322 is constructed from the plasmids of *E. coli*, pBR318 and pBR320. It contains origin of replication (*ori*) that was derived from a plasmid related to naturally occurring plasmid *ColEl*. Therefore, its replication may be more faster than bacterial DNA. It also possesses genes conferring resistance to antibiotics e.g. ampicillin (*amp<sup>r</sup>*) and tetracycline (*tet<sup>r</sup>*), and unique recognition sites for 20 restriction endonucleases. Six of these 20 sites such as *EcoRIV*, *BamHI*, *SphI*, *Sall*, *XmaIII* and *NruI* are present within the gene coding for tetracycline resistance, two sites (*HindII* and *Clal*) are located within the promoter of the tetracycline resistance gene, and the three sites (*PstI*, *PvuI* and *Scal*) within the beta-lactamase gene that provide resistance to ampicillin. If any foreign DNA is cloned into any of these 11 sites, insertional inactivation of any of the antibiotic resistance markers takes place. For example, if a foreign DNA molecule is inserted at *tet<sup>r</sup>* gene cluster, the property of tetracycline resistance will be lost. The recombinant plasmid will allow the cells to grow only in the presence of ampicillin but will not protect them against tetracycline. The presence of an antibiotic resistant gene in a plasmid of bacterium will confer resistance to that antibiotic.

Therefore, the antibiotic resistant cells can be selected by culturing the cells on nutrient medium supplemented with ampicillin or tetracycline.

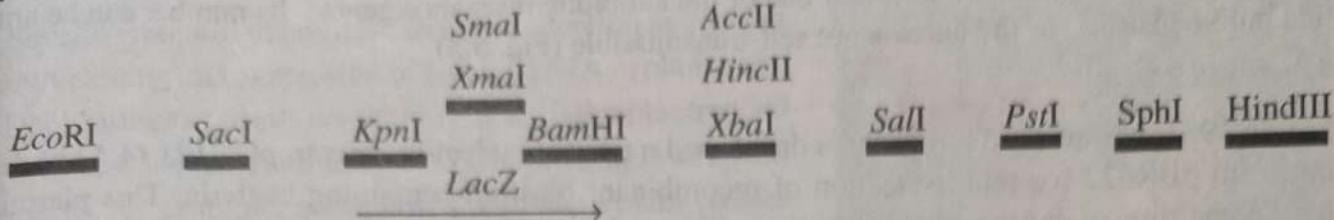
## 2. pUC Vectors

In 1983, Messings and co-workers developed the pUC vectors at the University of California. These plasmids consist of *ColE1 ori* (origin of replication) gene, *amp<sup>R</sup>* gene (for ampicillin resistance), *lacZ'* gene (the shortest derivative of *lacZ* operon for *E. coli* β-lactamase). The pUC19 is one of the popular plasmid vectors of pUC family (Fig. 5.2). The engineered version (*lacZ'*) of its *lacZ* gene consists of multiple restriction sites within the first part of the coding region of the gene. This region is known as **multiple cloning site** (MCS). If a target DNA is inserted at any of these sites, the *lacZ* gene is inactivated. Hence, such recombinant vector will form white colony on a suitable plate as compared to nonrecombinant vector. In cloning experiment the choice for application of restriction



**Fig. 5.1.** A map of pBR322 (4.36 Kb) showing a number of restriction sites and regions encoding for resistance to ampicillin (*amp<sup>r</sup>*) and tetracycline (*tet<sup>r</sup>*), and origin of replication (*ori*)

enzymes is increased after using as MCS. A MCS at the 5'-end of *lacZ* is given below:

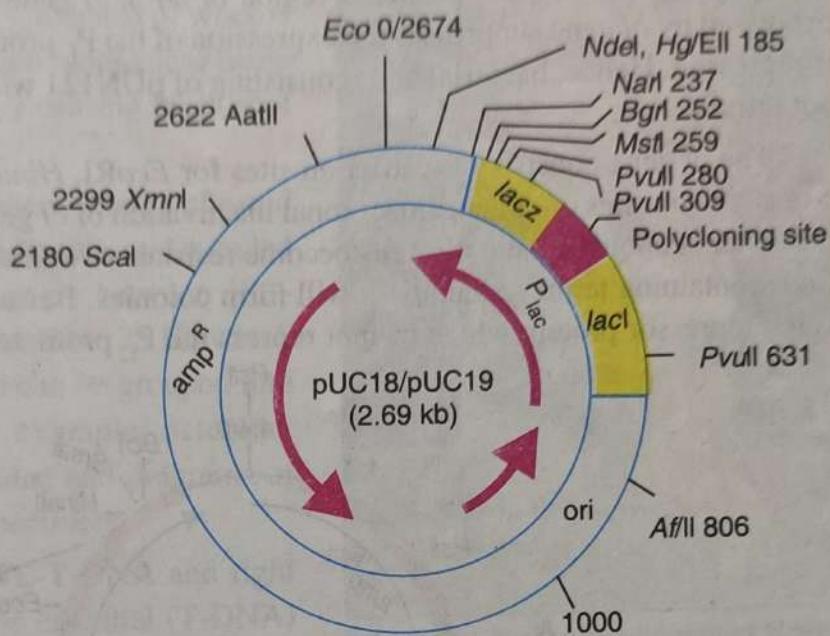


There are several plasmids of pUC series such as pUC8, pUC9, pUC12, pUC13, pUC18, pUC19, etc. The pUC19 is called as general purpose plasmid cloning vector. Such vectors are designated for cloning of relatively small (<10 kb) DNA fragments often in *E. coli* at MCS. It consists of 2,686 base pairs and possesses an *amp<sup>R</sup>* gene, *lacZ* gene (for β-galactosidase) and *lacI* gene (for production of repressor protein to regulate *lacZ*). In pUC19, MCS is incorporated into *lacZ*' gene without interfering the function of other genes. The MCS has the unique sites for several restriction endonucleases such as *Eco*RI, *Sac*I, *Kpn*I, *Xma*I, *Sma*I, *Bam*HI, *Xba*I, *Sal*I, *Hind*II, *Acc*I, *Bsp*MI, *Pst*I, *Hind*III. It also carries one *ori* gene from pBR322. This plasmid works in prokaryotes.

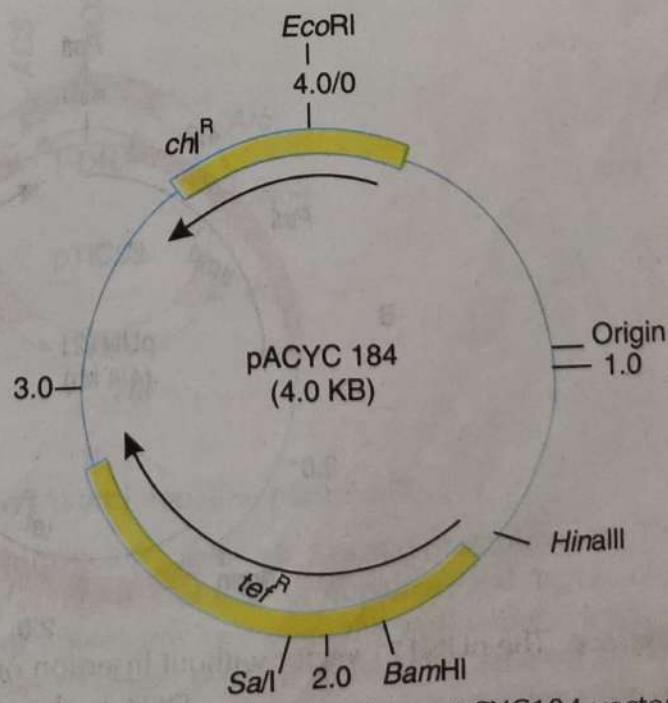
When the cells carrying unchanged pUC19 are grown in the presence of isopropylthiogalactoside (ITG) (an inducer of *lac* operon), the production of *lacI* gene cannot bind to the promoter-operator region of the *lacZ* gene. Therefore, the *lacZ* gene encoded in the plasmid is transcribed. The *lacZ*' protein combines with chromosomal DNA-transcribed protein and forms an hybrid β-galactosidase. The MCS is tagged in the *lacZ*' gene of pUC19 plasmid which does not interfere the production of hybrid β-galactosidase. On the other hand if a substrate 5-bromo-4-chlorionolyl-β-galactosidase (X-gal) is present in the medium, the hybrid β-galactosidase hydrolyses it to produce a blue product. Therefore, colonies containing unmodified pUC19 appear blue in colour. When the DNA insert is cloned in *lacZ*' gene, this gene will be inactivated and will not express β-galactosidase. Consequently, the colonies formed by *E. coli* cells containing recombinant plasmid will be white in colour. This plasmid vector is designed for blue-white screening.

### 3. pACYC184

The plasmid p15A was used to develop the plasmid pACYC184. It is a small sized plasmid (4 kb). As described in pBR322, it consists of several restriction sites, *chl<sup>R</sup>* (for chloramphenicol resistance)



**Fig. 5.2.** Genetic map of the plasmid cloning vector pUC19.



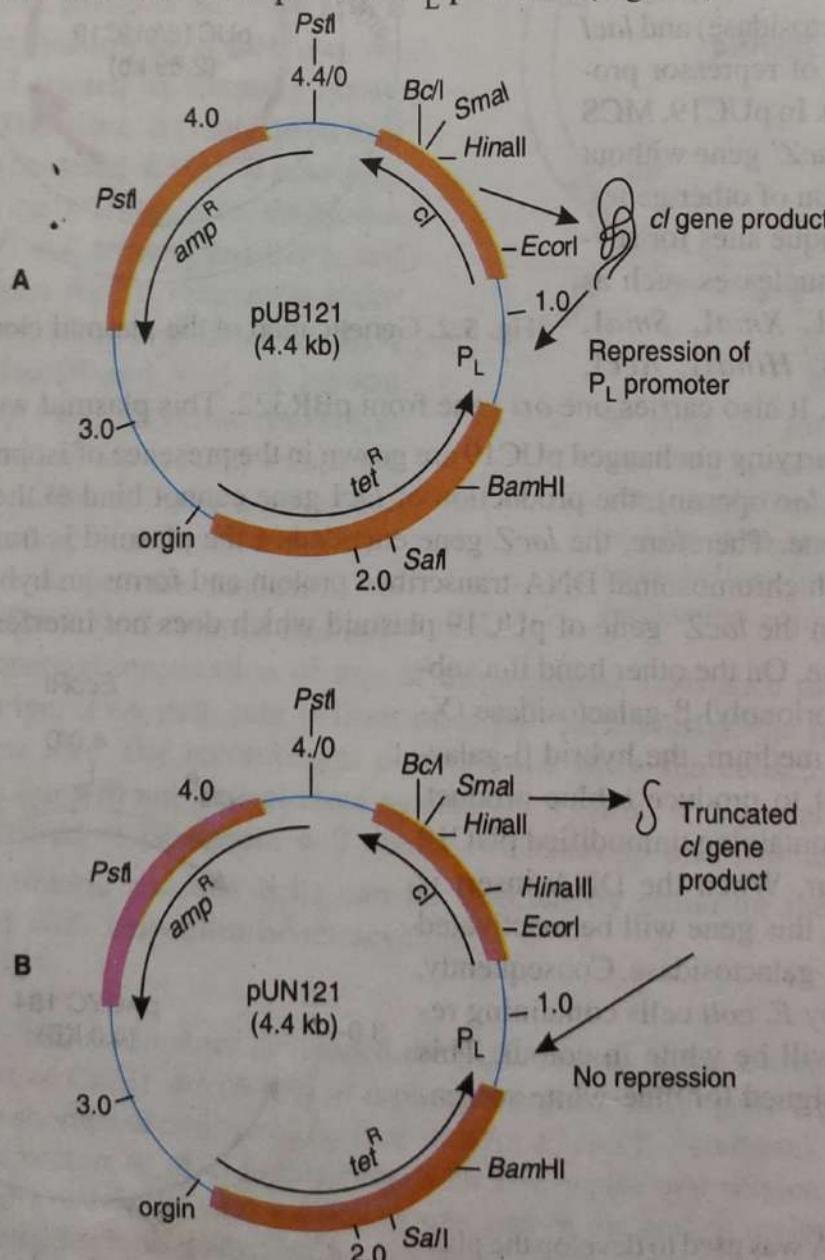
**Fig. 5.3.** Physical map of the pACYC184 vector.

and  $tet^R$  (for tetracycline resistance) genes. The plasmid containing cloned DNA fragments can be detected by insertional inactivation of one of the antibiotic resistance genes. Its number can be amplified unlike pBR322 as the latter is not self-transmissible (Fig. 5.3).

#### 4. pUN121

In 1983, Nilsson and co-workers developed a positive selection vector, pUN121 (4.4 kb) deriving from pBR322 for rapid selection of recombinant plasmid-containing bacteria. This plasmid is used at the time of construction of genomic library. It consists of  $amp^R$  gene  $P_L$  promoter of phage  $\lambda$  (exchanging with original promoter region of  $tet^R$ ),  $cI$  gene of phage  $\lambda$  cloning at pUN121. Protein expressed by  $cI$  gene suppresses the expression of the  $P_L$  promoter. Therefore, expression of  $tet^R$  gene is repressed. Hence, bacterial cells consisting of pUN121 will show resistance against ampicillin but not tetracycline.

The  $cI$  gene consists of restriction sites for *Eco*RI, *Hind*III, *Bcl*II and *Sma*I. If a foreign DNA is cloned at any sites of  $cI$  gene, insertional inactivation of  $cI$  gene takes place. Hence, suppression of  $P_L$  promoter is abolished and the cells become resistant to tetracycline. If such cells are grown on medium containing tetracycline, these will form colonies. Because in these cells  $cI$  gene produces truncated repressor protein which cannot repress the  $P_L$  promoter (Fig. 5.4).



**Fig. 5.4.** The pUN121 vector without insertion of foreign DNA (A). The pUN121 in which a foreign DNA is cloned in  $cI$  gene (B).

### 5. Agrobacterium-based Plasmid Vectors

Several groups of plants have received attention as a means for DNA cloning and expression of foreign DNA in plant cells. The DNA viruses of plants are of two types: (i) double stranded DNA viruses, such as caulimoviruses, and (ii) single stranded DNA virus, for example, geminivirus. Although plant viruses have major scope for the development as cloning vector, yet this aspect has not been studied so far. A good deal of work is done on soil-borne bacteria, *Agrobacterium tumefaciens* causing crown gall disease and *A. rhizogenes* causing hairy root disease on the stems of numerous plants.

**(a) Ti-Plasmid.** *A. tumefaciens* contains a large plasmid which induces tumor in plants, therefore, the plasmid is called as Ti-plasmid. The size of Ti-plasmid ranges between 180–250 Kb. It contains T-DNA region of about 23 to 25 Kb which is transferred into plant cells. Ti-plasmid can be grouped into three on the basis of opine types, for example, octopine, nopaline and agropine. Structural formulae and diagrams of these are given in Figs. 5.5 and 5.6, respectively.

Left side of T-DNA is known as TL-T-DNA and right side of the same as TR-T-DNA. Both the essential (T-DNA) and non-essential regions of Ti-plasmids (both octopine and nopaline types) contain many genes. One or more unusual amino acid derivatives encoded by DNA is known as opines. Opines are neither found normally in plants nor required by plants.



A tumor - inducing plasmid from a common bacterium (*Agrobacterium tumefaciens*) causes crown gall tumors which is used by scientists to work in cultures of plant cells to move desirable genes into the plant chromosomes.

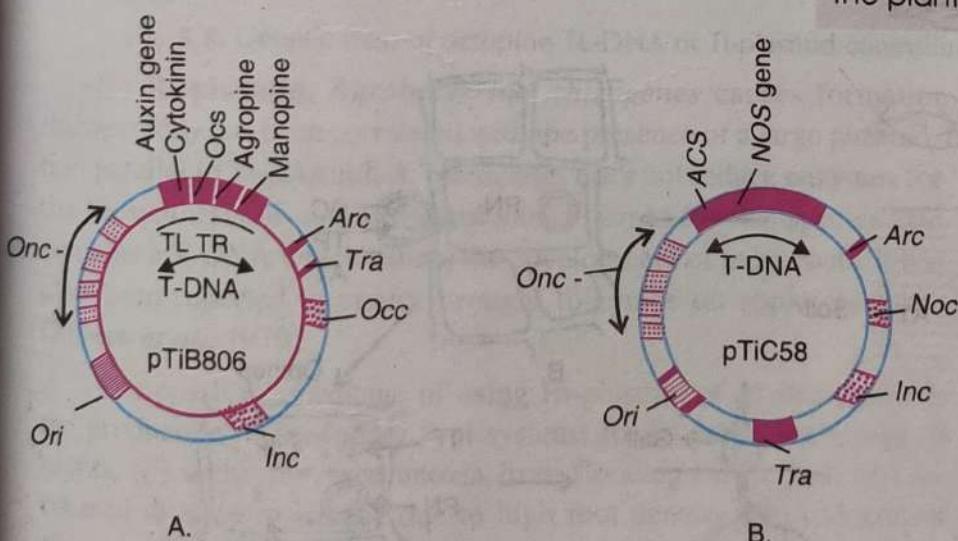
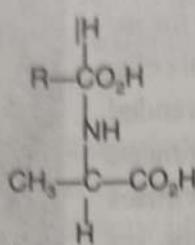


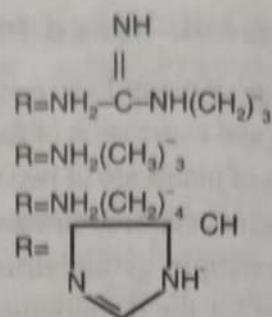
Fig. 5.5. A map of octopine plasmid (A) and nopaline plasmid (B).

*A. tumefaciens* uses it as a source of carbon and nitrogen for their growth and multiplication. The genes responsible for the biosynthesis (*ocs* gene for octopine and *nos* for nopaline) and degradation (*occ* for octopine and *noc* for nopaline catabolism) of opines are located at the well defined regions. *Arc* genes are required for arginin catabolism. The plasmid also carries transfer (*tra*) gene to help the transfer T-DNA from one bacterium to other bacterial or plant cells, *onc* gene for oncogenicity, *ori* gene for origin of replication and *inc* gene for incompatibility.

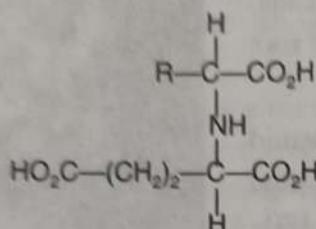
## A. Octopines



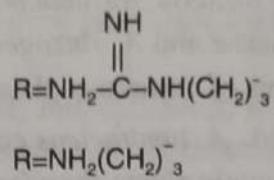
Octopine  
Octopinic acid  
Lysopine  
Histopine



## B. Nopalines

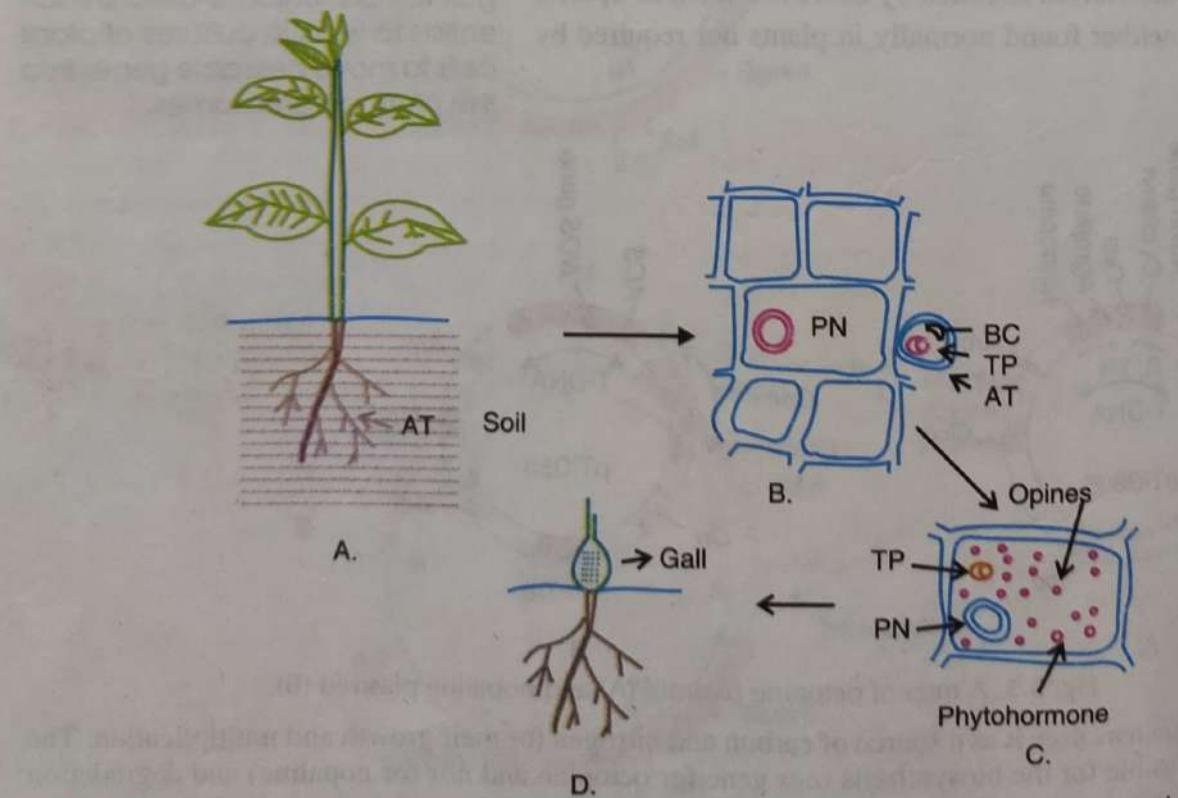


Nopaline  
Nopalinic acid or Ornaline



**Fig. 5.6.** Structural formulae of octopines (A) and nopalines (B).

(ii) **Mechanism of T-DNA transfer.** *A. tumifaciens* lives in soil and attack many dicotyledonous plants (Fig. 5.7 A) most probably at the level of soil surface. Formation of wound is necessary for establishment of bacterial plasmid into plant cells. Lipopolysaccharide, a compound secreted by bacterial cell wall, helps in its attachment with polygalacturonic acid fractions of plant cell wall. From the wounded cell walls of plants, a phenolic compound of low molecular weight (acetosyringone) is secreted which induces the *vir* genes of Ti-plasmids. *Vir* genes encode an enzyme which nick the double stranded T-DNA on the same strand at two points, and produces single stranded DNA molecules. A cut is made on right border of T-DNA and a single stranded T-DNA fragment of 5'→3' direction is generated. It is carried into plant cells (B).



**Fig. 5.7.** Process of infection by *Agrobacterium tumifaciens* (AT) on a dicot plant and induction of tumour or gall (A-D) TP, T-DNA of plasmid, PN, plant nucleus; BC, bacterial chromosome.

T-DNA of Ti-plasmid is stably integrated with plant DNA. The sequencing of nucleic acid at the junction between plant DNA and T-DNA has been done. T-region in both octopine and nopaline plasmids is flanked by a direct repeat of 25 base pairs. *Ops* gene encodes enzymes for the synthesis of opines in transformed cells which are required for proliferation of infecting bacteria. TL-T-DNA encodes two enzymes that are involved in biosynthesis of two phytohormones, auxin and cytokinin (C) (Ream, 1989). This results in disorganized proliferation of cells commonly known as callus, gall or tumour (D). Therefore, the galls are colonized by the bacteria.

**(ii) Tumour morphology.** Several gene loci for controlling the morphology of tumour have been identified on TL-DNA of two octopine plasmids. These loci are grouped into three categories: *tml* (which causes larger tumour), *tmr* (which induce tumour with large number of roots) and *tms* (which causes tumour with large number of shoots) (Fig. 5.8). These effects are due to the products of loci of TL-DNA on the cytokinin/auxin ratios within the transformed tissues. The function of *tml* is not known so far.

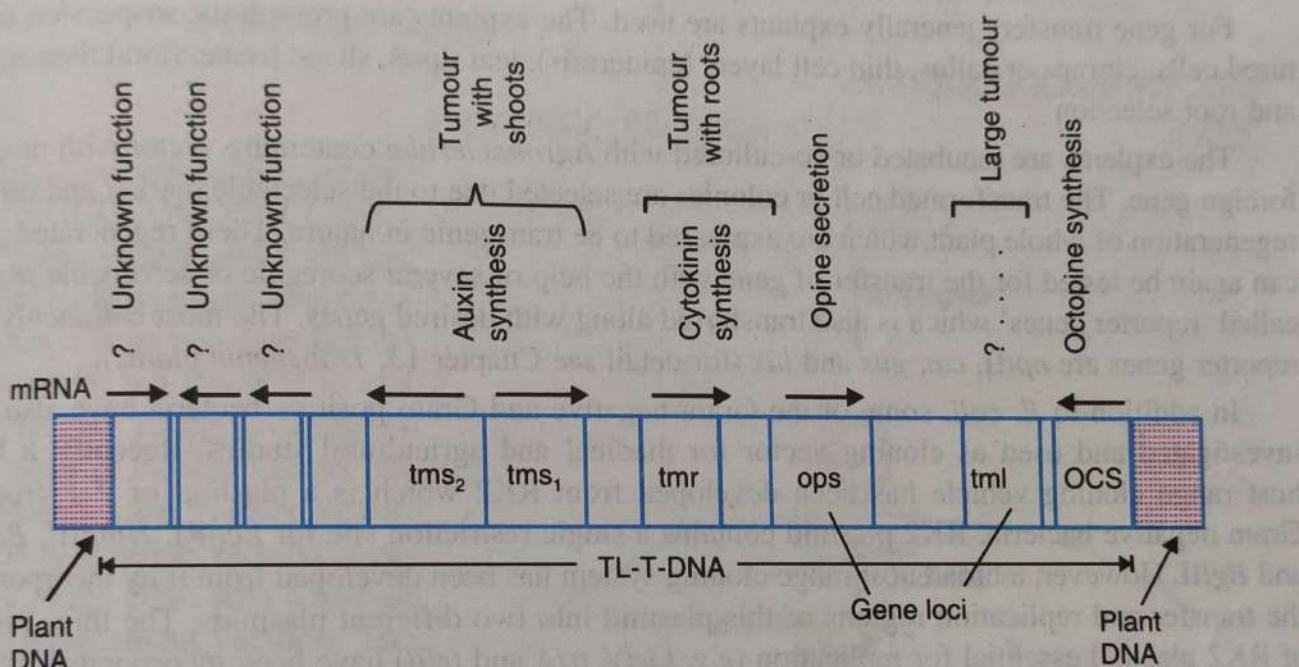
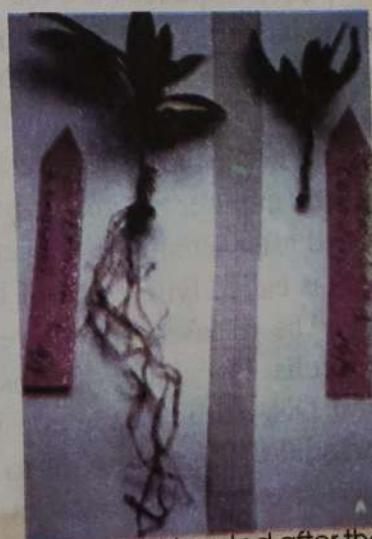


Fig. 5.8. Genetic map of octopine TL-DNA of Ti-plasmid controlling tumour morphology.

**(b) Ri-plasmid.** *Agrobacterium rhizogenes* causes formation of adventitious roots. This rhizogenicity has been correlated with the presence of a large plasmid, the Ri-plasmid, for root induction parallel to Ti-plasmid. *A. rhizogenes* does not induce enzymes for the biosynthesis of auxin or cytokinin. Plasmid for this species also contains a T-DNA which causes the development of hairy roots. It has also been reported to confer drought tolerance on apple seedlings (Moore *et al.*, 1979).

The possible advantage of using Ri-plasmid of *A. rhizogenes* is the production of secondary root systems for (i) better anchorage of plants, (ii) ability to resist anoxia from flooding of the soil, (iii) increased drought resistance due to high root density, and (iv) greater chance of interaction with soil-borne mycorrhizal fungi (Smith and Walker, 1981). It may be applied to establish the secondary root systems in certain plants (fruit trees) by inoculation of soil as well as beneficial for those plants which are propagated through cutting and rooting as stimulants for root formation on cuttings.



Almond plant rooted after the infection with *A. rhizogenes*.

(c) **Properties of Ti-and Ri based plasmids.** *Agrobacterium* plasmids have been used for introduction of genes of desirable traits into plants. The properties of Ti- and Ri based plasmids are as below:

- (i) They do not cause disease because of being disarmed by removing disease causing gene.
- (ii) They possess sites for insertion of foreign gene which needs to be introduced.
- (iii) They possess selectable markers i.e., genes which help in selecting the transformed cells.

The selectable marker is a powerful promoter e.g. 35S (for high level of expression of marker) derived from cauliflower mosaic virus (CaMV) and polyadenylation signal (for adding poly A to mRNA). The most commonly used selectable marker is neomycin phosphotransferase (*nptII*) that confers resistance to neomycin and allows the cells to grow in medium containing neomycin. Thus, after inserting a foreign gene, a vector is designed. The vector is transferred to *A. tumifaciens* or *A. rhizogenes* depending on cases. The transformed bacterium is used to infect the designed host cell where the transformation of gene is to be brought about.

For gene transfer, generally explants are used. The explants are protoplasts, suspension of cultured cells, clumps of callus, thin cell layers (epidermis), leaf discs, sliced tissue, floral tissues, stem and root selection.

The explants are incubated or co-cultured with *Agrobacterium* containing vector with modified foreign gene. The transformed cell or colonies are selected due to the selectable marker and used for regeneration of whole plant which are expressed to be transgenic in nature. These regenerated plants can again be tested for the transfer of gene with the help of several scoreable or screenable markers called 'reporter genes' which is also transferred along with desired genes. The most commonly used reporter genes are *nptII*, *cat*, *gus* and *lux* (for detail see Chapter 13, *Transgenic plants*).

In addition to *E. coli*, some of the Gram-negative and Gram positive-bacteria have also been investigated and used as cloning vector for medical and agricultural studies. Recently, a broad host range cloning vehicle has been developed from RK2 which is a plasmid of P-1 group of Gram negative bacteria. RK2 plasmid contains a single restriction site for *EcoRI*, *HindIII*, *BamHI* and *BglII*. However, a broad host range cloning system has been developed from it by incorporating the transfer and replication regions of this plasmid into two different plasmids. The three regions of RK2 plasmid essential for replication (e.g. *OriV*, *trfA* and *trfB*) have been incorporated into the plasmid cloning vector, the pRK290 (Bernard and Helinski, 1980).

## B. BACTERIOPHAGE VECTORS

Bacteriophages are the viruses that infect bacterial cells after injecting their genetic material (DNA or RNA) and kill them. The viral DNA replicates and expresses inside the bacterial cells, and produce a number of phage particles released after bursting the bacterial cells. This is called lytic cycle of bacteriophage. The released phages re-infect the live cells. The ability of transferring the viral DNA from phage capsid specific bacterial cell gave insight to the scientists to exploit bacteriophages and design them as cloning vectors. The two bacteriophages e.g. phage  $\lambda$  and M13 have been modified, extensively studied and commonly used as cloning vectors.

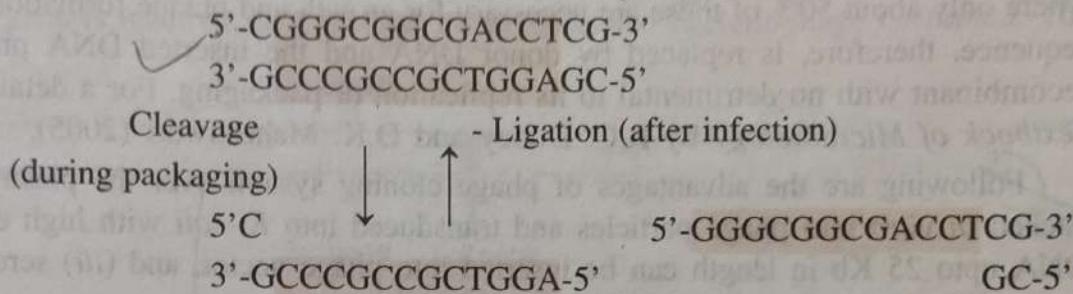


Bacteriophage Lambda Particles.

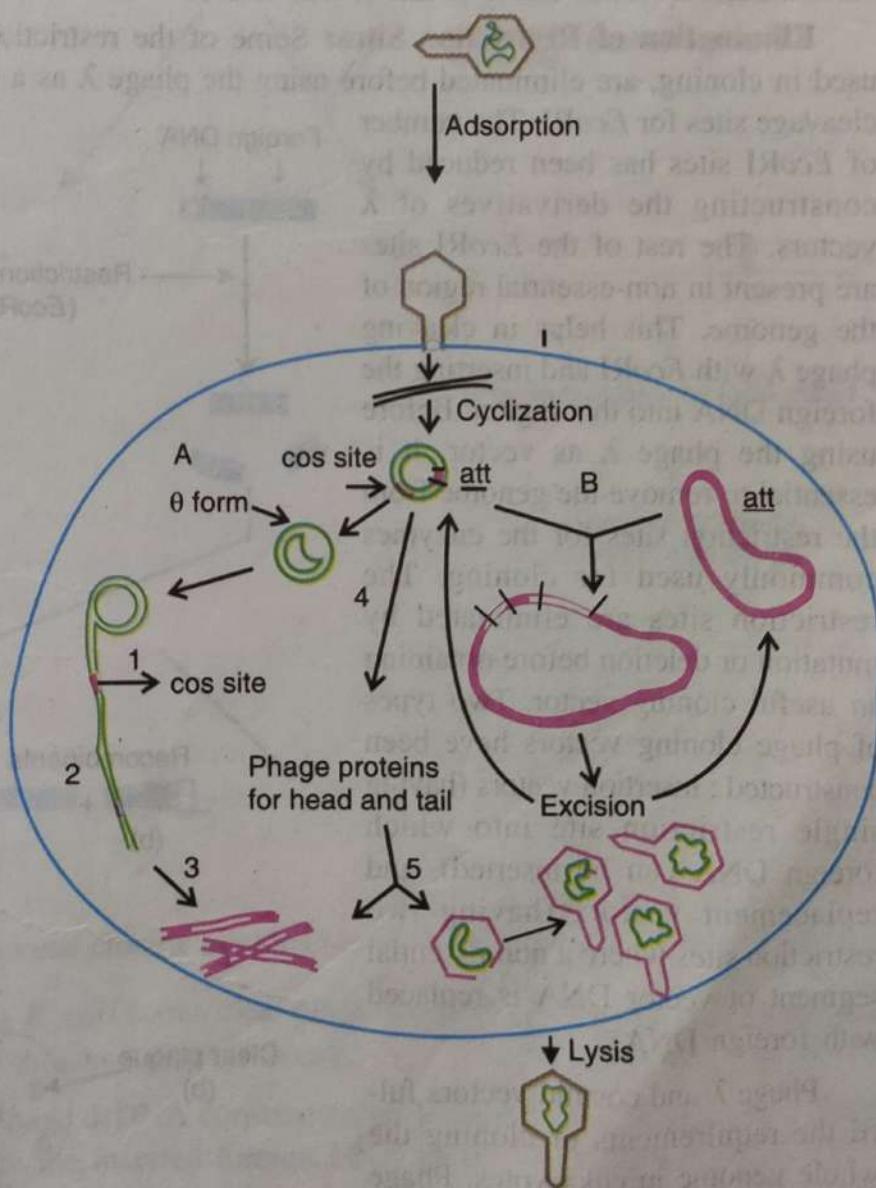
## 1. Phage $\lambda$

Unlike plasmid vectors, the phage vectors are required for cloning of large DNA fragment and, therefore, the gene bank or genomic libraries can be constructed. They are an alternative to bacterial plasmids, and related cosmids. The process of infection and replication of phage in *E.coli* cell are shown in (Fig. 5.9.)

Phage  $\lambda$  contains a proteinaceous head and a long tail attached to the head. In the head it possesses 50 genes in its 48.514 kb (kilobase pairs) genome of which about half of genes are essential. On attachment with tail to cell wall of *E. coli* it injects its linear DNA into the cell. The linear double stranded DNA molecule cyclizes through the single strand of 12 nucleotides commonly known as *cos* sited at its ends. The *cos* sites are the key feature of the DNA.



The 12 nucleotide long projections (5'GGGC<sub>n</sub>CGACCT-3') show cohesiveness and form the COS site (cohesive site). Phage genome has a large non-essential region which is not involved in cell lysis (Fig. 5.9). Taking advantage of it two types of cloning vectors can be produced, either by inserting foreign DNA (insetion vectors) or replacing phage DNA with a foreign DNA (replacement vectors). The upper limit of foreign DNA to be packed is about 23 kb. Replication cycle of phage  $\lambda$  is accomplished into two pathways : the lytic and lysogenic pathways (Wu and Taylor, 1971). In the lytic pathway, early in the infection sites the circular DNA replicates as theta ( $\theta$ ) forms. By a rolling circle mechanism it produces the long concatemeric molecules joined end to end, and composed of several linearly arranged genomes. At the same time phage DNA directs the synthesis of many proteins required to produce empty heads where DNA is packed after the cleavage of concatemeric DNA at its *cos* site to yield fragments of such sizes as to fit in their heads.



**Fig. 5.9.** Replication of phage lambda in a *Escherichia coli* cell. A - lytic cycle; 1. rolling circle replication; 2. production of concatemers; 3. cleavage at cos site; 4. transcription and translation; 5. packaging. B - lysogenic cycle.

Eventually, a tail is attached to the head and finally the mature phage particles are released out of the bacterial cell.

The lysogenic pathway for replication is another alternative mode of propagation where it becomes stably integrated into the host chromosome and replicated along with the bacterial chromosome. Phage genome integrates by an attachment site (*att*) with a partially homologous site on the *E. coli* chromosome, where it replicates as a chromosomal DNA segment. In this case a protein is produced by *cI* genes which represses all the genes responsible for lytic pathways. In this pathway no phage structural proteins are synthesized. The interactions of two proteins, the *cI* genes expressed protein (by phage genome) and *cro* gene expressed protein (by *E. coli* chromosomes) decide between the events of the lytic and lysogenic pathways. Phage  $\lambda$  has about 50 genes where only about 50% of these are necessary for growth and plaque formation. Non-essential DNA sequence, therefore, is replaced by donor DNA and the inserted DNA propagated as phage  $\lambda$ , recombinant with no detrimental to its replication or packaging. For a detailed description see *A Textbook of Microbiology* by R.C. Dubey and D.K. Maheshwari (2005).

Following are the advantages of phage cloning system over the plasmids: (i) DNA can be packed *in vitro* into phage particles and transduced into *E. coli* with high efficiency, (ii) foreign DNA upto 25 Kb in length can be inserted into phage vector, and (iii) screening and storage of recombinant DNA is easier (Dahl *et al.*, 1981).

**Elimination of Restriction Sites:** Some of the restriction sites for the enzymes, commonly used in cloning, are eliminated before using the phage  $\lambda$  as a cloning vector. Wild type  $\lambda$  has five cleavage sites for *EcoRI*. The number of *EcoRI* sites has been reduced by constructing the derivatives of  $\lambda$  vectors. The rest of the *EcoRI* sites are present in non-essential region of the genome. This helps in cleaving phage  $\lambda$  with *EcoRI* and inserting the foreign DNA into this region. Before using the phage  $\lambda$  as vector, it is essential to remove the genome from the restriction sites for the enzymes commonly used for cloning. The restriction sites are eliminated by mutation or deletion before obtaining an useful cloning vector. Two types of phage cloning vectors have been constructed : insertion vectors (having single restriction site into which foreign DNA can be inserted), and replacement vectors (having two restriction sites where a non-essential segment of vector DNA is replaced with foreign DNA).

Phage  $\lambda$  and cosmid vectors fulfil the requirements of cloning the whole genome in eukaryotes. Phage  $\lambda$  contains about 20–25 kb long segment and cosmid contains 45 kb long

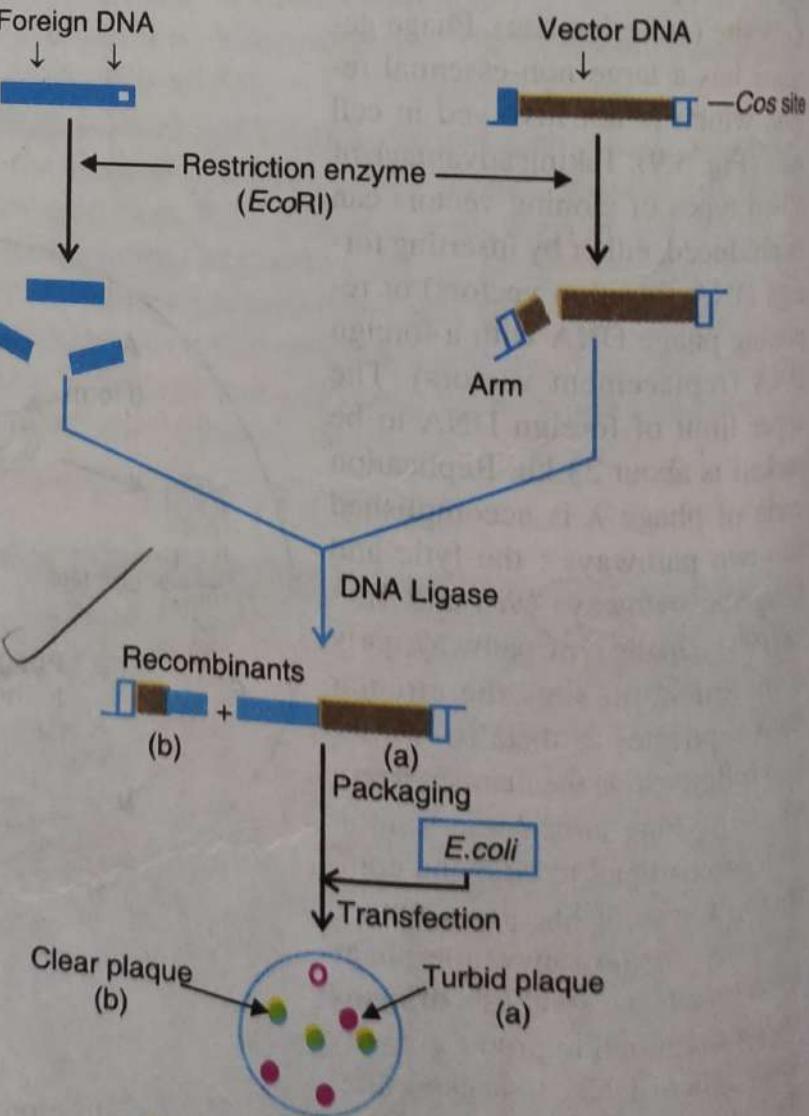
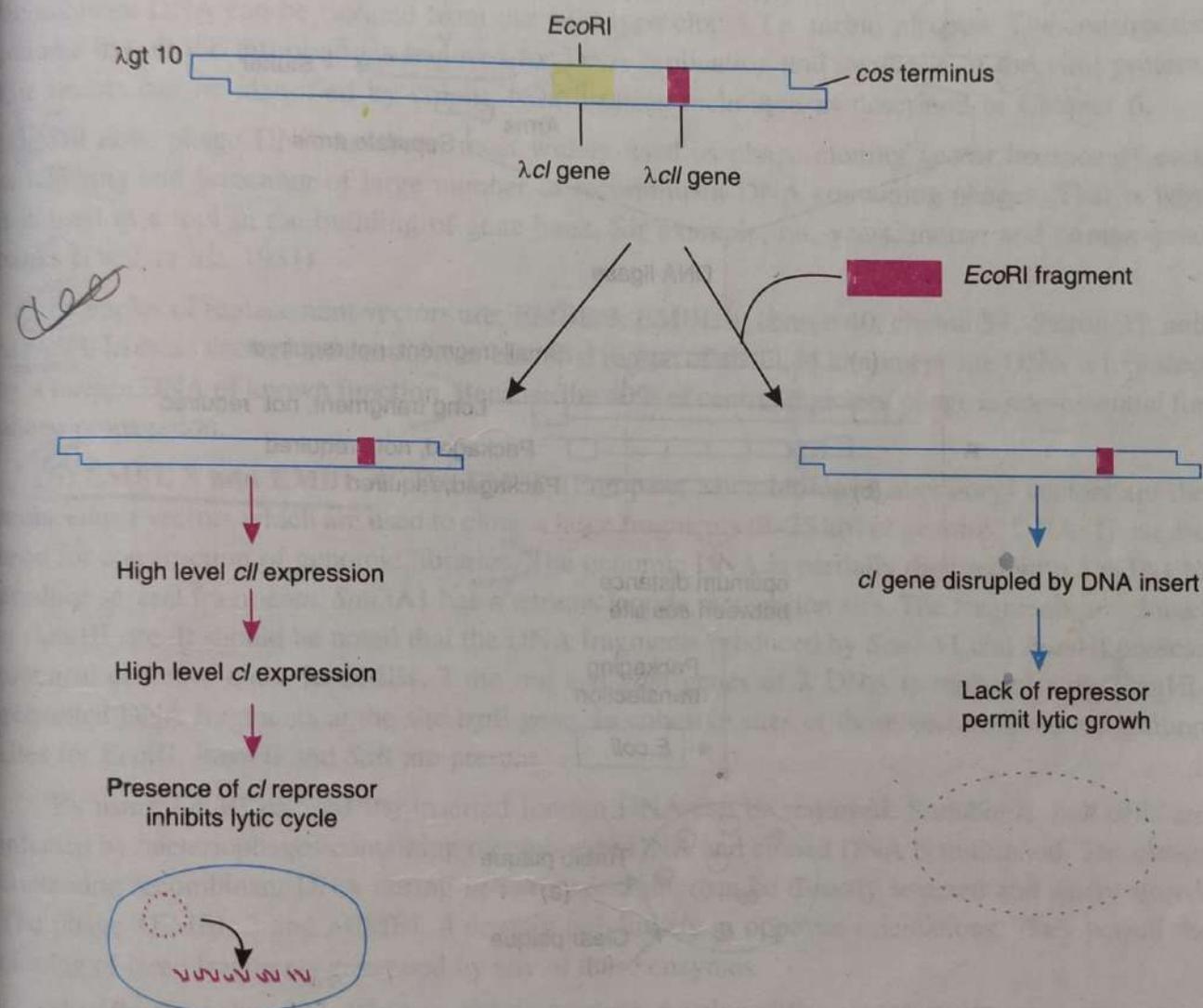


Fig. 5.10. Cloning of an insertion vector.

segment. Several phage  $\lambda$ -based cloning vectors have been constructed, for example  $\lambda$ gt10 and  $\lambda$ gt11, EMBL3 and EMBL 4, Charon, etc.

**(a) Insertion vector :** Insertion vectors have unique cleavage site into which relatively small piece of foreign DNA is inserted. Foreign DNA fragment does not affect the function of phage. The upper and lower limits of size of DNA that may be packed into phage particles is between 35-53 Kb. Therefore, the minimum size of vector must be above 35 Kb. It means the maximum size of a foreign DNA to be inserted is about 18 Kb. Cloning of an insertion vector is shown in Fig. 5.10.

(i)  **$\lambda$ gt10:** Phage  $\lambda$  is modified to construct  $\lambda$ gt10 that may clone cDNA fragment. The  $\lambda$ gt10 is a dsDNA of 43 kb which may clone 7 kb long foreign DNA fragment (Fig. 5.11). After inserting DNA,  $cI^+$  (repressor) gene is inactivated; therefore,  $cI^-$  recombinant bacteriophage is formed. The



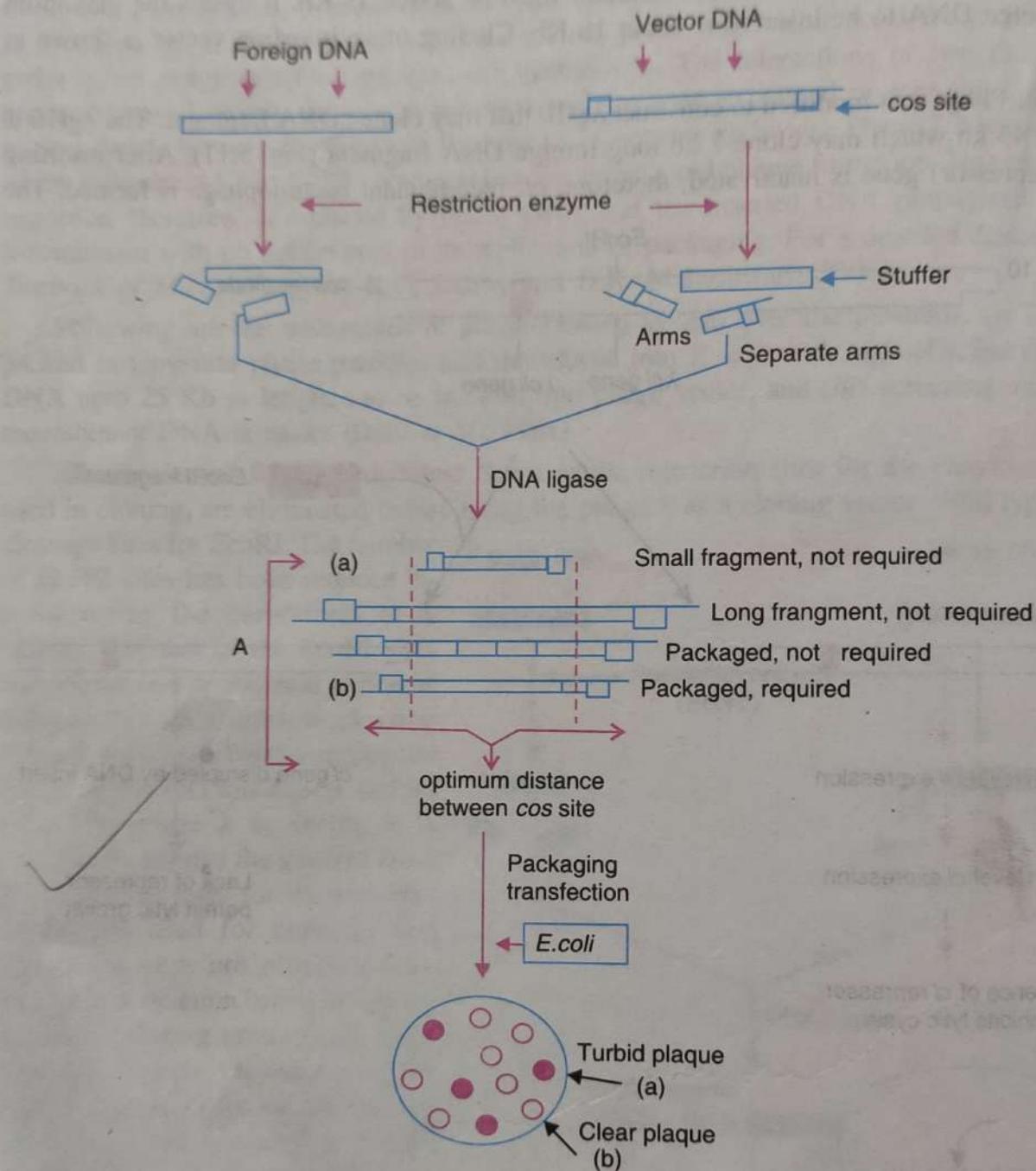
**Fig. 5.11.** Insertional cloning by using phage  $\lambda$ gt10 vector.

recombinant  $cI^-$   $\lambda$ gt10 after infecting *E. coli* forms clear plaques which can easily be screened from cloudy plaques formed by non-recombinant  $\lambda$ gt10  $cI^+$  vector.

(ii)  **$\lambda$ gt11:** It is a 43.7 kb long phage dsDNA constructed for cloning mostly cDNA (less than 6 kb). It is an expression vector where the inserted foreign DNA encodes as  $\beta$ -galactosidase fusion protein. After insertion of a foreign DNA, the recombinant  $\lambda$ gt11 is converted to  $\lambda$ gt11  $gal^-$  and non-recombinant  $\lambda$ gt11 remains  $gal^+$ . Therefore, by using IPTG (inducer) and X-gal (substrate) with suitable *E. coli* host cells, recombinant  $\lambda$ gt11 can be screened. Recombinant  $\lambda$ gt11  $gal^-$  forms white

(clear) colonies while non-recombinant  $\lambda$ gt11 gal<sup>+</sup> forms blue colonies.

**(b) Replacement Vector :** The replacement vectors have cleavage sites present on either side of a length of non-essential DNA of phage. As a result of cleavage left and right arms are formed, each arm has a terminal cos site and longer a stuffer region, the non-essential region, which can be substituted by foreign DNA fragment Fig. 5.12.



**Fig. 5.12.** Cloning of a replacement vector. A-optimization of the suitable sized ligation products for efficient packaging a phage.

The maximum size of inserted DNA fragment depends on how much of the phage DNA is non-essential. It has been found that about 25-30 Kb of genome codes for essential products for lytic cycle. The remaining 20-25 Kb of genome could be replaced with the foreign DNA fragments of known essential products. The substituted vectors are gt, WES and  $\lambda$  1059 (Dahl *et al.*, 1981).

Non-essential part of the  $\lambda$  genome can be separated from the arms by electrophoresis or velocity gradient ultracentrifugation that facilitates to make size differences. Formation of multiple

inserts can be checked by using alkaline phosphatase before ligation with insert fragment. Recombinant DNA formed by multiple inserts has too large genome to be packed in viral head. However, after ligation the recombinant DNA molecules have left arm plus large insert plus right arm linked by their *cos* sites at the arms. Optimum distance from *cos* sites governs efficiency for packaging in *E. coli*. As a result of ligation the size of recombinant DNA fragment may be less or more than the required size or may have more than two *cos* sites or many small fragments of foreign DNA. Those having the range of viral head can be packed *in vitro* using a preparation of head and tail proteins. The viruses thus constructed are allowed to multiply in *E. coli*. Development of plaque turbidity is a useful criteria for the selection of recombinant phages. Plaque turbidity is determined by the presence of nonlysed bacteria. The recombinant phages give clear plaques due to inactivation of *cl* gene. By using these methods the clones containing recombinant DNA can be isolated from the wild type clones *i.e.* turbid plaques. The constructed genome has all the informations required for DNA replication and synthesis of the viral protein. The inserts can be identified by colony hybridization technique as described in Chapter 6.

~~So far~~ Till now, phage DNA has been most widely used as phage cloning vector because of ease in handling and screening of large number of recombinant DNA containing phages. That is why it is used as a tool in the building of gene bank, for example, rat, yeast, mouse and human gene banks (Dahl *et al.*, 1981).

Examples of replacement vectors are: EMBL 3, EMBL 4, charon 40, charon 34, charon 35, and  $\lambda$ -DASH. In these vectors the central non-essential region of about 44 kb long phage DNA is replaced by a foreign DNA of known function. Because the 40% of central region of phage is non-essential for phage propagation.

(i) **EMBL 3 and EMBL 4:** The EMBL (European Microbiology Laboratory) vectors are the replacement vectors which are used to clone a large fragments (9–25 kb) of genomic DNA. These are used for construction of genomic libraries. The genomic DNA is partially digested with *Sau3A*1 to produce several fragments. *Sau3A*1 has a tetranucleotide recognition site. The fragments are cloned at *Bam*HI site. It should be noted that the DNA fragments produced by *Sau3A*1 and *Bam*HI possess identical cohesive ends. In EMBL 3 the *red* and *gam* genes of  $\lambda$  DNA is replaced with *Bam*HI-generated DNA fragments at the site *trpE* gene. In cohesive sites of these vectors three recognition sites for *Eco*RI, *Bam*HI and *Sal*I are present.

By using *Eco*RI or *Sau*I the inserted foreign DNA can be retrieved. Suitable *E. coli* cells are infected by bacteriophages containing recombinant DNA and cloned DNA is multiplied. The clones containing recombinant DNA during *in vitro* packaging can be directly selected and easily stored. The phage  $\lambda$ EMBL 3 and  $\lambda$ EMBL 4 contain polylinkers in opposite orientations. They permit the cloning of large fragments generated by any of these enzymes.

(ii) **Charons:** In 1977, Blattner and co-workers developed the charon series insertion vectors. Charon 7 consists of *imm434* and serves as an insertion vector for both *Hind*III and *Eco*RI (Fig. 5.13). Whereas charon 12 contains *lacZ* gene. Insertion of DNA into the single *Eco*R1 will inactivate the *lacZ* gene of charon 12. The active enzyme results in formation of blue colonies or plaque, whereas after inactivation of *lacZ* colourless colonies are formed (Fig. 5.13).

There are replacement vectors which have been described to clone large fragment of DNA molecules. Its stuffer consists of DNA molecule polymerised head to tail. It can be cleaved by a single restriction enzyme *Nae*. Through *in vivo* recombination charon 40 can be used for screening the recombinant genomic DNA library of eukaryotic DNA. The other charon series vectors are charon 4A, charon 34 and charon 35.

inserts can be checked by using alkaline phosphatase before ligation with insert fragment. Recombinant DNA formed by multiple inserts has too large genome to be packed in viral head. However, after ligation the recombinant DNA molecules have left arm plus large insert plus right arm linked by their *cos* sites at the arms. Optimum distance from *cos* sites governs efficiency for packaging in *E. coli*. As a result of ligation the size of recombinant DNA fragment may be less or more than the required size or may have more than two *cos* sites or many small fragments of foreign DNA. Those having the range of viral head can be packed *in vitro* using a preparation of head and tail proteins. The viruses thus constructed are allowed to multiply in *E. coli*. Development of plaque turbidity is a useful criteria for the selection of recombinant phages. Plaque turbidity is determined by the presence of nonlysed bacteria. The recombinant phages give clear plaques due to inactivation of *cl* gene. By using these methods the clones containing recombinant DNA can be isolated from the wild type clones *i.e.* turbid plaques. The constructed genome has all the informations required for DNA replication and synthesis of the viral protein. The inserts can be identified by colony hybridization technique as described in Chapter 6.

~~As~~ Till now, phage DNA has been most widely used as phage cloning vector because of ease in handling and screening of large number of recombinant DNA containing phages. That is why it is used as a tool in the building of gene bank, for example, rat, yeast, mouse and human gene banks (Dahl *et al.*, 1981).

Examples of replacement vectors are: EMBL 3, EMBL 4, charon 40, charon 34, charon 35, and  $\lambda$ DASH. In these vectors the central non-essential region of about 44 kb long phage DNA is replaced by a foreign DNA of known function. Because the 40% of central region of phage is non-essential for phage propagation.

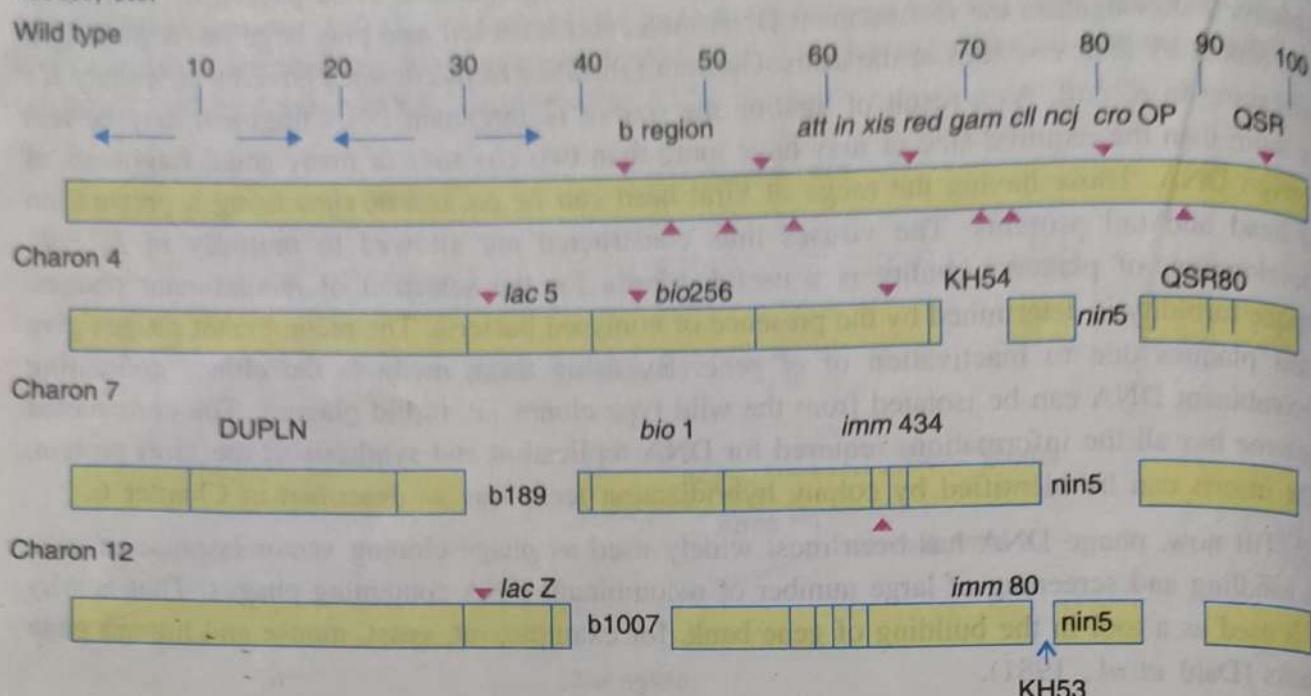
(i) **EMBL 3 and EMBL 4:** The EMBL (European Microbiology Laboratory) vectors are the replacement vectors which are used to clone a large fragments (9–25 kb) of genomic DNA. These are used for construction of genomic libraries. The genomic DNA is partially digested with *Sau3A*1 to produce several fragments. *Sau3A*1 has a tetranucleotide recognition site. The fragments are cloned at *Bam*HI site. It should be noted that the DNA fragments produced by *Sau3A*1 and *Bam*HI possess identical cohesive ends. In EMBL 3 the *red* and *gam* genes of  $\lambda$  DNA is replaced with *Bam*HI-generated DNA fragments at the site *trpE* gene. In cohesive sites of these vectors three recognition sites for *Eco*RI, *Bam*HI and *Sall* are present.

By using *Eco*RI or *Sau*I the inserted foreign DNA can be retrieved. Suitable *E. coli* cells are infected by bacteriophages containing recombinant DNA and cloned DNA is multiplied. The clones containing recombinant DNA during *in vitro* packaging can be directly selected and easily stored. The phage  $\lambda$ EMBL 3 and  $\lambda$ EMBL 4 contain polylinkers in opposite orientations. They permit the cloning of large fragments generated by any of these enzymes.

(ii) **Charons:** In 1977, Blattner and co-workers developed the charon series insertion vectors. Charon 7 consists of *imm434* and serves as an insertion vector for both *Hind*III and *Eco*RI (Fig. 5.13). Whereas charon 12 contains *lacZ* gene. Insertion of DNA into the single *Eco*R1 will inactivate the *lacZ* gene of charon 12. The active enzyme results in formation of blue colonies or plaque, whereas after inactivation of *lacZ* colourless colonies are formed (Fig. 5.13).

There are replacement vectors which have been described to clone large fragment of DNA molecules. Its stuffer consists of DNA molecule polymerised head to tail. It can be cleaved by a single restriction enzyme *Nae*. Through *in vivo* recombination charon 40 can be used for screening the recombinant genomic DNA library of eukaryotic DNA. The other charon series vectors are charon 4A, charon 34 and charon 35.

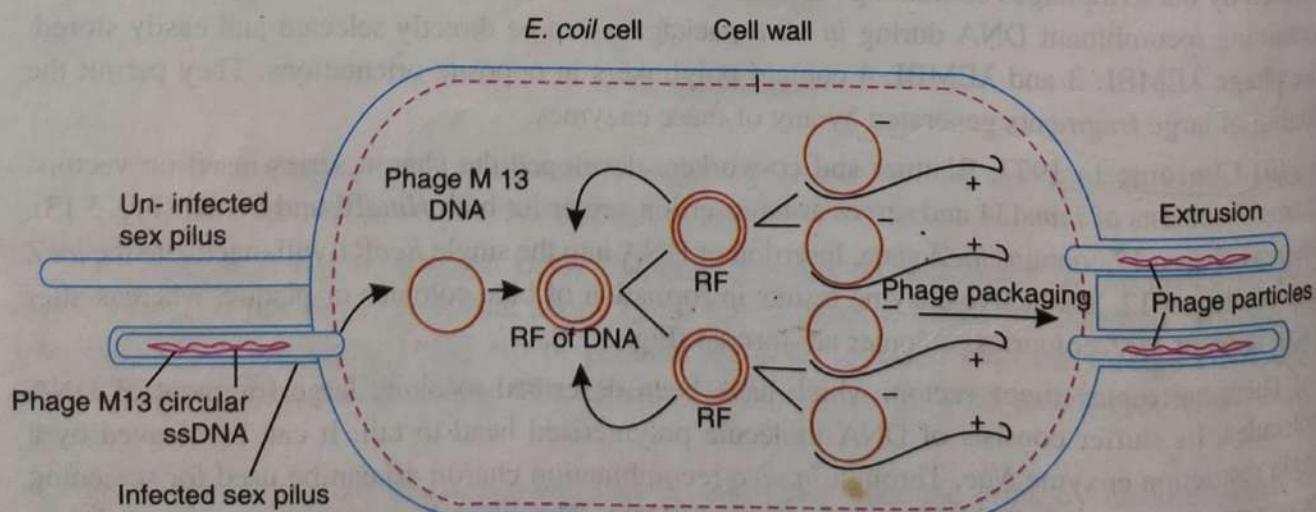
Some of the examples of insertion vectors are phage  $\lambda$ gt10 and  $\lambda$ gt11 vectors, charon series vectors, etc.



**Fig. 5.13.** A physical map of wild type phage  $\lambda$  and some other commonly used vectors, charons EcoRI sites (five dark triangles) are present above the map and HindIII (six dark triangles) are present below the map.

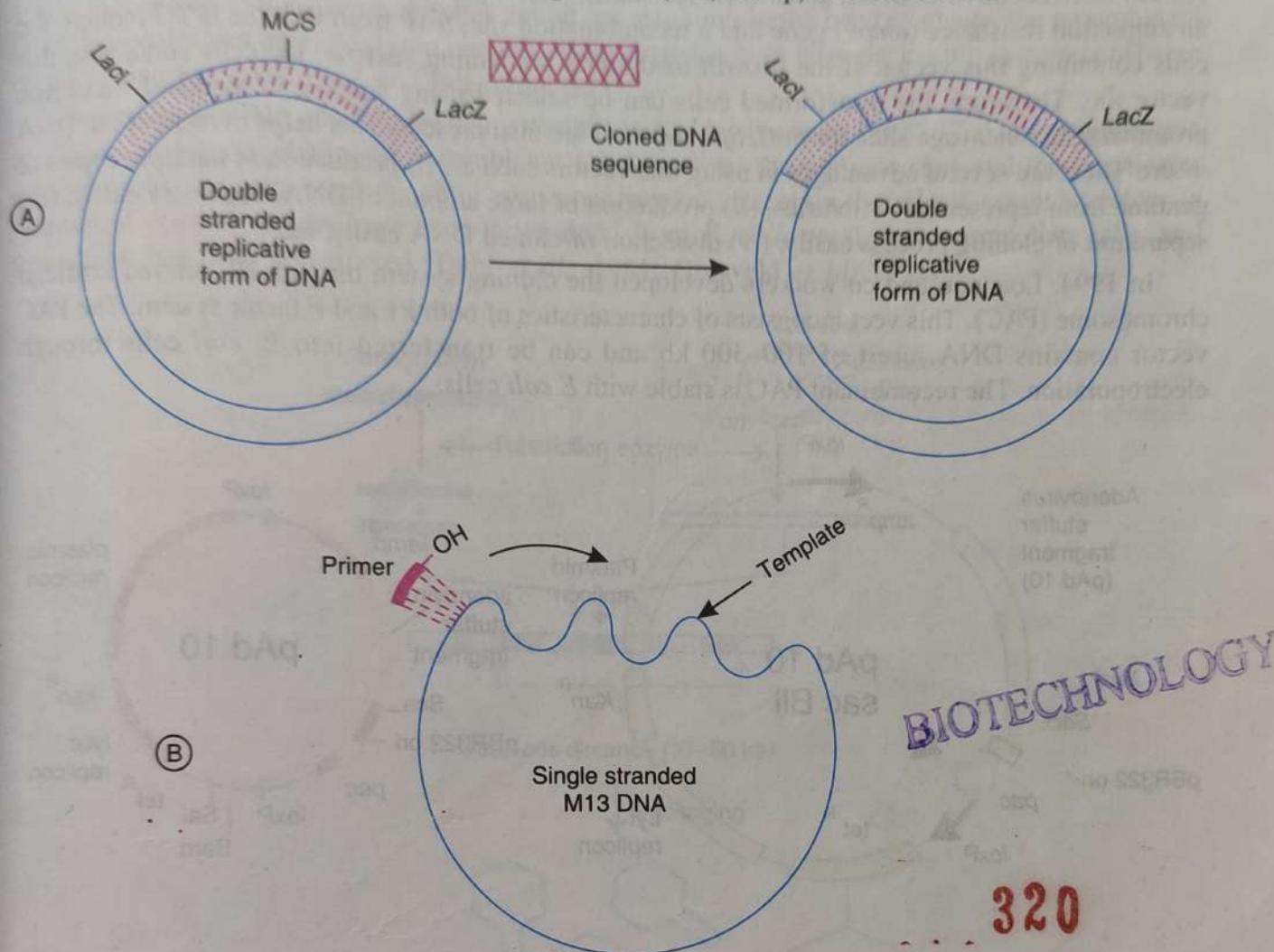
## 2. Phage M13

M13 is a filamentous phage of *E. coli* which infects only such cells which contain sex pili. It is closely related to the other coliphages such as fd and f1. M13 consists of single stranded circular DNA molecule containing 6,407 base pairs. It lands on sex pilus, binds to receptors specified by F plasmid, transfects and delivers its DNA into the cell through the lumen of pilus (Fig. 5.14). After delivery, the DNA is converted from single stranded molecule to double stranded molecule. It is called *replicative form* (RF) of DNA. The RF of M13 replicates and forms 50-100 RF molecules per cell. Finally, single stranded DNA molecules are produced from RF DNA. Phage M13 encodes single strand binding proteins that bind to the viral single stranded DNA which are synthesised by rolling circle mechanism. Finally, the progeny molecules extrude from the cell as M13 particles. The progeny molecules are packed upon extrusion through the cell wall. The cells are not killed but grow slowly.



**Fig. 5.14.** Multiplication of M13 phage in *E. coli* cells.

The M13 system has two advantages: (i) the double stranded RF DNA can be isolated and used as plasmid (Fig. 5.15 A), and (ii) the single stranded M13 DNA can be used as a template for sequencing DNA (Fig. 5.15 B). Foreign DNA of about 500 bp can be cloned into a multiple cloning sequence that forms part of a cloned modified *lacZ* gene on the double stranded RF of M13. Different strategies are followed when large sized foreign DNA (< 2,000 bp) is to be cloned.



**Fig. 5.15.** M13 as cloning and sequencing vectors. A- M13 double stranded DNA as cloning vector; B- M13 single stranded DNA used for sequencing.

DNA from the ligation reaction is mixed with the competent *E. coli* cells. Then the cells are plated on nutrient medium containing X-gal as substrate and the plates are incubated for 24 hours. Thereafter, both white (colourless) and blue colonies (M13 plaques) develop on the medium containing X-gal. X-gal turns blue which is hydrolysed by hybrid  $\beta$ -galactosidase where *lacZ* gene is not disrupted i.e. *lacZ* gene remains functional and lacks any inserted DNA. On the other hand the colourless colonies consist of inserted DNA that has disrupted *lacZ* gene and did not produce  $\beta$ -galactosidase. Therefore, X-gal could not be disintegrated.

The M13 particles are isolated from the white colonies. It is used as a source of M13 single stranded DNA which is further employed for DNA sequencing (Fig. 5.15 B). A primer oligonucleotide is added to the M13 single stranded cloned DNA construct that hybridizes to the vector DNA near the insertion site of the cloned DNA. The dideoxynucleotide reaction is carried out; the sequence of the insert is read from the autoradiographs.

### 3. P1 Phage Vector

P1 vector is used for cloning large DNA segment (about 1000 kb) producing  $10^5$  clones per  $\mu\text{g}$  of DNA insert. On the basis of cloning efficiency and capability, it is put between YAC and cosmids. By using 5–10 ml of log phase culture of *E. coli* cells, the vector with DNA insert can be amplified in *E.*

*coli*. By doing so, several micrograms of cloned DNA is obtained from such amount of *E. coli* cells. Fig. 5.16 shows two important and most commonly used vectors: pNS582tet14ad10 (pAd10) and pAd10sacBII.

The pAd10sacBII vector consists of ColE1 replicon, P *Sac* site, *Bam*H-I-*Scal* stuffer fragment (11 kb) from adenovirus DNA, kanamycin resistance (*kan*<sup>R</sup>) gene isolated from a transposon Tn903, an ampicillin resistance (*amp*<sup>R</sup>) gene and a recombination site, *loxP* from P1. The IPTG induces the cells containing this vector. If the growth medium is containing sucrose, the cells containing this vector die. Therefore, the transformed cells can be selected using sucrose in media. T7 and Sp6 promoters, rare cleavage sites for *Sall*, *Sfi*I and *No*I are also present. This helps to recover the DNA insert. There are several advantages in using this vectors such as: (i) production of multiple copies of genome from representative libraries, (ii) production of large amount of DNA clone very easily, (iii) separation of cloning process easily, (iv) dissection of cloned DNA easily, etc.

In 1994, Loannou and co-workers developed the cloning system through P1-derived artificial chromosome (PAC). This vector consists of characteristics of both P1 and F factor system. The PAC vector contains DNA insert of 100–300 kb and can be transferred into *E. coli* cells through electroporation. The recombinant PAC is stable with *E. coli* cells.

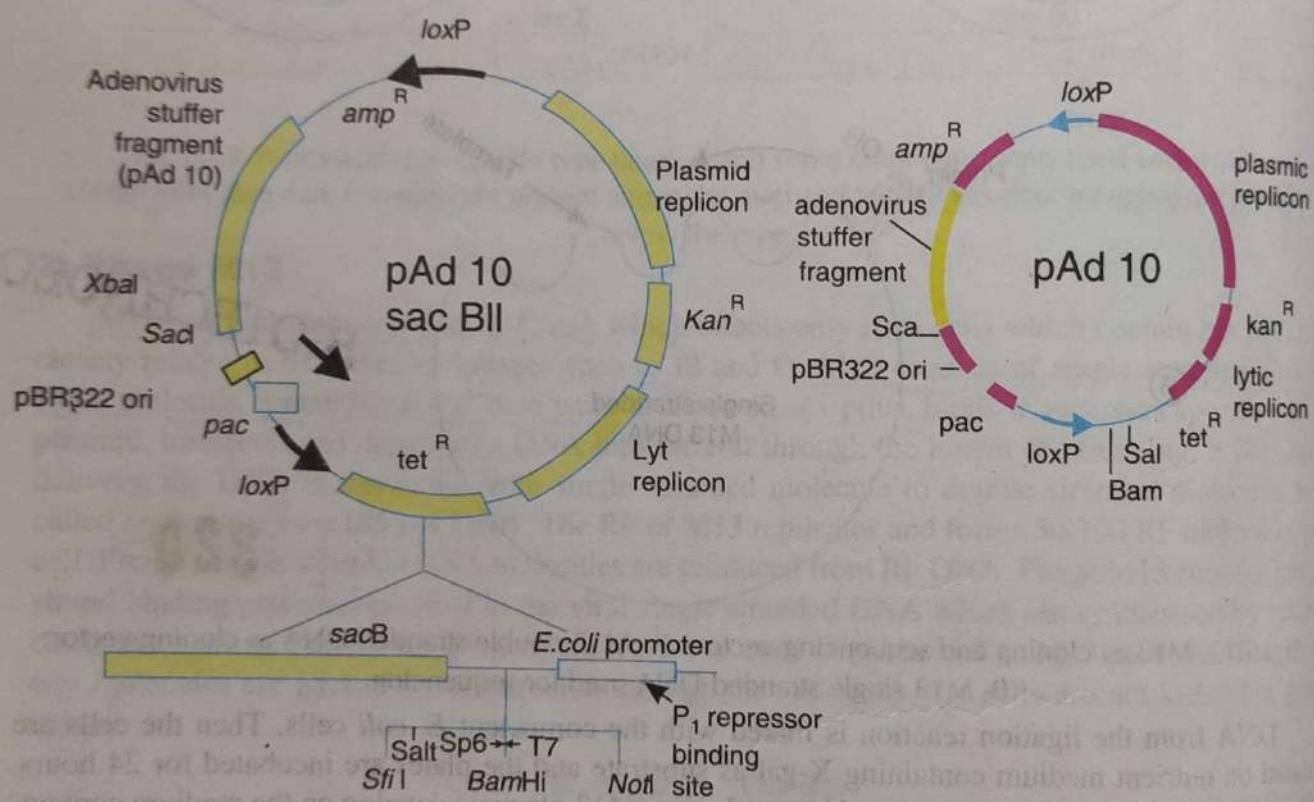


Fig. 5.16. P1 phage cloning vectors; A- pAd10sacBII vector; B- pAd10 vector.

### C. COSMIDS

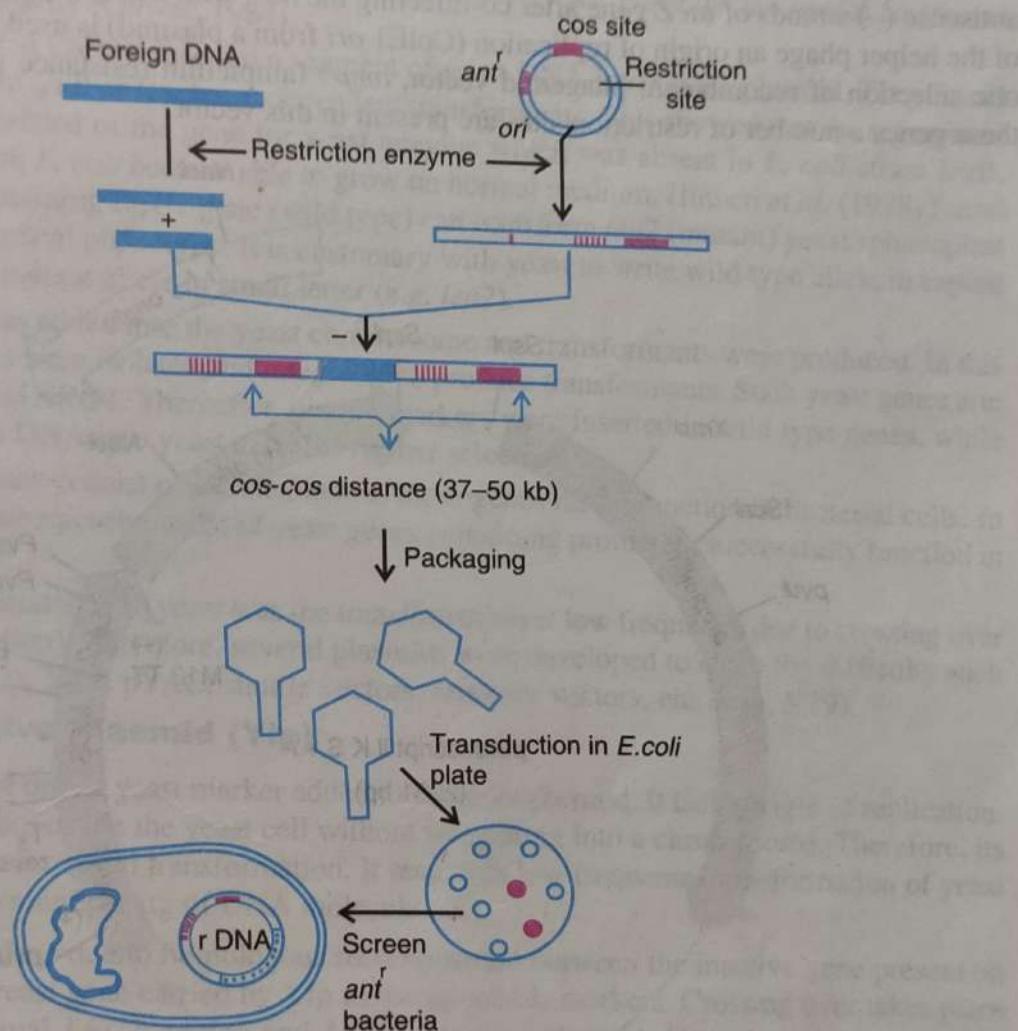
Based on the properties of DNA and *Col* E1 plasmid DNA, a group of Japanese workers (Fukumaki *et al.*, 1976) showed that the presence of a small segment of phage  $\lambda$  DNA containing cohesive end on the plasmid molecule is a sufficient prerequisite for *in vitro* packaging of this DNA into infectious particles. The cosmids can be defined as the hybrid vectors derived from plasmids which contain *cos* site of phage  $\lambda$  (cosmid = *cos* site + plasmid). For the first time it was developed by Collins and Hohn (1978).

Cosmids lack genes encoding viral proteins; therefore, neither viral particles are formed with the host cell nor cell lysis occurs. Special features of cosmids similar to plasmids are the presence of: (i) origin of replication, (ii) a marker gene coding for antibiotic resistance, (iii) a special cleavage site for the insertion of foreign DNA, and (iv) the small size.

Character dissimilar to plasmid is the presence of extra phage DNA, the *cos* site, which has about 12 bases. It helps the whole genome in circularization and ligation.

The cosmids have a length of about 5 Kb, the upper size limit of the foreign DNA fragment that may be inserted in cosmids and packed into phage particles is, therefore, approximately 45 Kb, much larger than it would be possible to clone in phage  $\lambda$  or plasmid vector (Dahl *et al.*, 1981). According to the size of *cos* sites and upper size limit in the head of phage, the recombinant DNA molecules can be packed into bacteriophage particles in *in vitro* packaging system consisting of packaging enzymes, head and tail proteins.

Procedure of DNA cloning by using cosmid vector is shown in Fig. 5.17. Upon transfection of *E. coli* by bacteriophage, the recombinant DNA cyclizes through *cos* sites and then replicates as a plasmid and expresses the drug resistance marker. Recently, based on cosmid vectors, a number of cosmid vectors have been determined from *E. coli*, yeast, and mammalian cells, and gene bank has been constructed (Dahl *et al.*, 1981; Grosveld *et al.* 1982).



**Fig. 5.17.** Cloning of a cosmid vector. Transduced bacteria contain a cosmid which show resistance to specific antibiotic. Such bacteria can be screened. *Ant<sup>r</sup>*, antibiotic resistance; *ori*, origin of replication.

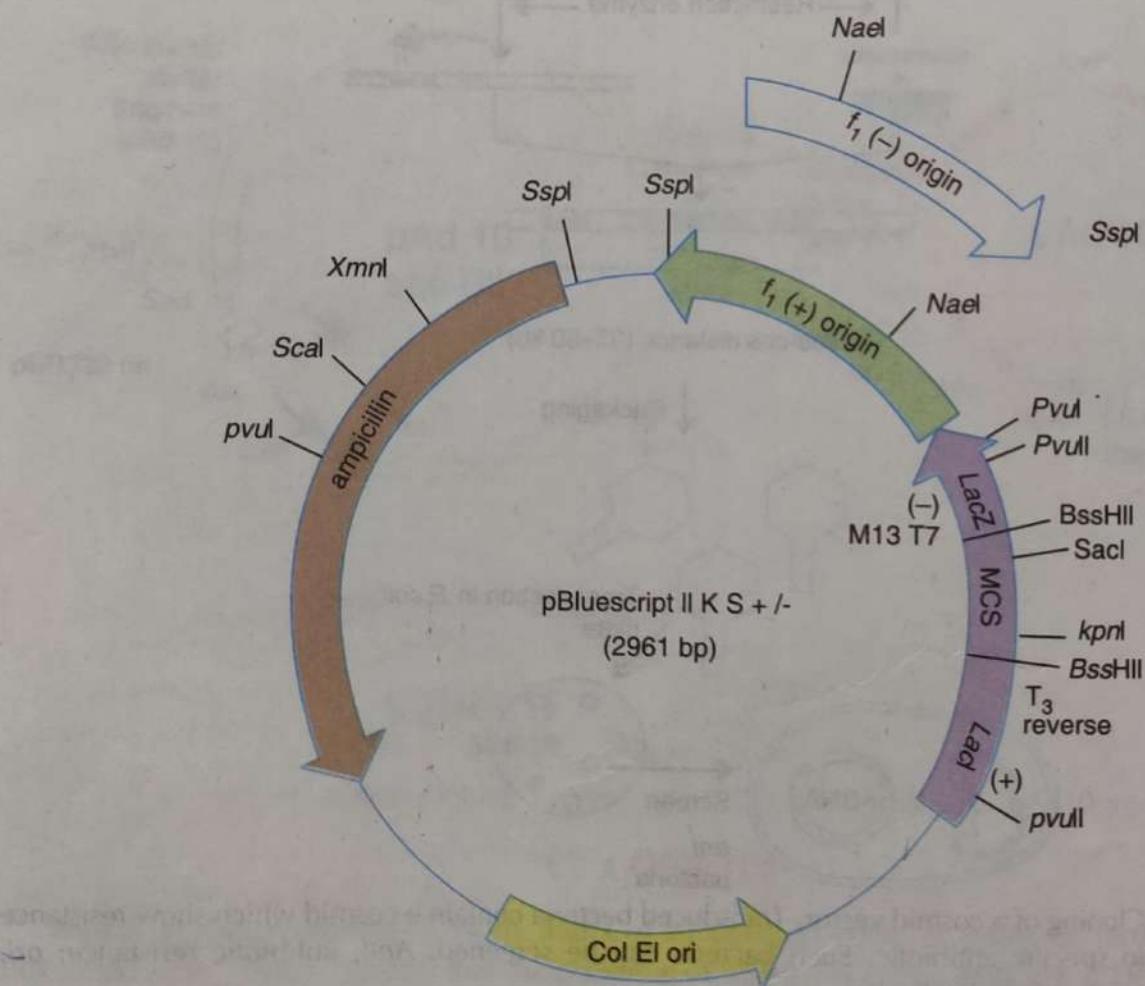
## D. PHAGEMID VECTORS

### 1. Phagemids

Some times generation of a single stranded DNA becomes important for DNA sequencing and site-directed mutagenesis. Using a single stranded DNA containing bacteriophage M13, a series of cloning vectors were developed. The bacteriophage f1 is closely related to M13 which also infects *E. coli*. Phagemids are plasmids that contain origin of replication (*ori*) for single strand DNA containing

bacteriophage such as f1. *E. coli* maintains a plasmid as double stranded DNA due to plasmid *ori* gene. If *E. coli* cells are infected by the helper f1 phage, the *ori* of f1 is activated. It switches to a mode of replication and produces single stranded DNA which is packed into phage particles as they are extruded from the host cell. Fig. 5.18 shows a phagemid pBlueScript II KS (+/-) which is used for generation of single stranded DNA molecules.

**(a) pBlueScript II KS (+/-):** This vector is 2961 bp long derivative from pUC19. The 'KS' denotes the orientation of polylinker *i.e.* transcription of *lacZ* genes proceeds from the *KpnI* restriction site towards *SacI* restriction site. It consists of multiple cloning site (MCS) flanked by *T3* and *T7* promoters in opposite directions on the two strands (Fig. 15.18). The MCS is set up for *lacZ* blue-white screening. A *lacI* promoter complements with *lacZ* of *E. coli* upstream of *lacZ* region. This feature helps in selection of recombinant vector based on the criteria of development of blue-white colonies (white colonies are formed if foreign DNA is inserted in the vector). It consists of f1 (+) and f1 (-) origin of replication obtained from a filamentous phage f1. It is used to get sense (+) and antisense (-) strands of *lacZ* gene after co-infecting the host with a helper phage. But in the absence of the helper phage an origin of replication (ColE1 *ori* from a plasmid) is used. Moreover, for antibiotic selection of recombinant phagemid vector, *amp<sup>R</sup>* (ampicillin resistance gene) is also used. In these genes a number of restriction sites are present in this vector.



**Fig. 5.18.** The phagemid pBlueScript II KS (+/-) vector

## D. YEAST PLASMID VECTORS

Yeast (*Saccharomyces cerevisiae*) is a unicellular eukaryotic microorganism has its long historical importance. It has been used for production of alcohol and alcoholic beverages. It can easily be cultured on artificial media. However, its genetic material is well organised into chromosomes present inside the nucleus.

Yeast cells contain their own plasmid (Begg, 1978) known as '2  $\mu$ m circle plasmid' which is used as vector for foreign genes. It is present in many stains in 50–100 copies per cell. The foreign DNA taken up by yeast is integrated by a specific crossing over to yeast DNA containing a homologous sequence on chromosome recognised by the yeast as the origin of replication (*ori*).

In some cases segments of yeast DNA linked with foreign DNA have been demonstrated to transform very efficiently (about  $10^4$  cells) (Hsiao and Carbon, 1979). The transforming molecules replicate autonomously within the cell. DNA sequences of such property are known as *Ars* (autonomously replicating sequence) fragment. *Ars* fragments have also been isolated from a number of eukaryotic organisms such as *Coenraorabditis elegans*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Zea mays*, etc. (Stinchcomb *et al.*, 1980). However, plasmids have been constructed by using *ars* fragments when joined at centromeric sequences (*cen* region of about 6–10 Kb DNA) to ensure the replication and stability in yeast. *Cen* is needed for DNA molecule to segregate correctly during meiosis and mitosis. Some times the minichromosomes are lost during meiosis when they are attached by chance to spindle apparatus. Thus, *cen* regions of yeast plasmids are useful fragments because they confer stability (Glover, 1984).

In 1977, Ratzkin and Carbon inserted a segment of yeast DNA into ColE1 plasmid. They found some of the *E. coli* strain *leuB* as leucine-independent transformants. This showed that the segment of yeast chromosome consisted of the gene for yeast enzyme which was absent in *E. coli* strain *leuB*. After ligation the mutant *E. coli* became able to grow on normal medium. Hinnen *et al.* (1978) found that ColE1 plasmid containing *LEU2* gene (wild type) can transform *leu2* (mutant) yeast sphaeoplast to form leucine-independent phenotype. It is customary with yeast to write wild type allele in capital letter (e.g. *LEU2*) and mutant allele in small letter (e.g. *leu2*).

The *LEU2* gene was added into the yeast chromosome and transformants were produced. In this way several yeast genes were isolated and integrated to produce transformants. Such yeast genes are: *LEU2*, *URA3*, *TRP1* and *ARG4*. Thereafter, unique markers were inserted in wild type genes, while introducing the foreign DNA into yeast cells for further selection.

The eukaryotic genes consist of introns; hence, these genes rarely function in bacterial cells. In contrast being eukaryotic microbe, most of yeast genes containing promoters successfully function in bacterial cells.

The difficulty associated with yeast was the transformation at low frequency due to crossing over (integrative transformation). Therefore, several plasmids were developed to solve the difficulty such as YIp, YE<sub>p</sub>, YRp, YCp, YLp, pYAC, shuttle vectors, retriever vectors, etc. (Fig. 5.19).

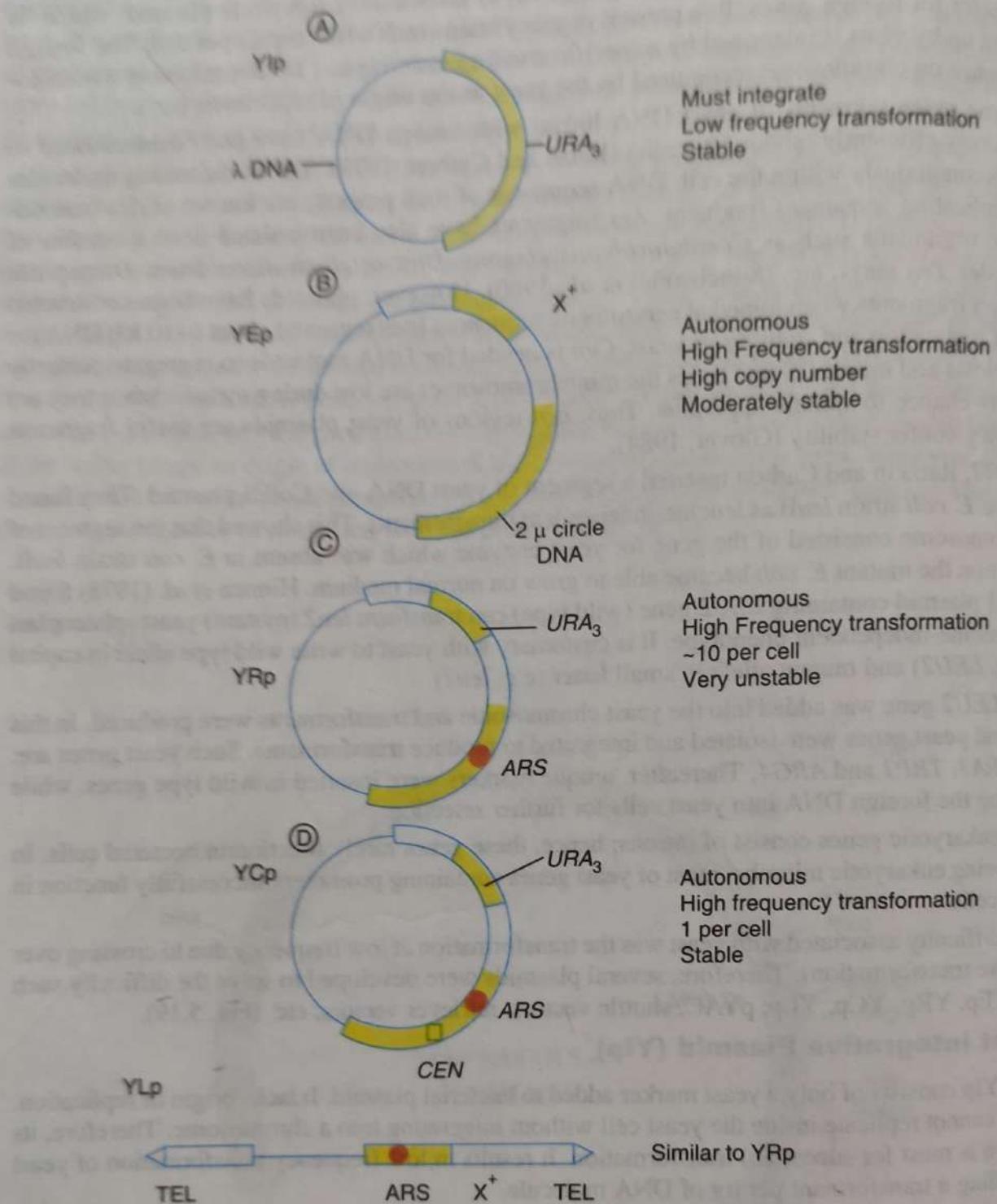
## 1. Yeast Integrative Plasmid (YIp)

The YIp consists of only a yeast marker added to bacterial plasmid. It lacks origin of replication. Hence, it cannot replicate inside the yeast cell without integrating into a chromosome. Therefore, its integration is must for successful transformation. It results in low frequency transformation of yeast cells yielding a transformant per  $\mu$ g of DNA molecule.

Integration takes place due to homologous recombination between the inactive gene present on one chromosome and yeast gene carried by YIp as the selectable markers. Crossing over takes place between the chromosomal *LEU2*<sup>-</sup> region and *LEU2*<sup>-</sup>-carrying plasmid. Consequently the flanking region linked to plasmid is integrated to *LEU2*<sup>-</sup> region. Autonomously replicating system (ARS) is absent in that plasmid; therefore, YIp cannot replicate as a plasmid (Fig. 5.19 A).

## 2. Yeast Episomal Plasmid (YE<sub>p</sub>)

The YE<sub>p</sub> is the 2  $\mu$ m circle-based vector which replicates autonomously without integrating into a yeast chromosome. The 2  $\mu$ m plasmid is 6 kb in size with so high frequency for transformation. The 2  $\mu$ m DNA consists of origin of replication and *rep* genes. This plasmid results in high frequency transformation *i.e.* about  $2 \times 10^4$  transformants per  $\mu$ g DNA. The number of restriction sites is limited. It is found in high copy number in the cells. It will lose its stability if 2  $\mu$  sequence is lost. It is moderately stable in the cell (Fig. 5.19 B).



**Fig. 5.19.** Different types of yeast plasmid cloning vectors.

### 3. Yeast Replicating Plasmid (YRp)

This vector has been developed on the basis of chromosomal elements to support autonomous replication of plasmid in yeast cells. *ARS* was isolated from many region of genomes of yeast and other eukaryotes and *ARS*-based vectors were constructed. *ARS*-based vectors are very unstable but can form high copy number inside the cells (Fig. 5.19 C).

### 4. Yeast Centromeric Plasmid (YCp) or Yeast Cntromere (CEN)

It is also yeast-based plasmid consisting of a sequence of chromosomal centromeric DNA of yeast. It acts as true mini chromosome which becomes circularised. It increases the stability of plasmid and segregates during mitosis and meiosis. Copy number of this plasmid is useful when low copy number of cloned genes are required (Fig. 5.19 D).

### 5. Yeast Linear Plasmid (YLP)

It is similar to ARS centromeric plasmid. It is not circular but linear plasmid constructed by adding telomeric sequences to both ends of a cleaved ARS-based plasmid. Telomere is the end of chromosome. Telomere was isolated from a ciliated protozoan, *Tetrahymena*. This linear plasmid is present in single copy as a chromosome and are very stable and cannot replicate in *E. coli*. Therefore, it has drawn less attention but has raised much possibility to put together the complete set of genes in artificial chromosome of eukaryotes (Fig. 5.19 E).

### E. YEAST ARTIFICIAL CHROMOSOMES (YAC)

Earlier you have studied that the plasmid can accommodate about 10 kb foreign DNA, phage  $\lambda$  can accommodate about 20 kb DNA insert, cosmid can accommodate about 40 kb long four DNA segments. It was realised that for stable replication and segregation in eukaryotes, particularly in yeast, a small and well defined sequence should be constructed as recombinant chromosome. With this objective David and co-workers developed, in 1987, **yeast artificial chromosomes (YAC)** by using new approaches where DNA segment of several thousand base pairs (about 1 Mb) can be cloned. Initially it was used for investigation of the maintenance of chromosomes in the cell. Later on it was used as vectors for carrying very large cloned fragments of DNA. It has also been used for physical mapping of human chromosome in 'Human Genome Project'.

A typical YAC consists of centromere element (*CEN*) for chromosomal segregation during cell division, telomere, and origin of replication (*ori*) were isolated and joined on plasmids constructed in *E. coli*. Structure of typical YAC3 vector is shown in Fig. 5.20.

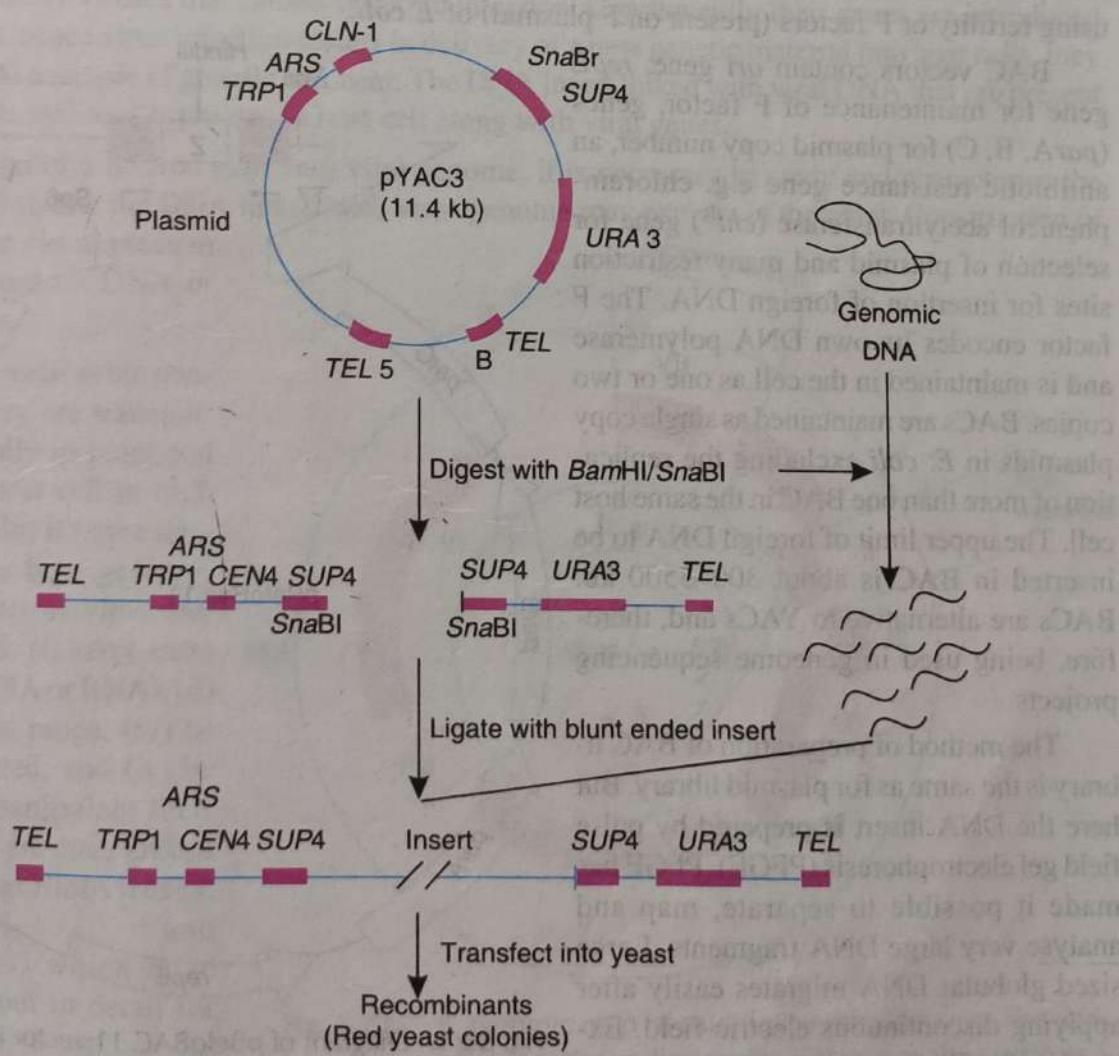


Fig. 5.20. Yeast artificial chromosome (YAC3) cloning vector.

The YAC3 vector contains the *E. coli* origin of replication (*ori<sup>E</sup>*) and a selectable marker (*amp*), a yeast DNA sequence, genes each for uracil biosynthesis pathway (*URA3*) providing centromeric function (*CEN*), autonomously replication sequence (*ARS*) (which acts as yeast *ori*), tryptophan synthesis pathway (*TRP*) and telomeric (T) sequence (which is extended by the enzyme telomerase inside the yeast cell) (Fig. 5.20). The *TRP1* and *URA3* are yeast selectable markers, one for each end, which ensure that properly reconstituted YACs survive in yeast cells, *Sup4* is inactivated in recombinants and acts as a basis for red-white selection unlike blue-white selection in *E. coli*. There are recognition sites for restriction enzymes such as *SmaI* and *BamHI*.

For cloning purpose, YAC is digested with restriction enzyme (*BamHIII* and *EcoRI*) and recombinants are produced by inserting a large fragment of genomic DNA. This molecule can be maintained in yeast as YAC. The transformants that contain YAC can be identified by red/white colour selection. Non-transformed yeast contain white colonies. Red colonies of yeast contain recombinant YAC molecules. Due to insertion of DNA molecule into *EcoRI* site, *Sup4* gene is inactivated and no protein will be expressed. This facilitates to develop red colonies.

## F. BACTERIAL ARTIFICIAL CHROMOSOME (BAC)

Mel Simon and Co-workers developed the BAC which is another cloning vector system in *E. coli* and an alternative to YAC vector vectors. BACs are maintained in *E. coli* as a large single copy. These are used as vectors for cloning very large (>50 kb) sequences of DNA. Some times these fragments turn out to comprise non-contiguous (non-adjacent) segments of the genome and frequently lose parts of the DNA during propagation. Hence, these become unstable. BAC are constructed by using fertility or F factors (present on F plasmid) of *E. coli*.

BAC vectors contain *ori* gene, *repE* gene for maintenance of F factor, genes (*parA*, *B*, *C*) for plasmid copy number, an antibiotic resistance gene e.g. chloramphenicol acetyltransferase (*chl<sup>R</sup>*) gene for selection of plasmid and many restriction sites for insertion of foreign DNA. The F factor encodes its own DNA polymerase and is maintained in the cell as one or two copies. BACs are maintained as single copy plasmids in *E. coli* excluding the replication of more than one BAC in the same host cell. The upper limit of foreign DNA to be inserted in BAC is about 300–3500 kb. BACs are alternative to YACs and, therefore, being used in genome sequencing projects.

The method of preparation of BAC library is the same as for plasmid library. But here the DNA insert is prepared by pulse field gel electrophoresis (PFGE). PFGE has made it possible to separate, map and analyse very large DNA fragments. Large sized globular DNA migrates easily after applying discontinuous electric field. Example of a BAC vector is pBeloBAC 11

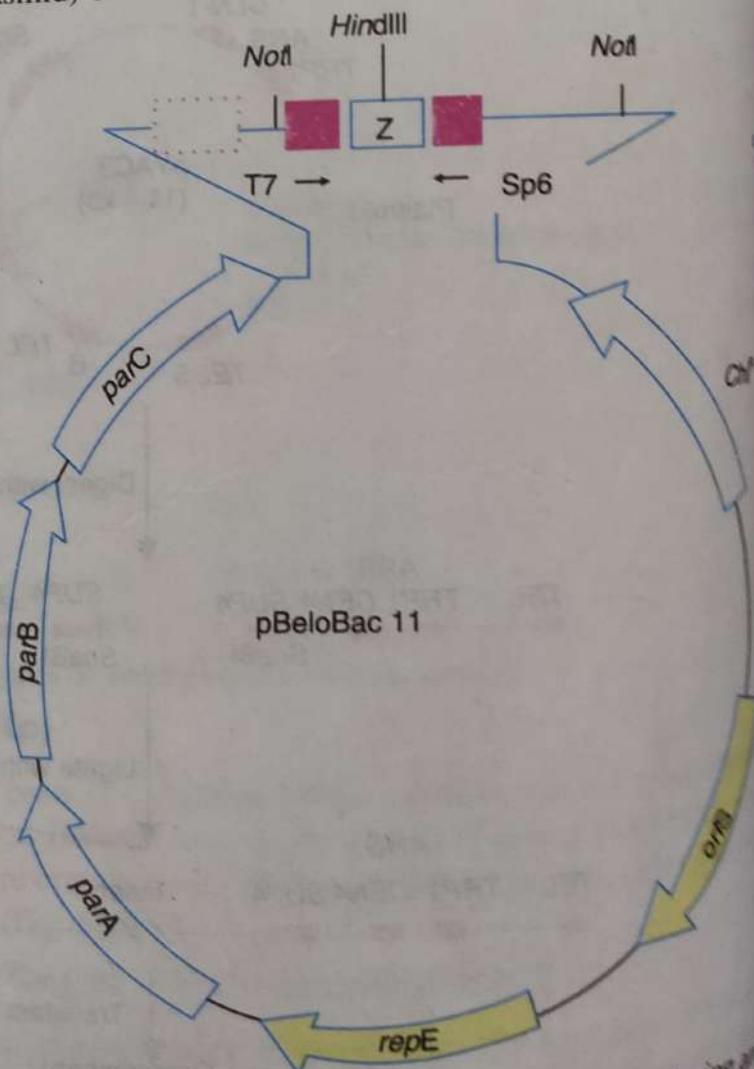


Fig. 5.21. Diagram of pBeloBAC 11 vector and cloning and selection of genomic DNA by using this vector.

(Fig. 5.21) which is partially digested while cloning a DNA fragment. Genomic DNA is partially digested with *Hind*III. Size fractionation of DNA is made by PFGE. Similarly, pBeloBAC 11 is treated with *Hind*III and phosphatase. Such digested vector and fragments of genomic DNA (100–300 kb) are ligated. The recombinant plasmid containing DNA insert is transferred into *E. coli* cells through electroporation (See Chapter 6). *E. coli* can be transformed very efficiently by BAC. Then *E. coli* cells are plated on LB medium containing chloramphenicol, IPTG and X-gal for selection. White colonies of *E. coli* are picked up based on chloramphenicol resistance.

In 1989, by using this vector DNA segments from bithorax gene of *Drosophila* have already been cloned. BAC clones are stable for many generations. In contrast, it lacks positive selection for clones that contain DNA inserts. Therefore, the yield of DNA is very low.

## G. PLANT AND ANIMAL VIRUSES AS CLONING VECTORS

Using gene cloning techniques based on *E. coli* as host, many eukaryotic genes and their control sequences have been isolated and analysed. Many applications of genetic engineering require vectors for the expression of genes on diverse eukaryotic organisms for engineering of new plants, gene therapy and large scale production of eukaryotic proteins. Animal and plant-based vectors have designed for this purpose have been discussed in this section.

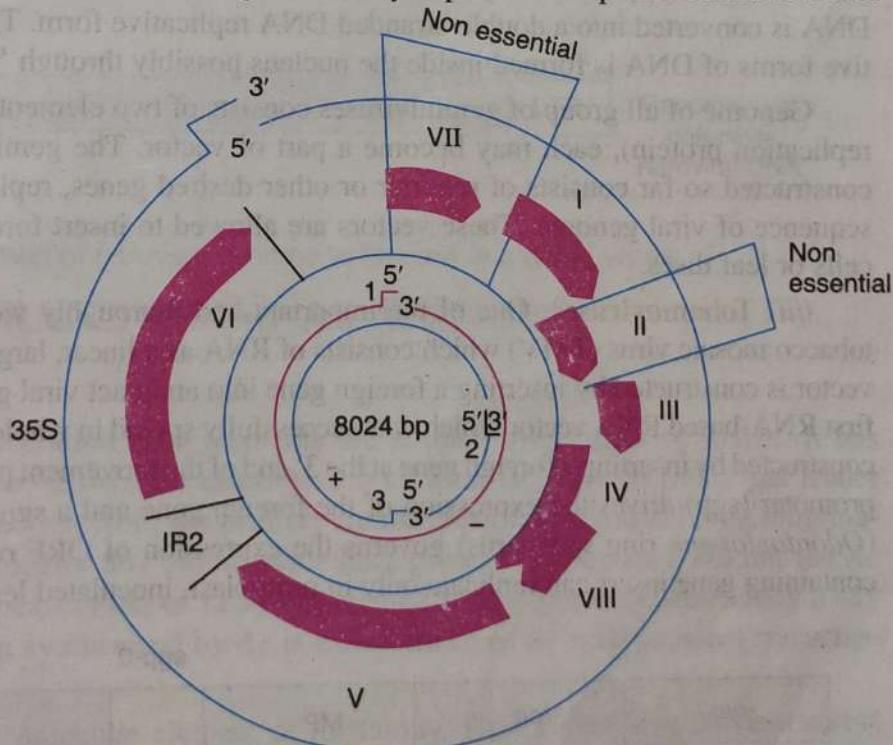
### 1. Plant Viral Vectors

Plant virus-based vectors can be used for transfer of genes with high efficiency. Because virus genome containing DNA insert cause viruses infection in host plant and amplified naturally. However, there are many viruses that causes systemic infection. Consequently their genes are introduced into all the cells. Since virus infection results in delivery of a new genetic material into host cells, they provide a natural example of genetic engineer. The DNA insert linked with viral DNA that are present in host cell replicates and expresses in host cell along with viral genes.

For insertion of a desired gene into viral genome, it is necessary to study and characterise the virus into detail so that the DNA linked with virus genome may express in the plant. Construction of viral vector that can express in host cell is based on DNA or RNA molecules.

Plant virus vectors are non-integrative. They are transmitted systematically in plant and assembled in host cell in high copy number. But it never integrates with the host genome. Therefore, to act as viral vector, they should: (i) carry extra nucleic acid (DNA or RNA), (ii) have broad host range, (iii) be easily transmitted, and (iv) be preferred to manipulate their genome. There are three groups of viruses (caulimoviruses, geminiviruses and tobamoviruses) which have been worked out in detail for construction of vector.

**(a) Caulimoviruses:** This



**Fig. 5.22.** A genomic map of cauliflower mosaic virus. The dark boxes are the coding region. DNA is in the centre as a thin line and outside as 35S transcript.

group includes cauliflower mosaic virus (CaMV) which has extensively been studied. This virus infects mainly the members of Cruciferae. Genome of CaMV is made up of double stranded DNA which was first manipulated by recombinant DNA technology. The CaMV is transmitted naturally simply by rubbing of leaves or by aphids. Viral particles transmitted are isometric, 50 nm in diameter and isolated from inclusion bodies. Viral particles spread rapidly in 3–4 weeks throughout the plant and found at high copy number (about  $10^6$  virions per cell).

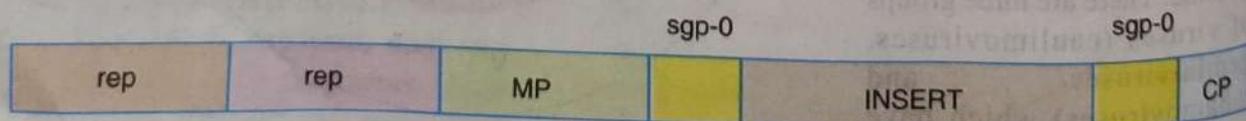
The CaMV DNA remains as linear open circular molecule of 8 kb twisted or knotted forms (Fig. 5.22). It consists of tightly packed six major and two minor (total I–VIII) open reading frame (ORF) on one coding strand. It replicates through reverse transcription. After infecting the host cell DNA enters inside the nucleus where a supercoiled minichromosome is formed after digestion of a single-stranded overhangs and digestion of gaps. Inside the nucleus of plant cell, the minichromosome acts as a template for nuclear RNA polymerase II forming RNA transcripts. Then the transcripts are transported to cytoplasm where it undergoes either translation or replication into (-)DNA strand using reverse transcriptase. The (+) DNA strand is synthesised from two primer-binding sites that are present near gaps 2 and 3. The RNA synthesis starts from gap 3' to 5' end of (-)DNA strand, whereas synthesis started from gap 3 continues to gap 2. Finally, DNA molecules are packed to form viral particles. These particles re-start the similar events after re-entering the cell of a healthy tissue of other plants.

The two genomic regions of the virus (ORF I and ORF VII) are non-essential. ORF II synthesises the factors for insect transmission and ORF VII has un-known function. These two regions can be replaced by a foreign DNA and vector constructed. By using this approach, in 1984, Brisson and co-workers transferred CaMV vector containing the bacterial dihydrofolate reductase, into turnip plant cells. This gene expressed into the plants.

**(b) Geminiviruses:** This group consists of single-stranded DNA containing plant viruses that cause diseases in maize cereals, cassava, tomato, *Digitaria*, *Chloris*, etc. It has geminated (two paired particles) morphology which differentiate it from the other groups of viruses. The capsids are icosahedral, of  $10\text{--}20 \times 20$  nm dimension enclosing covalently closed one or two circular single-stranded DNA molecule(s) (2.6–3.0 kb). After infection DNA molecule enters in the nucleus. The single stranded DNA is converted into a double-stranded DNA replicative form. Thereafter, many copies of replicative forms of DNA is formed inside the nucleus possibly through ‘rolling circle mechanism’.

Genome of all group of geminiviruses consists of two elements (i.e. an origin of replication and replication protein), each may become a part of vector. The geminivirus vectors which have been constructed so far consists of reporter or other desired genes, replacing the coat protein expressing sequence of viral genome. These vectors are allowed to insert foreign gene in protoplasts, cultured cells or leaf discs.

**(iii) Tobamoviruses:** One of the important and thoroughly worked out viruses of this group is tobacco mosaic virus (TMV) which consists of RNA as a linear, large, undivided genome. RNA-based vector is constructed by inserting a foreign gene into an intact viral genome. For example TB2 was the first RNA-based RNA vector which was successfully spread in whole plant (Fig. 5.23). TB2 vector was constructed by inserting a foreign gene at the 3' end of the movement protein ORF. The native subgenome promoter (sgp) drives the expression of the foreign gene and a sgp from the related virus i.e. ORSV (*Odontoglossum* ring spot virus) governs the expression of ORF of coat protein. RNA viral vectors containing gene insert can replicate only in protoplast, inoculated leaves or differentiated tissue.



**Fig. 5.23:** Genetic map of tobacco mosaic virus-based TB2 plant RNA viral vector; insert = *nptII*;  $\alpha$ -trichosanthin; CP = viral coat protein; MP = viral movement protein; rep = replication.

## 2. Animal Viral Vectors

In nature there are several viruses which cause diseases. A virus gets adsorbed to the body surface of suitable host and infects the cell. This ability of animal viruses has been exploited and virus-based vectors have been designed to introduce foreign DNA of known function into cultured eukaryotic cell. In 1979, for the first time a simian virus 40 (SV 40)-based cloning vector was constructed and used in cloning experiment using mammalian cells. Since then several vectors were constructed using adenovirus, papillomaviruses, retroviruses to clone foreign DNA into the mammalian cells, and baculoviruses in insect cells. In recent years, retrovirus-based vectors are commonly used for gene cloning, because they can infect several types of cells. Since they are single stranded, their high concentrations are required for fully differentiated cells like neurons, hepatocytes, etc. The *gag*, *pol* and *env* genes of retroviruses (which are required for replication and assembly of viral particles) can be replaced with foreign DNA. The recombinant DNA is introduced into mammalian cells in tissue culture (Fig. 5.24). For detailed discussion see Chapter 7.

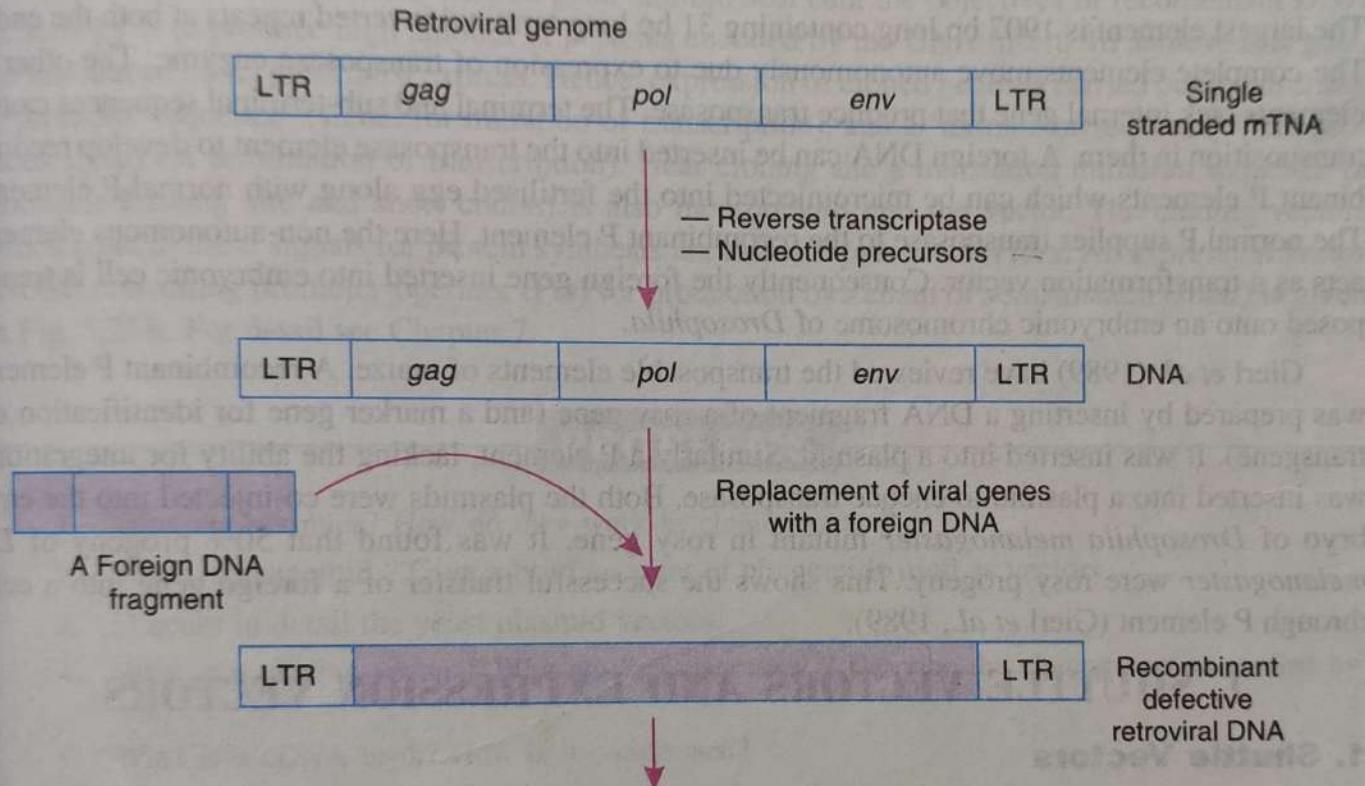


Fig. 5.24. Construction of retroviral genome to be used as cloning vector.

## H. TRANSPOSONS AS VECTOR

### 1. Ac-Ds Elements

The activator (*Ac*) and dissociation (*Ds*) elements are simple transposons of maize. It has 4.565 kb long autonomous sequence which transcribes 3.5 kb long GC-rich un-translated leader sequence. Four introns are spliced out from the primary transcript and 3.312 nucleotide long mature mRNA is obtained. Consequently 807 amino acids long protein is translated. Within the *Ac* element the terminal inverted repeats (TIR) of 11 bp are located at both the ends which play a key role in transposition. The protein synthesised by *Ac* is called transposase which causes transcription of *Ac* elements (Fig. 5.25).

The *Ds* (dissociation) is the defective element of *Ac* family. Due to deletion, non-autosomal elements are obtained from the autosomal elements. It activates the *trans*-acting transposase. The 11 bp long TIR is common to all *Ds* elements. The enzyme transposase cannot be expressed by the *Ds* element due to deletions.

But the transposase that is produced by Ac element is diffused and *trans*-acting. It causes transposition of *Ds* elements. By detecting part of the region transposons can be used as vector for cloning a foreign DNA.

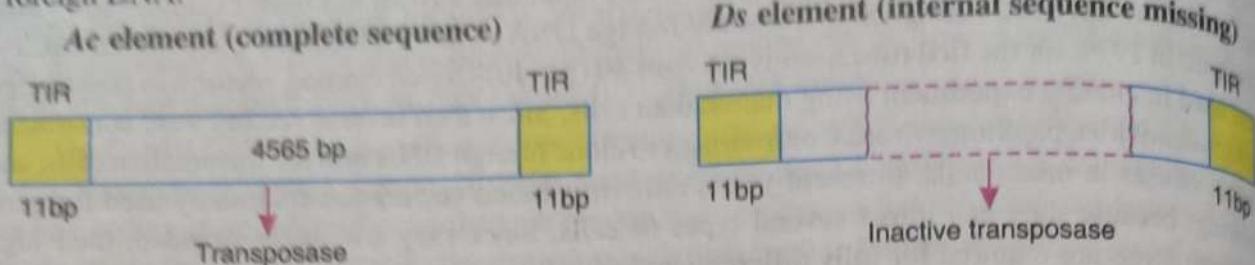


Fig. 5.25. Diagram of Ac-Ds elements of transposons in maize.

## 2. P Element

The transposons of *Drosophila* are known as P elements. The size of different P elements vary. The largest element is 1907 bp long containing 31 bp long terminal inverted repeats at both the ends. The complete elements move autonomously due to expression of transposase enzyme. The other P elements lack internal gene that produce transposase. The terminal and sub-terminal sequences cause transposition in them. A foreign DNA can be inserted into the transposase element to develop recombinant P elements which can be microinjected into the fertilised egg along with normal P element. The normal P supplies transposase to the recombinant P element. Here the non-autonomous element acts as a transformation vector. Consequently the foreign gene inserted into embryonic cell is transposed onto an embryonic chromosome of *Drosophila*.

Gierl *et al.* (1989) have reviewed the transposable elements of maize. A recombinant P element was prepared by inserting a DNA fragment of a rosy gene (and a marker gene for identification of transgene). It was inserted into a plasmid. Similarly a P element, lacking the ability for integration, was inserted into a plasmid to encode transposase. Both the plasmids were co-injected into the embryo of *Drosophila melanogaster* mutant in rosy gene. It was found that 50% progeny of *D. melanogaster* were rosy progeny. This shows the successful transfer of a foreign gene into a cell through P element (Gierl *et al.*, 1989).

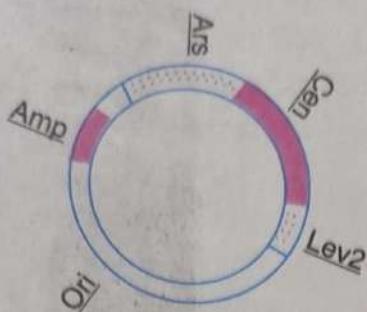
# I. SHUTTLE VECTORS AND EXPRESSION VECTORS

## 1. Shuttle Vectors

The prokaryotic vectors cannot exist and work in eukaryotic cells because the system of two groups of organisms vary. You know that prokaryotes lack introns, while the eukaryotes consist of introns. Therefore, several vectors have been constructed which may exist both in prokaryotic (*E. coli*) and eukaryotic cells. Such vectors possess two origin for replication i.e. *ori<sup>E</sup>* and *ori<sup>Euk</sup>* and, therefore, are called **shuttle vector**. The *ori<sup>E</sup>* functions in *E. coli* and *ori<sup>Euk</sup>* functions in eukaryotic cells like yeast. Besides, these vectors also contain the antibiotic resistance genes (e.g. *amp<sup>R</sup>*) that act as selectable marker gene. One of the examples of shuttle vectors is the yeast episomal plasmid (YEP) (Fig. 5.26A). The YEP is a 2  $\mu\text{m}$  circle plasmid-based vector which contains 2  $\mu\text{m}$  origin of replication, *E. coli* shuttle sequence and a selectable marker i.e. *leu2* yeast gene. The gene for  $\beta$ -lactamase is an example of a prokaryotic gene that is expressed in yeast. Kado and Tait (1983) have described the properties of shuttle vector as below :

- (i) The vector must replicate in many organism (e.g. bacteria, yeast and plants) to facilitate the isolation and characterization of genes;
- (ii) The vector must be easily recognised by selectable markers;
- (iii) The vector should be small in size to accommodate DNA inserts;
- (iv) Cloned genes should be easily detected;

- (v) Useful quantities of vector must be easily obtained.;
- (vi) The vector must be stable, non-pathogenic and non-stress-inducing;



**Fig. 5.26.** Diagram of shuttle vector (A), and expression vector (B).

## 2. Expression Vectors

In addition to incorporation desired gene into the host cell, the objectives of recombinant DNA technology is to produce high amount of proteins encoded by the DNA insert. To achieve this goal, the introduced novel gene must express. Hence, expression of cloned genes is carried out by inserting a ‘promoter sequence’ (signal for initiation of transcription), and a ‘terminator sequence’ (that provides signal for termination of transcription). Near cloning site a translation initiation sequence (a ribosome binding site and short codon) is also incorporated into the vector. The cloning vectors which contain these signals for protein synthesis are called *expression vectors*. An *expression vector* pSOMI containing promoter-operator (PO) for production of a chain of somatostatin (*somI*) is given in Fig. 5.26B. For detail see Chapter 7.

## PROBLEMS

1. What are plasmids? How do they work as cloning vector?
2. What is phagemid ? Give a brief account of phagemids used as vectors.
3. Discuss in detail the yeast plasmid vectors.
4. What is a cloning vector ? Why are they necessary ? Describe the cloning vectors studied by you.
5. What is a cDNA bank? How is it constructed?
6. Write short notes on the following:  
 (i) Cloning vectors, (ii) Phage  $\lambda$ , (iii) Cosmid and phasmid vectors, (iv) Insertion and replacement vectors (v) pBR322, (vi) *Escherichia coli*, (vii) *Agrobacterium*. (viii) Restriction-modification system (ix) BAC vector (x) YAC Vectors (xi) P1 vector (xii) Plant virus vector (xiii) Animal virus vectors (xiv) Transposons as vector.