Unit - 2

Vectors Must Possess Certain Essential Qualities

Vectors are cloning vehicles. These are the DNA molecules which carry the target DNA into host cells and which replicate leading to amplification of target DNA and its passage into the progeny host cells. Naturally, vectors must possess certain essential qualities. These qualities are given below.

- ⇒ The vector must possess an origin of replication. What this means is that it should be capable of replication.
- ⇒ The vector should replicate several times per cell cycle giving rise to a **high copy number** of the vector per cell. The higher the copy number, the higher is the amplification.
- ⇒ Ideally, it should possess restriction sites for a range of restriction enzymes. If the vector has this property, it can accept target DNAs cleaved by different restriction enzymes.
- ⇒ The vector should be comparatively small. A small size allows for easy manipulation.
- ⇒ The vector must be easily introduced into a host cell.
- ⇒ The vector should have one or more selectable markers. Presence of a selectable marker makes it easy to detect the presence of the vector in a host cell.

Three types of vectors are most commonly used - plasmids, λ phage, and viruses. All the three possess the desired properties stated above and, if there is a shortcoming on one or the other property, then these vectors have been engineered to overcome the shortfall. Of course, there are other vectors. In most cases, these other vectors have been given rise to by cutting and pasting elements of one type of vector with elements from the other. For example, we have a vector called a **phagemid** - as the name indicates, this vector is given rise to by combining features of a plasmid and a phage. Most of these other vectors are for specialized uses and we will be discussing them after we have discussed the generalities of molecular cloning. In this section, we will content ourselves with a discussion only of the three main vector types and some of their variations.

Plasmids have Basic Attributes of a Good Vector

Plasmids come in different sizes. The smallest are less than 1 kb and the larger ones can be in excess of 500 kb. They are self-replicating, circular DNA duplexes that are maintained in bacteria as independent extrachromosomal entities. All plasmids contain the genetic machinery, such as a replication origin, to allow their autonomous propagation in a bacterial or yeast cell. Almost all bacterial genera possess plasmids. Plasmids are of several types. Some plasmids contain information for their own transfer from one cell to another (bacterial conjugation; F plasmids); some others have genes that confer antibiotic resistance upon the cells in which they exist (R plasmids); yet others contain genes which encode enzymes involved in utilizing unusual metabolites (degradative plasmids); some other plasmids do not seem to possess functional coding genes (cryptic plasmids). The antibiotic resistance function of the plasmids is alarming for the general populace as well as the medical fraternity because these are considered to be central to the unabated spread of antibiotic resistance among the pathogens and consequent difficulty in treatment of diseases. The plasmids involved in conjugation also pose dangers of the same kind by being responsible for horizontal spread of resistance characteristics. *However, for the scientists doing genetic research, the functions encoded in a plasmid are of extreme importance, since it is because of these functions that the presence of a plasmid can be easily detected in a cell. The genes for the various functions, and the phenotypes which are a consequence of these genes, become the best selectable markers for selecting cells that have received either the plasmid or the recombinant DNA molecule. It is because of the extreme usefulness of these functions that scientists have generated several chimeric plasmids by engineering into them certain functions so that they could become better cloning vehicles.*

Plasmids differ from each other in terms of the copy number in which they exist in the host cells. For some plasmids, the copy number is small - just about 1 to 4 per host cell. These plasmids are called **low-copy-number plasmids** and they replicate just about once per cell division. The replication of these plasmids is said to be under **stringent control**. Other plasmids, used in gene cloning, exist in 10 to 700 copies per cell. These are called **high-copy-number plasmids** and their replication is said to be under **relaxed control**. Moreover, if one wants that the number of plasmids per cell increases even further, all one has to do is to treat the cell with chloramphenicol. This antibiotic inhibits protein synthesis and therefore blocks cell division. In these cells, the plasmids continue to replicate and at times even two to three thousand copies of a given plasmid can exist in a cell.

If a plasmid exists in a cell and you try to put another type of plasmid in the same cell, it is not necessary that they coexist. If two or more types of plasmids cannot coexist in a cell, they are said to belong to a **single incompatibility group**. Two plasmids belonging to a single incompatibility group cannot be maintained in the same cell. But plasmids belonging to different incompatibility groups can live happily together in the same cell. Thus, one can come across a cell containing as many as ten different types of plasmids, all belonging to different incompatibility groups. But while they coexist, they do not affect each others copy number - each still maintains its unique copy number.

Another peculiarity that some plasmids possess is with respect to their capability to replicate in a given host cell. Some plasmids replicate in a species specific manner - they replicate only if they are in a particular species (**narrow-host-range plasmids**). Other plasmids are less fussy about where they multiply (**broad-host-range plasmids**).

Plasmids have the basic attributes to be good vectors. However, naturally occurring plasmids lack several features that are required in a high-quality cloning vehicle. Let's see what are the features a plasmid should preferably possess.

⇒ Small size.

⇒ Small size is synonymous with easy manipulation - large plasmids break apart more easily.

⇒ Small plasmids often have a high-copy-number.

⇒ Small size automatically means more insert per mg plasmid.

⇒ Small sized plasmids can easily get into the host cell, i.e. , transformation is easy.

Thus, only those plasmids which are of a size less than 15 kb are preferred. A large number of plasmids are much larger than this limit and thus are not good cloning vehicles.

⇒ Presence of multiple selectable markers.

⇒ As we have said earlier, selectable markers make it easy to detect the cells possessing a plasmid.

⇒ If the plasmid possesses a single marker, it is easy to detect its presence in the host cell. But if it has more selectable markers, one can differentiate between host cells containing just the resealed plasmid and the host cells containing plasmids bearing the target DNA (we will talk about this aspect shortly).

Not all plasmids are rich in selectable markers.

⇒ Range of unique restriction sites.

⇒ Unique means single. It is preferable that the plasmid possesses a single site for a given restriction enzyme. If this is so, use of the restriction enzyme will linearize the plasmid. If the plasmid has more than one site for a given restriction enzyme, use of that enzyme will break the plasmid into several fragments. There is no guarantee that upon ligation these fragments will join in the correct manner. For example, if a plasmid is broken into four fragments - A, B, C, and D - upon using a particular restriction enzyme, the fragments can ligate as ACBD, ADBC, ADCB, BDCA .... and so on. A scrambled plasmid may be of no use for cloning purposes.

This is where unique restriction sites become important.

⇒ High copy number.

⇒ If you have to amplify the target DNA, the vector you put it into must replicate profusely.

While the above is a list of what a plasmid vector should possess, there are also certain features that a plasmid vector should not possess.

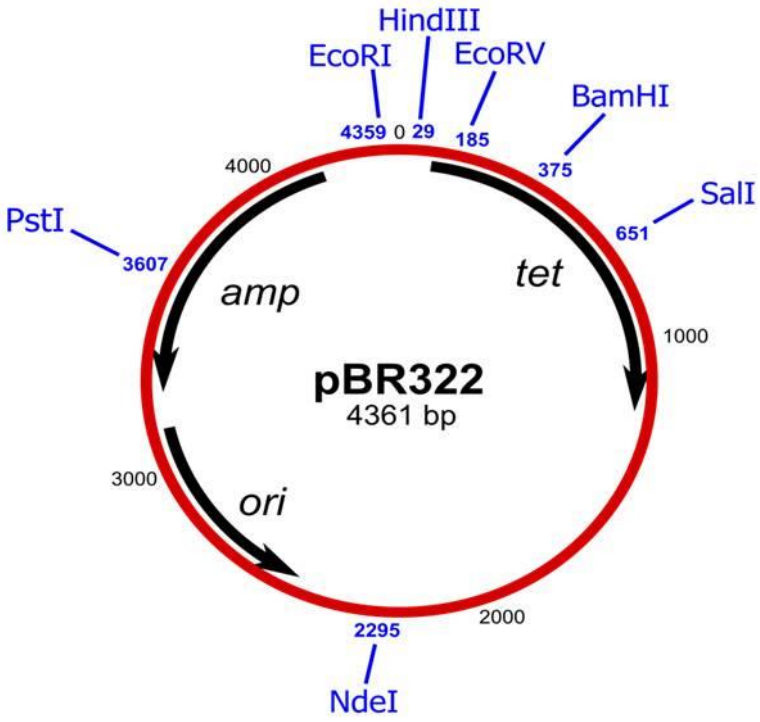
⇒ Some plasmids are autotransmissible. These plasmids can spread horizontally owing to the presence of either *tra* genes or the *mob* genes. A clone is made with considerable time, effort, and money. One cannot therefore use plasmids which make dispensation of the clone so easy. Thus the plasmid should not contain the above functions.

All the desirable features considered above are not present in naturally occurring plasmids. Plasmid cloning vectors have therefore to be genetically engineered.

- **Types of vectors**

1. pBR322
2. pUC18
3. Bacteriophage vectors (improved λ vector)
4. Cosmids
5. YAC (Yeast Artificial Chromosome)

1. pBR322



pBR322 is one of earlier chimeric plasmids which was profusely used as a general purpose cloning vector. Before we go ahead, let's see what the abbreviation pBR322 stands for. The lower case 'p' here signifies plasmid. B and R stand for the names of the scientists who designed this vector - F. *B*olivar, and R. *R*odriguez. The numerals - 322 - carry no meaning to us but must have been important to the two investigators in course of development of this vector. The significant features of this vector are given below.

- ⇒ The plasmid is 4361 bp long.
- ⇒ There are two selectable markers in this vector. Both are in the form of antibiotic resistance genes. One is for ampicillin resistance (Amp^r) and the other for tetracycline (Tet^r).
- ⇒ There are 5 unique restriction sites within the plasmid. Out of the five, three - *Bam*HI, *Hind*III, and *Sal*I - lie within the Tet^r gene. The fourth site - *Pst*I - is within the Amp^r gene. The fifth is *Eco*RI and does not interrupt any of the two resistance genes.
- ⇒ The plasmid is capable of replication only in *E. coli*. Thus, its transfer to any other bacteria is difficult.
- ⇒ It is maintained in high copy number in the host cell.

One look at the above features tells you that the plasmid possesses all the attributes that we thought should be present in a vector. But let's see how it works.

Restriction digestion with any of the five enzymes for which unique sites are there in this vector linearizes the vector. The target DNA, which has been processed with the help of the same restriction enzyme, is now mixed with the linearized vector. As we have seen earlier, it is not necessary that this mixing would always result in a recombinant DNA molecule. On the contrary, circularization of the vector is kinetically the most favored reaction. The other undesirable reaction which can take place is the tandem joining of several target DNA fragments. We are not bothered about the latter because such tandem fragments do not normally possess a replication origin. And as far as the circularization of the plasmid is concerned we can lessen its occurrence by treating the linear plasmid beforehand with alkaline phosphatase removing the 5′-

phosphate. Thus, the plasmid can circularize only after it has incorporated a target DNA fragment containing a 5′-phosphate group.

Figure 17.8 is a diagrammatic depiction of the features present in pBR322.

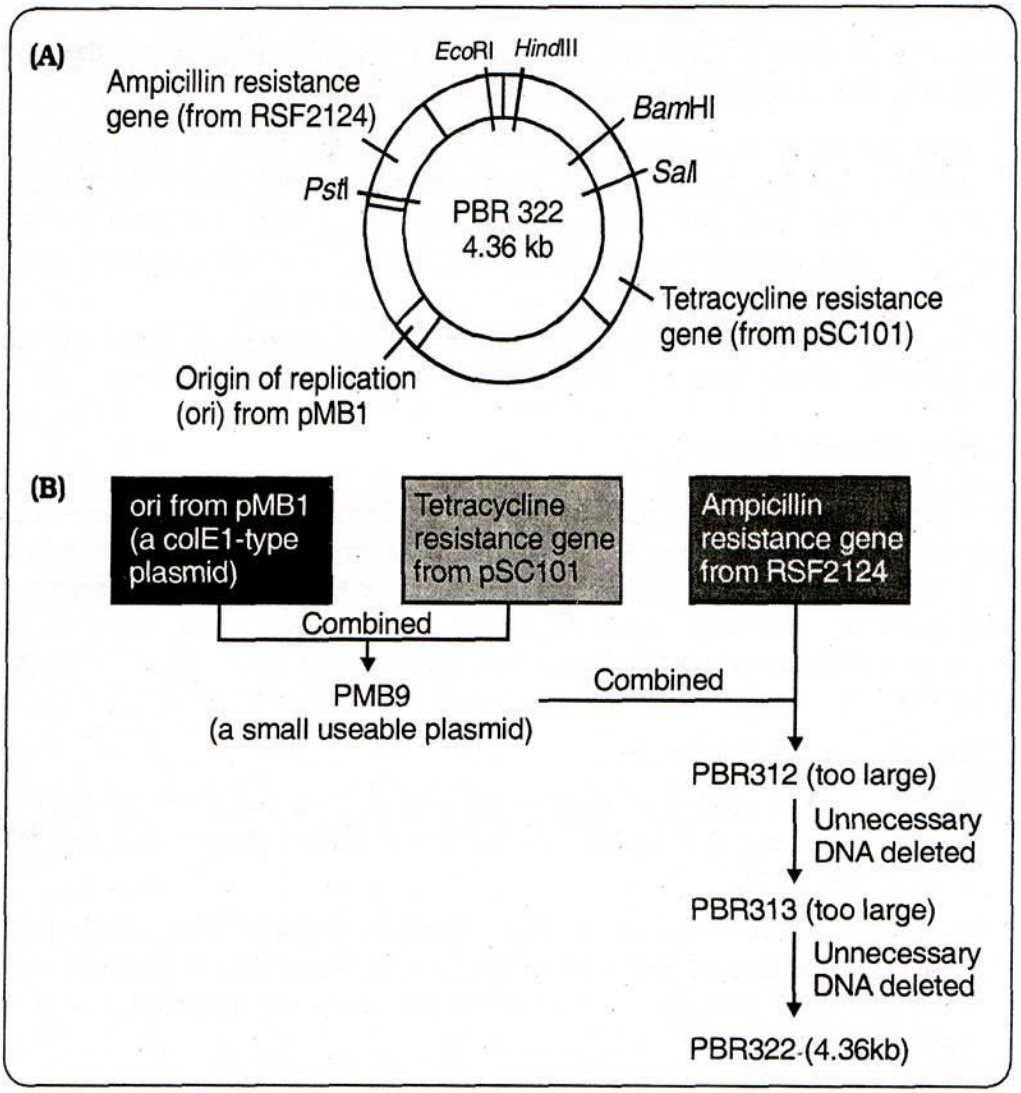


Figure 17.8. (A) Genetic map of the plasmid cloning vector pBR322. The plasmid contains two selectable markers - Amp^r and Tet^r. Note that except the *Eco*RI site all other restriction sites are within antibiotic resistance genes. (B) The steps in the construction of the artificial, genetically engineered, plasmid pBR322.

Recombinant Transformed Cells Can be Selected by Simple Antibiotic Sensitivity Tests

Let's look at the process of transformation now. One of the most used method to introduce DNA into *E. coli* cells is the so called 'heat shock' method. What we do here is we treat the bacteria with $CaCl_2$. Now we incubate these pretreated cells with DNA. This step allows DNA to bind to the bacterial surface. Now we raise the temperature of the medium to ~42° C. This rise is temperature is supposed to drill holes into the bacterial membrane allowing the uptake of DNA.

This transformation process is inefficient. We say this because just about ~1 cell

per thousand is transformed. Moreover, not all the cells which are transformed have received the recombinant DNA. Some have taken up the recircularized plasmid that escaped the phosphatase treatment. Some others have received the tandem joined copies of the target DNA fragments. Some others have truly received the recombinant plasmid. It is in the last group of cells that we are interested in. Our task now is finding out cells that bear this recombinant DNA. This process is called **selection**. Form a mental picture of the task - you have to search for an extremely small number of cells within a majority that are either not transformed or have been transformed by the recircularized plasmid. This is like searching for a needle in a haystack. And this is where the selectable markers - the antibiotic resistance genes - come in handy and make the job extremely simple.

Let's understand the process of selection with the help of an example. Let us assume that we have inserted the target DNA in the *PstI* site. Look at Figure 17.8 - *PstI* site lies within the Amp^r gene (at this stage discontinue reading and try to design a scheme for finding out which cells have received the recombinant DNA; its easy actually). The selection process involves just two steps. And these two steps are suggested by the following logical deductions

- ⇒ We have used *PstI* site for insertion. This site is within the ampicillin resistance gene.
- ⇒ Therefore, the plasmids which have received a target DNA fragment will have an interrupted Amp^r gene. But their Tet^r gene will remain functional.
- ⇒ The plasmids which have not received the target DNA will have both resistance genes functional.
- ⇒ Therefore *cells transformed only with the plasmid will be resistant to both antibiotics; cells transformed with recombinant plasmid will be sensitive to ampicillin but resistant to tetracycline; cells which have not received the plasmid will be sensitive to both antibiotics.*

The first step is therefore to plate all the cells after transformation onto a medium containing tetracycline. The cells which grow in this medium are sure to have received the plasmid. Cells not having the plasmid will not grow. Thus, in a single step we have selected the cells containing the plasmid. Now, in the second step, we have to select for those cells which contain the recombinant plasmids. Using the replica plating technique now, we transfer cells from all the colonies to a petriplate which contains ampicillin. Those cells which have received only the plasmid will grow here too. The other cells, which have received recombinant plasmids, will not be able to grow. Replica plating technique transfers the colonies in a manner that the arrangement of the colonies in the previous plate is preserved. Thus, all we have now to do is note the positions of the colonies and find out which positions have failed to grow in the ampicillin containing medium. These are the colonies that have received the plasmid-target DNA construct. We can now go back and pick up these colonies from the tetracycline plate and pool all the cells of all the colonies together. These are cloned cells. The entire process of selection is diagrammed in Figure 17.9.

It may have been that we started with the entire genomic DNA, processed it with *PstI*, and then incubated all the fragments so generated with *PstI* treated pBR322. In this case each cell may contain a different fragment of the target DNA. Thus, if you are after a particular gene/fragment, then you have to search further for the cell which

contains that particular fragment. This procedure is called **screening**. In a later section we will see how screening is done.

Just one more thing. We have to maintain the plasmid-insert construct in its original form and see to it that it perpetuates to the progeny cells in the same form. If this construct is altered, our entire cloning experiment goes to zero. There are two potential processes that could harm our construct. One is the restriction enzymes of the host cell - our construct may contain sequences recognized by the host cell restriction enzymes. If this is the case, our construct will be digested. Another process which can alter our construct is recombination - our construct may contain sequences that are similar to certain other sequences in the host. In this case, homologous recombination can take place leading to alteration of our construct. How do we avoid this situation? The answer is simple. We do not take any cell to be the host. We first of all knock out the genes for restriction endonucleases from the cells we want to use as host. This takes care of the first problem. Homologous recombination cannot take place without the mediation of RecA. Therefore we knock out the gene *recA* from the host cell. Thus the host is a cell that lacks genes for restriction endonucleases and is *recA*$^-$.

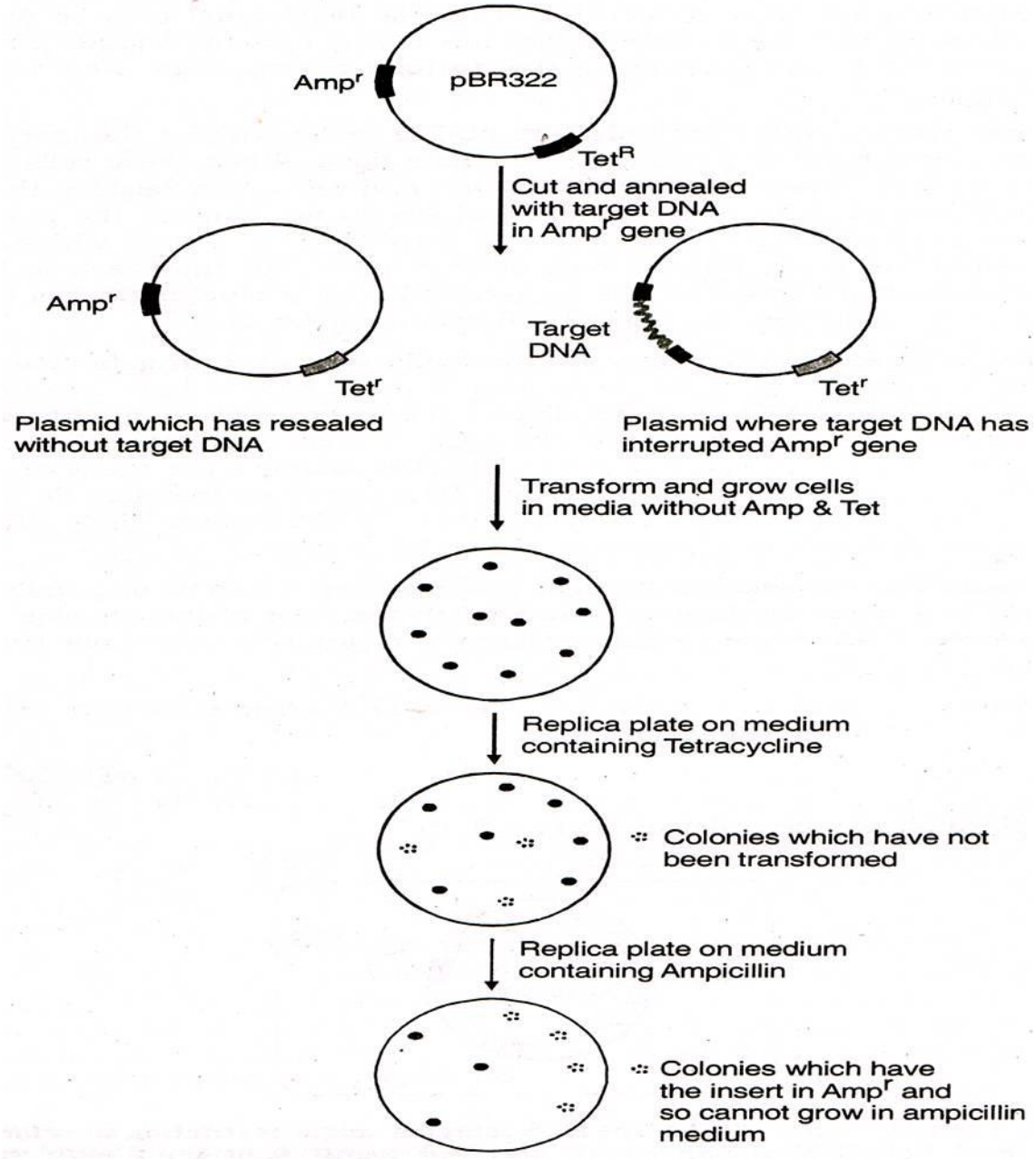


Fig. Selection of cells that contain pBR322 plasmid cloned DNA construct

2. pUC18

One disadvantage with pBR322 is that it has just a few unique restriction sites. Thus there is very limited choice of restriction enzymes with which we can process the target DNA. Another disadvantage is that the selection procedure is quite lengthy - there are two steps and each step takes time since it involves bacterial growth.

pUC18 remedies both these problems. It has 14 unique restriction sites within a short sequence which you can either call a polylinker or an MCS (*m*ultiple cloning sites). The replication origin is derived from pBR322. It has an Amp^r gene as a selectable marker. But the actual selection system is quite novel and actually it is this system which should interest you in knowing about pUC18.

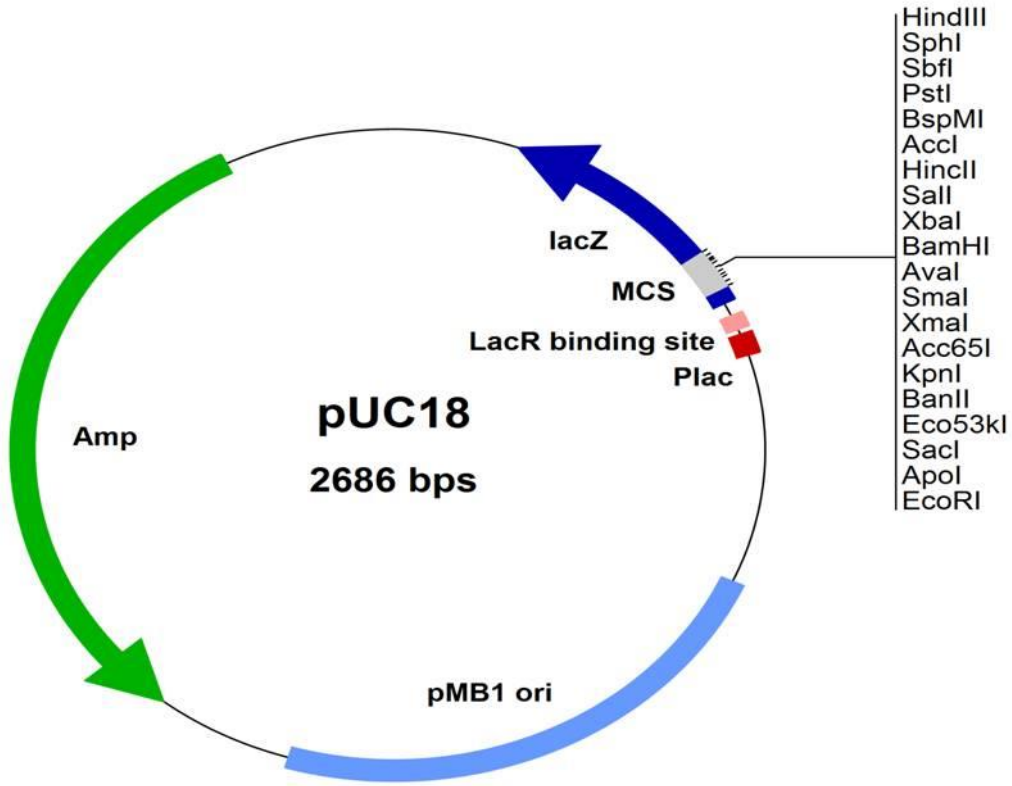


Figure 17.10. A genetic map of pUC18. The MCS contains unique restriction sites for *EcoRI*, *SacI*, *KpnI*, *XmaI*, *SmaI*, *BamHI*, *XbaI*, *SalI*, *HindII*, *AccI*, *PstI*, *BspMI*, *SphI*, and *HindIII*. This is the site where the target DNA is inserted. The vector contains *lacI* gene whose product represses the expression of the *lacZ'* gene in the absence of IPTG. Note the presence of the ampicillin resistance gene. All this is contained within a DNA sequence that is just 2686 bp long.

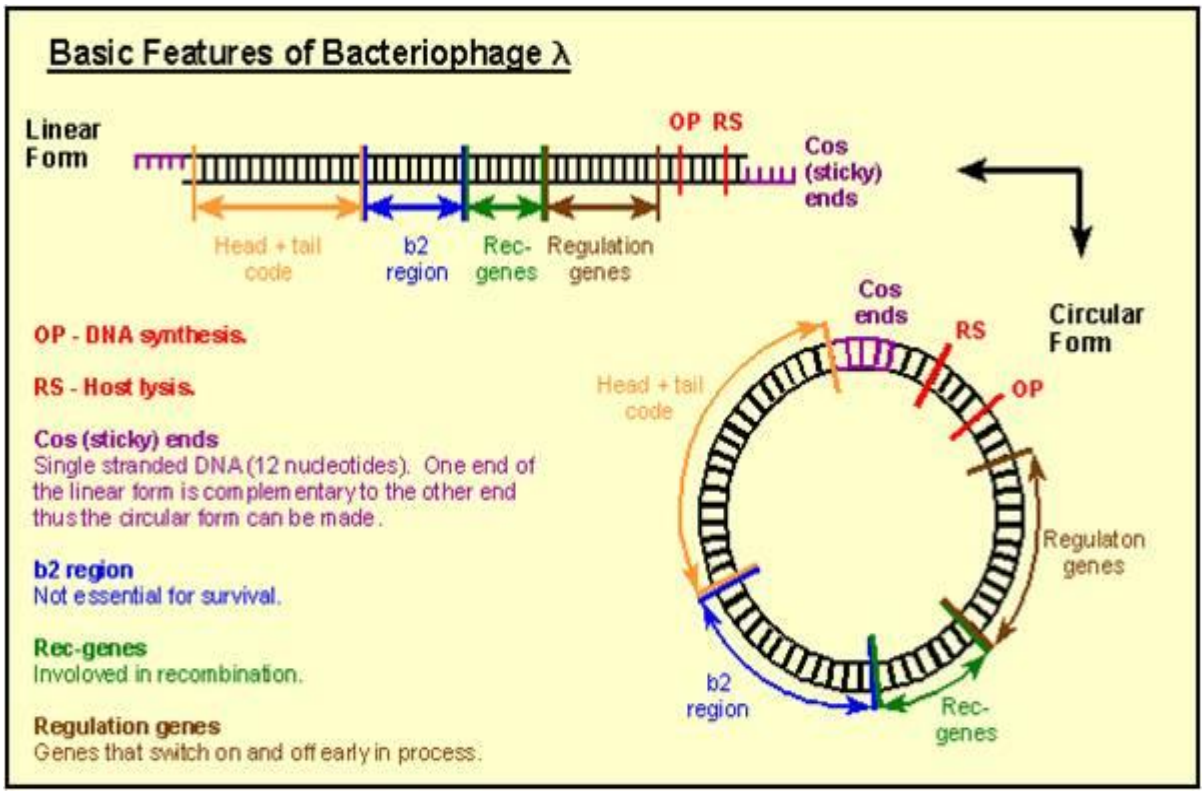
There is one more way in which pUC18 is much superior to pBR322. pUC18 is a higher copy number plasmid as compared to pBR322 because the regulatory gene *rop*, present in pBR322, has been deleted from pUC18.

3. Bacteriophage vectors (improved λ vector)

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

Several bacteriophages are used as cloning vectors, the most commonly used *E. coli* phages be λ (lambda) and M13 phages. Plasmid vectors have to be introduced into bacterial cells, which are t cloned and selected for the recovery of recombinant DNA. In contrast, the phage vectors are dire tested on an appropriate *bacterial lawn* (a continuous bacterial growth on an agar plate) where e phage particle forms a *plaque* (a clear bacteria-free zone in the bacterial lawn). Phage vectors pres two advantages over plasmid vectors. (1) They are more efficient than plasmids for cloning of large DNA fragments; the largest cloned insert size in a 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 t bacterial colonies for the identification of recombinant vectors.

λ Phage Vectors



The λ genome (size, 48,502 bp) contains an origin of replication; genes for head and tail proteins, for enzymes for DNA replication, lysis and lysogeny; and single-stranded protruding cohesive end 12 bases (5'GGGCGGCGACCT; the other end is complementary to it, *i.e.*, CCCGCCGCTGGA5'). The λ genome remains linear in the phage head, but within *E. coli* cells the two cohesive ends am 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 use of wild type λ genome as a vector presents two serious problems. The first problem concerns the size of DNA insert. *The λ DNA must be larger than 38 kb and smaller than 52 kb to be packaged into phage particles.* Thus *λ* genome can accommodate only ~3 kb DNA insert. This problem is resolved by deleting the central region of λ genome. Most of the genes for lysogeny are located in the segment between 20 and 35 kb; this region is called the *'nonessential' region* since it can be removed without affecting the ability of phage to infect *E. coli* cells. The whole or a part of segment is deleted to create λ, vectors that (1) accommodate larger DNA inserts and (2) ensure recombinant phage to be always lytic.

The *second difficulty* arises from the fact that the λ genome is so large that it contains more than one recognition site for virtually every restriction

enzyme. As a result, cleavage by restriction enzymes cannot be used to integrate the DNA insert into the λ genome. A process of natural selection was used to recover λ genomes lacking recognition sites for specific restriction enzymes. This was achieved as follows: An *E. coli* strain producing the given restriction enzyme, say, *EcoRI*, was infected by λ phage. Most λ genomes will be digested by *EcoRI*, but a few will survive and produce plaques. Several cycles of such a selection will yield mutant λ genomes lacking the given restriction sites, *e.g.*, EcoRI sites in this case.

Several vectors have been produced from wild type λ genome by mutation and recombination *in vivo* as well as by recombinant DNA techniques. These vectors have the following two basic features.

(1) The vector itself can be propagated as phage in *E. coli* cells enabling preparation of vector DNA.

(2) They contain restriction sites, which allow removal of the lysogenic segment and also provide insertion site for the DNA fragment to be cloned (Fig. 2.16). During annealing and ligation of the DNA insert with the λ vector, two or more recombinant vectors may join end-to-end producing a *concatemer*, which is the proper precursor for packaging of λ genomes into phage heads. The various λ vectors are classified into two groups as follows: (1) insertion vectors and (2) replacement vectors.

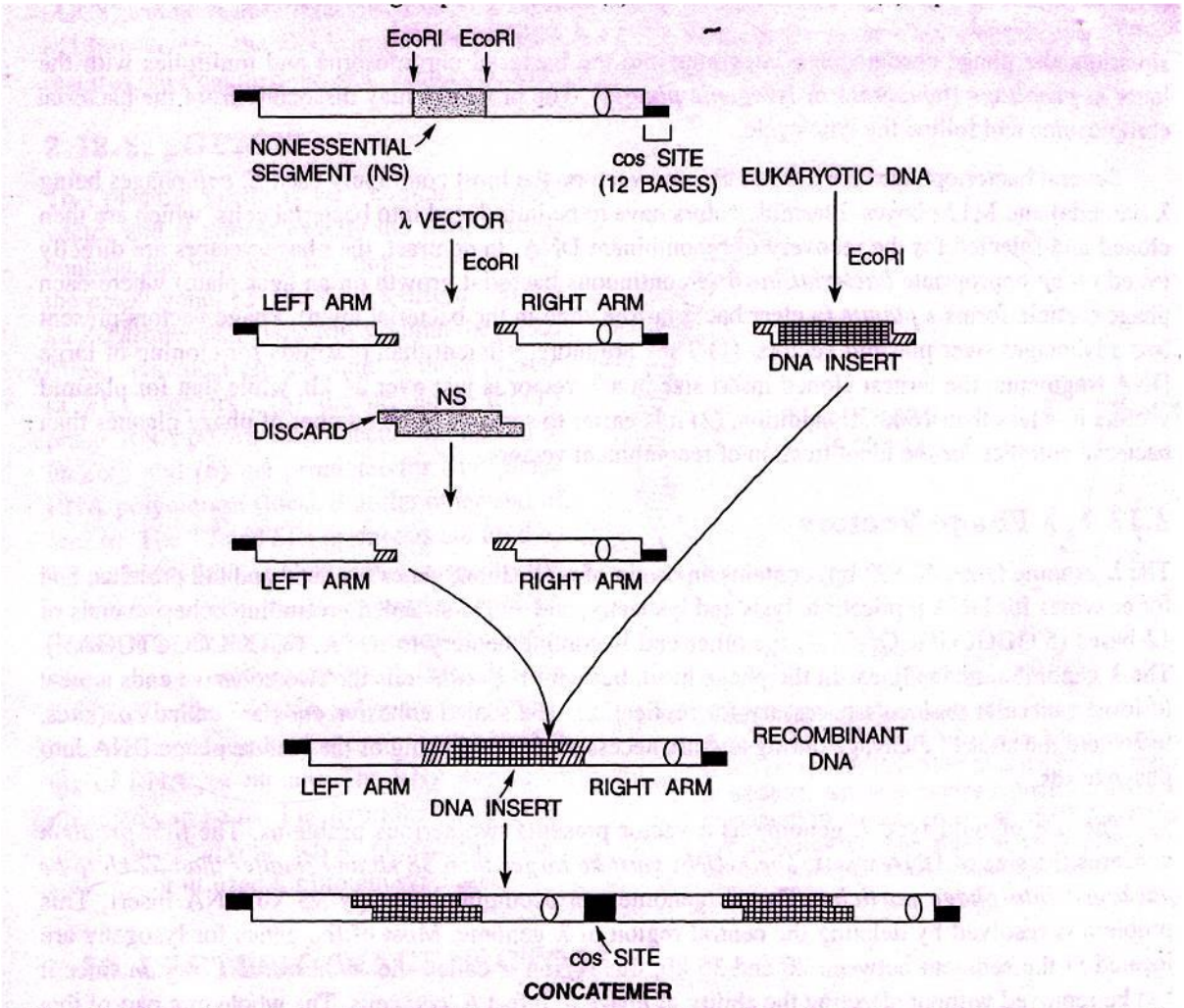


Fig. 2.16. Production of recombinant DNA from a λ vector and formation of concatemer during annealing and ligation. The open circle (O) within the λ genome represents the origin of DNA replication; the 'nonessential segment' has genes that are needed to establish lysogeny.

Insertion Vectors: In case of an *insertion vector* a large portion of the nonessential region is deleted and the two arms of λ genome are ligated. The vector contains at least one unique restriction site within which the DNA insert is integrated. Examples of such vectors are λgt10, λgtl1, λZAPII, etc. In case of λgtl0, DNA insert is placed within *cI* gene of λ, while in λgtl1 and λZAPII, it is integrated within the *lacZ* gene contained in the vector.

Replacement Vector: In these vectors, insertion of DNA fragment is accompanied with the deletion of all or the major part of the 'nonessential' region of λ genome; the deleted region is often called the *stuffer fragment.* All such vectors have two recognition sites for the restriction enzyme used for cloning; *these sites flank the stuffer fragment.* For example, in λWES.λB' two

EcoRI sites flank the stuffer fragment, while in λEMBL4 the stuffer fragment is flanked by pair of sites for *EcoRI, BamH1* and *Sai1* so that any of these enzymes can be used for cloning. Often the stuffer fragment contains additional restriction sites for the concerned enzyme. As a result, it is cut up into several pieces so that it is unable to reinsert into the vector in the place of the DNA insert. The size of deleted vector (= λ genome minus the stuffer region) is, ordinarily < 38 kb, so that it cannot be packaged into the λ phage particles. Replacement vectors can be used to clone much larger DNA inserts than in case of insertion vectors. For example, λEMBL4 can be used to clone up to 20 kb DNA.

Selection of Recombinant X Vectors: The recombinant λ vector has to be selected from nonrecombinant ones; this can be achieved in one of the following three ways.

1. Some λ vectors, *e.g.,* λgt10 (DNA insert of 8 kb), retain the lysogenic function so that they form turbid or cloudy plaques. In such vectors, the DNA fragment is inserted within the gene *cI* (lysis repressor) so that the recombinant DNA is cI^-, and forms clear plaques, which are easily distinguishable from the cloudy plaques produced by the unaltered vector. When such vectors are tested on an *E. coli* strain carrying the mutation *hfl*A150 (high frequency lysogeny), only the recombinant DNAs form plaques (all the unaltered vector molecules enter the lysogenic cycle).

2. Some λ vectors contain the promoter, operator and *lacZ* gene of *E. coli lac* operon. Insertion of DNA fragments in such vectors inactivates *lacZ* gene as the DNA insert is paced within *lacZ'* gene that has a polylinker sequence. Therefore, the unchanged vector forms (β-galactosidase, while the recombinant DNA does not. When such a vector is tested on a $lacZ^-$ *E. coli* lawn grown on media having X-gal, then unaltered vector produces blue plaque, while the recombinant vector produces colourless plaque and is readily identifiable. This strategy is employed by the λ*gt*ll vector; it is suited for all λ vectors that are >38 kb in their unaltered state.

Both λ*gt*10 and λ*gt*11 vectors have been designed to clone cDNA. λ*gt*11 is an expression vector, the protein encoded by the DNA insert being expressed as a (β-galactosidase fusion protein. The DNA insert size for λ*gt*10 and λ*gt*11 must not exceed 7 and 6 kb, respectively.

3. In case of replacement λ vectors, insertion of the DNA fragments is accompanied with the deletion of all or a major part of the nonessential segment specifying lysogeny. In general, such deleted vectors are smaller than 38 kb, the minimum genome size essential for packaging of the vector into phage heads. The recombinant DNAs, on the other hand, are much larger than 38 kb and are packaged into phage particles. Therefore, packaging into phage heads after DNA insertion offers an efficient selection strategy for recombinant DNA. Therefore, the vector and DNA insert sizes are carefully adjusted for an effective selection. Lambda vectors EMBL3 and EMBL4, both used for preparing genomic libraries of eukaryotes, employ this strategy. These vectors easily accept DNA inserts of up to 20 kb; they contain polylinkers having the reverse order, with respect to each other, or unique restriction endonuclease sites.

4. In the case some λ vectors, recombinants are selected by their *Spi⁻ phenotype.* Wild type λ phage cannot infect *E. coli* cells that contain in their chromosomes phage P2, which is related to λ phage. Therefore, wild type λ is said to have Spi^+ (sensitive to P2 prophage inhibition) phenotype. These λ cloning vectors are so designed that integration into them of the DNA inserts makes them Spi^-, *i.e.,* insensitive to inhibition by P2 prophage. When a mixture of intact and recombinant versions of such a vector is used to infect *E. coli* carrying P2 prophage, only the recombinants are able to infect these cells and form plaques. This strategy can be employed with λEMBL4, λGEM11 and λGM12 vectors as an alternative to the selection based on size.

Infection of E. coli cells with λ vector: The λ vectors can be used as naked DNA and introduced into *E. coli* cells just as is plasmid DNA *i.e.*, by heat shock following $CaCl_2$ treatment; this is called *transfection.* The phenomenon of transfection is identical to transformation, except that in this case phage DNA (including recombinant DNA) is introduced into the bacterial cells in the place of plasmid DNA. The λ DNA used for transfection must be in circular form, *i.e.*, its *cos* sites must be hydrogen bonded to each other. Alternatively, λ DNA may be packaged into phage particles and used to infect the host cells; plaque formation by the latter is 100-10,000 times of that obtained by the former. *In vitro* packaging of λ vectors is achieved by mixing the concatameric form of vector or recombinant DNAs with preformed empty phage heads, tails and terminase. Terminase specifically binds close to the *cos* sites; it cleaves the *cos* sites to generate the typical cohesive ends of the λ genome. Special techniques

and λ mutants are required for the preparation of empty heads, tails and DNA-free terminase.

Phage vectors present several unique advantages as follows. (1) Storage of phage particles is relatively much easier than that of plasmid DNA. (2) The shelf-life of phage particles is effectively infinite. (3) Screening of phage plaques by molecular hybridization often gives cleaner results than does that of bacterial colonies. In addition, (4) transformation of bacterial host is much easier with phage particles than that with plasmid DNA.

4. Cosmids

Cosmids are essentially plasmids that contain a minimum of 250 bp of λ DNA, which includes the following sequences from phage λ genome:
(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 λ phage particles.

A typical cosmid has
(1) replication origin,
(2) unique restriction sites and
(3) a selectable markers from a plasmid. Cosmid vectors are constructed using recombinant DNA techniques.

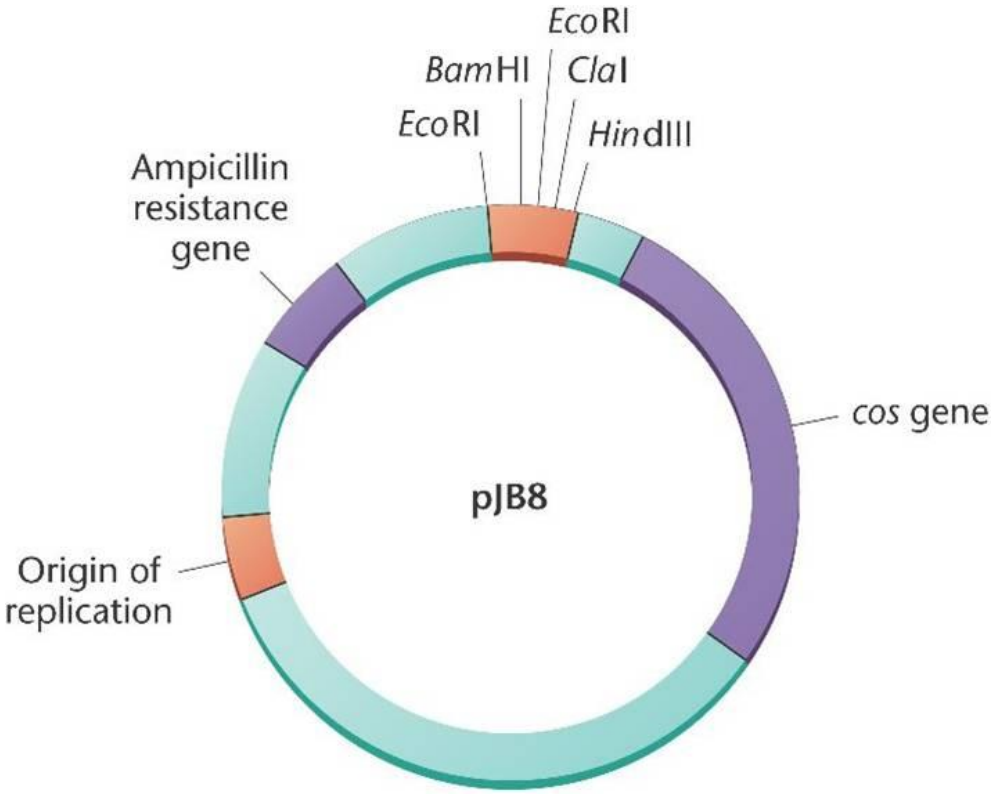
The cosmid vectors are opened by the appropriate restriction enzyme at a unique site, are then mixed with DNA inserts prepared by using the same enzyme and annealed. Among the several types of products, long cancatemers are present, which are the appropriate precursors for packaging in λ particles. This procedure selects for long DNA inserts since for packaging the distance between two *cos* sites must be between 38 and 52 kb. The DNA fragments used for cloning are usually produced by partial digestion with a restriction enzyme. This is because complete digestion almost always produces fragments that are too small for cloning in a cosmid. Cosmids can accommodate upto 40 kb long DNA inserts. Packaged cosmids infect host cells like λ particles, but once inside the host they replicate and propagate like plasmids. Because of the size limitation, only recombinant cosmids will be packaged in λ phage heads. The

transformed bacterial cells are selected on the medium containing the concerned selection agents.

The typical features of cosmids are as follows.

1. They can be used to clone DNA inserts of upto 40 kb.
2. They can be packaged into λ particles, which infect host cells; this is many-fold more efficient than plasmid transformation.
3. Selection for recombinant DNA 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 contributing plasmid.

Cosmids are particularly attractive for construction of genomic libraries of eukaryotes since they can be used for cloning large fragments. Since partial digests are used, two or more genomic fragments may join together in the ligation reaction.



This would create cosmid clones containing genomic fragments that were not located adjacent to each other in the genome; this would give incorrect information about genome organization. This problem can be overcome by separating the genomic fragments present in the partial digest on the basis of size and using only fragments of appropriate size in the ligation reaction. In

addition, the DNA fragments may be dephosphorylated in order to prevent their ligation with each other. But this approach is very sensitive to the exact ratio of DNA insert-to-vector DNA because vector-to-vector ligation can occur. Recombinant consmids having duplicated vector sequences are unstable during cloning.

Specific cloning cosmid procedures have been devised to overcome the above difficulties. An alternative solution is provided by creating cosmid vector c2XB, which has a *BamH1* cloning site and two *cos* sites separated by a blunt-end restriction site; the cloning procedure in this vector effectively prevents vector self-ligation. Modern cosmid vectors of the *pWE* and sCos series contain the following features:

(1) multiple cloning sites (MCS) for simple cloning using unselected DNA fragments,

(2) phage promoters flanking the MCS to generate RNA copies of the DNA inserts,

(3) unique *Not*I, *Sac*II or *Sfi*I (rare cutting restriction enzymes) flanking the MCS for recovery of the DNA insert as a single fragments from the vector. In addition, the vector may also contain mammalian expression modules encoding dominant selectable markers for gene transfer to mammalian cells.

5. YAC (Yeast Artificial Chromosome)

These are linear vectors that behave like yeast chromosome, hence they are called Yeast Artificial Chromosome (YAC). A typical YAC e.g. *p*YAC3, contains the following functional elements from yeast:

(1) an *ARS* sequence for replication.

(2) CEN4 sequence for centromeric function,

(3) telomeric sequences at the two ends for protection from exonuclease action, and

(4) one or two selectable marker genes, *viz., TRP1* and *URA3*, (strategy similar to other vectors);

(5) *SUP4*, *a* selectable marker into which the DNA insert is integrated; and

(6) the necessary sequences from *E. coli* plasmid for selection and propagation in *E. coli*. The *telomeric sequence* in yeast chromosomes is a 20-70 tandem repeat of the 6 base sequence 5'CCCCAA3' (its complementary sequence, 5'TTGGGG3' occurs in the other strand); there is a hairpin loop formation at the terminus, which makes the DNA duplex resistant to exonuclease action.

Vector *p*YAC3 is essentially a *p*BR322 plasmid into which the above described yeast sequences have been integrated. Subsequently, several YAC vectors have been constructed on the basic scheme of *p*YAC3. The vector itself is propagated in *E. coli,* while cloning is done in yeast. For cloning, the vector is restricted with a combination of *Bam*HI and *Sna*BI. *Bam*HI cleaves the vector at the junctions of the two *TEL* sequences with the fragment that is used to circularize the vector for propagation in *E. coli;* this fragment is discarded. The enzyme *Sna*BI recognizes the single sequence 5TACGTA3' located in *SUP4* and produces blunt-ended cleavage, thereby generating two arms of the YAC, each ending in a *TEL* sequence. The DNA insert, therefore, must have blunt ends; it is integrated within *SUP4* to generate the linear YAC. The recombinant YAC is introduced into *TRP1*$^{-}$ *URA3*$^{-}$ yeast cells by protoplast transformation; transformed cells are selected by plating them onto the minimal medium: only those cells are able to grow on this medium that have correctly constructed YAC containing one left and one right arm of each chromosome. Recombinant clones are identified due to the insertional inactivation of *SUP4* detected by a simple colour test: recombinant colonies are white, while nonrecombinant ones are red.

The *TEL* sequence of the vector is not the complete telomeric sequence, but it contains enough of this sequence to be able to support the creation of complete telomere once the YAC is inside a yeast cell. Thus a YAC is a shuttle vector that is propagated in circular form in *E. coli* and is used for cloning in yeast in a linear form. When a YAC is less than about 20 kb, the centromeric function is unable to control copy number during mitosis so that several copies of YAC accumulate per yeast cell. The centromeric function improves in YACs of 50 kb or more; YACs of 150 kb or more behave like regular yeast chromosomes. YACs are the predominant vector system used for cloning of very large fug to 100-1,400 kb) DNA segments for mapping of complex eukaryotic chromosomes. YACs are reported to suffer from many problems, including chimerism, tedious steps in YAC library construction and low yields of YAC insert DNA. The yeast genes present in different yeast vectors can

become integrated into the host genome; this is *called permanent transformation.* It generally occurs through homologous recombination between the gene present in a vector *(e.g., LEU2) and* that present in the yeast chromosomes *(e.g., LEU2$^-$).* Rarely, the gene may become inserted at a random chromosome site. The homologous recombination may occur by regular crossing over or it may involve gene conversion (a non-reciprocal recombination). Vectors have been devised for high frequency stable transformation; such vectors are introduced in yeast cells in linear form and contain at their both ends sequences that are homologous to those found at the target site (where the gene present in the vector is to be integrated) in the yeast genome. Such vectors permit integration of any specified-DNA sequence at the desired site in yeast genome, *i.e.,* they allow site-specific transformation (= integration) of genes.

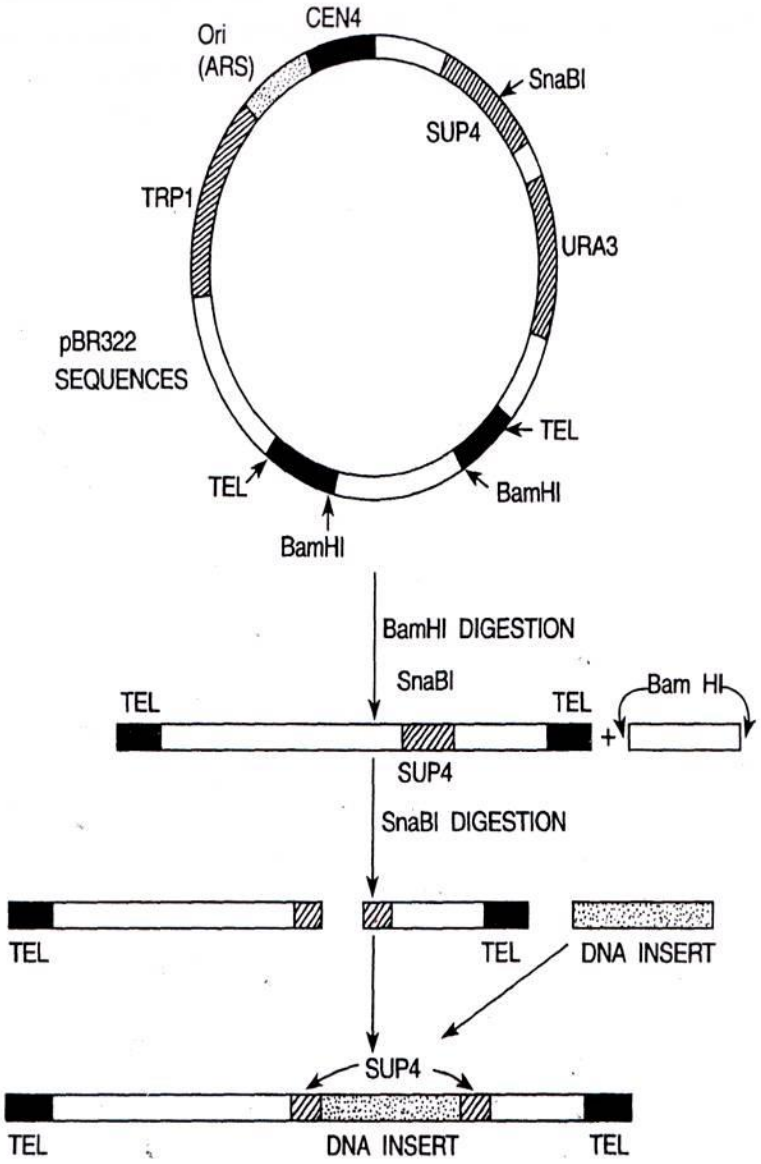


Fig. 2.24. Organization of YAC vector *p*YAC3, which is cloned in *E. coli* as a circular vector, but as a linear vector in yeast. *TEL*, telomeric sequences; *TRP1* and *URA3*, selectable marker genes for yeast; *ori* (*ARS*), origin of replication for yeast; *CEN4*, centromeric sequence from chromosome 4 of yeast; *SUP4*, a selectable marker into which DNA insert is placed. Enzymes *Bam*HI and *Sna*BI are used in combnation although here they are shown to be used sequentially.

==========

- **Uptake of DNA (Transfection methods)**

Several approaches have been used for introduction of DNA into animals cells / embryos, which are as follows

1. ***The calcium chloride treatment method***
2. ***Electroporation***
3. ***protoplast fusion***
4. ***Liposomes / lipofection***

1. The calcium chloride treatment method.

In this approach, the DNA preparation to be used for transfection is first dissolved in a phosphate buffer. Calcium chloride solution is then added to the DNA solution; this leads to the formation of insoluble calcium phosphate, which co-precipitates with the DNA. The calcium phosphate-DNA precipitate is added to the cells to be transfected. The precipitate particles are taken in by the cells by phagocytosis (Fig. 7.6). Initially, 1-2% of the cells were transfected by this approach. But the procedure has now been modified to obtain transfection of upto 20% of the cells. In a small proportion of the transfected cells, the DNA becomes integrated into the cell genome producing stable or permanent transfection. This general approach can be applied to virtually all mammalian cells, and a very large number of cells can be treated with little effort. But many cell lines do not like the calcium phosphate precipitate adhering to their surfaces or to their substrate (the surface of culture vessels).

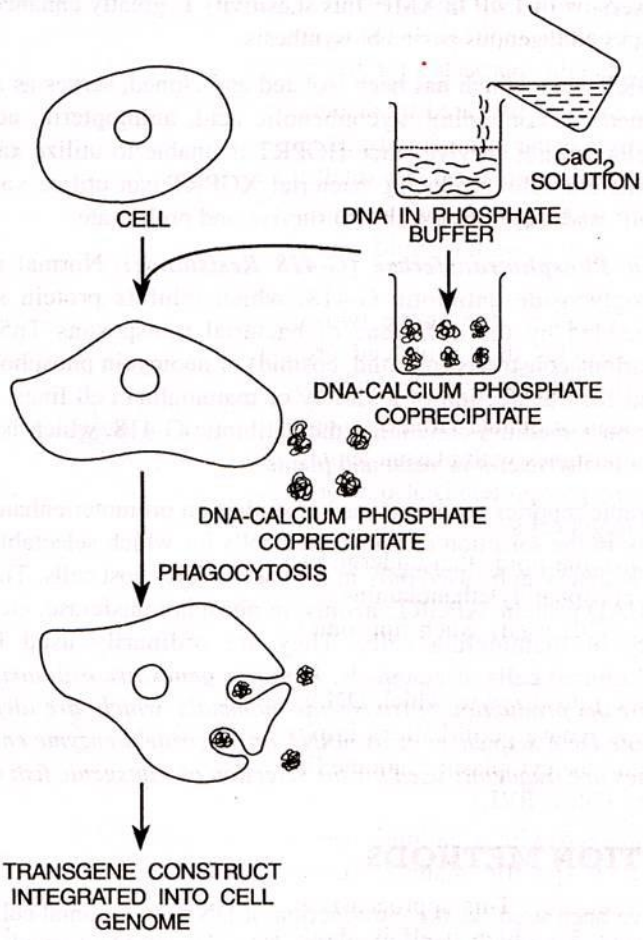


Fig. 7.6. A schematic representation of transfection by calcium phosphate precipitation.

2. Electroporation

In this approach, transfection mixture containing cells and DNA is exposed for a very brief period (few milliseconds) to a very high voltage gradient *(e.g.*, 4,000-8,000 V/cm). This induces transient pores in the cell membranes through which DNA seems to enter the cells. Treatment of cells with colcemid before they are electroporated increases the frequency of transfection. This is most likely due to the arrest of cells at metaphase and the associated absence of nuclear envelope or to an unusual permeability of the plasma membranes. Linearized DNA is far more efficient in transfection than circular supercoiled DNA. Electroporation technique has a general applicability, and many animal cell types that could not be transfected by other approaches were successfully transfected by this approach.

3. Protoplast fusion

The techniques for protoplast fusion are pretty well refined and highly effective for almost all the systems. A number of strategies have been used to induce fusion between protoplasts of different strains/species; of these the following three (Fig. 8.10) have been relatively more successful. Protoplasts of desired strains/species are mixed in almost equal proportion; generally, they are mixed while still suspended in the enzyme mixture. The protoplast mixture is then subjected to high pH (10.5) and high Ca^{2+} concentration (50 m mol/ lit) at 37°C for about 30 min **(high pH-high Ca^{2+} treatment).** This technique is quite suitable for some species, while for some others it may be toxic.

Polyethylene glycol (PEG) induced protoplast fusion is the most commonly used as it induces reproducible high frequency fusion accompanied with low toxicity to most cell types. The protoplast mixture is treated with 28-50% PEG (MW 1,500-6,000) for 15-30 min, followed by gradual washing of the protoplasts to remove PEG; protoplast fusion occurs during the washing. The washing medium may be alkaline (pH 9-10) and contain a high Ca^{2+} ion concentration (50 m mol / lit); this approach is a combination of PEG and high pH-high Ca^{2+} treatments, and is usually more effective than either

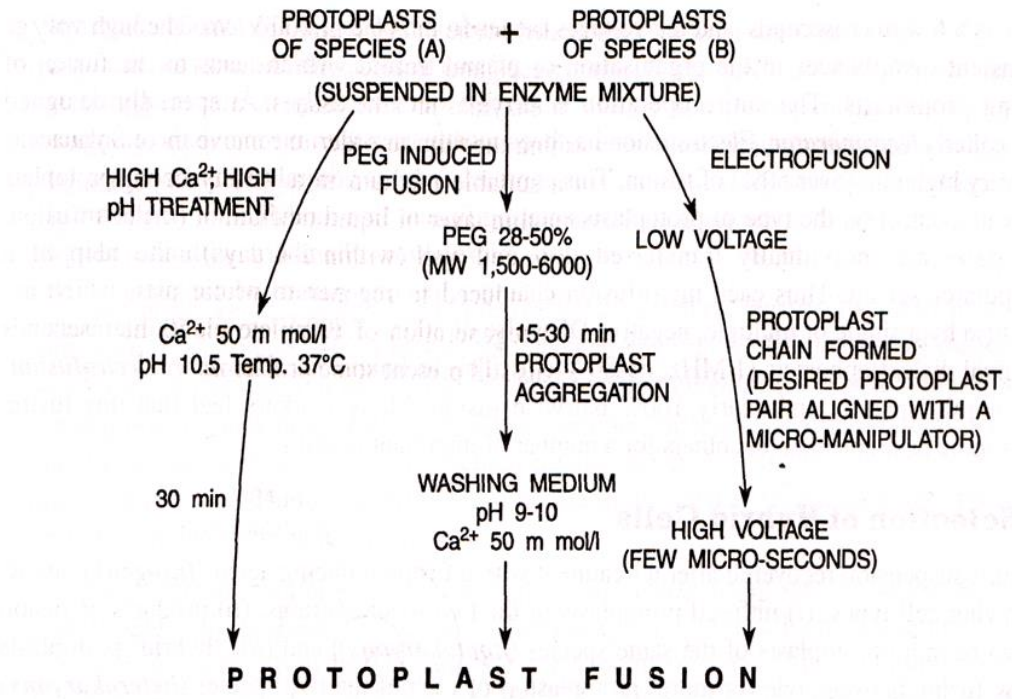


Fig. 8.10. A schematic representation of the 3 most successful protoplast fusion strategies.

treatment alone. PEG is negatively charged and may bind to cation like Ca^{2+}, which, in turn, may bind to the negatively charged molecules present in plasma lemma; they can also bind to cationic molecules of plasma membrane. During the washing process, PEG molecules may pull out the plasma lemma components bound to them. This would disturb plasma lemma organisation and may lead to the fusion of protoplasts located close to each other (Fig. 8.11).

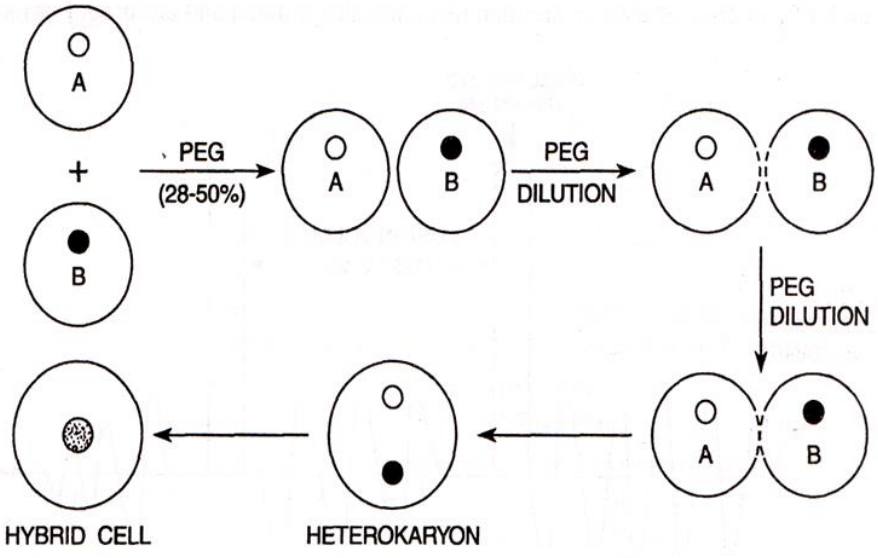


Fig. 8.11. PEG-induced protoplast fusion. Protoplasts are first brought close together (aggregation) by PEG. Fusion occurs during PEG dilution due to disturbances created in plasma membrane.

The above fusion techniques are nonselective in that they induce fusion between any two or more protoplasts. A more selective and less drastic approach is the ***electrofusion technique,*** which utilizes low voltage nonuniform alternating electric current pulses to bring the protoplasts in close contact (Fig. 8.12). Fusion of protoplasts is brought about by a short pulse of high voltage. The duration of

high voltage is a few microseconds, and the voltages ranges from 500 to 1,000 V/cm. The high voltage creates transient disturbances in the organisation of plasma lemma, which leads to the fusion of neighbouring protoplasts. The entire operation is carried out manually in a specially designed equipment, called electroporator. Electrofusion has been mostly used with the members of Solanaceae often with very high rate (over 50%) of fusion. This approach induces general fusion among protoplats and there is no control on the type of protoplasts entering fusion. In a modification of electrofusion, protoplast pairs are individually transferred into microfusion chambers with the help of a micromanipulator set up. Thus each microfusion chamber has one pair of protoplasts, which are induced to fuse by a single or multiple, negative DC-pulse of 800 to 1,800 V/cm for 50 microseconds — after mutual dielectrophoresis (1 MHz; 65-80 V/cm). This technique is called microelectrofusion; it leads to highly specific and nearly 100% pair wise fusion. Many workers feel that this fusion technique is more desirable than the others for a number of important reasons

4. Liposomes / lipofection

The delivery of DNA into cells using liposomes is called ***lipofection.*** ***Liposomes*** are small vesicles prepared from a suitable lipid. Initially, nonionic lipids were used for preparing liposomes so that DNA had to be introduced within the vesicles following specific encapsidation procedures. The use of cationic lipids for the construction of liposomes is a distinct advantage as DNA spontaneously and efficiently complexes with these liposomes making encapsidation procedures unnecessary. The cationic liposomes have a single lipid bilayer membrane (unilamellar), and they bind to the cells efficiently. Probably they fuse with the plasma membrane and thereby deliver the DNA (complexed with them) into the cells, which brings about transfection. Considerable work has been done on lipofection due to its potential application for targetting genes to specific human tissues for gene therapy. There are at least eight approaches for preparing liposomes for DNA delivery.

1. Usually, liposomes are prepared by dispersion of a phospholipid like phosphatidyl choline (PC) in water by mechanical methods, like

sonication, which tend to destroy DNA. DNA of up to 1 kb has been incorporated into small sonicated liposomes.

2. However, other techniques allow the entrapment of large DNA sequences into liposomes, *e.g.*, exposure of the anionic lipid phosphatidyl serine (PS) to Ca^{2+}, and two-phase techniques. Liposomes prepared by the above two approaches are phagocytosed by the cell. The phagocytosis vesicles thus produced ordinarily fuse with lysosomes leading to DNA degradation and low transfection frequencies.
3. The fusion protein of Sendai virus is incorporated into the liposome membrane; this enables the fusion of liposomes with plasma lemma and, thereby, a direct delivery of DNA into the cytoplasm. A receptor protein is also incorporated, which permits a controlled delivery of the DNA into the target cells.
4. The use of ionizable lipids that undergo phase change in response to the pH of cytoplasm, *e.g.*, dioleoyl phosphatidylethanolamine (DOPE), to construct liposomes (usually, a mixture of DOPE and PC is used). Such liposomes release DNA into the cytoplasm once they are phagocytosed.
5. Use of cationic liposomes to which DNA binds on the outside by electrostatic attraction. These liposomes cause perturbations in plasma membrane due to which they fuse and the DNA enters into the cytoplasm. Cationic liposomes are available commercially (marketed as 'Lipofectin' by Gibco-BRL).
6. DNA is complexed with a cationic peptide, *e.g.*, gramicidin, which interacts with lipid membranes in a specific manner. Incorporation of DOPE in the complex improves transfection frequency. This approach is about 10 times more efficient than cationic lipid liposomes (item 5), which itself is about 10 times more efficient than pH-sensitive or ionizable liposomes (item 4).
7. The liposomes may be targetted to cell surface receptors by the incorporation of ligand proteins into the liposome membrane. This is possible with the liposomes containing DNA inside the vesicle but not with cationic lipid and cationic peptide-DNA complexes. It has been recently proposed to use biotinylated bisanthracycline (which intercalates in double-stranded DNA) for attachment of specific ligand proteins via avidin to the DNA for delivering cationic lipid/peptide-complexed DNA into specific target cells.

8. Animal virus particles are enclosed in lipoprotein envelopes. Viral envelopes, separated from capsid proteins and viral genomes, have been used to package the desired DNA; these particles are called virosomes. Virosomes have been used to transfect cells in vitro, but in most cases the virosomes end up in lysosomes that results in degradation of up to 80% of the DNA. Sendai viral envelope has a fusion glycoprotein (F-protein) and a HN protein (haernagglutinin-neuraminidase protein). Sendai virus envelope has been reconstituted after removal of the HN protein since this protein targets the envelope to virtually all cell types. This reconstituted envelope contains only F-protein; therefore, it is called F-virosome. The F-protein mediates the fusion of viral envelope with cell plasma membrane. In addition, it confers on F-virosomes a very high specificity for the liver parenchyma cells since they contain receptors specific for the F-protein. F-virosomes have been used to deliver the reporter genes chloramphenicol actyltransferase and firefly luciferase specifically into liver parenchyma of mice. The transgenes showed multiple copy random integration, were expressed producing mRNAs and proteins and have been maintained for 6 months. F-virosomes deliver the DNA directly into the cytoplasm via fusion with cell plasma membrane, and appear 10-12 times more efficient than cationic liposomes in addition to being liver-specific in delivery. F-virosomes also have potential in targeted drug delivery to the liver cells.

Lipofection is the method of choice for transfection of mammalian cells in vitro. It has also been used to deliver DNA into live animals by direct injection or intravenous injection. Cationic liposomes have been used for intravenous or intratracheal injection in mice for the expression of marker genes in lungs. Targetted delivery has also been demonstrated by incorporating specific ligand proteins into the liposome membranes. Attempts are being made to deliver the cystic fibrosis gene via nasal or bronchial tissue for stimulating cytotoxic T-lymphocyte response in human patients.

The mechanism of movement of the DNA from cell cytoplasm into the nucleus is not known. In some cases, DNA movement to the nucleus is greatly facilitated by making DNA constructs that are capable of cytoplasmic translation, and by combining them with RNA polymerase or with a gene that produces RNA polymerase.

- **Selection of recombinant clones by Blue script / White script screening**

The selection system of pUC18 is based on the *lacZ'* gene encoded in the plasmid. Remember from our earlier discussion on the *lac* operon that the *lacZ* gene gives rise to the enzyme β-galactosidase. This enzyme is used by the bacteria to cleave lactose into glucose and galactose. But more interesting things can be achieved with β-galactosidase. If, rather than galactose, you give it a synthetic substrate called 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), then the enzyme hydrolyzes it to give rise to a blue product. It is on this colorful reaction that the selection system of pUC18 is based. If you have understood this much, we will go ahead and see the entire system and how it works.

⇒ The entire selection system of pUC18 revolves around two genes present in this plasmid - *lacZ'* and *lacI* - and the fact that the MCS is incorporated into the *lacZ'* gene without interfering with its expression.

⇒ *lacZ'* gene codes for a protein which all by itself cannot function as the enzyme β-galactosidase. But the host cell chromosome expresses another protein which can combine with *lacZ'* protein product to give rise to a fully functional β-galactosidase.

⇒ The protein product of *lacI* gene is a repressor and does not allow *lacZ'* gene to be transcribed. Therefore, β-galactosidase function is expressed in the host cell only when the cells are grown in the presence of isopropyl-β-D-thiogalactopyranoside (IPTG) which is an inducer of *lacZ'* transcription. IPTG binds to the repressor and does not allow it to inhibit transcription of *lacZ'* gene.

⇒ This apparently tasteless system is converted into a colorful scheme by a simple biochemical jugglery. The biochemical jugglery involves growing the cells not in the presence of lactose (which is the normal substrate for β-galactosidase) but in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), which is a synthetic substrate. Of course IPTG must also be present in the medium so that the β-galactosidase function is present. Another chemical which is present in the medium is the antibiotic ampicillin. Now let's see what happens.

⇒ When you transform the cells with pUC18 vector without the insert DNA, some cells take up the plasmid and the others don't. When these cells are grown in the medium given above (IPTG, X-Gal, and ampicillin besides the buffer and other necessary factors), the cells which do not contain the plasmid cannot grow because they are sensitive to ampicillin. The cells which contain the plasmid can grow. But not only do they grow, but they become blue! This is because when β-galactosidase hydrolyzes X-Gal, a blue product is formed which colors the cells and the colonies therefore appear blue.

⇒ Now let's see what happens to those cells which are transformed with pUC18 containing the target DNA. Remember that the MCS is built into the *lacZ'* gene. The presence of MCS does not disturb the proper expression of this gene, but if an insert is incorporated into this MCS, it does interfere with the expression. This is because the presence of target DNA changes the frame of reading. The cells transformed with plasmid-target DNA construct therefore do not express a functional β-galactosidase and therefore do not become blue. Thus, colonies transformed with this construct are white in color.

⇒ To sum up, the selection system is simple and clearcut and involves a single step. Transform the host cells and watch the color of the colonies. Blue means no target DNA, white means presence of target DNA. Cells not transformed at all do not grow.

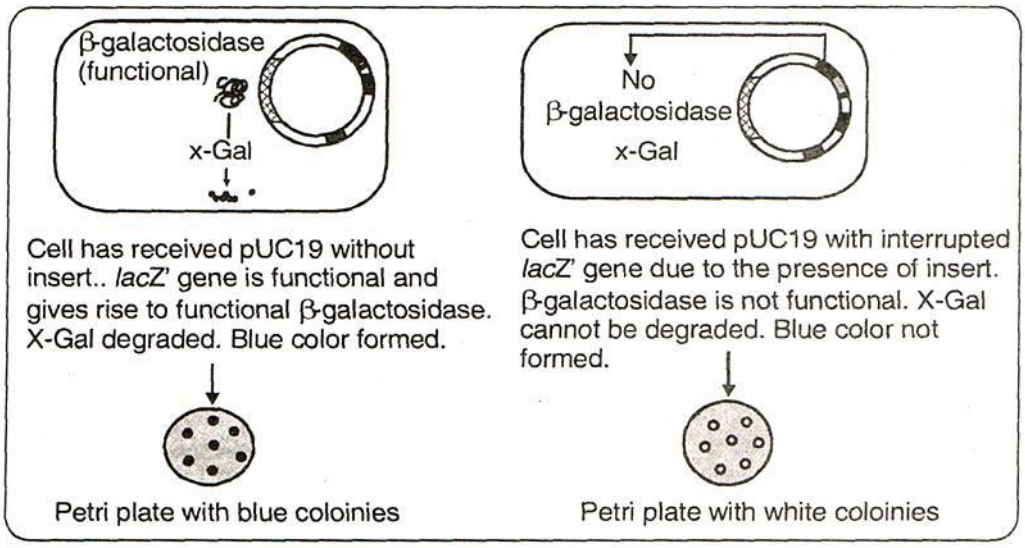


Figure 17.11. Interrupted *lacZ'* cannot give rise to a functional β-galactosidase. Thus the colonies which receive a target DNA-pUC19 construct are white in color while the colonies which receive recircularized plasmid are blue in color.

Blue/White Selection