

# Cloning Vector System

## **Cloning Vectors:**

For the cloning of any molecule of DNA it is necessary for the DNA to be incorporated into a cloning vector. These are DNA elements that may be stably maintained and propagated in a host organism for which the vector has replication functions.

A typical host organism is a bacterium, such as *Escherichia coli*, that grows and divides rapidly. Any vector with a replication origin in *E. coli* will replicate efficiently (together with any incorporated DNA). Thus, any DNA cloned into a vector will enable the amplification of the inserted foreign DNA fragment and also allow any subsequent analysis to be undertaken. In this way the cloning process resembles the PCR, although there are some major differences between the two techniques. By cloning, it is possible not only to store a copy of any particular fragment of DNA but also to produce unlimited amounts of it.

The vectors used for cloning vary in their complexity, ease of manipulation, selection and the amount of DNA sequence they can accommodate (the insert capacity). Vectors have in general been developed from naturally occurring molecules such as bacterial plasmids, bacteriophages or combinations of the elements that make them up, such as cosmids.

For gene library constructions there is a choice and trade-off between various vector types, usually related to the ease of the manipulations needed to construct the library and the maximum size of foreign DNA insert of the vector. Thus vectors with the advantage of large insert capacities are usually more difficult to manipulate.

**Although there are many more factors to be considered, which are indicated in the following treatment of vector systems:**

- 1. Plasmids,**
- 2. Bacteriophage *λ* as a Cloning Vector,**
- 3. Cosmids,**
- 4. Fosmids,**
- 5. Phagemid,**
- 6. Shuttle Vectors and**
- 7. Artificial Chromosomes.**

### **1. Plasmids:**

Plasmids are naturally occurring; circular, extra chromosomal DNA molecules. Natural strains of the common colon bacterium *Escherichia* isolated from various sources harbour diverse

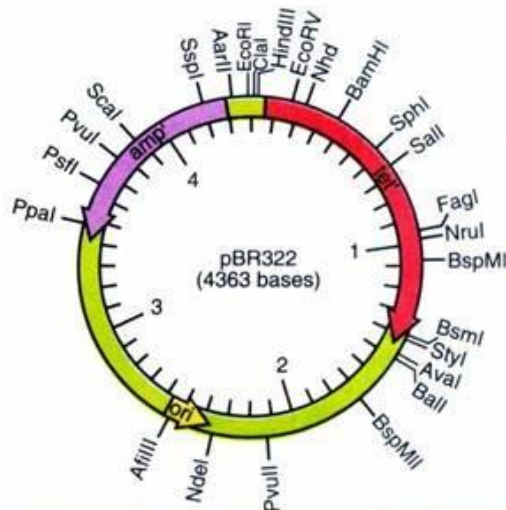
plasmids. Often these plasmids carry genes specifying novel metabolic activities that are advantageous to the host bacterium.

These activities range from catabolism of unusual organic substances to metabolic functions that endow the host cells with resistance to antibiotics, heavy metals, or bacteriophages. Plasmids that are able to perpetuate themselves in *E. coli*, the bacterium favoured by bacterial geneticists and molecular biologists, have become the darlings of recombinant DNA technology.

Because restriction endonuclease digestion of plasmids can generate fragments with overlap- ping or “sticky” ends, artificial plasmids can be constructed by ligating different fragments together. Such artificial plasmids were among the earliest recombinant DNA molecules. These recombinant molecules can be autonomously replicated, and hence propagated, in suitable bacterial host cells, provided they still possess a site signalling where DNA replication can begin (also called origin of replication or *ori* sequence).

### I. Plasmids as Cloning Vectors:

The idea arose that “foreign” DNA sequences could be inserted into artificial plasmids and that these foreign sequences would be carried into *E. coli* and propagated as part of the plasmid. That is, these plasmids could serve as cloning vectors to carry genes. (The word **vector** is used here in the sense of “**a vehicle or carrier.**”) Plasmids useful as cloning vectors possess three common features: a replicator, a selectable marker, and a cloning site (Fig. 4.1).



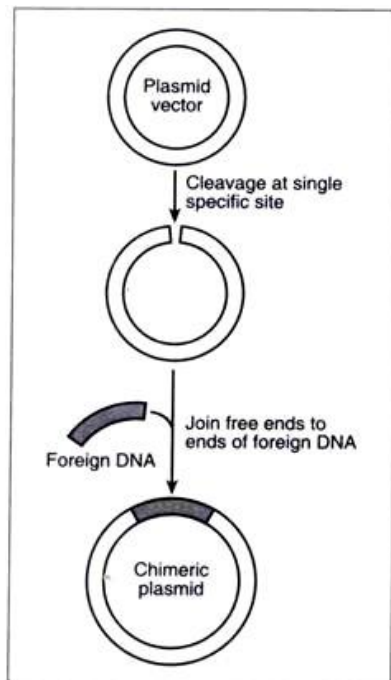
**Fig. 4.1:** One of the first widely used cloning vectors, the plasmid *pBR322*. This 4363-bp plasmid contains an origin of replication (*ori*) and genes encoding resistance to the drugs ampicillin (*amp<sup>r</sup>*) and tetracycline (*tet<sup>r</sup>*). The locations of restriction endonuclease cleavage sites are indicated.

A **replicator** is an origin of replication, or ori. The **selectable marker** is typically a gene conferring resistance to an antibiotic. Only those cells containing the cloning vector will grow in the presence of the antibiotic. Therefore, growth on antibiotic containing media “selects for” plasmid-containing cells.

Typically, the cloning site is a sequence of nucleotides representing one or more restrictions in endonuclease cleavage sites. Cloning sites are located where the insertion of foreign DNA neither disrupts the plasmid’s ability to replicate nor inactivates essential markers.

## II. Virtually any DNA Sequence can be cloned:

Nuclease cleavage at a restriction site opens, or linearizes, the circular plasmid so that a foreign DNA fragment can be inserted. The ends of this linearized plasmid are joined to the ends of the fragment so that the circle is closed again, creating a recombinant plasmid (Fig. 4.2). Recombinant plasmids are hybrid DNA molecules consisting of plasmid DNA sequences plus inserted DNA elements (called inserts).



**Fig. 4.2:** Foreign DNA sequences can be inserted into plasmid vectors by opening the circular plasmid with a restriction endonuclease. The ends of the linearized plasmid DNA are then joined with the ends of a foreign sequence, reclosing the circle to create a chimeric plasmid.

Such hybrid molecules are also called chimeric constructs or chimeric plasmids (The term chimera is borrowed from mythology and refers to a beast composed of the body and head of a

lion, the heads of a goat and a snake, and the wings of a bat.). The presence of foreign DNA sequences does not adversely affect replication of the plasmid, so chimeric plasmids can be propagated in bacteria just like the original plasmid. Bacteria often harbour several hundred copies of common cloning vectors per cell.

Hence, large amounts of a cloned DNA sequence can be recovered from bacterial cultures. The enormous power of recombinant DNA technology stems in part from the fact that virtually any DNA sequence can be selectively cloned and amplified in this manner. DNA sequences that are difficult to clone include inverted repeats, origins of replication, centromeres, and telomeres. The only practical limitation is the size of the foreign DNA segment: most plasmids with inserts larger than about 10 kbp are not replicated efficiently.

Bacterial cells may harbour one or many copies of a particular plasmid, depending on the nature of the plasmid replicator. That is, plasmids are classified as high copy number or low copy number. The copy number of most genetically engineered plasmids is high (200 or so), but some are lower.

**Table 4.1:** List of restriction endonucleases with their recognition sites

*Restriction Endonucleases*

About 1000 restriction enzymes have been characterized. They are named by *italicized* three-letter codes, the first one is a capital letter denoting the genus of the organism of origin, while the next two letters are the abbreviation of the particular species. Because prokaryotes often contain more than one restriction enzymes, the various representatives are assigned letter and number codes as they are identified. Thus, *EroRI* is the initial restriction endonuclease isolated from *Escherichia coli*, strain R. With one exception (*NciI*), all known type II restriction endonucleases generated fragments with 5'-PO<sub>4</sub> and 3'-OH ends.

Enzyme	Common Isoschizomers	Recognition Sequence	Compatible Cohesive Ends
<i>AtuI</i>		AG↓CT	Blunt
<i>ApyI</i>	<i>AtuI, EcoRII</i>	CC↓(A/T)GG	
<i>AsuII</i>		TT↓CGAA	<i>ClaI, HpaII, TaqI</i>
<i>AvaI</i>		G↓PyCGPuG	<i>SaI, XhoI, XmaI</i>
<i>AvrII</i>		C↓CTAGG	
<i>BalI</i>		TGG↓CCA	Blunt
<i>BamHI</i>		G↓GATCC	<i>BclI, BglII, MboI, Sau3A, XhoI</i>
<i>BclI</i>		T↓GATCA	<i>BamHI, BglII, MboI, Sau3A, XhoI</i>
<i>BglII</i>		A↓GATCT	<i>BamHI, BclI, MboI, Sau3A, XhoI</i>
<i>BstEII</i>		G↓GTNACC	
<i>BstXI</i>		CCANNNNN↓NTGG	
<i>ClaI</i>		AT↓CGAT	<i>AccI, AclI, AsyI, HpaII, TaqI</i>
<i>DdeI</i>		C↓TNAG	
<i>EcoRI</i>		G↓AATTC	
<i>EcoRII</i>	<i>AtuI, ApyI</i>	↓CC(A/T)GG	
<i>FnuDII</i>	<i>ThaI</i>	CG↓CG	Blunt
<i>HaeI</i>		(A/T)GG↓CC(T/A)	Blunt
<i>HaeII</i>		PuGCGC↓Py	
<i>HaeIII</i>		GG↓CC	Blunt
<i>HincII</i>		GTPy↓PuAC	Blunt
<i>HindIII</i>		A↓AGCTT	
<i>HpaI</i>		GTT↓AAC	Blunt
<i>HpaII</i>		C↓CGG	<i>AccI, AclI, AsuII, ClaI, TaqI</i>
<i>KpnI</i>		GGTAG↓G	<i>BamHI, BclI, BglII, XhoI</i>
<i>MboI</i>	<i>Sau3A</i>	↓GATC	
<i>MspI</i>		C↓CGG	
<i>MslI</i>		TGC↓GCA	Blunt
<i>NciI</i>		GC↓GGCCGC	

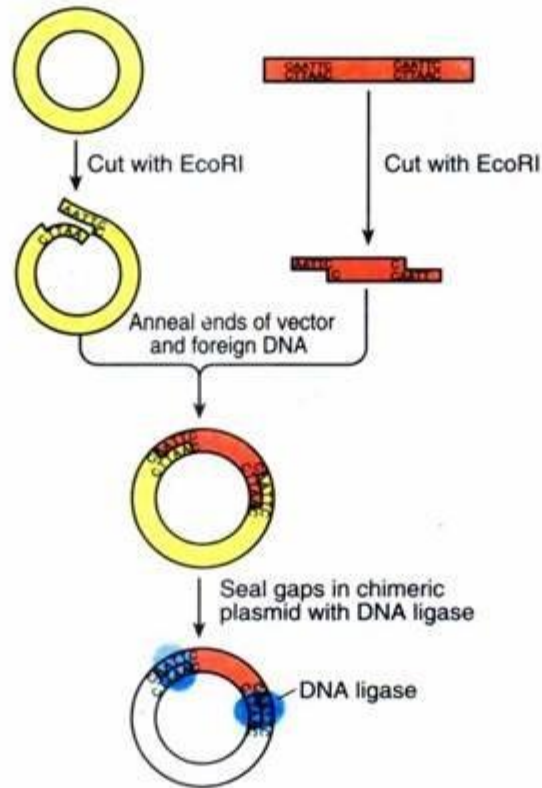
Enzyme	Common Isoschizomers	Recognition Sequence	Compatible Cohesive Ends
<i>Pst</i> I		CTGCA↓G	
<i>Sac</i> I	<i>Sst</i> I	GAGCT↓C	
<i>Sal</i> I		G↓TCGAC	<i>Ava</i> I, <i>Xho</i> I
<i>Sau</i> 3A		↓GATC	<i>Bam</i> HI, <i>Bcl</i> I, <i>Bgl</i> II, <i>Mbo</i> I, <i>Xho</i> II
<i>Sfi</i> I		GGCCNNNN↓NGGCC	
<i>Sma</i> I	<i>Xma</i> I	CCC↓GGG	Blunt
<i>Sph</i> I		GCATG↓G	
<i>Sst</i> I	<i>Sac</i> I	GAGCT↓G	
<i>Taq</i> I		T↓CGA	<i>Acc</i> I, <i>Acy</i> I, <i>Asu</i> II, <i>Clal</i> , <i>Hpa</i> II
<i>Xba</i> I		T↓CTAGA	
<i>Xho</i> I		C↓TCGAG	<i>Ava</i> I, <i>Sal</i> I
<i>Xho</i> II		( <sup>A</sup> ↓GATG( <sup>T</sup> )	<i>Bam</i> HI, <i>Bcl</i> I, <i>Bgl</i> II, <i>Mbo</i> I, <i>Sau</i> 3A
<i>Xma</i> I	<i>Sma</i> I	C↓CCGGG	<i>Ava</i> I

### III. Construction of Chimeric Plasmids:

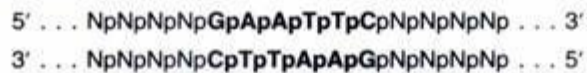
Creation of chimeric plasmids requires joining the ends of the foreign DNA insert to the ends of a linearized plasmid (Fig. 4.2). This ligation is facilitated if the ends of the plasmid and the insert have complementary, single stranded overhangs. Then these ends can base-pair with one another, annealing the two molecules together.

One way to generate such ends is to cleave the DNA with restriction enzymes that make staggered cuts; many such restriction endonucleases are available. For example, if the sequence to be inserted is an *Eco*RI fragment and the plasmid is cut with *Eco*RI, the single stranded sticky ends of the two DNAs can anneal (Fig. 4.3).

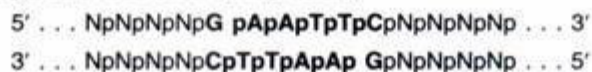




**Fig. 4.3:** Restriction endonuclease *EcoRI* cleaves double-stranded DNA. The recognition site for *EcoRI* is the hexameric sequence GAATTC:



Cleavage occurs at the G residue on each strand so that the DNA is cut in a staggered fashion, leaving 5'-overhanging single-stranded ends (sticky ends):



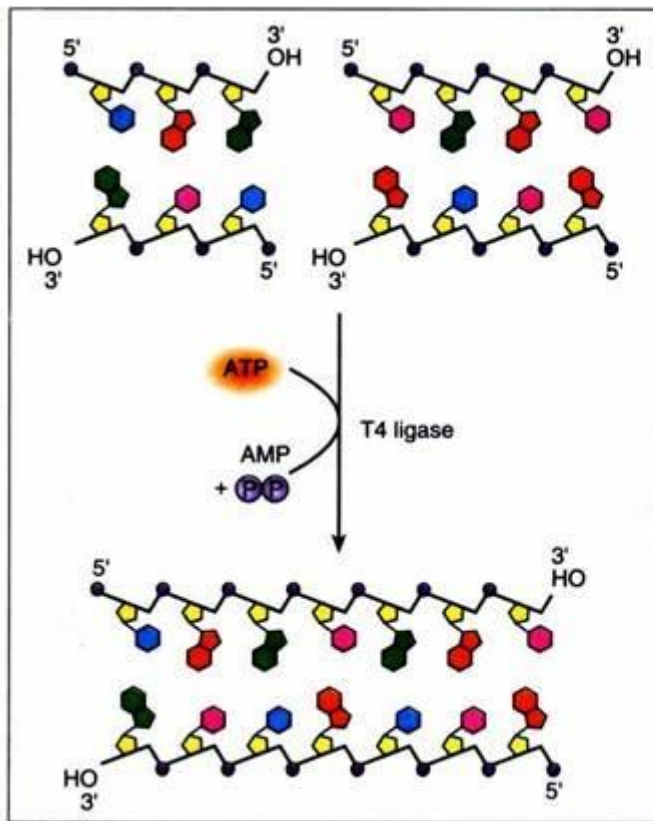
An *EcoRI* restriction fragment of foreign DNA can be inserted into a plasmid having an *EcoRI* cloning site by (a) cutting the plasmid at this site with *EcoRI*, annealing the linearized plasmid with the *EcoRI* foreign DNA fragment, and (b) sealing the nicks with DNA ligase.

The interruptions in the sugar- phosphate backbone of DNA can then be sealed with DNA ligase to yield a covalently closed, circular chimeric plasmid. DNA ligase is an enzyme that covalently links adjacent 3'-OH and 5'-PO<sub>4</sub> groups. An inconvenience of this strategy is that any pair of *EcoRI* sticky ends can anneal with each other.

o, plasmid molecules can re-anneal with themselves, as can the foreign DNA restriction fragments. These DNAs can be eliminated by selection schemes designed to identify only those bacteria containing chimeric plasmids. Blunt-end ligation is an alternative method for joining

different DNAs. This method depends on the ability of phage T4 DNA ligase to covalently join the ends of any two DNA molecules (even those lacking 3'- or 5'-overhangs) (Fig. 4.4).

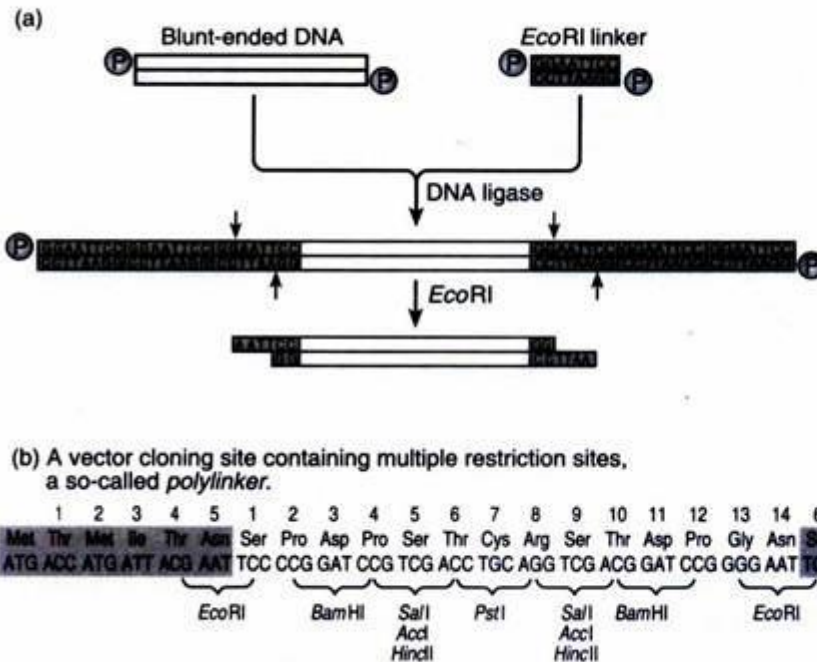
Some restriction endonucleases cut DNA so that blunt ends are formed. Because there is no control over which pair of DNAs are blunt end and ligated by T4 DNA ligase, strategies to identify the desired products must be applied.



**Fig. 4.4:** Blunt-end ligation using phage T4 DNA ligase, which catalyzes the ATP dependent ligation of DNA molecules. AMP and P<sub>i</sub> are by-products.

A great number of variations on these basic themes have emerged. For example, short synthetic DNA duplexes whose nucleotide sequence consists of little more than a restriction site can be blunt-end ligated onto any DNA. These short DNAs are known as linkers. Cleavage of the ligated DNA with the restriction enzyme then leaves tailor-made sticky ends useful in cloning reactions (Fig. 4.5). Similarly, many vectors contain a poly-linker cloning site, a short region of DNA sequence bearing numerous restriction sites.





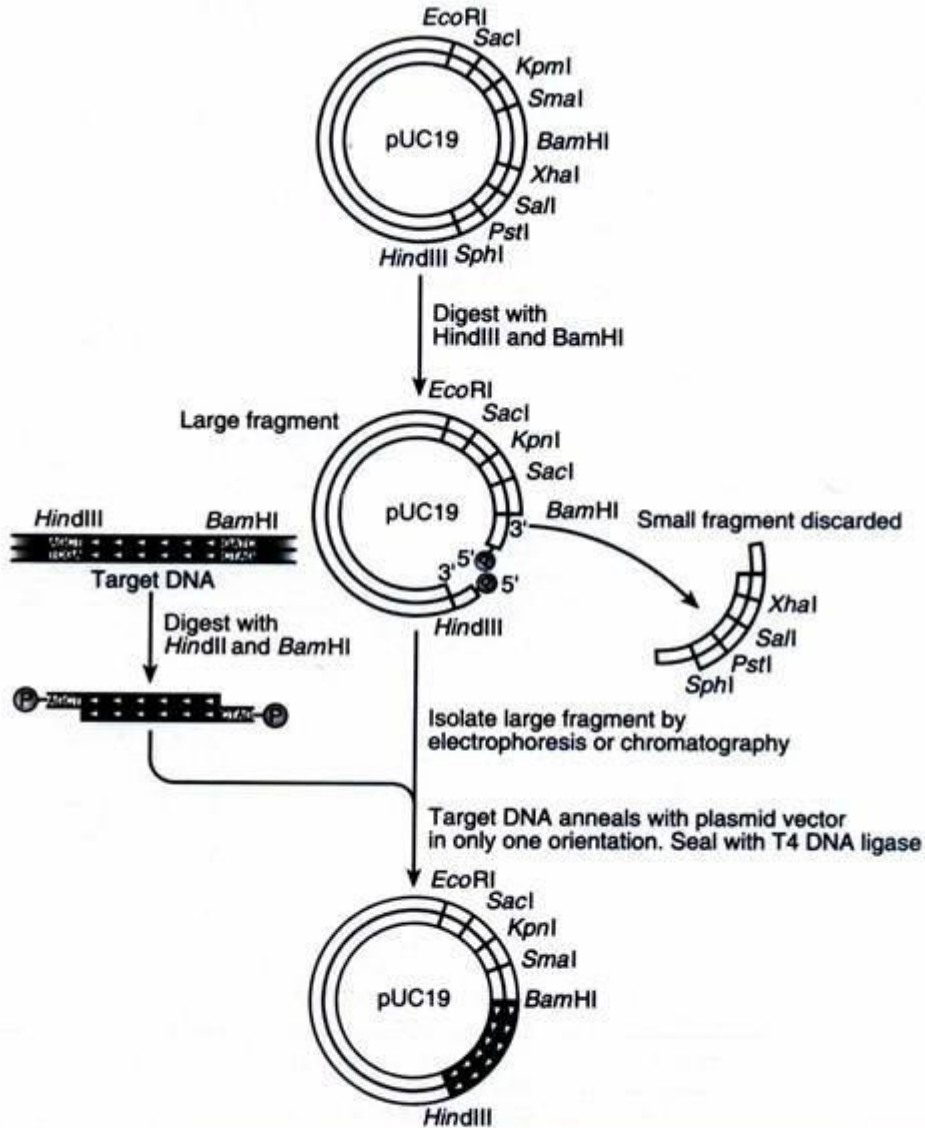
**Fig. 4.5:** (a) The use of linkers to create tailor-made ends on cloning fragments. Synthetic oligonucleotide duplexes whose sequences represent *EcoRI* restriction sites are blunt-end ligated to a DNA molecule using T4 DNA ligase. Note that the ligation reaction can add multiple linkers on each end of the blunt-ended DNA. *EcoRI* digestion removes all but the terminal one, leaving the desired 5'-overhangs. (b) Cloning vectors often have polylinkers consisting of a multiple array of restriction sites at their cloning sites, so restriction fragments generated by a variety of endonucleases can be incorporated into the vector. Note that the polylinker is engineered not only to have multiple restriction sites but also to have an uninterrupted sequence of codons, so this region of the vector has the potential for translation into protein. The sequence shown is the cloning site for the vectors M13mp7 and pUC7; the shaded amino acid residues are contiguous with the coding sequence of the *lacZ* gene carried by this vector.

#### IV. Promoters and Directional Cloning:

Note that the strategies discussed thus far create hybrids in which the orientation of the DNA insert within the chimera is random. Sometimes it is desirable to insert the DNA in a particular orientation. For example, an experimenter might wish to insert a particular DNA (a gene) in a vector so that its gene product is synthesized. To do this, the DNA must be placed downstream from a promoter.

A promoter is a nucleotide sequence lying upstream of a gene that controls expression of the gene. RNA polymerase molecules bind specifically at promoters and initiate transcription of adjacent genes, copying template DNA into RNA products.

One way to insert DNA so that it will be properly oriented with respect to the promoter is to create DNA molecules whose ends have different overhangs. Ligation of such molecules into the plasmid vector can only take place in one orientation, to give directional cloning (Fig. 4.6).



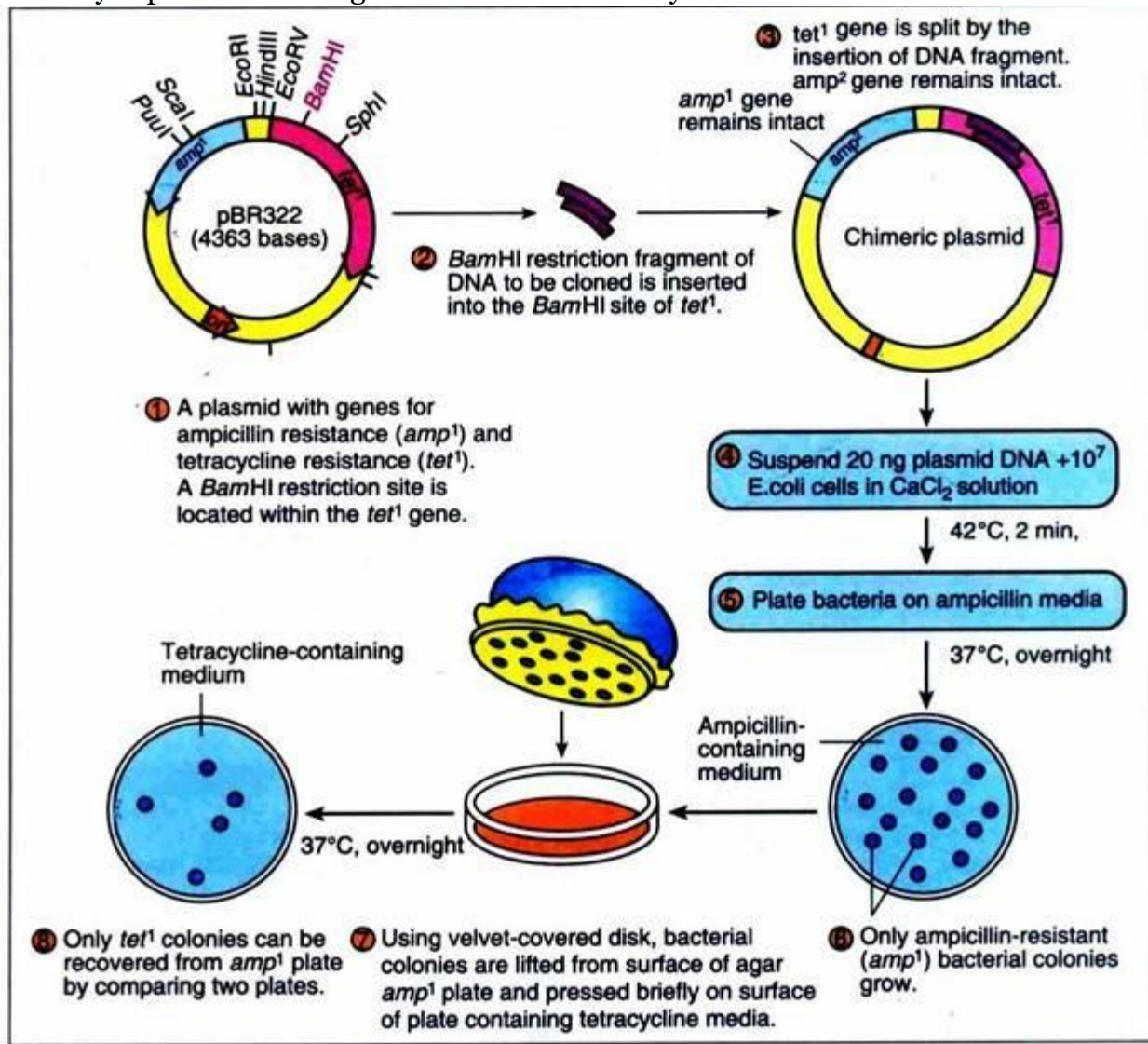
**Fig. 4.6:** Directional cloning. DNA molecules whose ends have different overhangs can be used to form chimeric constructs in which the foreign DNA can enter the plasmid in only one orientation. The foreign DNA is digested with two different restriction enzymes (*HindIII* and *BamHI*), and the plasmid is digested with the same two enzymes. Note that pUC19 has a polylinker or universal cloning site.

## V. Biologically Functional Chimeric Plasmids:

The first biologically functional chimeric DNA molecules constructed in vitro were assembled from parts of different plasmids in 1973 by Stanley Cohen, Annie Chang, Herbert Boyer, and Robert Helling. These plasmids were used to transform recipient *E. coli* cells (transformation means the uptake and replication of exogenous DNA by a recipient cell).

The bacterial cells were rendered somewhat permeable to DNA by  $\text{Ca}^{2+}$  treatment and a brief  $42^\circ\text{C}$  heat shock. Although less than 0.1% of the  $\text{Ca}^{2+}$  treated bacteria became competent for transformation following such treatment, transformed bacteria could be selected by their

resistance to certain antibiotics (Fig. 4.7). Consequently, the chimeric plasmids must have been biologically functional in at least two aspects: they replicated stably within their hosts and they expressed the drug resistance markers they carried.



**Fig. 4.7:** Vector. (1) Cleavage of pBR322 with restriction enzyme *Bam*H1, followed by (2) annealing and ligation of inserts generated by *Bam*H1 cleavage of some foreign DNA, (3) creates a chimeric plasmid. (4) The chimeric plasmid is then used to transform  $\text{Ca}^{2+}$ -treated heat-shocked *E. coli* cells, and the bacterial sample is plated on a Petri plate. (5) Following incubation of the petri plate overnight at 37°C, (6) colonies of *amp<sup>r</sup>* bacteria are evident. (7) Replica plating of these bacteria on plates of tetracycline-containing media (8) reveals which colonies are *tetr* and which are tetracycline sensitive (*tets*). Only the *tets* colonies possess plasmids with foreign DNA inserts.

In general, plasmids used as cloning vectors are engineered to be small, 2.5 kbp to about 10 kbp in size, so that the size of the insert DNA can be maximized. These plasmids have only a single origin of replication, so the time necessary for complete replication depends on the size of the plasmid.

Under selective pressure in a growing culture of bacteria, overly large plasmids are prone to delete any nonessential “genes,” such as any foreign inserts. Such deletion would thwart the purpose of most cloning experiments. The useful upper limit on cloned inserts in plasmids is about 10 kbp. Many eukaryotic genes exceed this size.

### **2. Bacteriophage $\lambda$ as a Cloning Vector:**

The genome of bacteriophage  $\lambda$  (lambda) is a 48.5 kbp linear DNA molecule that is packaged into the head of the bacteriophage. The middle one-third of this genome is not essential to phage infection, so  $\lambda$  phage DNA has been engineered so that foreign DNA molecules up to 16 kbp can be inserted into this region for cloning purposes.

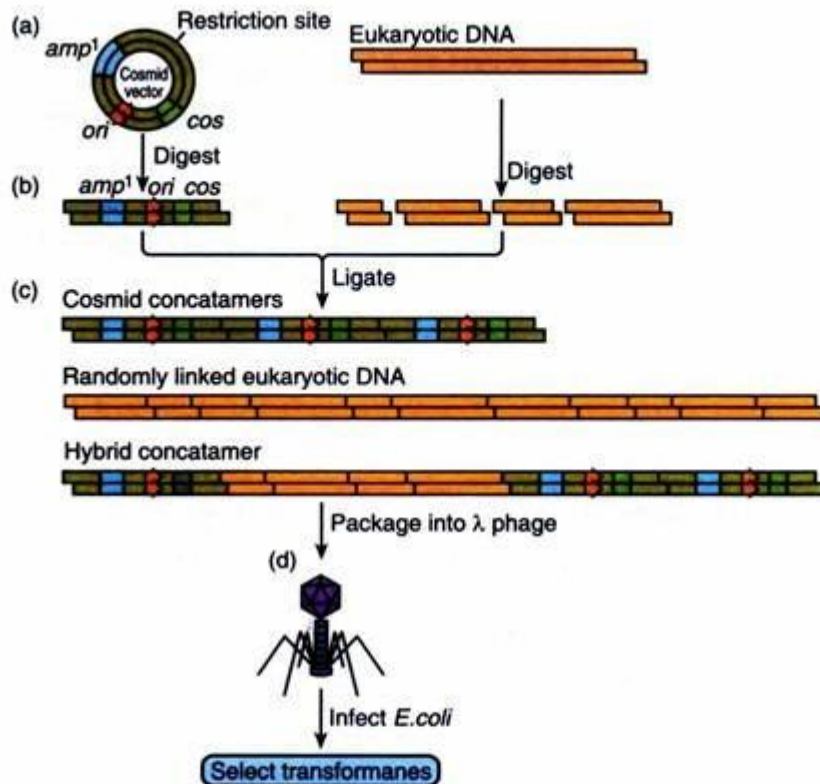
In **vitro** packaging systems are then used to package the chimeric DNA into phage heads which, when assembled with phage tails, form infective phage particles. Bacteria infected with these recombinant phage produce large numbers of phage progeny before they lyse, and large amounts of recombinant DNA can be easily purified from the lysate.

### **3. Cosmids:**

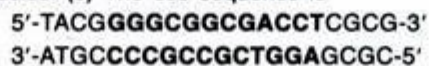
The DNA incorporated into phage heads by bacteriophage  $\lambda$  packaging systems must satisfy only a few criteria. It must possess a 14-bp sequence known as *cos* (which stands for cohesive end site) at each of its ends, and these *cos* sequences must be separated by no fewer than 36 kbp and no more than 51 kbp of DNA. Essentially any DNA satisfying these minimal requirements will be packaged and assembled into an infective phage particle.

Other cloning features such as an *ori*, selectable markers, and a poly-linker are joined to the *cos* sequence so that the cloned DNA can be propagated and selected in host cells. These features have been achieved by placing *cos* sequences on either side of cloning sites in plasmids to create cosmid vectors that are capable of carrying DNA inserts about 40 kbp in size (Fig. 4.8). Because cosmids lack essential phage genes, they reproduce in host bacteria as plasmids.





**Fig. 4.8:** Cosmid vectors for cloning large DNA fragments. (a) Cosmid vectors are plasmids that carry a selectable marker such as *amp<sup>r</sup>*, an origin of replication (*ori*), a polylinker suitable for insertion of foreign DNA, and (b) a *cos* sequence. Both the plasmid and the foreign DNA to be cloned are cut with a restriction enzyme, and the two DNAs are then ligated together. (c) The ligation reaction leads to the formation of hybrid concatamers, molecules in which plasmid sequences and foreign DNAs are linked in series in no particular order. The bacteriophage packaging extract contains the restriction enzyme that recognizes *cos* sequences and cleaves at these sites. (d) DNA molecules of the proper size (36 to 51 kbp) are packaged into phage heads, forming infective phage particles. (e) The *cos* sequence is



Endonuclease cleavage at the sites indicated by arrows leaves 12-bp cohesive ends.

#### 4. Fosmids:

Vector containing the single copy *E. coli* F-factor replicon, developed as an improved method for constructing libraries of cosmid-sized (approximately 40 kb) clones. The stability of inserts cloned into fosmid vectors has been shown to be substantially greater than in high copy vectors. Copy control fosmids, e.g., pCC1fos, contain both the *E. coli* F-factor replicon and the *oriV* high-copy origin of replication, thus providing the user the clone stability afforded by single-copy fosmid cloning and the high yields of DNA that can be realized from cosmid clones.

#### 5. Phagemid:

A phage-plasmid vector is able to replicate as single- or double-stranded DNA. Phagemids can be induced to produce phage particles containing single-stranded DNA. Example: pBluescript

series (Stratagene), which contains a filamentous f1 phage inter-genic region including the origin of replication.

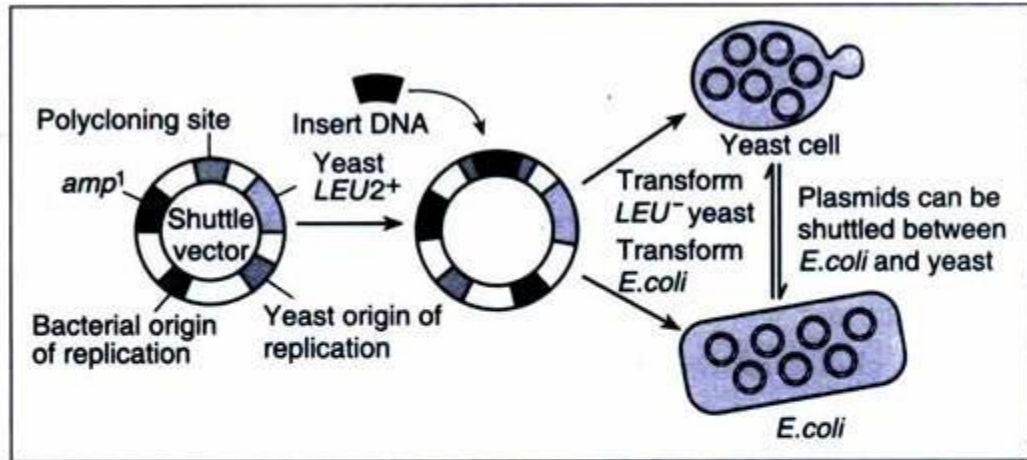
### **6. Shuttle Vectors:**

Plasmids used in eukaryotes require a eukaryotic origin of replication and marker genes that will be expressed by eukaryotic cells. At present plasmids are used for cloning in yeast and in plants. Although the yeast has a natural plasmid called the 2  $\mu$  circle, this is too large for use in cloning.

Plasmids such as yeast episomal plasmids (Yep) have been created by genetic manipulation using replication origins from the 2  $\mu$  circle, and by incorporating a gene that will complement a defective gene in the host yeast cell. Strain of such yeast with defective gene can be used as a selectable marker for the presence of that plasmid. Yeast, like bacteria, can be grown rapidly, hence well suited for use in cloning. Particular use has been the creation of shuttle vectors which has origin of replication for yeast and bacteria like E.coli.

Shuttle vectors are plasmids capable of propagating and transferring (“**shuttling**”) genes between two different organisms, one of which is typically a prokaryote (**E. coli**) and the other a eukaryote (for example, yeast). Shuttle vectors must have unique origins of replication for each cell type as well as different markers for selection of transformed host cells harbouring the vector (Fig. 4.9). Shuttle vectors have the advantage that eukaryotic genes can be cloned in bacterial hosts, yet the expression of these genes can be analyzed in appropriate eukaryotic backgrounds.





**Fig. 4.9:** A typical shuttle vector. This vector has yeast and bacterial origins of replication, *amp<sup>r</sup>* (ampicillin resistance gene for selection in *E. coli*) and *LEU2<sup>+</sup>*, a gene in the yeast pathway for leucine biosynthesis. The recipient yeast cells are *LEU2<sup>-</sup>* (defective in this gene) and thus require leucine for growth. *LEU2<sup>-</sup>* yeast cells transformed with this shuttle vector can be selected on medium lacking any leucine supplement.

### 7. Artificial Chromosomes:

DNA molecules 2 mega base pairs in length have been successfully propagated in yeast by creating yeast artificial chromosomes or YACs. Further, such YACs have been transferred into transgenic mice for the analysis of large genes or multi-genic DNA sequences in vivo, that is, within the living animal.

For these large DNAs to be replicated in the yeast cell, YAC constructs must include not only an origin of replication (known in yeast terminology as an autonomously replicating sequence or ARS) but also a centromere and telomeres.

Recall that centromeres provide the site for attachment of the chromosome to the spindle during mitosis and meiosis, and telomeres are nucleotide sequences defining the ends of chromosomes. Telomeres are essential for proper replication of the chromosome.