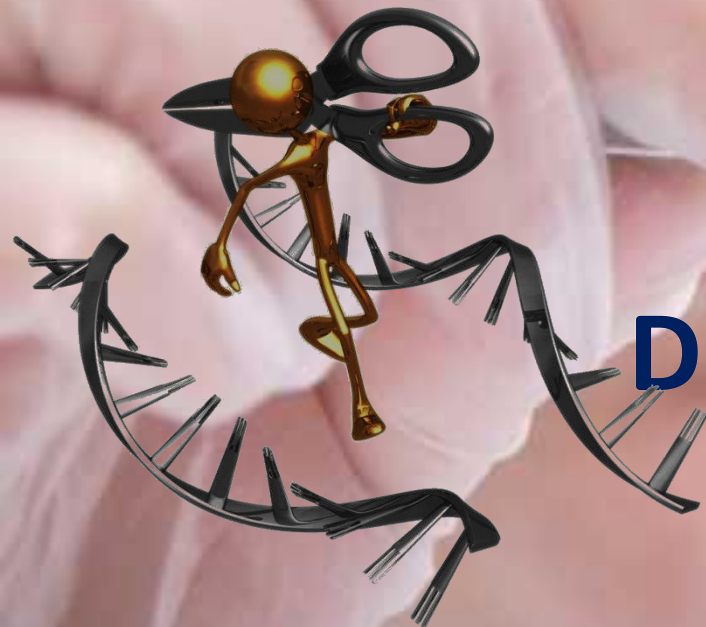
Genetic Engineering



Dr. S.K. Halder

What and Why?

- Genetic engineering, also called genetic modification, is the direct manipulation of an organism's genome using biotechnology.
- New DNA may be inserted in the host genome by first isolating and copying the genetic material of interest using molecular cloning methods to generate a DNA sequence, or by synthesizing the DNA, and then inserting this construct into the host organism.
- An organism that is generated through genetic engineering is considered to be a genetically modified organism (GMO).

Flash back….

- Genetic engineering as the direct manipulation of DNA by humans outside breeding and mutations has only existed since the 1970s.
- The term "genetic engineering" was first coined by Jack Williamson in his science fiction novel *Dragon's Island*, published in 1951, one year before DNA's role in heredity was confirmed by Alfred Hershey and Martha Chase, and two years before James Watson and Francis Crick showed that the DNA molecule has a double-helix structure.!!!
- The first GMOs: bacteria in 1973 and mice in 1974.
- Insulin-producing bacteria were commercialized in 1982 and genetically modified food has been sold since 1994.
- Glofish, the first GMO designed as a pet, was first sold in the United States December in 2003.

Table 14.1	Some Milestones in Biotechnology and Recombinant DNA Technology
1958	DNA polymerase purified
1970	A complete gene synthesized in vitro Discovery of the first sequence-specific restriction endonuclease and the enzyme reverse transcriptase
1972	First recombinant DNA molecules generated
1973	Use of plasmid vectors for gene cloning
1975	Southern blot technique for detecting specific DNA sequences
1976	First prenatal diagnosis using a gene-specific probe
1977	Methods for rapid DNA sequencing Discovery of "split genes" and somatostatin synthesized using recombinant DNA
1978	Human genomic library constructed
1979	Insulin synthesized using recombinant DNA First human viral antigen (hepatitis B) cloned
1981	Foot-and-mouth disease viral antigen cloned First monoclonal antibody-based diagnostic kit approved for use
1982	Commercial production by *E. coli* of genetically engineered human insulin Isolation, cloning, and characterization of a human cancer gene Transfer of gene for rat growth hormone into fertilized mouse eggs
1983	Engineered Ti plasmids used to transform plants
1985	Tobacco plants made resistant to the herbicide glyphosate through insertion of a cloned gene from *Salmonella* Development of the polymerase chain reaction technique
1987	Insertion of a functional gene into a fertilized mouse egg cures the shiverer mutation disease of mice, a normally fatal genetic disease
1988	The first successful production of a genetically engineered staple crop (soybeans) Development of the gene gun
1989	First field test of a genetically engineered virus (a baculovirus that kills cabbage looper caterpillars)
1990	Production of the first fertile corn transformed with a foreign gene (a gene for resistance to the herbicide bialaphos)
1991	Development of transgenic pigs and goats capable of manufacturing proteins such as human hemoglobin First test of gene therapy on human cancer patients
1994	The Flavr Savr tomato introduced, the first genetically engineered whole food approved for sale Fully human monoclonal antibodies produced in genetically engineered mice
1995	*Haemophilus influenzae* genome sequenced
1996	*Methanocaldococcus jannaschii* and *Saccharomyces cerevisiae* genomes sequenced
1997	Human clinical trials of antisense drugs and DNA vaccines begun; *E. coli* genome sequenced
1998	First cloned mammal (the sheep Dolly)
2002	*Plasmodium falciparum* genome sequenced
2003	Completion of the draft of the human genome
2005	Reconstruction of 1918 influenza virus

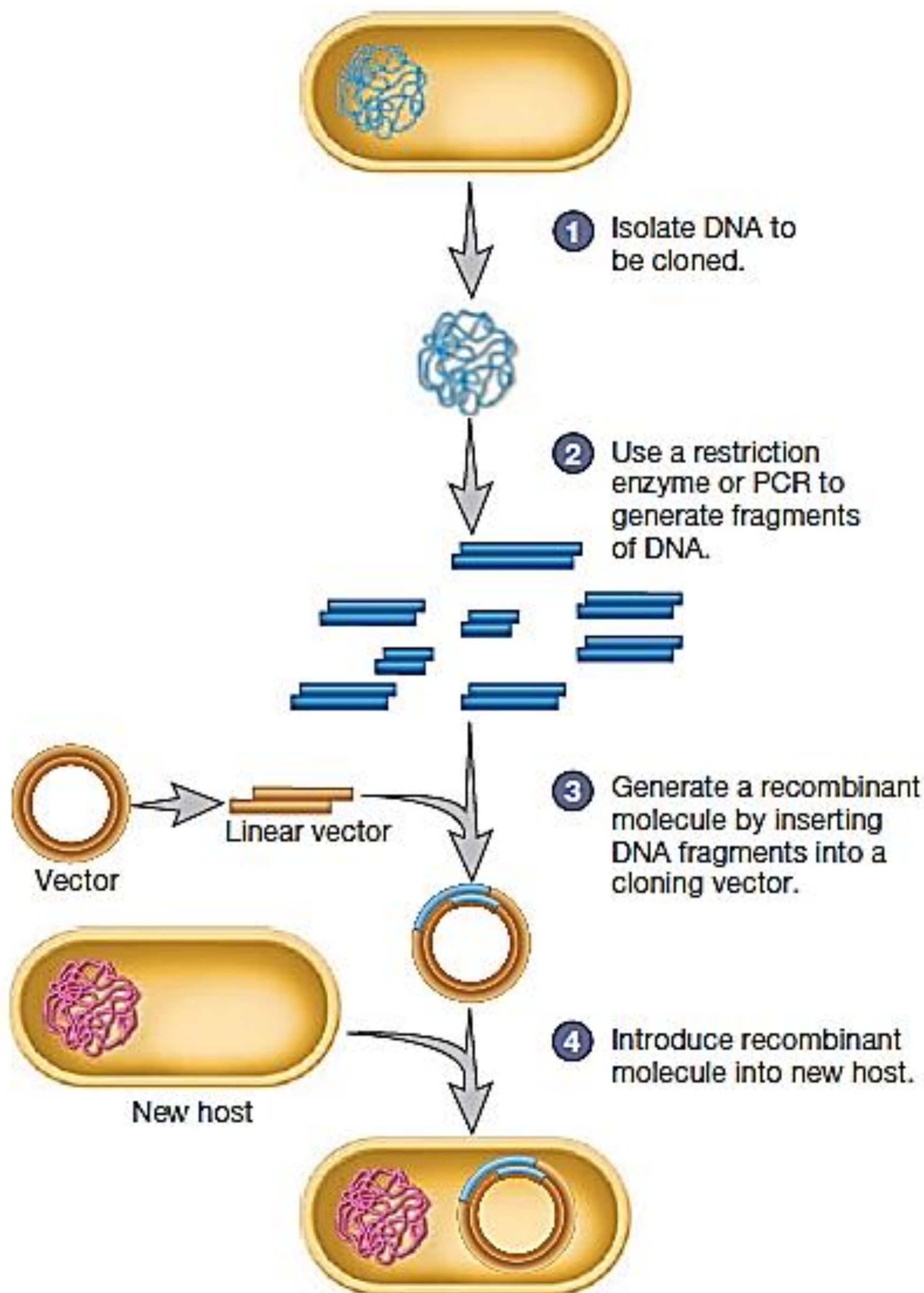
DNA cloning

Cloning *is the process of producing similar populations of genetically identical individuals that occurs in nature when organisms such as bacteria, insects or plants reproduce asexually. Cloning in biotechnology refers to processes used to create copies of DNA fragments (molecular cloning), cells (cell cloning), or organisms.*

- To obtain large amounts of pure DNA
  - Procedure
    - Isolate DNA
    - Use restriction enzymes to cut DNA
    - Ligate fragments into a cloning vector
    - Transform recombinant DNA into a host to replicate the DNA and pass copies into progeny.

Recombinant DNA Technology

- Currently it is relatively easy to cut out a specific piece of DNA, produce a large number of copies, (optional: determine its nucleotide sequence, slightly alter it) and then as a final step transfer it back into cell in.
- Intentional modification of organisms' genomes for practical purposes
  - Three goals
    - ***Eliminate undesirable phenotypic traits***
    - ***Combine beneficial traits of two or more organisms***
    - ***Create organisms that synthesize products humans need***



1 Isolate DNA to be cloned.
2 Use a restriction enzyme or PCR to generate fragments of DNA.
Vector
Linear vector
3 Generate a recombinant molecule by inserting DNA fragments into a cloning vector.
New host
4 Introduce recombinant molecule into new host.

Restriction Enzymes: Endonucleases

- Recognize a specific DNA sequence (restriction site).
- To cut DNA, all restriction enzymes make two incisions, once through each sugar-phosphate backbone (i.e. each strand) of the DNA helix.
- Break a phosphodiester linkage between a 3’ carbon and 5’ phosphate.
- The fragments generated by RE digestion is named restriction fragment.
- Over 3000 restriction enzymes have been studied in detail, and more than 600 of these are available commercially.

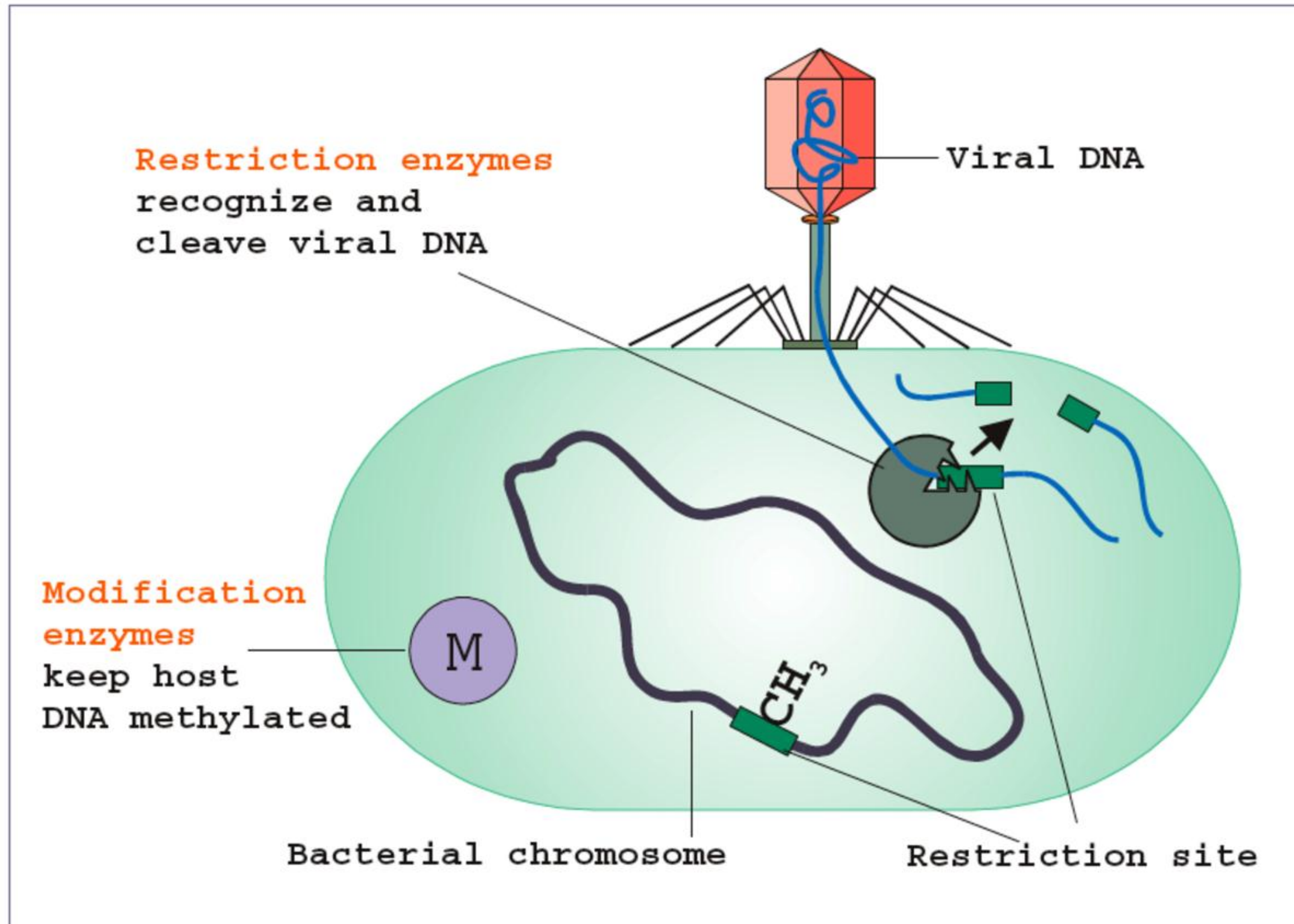
History of RE

- The term restriction enzyme originated from the studies of phage λ and the phenomenon of host-controlled restriction and modification of a bacterial virus
- The phenomenon was first identified in work done in the laboratories of Salvador Luria and Giuseppe Bertani in early 1950s
- It was found that a bacteriophage λ that can grow well in one strain of *Escherichia coli* (*E. coli* C), when grown in another strain, (*E. coli* K), its yields can drop significantly, by as much as 3-5 orders of magnitude.
- The restriction enzymes studied by Arber and Meselson were type I restriction enzymes which cleave DNA randomly away from the recognition site.
- In 1970, Hamilton O. Smith, Thomas Kelly and Kent Welcox isolated and characterized the first type II restriction enzyme, *Hin*dII, from the bacterium *Haemophilus influenzae*
- Daniel Nathans and Kathleen Danna showed that cleavage of simian virus 40 (SV40) DNA by restriction enzymes yielded specific fragments which can be separated using polyacrylamide gel electrophoresis, thus showing that restriction enzymes can be used for mapping of the DNA
- For their work in the discovery and characterization of restriction enzymes, the 1978 Nobel Prize for Physiology or Medicine was awarded to Werner Arber, Daniel Nathans, and Hamilton O. Smith.

- These enzymes are found in bacteria and archaea and provide a defense mechanism against invading viruses.
- Inside a prokaryote, the restriction enzymes selectively cut up *foreign* DNA in a process called *restriction*; while host DNA is protected by a modification enzyme (a methylase) that modifies the prokaryotic DNA and blocks cleavage.
- Together, these two processes form the **restriction modification system**.

Restriction Enzymes

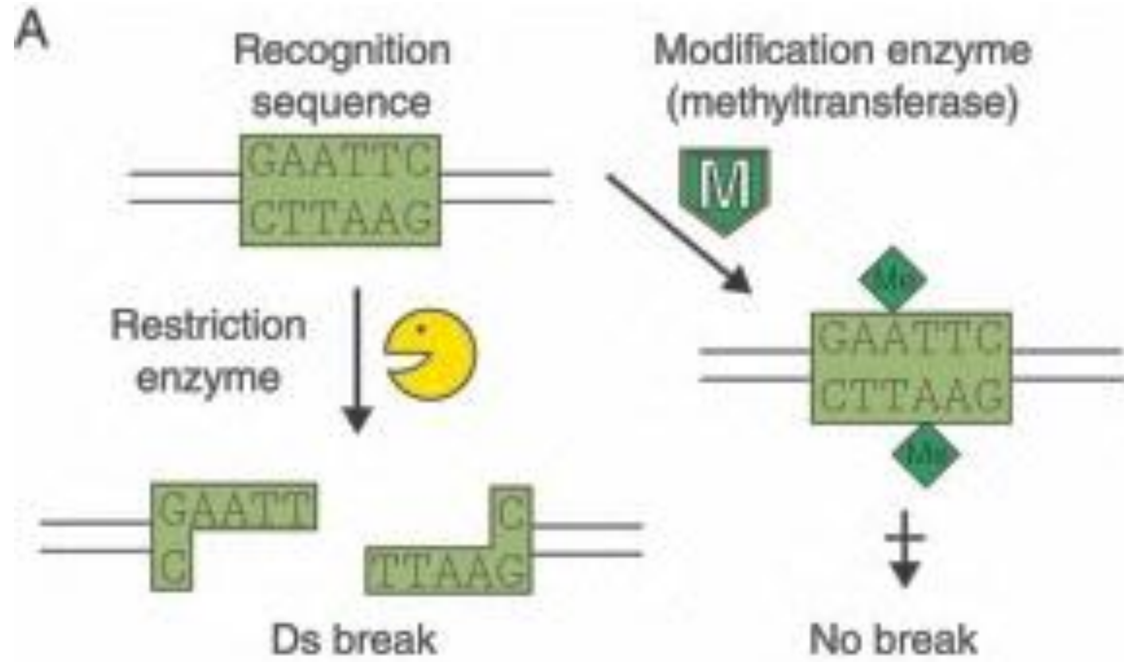
Bacterial defense against viral infection by restriction-modification complexes

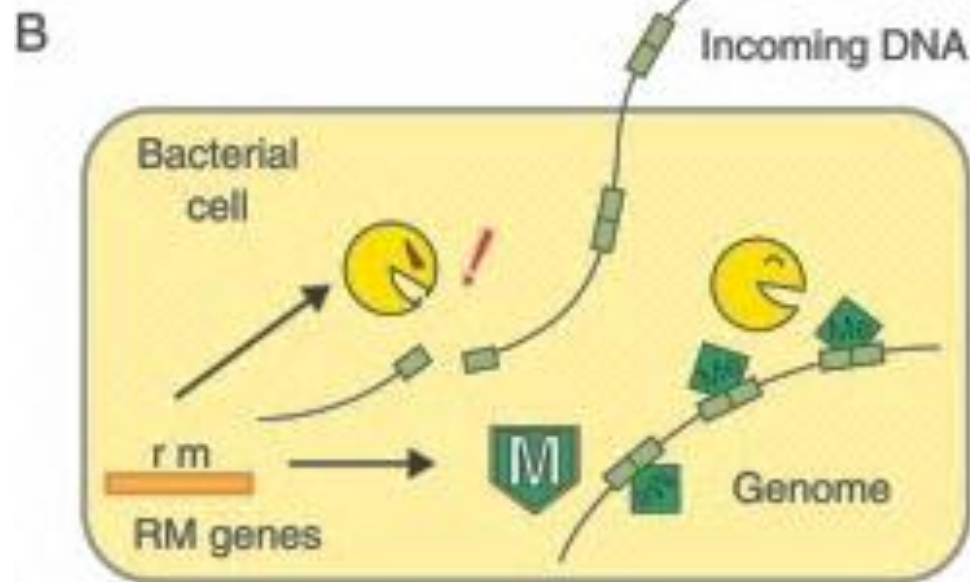


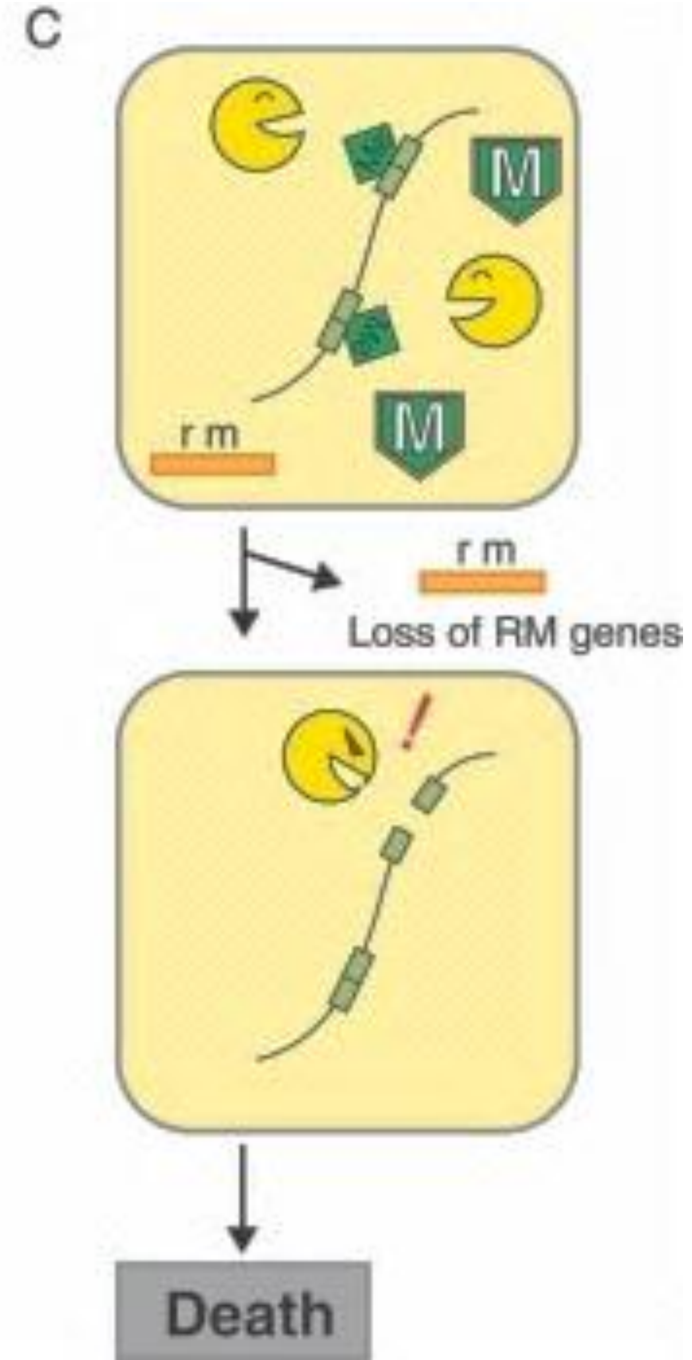
- Are denoted by three letter names derived from the bacterial strain they originate from.
- The recognize a sequence, usually are palindromes of 4-, 6- or 8-base pairs.
- Based on the probability, a specific short DNA sequence occur more frequently than a long one.
  - In 50% GC content, each base has a ¼ chance of occurring at a position.
  - The frequency of a particular restriction site is $(1/4)^n$.

Types

- **Type I** enzymes cleave at sites remote from recognition site (atleast 1000 bp apart); require both ATP and S-adenosyl-L-methionine to function; multifunctional protein with both Restriction and methylase activities.
- **Type II** enzymes cleave within or at short specific distances from recognition site; most require magnesium; single function (restriction) enzymes independent of methylase.
- **Type III** enzymes leave at sites a short distance from recognition site (about 25 bp apart); require ATP (but do not hydrolyse it); S-adenosyl-L-methionine stimulates reaction but is not required; exist as part of a complex with a modification methylase.
- **Type IV** enzymes target modified DNA, e.g. methylated, hydroxymethylated and glucosylhydroxymethylated DNA.

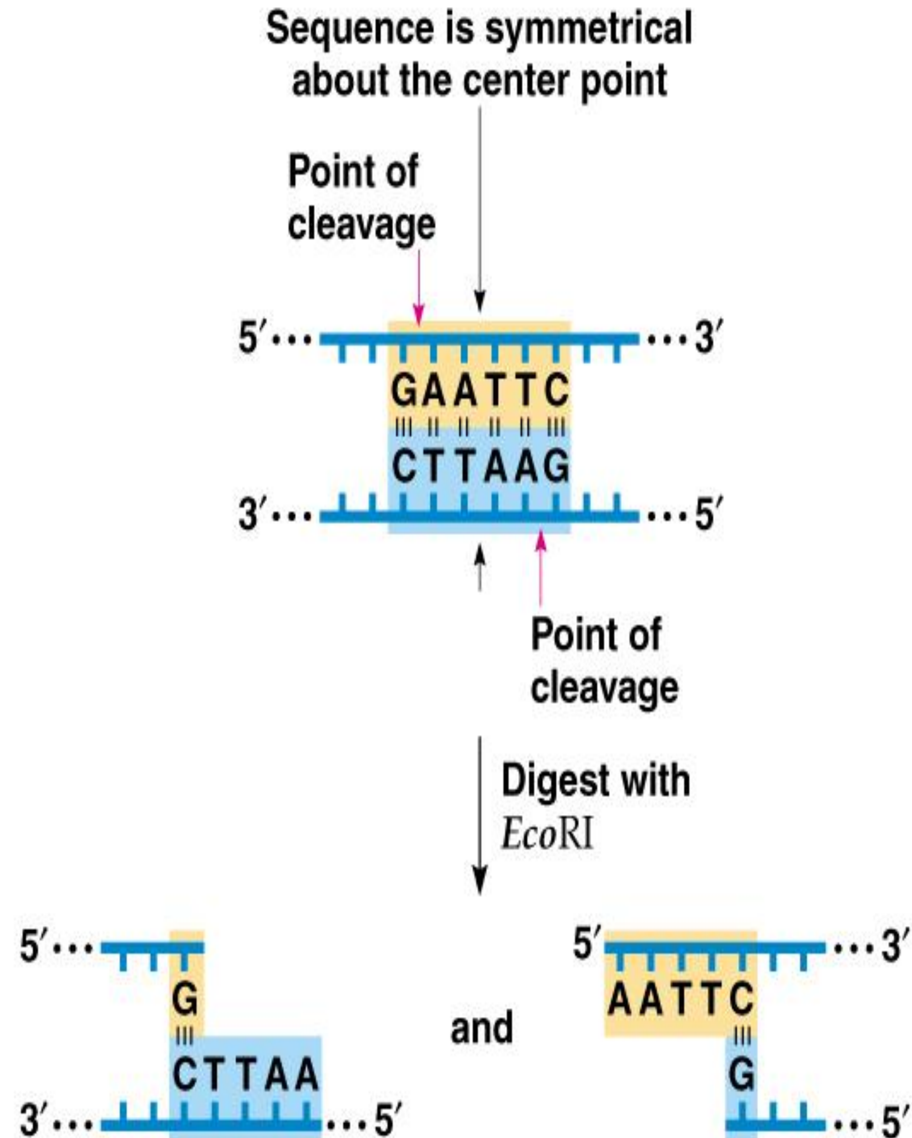

A
Recognition sequence
Modification enzyme (methyltransferase)
GAATTC
CTTAAG
M
Restriction enzyme
GAATT
C
C
TTAAG
Ds break
No break


B
Incoming DNA
Bacterial cell
r m
RM genes
M
Genome


C
M
M
r m
r m
Loss of RM genes
Death

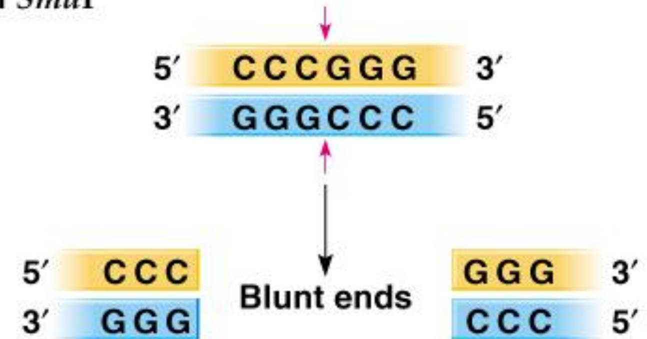
Type II

- They are a homodimer, with recognition sites are usually undivided and palindromic and 4–8 nucleotides in length.
- They recognize and cleave DNA at the same site, and they do not use ATP or AdoMet for their activity—they usually require only Mg^{2+} as a cofactor.
- They are divided into sub-groups

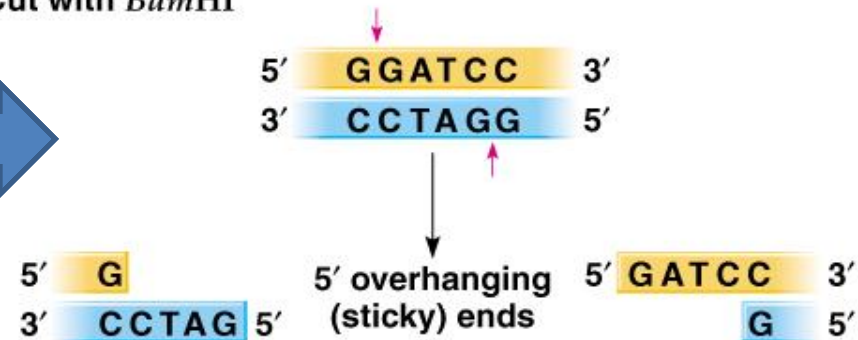


Restriction Enzyme II Sites

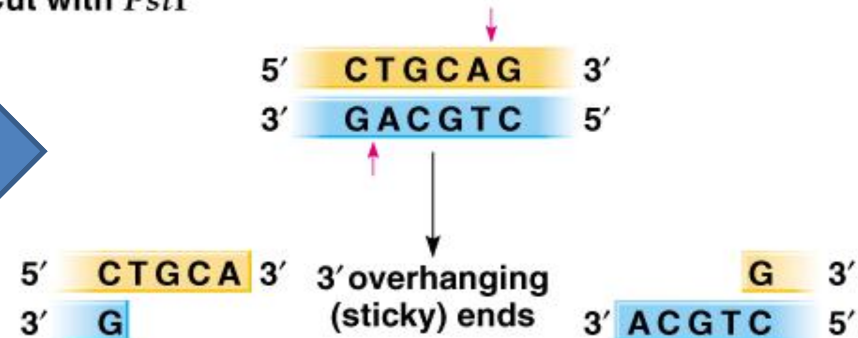
- Sma I (from *Serratia marcescens*) cuts a palindrome to give blunt ends.



- BamHI (from *Bacillus amyloliquefaciens H*) cuts to give a 5’ overhang.



- PstI (from *Providencia stuartii*) cuts to give a 3’ overhang.



Isoschizomers are pairs of restriction enzymes specific to the same recognition sequence. For example, SphI (CGTAC/G) and BbuI (CGTAC/G) are isoschizomers of each other.

An enzyme that recognizes the same sequence but cuts it differently is a **neoschizomer**. Neoschizomers are a specific type (subset) of isoschizomer. For example, SmaI(CCC/GGG) and XmaI (C/CCGGG) are neoschizomers of each other.

An enzyme that recognizes a slightly different sequence, but produces the same ends is an **isocaudomer**.

Significance: *This property of some isoschizomers allows identification of methylation state of the restriction site. For example, the restriction enzymes HpaII and MspI are isoschizomers, as they both recognize the sequence 5'-CCGG-3' when it is unmethylated. But when the second C of the sequence is methylated, only MspI can recognize it while HpaII cannot.*

Different types of enzymes involved in RDT

Type II restriction endonuclease

- Cleaves DNA at a specific base sequence

DNA ligase

- Binds two DNA molecules or fragments

DNA polymerase I (Klenow fragment)

- Fills single-stranded gaps in duplex DNA by stepwise addition of nucleotides to 3' ends

Reverse Transcriptase

- Makes a DNA copy of an RNA molecule

Polynucleotide Kinase

- Adds a phosphate to the 5'-OH end of a polynucleotide, to label it or permit ligation.

Terminal transferase

- Adds homopolymer tails to the 3'-OH ends of a linear duplex

Exonuclease III

- Removes nucleotide residues from the 3' ends of a DNA strand

Bacteriophage {lamda} exonuclease

- Removes nucleotides from the 5' ends of a duplex to expose 3' single-stranded ends

Alkaline phosphatase

- Removes terminal phosphates from the 5' end, the 3' end, or both

Nuclease S1

- Is an endonuclease enzyme that splits single-stranded DNA (ssDNA) and RNA into oligo- or mononucleotides

RNase H

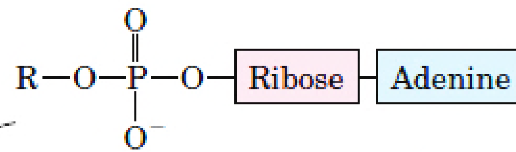
- Ribonuclease cleaves the 3’-O-P-5’ bond of RNA in a DNA/RNA duplex to produce 3’-hydroxyl and 5‘-phosphate terminated products.

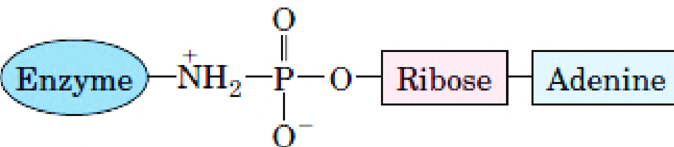
Some important enzymes for cloning

DNA Ligases

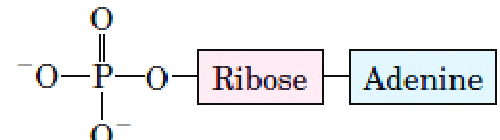
- DNA ligases close nicks in the phosphodiester backbone of DNA.
- Biologically, DNA ligases are essential for the joining of Okazaki fragments during replication, and for completing short-patch DNA synthesis occurring in DNA repair process.
- There are two classes of DNA ligases. The first uses NAD^+ as a cofactor and only found in bacteria. The second uses ATP as a cofactor and found in eukaryotes, viruses and bacteriophages. The smallest known ATP-dependent DNA ligase is the one from the bacteriophage T7.
- Recombinant DNA experiments require the joining of two different DNA segments or fragments in vitro. The cohesive ends generated by some RE will anneal themselves by forming hydrogen bonds. But the segments annealed thus are weak and do not withstand experimental conditions. To get a stable joining, the DNA should be joined by ligase.
- **DNA Ligase Mechanism**
  1. Formation of a covalent enzyme-AMP intermediate linked to a lysine side-chain in the enzyme.
  2. Transfer of the AMP nucleotide to the 5’ phosphate of the nicked DNA strand.
  3. Attack on the AMP-DNA bond by the 3’-OH of the nicked DNA sealing the phosphate backbone and resealing AMP.

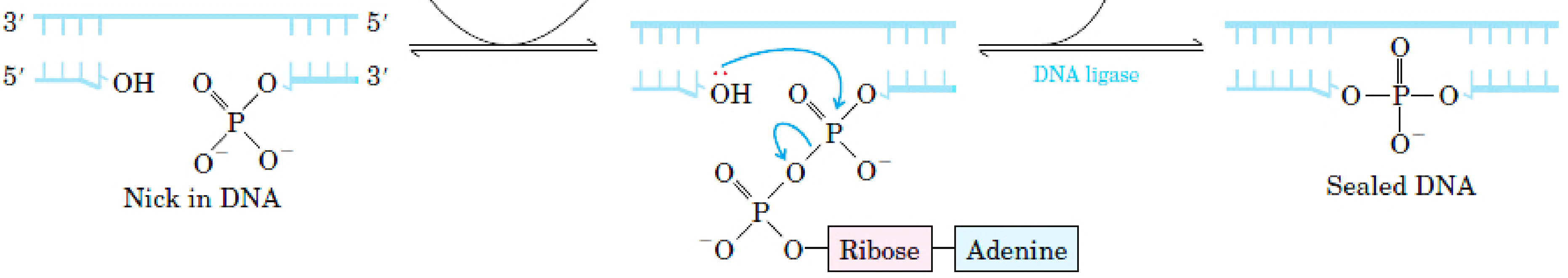
Enzyme
NH3+
DNA ligase
① Adenylylation of DNA ligase


R—O—P—O
O
O−
Ribose
Adenine
AMP from ATP (R = PPi)
or NAD+ (R = NMN)
PPi (from ATP)
or
NMN (from NAD+)


Enzyme
NH2—P—O
Ribose
Adenine
Enzyme-AMP
② Activation of 5′ phosphate in nick

Enzyme
NH3+


③ Displacement of AMP seals nick
−O—P—O
Ribose
Adenine
AMP


3′
5′
5′
3′
OH
Nick in DNA
DNA ligase
Ribose
Adenine
Sealed DNA

Alkaline phosphatase

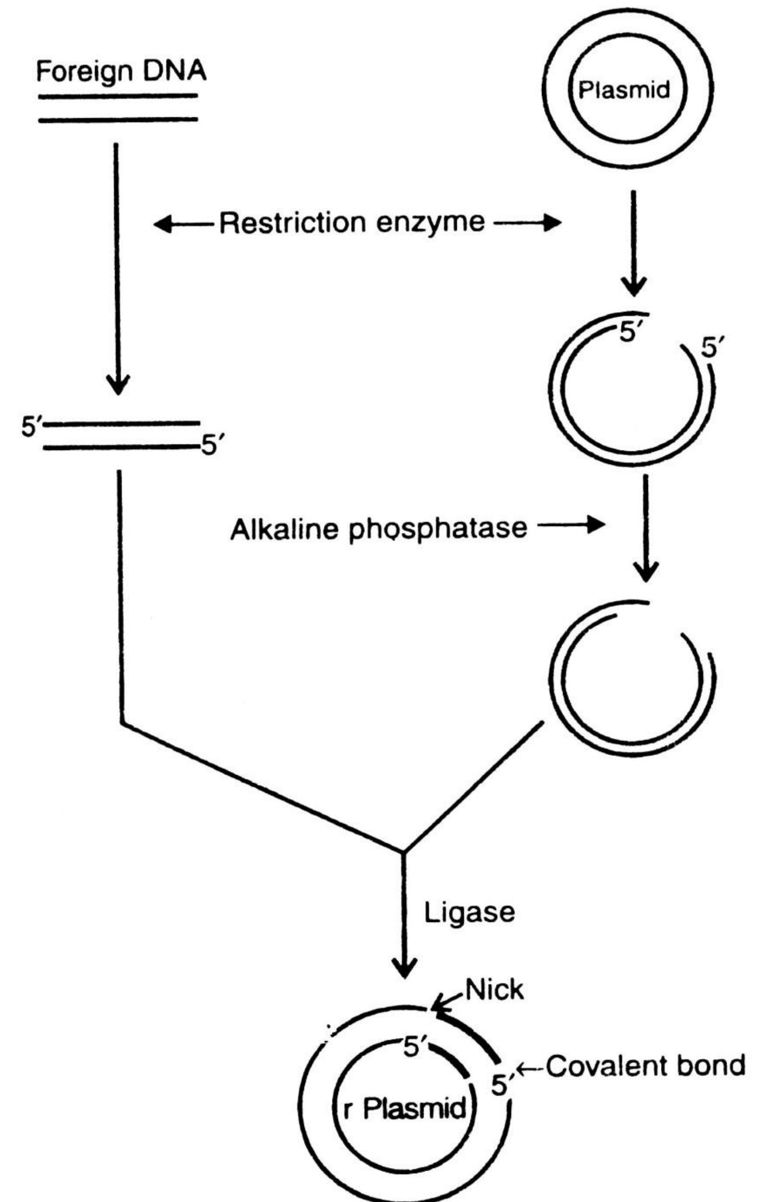
The enzyme alkaline phosphatase (AP) removes the phosphate group from the 5’ end of a DNA molecule leaving a free 5’ hydroxyl group hence it is used to prevent unwanted self-ligation of vector DNA molecules in cloning procedures.

This enzyme is isolated from bacteria (BAP) or calf intestine (CAP).

The hybrid or recombinant DNA obtained has a nick with 3' and 5' hydroxyl ends.

Ligase will only join 3' and 5' ends of recombinant DNA together if the 5' end is phosphorylated.

Thus, alkaline phosphatase and ligase prevent re-circularization of the vector and increase the frequency of production of recombinant DNA molecules.

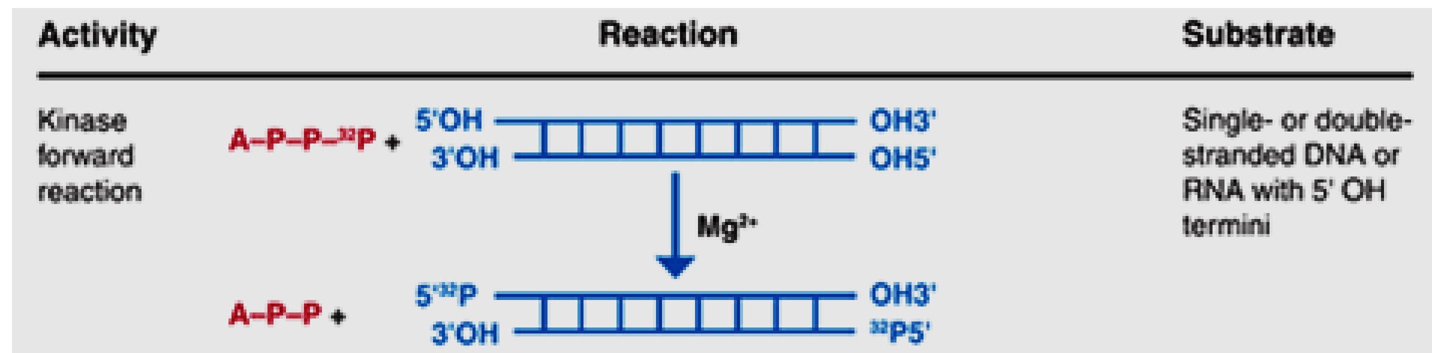


Polynucleotide 5'-hydroxyl-kinase

- In enzymology, a **polynucleotide 5'-hydroxyl-kinase** is an enzyme that catalyzes the chemical reaction

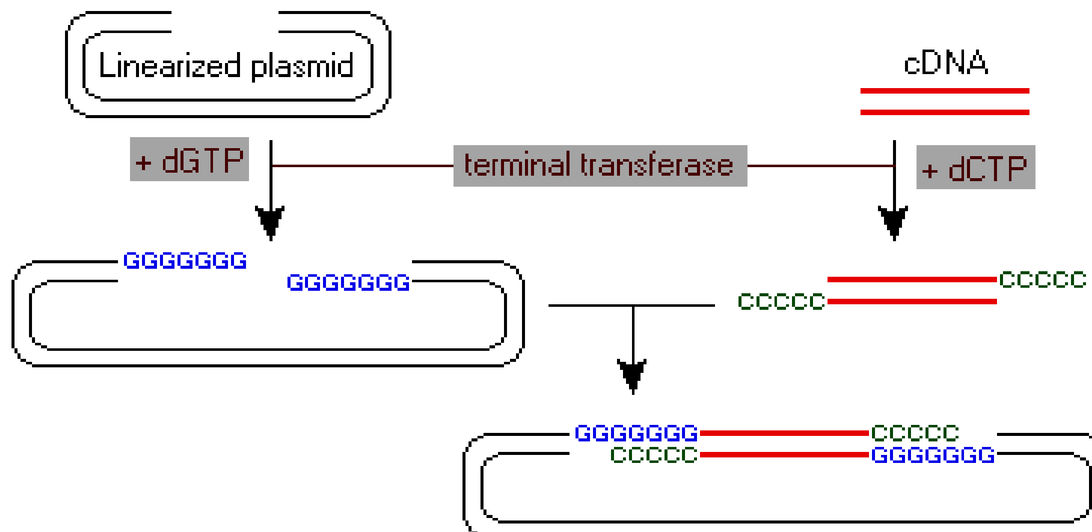
ATP + 5'-dephospho-DNA ADP + 5'-phospho-DNA

- Thus, the two substrates of this enzyme are ATP and 5'-dephospho-DNA
- Polynucleotide kinase is a T7 bacteriophage (or T4 bacteriophage) enzyme that catalyzes the transfer of a gamma-phosphate from ATP to the free hydroxyl end of the 5' DNA or RNA. The resulting product could be used to end-label DNA or RNA, or in a ligation reaction.



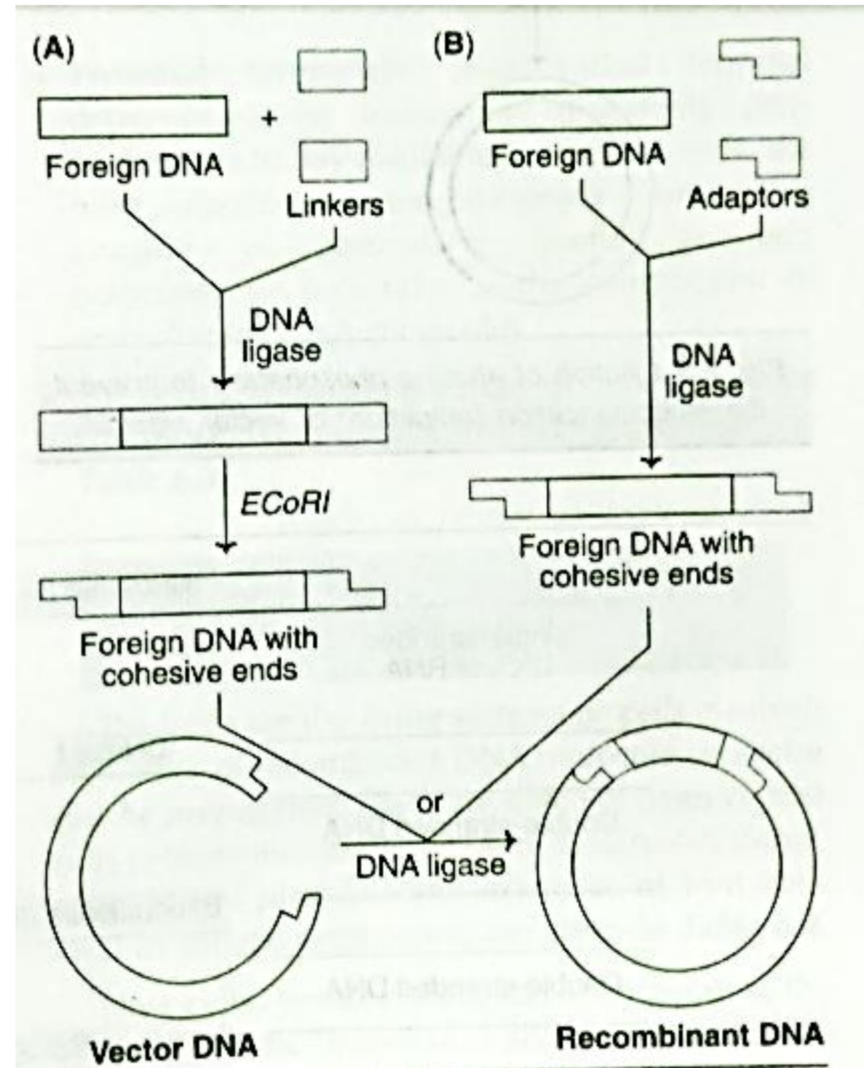
Terminal transferase and Homopolymer tailing

- **Terminal deoxynucleotidyl transferase** is a specialized DNA polymerase catalyzes the addition of nucleotides to the 3' terminus.
- Unlike most DNA polymerases, it does not require a template. The preferred substrate of this enzyme is a 3'-overhang, but it can also add nucleotides to blunt or dipped 3' ends.
- Cobalt is a necessary cofactor, however the enzyme catalyzes reaction upon Mg and Mn administration *in vitro*.
- It can be used to made target DNA and vector compatible. It can be used to add nucleotides that can then be used as a template for a primer in subsequent PCR.
- It can also be used to add nucleotides labeled with radioactive isotopes.



Linker/Adapter

- A **linker/adapter** in genetic engineering is a short, chemically synthesized, ds-DNA molecule which is used to link the ends of two other DNA molecules. It may be used to add sticky ends to cDNA allowing it to be ligated into the plasmid much more efficiently.
- Linker possess RE cleavage site, can ligate with blunt end DNA and cut with the same RE to produce DNA fragment with sticky end.
- **Pre-digested linker is adapter.**
- Adapters are used to link the ends of two DNA molecules that have different sequences at their ends. An adapter is used to join a DNA insert cut with one Restriction enzyme, say EcoRl, with a vector opened with another enzyme, Bam Hl.



Restriction mapping

- A restriction map is a map of known restriction sites within a sequence of DNA.
- Once the clone of interest has been isolated, the first stage of analysis is often the creation of a restriction map. Restriction mapping provides a compilation of the number, order, and distance between restriction endonuclease cutting sites along a cloned DNA fragment.
- In addition, restriction mapping plays an important role in characterizing DNA, mapping genes, and diagnostic tests for genetic diseases

Process

- To find the relative positions of restriction sites on a DNA, a technique involving single and double restriction digests is used.
- The experimental procedure first requires an aliquot of purified DNA.
- Digestion is then performed with each enzyme(s) chosen. Then double digestion with RE is performed.
- The resulting samples are subsequently run on an electrophoresis gel, typically on agarose gel.
- The lengths of the DNA fragments can be determined by comparing their position in the gel to reference DNAs of known lengths in the gel.
- The pattern of cutting in single and double digests indicates what the relationship is between the two sites.

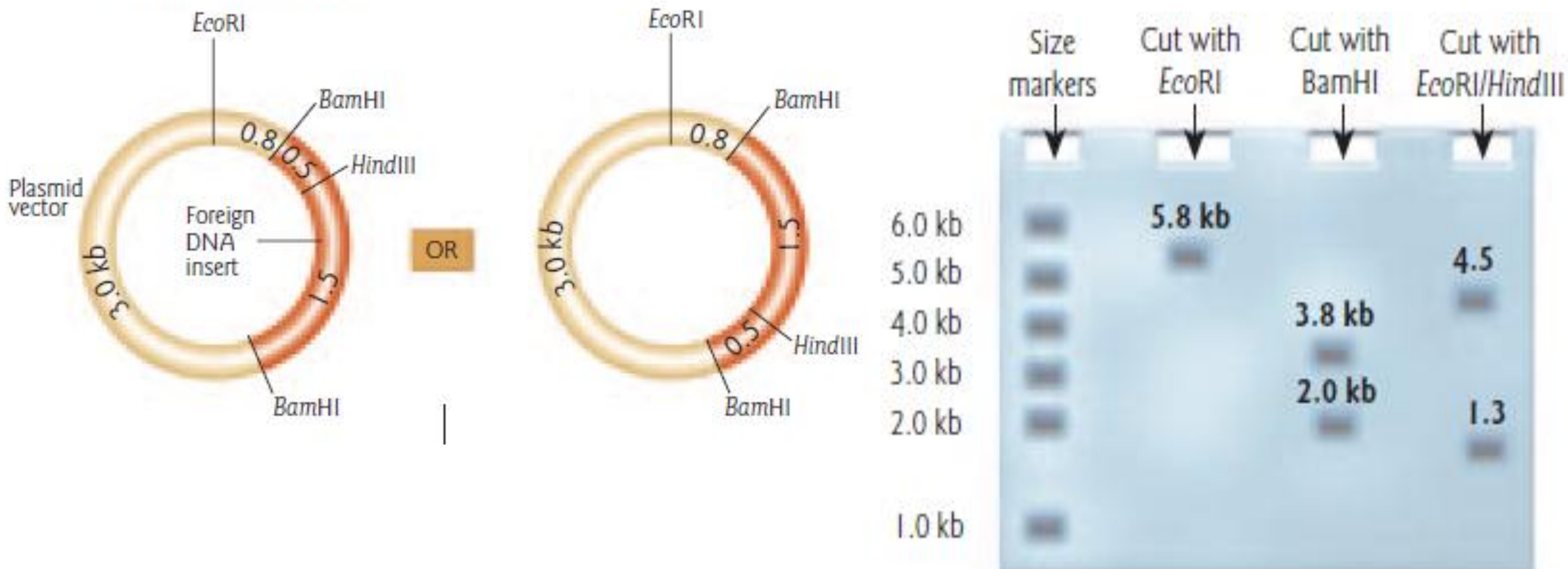


Figure 8.13 Analysis of recombinant DNA by restriction endonuclease digestion. Assume a 2.0 kb foreign DNA insert has been successfully ligated into the *Bam*HI site of a 3.8 kb plasmid vector. However, the orientation of the insert is unknown. Samples of the recombinant plasmid are digested with restriction endonucleases: one sample is digested with *Eco*RI, one with *Bam*HI, and one with both *Eco*RI and *Hin*dIII. The resulting fragments are separated by agarose gel electrophoresis. The sizes of the separated fragments can be measured by comparison with molecular weight standards in an adjacent lane. *Eco*RI linearizes the 5.8 kb plasmid which appears as a single band on the gel. *Bam*HI generates two fragments of 3.8 and 2.0 kb in size, representing the vector and the insert, respectively. Digestion with *Eco*RI and *Hin*dIII generates two fragments of 4.5 and 1.3 kb. These data indicate that the foreign DNA has been inserted in the desired orientation. If the DNA had been inserted in the opposite orientation, a double digest with *Eco*RI and *Hin*dIII would have generated fragment sizes of 3.5 and 2.3 kb.

A linear fragment of DNA (7.5 kb) is cleaved with the individual restriction enzymes *Hin*dIII and *Sma*I and then with a combination of the two enzymes. The fragments obtained are:

HindIII :	2.5 kb, 5 kb
Sma1:	2.0 kb, 5.5 kbs
HindIII and *SmaI*:	2.5 kb, 3.0 kb, 2.0 kb.

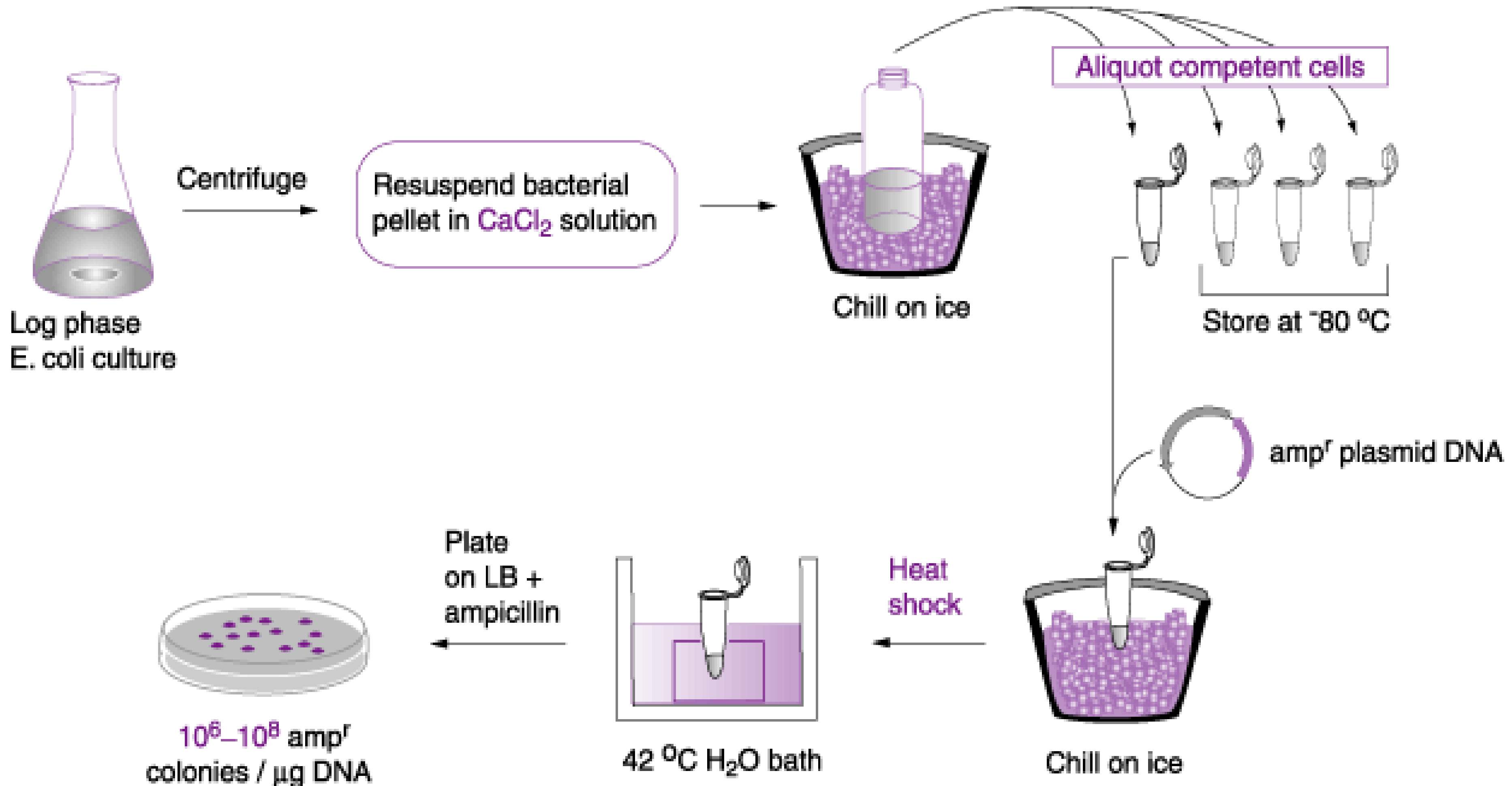
a. Draw the restriction map.

b. The mixture of fragments produced by the combined enzymes is cleaved with the enzyme *Eco*RI, resulting in the loss of the 3-kb fragment (band stained with ethidium bromide on an agarose gel) and the appearance of a band stained with ethidium bromide representing a 1.5-kb fragment. Mark the *Eco*RI cleavage site on the restriction map.

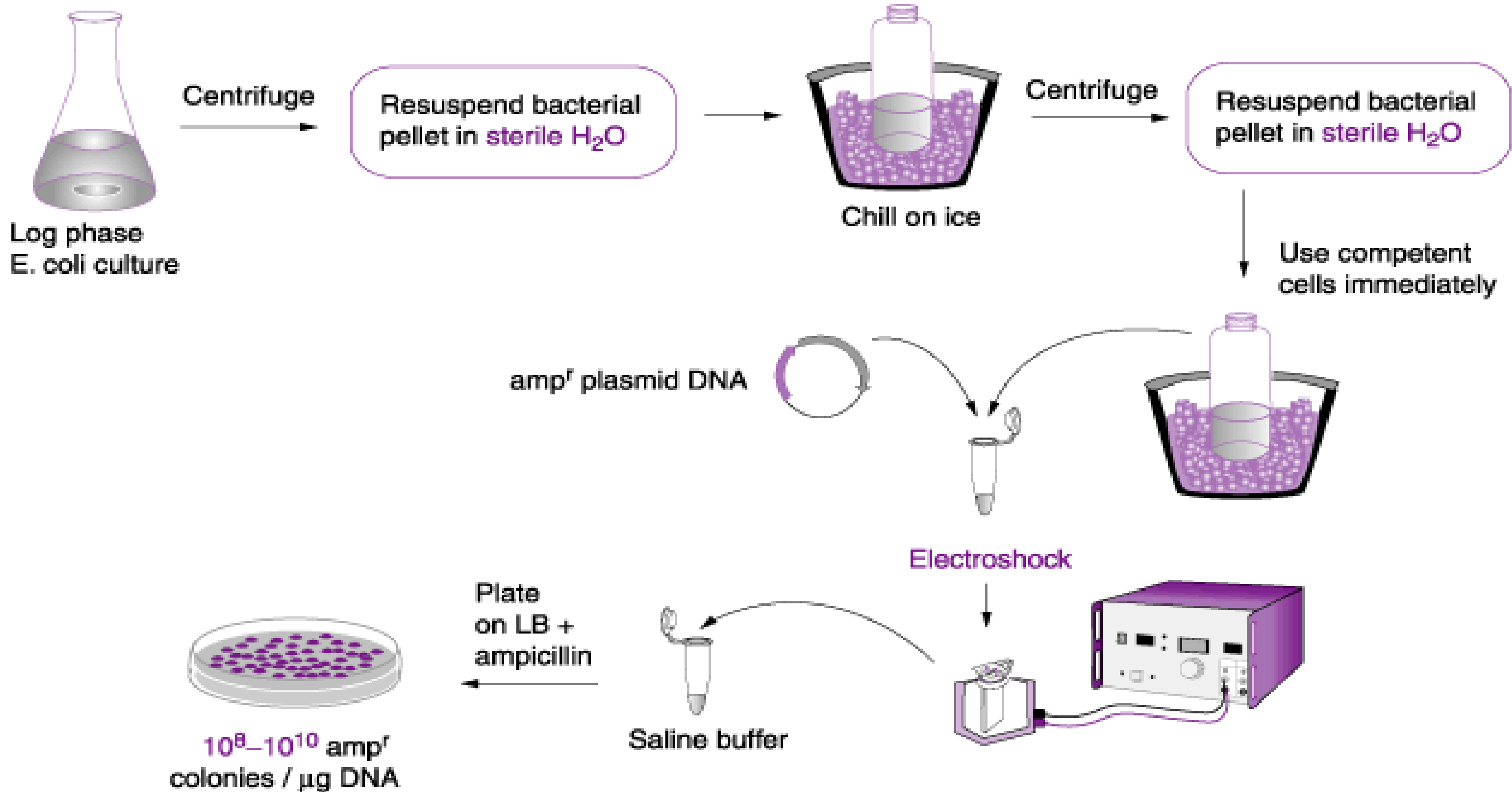
TRANSFORMATION OF LIGATION PRODUCTS

- The process of transferring exogenous DNA into cells is call "transformation"
- There are basically two general methods for transforming bacteria. The first is a chemical method utilizing $CaCl_2$ and heat shock to promote DNA entry into cells.
- A second method is called electroporation based on a short pulse of electric charge to facilitate DNA uptake.

CHEMICAL TRANSFORMATION WITH CALCIUM CHLORIDE



TRANSFORMATION BY ELECTROPORATION



Types of vectors used

- Cloning vector

A cloning vector is a small piece of DNA, taken from a virus, a plasmid, or the cell of a higher organism, that can be stably maintained in an organism, and into which a foreign DNA fragment can be inserted for cloning purposes.

- Expression vector

Expression vectors produce proteins through the transcription of the vector's insert followed by translation of the mRNA produced, they therefore require more components than the simpler transcription-only vectors. Expression in different host organism would require different elements, although they share similar requirements, for example a promoter for initiation of transcription, a ribosomal binding site for translation initiation, and termination signals.

- Shuttle vector

A shuttle vector is a vector (usually a plasmid) constructed so that it can propagate in two different host species. Therefore, DNA inserted into a shuttle vector can be tested or manipulated in two different cell types.

CLONING VECTORS

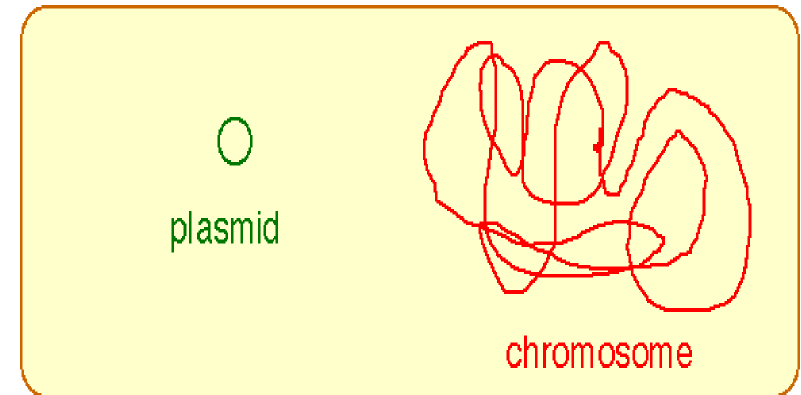
- **Cloning vectors are DNA molecules that are used to "transport" cloned sequences between biological hosts and the test tube.**

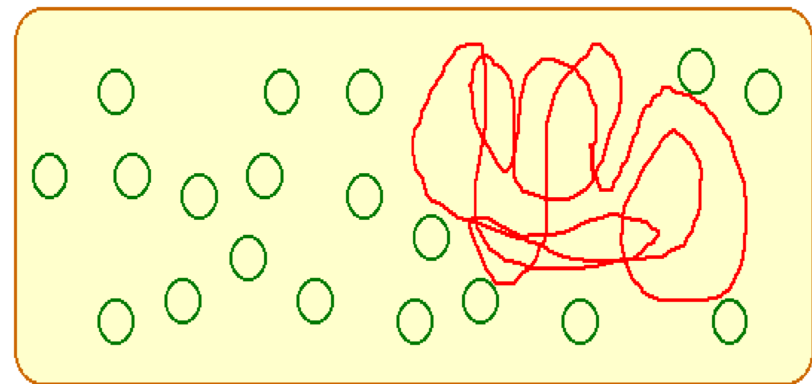
Cloning vectors share four common properties:

1. **Ability to promote autonomous replication.**
2. **Contain a genetic marker (usually dominant) for selection.**
3. **Unique restriction sites to facilitate cloning of insert DNA.**
4. **Minimum amount of nonessential DNA to optimize cloning.**

PLASMIDS

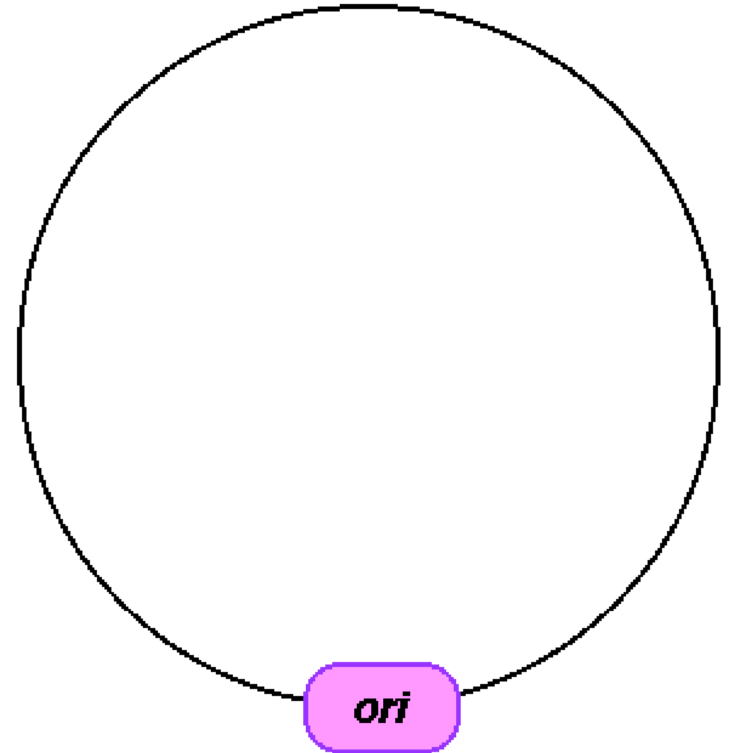
- Bacterial cells may contain extra-chromosomal DNA called plasmids.
- Plasmids are usually represented by small, circular DNA.
- Some plasmids are present in multiple copies in the cell





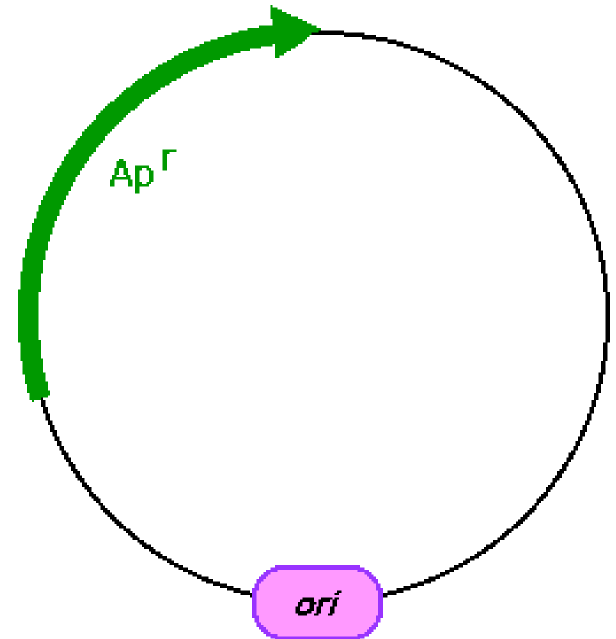
ORIGIN OF REPLICATION

- **Origin of replication** is a DNA segment recognized by the cellular DNA-replication enzymes.
- Without replication origin, DNA cannot be replicated in the cell.



SELECTIVE MARKER

- **Selective marker** is required for maintenance of plasmid in the cell.
- Because of the presence of the selective marker the plasmid becomes useful for the cell.
- Under the selective conditions, only cells that contain plasmids with selectable marker can survive
- Genes that confer resistance to various antibiotics are used.
- Genes that make cells resistant to ampicillin, neomycin, or chloramphenicol are used



Why Plasmids

- small size (easy to manipulate and isolate)
- circular (more stable)
- replication independent of host cell
- several copies may be present (facilitates replication)
- frequently have antibody resistance (detection easy)

pBR322

- **pBR322** is a plasmid and was one of the first widely used *E. coli* cloning vectors. Created in 1977 in the laboratory of Herbert Boyer at the University of California, Davis, it was named after the Mexican postdoctoral researchers who constructed it. The p stands for "plasmid," and BR for "Bolivar" and "Rodriguez."

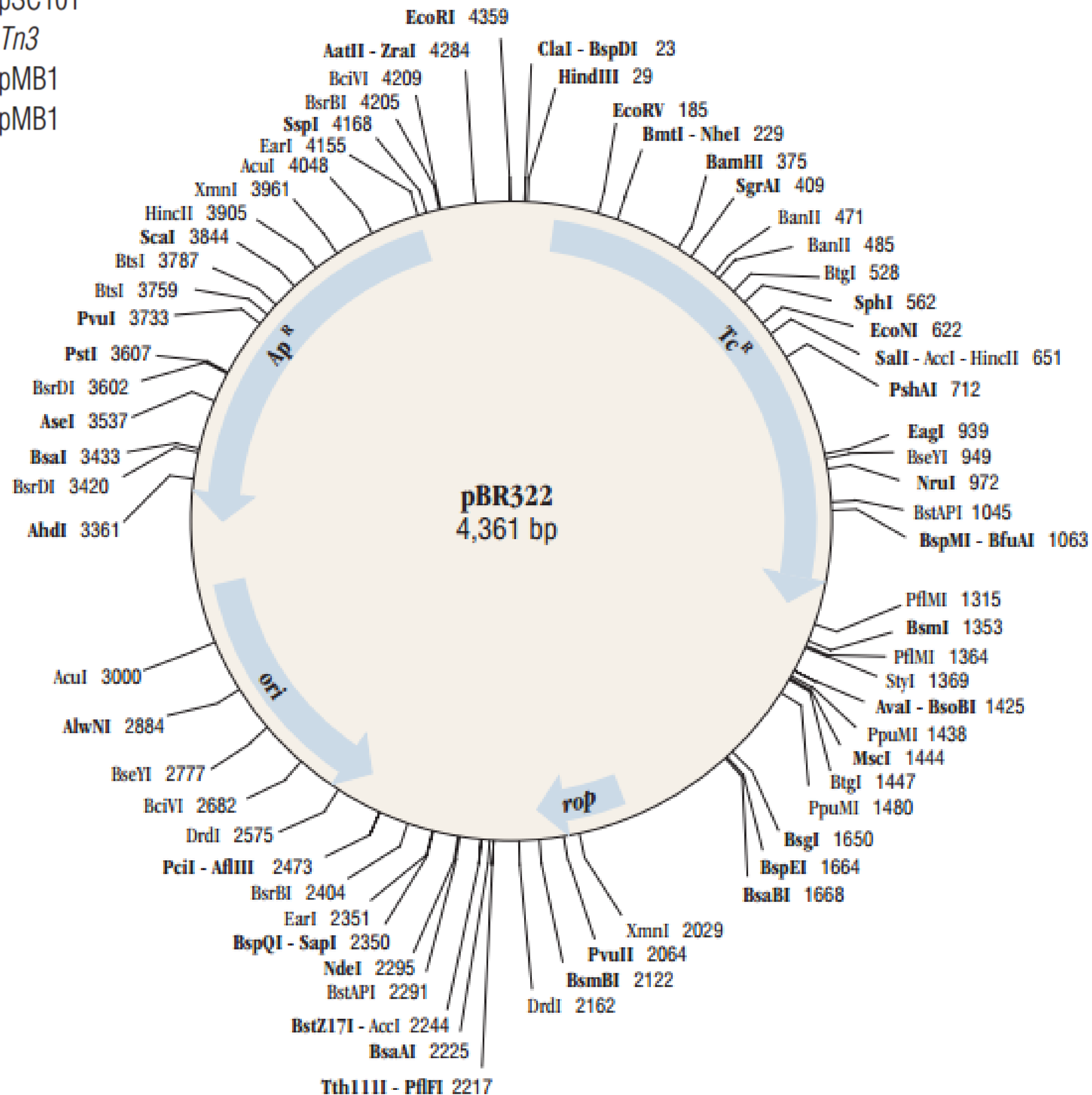
- pBR322 is 4361 base pairs in length and contains the replicon of plasmid pMB1, the ampR gene, encoding the ampicillin resistance protein (source plasmid RSF2124) and the tetR gene, encoding the tetracycline resistance protein (source plasmid pSC101).

- The plasmid has unique restriction sites for more than forty restriction enzymes. 11 of these 40 sites lie within the tetR gene. There are 2 sites for restriction enzymes HindIII and ClaI within the promoter of the tetR gene. There are 6 key restriction sites inside the ampR gene. The origin of replication or ori site in this plasmid is pMB1 (a close relative of ColE1).

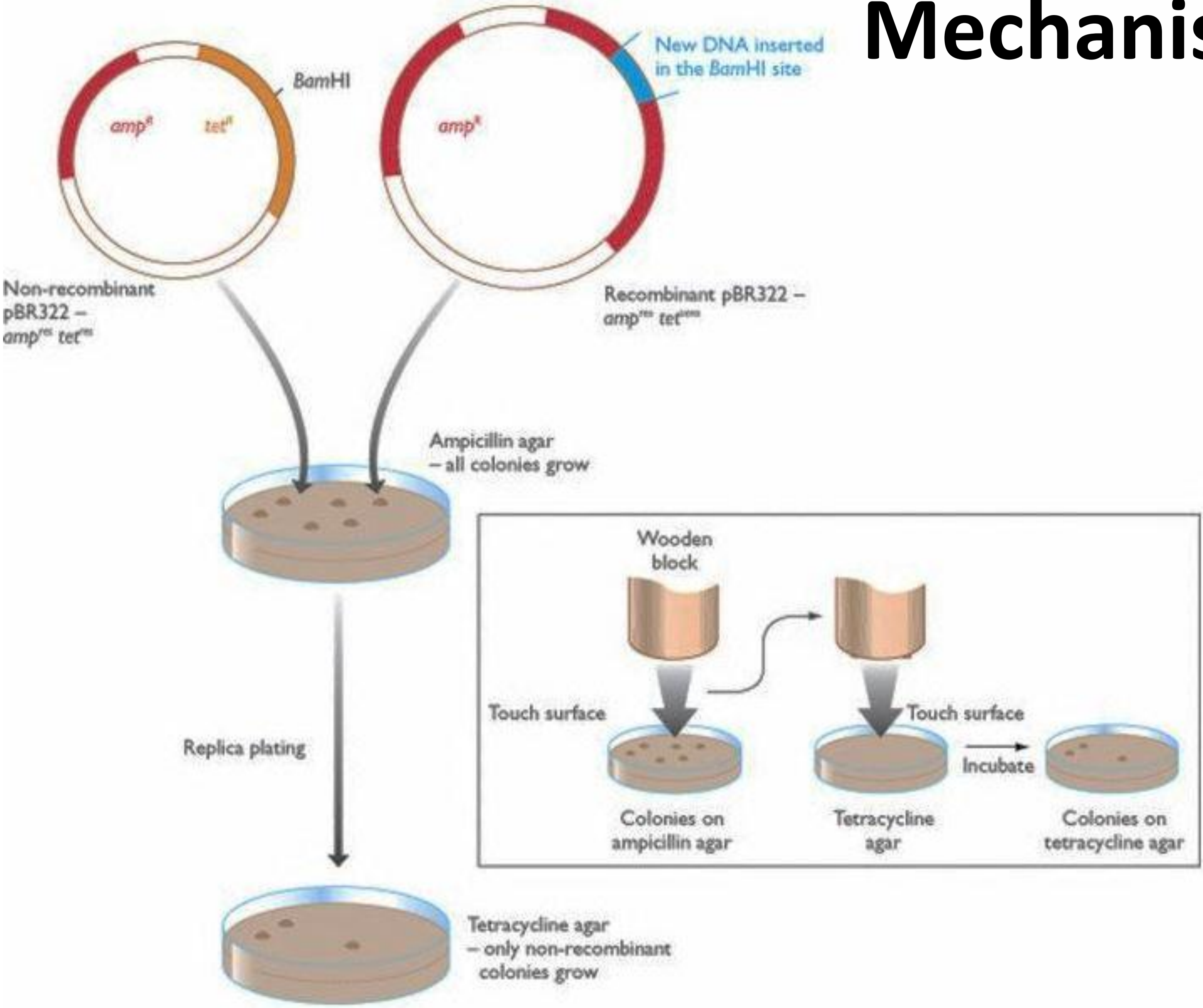
Feature	Coordinates	Source
tet (Tc[R])	86-1276	pSC101
bla (Ap[R])	4153-3293	*Tn3*
rop	1915-2106	pMB1
origin	3122-2534	pMB1

ori = origin of replication

Ap = ampicillin, Tc = tetracycline



Mechanism



Advancement of plasmid vector

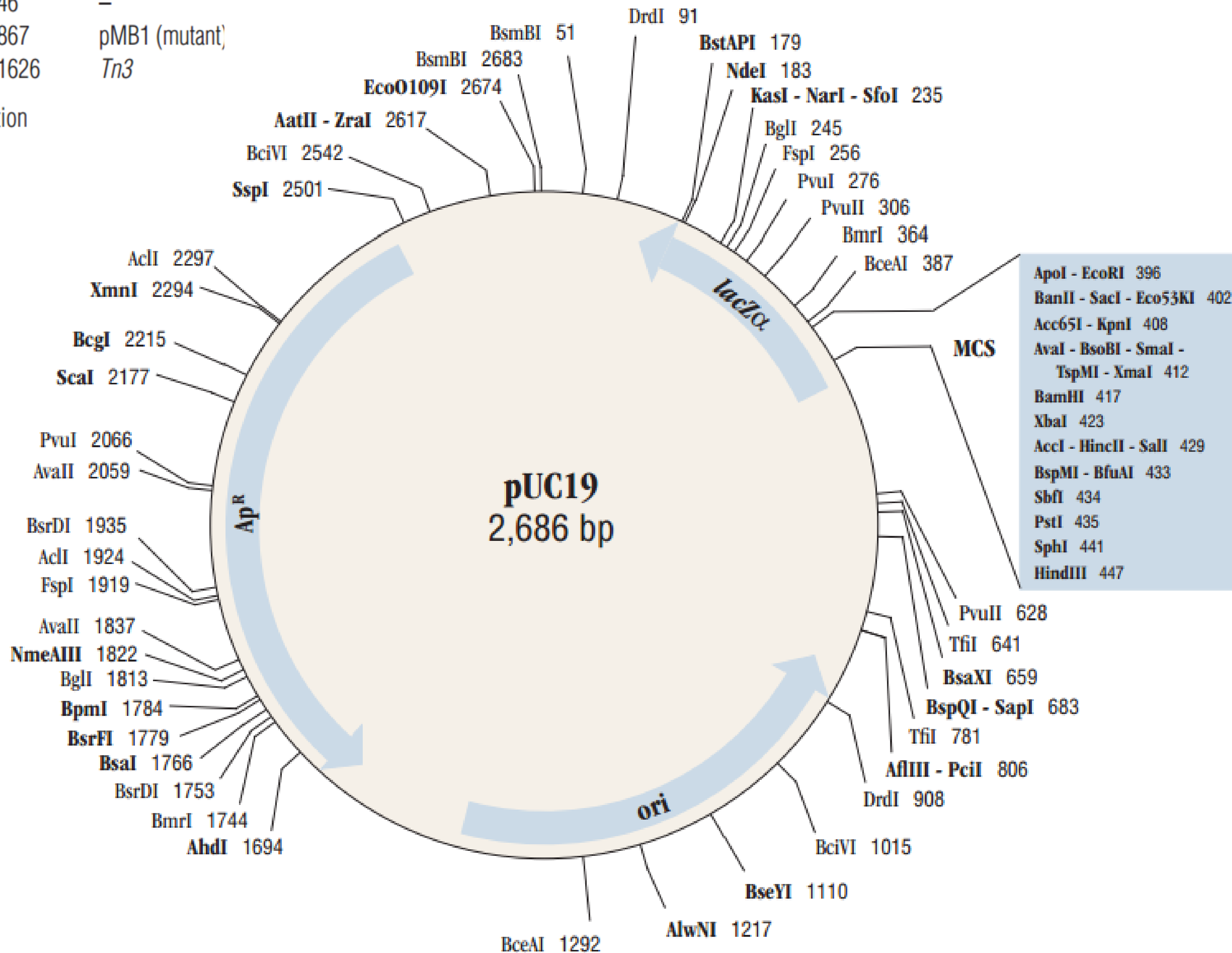
- Incorporation of MCS or polylinker
- Rapid detection of recombinant cell with insert

pUC 19

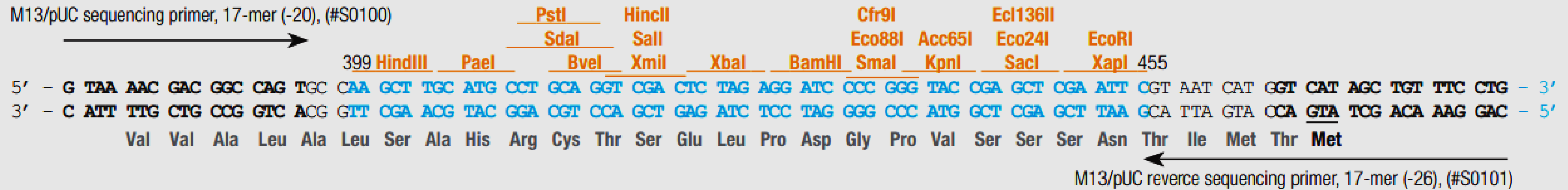
- pUC19 is one of a series of plasmid cloning vectors created by Messing and co-workers. The designation "pUC" is derived from the classical "p" prefix (denoting "plasmid") and the abbreviation for the University of California, where early work on the plasmid series had been conducted.
- It is a circular double stranded DNA and has 2686 base pairs. It has one ampR gene (ampicillin resistance gene), and an N-terminal fragment of β-galactosidase (lac Z) gene of E. coli. The multiple cloning site (MCS) region is split into the lac Z gene (codons 6–7 of lac Z are replaced by MCS), where various restriction sites for many restriction endonucleases are present.
- The ori site or replicon, rep is derived from pMB1 vector. pUC vector is small but has a high copy number. The high copy number of pUC plasmids is a result of the lack of the rop gene and a single point mutation in rep of pMB1.
- The lac Z gene codes for β-galactosidase. The recognition sites for HindIII, SphI, PstI, SalI, XbaI, BamHI, SmaI, KpnI, SacI and EcoRI restriction enzymes have been derived from the vector M13mp19.

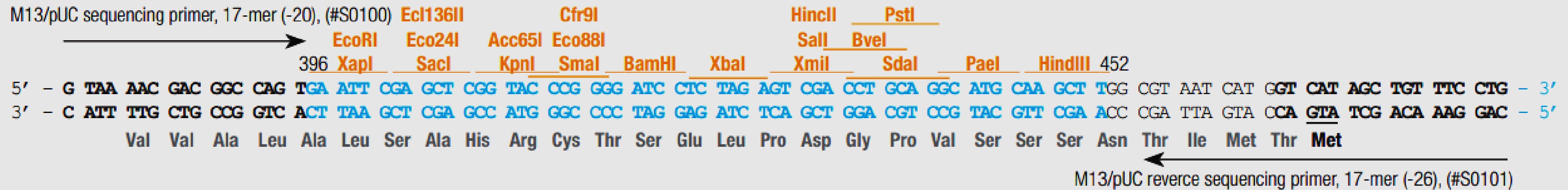
Feature	Coordinates	Source
lacZα	469-146	–
origin	1455-867	pMB1 (mutant)
bla (Ap[R])	2486-1626	*Tn3*

ori = origin of replication
Ap = ampicillin



MCS of pUC18 and 19





Mechanism

- The lac Z fragment, whose synthesis can be induced by IPTG, is capable of intra-allelic complementation with a defective form of β-galactosidase enzyme encoded by host chromosome (mutation lacZDM15 in E. coli JM109, DH5α and XL1-Blue strains).
- In the presence of IPTG in growth medium, bacteria synthesise both fragments of the enzyme. Both the fragments can together hydrolyse X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) and form blue colonies when grown on media where it is supplemented.
- Insertion of foreign DNA into the MCS located within the lac Z gene causes insertional inactivation of this gene at the N-terminal fragment of beta-galactosidase and abolishes intra-allelic complementation. Thus bacteria carrying recombinant plasmids in the MCS cannot hydrolyse X-gal, giving rise to white colonies, which can be distinguished on culture media from non-recombinant cells, which are blue.
- Therefore the media used should contain ampicillin, IPTG, and X-gal.

Vector

lacZ´ gene

mRNA

Functional *lacZ´* gene product

Blue colonies

Portion of *lacZ´* gene

Portion of *lacZ´* gene

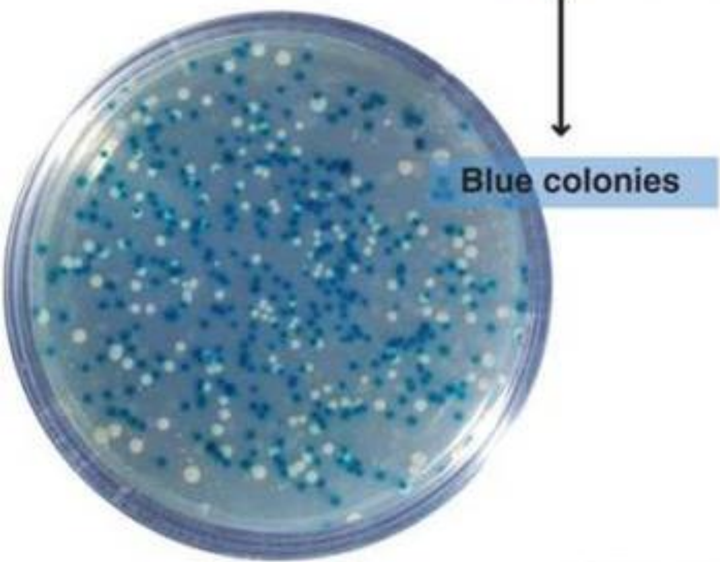
Insert

mRNA

Non-functional protein

White colonies

Bacterial cells plated on medium containing x-gal



Disadvantages using plasmids

- Cannot accept large fragments
- Sizes range upto 10 kb
- Standard methods of transformation are inefficient

Bacteriophage

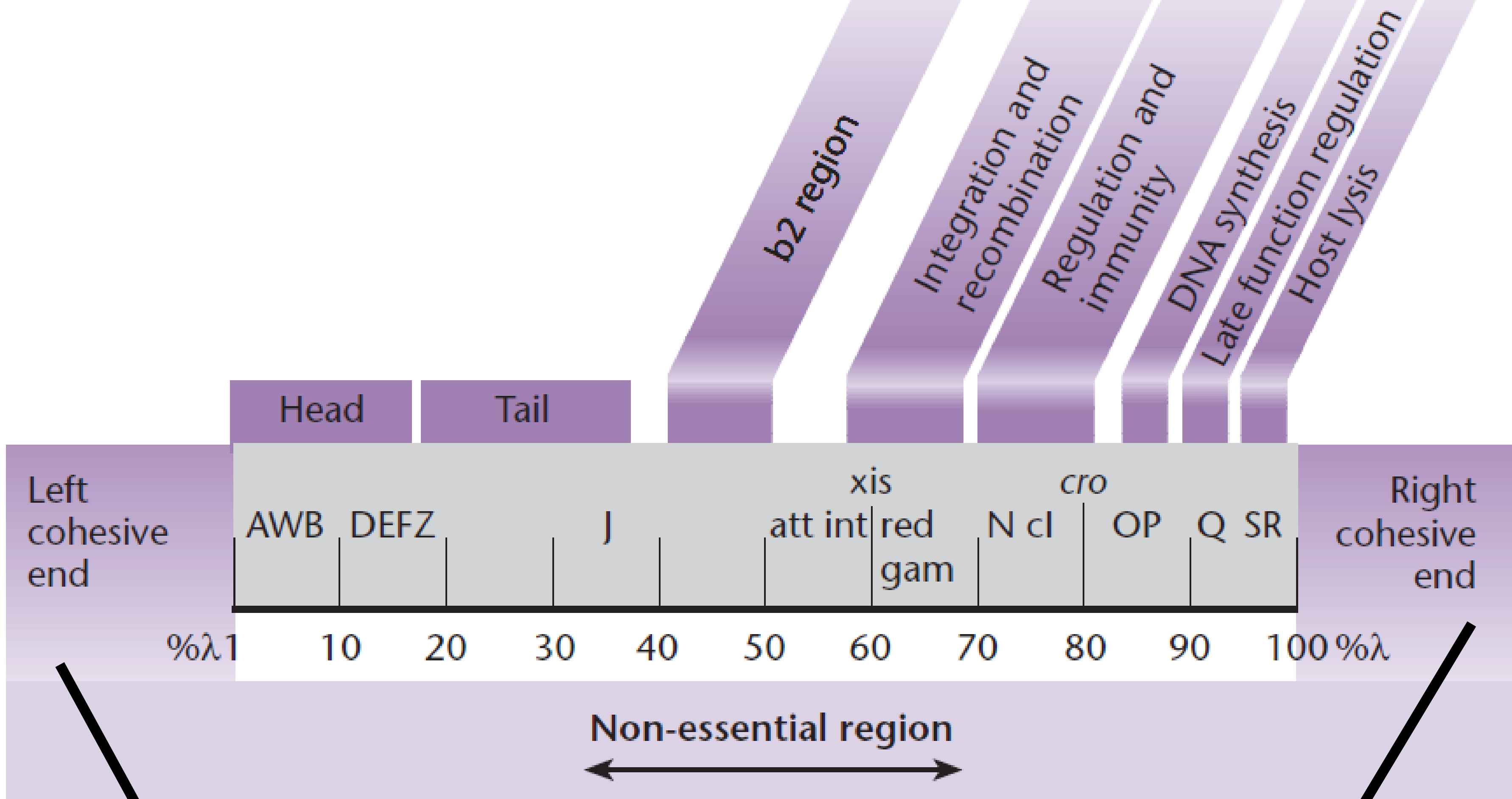
- The bacteriophages used for cloning are the phage λ and M13 phage. There is an upper limit on the amount of DNA that can be packed into a phage (a maximum of 53 kb), therefore to allow foreign DNA to be inserted into phage DNA, phage cloning vectors need to have some non-essential genes deleted, for example the genes for lysogeny in phage λ.

- There are two kinds of λ phage vectors - insertion vector and replacement vector. Insertion vectors contain a unique cleavage site whereby foreign DNA with size of 5–11 kb may be inserted. In replacement vectors, the cleavage sites flank a region containing genes not essential for the lytic cycle, and this region may be deleted and replaced by the DNA insert in the cloning process, and a larger sized DNA of 8–24 kb may be inserted.

BACTERIOPHAGE LAMBDA

- **Phage lambda** is a **bacteriophage** or **phage**, i.e. bacterial virus, that uses *E. coli* as host.
- Its structure is that of a typical phage: **head, tail, tail fibres.**
- **Lambda viral genome:** 48.5 kb linear DNA with a 12 base ssDNA "sticky end" at both ends; these ends are complementary in sequence and can hybridize to each other (this is the **cos** site: **co**he**s**ive ends).
- **Infection:** lambda tail fibres adsorb to a cell surface receptor, the tail contracts, and the DNA is injected.
- The DNA circularizes at the **cos** site, and lambda begins its life cycle in the *E. coli* host.

- Wild-type λ-DNA contains several target sites for most of the commonly used restriction endonucleases and so is not itself suitable as a vector.
- Derivatives of the wild-type phage have therefore been produced that either have a single target site at which foreign DNA can be inserted (**insertional vectors**) or have a pair of sites defining a fragment that can be removed (stuffer) and replaced by foreign DNA (**replacement vectors**).
- Since phage λ can accommodate only about 5% more than its normal complement of DNA, vector derivatives are constructed with deletions to increase the space within the genome.
- Many vector derivatives of both the insertional and replacement types.

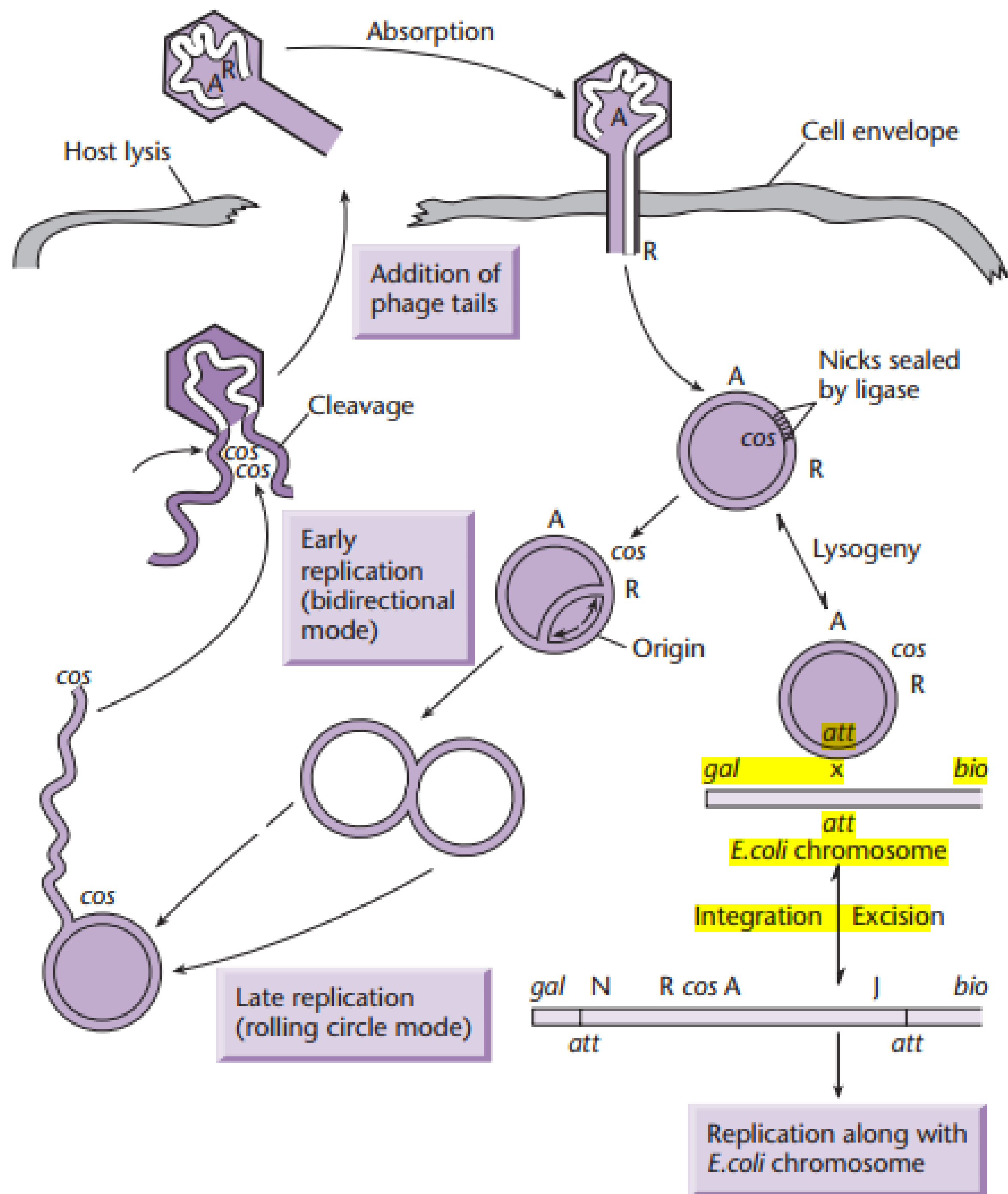


terminase cuts like this:

---G/GGGCGGCGACCTC---
---CCCCGCCGCTGGA/G---

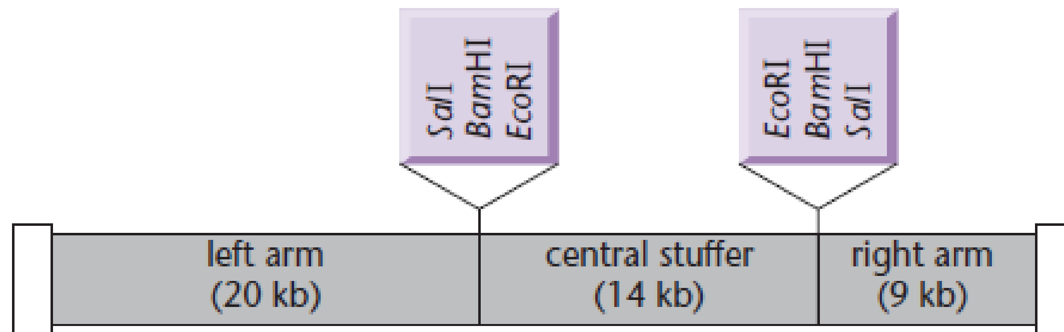
resulting in these **cos ends** in the linear Lambda DNA:

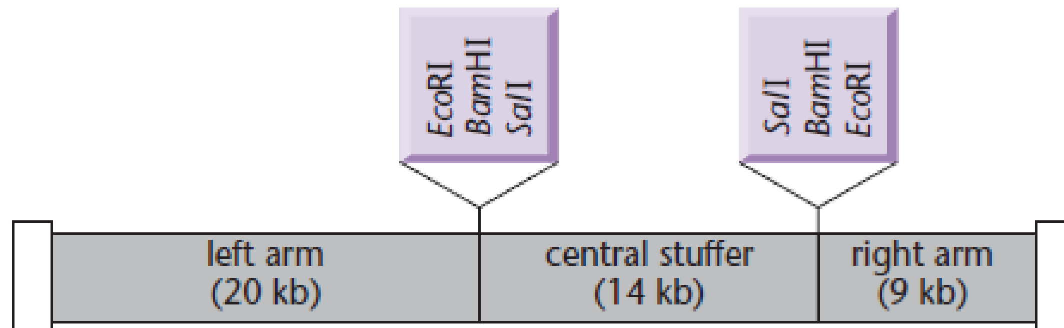
---G   GGGCGGCGACCTC---
---CCCCGCCGCTGGA   G---

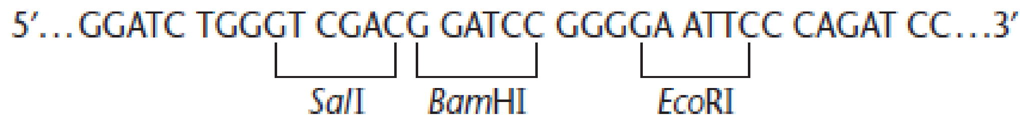


Absorption
Host lysis
Cell envelope
A
R
Addition of phage tails
Nicks sealed by ligase
cos
Cleavage
cos
cos
Early replication (bidirectional mode)
cos
Origin
Lysogeny
att
gal
x
bio
att
E.coli chromosome
Integration
Excision
gal N R cos A J bio
att
att
Late replication (rolling circle mode)
Replication along with E.coli chromosome

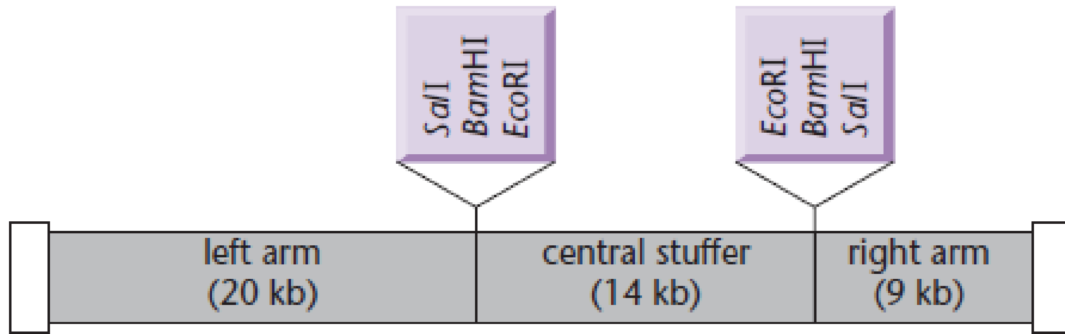
Screening/selection



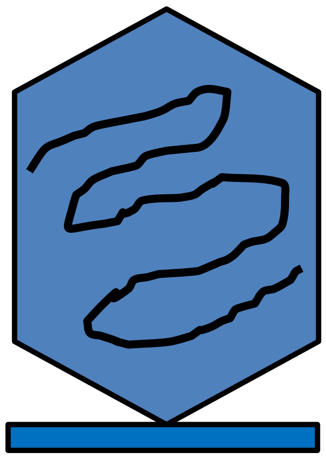


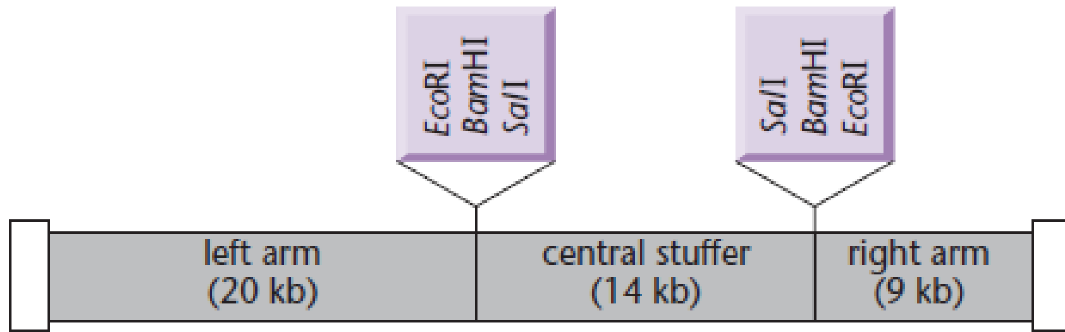


There is also a lower size limit for DNA that can be packed into a phage, and vector DNA that is too small cannot be properly packaged into the phage. This property can be used for selection - vector without insert may be too small, therefore only vectors with insert may be selected for propagation.

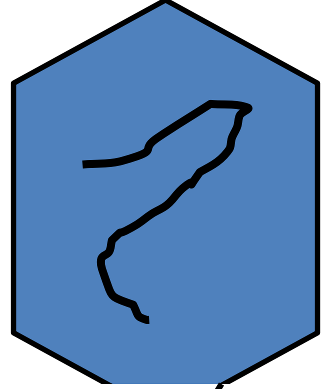


EMBL3 vector



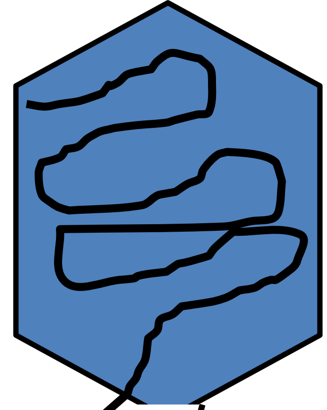


EMBL4 vector



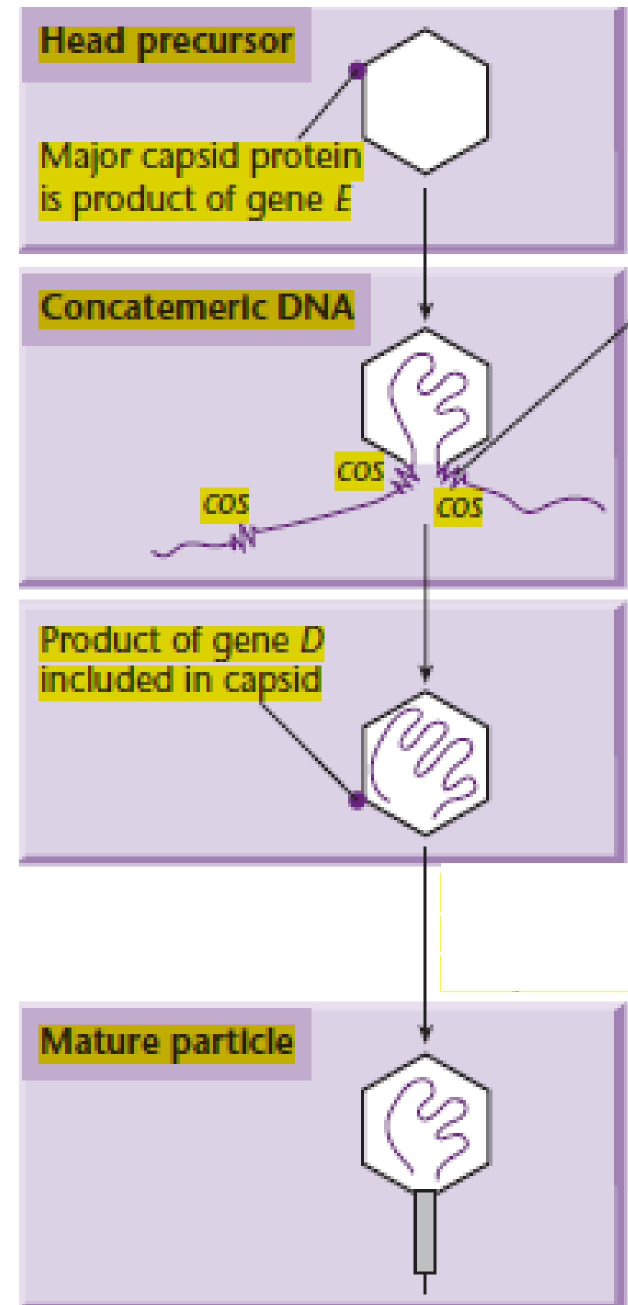
5′…GGATC TGGGT CGACG GATCC GGGGA ATTCC CAGAT CC…3′

*Sal*I *Bam*HI *Eco*RI



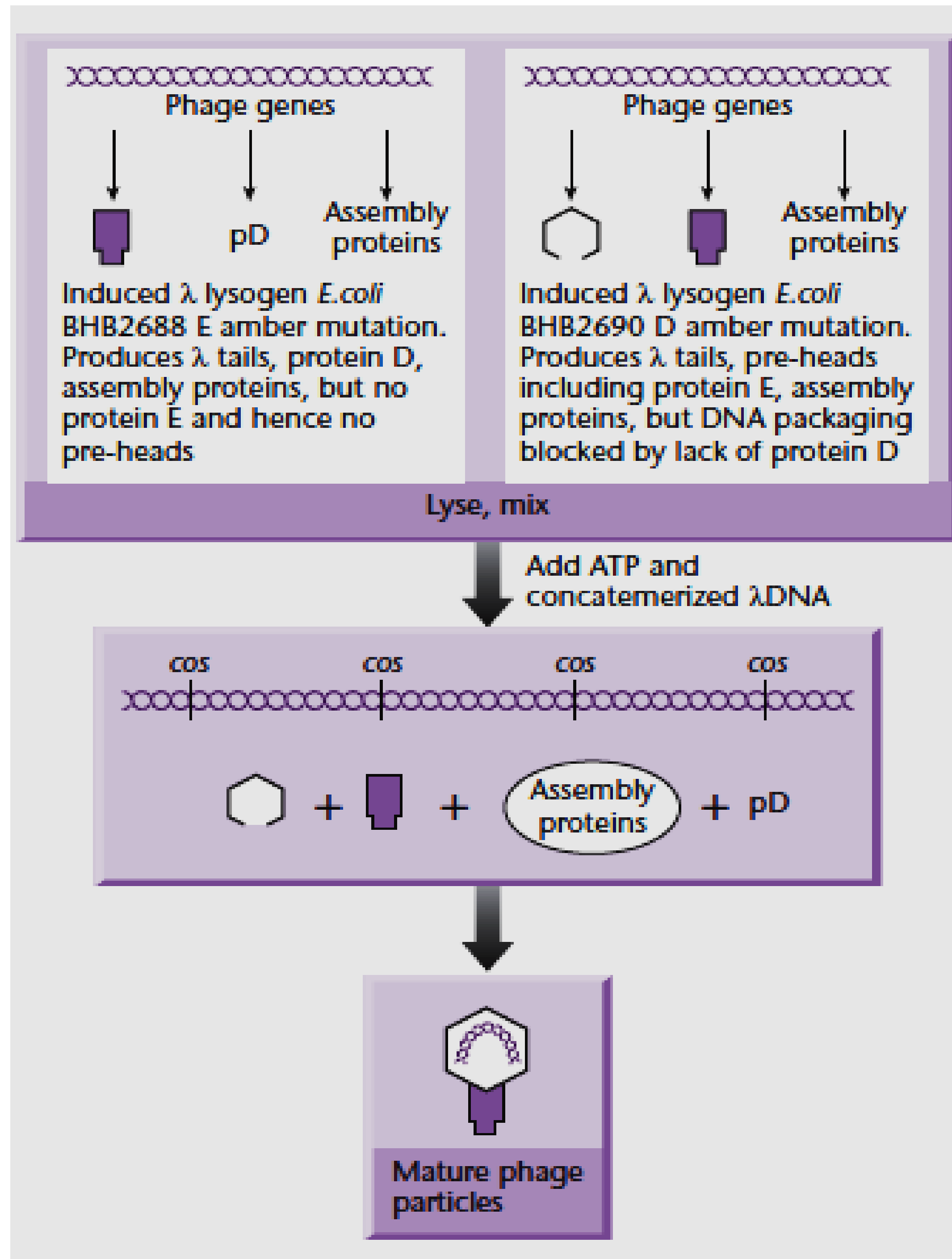
In-vitro phage packaging

- Phage DNA in concatemeric form, produced by a rolling-circle replication mechanism, is the substrate for the packaging reaction. In the presence of phage head precursor (the product of gene *E* is the major capsid protein) and the product of gene *A*, the concatemeric DNA is cleaved into monomers and encapsidated.
- Nicks are introduced in opposite strands of the DNA, 12 nucleotide pairs apart at each *cos* site, to produce the linear monomer with its cohesive termini. The product of gene *D* is then incorporated into what now becomes a completed phage head.
- The products of genes *W*and *FII*, among others, then unite the head with a separately assembled tail structure to form the mature particle.



In-vitro packaging

- this is most efficiently performed in a very concentrated mixed lysate of two induced lysogens, one of which is blocked at the pre-head stage by an amber (UAG) mutation in gene *D* and therefore accumulates this precursor, while the other is prevented from forming any head structure by an amber mutation in gene *E*
- In the mixed lysate, genetic complementation occurs and exogenous DNA is packaged.
- Concatemeric DNA is the substrate for packaging prepared covalently joined concatemers are, of course, produced in the ligation reaction by association of the natural cohesive ends of λ
- The in vitro system may package added monomeric DNA, which presumably first concatemerizes non-covalently.



DNA cloning with single-stranded DNA vectors

- M13 is filamentous coliphage containing a circular single-stranded DNA molecule
- The phage particles have dimensions 900 nm × 9 nm and contain a single-stranded circular DNA molecule, which is 6407 nucleotides long.
- The filamentous phages only infect strains of enteric bacteria harboring F pili.
- Replication of phage DNA does not result in host-cell lysis. Rather, infected cells continue to grow and divide, albeit at a slower rate than uninfected cells, and extrude virus particles. Up to 1000 phage particles may be released into the medium per cell per generation

- The single-stranded phage DNA enters the cell by a process in which decapsidation and replication are tightly coupled.
- The capsid proteins enter the cytoplasmic membrane as the viral DNA passes into the cell while being converted to a double-stranded replicative form (RF).
- The RF multiplies rapidly until about 100 RF molecules are formed inside the cell.
- Replication of the RF then becomes asymmetric, due to the accumulation of a viral-encoded single stranded specific DNA-binding protein. This protein binds to the viral strand and prevents synthesis of the complementary strand.
- Then, only viral single strands are synthesized. These progeny single strands are released from the cell as filamentous particles following morphogenesis at the cell membrane. As the DNA passes through the membrane, the DNA-binding protein is stripped off and replaced with capsid protein.

Gene 8 protein

Gene 3 protein

M13 filamentous phage particle (870 × 6 nm). Single-stranded DNA circle of 6407 nucleotides

F pilus of F+ *E.coli*

Stage 1
Absorption and conversion of single-stranded DNA to double-stranded (replicative form, RF) DNA
0–1 minute post-infection

5 2 (+)

RNA *pol*

4 3

DNA *pol*III

Stage 2
RF → RF rolling circle replication

~20 minutes post-infection

Nicking by gene 2 protein

5 6

8 (+) 7

Cutting by gene 2 protein

Stage 3
Gene 5 protein switches DNA synthesis to production of single strands

~20 minutes post-infection to indefinitely

Gene 5 protein

(+)

Virus particles extruded from infected cell without lysis

Vectors with ssDNA uses

- Sequencing by the original dideoxy method required single-stranded DNA
- Oligonucleotide-directed mutagenesis
- Primer/probe preparation

Opportunity and possibility

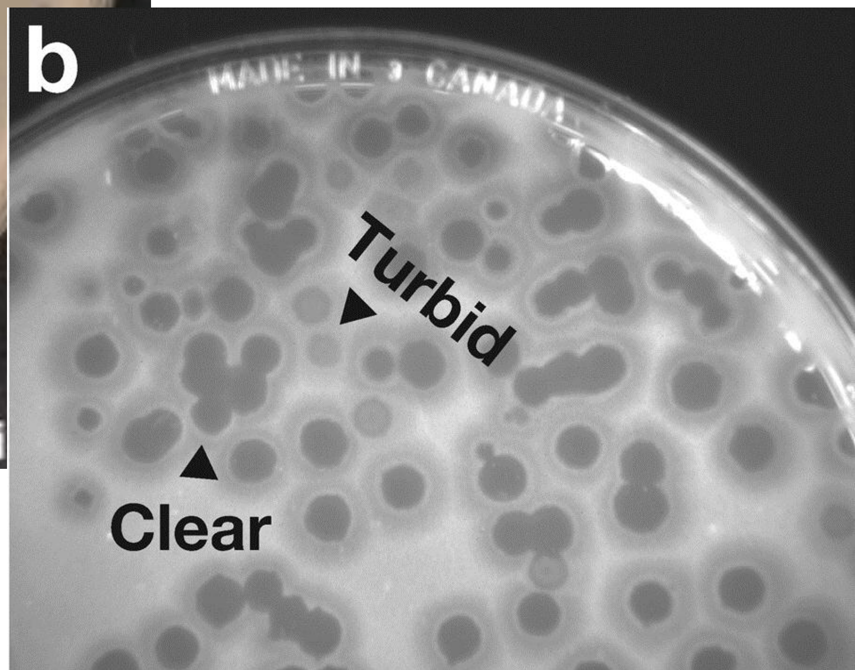
- This RF can be purified and manipulated *in vitro* just like a plasmid. Secondly, both RF and single-stranded DNA will transfect competent *E. coli* cells to yield either plaques or infected colonies, depending on the assay method.

- The normal M13 genome is 6.4 kb in length, but most of this is taken up by ten closely packed genes, each essential for the replication of the phage.
- There is only a single 507-nucleotide intergenic sequence (from position 5498 to 6005 of the DNA
- sequence) into which new DNA could be inserted without disrupting one of these genes, and this region includes the replication origin which must itself remain intact.
- Clearly there is only limited scope for modifying the M13 genome.

- The first example of M13 cloning made use of one of 10 *Bsu*I sites in the genome, two of which are in the intergenic region
- M13 RF was partially digested with BsuI and linear full-length molecules isolated by agarose gel electrophoresis
- These linear monomers were blunt-end-ligated to a HindII restriction fragment comprising the *E. coli* lac regulatory region and the genetic information for the α-peptide of β-galactosidase.
- The complete ligation mixture was used to transform a strain of E. coli with a deletion of the β-galactosidase α-fragment and recombinant phage detected by intragenic complementation on media containing IPTG and Xgal.

- Insertion of DNA fragments into the *lac* region of M13 mp1 destroys its ability to form blue plaques, making detection of recombinants easy.
- The *lac* region only contains unique sites for *Ava*II, *Bgl*I, and *Pvu*I and three sites for *Pvu*II, and there are no sites anywhere on the complete genome for the commonly used enzymes such as *Eco*RI or *Hin*dIII.
- To remedy this defect, *in vitro* mutagenesis to change a single base pair, thereby creating a unique *Eco*RI site within the *lac* fragment. This variant was designated M13 mp2.
- This phage derivative was further modified to generate derivatives with polylinkers upstream of the lac α-fragment.

ASM MicrobeLibrary.org © Lee and MacWi



Updated Vectors

- A number of purpose-built vectors were developed
- First, many of them combine elements from both plasmids and phages and are known as phasmids or, if they contain an M13 ori region, phagemids. One group of phasmids that is widely used is the λZAP series of vectors used for cDNA cloning
- Many different features that facilitate cloning and expression can be found combined in a single vector

- There are two general uses for cloning vectors: cloning large pieces of DNA and manipulating genes.
- When mapping and sequencing genomes, the first step is to subdivide the genome into manageable pieces. The larger these pieces, the easier it is to construct the final picture, hence the need to clone large fragments of DNA
- Large fragments are also needed if it is necessary to “walk” along the genome to isolate a gene

Cosmid

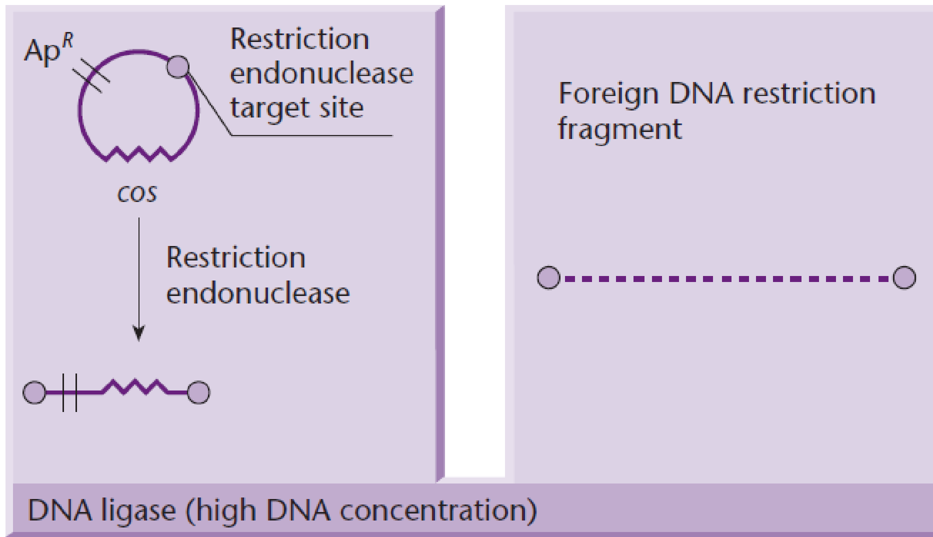
- A cosmid is a type of hybrid plasmid that contains a Lambda phage cos sequence.
- Cosmid (cos sites of phage + plasmid = cosmid) DNA sequences are originally from the lambda phage and plasmid.
- They are often used as a cloning vector in genetic engineering.
- Cosmids can be used to build genomic libraries.
- They were first described by Collins and Hohn in 1978.

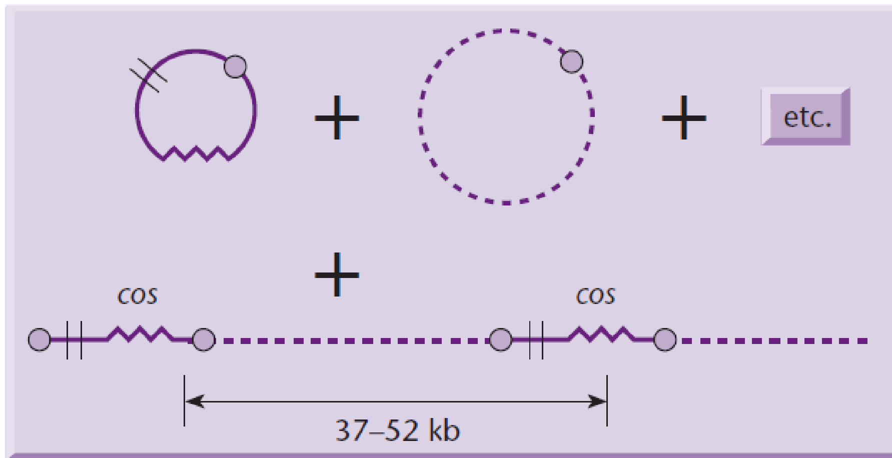
- Cosmids can contain 37 to 52 (normally 45) kb of DNA, while normal plasmids are able to carry only 1–20 kb.
- They can replicate as plasmids if they have a suitable origin of replication: for example SV40 ori in mammalian cells, ColE1 ori for double-stranded DNA replication or f1 ori for single-stranded DNA replication in prokaryotes.
- Unlike plasmids, they can also be packaged in phage capsids, which allows the foreign genes to be transferred into or between cells by transduction.
- This is made possible by the cohesive ends. In this way, they are similar to using the lambda phage as a vector, but only that all the lambda genes have been deleted with the exception of the cos sequence.
- Cos sequences are ~200 base pairs long and essential for packaging. They contain a cosN site where DNA is nicked at each strand, 12bp apart, by terminase. This causes linearization of the circular cosmid with two "cohesive" or "sticky ends" of 12bp. The cosB site holds the terminase while it is nicking and separating the strands. The cosQ site of next cosmid is held by the terminase after the previous cosmid has been packaged, to prevent degradation by cellular DNases.

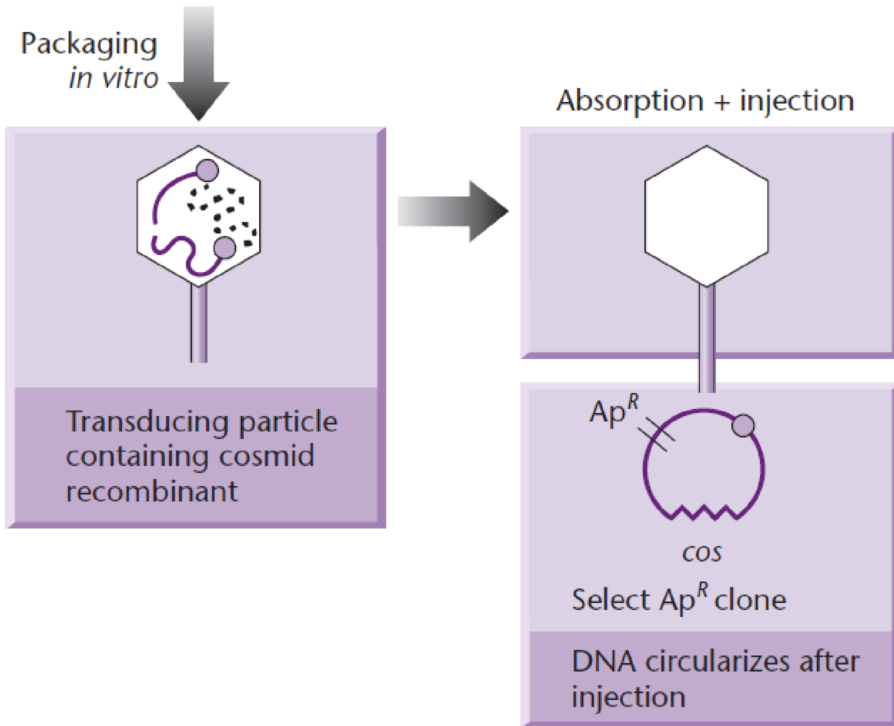
- Cosmids are predominantly plasmids with a bacterial oriV, an antibiotic selection marker and a cloning site, but they carry one, or more recently two, cos sites derived from bacteriophage lambda.
- Depending on the particular aim of the experiment broad host range cosmids, shuttle cosmids or 'mammalian' cosmids (linked to SV40 oriV and mammalian selection markers) are available.
- The loading capacity of cosmids varies depending on the size of the vector itself but usually lies around 40–45 kb.
- The cloning procedure involves the generation of two vector arms which are then joined to the foreign DNA.
- Selection against wildtype cosmid DNA is simply done via size exclusion.
- Cosmids, however, always form colonies and not plaques. Also the clone density is much lower with around 10^5 - 10^6 CFU per µg of ligated DNA.

- After the construction of recombinant lambda or cosmid libraries the total DNA is transferred into an appropriate E.coli host via a technique called in vitro packaging.
- The necessary packaging extracts are derived from E.coli cI857 lysogens (red- gam- Sam and Dam (head assembly) and Eam (tail assembly) respectively).
- These extracts will recognize and package the recombinant molecules in vitro, generating either mature phage particles (lambda-based vectors) or recombinant plasmids contained in phage shells (cosmids).

Cosmid

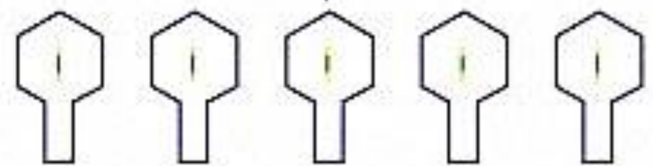


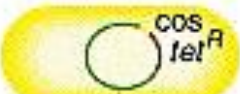




λ phage DNA concatenate
Cohesive ends (cos)
Clone cos DNA into plasmid.
Genomic DNA
cos sites
tet^R
*Bgl*II
ori
Partially cleave with *Sau*3A. Isolate 35–45-kb fragments.
Cosmid vector
Linearize at *Bgl*II site.
tet^R
cos sites
Ligate

tet^R
Vector DNA
cos sites
Genomic DNA
In vitro packaging recognizes two cos sites that are 35–45 kb apart.


Infect *E. coli*.
Select for tet resistance.
Recombinant cosmid replicates as a plasmid.


cos
tet^R
E. coli cell
Genomic DNA

Bacterial artificial chromosome

- A bacterial artificial chromosome (BAC) is a DNA construct, based on a functional fertility plasmid (or F-plasmid), used for transforming and cloning in bacteria, usually E. coli.
- F-plasmids play a crucial role because they contain partition genes that promote the even distribution of plasmids after bacterial cell division. The bacterial artificial chromosome's usual insert size is 150-350 kbp.
- BACs are often used to sequence the genome of organisms in genome projects, for example the Human Genome Project.
- A short piece of the organism's DNA is amplified as an insert in BACs, and then sequenced.
- Finally, the sequenced parts are rearranged in silico, resulting in the genomic sequence of the organism.
- BACs were replaced with faster and less laborious sequencing methods like whole genome shotgun sequencing and now more recently next-gen sequencing.

- BACs are capable of maintaining human and plant genomic fragments of greater than 300 kb for over 100 generations with a high degree of stability.
- The first BAC vector, pBAC108L, lacked a selectable marker for recombinants. Thus, clones with inserts had to be identified by colony hybridization.
- Two widely used BAC vectors, pBeloBAC11 and pECBAC1, are derivatives of pBAC108L in which the original cloning site is replaced with a lacZ gene have MCS.
- pBeloBAC11 has two EcoRI sites, one in the lacZ gene and one in the CM^R gene, whereas pECBAC1 has only the EcoRI site in the lacZ gene. Further improvements to BACs have been made by replacing the lacZ gene with the sacB gene.
- Insertional inactivation of sacB permits growth of the host cell on sucrose-containing media, i.e. positive selection for inserts.

Common gene components

repE

- for plasmid replication and regulation of copy number.

parA , parB and parC

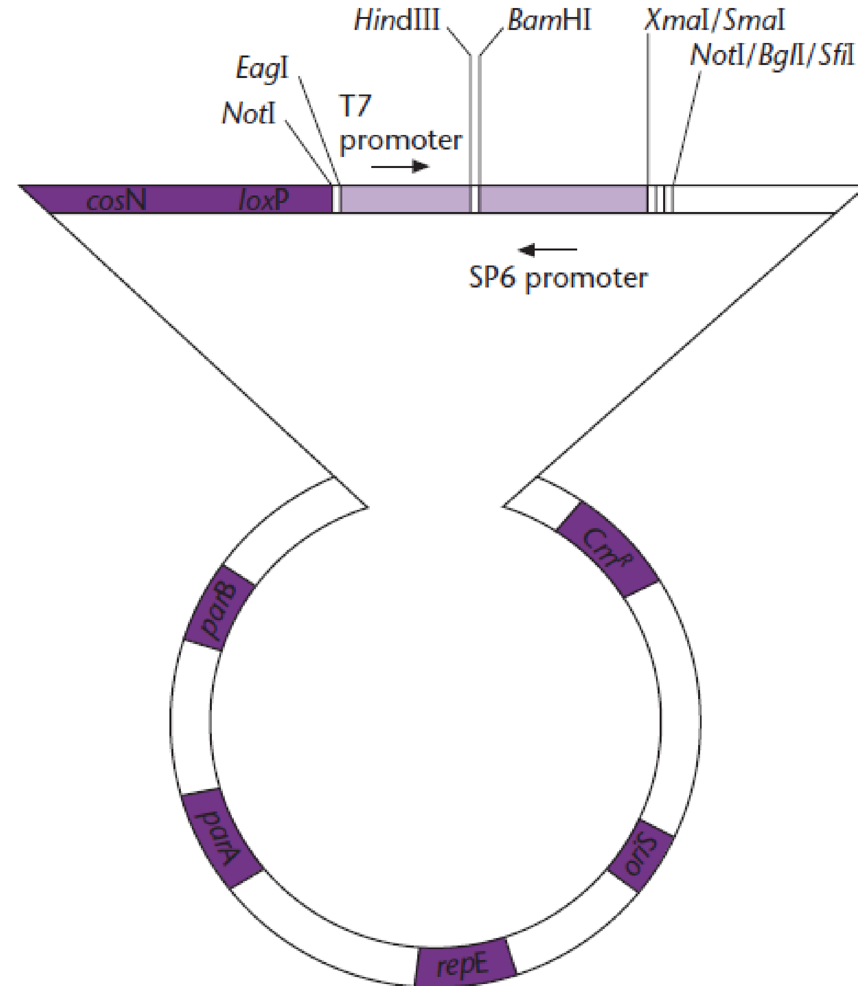
- for partitioning F plasmid DNA to daughter cells during division and ensures stable maintenance of the BAC.

A selectable marker

- for antibiotic resistance; some BACs also have lacZ at the cloning site for blue/white selection.

T7 & Sp6

- phage promoters for transcription of inserted genes.



F plasmid

cloning sites

C m R

origin

BAC vector

restriction endonuclease

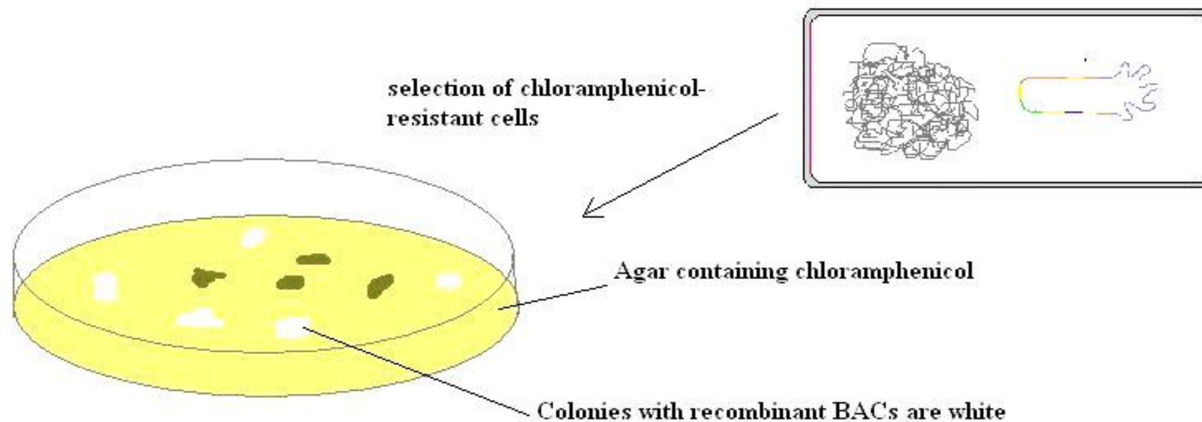
large foreign DNA fragment

DNA ligase

Recombinant BAC

Figure: *BACs as a cloning vector*

electroporation



Usefulness

- BACs are being greatly used in modeling genetic diseases in order to study their effects in the experimentation on transgenic mice. Because complex genes often have many regulatory sequences upstream of the encoding sequence, including many promoter sequences that will control a gene's express level, BACs have been found very useful in this area of study.
- BACs have also been used to detect genes or large sequences of interest, and then used to map them onto the human chromosome using BAC arrays.
- BACs can also act as clones of large DNA viruses as well as RNA viruses.
- The infectious property that the BACs contain has aided in the study of viruses such as herpesviruses, poxviruses, and coronaviruses

Yeast artificial chromosomes (YACs)

- Yeast artificial chromosomes (YACs) are genetically engineered chromosomes derived from the DNA of the yeast, Saccharomyces cerevisiae, which is then ligated into a bacterial plasmid. By inserting large fragments of DNA, from 100–1000 kb, the inserted sequences can be cloned and physically mapped using a process called chromosome walking.
- The primary components of a YAC are the ARS, centromere, and telomeres from *S. cerevisiae*.
- Additionally, selectable marker genes, such as antibiotic resistance and a visible marker, are utilized to select transformed yeast cells.
- A YAC is built using an initial circular DNA plasmid, which is typically cut into a linear DNA molecule using restriction enzymes; DNA ligase is then used to ligate a DNA sequence or gene of interest into the linearized DNA, forming a single large, circular piece of DNA.

Components

- A YAC can be considered as a functional artificial chromosome (self replicating element), since it includes three specific DNA sequences that enable it to propagate from one cell to its offspring:

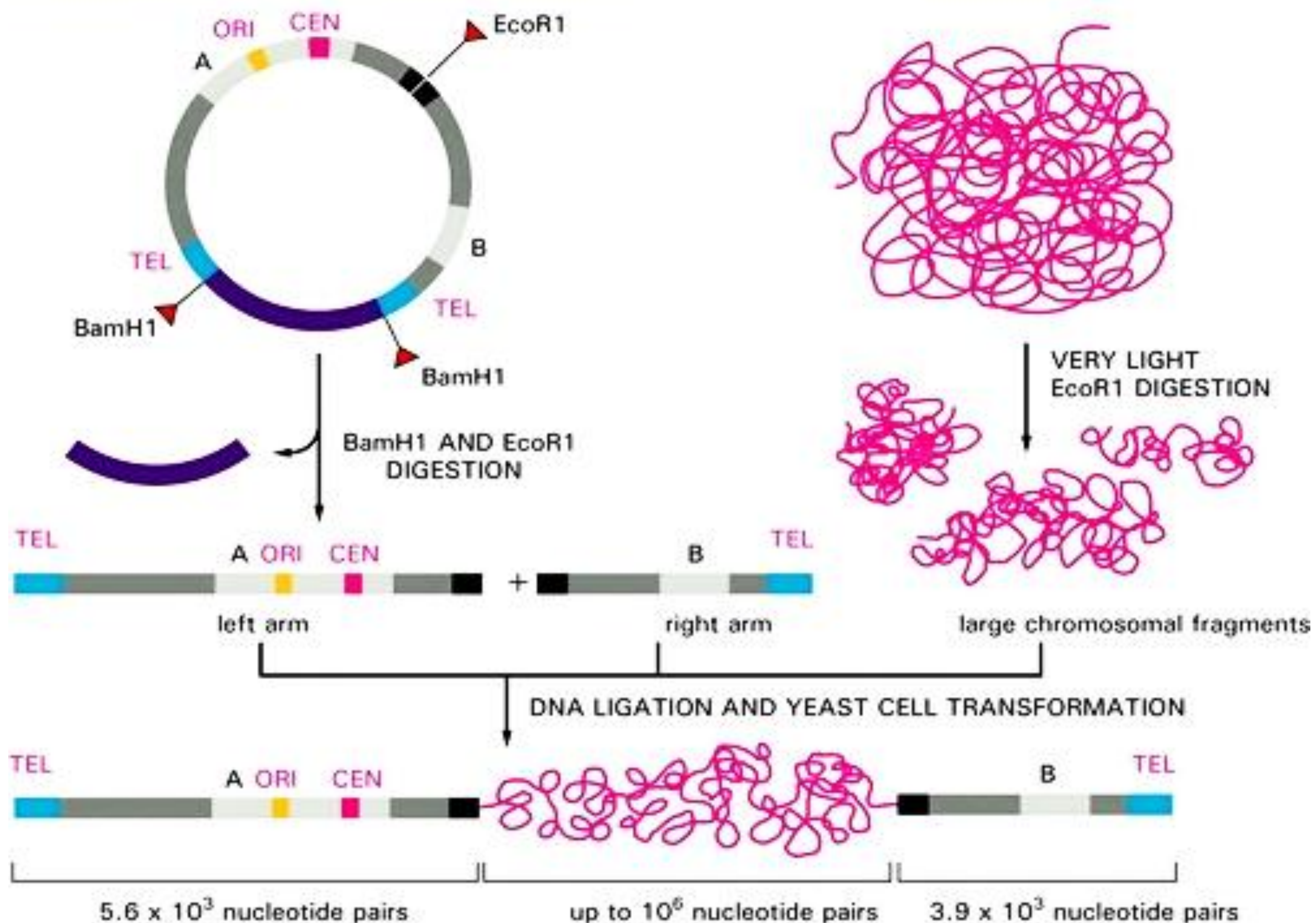
- TEL: The telomere which is located at each chromosome end, protects the linear DNA from degradation by nucleases.
- CEN: The centromere which is the attachment site for mitotic spindle fibers, "pulls" one copy of each duplicated chromosome into each new daughter cell.
- ORI: Replication origin sequences which are specific DNA sequences that allow the DNA replication machinery to assemble on the DNA and move at the replication forks.
- Ligation of selective marker into plasmid vector: this allows for the differential selection of colonies with, or without the marker gene. An antibiotic resistance gene, such as LEU2, allows the YAC vector to be amplified and selected for in E. coli by rescuing the ability of mutant E. coli to synthesize leucine in the presence of the necessary components within the growth medium.

- TRP1 and URA3 genes are other YAC vector cloning site for foreign DNA is located within the SUP4 gene. This gene compensates for a mutation in the yeast host cell that causes the accumulation of red pigment. The host cells are normally red, and those transformed with YAC only, will form colorless colonies. Cloning of a foreign DNA fragment into the YAC causes insertional inactivation of the gene, restoring the red color. Therefore the colonies that contain the foreign DNA fragment are red.

- Ligation of necessary centromeric sequences for mitotic stability

- Ligation of Autonomously Replicating Sequences (ARS) providing an origin of replication to undergo mitotic replication allows the plasmid to replicate extrachromosomally, but renders the plasmid highly mitotically unstable, and easily lost without the centromeric sequences

- Ligation of artificial telomeric sequences to convert circular plasmid into a linear piece of DNA


YEAST ARTIFICIAL CHROMOSOME VECTOR
HUMAN DNA
ORI
CEN
EcoR1
A
B
TEL
BamH1
TEL
BamH1
VERY LIGHT
EcoR1 DIGESTION
BamH1 AND EcoR1
DIGESTION
TEL
A ORI
CEN
B
TEL
left arm
right arm
large chromosomal fragments
DNA LIGATION AND YEAST CELL TRANSFORMATION
TEL
A ORI
CEN
B
TEL
5.6 x 10³ nucleotide pairs
up to 10⁶ nucleotide pairs
3.9 x 10³ nucleotide pairs
ARTIFICIAL YEAST CHROMOSOME WITH INSERTED HUMAN DNA

This is a photo of two copies of the Washington University Human Genome YAC Library. Each of the stacks is approximately 12 microtiter plates. Each plate has 96 wells, each with different yeast clones.

Construction of a genomic library

Plasmid

Vector

Cleavage site

Donor DNA

(a)

Cleavage by *Eco*RI endonuclease

Cleavage sites

Cleavage by *Eco*RI endonuclease

AATT AATT AATT

TTAA TTAA TTAA

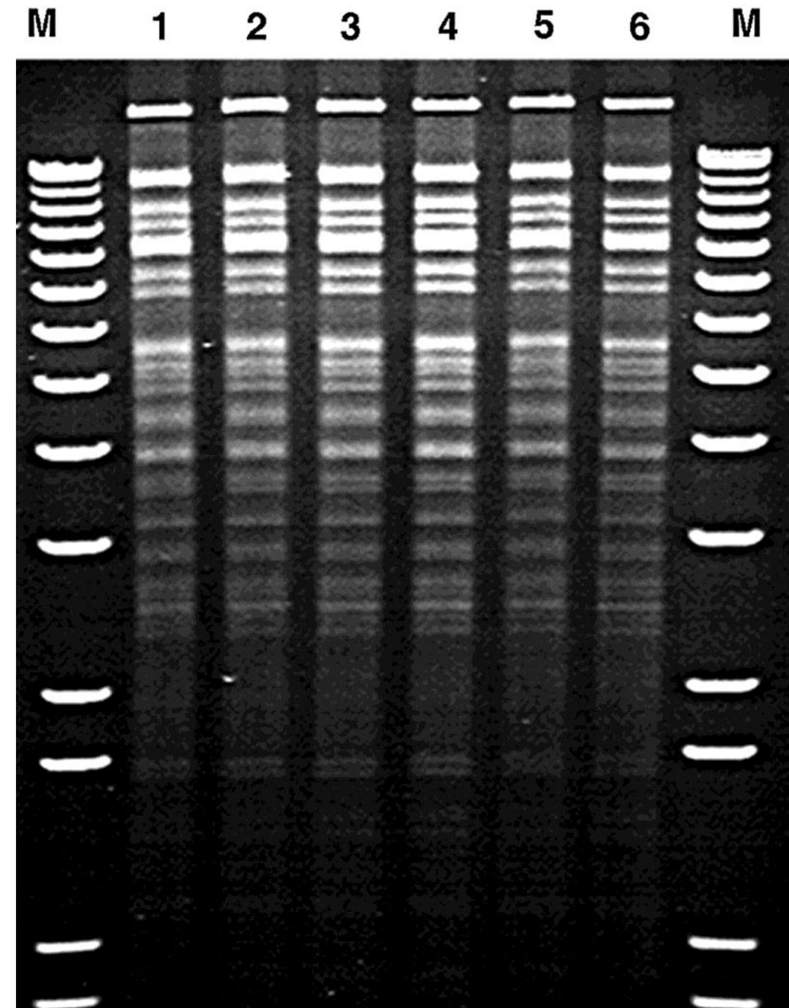
Hybridization

(b)

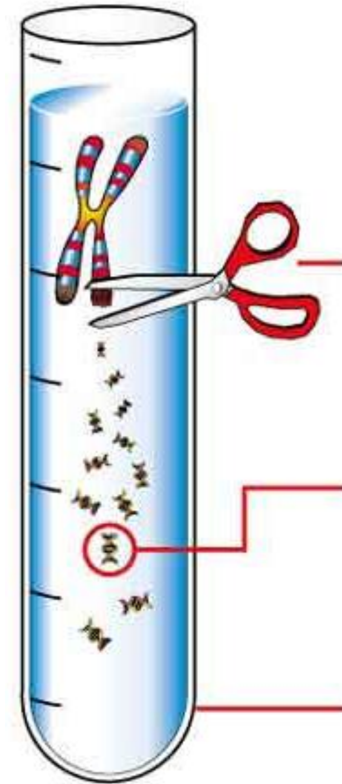
DNA ligase

(c)

Recombinant plasmid



There are several kinds of vectors available with various insert capacities. Generally, libraries made from organisms with larger genomes require vectors featuring larger inserts, thereby fewer vector molecules are needed to make the library. Researchers can choose a vector also considering the ideal insert size to find a desired number of clones necessary for full genome coverage



Vector type	Insert size (thousands of bases)
Plasmids	up to 15
Phage lambda (λ)	up to 25
Cosmids	up to 45
Bacteriophage P1	70 to 100
P1 artificial chromosomes (PACs)	130 to 150
Bacterial artificial chromosomes (BACs)	120 to 300
Yeast artificial chromosomes (YACs)	250 to 2000

Application

- Genomic libraries are commonly used for sequencing applications.
- They have played an important role in the whole genome sequencing of several organisms, including the human genome and several model organisms.
- Genomic library helps in identification of the novel pharmaceutically important genes.
- Genomic library helps in identification of new genes which were silent in the host.
- It helps us in understanding the complexity of genomes.

C-DNA library

© 2009 Encyclopædia Britannica, Inc.

Copyright © 2009 Pearson Education, Inc.

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

- To screen for a clone in a library usually want a 99% probability that your clone is found there.
- Frequency is the size of the DNA fragment in the library/the size of the haploid genome. For a lambda library 17 kb (1.7×10^4) is the average size of library. The size of the genome is

  3×10^9 bp

  $F = 1.7 \times 10^4 / 3 \times 10^9$ bp

  $N = \ln (1-.99) / \ln (1- [1.7 \times 10^4 / 3 \times 10^9])$

  $N = \ln .01 / \ln (1 - 0.56 \times 10^{-5})$

  $N = -4.6061702 / -0.0000056$

  $N = 822,351$ clones