

Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 29

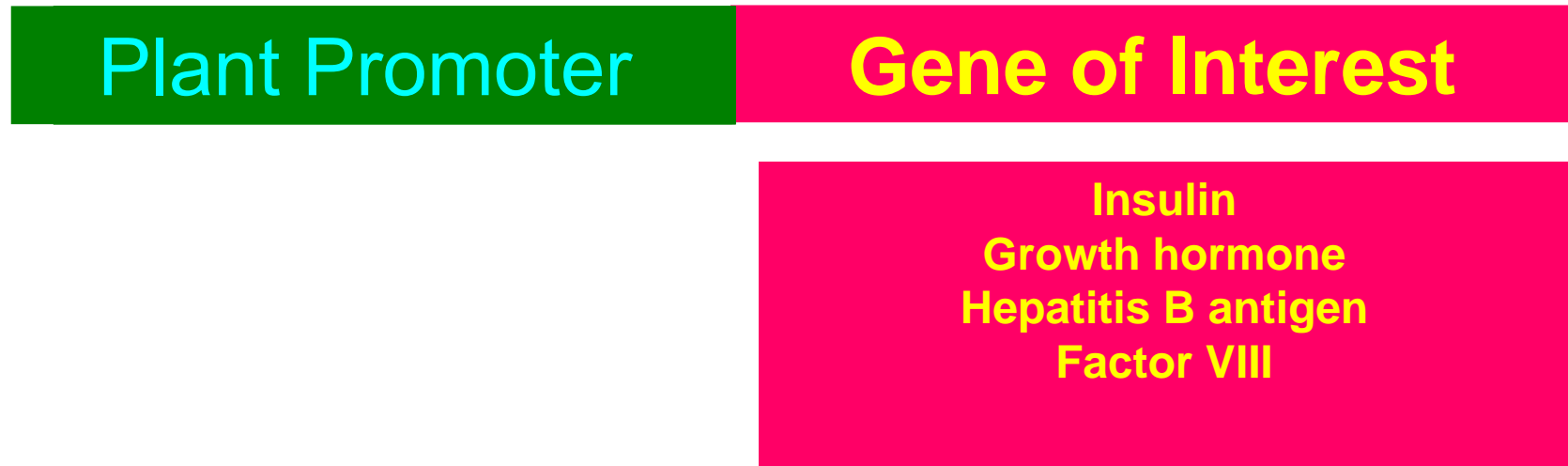
Cloning and Expression vectors

The knowledge that expression of a protein coding gene can be induced by placing this gene downstream of a promoter led to the development of a number of **expression vectors**: both prokaryotic as well as eukaryotic.

Promoter of choice

Gene of Interest

Making recombinant proteins



Expression vectors became the basic tools for biotechnology for the production of recombinant proteins.

What is an expression vector?

An expression vector, is usually a **plasmid** that is used to introduce a specific gene into a target cell.

Once inside the host cell, the gene encoded by the expression vector is transcribed by the host transcription machinery and the RNA thus synthesized is translated by the host translation machinery leading to the synthesis of the specific protein.

An expression plasmid contains regulatory sequences that act as enhancer and promoter regions and lead to efficient transcription of the gene carried on the expression vector.

The goal of a well-designed expression vector is the production of large amounts of mRNA, and therefore proteins.

The design and development of expression vectors and their use in biotechnology for the benefit of mankind was closely linked to key discoveries in the area of:

Recombinant DNA technology or Genetic engineering.

In 1972, Paul Berg in Stanford University, used a restriction enzyme to isolate a gene from a human cancer-causing monkey virus (SV40) and used an enzyme called DNA ligase to join this virus DNA with a molecule of DNA from the bacterial virus lambda, creating the first recombinant DNA molecule.

Berg realized the dangers of his experiment and temporarily terminated it before the recombinant DNA molecule was added to *E. coli*, where it would have been quickly reproduced.

Following the termination of his experiment, he proposed a one year moratorium on recombinant DNA studies while safety issues were addressed.

Berg later resumed his studies into recombinant DNA techniques, and was awarded the 1980 Nobel Prize in chemistry.



Jackson, D.A., Symons, R.H. and Berg, P., "Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA containing lambda phage genes and the galactose operon of *Escherichia coli*," *Proc. Nat. Acad. Sci. USA* **69**, pp. 2904-2909 (1972).



Stanley Cohen



Herbert Boyer

Around the same time, Paul Berg was generating recombinant DNA. Stanley Cohen and Herbert Boyer met at a scientific conference in Hawaii on plasmids, which are small circular extra chromosomal DNA molecules encoding genes such as those coding for antibiotic resistance in bacteria.

Boyer's lab had isolated an enzyme (now known as **RESTRICTION ENZYMES**) that could be used to precisely cut DNA into segments which can then be ligated to other strands of DNA using another enzyme called **DNA LIGASE**

Cohen had developed a method to introduce antibiotic-carrying plasmids into certain bacteria, as well as a method of isolating and cloning genes carried by the plasmids.

By combining their expertise, Boyer and Cohen realized that by combining their expertise, they can introduce genes into bacterial cells and express them.



Stanley Cohen



Herbert Boyer

In 1973-1974 Stanley N. Cohen of Stanford and Herbert W. Boyer of the University of California, San Francisco, described a laboratory process for joining and replicating DNA from different species.

In 1974, Stanford and UC applied for a patent ([U.S. patent no. 4,237,224](#)) on the recombinant DNA process; the U.S. Patent Office granted it in 1980.

Two of the most profitable patents in the biotechnology area are those of Cohen and Boyer, which are owned by Stanford University.

These patents cover the fundamental technology used throughout molecular biology, including recombinant DNA research. As of early 1995, these patents generated an income of \$139 million in royalty payments.

Boyer went on to start **Genentech**, the world's first biotechnology company.

Genentech went on to produce human insulin in bacteria in September 1978 and licensed the production method to Eli Lilly and Company.

By 1980, a new company, Biogen, produced interferon through recombinant DNA

By 1988, five proteins from genetically engineered cells were approved as drugs by the United States Food and Drug Administration (FDA): synthetic insulin, human growth hormone, hepatitis B vaccine, alpha-interferon, and tissue plasminogen activator (TPa), for lysis of blood clots.

However, by the end of the 1990s, however, 125 more genetically engineered drugs would be approved.

Over \$350 billion has been invested in biotech since the emergence of the industry, and global revenues rose from \$23 billion in 2000 to more than \$50 billion in 2005.

The growth of recombinant DNA technology and biotech industry was facilitated by the United States Supreme Court ruling in 1980 in favor of microbiologist Ananda Chakrabarty in the case of a US patent office request for a first patent granted to a genetically modified living organism (GMO) in history.

First biotech companies formed:

1976 - Genentech

1978 - Biogen

1980 - Amgen

1981 - Immunex

1981 - Chiron

1981 - Genzyme

- ✓ **More than 325 million people worldwide have been helped by 160 approved biotech drugs and vaccines**
- ✓ **>350 more biotech drugs and vaccines now in clinical trials targeting more than 200 diseases**
- ✓ **Biotechnology is responsible for hundreds of diagnostic tests, including HIV tests and home pregnancy tests, DNA fingerprinting...**

**LET US UNDERSTAND HOW DOES ONE DESIGN VECTORS
FOR CLONING GENES
AND
MAKING RECOMBINANT PROTEINS**

PLASMID AND PHAGE VECTORS

A vector is used to amplify a single molecule of DNA into many copies.

A DNA fragment must be inserted into a cloning vector.

A cloning vector is a DNA molecule that has an origin of replication and is capable of replicating in a bacterial cell.

General Features of Vectors

Vectors must be able to replicate autonomously in the host cell and must have at least one usable restriction enzyme site for the introduction of foreign DNA.

In addition, a vector should also have:

- selectable markers (to select for cells that have taken up the vector DNA);
- a range of restriction enzyme sites available for cloning,
- insertional inactivation indicator genes (i.e., genes that are insertionally disrupted by the cloning process to indicate that cloning has occurred);
- a promoter (preferably a strong one) upstream of the cloning site;
- a terminator (strong enough to match the strength of the promoter) downstream of the cloning site (this makes the vector an expression vector);
- an origin of replication

Plasmids

Plasmids are circular, double-stranded DNA molecules that exist in bacteria and in the nuclei of some eukaryotic cells.

They can replicate independently of the host cell.

The size of plasmids ranges from a few kb to near 100 kb

They can hold up to 10 kb fragments

Plasmids may encode a wide variety of genetic determinants, which permit their bacterial hosts to survive better in an adverse environment or to compete better with other microorganisms occupying the same ecological niche.

The plasmids have enormous medical importance since some of them encode genes for antibiotic resistance, as well as specific virulence traits.

Plasmids rely on host-encoded factors for their replication. Plasmid replication initiates in a predetermined *cis*-site called *ori*.

Plasmid Vectors

Psc101

pBR322

pUC8/9, 18/19

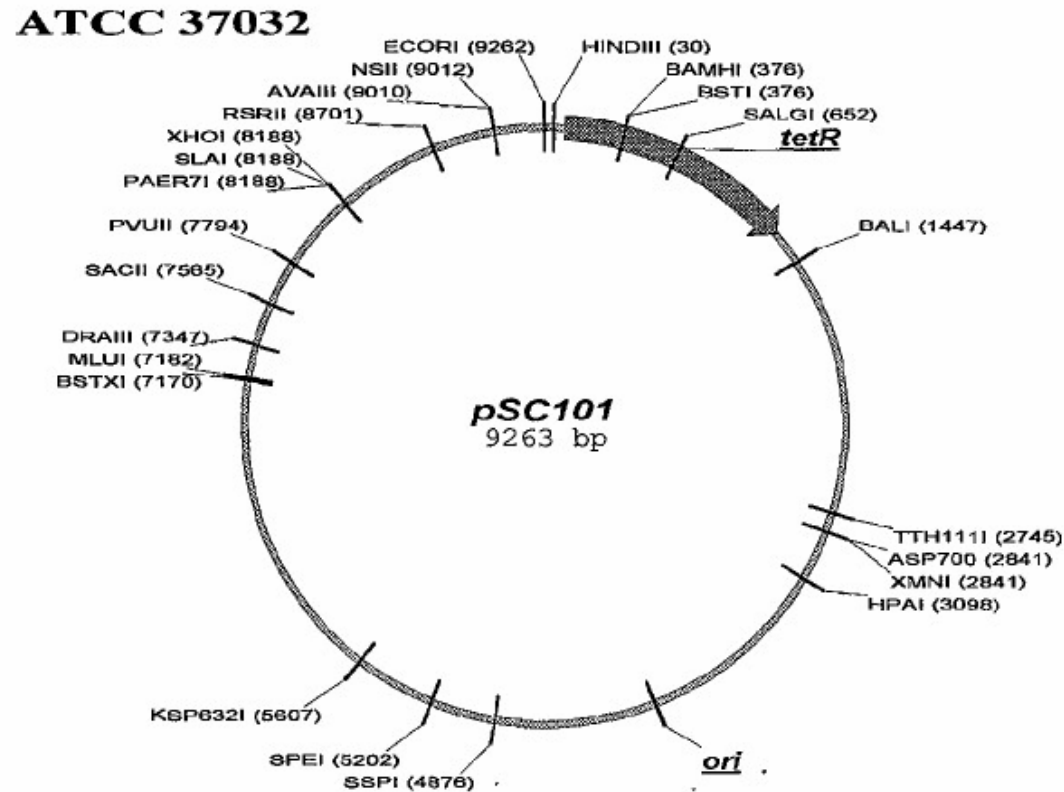
Shuttle vectors

FIRST GENERATION PLASMID VECTORS

pSC101

This is the earliest plasmid vector made by [Stanley Cohen](#).

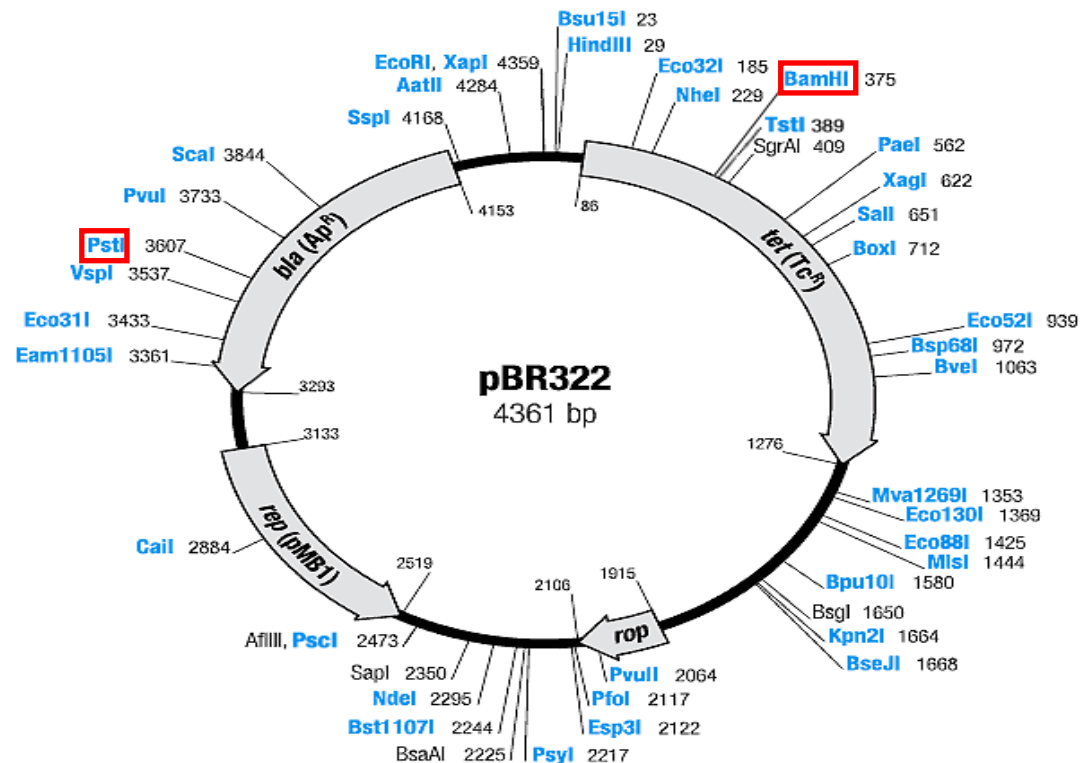
This was an 8.9 kb low copy number plasmid which had a single restriction site and a tetracycline resistance marker.



SECOND GENERATION PLASMID VECTORS

pBR322

- This was the next generation plasmid of 4.0 kb size
- It was a low copy number plasmid, which had many restriction enzyme sites (EcoRI, BamHI, PstI, HindIII, and Sall). These were located in either of two antibiotic resistance markers, ampicillin and tetracycline.
- Cloning into one of the restriction enzyme sites would inactivate one of the antibiotic resistance markers, leaving the other for selection of transformation (insertional inactivation).
- Screening for the absence of the second antibiotic marker was putative evidence of a successful cloning event. The lower molecular weight allowed cloning of larger fragments.



The plasmid pBR322 used to be one of the most commonly used *E.coli* cloning vectors.

the replicon *rep* responsible for the replication of plasmid

rop gene coding for the Rop protein, which promotes conversion of the unstable RNA I - RNA II complex to a stable complex and serves to decrease copy number

bla gene, coding for beta-lactamase that confers resistance to ampicillin

tet gene, encoding tetracycline resistance protein

THIRD GENERATION PLASMID VECTORS

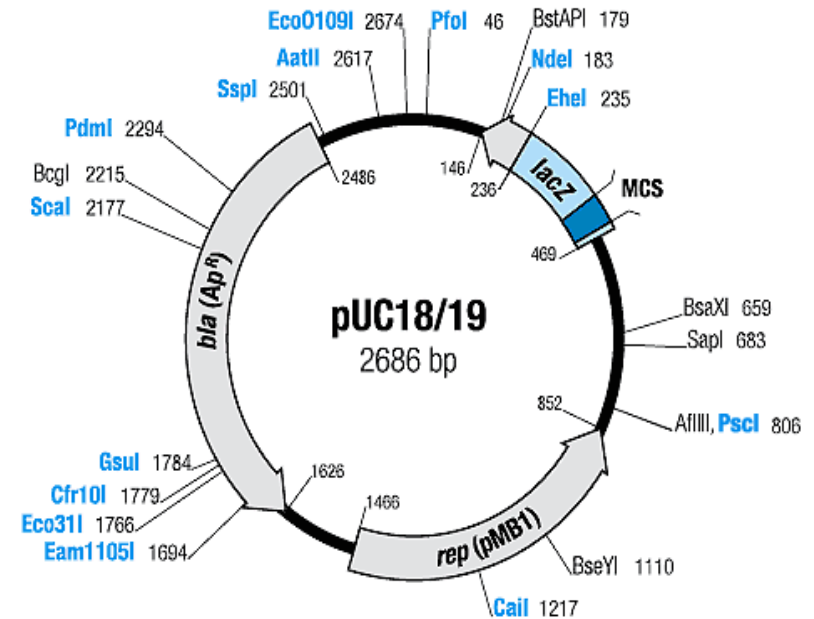
pUC8/9, 18/19

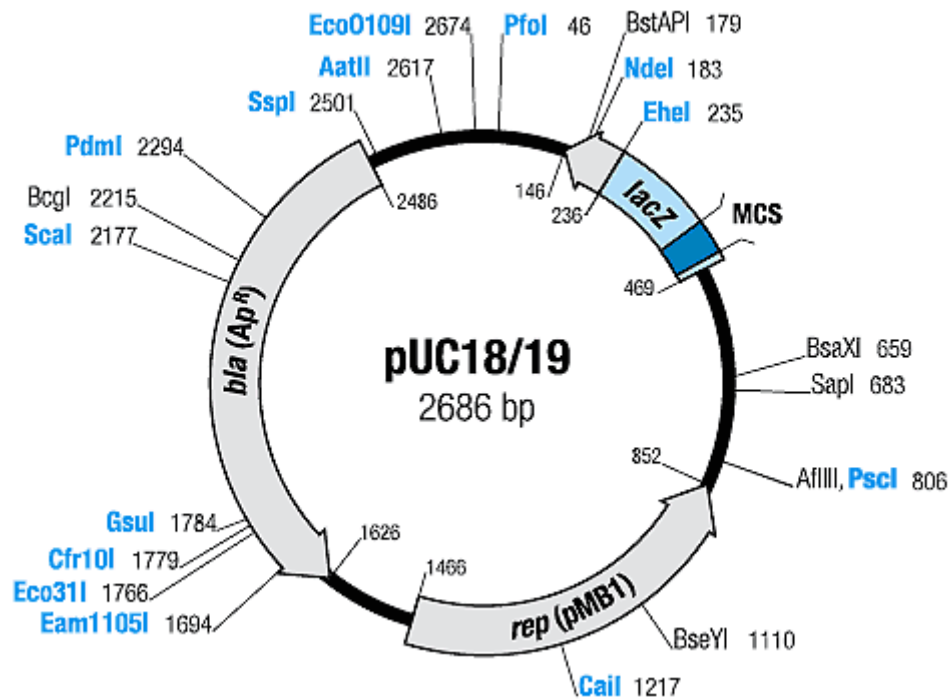
The pUC plasmids were engineered from the pBR322 origin of replication to include the alpha portion of the beta-galactosidase gene (*lacZ'*), complete with its promoter. The beta portion of *lacZ* was included in the chromosome of the host, so that the host containing the plasmid was Lac⁺.

The restriction enzyme sites (6 in pUC8/9; 10 in pUC18/19) are clustered together on an oligonucleotide (called the polylinker, or multiple cloning site), which is in-frame within the first few amino acids of the *lacZ'* reporter gene (thus providing insertional inactivation, Lac⁺ → Lac⁻).

The native *lacZ* promoter (*P_{lac}*) is situated just upstream of the cloned gene, allowing expression (although not over-expression) of genes on inserts that are correctly oriented. Most of the nonessential DNA has been removed to provide the ability to clone larger fragments.

- An ampicillin marker is included for selection of transformants.





M13/pUC sequencing primer (-20), 17-mer → 399 HindIII PaeI SdaI BveI PstI HincII Sall XmiI XbaI BamHI Cfr9I Eco88I SmaI Acc65I KpnI Eco136II Eco24I SacI EcoRI XapI 455

5' G TAA AAC GAC GGC CAG TGC CAA GCT TGC ATG CCT GCA GGT CGA CTC TAG AGG ATC CCC GGG TAC CGA GCT CGA ATT CGT AAT CAT GGT CAT AGC TGT TTC CTG 3'

3' C ATT TTG CTG CCG GTC ACG GTT CGA ACG TAC GGA CGT CCA GCT GAG ATC TCC TAG GGG CCC ATG GCT CGA GCT TAA GCA TTA GTA CCA GTA TCG ACA AAG GAC 5'

LacZ ← Val Val Ala Leu Ala Leu Ser Ala His Arg Cys Thr Ser Glu Leu Pro Asp Gly Pro Val Ser Ser Ser Asn Thr Ile Met Thr Met

M13/pUC reverse sequencing primer (-26), 17-mer ←

M13/pUC sequencing primer (-20), 17-mer → 396 EcoRI XapI Eco136II Eco24I SacI Acc65I KpnI Cfr9I Eco88I SmaI BamHI XbaI HincII Sall XmiI BveI PstI SdaI PaeI HindIII 452

5' G TTG TAA AAC GAC GGC CAG TGA ATT CGA GCT CGG TAC CCG GGG ATC CTC TAG AGT CGA CCT GCA GGC ATG CAA GCT TGG CGT AAT CAT GGT CAT AGC TGT TTC CTG 3'

3' C AAC ATT TTG CTG CCG GTC ACT TAA GCT CGA GCC ATG GGC CCC TAG GAG ATC TCA GCT GGA CGT CCG TAC GTT CGA ACC GCA TTA GTA CCA GTA TCG ACA AAG GAC 5'

LacZ ← Leu Val Val Ala Leu Ser Asn Ser Ser Pro Val Arg Pro Asp Glu Leu Thr Ser Arg Cys Ala His Leu Ser Pro Thr Ile Met Thr Met

M13/pUC reverse sequencing primer (-26), 17-mer ←

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.

bla gene, coding for beta-lactamase that confers resistance to ampicillin (source - plasmid pBR322). It differs from that of pBR322 by two point mutations.

The region of *E.coli* operon *lac* containing CAP protein binding site, promoter Plac, *lac* repressor binding site and 5'-terminal part of the *lacZ* gene encoding the N-terminal fragment of beta-galactosidase.

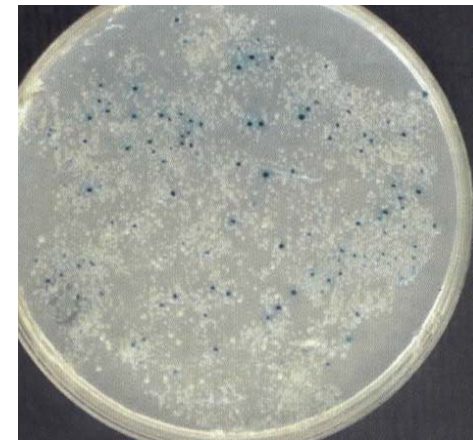
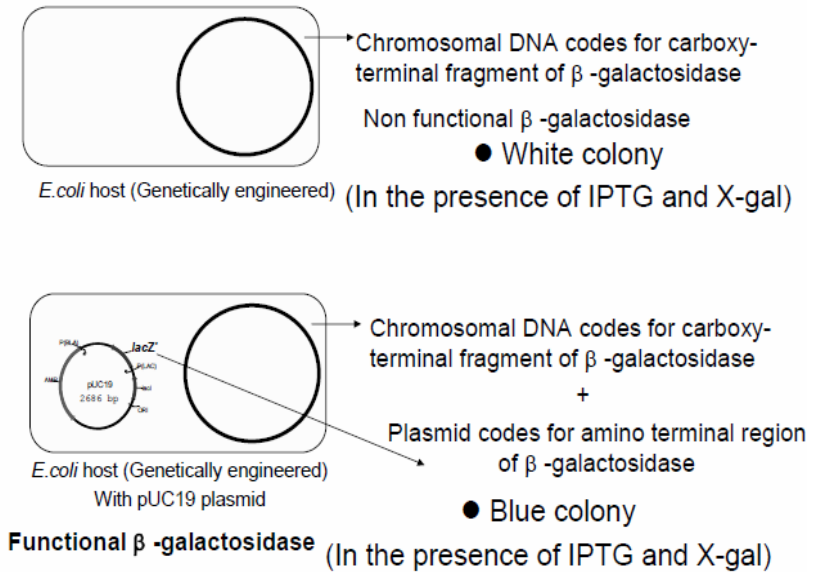
This fragment, whose synthesis can be induced by IPTG, is capable of intra-allelic α complementation with a defective form of beta-galactosidase encoded by host (mutation *lacZ*^DM15).

In the presence of IPTG, bacteria synthesise both fragments of the enzyme and form blue colonies on media with X-Gal.

Insertion of DNA into the MCS located within the *lacZ* gene (codons 6-7 of *lacZ* are replaced by MCS) inactivates the N-terminal fragment of beta-galactosidase and abolishes alpha-complementation.

Bacteria carrying recombinant plasmids therefore give rise to white colonies.

α -complementation



Shuttle vectors

Shuttle vectors are plasmid vectors that have origins of replication for more than one cloning host.

For example, pMK3/4 has a gram-positive ori for cloning in *Bacillus subtilis*, and a gram-negative ori for cloning in *E. coli*.

In this way, genetic engineering may be done in *E. coli* where it is easier and transferred to *B. subtilis* for final expression and excretion.

Many **eukaryotic expression vectors** are shuttle vectors since they are first produced in *E. coli* and then introduced into the eukaryotic host.

Limitations of Plasmid vectors

- 1) Can accommodate only small inserts (<10 Kb)
- 2) Low efficiency of transformation
- 3) Large colony size
Can screen fewer recombinants per plate
(~5,000 colonies /plate)

Virus (Bacteriophage) Vectors

Lambda

M13

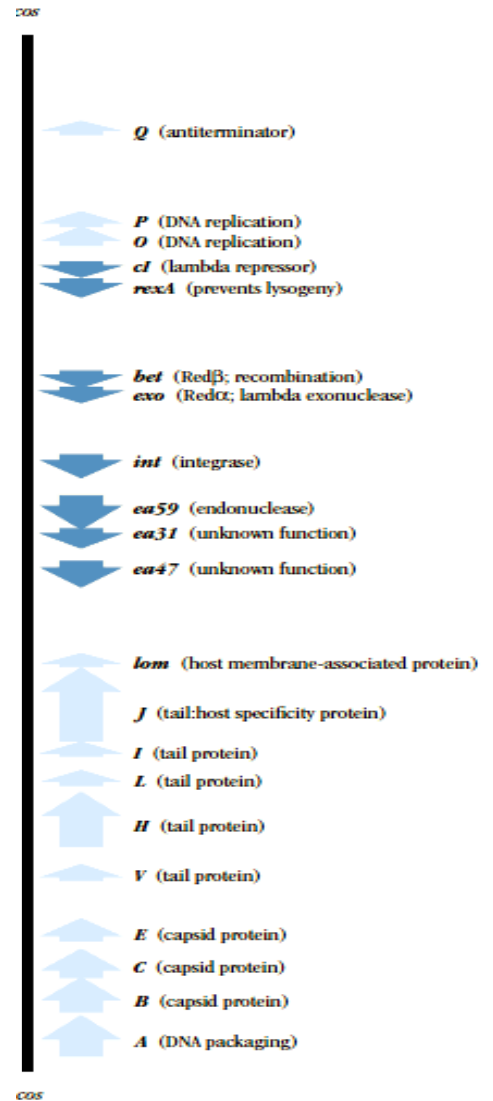
Lambda (λ)

One of the best understood bacteriophages, lambda, was a natural candidate for development into a molecular cloning vector.

Lambda (λ) is a large, temperate *E. coli* bacteriophage with a linear, largely double-stranded DNA genome. At each end, the 5' strand overhangs the 3' strand by 12 bases. These single-stranded overhangs are complementary and anneal to form a *cos* site following entry into a host cell. Once annealed, the genome is a circular, completely double-stranded molecule which serves as a template for rolling-circle replication.

The **Charon phages** contain replaceable regions that are exchanged for the cloned target DNA and simultaneously remove reporter gene(s), such as *lac* or *bio*, allowing screening for putative candidate clones.

Another mechanism to detect the cloning event makes use of the fact that lambda requires a certain size of genome in order to package DNA into the phage head. Removal of the replaceable region leaves a genome too small to package; i.e., it must be replaced with DNA to produce a viable phage. Various size replaceable regions allow a range of insert DNAs.



Lambda map. 48,502 bp
(GenBank Acc #:NC_001416)

Many lambda-based vectors are sold by companies such as:

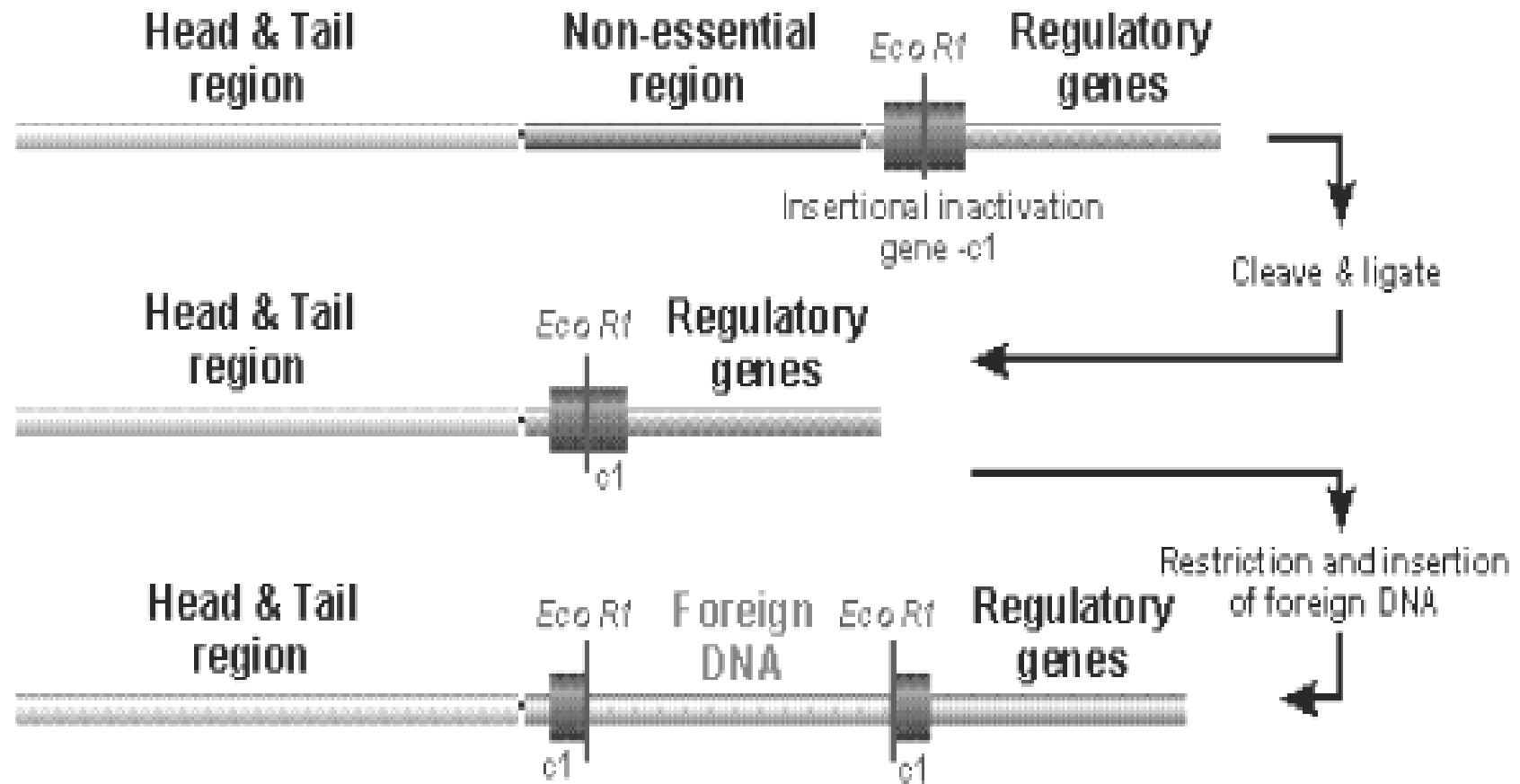
New England Biolabs,

Clontech,

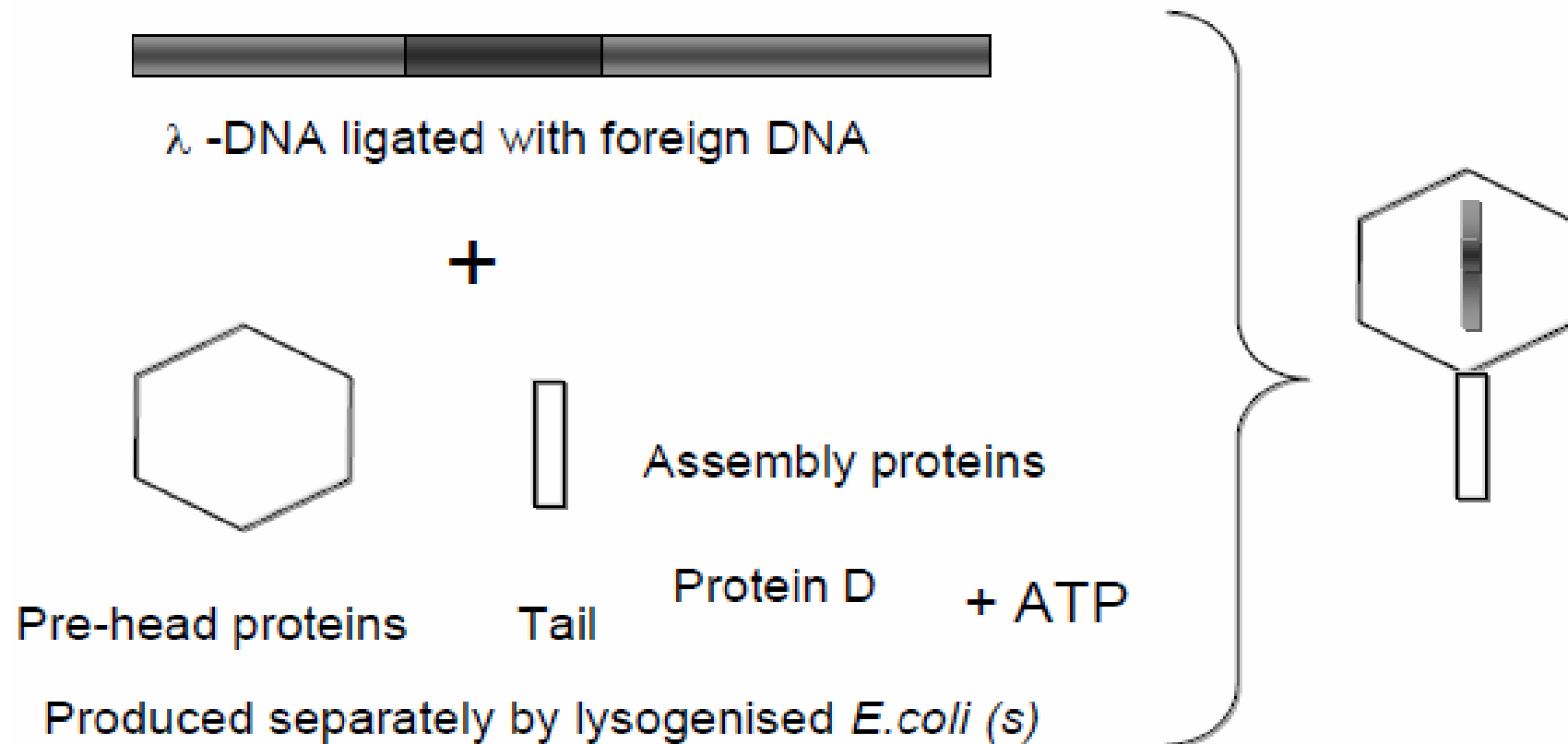
Stratagene etc.

Lambda Insertion Vectors

Example - λ gt10



in-vitro Packaging of λ -DNA



Advantages of λ virus vectors:

- 1) Accepts large inserts (25 Kb)
- 2) Efficient infection (compared to plasmids)
- 3) Produce tiny plaques
 Can screen (~50,000 plaques/plate)

Disadvantages of λ virus vectors:

- 1) Not as easy as working with plasmids

1. Echols, H. and Murialdo, H. (1978) *Microbiol. Rev.* 42, 577–591.
2. 2. Szybalski, E.H. and Szybalski, W. (1979) *Gene* 7, 217–270.
3. 3. Daniels, D.L., de Wet, J.R. and Blattner, F.R. (1980) *J. Virol.* 33, 390–400.
4. 4. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B. (1982) *J. Mol. Biol.* 162, 729–773.
5. Daniels, D.L. et al. (1983) in: Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (eds.) *Lambda II: Appendix*, Cold Spring Harbor Press, New York.

M13

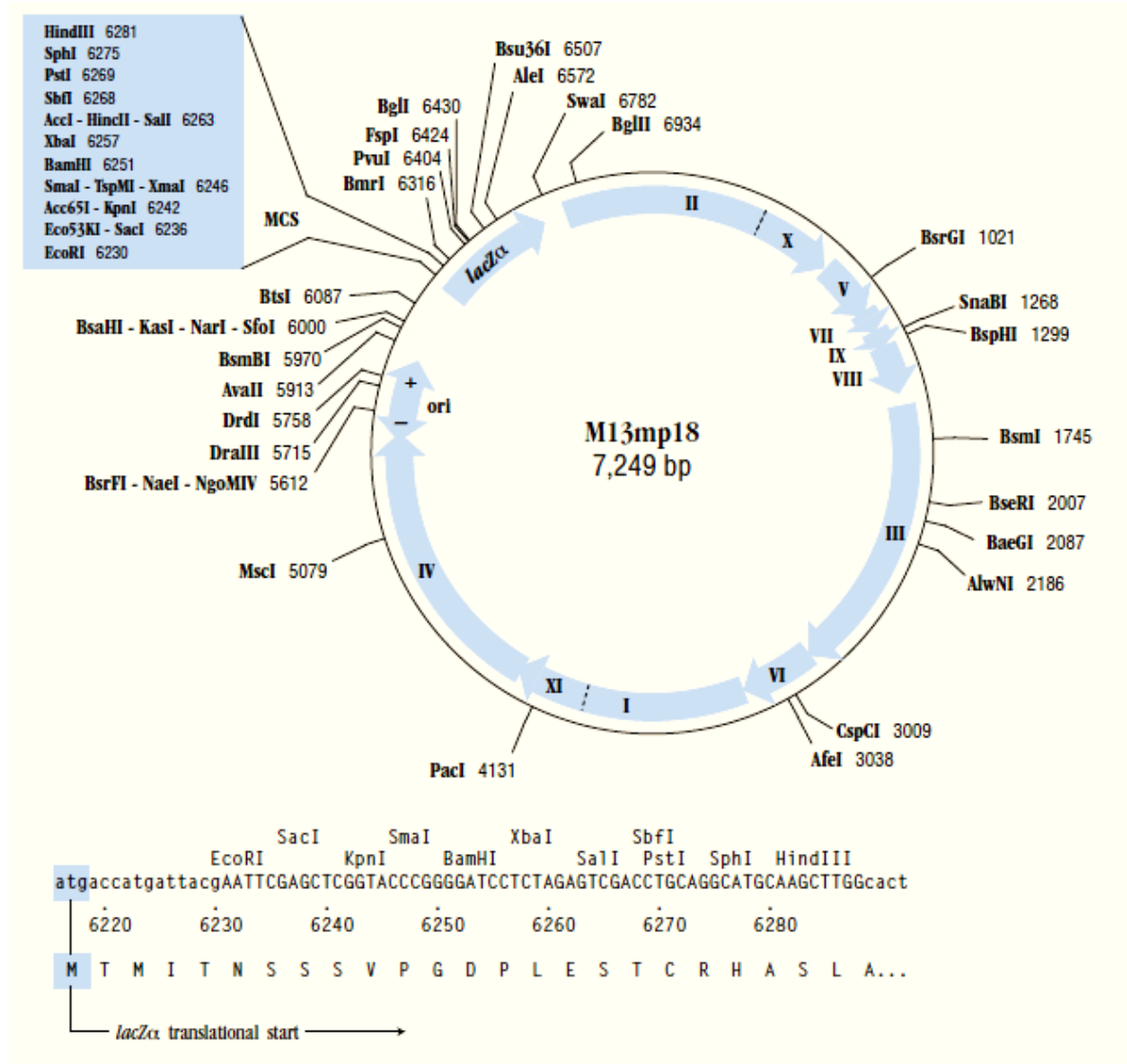
- M13 is a single-stranded DNA virus, but has a double-stranded replicative form of DNA

- It has been engineered to contain a reporter gene (*lacZ'*) and multiple cloning sites to make the M13mp series of vectors.

- The MCS is within the *lacZ'* gene and cloning results in insertional inactivation ($lac^+ \rightarrow Lac^-$).

- The fact that the phage is packaged as single strands makes the cloned DNA easy to sequence or mutate by oligonucleotide site-directed mutagenesis.

- The packaging origin may be reversed to allow packaging of either strand of the double stranded form.



http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/dna_sequences_maps.asp

Cosmid vectors

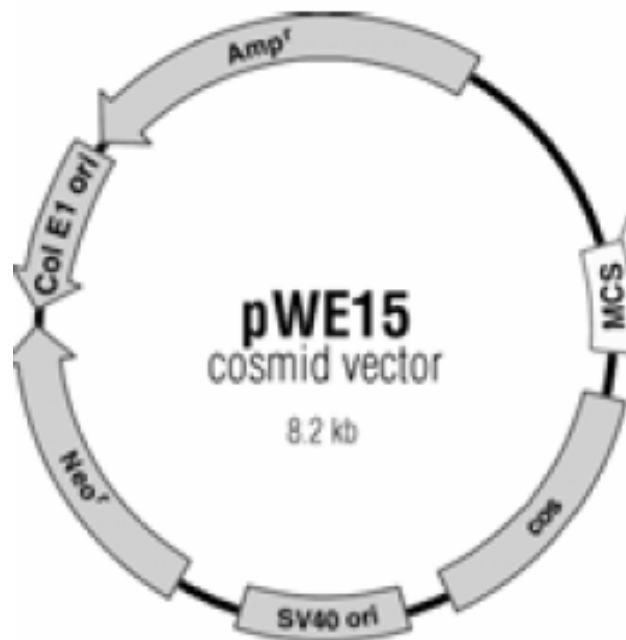
These are plasmids that contain the cohesive ends (cos) of bacteriophage lambda.

These allow packaging of DNA into lambda phage heads.

Cosmids:

- 1) Package large inserts and thus are ideal for genomic libraries
- 2) Have a selectable marker
- 3) Have an *ori*
- 4) Have a polylinker with multiple cloning sites
- 5) have *cos* sequences (needed for packaging into empty virus heads)
- 6) Are packaged into defective phage (virus) particles
- 7) Produce colonies (NOT plaques!)

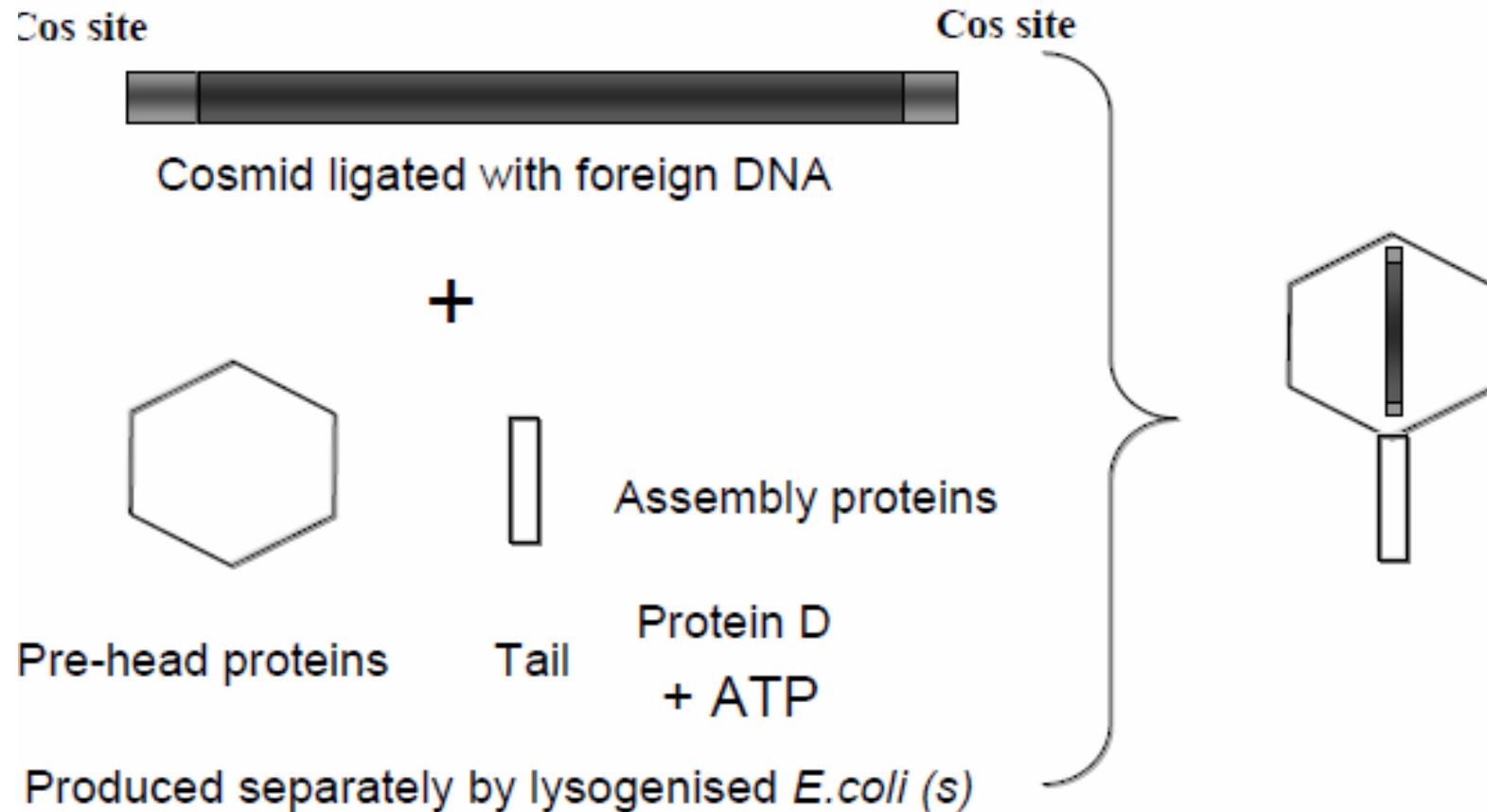
Cosmid



© Stratagene

- Plasmid with lambda phage cos site
- Upto 40 kb inserts can be cloned in cosmid
- Foreign DNA cloned, packaged and infected into *E.coli*
- but in *E.coli*, it propagates like a plasmid
- pWE15 carries neomycin marker and SV40 promoter for neomycin gene expression in eukaryotic cells

in-vitro Packaging of Cosmid



Phasmids

Phasmids are hybrid combinations of phages and plasmids.

Cosmids are an example.

Another example is pUC118/119, which has incorporated the M13 packaging origin site into pUC18/19. This allows plasmids (and their cloned DNA) to be packaged as ssDNA into M13 phage heads.

Bluescript M13, Bluescript SK (+/-), Bluescribe (Stratagene) and pTZ vectors (GE Healthcare) are other examples of these phasmids, but they use the ssDNA packaging origin of another ssDNA phage, f1.

Bacterial Artificial Chromosomes(BACs)

- BACs can hold up to 300 kbs.
- The F factor of E.coli is capable of handling large segments of DNA.
- Recombinant BACs are introduced into E.coli by electroportation (a brief high-voltage current). Once in the cell, the rBAC replicates like an F factor.
- Example: pBAC108L
- Has a set of regulatory genes, OriS, and repE which control F-factor replication, and parA and parB which limit the number of copies to one or two.
- A chloramphenicol resistance gene, and a cloning segment.

Yeast Artificial Chromosomes(YACs)

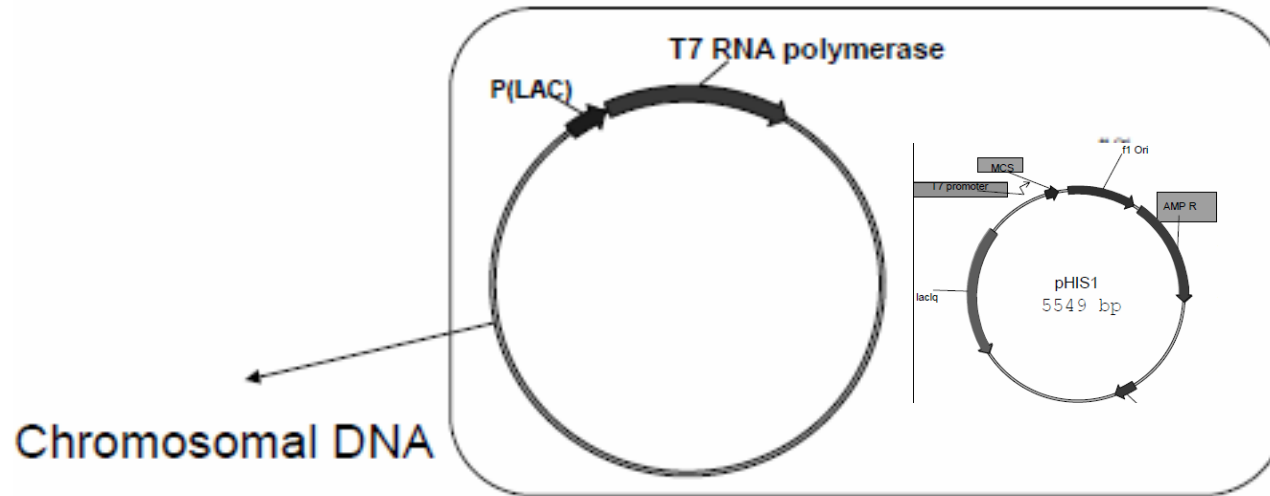
- YACs can hold up to 500 kbs.
- YACs are designed to replicate as plasmids in bacteria when no foreign DNA is present. Once a fragment is inserted, YACs are transferred to cells, they then replicate as eukaryotic chromosomes.
- YACs contain: a yeast centromere, two yeast telomeres, a bacterial origin of replication, and bacterial selectable markers.
- YAC plasmid → Yeast chromosome
- DNA is inserted to a unique restriction site, and cleaves the plasmid with another restriction endonuclease that removes a fragment of DNA and causes the YAC to become linear. Once in the cell, the rYAC replicates as a chromosome, also replicating the foreign DNA.

Expression Vectors

- Expression vectors may be constructed from either plasmid or virus vectors.
- The purpose of these vectors is to overexpress the protein from a cloned gene in a particular host.
- This is accomplished by placing a strong promoter upstream of the cloned gene and a strong terminator at the end of the gene. Phage gene promoters, such as lambda L (leftward promoter) or a T7 promoter, have been used respectively in pAS vectors and pET vectors or pT7 vectors.
- Modified prokaryote promoters, such as Ptac, have been used to construct expression vectors

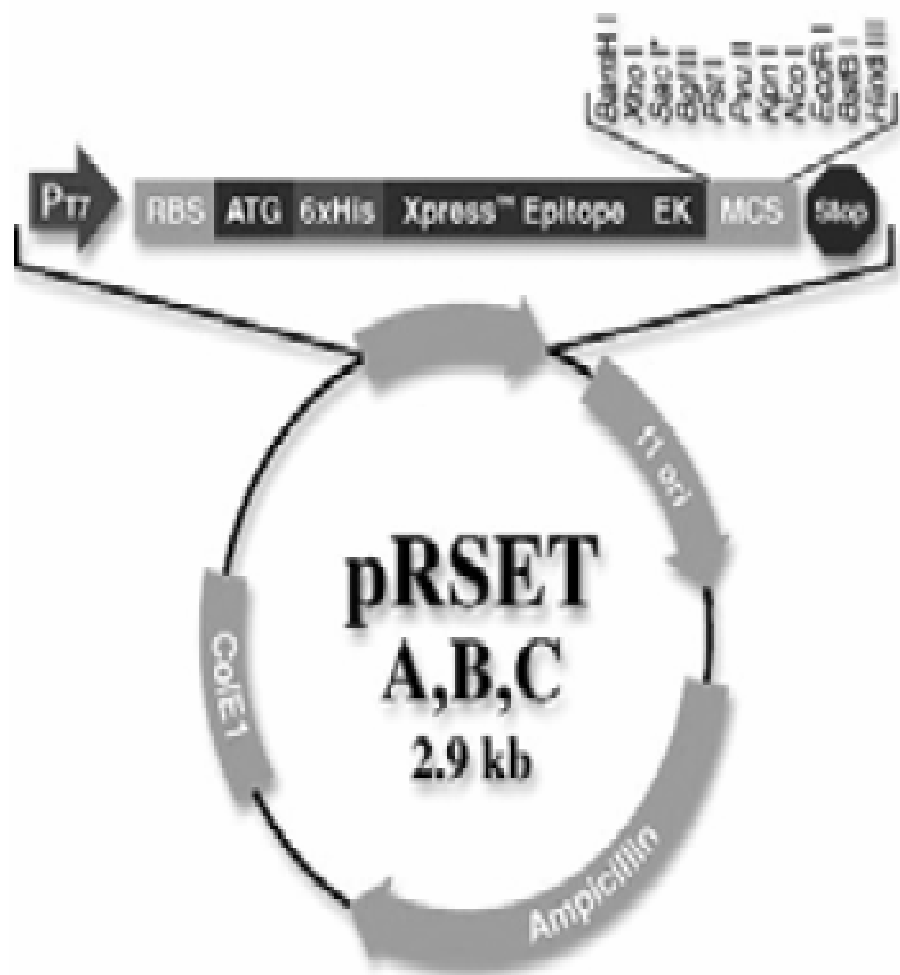
T7 promoter-based expression vectors

Mechanism of T7 expression system



BL21 (DE3) *E. coli* host

- T7 RNA polymerase gene is from T7 Phage, not usually present in *E. coli*.
- T7 RNA pol gene is integrated into the chromosome of *E. coli* using a temperate phage DE3
- T7 RNA polymerase gene is under the control of a lac promoter. Thus, induction by IPTG, causes the host to produce T7 RNA polymerase
- The activity of T7 RNA polymerase is higher than *E. coli* RNA polymerase, hence the protein expression by T7 expression vector is always higher



- High-level expression from the bacteriophage T7 promoter
- N-terminal polyhistidine (6xHis) tag for rapid purification by metal affinity chromatography
- N-terminal Xpress epitope for protein detection with the Anti-Xpress Antibody
- Enterokinase cleavage site for removal of fusion tag
- f1 origin for ssDNA rescue to allow easy sequencing and mutagenesis

<p>Cloning vector (pUC19)</p>	<p>Resistance marker -Ampicillin Origin of replication Multiple cloning site(MCS) lacZ lacI</p>	<p>Cloning a gene Screening by α-complementation and insertional inactivation ..Blue/White colony screening</p>
<p>Expression vector pHis, pRSET</p>	<p>R, O, M Ampicillin resistance Strong promoter (T7) 6 HIS- affinity tag Enzyme cleavage site F1 origin of replication</p>	<p>Recombinant protein production and purification using nickel column</p>
<p>Transcription vector- (pSP72)</p>	<p>R, O, M Ampicillin resistance T7 promoter SP6 promoter</p>	<p>Sense and antisense RNA probes by <i>Invitro</i>- transcription <i>Invitro</i>-translation</p>
<p>PCR cloning vector (pTOPO 2.1)</p>	<p>R, O, M Ampicillin & Kanamycin resistance "T" overhang F1 origin of replication T7 promoter M13 reverse priming site</p>	<p>Cloning PCR product Invitro transcription -complementation and insertional inactivation Blue/White colony screening</p>

<p>Secretion vector pET21</p>	<p>R, O, M Ampicillin resistance F1 origin of replication T7 promoter PeIB signal peptide 6 HIS- affinity tag Enzyme cleavage site</p>	<p>Recombinant protein production, secretion and purification using nickel column</p>
<p>Cosmid pWE15</p>	<p>R, O, M Cos site (Does not have the lambda phage gene except COS site) Can be package with head and tail</p>	<p>Upto 40 kb inserts can be cloned Used for making gene libraries</p>
<p>Lambda insertion vector λgt10</p>	<p>MCS in the c1 repressor gene Can be package with head and tail</p>	<p>Upto 8-23 kb inserts can be cloned Used for making gene libraries</p>
<p>Filamentous phage</p>	<p>Single stranded DNA Infect E coli having pilus (F')</p>	<p>useful for sequencing, in-vitro mutagenesis Upto 3 kb inserts can be cloned</p>
<p>Phagemid</p>	<p>Phage and plasmid properties Contains M13 initiator and terminator</p>	<p>Used for making gene libraries Converting into plasmids with out subcloning</p>

Eukaryotic Expression Vectors

Generally similar to prokaryotic ones

Commonly start with prokaryotic vector

- Construction done in *E. coli*

- Shuttle vector

- Origin

- MCS

- Selectable marker

- High expression regulatable promoter

- But also generally an intron included

- Especially for cDNA clones

Eukaryotic expression vector

Advantages

- Post translational modification eg. Glycosylation
- Disulfide bond formation
- Proper protein folding
- Endotoxin free recombinant protein

Making dollars out of DNA. The first major patent in biotechnology and the commercialization of molecular biology, 1974-1980.

Hughes SS. *Isis*. 2001 Sep;92(3):541-75.

The story of the Cohen–Boyer patents

Rajendra K. Bera CURRENT SCIENCE, VOL. 96, NO. 6,
25 MARCH 2009