

Genetic engineering and genomics

Gulshan Kumar
School of Fisheries

Recombinant DNA

- Clone is an identical copy. This term originally applied to cells of a single type, isolated and allowed to reproduce to create a population of identical cells.
- DNA cloning involves separating a specific gene or DNA segment from a larger chromosome, attaching it to a small molecule of carrier DNA, and then replicating this modified DNA thousands or millions of times through both an increase in host cell number and the creation of multiple copies of the cloned DNA in each cell.
- The result is selective amplification of a particular gene or DNA segment. Cloning of DNA from any organism entails five general procedures:

- 1. Cutting DNA at precise locations. Sequence specific endonucleases (restriction endonucleases) provide the necessary molecular scissors.
- 2. Selecting a small molecule of DNA capable of self-replication. These DNAs are called cloning vectors (a vector is a delivery agent). They are typically plasmids or viral DNAs.
- 3. Joining two DNA fragments covalently. The enzyme DNA ligase links the cloning vector and DNA to be cloned. Composite DNA molecules comprising covalently linked segments from two or more sources are called recombinant DNAs.

- 4. Moving recombinant DNA from the test tube to a host cell that will provide the enzymatic machinery for DNA replication.
- 5. Selecting or identifying host cells that contain recombinant DNA.
- The methods used to accomplish these and related tasks are collectively referred to as recombinant DNA technology or, more informally, genetic engineering.

Cloning
vector
(plasmid)

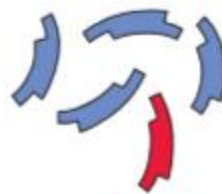


Eukaryotic
chromosome

① Cloning vector
is cleaved with
restriction
endonuclease.



② DNA fragment of interest
is obtained by cleaving
chromosome with a
restriction endonuclease.



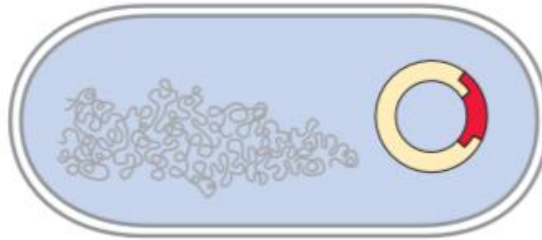
③ Fragments are ligated
to the prepared cloning
vector.

DNA ligase

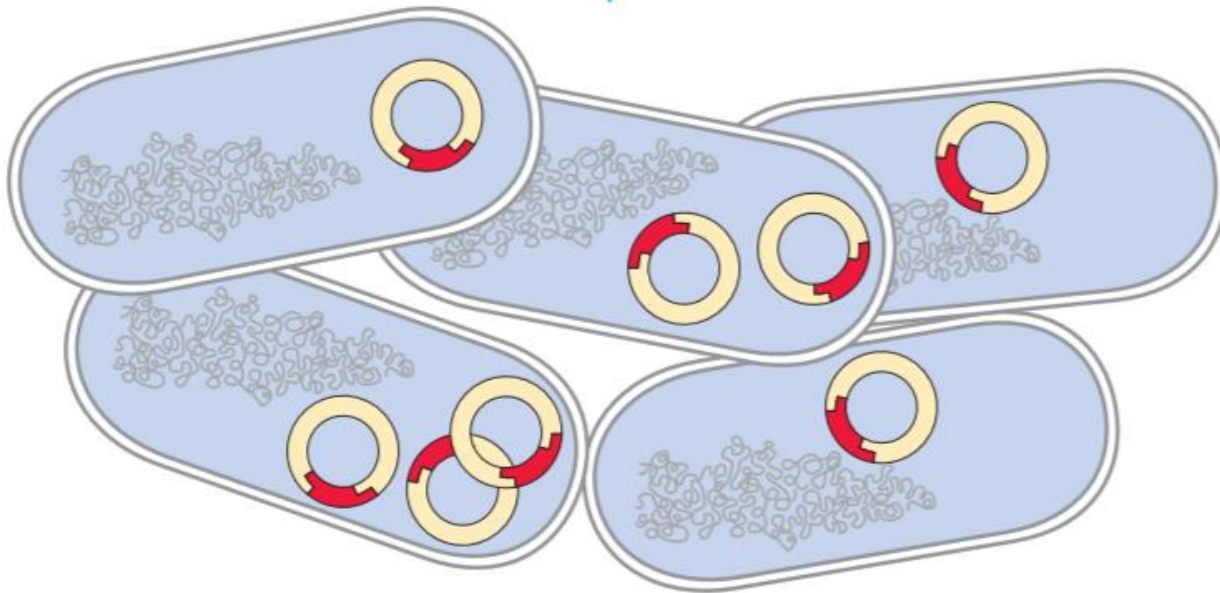


Recombinant
vector

④ DNA is introduced into the host cell.



⑤ Propagation (cloning) produces many copies of recombinant DNA.



Gene isolation

- Gene of interest is the key molecular cloning. The product of recombinant DNA technology (rDNA) is solely dependent on the gene of interest.
- The gene of interest has to be chosen as the target of the experiment. If the function of gene and its sequence is known, the process of gene isolation is straightforward.
- If we don't know the gene responsible for our target product, elaborate experiments involving loss of function and gain of function have to be performed to clearly determine the physiological function of a gene.

- If we know which gene is responsible for target phenotype or product, it is easy to isolate by using RE or PCR.
- **Method of isolation of gene:**
- **1. Using specific RE and probe hybridisation:-**
Specific RE targeting flanking regions of gene of interest can be use to cut the genome to isolate the gene.
- Sequence specific probes are needed to hybridise the target gene. Probe hybridised gene can be isolated based on label of the probe.

- **2. Using PCR: -**
- PCR can be done using gene specific primers to amplify the gene of interest. The product of PCR can be gel fractionated and purified for cloning.
- Most straightforward way for eukaryotic genes is using cDNA as a template for PCR.
- Target mRNA can be reverse transcribed by using reverse transcriptases to form cDNA.
- **3. Gene synthesis approaches:**
- Now a days technology for synthesizing larger fragments of DNA is commonly available and can be used straightaway for synthesis of target gene.

Restriction enzymes

- Particularly important to recombinant DNA technology is a set of enzymes made available through decades of research on nucleic acid metabolism. Two classes of enzymes lie at the heart of the classic approach to generating and propagating a recombinant DNA molecule.
- First, restriction endonucleases (also called restriction enzymes) recognize and cleave DNA at specific sequences (recognition sequences or restriction sites) to generate a set of smaller fragments. Second, the DNA fragment to be cloned is joined to a suitable cloning vector by using DNA ligases to link the DNA molecules together.

- Restriction endonucleases are found in a wide range of bacterial species. Werner Arber discovered in the early 1960s that their biological function is to recognize and cleave foreign DNA (the DNA of an infecting virus,
- In the host cell's DNA, the sequence that would be recognized by its own restriction endonuclease is protected from digestion by methylation of the DNA, catalyzed by a specific DNA methylase. The restriction endonuclease and the corresponding methylase are sometimes referred to as a **restriction-modification system**.

- There are three types of restriction endonucleases, designated I, II, and III. Type I and III has recognition and restriction site away from each other and are ATP dependent. So they are not used in molecular cloning.
- Type II RE has desirable features and are used for molecular cloning.
- Type II restriction endonucleases, first isolated by Hamilton Smith in 1970, are simpler, require no ATP, and cleave the DNA within the recognition sequence itself.
- Thousands of restriction endonucleases have been discovered in different bacterial species, and more than 100 different DNA sequences are recognized by one or more of these enzymes.
- The recognition sequences are usually 4 to 6 bp long and palindromic . Table given in next slide lists sequences recognized by a few type II restriction endonucleases.

- Some restriction endonucleases make staggered cuts on the two DNA strands, leaving two to four nucleotides of one strand unpaired at each resulting end. These unpaired strands are referred to as **sticky ends**, because they can base-pair with each other or with complementary sticky ends of other DNA fragments.
- Other restriction endonucleases cleave both strands of DNA at the opposing phosphodiester bonds, leaving no unpaired bases on the ends, often called **blunt ends**

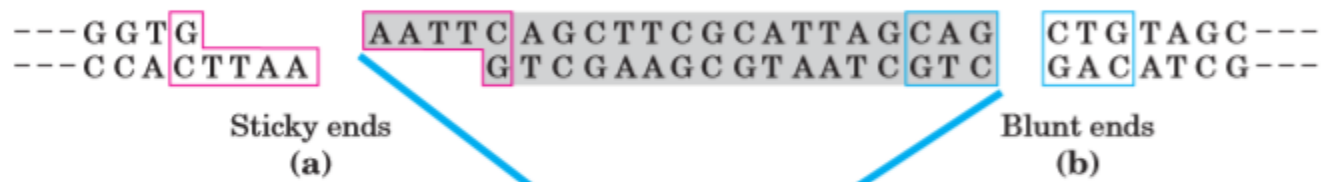
TABLE 9-2
Recognition Sequences for Some Type II Restriction Endonucleases

<i>Bam</i> HI	$\begin{array}{c} \downarrow \qquad \qquad \qquad * \\ (5') \text{ G G A T C C } (3') \\ \text{ C C T A G G} \\ * \qquad \qquad \qquad \uparrow \end{array}$	<i>Hind</i> III	$\begin{array}{c} \downarrow \\ (5') \text{ A A G C T T } (3') \\ \text{ T T C G A A} \\ \qquad \qquad \qquad \uparrow \end{array}$
<i>Cla</i> I	$\begin{array}{c} \downarrow \qquad \qquad \qquad * \\ (5') \text{ A T C G A T } (3') \\ \text{ T A G C T A} \\ * \qquad \qquad \qquad \uparrow \end{array}$	<i>Not</i> I	$\begin{array}{c} \downarrow \\ (5') \text{ G C G G C C G C } (3') \\ \text{ C G C C G G C G} \\ \qquad \qquad \qquad \uparrow \end{array}$
<i>Eco</i> RI	$\begin{array}{c} \downarrow \qquad * \\ (5') \text{ G A A T T C } (3') \\ \text{ C T T A A G} \\ \qquad \qquad * \qquad \uparrow \end{array}$	<i>Pst</i> I	$\begin{array}{c} \qquad \qquad * \qquad \downarrow \\ (5') \text{ C T G C A G } (3') \\ \text{ G A C G T C} \\ \uparrow \qquad \qquad * \end{array}$
<i>Eco</i> RV	$\begin{array}{c} \downarrow \\ (5') \text{ G A T A T C } (3') \\ \text{ C T A T A G} \\ \qquad \qquad \qquad \uparrow \end{array}$	<i>Pvu</i> II	$\begin{array}{c} \downarrow \\ (5') \text{ C A G C T G } (3') \\ \text{ G T C G A C} \\ \qquad \qquad \qquad \uparrow \end{array}$
<i>Hae</i> III	$\begin{array}{c} \downarrow * \\ (5') \text{ G G C C } (3') \\ \text{ C C G G} \\ * \qquad \qquad \qquad \uparrow \end{array}$	<i>Tth</i> 111I	$\begin{array}{c} \downarrow \\ (5') \text{ G A C N N N G T C } (3') \\ \text{ C T G N N N C A G} \\ \qquad \qquad \qquad \uparrow \end{array}$



EcoRI
 restriction
 endonuclease

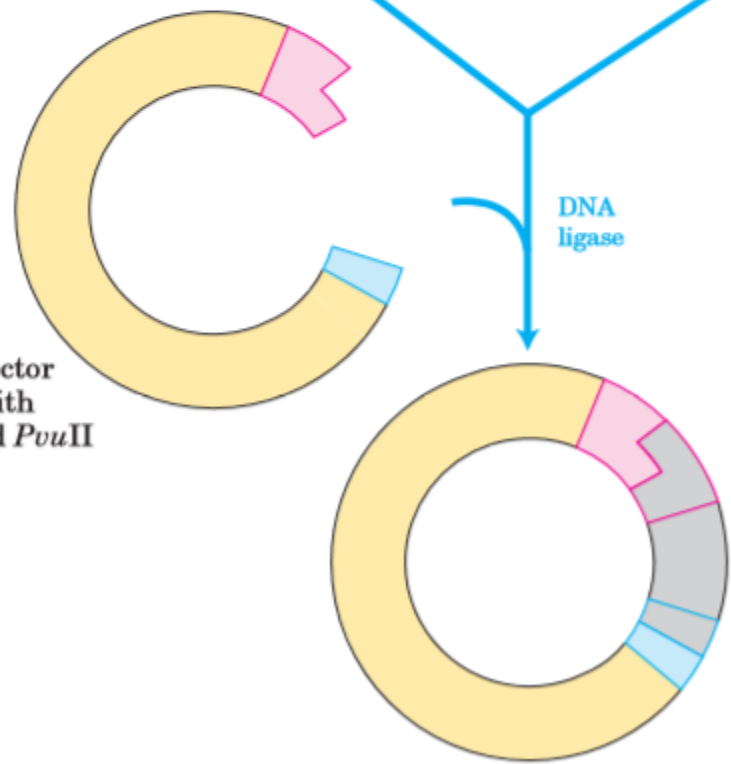
PvuII
 restriction
 endonuclease



Plasmid
 cloning vector
 cleaved with
EcoRI and *PvuII*

DNA
 ligase

Please ignore the shape



Cloning vectors

- A vector delivers recombinant DNA in clonable form to a host cell, and insures its subsequent amplification in the host cells.
- There are three popular cloning vectors commonly used in experiments with *E. coli*—plasmids, bacteriophages, and bacterial artificial chromosomes—and a vector used to clone large DNA segments in yeast (Yeast artificial chromosome).

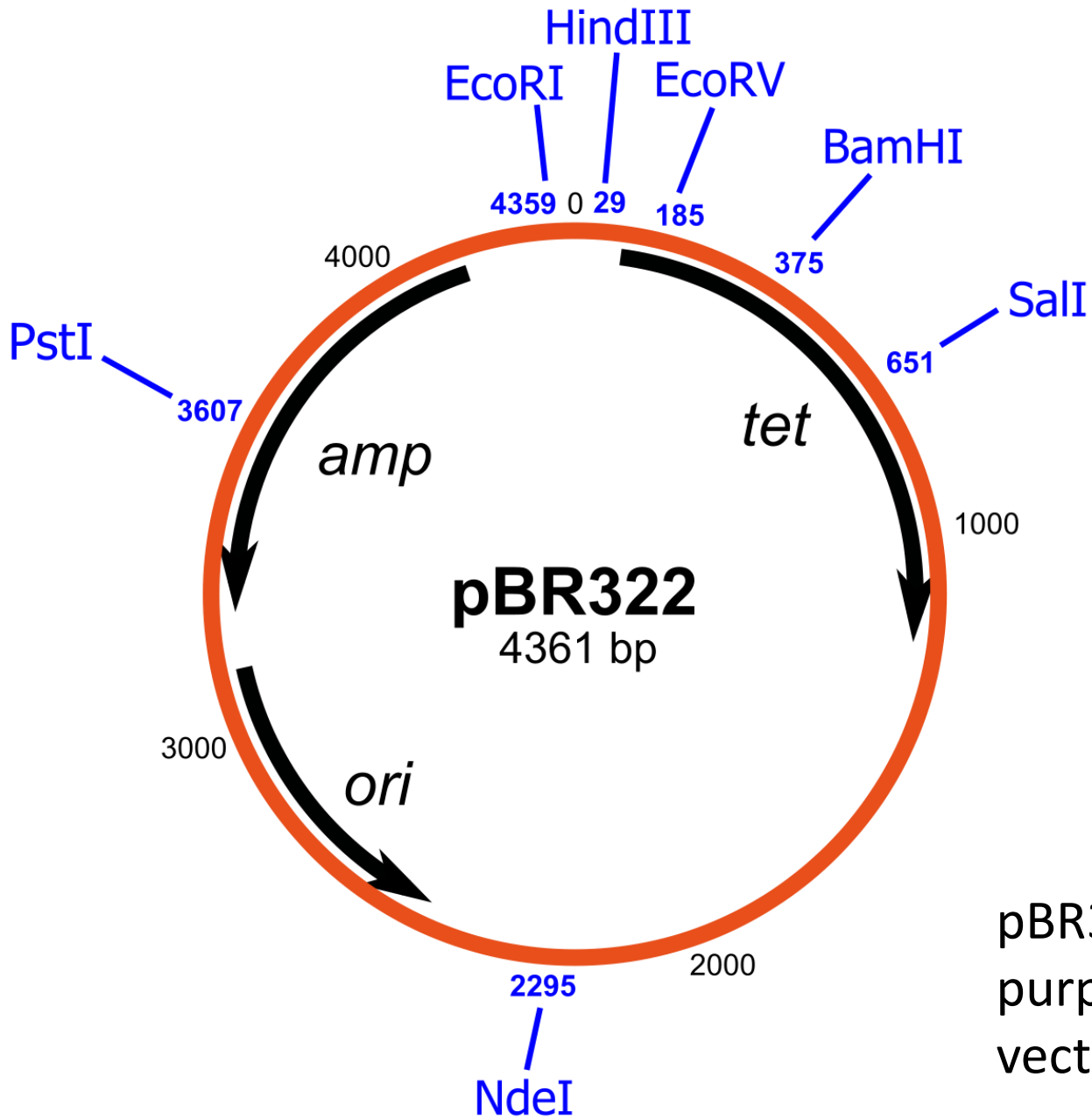
- **Plasmids**
- Plasmids are circular DNA molecules that replicate separately from the host chromosome.
- Naturally occurring bacterial plasmids range in size from 5,000 to 400,000 bp. T
- they can be introduced into bacterial cells by a process called **transformation**.
- The cells (generally E. coli) and plasmid DNA are incubated together at 0 C in a calcium chloride solution, then subjected to a shock by rapidly shifting the temperature to 37 to 43 C.
- For reasons not well understood, some of the cells treated in this way take up the plasmid DNA. In an alternative method, cells incubated with the plasmid DNA are subjected to a high-voltage pulse. This approach, called **electroporation**, transiently renders the bacterial membrane permeable to large molecules

- Regardless of the approach, few cells actually take up the plasmid DNA, so a method is needed to select those that do.
- The usual strategy is to use a plasmid that includes a gene that the host cell requires for growth under specific conditions, such as a gene that confers resistance to an antibiotic.
- Only cells transformed by the recombinant plasmid can grow in the presence of that antibiotic, making any cell that contains the plasmid “**selectable**” under those growth conditions.
- Such a gene is called a **selectable marker**. Investigators have developed many different plasmid vectors suitable for cloning by modifying naturally occurring plasmids. The now classic E. coli plasmid pBR322 offers a good example of the features useful in a cloning vector.

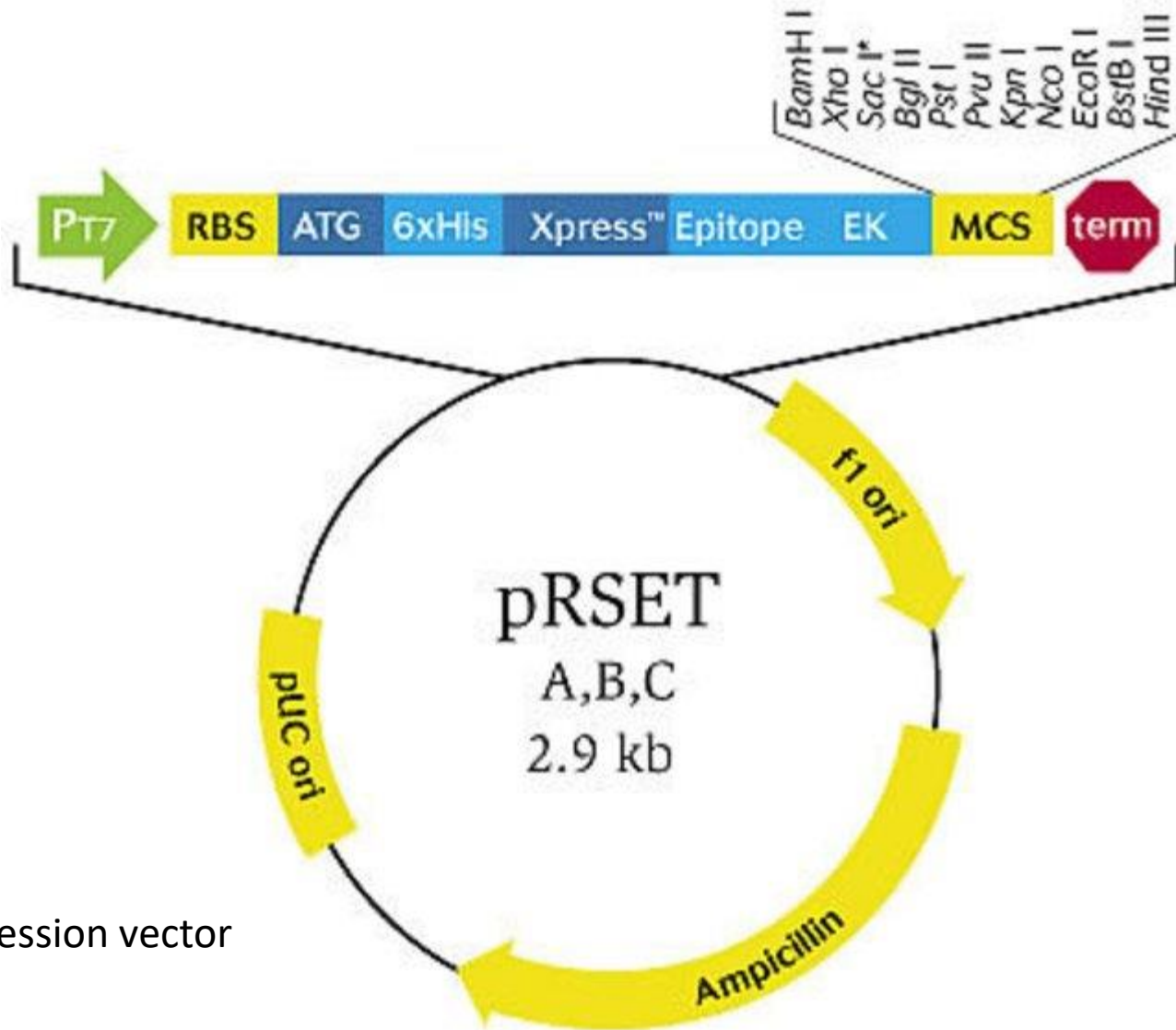
Important features of a cloning vector

- **1. An origin of replication, ori**, a sequence where replication is initiated by cellular enzymes. This sequence is required to propagate and maintain it at a level of 10 to 20 copies per cell.
- 2. Two genes that confer resistance (called selectable markers) to different antibiotics (tetR, ampR), allowing the identification of cells that contain the intact plasmid or a recombinant version of the plasmid or any vector (transformants and recombinants).
- 3. Several unique recognition sequences, that are targets for different restriction endonucleases, providing sites where the plasmid can later be cut to insert foreign DNA.
- 4. Small size, which facilitates entry of the plasmid into cells and the biochemical manipulation of the DNA.

- **Bacteriophage vectors**
- Bacteriophages are the viruses infecting bacterial cells.
- They can transfer our gene of interest into target bacterial cell by the process of transduction.
- They can transfer larger fragments than plasmid DNA.
- Bacteriophage λ is extensively modified so that it can have essential features of a vector and used in molecular cloning.



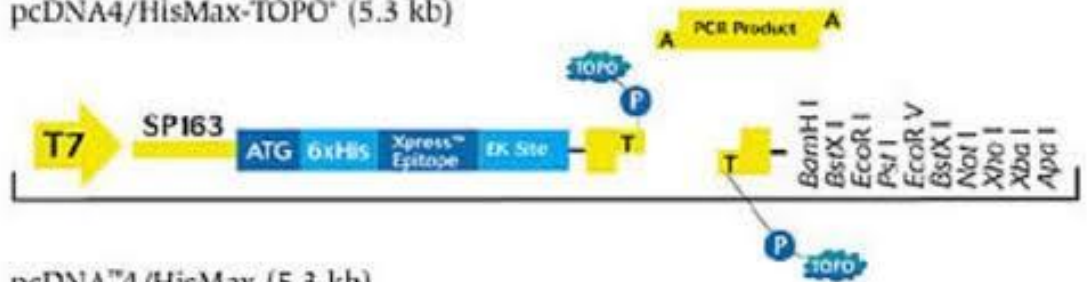
pBR322 a general purpose cloning vector.



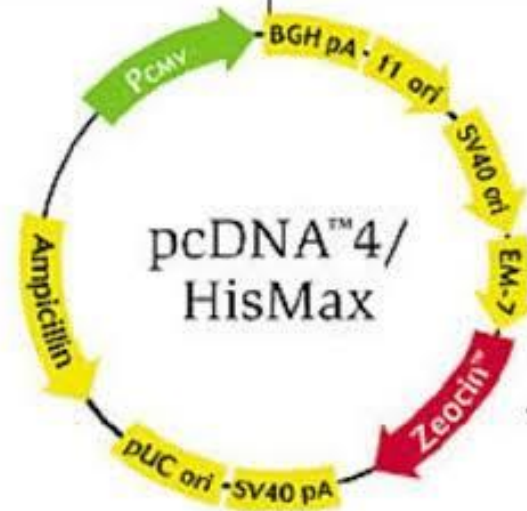
E coli expression vector

Figure 2 - pcDNA™ 4/HisMax vectors

pcDNA4/HisMax-TOPO* (5.3 kb)



pcDNA™4/HisMax (5.3 kb)



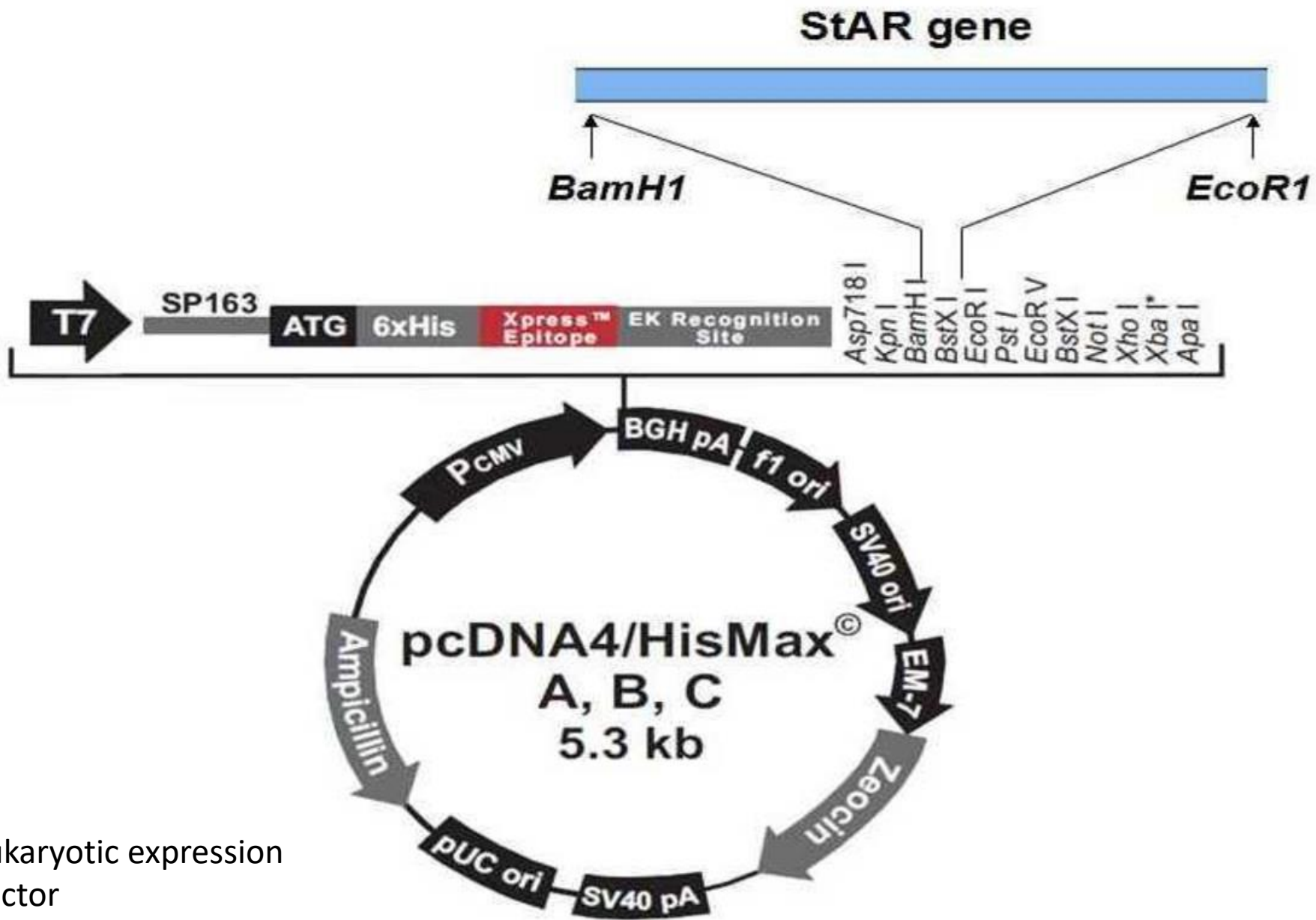
* There is a stop codon following the Xba I site in version A.



Represents covalently bound topoisomerase I

See polylinker sequences on the web.

Eukaryotic expression vector



Eukaryotic expression
vector

- **Bacterial Artificial Chromosomes (BACs)**
Bacterial artificial chromosomes are simply plasmids designed for the cloning of very long segments (typically 100,000 to 300,000 bp) of DNA.
- It can be used in E coli host.
- **Yeast Artificial Chromosomes (YACs)**
- Like E coli, Yeast (*Saccharomyces cerevisiae*) is used as a host for molecular cloning works targeted to eukaryotes.

- It is easy to maintain and its genetics is thoroughly studied.
- Research with large genomes and the associated need for high-capacity cloning vectors led to the development of yeast artificial chromosomes (YACs).
- YAC vectors contain all the elements needed to maintain a eukaryotic chromosome in the yeast nucleus.
- Insert size or vector capacity for YAC is 100 Kb to 10 Mb.

Selection of transformants and recombinants

- At the time of ligation out of total no of vector molecules given in the reaction mixture only a small proportion will be ligated to the insert to form recombinant vector.
- Some times two ends of vector may get ligated without insert forming empty vectors.
- Unligated vectors may also remain in reaction mixture due to incomplete restriction digestion.
- At the time of transformation not all the cells present in media will take DNA. Cells taking DNA is called transformed cells or transformants.

- Transformed cells can have either recombinant plasmid or empty one.
- Transformed cells with recombinant vector having insert at their MCS is needed for making recombinant products.
- For selecting recombinant cells out of all the transformed and nontransformed cells an elaborate selection strategy has to be followed.
- Selection methodology of recombinant cells depends on selectable markers given in the vector.

- Based on that three ways of selection may be applied.
- 1. Antibiotic selection for both transformant and recombinants:-
- Two antibiotic resistance genes one outside the MCS and other inside MCS is given.
- This selection process is two stage process in which transformed cell mixture is first plated in a agar plate having antibiotic for which the resistance gene was given outside the MCS.
- In this plate both recombinant and nonrecombinant cells will form colonies. Only nontransformed cell will die as they do not have any antibiotic resistance gene.

- The colonies got in the first plate has to be transferred in same orientation into another plate having both the antibiotic in media.
- The colonies which unable to grow in the second plate is the recombinant desired colonies.
- These colonies can be used for further applications.

- **2. Enzymatic selection (Blue-white screening)**
- Here in MCS region in place of an antibiotic resistance gene, ORF of α - part of lacZ gene (β -galactosidase is given. The β -fragment is supplied by the bacterial genome.
- The antibiotic resistance gene outside the MCS is present here also.
- The enzyme encoded by the gene can convert X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) into a blue color compound in presence of IPTG (Isopropyl β - d-1-thiogalactopyranoside) as an inducer.
- As recombinant vector contain insert in the MCS, the ORF of lacz gene get disrupted and so functional enzyme can not be formed. The colonies having recombinant colonies will remain white while nonrecombinant will become blue.
- Selection for transformants is like previous one with antibiotics.
- This method is having advantages on previous one as it is a single step selection.

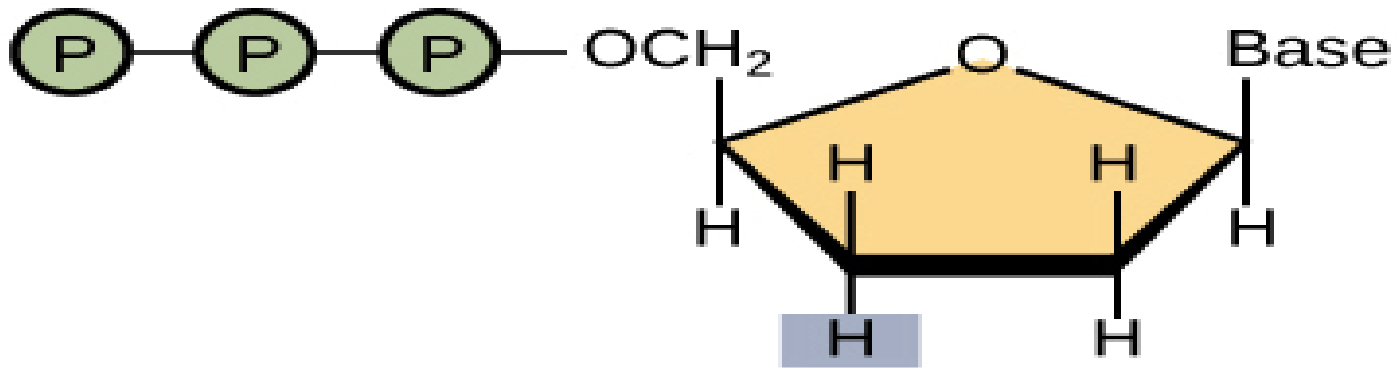
- **3. Selection with PCR:-**

- The selection for transformant is generally done with antibiotic containing media.
- From the selected colonies recombinant can be identified by screening using PCR with specific primers targeting to cloned ORF or universal primer binding site given in the vector..

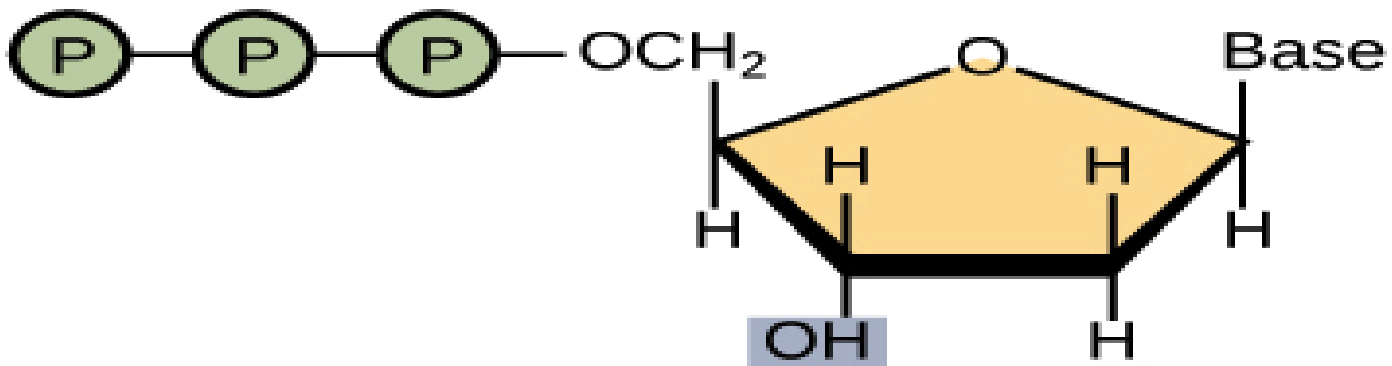
- **DNA sequencing**
- **DNA sequencing** is the process of determining the sequence of nucleotides (As, Ts, Cs, and Gs) in a piece of DNA.
- In **Sanger sequencing**, the target DNA is copied many times, making fragments of different lengths. Fluorescent “chain terminator” nucleotides mark the ends of the fragments and allow the sequence to be determined.
- **Next-generation sequencing** techniques are new, large-scale approaches that increase the speed and reduce the cost of DNA sequencing.

- What is sequencing?
- You may have heard of genomes being sequenced. For instance, the human genome was completed in 2003, after a many-year, international effort. But what does it mean to sequence a genome, or even a small fragment of DNA?
- **DNA sequencing** is the process of determining the sequence of nucleotide bases (As, Ts, Cs, and Gs) in a piece of DNA. Today, with the right equipment and materials, sequencing a short piece of DNA is relatively straightforward.

- Sequencing an entire genome (all of an organism's DNA) remains a complex task. It requires breaking the DNA of the genome into many smaller pieces, sequencing the pieces, and assembling the sequences into a single long "consensus." However, thanks to new methods that have been developed over the past two decades, genome sequencing is now much faster and less expensive than it was during the Human Genome Project¹¹start superscript, 1, end superscript.



Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

- **Sanger sequencing: The chain termination method**
- Regions of DNA up to about 900 base pairs in length are routinely sequenced using a method called **Sanger sequencing** or the **chain termination method**. Sanger sequencing was developed by the British biochemist Fred Sanger and his colleagues in 1977.

- Other than conventional PCR components viz, template, primers, dNTPs and DNA polymerase; a Sanger sequencing reaction also contains a unique ingredient: Dideoxy, or **chain-terminating**, versions of all four nucleotides (ddATP, ddTTP, ddCTP, ddGTP), each labeled with a different color of dye.
- Dideoxy nucleotides are similar to regular, or deoxy, nucleotides, but with one key difference: they lack a hydroxyl group on the 3' carbon of the sugar ring. In a regular nucleotide, the 3' hydroxyl group acts as a "hook," allowing a new nucleotide to be added to an existing chain.
- Once a dideoxy nucleotide has been added to the chain, there is no hydroxyl available and no further nucleotides can be added. The chain ends with the dideoxy nucleotide, which is marked with a particular color of dye depending on the base (A, T, C or G) that it carries.

- **Next-generation sequencing**
- The most recent set of DNA sequencing technologies are collectively referred to as **next-generation sequencing (NGS)**. .
- There are a variety of next-generation sequencing techniques that use different technologies. However, most share a common set of features that distinguish them from Sanger sequencing:
- **Highly parallel:** many sequencing reactions take place at the same time
- **Micro scale:** reactions are tiny and many can be done at once on a chip

- **Fast:** because reactions are done in parallel, results are ready much faster
- **Low-cost:** sequencing a genome is cheaper than with Sanger sequencing
- **Shorter length:** reads typically range from 50-700 nucleotides in length
- **Pyrosequencing is a method among the several of NGS methods of DNA sequencing that detects light emitted during the sequential addition of nucleotides during the synthesis of a complementary strand of DNA.**
- **Some examples of NGS platforms commonly used today are**
 - Illumina/Solexa
 - Roche/454,
 - ABI/SOLiD
 - Ion-torrent

- **Genomics**
- **An organism's complete set of DNA is called its genome. Virtually every single cell in the body contains a complete copy of the approximately 3 billion DNA base pairs, or letters, that make up the human genome.**
- Genomics is the study of whole genomes of organisms, and incorporates elements from genetics. Genomics uses a combination of recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyse the structure and function of genomes. It differs from 'classical genetics' in that it considers an organism's full complement of hereditary material, rather than one gene or one gene product at a time. Moreover, genomics focuses on interactions between loci and *alleles* within the genome and other interactions such as, *pleiotropy* and *heterosis* . Genomics harnesses the availability of complete DNA sequences for entire organisms and was made possible by both the pioneering work of Fred Sanger and the more recent next-generation sequencing technology.

- The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb [megabase]) in 1995
- The following year a consortium of researchers from laboratories across North America, Europe, and Japan announced the completion of the first complete genome sequence of a eukaryote, *S. cerevisiae* (12.1 Mb), and since then genomes have continued being sequenced at an exponentially growing pace. As of October 2011, the complete sequences are available for: 2,719 viruses, 1,115 archaea and bacteria, and 36 eukaryotes, of which about half are fungi.
- The ***Human Genome Project (HGP)*** was an international 13-year effort, 1990 to 2003.

- **Applications**
- **Genomic medicine**
- Next-generation genomic technologies allow clinicians and biomedical researchers to drastically increase the amount of genomic data collected on large study populations. When combined with new informatics approaches that integrate many kinds of data with genomic data in disease research, this allows researchers to better understand the genetic bases of drug response and disease.
- **Synthetic biology and bioengineering**
- The growth of genomic knowledge has enabled increasingly sophisticated applications of synthetic biology. In 2010 researchers at the J. Craig Venter Institute announced the creation of a partially synthetic species of bacterium, *Mycoplasma laboratorium*, derived from the genome of *Mycoplasma genitalium*.
- **Conservation genomics**
- Conservationists can use the information gathered by genomic sequencing in order to better evaluate genetic factors key to species conservation, such as the genetic diversity of a population or whether an individual is heterozygous for a recessive inherited genetic disorder.