RECOMBINANT DNA TECHNOLOGY AND BIOTECHNOLOGY

Tools of Recombinant DNA Technology

Rakesh Bhatnagar
Professor
Centre for Biotechnology
Jawaharlal Nehru University
JNU New Campus
New Delhi 110067

27-Mar-2006 (Revised 24-Jul-2007)

CONTENTS

Keywords

Recombinant DNA; Plasmids; Bacteriophages; Agarose Gel Electrophoresis; Restriction endonuclease; Isochizomers; DNA ligase; DNA polymerase;Bacteriophage polymerase; Reverse Transriptase; Nucleases; Alkaline phosphatase; Poymerase Chain Reaction (PCR); Plasmid vectors; Gene cloning; Bacteriophage vector; Lambda replacement vector; Insertion vector; Shuttle vector; Plant vector; Ti plasmid.

What is a Recombinant DNA

DNA molecules constructed outside the living cells that is *in vitro* by joining natural or synthetic DNA segments that can replicate in a living cell

Goals of Recombinant DNA Technology

a) To isolate and characterize a gene
b) To make desired alterations in one or more isolated genes
c) To return altered genes to living cells

Basic Tools of Recombinant DNA Technology

Nucleic Acid Enzymes

DNA and RNA polymerases, reverse transcriptase, DNA ligases, Restriction endonucleases and many more.

Plasmids and Bacteriophages

a) Bacteriophages replicate via the lytic phase cycle and the phage genome is injected into the cell, phage genes are expressed and phage proteins and DNA are made, progeny phage are packaged, and the cell is lysed. Two genetically different phage that infect the same host cell may recombine during the lytic cycle
b) Some PHAGE can also replicate via the Lysogenic cycle. The phage genome is integrated into the host chromosome and is inherited into the chromosomes of all daughter bacteria. This "prophage" can be induced to enter the lytic cycle and kill its host by a variety of stresses like UV light
c) Plasmids: Circular DNAs that replicate autonomously

Visualization of DNA: Agarose Gel Electrophoresis

Electrophoresis is a technique used to separate and purify macromolecules like nucleic acids that differ in size, charge or conformation. It is one of the most widely-used techniques in biochemistry and molecular biology. When charged molecules are placed in an electric field, they migrate toward either the positive or negative pole according to their charge. In contrast to proteins, which can have either a net positive or net negative charge, nucleic acids have a consistent negative charge imparted by their phosphate backbone, and migrate towards the positive end of the electrode, anode.

The nucleic acids are electrophoresed on a matrix called as Gel. Most commonly, the gel is cast in the shape of a thin slab, with wells for loading the sample. The gel is immersed within an electrophoresis buffer that provides ions to carry a current and maintain the pH at a relatively constant value. The gel is composed of agarose, a polysaccharide extracted from seaweed. It is typically used at concentrations of 0.5 to 2%. The higher the agarose concentration the gel is of less porosity. The agarose gels have a large range of separation, but relatively low resolving power. By varying the concentration of agarose, fragments of DNA from about 200 to 50,000 bp can be separated using standard electrophoretic techniques.

The distance DNA has migrated in the gel can be judged by visually monitoring migration of the tracking dyes. Bromophenol blue and xylene cyanol dyes are used usually and they migrate through agarose gels at roughly the same rate as double-stranded DNA fragments of 300 and 4000 bp, respectively. When adequate migration has occured, DNA fragments are visualized by staining with ethidium bromide, a fluorescent dye which intercalates between the bases of DNA and RNA. Fragments of linear DNA migrate through agarose gels with a mobility that is inversely proportional to the $\log_{10}$ of their molecular weight. However, circular forms of DNA migrate in agarose distinctly differently from linear DNAs of the same mass. Uncut plasmids will migrate more rapidly than the same plasmid when linearized and most preparations of uncut plasmid contain two topologically-different forms of DNA, corresponding to supercoiled forms and nicked circles. Several additional factors have important effects on the mobility of DNA fragments in agarose gels like:

Agarose concentration: By using gels with different concentrations of agarose, we can resolve different sizes of DNA fragments. Higher concentrations of agarose facilitate separation of small DNAs, while low agarose concentrations allow resolution of larger DNAs

Voltage: As the voltage applied to a gel is increased, larger fragments migrate proportionally faster than the small fragments. For that reason, the best resolution of fragments larger than about 2 kb is attained by applying no more than 5 volts per cm to the gel where, cm value is the distance between the two electrodes, not the length of the gel).

Electrophoresis buffer: Several different buffers have been recommended for electrophoresis of DNA. The most commonly used for duplex DNA are TAE (Tris-acetate-EDTA) and TBE (Tris-borate-EDTA). DNA fragments will migrate at somewhat different rates in these two buffers due to differences in ionic strength. Buffers not only establish a pH, but provide ions to support conductivity. If water is used instead of buffer there will be no migration of DNA in the gel and conversely, if concentrated buffer is used like a 10X solution instead of 1X, heat will be generated in the gel which is enough to melt it.

Ethidium bromide: Ethidium bromide is a fluorescent dye that intercalates between bases of nucleic acids and allows very convenient detection of DNA fragments in gels. It is added to the DNA sample before loading to enable visualization of the fragments within the gel or can be added in the electrophoresis buffer. The binding of ethidium bromide to DNA alters its mass and rigidity, and therefore its mobility.

Restriction Endonuclease – An overview

Phage growth restriction – Methylase & Nuclease concept

It has been observed that phage particles that grow well and efficiently infect one strain of bacteria are often unable to grow well and infect other strains of the same bacterial species. Studies carried out showed that the phage particles that efficiently grow and infect host cells have DNA molecules which are chemically modified by the addition of $-CH_3$ groups to some of their A/C bases, while the DNA of phage particles, which poorly infect the bacterial strain, does not show this pattern of chemical modification of methylation. To summarize, phage particles with unmethylated DNA do not grow and infect efficiently. Therefore, their DNA molecules gets cleaved and degraded by the enzyme of host cells, which recognize it as foreign. In contrast, methylated DNA is recognized as self and is protected from host-mediated destruction. This phenomenon of degrading unmethylated DNA and protecting the self methylated DNA is responsible for the phenomenon of phage growth restriction. In

addition to this particular methylase, individual bacterial strains also contain accompanying specific endonuclease activity. In late 1960s, Stewart Linn and Werner Arber isolated two types of enzymes responsible for this observed phage restriction phenomenon. Later on, in 1978, Smith, Nathans and Arber were awarded Nobel Prize for Physiology and Medicine for the discovery of endonucleases. The first type of enzyme was called a methylase while other was called a restriction nuclease. The endonuclease cleaves at or near the methylation recognition site. Different strains have difference in residues which are methylated. Typical sites of methylation include N6 position of adenine, N4 of cytosine or C5 of Cytosine. And, only a fraction of bases are methylated, that is, not every A/C will be methylated. This suggests that there are only few unique bases which are substrates of methylase. Thus, this unique system was coined as Restriction Modification system in bacteria. It acts as a small immune system for protection from infection by foreign DNA, example viral DNA, in individual bacterial strains and provides a reason why the restriction endonuclease doesn't chew its own DNA.

Eg. *ECOR I* methylase, the sequence GAATTC will be internally methylated at adenine base (shown in asterisks) by the enzyme. The *ECOR I* endonuclease within the host bacteria will not cleave the methylated DNA. The foreign viral DNA, which is not methylated at this particular sequence, GAATTC, will be recognized as foreign invasive DNA and will be cleaved by *ECOR I* endonuclease. This cleavage of foreign DNA abandons any viral infection, thus protecting the host from any fatal consequence. Such endonucleases are termed as restriction endonucleases since they restrict the DNA within the cell to being "self".

If the attacking phage was previously grown on the same strain, which is infecting now, this system will be rendered ineffective. In this case, the phage's DNA will already be methylated at appropriate sequence and will be recognized as self. Structural and biochemical studies have indicated that for the type II R/M system, the methylase recognizes and methylates one strand of DNA duplex, whereas restriction endonucleases recognize both the strands, that is, both should be non-methylated for recognition. It is able to do this because it is a homodimer. The methyl group protrudes into the major groove of DNA at the binding site and prevents the restriction enzyme from acting upon it. Together, a Restriction enzyme and its cognate modification enzyme, works in the bacteria to prevent it from infections, acting as miniature microbial immune system. In some R-M systems, the restriction enzyme and the modification enzyme(s) are separate proteins that act independent of each other while in some, the two activities occur as separate subunits / domains of a larger, combined R-M enzyme.

Restriction enzyme or Endonuclease is an enzyme that cuts double stranded DNA. Precisely, it cleaves the sugar-phosphate backbone of DNA without damaging the bases. A restriction endonuclease works by scanning the length of DNA molecule and once it encounters its particular specific recognition sequence; it will bind to the DNA and makes one incision in each of the two sugar- phosphate backbones of the double helix. Thus, rather than cutting DNA indiscriminately, a RE cuts only at a particular nucleotide sequence, termed as Recognition sequence. Therefore, scientists could cut the DNA molecule in a predictable and reproducible way. Restriction enzymes are traditionally classified into 3 types on the basis of subunit composition, cleavage position, sequence specificity and requirement for cofactors.

Type I

Complex, multisubunit, cut DNA at random, far from their recognition sequences. They may have biological significance but have a very little practical value since they do not produce discrete restriction fragments.

Type II

Cut DNA at very precise and defined positions close to or within the recognition sequences. It is the only class of restriction endonuclease used in the laboratory for DNA analysis and gene cloning. They require only Magnesium for activity and corresponding model of enzymes require only s-adenosyl methionine (SAM). They are smaller with subunits in the range of 200-350 amino acids. The most common are the ones which cleave within symmetric recognition sequence. e.g. *Hha* I, *Hind* III and *Not* I. Next common of Type II, the ones which cleave outside their Recognition sequence to one side. They are intermediate in size with 400-650 amino acids and recognize sequences that are continuous and asymmetric. Third major type II Restriction enzymes are referred to as Type IV, are large, 850-1250 amino acids, where two enzymatic activities reside in the same protein chain, cleave outside the recognition sequence and others those recognize discontinuous sequences and cleave on both sides releasing a small fragment containing recognition sequence.
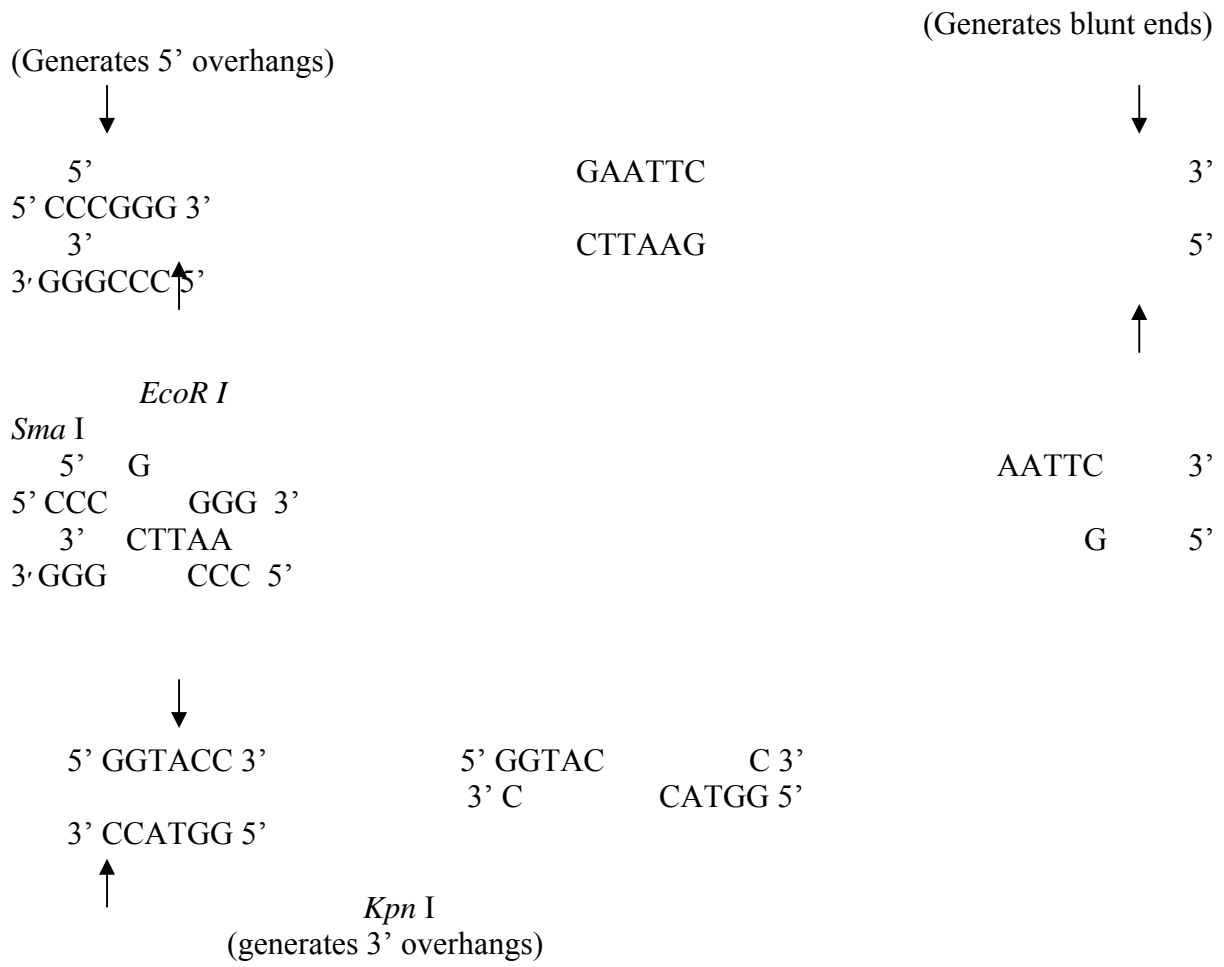
Type III

Large, cleave outside the recognition sequence and require two sequences in opposite orientations within the same DNA molecule to accomplish cleavage. They rarely give complete digests. They have basically no practical requirement and there is none so available commercially.

Source and site of Action

Patterns of DNA cutting
Restriction Endonuclease recognizes short, specific palindromic sequences of DNA and cleaves them. Some restriction enzymes symmetrically leave blunt ends without overhangs like Sma I as shown below while others cleave DNA backbones in positions that are not directly opposite to each other. e.g *EcoR I*. Arrows show the site of cleavage. Once the cuts have been made, the resulting fragments are held together only by the relatively weak hydrogen bonds holding complimentary bases with each other. These enzymes generate either of two types of ends, 5ʹ or 3ʹ, called overhangs or sticky ends.

(Generates 5' overhangs)   (Generates blunt ends)

↓   ↓

5'   GAATTC   3'
5' CCCGGG 3'
3'   CTTAAG   5'
3' GGGCCC↑5'
                                                                                     ↑

EcoR I

Sma I
5'   G   AATTC   3'
5' CCC   GGG 3'
3'   CTTAA   G   5'
3' GGG   CCC 5'

↓

5' GGTACC 3'   5' GGTAC   C 3'
                         3' C   CATGG 5'

3' CCATGG 5'

↑
Kpn I
(generates 3' overhangs)

Nomenclature

Restriction endonucleases are named based on bacteria from which they are isolated in the following manner

E	Escherichia (Genus)
Co	Coli (Species)
R	RY13 (Strain)
I	First Identified order (Id) in Bacteria

Name of RE	Organism from which isolated	Recognition Sequence	
EcoR I	E.coli	GAATCC	6bp cutter, leaves 5' overhangs
BamH I	Bacillus amyloliquefaciens	GGATCC	6bp cutter, leaves 5' overhangs
Hind III	Haemophilus influenzae	AAGCTT	6bp cutter, leaves 5' overhangs
Alu I	Arthobacter luteus	AGCT	4bp cutter, Leaves blunt end
Not I	Nocardia olitidiscaviarum	GCGGCCGC	8bp cutter, 5' overhangs

Recognition Sequence typically is only 4-12 nucleotides long. Thus probability that they are found in chromosomes of any organism is quite possible and frequent. Thus, if there is 25% probability for a specific base at any given site (because 4 bases, A/T/G/C), then the frequency with which different RE site will occur in a gene will be 4^n, where n is nucleotide specificity. Eg. *Alu* I is a 4bp cutter, therefore, $(4)^4 = 256$. Therefore, on an average, any DNA will contain an *Alu* I site after every 0.25 kilobases, whereas *Not* I site occurs once about every 65.5 kilo bases (4^8). Thus *Not* I generates bigger fragments while *Alu* I would be expected to digest DNA into many small fragments. Thus, length of Recognition sequence dictates how frequently the enzyme will cut in a random DNA sequence.

Isochizomers

Isochizomers are different Restriction endonucleases having same recognition site. In some cases, they cut identically within their recognition site, but sometime they do not. They have different optimum reaction conditions, stabilities and cost that give us an option of what to purchase. Some Restriction endonucleases recognizes only one sequence but never other, called as Ambiguous Recognition Sequence. Eg. *BamH* I recognize GGATCC, while *Hinf I* recognizes a 5bp sequence, with an eligibility of sequence starting with GA and ending with TC and having any base in between GANTC. Some REs recognition site has a site for cleavage by other Restriction Endonuclease. e.g: *BamH* I site GGATCC have site recognized and cleaved by *Sau3A* I GATC. Thus all *BamH* I sites can consequently be cut by *Sau3A* I.

Conditions and requirements that influence Enzyme Activity

One unit activity:- is defined as amount of enzyme required to digest 1μg of reference DNA in 60 minutes at 37^0C. Usually reference DNA is λ phage DNA.

Buffer:- The most critical determinant of the enzyme activity is the ionic concentration (NaCl content) of the buffer. There are commercially 3 buffer systems available for the activity, low salt buffer (20mM), medium (100mM) or high (250mM) NaCl buffers. The buffer is usually 10mM Tris-HCl, pH 8.0, supplemented with Magnesium salt (often 50mM $MgCl_2$), a reducing agent (usually 1mM DTT), and some enzymes require BSA (100μg/ml) and salt (NaCl). The reaction does not require ATP.

Quality of DNA :- Preparation of DNA that has to be cleaved should be free of contaminants such as phenol, $CHCl_3$, alcohol, EDTA, detergents, excessive salts, protein, RNA, all of which can pose a hindrance to the enzyme's activity.

Stopping a Reaction:- To terminate Restriction enzymes activity, heat inactivation is a way but for the enzymes which it does not work, phenol / $CHCl_3$ extraction is another means of inactivation of Restriction enzymes.

Incubation Temperature and Time: - Recommended temperature for most of the Restriction enzymes is 37^0C except for the ones which are isolated from thermophilic bacteria which require higher temperature ranging from 50-65^0C. While some restriction enzymes have short ½ lives at 37^0C and for those lower temperatures are required. Incubation time as per the unit definition is 1 hour but it can be manipulated. Example, it may be shortened if more units of Restriction enzyme is added or longer times are used to allow a reaction to proceed to completion with fewer enzymes units.

Cleavage near ends :- When Recognition sequences are located at or very close to the ends of the linear fragments, most enzymes require few bases on either side of Recognition sequence in order to bind and cleave. Thus there is some specificity on end requirement for a variety of enzymes.

DNA methylation :- Almost all strains of *E.coli* used for propagating cloned DNA contain 2 site-specific DNA methylase:-

Dam Methylase	Dcm Methylase
Adds a –CH_3 group to A in separate G^mATC	Adds methyl group to internal cytosine C^mC (A/T)GG

Importance of this phenomenon lies in the fact that a number of Restriction enzymes will not cleave methylated DNA. Eg. *Mbo* I and *Sau3A* I are isochizomers recognizing sequence GATC. Dam methylase methylates it to G^mATC, while *Mbo* I does not cleave it, *Sau3A* I is unaffected by the sequence methylation.

Relaxed or Altered activity of Restriction enzymes - Star activity
If non standard conditions are used for DNA digestion, cleavage can occur at sites different from canonical recognition sequence. This aberrant non specific DNA cleavage is called as star activity. Some non standard conditions that contribute to this star activity are as follows:-

  ➢ High pH (>8)
  ➢ Low ionic strength of buffer (<25mM)
  ➢ >5% glycerol concentration (Enzymes are stored in 50 % glycerol)
  ➢ High enzyme concentration (>100U for 1μg DNA)
  ➢ Presence of organic solvents in DNA preparation Eg. DMSO, Ethylene Glycol
  ➢ Substitution of Mg^{+2} with other divalent cations like Mn^{+2}, Cu^{+2}, Co^{+2}, Zn^{+2}

Thus, to avoid star activity, fewer units should be used which not only prevents over digestion but reduces final glycerol concentration in the reaction. High ionic strength should be used provided enzyme is not inhibited by high salt. pH should be lowered accordingly.

Restriction Mapping:-
A restriction map is a profile of restriction endonuclease cleavage in a piece of DNA. Generation of restriction map is the first step in characterizing an unknown DNA for the presence of recognition sequences. Restriction Endonucleases which are infrequent cutters are used for generating the maps. Fragments of varying lengths using various combinations of enzymes are generated and then analyzed on agarose gel. Single digestions are used to determine which fragments are in the unknown DNA, and double digest to predict the order and orientation of fragments. Success in using this technique relies on obtaining complete digestion of the DNA with each of the enzymes used since partial digest will yield false fragments which will lead to confusion.
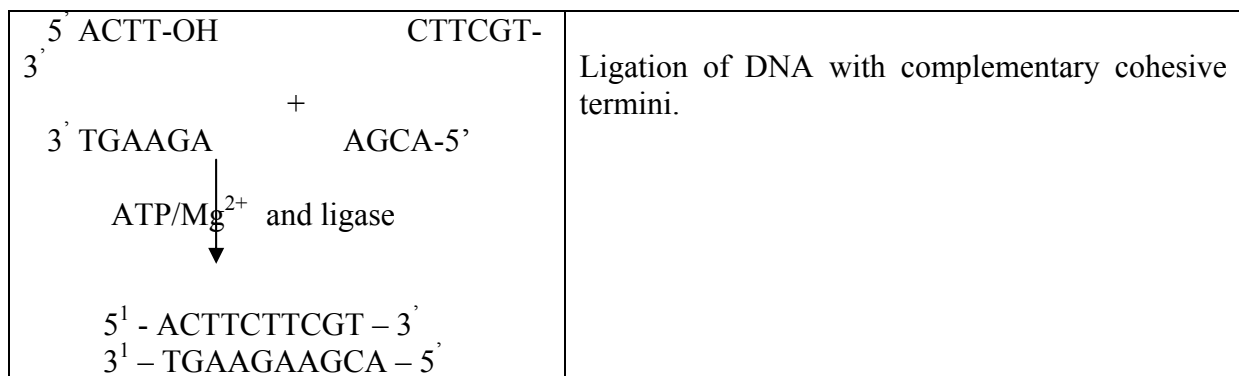
DNA Modifying Enzymes

1. DNA Ligase

Ligases are vital enzymes required for physiological cellular processes such as DNA replication, repair of damaged DNA and recombination. This enzyme facilitates formation of
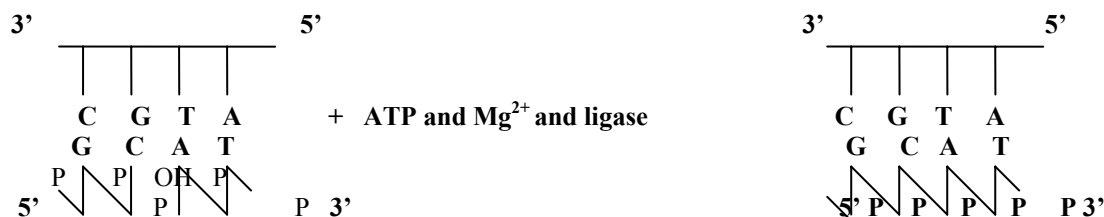
phosphodiester bonds between adjacent 3'OH and 5' phosphate termini, thereby joining the nicks in double strandeded DNA. There are 2 categories of ligase depending on the requirement for cofactors. Eukaryotic and virally encoded enzymes are ATP dependant, whereas most prokaryotic enzymes require NAD^+ for their activity as a cofactor.

ATP dependant DNA ligase

Ligase isolated from bacteriophage T7 is a 41Kda monomer and is found to form a complex with ATP. The structure involves an N terminal ATP binding domain and a C terminal domain which is the DNA binding domain. The N-terminal DNA is itself an active ligase but with the reduced ligase activity. Both domains are able to bind to single stranded DNA but smaller C terminal domain is only able to bind double stranded DNA. The affinity for nicked DNA comes only from reconstitution of these two domains. This enzyme also catalyses covalent joining of two segments to one uninterrupted strand in a DNA duplex, provided that no nucleotides are missing at the junction (repair reaction). For its catalytic activity, the enzyme requires ATP and Mg^{+2}. DNA which does not have phosphate residues can be rendered capable of ligation by Phosphorylation with Kinase.

$5'$ ACTT-OH CTTCGT-$3'$ + $3'$ TGAAGA AGCA-5' ATP/Mg^{2+} and ligase ↓ 5^1 - ACTTCTTCGT – $3'$ 3^1 – TGAAGAAGCA – $5'$	Ligation of DNA with complementary cohesive termini.

Repair Reaction:-

3' _____ 5'   3' _____ 5'

C G T A   + ATP and Mg^{2+} and ligase   C G T A
G C A T   G C A T
5' P P OH P   5' P P P P   P 3'
P   P   P 3'

Bacteriophage T4 DNA ligase is a polypeptide with a Molecular weight of 68000 Daltons. Mg^{+2} ions in 10mM concentration sulfhydryl agents (DTT, 2-mercaptoethanol) are also required for its catalytic activity. High NaCl concentration can be inhibitory. For intermolecular ligations, PEG (Polyethylene glycol) (1-10%) enhances the efficiency of ligation reaction. Addition of PEG can also facilitate blunt end ligation, whose efficiency is as such lower. Since there is no base pairing to hold fragments together temporarily, concentration of DNA and ligase must be high. The reaction which is driven by ATP is often carried out at 10^0C to lower the kinetic energy of molecules, and so reduce the chances of

base paired sticky ends parting before they have been fixed by ligase. But, then long reaction times are needed to compensate for the lowered activity of DNA ligase at lower temperatures.

NAD$^+$ dependent DNA ligase

Isolated from *E. coli*, is quite different from ATP dependent T4 DNA ligase. In this, there are two domains with the larger N-terminal domain retaining full activity while the C terminal domain having all the DNA binding activity. Thus, the scenario is different, in the context that, there both the domains contribute to the enzyme activity.

Units of Ligase Activity

Weiss unit: - is defined as the amount of ligase that catalyzes the exchange of 1nanomole of P^{32} from inorganic pyrophosphate to ATP in 20min at 37^{0}C.

2. DNA Polymerases

E.coli DNA polymerase I

E. coli DNA polymerase I is a single polypeptide chain of 109000Da, encoded by *E. coli pol A* gene and requires magnesium as a cofactor. Polymerase I carry out three enzymatic reactions that are performed by three distinct functional domains of the holoenzyme. Proteolytic deletions can be generated to study the function of each domain in terms of the activity performed. The domains of *E. coli* polymerase I are summarized below in a table which mentions the function of each subunit separately.

Domain	Catalytic Activity	Characteristic Function
C-terminal (46 kDa)	5$^'$-3$^'$ DNA Polymerase	Addition of nucleotide residue to 3$^'$ –OH termini of RNA/DNA primes
Central Domain (22 kDa)	3$^'$-5$^'$ exonuclease	Cleavage of nucleotide residue from 3$^'$ –OH, generating 3$^'$ recess
-NH$_2$ terminal	5$^'$-3$^'$ exonuclease	Cleavage of nucleotide from base paired 5$^'$ termini
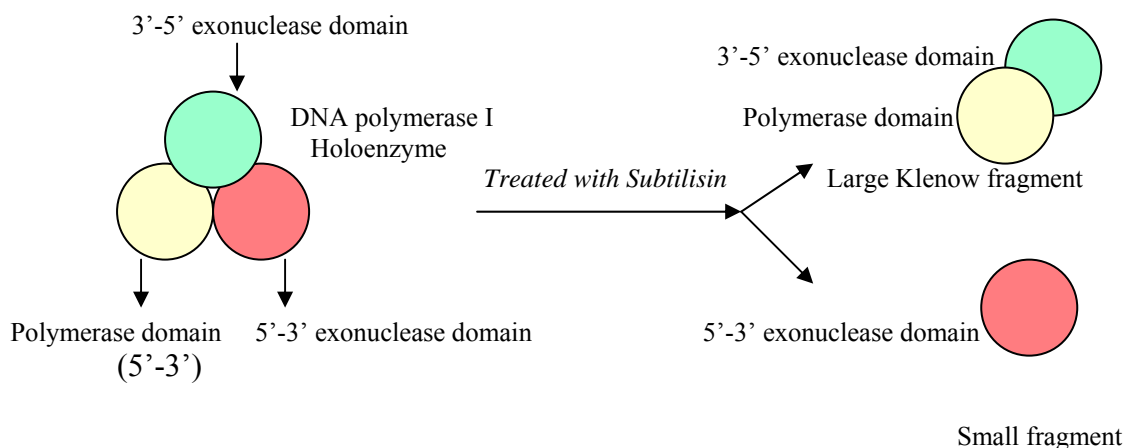
Polymerase I on mild cleavage with subtilisin releases two fragments, the larger one, known as Klenow fragment, carries both DNA polymerase and $3^{'}$-$5^{'}$ exonuclease while the smaller one possesses $5^{'}$-$3^{'}$ exonuclease activity of the holoenzyme. The exonuclease activity performs a proof reading function by rejecting DNA, which is mismatched.
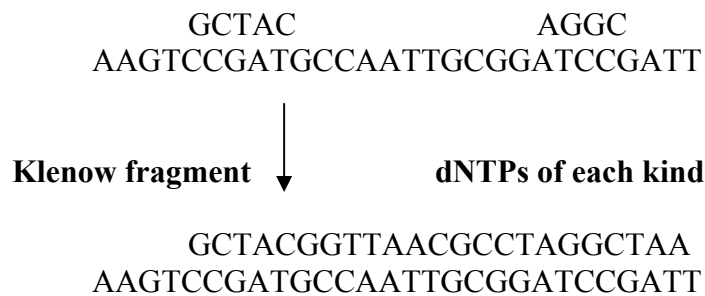
Units of DNA polymerase
One unit of DNA polymerase is the amount of enzyme required to catalyze the conversion of 10 nanomoles of total dNTPs to an acid-insoluble form in 30minutes at 37 ^{0}C using poly d [A-T] as the template primer. Mg^{+2} ions are preferred for accurate replication, whereas Mn^{+2} if present, increases the frequency of errors and mismatches.

**Application of Klenow fragment in Molecular Biology**

1.   Synthesis of double stranded DNA from single stranded template
The primary function of DNA polymerase is to synthesize complimentary strands during DNA replication. DNA polymerase requires a primer to provide $3^{'}$–OH group to which newer nucleotides can be added. The primers used are generally 6-20 bases in length, termed as oligonucleotides, which are complimentary to a specific region of template DNA. Shown below is the display of the catalytic activity carried out by the enzyme.

<div align="center">

GCTAC   AGGC
AAGTCCGATGCCAATTGCGGATCCGATT

</div>

<div align="center">

Klenow fragment ↓   **dNTPs of each kind**

</div>

<div align="center">

GCTACGGTTAACGCCTAGGCTAA
AAGTCCGATGCCAATTGCGGATCCGATT

</div>

2.   Filling in recessed $3^{'}$ ends of DNA fragments
Klenow fragment is also used to create blunt ends on fragments created by restriction enzymes that leave $5^{'}$ overhangs.

	Klenow and dNTPs of each kind	
5'AGGCAG3' 3'TCCGTCGAACT5'	→	5'AGGCAGCTTGA3' 3'TCCGTCGAACT5'
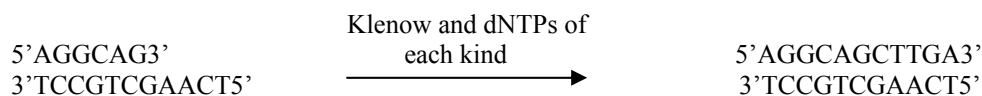
3.   Digestion of Protruding $3^{'}$ overhangs
This is another way of producing blunt ends on a DNA, which is created by restriction enzymes that produce $3^{'}$ overhangs. Removal of nucleotides from $3^{'}$ ends will continue, but in the presence of nucleotides, the polymerase activity balances the exonuclease activity, yielding blunt ends.

4.   Generating novel cohesive ends
The DNA digested with restriction enzymes generates cohesive ends that can be end-filled. The end-filling reaction can be controlled by omitting one, two or three of the four dNTPs from the reaction and thereby generate partially filled termini.
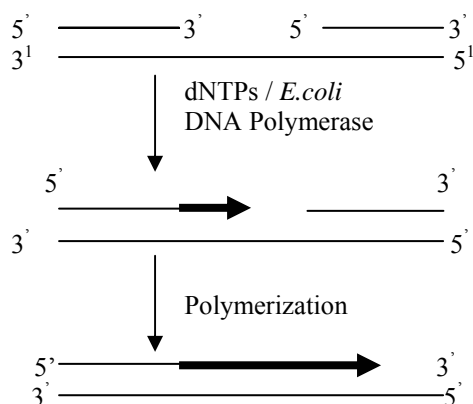
3. T4 Bacteriophage polymerase

T4 Bacteriophage polymerase is similar to *E. coli* DNA polymerase I larger subunit, Klenow fragment. Thus, it has $5'$-$3'$ DNA polymerase, $3'$-$5'$ exonuclease activities but $5'$-$3'$ exonuclease activity is missing. It is also used for similar type of work like creating blunt ends in a DNA with $5'$ or $3'$ overhangs. The functional difference between the two enzymes which have practical significance is as follows:-

1. $3'$-$5'$ exonuclease of T4 is 200 times more potent than that of Klenow which is a sluggish enzyme, making it a preferred enzyme for generating blunt ends in a DNA with $3'$ overhangs.
2. Like Klenow displace downstream oligonucleotides with $5'$-$3'$ exonuclease activity as it extends the chain, T4 polymerase does not. This makes it a preferred choice for oligonucleotides directed mutaqenesis.

Use of Polymerase I- Nick Translation

Labelling of DNA by nick translation

Enzyme binds to a nick or gap in duplex DNA. The 5'-3' exonuclease activity of Polymerase I remove nucleotide from one strand and create a template for the growing chain. The original nick is therefore translated along the DNA molecule by the combined action of $5'$-$3'$ exonuclease and $5'$-$3'$ polymerase. This is the oldest method of nucleic acid labeling and is still used commonly. The DNA preparation to be labeled is treated with DNase I for a very short time; this induces nicks in the DNA molecule at random positions. A nick is a point in a DNA duplex where the phosphodiester bond is broken and has a free 3' OH group. The nicked DNA is subjected to *E. coli* DNA polymerase I which successively adds new nucleotides to the free 3'OH group. As the new bases are synthesized, the existing strand is displaced progressively and is digested away by the enzyme 5'-3' exonuclease activity of the polymerase enzyme. In this reaction, usually one of the deoxynucleoside triphosphates is radiolabel led or even all the four can be labeled, thus incorporating labeled nucleotides into the newly synthesized segment. Since this segment is synthesized by complimentary base-pairing, it has the same sequence as the strand it replaces. This process is called '**nick translation**' because there is a movement (translation) of the nick along the DNA duplex due to the activities of DNA polymerase I.

Nick translation is usually carried out at lower temperature to reduce the synthesis of snap back DNA, which is produced when 3'-OH terminus of growing strand loops back on itself

4. T7 DNA Polymerase

The T7 DNA polymerase from T7 bacteriophage has $3^{'}$-$5^{'}$ exonuclease and DNA polymerase activity but lacks $5^{'}$-$3^{'}$ exonuclease domain, which is similar to T4 DNA polymerase. The processivity of this enzyme is quite good that is, the average length of DNA synthesized before the enzyme dissociates from the template, is considerably greater than for other enzymes. Thus, the average length of DNA synthesized by a single molecule of bacteriophage T7 polymerase is much greater than that of DNAs synthesized by other DNA polymerases. The binding and polymerization domain is occupied by the carboxy terminus while the potent 3'-5'exonuclease activity resides on the amino terminus.

The exonuclease activity is completed inactivated by incubating the enzyme with a reducing agent, molecular oxygen, and low concentrations of ferrous ions, for several days. Over 99% of the exonuclease activity is abolished without affecting the polymerization activity by these agents, which cause mutations and site specific modifications. The resulting chemically modified enzyme is marketed under the trade name **Sequenase** is ideal for determining the sequence of long tracts of DNA by the dideoxy mediated chain termination method.

5. Thermostable DNA polymerase

Taq DNA polymerase:-
Taq DNA polymerase is a DNA dependent DNA polymerase, first isolated from the hot spring bacterium, *Thermus aquatics* in 1976 and in 1989. Due to its wide use in molecular biology (primarily PCR), it is termed as 'Molecule of the Year'. This thermophilic DNA polymerase encodes an 832-amino acid, 94 kDa protein, which consists of two domains.

1. -NH$_2$ domain: Similar to $5^{'}$-$3^{'}$ exonuclease domain of members of polymerase I family of DNA polymerase
2. -C terminal domain contains a catalytically inactive $3^{'}$-$5^{'}$ exonuclease and a polymerase sub domain, similar to klenow of DNA polymerase I.

The thermal stability of Taq DNA polymerase is attributed to its hydrophobic core and stable electrostatic interactions and high density of proline residues on the surface of the enzyme. The optimal activity is at 75-80^{0}C temperature and at 60^{0}C, the activity is reduced by a factor of 2 and at 37^{0}C, its activity is reduced to only 10%. To initiate DNA synthesis, like other DNA polymerases, it also requires a primer that is annealed to the template strand and caries an extensible $3^{'}$-OH group. Taq DNA polymerase requires Mg2+ for its optimal activity. Phosphate buffers inhibit Taq DNA polymerase and therefore should be avoided. The reaction is usually carried out in the presence of Tris buffer at pH 8.3. Because of the lack of a proofreading function, the rate of misincorporation of dNTPs is high in PCR reactions which are catalyzed by Taq polymerase (or any other DNA polymerase that does not have editing domains). Several mutant forms of native polymerases, are also available like Pfu And vent. Both of them have proofreading activities contributed by $3^{'}$-$5^{'}$ exonuclease. Pfu polymerase is therefore known to generate lowest errors while vent is probably intermediate between Taq and Pfu.

Pfu polymerase is isolated from *Pyrococus furiosus*
Vent is isolated from *Thermococcus litoralis* (also known as Tli polymerase)

6. Reverse Transcriptase (RT)

Reverse Transcriptase is a RNA dependent DNA polymerase, encoded by reteroviruses, where there is a requisite of converting viral RNA genome into DNA, prior to its integration into the host cells. Reverse transcriptase has two independent functional activities in two separate domains.

1. DNA polymerase activity: As its native function, Reverse Transcriptase copies only RNA, but *in vitro*, it can transcribe both single stranded RNA and single stranded DNA templates with equivalent efficiency. And in both cases, RNA or DNA primer is required to initiate synthesis.
2. RNaseH activity: Ribonuclease that degrades RNA from RNA-DNA hybrids. The products generated includes ribonucleotides which are 4-20 nucleotides in length and contain 5'phosphates and 3' OH terminus

Reverse transcriptase is used chiefly to transcribe mRNA into double stranded cDNA that can be inserted into prokaryotic vectors. Reverse transcriptase lacks 3'-5' exonuclease activity which acts as an editing function in *E. coli* DNA polymerase I, and is therefore more prone to errors

Usually Reverse Transcriptase for laboratory purposes is isolated from one of the following reteroviruses, either by purification from virus i.e the native host or from heterologous hosts as a recombinant protein.

a.) *Moloney Murine leukemia virus* - Single polypeptide
b.) *Avian myelobalstosis virus*- Two polypeptide chains

Both of these differ from one another in a number of aspects. Mo-MLV Reverse Transcriptase contains two non overlapping domains, larger $-NH_2$ domain with DNA polymerase activity, while C-terminal domain encompasses a comparatively weak RNaseH activity which has an advantaqe over the avian form, when synthesizing cDNAs complimentary to long mRNAs. Because at the beginning of cDNA synthesis, there is a competition between degradation of template mRNA and initiation of DNA synthesis. The presence of RNaseH activity leads to premature termination of RNA-DNA duplex. Therefore, certain mutants are made that retains DNA polymerase activity but lacked RNase H activity. Serendipitously, it has been found that due to these mutations, the enzyme can work at elevated temperatures, as high as 50^0C, whereas RNase H deficient mutant forms of Avian enzyme can be used at 60^0C while the murine enzyme is rapidly degraded at these temperatures. These mutated thermostable forms of Reverse Transcriptase facilitated their use in RT PCR which itself has myriad applications in Molecular Biology. The avian enzyme works more efficiently at pH 8.3 than at pH 7.6, at which the murine enzyme works.

7. Nucleases

Enzyme	Characteristic Feature (Activity catalyzed)
Nuclease S I	Heat stable, isolated from fungi *Asperigillus oryzae* ,works optimally at high ionic strength, low pH (~4.2) in the presence of Zn^{+2}, degrades single stranded DNA / RNA, chief product is 5' mononucleotide
Exonuclease VII	Single strand specific DNase that hydrolyses denatured DNA, Single stranded regions extending from the termini of duplex DNA, chief product is oligonucleotide (2-25 nucleotides), does not require divalent ions, has several applications
Mung bean Nuclease	To eliminate primers from completed PCRs, isolated from mung bean sprouts, degrades Single stranded DNA / RNA generating 5' mono and oligonucleotides, very specific for single stranded DNA but if used in high concentration or with long incubation time, can cleave double stranded DNA also. Used to convert protruding termini on double stranded DNAs to blunt ends. It is easier to control than SI nuclease. SI will cleave the DNA strand opposite to a nick while mung bean nuclease will attack the nick only when it is enlarged to a gap of several nucleotides, prefers to remove nucleotide from 5' end of the substrates and therefore not useful when DNA carries 3' overhangs
BAL 31	Isolated from marine bacterium *Alteromonas espejiana* , Ca^{+2} dependent exonuclease, has 3'-5' exonucleolytic activity that removes mononucleotides from double stranded DNA, a 5'-3' exonuclease activity for single stranded DNA. The combination of both exonuclease causes both stands of double stranded DNA to be shortened from both the ends. It is used to generate uni and bi-directional deletions from the duplex DNA cloned in any vector.
Exonuclease III	Isolated from *E.coli* removes 5' mononucleotides from 3'- OH of duplex DNA, leaving protruding 5' termini. Preferred substrate is DNA with blunt or 5' protruding ends. It has several applications :- a) Generating single stranded templates for sequencing b) To produce unidirectional sets of progressive deletions in populations of linear DNA molecules with one resistant (3') terminus and one susceptible (blunt or 5') terminus, inactive on single stranded DNA.
Ribonuclease T1	Isolated from *Aspergillus*, endonuclease that cleave RNAs at 3'–(p) of G residues, producing oligo nucleotide terminal guanosine 3' phosphates. It is used to remove unannealed regions of RNA from RNA-DNA hybrids.
Ribonuclease A	Endoribonuclease that cleave single stranded RNA at 3' pyrimidine residues, generating 3' phosphorylated mononucleotides and oligonucleotides.
Deoxyribonuclease	Endonuclease which cleaves double stranded or single stranded DNA preferentially adjacent to pyrimidine (c or T) residues generating 5'-phosphorylated di, tri and tetra nucleotides, works optimally in the presence of Mg2+ cleaving duplex DNA independently, while in the presence of Mn^{+2}, the enzyme cleaves both the strands of the DNA at approximately same site, producing blunt ends or fragments with 1-2 base overhangs. It doesn't work on RNA.

8. Alkaline Phosphatase (AP)

Alkaline Phosphatase is an important tool in molecular biological processes like cloning. It removes $3^{'}$- phosphate groups from a variety of substrates. Although in laboratory, it is used to catalyze the removal of terminal $5^{'}$-(P), residues from single stranded or double stranded DNA and RNA. The resulting $5^{'}$-OH termini can no longer take part in ligation reactions, thus prevents self religation of vectors, reducing the background of transformed bacterial colonies that carry empty plasmids. This enzyme works optimally at alkaline pH (range of 8-9 in the presence of low Zn^{+2} concentrations) and hence derived the name.

Alkaline Phosphatase is isolated from various sources:-

a) Bacterial Alkaline phosphatase
Secreted in monomeric form into the Periplasmic space of *E.coli*, where it form dimers and gets catalytically activated. It's a remarkably stable enzyme and is resistant to inactivation by heat and detergent. Thus, bacterial alkaline phosphatase is the most difficult to destroy in the reaction mix.

b) Calf Intestinal Phosphatase
Calf intestinal phosphatase is a dimeric glycoprotein isolated from bovine intestine. This has much more practical significance than bacterial alkaline phosphatase, since it can be readily inactivated from the reaction mixture using proteinase K or by heating at $65^{0}C$ for 30 minutes or $75^{0}C$ for 15 minutes in the presence of 10mM EGTA.

c) Shrimp alkaline phosphatase
Extracted from cold water shrimp, can be inactivated readily by heating at $65^{0}C$ for 15 min.

9. Terminal Deoxy Nucleotidyl Transferase

Terminal transferase is an unusual DNA polymerase found only in prelymphocytes and in early staqes of lymphoid differentiation. Synthesis of single stranded tails at the 3' ends of either single stranded DNA or double stranded DNA with protruding 3' termini, by the enzyme Terminal Deoxy Nucleotidyl Transferase, is called tailing, can be used to generate protruding ends of defined sequence to facilitate cloning of fragments. It can be used to generate protruding ends of defined sequence, e.g poly A tails on the 3' ends of the DNA insert and poly T tails on 3' ends of the vector. Thus, the protruding ends of the DNA insert and vector will base pair under appropriate annealing conditions. Mg^{2+} cation is preferred when the nucleotide to be added is a purine while Co^{2+} is preferred for the addition of pyrimidines. The enzyme strongly prefers DNA with protruding 3' terminus, although blunt or recessed 3' termini are also used, but less efficiently, in buffers of low ionic strength with Co^{2+}, Mg^{2+} or Mn^{2+} as bivalent cations. As many as thousands of deoxynucleotides can be incorporated using this enzyme on a template of DNA. Single nucleotide can be added to the 3' termini of DNA if modified bases like dideoxynucleotides or cordycepin triphosphates are used instead of natural deoxynucleotide triphosphates.

Polymerase Chain Reaction (PCR)
PCR is a robust, speedy and flexible method, conceived by and put in practice by Kary Mullis in 1985. It is used to generate microgram quantities of DNA (up to billions of copies) of the desired DNA (or RNA), from a single copy in few hours. The PCR process has been

completely automated and compact thermal cyclers are available in markets. The fundamental of PCR reaction with its components are as follows:

1. A thermostable **DNA polymerase** to catalyze template dependent synthesis of DNA
   For routine purpose, Taq polymerase is the enzyme of choice. However, when great fidelity is required, when working with larger lengths of target amplicons, mutant forms of Taq, having proof reading exonuclease activity are used to generate faithful copies of a gene or else cocktail of both enzymes are used to compensate for the low processivity of proof reading enzymes. eg. A mixture of Taq & Pfu polymerase has desirable features of both high processivity (a feature of Taq polymerase) and fidelity which is contributed by Pfu in one Reaction mixture, one can generate high yield of accurate amplified products up to 35 kb.

2. A pair of synthetic **oligonucleotides** to prime DNA synthesis.
   Appropriate designing of primers is essential for accurate. Appropriate designing of primers is essential for accurate amplification avoiding unwanted sequence accumulation. The concentration of primer is also crucial. Lower concentrations can limit the reaction while high concentrations favor mispriming & non-specific amplification.

3. **dNTPs**
   Equimolar amounts of all the four nucleotides, dATP/ dTTP/ dCTP/ dGTP are required to carry out a PCR Reaction.

4. **Divalent cations**
   Divalent cations are required by all thermostable polymerases for their optimal activity. The most routinely used Taq Polymerase required Mg^{+2}. Since dNTPs & oligonucleotides are known to sequester Mg^{+2} ions, the molar concentrations of phosphate groups of dNTPs & primers should be standardized. Thus optimal concentration of ions for each primers combination & template is crucial.

5. **Monovalent cations**
   Standard PCR buffers are supplemented with 50 mM KCl.

6. **Buffer**
   Tris-HCl of pH 8.3 and 8.8 is included in standard PCR reaction at a ionic strength of 10 mM. When incubated at $72^{o}C$, a temperature at which Taq Polymerase carries out polymerisation reaction, the pH of Reaction mixture drops by more than 1 unit, producing a buffer of pH 7.2.

7. **Template DNA**
   A template is the DNA containing target sequence to be amplified. The preferred substrate for DNA amplification is linear DNA rather than closed circular DNA template. The size is not an important parameter but working with large size DNA may pose some obstruction. Ideally, one single copy is enough to initiate a PCR reaction but several thousand copies are seeded while putting a PCR Reaction. The recommended amounts of template DNA in case of yeast DNA, bacterial DNA and plasmid DNA that should be used per reaction are 10ng, 1ng & 1 pg respectively.

PCR Iterative programming

A PCR cycle comprises of basically three repeated reaction including denaturation of the template DNA followed by primer annealing to target sequence & extension of the primers by DNA polymerase.

DNA denaturation

The optimum temperature for DNA denaturation is determined by the % GC content of each particular template DNA. The more is the proportion of GC, higher temperatures are required to denature the parental duplex strands. At lower temperature, only A-T regions start melting. Longer DNA templates require longer denaturation time to separate them completely. Routinely 95^0C is the temperature used for denaturation, since it is the maximum temperature at which Taq Polymerase can exhibit its activity. For GC rich DNA, thermostable polymerase isolated from Archae that can sustain even much higher temperature, without any deleterious effects, are preferred.

Annealing

The temperature which is required to anneal primers to the template DNA is critical. It should not be too high to avoid poor annealing and even not too low which facilitate nonspecific annealing, resulting in accumulation of unwanted sequences. As such there is no proper formula to calculate annealing temperature. Conditions are optimized by performing several hit & trial PCRs at temperature ranging from 2-10^{0}C below the melting temperature, calculated from the oligonucleotide primer sequence. The most commonly used formula to determine the Tm from primer sequence is: 4(G+C) + 2(A+T), where G/C/A/T denotes, number of times these bases are present. When the temperature is lowered to carry out annealing Reaction, there are chances of renaturation of template DNA.

Extension

Extension reaction is usually carried out at temperature favorable to the themostable polymerase used in the Reaction. For Taq polymerase 72-78^{0}C is the optimum temperature. In 1st two cycles, polymerase extends beyond the sequence complementary to the binding site of the primer but after the 3rd cycle, only sequence that is defined & limited by the binding of primers are accumulated in a geometrical progression. As a thumb rule, for every 1000 base pair, 1 minute is used.

No. of cycles

Number of cycles that are to be used to amplify a particular region from the target DNA depends on the initial concentration of template present and the efficiency of primer extension and amplification.

Variations from the conventional PCR

Hot start PCR

It a general modification in the protocol followed for setting a conventional PCR to maximize the yield of desired amplified product and simultaneously to suppress any non specific amplification. This is accomplished by withholding an essential component that plays a critical role in the reaction e.g DNA polymerase. It is not added until the temperature is raised high enough that inhibits non specific priming or oligomerisation of primers. Then to the preheated mixture, Taq DNA polymerase is supplemented to facilitate the extension reaction. Initial methods also involve use of monoclonal antibodies against polymerase. As the temperature rises, the antibody dissociates from the enzyme & renders it active. Another

method is the use of enzyme 'Ampli Taq Gold' which requires heating to 93^{0}C for 10 minutes to become fully active.
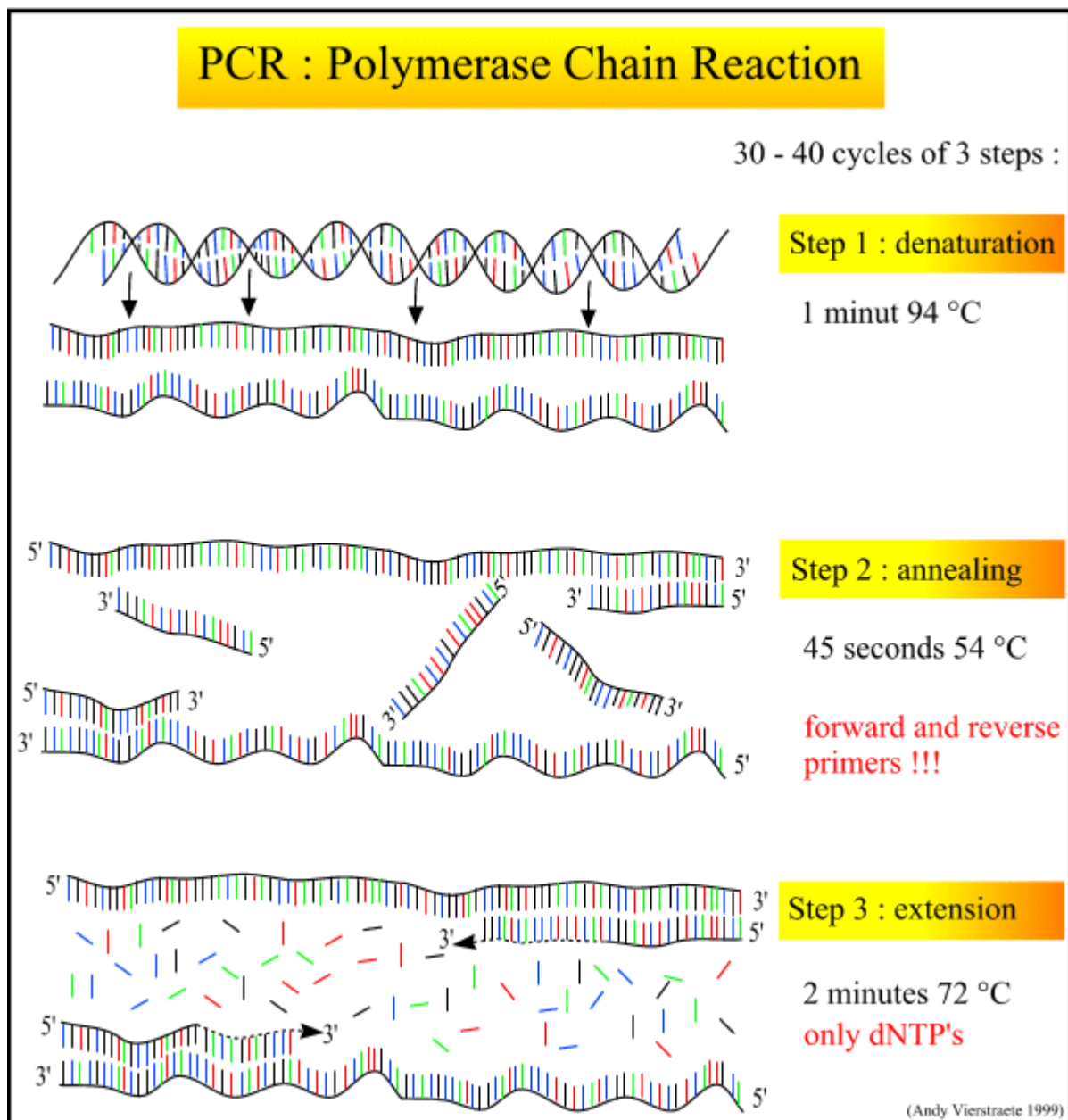

Figure 1: Because both strands are copied during PCR, there is an **exponential** increase of the number of copies of the gene. Suppose there is only one copy of the desired gene before the cycling starts, after one cycle, there will be 2 copies, after two cycles, there will be 4 copies, and three cycles will result in 8 copies and so on. Thus, Polymerase chain reaction (PCR), is a common method of creating copies of specific fragments of DNA. PCR rapidly amplifies a single DNA molecule into many billions of molecules. In one application of the technology, small samples of DNA, such as those found in a strand of hair at a crime scene, can produce sufficient copies to carry out forensic tests.
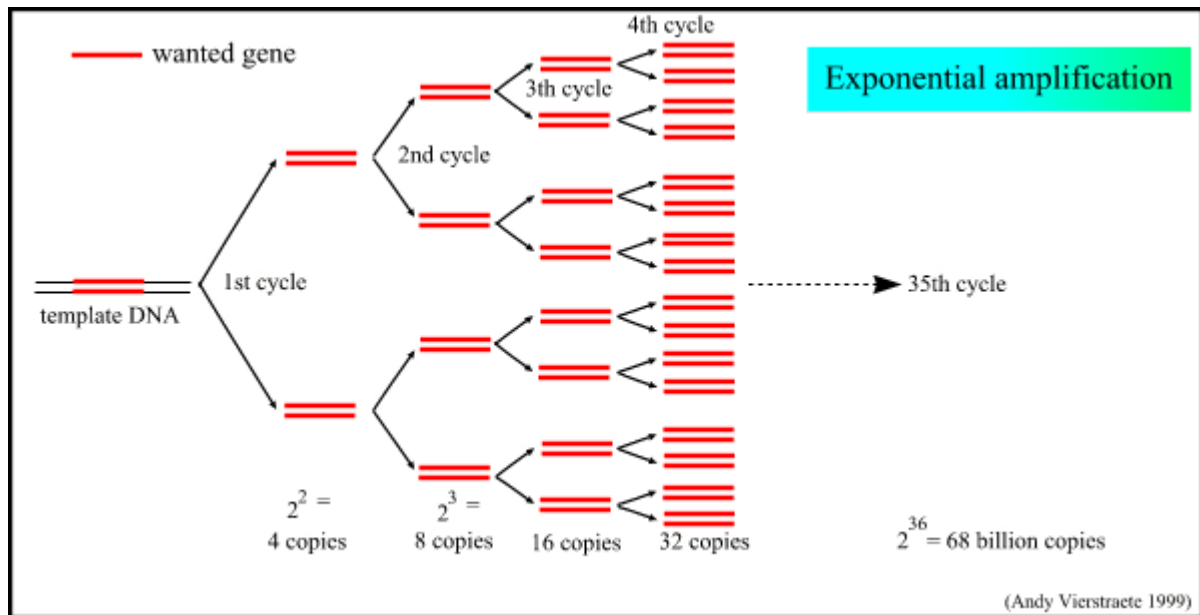
The exponential amplification of the desired gene in a PCR reaction

Touch down PCR

It is another variation of conventional PCR that is followed to optimize yields of specific amplified DNA when the Tm of hybrids between oligonucleotides (primers) and their target sequences is not known with certainty. In this, a range of annealing temperature is utilized in PCR. The annealing temperature during the 1st two cycles of amplification is set to 3^0C above the calculated Tm of a perfect hybrid between most GC rich oligonucleotide in the Reaction & its target DNA. Then the annealing temperature is reduced by 1^0C each cycle. When appropriate temperature is encountered, amplification of target sequence begins. But lowering of temperature continues till the temperature at which nonspecific priming can also occur, but till that time, the desired product will be in hegemony.

Nested PCR

It is used to amplify specific sequences of DNA from a large complex DNA. It utilizes two pairs of PCR primers for amplification of a single locus. The first pair amplifies the locus as done by any PCR. The second pair of primers called as nested primers bind within the 1st PCR product & produces a second PCR product that will be shorter than the first one. The advantage of using nested PCR is, if wrong locus were amplified by mistake, the probability of inaccurate second amplification will be low.

Anchored PCR

When sequence of only one end of the desired segment or gene is known, the primer complimentary to the 3' end is used to produce several copies of only one strand of the gene. Now a poly G (or any other homopolymer) tail is added to the 3' ends (the end for which the primer is not available) of the single stranded DNA copies produced by PCR. This allows the use of complimentary homopolymer poly C, to be used as primer for copying the DNA generated by PCR, and give rise to the complete DNA duplex that can be amplified.

Asymmetric PCR

It is used to generate single stranded copies of DNA sequence which can be directly used for DNA sequencing. To achieve this, the quantities of two primers for the 3' end are so adjusted

in the reaction mixture that one of them is exhausted before the termination of PCR reaction. As a result, in the last cycles, only a single strand of DNA is copied which are the ideal starting materials for DNA sequencing.

Inverse PCR

Inverse PCR is used to clone sequences flanking a known sequence. Flanking sequences are digested and ligated to make a circular DNA. PCR primers pointing away from the known sequences are used to amplify the flanking sequences. Inverse PCR (IPCR), described by Ochman et al in 1988, is a method for the rapid *in vitro* amplification of DNA sequences that flank a region of known sequence. The method uses the polymerase chain reaction (PCR), but it has the primers oriented in the reverse direction of the usual orientation. Inverse PCR has many applications in molecular genetics, for example, the amplification and identification of sequences flanking transposable elements.

Degenerate PCR

Degenerate PCR is in most respects identical to ordinary PCR, but with one major difference. Instead of using specific PCR primers with a given sequence, you use mixed PCR primers. That is, if you do not know exactly the sequence of the gene you are going to amplify, you insert "wobbles" in the PCR primers where there is more than one possibility. Degenerate PCR primers used to amplify unknown DNA sequences based on similar known sequences. Primers which have a number of options at several positions in the sequence so as to allow annealing and amplification of a variety of related sequences.

Introduction of restriction endonuclease sites in the PCR product

In another variation, the recognition sequence for a given restriction endonuclease are added to the 5' ends of both the primers used in the PCR amplification of the desired DNA segment. The recognition sequence will remain as single stranded tails when the primers are annealed to the original template DNA. This results in amplification of desired segments of DNA along with inbuilt endonuclease sites which are used during subsequent cloning of the amplified segment in different vectors.

PCR to introduce mutations at specific sites- Site Directed Mutagenesis

PCR can be used to induce desired mutations at specified sites in any gene. The sequence of the site where mutations has to be created is used to make oligonucleotide primers. The mutations in the desired gene segment are introduced by designing primers with desired changes in their base sequence. These modified primers are used in PCR amplification of the desired gene. Thus, the new copies of the gene so generated, carry the appropriate mutations at specific sites, which are their in the base sequence of primers.

Vectors

Plasmids

The word "plasmid" was introduced by 'Joshua Lederberg' in 1952 and is defined as an extrachromosamal circular DNA molecule which are not part of the bacterial genome. A plasmid is an independent, circular, self replicating DNA molecule that carries few genes only. A plasmid is usually not essential for the host cellexcept under some specific conditions. The numbers of plasmids within a cell generally remains constant from generation to generation and have evolved mechanisms to partition plasmid molecules accurately to daughter cells during replication. Plasmids are usually dependent on host synthesized enzymes and proteins for their replication and transcription. Their size ranges

from 1-2000 kilo base pairs and they carry additional functions that are advantageous to the host such as they produce enzymes which degrade antibiotics or heavy metals or complex organic compounds or produce colicins, enterotoxins etc or produce restriction and modifying enzymes. Replication of plasmid DNA is coupled with the host replication machinery in two different ways:

Stringent control: in which one or two plasmids are made during every round of host DNA replication
Relaxed control: where 10-200 copies of the plasmid is made during each round of bacterial replication.

The three most widely studied types of plasmids are:
1. F plasmids: responsible for conjugation
2. R plasmids: carry genes for resistance to antibiotics
3. Col plasmids: code for colicins, the proteins that kill sensitive E.coli cells and also carry genes that provide immunity to the particular colicin

Plasmid incompatibility

Plasmid incompatibility arises when two different plasmids share same replication machinery. Then the two plasmids compete with each other during replication and subsequent step of portioning into daughter cells. Thus, plasmids carrying the same replicon belong to the same incompatibility group. These plasmids have to be maintained by using antibiotics as selection pressure.

Plasmids as cloning vector

A vector is any DNA that can take foreign DNA by ligation using complementary site in the plasmid. These sites are generated by digesting the DNA & vector with the same Restriction enzymes. It should be ensured that the restriction enzymes used for cloning purpose should be present only once in the plasmid so that a single linear molecular can be generated and sites should not be present in the foreign DNA to be inserted. Vectors contain genetic markers that confer strong growth advantages under selective conditions. The plasmid DNA containing foreign DNA as insert is usually introduced artificially by transformation. Selection is then carried out by screening for the colonies those have acquired the plasmid and hence grow on Antibiotic containing plates. Routinely used Antibiotics include Ampicillin, Kanamycin, chloramphenicol, carbenicillin & tetracycline.

Most versatile cloning vectors should have small size, high copy number and are able to accept fragments of foreign DNA generated by cleavage with a wide variety of Restriction enzymes. Smaller size of plasmid facilitates better stability and higher efficiency of transformation. Earlier used plasmids in cloning, either replicate poorly or carried unsuitable selectable markers and very few restriction sites. The first plasmid with all desirable features was pBR322 that has relaxed mode of replication with high copy number, ampicillin & tetracycline as selectable markers with important restriction enzyme sites. *Pst* I and *Pvu* I are located within ampicillin resistance gene and *BamH* I and *Sal* I are located within tetracycline resistance gene. The presence of restriction sites within the antibiotic markers tetracycline and ampicillin permits an easy selection for cells transformed with pBR322. insertion of foreign DNA fragment into the plasmid using restriction enzymes *Pst* I or *Pvu* I, places the DNA insert within ampicillin resistance gene and makes it non functional. Bacterial cells

containing such a recombinant will be unable to grow in the presence of ampicillin but grow well on tetracycline containing plates. Similarly when *BamH* I or *Sal* I is used, the DNA insert is placed within the tetracycline resistance gene making it non functional. Such recombinants grow well on ampicillin but not on tetracycline. Thus it allows an easy way of selection of a single bacterial cell having recombinant pBR322 from 10^8 other wild type pBR322 containing cells. The name PBR denotes the following: p signifies plasmid, B is from Boliver and R is from Rodriguez, the two scientists who developed pBR322. Nowadays, it is the most widely used cloning vehicle and many descendant forms are extracted from it. Another series of vectors, pUC vectors (derived its name from the place where it was developed, University of California) revolutionized the cloning strategies. The number of restriction enzyme sites was expanded. Moreover, it is the first plasmid to contain a series of synthetic cloning sites, termed as polylinkers, MCS (Multiple Cloning Sites) or polycloning sites eg. pUC19 has an array of 13 restriction enzyme sites which are very unique and can be used efficiently for cloning.

Advantages of pBR322 over pUC series of vector

pBR322 carried two different selectable markers, tetracycline and ampicillin. Insertion of foreign DNA sequences into one of these sites inactivates either of the two markers. The recombinant plasmids can therefore be distinguished from the empty parental vector by virtue of its ability to grow only one Antibiotic. This phenomenon is called as **Insertional inactivation** which is not possible in pUC vectors which carry only one Antibiotic resistance gene, usually for Ampicillin and a clustered set of MCS. But pUC series of vector demonstrate a phenomenon called as **α complementation.** The recombinant plasmids can readily be distinguished from parental pUC plasmid by visualizing the color of bacterial colonies. PUC vectors and their derivatives carry a short segment of *E.coli* DNA that has regulatory sequences of *lacz* gene and code for $-NH_2$ termini of β-galactosidase (called as α fragment). The fragment expressed by pUC vector (α), when introduced in *E.coli* strains that contain inactive C- terminal fragment (ω-fragment), complements for the β-galactosidase activity. But if a segment of foreign DNA is cloned into the site of pUC vector, carrying α fragment, the coding region gets disrupted and α complementation is abolished. Thus, bacterial colonies containing recombinant plasmids are therefore ampicillin resistant but contain no β –galactosidase activity. Contrary, to this, the parental plasmid carrying colonies, will also be ampicillin resistant but are able to hydrolyze chromogenic substrate X-gal (5-bromo-4-chloro-3-indole- β –D-galactoside**)** by cleavage of β –galactoside bond. Thus, the plate will contain two visually distinct colonies: The parental non recombinant deep blue colonies and recombinant white colonies.

Plasmid vectors with origin of replication derived from single stranded Bacteriophages

M13 or f1 are single stranded filamentous bacteriophage. Many routinely used vectors contain origin of DNA replication derived from these bacteriophages. Such vectors are often called **phagemids.** They confer advantage of two separate modes of replication, one as conventional double stranded DNA and other as a template to produce single stranded copies of one of the phagemid strand. Thus single strands of DNA can be packed into the phage particles. The production of single stranded DNA is induced when bacteria carrying phagemid are infected with a helper bacteriophage that can supplement for the genes required for production of virions *in trans*. The orientation of the origin determines which of the two strands will be encapsulated in the virus particles. There are many plasmid vectors that carry promoters derived from bacteriophages like T3, T7 or SP6, adjacent to the MCS. Hence, the

foreign DNA so inserted can be transcribed *in vitro* in the presence of specific RNA polymerase and Ribonuclease precursors. Presence of bacteriophage promoters on both sides of MCS allows either of the stands to be transcribed.

Expression plasmid vectors

There is an array of plasmid vectors that can generate large amounts of mRNA *in vivo* from cloned foreign genes. The requisite features of a plasmid to serve as an expression vector are:
  a)   Minimal level of basal expression under repressed condition.
  b)   Shine Dalgarno sequence upstream of ATG initiation codon.
  c)   Presence of tags which are recognized by specific antibodies. Epitope tagged proteins thus can be visualized by probing with antibodies directed against tags and can be purified from rest of the host protein using affinity chromatography.
  d)   An appropriate system for fast induction of expression of cloned gene.

Cloning

PCR Mediated Gene Cloning

There are three strategies for cloning PCR products-
1) **T/A cloning** is the easiest cloning method. T/A cloning takes advantage of the terminal transferase activity of *Taq* polymerase and other non-proofreading DNA polymerases which adds a single 3'-A overhang to each end of the PCR product. The resulting PCR product is then ligated into a linear vector with a 3´ terminal 'T' or 'U' at both ends.

2) **Directional cloning**. A restriction enzyme target site is introduced into each of the PCR primers. The resulting PCR product and cloning vector are digested with the restriction enzymes to generate complementary ends at the PCR product and the vector which are then ligated.

3) **Blunt-end cloning**. Blunt-end PCR product generated by proof-reading polymerase such as the Pfu DNA Polymerase can also be cloned into a blunt-end vector.

The cloning of PCR-amplified fragments into a linear vector is typically a rapid and efficient process. However, not all PCR fragments will clone with the same efficiency into the same vector. These differences may be due to fragment size, insert toxicity, and the complexity of the insert. The size of the fragment being cloned is a primary contributor to the overall cloning efficiency. Large fragments of DNA ($\geq$ 5 kb) are amenable to cloning in high-copy number vectors, yet at a much lower efficiency. Optimization of molar concentration ratios of the vector to insert is critical to ensure efficient cloning. Successful cloning ratios may range from 1:1 to 1:10. For example, if the vector is 3 kb and the insert is 1 kb, one-third the amount of insert needs to be added to attain a 1:1 molar ratio.

Cloning using plasmid vectors

Cloning in brief is cleaving closed circular plasmid DNA with one or more Restriction enzymes and ligating it *in vitro* to foreign DNA bearing compatible terminus. This is then transformed into an appropriate *E.coli* strain and the transformants are then screened for

recombinants by either PCR using gene specific primers or restriction enzyme digestion to check for the presence of desired gene fragment in the vector.

Cloning of DNA with overhangs

Restriction enzymes are known to cleave the DNA within their recognition sequences but asymmetrically and thus generate 1-6 base pair long overhangs or protruding ends. When the ends protruding from foreign DNA fragment and the vector are compatible, they can anneal by hydrogen bonding and form a circular recombinant plasmid that can be easily transformed. Annealing reaction involves bringing of 5′ phosphate and 3′ –OH residues together and finally formation of a phosphodiester bond catalyzed by the enzyme DNA ligase.

Directional cloning

1. Both ends cohesive (protruding) and separately matched
This situation arises when the foreign DNA fragment is produced by digestion with two restriction enzymes having different Recognition sequences. Hence the foreign DNA fragment will be non complementary and unable to ligate to each other. But it will ligate to the plasmid vector which is cleaved with the same two restriction enzymes. This will generate circular recombinants containing a single insert in a predefined orientation. Also in this situation, only the recombinant vector, containing the insert, will circularize since the ends of the vectors are non compatible to ligate to each other, as they are generated by cutting with two different restriction enzymes. The scheme of directional cloning is shown below in Figure2.

2. Both ends cohesive and compatible
This situation arises when both vector and insert are cleaved using the same restriction enzyme. Both the vector DNA and the foreign DNA are mixed under annealing conditions which allow pairing between cohesive ends of the vector and the insert DNA. The DNA insert can go in either of the two orientations. The orientation of the inserted DNA is significant when the expression of the insert is required. In addition to the formation of recombinant vector few undesirable products will also be formed like: the cohesive ends of the vector will itself pair to form unaltered circular vector. Similarly, the two ends of the insert DNA can also join to form concatemers, which may not pose a problem since it lacks an origin of replication and will therefore dilute out. The generation of re-circularized vector DNA can be prevented by treating the digested vector with alkaline phosphatase which removes 5' phosphate present at the ends of the vector.

Cloning blunt ends
Ligation of foreign DNA to a linearized plasmid, both carrying blunt ends, is comparatively an inefficient reaction. Only at high concentrations of vector, insert and enzyme, blunt ended molecules can be ligated and phosphodiester bonds can be made between these residues. The DNA insert will get ligated in either of the orientation. Usually PEG 8000 at the concentration of 5 % is used to increase macromolecular crowding.
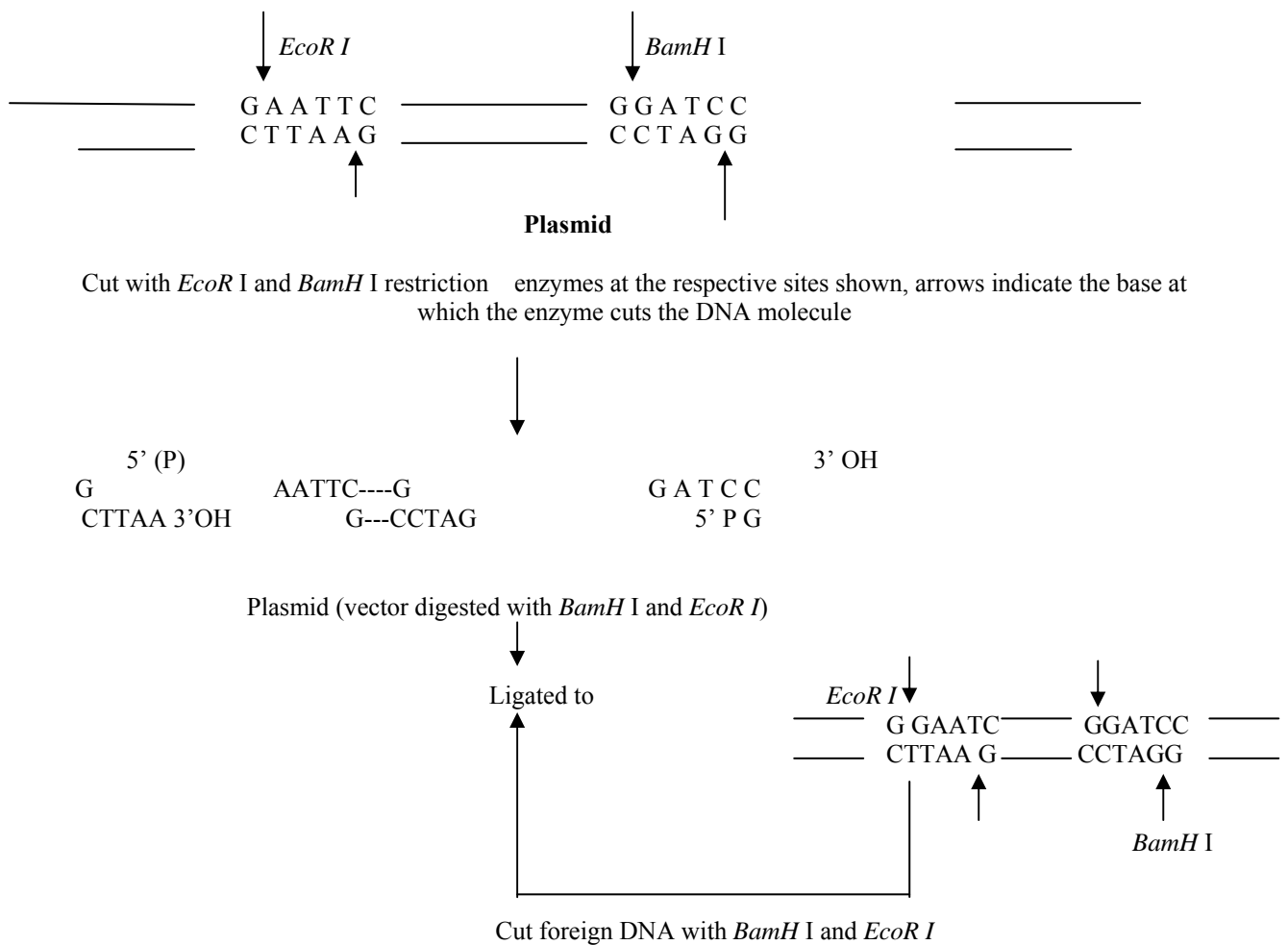
Cut with *EcoR* I and *BamH* I restriction enzymes at the respective sites shown, arrows indicate the base at which the enzyme cuts the DNA molecule

Figure2

Methods for Selection and/or Screening of Desired Recombinants

Selection refers to applying conditions that allow the desired cells or phages (containing vector or vector and insert) to replicate while preventing others from replicating. Typical selections include:

- antibiotic sensitivity and resistance;
- nutrient requirements;
- plaque formation.

Screening allows all cells to grow, but tests the resulting clones for particular properties. Often these are indicative of the presence of an insert in the vector. Screening properties used include:

- antibiotic sensitivity and resistance;
- nutrient requirements;
- plaque type;
- blue-white selection (-galactosidase activity);
- specific (by nucleic acid hybridization, PCR with gene specific primers or antibodies).

Screening of Clones

By PCR

This is the fastest way to screen bacterial colonies. A primer pair is chosen which has a sequence derived from vector (flanking primer) and a gene specific primer. This allows going to the replica agar plate on the same day and setting up miniprep cultures of the likely candidate colonies.

By Hybridization with radiolabeled probe

1. **Lysis of Colonies and Binding of DNA to Nitrocellulose Filters**
    The colonies are lifted from agar plates onto nitrocellulose membranes, lysed and neutralized and baked with the bacterial DNA fixed to the membranes.
2. **Labeling of DNA-Probe with ^{32}P**
    Using DNA polymerase, templates, and primers in combination with radiolabeled nucleotide (either $[a$-^{32}P]dATP or $[a$-^{32}P]dCTP), a radioactive probe is synthesized. The procedure works with either exact size templates or any (random or specific) primers or with templates within a vector and specific primers.
3. **Hybridization to Nitrocellulose Filters Containing Replicas of Bacterial Colonies**
    After the DNA has been fixed to the filters by baking, wash step removes bacterial debris off the filter surface. Filters then are prehybridized to block binding sites, and the radiolabeled probe added thereafter will compete for specific binding sites. After washing, the air died filters are exposed to an X-ray film which shows positive colonies.

Host vector relationship in cloning of the genes

Any DNA cloning procedure consists of:
- Generation of DNA fragment for cloning: the source can be genomic DNA, cDNA generated from mRNA or chemically synthesized gene fragment.
- Modification of the termini of DNA and vector to make them compatible with each other
- Ligation of the DNA fragment and vector
- Introduction of the recombinant DNA into a suitable host
- Selection and screening of the recombinant transformants

Cloning can be viewed from two different aspects:
Homologous gene cloning: cloning of a gene into the host from which the genetic information was derived
Heterologous gene cloning: cloning of a gene into a new host strain

Following are the hosts that are widely used for cloning of genes:
- Bacteria (E. coli)
- Yeast (S. cerevisiae)
- Insects (D. melanogaster)
- Plants (Arabadopsis thaliana)
- Mammals (Mus musculus)

A cloning host must:
- be easy to handle and propagate
- have a defined genotype
- accept a range of vectors

DNA cloning requires a bacterial host strain with high transformation efficiency, which does not cleave the recombinant DNA, provides good yields of high quality plasmid or phage DNA, and consists of components for various vector screening systems. An expression strain, on the other hand, must have the appropriate regulatory proteins for expression of cloned genes—a requirement that may not be compatible with attributes of cloning strains. To improve yields of expressed proteins from the cloned genes, expression strains are often deficient in some proteases. For the pET system, expression hosts are λDE3 lysogens, which have the inducible gene for T7 RNA polymerase.

E. coli is often used as the host cell for expression of foreign genes because the control of gene expression has been studied most extensively and the organism is most easy to grow and manipulate. The inclusion of regulatable promoters in expression vectors allows the expression of foreign gene products that are even toxic to the cells/host. The expression of the eukaryotic genes in *E. coli* requires that the coding region should be placed downstream from a strong *E. coli* promoter since eukaryotic promoters are not easily or weakly recognized by the *E. coli* RNA polymerase. Few strong promoters which are used generally for the expression of foreign genes in *E. coli* are bacteriophage lambda P_L, bacteriophage lambda P_R, lac, trp, tac (trp-lac hybrid). Few limitations of using *E. coli* for the expression of foreign genes are:
- No enzyme to process eukaryotic introns in prokaryotes and therefore only cDNAs and intronless genes can be expressed in bacterial cells
- No post translational modification like glycosylation and phosphorylation and any protein which requires these modifications for its biological activity will be virtually inactive.
- Bacterial cells lack proteolytic enzymes found in eukaryotic cells that are able to process the precursor peptides and release mature form of the proteins.

A number of features have been incorporated into *E. coli* DNA cloning strains to facilitate transformation of the initial ligation reactions, screening for recombinants, stability of the recombinant plasmid, and high quality plasmid DNA preparations.

Genetic features to look for are:

1. *lacZDM15* encodes an enzymatically inactive b-galactosidase lacking amino acids 11-41. Some cloning vectors encode the deleted b-galactosidase amino acids (*lacZa* or *lacZ'*), which, when expressed, can associate with the inactive protein in the host to generate active enzyme. The combination of host strain and vector allow blue/white screening for recombinant plasmids by a-complementation.
2. *recA* encodes a protein essential for homologous recombination in *E. coli*. Plasmids in host strains deficient in *recA* are less likely to incur deletions or other rearrangements. Plasmid DNA prepared from these strains contains monomer circles and lack the multimers common in preparations from $RecA^+$ strains.
3. *endA* encodes a thermostable DNase that can degrade plasmid DNA during purification. Host strains deficient for *endA* allow higher yields of plasmid DNA.
4. *hsdRMS* encodes the enzymes responsible for the *Eco* restriction-modification system. For transformation with unmodified recombinant plasmid DNA, strains must be restriction minus.
5. *mcrAB* encode enzymes that restrict foreign DNA.

Bacteriophage vectors

A **phage** (also called *bacteriophage*) is a small virus that infects only bacteria . Like viruses that infect eukaryotes , phages consist of an outer protein coat and the enclosed genetic material (which consists of double-stranded DNA in 95% of the phages known and few have RNA as the genetic material) of 5-650 kbp (kilo base pairs) with a length of 24-200 nm . The vast majority of phages (95%) have a tail which has the ability to bind to specific molecules on the surface of their target bacteria and thus inject their genetic material into the host. Phages were discovered independently by Frederick Twort in 1915 and by Felix d' Herelle in 1917. Phages infect only specific bacteria. Some phages are virulent, that is upon infecting a bacterial cell, they immediately start reproducing, and within a short time, lyse (destroy) the bacterial cell, releasing new phage particles. These young phages burst from the host cell (killing it) and infect more bacteria. These types of phages are called as Lytic phages. Some phages (so-called *temperate phages*) can instead enter a relatively harmless state, either integrating their genetic material into the chromosomal DNA of the host bacterium (much like endogenous retrovirus in animals) or establishing themselves as plasmids. These endogenous phages, referred to as *prophages*, are then copied with every cell division together with the DNA of the host cell. They do not kill the cell, but monitor (*via* some proteins they code for) the status of their host. When the host cell shows signs of stress, the endogenous phages become active again and start their reproductive cycle, resulting in the lysis of the host cell. An example is phage λ of *E. coli* . Sometimes, prophages even provide benefit to the host bacterium while they are dormant, by adding new functions to the bacterial genome , a phenomenon called **Lysogenic Conversion** . A famous example is the harmless *Vibrio* bacteria strain, which is turned into *Vibrio cholerae* by a phage, causing cholera. Phages play an important role in molecular biology as cloning vectors to insert DNA into bacteria.

Lambda-Replacement Vector

Bacteriophage λ with its large and complex genome, for use as a vector, would seem to be impossible. Wild type DNA has multiple sites for most useful restriction enzymes but they are located in the genomic DNA essential for lytic growth of virus. Initially it was thought, that since bacteriophage λ particles will not accommodate DNA larger than the viral genome, therefore suitable only for small foreign fragments. Methods have developed large number of vectors that can accept and propagate fragments of foreign DNA generated by a variety of different Restriction enzymes. The λ genome remains linear inside the phage head but circularizes as soon as bacteriophage DNA goes inside the host *E.coli*. The single stranded protruding ends of 12 bases, called as cohesive site, (Cos site), anneal to form a closed circular molecule necessary for replication. The sealed cohesive ends are the sites of cleavage during packaging of mature DNA into phage heads. The λ DNA must be greater than 38kb and smaller than 52kb to be packaged into phage head. The genes for lysogeny are located in this segment which is removed to make high capacity vectors which can accommodate larger size of foreign DNA.

Vectors that have a pair of cloning sites flanking a segment of non-essential DNA are known as replacement vectors. Removal of the central non essential stuffer fragment reduces the size of vector to a size that cannot be packaged into bacteriophage particles. Particles can be packed only by ligating a segment of foreign DNA. The range of DNA that can be packaged lies between 8kb–24 kb. If one wants to clone longer DNA fragments, deletion of essential bacteriophage genes, is required, which then have to be supplied in trans. Cloning in bacteriophage vectors is advantageous since for successful packaging, there should be enough

DNA. No infectious particles are obtained that lack an insert or contain an insert that is below the lower size limit for a particular vector. Inserts with DNA too large are also not successfully cloned. Therefore packaging into phage heads of the vector DNA after DNA insertion offers an efficient selection strategy for recombinant vectors. Thus, the size of the vector and insert DNA should be carefully adjusted for an effective selection. Few examples of vectors that employ this strategy are bacteriophage λ vectors EMBL3 and EMBL4.
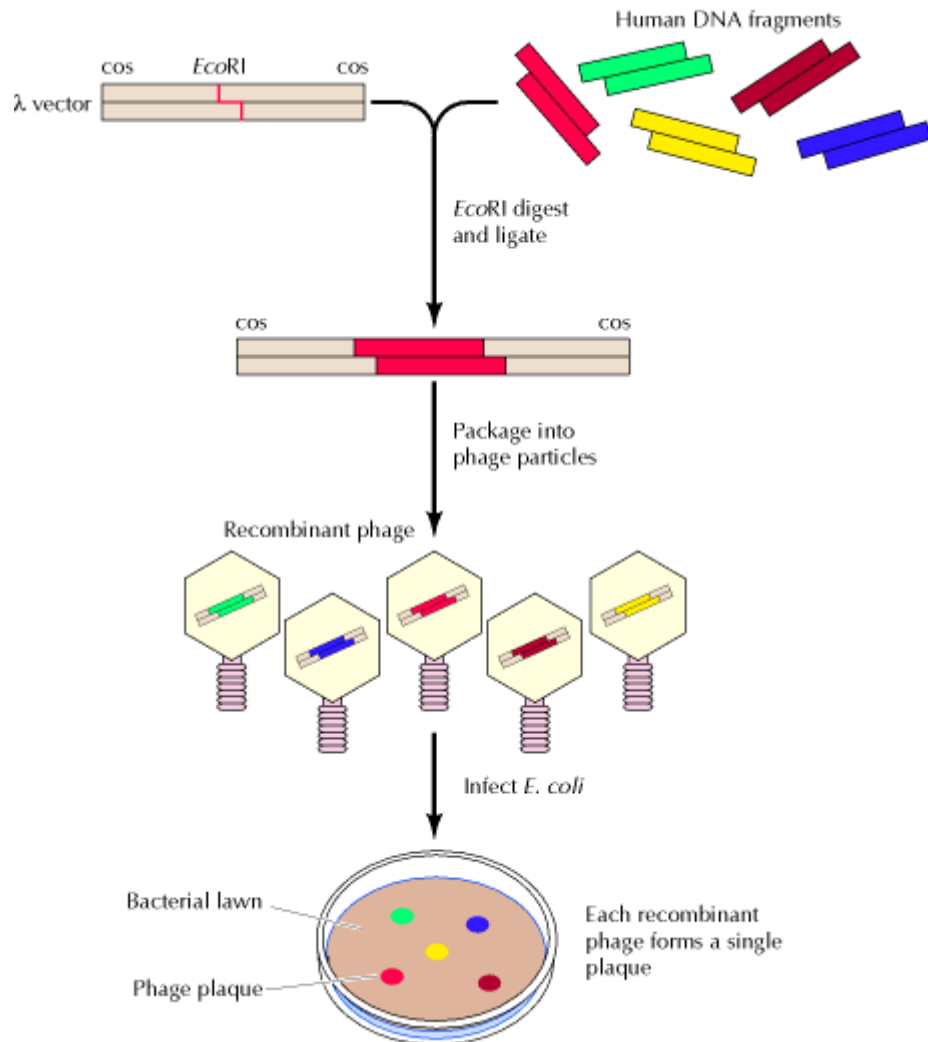
Cloning in bacteriophage vectors

The vector contains a restriction site (e.g.,an *Eco*RI site) for insertion of cloned DNA. In addition, cos sites which are required for packaging DNA into phage particles, are present on both ends of the vector DNA. Insert DNA is ligated to the vector, and the recombinant molecules are packaged into phage particles when mixed with phage proteins. The recombinant phages are then used to infect *E. coli*. Each recombinant phage, which carries a unique insert of cloned DNA, forms a single plaque in the infected bacterial culture. Progeny phage carrying unique DNA inserts can then be isolated from individual plaques and grown in large quantity.

This type of vector has a loading capacity of 10-23 kb and consists of a full length lambda-molecule with two identical restriction sites flanking a non-essential region called stuffer fragment. The idea behind it is that the stuffer fragment is deleted and replaced by foreign DNA. Again a selection system is required to differentiate between wild type and recombinant phage.

This is done by placing relevant genes onto the stuffer fragment the loss of which gives rise to a detectable phenotypic signal. Two examples are given:

1. *LacZ* inactivation: This is done by inserting the *lacZ* gene onto the stuffer fragment. Loss of the fragment and therefore *lacZ* can be monitored by the blue/white colour discrimination.

2. Spi-phenotype: In this case the stuffer fragment carries the phage genes red and gam necessary for lambda-DNA replication. Whereas the wild type lambda is not capable of infecting a *E.coli* P2 lysogen due to the presence of red and gam (Spi phenotype in short for sensitive towards P2 prophage inhibition), loss of red and gam in all recombinant phage particles results in plaque formation on a *E.coli* P2 lysogen; hence all plaques should contain recombinant phage molecules! The most popular lambda-replacement vector for library construction is EMBL3A with *Sal* I, *BamH* I and *EcoR* I sites flanking the stuffer fragment.

Insertion Vectors

Vectors that have a single target site for insertion of foreign DNA are known as insertion vectors. 20% DNA that is not required for lytic growth is removed and therefore insertion of foreign DNA resumes the size back to something like its full length and can be packaged *in vitro*. Maximum size of DNA that can be accommodated varies from 9- 11 kb.

For DNA of larger sizes, high capacity vectors are designed like:-

Vector	Capacity	Origin of replication	Host
Cosmid	30-45 kb	Col E1 Replicon	*E. coli*
BAC	120-300	Foreign	*E. coli*
YAC	250-400	ARS	Yeast

Cosmids

Smallest capacity for foreign DNA, contain one or two copies of small region of bacteriophage λ that is the cohesive end site (Cos), containing all the cis-acting elements required for packaging of viral DNA into virions. The vector also contains sequences required for terminase so that it can bind and cleave and under appropriate conditions the DNA is packaged into empty λ phage particles. It should have unique restriction sites, a replication origin and selectable markers. In vectors such as supercos-1, two cos sites are there within *Xba* I site. The cosmid vectors are opened up by appropriate restriction enzymes and then mixed with the DNA inserts prepared by using same enzymes under annealing conditions. Long concatemers so generated are products for packaging into the λ heads. This selects the DNA to be inserted since for packaging in λ particles, the distance between two cos sites must be between 38 to 52 kb. Cosmids can accommodate up to 45 kbp long DNA inserts. Cosmids infect the host cells like λ particles but once inside the host, it will replicate and propagate as a plasmid. They have been developed in the late 1970s and have been improved significantly since. They are predominantly plasmids with a bacterial origin, an antibiotic selection marker and a cloning site, but they carry one, or more recently two cos sites derived from bacteriophage lambda. 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 wild type Cosmid DNA is simply done via size exclusion. The strategy used for cloning in Cosmid vectors is shown below:
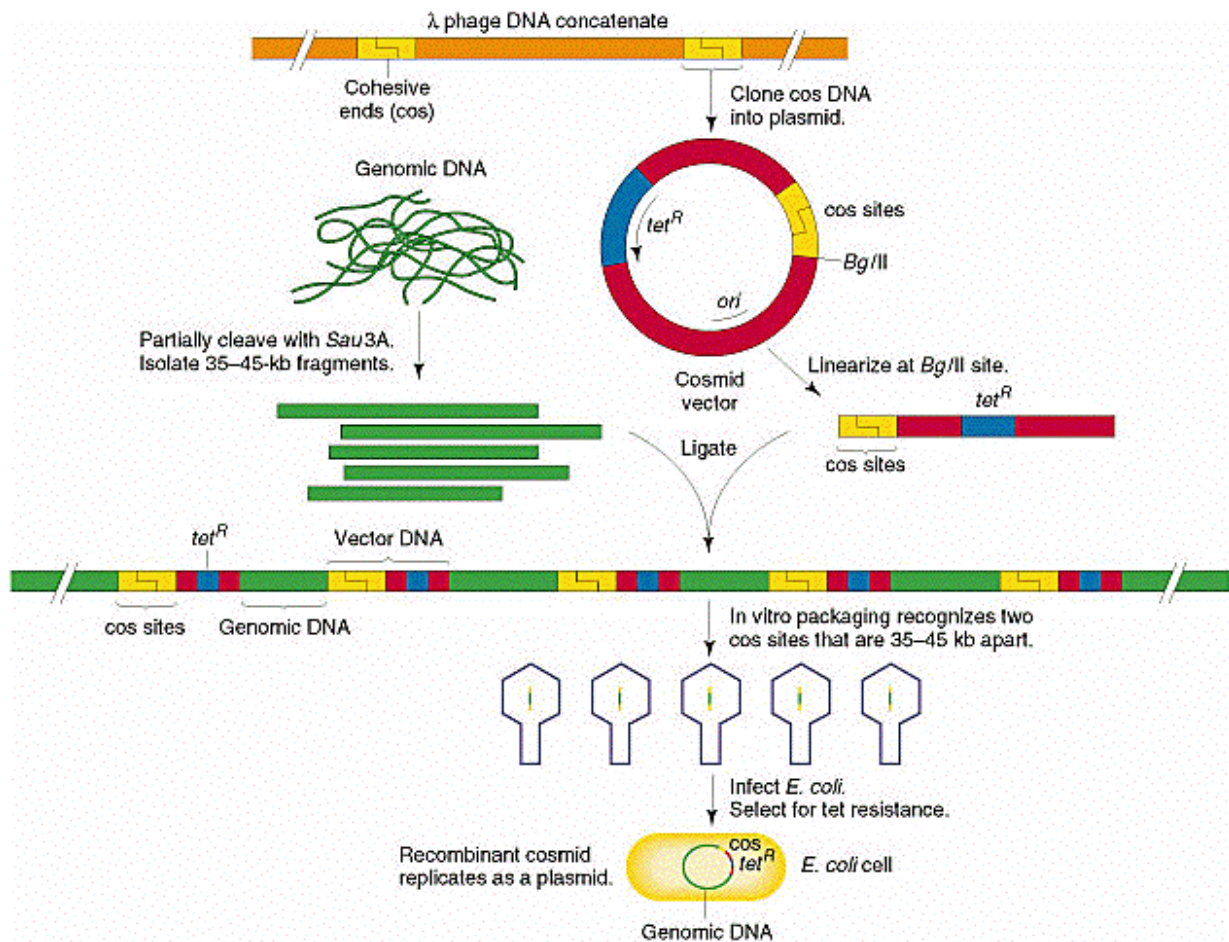
**Cloning using Cosmid vectors**

The cosmid is cut at a *Bgl* II site next to the cos site. Donor genomic DNA is cut using *Sau3*A, which gives sticky ends compatible with *Bgl* II. A tandem array of donor and vector DNA results from mixing. Phage is packaged *in vitro* by cutting at the cos site. The cosmid with inserts recircularizes once it is inside the bacterial cell.

Bacteriophage Artificial Chromosomes (BACs)

These plasmids are circular DNA molecules carrying conventional antibiotic resistance marker, origin of replication derived from the F factor of *E.coli*, an ATP driven helicase (repE) to facilitate DNA replication and three loci (parA, parB and parC) for proper partitioning of the plasmid to daughter cells. BAC vectors have no packing constraints and there is no fixed limit to the size of genomic DNA that they accept. Usually the size of DNA is approximately 120-kilo base pairs.

Yeast Artificial Chromosome (YACs)

These are linear DNA molecules similar to yeast chromosome. Recombinant YACs are made by ligating large fragments of genomic DNA and then the resultant plasmid is introduced into yeast by transformation. The vector carries selection marker, DNA sequences called as telomere, so that the product can be stabilized inside the yeast cell, an origin of replication called autonomous replication origin, ARS. Large size of DNA can be inserted into YAC vectors, usually between 250kilobases to 400kilobasepairs. Large size of mammalian genomic libraries is also made with approximately 1 megabasepairs of foreign inserts. Insertion of foreign DNA into the cloning site inactivates a mutant expressed in vector DNA and formation of red rather than white colonies by yeast strain is observed. Thus transformants are identified as red colonies, which grow in yeast that is mutant for TRP1 and URA3, which ensure that the cell has received an artificial chromosome and with both the telomeres since it is complimented for both the mutations. And the colony also contains foreign DNA because it is red in color.
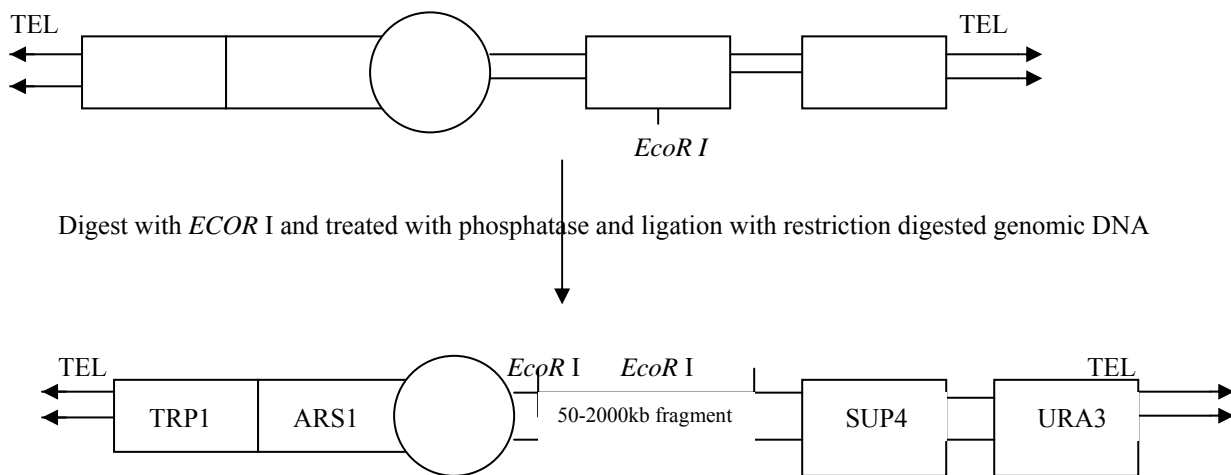
Shuttle vectors

Cloning of foreign DNA is usually carried out primarily in E.*coli* since the organism is most thoroughly studied. But subsequent work often requires the foreign segment to be delivered to different host cells like eukaryotes. A number of vectors are devised to satisfy this requirement. These vectors are termed as shuttle vectors. These vectors have origins of replication of various hosts. The also contain fragments of eukaryotic viruses to facilitate entry into the cell or expression or integration in the cell itself. Thus shuttle vectors allow DNA to be transferred between two different species where it can be propagated by utilizing both the origins of replication. Usually the origins of replication are derived from bacterial and eukaryotic systems. Shuttle vectors also carry antibiotic resistance genes, which are functional in eukaryotes e.g. Neomycin (G418), Hygromycin, Methotrexate etc. All the DNA manipulation and characterization are done in prokaryotic system and then the manipulated DNA is introduced into the eukaryotic systems for protein expression and functional analysis. Eukaryotic host systems are better for expression of protein for few reasons:

1. Proper folding of the protein to attain functional activity

2. Posttranslational modification of proteins for which prokaryotes does not possess any machinery. The most conventional and convenient model system for expression of eukaryotic proteins is yeast, *Pichia pastoris*, which is both genetically and physiologically well characterized.

Plant Vectors

A vector is a circular DNA molecule capable of independent existence and replication within a host cell. In case of plants, Ti and Ri are the two most commonly used plasmids which are used as vectors. Plant cells as such do not possess any endogenous plasmids. But two plasmids called pTi and pRi, are present naturally in the bacteria, *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*, respectively. These plasmids provide a naturally occurring transformation system. A part of the plasmid DNA, called as T-DNA, is transferred into the genomes of most dicot and some monocot plants. pTi stands for tumor inducing plasmid and pRi stands for root inducing plasmid. The infection of Agrobacterium tumefaciens is mediated by transfer of a segment of pTi called as T-DNA into the plant cell. Various bacterial chromosomal genes, such as *chv*B, *exo* genes, cell genes, are concerned with the biosynthesis of cell attachment polysaccharides due to which the bacterial cells adhere firmly to the plant cells. While two chromosomal genes are expressed constitutively in bacterial cells that is expressed at all the times inside a cell, which are responsible for virulence associated aspects, *Agrobacterium tumefaciens* Ti plasmid produces tumor like growth from which roots / shots may sometimes be produced. The infected cells are able to grow in culture on a medium devoid of any growth regulator while uninfected normal plant cells need exogenous auxin or cytokinin. These plasmids also carry genes for IAA (Indole Acetic Acid - auxin) and cytokinin production which is the reason for indefinite growth on a growth regulator free culture medium. When pTi is introduced into *Rhizobium trifolii*, it gains the ability to produce galls and to utilize opines. The crown gall root cells also synthesize unique nitrogenous compounds called opines, which are not produced by normal plant cells, which are not infected, nor are they utilized. The infected cells use opines as their carbon and nitrogen source. The type of opine produced depends on the bacterial strain. *Agrobacterium tunefaciens* strain produces either octopine or nopaline which the *Agrobacterium rhizogenes* produce either agropine or mannopine.

Ti plasmid

Ti plasmid is a large mega plasmid conjugative plasmid of ~200kb. pTi is lost when *Agrobacterium* is grown above 28°C, such cured bacteria do not induce crown galls that is, they become avirulent. pTi and pRi, although do not share sequence homology but are unique in following respects:-

a) They contain some genes, which are located within their T-DNA which has regulatory sequences recognized by plant cells, while their remaining genes have prokaryotic regulatory sequences. As a result, the former are expressed only in plant cells but not in the *Agrobacterium*, while the latter are only expressed in the bacterium.

b) These plasmids naturally transfer a part of their DNA, called as T-DNA, into host plant cells. The T-DNA usually contains following important functional regions.
   1. T-DNA contains oncogenes and opine synthesis genes and is transferred into host plant.
   2. Vir region which regulates the transfer of T-DNA
   3. Opine catabolism genes for utilization of opines.

4. Origin of replication for propagation in *Agrobacterium.*

The T-DNA contains a 24bp direct repeat border sequence and contains the genes necessary for tumor / possess gene for auxin and cytokinin biosynthesis. All the genes present in T-DNA have eukaryotic regulatory sequences. As a result, these genes are expressed only in plant cells but never express in *Agrobacterium.* The vir region mediates the transfer of T-DNA into plant genomes and hence is essential for virulence. The genes of vir region are not transferred but induce the transfer of T-DNA. Also, the genes present in T-DNA are not responsible for its transfer, but the 24 bp direct repeat at both the left and right ends of T-DNA is essential for the transfer. The exact mechanism of transfer of T-DNA is not known clearly known but is brought by the vir region. The phenols produced by wounded plant tissue initiates the transfer process. The T-DNA is transferred into the plant cells as single stranded DNA, which increases the efficiency of its transformation. But, as soon as it enters into the plant cell, it is immediately converted into a double stranded form. This form integrates at random sites in the host plant genome by a phenomenon called illegitimate recombination, which are due to sequence of homology in short segments of the host DNA. This integration is usually in low copy numbers. Few vectors are derived from pTi (wild type) due to some problems posed by wild type plasmid eg. The presence of oncogenes causes a disorganized growth, their large size and lack of cloning sites within the T-DNA, which are needed for the insertion of DNA segments that has to be cloned.

Disarming

Disarming is the removal of oncogenes from T-DNA and it has resolved many problems. This disarmed plasmid can still transfer its T-DNA into plant cells. But the cells containing this modified T-DNA will be non tumorous, produce opines and generate plantlets. Only the border sequences are necessary for the transfer of any DNA insert placed between them. Thus pTi and pRi which are disarmed are more in use for gene transfer experiments. But in some plants, the efficiency of transformation by disarmed pTi is much lower than the wild type pTi.

Introduction of DNA into plant cells without the involvement of a biological agent like *Agrobacterium* leading to a stable transformation is known as **Direct Gene Transfer**. The spontaneous uptake of DNA is quite low. The various methods those are utilized for direct gene transfer are:
1. Chemical methods like PEG, Calcium phosphate
2. Electroporation
3. Particle gun delivery
4. Lipofection
5. Microinjection and Macroinjection