SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – 1 – GENETIC ENGINEERING – SBTA1501

History,milestones and recent developments in Genetic Engineering G

Genes and Genetic Engineering

A **gene** is a basic unit of heredity in a living organism. It is "a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions ".

Central Dogma of Molecular Biology: The flow of genetic information in the cell starts at DNA, which replicates to form more DNA. Information is then 'transcribed" into RNA, and then it is "translated" into protein.

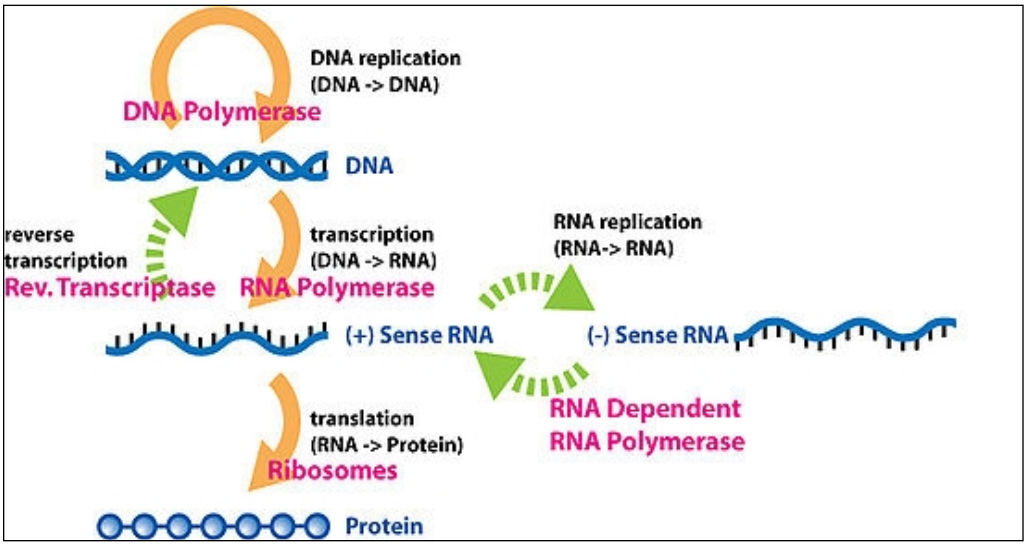


Fig : 1.Central dogma of Protein synthesis

Genetic Engineering: The technology entailing all processes of altering the genetic material of a cell to make it capable of performing the desired functions, such as producing novel substances. In other words, Genetic engineering is the deliberate controlled manipulation of genes in an organism in order to upgrade that organism.

Although there are many diverse and complex techniques involved, the basic principles of genetic manipulation are reasonably simple. The premise on which the technology is based is that genetic

information, encoded by DNA and arranged in the form of genes, is a resource that can be manipulated in various ways to achieve certain goals in both pure and applied science and medicine.

There are many areas in which genetic manipulation is of value, including the following:

- Basic research on gene structure and function
- Production of useful proteins by novel methods
- Generation of transgenic plants and animals
- Medical diagnosis and treatment
- Genome analysis by DNA sequencing

A Brief History of Genetic Engineering

Some major steps in the development of Genetic Engineering

1869	**Miescher** first isolates DNA from white blood cells harvested from pus-soaked bandages obtained from a nearby hospital.
1944	**Avery** provides evidence that DNA, rather than protein, carries the genetic information during bacterial transformation.
1953	**Watson and Crick** propose the double-helix model for DNA structure based on x-ray results of **Franklin and Wilkins**.
1955	**Kornberg** discovers DNA polymerase, the enzyme now used to produce labeled DNA probes.
1961	**Marmur and Doty** discover DNA renaturation, establishing the specificity and feasibility of nucleic acid hydridization reactions.
1962	**Arber** provides the first evidence for the existence of DNA restriction nucleases, leading to their purification and use in DNA sequence characterization by **Nathans and H. Smith**.
1966	**Nirenberg**, **Ochoa**, and **Khorana** elucidate the genetic code.
1967	**Gellert** discovers DNA ligase, the enzyme used to join DNA fragments together.
1972–1973	DNA cloning techniques are developed by the laboratories of **Boyer**, **Cohen**, **Berg**, and their colleagues at Stanford University and the University of California at San Francisco.
1975	**Southern** develops gel-transfer hybridization for the detection of specific DNA sequences.
1975–1977	**Sanger and Barrell** and **Maxam and Gilbert** develop rapid DNA-sequencing methods.

1981–1982	**Palmiter and Brinster** produce transgenic mice; **Spradling and Rubin** produce transgenic fruit flies.
1982	**GenBank**, NIH's public genetic sequence database, is established at Los Alamos National Laboratory.
1985	**Mullis** and co-workers invent the polymerase chain reaction (PCR).
1987	**Capecchi** and **Smithies** introduce methods for performing targeted gene replacement in mouse embryonic stem cells.
1989	**Fields and Song** develop the yeast two-hybrid system for identifying and studying protein interactions
1989	**Olson** and colleagues describe sequence-tagged sites, unique stretches of DNA that are used to make physical maps of human chromosomes.
1990	**Lipman** and colleagues release BLAST, an algorithm used to search for homology between DNA and protein sequences.
1990	**Simon** and colleagues study how to efficiently use bacterial artificial chromosomes, BACs, to carry large pieces of cloned human DNA for sequencing.
1991	**Hood and Hunkapillar** introduce new automated DNA sequence technology.
1995	**Venter** and colleagues sequence the first complete genome, that of the bacterium *Haemophilus influenzae.*
1996	**Goffeau** and an international consortium of researchers announce the completion of the first genome sequence of a eucaryote, the yeast *Saccharomyces cerevisiae.*
1996–1997	**Lockhart** and colleagues and **Brown and DeRisi** produce DNA microarrays, which allow the simultaneous monitoring of thousands of genes.
1998	**Sulston and Waterston** and colleagues produce the first complete sequence of a multicellular organism, the nematode worm *Caenorhabditis elegans.*
2001	Consortia of researchers announce the completion of the draft human genome sequence.

Steps involved in rDNA Technology

Isolation and Purification of Genomic DNA

Every gene manipulation procedure requires genetic material like DNA and RNA. Nucleic acids occur naturally in association with proteins and lipoprotein organelles. The dissociation of a nucleoprotein into nucleic acid and protein moieties and their subsequent separation, are the essential steps in the isolation of all species of nucleic acids. Isolation of nucleic acids is followed by quantization of nucleic acids generally done by either spectrophotometric or by using fluorescent dyes to determine the average concentrations and purity of DNA or RNA present in a mixture. Isolating the genetic material (DNA) from cells (bacterial, viral, plant or animal) involves three basic steps-

- Rupturing of cell membrane to release the cellular components and DNA
- Separation of the nucleic acids from other cellular components
- Purification of nucleic acids

Genomic DNA is found in the nucleus of all living cells with the structure of double-stranded DNA remaining unchanged (helical ribbon). The isolation of genomic DNA differs in animals and plant cells. DNA isolation from plant cells is difficult due to the presence of cell wall, as compared to animal cells. The amount and purity of extracted DNA depends on the nature of the cell.

The method of isolation of genomic DNA from a bacterium comprises following steps

1. Bacterial culture growth and harvest.
2. Cell wall rupture and cell extract preparation.
3. DNA Purification from the cell extract.
4. Concentration of DNA solution.

Growth and harvest of bacterial culture

Bacterial cell culture is more convenient than any other microbe, as it requires only liquid medium (broth) containing essential nutrients at optimal concentrations, for the growth and division of bacterial cells. The bacterial cells are usually grown on a complex medium like Luria-Bertani (LB), in which the medium composition is difficult to decipher. Later, the cells are separated by centrifugation and resuspended in 1% or less of the initial culture volume.

Preparation of cell extract

Bacterial cell is surrounded by an additional layer called cell wall, apart from plasma membrane with some species of *E. coli* comprising multilayered cell wall. The lysis of cell wall to release the genetic material i.e. DNA can be achieved by following ways-

- Physical method by mechanical forces.
- Chemical method by metal chelating agents i.e. EDTA and surfactant i.e. SDS or enzyme (e.g. lysozyme).

Lysozyme

- present in egg-white, salivary secretion and tears.
- catalyzes the breakdown of cell wall i.e. the peptidoglycan layer.

EDTA (Ethylene diamine tetra-acetic acid)

- a chelating agent necessary for destabilizing the integrity of cell wall.
- Inhibits the cellular enzymes that degrade DNA.

SDS (Sodium dodecyl sulphate)

- helps in removal of lipid molecules and denaturation of membrane proteins.

Generally, a mixture of EDTA and lysozyme is used. Cell lysis is followed by centrifugation to pellet down the cell wall fractions leaving a clear supernatant containing cell extract.

Purification of DNA

In addition to DNA, a cell extract contains significant quantities of protein and RNA which can be further purified by following methods-

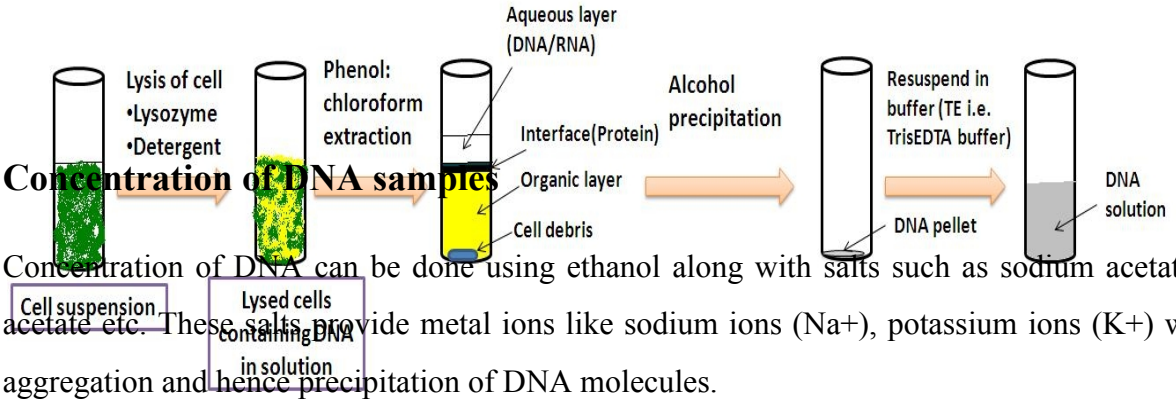
Organic extraction and enzymatic digestion for the removal of contaminants

It involves the addition of a mixture of phenol and chloroform (1:1) to the cell lysate for protein separation. The proteins aggregate as a white mass in between the aqueous phase containing DNA and RNA, and the organic layer. Treatment of lysate with pronase or protease, in addition to phenol/chloroform, ensures complete removal of proteins from the extract. The RNA can be effectively removed by using Ribonuclease, an enzyme which rapidly degrades RNA into its ribonucleotide subunits. Repeated phenol extraction is not desirable, as it damages the DNA.

Using ion-exchange chromatography

This involves the separation of ions and polar molecules (proteins, small nucleotides and amino acids) based on their charge. DNA carrying negative charge binds to the cationic resin or matrix which can be eluted from the column by salt gradient. Gradual increase in salt concentration detaches molecules from the resin one after another.

Fig:2 DNA Isolation method



Concentration of DNA samples

Concentration of DNA can be done using ethanol along with salts such as sodium acetate, potassium acetate etc. These salts provide metal ions like sodium ions (Na+), potassium ions (K+) which help in aggregation and hence precipitation of DNA molecules.

Advantage

It leaves short-chain and monomeric nucleic acid components in solution. Ribonucleotides produced by the ribonuclease treatment are separated from DNA.

Isolation and Purification of Plasmid DNA

Plasmids are circular, double stranded extra cellular DNA molecules of bacterium and most commonly used in recombinant DNA technology. The isolation of plasmid DNA involves three major steps-

1. Growth of the bacterial cell.
2. Harvesting and lysis of the bacteria.
3. Purification of the plasmid DNA.

Growth of the bacterial cell

It involves growth of the bacterial cells in a media containing essential nutrients.

Harvest and lysis of bacteria

Lysis of bacteria results in the precipitation of DNA and cellular proteins. Addition of acetate-containing neutralization buffer results in the precipitation of large and less supercoiled chromosomal

DNA and proteins leaving the small bacterial DNA plasmids in solution.

Purification of Plasmid DNA

This step is same for both plasmid and genomic but former involves an additional step i.e. the separation of plasmid DNA from the large bacterial chromosomal DNA.

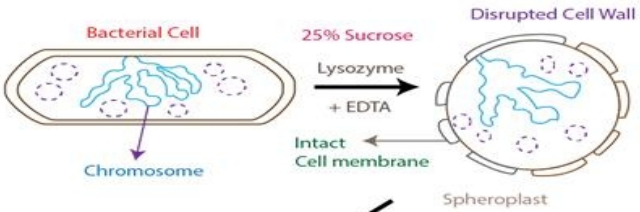
Methods for separation of plasmid DNA

Separation of plasmid DNA is based on the several features like size and conformation of plasmid DNA and bacterial DNA. Plasmids are much smaller than the bacterial main chromosomes, the largest plasmids being only 8% of the size of the *E. coli* chromosome. The separation of small molecules (i.e. plasmids) from larger ones (i.e. bacterial chromosome) is based on the fact that plasmids and the bacterial chromosomes are circular but bacterial chromosomes break into linear fragments during the preparation of the cell extract resulting in separation of pure plasmids. The methods of separation of plasmid DNA are described as below-

Separation based on size difference

- It involves lysis of cells with lysozyme and EDTA in the presence of sucrose (prevents the immediate bursting of cell).
- Cells with partially degraded cell walls are formed that retain an intact cytoplasmic membrane called as sphaeroplasts.
- Cell lysis is then induced by the addition of a non-ionic detergent (e.g. Triton X-100) or ionic detergents (e.g. SDS) causing chromosomal breakage.
- Bacterial chromosome attached to cell membrane, upon lysis gets removed with the cell debris.
- A cleared lysate consisting almost entirely of plasmid DNA is formed with very little breakage of the bacterial DNA.

Fig 3:Separation of Plasmid DNA



Separation based on conformation

Plasmids are supercoiled molecules formed by partial unwinding of double helix of the plasmid DNA during the plasmid replication process by enzymes called topoisomerases. The supercoiled conformation can be maintained when both polynucleotide strands are intact, hence called covalently closed-circular (ccc) DNA. If one of the polynucleotide strands is broken, the double helix reverts to its normal relaxed state taking an alternative conformation, called open-circular (oc). Super coiling is important in plasmid preparation due to the easy separation of supercoiled molecules from non-supercoiled ones.

The commonly used methods of separation based on conformation are as follows-

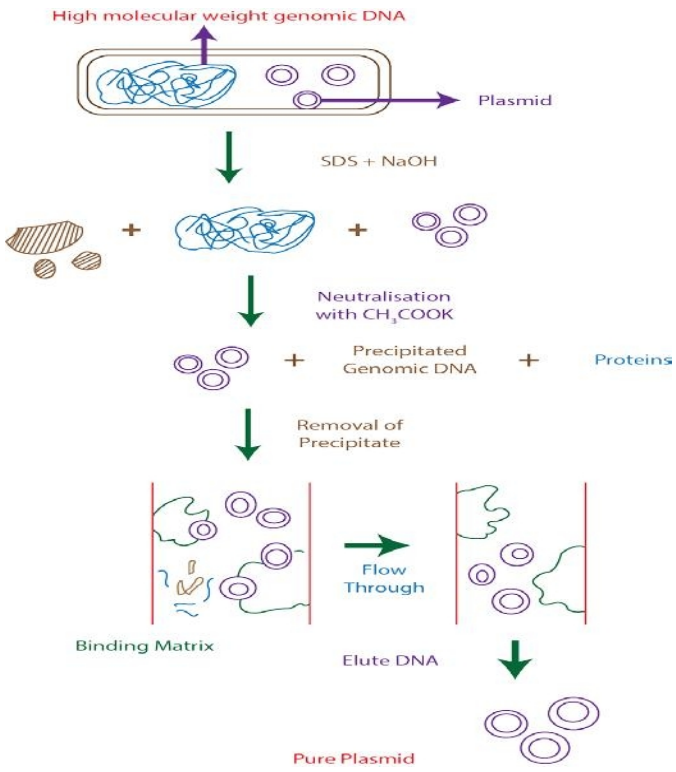
Alkaline denaturation method

- This method is based on maintaining a very narrow pH range for the denaturation of non-supercoiled DNA but not the supercoiled plasmid.
- Addition of sodium hydroxide to cell extracts or cleared lysate (pH12.0-12.5) results in disruption of the hydrogen bonds of non-supercoiled DNA molecules.
- As a result, the double helix unwinds and two polynucleotide chains separate.
- Further addition of acid causes the aggregation of these denatured bacterial DNA strands into a tangled mass which can be pelleted by centrifugation, leaving plasmid DNA in the supernatant.

Advantage

- Most of the RNA and protein under defined conditions (specifically cell lysis by SDS and neutralization with sodium acetate) can be removed by the centrifugation.
- No requirement of organic extraction.

Fig4 : Isolation of plasmid DNA by alkali lysis method



Ethidium bromide-cesium chloride density gradient centrifugation

- Density gradient centrifugation can separate DNA, RNA and protein. It is a very efficient method for obtaining pure plasmid DNA.
- A density gradient is produced by centrifuging a solution of cesium chloride at a very high speed which pulls the CsCl ions towards the bottom. This process is referred as isopycnic centrifugation.
- The DNA migrates to the point at which it has density similar to that of CsCl i.e.1.7 g/cm^3 in the gradient.
- In contrast, protein molecules having lower buoyant densities float at the top of the tube whereas RNA gets pelleted at the bottom.

Density gradient centrifugation in the presence of ethidium bromide (EtBr) can be used to separate supercoiled DNA from non-super coiled molecules. Ethidium bromide is an intercalating dye that binds to DNA molecules causing partial unwinding of the double helix. Supercoiled DNA have very little freedom to unwind due to absence of free ends and bind to a limited amount of EtBr resulting in very less decrease in buoyant density (0.085 g/cm^3) than that of linear DNA (0.125 g/cm^3). As a result, they form a distinct band separated from the linear bacterial DNA. The EtBr bound to DNA is then extracted by n-butanol and the CsCl is removed by dialysis.

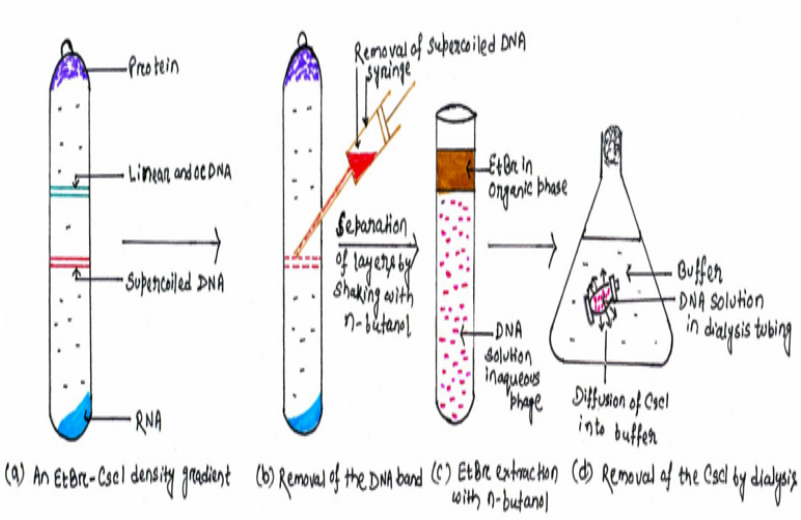


Fig 5; Density Gradient method .

Bacteriophage DNA preparation

In bacteriophage DNA preparation, a cell extract is not the starting material, because bacteriophage particles can be obtained in large numbers from the extracellular medium of an infected bacterial culture. When such a culture is centrifuged, the bacteria are pelleted, leaving the phage particles in suspension. The phage particles are then collected from the suspension and their DNA extracted by a single deproteinization step to remove the phage capsid.

Isolation and Purification of RNA

RNA (Ribonucleic acid) is a polymeric substance consisting of a long single-stranded chain of phosphate and ribose units with the nitrogen bases adenine, guanine, cytosine and uracil bonded to the ribose sugar present in living cells and many viruses. The steps for preparation of RNA involve homogenization, phase separation, RNA precipitation, washing and re-dissolving RNA.

The method for isolation and purification of RNA are as follows-

1) Organic extraction method
2) Filter-based, spin basket formats
3) Magnetic particle methods
4) Direct lysis method.

Organic extraction method

This method involves phase separation by addition and centrifugation of a mixture of a solution containing phenol, chloroform and a chaotropic agent (guanidinium thiocyanate) and aqueous sample. Guanidium thiocyanate results in the denaturation of proteins and RNases, separating rRNA from ribosomes. Addition of chloroform forms a colorless upper aqueous phase containing RNA, an interphase containing DNA and a lower phenol-chloroform phase containing protein. RNA is collected from the upper aqueous phase by alcohol (2-propanol or ethanol) precipitation followed by rehydration.

One of the advantages of this method is the stabilization of RNA and rapid denaturation of nucleases. Besides advantages, it has several drawbacks such as it is difficult to automate, needs labor and manual

intensive processing, and use of chlorinated organic reagents

Direct lysis methods

This method involves use of lysis buffer under specified conditions for the disruption of sample and stabilization of nucleic acids. If desired, samples can also be purified from stabilized lysates. This method eliminates the need of binding and elution from solid surfaces and thus avoids bias and recovery efficiency effects.

Advantages

- Extremely fast and easy.
- Highest ability for precise RNA representation.
- Easy to work on very small samples.
- Amenable to simple automation.

Drawbacks

- Unable to perform traditional analytical methods (e.g. spectrophotometric method).
- Dilution-based (most useful with concentrated samples).
- Potential for suboptimal performance unless developed/optimized with downstream analysis.
- Potential for residual RNase activity if lysates are not handled properly.

Electrophoresis

Electrophoresis is a method that separates macromolecules either nucleic acids or proteins-on the basis of size, electric charge, and other physical properties. Electro refers to the energyof electricity. Phoresis, from the Greek verb phoros, means “to carry across.” Thus, gel electrophoresis refers to the technique in which molecules are forced across a span of gel, motivated by an electrical current. Activated electrodes at either end of the gel

provide the driving force. A molecule’s properties determine how rapidly an electric field can move the molecule through a gelatinous medium. Many important biological molecules such as amino acids,peptides, proteins, nucleotides, and nucleic acids, posses ionisable groups and, therefore, at any given pH, exist in solution as electrically charged species either as cations (+) or

anions (-). Depending on the nature of the net charge, the charged particles will migrate either to the cathode or to the anode.

Gel electrophoresis is a technique used for the separation of nucleic acids and proteins. Separation of large (macro) molecules depends upon two forces: charge and mass. When a biological sample, such as proteins or DNA, is mixed in loading dye and applied to a gel, these two forces act together. The electrical current from one electrode repels the molecules while the other electrode simultaneously attracts the molecules. The frictional force of the gel material acts as a "molecular sieve," separating the molecules by size. During electrophoresis, macromolecules are forced to move through the pores when the electrical current is applied. Their rate of migration through the electric field depends on the strength of the field, size and shape of the molecules, relative hydrophobicity of the samples, and on the ionic strength and temperature of the buffer in which the molecules are moving. After staining, the separated macromolecules each lane can be seen in a series of bands spread from one end of the gel to the other.

Gel Electrophoresis is one of the staple tools in molecular biology and is of critical value in many aspects of genetic manipulation and study. One use is the identification of particular DNA molecules by the band patterns they yield in gel electrophoresis after being cut with various restriction enzymes. Viral DNA, plasmid DNA, and particular segments of chromosomal DNA can all be identified in this way. Anothermuse is the isolation and purification of individual fragments containing interesting genes, which can be recovered from the gel with full biological activity. Gel electrophoresis makes it possible to determine the genetic difference and the evolutionary relationship among species of plants and animals.

Agarose Gel

There are two basic types of materials used to make gels: agarose and polyacrylamide. Agarose is a natural colloid extracted from sea weed. It is very fragile and easily destroyed by handling. Agarose gels have very large "pore" size and are used primarily to separate very large molecules with a molecular mass greater than 200 kDa. Agarose gels can be processed faster than polyacrylamide gels, but their resolution is inferior. That is, the bands formed in the agarose gels are fuzzy and spread far apart. This is a result of pore size and it cannot be controlled. However, for the DNA analysis agarose gel electrophoresis is commonly used to separate the DNA band.

Agarose is a linear polysaccharide (average molecular mas about 12,000) made up of the basic repeat unit agarobiose, which comprises alternating units of galactose and 3,6-anhydrogalactose. Agarose is usually used at concentrations between 1% and 3%. Agarose gels are formed by suspending agarose powder in aqueous buffer, then boiling the mixture until a clear solution
forms. This is poured and allowed to cool to room temperature to form a rigid gel.

Electrophoresis of Nucleic Acids

Gel electrophoresis is the process by which scientists can sort pieces of DNA cut with restriction enzymes by size. An agarose or polyacrylamide gel is loaded with the DNA fragments and
cu rrent is passed through the gel. Since DNA is negatively charged, it will migrate towards the positive pole. The DNA will not migrate at the same rate, however. Larger pieces of DNAcollide with the gel matrix more often and are slowed down,while smaller pieces of DNA move through more quickly. Since different genes have different nucleotide sequences, restriction enzymes will cut them at different places, generating different size DNA fragments. By using gel electrorphoresis, biologists can tell which gene is which based upon the sizes of
the fragments generated when a gene is treated with a restriction enzyme.

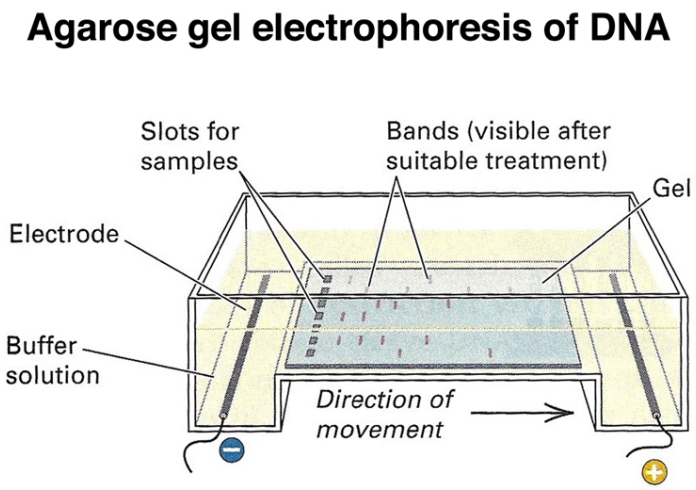


Fig 6.Agarose gel electrophoresis

Restriction enzymes – cutting DNA

The **restriction enzymes**, which cut DNA at defined sites, represent one of the most important groups of enzymes for the manipulation of DNA. These enzymes are found in bacterial cells, where they function as part of a protective mechanism called the **restriction-modification system**. In this system the restriction enzyme hydrolyses any exogenous DNA that appears in the cell. To prevent the enzyme acting on Restriction enzymes act as a 'protection' system for bacteria in that they hydrolyse exogenous DNA that is not methylated by the host modification enzyme. the host cell DNA, the modification enzyme of the system (a methylase) modifies the host DNA by **methylation** of particular bases in the restriction enzyme's **recognition sequence**, which prevents the enzyme from cutting the DNA.

Restriction enzymes are of three types (I, II, or III). Most of the enzymes commonly used today are type II enzymes, which have thesimplest mode of action. These enzymes are **nucleases** as they cut at an internal position in a DNA strand (as opposed to beginning degradation at one end) they are known as **endonucleases**. Thus, the correct designation of such enzymes is that they are type II restriction endonucleases, although they are often simply called restriction enzymes. In essence they may be thought of as **molecular scissors**.

Type II restriction endonucleases

Restriction enzyme nomenclature is based on a number of conventions. The generic and specific names of the organism in which the enzyme is found are used to provide the first part of the designation, which comprises the first letter of the generic name and the first two letters of the specific name. Thus, an enzyme from a strain of ***Escherichia coli*** is termed *Eco*, one from ***Bacillus amyloliquefaciens*** is *Bam*, and so on. Further descriptors may be added, depending on the bacterial strain involved and on the presence or absence of extrachromosomal elements. Two widely used enzymes from the a forementioned bacteria are *Eco*RI and *Bam*HI. The value of restriction endonucleases lies in their specificity. Each particular enzyme recognises a specific sequence of bases in the DNA, the most common recognition sequences being four, five, or six base pairs in length. Thus, given that there are four bases in the DNA, and assuming a random distribution of bases, the expected frequency of any

particular sequence can be calculated as 4*n*, where *n* is the length of the recognition sequence. This predicts that tetranucleotide sites will occur every 256 base pairs, pentanucleotide sites every 1024 base pairs, and hexanucleotide sites every 4096 base pairs. There is, as one might expect, considerable variation from these values, but generally the fragment lengths produced will lie around the calculated value. Thus, an enzyme recognising a tetranucleotide sequence (sometimes called a 'four-cutter') will produce shorter DNA fragments than a six-cutter. Some of the most commonly used restriction enzymes are listed in Tab No : 1

Table no: 1 : List of Restriction enzymes

Enzyme	Recognition sequence	Cutting sites Ends
*Bam*HI	5_-GGATCC-3_	G▼GAT C C 5_ C C T A G▲G
*Eco*RI	5_-GAATTC-3_	G▼A AT T C 5_ C T T A A▲G
*Hae*III	5_-GGCC-3	_ GG▼C C Blunt C C▲GG
*Hpa*I	5_-GTTAAC-3_	G T T▼A A C Blunt C A A▲T T G
*Pst*I	5_-CTGCAG-3_	C T GC A▼G 3_ G▲A C G T C
*Sau*3A	5_-GATC-3_	▼G A T C 5_ C T A G▲
*Sma*I	5_-CCCGGG-3_	C C C▼GG G Blunt

Use of restriction endonucleases

Restriction enzymes are very simple to use -- an appropriate amount of enzyme is added to the target DNA in a buffer solution, and the reaction is incubated at the optimum temperature (usually 37◦C) for a suitable length of time. Enzyme activity is expressed in units, with one unit being the amount of enzyme that will cleave one microgram of DNA in one hour at 37◦C. Although most experiments require complete digestion of the target DNA, there are some cases where various combinations of enzyme concentration and incubation time may be

used to achieve only partial digestion. The type of DNA fragment that a particular enzyme produces depends on the recognition sequence and on the location of the cutting site within this sequence. As we have already seen, fragment length is dependent on the frequency of occurrence of the recognition sequence.

The actual cutting site of the enzyme will determine the type of ends that the cut fragment has, which is important with regard to further manipulation of the DNA.

Three types of fragment may be produced:

(1) **Blunt ends** (sometimes known as flush-ended fragments),

(2) Fragments with protruding 3_ ends, and

(3) Fragments with protruding 5_ ends.

Enzymes such as *Pst*I and *Eco*RI generate DNA fragments with cohesiveor '**sticky**' **ends**, as the protruding sequences can base pair with complementary sequences generated by the same enzyme. Thus, by cutting two different DNA samples with the same enzyme and mixing the fragments together, recombinant DNA can be produced. This is one of the most useful applications of restriction enzymes and is a vital part of many manipulations in genetic engineering.

Restriction mapping

Most pieces of DNA will have recognition sites for various restriction enzymes, and it is often beneficial to know the relative locations of some of these sites. The technique used to obtain this information is known as **restriction mapping**. This involves cutting a DNA fragment with a selection of restriction

enzymes, singly and in various combinations. The fragments produced are run on an agarose gel and their sizes determined. From the data obtained, the relative locations of the cutting sites can be worked out. A fairly simple example can be used to illustrate the technique, as outlined in the following. Let us say that we wish to map the cutting sites for the restriction enzymes *Bam*HI, *Eco*RI, and *Pst*I, and that the DNA fragment of interest is 15 kb in length. Various digestions are carried out, and the fragments arising from these are analysed and their sizes determined. As each of the single enzyme reactions produces two DNA fragments, we can concludethat the DNA has a single cutting site for each enzyme. Thedouble digests enable a map to be drawn up, and the triple digest confirms this.

DNA modifying enzymes

Restriction enzymes and DNA ligase provide the cutting and joining functions that are essential for the production of recombinant DNA molecules. Other enzymes used ingenetic engineering may be loosely termed **DNA modifying enzymes**, with the term used here to include degradation, synthesis, and alteration of DNA.

Nucleases

Nuclease enzymes degrade nucleic acids by breaking the phosphodiester bond that holds the nucleotides together. Restriction enzymes are good examples of endonucleases, which cut within a DNA strand.

A second group of nucleases, which degrade DNA from the termini of the molecule, are known as **exonucleases**.Apart from restriction enzymes, there are four useful nucleases that are often used in genetic engineering. These are **Bal 31** and **exonuclease III** (exonucleases), and **deoxyribonuclease I** (DNase I) and **S1-nuclease** (endonucleases). These enzymes differ in their precise mode of action and provide the genetic engineer with a variety of strategies for attacking DNA. In addition to DNA-specific nucleases, there are **ribonucleases** (RNases), which act on RNA. These may be required for many of the stages in the preparation and analysis of recombinants and are usually used to get rid of unwanted RNA in the preparation. However,

as well as being useful, ribonucleases can pose some unwanted problems. They are remarkably difficult to inactivate and can be secreted in sweat. Thus, contamination with RNases can be a problem in preparing recombinant DNA, particularly where cDNA is prepared from an mRNA template. In this case it is vital to avoid RNase contamination by wearing gloves and ensuring that all glass and plastic equipment is treated to avoid ribonuclease contamination.

Polymerases

Polymerase enzymes synthesise copies of nucleic acid molecules and are used in many genetic engineering procedures. When describing a polymerase enzyme, the terms 'DNA-dependent' or 'RNA-dependent' may be used to indicate the type of nucleic acid template that the enzyme uses. Thus, a DNA-dependent DNA polymerase copies DNA into DNA, an RNA-dependent DNA polymerase copies RNA into DNA, and a DNA-dependent RNA polymerase transcribes DNA into RNA. These enzymes synthesise nucleic acids by joining together nucleotides whose bases are complementary to the template strand bases. The synthesis proceeds in a 5_→3_ direction, as each subsequent nucleotide addition requires a free 3_-OH group for the formation of the phosphodiester bond. This requirement also means that a short double-stranded region with an exposed 3_-OH (a primer) is necessary for synthesis to begin.

The enzyme **DNA polymerase I** has, in addition to its polymerase function, 5_→3_ and 3_→5_ exonuclease activities. The enzyme catalyses a strand-replacement reaction, where the 5_→3_ exonuclease function degrades the non-template strand as the polymerase synthesizes the new copy. A major use of this enzyme is in the nick translation procedure for radiolabelling DNA. The 5_→3_ exonuclease function of DNA polymerase I can be removed by cleaving the enzyme to produce what is known as the **Klenow fragment**. This retains the polymerase and 3_→5_ exonuclease activities. The Klenow fragment is used where a single-stranded DNA molecule needs to be copied; because the 5_→3_ exonuclease function is missing, the enzyme cannot degrade the non-template strand of dsDNA during synthesis of the new DNA. The 3_→5_ exonuclease activity is supressed under the conditions normally used for the reaction. Major uses for the Klenow fragment include radiolabelling by primed synthesis and DNA sequencing by the dideoxy method in addition to the copying of single-stranded DNAs during the production of recombinants.

Reverse transcriptase (RTase) is an RNA-dependent DNA polymerase, and therefore produces a DNA strand from an RNA template. It has no associated exonuclease activity. The enzyme is used mainly for copying mRNA molecules in the preparation of cDNA (**complementary**
or **copy DNA**) for cloning

Enzymes that modify the ends of DNA molecules

The enzymes **alkaline phosphatase**, **polynucleotide kinase**, and **terminal transferase** act on the termini of DNA molecules and provide important functions that are used in a variety of ways

Bacterial alkaline phosphatase (there is also a similar enzyme, calf intestinal alkaline phosphatase)removes phosphate groups from the 5_ ends of DNA, leaving a 5_-OH group. The enzyme is used to prevent unwanted ligation of DNA molecules, which can be a problem in certain cloning procedures.It is also used prior to the addition of radioactive phosphate to the 5_ ends of DNAs by polynucleotide kinase

Terminal transferase (terminal deoxynucleotidyl transferase) repeatedly adds nucleotides to any available 3_ terminus. Although it works best on protruding 3_ ends, conditions can be adjusted so that blunt-ended or 3_-recessed molecules may be utilised. The enzyme is mainly used to add homopolymer tails to DNA molecules prior to the construction of recombinants.

DNA ligase – joining DNA molecules

DNA ligase is an important cellular enzyme, as its function is to repair broken phosphodiester bonds that may occur at random or molecular glue, which is used to stick pieces of DNA together. This function is crucial to the success of many experiments, and DNA ligase is therefore a key enzyme in genetic engineering.

The enzyme used most often in experiments is T4 DNA ligase, which is purified from *E. coli* cells infected with bacteriophage T4. Although the enzyme is most efficient when sealing gaps in fragments that are held together by cohesive ends, it will also join blunt-ended DNA molecules together under appropriate conditions. The enzyme works best at 37◦ C, but is often used at much lower temperatures (4--15◦C) to prevent thermal denaturation of the short base-paired regions that hold the cohesive ends of DNA molecules together.

The ability to cut, modify, and join DNA molecules gives the genetic engineer the freedom to create recombinant DNA molecules.

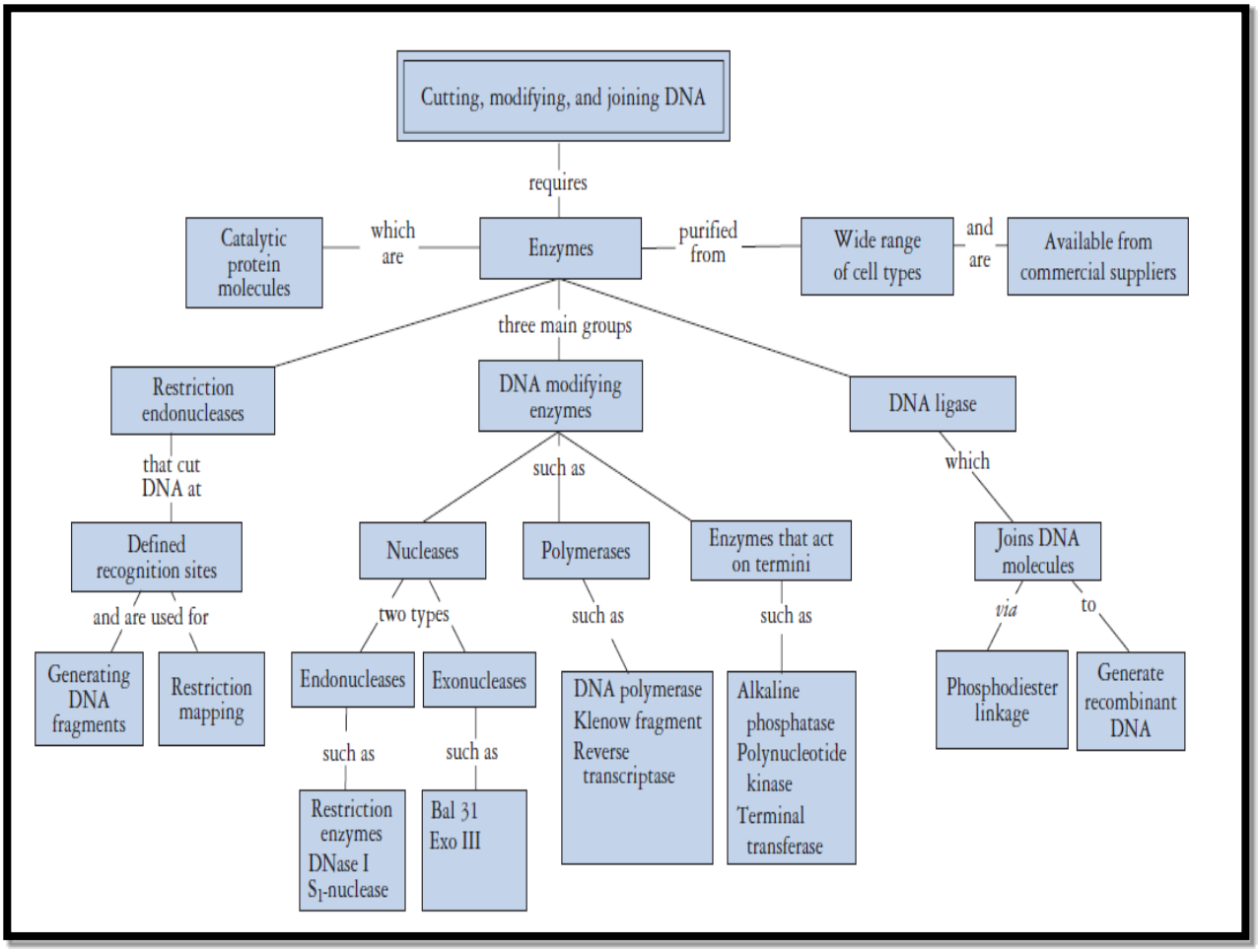


Table : 2 Types of Restriction Enzymes

Modifications of Cut Ends in DNA

The 3′-ends of DNA strands generated by cleavage always carry a free hydroxyl (—OH) group, while their 5′-ends always bear a phosphate group. Often the ends produced by restriction enzymes have to be modified for further manipulation of the fragments; some of the modifications are summarised below.

1. Removal of the 5’phosphate group of vector DNA by alkaline phosphatase treatment in order to prevent vector circularization during DNA inserts integration.

2. Addition of a phosphate group to a free 5’hydroxyl group by T4 polynucleotide kinase.

3. Removal of the protruding ends by digestion with, say, S1 nuclease; this enzyme digests both 3′- and 5′- protruding ends.

4. Making the single-stranded protruding ends double-stranded by extending the recessed (shorter) strand with, say, Klenow fragment of E. coli DNA polymerase I. This strategy is preferred to SI nuclease digestion for various reasons. (Both the strategies 3 and 4 generate blunt ends, which can be ligated by T_4 polynucleotide ligase.)

5. Removal of nucleotides from the 5′-ends using A exonuclease.

6. Removal of nucleotides from the 3′-ends using E. coli exonuclease III.

(Both the strategies 5 and 6 convert blunt ends into protruding single-stranded ends of undefined base sequence).

7. Treatment of double-stranded DNAs with exonuclease Ball 1, which simultaneously digests both the strands (from both the ends) of a DNA molecule; this treatment produces shortened DNA fragments with blunt ends.

8. Synthesis of single-stranded tails (protruding ends) at the 3′-ends of blunt-ended fragments by the enzyme terminal deoxynucleotidy1 transferase; this is called tailing. This reaction 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 the 3′-ends of the vector; the protruding ends of the DNA insert and the vector will, therefore, base pair under annealing conditions.

In practice, the tails on vector and DNA inserts differ in length. As a result, short gaps remain when the ends of DNA insert pair with those of vector. Therefore, Klenow fragment is first used to fill in this gap before DNA ligase is used to ligate them together.

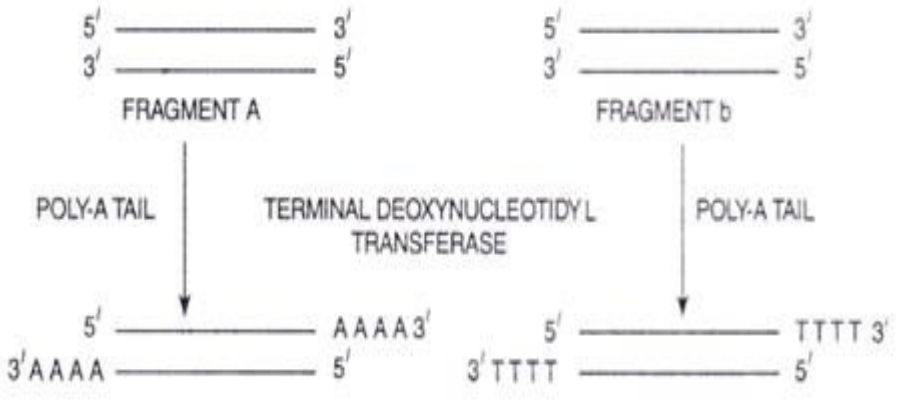


Fig 7.Tailing of blunt-ended DNA fragments A and B using terminal nucleotidyl transferase. Poly-A tail is added to the 3′ ends of fragment A, while poly-T tail is added to the 3′ ends of fragment B so that the protruding ends of fragment A are complementary to those of B.

9. Linker and/or adapter molecules can be joined to the cut ends. Linkers are short, chemically synthesized, double-stranded oligonucleotides, which contain within them one or more restriction endonuclease sites, e.g., linker 5'CCGAATTCGG (only one strand of the linker is shown here) contains one EcoRl site. Linkers are fused with blunt-ended DNA fragments; cleavage of the linker with the appropriate restriction enzyme creates suitable cohesive protruding ends .

Linkers have the following two applications: creation of cohesive ends (1) on blunt-ended DNA fragments, and (2) on fragments having unmatched or undefined sequences in their protruding ends. In the latter situation, the DNA fragments are first made blunt-ended using either strategy 3 or 4, following which the selected linkers are ligated to them by T4 ligase.

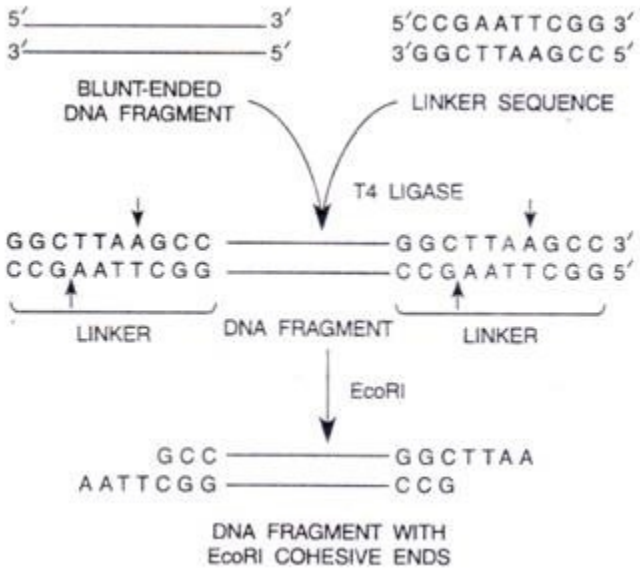


Fig : 8 Creation of cohesive ends on blunt-ended DNA fragments. Suitable linkers are ligated to the blunt ends by T4 DNA ligase. The linker is then cleaved with the appropriate restriction enzyme to generate sticky ends.

A potential drawback of linkers is as follows. The linkers must be cleared by the concerned restriction enzyme to generate the sticky ends. In case, one or more recognition sites for this enzyme are present within the DNA insert, it will also be cleaved into two or more pieces. In such a case, adapters are employed.

10. Adapters are short, chemically synthesized DNA double strands, which already have one or both sticky ends . When a blunt end is converted into a sticky end, the adapter has one blunt end and one sticky end corresponding to the concerned restriction enzyme. The blunt end of the adapter is ligated to the blunt ends of the DNA insert, which are now converted into sticky ends.

In order to prevent ligation of a further adapter molecule to a sticky end so produced, the 5′-terminus at the sticky end of the adapter molecules is occupied by a —OH rather than the normal phosphate group. After the adapters have been attached to the DNA insert, their 5′-ends are phosphorylated by polynucleotide kinase so that the DNA insert can be ligated to the vector.

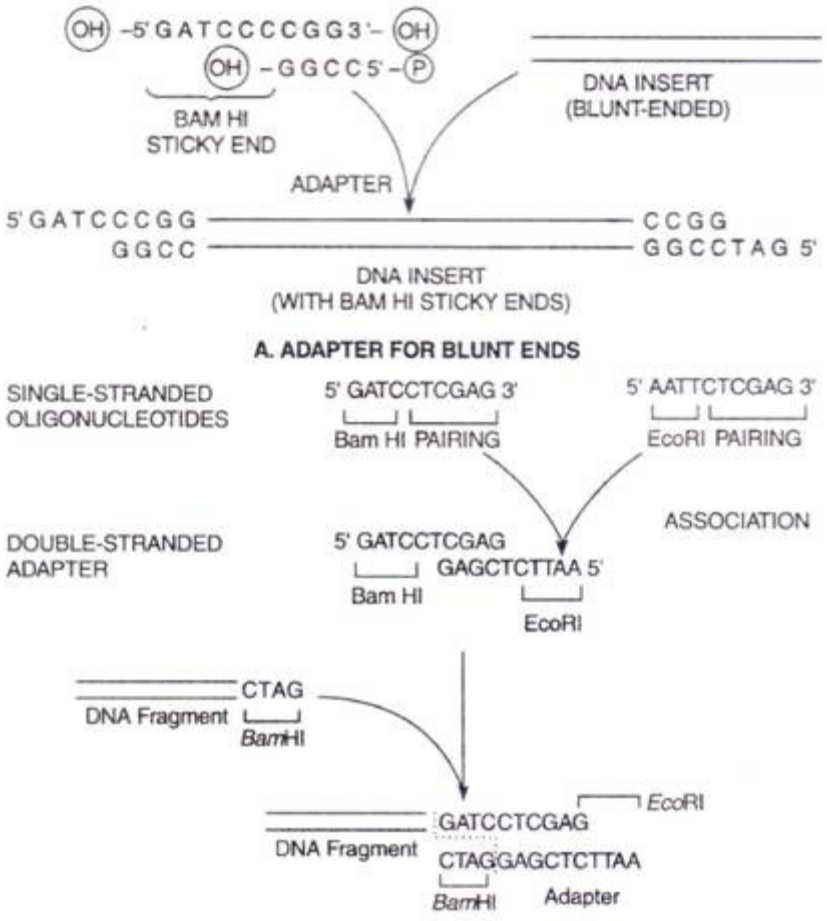


Fig : 9 Examples of adapters. A. An adapter used to create BamHl sticky ends at the blunt ends of a DNA insert. B. A conversion adapter produced by associating two oligonucleotides, each having a different recognition sequence at its 5′-end but a complementary sequence for base pairing at the 3′-end.

Linkers

Linkers are short stretches of double stranded DNA of length 8-14 bp that have recognition site for restriction enzymes. Linkers are ligated to blunt end DNA by ligase enzyme. The linker ligation is more efficient as compared to blunt-end ligation of larger molecules because of the presence of high concentration of these small molecules in the reaction. The ligated DNA can be digested with appropriate restriction enzyme generating cohesive ends required for cloning in a vector. The restriction sites for the enzyme used to generate cohesive ends may be present within the target DNA fragment which may limit their use for cloning.

Adapters

These are short stretches of oligonucleotide with cohesive ends or a linker digested with restriction enzymes prior to ligation. Addition of adaptors to the ends of a DNA converts the blunt ends into cohesive ends.

ii. Restriction enzymes that generate sticky ends

Genomic DNA can be digested with commonly available restriction enzymes that generate sticky ends. For example, digestion of genomic DNA with the restriction enzyme *Sau3AI* (recognition sequence 5'-GATC-3') generates DNA fragments that are compatible with the sticky end produced by *BamHI* (recognition sequence 5'-GGATCC-3') cleavage of a vector. Once the DNA fragments are produced, they are cloned into a suitable vector.

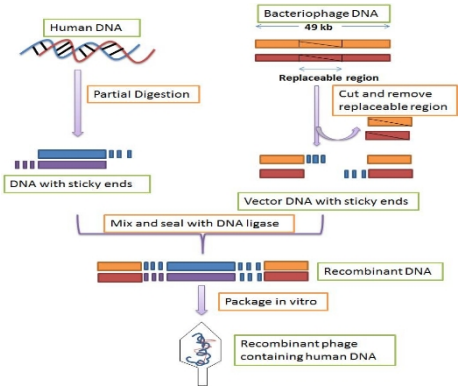


Figure 10. Construction of genomic library

Cloning of genomic DNA

Various vectors are available for cloning large DNA fragments. λ phage, yeast artificial chromosome, bacterial artificial chromosome etc. are considered as suitable vectors for larger DNA and λ replacement vectors like *λDASH* and *EMBL3* are preferred for construction of genomic DNA library. T4 DNA ligase is used to ligate the selected DNA sequence into the vector.

Table : 3 Lists of vectors and their Features

Vector	Insert size	Features
λ phages	Up to 20-30 kb	Genome size-47 kb, efficient packaging system, replacement vectors usually employed, used to study individual genes.

Cosmids	Up to 40 kb	Contains *cos* site of λ phage to allow packaging, propagate in *E. coli* as plasmids, useful for sub-cloning of DNA inserts from YAC, BAC, PAC etc.
Fosmids	35-45 kb	Contains F plasmid origin of replication and λ*cos* site, low copy number, stable.
Bacterial artificial chromosomes (BAC)	Up to 300kb	Based on F- plasmid, relatively large and high capacity vectors.
P1 artificial chromosomes (PACs)	Up to 300 kb	Derived from DNA of P1 bacteriophage, combines the features of P1 and BACs, used to clone larger genes and in physical mapping, chromosome walking as well as shotgun sequencing of complex genomes.
Yeast artificial chromosomes (YAC)	Up to 2000kb	Allow identification of successful transformants (BAC clones are highly stable and highly efficient)

(1) λ replacement vectors

The *λEMBL* series of vectors are widely used for genomic library construction. The multiple cloning sites of these vectors flanking the stuffer fragment contain opposed promoters for the T3 and T7 RNA polymerases. The restriction digestion of the recombinant vector generates short fragments of insert DNA left attached to these promoters. This generates RNA probes for the ends of the DNA insert. These vectors can be made conveniently, directly from the vector, without recourse to sub-cloning.

(2) High-capacity vectors

The high capacity cloning vectors used for the construction of genomic libraries are cosmids, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs) and yeast artificial chromosomes (YACs). They are designed to handle longer DNA inserts, much larger than

for λ replacement vectors. So they require lower number of recombinantsto be screened for identification of a particular gene of interest.

The recombinant vectors and insert combinations are grown in *E. coli* such that a single bacterial colony or viral plaque arises from the ligation of a single genomic DNA fragment into the vector.

Number of clones required for a library

The number of clones to be pooled depends upon the size of the genome *f* and average size of the cloned DNA.

Let (*f*) be the fraction of the genome size compared to the average individual cloned fragment size, would represent the lowest possible number of clones that the library must contain.

The minimum number of clones required can be calculated as-

f= genome size/ fragment size

For the *E. coli* genome (4.6 Mb) with an average cloned fragment size of 5 kb, *f* will be 920.

The number of independent recombinants required in the library must be greater than *f,* as sampling variation leads to the several times inclusion and exclusion of some sequences in a library of just *f* recombinants. In 1976, Clarke and Carbon derived a formula to calculate probability (P) of including any DNA sequence in a random library of N independent recombinants.

The actual number of clones required can be calculated as-

$$N = \frac{\ln (1-P)}{\ln (1-f)}$$

where N= number of clones and P= probability that a given gene will be present.

Bigger the library better will be the chance of finding the gene of interest. The pooling together

of either recombinant plaques or bacterial colonies generates a primary library.

4-4.3. Amplified library

- The primary library created is usually of a low titer and unstable. The stability and titer can be increased by amplification. For this, the phages or bacterial colonies are plated out several times and the resulting progenies are collected to form an amplified library.
- The amplified library can then be stored almost indefinitely due to longshelf-life of phages.
- It usually has a much larger volume than the primary library, and consequently may be screened several times.
- It is possible that the amplification process will result in the composition of the amplified library not truly reflecting the primary one.
- Certain DNA sequences may be relatively toxic to *E. coli* cells. As a consequence bacteria harboring such clones will grow more slowly than other bacteria harboring non-toxic DNA sequences. Such problematic DNA sequences present in the primary library may be lost or under-represented after the growth phase required to produce the amplified library.

Subgenomic library

Subgenomic library is a library which represents only a fraction of the genome. Enhancing the fold of purification of target DNA is crucial for subgenomic DNA libraries which can be achieved by multiple, sequential digestion when information of the restriction map of the sequences of interest is known. After initial purification of a given fragment, the purification can further be increased by redigestion with another enzyme generating a smaller (clonable) fragment relative to original DNA.

Advantages of genomic libraries

- Identification of a clone encoding a particular gene of interest.
- It is useful for prokaryotic organisms having relatively small genomes.
- Genomic libraries from eukaryotic organisms are very important to study the genome sequence of a particular gene, including its regulatory sequences and its pattern of introns and exons.

Disadvantages of genomic library

- Genome libraries from eukaryotes having very large genomes contain a lot of DNA which does

not code for proteins and also contain non-coding DNA such as repetitive DNA and regulatory regions which makes them less than ideal.

- Genomic library from a eukaryotic organism will not work if the screening method requires the expression of a gene.

Applications

- To determine the complete genome sequence of a given organism.
- To study and generate transgenic animals through genetic engineering, serving as a source of genomic sequence.
- To study the function of regulatory sequences *in vitro.*
- To study the genetic mutations.
- Used for genome mapping, sequencing and the assembly of clone contigs.

Comparison of Genomic and cDNA Libraries

cDNA library has revolutionized the field of molecular genetics and recombinant DNA technology. It consists of a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism is represented as its cDNA insertion in a vector. cDNA libraries are used to express eukaryotic genes in prokaryotes. In addition, cDNAs are used to generate expressed sequence tags (ESTs) and splices variant analysis.

Table 4 Some of the differences of cDNA library with genomic library are presented

Feature	Genomic library	cDNA library
Sequences present	Ideally, All genomic Sequences	Only structural genes that are Transcribed
Contents affected by : (a) Developmental stage (b) Cell type	 No No	 Yes Yes
Features of DNA insert(s)		

representing a gene: (a) Size	As present in genome	Ordinarily, much smaller
(b) Introns	Present	Absent
(c) 5′- and 3′- regulatory sequences	Present	Absent
As compared to the genome (a) Enrichment of sequences	In amplified genomic libraries	For abundant mRNAs
(b) Redundancy in frequency	In amplified libraries	For rare mRNA species
(c) Variant forms of a gene	Not possible	For such genes, whose RNA transcripts are alternatively Spliced

UNIT – 2 – GENETIC ENGINEERING – SBTA1501

CLONING VECTORS

Cloning vector is a small molecule of DNA that can carry foreign DNA into a host cell and is capable of self-replication. Different types of cloning vectors are available, each vector designed to perform a specific function.

Choice of the right cloning vector is an important aspect of cloning. There are certain essential

properties which a cloning vector must possess:

- It should be able to replicate itself and the DNA segment it carries independently.
- It should be of small size, so that it is easy to manipulate.
- It should be easy to recover from the host cell.
- It should be easily transferred from one cell to another by simple methods.
- It should contain several unique restriction enzyme sites. These sites are present only once and clustered in one location and are known as multiple cloning sites or polylinker. Restriction enzyme site is cleaved with a restriction enzyme to open the cloning vector without disrupting any other region. Insert DNA fragment is digested with the same enzyme and ligated with the vector.
- There should be easy ways to detect their presence in host organisms. Bacterial cloning plasmids carry an antibiotic resistance gene to distinguish the host cells that carry vectors from the host cells that do not contain vectors.
- There should be easy ways to detect the presence of inserted DNA.

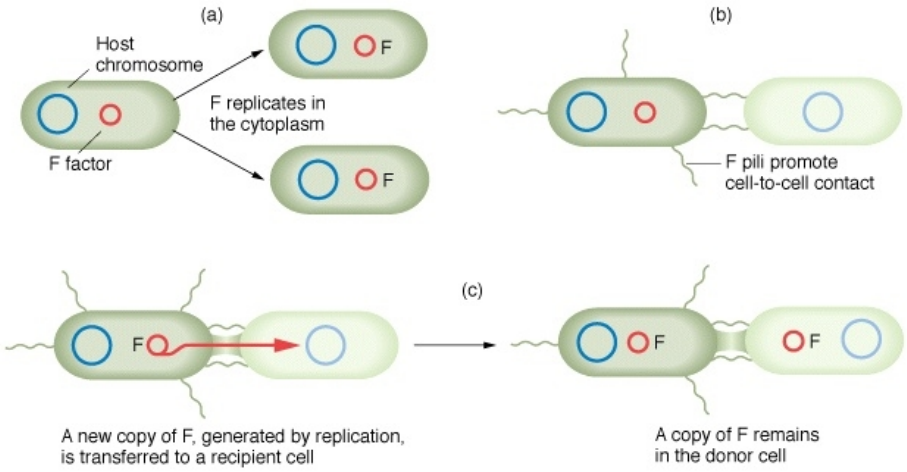
Plasmids

Plasmids are circular double stranded 2.5 to 5 Kb long DNA molecules, present in bacteria that replicate autonomously from the host chromosome and are inherited in extra-chromosomal state from one generation to the other. They are widespread throughout the prokaryotes and are capable of carrying 15 Kb of foreign DNA. Most of the bacterial plasmids fulfill first four requirements of a cloning vector.

Five types of plasmids known are:

Fertility plasmids or F plasmids – These are conjugative plasmids that allow F^+ bacterial cells that carry this plasmid to transfer the plasmid at a high frequency to F^- bacterial cells that lack this plasmid by the process of conjugation.

Fig: 1 Fertility of Plasmids



- **Resistance plasmids or R plasmids** – These are also conjugative plasmids that contain genes for resistance against antibiotics, metal ions, ultraviolet radiation and bacteriophages.

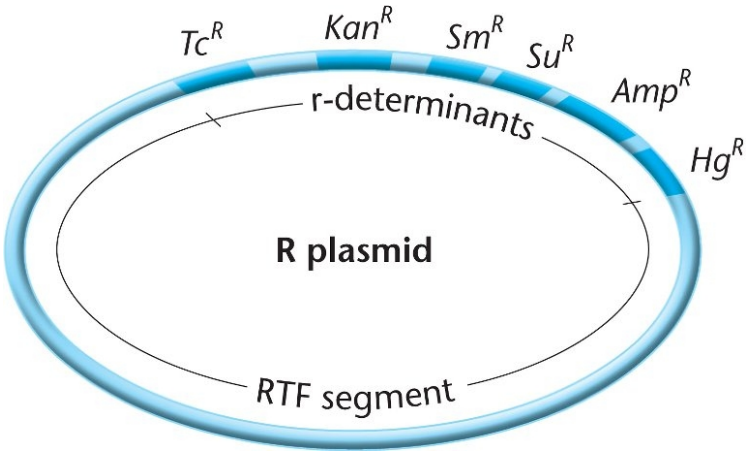


Figure:2 An **R plasmid** consists of a resistance transfer factor (**RTF**), which enables **conjugation**, and one or more **r-determinants**: genes conferring resistance to antibiotics.

- **Col plasmids** – These plasmids carry genes that code for bacteriocins, proteins required to kill other bacteria.

- **Degradative plasmids** – These plasmids contain genes that degrade chemicals and aid in digestion of those toxins, such as toluene and salicylic acid. E.g.; TOL plasmids

- **Virulence plasmids** – These plasmids are required by the bacterium for pathogenicity, the ability of the bacterium to cause disease in a host organism. E.g.; Ti plasmids

Plasmids as cloning vectors

Earlier cloning experiments used natural *in vivo* plasmids, i.e.ColE1 and pSC101. ColE1 is a small circular plasmid known for production of 57kD toxic protein known as colicin E1. These natural plasmids are of large size, possess stringent replication control and lack marker genes. Further studies on cloning led to the construction of synthetic plasmids to meet other requirements of a vector and to enhance the delivery of gene of interest. The first successful plasmid was pSC101.

Basic properties of plasmids are:

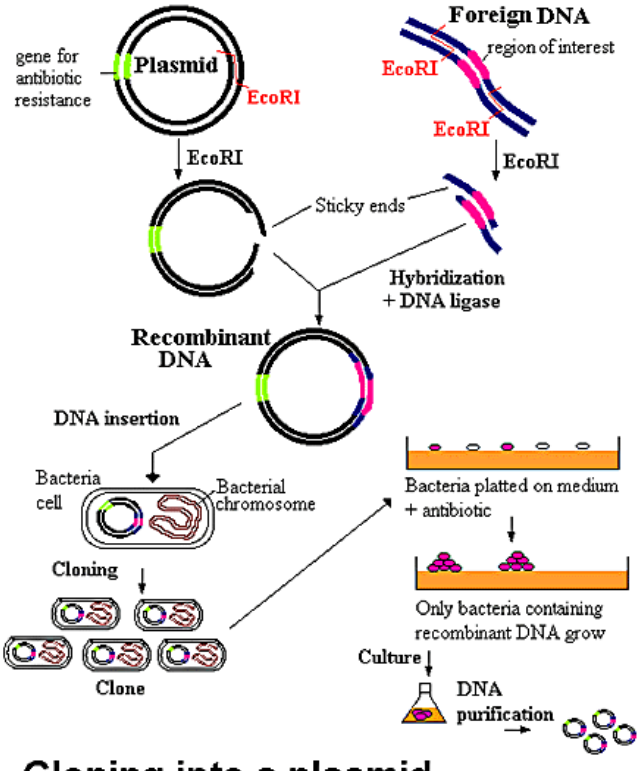
1. Plasmids vary in their copy number in each bacterial cell. Some plasmids occur in just one/few copies and are known as low copy plasmids whereas others exist in multiple copies per cell and are known as high copy plasmids. The number of copies of plasmid in each bacterial cell depends on type of replicon present in it. Replicon contains essential regulatory elements and origin of replication or *ori*.

   Plasmids carrying pMB1 or colicin E1 (colE1) replicons are under stringent control and maintain low copies (15-20) in each bacterial cell. pMB1 or colE1 replicons have been modified extensively over the years to increase their copy number. These high copy plasmids like pBR322 are more useful for cloning purposes as higher yield of insert DNA can be obtained from higher

amount of plasmid DNA. Low copy plasmids like pSC101 are typically used when genes toxic to the host bacteria are cloned as higher copies will be lethal. They are also used for construction of bacterial artificial chromosome (BAC).

2. Some plasmids have narrow host range as they can replicate in restricted number of bacterial species, for e.g. pBR322, pUC18 are restricted to *E. coli* and closely related species only. Some plasmids can replicate in large number of bacterial species and are known as broad range plasmids for e.g. IncP of *E.coli*. These broad range plasmids are a useful tool in bacterial mapping studies.

3. Plasmids can be introduced easily into bacterial cells by a process called transformation. Transformed bacteria are grown on agar plate and bacterial colonies derived from individual cells that have taken up recombinant plasmid are grown. Transformation is generally inefficient as plasmids are stably established in small population of cells only. These plasmids also carry selectable markers which allow rare transformants to be selected with ease.



Cloning into a plasmid

Figure:3 Antibiotic selection gene in plasmid provides an easy means of selecting cells containing the plasmid.

4. Plasmids often carry genes that confer antibiotic resistance to the host organism. The antibiotic resistance genes encoded by plasmid DNA is an advantageous trait and used in the construction of vectors. These genes provide an easy means of selecting cells containing the plasmid. By growing colonies on agar plate containing antibiotic it is possible to isolate bacteria that have taken up plasmid during transformation as only resistant bacteria will be able to grow. These markers confer the cells containing vector, the ability to grow in presence of antibiotics (like ampicillin, kanamycin, tetracycline etc.). Bacteria carrying parental plasmid are unable to form colonies under selective conditions. For example, if the cloning vector carries a kanamycin resistance gene (*kan*^{r}), it will confer resistance to the *E. coli* cells against kanamycin if the transformation process was successful. Therefore it is possible to isolate bacteria that have taken up plasmid during transformation as only resistant bacteria will be able to grow.

5. Most plasmid vectors contain multiple cloning sites (MCS) or polylinker. This allows the investigator to cleave the plasmid at one place without disrupting plasmid's replication genes. Multiple cloning sites are sometimes present in between a gene which becomes non-functional when a DNA insert is cloned into the vector. Insertional inactivation of an antibiotic resistance gene permits simple selection of the desired clone. In some cases, plasmid vectors carry two different antibiotic resistance genes. Foreign DNA is cloned into one of the antibiotic resistance gene. Recombinant host cells will be sensitive to one antibiotic and resistant to the other

pBR322

One of the most commonly used and versatile plasmids was developed by Francisco Bolivar and Rodriguez in 1977 and is named eponymously as pBR322 (p stands for plasmid, B for Bolivar and R for Rodriguez).This plasmid has been completely sequenced and is therefore widely used as a cloning vector. Exact length of every inserted fragment can be calculated.

- It has a small size of 4.361 Kb, facilitating its easy entry into cells. Entry of a vector becomes

less successful as vector size increases.

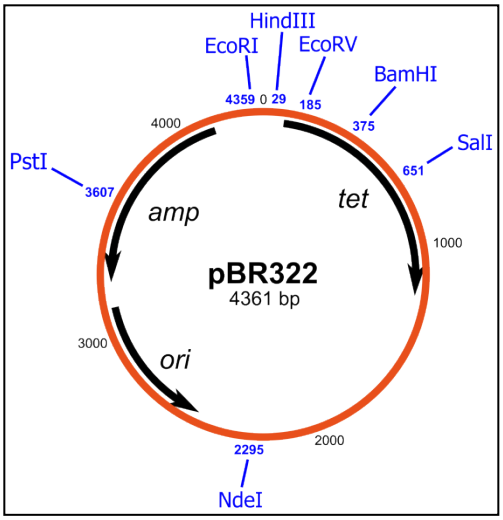
- The plasmid contains: (i) replicon, pMB1 (a close relative of ColE1) for replication of plasmid, (ii) the *rop* gene, encoding for the Rop protein, which promotes formation of stable RNA I – RNA II complex from the unstable complex and decreases copy number, (iii) the *bla* gene, encoding β-lactamase, that provides resistance to ampicillin, (iv) the tet^R gene, that confers tetracycline resistance.

- There are over 40 enzymes with unique restriction sites on the pBR322 genome. Plasmid can be cleaved at these sites and DNA fragments can be inserted. Genes
  that confer resistance to antibiotics i.e., tetracycline (tet^R) and ampicillin (amp^R), allow the identification of cells containing the intact plasmid or a recombinant version of the plasmid. There are 11 restriction sites that lie within tet^R gene. 6 restriction sites are inside amp^R gene. Insertion of DNA segment within the tet^R or amp^R site inactivates the gene conferring tetracycline or ampicillin resistance. Host cells without the vector are sensitive to both antibiotics. Transformed cells containing

  pBR322 with insertion in tet^R gene are sensitive to tetracycline but resistant to ampicillin. Transformed cells containing pBR322 with insertion in amp^R gene are sensitive ampicillin, but resistant to tetracycline. Insertional inactivation of antibiotic resistance gene allows the recombinants to be readily selected. Cells containing pBR322 lacking the DNA insert are resistant to both antibiotics.

- It has been widely used as a model system for the study of prokaryotic transcription and translation.

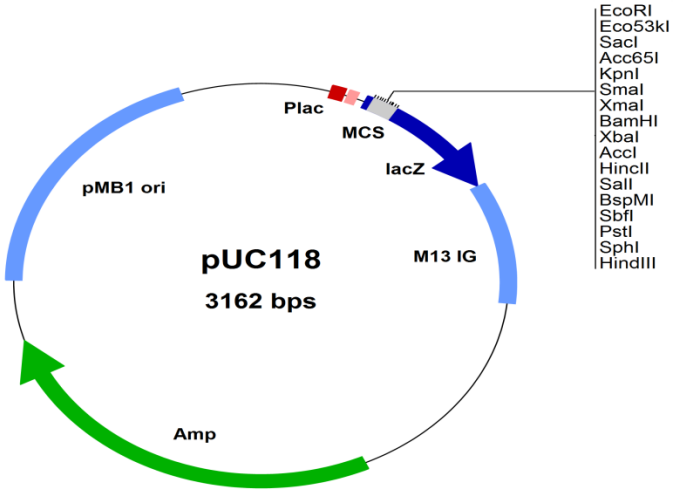
Fig : 4 pBR 322



pUC series

pUC series vectors were derived from plasmid pBR322 by Messing and co-workers to fulfill special cloning requirements. These plasmid vectors are called so because initial work on this vector was carried in the University of California and prefix "p" stands for "plasmid". pUC vectors are small, high copy number plasmids, with 500-700 copies per cell. The circular double stranded DNA of pUC18/19 has 2,686 base pairs.

Fig : 5 pUC 118



pUC18/19 plasmids contain:

(i) the pMB1 replicon for replication of the plasmid. Lack of the *rop* gene and a single point mutation in the replicon rep of pMB1 results in high copy number of pUC plasmids in comparison to pBR322

(ii) the *bla* gene coding for β-lactamase, conferring ampicillin resistance. This gene differs from that of pBR322 plasmid by two point mutations. (3) *bla* gene is fused with *E. coli lacZ'* gene which consists of CAP protein binding site, promoter plac, operator, lac repressor binding site and the 5' terminal region including 146 codons that code for deleted version of the gene *lacZ* encoding β-galactosidase. Codons 6–7 of *lacZ* are replaced by the multiple cloning site (MCS) region, where various recognition sites for many restriction enzymes are present.

pUC18 and puC19 are similar except that they contain multiple cloning sites (MCS) arranged in opposite orientations.

The pUC vectors are the first plasmid vectors to contain polylinker or multiple cloning sites (MCS). The MCS increases the number of potential cloning strategies by extending the range of enzymes. It provides extra flexibility during cloning procedure by allowing foreign DNA to be inserted at any of several restriction sites. The potential drawback of polylinker is inability to use insertional inactivation of antibiotic resistance to screen for recombinants.

Functional *lacZ* gene in the pUC vectors allows the researcher to differentiate between recombinant and parental host bacteria by visual screening of bacterial colonies on media containing the lactose analogue, X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside). This fragment is capable of intra-allelic (α) complementation with a defective form of β-galactosidase synthesized by the host genome. In the presence of IPTG, bacteria can synthesize both the fragments of the enzyme and form blue colonies on media containing X-Gal.β-galactosidase can hydrolyze X-Gal producing a blue colored substance and colonies that carry the plasmid are therefore blue colored. Multiple cloning site is located on the plasmid immediately downstream to the start codon of *lacZ gene*. Insertion of foreign DNA into the multiple cloning site disrupts the N-terminal fragment of β-galactosidase resulting in its inability to form blue color and therefore white colonies are formed.

Bacteriophages as vectors

Max Delbruck, in the 1940s laid foundations for the study of bacteriophages. Bacteriophages or phages are viral genomes that attack bacteria. Structurally bacteriophages can be tailless, with head and tail or filamentous. Phage genome which can be single or double-stranded DNA or RNA is encapsulated in an icosahedral protein shell known as head or capsid, in tailless or tailed forms.

Phages may be designated as either temperate or virulent, depending on their life cycles. When a phage enters a bacterial cell it can produce more phages of its kind and kill the bacterial cell. This is called the lytic growth cycle of a phage. It can also integrate into the chromosome of host and remain in a quiescent state without killing the bacterial cell. This is the lysogenicgrowth cycle of a phage. Virulent phages exhibit only the lytic life cycle. Temperate phages exhibit both lysogenic life cycles and lytic cycle. Bacteriophages can be used as vectors as larger non-viral DNA segments can be packaged in the virus particle. Phage vectors infect cells more efficiently than plasmids transform the cells; therefore yield of clones with phage vectors is usually more than the yield of plasmid transformed cells.

Lambda (λ) Phage

Lambda bacteriophages are the most important and widely used vectors having linear genomes. Wild-type lambda phages have 48.5 Kb linear double-stranded DNA
genome having 73 ORFs (open reading frames). The linear genome bears short (12 bp) single-stranded sequences at its ends known as cos sites which are complementary to each other. These cohesive or 'sticky' ends enable circularization of the λ genome after infection during packaging. Phage vectors bind to receptors on the bacterial surface. This is known as adsorption. The phage DNA is injected into the cell and the phage life cycle can start. The phage genome circularizes and the phage can initiate either the lytic or lysogenic cycle, depending on nutritional and metabolic status of the host cell. The ratio of phage to bacteria during adsorption also influences the lytic or lysogenic growth cycle. Specific environmental changes can also trigger the lysis of host cell.

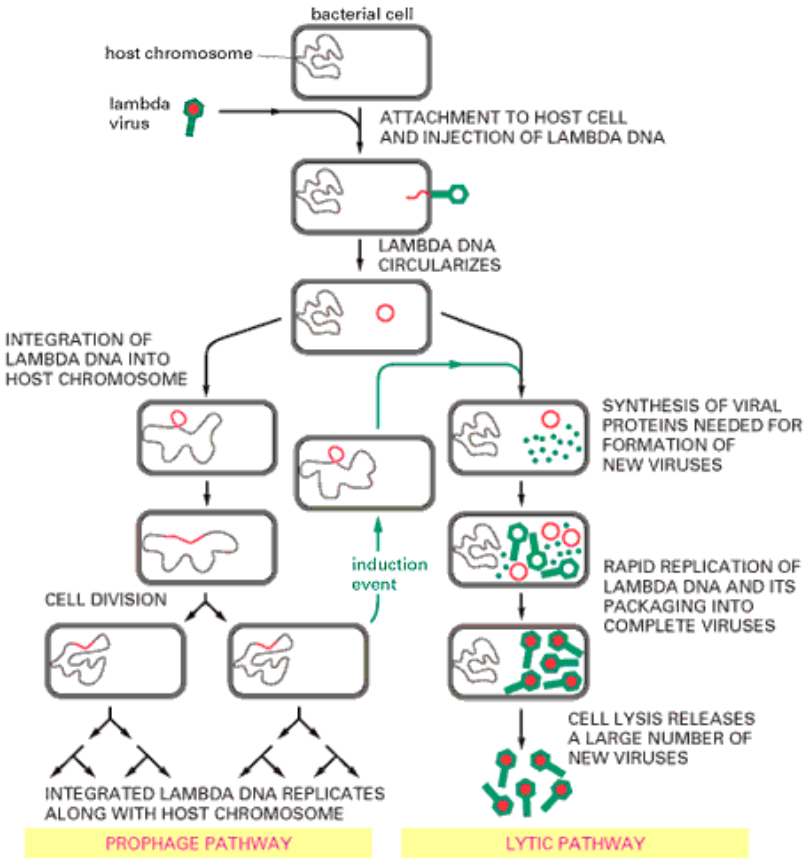


Fig :6 Lytic and Lysogenic pathway of lambda phages

Important features contributing to the utility of λ phages as vectors are:

Only two-thirds of the genome is essential and about one-third of the genome is non-essential which can be discarded and replaced with foreign DNA. The middle region of the lambda genome is dispensable as it controls the lysogenic properties of the phage, therefore this region can be deleted without disrupting the functions required for the lytic growth cycle. This region is substituted with the polylinker.

The packaging mechanism in λ phage facilitates insertion of recombinant DNA as packaging into infectious phage particles is possible only if DNA is between 40 to 53 Kb long. λ bacteriophage vectors can be digested to release three fragments, of which two fragments harbor essential genes comprising ~ 30 Kb. The third middle fragment can be discarded when additional DNA is inserted for cloning, producing viable phage particles. Therefore bacteriophage vectors allow the insertion of DNA fragments of up to 23 Kb. Once the bacteriophage fragments are ligated to insert DNA fragments of appropriate

size, the resulting recombinant DNAs can be packaged into phage particles. This requires adding packaged phage to crude bacterial cell extracts containing all the proteins required to assemble a complete phage, also known as *in vitro* packaging. Nearly all infective viable phage assemblies contain a foreign DNA fragment. They can be subsequently transferred into *E. coli* cells in an efficient manner.

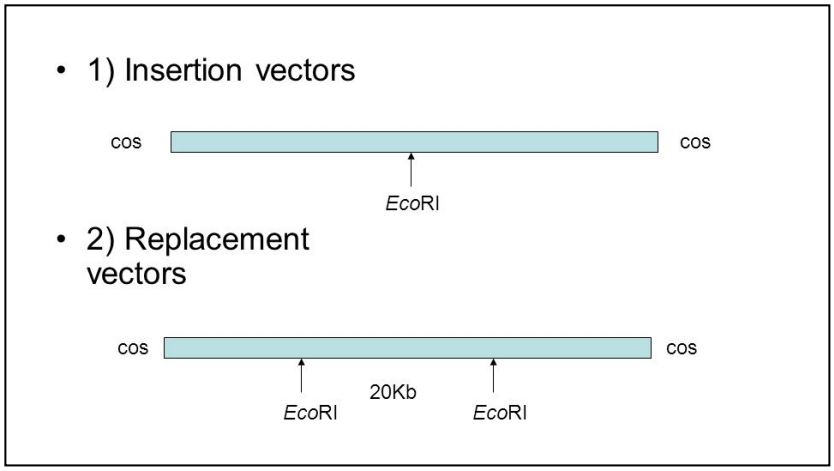
Additional advantage of using these modified phages as vectors is that they can enter bacteria much more easily than plasmids. One of the major limitations of λ vectors is that the head packaging imposes a physical limitation on the amount of DNA that can be incorporated into it during phage assembly. This restricts the size of foreign DNA fragments that can be cloned into λ phage. During packaging, infectious phage particles can be constituted from DNA that is approximately between 40 to 53 Kb long. Thus, a wild-type phage genome can accommodate only around 23 Kb of cloned foreign DNA. This drawback has been taken care of by careful construction of vectors to accommodate fragments of DNA that are close to the theoretical maximum limit for the particular vector.

λ vectors can be of two types: (1) insertion vectors and (2) replacement or substitution vectors.

λ Insertion vectors

As the name suggests, in these vectors no phage DNA is removed and foreign DNA fragments are inserted. They have a single recognition site for one or more restriction enzymes that cuts it only once. The foreign DNA fragments are inserted into this site in the genome. Since phage DNA is not removed there is a limit on the size of insert that can be cloned in this vector. They are generally utilized for construction of cDNA libraries from eukaryotic mRNA sequences. λt10 and Charon16A are examples of λ insertion vectors.

Fig 7 Insertion and Replacement vectors



λ Replacement or Substitution vectors

Insertion vectors have restricted scope for cloning large pieces of DNA. Therefore, replacement vectors were developed in which a middle dispensable 'filler' fragment is discarded and replaced with the insert DNA. These vectors contain restriction sites flanking the non-essential region of phage. They can accommodate larger inserts than insertion vectors. The middle region which is replaced by foreign DNA usually contains a gene that makes the phage non-viable in an appropriate bacterial host. Therefore vector molecules which do not contain foreign DNA can be selected against vector molecules containing foreign DNA, which will be viable. They are generally utilized for construction of genomic libraries. EMBL4 and Charon40 are examples of *λ* replacement vectors.

M13 Phage

M13 phage, a filamentous virus is 9 nm wide and 900 nm long and has a smaller genome than λ phage. Its genome is a single-stranded circular DNA. The 6.4 Kb single-stranded circular DNA of phage is encapsulated by 2,710 identical protein subunits.

Phage M13 can enter *E. coli* through the bacterial sex pilus, which is a protein appendage allowing the transfer of DNA between bacteria. Phage genome enters the bacterial cell where single stranded phage genome is converted into an intermediate circular double-stranded replicative form. This replicative form (RF) contains (+) and (-) strands and is essentially similar to plasmid. It acts as a template for replication of the single-stranded genome of the virus particle [the (+) strand] by rolling circle

replication. The + strand is then packaged into new virus particles. Thousands of progeny M13 phages are produced per generation.

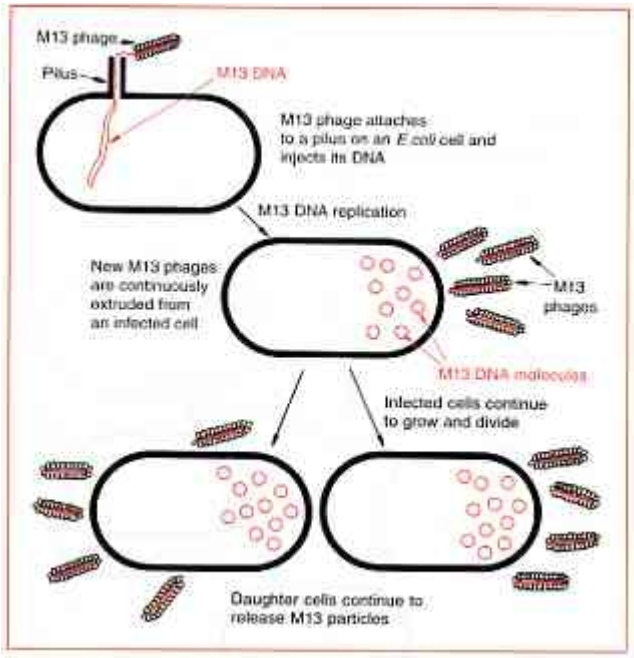


Fig: 8: M13 phage production

M13 phage is a non-lytic virus as it does not kill its bacterial host and therefore large quantities of M13 phages can be cultured and harvested easily. An M13 phage can be isolated and manipulated as a vector for cloning in a plasmid like manner. The circular double stranded replicative form DNA is digested at a single site by restriction endonucleases. The cut is made in the polylinker region of the phage. The double-stranded insert DNA fragment is produced by digestion with the same restriction enzyme. Insert DNA is then ligated to the digested vector. The foreign DNA can be inserted in any orientations as both ends of DNA molecules are same. Therefore, half of the newly synthesized (+) strands packaged into virus particles contain one of the strands of the insert DNA, while the other half will contain the other strand. Infection of *E. coli* by a single phage will produce large amount of single-stranded M13 DNA containing the same strand of the insert DNA. Instead of bacterial colonies, phage plaques will be formed and infected bacteria can be easily isolated from them. M13 phage harbors *lacZ* ' gene allowing blue/white selection system for screening of recombinants. Single-stranded DNA cloned into M13

phage can be subjected to sequence analysis.

The replicative form of M13 phage can be isolated and manipulated as a vector for cloning in a plasmid like manner. The use of vectors like M13, occurring in a single-stranded form is very useful in large scale sequencing projects since cloning, amplification and strand separation steps can be performed in combination. These are also used in labeling experiments, site-directed mutagenesis and phage display libraries where a foreign peptide is fused in frame to a gene coding for one of the coat proteins.

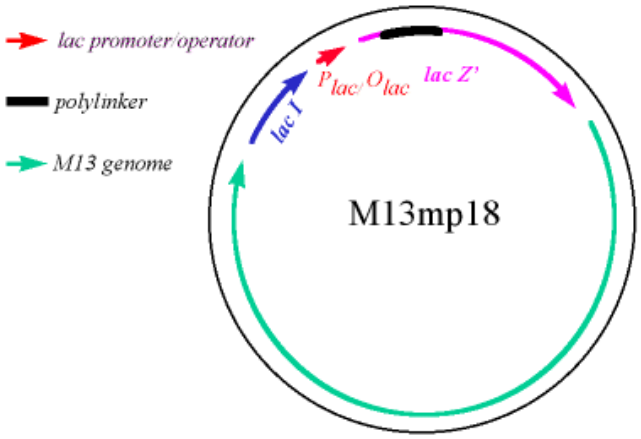


Fig: 9: M13mp18 phage production

Another advantage of M13 phage is that it contains multiple cloning sites which are similar to pUC18 and pUC19 so that genes cloned in pUC18 and pUC19 can be sub-cloned at the same sites in M13 phages.

M13 phages do not contain any non-essential genes. The 6.4 Kb genome of phage is used very efficiently and only 507 bp intergenic region is available for insertion of foreign DNA. Polylinker and *lacZ* peptide sequence are also inserted into this region.

M13 phages also have certain disadvantages like:

- These vectors are not suitable for long-term propagation of recombinant DNA as inserts longer than 1 Kb cannot be stably maintained.

- Though there are no packaging constraints but the cloning efficiency decreases when foreign DNA fragments longer than 1.5 Kb are inserted.

Phagemids/ Phasmids

Both plasmids and bacteriophages have certain advantages over each other as well as disadvantages. Therefore phagemids/phasmids were constructed to maintain advantages of both vectors. They are plasmid vectors with M13 origin of replication (pUC18 or pUC19 or pBR322 + M13 *ori*). They can be replicated in the host cells like plasmids maintaining double stranded replication and high copy number so that large amount of foreign DNA can be recovered. They can also package their single stranded DNA in phage particles because of the presence of M13 origin of replication. Phagemids can be introduced into bacterial host cells by transformation or electroporation. A positive selection marker is used to select bacteria containing the phagemid. Fragments of several kilobases of DNA in length can be isolated in single stranded form from the phagemids. They are also utilized for sequencing purposes and for generating templates for site-directed mutagenesis. λ ZAP is an example of phagemid.

Cosmids

The efficiency of cloning decreases in plasmids as fragment size increases and the length of non-essential region in λ phage vectors limits the length of inserted DNA in them. The requirement for accommodating larger fragments of genomic DNA led to the development of cosmid vectors by Collins and Hohn in the late 1970s. *In vitro* packaging in λ phage vectors is sequence independent, but is dependent on the *cos* sites separated by DNA of packagable size. This feature has been exploited for the construction of cosmid vectors.

The term cosmid includes ‘cos’ from cos sites and ‘mid’ from plasmid vectors as these are plasmid vectors containing cos sites of λ phage vectors. Cosmids are circular double-stranded DNA molecules containing a prokaryotic origin of replication, a selection marker, cloning sites for restriction endonucleases and λ cos sites. They are small (4-6 Kb) and can hold insert DNA fragments up to 47 Kb

in length. Cosmid vectors are used for cloning in a manner essentially similar to plasmid vectors and can clone larger DNA fragments than plasmid or phage. Insert DNA is ligated between two cos sites using restriction endonucleases.DNA is packaged *in vitro* and introduced into *E. coli*.

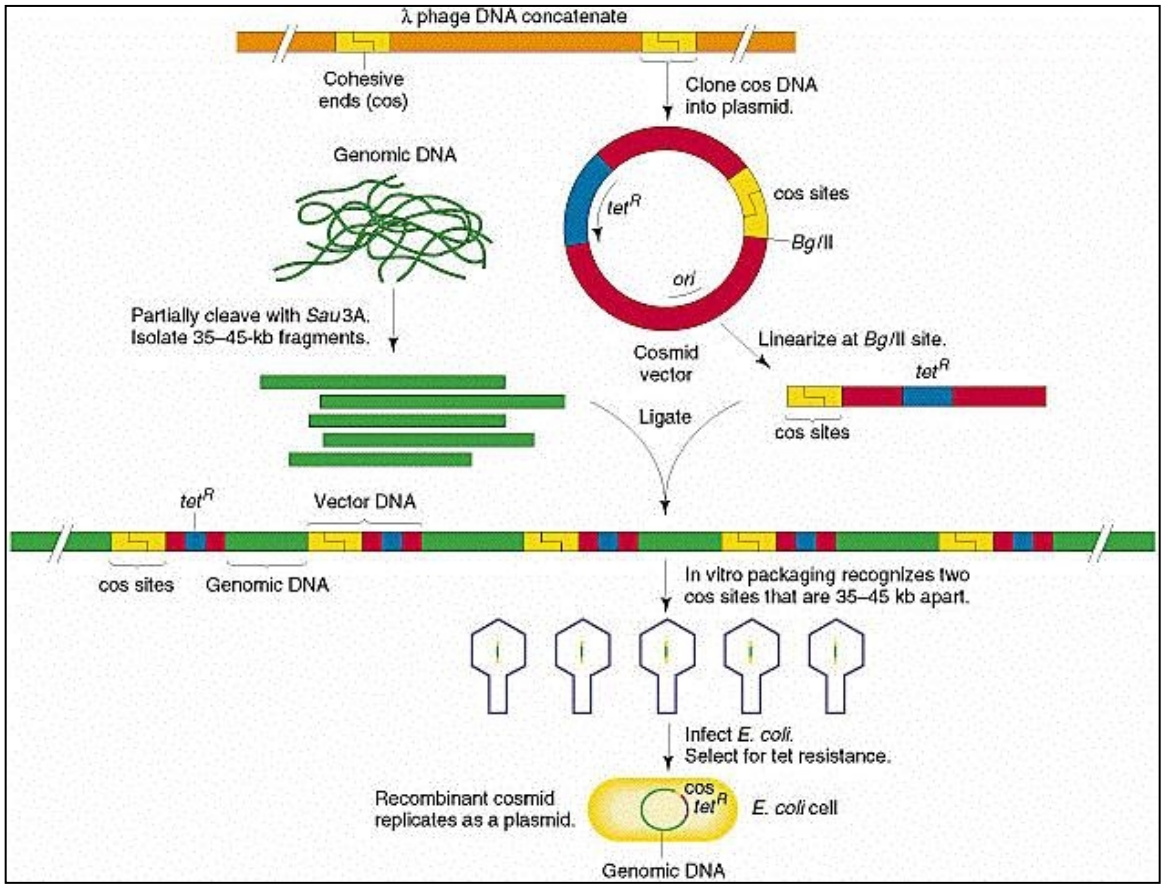


Fig: 10: Cosmids production

Since they lack phage genes, they act as plasmids when transferred into *E. coli.* Inside the bacterial cell, the cosmid circularizes and replicates as a plasmid maintaining 15-20 copies per cell. Yeast genome was sequenced using cosmid vectors. The cosmids require tedious working protocols for cloning and screening. Another disadvantage of cosmids is that their enormous insert size can result in recombination within the cloned sequences. But once desired recombinant clone is identified, isolation of DNA is easy.

Shuttle Vectors

They are plasmid vectors that can replicate autonomously in two different hosts. They are also known as

bifunctional vectors. Shuttle vectors are plasmids with the ability of replicating and transferring ("shuttling") between two different host organisms. They can also shuttle between a prokaryote *(E. coli)* and a eukaryote (yeast). These vectors possess two origins of replication, each origin unique for each host organism and different selection marker indicating transformed host cells containing the vector.

Eukaryotic genes can be cloned in bacterial hosts using the shuttle vectors and the expression of cloned genes can be analyzed in appropriate eukaryotic hosts. Initially cloning and amplification of inserted DNA is done in *E. coli* and then the shuttle vector is transferred to a eukaryotic host to study the expression of gene of interest. The most commonly used shuttle vector is yeast shuttle vector. These vectors are therefore used for

functional complementation assay of screening a yeast genomic library in which a functional protein complements a recessive mutation. Functional complementation is tested for yeast genes cloned in plasmid vector and therefore shuttle vectors are used to facilitate screening of recombinant plasmids. It contains essential elements to allow cloning of yeast DNA fragments in *E. coli*. Additionally, it contains origin for DNA replication in yeast (ARS, autonomously replicating sequence), yeast centromere (CEN) to allow faithful segregation of plasmid during yeast cell division and a yeast gene encoding a selectable marker URA3 for orotidine-5'phosphate decarboxylase, an enzyme which is required for the synthesis of uracil. Yeast gene sequences are partially cleaved to release overlapping restriction fragments. Shuttle vector is cleaved with the same restriction enzyme within the multiple cloning site to produce sticky ends complementary to the gene sequence. Vector is transformed to *E. coli* cells, and cells that grow after selection for ampicillin resistance contain single type of yeast cDNA fragment.

Yeast Artificial Chromosomes (YACs)

YACs are the largest capacity yeast vectors that can accommodate megabase range foreign DNA fragments within yeast cells. The capacity of YACs to harbor foreign DNA is much more (between 0.2 and 2.0 Mb) than other vectors discussed so far. A YAC vector includes a cloning region containing one or more restriction sites present only once in the vector, a yeast centromere sequence (*CEN*) to allow regulated segregation during mitosis, a yeast telomere (*TEL*) at each end of chromosome, yeast origin of replication sequence (*ARS* – autonomously replicating sequence) which allows the vector to propagate

in a yeast cell, a bacterial origin of replication (*ori*) which allows the circular version of the parental vector to replicate in *E. coli* and a selection marker for bacteria as well as yeast. Yeast ARS does not function in bacteria and bacterial *ori* will not function in yeast. Even, bacterial and eukaryotic promoters are variable, implying that the bacterial RNA polymerase is unable to transcribe the yeast selection genes namely *URA3* (involved in uracil biosynthesis) and *TRP1* gene (involved in tryptophan biosynthesis), and therefore yeast selection markers will function only in yeast. Similarly yeast RNA polymeraseII cannot transcribe the antibiotic resistance gene of bacteria. The circular YAC is digested with two restriction endonucleases, one cleaves in the multiple cloning site and other restriction enzyme cuts between the two *telomeres* producing right and left arms. Insert DNA is digested with the same restriction enzyme which was used to cleave the YAC multiple cloning site. Insert DNA is ligated to the two arms of YAC and the recombinant molecules are transformed into yeast. The transformants can be selected for selection markers to ensure that recombinant YACs contain both the left and right arms. They are helpful in creating physical maps of larger genomes like human genome. There are certain disadvantages of YAC vectors. Though the principle of cloning in YAC is similar to plasmid or cosmid, but the process of cloning is too complicated to carry out. It is difficult to isolate YAC DNA from host cell because of its similar size to host chromosome. Even higher yields of YAC DNA cannot be obtained from host cells.YAC vectors sometimes accept two or more different DNA fragments, creating a chimeric YAC. Additionally insert DNA is frequently modified or deleted by the host cell. YACs harbor very large sized insert, which is prone to breakage resulting in rearrangement and recombination with other DNA in the host cell. These alterations in the genome create problems during assembly of the genome.

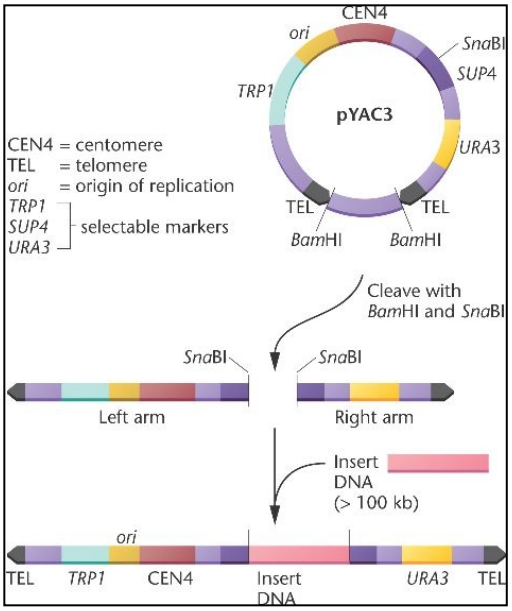


Fig: 11: YAC phage production

Bacterial Artificial Chromosomes (BACs)

Bacterial artificial chromosomes were created from natural plasmids of *E. coli* called the fertility factor (F factor). Studies on F factor revealed that these circular molecules are capable of replication in bacteria and can accommodate millions of base pairs. Hiroaki Shizuya turned parts of F factor into a vector in 1992. These vectors are customized for the cloning of very long segments (~ 100to 300 Kb) of DNA. They generally include a very stable origin of replication (*ori*) that maintains the plasmid at one or two copies per cell and selectable markers conferring resistance to the antibiotic chloramphenicol. BAC vectors were also modified to incorporate *lacZ* gene to allow easy identification of recombinants. Gene *sacB* encoding levansucrase was also incorporated into BAC vectors. The protein levansucrase turns sucrose into levan, a toxic product for bacteria. Bacteria cultured on media containing sucrose will die if sacB is intact as in non-recombinant cells. If the vector carries an insert DNA, sacB is disrupted and is unable to produce levansucrase, and the bacteria can survive in the media containing sucrose. The large circular recombinant DNAs are transferred to the host bacteria by electroporation. As the bacterial cell will grow, BACs will replicate and can be isolated. BACs are useful for cloning larger genes

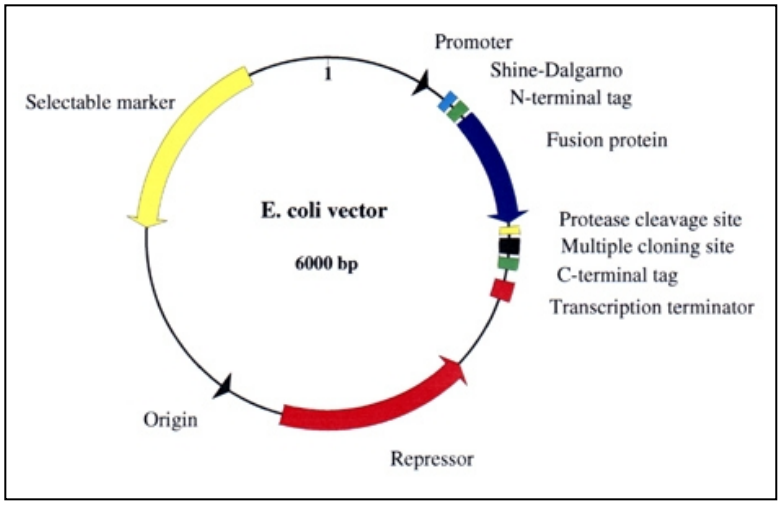
including the upstream regulatory regions controlling gene expression. These are used in genome sequencing projects to clone several genes together since several thousands of base pairs can be cloned into the BAC vector. BACs donor undergo recombination in the host cell and are helpful tools in physical mapping studies of genomes. These vectors are also utilized for cloning infectious DNA and RNA pathogens and development of vaccines. BACs are not suitable for cloning DNA sequences toxic to *E. coli* and AT-rich DNA fragments

Expression Vectors

Cloning vector is a small molecule of DNA that can carry foreign DNA into a host cell and is capable of self-replication. It does not necessarily help to express a protein which the inserted foreign DNA encodes in the cell. Expression vectors are constructed specifically to express the protein in the host cell.

Prokaryotic vectors are useful for producing recombinant heterologous proteins from cloned eukaryotic complementary DNAs. However, sometimes eukaryotic proteins synthesized in bacteria either lack biological activity or are unstable. Quite a number of times, bacterial compounds that are either toxic or cause increase in body temperature in humans and animals contaminate the final product. Many proteins require post–translational modifications to be functional and therefore, prokaryotic organisms are unable to produce authentic version of eukaryotic proteins.

Fig: 12: E.Coli vector production



To overcome these problems, researchers have developed eukaryotic expression system in yeast, insects, and mammalian cells, each with its own merits and demerits. These expression systems are useful for production of uncontaminated therapeutic agents for animals or humans, stable biologically active proteins for structural, biophysical and biochemical studies, and large quantities of proteins for industrial purposes. They have been used for production of insulin, enzymes, growth hormones, antibiotics, vaccines, antibodies. They are the basic tools of transgenic research to produce transgenic plants and animals for e.g. golden rice, insect resistant plants; glofish.The usage of expression system depends on the quality and quantity of recombinant protein that is produced, the cost of production and purification and ease of use. Unfortunately there is no single eukaryotic host cell that can produce functional protein from every cloned gene.

A basic eukaryotic expression vector system has all the elements of a cloning vector (an origin of replication, a selectable vector, multiple cloning site). Additionally it possesses a strong inducible eukaryotic promoter to drive the transcription of cloned gene, eukaryotic transcriptional stop signal, translational start sequence like ribosome binding site, Shine dalgarno sequence for expression in prokaryotes/ kozak consensus for expression in eukaryotes, a translation termination sequence, a sequence to add multiple adenines to mRNA (polyadenylation), a purification tag to ease out purification process. Common yeast offers several advantages for use as a host cell for expression of eukaryotic cloned genes. Firstly it is a single celled organism and its genetics and physiology are well understood. It can be easily cultured. Several tightly regulatable, strong inducible promoters have earlier been isolated from yeast. It is also capable of carrying out many post translational modifications. Yeast

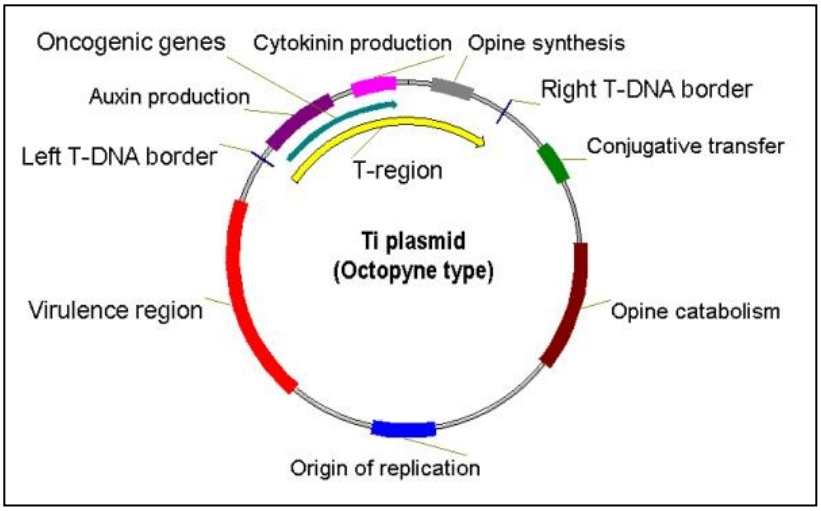
secretes very few proteins and therefore product protein can be readily purified. Lastly, it is recognized as safe organism by US food and drug administration. Many recombinant proteins have been produced by yeast expression system including vaccines (Hepatitis B virus surface antigen, HIV-1 envelope protein), human therapeutic agents (insulin, proinsulin, human growth factor, platelet derived growth factor, hirudin etc). However, with yeast as the host cell, recombinant heterologous proteins are hyperglycosylated and sometimes yields are low. To overcome these problems, methylotrophic yeast *Pichia pastoris* was developed because of the low occurrence of hyperglycosylation, the ease of obtaining higher yield and strong response of promoter.

Ti Plasmid

Ti (tumor-inducing) plasmid based vectors are high efficiency vectors and were developed with the aim of introducing genes into plants. *Agrobacterium tumefaciens* causes crown gall disease in plants in which cells grow in an undifferentiated and uncontrolled manner to form a tumor. This gram-negative soil bacterium possesses tumor-causing ability because of the presence of a large (206.479 Kb) double-stranded circular extra chromosomal element called as Ti plasmid.

It bears 196 genes including one structural RNA and encodes 195 proteins. This plasmid has a 20 Kb T-DNA region which can be transferred to the plant cell and is integrated into the plant DNA. It is also stably transmitted through divisions of meiosis and mitosis. This transfer of T-DNA from *Agrobacterium* to plant cells depends on 25 bp repeated sequences at both ends of the T-DNA known as left and right border repeats and on virulence (*vir*) genes which are grouped into operons. The T-DNA carries enzymes that convert plant metabolites to hormones cytokinin and auxin, which stimulate tumor formation. It also carries genes for opine synthesis. Opines serve as a food for the bacterium and are synthesized in higher concentrations in tumors cells. Researchers engineered this Ti plasmid by disrupting these tumor causing and opine synthesis genes and replaced them with selectable marker genes conferring resistance to antibiotics.

Fig: 13: Ti Plasmid production



Ti plasmid based vectors were constructed with the T-DNA carrying following components :(i)*ori* for origin of replication, allowing the plasmid to replicate; (ii)Right border sequence which is necessary for transfer of Ti plasmid into plant genome; (iii)A multiple cloning site to ease the insertion of gene of interest into the region between T-DNA border sequences.

Recombinant genes can then be integrated into the T-DNA of the Ti plasmid, and the plasmid can be used to infect plant cells. The infected cells are placed on a culture medium containing growth factors and the selectable antibiotic. Only the cells harboring T-DNA can grow in the presence of the antibiotic.

Plant viruses are non-integrative vectors:

The plant viruses do not integrate into the host genome in contrast to the vectors based on T-DNA of A. tumefaciens which are integrative. The viral genomes are suitably modified by introducing desired foreign genes. These recombinant viruses are transferred, multiplied and expressed in plant cells. They spread systemically within the host plant where the new genetic material is expressed.

Criteria for a plant virus vector:

An ideal plant virus for its effective use in gene transfer is expected to posses the following characteristics:

i. The virus must be capable of spreading from cell to cell through plasmodesmata.

ii. The viral genome should be able to replicate in the absence of viral coat protein and spread from cell to cell. This is desirable since the insertion of foreign DNA will make the viral genome too big to be packed.

iii. The recombinant viral genome must elicit little or no disease symptoms in the infected plants.

iv. The virus should have a broad host range.

v. The virus with DNA genome is preferred since the genetic manipulations involve plant DNA.

The three groups of viruses caulimoviruses, Gemini viruses and RNA viruses that are used as vectors for gene transfer in plants are briefly described.

Caulimoviruses as Vectors:

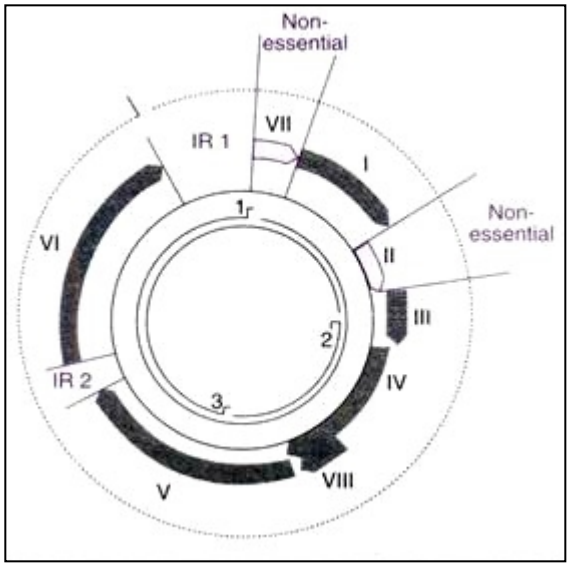
The caulimoviruses contain circular double- stranded DNA, and are spherical in shape. Caulimoviruses are widely distributed and are responsible for a number of economically important diseases in various crops. The caulimovirus group has around 15 viruses and among these cauliflower mosaic virus (CaMV) is the most important for gene transfer. The other caulimoviruses include carnation etched virus, dahlia mosaic virus, mirabilis mosaic virus and strawberry vein banding virus.

Cauliflower mosaic virus (CaMV):

CaMV infects many plants (e.g. members of Cruciferae, Datura) and can be easily transmitted, even mechanically. Another attractive feature of CaMV is that the infection is systemic, and large quantities of viruses are found in infected cells.

A diagrammatic view of the CaMV genetic map is depicted in Fig. 49.8. The genome of CaMV consists of a 8 kb (8024 bp) relaxed but tightly packed circular DNA with six major and two minor coding regions. The genes II and VII are not essential for viral infection.

Fig: 14: CaMV production



Use of CaMV in gene transfer:

For appropriate transmission of CaMV, the foreign DNA must be encapsulated in viral protein. Further, the newly inserted foreign DNA must not interfere with the native assembly of the virus. CaMV genome does not contain any non-coding regions wherein foreign DNA can be inserted. It is fortunate that two genes namely gene II and gene VII have no essential functions for the virus. It is therefore possible to replace one of them and insert the desired foreign gene.

Gene II of CaMV has been successfully replaced with a bacterial gene encoding dihydrofolate reductase that provides resistance to methotrexate. When the chimeric CaMV was transmitted to turnip plants, they were systemically infected and the plants developed resistance to methotrexate.

Limitations of CaMV as a vector:

i. CaMV vector has a limited capacity for insertion of foreign genes.

ii. Infective capacity of CaMV is lost if more than a few hundred nucleotides are introduced.

iii. Helper viruses cannot be used since the foreign DNA gets expelled and wild-type viruses are produced.

Gemini Viruses as Vectors:

The Gemini viruses are so named because they have geminate (Gemini literally means heavenly twins) morphological particles i.e. twin and paired capsid structures. These viruses are characterized by possessing one or two single-stranded circular DNAs (ss DNA). On replications, ss DNA forms an intermediate double-stranded DNA.

The Gemini viruses can infect a wide range of crop plants (monocotyledons and dicotyledons) which attract plant biotechnologists to employ these viruses for gene transfer. Curly top virus (CTV) and maize streak virus (MSV) and bean golden mosaic virus (BGMV) are among the important Gemini viruses.

It has been observed that a large number of replicative forms of a Gemini virus genome accumulate inside the nuclei of infected cells. The single-stranded genomic DNA replicates in the nucleus to form a double-stranded intermediate.

Gemini virus vectors can be used to deliver, amplify and express foreign genes in several plants/ explants (protoplasts, cultured cells). However, the serious drawback in employing Gemini viruses as vectors is that it is very difficult to introduce purified viral DNA into the plants. An alternate arrangement is to take the help of Agrobacterium and carry out gene transfer.

RNA Plant Viruses as Vectors:

There are mainly two types of single-stranded RNA viruses:

1. Mono-partite viruses:

These viruses are usually large and contain undivided genomes for all the genetic information e.g. tobacco mosaic virus (TMV).

2. Multipartite viruses:

The genome in these viruses is divided into small RNAs which may be in the same particle or different particles, e.g. brome mosaic virus (BMV). HMV contains four RNAs divided between three particles. Plant RNA viruses, in general, are characterized by high level of gene expression, good efficiency to infect cells and spread to different tissues. But the major limitation to use them as vectors is the difficulty of joining RNA molecules in vitro.

Use of cDNA for gene transfer:

Complementary DNA (cDNA) copies of RNA viruses are prepared in vitro. The cDNA so generated can be used as a vector for gene transfer in plants. This approach is tedious and cumbersome. However, some success has been reported. A gene sequence encoding chloramphenicol resistance (enzyme-chloramphenicol acetyltransferase) has been inserted into brome mosaic virus genome. This gene expression, however, has been confined to protoplasts.

Limitations of Viral Vectors in Gene Transfer:

The ultimate objective of gene transfer is to transmit the desired genes to subsequent generations. With virus vectors, this is not possible unless the virus is seed-transmitted. However, in case of vegetatively propagated plants, transmission of desired traits can be done e.g. potatoes. Even in these plants, there is always a risk for the transferred gene to be lost anytime. For the reasons referred above, plant biotechnologists prefer to insert the desired genes of interest into a plant chromosome.

Viruses have a natural ability to absorb to the surface of host cells and infect. This property can be exploited to deliver rDNA into animal cells. The viral system is efficient in transfer, expressed well and replicate rapidly in the host cells. Several classes of viral vectors have been developed for use in human gene therapy and at least eight have been used in clinical trials.

The General Properties of Viral Vector Are:

1. Trans-gene may be incorporated into viral vectors as additional gene or as replacement to certain genes of viral genome by ligation or homology recombination. If virus can propagate independently it is called helper independent. If essential viral genes are replaced by trans-gene then virus need a replication gene in Trans position and virus is called ‘helper-dependent’.

2. It is necessary to prevent replication as well as recombination of helper virus. Icosahedral viruses such as adenovirus and retroviruses package their genome into preformed capsid, their volume is fixed with defined amount of DNA can be packaged. Rod shaped baculo-viruses form the capsid around the genome, so there are no such size constraints. There is no ideal virus, each has his own advantages or disadvantages.

Adenoviruses are DNA viruses with a linear, double stranded genome of approx. 36 kb. They are used in gene transfer because they show some advantageous features like stability, a high capacity for foreign DNA, a wide host range that includes non dividing cells and the ability to produce high titer stock (up to 10^{11} pfu/ml). They are suitable for transient expression in dividing cells because they do not integrate efficiently into the genome.

Adenoviruses are also used as gene therapy vectors because the virions are taken up efficiently by cells in vivo and adenovirus derived vaccines have been used in humans with no reported side effects.

Baculo-viruses have large double stranded DNA genomes. They efficiently infect arthropods, particularly insects. Nuclear polyhedrosis viruses, a group of baculoviruses, have an unusual infection cycle that involves the production of nuclear occlusion bodies. These are pro-teinaceous particles in which the virions are embedded allowing the virus to survive harsh environmental conditions such as desiccation.

Baculo-viruses are mainly used for high level transient protein expression in insects and insect cells. Two baculoviruses have been extensively developed as vectors, namely the *Autographa calofornica* multiple nuclear polyhedrosis virus and the *Bombyx mori* nuclear polyhedrosis virus.

UNIT – 3 – GENETIC ENGINEERING – SBTA1501

Selection and Screening of recombinants

Library screening is the process of identification of the clones carrying the gene of interest. Screening relies on a unique property of a clone in a library. The DNA libraries consist of a collection of probably many thousand clones in the form of either plaques or colonies on a plate.

Screening of libraries can be done by following approaches based on-

- Detecting a particular DNA sequence and
- Gene expression.

Insertional Inactivation

The inactivation of marker gene function in vector due the insertion of foreign gene.

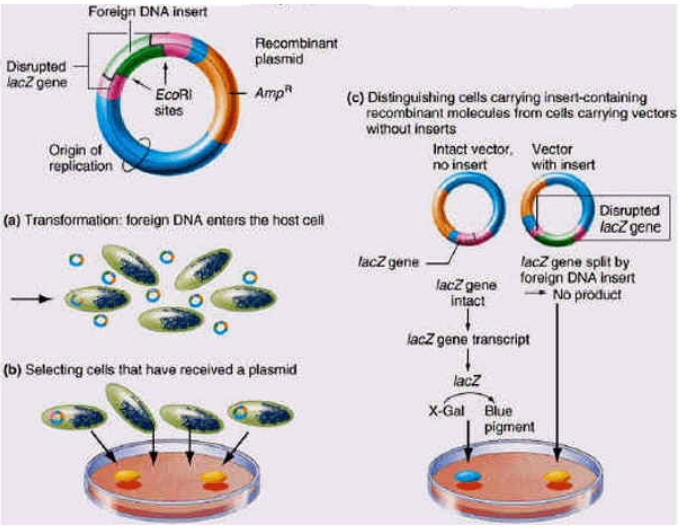


Fig : 1 Screening by hybridization

- Nucleic acid hybridization is the most commonly used method of library screening first developed by Grunstein and Hogness in1975 to detect DNA sequences in transformed colonies using radioactive RNA probes.
- It relies on the fact that a single-stranded DNA molecule, used as a probe can hybridize to its complementary sequence and identify the specific sequences.
- This method is quick, can handle a very large number of clones and used in the identification of cDNA clones which are not full-length (and therefore cannot be expressed).

The commonly used methods of hybridization are,

a) Colony hybridization
b) Plaque hybridization.

Colony hybridization

Colony hybridization is an application of nucleic acid hybridization that is combined with conventional environmental microbiological sampling and viable planting procedures. This process utilizes a gene probes, which is radio actively marker, which attaches to complimentary base pairs from a single strand of bacteria DNA. This has proven to be a useful procedure for efficiently screening bacteria clones by

RNA-DNA or DNA-DNA hybridization .

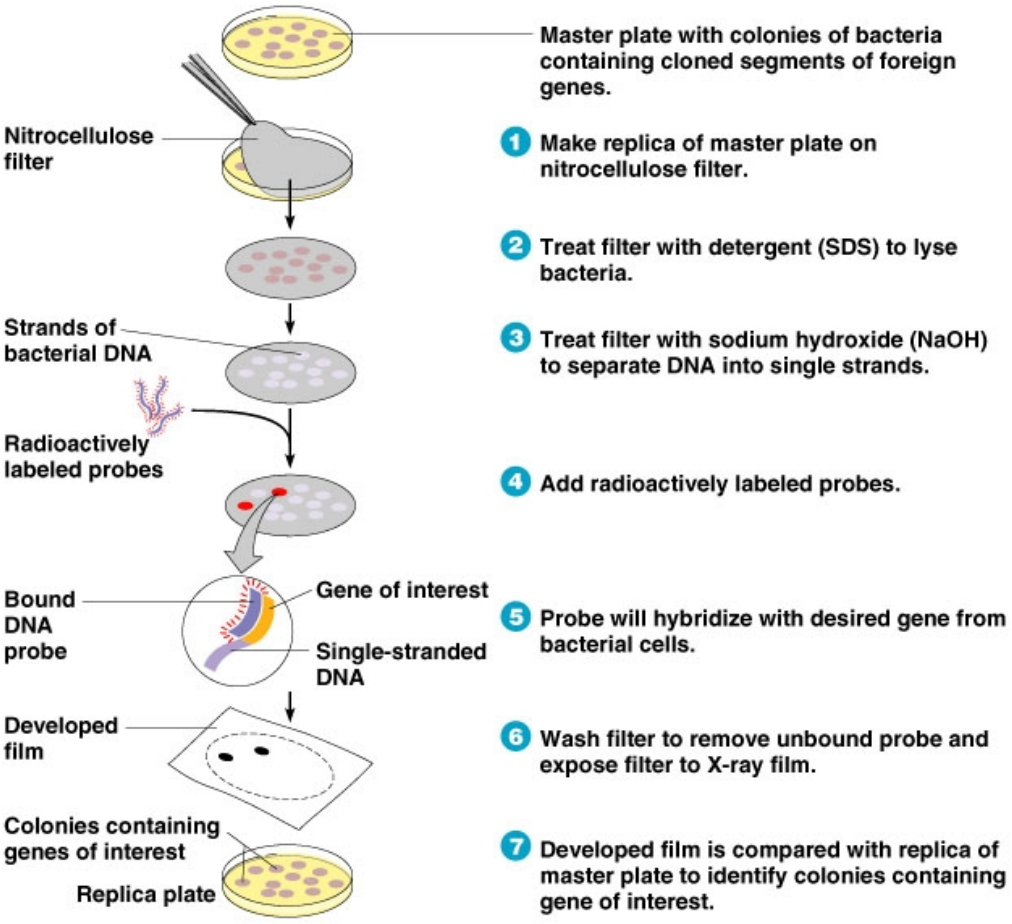


Fig : 2 colony Hybridization

Plaque Hybridization

Plaque hybridization, also known as *Plaque lift*, was developed by Benton and Davis in 1977 and employs a filter lift method applied to phage plaques. This procedure is successfully applied to the isolation of recombinant phage by nucleic acid hybridization and probably is the most widely applied method of library screening. The method of screening library by plaque hybridization is described below-

- The nitrocellulose filter is applied to the upper surface of agar plates, making a direct contact

between plaques and filter.

- The plaques contain phage particles, as well as a considerable amount of unpackaged recombinant DNA which bind to the filter.
- The DNA is denatured, fixed to the filter, hybridized with radioactive probes and assayed by autoradiography.

Advantages

- This method results in a 'cleaner' background and distinct signal (less background probe hybridization) for λ plaque screening due to less DNA transfer from the bacterial host to the nitrocellulose membrane while lifting plaques rather than bacterial colonies.
- Multiple screens can be performed from the same plate as plaques can be lifted several times

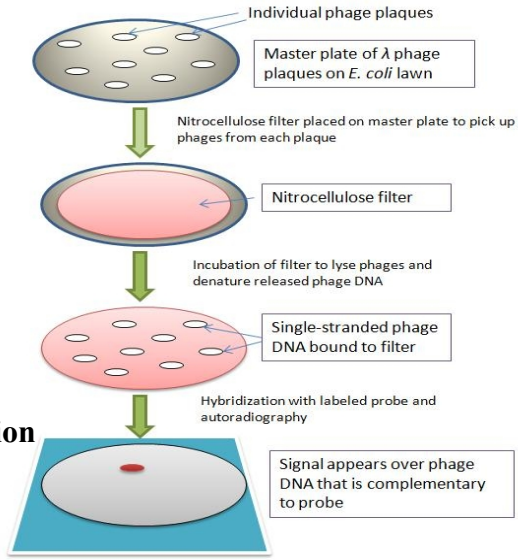


Fig : 3 Plaque Hybridization

Nucleic acid hybridization is a basic technique in molecular biology which takes advantage of the ability of individual single-stranded nucleic acid molecules to form double-stranded molecules.

Southern hybridization

The basic principle behind the southern hybridization is the nucleic acid hybridization. Southern hybridization commonly known as southern blot is a technique employed for detection of a specific DNA sequence in DNA samples that are complementary to a given RNA or DNA sequence. It was the first blotting technique to be devised, named after its pioneer E.M Southern, a British biologist. Southern blotting involves separation of restricted DNA fragments by electrophoresis and then transferred to a nitrocellulose or a nylon membrane, followed by detection of the fragment using probe hybridization.

Separated by electrophoresis is transferred from gel to a membrane which in turn is used as a substrate for hybridization analysis employing labeled DNA or RNA probes specific to target fragments in the blotted DNA. Southern hybridization helps to detect specific fragment against a background of many other restriction fragments. Southern blotting is a technique which is used to confirm the identity of a cloned fragment or for recognition of a sub-fragment of interest from within the cloned DNA, or a genomic DNA. Southern blotting is a prerequisite to techniques such as restriction fragment length polymorphism (RFLP) analysis.

Procedure:

1. The high-molecular-weight DNA strands are fractioned using restriction enzymes.

2. The DNA fragments are separated based on size by agarose gel electrophoresis.

3. The gel with the restricted fragments is then laid on a filter paper wick which serves as a connection between the membrane and the high salt buffer.

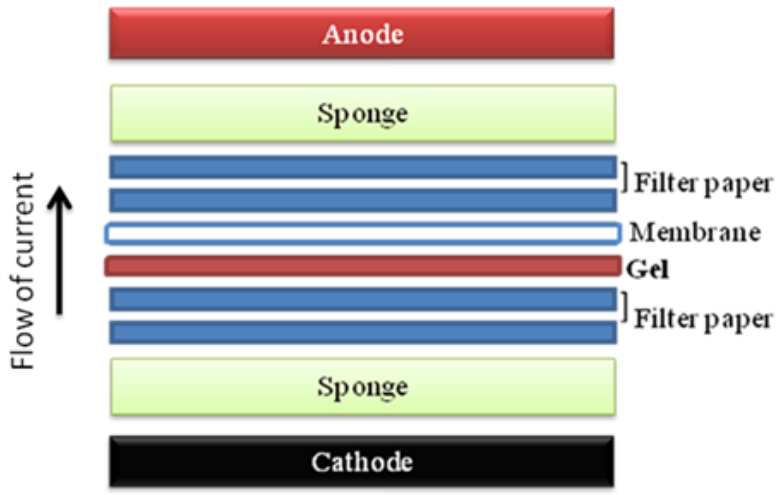
4. The nitrocellulose membrane is placed on top of the gel and a tower of filter papers is used to cover it and these are kept in place with a weight. The capillary action drives the buffer soaking through the filter paper wick, through the gel and the membrane and into the paper towels. Along with the buffer passing through the gel the DNA fragments are also carried with it into the membrane and they bind to

the membrane. This causes an effective transfer of fragments (up to 15 kb in length taking around 18hours or overnight.

For DNA fragments larger than 15 kb, before blotting an acid such as diluted HCl is used to treat the gel that depurinates the DNA fragments causing breakage of DNA into smaller pieces, resulting in more efficient transfer from the gel to membrane.

Now a day's blotting is also done by applying electric field. This **electro blotting** technique depends upon current and transfer buffer solution to nucleic acids onto a membrane. Following electrophoresis, a standard tank or semi-dry blotting transfer system is set up. A stack is put together in the following order from cathode to anode: sponge, three sheets of filter paper soaked in transfer buffer gel, PVDF or nitrocellulose membrane, three sheets of filter paper soaked in transfer buffer and then again sponge. Importantly the membrane should be located between the gel and the positively-charged anode, as the current and sample will be moving in that direction. Once the stack is prepared, it is placed in the transfer system, and suitable current is applied for a specific period of time according to the materials being used.

Figur 4; Electroblotting



5. For using alkaline transfer methods, the DNA gel is placed into an alkaline solution (like that of sodium hydroxide) causing denaturation of the double-stranded DNA. Denaturation in an alkaline environment enhances the binding between the negatively charged DNA and the positively charged

membrane, causing separation to single DNA strands for further hybridization to the probe, alongside destroying any residual RNA that may persist in DNA. The membrane is washed with buffer to remove unbound DNA fragments.

6. The membrane which contains the transferred fragments is heated in presence or absence of vacuum at 80°C for 2 hours or exposed to ultraviolet radiation (nylon membrane) for permanent attachment of the transferred DNA to the membrane.

7. The obtained membrane is then hybridized with a probe (a DNA fragment with a specific sequence whose presence in the target DNA is to be determined).

8. Labeling of the probe DNA is done for easy detection, usually radioactivity is incorporated or the molecule is tagged with a fluorescent or chromogenic dye. The hybridization probe may be made of RNA, instead of DNA in some cases where the target is RNA specific.

9. Washing of the excess probe from the membrane is done by SSC buffer after the hybridization step and the hybridization pattern is studied on an X-ray film by autoradiography or via color development on membrane if a chromogenic detection method is employed.

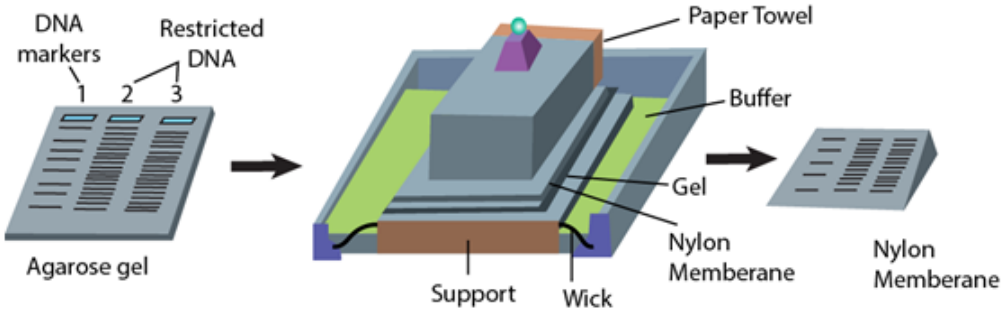


Fig 5 Steps of Southern Hybridization

Analysis of Southern Blot:

Hybridization of the probe to a specific DNA fragment on the membrane indicates the presence of a complementary fragment in the DNA sequence. Southern hybridization performed by digestion of genomic DNA using a restriction enzyme digestion, helps in determining the number of sequences (or gene copies) in the genome. For a probe hybridizing to a single DNA segment that has not been cut by

the restriction enzyme, a single band is observed and on the other hand multiple bands will likely be observed when the hybridization occurs between the probe and several highly similar target sequence (Due to sequence duplication). Alterations in the hybridization conditions like enhancing the hybridization temperature or decreasing salt concentration, helps in altering specificity and hybridization of the probe to sequences that are less than 100% similar.

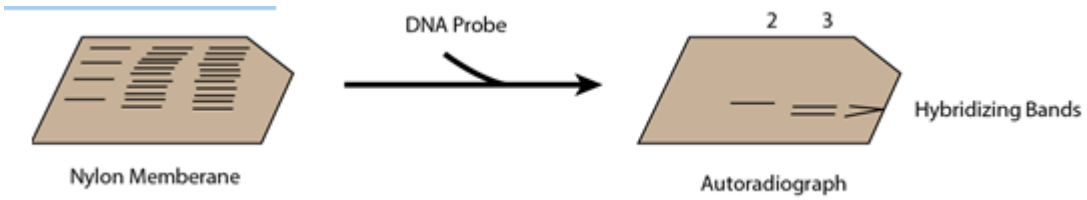


Fig 6 Southern hybridization analysis

Applications:

a) Clone identification: One of the most common applications of Southern blotting is identification and cloning of a specific gene of interest. Southern blotting is carried out for identification of one or more restriction fragments that contain the gene of interest in genomic DNA.. After cloning and tentative identification of the desired recombinant by employing colony or plaque hybridization, southern blotting is further is used to confirm the clone identification and possibly to locate a shorter restriction fragment, containing the sequence of interest.

b) Restriction fragment length polymorphism Analysis: Another major application of Southern hybridization is restriction fragment length polymorphism (RFLP) mapping, which is crucial in construction of genome maps.

Northern hybridization:

Northern blotting was developed by James Alwine, George Stark and David Kemp (1977). Northern blotting drives its name because of its similarity to the first blotting technique, which is Southern blotting, named after the biologist Edwin Southern. The major difference is that RNA being analyzed rather than DNA in the northern blot.

Expression of a particular gene can be detected by estimating the corresponding mRNA by Northern blotting. Northern blotting is a technique where RNA fragments are separated by electrophoresis and immobilized on a paper sheet.Identification of a specific RNA is then done by hybridization using a labeled nucleic acid probe. It helps to study gene expression by detection of RNA (or isolated mRNA) in a sample.

In Northern blotting, probes formed of nucleic acids with a sequence which is complementary to the sequence or to a part of the RNA of interest. The probe can be DNA, RNA or chemically synthesized oligonucleotides of minimum 25 complementary bases to the target sequence.

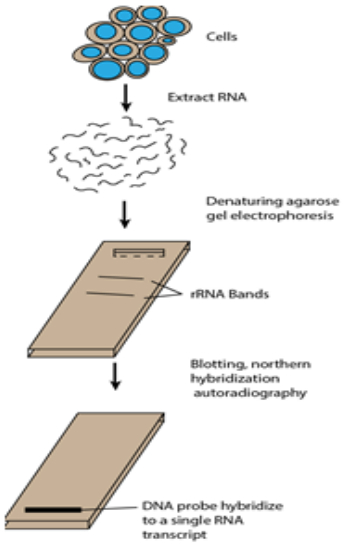


Fig 7 Steps of Northern Hybridization

Procedure:

The northern blotting involves the following steps:

1. Total RNA is extracted from a homogenized tissue sample or cells. Further eukaryotic mRNA can then be isolated by using of oligo (dT) cellulose chromatography to isolate only those RNAs by making use of a poly A tail.

2. The isolated RNA is then separated by gel electrophoresis.

3. The RNA samples separated on the basis of size are transferred to a nylon membrane employing a capillary or vacuum based system for blotting.

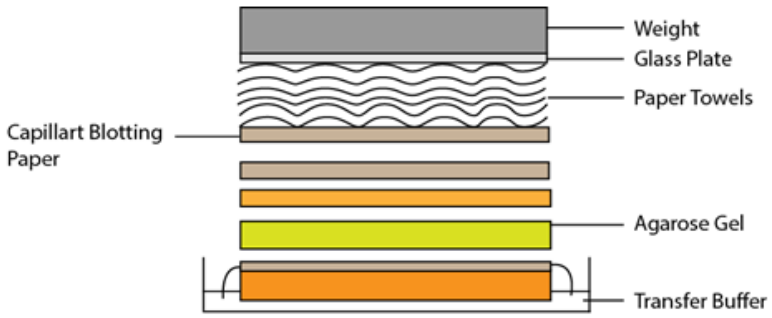


Fig 8 Setup for Northern blotting

4. Similar to Southern blotting, the membrane filter is revealed to a labeled DNA probe that is complementary to the gene of interest and binds.

5. The labeled filter is then subjected to autoradiography for detection.

The net amount of a specific RNA in a sample can be estimated by using Northern blot. This technique is widely used for comparing the amounts of a particular mRNA in cells under different conditions. The separation of RNA samples is often done on agarose gels containing formaldehyde as a denaturing agent as it limits the RNA to form secondary structure.

Analysis of Northern Blot:

RNA extract is electrophoresed in an agarose gel, using a denaturing electrophoresis buffer (containing formaldehyde) to ensure that the RNAs do not form inter- or intra-molecular base pairs, as base pairing would affect the rate at which the molecules migrate through the gel. After electrophoresis, the gel is blotted onto a nylon or nitrocellulose membrane, and hybridized with a labeled probe. If the probe is a cloned gene, the band that appears in the autoradiograph is the transcript of that gene. The size of the transcript can be determined from its position within the gel, and if RNA from different tissues is run in different lanes of the gel, then the possibility of differentially expressed gene can be examined.

Applications:

Northern blotting helps in studying gene expression pattern of various tissues, organs, developmental stages, pathogen infection, and also over the course of treatment.

Northern blotting is also used for the analysis of alternate spliced products of same gene or repetitive sequence motif by investigating the various sized RNA of the gene. This is done when only probe type with variation in one location is used to bind to the target RNA molecule.

Variations in size of a gene product may also help to identify deletions or errors in transcript processing, by altering the probe target that can be used along the known sequence and make it possible to determine the missing region of the RNA.

Dot Blot and Slot Blot Hybridization

These two techniques represents the simplification of Southern and Western blots saving the time involved in procedures of chromatography, electrophoresis, restriction digestion and blotting of DNA or proteins from the gel to membrane. Here nucleic acid mixture is directly applied (blotted) on to the nylon or nitrocellulose membrane where hybridization between probe and target takes place, denatured to single-stranded form and baked at 80°C to bind DNA target to membrane. In dot-blot, target is blotted as circular blots whereas in slot-blots, it is in the form of rectangular blots. Due to this, slot-blot offers greater precision in observing different hybridization signals. After blotting, membrane is allowed to dry and non-specific sites are blocked by soaking in blocking buffer containing BSA. It is then followed by hybridization of labeled probe for detection of specific sequences or gene.

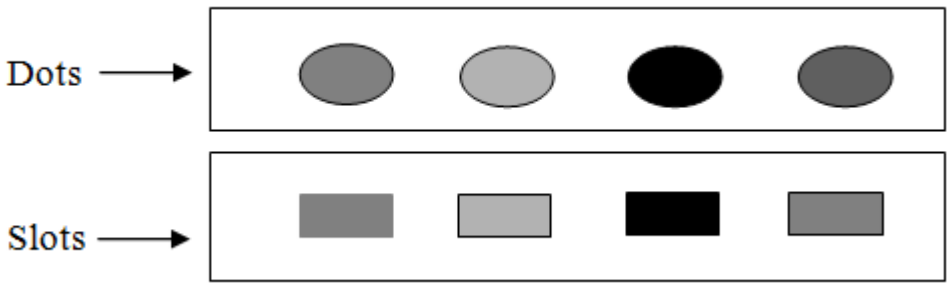


Fig 9. Slot Blot steps

Dots and slots in dot and slot blot hybridization

These procedures can only detect presence and absence of particular sequence or gene. It cannot distinguish between two molecules of different sizes as they appear as single dot on membrane. It also has application in detecting alleles that differ in single nucleotide with the help of allele-specific oligonucleotides.

Western Blotting

The first step in a Western blotting procedure is to separate the macromolecules using gel electrophoresis. After electrophoresis, the separated molecules are transferred or blotted onto a second matrix, generally a nitrocellulose or polyvinylidene difluoride (PVDF) membrane. Next, the membrane is blocked to prevent any nonspecific binding of antibodies to the surface of the membrane. Most commonly, the transferred protein is complexed with an enzyme-labeled antibody as a probe. An appropriate substrate is then added to the enzyme and together they produce a detectable product such as a chromogenic precipitate on the membrane for colorimetric detection. The most sensitive detection methods use a chemiluminescent substrate that, when combined with the enzyme, produces light as a byproduct. The light output can be captured using film, a CCD camera or a phosphorimager that is designed for chemiluminescent detection. Alternatively, fluorescently tagged antibodies can be used, which are directly detected with the aid of a fluorescence imaging system. Whatever system is used, the intensity of the signal should correlate with the abundance of the antigen on the membrane.

Proteins are commonly separated using polyacrylamide gel electrophoresis (PAGE) to characterize individual proteins in a complex sample or to examine multiple proteins within a single sample. When combined with Western blotting, PAGE is a powerful analytical tool providing information on the mass, charge, purity or presence of a protein. Several forms of PAGE exist and can provide different types of information about the protein(s)

Following electrophoresis, the protein must be transferred from the electrophoresis gel to a membrane. There are a variety of methods that have been used for this process, including diffusion transfer, capillary transfer, heat-accelerated convectional transfer, vacuum blotting transfer and electroelution. The transfer method that is most commonly used for proteins is electroelution or electrophoretic transfer because of its speed and transfer efficiency. This method uses the electrophoretic mobility of proteins to

transfer them from the gel to the membrane. Electrophoretic transfer of proteins involves placing a protein-containing polyacrylamide gel in direct contact with a piece of nitrocellulose or other suitable, protein-binding support and "sandwiching" this between two electrodes submerged in a conducting solution. When an electric field is applied, the proteins move out of the polyacrylamide gel and onto the surface of the membrane, where the proteins become tightly attached. The result is a membrane with a copy of the protein pattern that was originally in the polyacrylamide gel.

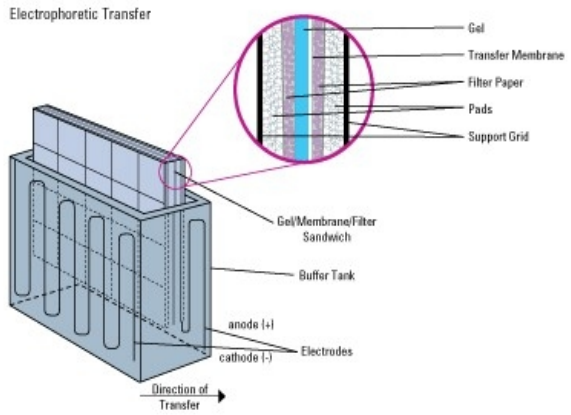


Fig 10. Western Blot apparatus

Transfer efficiency can vary dramatically among proteins, based upon the ability of a protein to migrate out of the gel and its propensity to bind to the membrane under a particular set of conditions. The efficiency of transfer depends on factors such as the composition of the gel, complete contact of the gel with the membrane, the position of the electrodes, the transfer time, size and composition of proteins, field strength and the presence of detergents and alcohol in the buffer.

After transfer and before proceeding with the Western blot, total protein on the membrane is often stained with a dye, such as Ponceau S or amido black 10B, to check the transfer efficiency; the gel may also be stained to confirm that protein has been moved out of the gel, but this does not ensure efficient binding of protein to the membrane. Because dyes may interfere with antibody binding and detection, a protein stain that is easily removable is ideal. Ponceau S stain is the most widely used reagent for reversibly staining proteins on a membrane, although it has limited sensitivity, does not photograph well and fades quickly, which makes documentation difficult. Superior alternatives for staining protein on

nitrocellulose or PVDF membranes are available, which allow the detection of low-nanogram levels of protein, are easily photographed and do not fade until removed.

Blocking Nonspecific Sites

The membrane supports used in Western blotting have a high affinity for proteins. Therefore, after the transfer of the proteins from the gel, it is important to block the remaining surface of the membrane to prevent nonspecific binding of the detection antibodies during subsequent steps. A variety of blocking buffers ranging from milk or normal serum to highly purified proteins have been used to block free sites on a membrane.

Primary and Secondary Antibodies

Although other methods are used, Western blotting is typically performed by probing the blocked membrane with a primary antibody that recognizes a specific protein or epitope on a group of proteins (e.g., SH2 domain or phosphorylated tyrosine).

In general, the primary antibody which recognizes the target protein in a Western blot is not directly detectable. Therefore, tagged secondary antibodies or other detection reagents are used as the means of ultimately detecting the target antigen (indirect detection). A wide variety of labeled secondary detection reagents can be used for Western blot detection. The choice of which depends upon either the species of animal in which the primary antibody was raised (the host species) or any tag on that antibody (i.e., biotin or DIG). For example, if the primary antibody is an unmodified mouse monoclonal antibody then the secondary antibody must be an anti-mouse IgG secondary antibody obtained from a non-mouse host.

Detection Methods

Enzymatic labels are most commonly used for Western blotting and, although they require extra steps, can be extremely sensitive when optimized with an appropriate substrate. Alkaline phosphatase (AP) and horseradish peroxidase (HRP) are the two enzymes used most extensively as labels for protein detection. An array of chromogenic, fluorogenic and chemiluminescent substrates are available for use with either enzyme. Alkaline phosphatase offers a distinct advantage over other enzymes in that its

reaction rate remains linear allowing sensitivity to be improved by simply allowing a reaction to proceed for a longer time period. Unfortunately, the increased reaction time often leads to high background signal resulting in low signal:noise ratios. Horseradish peroxidase (HRP) conjugated antibodies are considered superior to antibody-AP conjugates with respect to the specific activities of both the enzyme and antibody due the smaller size of HRP enzyme and compatibility with conjugation reactions. In addition, the high activity rates, good stability, low cost and wide availability of substrates make HRP the enzyme of choice for most applications.

Screening methods based on gene expression

Immunological screening

This involves the use of antibodies that specifically recognize antigenic determinants on the polypeptide. It does not rely upon any particular function of the expressed foreign protein, but requires an antibody specific to the protein.

Earlier immunoscreening methods employed radio-labeled primary antibodies to detect antibody binding to the nitrocellulose sheet. It is now superseded by antibody sandwiches resulting in highly amplified signals. The secondary antibody recognizes the constant region of the primary antibody and is, additionally, conjugated to an easily assayable enzyme (*e.g.* horseradish peroxidase or alkaline phosphatase) which can be assayed using colorimetric change or emission of light using X-ray film .

- In this technique, the cells are grown as colonies on master plates and transferred to a solid matrix.
- These colonies are subjected to lysis releasing the proteins which bind to the matrix.
- These proteins are treated with a primary antibody which specifically binds to the protein (acts as antigen), encoded by the target DNA. The unbound antibodies are removed by washing.
- A secondary antibody is added which specifically binds to the primary antibody removing the unbound antibodies by washing.
- The secondary antibody carries an enzyme label (e.g., horse radishperoxidase or alkaline phosphatase) bound to it which converts colorless substrate to colored product. The colonies

with positive results (i.e. colored spots) are identified and subcultured from the master plate.

The main difficulty with antibody-based screening is to raise a specific antibody for each protein to be detected by injecting a foreign protein or peptide into an animal. This is a lengthy and costly procedure and can only be carried out successfully with proteins produced in reasonably large amounts.

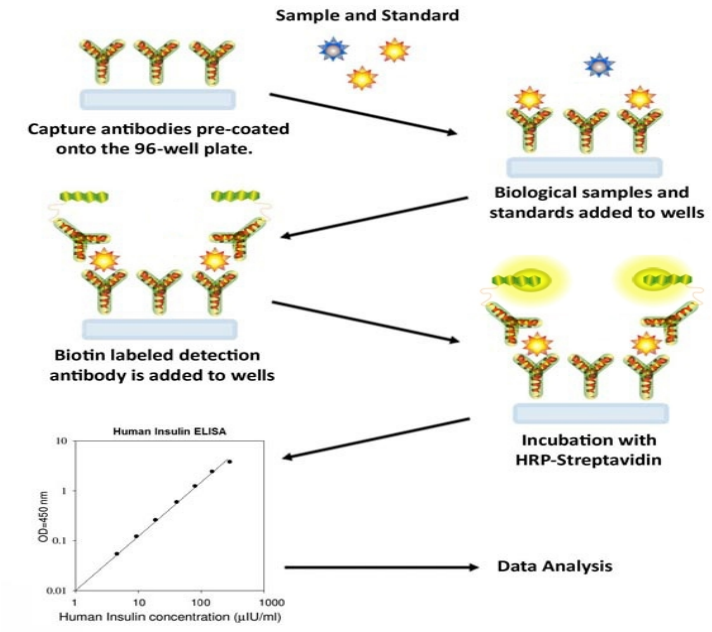


Fig 11.Immunological screening

Number of clones required for a library

The number of clones to be pooled depends upon the size of the genome *f* and average size of the cloned DNA.

Let (*f*) be the fraction of the genome size compared to the average individual cloned fragment size, would represent the lowest possible number of clones that the library must contain.

The minimum number of clones required can be calculated as-

***f*= genome size/ fragment size**

For the *E. coli* genome (4.6 Mb) with an average cloned fragment size of 5 kb, *f* will be 920.

The number of independent recombinants required in the library must be greater than *f,* as sampling variation leads to the several times inclusion and exclusion of some sequences in a library of just *f* recombinants. In 1976, Clarke and Carbon derived a formula to calculate probability (P) of including any DNA sequence in a random library of N independent recombinants.

The actual number of clones required can be calculated as-

$$N = \frac{\ln (1-P)}{\ln (1-f)}$$

where N= number of clones and P= probability that a given gene will be present.

Bigger the library better will be the chance of finding the gene of interest. The pooling together of either recombinant plaques or bacterial colonies generates a primary library.

4-4.3. Amplified library

- The primary library created is usually of a low titer and unstable. The stability and titer can be increased by amplification. For this, the phages or bacterial colonies are plated out several times and the resulting progenies are collected to form an amplified library.
- The amplified library can then be stored almost indefinitely due to longshelf-life of phages.
- It usually has a much larger volume than the primary library, and consequently may be screened several times.
- It is possible that the amplification process will result in the composition of the amplified library not truly reflecting the primary one.
- Certain DNA sequences may be relatively toxic to *E. coli* cells. As a consequence bacteria harboring such clones will grow more slowly than other bacteria harboring non-toxic DNA sequences. Such problematic DNA sequences present in the primary library may be lost or

under-represented after the growth phase required to produce the amplified library.

Subgenomic library

Subgenomic library is a library which represents only a fraction of the genome. Enhancing the fold of purification of target DNA is crucial for subgenomic DNA libraries which can be achieved by multiple, sequential digestion when information of the restriction map of the sequences of interest is known. After initial purification of a given fragment, the purification can further be increased by redigestion with another enzyme generating a smaller (clonable) fragment relative to original DNA.

Advantages of genomic libraries

- Identification of a clone encoding a particular gene of interest.
- It is useful for prokaryotic organisms having relatively small genomes.
- Genomic libraries from eukaryotic organisms are very important to study the genome sequence of a particular gene, including its regulatory sequences and its pattern of introns and exons.

Disadvantages of genomic library

- Genome libraries from eukaryotes having very large genomes contain a lot of DNA which does not code for proteins and also contain non-coding DNA such as repetitive DNA and regulatory regions which makes them less than ideal.
- Genomic library from a eukaryotic organism will not work if the screening method requires the expression of a gene.

Applications

- To determine the complete genome sequence of a given organism.
- To study and generate transgenic animals through genetic engineering, serving as a source of genomic sequence.
- To study the function of regulatory sequences *in vitro.*
- To study the genetic mutations.
- Used for genome mapping, sequencing and the assembly of clone contigs.

Comparison of Genomic and cDNA Libraries

cDNA library has revolutionized the field of molecular genetics and recombinant DNA technology. It consists of a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism is represented as its cDNA insertion in a vector. cDNA libraries are used to express eukaryotic genes in prokaryotes. In addition, cDNAs are used to generate expressed sequence tags (ESTs) and splices variant analysis. Some of the differences of cDNA library with genomic library are presented in Table :1

Feature	Genomic library	cDNA library
Sequences present	Ideally, all genomic sequences	Only structural genes that are Transcribed
Contents affected by :		
(a) Developmental stage	No	Yes
(b) Cell type	No	Yes
Features of DNA insert(s) representing a gene:		
(a) Size	As present in genome	Ordinarily, much smaller
(b) Introns	Present	Absent
(c) 5′- and 3′- regulatory sequences	Present	Absent
As compared to the genome		
(a) Enrichment of sequences	In amplified genomic libraries	For abundant mRNAs
(b) Redundancy in frequency	In amplified libraries	For rare mRNA species
(c) Variant forms of a gene	Not possible	For such genes, whose RNA transcripts are alternatively Spliced

Features of genomic and cDNA library

Recombinant Protein -What is it?

Expression of a gene from one organism expressed in another organism.

Heterologous Gene/Protein expression

Often times but not always “tagged” with extra amino acids to identify the foreign protein and to assist in its purification ü Proteins used in research, clinical, and agriculture environments

100 µg of a purified expressed protein can sell for $500 to $2000. ü Insulin, interferon, theraputic antibodies are all examples Creating a Recombinant Protein Ø Recombine gene from one organism with the plasmid expression vector for another organism (host system) ü i.e. cloned gene from human is cut into DNA to be transformed into bacteria

What is a plasmid

At their most basic level, plasmids are small circular pieces of DNA that replicate independently from the host's chromosomal DNA. They are mainly found in bacteria, but also exist naturally in archaea and eukaryotes such as yeast and plants. In nature, plasmids provide one or more functional benefits to the host such as resistance to antibiotics, degradative functions, and/or virulence. All natural plasmids contain an origin of replication (which controls the host range and copy number of the plasmid) and typically include a gene that is advantageous for survival, such as an antibiotic resistance gene. In contrast, plasmids utilized in the lab are usually artificial and designed to introduce foreign DNA into another cell. Minimally, lab-created plasmids have an origin of replication, selection marker, and cloning site. The ease of modifying plasmids and the ability of plasmids to selfreplicate within a cell make them attractive tools for the life scientist or bioengineer. (taken from plasmid 101)

Critical Points for Expression Clone

Framing – ensure that your gene’s start site is in the correct reading frame ü Many expression vectors have multiple variants for each of the three possible frames

Ø Restriction sites – blunt or sticky ends – PCR your needs into the clone, then digest. Ø Ensure start site for plasmid is upstream of your gene

Ø If using a tag (more later) is it in frame and on the desired terminus of the protein Plasmid Points Ø Important regions of the vector

ü Antibiotic resistance – selectable marker

• For bacteria – to create more plasmid

• For host system – if creating stable transfectants ü Origin of Replication

• Used by bacteria to make copies of plasmid ü Second “helper” plasmid ü For bacteria and T7 ü Typically carried in host competent cell ü Used for high level of expression ü Fusion tag

• N or C terminus • Large or small size – does the tag alter the function of the protein

• Purification or detection tag ü Compartment sequences

• Short sequences to send protein to membrane or organelle or even secreted into cell culture media ü High or low expression – inducible or repressed Appropriate Termination of GOI ü Proteolytic Cleavage site from fusion partner or tag Biochemistry Lab Systems for Recombinant Protein Expression Lecture Notes Handout Antibiotics

Ø Addition of a plasmid must include a second gene coding for a protein adding antibiotic resistance ü Only cells with plasmid can grow on selective medium (in the presence of an antibiotic) ü Maintains pressure on the bacteria to keep the plasmids

Ø In general, antibiotics are agents that kill bacteria or inhibit their growth. ü Kanamycin – inhibits 30S ribosome-protein expression ü Ampicillin – inhibits bacterial cell wall synthesis ü Carbencillin – longer lasting derivative of Amp ü Chloramphenicol – Stops 50S ribosome peptidyl translocation Ø Used as frozen 500 – 1000X stocks.

Ø Additional gene in plasmid must code for protein which degrades drug ü Correct Multiple Cloning Site (MCS) The Host with the Most!

Ø Choice of host depends on several needs ü Costs and availability of supports (kits, vectors, reagents and expertise) ü Yield – large mass of protein presents specific challenges for scale up (bacteria/yeast vs. mammalian)

How will the protein be used after expression?

Toxicity of protein on host cell ü Post-translational modification?

Impact on host cell on protein solubility and hydrolysis ü Location of protein within the host cell – membrane, inclusion body, secretion into media ü Complexity of medium Expression Systems (host cells)

Ø Bacteria, Yeast, Insect Cells, Mammalian Cells Plant Cells – protoplasts, Cell free “test tube” lysate translation

Ø Animals / Eggs (antibodies, serum proteins) Why Chose E.coli an Expression System? Ø Advantages – ü First choice due to level of support (technical, literature, kits, plasmids, reagents)…

High yield and easier scale up (up to 100 mg per liter of culture) ü Cost of media and reagents is relatively inexpensive ü Many expression plasmids – many with gene expression regulation ü Fast growth condition – one to two days from starter culture to cell pellet

Amenable to shaker flasks or fermentation ü Up to 50% of protein can be GOI! Ø Disadvantages ü No post-translational modification

• Lipid modification, phosphorylation, glycosylation, decarboxylation, acetylation

• These may be critical requirements for functional protein ü Large proteins may be difficult to express (plasmid size is limited to about 15 kb) ü Membrane proteins may not fold or isolate well

• Lipid membrane of bacteria – mostly phosphatidylethanolamine, animal cells – phosphatidylecholine and other p-lipids

• Large number of proteins are toxic to E.coli

• May need to significantly alter plating and other expression issues ü Inclusion body packing of protein
• Insoluble, aggregated and non-native protein

• Often times takes place in high expressed or toxic proteins ü Codon Use differences, false stops and low expression Bacterial Expression Ø Basic Process of Expression ü Clone heterologous gene into E.coli vector ü Transform plasmid (with antibiotic resistance) into appropriate strain of bacteria

• Low efficiency method (competent chemical transformation)– if plenty of plasmid DNA is present Biochemistry Lab Systems for Recombinant Protein Expression Lecture Notes Handout

• High efficiency method- (electroporation) if low abundant or problems with selection of plasmid

• Bacterial strain for DNA purification (DH5a, XL1-Blue or others) which are low in recombinases (RecA-) - not appropriate for protein expression – so you must keep two strains available, one for DNA maintenance and another for expression (more on expression later) ü Expand Single Colony from plate into shaking suspension culture

• Large clumps or lawns will allow cells which do not carry resistance and thus your plasmid with GOI – a great recipe for disaster ü Expand culture to low density ü Induce cells while early log phase ü Harvest cells before density reaches critical stage ü Lyse cells and purify protein – what could be easier! E.coli Expression Strains Ø Key choice when starting expression Ø Characteristics to consider: ü Low protease activity (clpC, clpP, Ion, and ompT genes) ü Lower RNAse activity – enhances protein expression ü Level of tRNA for rare codons

• BL21 codon plus, Rosetta – contain additional plamids for tRNA ü Appropriateness for induction • IPTG (LacZ gene) – most will work but must have lac permease gene in cell strain (LacZY) to allow the compound to enter the cell

• Tuner strains from Novagen – have a mutated lac permease to allow a better control of IPTG entry and titration of induction

• Arabinose (PBAD – operator) need to have arabinose transporter expressed in cells (BL21-A1, KS272 or LMG194) Ø Key choice when starting expression Ø Characteristics to consider: ü Cys-Cys disulfide bridges – multiple bridges require special strains which have mutations in thioredoxin reductase and glutathione reductase (involved with –SH redox) to enhance disulfide bond formation – Origami strain (also low in proteases) ü Toxicity of recombinant protein – will need to use C41(DE3) or C43(DE3) strains • Decreased cell death and allow membrane protein expression from a number of organisms ü Vector specific strains – provide “helper” genes

• pLYSE pQE strains IPTG_Induction Ø Lac Operon – set of bacterial genes used by bacteria to transport and metabolize lactose when the sugar is present in media as a carbon source Ø IPTG – mimics effect of lactose ü Is not metabolized and is longer lasting ü Provides strong induction of Lac Operon Ø IPTG Induction Vectors/Promoters ü Replace the three genes with GOI Expression Vectors for E.coli Ø pET - Strong expression of GOI – up to 50% of total cell protein ü Dose-responsive promoter (IPTG) tunable strain may be used in concert ü Uses strong promoter of T7 RNA polymerase to drive expression of second promoter in the pET vector ü Expression of pET is indirectly tied to IPTG and MUST have either a second plasmid or a host cell already containing the T7 RNA pol gene (DE3) ü This provides a strong control of expression – not leaky Ø pBAD ü Induction by addition of arabinose ü Sugar binds to AraC protein ü Complex binds and initiates RNA polymerase ü Permits a dose response Biochemistry Lab Systems for Recombinant Protein Expression Lecture Notes Handout ü Good choice for toxic proteins ü Control expression for insoluble – inclusion body formation Ø pQE – Qiagen 6xHis tagged construct ü Low copy plasmid with a T5 promoter ü Two lac operon sequence for

repressor binding ü These can be leaky promoters (as they are so strong) and thus inhibition may need pREP4 (a second plasmid coding for lac repressor protein) ü IPTG inducable ü Some allow two genes

• Bis-cistronic Ø pGEX– Amersham– GE HealthSciences ü IPTG inducable ü Fusion protein – Glutathione S Transferase (GST) ü Many different vector variations ü Large fusion protein sometimes allows for good expression ü Easy to purify ü GST may need to be removed Inclusion Bodies Ø Region of bacteria which can fill with insoluble protein Ø Over expression of proteins or toxic proteins induce formation of inclusion bodies ü Aggregates may be mostly the expressed protein

• disulfide bonds incorrectly formed as the protein is expressed at a high rate may be cause ü Isolation of inclusion bodies requires lysing bacteria but not inclusion bodies, several centrifugation steps followed by a difficult lysing of body Ø Over expression of proteins or toxic proteins induce formation of inclusion bodies ü Denaturation and refolding of proteins can result in high yield of native protein ü Requires strong chaotropic detergents – urea, guanidine chloride ü Renaturation is an art form and no standard protocol exists ü Isolating non-functioning protein may work well for some uses – antigen injection for antibody production Cloning & Transforming in Yeast Cells Pichia Pastoris

Ø Yeast are single celled eukaryotes Ø Behave like bacteria, but have key advantages of eukaryotes Ø P. pastoris is a methylotrophic yeast that can use methanol as its sole carbon source (using alcohol oxidase)

Ø Has a very strong promoter for the alcohol oxidase (AOX) gene (~30% of protein produced when induced) Yeast as an expression system

Ø Pichia Pastoris or Saccharomyces Cervecisiae

Ø Reasonable cost – media and scale up are straight forward

Ø Takes longer to prepare and screen for expression than bacteria

Ø Glycosylation is possible but not as complex as human cells – ü Modification can be heterogenious and interfere with purification

Ø Less support for expression than bacteria Ø Pastoris expression induced by methanol Ø Cloning in GOI is straight forward – many cDNA genes available Expression in Baculovirus

Ø SF9 cells/Insect cells ü From the gut of moth ü Grown in suspension at mild temps ü Allow for scale up ü Often used for difficult to express clones in bacteria Biochemistry Lab Systems for Recombinant

Protein Expression Lecture Notes Handout ü Also used for modified proteins ü Expression level is variable between infections

Ø Making the Virus ü Clone GOI into transfer plasmid vector ü Second DNA (linear) includes viral recombination region ü Both are infected into SF9 cells where they recombine and are packaged as viral particals ü Viral particals are tested for infection levels (MOI) ü Infect Sf-9 for protein expression Types of Insect cell lines cells Doubling time Cell appearance Medium Origin Type of culture Sf 9 72 hrs Spherical, granular, regular in size, firm attachment to surface TNM-FH IPLBSF-21 cell lines of the fall army worm spodoptera frugiperda Grow well as monolayer and suspension Sf 21 24 hrs Spherical, granular, different in size, firm attachment to surface TNM-FH IPLBSF-21 cell lines of the fall army worm spodoptera frugiperda Grow well as monolayer and suspension High-five 18 hrs Spherical, granular, regular in size, loose attachment to surface Express five SFM Ovarian cells of cabbage looper Grow well as monolayer, also as suspension Insect Medium Ø Grace's Insect medium-unsupplemented but contains L-glutamine Ø Grace's Insect medium supplemented-contains additional TC yeastolate & Lactalbumin hydrolysate Ø Trichoplusia ni Medium formulation hink (TNM-FH)-contains 10% FBS Requirements for proper cell culture Ø Temperature- Optimal range is 27-28 C Ø pH- Optimal range is 6.1 to 6.4

Ø Aeration-Requires passive 02 diffusion for optimal growth & recombinant protein expression

Ø Osmolality- Optimum is 345-380 mOsm/kg Ø FBS- Working with suspension culture it is advisable to use (10-20% FBS) to gave protection from cellular shear forces Types of cell culturing

Ø Suspension culture Ø Monolayer culture

Three methods to dislodge monolayers in adherent cell culture - Sloughing -Trypsinization -Tapping the layer until monolayer loosens Expression in Baculovirus

Ø More time and labor intensive

Ø More costs – reasonable support Ø Advantages: ü Provides soluble post-translationally modified proteins Biochemistry Lab Systems for Recombinant Protein Expression Lecture Notes Handout ü Virus can handle large genes ü Works well for membrane or cytoplasmic proteins ü Can infect with multiple virus to express multi-subunit proteins ü High levels of expression possible 1-5 mg of protein / liter ü Scale up possible ü Low level of risk as virus is restricted to specific insect cell Mammalian Cell Expression

Ø Much more expensive and time intensive than E. coli Ø Much lower potential yield than yeast or bacterial expression However:

Ø Most authentic secretion, glycosyation, phosphorylation and other post-translational modification Ø Nearly the only way to fully express membrane proteins from mammalian cells Ø Best structural and functional features and similar to cognate native forms

Ø Infection with virus or transient transfection ü Selection must occur for producing large scale amounts of protein. Scale-able

Ø Suspension cells include COS (monkey), CHO (hamster), HEK293 (Human) ü Easy to transfect and select ü Grow in relatively simple media and can be conditioned to large flasks Ø Monolayer cells – fibroblasts but other cells will work Scale-able

Ø Suspension - After stable transfection, can move to shaker flasks, wave flasks or bioreactors Scale-able Ø Monolayers – limited to roller bottles or cell farms Potential Problems

Ø Cost and time – both are very high ü Labor to create cell lines ü Cost of infection or transfection ü Often low yield

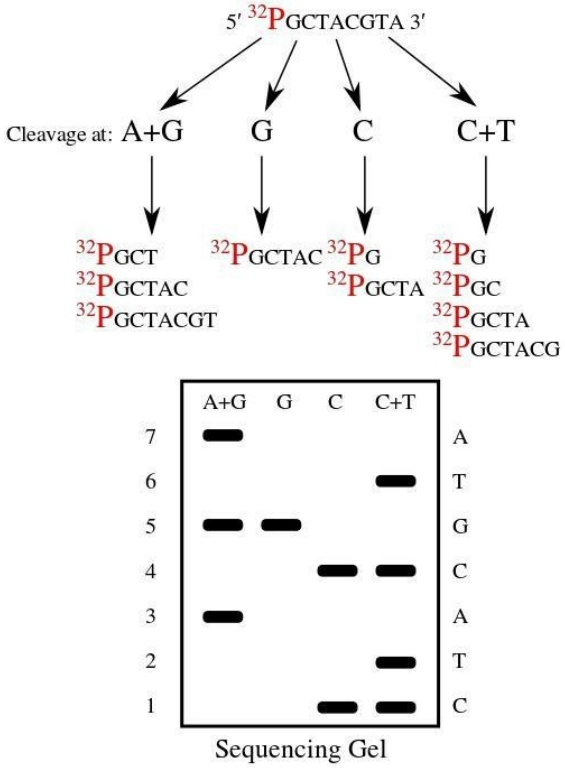
Ø Over expression of protein can be toxic to cell

UNIT – 4 – GENETIC ENGINEERING – SBTA1501

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. Knowledge of DNA sequences has become indispensable for basic biological research, and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics.

Maxam–Gilbert sequencing is a method of DNA sequencing developed by Allan Maxam and Walter Gilbert in 1976–1977. This method is based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified

Fig:1. MAXAM GILBERT SEQUENCING



Sequencing Gel

Maxam–Gilbert sequencing requires radioactive labeling at one 5′ end of the DNA fragment to be sequenced (typically by akinase reaction using gamma-^{32}P ATP) and purification of the DNA. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). For example, the purines(A+G) are depurinated using formic acid, the guanines (and to some extent the adenines) are methylated by dimethyl sulfate, and the pyrimidines (C+T) are hydrolysed using hydrazine. The addition of salt (sodium chloride) to the hydrazine reaction inhibits the reaction of thymine for the C-only reaction. The modified DNAs may then be cleaved by hot piperidine;(CH2)5NH at the position of the modified base. The concentration of the modifying chemicals is controlled to introduce on average one modification per DNA molecule. Thus a series of labeled fragments is generated, from the radiolabeled end to the first "cut" site in each molecule.

The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each showing the location of identical radiolabeled DNA molecules. From presence and absence of certain fragments the sequence may be inferred.

Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminatingdideoxynucleotides by DNA polymerase during in vitro DNA replication. Developed by Frederick Sanger and colleagues in 1977, it was the most widely used sequencing method for approximately 25 years. More recently, Sanger sequencing has been supplanted by "Next-Gen" sequencing methods, especially for large-scale, automated genome analyses.

However, the Sanger method remains in wide use, for smaller-scale projects, validation of Next-Gen results and for obtaining especially long contiguous DNA sequence reads (>500 nucleotides).

The classical chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleosidetriphosphates (dNTPs), and modified di-deoxynucleosidetriphosphates (ddNTPs), the latter of which terminate DNA strand elongation. These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, causing DNA polymerase to cease extension of DNA when a modified ddNTP is incorporated. The ddNTPs may be radioactively or fluorescently labeled for detection in automated sequencing machines.

The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides(dATP, dGTP, dCTP and dTTP) and the DNA polymerase. To each reaction is added only one of the fourdideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP), while the other added nucleotides are ordinary ones. Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat denatured and separated by size using gel electrophoresis. In the original publication of 1977,[2] the formation of base-paired loops of ssDNA was a cause of serious difficulty in resolving bands at some locations. This is frequently performed using a denaturing polyacrylamide-urea gel with each of the four reactions run in one of four individual lanes (lanes A, T, G, C). The DNA bands may then be visualized by autoradiography or UV light and the DNA sequence can be directly read off the X-ray film or gel image.

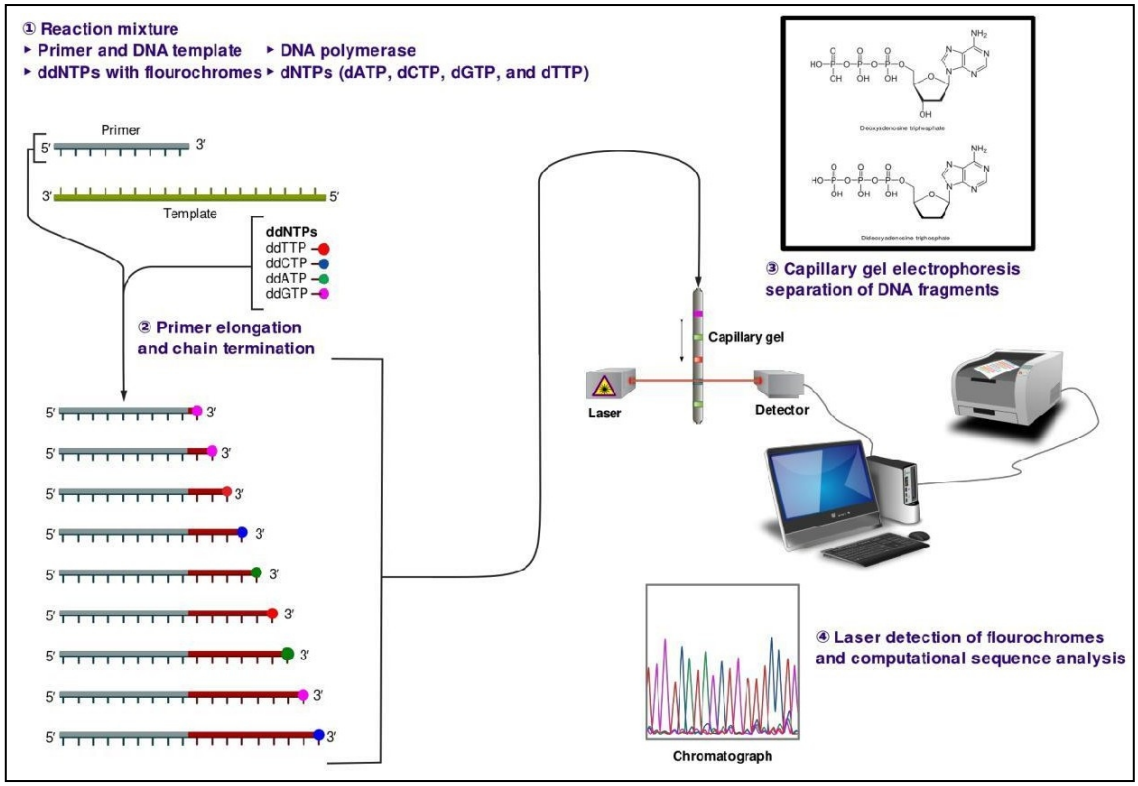


Fig : 2 DNA Sequencing Sangers

Pyrosequencing is a method of DNA sequencing (determining the order ofnucleotides in DNA) based on the "sequencing by synthesis" principle.

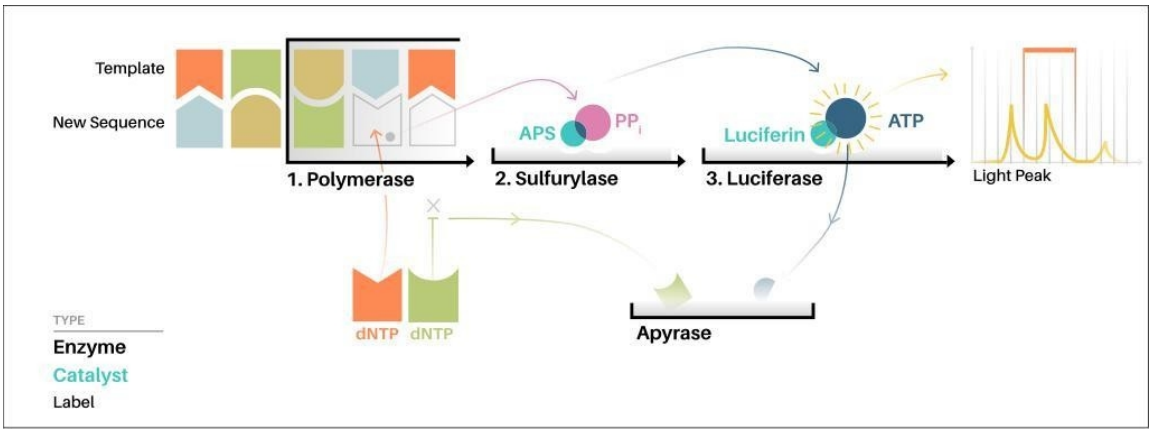
"Sequencing by synthesis" involves taking a single strand of the DNA to be sequenced and then synthesizing its complementary strand enzymatically. The pyrosequencing method is based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step. The template DNA is immobile, and solutions of A, C, G, and T nucleotides are sequentially added and removed from the reaction. Light is produced only when the nucleotide solution complements the first unpaired base of the template. The sequence of solutions which produce chemiluminescent signals allows the determination of the sequence of the

template.

The single-strand DNA (ssDNA) template is hybridized to a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase, and with the substrates adenosine 5´ phosphosulfate (APS) andluciferin.

- The addition of one of the four deoxynucleoside triphosphates (dNTPs) (dATPαS, which is not a substrate for a luciferase, is added instead of dATP to avoid noise) initiates the second step. DNA polymerase incorporates the correct, complementary dNTPs onto the template. This incorporation releases pyrophosphate (PPi).
- ATP sulfurylase converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. This ATP acts as a substrate for the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a camera and analyzed in a pyrogram.
- Unincorporated nucleotides and ATP are degraded by the apyrase, and the reaction can restart with another nucleotide.

Fig 3 PYROSEQUENCING

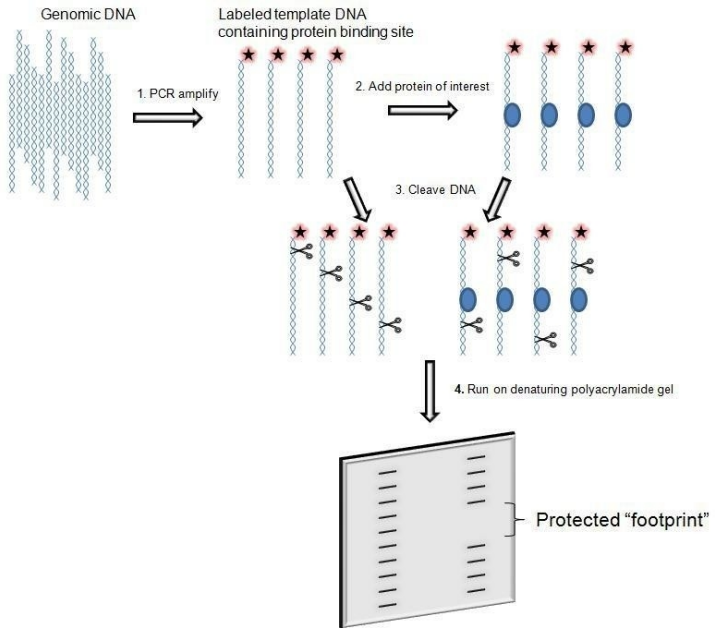


DNA footprinting

DNA footprinting is a method of investigating the sequence specificity of DNA-binding proteins in vitro. This technique can be used to study protein-DNA interactions both outside and within cells.

The regulation of transcription has been studied extensively, and yet there is still much that is not known. Transcription factors and associated proteins that bind promoters, enhancers, or silencers to drive or repress transcription are fundamental to understanding the unique regulation of individual genes within the genome. Techniques like DNA footprinting help elucidate which proteins bind to these associated regions of DNA and unravel the complexities of transcriptional control.

Fig 4; DNA FOOTPRINITING



In 1978, David Galas and Albert Schmitz developed the DNA footprinting technique to study the binding specificity of the lac repressor protein. It was originally a modification of the Maxam-Gilbert chemical sequencing technique.

The simplest application of this technique is to assess whether a given protein binds to a region of interest within a DNA molecule. Polymerase chain reaction (PCR) amplify and label region of interest that contains a potential protein-binding site, ideally amplicon is between 50 to 200 base pairs in length. Add protein of interest to a portion of the labeled template DNA; a portion should remain separate without protein, for later comparison. Add a cleavage agent to both portions of DNA template. The cleavage agent is a chemical or enzyme that will cut at random locations in a sequence independent manner. The reaction should occur just long enough to cut each DNA molecule in only one location. A protein that specifically binds a region within the DNA template will protect the DNA it is bound to from the cleavage agent. Run both samples side by side on a polyacrylamide gel electrophoresis. The portion of DNA template without protein will be cut at random locations, and thus when it is run on a gel, will produce a ladder-like distribution. The DNA template with the protein will result in ladder distribution with a break in it, the "footprint", where the DNA has been protected from the cleavage agent. Note: Maxam-Gilbert chemical DNA sequencing can be run alongside the samples on the polyacrylamide gel to allow the prediction of the exact location of ligand binding site.

A variety of cleavage agents can be chosen.

DNase I is a large protein that functions as a double-strand endonuclease. It binds the minor groove of DNA and cleaves the phosphodiester backbone. It is a good cleavage agent for footprinting because its size makes it easily physically hindered. Thus is more likely to have its action blocked by a bound protein on a DNA sequence. In addition, the DNase I enzyme is easily controlled by adding EDTA to stop the reaction. There are however some limitations in using DNase I. The enzyme does not cut DNA randomly; its activity is affected by local DNA structure

and sequence and therefore results in an uneven ladder. This can limit the precision of predicting a protein's binding site on the DNA molecule.

Hydroxyl radicals are created from the Fenton reaction, which involves reducing Fe^{2+} with H_2O_2 to form free hydroxyl molecules. These hydroxyl molecules react with the DNA backbone, resulting in a break. Due to their small size, the resulting DNA footprint has high resolution. Unlike DNase I they have no sequence dependence and result in a much more evenly distributed ladder. The negative aspect of using hydroxyl radicals is that they are more time consuming to use, due to a slower reaction and digestion time.

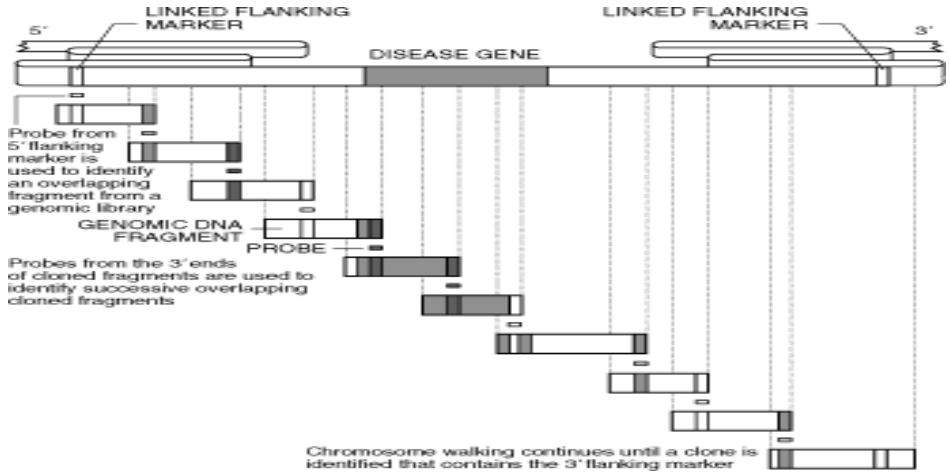
Ultraviolet irradiation can be used to excite nucleic acids and create photoreactions, which results in damaged bases in the DNA strand. Photoreactions can include: single strand breaks, interactions between or within DNA strands, reactions with solvents, or crosslinks with proteins. The workflow for this method has an additional step, once both your protected and unprotected DNA have been treated, there is subsequent primer extension of the cleaved products. The extension will terminate upon reaching a damaged base, and thus when the PCR products are run side-by-side on a gel; the protected sample will show an additional band where the DNA was crosslinked with a bound protein. Advantages of using UV are that it reacts very quickly and can therefore capture interactions that are only momentary. Additionally it can be applied to *in vivo* experiments, because UV can penetrate cell membranes. A disadvantage is that the gel can be difficult to interpret, as the bound protein does not protect the DNA, it merely alters the photoreactions in the vicinity.

Chromosome walking

Chromosome walking is a method of positional cloning used to find, isolate, and clone a particular allele in a gene library. An allele is a gene for a particular genetic trait passed on from adults to their offspring, such as the allele for brown eyes in a gene for eye color. Sometimes, the approximate location of a single allele in a string of deoxyribonucleic acid (DNA) may be known. To isolate a particular allele for a genetically transmitted disease, chromosome walking may need to explore for the desired specimen in an unmapped DNA sequence outside of previously mapped sequences.

To locate a particular disease gene, the walking starts at the closest gene that has already been identified, known as a marker gene. Each successive gene in the sequence is tested repeatedly for what are known as overlap restrictions and mapped for their precise location in the sequence. Eventually, walking through the genes reaches the mutant gene in an unmapped sequence that binds to a fragment of a gene of that particular disease. Once the gene is cloned, its function can be fully identified. Throughout this process, tests are done to fully identify the properties of each

Fig 5 Chromosome walking



successive clone, to map their locations for future use.

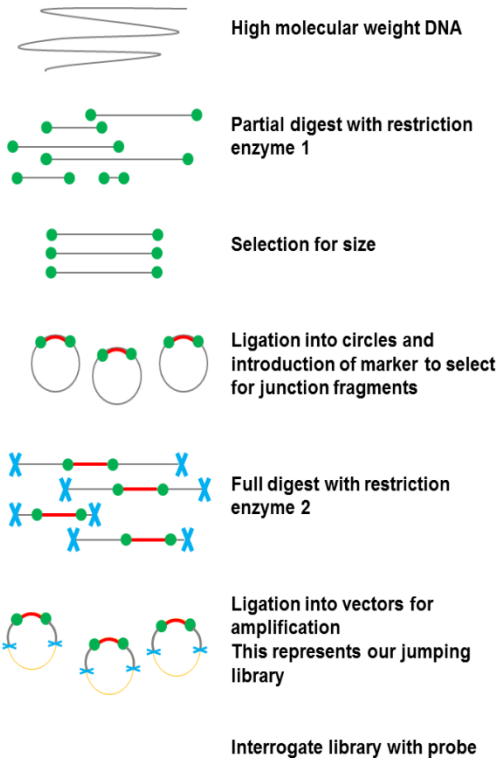
Chromosome jumping

Chromosome jumping is a tool of molecular biology that is used in the physical mapping of genomes. It is related to several other tools used for the same purpose, including chromosome walking.

Chromosome jumping is used to bypass regions difficult to clone, such as those containing repetitive DNA, that cannot be easily mapped by chromosome walking, and is useful in moving along a chromosome rapidly in search of a particular gene.

In chromosome jumping, the DNA of interest is identified, cut into fragments with restriction enzymes, and circularised (the beginning and end of each fragment are joined together to form a circular loop). From a known sequence, a primer is designed to sequence across the circularised junction. This primer is used to jump 100 kb-300 kb intervals: a sequence 100 kb away would have come near the known sequence on circularisation. Thus, sequences not reachable by chromosome walking can be sequenced. Chromosome walking can be used from the new jump position (in either direction) to look for gene-like sequences, or additional jumps can be used to progress further along the chromosome.

Fig 5 Chromosome jumpoing



Rapid amplification of cDNA ends (**RACE**) is a technique used to obtain the full length sequence of an RNA transcript found within a cell. RACE results in the production of a cDNA copy of the RNA sequence of interest, produced through reverse transcription, followed by PCR amplification of the cDNA copies (see RT-PCR). The amplified cDNA copies are then sequenced and, if long enough, should map to a unique genomic region. RACE can provide the sequence of an RNA transcript from a small known sequence within the transcript to the 5' end (5' RACE-PCR) or 3' end (3' RACE-PCR) of the RNA. This technique is sometimes called *one-sided PCR* or *anchored PCR*.

The first step in RACE is to use reverse transcription to produce a cDNA copy of a region of the RNA transcript. In this process, an unknown end portion of a transcript is copied using a known sequence from the center of the transcript. The copied region is bounded by the known sequence, at either the 5' or 3' end.

The protocols for 5' or 3' RACES differ slightly. 5' RACE-PCR begins using mRNA as a template for a first round of cDNA synthesis (or reverse transcription) reaction using an *anti-sense* (reverse) oligonucleotide primer that recognizes a known sequence in the middle of the gene of interest; the primer is called a *gene specific primer* (GSP). The primer binds to the mRNA, and the enzyme reverse transcriptase adds base pairs to the 3' end of the primer to generate a specific single-stranded cDNA product; this is the reverse complement of the mRNA. Following cDNA synthesis, the enzyme terminal deoxynucleotidyl transferase (TdT) is used to add a string of identical nucleotides, known as a homopolymeric tail, to the 3' end of the cDNA. (There are some other ways to add the 3'-terminal sequence for the first strand of the de novo cDNA synthesis which are much more efficient than homopolymeric tailing, but the sense of the method remains the same). APCR reaction is then carried out, which uses a second anti-sense gene specific primer (GSP2) that binds to the known sequence, and a sense (forward) universal primer (UP) that binds the homopolymeric tail added to the 3' ends of the cDNAs to amplify a cDNA product from the 5' end.

3' RACE-PCR uses the natural polyA tail that exists at the 3' end of all eukaryotic mRNAs for priming during reverse transcription, so this method does not require the addition of nucleotides by TdT. cDNAs are generated using an Oligo-dT-adaptor primer (a primer with a short sequence of deoxy-thymine nucleotides) that complements the polyA stretch and adds a special adaptor sequence to the 5' end of each cDNA. PCR is then used to amplify 3' cDNA from a known region using a sense GSP, and an anti-sense primer complementary to the adaptor sequence.

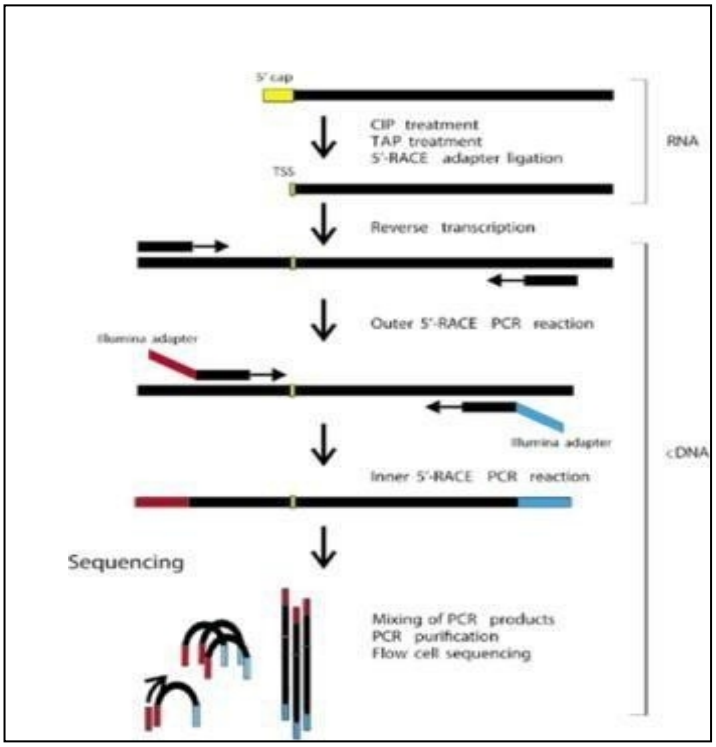


Fig : 6 RACE

Gene silencing is a general term used to describe the regulation of gene expression. In particular, this term refers to the ability of a cell to prevent the expression of a certain gene.

Gene silencing can occur during either transcription ortranslation and is often used in research. In particular, methods used to silence genes are being increasingly used to produce therapeutics to combat cancer and diseases, such as infectious diseases and neurodegenerative disorders.

Gene silencing is often considered the same as gene knockout. When genes are silenced, their expression is reduced. In contrast, when genes are knocked out, they are completely erased from the organism's genome and, thus, have no expression. Gene silencing is considered a gene knockdown mechanism since the methods used to silence genes, such as RNAi, CRISPR, or siRNA, generally reduce the expression of a gene by at least 70% but do not completely eliminate it. Methods using gene silencing are often considered better than gene knockouts since they allow researchers to study essential genes that are required for the animal models to survive and cannot be removed. In addition, they provide a more complete view on the development of diseases since diseases are generally associated with genes that have a reduced expression.

Antisense oligonucleotides

Antisense oligonucleotides were discovered in 1978 by Paul Zamecnik and Mary Stephenson. Oligonucleotides, which are short nucleic acid fragments, bind to complementary target mRNA molecules when added to the cell. These molecules can be composed of single-stranded DNA or RNA and are generally 13-25 nucleotides long. The antisense oligonucleotides can affect gene expression in two ways: by using an RNase H-dependent mechanism or by using a steric blocking mechanism. RNase H-dependent oligonucleotides cause the target mRNA molecules to be degraded, while steric-blocker oligonucleotides prevent translation of the mRNA molecule. The majority of antisense drugs function through the RNase H-dependent mechanism, in which RNase H hydrolyzes the RNA strand of the DNA/RNAheteroduplex. This mechanism is thought to be more efficient, resulting in an approximately 80% to 95% decrease in the protein and mRNA expression.

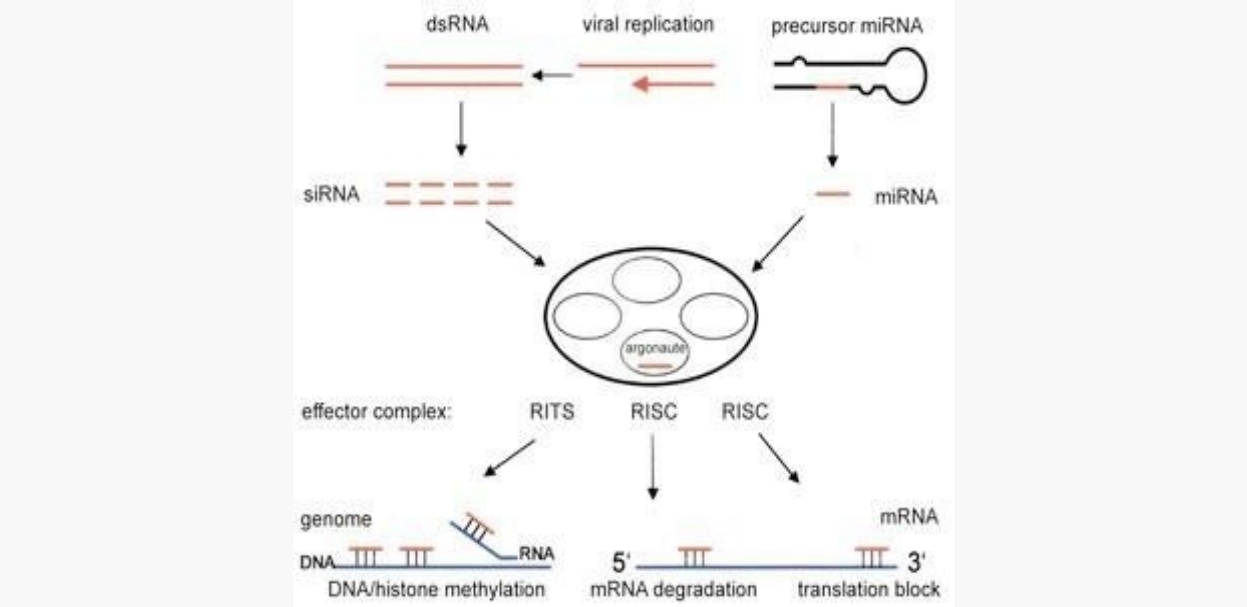
Ribozymes

Ribozymes are catalytic RNA molecules used to inhibit gene expression. These molecules work by cleaving mRNA molecules, essentially silencing the genes that produced them. Sidney Altman and Thomas Cech first discovered catalytic RNA molecules, RNase P and group II intron ribozymes, in 1989 and won the Nobel Prize for their discovery. Several types of ribozyme

motifs exist, includinghammerhead, hairpin, hepatitis delta virus, group I, group II, and RNase Pribozymes. Hammerhead, hairpin, and hepatitis delta virus (HDV) ribozyme motifs are generally found in viruses or viroid RNAs. These motifs are able to self-cleave a specific phosphodiester bond on an mRNA molecule. Lower eukaryotes and a few bacteria contain group I and group II ribozymes. These motifs can self-splice by cleaving and joining together phosphodiester bonds. The last ribozyme motif, the RNase P ribozyme, is found in *Escherichia coli* and is known for its ability to cleave the phosphodiester bonds of several tRNA precursors when joined to a protein cofactor.

The general catalytic mechanism used by ribozymes is similar to the mechanism used by protein ribonucleases. These catalytic RNA molecules bind to a specific site and attack the neighboring phosphate in the RNA backbone with their 2' oxygen, which acts as a nucleophile, resulting in the formation of cleaved products with a 2'3'-cyclic phosphate and a 5' hydroxyl terminal end. This catalytic mechanism has been increasingly used by scientists to perform sequence-specific cleavage of target mRNA molecules. In addition, attempts are being made to use ribozymes to produce gene silencing therapeutics, which would silence genes that are responsible for causing diseases.

Fig 7 : RNA interference



RNA interference (RNAi) is a natural process used by cells to regulate gene expression. It was discovered in 1998 by Andrew Fire and Craig Mello, who won the Nobel Prize for their discovery in 2006. The process to silence genes first begins with the entrance of a double-stranded RNA (dsRNA) molecule into the cell, which triggers the RNAi pathway The double-stranded molecule is then cut into small double-stranded fragments by an enzyme called Dicer. These small fragments, which includesmall interfering RNAs (siRNA) and microRNA (miRNA), are approximately 21-23 nucleotides in length. The fragments integrate into a multi-subunit protein called the RNA-induced silencing complex, which contains Argonaute proteins that are essential components of the RNAi pathway. One strand of the molecule, called the "guide" strand, binds to RISC, while the other strand, known as the "passenger" strand is degraded. The

guide or antisense strand of the fragment that remains bound to RISC directs the sequence-specific silencing of the target mRNA molecule. The genes can be silenced by siRNA molecules that cause the endonucleatic cleavage of the target mRNA molecules or by miRNA molecules that suppress translation of the mRNA molecule. With the cleavage or translational repression of the mRNA molecules, the genes that form them are essentially inactive. RNAi is thought to have evolved as a cellular defense mechanism against invaders, such as RNA viruses, or to combat the proliferation of transposons within a cell's DNA. Both RNA viruses and transposons can exist as double-stranded RNA and lead to the activation of RNAi. Currently, siRNAs are being widely used to suppress specific gene expression and to assess the function of genes

Construction of siRNA vectors

There are several methods for preparing siRNA, such as chemical synthesis, in vitro transcription, siRNA expression vectors, and PCR expression cassettes. Irrespective of which method one uses, the first step in designing a siRNA is to choose the siRNA target site.

General Design Guidelines

If you prefer to design your own siRNAs, you can choose siRNA target sites in a variety of different organisms based on the following guidelines. Corresponding siRNAs can then be chemically synthesized, created by in vitro transcription, or expressed from a vector or PCR product.

1. Find 21 nt sequences in the target mRNA that begin with an AAdinucleotide.

Beginning with the AUG start codon of your transcript, scan for AA dinucleotide sequences. Record each AA and the 3' adjacent 19 nucleotides as potential siRNA target sites. This strategy for choosing siRNA target sites is based on the observation by Elbashir et al. that siRNAs with 3' overhanging UU dinucleotides are the most effective. This is also compatible with using RNA pol III to transcribe hairpin siRNAs because RNA pol III terminates transcription at 4-6

nucleotide poly(T) tracts creating RNA molecules with a short poly(U) tail.

In Elbashir's and subsequent publications, siRNAs with other 3' terminal dinucleotide overhangs have been shown to effectively induce RNAi. If desired, you may modify this target site selection strategy to design siRNAs with other dinucleotide overhangs, but it is recommended that you avoid G residues in the overhang because of the potential for the siRNA to be cleaved by RNase at single-stranded G residues.

2. Select 2-4 target sequences.

Research has found that typically more than half of randomly designed siRNAs provide at least a 50% reduction in target mRNA levels and approximately 1 of 4 siRNAs provide a 75-95% reduction. Choose target sites from among the sequences identified in Step 1 based on the following guidelines:

- Researchers find that siRNAs with 30-50% GC content are more active than those with a higher G/C content.

- Since a 4-6 nucleotide poly(T) tract acts as a termination signal for RNA pol III, avoid stretches of > 4 T's or A's in the target sequence when designing sequences to be expressed from an RNA pol III promoter.

- Since some regions of mRNA may be either highly structured or bound by regulatory proteins, we generally select siRNA target sites at different positions along the length of the gene sequence. We have not seen any correlation between the position of target sites on the mRNA and siRNA potency.

- Compare the potential target sites to the appropriate genome database (human, mouse, rat, etc.) and eliminate from consideration any target sequences

with more than 16-17

contiguous base pairs of homology to other coding sequences. We suggest using BLAST, which can be found on the NCBI server at: www.ncbi.nlm.nih.gov/BLAST.

3. Design appropriate controls.

A complete siRNA experiment should include a number of controls to ensure the validity of the data. The editors of Nature Cell Biology have recommended several controls (2). Two of these controls are:

- **A negative control siRNA with the same nucleotide composition as your siRNA but which lacks significant sequence homology to the genome.** To design a negative control siRNA, scramble the nucleotide sequence of the gene-specific siRNA and conduct a search to make sure it lacks homology to any other gene.

- **Additional siRNA sequences targeting the same mRNA.** Perhaps the best way to ensure confidence in RNAi data is to perform experiments, using a single siRNA at a time, with two or more different siRNAs targeting the same gene. Prior to these experiments, each siRNAshould be tested to ensure that it reduces target gene expression by comparable levels.

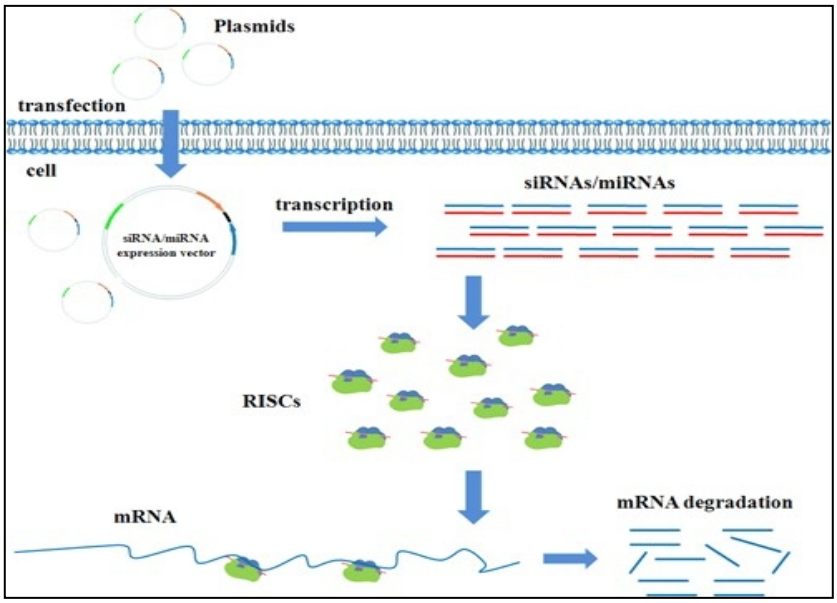


Fig 7 SiRNA Technology

Site-directed mutagenesis

Site-directed mutagenesis is one of the most important techniques in laboratory for introducing mutation into a DNA sequence. However, with decreasing costs of oligonucleotide synthesis, artificial gene synthesis is now occasionally used as an alternative to site-directed mutagenesis.

The basic procedure requires the synthesis of a short DNA primer. This synthetic primer contains the desired mutation and is complementary to the template DNA around the mutation site so it can hybridize with the DNA in the gene of interest. The mutation may be a single base change (a point mutation), multiple base changes, deletion, or insertion. The single-strand primer is then extended using a DNA polymerase, which copies the rest of the gene. The gene thus copied contains the mutated site, and is then introduced into a host cell as a vector and cloned. Finally, mutants are selected by DNA sequencing to check that they contain the desiredmutation.

The original method using single-primer extension was inefficient due to a low yield of mutants.

This resulting mixture contains both the original unmutated template as well as the mutant strand, producing a mixed population of mutant and non-mutant progenies. Furthermore the template used is methylated while the mutant strand is unmethylated, and the mutants may be counter-selected due to presence of mismatch repair system that favors the methylated template DNA, resulting in fewer mutants. Many approaches have since been developed to improve the efficiency of mutagenesis.

A large number of methods are available to effect site-directed mutagenesis, although most of them are now rarely used in laboratories since the early 2000s, as newer techniques allow for simpler and easier ways of introducing site-specific mutation into genes.

Kunkel's method

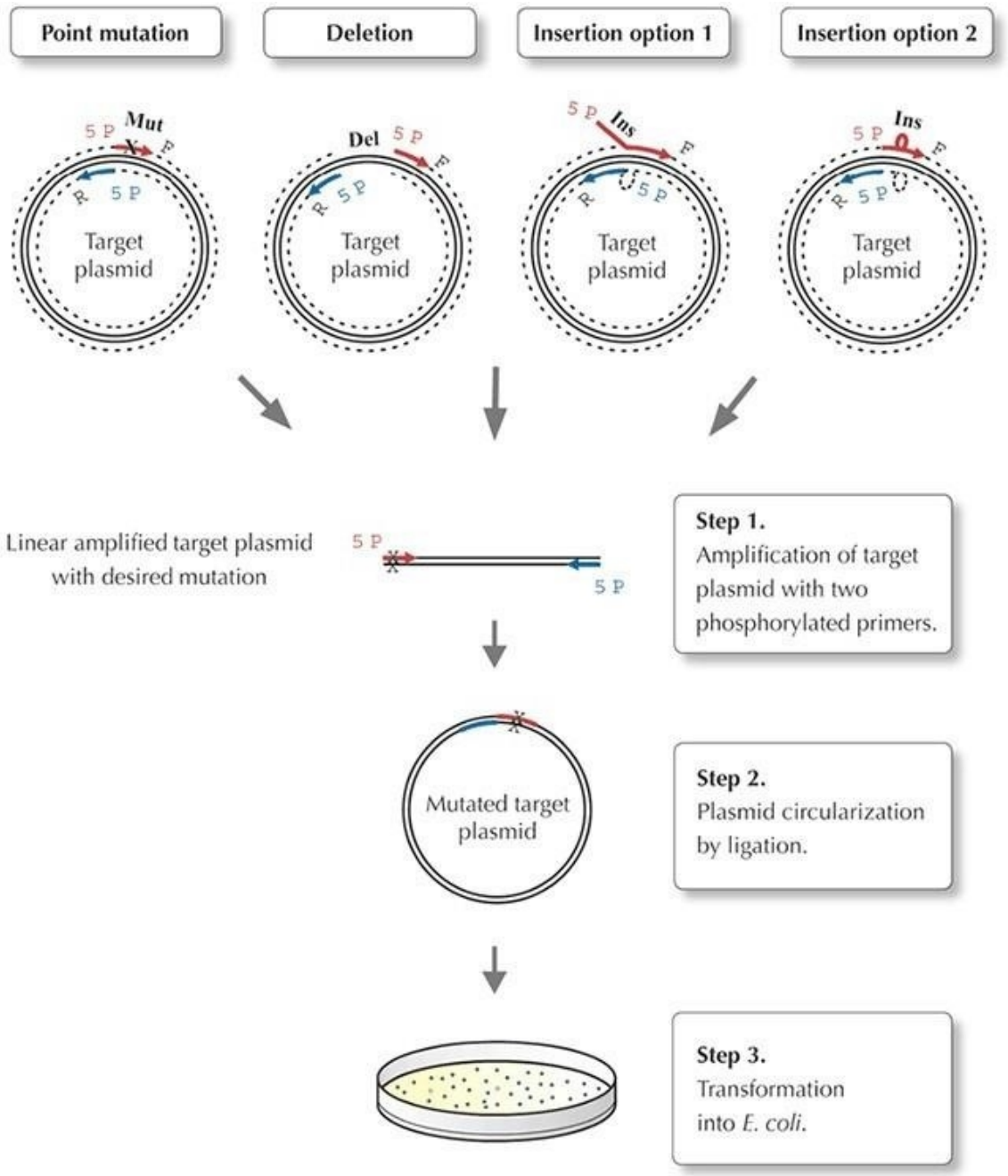
In 1987, Thomas Kunkel introduced a technique that reduces the need to select for the mutants.The DNA fragment to be mutated is inserted into a phagemid such as M13mp18/19 and is then transformed into an *E. coli* strain deficient in two enzymes, dUTPase (*dut*) and uracil deglycosidase (*ung*). Both enzymes are part of a DNA repair pathway that protects the bacterial chromosome from mutations by the spontaneous deamination of dCTP to dUTP. The dUTPase deficiency prevents the breakdown of dUTP, resulting in a high level of dUTP in the cell. The uracil deglycosidase deficiency prevents the removal of uracil from newly synthesized DNA. As the double-mutant *E. coli* replicates the phage DNA, its enzymatic machinery may, therefore, misincorporate dUTP instead of dTTP, resulting in single-strand DNA that contains some uracils (ssUDNA). The ssUDNA is extracted from the bacteriophage that is released into the medium, and then used as template for mutagenesis. An oligonucleotide containing the desired mutation is used for primer extension. The heteroduplex DNA, that forms, consists of one parental non-mutated strand containing dUTP and a mutated strand containing dTTP. The DNA is then transformed into an E. coli strain carrying the wildtype *dut* and *ung* genes. Here, the uracil-containing parental DNA strand is degraded, so that nearly all of the resulting DNA consists of the mutated strand.

Cassette mutagenesis

Unlike other methods, cassette mutagenesis need not involve primer extension using DNA polymerase. In this method, a fragment of DNA is synthesized, and then inserted into a plasmid. It involves the cleavage by a restriction enzyme at a site in the plasmid and subsequent ligation of a pair of complementary oligonucleotides containing the mutation in the gene of interest to the plasmid. Usually, the restriction enzymes that cut at the plasmid and the oligonucleotide are the same, permitting sticky ends of the plasmid and insert to ligate to one another. This method can generate mutants at close to 100% efficiency, but is limited by the availability of suitable restriction sites flanking the site that is to be mutated.

PCR site-directed mutagenesis

The limitation of restriction sites in cassette mutagenesis may be overcome using polymerase chain reaction witholigonucleotide "primers", such that a larger fragment may be generated, covering two convenient restriction sites. The exponential amplification in PCR produces a fragment containing the desired mutation in sufficient quantity to be separated from the original, unmutated plasmid by gel electrophoresis, which may then be inserted in the original context using standard recombinant molecular biology techniques. There are many variations of the same technique. The simplest method places the mutation site toward one of the ends of the fragment whereby one of two oligonucleotides used for generating the fragment contains the mutation. This involves a single step of PCR, but still has the inherent problem of requiring a suitable restriction site near the mutation site unless a very long primer is used. Other variations, therefore, employ three or four oligonucleotides, two of which may be non-mutagenic oligonucleotides that cover two convenient restriction sites and generate a fragment that can be digested and ligated into a plasmid, whereas the mutagenic oligonucleotide may be complementary to a location within that fragment well away from any convenient restriction site. These methods require multiple steps of PCR so that the final fragment to be ligated can contain the desired mutation.



Figur 8 Transformation

PCR (Polymerase Chain Reaction)

PCR (Polymerase Chain Reaction) is a revolutionary method developed by Kary Mullis in the 1980s. PCR is based on using the ability of DNA polymerase to synthesize new strand of DNA complementary to the offered template strand. Because DNA polymerase can add a nucleotide only onto a preexisting 3'-OH group, it needs a primer to which it can add the first nucleotide. This requirement makes it possible to delineate a specific region of template sequence that the researcher wants to amplify. At the end of the PCR reaction, the specific sequence will be accumulated in billions of copies (amplicons).

Components of PCR

DNA template

- the sample DNA that contains the target sequence. At the beginning of the reaction, high temperature is applied to the original double-stranded DNA molecule to separate the strands from each other.

DNA polymerase

- type of enzyme that synthesizes new strands of DNA complementary to the target sequence. The first and most commonly used of these enzymes is *Taq* DNA polymerase (from *Thermus aquaticus*), whereas *Pfu*DNA polymerase (from *Pyrococcus furiosus*) is used widely because of its higher fidelity when copying DNA. Although these enzymes are subtly different, they both have two capabilities that make them suitable for PCR: 1) they can generate new strands of DNA using a DNA template and primers, and 2) they are heat resistant.

Primers

- short pieces of single-stranded DNA that are complementary to the target sequence. The polymerase begins synthesizing new DNA from the end of the

primer.

Nucleotides (dNTPs or deoxynucleotide triphosphates)

• single units of the bases A, T, G, and C, which are essentially "building blocks" for new DNA strands.

The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA

fragments) containing sequences complementary to the target region along with a DNA polymerase, which the method is named after, are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase (an enzyme originally isolated from the bacterium *Thermus aquaticus*). This DNA polymerase enzymatically assembles a new DNA strand from DNA building-blocks, the nucleotides, by using single-stranded DNA as a template and DNA oligonucleotides (also called DNA primers), which are required for initiation of DNA synthesis. The vast majority of PCR methods use thermal cycling, i.e., alternately heating and cooling the PCR sample through a defined series of temperature steps.

Typically, PCR consists of a series of 20–40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2–3 discrete temperature steps, usually three (Figure below). The cycling is often preceded by a single temperature step at a high temperature (>90 °C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis, the concentration of divalent ions and dNTPs in the reaction, and the melting temperature (Tm) of the primers.

- *Denaturation step*: *This* step is the first regular cycling event and consists of heating the reaction to 94–98 °C for 20–30 seconds. It causes DNA melting of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single-stranded DNA molecules.

 Annealing step: *The* reaction temperature is lowered to 50–65 °C for 20–40 seconds allowing annealing of the primers to the single-stranded DNA template. This temperature must be low enough to allow for hybridization of the primer to the strand, but high enough for the hybridization to be specific, i.e., the primer should only bind to a perfectly complementary part of the template. If the temperature is too low, the primer could bind imperfectly. If it is too high, the primer might not bind. Typically the annealing temperature is about 3–5 °C below the Tm of the primers used. Stable DNA–DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA formation. It is very vital to determine the annealing temperature in PCR. This is because in PCR, efficiency and specificity are affected by the annealing temperature. An incorrect annealing temperature will cause an error in the test.

 Extension/elongation step: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75–80 °C, and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to amplify. As a rule-of-thumb, at its optimum temperature, the DNA polymerase polymerizes a thousand bases per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

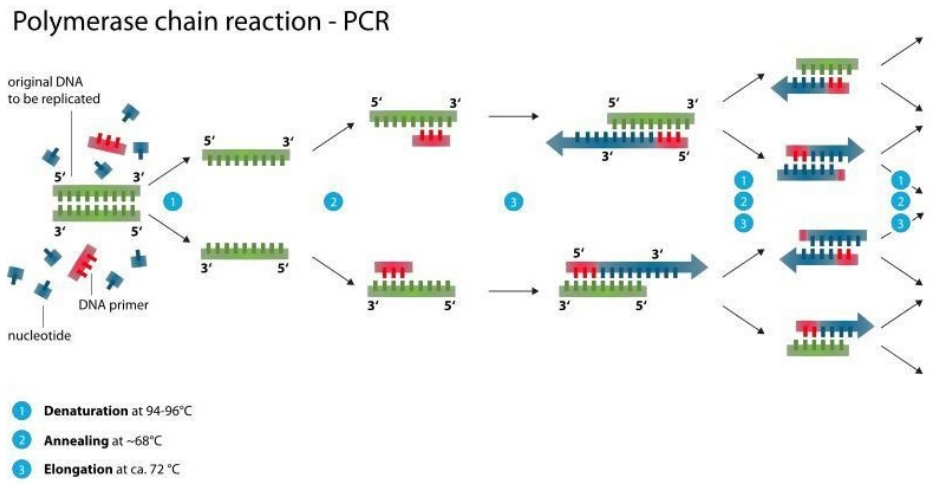


Fig. 9 PCR POLYMERASE CHAIN REACTION

Multiplex PCR

Multiplex polymerase chain reaction refers to the use of polymerase chain reaction to amplify several different DNA sequences simultaneously (as if performing many separate PCR reactions all together in one reaction). This process amplifies DNA in samples using multiple primers and

- temperature-mediated DNA polymerase in a thermal cycler. Multiplex-PCR consists of multiple primer sets within a single PCR mixture to produce amplicons of varying sizes that are specific to different DNA sequences. By targeting multiple sequences at once, additional information may be gained from a single test run that otherwise would require several times the reagents and more time to perform. Annealing temperatures for each of the primer sets must be optimized to work correctly within a single reaction, and amplicon sizes, i.e., their base pair length, should be different enough to form distinct bands when visualized by gel electrophoresis. Alternatively, if amplicon sizes overlap, the different amplicons may be differentiated and visualised using primers that have been dyed with different colour fluorescent dyes.

Nested PCR

Nested polymerase chain reaction is a modification of polymerase chain reaction intended to reduce non-specific binding in products due to the amplification of unexpected primer binding sites

---

Nested polymerase chain reaction involves two sets of primers, used in two successive runs of polymerase chain reaction, the second set intended to amplify a secondary target within the first run product.

1. The target DNA undergoes the first run of polymerase chain reaction with the first set of primers, shown in green. The selection of alternative and similar primer binding sites gives a selection of products, only one containing the intended sequence.
2. The product from the first reaction undergoes a second run with the second set of primers, shown in red. It is very unlikely that any of the unwanted PCR products contain binding sites for both the new primers, ensuring the product from the second PCR has little contamination from unwanted products of primer dimers, hairpins, and alternative primer target sequences.

RT-PCR

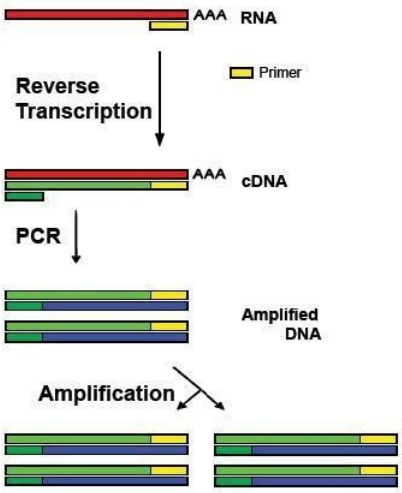
RT-PCR is used to qualitatively detect gene expression through creation of complementary DNA (cDNA) transcripts from RNA.

Although RT-PCR and the traditional PCR both produce multiple copies of particular DNA isolates through amplification, the applications of the two techniques are fundamentally different. Traditional PCR is used to exponentially amplify target DNA sequences. RT-PCR is used to clone expressed genes by reverse transcribing the RNA of interest into its DNA

complement through the use of reverse transcriptase. Subsequently, the newly synthesized cDNA is amplified using traditional PCR.

In addition to the qualitative study of gene expression, quantitative PCR can be utilized for quantification of RNA, in both relative and absolute terms, by incorporating qPCR into the technique. The combined technique, described as quantitative RT-PCR or real-time RT-PCR is often abbreviated as qRT-PCR, RT-qPCR, or RRT-PCR. Compared to other RNA quantification methods, such as northern blot, qRT-PCR is considered to be the most powerful, sensitive, and quantitative assay for the detection of RNA levels. It is frequently used in the expression analysis of single or multiple genes, and expression patterns for identifying infections and diseases.

Fig 10.RT-PCR



Real time PCR

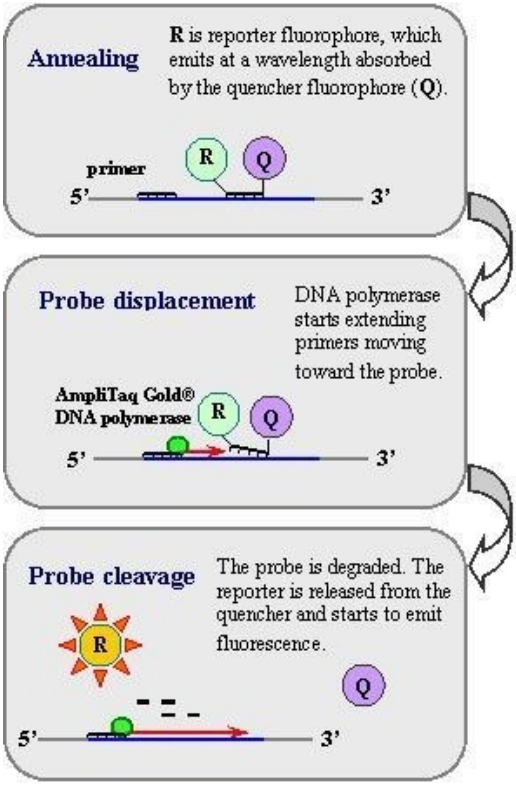
A real-time polymerase chain reaction is a laboratory technique ofmolecular biology based on the polymerase chain reaction (PCR). It monitors the amplification of a targeted DNA molecule during the PCR, i.e. in real-time, and not at its end, as in conventional PCR. Real-time PCR can be used quantitatively (Quantitative real-time PCR), semi-quantitatively, i.e. above/below a

certain amount of DNA molecules (Semi quantitative real-time PCR) or qualitatively (Qualitative real-time PCR).

TaqMan assay (named after *Taq* DNA polymerase) **was** one of the earliest methods introduced for real time PCR reaction monitoring and has been widely adopted for both the quantification of mRNAs and for detecting variation. The method exploits the 5' endonuclease activity of *Taq* DNA polymerase to cleave an oligonucleotide probe during PCR, thereby generating a

detectable signal. The probes are fluorescently labeled at their 5' end and are non-extendable at their 3' end by chemical modification. Specificity is conferred at three levels: via two PCR primers and the probe.

Fig 11.qRT-PCR



A DNA-binding dye binds to all double-stranded (ds) DNA in PCR, causing fluorescence of the dye. An increase in DNA product during PCR therefore leads to an increase in fluorescence

intensity measured at each cycle. However, dsDNA dyes such as SYBR Green will bind to all dsDNA PCR products, including nonspecific PCR products (such as Primer dimer). This can potentially interfere with, or prevent, accurate monitoring of the intended target sequence.

In real-time PCR with dsDNA dyes the reaction is prepared as usual, with the addition of fluorescent dsDNA dye. Then the reaction is run in a real-time PCR instrument, and after each cycle, the intensity of fluorescence is measured with a detector; the dye only fluoresces when bound to the dsDNA (i.e., the PCR product). This method has the advantage of only needing a pair of primers to carry out the amplification, which keeps costs down; however, only one target sequence can be monitored in a tube.

DNA fingerprinting

In the 1980's, the use of hypervariable regions of DNA called minisatellites was introduced. The molecular techniques associated with these regions, typically referred to as "DNA fingerprinting," have been a useful tool for pedigree analysis and in estimating the genetic variation in populations.

The chemical structure of everyone's DNA is the same. The only difference between people (or any animal) is the order of the base pairs. There are so many millions of base pairs in each person's DNA that every person has a different sequence.Using these sequences, every person could be identified solely by the sequence of their base pairs. However, because there are so many millions of base pairs, the task would be very time-consuming. Instead, scientists are able to use a shorter method, because of repeating patterns in DNA. These patterns do not, however, give an individual "fingerprint," but they are able to determine whether two

DNA samples are from the same person, related people, or non-related people. Scientists use a small number of sequences of DNA that are known to vary among individuals a great deal, and analyze those to get a certain probability of a match.

Random Amplified Polymorphic DNA (RAPD)

Random Amplified Polymorphic DNA (RAPD) markers are DNA fragments from PCR amplification of random segments of genomic DNA with single primer of arbitrary nucleotide sequence.

Unlike traditional PCR analysis, RAPD (pronounced "rapid") does not require any specific knowledge of the DNA sequence of the target organism: the identical 10-mer primers will or will not amplify a segment of DNA, depending on positions that are complementary to the primers' sequence. For example, no fragment is produced if primers annealed too far apart or 3' ends of the primers are not facing each other. Therefore, if a mutation has occurred in the template DNA at the site that was previously complementary to the primer, a PCR product will not be produced, resulting in a different pattern of amplified DNA segments on the gel.

Example

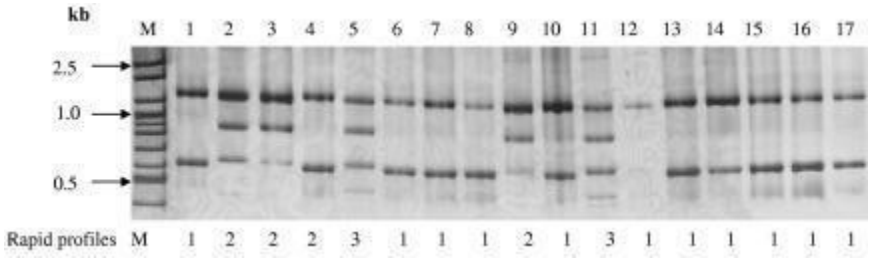


Fig 12.RAPD

RAPD is an inexpensive yet powerful typing method for many bacterial species.

Silver-stained polyacrylamide gel showing three distinct RAPD profiles generated by primer OPE15 for Haemophilus ducreyi isolates from Tanzania, Senegal, Thailand, Europe, and North America.

Selecting the right sequence for the primer is very important because different sequences will produce different band patterns and possibly allow for a more specific recognition of individual strains.

Limitations of RAPD

- Nearly all RAPD markers are dominant, i.e. it is not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies). Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, are detected only rarely.

- PCR is an enzymatic reaction, therefore the quality and concentration of template DNA, concentrations of PCR components, and the PCR cycling conditions may greatly influence the outcome. Thus, the RAPD technique is notoriously laboratory dependent and needs carefully developed laboratory protocols to be reproducible.

- Mismatches between the primer and the template may result in the total absence of PCR product as well as in a merely decreased amount of the product. Thus, the RAPD results can be difficult to interpret.

Restriction Fragment Length Polymorphism (RFLP)

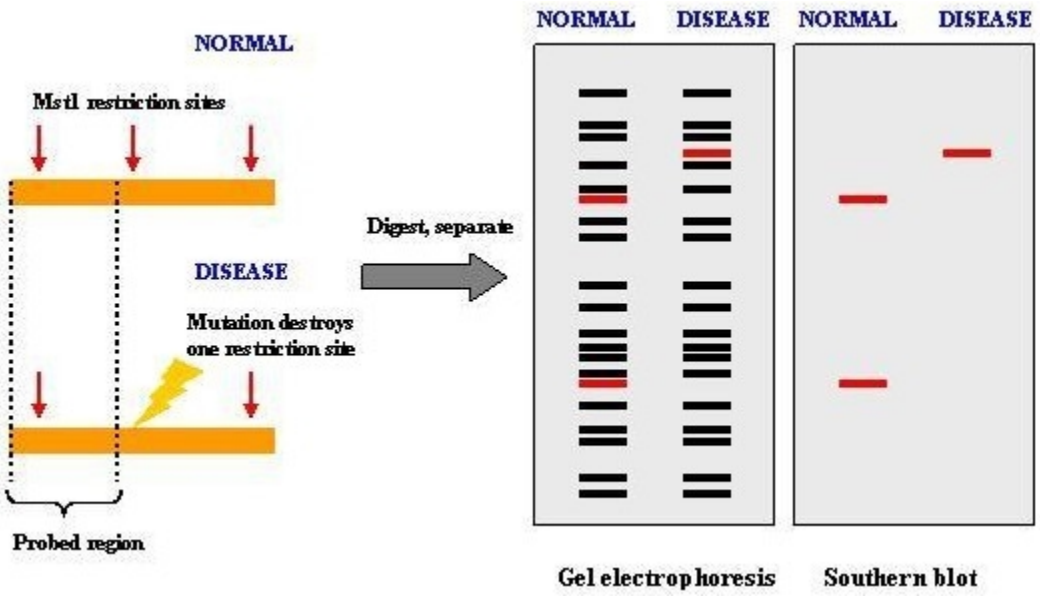
Restriction Fragment Length Polymorphism (RFLP) is a difference in homologous DNA sequences that can be detected by the presence of fragments of different lengths after digestion of the DNA samples in question with specific restriction endonucleases.RFLP, as a molecular marker, is specific to a single clone/restriction enzyme combination.

Most RFLP markers are co-dominant (both alleles in heterozygous sample will be detected) and highly locus-specific.

An RFLP probe is a labeled DNA sequence that hybridizes with one or more fragments of the digested DNA sample after they were separated by gel electrophoresis, thus revealing a unique blotting pattern characteristic to a specific genotype at a specific locus. Short, single- or low-copy genomic DNA or cDNA clones are typically used as RFLP probes.

The RFLP probes are frequently used in genome mapping and in variation analysis (genotyping, forensics, paternity tests, hereditary disease diagnostics, etc.).

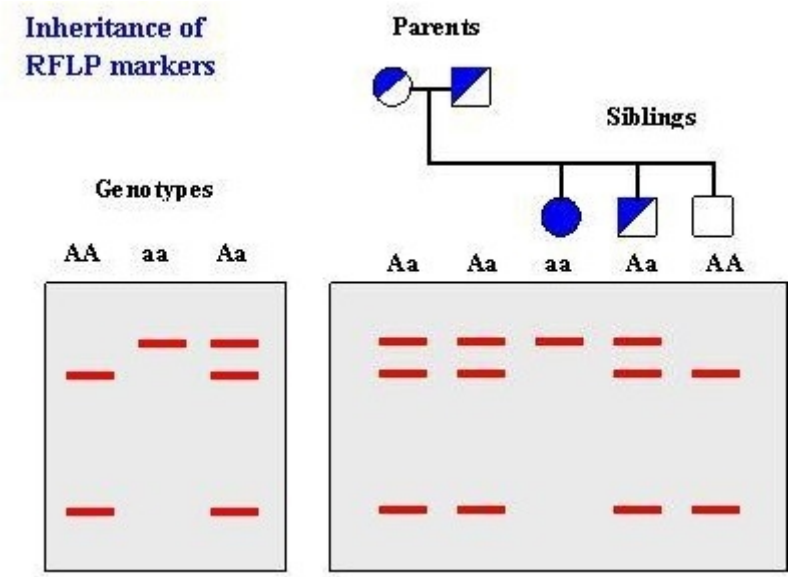
Fig 13.RFLP



SNPs or INDELs can create or abolish restriction endonuclease (RE) recognition sites, thus affecting quantities and length of DNA fragments resulting from RE digestion.

Genotyping

Fig 14.Inheritance of RFLP markers



Developing RFLP probes

- Total DNA is digested with a methylation-sensitive enzyme (for example, PstI), thereby enriching the library for single- or low-copy expressed sequences (PstI clones are based on the suggestion that expressed genes are not methylated).
- The digested DNA is size-fractionated on a preparative agarose gel, and fragments ranging from 500 to 2000 bp are excised, eluted and cloned into a plasmid vector (for example, pUC18).
- Digests of the plasmids are screened to check for inserts.
- Southern blots of the inserts can be probed with total sheared DNA to select clones that hybridize to single- and low-copy sequences.
- The probes are screened for RFLPs using genomic DNA of different genotypes digested with restriction endonucleases. Typically, in species with moderate to high polymorphism

rates, two to four restriction endonucleases are used such as EcoRI, EcoRV, and HindIII. In species with low polymorphism rates, additional restriction endonucleases can be tested to increase the chance of finding polymorphism.

Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphisms (AFLPs) are differences in restriction fragment lengths caused by SNPs or INDELsthat create or abolish restriction endonuclease recognition sites.

The AFLP technique is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA.

After final amplification, selectively amplified fragments are separated by gel electrophoresis and visualized autoradiographically. MseI-MseI fragments are excluded from the autorad because only EcoRI-directed primers are normally labeled. Typically, the autorad has 100-300 fingerprints with sizes ranging from 80 to 500 nucleotides. Only a subset (10-40) of these total bands is polymorphic between two related individuals, such as Arabidopsis thaliana Columbia and Landsberg erecta ecotypes.

Using 3-bp selective primer extensions gives 128 possible linker combinations. Therefore, 128 subsets of genomic DNA can be readily amplified. Thus, thousands of markers can be generated quite rapidly.

Weaknesses of AFLP

- Proprietary technology is needed to score heterozygotes and ++ homozygotes. Otherwise, AFLP must be dominantly scored.
- Developing locus-specific markers from individual fragments can be difficult.
- Need to use different kits adapted to the size of the genome being analyzed.

Sequence-Tagged Sites (STS)

Sequence-Tagged Site (STS)is a relatively short, easily PCR-amplified sequence (200 to 500 bp) which can be specifically amplified by PCR and detected in the presence of all other genomic sequences and whose location in the genome is mapped.

The STS concept was introduced by Olson et al (1989). In assessing the likely impact of the Polymerase Chain Reaction (PCR) on human genome research, they recognized that single-copy DNA sequences of known map location could serve as markers for genetic and physical mapping of genes along the chromosome. The advantage of STSs over other mapping landmarks is that the means of testing for the presence of a particular STS can be completely described as information in a database: anyone who wishes to make copies of the marker would simply look up the STS in the database, synthesize the specified primers, and run the PCR under specified conditions to amplify the STS from genomic DNA.

STS-based PCR produces a simple and reproducible pattern on agarose or polyacrylamide gel. In most cases STS markers are co-dominant, i.e., allow heterorozygotes to be distinguished from the two homozygotes.

The DNA sequence of an STS may contain repetitive elements, sequences that appear elsewhere in the genome, but as long as the sequences at both ends of the site are unique and conserved, researches can uniquely identify this portion of genome using tools usually present in any laboratory.

Thus, in broa

d sense, STS include such markers as microsatellites (SSRs, STMS or SSRPs), SCARs, CAPs,

and ISSRs.

Microsatellites

Polymorphic loci present in nuclear DNA and organellar DNA that consist of repeating units of 1-10 base pairs, most typically, 2-3 bp in length, also called Simple Sequence Repeats (SSR), Sequence-Tagged Microsatellite Sites (STMS) orSimple Sequence

Repeats Polymorphisms (SSRP). SSRs are highly variable and evenly distributed throughout the genome. This type of repeated DNA is common in eukaryotes. These polymorphisms are identified by constructing PCR primers for the DNA flanking the microsatellite region. The flanking regions tend to be conserved within the species, although sometimes they may also be conserved in higher taxonomic levels.

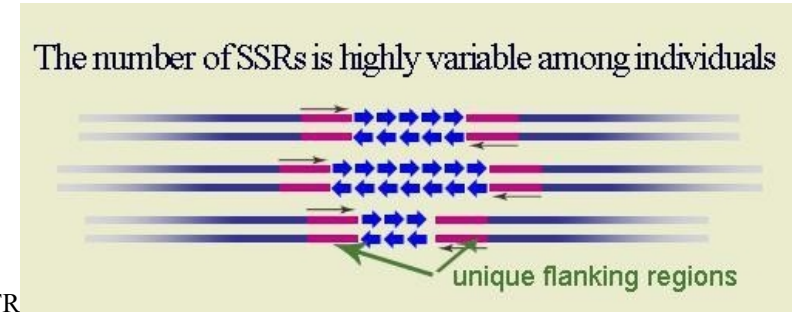
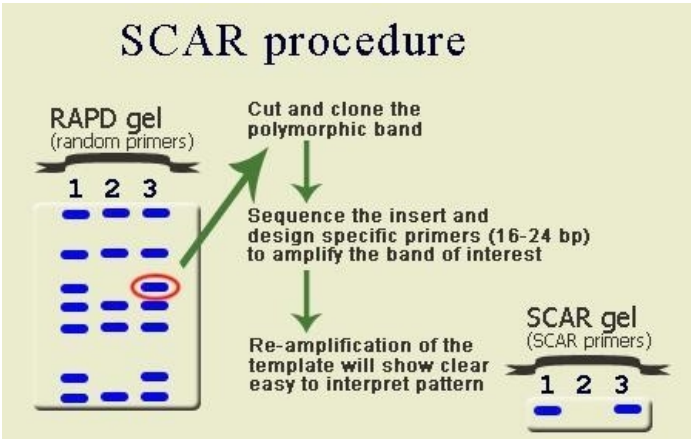


Fig 15.UFR

Sequence Characterized Amplified Region (SCAR)

DNA fragments amplified by the Polymerase Chain Reaction (PCR) using specific 15-30 bp primers, designed from nucleotide sequences established in cloned RAPD (Random Amplified Polymorphic DNA) fragments linked to a trait of interest. By using longer PCR primers, SCARs do not face the problem of low reproducibility generally encountered with RAPDs. Obtaining a co-dominant marker may be an additional advantage of converting RAPDs into SCARs.



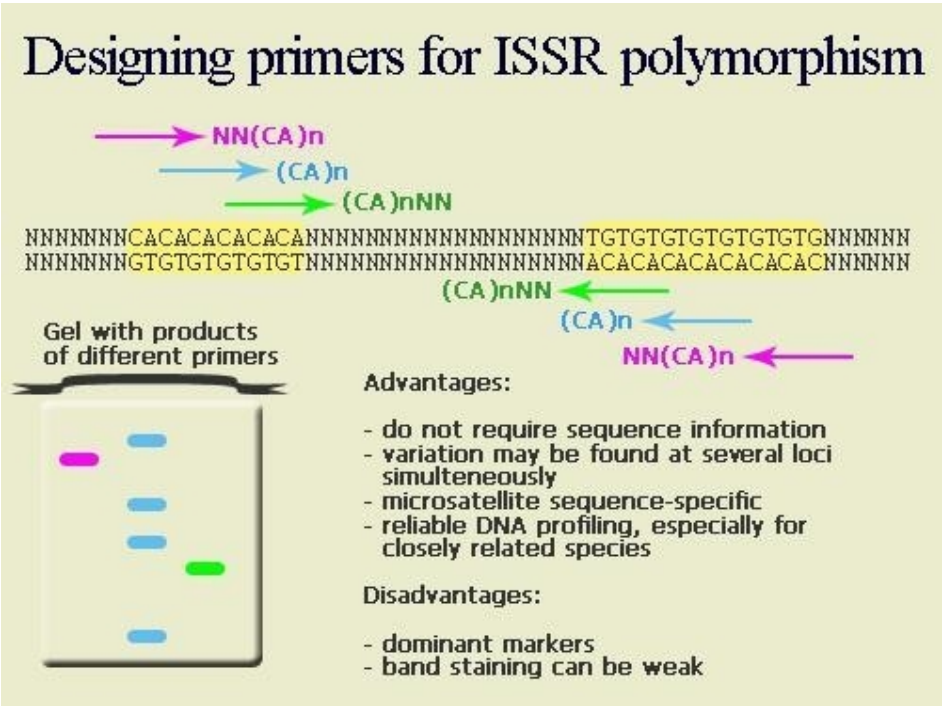
Cleaved Amplified Polymorphic Sequences (CAPS)

STS polymorphisms that can be detected by differences in restriction fragment lengths caused by SNPs or INDELs that create or abolish restriction endonuclease recognition sites in PCR amplicons produced by locus-specific oligonucleotide primers. In other words this technique aims to convert and amplified band that does not show variation by length of PCR product into a polymorphic one. More about CAPS in Overview of CAPS technology.

Inter-simple Sequence Repeats (ISSRs)

STS polymorphisms that are found between microsatellite repeats. Primers can be designed based on a microsatellite repeats exclusively, in which case this technique will target multiple loci due to known abundance of repeat sequences in the genome. Alternatively, primers can be extended outside or inside the ISSR in which case a unique region most likely will be amplified.

Fig 16. Designing primers for ISSr Polymorphism



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

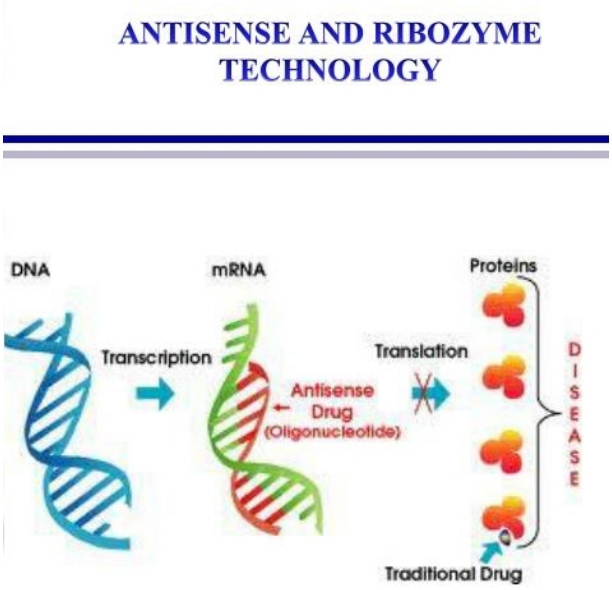
UNIT – 5 – GENETIC ENGINEERING – SBTA1501

Strategies in plant and Animal Genetic Engineering:

Antisense Technology

Among several possible strategies for inactivating a single, chosen gene the most approved one is Antisense technology. This modification does not involve actual subtraction but inactivation of gene or suppressing gene activity. When DNA strand is normally transcribed the RNA produced is called *sense RNA*, complementary to DNA but if orientation of gene to be transcribed is reversed with respect to promoter the RNA transcribed from it would be reversed too, the RNA produced so has sequence same to that of antisense strand of the normal gene, thus known as *Antisense RNA (asRNA).*

Fig : 1 Antisense RIBOZYME technology



The Basis of Antisense Technology is the use of antisense RNA, the gene which is to be cloned is ligated into the vector in reverse orientation thus on transcription the RNA synthesized is reverse complement of the mRNA transcribed from normal version of gene and being complementary to each other they will pair to form hybrid.

The application of Antisense technology is seen most in plant genetic engineering.

1) **Slow Ripening Tomato:** In Tomato in the later stages of ripening polygalactouronase gene is switched on coding for polygalactouronase enzyme which breaks down polygalactouronic acid

2) component of cell walls resulting in softening which results in spoilt tomato. Transgenic tomatoes were prepared containing antisense construct of the gene PG resulting in reduced expression of PG and slow ripening and fruit softening, improving shelf life.

2) **Inactivate ethylene synthesis**

3) **Modification of flower color in decorative plants**

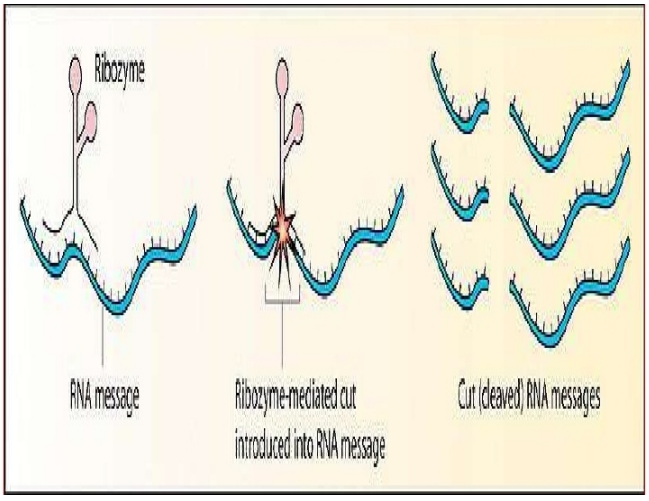
More application will undoubtedly follow in future.

Ribozyme Technology

Stating that all enzymes are proteins is wrong. In 1990, Tom Cech and Sidney Altman shared the Nobel Prize for their demonstration that RNA could act as an enzyme. **Ribozymes are antisense RNA molecules with enzymatic properties**, or RNAs that, like those in ribonuclease P, make up part of an enzyme that contains other compounds as well. They function by binding to the target RNA moiety through base pairing and inactivate it by cleaving the phosphodiester backbone at a specific cutting side. The various types of reactions performed by ribozymes are based on transesterification; these reactions include

Splicing, oligonucleotide chain extension, RNA ligation, endonuclease action and phosphatase action. The ribozyme action is generated by formation of particular secondary and tertiary structures that create active sites. A ribozyme has two sites: a substrate binding site and a guanosine-binding site.

Fig 2; 3D strctures of ROBOZYMES



Five classes of ribozymes have been described based on sequences as well as three-dimensional structures. They are:

1) *Tetrahymena* group I intron

2) Rnase P

3) Hammerhead ribozyme

4) Hairpin ribozyme

5) Hepatitis delta virus ribozyme

Ribozymes have the potential to become useful therapeutic agents and currently the vast majority of effort has been expended in the development of trans cleaving hammerhead and hairpin ribozymes as inhibitors of viral gene expression, in particular to cleave and destroy HIV-I RNAs to inhibit viral replication in infected cells. RNA enzymes can potentially be quite useful for a variety of gene therapy applications.

Human genome project and its application

The HGP is a multinational consortium established by government research agencies and funded publicly.Celera Genomics is a private company whose former CEO, J. Craig Venter,

ran an independent sequencing project.Differences arose regarding who should receive the credit for this scientific milestone .June 6, 2000, the HGP and Celera Genomics held a joint press conference to announce that TOGETHER they had completed ~97% of the human genome

Goals established for the Human Genome Project when it began in 1990

- Identify all of the genes in human DNA.
- Determine the sequence of the 3 billion chemical nucleotide bases that make up human DNA.
- Store this information in data bases.
- Develop faster, more efficient sequencing technologies.
- Develop tools for data analysis.
- Address the ethical, legal, and social issues (ELSI) that arise from the project.

The sequencing of the human genome holds benefits for many fields, from molecular medicine to human evolution. The Human Genome Project, through its sequencing of the DNA, can help us understand diseases including: genotyping of specific viruses to direct appropriate treatment; identification of mutations linked to different forms of cancer; the design of medication and more accurate prediction of their effects; advancement in forensic applied sciences; biofuels and other energy applications; agriculture, animal husbandry, bioprocessing; risk assessment; bioarcheology, anthropology andevolution. Another proposed benefit is the commercial development of genomics research related to DNA based products, a multibillion-dollar industry.

The sequence of the DNA is stored in databases available to anyone on the Internet. The U.S. National Center for Biotechnology Information (and sister organizations in Europe and Japan) house the gene sequence in a database known as GenBank, along with sequences of known and hypothetical genes and proteins. Other organizations, such as the UCSC Genome Browser at the University of California, Santa Cruz, and Ensembl present additional data and annotation and powerful tools for visualizing and searching it. Computer programs have been developed to analyze the data, because the data itself is difficult to interpret without such programs.

Gene therapy

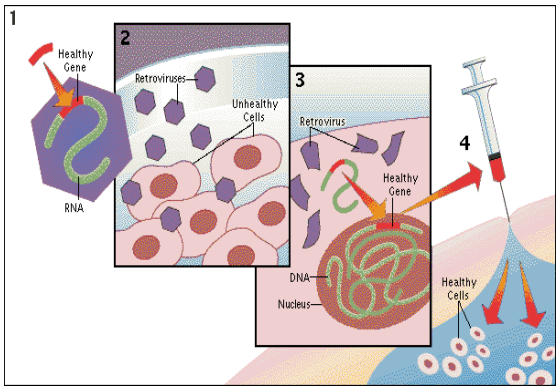
It is a technique for correcting defective genes that are responsible for disease development.There are four approaches:

1. **A normal gene inserted to compensate for a nonfunctional gene.**
2. **An abnormal gene traded for a normal gene**
3. **An abnormal gene repaired through selective reverse mutation**
4. **Change the regulation of gene pairs**

Steps involved;

1. **A vector delivers the therapeutic gene into a patient's target cell**
2. **The target cells become infected with the viral vector**
3. **The vector's genetic material is inserted into the target cell**
4. **Functional proteins are created from the therapeutic gene causing the cell to return to a normal state.**

Fig 3.Adenoviral method gene theraphy



Non-viral Options

1. **Direct introduction of therapeutic DNA**
    a. **But only with certain tissue**

b. **Requires a lot of DNA**

2. **Creation of artificial lipid sphere with aqueous core, liposome**
    a. **Carries therapeutic DNA through membrane**
3. **Chemically linking DNA to molecule that will bind to special cell receptors**
    a. **DNA is engulfed by cell membrane**
    b. **Less effective**
4. **Trying to introduce a 47th chromosome**
    a. **Exist alongside the 46 others**
    b. **Could carry a lot of information**
    c. **But how to get the big molecule through membranes?**

Current Status

FDA hasn't approved any human gene therapy product for sale

Reasons:

- **In 1999, 18-year-old Jesse Gelsinger died from multiple organ failure 4 days after treatment for omithine transcarboxylase deficiency.Death was triggered by severe immune response to adenovirus carrier .**
- **January 2003, halt to using retrovirus vectors in blood stem cells because children developed leukemia-like condition after successful treatment for X-linked severe combined immunodeficiency disease**

Problems with Gene Therapy

- **Short Lived**
    - **Hard to rapidly integrate therapeutic DNA into genome and rapidly dividing nature of cells prevent gene therapy from long time**
    - **Would have to have multiple rounds of therapy**

- **Immune Response**
  - **new things introduced leads to immune response**
  - **increased response when a repeat offender enters**

- **Viral Vectors**
  - **patient could have toxic, immune, inflammatory response**
  - **also may cause disease once inside**
- **Multigene Disorders**
  - **Heart disease, high blood pressure, Alzheimer's, arthritis and diabetes are hard to treat because you need to introduce more than one gene**
- **May induce a tumor if integrated in a tumor suppressor gene because insertional mutagenesis**

DNA vaccine

DNA vaccination is a technique for protecting an organism against disease by injecting it with genetically engineered DNA to produce an immunological response. Nucleic acid vaccines are still experimental, and have been applied to a number of viral, bacterial and parasitic models of disease, as well as to several tumour models. DNA vaccines have a number of advantages over conventional vaccines, including the ability to induce a wider range of immune response types.

Vaccines are among the greatest achievements of modern medicine – in industrial nations, they have eliminated naturally-occurring cases of smallpox, and nearly eliminated polio, while other diseases, such as typhus, rotavirus, hepatitis A and B and others are well controlled. Conventional vaccines, however, only cover a small number of diseases, and infections that lack effective vaccines kill millions of people every year, with AIDS, hepatitis C and malaria being particularly common.

DNA vaccines are *third generation vaccines*, and are made up of a small, circular piece of bacterial DNA (called a plasmid) that has been genetically engineered to produce one or two specific proteins (antigens) from a micro-organism. The vaccine DNA is injected into the cells of

the body, where the "inner machinery" of the host cells "reads" the DNA and converts it into pathogenic proteins. Because these proteins are recognised as foreign, they are processed by the host cells and displayed on their surface, to alert the immune system, which then triggers a range of immune responses. These DNA vaccines developed from “failed” gene therapy experiments. The first demonstration of a plasmid-induced immune response was when mice inoculated with a plasmid expressing human growth hormone elicited antibodies instead of altering growth.

Thus far, few experimental trials have evoked a response sufficiently strong enough to protect against disease, and the usefulness of the technique, while tantalizing, remains to be conclusively proven in human trials. However, in June 2006 positive results were announced for a bird flu DNA vaccine and a veterinary DNA vaccine to protect horses from West Nile virus has been approved In August 2007, a preliminary study in DNA vaccination against multiple sclerosis was reported as being effective.

Advantages and disadvantages of DNA vaccines

Advantages	**Disadvantages**
▪ Subunit vaccination with no risk for infection ▪ Antigen presentation by both MHC class I and class II molecules ▪ Able to polarise T-cell help toward type 1 or type 2 ▪ Immune response focused only on antigen of interest ▪ Ease of development and production ▪ Stability of vaccine for storage and shipping ▪ Cost-effectiveness ▪ Obviates need for peptide synthesis, expression and purification of recombinant proteins and the use of toxic adjuvants ,Long-term persistence of immunogen ▪ In vivo expression ensures protein more closely resembles normal eukaryotic structure, with accompanying post-translational modifications	▪ Limited to protein immunogens ▪ Potential for atypical processing of bacterial and parasite proteins

Plasmid vectors for use in vaccination

Vector design

DNA vaccines elicit the best immune response when highly active expression vectors are used. These are plasmids which usually consist of a strong viral promoter to drive the in vivo transcription and translation of the gene (or complementary DNA) of interest Intron A may sometimes be included to improve mRNA stability and hence increase protein expression.

Plasmids also include a strong polyadenylation/transcriptional termination signal, such as bovine growth hormone or rabbit beta-globulin polyadenylation sequences. Multicistronic vectors are sometimes constructed to express more than one immunogen, or to express an immunogen and an immunostimulatory protein.

Because the plasmid is the "vehicle" from which the immunogen is expressed, optimising vector design for maximal protein expression is essential. One way of enhancing protein expression is by optimising the codon usage of pathogenic mRNAs for eukaryotic cells. Pathogens often have different AT contents than the species being immunized, so altering the gene sequence of the immunogen to reflect the codons more commonly used in the target species may improve its expression.

Another consideration is the choice of promoter. The SV40 promoter was conventionally used until research showed that vectors driven by the Rous Sarcoma Virus (RSV) promoter had much higher expression rates. More recently, expression rates have been further increased by the use of the cytomegalovirus (CMV) immediate early promoter. Inclusion of the Mason-Pfizer monkey virus (MPV)-CTE with/without rev increased envelope expression. Furthermore the CTE+rev construct was significantly more immunogenic then CTE alone vector. Additional modifications to improve expression rates have included the insertion of enhancer sequences, synthetic introns, adenovirus tripartite leader (TPL) sequences and modifications to the polyadenylation and transcriptional termination sequences.

Vaccine insert design

Immunogens can be targeted to various cellular compartments in order to improve antibody or cytotoxic T-cell responses. Secreted or plasma membrane-bound antigens are more effective at

inducing antibody responses than cytosolic antigens, while cytotoxic T-cell responses can be improved by targeting antigens for cytoplasmic degradation and subsequent entry into the major histocompatibility complex (MHC) class I pathway. This is usually accomplished by the addition of N-terminal ubiquitin signals

The conformation of the protein can also have an effect on antibody responses, with "ordered" structures (like viral particles) being more effective than unordered structures. trings of minigenes (or MHC class I epitopes) from different pathogens are able to raise cytotoxic T-cell responses to a number of pathogens, especially if a TH epitope is also included.

DNA vaccines have been introduced into animal tissues by a number of different methods. The two most popular approaches are injection of DNA in saline, using a standard hypodermic needle, and gene gun delivery. A schematic outline of the construction of a DNA vaccine plasmid and its subsequent delivery by these two methods into a host is illustrated at Scientific American. njection in saline is normally conducted intramuscularly (IM) in skeletal muscle, or intradermally (ID), with DNA being delivered to the extracellular spaces. This can be assisted by electroporationby temporarily damaging muscle fibres with myotoxins such as bupivacaine; or by using hypertonic solutions of saline or sucrose Immune responses to this method of delivery can be affected by many factors, including needle type, eedle alignment, speed of injection, volume of injection, muscle type, and age, sex and physiological condition of the animal being injected. Gene gun delivery, the other commonly used method of delivery, ballistically accelerates plasmid DNA (pDNA) that has been adsorbed onto gold or tungsten microparticles into the target cells, using compressed helium as an accelerant.

Alternative delivery methods have included aerosol instillation of naked DNA on mucosal surfaces, such as the nasal and lung mucosa, and topical administration of pDNA to the eye and vaginal mucosa. Mucosal surface delivery has also been achieved using cationic liposome-DNA preparations, biodegradable microspheres, attenuated *Shigella* or *Listeria* vectors for oral administration to the intestinal mucosa and recombinant adenovirus vectors.

The method of delivery determines the dose of DNA required to raise an effective immune response. Saline injections require variable amounts of DNA, from 10 μg-1 mg, whereas gene gun deliveries require 100 to 1000 times less DNA than intramuscular saline injection to raise an effective immune response. Generally, 0.2 μg – 20 μg are required, although quantities as low as 16 ng have been reported. These quantities vary from species to species, with mice, for example,

requiring approximately 10 times less DNA than primates. Saline injections require more DNA because the DNA is delivered to the extracellular spaces of the target tissue (normally muscle), where it has to overcome physical barriers (such as the basal lamina and large amounts of connective tissue, to mention a few) before it is taken up by the cells, while gene gun deliveries bombard DNA directly into the cells, resulting in less "wastage".

Another approach to DNA vaccination is expression library immunization (ELI). Using this technique, potentially all the genes from a pathogen can be delivered at one time, which may be useful for pathogens which are difficult to attenuate or culture. ELI can be used to identify which of the pathogen's genes induce a protective response. This has been tested with *Mycoplasma pulmonis*, a murine lung pathogen with a relatively small genome, and it was found that even partial expression libraries can induce protection from subsequent challenge.

Table : 1 Summary of Plasmid DNA delivery methods

Method of Delivery		**Formulation of DNA**	**Target Tissue**	**Amount of DNA**
Parenteral	Injection (hypodermic needle)	Aqueous solution in saline	IM (skeletal); ID; (IV, subcutaneous and intraperitoneal with variable success)	Large amounts (approximately 100-200 µg)
	Gene Gun	DNA-coated gold beads	ED (abdominal skin); vaginal mucosa; surgically exposed muscle and other organs	Small amounts (as little as 16 ng)
	Pneumatic (Jet) Injection	Aqueous solution	ED	Very high (as much as 300 µg)
Topical application		Aqueous solution	Ocular; intravaginal	Small amounts (up to 100 µg)

Cytofectin-mediated	Liposomes (cationic); microspheres; recombinant adenovirus vectors; attenuated Shigella vector; aerosolised cationic lipid formulations	IM; IV (to transfect tissues systemically); intraperitoneal; oral immunization to the intestinal mucosa; nasal/lung mucosal membranes	variable

Advantages and disadvantages of commonly used DNA vaccine delivery methods

Method of Delivery	Advantage	Disadvantage
Intramuscular or Intradermal injection	▪ No special delivery mechanism ▪ Permanent or semi-permanent expression ▪ pDNA spreads rapidly throughout the body	▪ Inefficient site for uptake due to morphology of muscle tissue ▪ Relatively large amounts of DNA used ▪ Th1 response may not be the response required
Gene Gun	▪ DNA bombarded directly into cells ▪ Small amounts DNA	▪ Th2 response may not be the response required ▪ Requires inert particles as carrier
Jet injection	▪ No particles required ▪ DNA can be delivered to cells mm to cm below skin surface	▪ Significant shearing of DNA after high-pressure expulsion ▪ 10-fold lower expression, and lower immune response ▪ Requires large amounts of DNA (up to 300 µg)
Liposome-mediated delivery	▪ High levels of immune response can be generated ▪ Can increase transfection of intravenously delivered pDNA ▪ Intravenously delivered	▪ Toxicity ▪ Ineffectiveness in serum ▪ Risk of disease or immune reactions

	liposome-DNA complexes can potentially transfect all tissues - Intranasally delivered liposome-DNA complexes can result in expression in distal mucosa as well as nasal muscosa and the generation of IgA antibodies	

Transgenic plants

Transgenic plants are crops which have been genetically modified with genes from another organism to make the plants more agriculturally productive. Transgenic plants are only those with genes from other species, whereas genetically modified plants can have both new genes and a re-arrangement of the genes already found in the plant. Traditional breeding methods are one form of genetic modification.

Transgenic plants have been developed for a variety of reasons: longer shelf life, disease resistance, herbicide resistance, pest resistance, and improved product quality. The first transgenic crop approved for sale in the US, in 1994, was the FlavrSavr tomato, which was intended to have a longer shelf life. There are many controversial issues surrounding the use of transgenic crops. One of the most far-reaching issues is what could happen if these crop plants were to 'escape' from the fields and enter into the environment. This article addresses the potential effects of transgenic plants on their wild relatives, rather than their possible effects on other types of organisms, like Monarch butterflies.

Today there are more than 67.7 million hectares (677,000 km²) of transgenic plants being grown throughout the world[1]. There are three general types of transgenic plants; those with genes to improve the quality of the product, those with genes to allow them to resist disease or herbivory (consumption by herbivores, usually insects), and plants with genes that allow them to be resistant to the effects of specific herbicides.

Transgenic crops are grown worldwide, although the greatest concentration of transgenic crops is

in the United States, at 63% of the world total in 2003. At that time, 81% of the soybeans, 73% of thecotton and 40% of the corn being grown were transgenic. At that time most of the transgenic crops had genes either for herbicide resistance or for insect resistance.

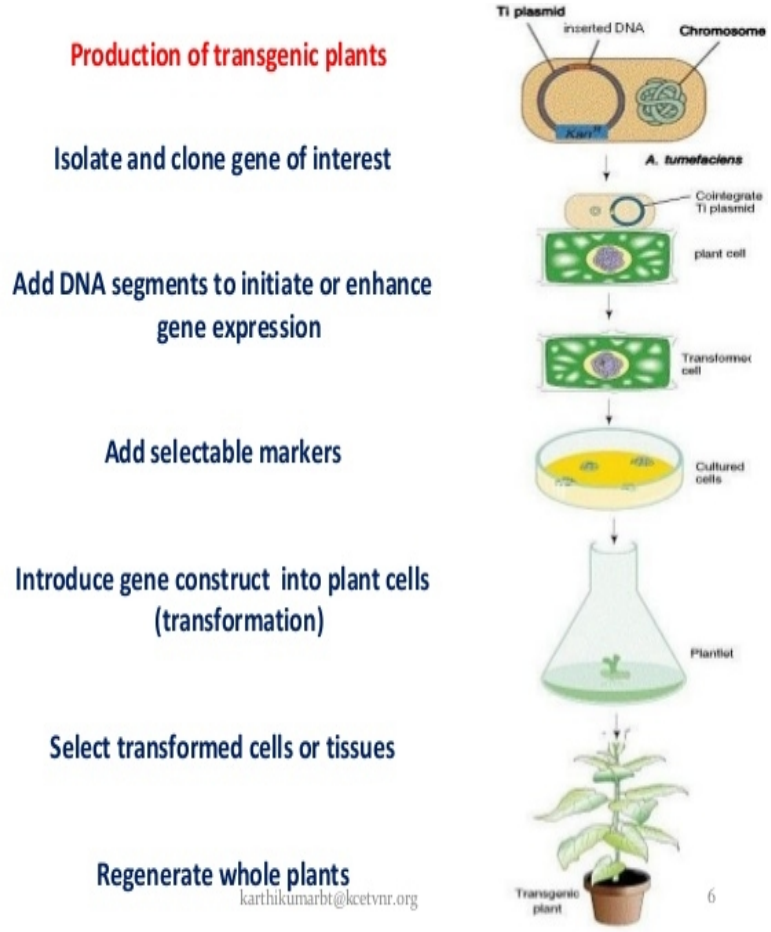


Fig 3 Production of Transgenic plants

Transgenic animals are **animals** (most commonly mice) that have had a foreign gene deliberately inserted into their genome. Such **animals** are most commonly created by the micro-injection of DNA into the pronuclei of a fertilised egg which is subsequently implanted into the oviduct of a pseudopregnant surrogate mother.

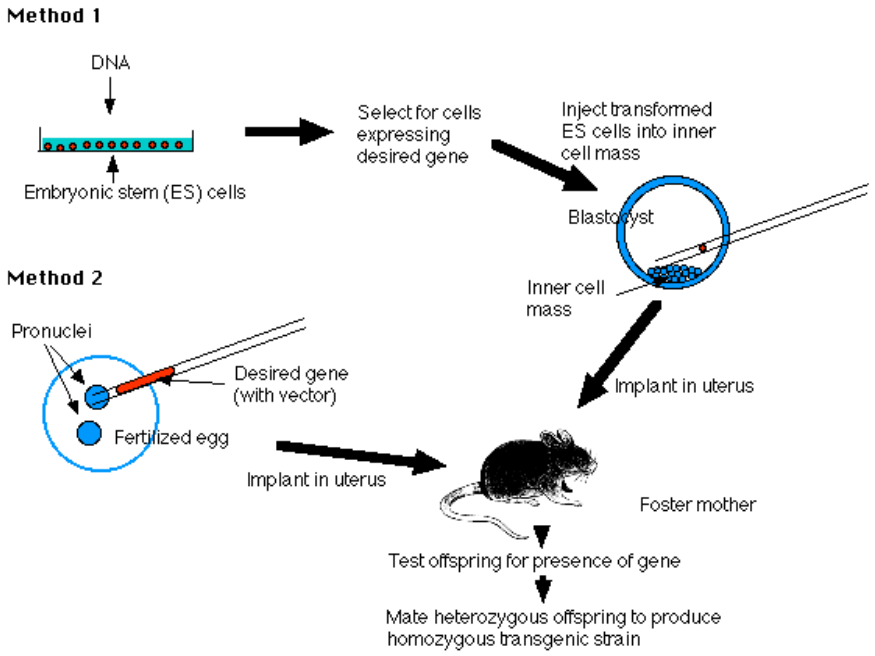


Fig 4,Production of Transgenic animals

Gene manipulation technology is the most important tool considered as the back bone of modern biotechnology. Presently diverse techniques are involved in the production of insulin, growth hormone and monoclonal antibodies. These are the modern medicines produced by the genetically engineered organisms (FDA approved GRAS –generally regarded as safe organisms). Production of human insulin by recombinant *E. coli* is considered as a significant outcome of recombinant DNA technology, more complex proteins of medical uses can also be produced by metabolic and cellular engineering of microorganisms. But production of proteins and other derivatives in its native, functional and intrinsic condition is the ultimate challenge of recombinant technology.

Production of Insulin:

Insulin is a peptide hormone mainly used in treatment of diabetes mellitus to control elevated blood glucose level. Banting and Best named it originally as 'isletin' and was later renamed as insulin by Macleod, a word that had been suggested in 1910. This hormone is secreted by the β-cells of the pancreas and consists of two polypeptide chains, A and B which are linked by two inter-chain and one intra-chain disulphide bridge. Insulin is synthesized as a single-chain precursor, pro-insulin, and produced by the proteolytic processing of pro-insulin in the pancreas .

Originally insulin was first identified from dog pancreas which was commercially produced from various sources like foetal calf pancreas obtained from slaughter houses. Now human insulin protein is mass-produced through geneticengineering processes. Recombinant DNA technology has been a great enabler in producing human insulin outside the body for being used as a therapeutic. Insulin is the first human hormone produced in bacteria to be tested in humans for medical purposes.

There are many methods for the production of recombinant human insulin in both bacteria and yeast. One typical scheme for preparing human insulin utilizes pro-insulin that is produced in *E. coli* cytoplasm as an inclusion body of a fusion protein.

Manufacturing of insulin using microbes as a cell factory involves the following steps –

1. Isolation of gene: The gene for producing human insulin protein is isolated.
2. Preparation of target DNA: Circular piece of DNA called plasmid is obtained from bacteria.
3. Insertion of DNA into plasmid: The gene for insulin is inserted into the plasmid construct. The human insulin gene is now recombined with bacterial DNA.
4. Plasmid insertion: The bacterial DNA having insulin gene is inserted back into bacteria.
5. Plasmid multiplication: The bacterial cells having insulin gene are allowed to grow and multiply and during this process bacterial cells start to produce recombinant insulin. During division newly synthesized copy of cell are produced.
6. Human insulin produced by bacteria is purified.

Production of Growth Hormones:

Growth hormone is one of the most important hormones in human body. The core center for production of growth hormone is pituitary gland. The action of growth hormone is either direct or indirect on the human physiological process. But in some children, malfunction of growth hormone results in abnormal growth of the individual. In case of these conditions recombinant growth hormone is useful for the treatment.

Human growth hormone has versatile functions:

• Activates the production of protein in cells by releasing some essential factors.

• Helps in fastening the production of DNA and RNA.

• Accelerates the generation of red blood cells and augments the flow of blood to the kidneys and the rate at which the kidney does its vital filtration work.

• Plays a major role in maintaining the level of fats in the body.

Table 2: Activates bone growth and skeletal development indirectly by producing intermediate factor IGF-1.

Hormone	**Production host**	**Engineering approach**
Gonadotropin-releasing hormone	*Escherichia coli*	Heterologous expression of the recombinant gonadotropin-releasing hormone in *E. coli* using a T7 RNA polymerase-based expression system and evaluation of various culture conditions on the plasmid stability and the product yield.
Human growth hormone	*Escherichia coli*	Activation of the promoter lambda PL by temperature shift for production of human growth hormone without contaminants.
Human parathyroid hormone	*Escherichia coli*	Using recombinant *E. coli* strain BL21 (DE3) harboring the plasmid. pET32aBI1 encoding the fusion gene of thioredoxin and human parathyroid hormone

GUIDELINES

Definition of recombinant DNA: Recombinant deoxyribonucleic acid (rDNA) by definition involves *in vitro* introduction of different segments of DNA (one being the vector and the others normally unrelated DNA sequences) that are capable of replication in a host cell either

autonomously or as an integral part of host's genome and maintenance of their continued propagation. This will include all types of cell fusion, microinjection of DNA or RNA or parts or all of chromosomes, genetic engineering including self cloning and deletion as well as cell hybridation, transformation and other types of virus or pathogen introduction into unnatural hosts.The organisms involved may belong to these categories:

i) Intergeneric organisms

ii) Well defined organisms with non-coding regulatory regions

i) Biological agents whose source of DNA is a pathogen

ii) Organisms that are generally recognised as non-pathogenic and may imbibe the characteristics of a pathogen on genetic manipulation.

Classification of a pathogenic microorganisms

The classification of infective microorganisms are drawn up under 4 risk groups in increasing order of risk based on the following parameters:

- pathogenecity of the agent
- modes of transmission and host range of the agent
- availability of effective preventive treatments or curative medicines
- capability to cause diseases to humans/animals/plants
- epidemic causing strains in India

The above mentioned parameters may be influenced by levels of immunity, density and movement of host population, presence of vectors for transmission and standards of environmental hygiene.

An inventory of pathogenic organisms classified in different groups is provided in Chapter V: A1. The scientific considerations for assessment of potential risks in handling of pathogenic organisms include the following:

Characterisation of donor and recipient organisms

i) Characterisation of the modified organism

ii) Expression and properties of the gene product

Based on the risk assessment information, the probability of risk could be further assigned

certain quantitative values (Chapter V: A7) for categorisation of experiments in terms of the following:

i) access factor of the organism
ii) expression factor of DNA
iii) damage factor of the Biologically active substance

Containment

Containment facilities for different Risk Groups as per the recommendations of World Health Organization (WHO).The term "Containment" is used in describing the safe methods for managing infectious agents in the laboratory environment where they are being handled or maintained.

Purpose of containment

To reduce exposure of laboratory workers, other persons, and outside environment to potentially hazardous agents.

Types of containment

Biological containment (BC): In consideration of biological containment, the vector (plasmid, organelle, or virus) for the recombinant DNA and the host (bacterial, plant, or animal cell) in which the vector is propagated in the laboratory will be considered together. Any combination of vector and host which is to provide biological containment must be chosen or constructed to limit the infectivity of vector to specific hosts and control the host-vector survival in the environment. These have been categorized into two levels - one permitting standard biological containment and the other even higher that relates to normal and disabled host-vector systems respectively (Chapter V: A3).

Physical Containment (PC): The objective of physical containment is to confine recombinant organisms thereby preventing the exposure of the researcher and the environment to the harmful agents. Physical containment is achieved through the use of i) Laboratory Practice, ii) Containment Equipment, and iii) Special Laboratory Design. The protection of personnel and the immediate laboratory environment from exposure to infectious agents, is provided by good microbiological techniques and the use of

appropriate safety equipment, (Primary Containment).

The protection of the environment external to the laboratory from exposure to infectious materials, is provided by a combination of facility design and operational practices, (Secondary Containment).

Elements of Containment: The three elements of containment include laboratory practice and technique, safety equipment and facility design.

i) Laboratory practice and technique**:**

- Strict adherence to standard microbiological practices and techniques
- Awareness of potential hazards
- Providing/arranging for appropriate training of personnel
- Selection of safety practices in addition to standard laboratory practices if required
- Developing of adopting a biosafety or operations manual which identifies the hazards

ii) Safety equipment (*primary barriers*): Safety equipment includes biological safety cabinets and a variety of enclosed containers (e.g. safety centrifuge cup). The biological safety cabinet (BSC) is the principal device used to provide containment of infectious aerosols generated by many microbiological procedures. Three types of BSCs (Class I, II, III) are used in microbiological laboratories. Safety equipment also includes items for personal protection such as gloves, coats, gowns, shoe covers, boots, respirators, face shields and safety glasses, etc.

iii) Facility Design (*Secondary barriers*): The design of the facility is important in providing a barrier to protect persons working in the facility but outside of the laboratory and those in the community from infectious agents which may be accidentally released from the laboratory. There are three types of facility designs: viz, the Basic Laboratory (for Risk Group I and II), the Containment Laboratory (for Risk Group III) and the Maximum Containment Laboratory (for Risk Group IV).

Bio-safety levels: It consists of a combination of laboratory practices and techniques,

safety equipment and laboratory facilities appropriate for the operations performed and the hazard posed by the infectious agents. The guidelines for Microbiological and Biomedical Laboratories suggest four Biosafety levels in incremental order depending on the nature of work. Additional flexibility in containment levels can be obtained by combination of the physical with the biological barriers. The proposed safety levels for work with recombinant DNA technique take into consideration the source of the donor DNA and its disease-producing potential. These four levels corresponds to (P1<P2<P3<P4) facilities approximate to 4 risk groups assigned for etiologic agents.

These levels and the appropriate conditions are enumerated as follows:

Biosafety Level 1: These practices, safety equipment and facilities are appropriate for undergraduate and secondary educational training and teaching laboratories and for other facilities in which work is done with defined and characterised strains of viable microorganisms not known to cause disease in healthy adult human. No special accommodation or equipment is required but the laboratory personnel are required to have specific training and to be supervised by a scientist with general training in microbiology or a related science.

Biosafety Level 2: These practices, safety equipment and facilities are applicable in clinical, diagnostic, teaching and other facilities in which work is done with the broad spectrum of indigenous moderate-risk agents present in the community and associated with human disease of varying severity. Laboratory workers are required to have specific training in handling pathogenic agents and to be supervised by competent scientists. Accommodation and facilities including safety cabinets are prescribed, especially for handling large volume are high concentrations of agents when aerosols are likely to be created. Access to the laboratory is controlled.

Biosafety level 3: These practices, safety equipment and facilities are applicable to clinical, diagnostic, teaching research or production facilities in which work is done with indigenous or exotic agents where the potential for infection by aerosols is real and the disease may have serious or lethal consequences. Personnel are required to have specific training in work with these agents and to be supervised by scientists experienced in this

kind of microbiology. Specially designed laboratories and precautions including the use of safety cabinets are prescribed and the access is strictly controlled.

Biosafety level 4: These practices, safety equipment and facilities are applicable to work with dangerous and exotic agents which pose a high individual risk of life-threatening disease. Strict training and supervision are required and the work is done in specially designed laboratories under stringent safety conditions, including the use of safety cabinets and positive pressure personnel suits . Access is strictly limited.

A specially designed suit area may be provided in the facility. Personnel who enter this area wear a one-piece positive pressure suit that is ventilated by a life support system. The life support system is provided with alarms and emergency break-up breathing air tanks. Entry to this area is through an airlock fitted with air tight doors. A chemical shower is provided to decontaminate the surface of the suit before the worker leaves the area. The exhaust air form the suit area is filtered by two sets of HEPA filters installed in the series. A duplicate filtration unit, exhaust fan and an automatically starting emergency power source are provide. The air pressure within the suit area is lower than that of any adjacent area. Emergency lighting and communication systems are provided. All penetrations into the inner shell of the suit area are sealed. A double door autoclave is provided for decontamination of disposable waste materials from the suit area.

Guidelines for rDNA research activities: The guidelines stipulate three categories of research activities, These are:

Category I: Which are exempt for the purpose of intimation and approval of competent authority.

The experiments involving self cloning, using strains and also inter-species cloning belonging to organism in the same exchanger group (Vide Chapter-V A4, A5).

(i) Organelle DNA including those from chloroplasts and mitochondria.

(ii) Host-vector systems consisting of cells in culture and vectors, either non-viral or viral containing defective viral genomes (except from cells known to harbour class III, IV and special category etiologic agents listed under Chapter V: A1.

Category II: Those requiring prior intimation of competent authority.

Experiments falling under containment levels II, III and IV.

(i) Experiment wherein DNA or RNA molecules derived from any source except for eukaryotic viral genome may be transferred to any non-human vertebrate or any invertebrate organisms and propagated under conditions of physical containment PC1 and appropriate to organism under study.

(ii) Experiments involving non pathogen DNA vector systems and regeneration from single cells.

(iii) Large scale use of recombinants made by self cloning in systems belonging to exempt category (e.g. *E.coli, Saccharomyces,* and *B. subtilis*)

Category III: Those requiring review and approval of competent authority before commencement.

(i) Toxin gene clonings : A list of toxins classified based on their potential toxicity is listed in Chapter V - A6. The number of plasmid toxin gene clonings at present going on are only three viz. *B. subtilis* and *B. sphericus* toxin genes are cloned in *B. subtilis* and cholera toxin genes and *B. thuringiensis* crystal protein genes cloned in *E.coli* K12. These toxins gene cloning are being done under PC1 and BC 1 Containment conditions. All toxin gene cloning experiments producing LD50 less than 50 ug/kg of body weight of vertebrates (Chapter V-A6) or large scale growing may be referred to Institutional Biosafety Committee (IBSC) for clearance.

(ii) Cloning of genes for vaccine production: e.g. Rinderpest and leprosy antigens. Rinderpest has been classified under Risk Group II in view of the common incidence of the disease in India, though it is listed under special category in the Centres for Disease Control & National Institute of Health (CDC-NIH) system. Similarly, leprosy afflicts a large segment of population which calls for concerted programme to control the disease by vaccination and detection at early stages through immunodiagnostic tests. The containment should be decided by Review Committee on Genetic Manipulation (RCGM) on a case by case basis on experiment utilising DNA from non-defective genomes of organisms recognised as pathogen. In view of no demonstrated risk from handling free *M. laprae* antigens, inactivated whole cells as well as antigens can be assigned to Risk Group I. The details of the rDNA technology

in development of vaccines for human and animal health giving containment conditions for observance of safeguards in large scale operations are given in Chapter V-B.

(iii) Cloning of mosquito and tick DNA experiments should be prescribed on a case by case basis since these are natural vectors for certain endemic viral and parasitic diseases.

(iv) Genes coding for antibiotic resistance into pathogenic organisms which do not naturally possess such resistance.

(v) Introduction into cultured human cells of recombinant DNA molecules containing complete genes of potentially oncogenic viruses or transformed cellular genes.

(vi) Introduction into animal cells of unidentified DNA molecules derived from cancer cells or in vitro transformed cells.

(vii) Experiments involving the use of infectious animal and plant viruses in tissue culture systems.

(viii) Experiments involving gene transfer to whole plants and animals.

(ix) Cell fusion experiments of Animal cells containing sequences from viral vectors if the sequence lead to transmissible infection either directly or indirectly as a result of complementation or recombination in the animals. For experiments involving recombinant DNA of higher class organisms using whole animals will be approved on case by case following IBSC review.

(x) Transgenosis in animal experiments : Transgenosis method is used to transform animal cells with foreign DNA by using viruses as vectors or by microinjection of DNA into eggs and pre-embryos. The expression of an inserted gene can be influenced both by the regulatory sequences associated with the gene and the sequences present at the site of integration of host genome. At present, there is no way to control where a gene is inserted into the chromosome of either an animal or plant cell. Yet this site of insertion can affect not only the expression of the interested gene but also the regulation of the host cells- DNA e.g. by non-specific activation of cellular protooncogenes.

(xi) All experiments involving the genetic manipulation of plant pathogens and the use of such genetically manipulated plant pathogens would require approval of competent authority (IBSC).

(xii)Transfer of genes with known toxicity to plants using *Agrobacterium tumefaciens* or other vectors. Attempts are under way using Ti-plasmid, *A. tumefaciens* and other vectors to transfer toxin-encoding genes that enable plants to make their own insecticide, resist infections or tolerate a variety of environmental stresses. Case by case clearance is needed though exemption may be made for the use of well characterized vectors and non-toxic genes.

(xiii)In case of plant viruses, permission may be obtained only when it is known that there is a chance of non-species specific spread of infection to plants that could produce changes in pathogenicity, host range or vector transmissibility. The growth of whole plants, propagation of genetically manipulated organisms in plants, regeneration of plants from cells transformed by manipulated plant pathogen vector would require containment conditions that are elaborated in Chapter V: C2.

(xiv)Experiments requiring field testing and release of rDNA engineered microorganisms and plants (Chapter V: C3).

(xv)Experiments involving engineered microbes with deletions and certain rearrangements.

(xvi)Diagnostics: No major risk can be foreseen on diagnostics involving in vitro tests. But for diagnostics involving in vivo tests, specific containment levels have to be prescribed on case by case basis. For example, tuberculin moiety could be cloned and used for in vivo hypersensitivity test as a diagnostic method.

(xvii)Gene therapy for hereditary diseases of genetic disorders.

MECHANISM OF IMPLEMENTATION OF BIOSAFETY GUIDELINES

For implementation of the guidelines it is necessary to have an institutional mechanism to ensure the compliance of requisite safeguards at various levels. The guidelines prescribe specific actions that include establishing safety procedures for rDNA research, production and release to the environment and setting up containment conditions for certain experiments. The guidelines suggest compliance of the safeguards through voluntary as well as regulatory approach. In this connection, it is proposed to have a mechanism of advisory and regulatory bodies to deal with the specific and discretionary actions on the following:

a. Self regulation and control in the form of guidelines on recombinant research activities; and

b. Regulation of large scale use of engineered organisms in production activity and release of organisms in environmental applications under statutory provisions.

The institutional mechanism as proposed for implementation of guidelines is shown in organogram in Figure 2. Mainly it consists of the following:-

i) Recombinant DNA Advisory Committee (RDAC)
ii) Institutional Biosafety Committee (IBSC)
iii) Review Committee on Genetic Manipulation (RCGM)
iv) Genetic Engineering Approval Committee (GEAC)

Scope and functions of advisory committee and statutory body

1. Recombinant DNA Advisory Committee (RDAC): **The Committee should take note of developments at national and international levels in Biotechnology towards the currentness of the safety regulation for India on recombinant research use and applications. It would meet once in 6 months or sooner for this purpose.**

The specific terms of reference for Recombinant Advisory Committee include the following :

i) To evolve long term policy for research and development in Recombinant DNA research.
ii) To formulate the safety guidelines for Recombinant DNA Research to be followed in India.
iii) To recommended type of training programme for technicians and research fellows for making them adequately aware of hazards and risks involved in recombinant DNA research and methods of avoiding it.

2. Implementation Committees:

2.1 Institutional Biosafety Committee (IBSC)

Institutional Biosafety Committee (IBSC) are to be constituted in all centres engaged in genetic engineering research and production activities. The Committee will constitute the following:

Head of the Institution or nominee

(i) 3 or more scientists engaged in DNA work or molecular biology with an outside expert in the relevant discipline.

(ii) A member with medical qualifications - Biosafety Officer (in case of work with pathogenic agents/large scale use).

(iii)One member nominated by DBT.

2.2 The Institutional Biosafety Committee shall be the nodal point for interaction within institution for implementation of the guidelines. Any research project which is likely to have biohazard potential (as envisaged by the guidelines) during the execution stage or which involve the production of either microorganisms or biologically active molecules that might cause bio-hazard should be notified to IBSC. IBSC will allow genetic engineering activity on classified organisms only at places where such work should be performed as per guidelines. Provision of suitable safe storage facility of donor, vectors, recipients and other materials involved in experimental work should be made and may be subjected to inspection on accountability.

The biosafety functions and activity include the following:

Registration of Bio-safety Committee membership composition with RCGM and submission of reports.

IBSC will provide half yearly report on the ongoing projects to RCGM regarding the observance of the safety guidelines on accidents, risks and on deviations if any. A computerised Central Registry for collation of periodic report on approved projects will be set up with RCGM to monitor compliance on safeguards as stipulated in the guidelines.

i) Review and clearance of project proposals falling under restricted category that meets the requirements under the guidelines.

IBSC would make efforts to issue clearance quickly on receiving the research proposals from investigators.

ii) Tailoring biosafety programme to the level of risk assessment.

iii) Training of personnel on biosafety.

iv) Instituting health monitoring programme for laboratory personnel.

Complete medical check-up of personnel working in projects involving work with potentially dangerous microorganisms should be done prior to starting such projects. Follow up medical checkups including pathological tests should be done periodically, at least annually for scientific workers involved in such projects. Their medical records should be accessible to the RCGM. It will provide half yearly reports on the ongoing projects to RCGM regarding the observance of the safety guidelines on accidents, risks and on deviations if any.

v) Adopting emergency plans.

3. Review Committee on Genetic Manipulation (RCGM): The RCGM will have the following composition:

i) Department of Biotechnology
ii) Indian Council of Medical Research
iii) Indian Council of Agricultural Research
iv) Council of Scientific & Industrial Research
v) Three Experts in Individual capacity
vi) Department of Science & Technology

The RCGM will have the functions:

To establish procedural guidance manual - procedure for regulatory process with respect to activity involving genetically engineered organisms in research, production and applications related to environmental safety.

i) To review the reports in all approved ongoing research projects involving high risk category and controlled field experiments, to ensure that safeguards are maintained as per guidelines.
ii) To recommended the type of containment facility and the special containment conditions to be followed for experimental trials and for certain experiments.
iii) To advise customs authorities on import of biologically active material, genetically engineered substances or products and on excisable items to Central Revenue and Excise.

iv) To assist Department of Industrial Development, Banks towards clearance of applications in setting up industries based on genetically engineered organisms.

v) To assist the Bureau of Indian Standards to evolve standards for biologics produced by rDNA technology.

vi) To advise on intellectual property rights with respect to rDNA technology on patents.

3.1 The RCGM would have a Research Monitoring function by a group consisting of a smaller number of individuals (3 or 4). The monitoring group would be empowered to visit experimental facilities in any laboratory in India where experiments with biohazard potential are being pursued in order to determine the Good Laboratory practice and conditions of safety are observed.

3.2 In addition, if the RCGM has reasons to believe that there is either actual or potential danger involved in the work carried out by any laboratory (which might or might not have obtained prior clearance for the project), the monitoring group would be empowered to inspect the facility and assess the cause of any real or potential hazard to make appropriate recommendation to the RCGM. RCGM would be empowered to recommend alteration of the course of experiments based on hazard considerations or take steps to cancel the project grant, in case of deliberate negligence and to recommend appropriate actions under the provisions of Environmental Protection Act (EPA) where necessary.

4. Genetic Engineering Approval Committee (GEAC): Genetic Engineering Approval Committee (GEAC) will function under the Department of Environment (DOEn) as statutory body for review and approval of activities involving large scale use of genetically engineered organisms and their products in research and development, industrial production, environmental release and field applications.

The functions include giving approval from environmental angle on:

Import, export, transport, manufacture, process, selling of any microorganisms or genetically engineered substances or cells including food stuffs and additives that contains products derived by Gene Therapy.

i) Discharge of Genetically engineered/classified organisms/cells from Laboratory, hospitals and related areas into environment.

ii) Large scale use of genetically engineered organisms/classified microorganisms in industrial production and applications. (Production shall not be commenced without approval).

iii) Deliberate release of genetically engineered organisms. The approval will be for a period of 4 years.

The composition of the Committee would be as follows:

Chairman - Additional Secretary, Department of Environment

Co-Chairman - Expert Nominee of Secretary, DBT.

1. Representatives of concerned Agencies and Departments:
   - Ministry of Industrial Development
   - Department of Science & Technology
   - Department of Ocean Development
   - Department of Biotechnology
3. Expert Members:
   - Director-General, Indian Council of Agricultural Research
   - Director General, Indian Council of Medical Research
   - Director-General, Council of Scientific & Industrial Research
   - Director-General, Health Services (Ministry of Health & Family Welfare)
   - Plant Protection Adviser (Ministry of Agriculture)
   - Chairman, Central Pollution Control Board
   - Outside experts in individual capacity.
   - Member Secretary - Official of, DOEn

4.1 GEAC will have the Biotechnology Coordination Committees under it which will functions as legal and statutory body with judicial powers to inspect, investigate and take punitive action in case of violations of statutory provisions under EPA.

Review and control of safety measures adopted while handling large scale use of genetically engineered organisms/classified organisms in research, developmental and industrial production activities.

i) Monitoring of large scale release of engineered organisms/products into environment, oversee field applications and experimental field trials.
ii) To provide information/data inputs to RCGM upon surveillance of approved projects under industrial production, and in case of environmental releases with respect to safety, risks and accidents.

4.2 Statutory rules and regulations to be operated by the GEAC would be laid down under the Environment Protection Act, 1986.