KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)

Coimbatore – 641 021.

Syllabus

SEMESTER IV

18BTU402 RECOMBINANT DNA TECHNOLOGY 4H - 4C

Total hours/week: L:4 T:0 P:0 Marks: Internal: 40 External: 60 Total: 100

Scope: This syllabus includes current technical procedures for rDNA technology and its applications.

Objective: The application of rDNA technology ranges from developing genetically modified crops and livestock, production of novel drugs, vaccines, and diagnostics.

UNIT-I

Introduction to r-DNA technology: Basic tools and applications – isolation and purification of nucleic acids, Enzymes used in cloning - restriction enzymes, ligases, polymerases, kinases, phosphatases. Gene recombination and gene transfer – transformation, episomes, plasmids and other cloning vectors (bacteriophage-derived, artificial chromosomes), microinjection, electroporation, ultrasonication.

UNIT-II

Selection and screening of recombinant clones: Probes – radio labeled and non radiolabeled, guessmers and degenerate probes. Sequence dependent and independent screening, southern, northern hybridization, colony and plaque hybridization, *in situ* chromosomal hybridization, chromosome walking, Genome mapping, DNA fingerprinting, Polymerase chain reaction (PCR), RT- (Reverse transcription) PCR.

UNIT-III

Expression and Characterization of Cloned DNA: Expression vectors, optimization of protein expression in heterologous systems, Fusion proteins, *In vitro* translation systems. Preparation and comparison of Genomic and cDNA library.

UNIT-IV

Random and Site-directed Mutagenesis: Primer extension and PCR based methods of site directed mutagenesis, Random mutagenesis, Gene shuffling, production of chimeric proteins, Protein engineering concepts and examples (any two).

UNIT-V

Applications of Genetic Engineering: In plants: use of *Agrobacterium tumefaciens* and *A. rhizogenes*, Ti plasmids, Strategies for gene transfer to plant cells, Direct DNA transfer to plants, Gene targeting in plants, Use of plant viruses as episomal expression vectors.

In animals: Production and applications of transgenic mice, role of ES cells in gene targeting in mice, Therapeutic products - blood proteins, human hormones, immune modulators and vaccines (one example each). Ethical, legal and social issues.

References

1. Brown, T.A., (2006). *Gene Cloning and DNA Analysis* (5th ed.). Oxford: UK, Blackwell Publishing.
2. Clark, D.P., & Pazdernik, N.J. (2009). *Biotechnology-Applying the Genetic Revolution*. USA: Elsevier Academic Press.
3. Glick, B.R., & Pasternak, J.J. (2003). *Molecular Biotechnology- Principles and Applications of recombinant DNA*. Washington: ASM Press.
4. Primrose, S.B., & Twyman, R.M. (2006). *Principles of Gene Manipulation and Genomics* (7th ed.). Oxford: UK, Blackwell Publishing.

KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University Established Under Section 3 of UGC Act 1956)

Coimbatore – 641 021.

LECTURE PLAN

DEPARTMENT OF BIOTECHNOLOGY

STAFF NAME: Dr. BARATHKUMAR, S.
SUBJECT NAME: RECOMBINANT DNA TECHNOLOGY SUB.CODE:18BTU402
SEMESTER: IV CLASS: II B.Sc.

Unit I		
1 hour	**Overview of r-DNA technology**	T1- pg 1-20
	Basic techniques	
	Isolation and purification of nucleic acids	T1- pg 27-51
	Genomic DNA isolation	
	Plasmid and RNA isolation	
1 hour	**Enzymes used in cloning**	T1- pg 52-56
	Restriction enzymes & types, nomenclature, recognition, cleavage pattern	
	Ligases – History, types, Mechanism, Application	T1- pg 56-57
1 hour	DNA polymerases – History, types, application	T1- pg 57-58
	RNA polymerases –History, types, application	
	Taq DNA polymerases	T1- pg 238
1 hour	Reverse transcriptase – Features, application	T1- pg 58
1 hour	**Modifying enzymes**	
	Phosphatases, Kinases, Transferases	T1- pg 59
	DNase	T1- pg 53-54
	RNase	T1- pg 54-56
1 hour	Vectors - Episomes, plasmids	
1 hour	Bacteriophage based – Lamda phage	T1- pg 19-20
	M13 Phage based	T1- pg 114-120
	Phagemids	T1- pg 121-125
	Cosmids – pJB8	T1- pg 128-130
1 hour	YAC vector – Feature, Types	T1- pg 139-146
	BAC vector – Features, Types	T1- pg 131-132
1 hour	Gene Recombination and gene transfer	
	Electroporation, ultrasonication, Microinjection	
1 hour	Recapitulation – Unit I	
1 hour	Class test	
Total	12 hours	

UNIT II		
1 hour	Probes, probe preparation methods	J1
1 hour	Types of Probes – Radiolabelled &	
1 hour	Non-radiolabelled	
1 hour	Guessmers	
1 hour	Degenerate probes	T2- pg 114-148
1 hour	**Screening of recombinant clones -** methods	T1- pg 158-163
	Screening of rClones with specific DNA insert	T1- pg 163-166
1 hour	Sequence dependent screening	T1- pg 166-172
	1.Colony hybridization	
1 hour	2.Plaque hybridization	
	3.Chromosome walking	T1- pg 191-192
1 hour	Sequence independent screening	T1- pg 177-178
	1. Immunological methods	
1 hour	2.southern-western hybridization	T1- pg 184
	3.*In-situ* chromosomal hybridization	T1- pg 202-206
1 hour	4.Ligand based	
1 hour	Genome mapping, DNA Finger printing	T3 – pg 445-447, T1- pg 202-204
1 hour	PCR & RT-PCR	T1- pg 238-249
1 hour	Recapitulation – Unit II	
1 hour	Class test	
Total	15 hours	

Unit III		
1hour	**Expression of cloned DNA**	T1- pg 275-287
1hour	Expression Vectors – features, types	T1- pg 255-265 T1- pg 282-285
1 hour	Expression systems	
	Types and elements of expression vectors	
1 hour	Types of expression vectors	
1 hour	Expression system – Optimization	
1 hour	RNAi Vectors	T1- pg 264
1 hour	*In vitro* translation systems	T1- pg 114-115
1 hour	Generation of Genomic DNA libraries	T1- pg 165-167
1 hour	Generation of cDNA libraries	T1- pg 167-168
1 hour	Recapitulation – Unit III	
Total	10 hours	

Unit IV		
1hour	Random and site directed mutagenesis	R2-141-150
1 hour	- PCR based method	R2-144
1 hour	- Random mutagenesis	
1 hour	Gene shuffling	W2
1 hour	Production of chimeric proteins	W3
1 hour	Introduction to protein engineering	W4
1 hour	Recapitulation – Unit IV	
1 hour	Class test	
Total	8 hours	

Unit V		
1hour	Application of Genetic Engineering – Plants	R2- 183,289-290
1hour	*Agrobacterium tumifaciencs* and *A.rhizogenes*	
1hour	- Strategies for gene transfer	
1hour	- Gene targeting in plants	
1hour	- Plant virus usage as episomal vectors	R2-294-298
1hour	Application of Genetic Engineering – Animals	
1hour	Transgenic mice – production and application	
1hour	- Role of ES cells in gene targeting in mice	
1hour	Production of therapeutic products	
1hour	- Blood proteins	
1hour	- Human hormones	
1hour	- Immune modulators	
1hour	- vaccines	
1hour	Ethical, legal and social issues in rDNA technology	T4- pg 407-410, T3-pg 633, 664, 682
1 hour	Previous year ESE Question paper Discussion	
1 hour	Previous year ESE Question paper Discussion	
1 hour	Previous year ESE Question paper Discussion	
Total	17 hours	

References:

1. R1) T.A.Brown (2001) Gene Cloning and DNA analysis - An Introduction, 6th Edition, Oxford UK, Black well publishers.
2. R2) Primrose, S.B., Twyman, R.M.(2006). Principles of Gene manipulation and Genomics, 7th edition, Oxford UK, Black well publisher.
3. T1) R.C.Dubay (1998). A textbook of Biotechnology, 5th edition, Chand publications.
4. T2) B.D.Singh (2005) Biotechnology, 5th Edition, Kalyani publishers.
5. J1) Zhang et. al.,(1991) Efficient transformation of tobacco by ultrasonication. Biotechnol.(9)996-7.
6. J2) Vasavirama, K. (2013). Molecular probes and their applications, IJLPS,2(2).
7. W1) WWW.Sciencedirect.com/topics/biochemistry-genetics and molecular biology/fussion proteins.
8. W2) http://springer.com/protocol/10.1385/1-59259-395-x185.
9. W3) WWW.nature.com/articles/hrg3268.

10. W4) WWW.Sciencedirect.com/topics/biochemistry-genetics and molecular biology/ protein engineering.

Unit I

SYLLABUS

Basic tools and applications – isolation and purification of nucleic acids, Enzymes used in cloning - restriction enzymes, ligases, polymerases, kinases, phosphatases. Gene recombination and gene transfer – transformation, episomes, plasmids and other cloning vectors (bacteriophage-derived, artificial chromosomes), microinjection, electroporation, ultrasonication.

Recombinant DNA (rDNA) molecules

- They are DNA sequences that result from the use of laboratory methods (molecular cloning) to bring together genetic material from multiple sources, creating sequences that would not be found in biological organisms.

Recombinant DNA is possible because

- DNA molecules from all organisms share the same chemical structure;
- they differ only in the sequence of nucleotides within that identical overall structure. Consequently, when DNA from a foreign source is linked to host sequences that can drive DNA replication and then introduced into a host organism, the foreign DNA is replicated along with the host DNA.

Recombinant DNA molecules are sometimes called **chimeric DNA**, because they are usually made of material from two different species, like the mythological chimera.

- The DNA sequences used in the construction of recombinant DNA molecules can originate from any species.
- For example, plant DNA may be joined to bacterial DNA, or human DNA may be joined with fungal DNA.

- In addition, DNA sequences that do not occur anywhere in nature may be created by the chemical synthesis of DNA, and incorporated into recombinant molecules.
- Using recombinant DNA technology and synthetic DNA, literally any DNA sequence may be created and introduced into any of a very wide range of living organisms.
- Proteins that result from the expression of recombinant DNA within living cells are termed recombinant proteins.
- When recombinant DNA encoding a protein is introduced into a host organism, the recombinant protein will not necessarily be produced.
- Expression of foreign proteins requires the use of specialized expression vectors and often necessitates significant restructuring of the foreign coding sequence.
- It is important to note that recombinant DNA differs from genetic recombination in that the former results from artificial methods in the test tube, while the latter is a normal biological process that results in the remixing of existing DNA sequences in essentially all organisms.

Creating recombinant DNA

- Molecular cloning is the laboratory process used to create recombinant DNA.
- It is one of two widely-used methods (along with polymerase chain reaction, PCR) used to direct the replication of any specific DNA sequence chosen by the experimentalist.
- The fundamental difference between the two methods is that molecular cloning involves replication of the DNA within a living cell, while PCR replicates DNA in the test tube, free of living cells.
- Formation of recombinant DNA requires a cloning vector, a DNA molecule that will replicate within a living cell.
- Vectors are generally derived from plasmids or viruses, and represent relatively small segments of DNA that contain necessary genetic signals for replication, as well as additional elements for convenience in inserting foreign DNA, identifying cells that contain recombinant DNA, and, where appropriate, expressing the foreign DNA.

- The choice of vector for molecular cloning depends on the choice of host organism, the size of the DNA to be cloned, and whether and how the foreign DNA is to be expressed.
- The DNA segments can be combined by using a variety of methods, such as restriction enzyme/ligase cloning or Gibson assembly.

In standard cloning protocols, the cloning of any DNA fragment essentially involves **seven steps:**

(1) Choice of host organism and cloning vector,

(2) Preparation of vector DNA,

(3) Preparation of DNA to be cloned,

(4) Creation of recombinant DNA,

(5) Introduction of recombinant DNA into the host organism,

(6) Selection of organisms containing recombinant DNA,

(7) Screening for clones with desired DNA inserts and biological properties.

Isolation and purification of nucleic acids

Three major types of techniques, or combinations of them, are employed in the isolation of nucleic acids:

[1] differential solubility,
[2] absorption methods,
[3] density gradient centrifugation.

- The choice of method will depend on the type of DNA being isolated and the application.
- A major goal of nucleic acid isolation is the removal of proteins.
- The separation of nucleic acids from proteins is generally easily accomplished due to their different chemical properties.

- In particular, the highly charged phosphate backbone makes the nucleic acids rather hydrophilic as compared to proteins which are more hydrophobic.

Most nucleic acid isolation protocols involve a

[1] cell lysis step,
[2] enzymatic treatments,
[3] differential solubility (eg., phenol extraction or absorption to a solid support),
[4] precipitation.

[1] Cell Lysis

- Nucleic acids must be solubilized from cells or other biological material.
- This solubilization is usually carried out under denaturing conditions such as: SDS, alkali, boiling or chaotropic agents.
- These denaturing conditions efficiently solubilize the nucleic acids and generally do not adversely affect them.
- In addition, the denaturing conditions promote the removal of proteins during the subsequent steps and inhibit the activity of nucleases which will degrade the nucleic acids.

[2] Enzymatic Treatment

- Another approach in the isolation of nucleic acids is to degrade unwanted components.

For example, inclusion of proteases (eg., proteinase K) in the lysate will promote the removal of proteins.

- Proteinase K is still active at 55°C in the presence of 0.5% SDS.
- The elevated temperature and SDS improve solubility and inhibit any DNAse activity that may be present in the lysate.
- Nucleases can also be used to remove unwanted nucleic acids.

For example, many DNA extraction protocols include a RNase treatment step, and visa versa.

- It is important that the RNase be free of DNase activity.

DNase-free RNase is easily prepared by boiling commercial RNase for 10 minutes.

- The stability of RNase makes the preparation of RNase-free DNase more difficult.
- RNase-free DNase should be purchased from a reliable vendor or tested before it use.

[3] Phenol Extraction

- Phenol is an organic solvent that is used to separate proteins from nucleic acids.
- Proteins are hydrophobic and partition in the organic phase.
- Nucleic acids are highly charged and partition in the aqueous phase.

The advantages are of phenol extraction are that

- it is easy to carry out and can be adapted to many applications.
- It is also easily applied over a wide range of volumes (40 μl to several ml).
- In particular, phenol extraction is widely used for the isolation of high molecular weight genomic DNA.

Phenol extraction is accomplished by mixing the sample with an equal volume of phenol which has been previously saturated with a Tris buffer at pH 8 containing EDTA and NaCl.

- The phenol should be molecular biology grade phenol should and store at -20$_o$ until preparing the saturated solution.
- The saturated solution is stored at 4$_o$.
- Phenol is easily oxidized, as evidenced by yellowing, and the oxidation products can break DNA. Oxidized phenol should be discarded.
- Depending on the application, the two phases are completely mixed by vortexing, or gently mixed (eg., high molecular weight DNA).
- The phases are separated by centrifugation and the upper aqueous phase, which contains the nucleic acids, is retained.
- Proteins will often be visible as flocculent material at the top of the phenol phase.

- The two phases need to be carefully separated in that the nucleic acids and proteins tend to be at the interface.
- Leaving too much of the aqueous layer behind will lead to undue loss of material and aspirating too close to the interface can include protein.
- The aqueous phase can be re-extracted with phenol to remove more protein.
- Phenol is a hazardous waste material that needs to be disposed of properly.
- A common variation of phenol extraction is a mixture of phenol:chloroform: isoamyl alcohol (25:24:1).
- The more organic chloroform removes lipids, denatures more protein and mixes less with the aqueous phase leading to more efficient extraction.

[4] Ethanol Precipitation

- Nucleic acids can be precipitated from dilute solutions with ethanol.
- This precipitation can be a concentration step or a means to change buffers, especially after phenol extraction.
- Typically either sodium acetate or potassium acetate, pH 5.0-5.5, is added to a final concentration of approximately 0.3 M.
- The sodium and acidic pH will neutralize the highly charged phosphate backbone and to promote hydrophobic interactions.
- Two-to-two and a half volumes of ethanol are added and the sample is incubated as -20°C.
- If the nucleic acids are small in size and/or in low concentrations an extended incubation (severalhours to overnight) is needed.
- The precipitated DNA is collected by centrifugation.
- The pellet is rinsed with 70% ethanol to remove any excess salt, dried and dissolved in the appropriate buffer.
- A variation is to substitute ammonium acetate if the 'hard' salts are a problem.

- Another modification is to use an equal volume of isopropanol (instead of 2-2½ volumes of ethanol)which minimizes the increase in sample volume.

Isolation of High Molecular Weight Genomic DNA

- High molecular weight chromosomal DNA is usually isolated by multiple rounds of phenol extraction and enzyme treatments as discussed above.
- Shear forces, which can break long DNA molecules, need to be avoided during all steps and samples should never be vortexed.
- Therefore, the phenol extraction is carried with gentle rocking for several hours.
- These precautions against shear forces are not necessary in the isolation of low molecular weight DNA.
- Another common modification at the ethanol precipitation step is 'spool out' the high molecular weight genomic DNA on the end of a sealed Pastuer pipet.
- The precipitated DNA is wrapped around the end of the pipet is then allowed to partially dry and then dissolved in the appropriate buffer.
- This minimizes the contamination with RNA and low molecular weight DNA fragments.

Plasmid Minipreps and Adsorption Methods

- Historically, phenol extractions were used for the isolation of most forms of nucleic acids.
- It is now more common to use techniques based upon adsorption chromatography for the isolation of smaller DNA molecules, such as plasmids.
- Various kits are available for the rapid isolation of small quantities of plasmid DNA.
- The procedure consists of solubilizing the bacteria in an alkali solution followed by neutralization with sodium acetate.
- The neutralization results in the precipitation of some of the protein and the genomic DNA which is removed by centrifugation.
- The soluble material is then mixed with a resin in the presence of chaotropic agents (usually guanidine hydrochloride).
- The resins are usually either based on silica or diatomaceous earth.

- Under these conditions DNA binds to the matrix, but proteins and RNA do not.
- The DNA is eluted in a low salt buffer.
- These methods are rapid and yield a highly purified plasmid DNA which can generally be used directly in most applications without further processing.
- Another common application for an adsorption method is the isolation of DNA fragments following gel electrophoresis.
- In this case the agarose gel piece containing the DNA is dissolved in NaI, a chaotropic salt, and the DNA adsorbed to silica.
- The DNA is then eluted with a low salt buffer and sometimes gentle heating.

Isolation of RNA

Most RNA isolation protocols also involve phenol extractions and are similar to DNA isolations. However, there are some differences and special considerations.

In particular,

- precautions against RNase activity must be taken.
- RNase is an extremely stable and active enzyme.
- Gloves should be worn at all times and sterile plasticware should be used whenever possible to avoid introducing exogenous RNase to the sample.
- Glassware needs to treated with DEPC-water and autoclaved to inactivate any RNase.
- Buffers should be prepared from DEPC-water or RNase inhibitors included.
- The cell lysis and solubilization of RNA will typically be carried out in guanidine salts(especially guanidine thiocyanate).
- Guanidine is a strong chaotropic agent and will inhibit RNases.
- The strong denaturing effect of this salt will also promote better phenol extraction.
- The phenol extraction is the same as the DNA isolation except that the phenol is usually saturatedwith a buffer of pH 5-6.
- The lower pH will result in some DNA partitioning in the organic phase.

- If DNA contamination is a problem, it is possible to purchase RNase-free DNase.
- Long term storage is best as a precipitate in 70% ethanol at –20°C.

Precipitation of RNA with LiCl

- LiCl has been used to selectively precipitate RNA.
- Large RNAs (rRNA, mRNA) are insoluble at high ionic strength, whereas small RNAs (tRNA and 5S rRNA) and DNA generally remain soluble.
- Following either phenol or guanidine extraction, an equal volume of 8 M LiCl is added.
- The sample is mixed vigorously and incubated at -20°C.
- The precipitate is collected by centrifugation and reprecipitated if necessary.

Affinity chromatography

- Most eukaryotic mRNA contains a stretch of A residues at its 3' endwhich added post-transcriptionally.
- It is possible toisolate mRNA by affinity chromatography on oligo dT columns.
- The RNA solution is passedover an oligo-dT column under conditions which promote base pairing. Only RNA with a polyA tail binds.
- The polyA RNA is eluted under conditions (usually low salt and high temperature) which breaks the base pairing.

DENSITY GRADIENT CENTRIFUGATION

- Density gradient centrifugation can also be used in the analysis and isolation of nucleic acids. Double-stranded DNA, single-stranded DNA, RNA and proteins all have different densities and therefore can be separated by isopycnic (i.e., equilibrium) centrifugation.
- CsCl is the standard medium for the density gradient centrifugation of nucleic acids and are especially useful for the purification of large amounts of highly purified DNA.
- The gradients are carried out in the presence of ethidium bromide which fluoresces when bound to DNA.

- The DNA bands are detected by illumination with ultraviolet light and easily recovered with a syringe and needle by puncturing the wall of the disposable tube and aspirating the DNA.
- The CsCl can be removed by dialysis or by precipitating the DNA.
- The %G:C content affects the density of DNA.
- This can result in multiplebands on CsCl gradients if DNA composed different G:C compositions arecentrifuged.
- For example, minor bands, called satellite DNA, are often observed when total DNA from an organism isanalyzed by CsCl gradients.
- These satellite bands are usually due to highly repetitive DNA or organellar DNA.
- Mitochondrial DNA is noted for its high A:T content.
- These satellite DNA bands can be purified from the genomic DNA by density gradient centrifugation.
- Nucleic acids can also be separated according to size by rate zonal centrifugation on sucrose gradients.
- However, this method is not widely used since gel electrophoresis is generally a more convenient method for the size fractionation of nucleic acids.

Enzymes used in rDNA technology (DNA manipulating enzymes)

Based on the type of reaction that they catalyse, five classes of DNA manipulative enzymes are available

1. Nucleases – enzymes that cut, shorten or degrade nucleic acid.
2. Ligases – join nuclic acid molecules together.
3. Polymerases – makes copies of nucleic acid molecules.
4. Modifying enzymes – remove or add chemical groups.
5. Topoisomerases – introduce or remove supercoils from covalently closed circular DNA.

Nuclease

A **nuclease** is an enzyme capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids.

In the late 1960s, **Stuart Linn** and **Werner Arber** isolated examples of the two types of enzymes responsible for phage growth restriction in Escherichia coli (*E. coli*) bacteria.

One of these enzymes added a methyl group to the DNA, generating methylated DNA called "methylase, the other enzyme cleaves unmethylated DNA at a wide variety of locations along the length of the molecule called as "restriction nuclease". These enzymatic tools were important to scientists who were gathering the tools needed to "cut and paste" DNA molecules.

Types of nucleases

1. **Structure specific nuclease**
2. **Sequence specific nuclease**

Sequence specific nuclease

H.O. Smith, K.W. Wilcox, and T.J. Kelley, in 1968, isolated and characterized the first restriction nuclease whose functioning depended on a specific DNA nucleotide sequence. Working with *Haemophilus influenzae* bacteria, this group isolated an enzyme, called *Hind*II, that always cut DNA molecules at a particular point within a specific sequence of six base pairs.

Types of Sequence specific nuclease

1. Exonuclease
2. Endonuclease

Exonuclease

- A nuclease that functions by removing nucleotides from the ends of the DNA molecule is called an exonuclease.

Endonuclease

- A restriction endonuclease functions by "scanning" the length of a DNA molecule.
- Once it encounters its particular specific recognition sequence, it will bind to the DNA molecule and makes one cut in each of the two sugar-phosphate backbones.
- The positions of these two cuts, both in relation to each other, and to the recognition sequence itself, are determined by the identity of the restriction endonuclease used to cleave the molecule in the first place.
- Different endonucleases yield different sets of cuts, but one endonuclease will always cut a particular base sequence the same way, no matter what DNA molecule it is acting on.
- Once the cuts have been made, the DNA molecule will break into fragments.

Restriction Endonucleases –

- Endonucleases are enzymes that produce internal cuts, called cleavage, in DNA molecules. Many endonucleases cleave DNA molecules at random sites.
- But a class of endonucleases cleaves DNA only within or near those sites, which have specific base sequences; such endonucleases are known as restriction endonucleases, and the sites recognised by, them are called recognition sequences or recognition sites.
- The recognition sequences are different and specific for the different restriction endonucleases or restriction enzymes.
- Restriction enzymes were discovered due to and named after the phenomenon of host restriction of bacterial phages.

The presence restriction enzymes was postulated by **W. Arber** during 1960s, while the first true restriction endonuclease was isolated in **1970**.

Smith, Nathans and Arber were awarded the Nobel Prize for Physiology and Medicine in **1978** for the discovery of endonucleases.

- Restriction endonucleases are indispensable for DNA cloning and sequencing.

- They serve as the tools for cutting DNA molecules at predetermined sites, which is the basic requirement for gene cloning or recombinant DNA technology.

Types of Restriction Endonucleases –

There are three distinct types of restriction endonucleases.

Type I restriction endonucleases –

- They are complex endonucleases, and
- They have recognition sequences of about 15 bp;
- They cleave the DNA about 1000 bp away from the 5'-end of the sequence "TCA" located within the recognition site, e.g., EcoK, EcoB etc.

Type II restriction endonucleases

- They are remarkably stable and induce cleavage either, in most cases, within their recognition sequences or very close to them.
- More than 350 different type II endonucleases with over 100 different recognition sequences are known.
- They require Mg^{2+} ions for cleavage.
- The first type II enzyme to be isolated was **Hind II in 1970**.
- Only type II restriction endonucleases are used for restriction mapping and gene cloning.

Type III restriction endonucleases

- are intermediate between the type I and type II enzymes;
- they cleave DNA in the immediate vicinity of their recognition sites, e.g., EcoPl, EcoP15, Hind III etc.

Nomenclature of Restriction Endonucleases –

The nomenclature of restriction endonucleases follows a general pattern.

(1) The first letter of the name of genus in which a given enzyme is first discovered is written in capital.

(2) This is followed by the first two letters of species name of the organism. These three letters are generally written in italics, e.g., Eco from Escherichia coli, Hin from Haemophilus influenzae, etc.

(3) Strain or type identification is depicted next in Roman. e.g., Ecok;

(4) When an organism produces more than one enzyme, they are identified by sequential Roman numerals, e.g., the different enzymes produced by H. influenzae strain Rd are named Hind II, Hind III, etc.

Some Restriction Endonucleases -

Restriction endonuclease	**Source (organism an strain)**	**Recognition sequence**
AluI	*Arthrobacter luteus*	AG/CTTC/GA
BamHI	*Bacillus amyloliquefaciens* H	G/GATCCCCTAG/G
EcoRI	*Escherichia coli* Ry13	G/AATTCCTTAA/G
HindIII	*H. influenzae* Rd	A/AGCTTTTCGA/A
Sau3A	*Staphylococcus aureus* 3A	/GATCCTAG/
TaqI	*Thermus aquaticus* YTI	T/CGAAGC/T

Recognition Sequences For Type II Endonucleases

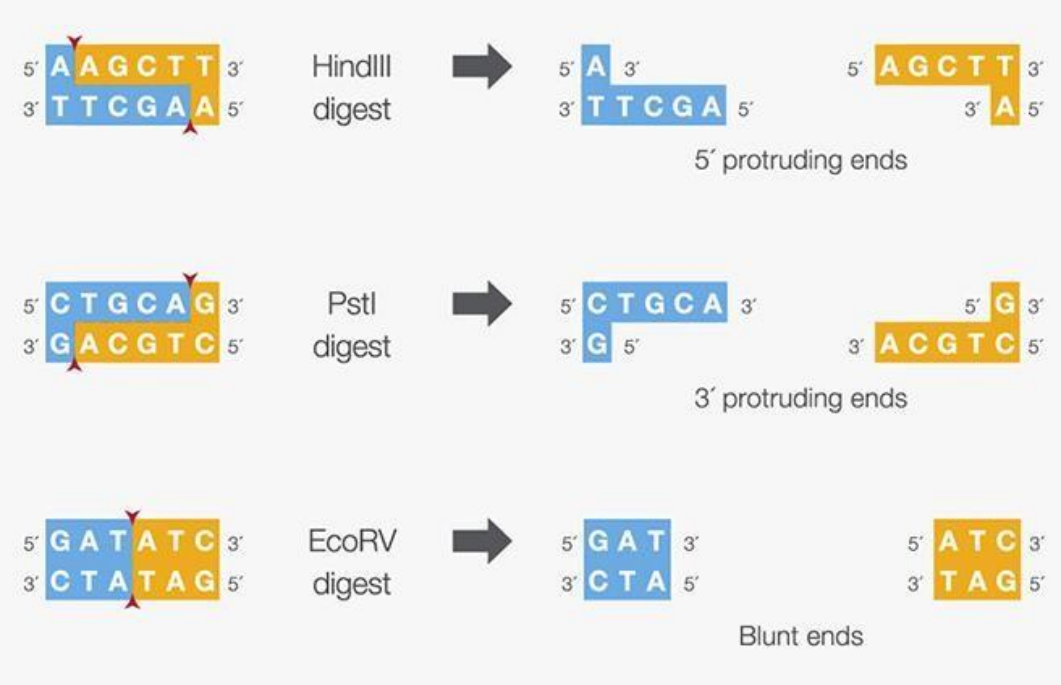
- The recognition sequences for Type II endonucleases form palindromes with rotational symmetry.
- In a palindrome, the base sequence in the second half of a DNA strand is the mirror image of the sequence in its first half; consequently, the complementary DNA strand of a double helix also shows the same situation.
- But in a palindrome with rotational symmetry, the base sequence in the first half of one strand of a DNA double helix is the mirror image of the second half of its complementary strand.
- Thus in such palindromes, the base sequence in both the strands of a DNA duplex reads the same when read from the same end (either 5' or 3') of both the strands.
- Most of the type II restriction endonucleases have recognition sites of 4, 5 or 6 bp (base pairs), which are predominantly GC-rich.
- Longer palindromic target sequences are also known, and so are nonpalindromic ones (specific for some enzymes).
- Some restriction enzymes have ambiguities in their recognition sites, e.g. EcoRII, so that they may recognise upto 4 different target sequences.

Cleavage Pattern of Type II Restriction Endonucleases –

Most type II restriction endonucleases cleave the DNA molecules within their specific recognition sequences, but some produce cuts immediately outside the target sequence, e.g., NlaIII, Sau3A, etc.

These cuts are either

(1) Staggered or Sticky end cutter
(2) Blunt end cutter



A pictorial example of cleavage pattern of restriction enzymes

- Most enzymes produce staggered cuts in which the two strands of a DNA double helix are cleaved at different locations; this generates protruding (3'- or 5'-) ends, i.e., one strand of the double helix extends some bases beyond the other:
- Due to the palindromic (symmetrical) nature of the target sites, the two protruding ends generated by such a cleavage by a given enzyme have complementary base sequence.
- As a result, they readily pair with each other; such ends are called cohesive or sticky ends.
- An important consequence of this fact is that when fragments generated by a single restriction enzyme from different DNAs are mixed, they join together due to their sticky ends.
- Therefore, this property of the restriction enzymes is of great value for the construction of recombinant DNAs.

- Some restriction enzymes, on the other hand, cut both the strands of a DNA molecule at the same site so that the resulting termini or ends have blunt or flush ends in which the two strands end at the same point.
- The blunt cut ends also can be effectively utilized for construction of recombinant DNAs following one of several strategies.

DNA ligase (EC 6.5.1.1)

- is a specific type of enzyme that repairs single-stranded discontinuities in double stranded DNA molecules, in simple words strands that have double-strand break (a break in both complementary strands of DNA).
- Purified DNA ligase is used in gene cloning to join DNA molecules together.
- The alternative, a single-strand break, is fixed by a different type of DNA ligase using the complementary strand as a template, but still requires DNA ligase to create the final phosphodiester bond to fully repair the DNA.
- DNA ligase has applications in both DNA repair and DNA replication.
- In addition, DNA ligase has extensive use in molecular biology laboratories for Genetic recombination experiments

History

- The first DNA ligase was purified and characterized in 1967.
- The common commercially available DNA ligases were originally discovered in bacteriophage T4, *E. coli* and other bacteria.

Ligase mechanism

The mechanism of DNA ligase is to form two covalent phosphodiester bonds between 3' hydroxyl ends of one nucleotide, ("acceptor") with the 5' phosphate end of another ("donor").

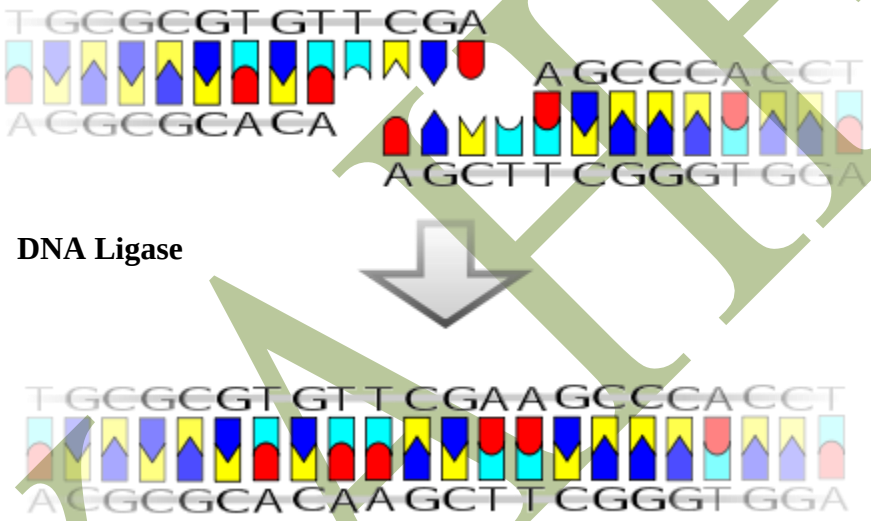
ATP is required for the ligase reaction, which proceeds in three steps:

(1) adenylation (addition of AMP) of a residue in the active center of the enzyme, pyrophosphate is released;

(2) transfer of the AMP to the 5' phosphate of the so-called donor, formation of a pyrophosphate bond;

(3) formation of a phosphodiester bond between the 5' phosphate of the donor and the 3' hydroxyl of the acceptor.

Ligase will also work with blunt ends, although higher enzyme concentrations and different reaction conditions are required.



A pictorial example of how a ligase works (with sticky ends)

In mammals, there are **four specific types of ligase**.

- **DNA ligase I:** ligates the nascent DNA of the lagging strand after the Ribonuclease H has removed the RNA primer from the Okazaki fragments.
- **DNA ligase II:** alternatively spliced form of DNA ligase III found in non-dividing cells.
- **DNA ligase III:** complexes with DNA repair protein XRCC1 to aid in sealing DNA during the process of nucleotide excision repair and recombinant fragments.
- **DNA ligase IV:** complexes with XRCC4. It catalyzes the final step in the non-homologous end joining DNA double-strand break repair pathway. It is also required for V(D)J

recombination, the process that generates diversity in immunoglobulin and T-cell receptor loci during immune system development.

Applications of DNA ligase in molecular biology research

- DNA ligases have become an indispensable tool in modern molecular biology research for generating recombinant DNA sequences.
- For example, DNA ligases are used with restriction enzymes to insert DNA fragments into plasmids.
- Most experiments use T4 DNA Ligase (isolated from bacteriophage T4), which is most active at 25°C.
- However, for optimal ligation efficiency with cohesive-ended fragments ("sticky ends"), the optimal enzyme temperature needs to be balanced with the melting temperature T_m (also the annealing temperature) of the sticky ends being ligated.
- If the ambient temperature exceeds T_m, the homologous pairing of the sticky ends would not be stable because the high temperature disrupts hydrogen bonding.
- Ligation reaction is most efficient when the sticky ends are already stably annealed, disruption of the annealing ends would therefore results in low ligation efficiency.
- The shorter the overhang, the lower the T_m, typically a 4-base overhang has a T_m of 12-16°C.
- Since blunt-ended DNA fragments have no cohesive ends to anneal, the melting temperature is not a factor to consider within the normal temperature range of the ligation reaction. However, the higher the temperature, the less chance that the ends to be joined will be aligned to allow ligation (molecules move around the solution more at higher temperatures). The limiting factor in blunt end ligation is not the activity of the ligase but rather the number of alignments between DNA fragment ends that occur.
- The most efficient ligation temperature for blunt-ended DNA would therefore be the temperature at which the greatest number of alignments can occur.
- Therefore, the majority of blunt-ended ligations are carried out at 14-16°C overnight.

DNA polymerase

- A **DNA polymerase** is an enzyme that helps catalyze in the polymerization of deoxyribonucleotides into a DNA strand.
- DNA polymerases are best known for their feedback role in DNA replication, in which the polymerase "reads" an intact DNA strand as a template and uses it to synthesize the new strand.
- This process copies a piece of DNA.
- The newly polymerized molecule is complementary to the template strand and identical to the template's original partner strand.
- DNA polymerases use magnesium ions as cofactors.
- Human DNA polymerases are 900-1000 amino acids long.

Function of DNA polymerase

- DNA polymerase can add free nucleotides to only the 3' end of the newly forming strand. This results in elongation of the new strand in a 5'-3' direction.
- No known DNA polymerase is able to begin a new chain (*de novo*).
- DNA polymerase can add a nucleotide onto only a preexisting 3'-OH group, and, therefore, needs a primer at which it can add the first nucleotide. Primers consist of RNA and/or DNA bases.
- In DNA replication, the first two bases are always RNA, and are synthesized by another enzyme called primase.
- An enzyme known as a helicase is required to unwind DNA from a double-strand structure to a single-strand structure to facilitate replication of each strand consistent with the semiconservative model of DNA replication.
- Error correction is a property of some, but not all, DNA polymerases.
- This process corrects mistakes in newly synthesized DNA.
- When an incorrect base pair is recognized, DNA polymerase reverses its direction by one base pair of DNA.

- The 3'-5' exonuclease activity of the enzyme allows the incorrect base pair to be excised (this activity is known as *proofreading*).
- Following base excision, the polymerase can re-insert the correct base and replication can continue.

Various DNA polymerases are extensively used in molecular biology experiments.

DNA Polymerase I (or **Pol I**)

- is an enzyme that participates in the process of DNA replication in prokaryotes.
- It is composed of 928 amino acids, and is an example of a processive enzyme - it can sequentially catalyze multiple polymerisations.
- Discovered by Arthur Kornberg in 1956,
- it was the first known DNA polymerase (and, indeed, the first known of any kind of polymerase).
- It was initially characterized in *E. coli*, although it is ubiquitous in prokaryotes.
- In *E. coli* and many other bacteria, the gene which encodes Pol I is known as ***polA***.

Pol I possesses three enzymatic activities:

1. A 5' -> 3' (forward) DNA polymerase activity, requiring a 3' primer site and a template strand
2. A 3' -> 5' (reverse) exonuclease activity that mediates proofreading
3. A 5' -> 3' (forward) exonuclease activity mediating nick translation during DNA repair.
4. In the replication process, DNA Polymerase I removes the RNA primer (created by Primase) from the lagging strand and fills in the necessary nucleotides between the Okazaki fragments in 5' -> 3' direction, proofreading for mistakes as it goes.
5. It is a template-dependent enzyme - it only adds nucleotides that correctly base pair with an existing DNA strand acting as a template.

Research applications of DNA polymerase I

1. DNA polymerase I obtained from *E. coli* is used extensively for molecular biology research.
2. However, the 5' -> 3' exonuclease activity makes it unsuitable for many applications.
3. Fortunately this undesirable enzymatic activity can be simply removed from the holoenzyme to leave a useful molecule called the Klenow fragment, widely used in molecular biology.
4. Exposure of DNA polymerase I to the protease subtilisin cleaves the molecule into a smaller fragment, which retains only the DNA polymerase and proofreading activities.

DNA polymerase II (also known as **DNA Pol II** or **Pol II**)

- is a prokaryotic DNA polymerase most likely involved in DNA repair.
- The enzyme is 90 kDa in size and
- is coded by the polB gene.
- DNA Pol II can synthesize DNA new base pairs at an average rate of between 40 and 50 nucleotides/second.
- Strains lacking the gene show no defect in growth or replication.
- Synthesis of **Pol II** is induced during the stationary phase of cell growth.
- This is a phase in which little growth and DNA synthesis occurs.
- It is also a phase in which the DNA can accumulate damage such as short gaps, which act as a block to DNA Pol III.
- Under these circumstances, Pol II helps to overcome the problem because it can reinitiate DNA synthesis downstream of gaps.
- Pol II has a low error rate but it is much too slow to be of any use in normal DNA synthesis.
- Pol II differs from Pol I in that it lacks a 5'->3' exonuclease activity, and cannot use a nicked duplex template.

DNA polymerase III holoenzyme

- is the primary enzyme complex involved in prokaryotic DNA replication.
- It was discovered by **Thomas Kornberg** and **Malcolm Gefter in 1970**.
- The complex has high processivity (i.e. the number of nucleotides added per binding event) and, specifically referring to the replication of the *E.coli* genome, works in conjunction with four other DNA polymerases (Pol I, Pol II, Pol IV, and Pol V).
- Being the primary holoenzyme involved in replication activity,
- the DNA Pol III holoenzyme also has proofreading capabilities that correct replication mistakes by means of exonuclease activity working 3'->5'.
- DNA Pol III is a component of the replisome, which is located at the replication fork.

The replisome is composed of the following:

- 2 **DNA Pol III enzymes**, each comprising **α**, **ε** and **θ** subunits.
  - the α subunit has the polymerase activity.
  - the ε subunit as 3'-5' exonuclease activity.
  - the θ subunit stimulates the ε subunit's proofreading.
- 2 **β** units which act as sliding DNA clamps, they keep the polymerase bound to the DNA.
- 2 **τ** units which acts to dimerize two of the core enzymes (α, ε, and θ subunits).
- 1 **γ** unit which acts as a clamp loader for the lagging strand Okazaki fragments, helping the two β subunits to form a unit and bind to DNA. The γ unit is made up of 5 γ subunits which include 3 γ subunits, 1 δ subunit, and 1 δ' subunit. The δ is involved in copying of the lagging strand.
- **X** and **Ψ** which form a 1:1 complex and bind to γ or η.

Activity of DNA polymerase III

1. synthesizes base pairs at a rate of around 1000 nucleotides per second.
2. DNA Pol III activity begins after strand separation at the origin of replication.

3. Because DNA synthesis cannot start *de novo*, an RNA primer, complementary to part of the single-stranded DNA, is synthesized by primase (an RNA polymerase)

Taq DNA polymerase

- is a thermostable DNA polymerase named after the thermophilic bacterium *Thermus aquaticus* from which it was originally isolated by **Thomas D. Brock** in **1965**.
- It is often abbreviated to "***Taq* Pol**" (or simply "***Taq***"), and is frequently used in polymerase chain reaction (PCR), a method for greatly amplifying short segments of DNA.
- *T. aquaticus* is a bacterium that lives in hot springs and hydrothermal vents, and *Taq* polymerase was identified as an enzyme able to withstand the protein-denaturing conditions (high temperature) required during PCR.
- Therefore it replaced the DNA polymerase from *E. coli* originally used in PCR.
- *Taq*'s optimum temperature for activity is 75–80°C,
  with a half-life of greater than 2 hours at 92.5°C, 40 minutes at 95°C and 9 minutes at 97.5°C,

  and can replicate a 1000 base pair strand of DNA in less than 10 seconds at 72°C.

***Taq*'s drawbacks**

  - Its relatively low replication fidelity.
  - It lacks a 3' to 5' exonuclease proofreading activity,and
  - has an error rate measured at about 1 in 9,000 nucleotides.
- Some thermostable DNA polymerases have been isolated from other thermophilic bacteria and archaea, such as *Pfu* DNA polymerase, possessing a proofreading activity, and are being used instead of (or in combination with) *Taq* for high-fidelity amplification.
- *Taq* makes DNA products that have A (adenine) overhangs at their 3' ends.

This may be useful in TA cloning, whereby a cloning vector (such as a plasmid) that has a T (thymine) 3' overhang is used, which complements with the A overhang of the PCR product, thus enabling ligation of the PCR product into the plasmid vector.

RNA polymerase (RNAP or RNApol)

- is an enzyme that produces RNA.
- In cells, RNAP is necessary for constructing RNA chains using DNA genes as templates, a process called transcription.
- RNA polymerase enzymes are essential to life and are found in all organisms and many viruses.
- In chemical terms, RNAP is a nucleotidyl transferase that polymerizes ribonucleotides at the 3' end of an RNA transcript.

History of RNA polymerase

- RNAP was discovered independently by **Sam Weiss, Audrey Stevens, and Jerard Hurwitz** in **1960**.
- The 2006 Nobel Prize in Chemistry was awarded to **Roger Kornberg** for creating detailed molecular images of RNA polymerase during various stages of the transcription process.
- RNAP accomplishes *de novo* synthesis.
- It is able to do this because specific interactions with the initiating nucleotide hold RNAP rigidly in place, facilitating chemical attack on the incoming nucleotide.
- Such specific interactions explain why RNAP prefers to start transcripts with ATP (followed by GTP, UTP, and then CTP).
- In contrast to DNA polymerase, RNAP includes helicase activity, therefore no separate enzyme is needed to unwind DNA.

Products of RNAP include:

- Messenger RNA (mRNA)—template for the synthesis of proteins by ribosomes.

- Non-coding RNA or "RNA genes"—a broad class of genes that encode RNA that is not translated into protein. The most prominent examples of RNA genes are transfer RNA (tRNA) and ribosomal RNA (rRNA), both of which are involved in the process of translation. However, since the late 1990s, many new RNA genes have been found, and thus RNA genes may play a much more significant role than previously thought.
    - Transfer RNA (tRNA)—transfers specific amino acids to growing polypeptide chains at the ribosomal site of protein synthesis during translation.
    - Ribosomal RNA (rRNA)—a component of ribosomes.
    - Micro RNA—regulates gene activity.
    - Catalytic RNA (Ribozyme)—enzymatically active RNA molecules.

RNA polymerase in bacteria

In bacteria, the same enzyme catalyzes the synthesis of mRNA and ncRNA.

RNAP is a relatively large molecule. The core enzyme has 5 subunits (~400 kDa):

- α_2: The two α subunits assemble the enzyme and bind regulatory factors. Each subunit has two domains: αCTD (C-Terminal domain) binds the UP element of the extended promoter, and αNTD (N-terminal domain) binds the rest of the polymerase. This subunit is not used on promoters without an UP element.
- β: this has the polymerase activity (catalyzes the synthesis of RNA), which includes chain initiation and elongation.
- β': binds to DNA (nonspecifically).
- ω: restores denatured RNA polymerase to its functional form in vitro. It has been observed to offer a protective/chaperone function to the β' subunit in *Mycobacterium smegmatis*. Now known to promote assembly.[6]

- In order to bind promoter-specific regions, holoenzyme requires another subunit, sigma (ζ).
- The sigma factor greatly reduces the affinity of RNAP for nonspecific DNA while increasing specificity for certain promoter regions, depending on the sigma factor.

- That way, transcription is initiated at the right region.
- The complete holoenzyme therefore has 6 subunits: $\alpha_2\beta\beta'\zeta\omega$ (~480 kDa).
- The structure of RNAP exhibits a groove with a length of 55 Å (5.5 nm) and a diameter of 25 Å (2.5 nm). This groove fits well the 20 Å (2 nm) double strand of DNA. The 55 Å (5.5 nm) length can accept 16 nucleotides.
  - When not in use, RNA polymerase binds to low-affinity sites to allow rapid exchange for an active promoter site when one opens.
  - RNA polymerase holoenzyme, therefore, does not freely float around in the cell when not in use.

RNA polymerase in eukaryotes

Eukaryotes have several types of RNAP, characterized by the type of RNA they synthesize:

1. **RNA polymerase I** - synthesizes a pre-rRNA 45S (35S in yeast), which matures into 28S, 18S and 5.8S rRNAs which will form the major RNA sections of the ribosome.
2. **RNA polymerase II** - synthesizes precursors of mRNAs and most snRNA and microRNAs. This is the most studied type, and due to the high level of control required over transcription a range of transcription factors are required for its binding to promoters.
3. **RNA polymerase III** - synthesizes tRNAs, rRNA 5S and other small RNAs found in the nucleus and cytosol.
4. **RNA polymerase IV** - synthesizes siRNA in plants.
5. **RNA polymerase V** - synthesizes RNAs involved in siRNA-directed heterochromatin formation in plants.

There are other RNA polymerase types in mitochondria and chloroplasts. And there are RNA-dependent RNA polymerases involved in RNA interference.

RNA polymerase in archaea

- Archaea have a single RNAP that is closely related to the three main eukaryotic polymerases (Pol I,II,III).
- Thus, it has been speculated that the archaeal polymerase resembles the ancestor of the specialized eukaryotic polymerases.

RNA polymerase in viruses

- Many viruses also encode for RNAP.
- the most widely studied viral RNAP is found in bacteriophage T7.
- The single-subunit T7 RNA polymerase is related to that found in mitochondria and chloroplasts, and shares considerable homology to DNA polymerase.
- It is believed that most viral polymerases therefore evolved from DNA polymerase and are not directly related to the multi-subunit polymerases described above.
- The viral polymerases are diverse, and include some forms that can use RNA as a template instead of DNA.
- This occurs in negative strand RNA viruses and dsRNA viruses, both of which exist for a portion of their life cycle as double-stranded RNA.
- However, some positive strand RNA viruses, such as polio, also contain these RNA-dependent RNA polymerases.

Reverse transcriptase –

In the fields of molecular biology and biochemistry, a **reverse transcriptase**, also known as **RNA-dependent DNA polymerase**,

- is a DNA polymerase enzyme that transcribes single-stranded RNA into single-stranded DNA.

- It also is a **DNA-dependent DNA polymerase** which synthesizes the second strand of DNA of the double helix DNA once the RNA has been reverse transcribed, and degraded via it's RNaseH activity, into a single strand cDNA.
- Normal transcription involves the synthesis of RNA from DNA; hence, reverse transcription is the *reverse* of this.

Well studied reverse transcriptases include:

- **HIV-1 reverse transcriptase** from human immunodeficiency virus type 1 (PDB 1HMV)
- **M-MLV reverse transcriptase** from the Moloney murine leukemia virus
- **AMV reverse transcriptase** from the avian myeloblastosis virus
- **Telomerase reverse transcriptase** that maintains the telomeres of eukaryotic chromosomes

History of reverse transcriptase

Reverse transcriptase was discovered by **Howard Temin** and independently by **David Baltimore** in **1970**. The two shared the 1975 Nobel Prize in Physiology or Medicine with Renato Dulbecco for their discovery.

Function of reverse transcriptase in viruses

- The enzyme is encoded and used by reverse-transcribing viruses, which use the enzyme during the process of replication.
- Reverse-transcribing RNA viruses, such as retroviruses, use the enzyme to reverse-transcribe their RNA genomes into DNA, which is then integrated into the host genome and replicated along with it.
- Reverse-transcribing DNA viruses, such as the hepadnaviruses, can allow RNA to serve as a template in assembling, and making DNA strands.
- HIV infects humans with the use of this enzyme. Without reverse transcriptase, the viral genome would not be able to incorporate into the host cell, resulting in the failure of the ability to replicate.

Process of reverse transcription

- Reverse transcriptase creates single-stranded DNA from an RNA template.
- In virus species with reverse transcriptase lacking DNA-dependent DNA polymerase activity, creation of double-stranded DNA can possibly be done by host-encoded DNA polymerase δ, mistaking the viral DNA-RNA for a primer and synthesizing a double-stranded DNA by similar mechanism as in primer removal, where the newly synthesized DNA displaces the original RNA template.
- The process of reverse transcription is extremely error-prone and it is during this step that mutations may occur. Such mutations may cause drug resistance.

Retroviral Reverse Transcription

- Retroviruses, also referred to as class VI ssRNA-RT viruses,
- are RNA reverse transcribing viruses with a DNA intermediate.
- Their genomes consist of two molecules of positive sense single stranded RNA with a 5' cap and 3' polyadenylated tail. Examples of retroviruses include *Human Immunodeficiency Virus* (HIV) and *Human T-Lymphotropic virus* (HTLV).

Creation of double-stranded DNA occurs in the cytosol[5] as a series of steps:

1. A specific cellular tRNA acts as a primer and hybridizes to a complementary part of the virus genome called the primer binding site or PBS
2. Complementary DNA then binds to the U5 (non-coding region) and R region (a direct repeat found at both ends of the RNA molecule) of the viral RNA
3. A domain on the reverse transcriptase enzyme called RNAse H degrades the 5’ end of the RNA which removes the U5 and R region
4. The primer then ‘jumps’ to the 3’ end of the viral genome and the newly synthesised DNA strands hybridizes to the complementary R region on the RNA
5. The first strand of complementary DNA (cDNA) is extended and the majority of viral RNA is degraded by RNAse H

6. Once the strand is completed, second strand synthesis is initiated from the viral RNA
7. There is then another 'jump' where the PBS from the second strand hybridizes with the complementary PBS on the first strand
8. Both strands are extended further and can be incorporated into the hosts genome by the enzyme integrase

In eukaryotes

- Self-replicating stretches of eukaryotic genomes known as retrotransposons utilize reverse transcriptase to move from one position in the genome to another via a RNA intermediate. They are found abundantly in the genomes of plants and animals.
- Telomerase is another reverse transcriptase found in many eukaryotes, including humans, which carries its own RNA template; this RNA is used as a template for DNA replication.

In prokaryotes

- Reverse transcriptases are also found in bacterial Retron msr RNAs, distinct sequences which code for reverse transcriptase,
- and are used in the synthesis of msDNA.
- In order to initiate synthesis of DNA, a primer is needed. In bacteria, the primer is synthesized during replication.

Structure of reverse transcriptase

- Reverse transcriptase enzymes include an RNA-dependent DNA polymerase and a DNA-dependent DNA polymerase, which work together to perform transcription.
- In addition to the transcription function, retroviral reverse transcriptases have a domain belonging to the RNase H family which is vital to their replication.

Applications of reverse transcriptase

Antiviral drugs

- As HIV uses reverse transcriptase to copy its genetic material and generate new viruses (part of a retrovirus proliferation circle), specific drugs have been designed to disrupt the process and thereby suppress its growth.
- Collectively, these drugs are known as reverse transcriptase inhibitors and include the nucleoside and nucleotide analogues zidovudine (trade name Retrovir), lamivudine (Epivir) and tenofovir (Viread), as well as non-nucleoside inhibitors, such as nevirapine (Viramune).

Molecular biology

- Reverse transcriptase is commonly used in research to apply the polymerase chain reaction technique to RNA in a technique called reverse transcription polymerase chain reaction (RT-PCR).
- The classical PCR technique can be applied only to DNA strands, but, with the help of reverse transcriptase, RNA can be transcribed into DNA, thus making PCR analysis of RNA molecules possible.
- Reverse transcriptase is used also to create cDNA libraries from mRNA.
- The commercial availability of reverse transcriptase greatly improved knowledge in the area of molecular biology, as, along with other enzymes, it allowed scientists to clone, sequence, and characterise DNA.
- Reverse transcriptase has also been employed in insulin production.
- By inserting eukaryotic mRNA for insulin production along with reverse transcriptase into bacteria, the mRNA can insert itself into the prokaryote's genome, and large amounts of insulin can be created, sidestepping the need to harvest pig pancreas and other such traditional sources.

- Inserting eukaryotic DNA (instead of mRNA) into bacteria would not work because it is fragmented, with introns, and would not transcribe successfully using the bacteria's ribosomes.

Polynucleotide kinase (or PNK)

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

Alkaline phosphatase (**ALP**, **ALKP**) (EC 3.1.3.1)

- is a hydrolase enzyme responsible for removing phosphate groups from many types of molecules, including nucleotides, proteins, and alkaloids.
- The process of removing the phosphate group is called *dephosphorylation*.
- As the name suggests, alkaline phosphatases are most effective in an alkaline environment. It is sometimes used synonymously as **basic phosphatase**.

In bacteria,

- alkaline phosphatase is located in the periplasmic space, external to the cell membrane.
- Since this space is much more subject to environmental variation than the actual interior of the cell, bacterial alkaline phosphatase is comparatively resistant to inactivation, denaturation, and degradation, and also has a higher rate of activity.
- The optimal pH for the activity of the *E. coli* enzyme is 8.0 while the bovine enzyme optimum pH is slightly higher at 8.5.

Common alkaline phosphatases used in research include:

- **Shrimp alkaline phosphatase** (SAP), from a species of Arctic shrimp (*Pandalus borealis*)
- **Calf-intestinal alkaline phosphatase** (CIP)
- **Placental alkaline phosphatase** (PALP) and its C terminally truncated version that lacks the last 24 amino acids (constituting the domain that targets for GPI membrane anchoring) - the secreted alkaline phosphatase (SEAP)

Use in research

- Typical use in the lab for alkaline phosphatases includes removing phosphate monoester to prevent self ligation.
- Alkaline phosphatase has become a useful tool in molecular biology laboratories, since DNA normally possesses phosphate groups on the 5' end.
- Removing these phosphates prevents the DNA from ligating (the 5' end attaching to the 3' end), thereby keeping DNA molecules linear until the next step of the process for which they are being prepared; also, removal of the phosphate groups allows radiolabeling (replacement by radioactive phosphate groups) in order to measure the presence of the labeled DNA through further steps in the process or experiment.
- Another important use of alkaline phosphatase is as a label for enzyme immunoassays.
- One common use in the dairy industry is as a marker of pasteurisation in cows' milk.
- This molecule is denatured by elevated temperatures found during pasteurisation, and can be tested for via colour change of a para-Nitrophenylphosphate substrate in a buffered solution (Aschaffenburg Mullen Test).
- Raw milk would typically produce a yellow colouration within a couple of minutes, whereas properly pasteurised milk should show no change.

Terminal deoxynucleotidyl transferase

Terminal deoxynucleotidyl transferase (TDT), also known as **DNA nucleotidylexotransferase** (DNTT) or **terminal transferase**,

- is a specialized DNA polymerase expressed in immature, pre-B, pre-T lymphoid cells, and acute lymphoblastic leukemia/lymphoma cells.
- TdT adds N-nucleotides to the V,D, and J exons during antibody gene recombination enabling the phenomenon of junctional diversity.
- In humans, terminal transferase is encoded by the *DNTT* gene.

- TdT is notably absent in fetal liver HSCs significantly impairing junctional diversity in B-cells during the fetal period.

Function of TDT

- TdT catalyses the addition of nucleotides to the 3' terminus of a DNA molecule.
- Unlike most DNA polymerases it does not require a template.
- The preferred substrate of this enzyme is a 3'-overhang, but it can also add nucleotides to blunt or recessed 3' ends.
- Cobalt is a necessary cofactor, however the enzyme catalyzes reaction upon Mg and Mn administration *in vitro*.

Uses of TDT

- Terminal transferase has applications in molecular biology.
- It can be used in RACE to add nucleotides which can then be used as a template for a primer in subsequent PCR.
- It can also be used to add nucleotides labeled with radioactive isotopes, for example in the TUNEL assay (**T**erminal deoxynucleotidyl transferase d**U**TP **N**ick **E**nd **L**abeling) for the demonstration of apoptosis (which is marked, in part, by fragmented DNA).
- Also used in the immunofluorescence assay for the diagnosis of acute lymphoblastic leukemia.
- In immunohistochemistry, antibodies to TdT can be used to demonstrate the presence of immature T and B cells and multipotent haematopoietic stem cells, which possess the antigen, while mature lymphoid cells are always TdT-negative.
- While TdT-positive cells are found in small numbers in healthy lymph nodes and tonsils, the malignant cells of acute lymphoblastic leukaemia are also TdT positive, and the antibody can therefore be used as part of a panel to diagnose this disease and to distinguish it from, for example, small cell tumours of childhood.

Deoxyribonuclease (Dnase)

- is any enzyme that catalyzes the hydrolytic cleavage of phosphodiester linkages in the DNA backbone.
- Thus, deoxyribonucleases are one type of nuclease.
- A wide variety of deoxyribonucleases are known, which differ in their substrate specificities, chemical mechanisms, and biological functions.

Modes of action

- Some DNases cleave only residues at the ends of DNA molecules (exodeoxyribonucleases, a type of exonuclease).
- Others cleave anywhere along the chain (endodeoxyribonucleases, a subset of endonucleases).
- Some are fairly indiscriminate about the DNA sequence at which they cut, while others, including restriction enzymes, are very sequence-specific.
- Some cleave only double-stranded DNA; others are specific for single-stranded molecules; and still others are active toward both.
- DNase enzymes can be inhaled using a nebuliser by cystic fibrosis sufferers.
- DNase enzymes help because white blood cells accumulate in the mucus, and, when they break down, they release DNA, which adds to the 'stickiness' of the mucus. DNase enzymes break down the DNA, and the mucus is much easier to clear from the lungs.

Types of deoxyribonucleases

The two main types of DNase found in metazoans are known as

1. deoxyribonuclease I and
2. deoxyribonuclease II.

Other types of DNase include Micrococcal nuclease.

Assay of deoxyribonucleases

- DNA absorbs UV light with a wavelength of maximal absorbance near 260 nm.
- This absorption is due to the pi electrons in the aromatic bases of the DNA.
- In dsDNA, or even regions of RNA where double-stranded structure occurs, the bases are stacked parallel to each other, and the overlap of the base molecular orbitals leads to a decrease in absorbance of UV light.
- This phenomenon is called the hyperchromic effect.
- When DNAse liberates nucleotides from dsDNA, the bases are no longer stacked as they are in dsDNA, so that orbital overlap is minimized and UV absorbance increases.
- This increase in absorbance underlies the basis of Kunitz unit of DNAse activity.
- One Kunitz unit is defined as the amount of enzyme that causes an increase in absorbance at 260 nm of 0.001 per mL when acting upon highly polymerized DNA at 25 ºC and pH 5.0 under specified conditions.
- A standard enzyme preparation should be run in parallel with an unknown because standardization of DNA preparations and their degree of polymerization in solution is not possible.

Ribonuclease (RNase)

- is a type of nuclease that catalyzes the degradation of RNA into smaller components.
- Ribonucleases can be divided into
    i. endoribonucleases
    ii. exoribonucleases,
- comprise several sub-classes within the EC 2.7 (for the phosphorolytic enzymes) and 3.1 (for the hydrolytic enzymes) classes of enzymes.

Function

- All organisms studied contain many RNases of many different classes, showing that RNA degradation is a very ancient and important process.
- As well as cleaning of cellular RNA that is no longer required,
- RNases play key roles in the maturation of all RNA molecules, both messenger RNAs that carry genetic material for making proteins, and non-coding RNAs that function in varied cellular processes.
- In addition, active RNA degradation systems are a first defense against RNA viruses, and provide the underlying machinery for more advanced cellular immune strategies such as RNAi.
- Some cells also secrete copious quantities of non-specific RNases such as A and T1.
- RNases are, therefore, extremely common, resulting in very short lifespans for any RNA that is not in a protected environment.
- It is worth noting that all intracellular RNAs are protected from RNase activity by a number of strategies including 5' end capping, 3' end polyadenylation, and folding within an RNA protein complex (ribonucleoprotein particle or RNP).
- RNases play a critical role in many biological processes, including angiogenesis and self-incompatibility in flowering plants (angiosperms).
- Also, RNases in prokaryotic toxin-antitoxin systems are proposed to function as plasmid stability loci, and as stress-response elements when present on the chromosome.
- Similar to restriction enzymes, which cleave highly specific sequences of double-stranded DNA, a variety of endoribonucleases that recognize and cleave specific sequences of single-stranded RNA have been recently classified.
- RNases play a critical role in many biological processes, including angiogenesis and self-incompatibility in flowering plants (angiosperms).

- Also, RNases in prokaryotic toxin-antitoxin systems are proposed to function as plasmid stability loci, and as stress-response elements when present on the chromosome.

Major types of endoribonucleases

- **RNase A**

  is an RNase that is commonly used in research.

  RNase A (e.g., bovine pancreatic ribonuclease A) is one of the hardiest enzymes in common laboratory usage; one method of isolating it is to boil a crude cellular extract until all enzymes other than RNase A are denatured.

  It is specific for single-stranded RNAs.

  It cleaves 3'end of unpaired C and U residues, leaving a 3'-phosphorylated product, via a 2',3'-cyclic monophosphate.

- **RNase H**

  is a ribonuclease that cleaves the RNA in a DNA/RNA duplex to produce ssDNA.

  RNase H is a non-specific endonuclease and catalyzes the cleavage of RNA via a hydrolytic mechanism, aided by an enzyme-bound divalent metal ion. RNase H leaves a 5'-phosphorylated product.

- **RNase I**

  cleaves 3'-end of ssRNA at all dinucleotide bonds leaving a 5' hydroxyl, and 3' phosphate, via a 2',3'-cyclic monophosphate intermediate.

- **RNase III**

  is a type of ribonuclease that cleaves rRNA (16s rRNA and 23s rRNA) from transcribed polycistronic RNA operon in prokaryotes.

It also digests double strands RNA (dsRNS)-Dicer family of RNAse, cutting pre-miRNA (60–70bp long) at a specific site and transforming it in miRNA (22–30bp), that is actively involved in the regulation of transcription and mRNA life-time.

- **RNase L**
  - is an interferon-induced nuclease that, upon activation, destroys all RNA within the cell
- **RNase P**
  - is a type of ribonuclease that is unique in that it is a ribozyme – a ribonucleic acid that acts as a catalyst in the same way as an enzyme.
  - Its function is to cleave off an extra, or precursor, sequence on tRNA molecules.
  - RNase P is one of two known multiple turnover ribozymes in nature (the other being the ribosome).
  - A form of RNase P that is a protein and does not contain RNA has recently been discovered.
- **RNase PhyM**
  - is sequence specific for single-stranded RNAs. It cleaves 3'-end of unpaired A and U residues.
- **RNase T1**
  - is sequence specific for single-stranded RNAs.
  - It cleaves 3'-end of unpaired G residues.
- **RNase T2**
  - is sequence specific for single-stranded RNAs.
  - It cleaves 3'-end of all 4 residues, but preferentially 3'-end of As.
-
- **RNase U2**
  - is sequence specific for single-stranded RNAs.
  - It cleaves 3'-end of unpaired A residues.
  -

- **RNase V1**
  - is non-sequence specific for double-stranded RNAs.
  - It cleaves base-paired nucleotide residues.
- **RNase V**

Major types of exoribonucleases

- **Polynucleotide Phosphorylase (PNPase)** functions as an exonuclease as well as a nucleotidyltransferase.
- **RNase PH** functions as an exonuclease as well as a nucleotidyltransferase.
- **RNase II** is responsible for the processive 3'-to-5' degradation of single-stranded RNA.
- **RNase R** is a close homolog of RNase II, but it can, unlike RNase II, degrade RNA with secondary structures without help of accessory factors.
- **RNase D** is involved in the 3'-to-5' processing of pre-tRNAs.
- **RNase T** is the major contributor for the 3'-to-5' maturation of many stable RNAs.
- **Oligoribonuclease** degrades short oligonucleotides to mononucleotides.
- **Exoribonuclease I** degrades single-stranded RNA from 5'-to-3', exists only in eukaryotes.
- **Exoribonuclease II** is a close homolog of Exoribonuclease I.

Cloning Vectors

- A cloning vector is a DNA molecule that has the ability to replicate in an appropriate host cell, and into which the DNA insert is integrated for cloning.
- Therefore, a vector must have an origin of DNA replication (denoted as ori) that functions in the host cell.
- Any extrachromosomal small genome, e.g., plasmid, phage and virus, may be used as a vector.

Properties of Good Vector

- It should be able to replicate autonomously.

- When the objective of cloning is to obtain a large number of copies of the DNA insert, the vector replication must be under relaxed control so that it can generate multiple copies of itself in a single host cell.
- It should be easy to isolate and purify.
- It should be easily introduced into the host cells, Le., transformation of the host with the vector should be easy
- The vector should have suitable marker genes that allow easy detection and/or selection of the transformed host cells.
- When the objective is gene transfer, it should have the ability to integrate either itself or the DNA insert it carries into the genome of the host cell.
- The cells transformed with the vector containing the DNA insert (recombinant DNA) should be identifiable be selectable from those transformed by the unaltered vector.
- A vector should contain unique target sites for as may restriction enzymes as possible into which the DNA insert can be integrated.
- When expression of the DNA insert is desired, the vector should contain atleast suitable control elements, e.g., promoter, operator and ribosome binding sites.

It should be kept in mind that

(1) the DNA molecules used as vectors have coevolved with their specific natural host species, and hence are adapted to function well in them and in their closely related species. Therefore, the choice of vector depends largely on the host species into which the DNA insert of gene is to be cloned.

(2) most naturally occurring vectors do not have all the required functions; therefore, useful vectors have been created by joining together segments performing specific functions (called modules) from two or more natural entities.

Cloning and Expression Vectors

All vectors used for propagation of DNA inserts in a suitable host are called **cloning vectors**. But when a vector is designed for the expression of, i.e., production of the protein specified by, the DNA insert, it is termed as **expression vector**.

- As a rule, such vectors contain at least the regulatory sequences, i.e., promoters, operators, ribosomal binding sites, etc., having optimum function in the chosen host.
- When an eukaryotic gene is to be expressed in a prokaryote, the eukaryotic coding sequence has to be placed after prokaryotic promoter and ribosome building site since the regulatory sequences of eukaryotic are not recognised in prokaryotes.
- In addition, eukaryotes genes, as a rule, contain introns (noncoding regions) present within their coding regions.
- These introns must be removed from the DNA insert to enable the proper expression of eukaryotic genes since prokaryotes lack the machinery needed for their removal from the RNA transcripts.
- When eukaryotic genes are isolated as cDNA, they are intron-free and, hence, suitable for expression in prokaryotes.

E.coli vectors

- Bacteria are the hosts of choice for DNA cloning.
- Among them, *E. coli* occupies a prominent position since cloning and isolating DNA inserts for structural analysis is the easiest in this host.
- Therefore, the initial cloning experiments are generally carried out in *E. coli*.
- The E. coli strain K12 is the most commonly used;
- it has several substrains, e.g., C600, RRI, HB101, etc., each of which has some specific features important in cloning.

- For example, the substrain RRI has, in addition to certain other features, the mutation hsdR, which inactivates the restriction enzyme endogenous to *E.coli* K12; this minimises the degradation of recombinant DNA introduced into it.

Properties of Good Host

A good host should have the following features:

(1) be easy to transform,

(2) support the replication of recombinant DNA,

(3) be free from elements that interfere with replication of recombinant DNA,

(4) lack active restriction enzymes, e.g., E. coli K12 substrain HB101,

(5) should not have methylases since these enzymes would methylate the replicated recombinant DNA. which, as a result, would become resistant to useful restriction enzymes, and

(6) be deficient in normal recombination function so that the DNA insert is not altered by recombination events.

***E. coli* supports several types of vectors, which can be grouped as follows:**

(1) plasmids,

(2) bacteriophages (both natural),

(3) cosmids,

(4) phasmids,

(5) shuttle vectors

Plasmid vector

- Plasmids are widely used as cloning vector,
- Plasmids are replicons which are stably inherited in an extrachromosomal state.
- Most plasmids exist as double-stranded circular DNA molecules.
- If both strands of DNA are intact circles the molecules are described as covalently closed circles or CCC DNA.
- If only one strand is intact, then the molecules are described as open circles or OC DNA.
- When isolated from cells, covalently closed circles often have a deficiency of turns in the double helix,
- such that they have a supercoiled configuration.
- Plasmids are widely distributed throughout the prokaryotes, vary in size from less than 1 x 10^6 daltons to greater than 200 x 10^6, and are generally dispensable.
- Plasmids to which phenotypic traits have not yet been ascribed are called *cryptic* plasmids.

Plasmids can be categorized into one of two major type – depending upon whether or not they carry a set of transfer genes, called the *tra* genes, which promote bacterial conjugation.

1. conjugative
2. non-conjugative

Plasmids can also be categorized on the basis of copies per cell being maintained as

1. *relaxed* plasmids - multiple copies per cell
2. *stringent* plasmids - a limited number of copies per cell

pBR322 - An ideal plasmid vector must have the following functions:

(1) minimum amount of DNA,
(2) relaxed replication control,
(3) at least two selectable markers,
(4) only one (unique) recognition site for at least one restriction endonuclease, and

(5) for easy selection of the recombinant DNA, this unique restriction site must be located within one of the two selectable markers.

The name pBR denotes the following:

p signifies plasmid,

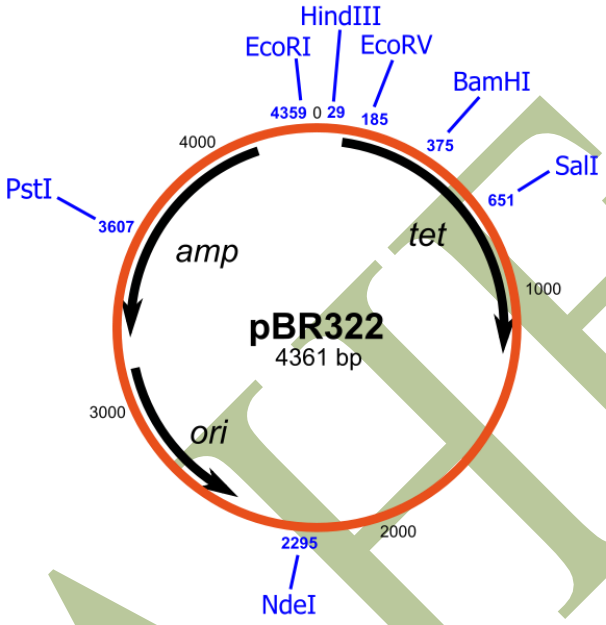
B is from Boliver, and

R is from Rodriguez, the two initials of the scientist who developed pBR322.

pBR322

- is the most popular and
- most widely used plasmid of 4363 bp;
- It has the replication module of *E. coli* plasmid Col El.
- This module has been incorporated in many other plasmid vectors since it permits plasmid replication even when chromosome replication and cell division are inhibited by amino acid starvation or chloramphemicol.
- Under such conditions, each cell accumulates several thousand copies of the plasmid so that one litre of bacterial culture easily yields a milligram of plasmid DNA.
- It has two selectable markers (tetracycline, tetr, and ampicillin, amp', resistance genes), and unique recognition sites for 12 different restriction enzymes (two unique sites, PstI and PvuI, are located within the amp' gene, and 4, e.g., BamHI, SalI, etc., are within tetr gene). The presence of restriction sites within the markers tetr and ampr permits an easy selection for cells transformed with the recombinant pBR322.
- Insertion of the DNA fragment into the plasmid using restriction enzyme PstI or PvuI places the DNA insert within the gene amp'; this makes amp' nonfunctional.
- Bacterial cells containing such a recombinant pBR322 will be unable to grow in the presence of ampicillin, but will grow on tetracycline.

- Similarly, when restriction enzyme BamHI or SalI is used, the DNA insert is placed within the gene tetr making it nonfunctional.



- Bacterial cells possessing such a recombinant pBR322 will, therefore, grow on ampicillin but not on tetracycline.
- This feature allows an easy selection of a single bacterial cell having recombinant pBR322 from among 108 other types of cells.
- Transformed E. coli cells are first plated on an agar medium containing the antibiotic within the resistance gene for which the DNA fragment is not inserted, i.e., for which the bacterial cells having the recombinant DNA are expected to be resistant.
- This eliminates nontransformed bacterial cells; the resulting bacterial colonies will posses either recombinant or unaltered pBR322.
- The colonies so obtained are then replicaplated on agar plates containing the other antibiotic (within the resistance gene for which the DNA insert is placed);
- all the colonies that develop on this plate will contain the unaltered pBR322.

- Therefore, the antibiotic sensitive colonies are identified and recovered from the master plate; these colonies will have the recombinant pBR322. This entire process may take up to 2 days.

pUC18-19

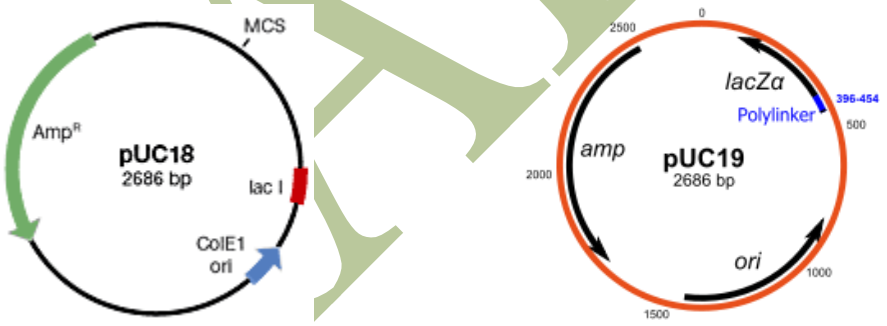
- pUC gets its name from University of California.
- It is a derivative of pBR322 and is much smaller (-2.7 kb);
- it has all the essential parts of pBR322, e.g.,

(1) ampicillin resistance gene and

(2) Col E1 origin.

(3) The second scrobale marker is due to *E. coli* gene lacZα encoding the a fragment of β-galactosidase, the enzyme that hydrolyses lactose.

(4) The *E.coli* strains, e.g., JMIO3, JM1O9, used as hosts for the pUC series vectors have the lacZα deleted from their lacZ genes.



- When pUC enters such an *E.coli* cell, the host genome and the plasmid encode for different parts of the β-galactosidase enzyme, which interact with each other to produce the active enzyme enabling these cells to hydrolyse lactose. β -galactosidase also hydrolyses X-gal (5-Bromo-4-chloro-3-indolyl-p-D-galactoside) to yield a blue dye.
- Therefore appropriate lacZ- *E. coli* cells transformed by the pUC vectors behave as lacZ+ and produce blue coloured colonies on a X-gal containing medium.

- A poly linker sequence located within the lacZα provides several (10 in case of pUC18/pUC19) unique restriction sites for DNA insertion.
- The polylinker sequence by itself does not interfere with lacZα ' expression.
- But when a DNA insert is placed within it, lacZα expression is prevented.
- Vectors pUC18 and pUC19 are identical, except for the orientation of the polylinker sequence, which is oriented in the opposite directions in the two vectors.
- The unique restriction sites used for integration of DNA inserts into pUC vectors interrupt the lacZα fragment so that appropriate
- E. coli cells possessing recombinant pUC DNA are β-galactosidase deficient and, as a result,
- produce white colonies on X-gal medium.
- Therefore, appropriate E.coli cells transformed with pUC recombinant DNA are grown on ampicillin, X-gal and IPG (isopropyl- β D-thiogalactoside; it serves as inducer of β -galactosidase, while X-gal itself can not) containing medium to eliminate non transformed cells.

  The white colonies are selected as they contain the recombinant DNA (in contrast, blue colonies will contain the unaltered vector). The other vectors in pUC series are pUC 8, pUC 9, pUC 12, pUC 13, etc.

The pUC series vectors offer the following adantages over pBR322:

(1) each E. coli cell produces up to 700 copies without any treatment,
(2) cells containing recombinant DNA are selected in a single step,
(3) the sites for DNA insert integration are confined to the poly linker, which permits the use of two restriction anzymes to open the vector, and
(4) they also allow sequencing of the DNA insert.

Bacteriophage Vectors

- Bacteriophages are viruses that attack bacteria.
- Most phages lyse the bacterial cells they infect (lytic phages).

- But many others can choose to follow either a lytic or a lysogenic cycle; in the latter situation, the phage chromosome integrates into the bacterial chromosome and multiplies with the latter as prophage (temperate or lysogenic phages).
- The prophage may dissociate from the bacterial chromosome and follow the lytic cycle.

Several bacteriophages are used as cloning vectors,

- the most commonly used E. coli phages being λ (lambda) and M13 phages.

Plasmid vectors have to be introduced into bacterial cells, which are then cloned and selected for the recovery of recombinant DNA.

In contrast, the phage vectors are directly tested on an appropriate bacterial lawn (a continuous bacterial growth on an agar plate) where each phage particle forms a plaque (a clear bacteria-free zone in the bacterial lawn).

Phage vectors present two advantages over plasmid vectors.

(1) They are more efficient than plasmids for cloning of large DNA fragments; the largest cloned insert size in a λ vector is just over 24 kb, while that for plasmid vectors it is less than 15 kb. In addition,

(2) it is easier to screen a large number of phage plaques than bacterial colonies for the identification of recombinant plaques/clones.

Lambda (λ) Phage Vectors –

- The λ genome (total 48,502 bp) contains an
- origin of replication,
- genes for head and tail proteins and
- enzymes for DNA replication, lysis and lysogeny, and
- single-stranded protruding cohesive ends of 12 bases (5' GGGCGGCGACCT; the other end is complementary to it, i.e., CCCGCCGCTGGA 5').

- The λ genome remains linear in the phage head, but within E. coli cells the two cohesive ends anneal to form a circular molecule necessary for replication.
- The sealed cohesive ends are called cos sites, which are the sites of cleavage during and are necessary for packaging of the mature phage DNA into phage heads.
- The λ DNA must be larger than 38 kb and smaller than 52 kb to be packaged into phage particles.
- The genes for lysogeny are located in the segment between 20 and 38 kb; the whole or a part of this segment is deleted to create λ vectors to

(1) accommodate larger DNA inserts and
(2) to ensure that the recombinant phage is always lytic.

Several vectors were produced from wild type λ genome by mutation and recombination in vivo as well as by recombinant DNA techniques.

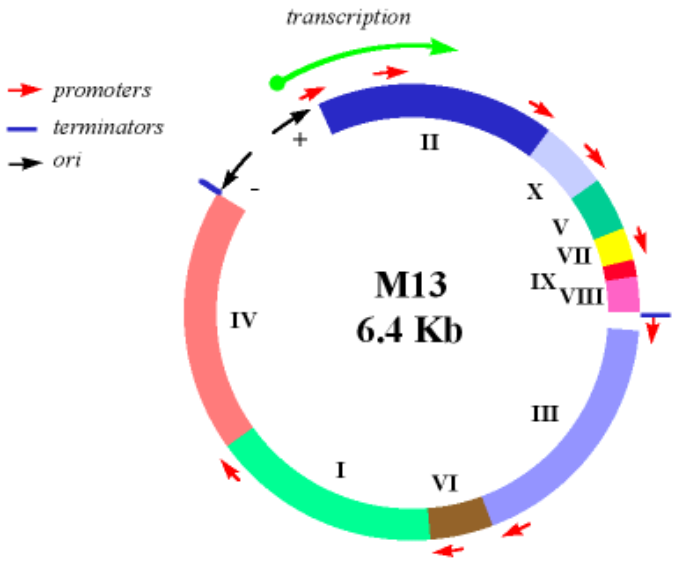
These vectors have the following two basic features.

(1) The vector itself can be propagated as phage in E. coli cells enabling preparation of vector DNA.
(2) They contain restriction sites, which allow the removal of the lysogenic segment and also provide insertion site for the DNA fragment.
(3) During annealing and ligation of the DNA insert with the λ vector, two or more recombinant DNAs may join end-to-end producing a concatemer, which is the proper precursor for packaging of λ genome into phage heads.

Phage M13 Vectors –

- These vectors are used for obtaining single-strand copies of cloned DNA, which are especially suited for DNA sequencing.
- They are derived from the 6.4 kb genome of the E. coli filamentous bacteriophage M13.

- This phage has a single-stranded linear DNA genome in phage particles, which converts into a double stranded circular replicative intermediate within the host cells.
- M13 infects only F+ cells; it does not kill the cells, but forms turbid plaques due to growth retardation of infected cells.
- Ordinarily, the double stranded form is used to produce recombinant molecules since single-stranded DNAs are not cleaved by type II restriction endonucleases; this form is readily isolated from M13-infected E. coli cells.



However, the single-stranded form of M 13 is used to recover single-stranded copies of the DNA inserts; this form of vector is available from the phage particles abundant in the growth medium.

Properties of M13 Vectors –

M13 genome has been used to produce M13mp series of vectors, e.g., M13mp 8, M13mp 9, etc.

The desirable features of M13 vectors are as follows:

(1) very large inserts can be cloned since packaging does not depend on genome size (as is the case with λ vectors).

(2) Pure single-strand copies of double-strand DNA inserts are obtained in abundance.

(3) Since DNA inserts are accepted in either of the orientations (which is also the case for plasmid and λ vectors), some recombinant clones will produce single-strand copies of one strand of the DNA double-strand, while others would produce copies of the complementary strand of the DNA insert.

The phage particles in a single plaque, as a rule, will yield copies of the same single-strand. This property is very useful for a precise DNA sequencing (using both the strands of a DNA molecule) and for the synthesis of specific radio-labelled DNA probes.

(4) Bacterial cells infected by these vectors remain viable as in the case of plasmid vectors; this allows easy maintenance of the vector. Finally,

(5) they form plaques like λ phage vectors making selection of the recombinant DNAs rather easy, and

(6) the recombinant DNA is obtained within stable bacteriophage particles.

Two types of λ phage vectors

1. **Insertional vectors**
2. **Replacment vectors**

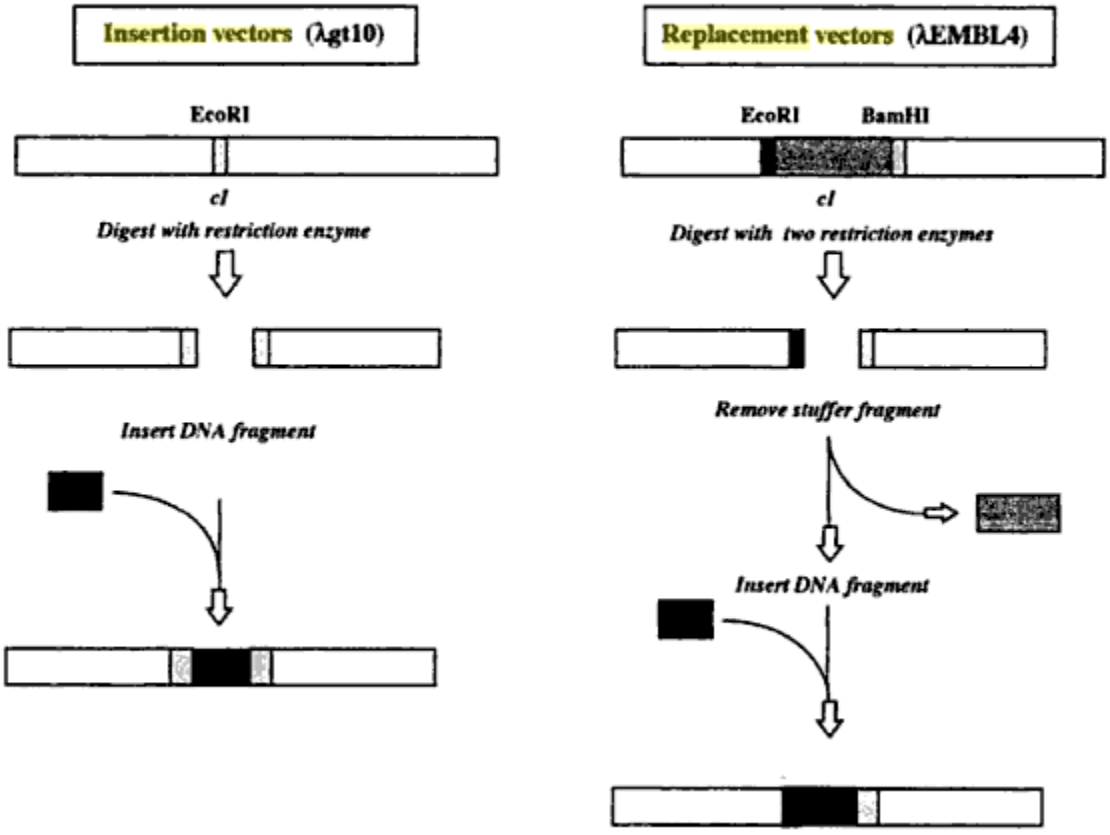
The λ insertional vectors – accept less DNA than the replacement type, the foreign DNA is simply inserted into a region of the phage genome with appropriate restriction sites.

Example - **λgt10, λcharon16A**

In λ replacement vectors – a central region of DNA not essential for lytic growth is removed. This creates two DNA fragments, called right and left arms. The central stuffer fragment is replaced by inserting foreign DNA between the arms to form a functional recombinant λ phage.

Example – λembl, λZAP

General schemes used for cloning in λ insertion and λ replacement vectors



Cosmid Vectors

Cosmids are essentially plasmids that contain a minimum of 250 bp of λ DNA, which includes

(1) the cos site (the sequence yielding cohesive ends) and

(2) sequences needed for binding of and cleavage by terminase so that under appropriate conditions they are packaged in vitro into empty λ phage particles.

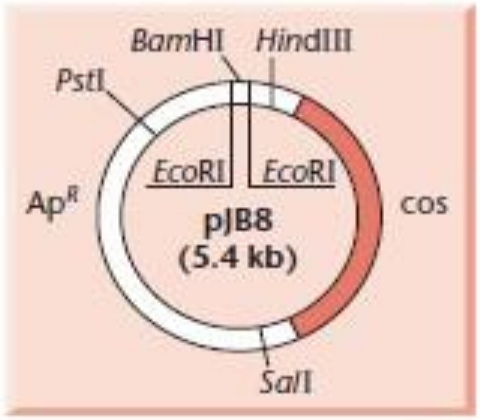
A typical cosmid has

(1) replication origin,

(2) unique restriction sites and

(3) selectable markers from the plasmid; therefore, selection strategy for obtaining the recombinant DNA is based on that for the contributing plasmid.

- Cosmid vectors are constructed using recombinant DNA techniques.
- The cosmid vectors are opened by the appropriate restriction enzyme at a unique site, are then mixed with DNA inserts prepared by using the same enzyme and annealed.
- Among the several types of products, long cancatemers are present, which are the appropriate precursors for packaging in λ. particles.
- This procedure selects for long DNA inserts since for packaging the distance between two cos sites must be between 38 and 52 kb.
- Cosmids can accommodate upto 45 kb long DNA inserts.
- Packaged cosmids infect host cells like λ particles, but once inside the host they replicate and propagate like plasmids



The typical features of cosmids areas follows:

(1) they can be used to clone, DNA inserts of upto 45 kb.

(2) They can be packaged into A. particles that infect host cells, which is many-fold more efficient than plasmid transformation.

(3) Selection for recombinant DNA is based on the procedure applicable to the plasmid making up the cosmid.

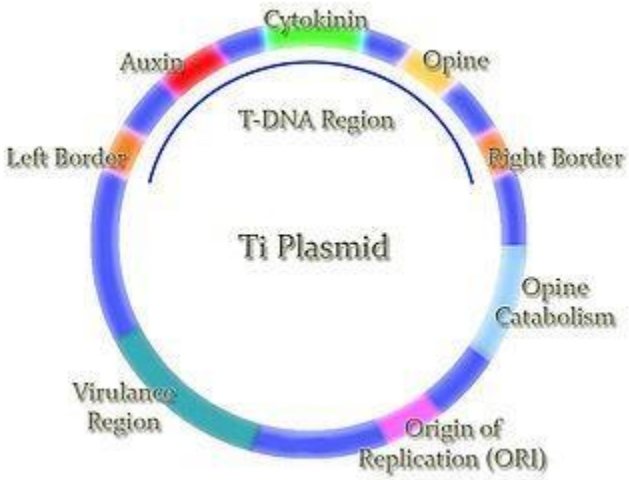
(4) Finally, these vectors are amplified and maintained in the same manner as the contributing plasmid.

Phasmid Vectors

- These vectors are shortened linear A. genomes containing DNA replication and lytic functions plus the cohesive ends of the phage; their middle nonessential segment is replaced by a linearized plasmid with intact replication module.
- In practice, a plasmid vector contains several tandem copies of the plasmid to make it longer than 38 kb, the minimum size needed for packaging in λ particles.
- During construction of the recombinant DNA, one or more copies of the plasmid are deleted from and the DNA insert is integrated into the vector, but generally one copy of the plasmid is retained in the recombinant DNA.
- Phasmids, both recombinant and unaltered, are packaged in A. particles in vitro and used for infection of appropriate *E. coli* cells.
- If a plasmid lacks the λ gene el, which produces the lysis repressor, it multiplies like a phage and produces plaques on a bacterial lawn. But if el gene is present, the plasmid replicates like a plasmid.
- Further, a plasmid may contain a mutant cI gene, which produces a temperature sensitive CI protein (inactive at higher temperatures); such vectors replicate as plasmids at lower temperatures, but behave like phage at higher temperatures. This feature is quite useful in some experiments.

Vectors For Plants

- Plants cells do not contain any plasmid.
- But two plasmids, called pTi and pRi, and present in the bacteria *Agrobacterium tumefaciens* and *A. rhizogenes*, respectively, provide a naturally occurring transformation system.
- These plasmids transfer a part of their DNA, called T-DNA, into the genomes of most dicot and some monocot plants.
- These plasmids, especially the Ti plasmid, have been used to develop a variety of vectors.
- In addition, genomes of many plant viruses are being developed as vectors.



- The purpose of plant vectors is almost always a stable transformation ordinarily in the form of integration in plant genomes.
- But in the case of virus vectors, the objective is to produce large quantities of the protein encoded by the DNA insert.

Animal viral vectors

Retroviruses

- They are one of the main type of current gene therapy approaches.
- The recombinant retroviruses such as the Moloney murine leukemia virus have the ability to integrate into the host genome in a stable fashion.
- They contain a reverse transcriptase that allows integration into the host genome.
- They have been used in a number of FDA-approved clinical trials such as the SCID-X1 trial.
- Retroviral vectors can either be replication-competent or replication-defective.
- Replication-defective vectors are the most common choice in studies because the viruses have had the coding regions for the genes necessary for additional rounds of virion replication and packaging replaced with other genes, or deleted.
- These virus are capable of infecting their target cells and delivering their viral payload, but then fail to continue the typical lytic pathway that leads to cell lysis and death.
- Conversely, replication-competent viral vectors contain all necessary genes for virion synthesis, and continue to propagate themselves once infection occurs.
- Because the viral genome for these vectors is much lengthier, the length of the actual inserted gene of interest is limited compared to the possible length of the insert for replication-defective vectors. Depending on the viral vector, the typical maximum length of an allowable DNA insert in a replication-defective viral vector is usually about 8–10 kB.

Drawback to use of retroviruses the Moloney retrovirus involves the requirement for cells to be actively dividing for transduction.

- As a result, cells such as neurons are very resistant to infection and transduction by retroviruses.
- There is concern that insertional mutagenesis due to integration into the host genome might lead to cancer or leukemia.

- This concern remained theoretical until gene therapy for ten SCID-X1 patients using Maloney murine leukemia virus resulted in two cases of leukemia caused by activation of the LMO2 oncogene due to nearby integration of the vector.

Shuttle Vectors

- These vectors have been designed to replicate in cells of two different species; therefore, they contain two origins of replication, one specific for each host species, as well as those genes necessary for their replication and not provided by the host cells.
- These vectors are created by recombinant techniques.
- Some of them can be grown in two different prokaryotic species, while others can propagate in a prokaryotic species, usually E. coli, and a eukaryotic one, e.g. yeast, plants, animals.
- Since these vectors can be grown in one host and then moved into another without any extra manipulation, they are called shuttle vectors.

A shuttle vector designed to replicate in *E. coli* and *Streptomyces* has been constructed as follows:
(1) the modules for DNA replication in *Streptomyces* and methylenomycin A resistance are derived from a streptomyces plasmid, and

(2) the replication module for maintenance in *E. coli* and a gene for antibiotic resistance are taken from an *E. coli* plasmid.

- This shuttle vector allows the initial cloning of *Streptomyces* DNA inserts in *E. coli* and their subsequent functional tests in Streptomyces.
- Shuttle vectors have been designed to specifically satisfy this need, i.e., the initial cloning of DNA inserts in *E. coli* and subsequent functional tests in the species to which the DNA inserts belong.
- Most of the eukaryotic vectors are, in fact, shuttle vectors.

. Transformation of rDNAs to host cell

Transfer of rDNA into a suitable host cell is an important step in recombinant DNA technology looking for maintenance and expression of a foreign DNA in that cell.

The rDNA is safe in the cell and replicates independently of the chromosomal DNA of the cell. While doing so, the desired foreign gene present in the rDNA expresses its characters in the host cell. The host cell may be a bacterium or plant cell or animal cell. The cell containing an rDNA is known as transformed cell or transformant or recombinant.

Scientists have developed several methods to transfer genes into different types of host cells. The right method to be selected however depends on the type of gene cloning vector and nature of host cell, i.e. whether it is a bacterial cell or plant cell or animal cell.

The rDNAs can be introduced into host cells by *Transformation, Biolistics, Transduction, Electroporation, Transfection, Liposome fusion, Microinjection.*

Gene Transfer into Bacterial Cells

The rDNA can be delivered into the bacterial cells by *Transformation, transduction and electroporation.* These methods are briefly described below.

Transformation

- Direct intake of DNA fragments in the medium by bacterial cells is called transformation. This method was first adopted to transfer rDNA into *E.coli* cells by Mandell and Higa in 1979.
- The ability of bacterial cells to intake DNAs from the medium is said to becompetence.
- Competence can be increased by physical or chemical treatment.
- Competent bacterial cells can intake rDNAs with the size liss than 15 kbp from the culture.
- To introduce rDNAs into *E.coli* cells, the rDNA is added to the bacterial culture and the culture is treated with 50mM calcium chloride ($CaCl_2$) solution at room temperature.

- $CaCl_2$ adheres the rDNAs onto the surface of *E.coli* cells.
- It modifies the bacterial cell wall to intake rDNAs.
- The bacterila culture is then heated gently upto 42°C to induce E.coli cells to intake the rDNAs.
- *E.coli* cells intake the rDNA and become recombinants.
- Transformation method is also used to transfer rDNA into Cyanobacteria, microalgae, yeast and fungi.

Transduction

- Introduction of a foreign DNA in to a bacterial cell through a genetically modified virus or bacteriophage is called transduction.
- rDNAs of 38-50kbp size can be introduced into bacteria using this technique, but the bacteria should be the natural host of the virus.
- Phage derived vectors and cosmids have two cos-sites, one on either side of the DNA.
- These two cos-sides help the DNA to get packed inside the protein coat of lamda phage .
- The desired DNA is inserted into the viral vector DNA to construct an rDNA.
- The rDNA is packed in the protein coat of virus to get infective recombinant virus particles.
- When these virus particles are allowed to infect bacteria in a culture, the virus infects its rDNA into the bacterium.
- Thus, the bacterium receives foreign DNA in the rDNA from virus infection.

Electroporation

- Electroporation is a process of changing the permeability of the cell membrane of cells to uptake macromolecules or organelles in the medium.
- Its is done with an electrical instrument called electroporator.
- The electroporation unit consists of an electroporator creates electric current and sends it through the electrodes.

- It generates 50-240 volts through the electrodes. the aluminium electrodes conduct the electric pluse through the medium in the cuvette.
- The electrodes are fixed on the inner side of the cuvette.
- The cuvette is a small vessel of 10-50 ml capacity made of plastic or glass.
- The cell and DNA solution are mixed together and pipetted into the cuvette.
- The electoporator is setup to produce a current of 50-240 volts for 1-10 minutes.
- The electric current induces some permeability to the cell membrane and hence the cells uptakes the DNA from the medium.
- Electroporation is useful to transfer bacteria, fungi, plant cells and animal cells. The frequency of uptake can be increased by adding colcemid to the reaction mixture. To introduce rDNAs into E.coli cells an electric pulse of 2.5 kilo volt/cm is generated with the electrodes for 4.6 milliseconds. it is a rapid method to create more number of recombinants within a short time.

Gene transfer in to plant cell

A few important methods are available to delivered rDNA into plant cells and tissue.

1. Transfection,
2. electroporation,
3. ultrasonication,
4. partical bombardment,
5. agro infection and
6. liposome fusion.

These methods are discussed below:

Transfection

- The direct intake of naked DNA from the culture medium by eukaryotic cells, is called **transfection** or **direct transformation** it is very similar to transformation in bacteria. Transfection takes place in plant cells and animal cell in vitro.

- Transformation in eukaryotic cells refers to differentiation of normal cells into cancer cells. Hence the word transfection is used to explain the same process in eukaryotes.
- young plants can uptake DNA fragments added to the neutrient solution by their route system. However, only a few cells get the DNA while most cells are been non-recombinants. Hence this method is not useful for gene manipulation.
- J.Paszkovski et al. (1984) made the isolated protoplast of tobacco to uptake rDNA having kanamycin resistance gene ,from the culture medium.They used PEG to induce competence in the protoplast.
- The protoplast took the rDNA from the medium and grew well in the medium containing kanamycin.
- The plants regenerate from the recombinants protoplast were resistant to kanamycin.this method provides the rDNA to all cells of the plant.
- How ever , large DNAs cannot be introduced into plant cells by this method.

Electoporation

- Electoporation is a process of changing the permeability of cell membrance of cells to uptake macromolecules or organelles in medium.
- It is done with an **electrical instrument** called **electroporator.**
- The electroporation unit consist of an **electroporator, two aluminium electro** and a **cuvette.**
- The electroporator creates electric current and sends it through the electrodes.
- It generates 50 to 240 volts through the electrodes.
- The alliminium electrodes conduct the electric pulse through the medium into the cuvette.
- The electrodes are fixed on the inner side of the cuvette.
- The cuvette is a small vessel of 10 to 50 ml capacity made of plastic or glass .
- The cell and DNA solution are mixed together and pipette into the cuvette.
- The electroporator is set up to produce a current of 50 to 240 volts per 1 to 10 minutes.

- The electric current induces some permeability to the cell membrane and hence the cells uptake the DNA from the medium.
- Electroporation is useful to transform bacteria, fungi, plant cells and animal cells

The rDNA and protoplast suspension are mixed together and kept in the cuvette. Then a single electric pulse of about 4 to 8 kilo volts/cm is generated between two electrodes for 4 to 5 milli seconds. The electric pulse forms transient pores in the cell membrane and makes it permeable to the rDNAs.

The protoplast uptakes one or more copies of rDNA from the suspension by **pinocytosis.** The protoplast can uptake rDNA of 3 to 140kbp size.

The frequency of uptake can be further increased by adding colcemid to the reaction mixture.

Limitation

1. Electroporation method cannot be adapted to introduce rDNAs into intact plants.
2. Animal cells are sensitive to electric treatment.
3. If the electric strength is too strong ,plant protoplast may loose their viability.

Ultrasonication

- The method of subjecting plant cells to sound waves at the frequency of 20 to 35.1KHz is called as ultrasonication.
- These ultrasonic waves cannot be audiable by human ears, but they do some changes on the surface of plant cells and cell membrane.
- Ultrasonic waves are generated by electronic devices called ultrasonicator.
- The ultrasonic method of gene delivery was discovered by wyber et al in 1999.
- Ultrasonication method is employed in the delivery of desired genes into seedlings, cotyledons, leaflets, immature embryos, callus and embryogenic cell suspensions.

- When a plant material is exposed to ultrasonic waves, the permeability of cell wall and plasma membrane is altered suitably as the cells uptake DNA or macromolecules.
- Plant cell transformation using ultrasonication is easy and simple to perform in ordinary laboratory. However, sonication could cause cell damage and rupture which are not likely during plant transformation.
- Hence, it is necessary to optimise the condition to facilitated uptake of DNA without damaging the cells.

Mechanism of DNA uptake

- **Y.Leu** et.al (2005) explained the mechanism of DNA uptake by cells during sonication.
- When the cells are sonicated, the ultrasonic waves increases the temperature of the medium and produces acoustic cavitation of the cell surface.
- This cavitation increases the permeability of the cell wall and cell membrane.
- Meantime, the rate of mass transfer also increases in the medium.
- This increased mass transfer activity forces the DNA or macromolecule towards the cavitation so that the DNA kit and penetrate the cell membrane to reach the cell.
- Inside the cell, sonication is stopped the cell starts to regain its original structure by resynthesis of cell wall and membrane.

Ultrasonicators

Many types of ultrasonicators are available for plant transformation.

Thay are of two types – **Bath sonicators and probe sonicators.**

Bath sonicators

- The bath sonicators are small automated cabinets equipped with ultra sound producing devices. They can produce ultra sounds with wavelength upto 55KHz.
- They are used to transform cotyledons, leaf tissues, shoot apex and meristem.

Probe sonicators

- The probe sonicator has one or two probes which produce ultrasonic waves.
- This type of wavelengths upto 20KHz.
- The probe sonicators are employed for the transformation of calli and cell suspensions.
- Probe of the sonicator is inserted in the tube containing plant material and the machine is switched on to prodcuce ultrasound.
- It needs less time of sonication than bath sonicator.

Vessels for Sonication

- Generally, borosilicate tubes (13x100nm) are preferred for the sonication of cotyledons and embryos in bath sonicators.
- Leaf tissues, meristem, apices and seedlings are kept in polypropylene tubes (50 ml) for sonication

Sonication

- The plant material to be transformed is taken in a borosilicate tube containing liquid nutrient medium.
- Then, probe of the ultrasonicator is placed on the surface of the medium at a fixed position.
- This altered permeability makes the cells to uptake DNA from the medium.

Duration of Sonication

- Ultrasonication method has been used to deliver foreign genes into plants such as tobacco, wheat, sugar beet, maize, rice, etc.
- Monocot, dicot and gymnosperm plants can be transformed with this method.
- The duration of sonication, however, varies depending on the plant specimen to be transformed in the experiment-
    - Plant cells/tissues are sonicated for 2 seconds to 5 minutes.
    - Cotyledons are sonicated for 1-300seconds.
    - Leaf tissues are sonicated for 10 to 60 seconds.

- Shoot tips, apices and meristem are sonicated for 5 to 300 seconds.
- Embryonic suspensions are sonicated for 5 to 120 seconds.
- Seedlings are sonicated for 10 to 100 seconds.

After this prescribed time point, the machine is switched off to stop sonication,

Transformation:

- The foreign gene is cloned in plasmid carrying genetic markers such as **cat, nptII** and **Gus** by using suitable restrication and DNA ligase.
- The rDNA may be added to the plant material in the suspension **before sonication or during sonication or after sonication.**
- The sonicated plant material intake the DNA from the solution to become **transformed** material in 1 hour.

Particle bombardment gun method:

- Shooting the plant or animal cells by DNA coated tungsten particle for getting DNAs into the cells, is called **particle bombardment** or **biolistics.**
- This method is also called microprojectile bombardment. It is a suitable method for introducing rDNA into plant cells, fungal cells, animal cells and cell organells such as chloroplast and mitochondria.
- The rDNA is mixed with **tungsten or gold particles** of 1 to 4 µm diameter and it is treated with **$CaCl_2$** or **spermidine** or **PEG.**
- **$CaCl_2$** precipitates the rDNA on to the metal particles.
- The DNA coated tungsten or gold particles are said to be **microprojectiles.**
- The microprojectiles are fired into the plant cells or animal cells with the speed of 300 to 600 m/s(meter/sec) using an instrument called **microprojectile gun or particle gun or shot gun.**
- Gun powder, helium or electric power is used to provide propelling force to the gun to guide the microprojectile into the cells.

- The microprojectiles penetrate into the cell wall and cell membrane and deliver the rDNA into the cells.
- The rDNAs get integerated with the cell DNA and hence recombinant cells are formed.
- Biolistics can be adopted to introduce rDNA into intact plant cells having cell walls.
- It is a useful method to trnansfer novel genes into immature embryos, embryonic calli, root sections and epidermal cells of many monocot and dicot plants.
- Biolistics is an effective method to transfer rDNA into mammalian cells and organs.
- The skin and ears of live forms are short with microprojectiles coated with rDNAs for getting the DNA into the cells.

Pollen transformation through particle bombardment:

- Delivery of rDNA into pollen grains using a particle bombardment gun is called **pollen transformation.**
- Pollen grains do not intake naked DNAs from the culture by means of transfection because of the presence of thick pollen wall.
- Again, gene transfer through callus transformation is time consuming and requires special tissue culture techniques and skilled persons.
- Therefore, scientists have developed pollen transformation through particle bombardment gun to transfer desired genes to plants.

In 1997, **Ramaiah** and **Skinner** developed transgenic alfalfa plant by pollen transformation through particle bombardment. This method involves the following steps:

1. Mature pollen grains are collected from the flowers of alfalfa plant (*Medicago sativa)* by teasing the stamens with a needle.
2. The pollen grains are kept in a test tube and the tube is kept on ice for 1 hour to reduce the metabolic activities of the pollens.
3. Then the pollen grains are kept suspended in 13% sucrose solution for 10 minutes. The sucrose solution for 10 minutes. The sucrose solution activates the pollen grains to start

germination and it makes the pollen wall permeable to microprojectiles during bombardment.

4. The desired DNA is cloned in **pBI 121 plasmid** using proper restriction enzyme and DNA ligase. The pBI121 has GUS reporter gene as a genetic marker useful for screening the recombinants.
5. The gene cloned PBI 121 is mixed with tungsten or gold particles 1-4µm diameter and $CaCl_2$ precipitates the rDNA onto the metal particles to form DNA coated particles called ***microprojectiles.***
6. The microprojectiles are loaded in a particle bombardment gun.
7. The pollen grains (in step 3) are place in a petridish and the microprojectiles are shoot into pollen by pressing the firing pin of the particle bombardment gun.
8. The suspension containing the pollen grains are taken in a test tube is kept on ice for 1 hr.
9. Then the suspension is centrifuged to remove the sucrose solution. The pellet of pollen thus obtained is again suspended in 10%sucrose solution for 30minutes.
10. After removing the sucrose solution, the pollen grains are taken on a fine brush and dusted on the tip of pistil of emasculated flower. The transformed pollens fertilize the eggs to form transformed seeds.
11. Seeds are obtained from the artificially pollinated flowers and sown in separate rows to grow into seedlings in a green house.
12. After 30days, the seedlings are tested for GUS activity to indicate transformants. To test GUS activity, protoplast isolated from each and every seedling and treated with 5-bromo 4-chloro-3-indolyl-β-D-glucuronic acid. Protoplasts of trasformants become blue coloured during incubation. This can be observed by viewing the protoplasts under a microscope.
13. The GUS positive seedlings are the genetically manipulated plants which give transgenic product.

Nearly, 30% of seedling raised from the seeds give rise to transgenic plants

while the others remain untransformed ones. The transformed plants retain GUS activity for ten generations and then it may decrease slowly. The reason for low GUS activity after ten generations is that the plasmid may be lost spontaneously during repeated cell divisions. So transgenic plants raised through pollen transformation can only be maintained for about 10 generations.

Agroinfection

- Transfer of foreign DNA to plant cells by Agrobacterium is called **Agroinfection**.
- *Agrobacterium* cannot pick up recombinant Ti plasmid from the nutrient medium.
- Therefore, the rDNA is introduced in to E.coli, it is transferred to *Agrobacterium* by three parental mating.
- The resulting recombinant *Agrobacteria* are co-cultured with plant cells or tissues for two days.
- During this time, the agrobacteria infects the plant cells and deliver their recombinant Ti plasmid in to the cells.
- The T-DNA integared with the cell DNA and hence recombinant plant cells are formed.
- After proper selection and screening , plants are regenerated from the recombinant cells.

Limitation

Argoinfection takes place only in dicot plants.

Liposome fusion

- Liposomes are small spherical vesilces made of phospholipids.
- The rDNA to be delivered into plant cell is entrapped in liposomes which are then allowed to fuse with protoplast of that plant cell using PEG.
- When the liposome fuses with the cell membrane of the plant proptoplast, the rDNA contained in it is delivered into the protoplasm of the cell.
- Several genes cloned in plasmid vectors have been transferred to protoplast of carrot, tobacco,petunia, tomato, etc. by using liposomes.

Gene transgferred through liposome fusion **has following advantages**

- Liposomes do not show any toxicity to protoplst.
- They protect the rDNA from the cellular nuclease enzyme.
- The rDNA can be stored in liposome for a long time.
- Liposomes can be used for different types of cells.
- Preparation of liposomes with rDNAs is easy.
- Risk potential is very low.

Gene transfer into animals

- The rDNA can be delivered into animal cells by
- transduction,
- electroporation,
- particle bombardment,
- retiroviral method
- microinjection.

These methods are discussed below :

Transduction

- The direct intake of naked DNA from the culture mediumby eukaryotic cells, is called **transfection.**
- It is very similar to transformation in bacteria.
- Transfection takes place in plant cells and animal cells invitro.
- Foreign DNA can be introduced directly into animal cells and organs by transfection. Example –*He La cells, leukaemia L1210 cells, liver cells, lymphocytes, oocytes of mammals and vertebrates.*
- The DNA is dissolved in **phosphate buffer. Then calcium chloride** solution is added.
- This leads to the formation of **calcium phosphate.**

- it combines with DNA to form **calcium phosphate DNA precipitate.**
- This precipitate is added to the cells. The precipitate particles adhere with the cell surface.
- The cell engulfs the particles with (DNA) by phagocytosis.
- The DNA, entering the cell, is integerated with the cells genome.

Liposome mediated gene transfer

- **The liposome** is a small spherical vesicle made of phosphor lipids. It is used to transfer genes to animal cells.
- It is formed when a lipid is agitated with water. It contains many concentric layers of phosphor lipids.
- In the liposome. Polar heads face outward and the non polar tail face to the centre. If there is more than one lipid layer, the adjacent layers are separated by water layer.
- The size of the liposomes varies from 25nm to a few microns in diameter.
- Ribosomes fuses with cell membrane and discharges its contents int the cell. Hence it is used as a gene transfer system.
- The rDNA, water and phosphatidyl choline are mixed together in a test tube and the tube is shaken well.
- During this process, the lipid bilayers develop around the rDNA present in water and form a liposome. In the liposome in the tube are added to a culture of animal cells.
- Liposomes fuse with cell membrane and discharge their contents into the cells.
- The rDNA present in the contents integrates with the cell DNA.
- However, the DNA may be destroyed in the cytoplasm before reaching the nuclease.
- The safe delivery of rDNA into the nuclease, the liposome is impregnated with some nuclear proteins and then another lipid layer is developed around it.
- The outer lipid layer fuses with the cell membrane and the rest is released into the cell.
- As the inner lipid layer has nucleoprotein,cellular enzyme do not attach it .

- The inner lipid layer fuses with nuclear membrane and discharges its contents into the nucleus.
- The frequency of liposome fusion can be increased by adding addition of polyethylene glycol(PEG) .
- Liposomes are also used to deliver certain drugs or rDNAs or enzymes in to cells of particular organs.

Advantages of liposome

- Any large DNA can be delivered into the cells using liposomes.
- Liposomes never disrupts the integrity of cells.
- They protect foreign DNAs from cellular nuclease enzymes.
- Liposomes with desired properties can be made by impregnating them with specific proteins. Liposomes never ineterfere with the human system.

Particle bombardment (Biolistics)

- *Shooting the plant or animal cells by DNA coated gold or tungsten particles for introducing DNA into the cells is called* ***particle bombardment or biolistics****.*
- This method is also called as *microprojectile bombardment.*
- By this method rDNA can be introduced in to plant cells, fngal cells, animal cells and cell organelles such as chloroplast and mitochondria.
- In this method ,no vector is introduced in to the rDNA in to the cells.The rDNA is literally shot in to cells.
- The instrument used to shoot the DNA in to cell are called **gene gun or microprojectile** gun
- The gene gun is doing the work of vector.

The gene gun consists of following components

  - Gas acceleration tube
  - Firing pin

- A blank charger
- Macro projectile
- Micro projectile
- Stopping screen
- Target cell

- All these components are enclosed in a vaccum chamber.
- The rDNA is mixed with tungsten or gold particle of microscopic size and it is treated with $CaCl_2$ or spermidine or PEG.
- $CaCl_2$ precipitate the rDNA on to the metal particles.
- The rDNA coated tungsten or gold particle said to be microprojectile.
- The microprojectiles are positioned on a holder about 5 cm above the cells, within the chamber of the gun.
- The chamber is sealed and helium gas is pumped into a small compartment situated above the gold particles.
- Once a predetermined pressure is reached, the gas is released and the gold particles are shot into the cells.
- The gold particle acts as **bullets**, penetrating the cells wall and delivering the rDNA into the cells, interior.
- The microprojectiles are fired in to the plant cells or animal cells with the speed of 300-600 m/s usimg the gene gun.
- **Gun powder, pressurized helium gas** or **electric power** is used to provide propelling force for the gun to drive the microprojectiles into the cells.
- The rDNA gets intergrated with the cell DNA and hence recombinant cells are formed.
- Biolistics can be adopted to introduce rDNA into intact plant cell having cell walls.
- It is a useful method to transfer novel ganes into immature embryos, embryonic calli, root section and epidermal cells of many monocot and dicot plants.
- Biolistics is an effective method to transfer rDNAs into mammalian cells and organs.

- The skin and ears of live mouse are shot with microprojectiles coated with rDNAs for getting the DNAs into the cells.

Virus vector method

- *Introduction of foreign DNA into eggs with the help of a retrovirus vector is rectroviral infection.*
- Small foreign genes, less than 8 kbp, can be transferred to receipt cells through retrovirus vector. Murine leukaemia virus(MuLV) is the most common retrovirus used to introduce genes into animal cells.
- The rDNA is constructed and introduced into a mouse cell line by using calcium phosphate mediated transfection.
- The cell line is then infected with a mutent MuLVwhich has no signal for invivo packaging.
- RNA of the helper virus produces vbiral capsid.
- The rDNA produces recombinant viral RNA. The latter gets package in the viral capsid to form infective recombinant MuLV particles
- Eggs of mouse are collected and fertilized invitro to produce zygotes. On reaching 8-celled stage ,the mouse embryos are infected eith the recombinant MuLV. Inside the infected cell the recombinant RNA is reverse transcriped into double stranded DNA. The DNA intagates with the cel DNA and hence recombinant cells are formed.thus , the virus delivers the rDNA into the animal cells.
- The mouse embryos are then transferred to a medium containing neomycin . The embryos has recombinants cells alone survive in the medium and the others die.

Non viral method or microinjection

- Microinjection refers to injection of DNAs or cell organelles directly in to cell using a injection needle.
- By this method DNAs, RNAs, proteins or cell organelles into animal cells, eggs, zygotes and plant protoplast.

- A Sterioscopic dissecting microscope, Micropipette, Injection is needed.
- A Fertilized egg is transferred on to microscopic slide under the microscope. The cell is held in position using a sucking pipette. One end of sucking pipette is positioned on the surface of cell and a gentle suction pressure is applied on its other end.
- The rDNA is sucked into the glass injection needle. It is gently inserted into the zygote by viewing through the microscope.
- The rDNA is delivered into the zygote and the needle is drawn back carefully.
- The rDNA gets integrated onto the needle of the zygote.
- The microinjected embryos are cultured *in-vitro* upto the morula or blastocyst stage and then implanted into a female to produce transgenic organism.
- The surviving embryos are transferred into the uterus of a surrogate mother. These embryos develop into *transgenic mice.*

Limitations:

1. rDNAs can be microinjected only intoa few cells at time.
2. Microinjection needs technical skill experienced workers.
3. It needs a micropanipulator.

Possible Questions:

1.Explain in detail about nucleic acid isolation procedure.

2.Give a detailed account on DNA modifying enzymes.

3.Explain in detail about Nuclease enzymes and its types

4.Give a detailed account on ligase enzyme and its mechanism.

5.Explain in detail about DNApolymerase enzyme and its types.

6.Give a detailed note on DNA manipulating enzymes.

7.Explain in detail about reverse transcriptase enzyme.

8.What are topoisomerases? Explain its role in cloning.

9.Elaborately explain about cloning process.

10. What is cloning? Give a detailed note on steps involved in cloning.

Unit II

SYLLABUS

Probes – radio labeled and non radio-labeled, guessmers and degenerate probes. Sequence dependent and independent screening, southern, northern hybridization, colony and plaque hybridization, *in situ* chromosomal hybridization, chromosome walking, Genome mapping, DNA fingerprinting, Polymerase chain reaction (PCR), RT- (Reverse transcription) PCR.

Molecular Probes

A stretch of DNA or RNA sequence that can detect a target sequence in the genome is called as probe. Ever since their development, they are constantly being modified to enhance their utility and to bring about automation in the process of genome analysis.

Although, initially these probes were developed and used for genetic engineering research but are now frequently used for a variety of purposes including –

- Diagnosis of infectious diseases
- Identification of food contaminants
- Variety of microbiological tests
- Forensic tests and
- To identify different varieties of crop species
- In molecular biology laboratories these are frequently used for identification and isolation of genes or related sequences.

Probe Sources

Any nucleic acid can be used as a probe provided it can be labeled to permit identification and quantitation of the hybrid molecules formed between the probe and sequence to be identified.

In practice, **double and single standard DNAs, mRNAs, and other RNAs synthesized in vitro** are all used as probes.

DNA/ RNA probe assays are **faster and sensitive** so that many conventional diagnostic tests for viruses and bacteria involving culturing of the organisms are being fast replaced by molecular probe assays. While culture tests can take days or even months, molecular probe assays can be performed with in few hours or minutes.

Molecular probes can be broadly categorized into DNA probes and RNA probes, sometimes cDNA probes and synthetic oligonucleotide probes can also be used for various purposes.

Preparation of Probes

Different types of probes can be prepared in various ways.

DNA Probes

Extract the DNA from an animal or plant tissue. Digest extracted DNA with a restriction enzyme such as EcoRI or Hind III which cuts DNA at specific sites or positions where a specific sequences recognized by the enzyme is found. Run the digested DNA on an agarose or polyacrylmide gel electrophoresis to separate fragments of different sizes. Isolate DNA of specific fragment from a particular band identified through southern blots by hybridization with specific labeled mRNA or cDNA molecules. Clone this DNA in a vector. Allow chimeric vector to infect bacteria for multiplication where it can make billions of copies.

Uses

DNA probes prepared in this manner can be used for southern blotting and RFLP analysis.

Limitations

Probe-probe hybridization

RNA Probes

High specific activity RNA probes or riboprobes may also be synthesized from DNA templates cloned in expression vectors such as SP6 (which infects Salmonella typhimurium) and T7 phage (infects E.coli). This is achieved through RNA synthesized in vitro and labeled simultaneously with labeled nucleotides. Usually SP6 and T7 systems can be and have been utilized to express whole RNAs, but for making a probe, only a short labeled RNA is sufficient. To enable such probes to be transcribed into uniform lengths, it is practice to linearize the plasmid by cleaving it with a

restriction enzyme. The vector thus carries the input in the following order phage promoter-Enzyme 1- Enzyme 2. The template DNA is inserted at the number 1 site and composite is treated with restriction enzyme 2. So now this creates a linear DNA with a promoter, the template DNA and automatic termination site at the end cleaved with enzyme 2. The mRNAs fall of when they reach this end of the vector. Such templates reflect to run of templates and are uniform in size and are easier to isolate from the reaction mixture. RNA probes prepared in this manner can be used for northern blotting and in situ hybridization.

Advantages

RNA probes offer several advantages over DNA probes. Since these are single stranded and provide improved signal or hybridization blots.

There is lack of competition of probe/ probe hybridization.

Limitation

Even though, some advantages are there wide spread presence of ribonuclease creates some problems in their preparation and use. So RNA probes are more sensitive to degradation than equivalent DNA probes, therefore extreme care must be taken in the preparation of RNA probes by keeping all glass ware free of ribonuclease.

cDNA Probes

A DNA sequence corresponding to a part of a specific gene can be obtained by reverse transcription of mRNA. cDNA thus obtained can be cloned and used as a probe.

Synthetic oligonucleotides as probes

DNA probes with known nucleotide sequence can also be synthesized chemically using automated DNA synthesizers. These synthetic probes will be efficient only when they are not more than 20- 40 nucleotides in length.

Labeling of probes

The detection of homologous sequences after hybridization with the probe is like finding a needle in the hay-stock. Therefore, for the success of DNA probe assay it is necessary to develop simple, safe and sensitive techniques for their use. As probes transmit no signal of their own they have to be

either labeled with radioactive isotopes or coupling of non-radioactive signal molecules to the probes without impairing the hybridization ability of these probes.

Signal moleculules

The signal molecules are fluorescent antibodies, enzymes that produce color changes in dyes and chemiluminescent catalysts.

Methods for Labeling of Probes

There are two methods for labeling of probes. i.e.

- End labeling and
- Nick translation.

End labeling

- In this technique probe is isolated and end labeled by removing the 5'-terminal phosphate using alkaline phosphatase first and adding a 32P-labeled phosphate with the help of a kinase.
- End labeled probes are far less labeled than the transcribed ones with labels at several nucleotides in the strand.

Nick Translation

- It is one of the commonly used techniques for producing a radioactive probe. A purified phage or plasmid vector containing a cloned genomic or cDNA sequence is treated with a small amount of pancreatic DNase which hydrolyzes the phosphodiester bonds between nucleotides.
- At very low concentration the DNase produces only scattered –nicks‖ in one or other strand of the duplex DNA. DNA polymerase and radioactively labeled deoxynucleotides are also added to the DNA sample.
- Using the unharmed strand as template, the DNA polymerase synthesizes a new second strand using exposed 3' end at a nick site as primer, which then displaces the existing DNA from the 5‘ end of the nick. Radioactive nucleotides are incorporated into the new strand, so, a single standard probe is created when the duplex DNA is denatured.

Choice of Label

Probes can be labeled either by radioactive isotopes or can also be labeled with nonradioactive molecules such as biotin, digoxegenin etc.

Radiolabeled Probes

- The most sensitive detection method employed is radioactive label. Traditionally radioactively labeled probes are used for a variety of experiments. The autoradiographic detection of labeled probes depends on the isotope used and the specific activity which has to be high enough to permit detection after hybridization within a reasonable exposure of time or with a good signal.
- Several isotopes are available for radioactive labeling. The use of 32P allows rapid detection of signal yet cellular localization is suboptimal because of the long path of the [32P] β-rays.
- Autoradiographs of high quality and improved cellular localization employ 35S labeled probes. 35S emits γ-rays of much shorter path length. One of the major disadvantages of 35 S is non-specific binding of label to cell or tissues. It has been suggested that pre-hybridization in the presence of non-labeled thio alpha UTP at pH 5.5 reduces this non-specific binding. It is also possible to label nucleic acids with tritium or by iodonization with [125 I] iodine. Tritium labeled probes give intracellular localization, but for low abundance target molecules may require exposure of up to 100 days. This may result in high background and is unacceptably long for most purposes. Disadvantage with radiolabel is Instability of label.

Non-Radiolabeled Probes

a) Biotin Labeled Probes

Recent advances in nucleic acid technology now offer alternatives to radio activity labeled probes. One of such procedure that is becoming popular is biotin labeling of nucleic acids. This system exploits the affinity which the glycoprotein avidin has for biotin. Avidin is commonly found in egg white. Biotinylated probes are prepared through a nick-translation reaction by replacing nucleotides with biotinylated derivatives. After hybridization and washing, detection

of hybrids is done by a series of cytochemical reactions which finally give a blue color whose intensity is proportional to the amount of biotin in the hybrid.

Advantages

- These assays employ non-toxic materials, whose half-life is longer.
- These probes can be prepared in advance in bulk and stored at -20°C for repeated uses.
- Detection of hybrids is much faster than by radioactive probes.

Limitations

- A limitation of this technology is that very small probe contains only a small number of biotinylated sites limiting the intensity of signal obtained. It has been solved by adding long _tails' of biotinylated nucleotides to the probes through enzymatic methods. Sometimes the probe does not need to be labeled with biotin but only coupled with a tail
- Another disadvantage of biotin labeled probes is that cytochemical visualization reactions lead to precipitation of insoluble material which cannot be removed and therefore the filter cannot be reused, whereas with radio labeled probes, the filters can be used repeatedly for hybridization with a number of probes one at a time.

b) Digoxigenin

Labeled Probes Digoxigenin is another chemical derived from plants and used for non-radioactive labeling of probes. An antibody associated with an enzyme (antidigoxigenin - alkaline phosphatase conjugate) is used for the detection of the presence of digoxigenin. The probes may be labeled with digoxigenin - II - dUTP supplied with a digoxigenin kit (these kits are available from any commercial firm, eg. BoehringerMannheem). The labeled and denatured probe may be used for hybridization. After hybridization the membrane or the slide may be transferred into detection buffer containing 20 ug/ ml of antidigoxigenin fluorescein and 5% (W/V) BSA (bonine serum albumin). This is incubated for 1 h at 37°C and then the membrane or slide is washed in detection buffer three times (8 min each at 37°C) and alkaline phosphatase

activity was detected using 0.17 mg/ml BCIP (5- bromo 4-chloro 3-indoyl phosphate) and 0.33 mg/ ml NBL (nitro blue tetrazolium) as dye substrate.

c) Alternatives to Biotin and Digoxigenin Labeling

- The techniques of non-radio isotopic labeling have been further expanded and new methods have been devised for attaching other ligands (eg: hapten determinants, 2, 4 dinitrophenol, arsenative derivatives etc) to nucleotides without hampering their ability to be incorporated into DNA.
- These alternatives require binding of attached ligands to specific proteins that can be tagged with enzymes or fluorescent molecules. It is possible to monitor many probes simultaneously by using several different ligands since each ligand would yield a different signal.
- A chemiluminscent probe system has also been developed in which two different probes complementary to a continuous segment of DNA hybridize to adjacent segments of a gene.
- The first label is a chemiluminescent complex that emits light at a specific wavelength this emission excites the label molecule on the second probe to emit light at a different wave length which can be detected using a photomultiplier device: This process called non-radioactive energy transfer can occur only if the two probes hybridize correctly and the two labels are close to each other.
- This system has great fidelity and provides basic technology for a homogeneous assay. In addition DNA does not need to be immobilized and no washing steps are necessary which may be an additional advantage for large scale testing.

Applications of Molecular Probes

Molecular probes are used in restriction fragment length polymorphisms (RFLPs) and related Analysis

a) RFLPs for Evolutionary Studies

The restriction fragment length polymorphisms (RFLPs) can be studied in a set of related species using a random or a specific DNA probe. The similarities and differences can be

used to infer phylogenetic relationships. This has actually been done in a number of cases both in plantsand in animals.

b) RFLP Maps and Linkage of RFLPs with Specific Genes

RFLPs have been used to prepare chromosome maps in humans, mice, fruit fly and in plants including maize, tomato, lettuce, and rice. The use of Mendelian markers for genetic mapping is sometimes limited due to non-availability of mutants. The list of markers can be increased or extended by using molecular markers which are examined in the form of RFLPs. Once a large number of RFLPs are available in a species the parents, F1 , F2 generations can be used to study their inheritance and linkage relationship and genetic linkage maps can be prepared. In this way they can be used for plant and animal breeding (Mburu and Hanotte et al., 2005; Guimaraes et al., 2007; Parasnis et al., 1999; Sen et al., 1997; Pujar et al., 1999; Sant et al., 1999)

c) RFLP Markers to Map the Genes in Diseased Persons and Identification of Disease

DNA polymorphisms (many forms) are differences in DNA sequence that result from point mutations, random deletions or insertions or the presence of varying number of repeated copies of a DNA fragments (tandem repeats). A polymorphism in the coding region of a gene may be detected as an alteration in the amino acid sequence of the encoded protein. It is now possible to detect polymorphisms in unexpressed regions of DNA by this analysis. Genetic disorders like sickle cell anemia, Thalassemia‘s, Huntington‘s disease and cystic fibrosis were identified through RFLP mapping analysis which demonstrates power of RFLP/Linkage analysis.

Use of Molecular Probes in Molecular Cytogenetics

a) Isolation of Genes Using DNA or RNA Probes

Specific molecular probes can be used for isolation of specific genes. These probes may be available either from same species or from another species can be used for isolation of genes. If probes obtained from one species used for isolation of gene from the same

species they are called as homologous probes. If probes obtained from another species used for isolation of genes in other species they are defined as heterologous probes.

These heterologous probes have been found to be effective in identifying gene clones during colony hybridization or plaque hybridization or on southern blots. For instance, the gene for chalcone-synthase (CHS) has been isolated from Antirrhinum majus and Petunia hybrida using heterologous cDNA probes from parsley similarly heterologous Antirrhinum cDNA probe was used for isolation of CHS gene from barley and heterologous probes from maize were used for isolation of barley genes Wx (Waxy Genes) and Al (aleurone gene). Heterologous probes should ordinarily be used with cDNA library and not with the genomic library since in the latter case unrelated genes or pseudogenes (which do not express) may be isolated and cloned. These heterologous clones if available in expression vector [Eg. PGEM Blue] can also be used for getting RNA probes which have been bound to be more sensitive and efficient.

b) In situ Hybridization

In situ hybridization (ISH) is a technique which permits detection of DNA or RNA sequences in cell smears, tissue sections and metaphase chromosome spreads. The method is based on the formation of double stranded hybrid molecules which form between a DNA or RNA target sequence and the complementary single standard labeled probe. In a number of cases rye chromosomes in wheat background have been identified using this technique.

Satellite chromosomes with NORs (nuclear organizing regions) are recognized using probes for ribosomal DNA in wheat, barley etc. Telomeres have been identified in human and other eukaryotes using a telomeric sequence as a molecular probe showing that telomeres of all chromosomes carry the same sequence. In wheat and related species chromosomes of D genome can be identified using a D genome specific probe (PASI separated from Aegilops squarrosa).

This method is particularly useful if target sequences are distributed in a non-random way in tissues for the visualization of heterogenicity and the study of cell

differentiation. The method of ISH for RNA-RNA hybridization is generally applicable to rapid screening of small numbers of cultured cells for expression of oncogenic mRNA. It circumvents the problem of extracting rare RNAs in sufficient amounts for detection. Similar methods have been used to visualize homeotic gene expression in developing larvae. DNA-DNA ISH is also suitable for the detection of viral genomes in sections of routinely processed archival paraffin blocks of human tissues.

Screening and expression of rDNA

Identification of the Desired Clone

Identification of the bacterial colony containing the desired DNA fragments from among those making up the library.

Screening strategies

The identification of a specific clone from a DNA library can be carried out by

- exploiting the sequence of the clone
- the structure/function of its expressed product.

Exploiting the sequence of the clone

- applies to any type of library, genomic or cDNA, and
- can involve either nucleic acid hybridization or the PCR.

Sequence-dependent screening

methods are

1. Screening by hybridization
2. Chromosome walking
3. Screening by PCR

Screening by hybridization

Nucleic acid hybridization is the most commonly used method of library screening because

- it israpid,
- it can be applied to very large numbers of clones and, in the case of cDNA libraries,
- can be used to identify clones that are not full-length (and therefore cannot be expressed).

Grunstein and Hogness (1975) developed a screening procedure

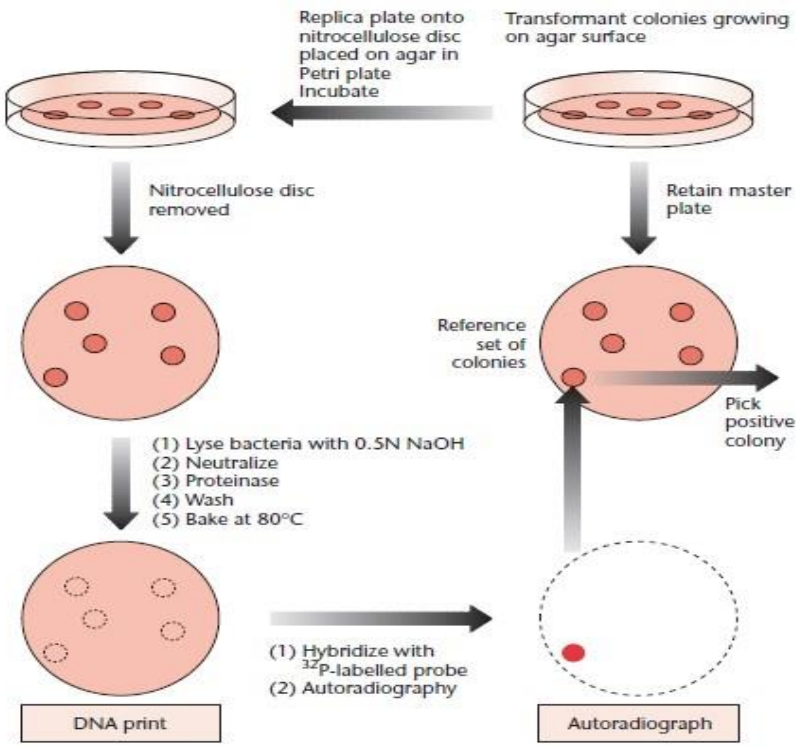
- to detect DNA sequences in transformed colonies by hybridization *in situ* with radioactive RNA probes.
- Their procedure can rapidly determine which colony among thousands contains the target sequence.

1. The colonies to be screened are first replica-plated on to a nitrocellulose filter disc that has been placed on the surface of an agar plate prior to inoculation (Fig).
2. A reference set of these colonies on the master plate is retained.
3. The filter bearing the colonies is removed and treated with alkali so that the bacterial colonies

are lysed and the DNA they contain is denatured.

4. The filter is then treated with proteinase K to remove protein and leave denatured DNA bound to the nitrocellulose, for which it has a high affinity, in the form of a _DNA print‘ of the colonies.
5. The DNA is fixed firmly by baking the filter at 80°C.
6. The defining, labelled RNA is hybridized to this DNA and the result of this hybridization is monitored by autoradiography.
7. A colony whose DNA print gives a positive autoradiographic result can then be picked from the reference plate.

Grunstein–Hogness method for detection of recombinant clones by colony hybridization



Benton and Davis (1977) devised a method called ***plaque lift***, in which

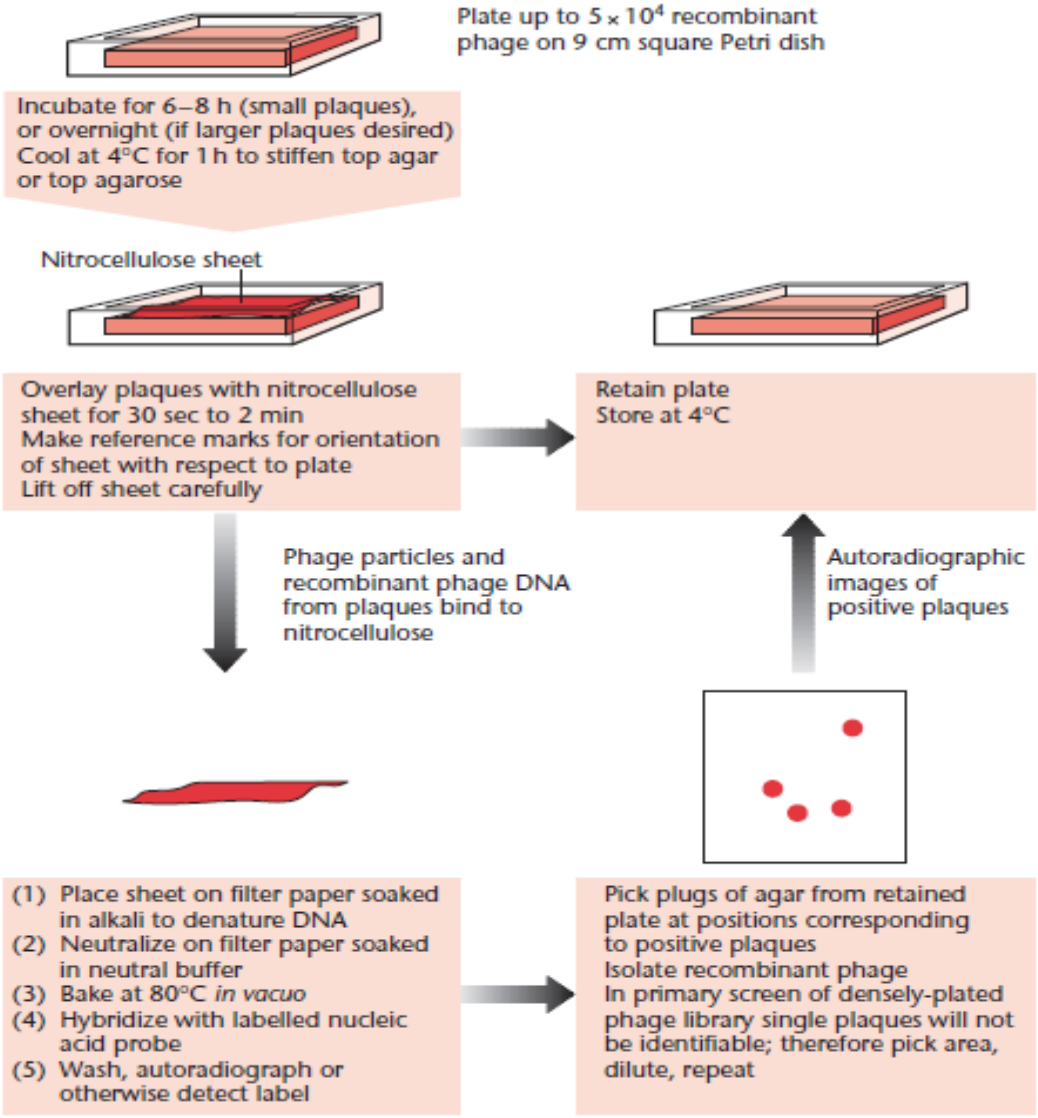
1. the nitrocellulose filter is applied to the upper surface of agar plates, making direct contact between plaques and filter.
2. The plaques contain phage particles, as well as a considerable amount of unpackaged recombinant DNA.
3. Both phage and unpackaged DNA bind to the filter and can be denatured, fixed and hybridized.

Advantages

- several identical DNA prints can easily be made from a single-phage plate:

- this allows the screening to be performed in duplicate,
- increased reliability,
- allows a single set of recombinants to be screened with two or more probes.

The Benton and Davis (1977) procedure is the most widely applied method of library screening, successfully applied in thousands of laboratories to the isolation of recombinant phage by nucleic acid hybridization.



In place of RNA probes, DNA or synthetic oligonucleotide probes can be used.

A number of alternative labelling methods are also available that avoid the use of radioactivity.

These methods involve the incorporation of chemical labels into the probe, such as digoxigenin or biotin, which can be detected with a specific antibody or the ligand streptavidin, respectively.

Hybridization probes & Labelling

DNA and RNA probes

- In molecular biology, a **hybridization probe** is a fragment of DNA or RNA of variable length (usually 100-1000 bases long), which is used in DNA or RNA samples to detect the presence of nucleotide sequences (the DNA target) that are complementary to the sequence in the probe.
- The probe thereby hybridizes to single-stranded nucleic acid (DNA or RNA) whose base sequence allows probe-target base pairing due to complementarity between the probe and target.
- The labeled probe is first denatured (by heating or under alkaline conditions such as exposure to sodium hydroxide) into single stranded DNA (ssDNA) and then hybridized to the target ssDNA (Southern blotting) or RNA (northern blotting) immobilized on a membrane or in situ.
- To detect hybridization of the probe to its target sequence, the probe is tagged (or labelled) with a molecular marker of either radioactive or (more recently) fluorescent molecules; commonly used markers are ^{32}P (a radioactive isotope of phosphorus incorporated into the phosphodiester bond in the probe DNA) or Digoxigenin, which is non-radioactive antibody-based marker.
- DNA sequences or RNA transcripts that have moderate to high sequence similarity to the probe are then detected by visualizing the hybridized probe via autoradiography or other

imaging techniques. Normally, either X-ray pictures are taken of the filter, or the filter is placed under UV light.

- Detection of sequences with moderate or high similarity depends on how stringent the hybridization conditions were applied — high stringency, such as high hybridization temperature and low salt in hybridization buffers, permits only hybridization between nucleic acid sequences that are highly similar, whereas low stringency, such as lower temperature and high salt, allows hybridization when the sequences are less similar.

Hybridization probes used in DNA microarrays refer to DNA covalently attached to an inert surface, such as coated glass slides or gene chips, and to which a mobile cDNA target is hybridized.

Depending on the method the probe may be synthesized using phosphoramidite method or generated and labeled by PCR amplification or cloning (older methods).

In order to increase the in vivo stability of the probe RNA is not used, instead RNA analogues may be used, in particular morpholino. Molecular DNA- or RNA-based probes are now routinely used in screening gene libraries, detecting nucleotide sequences with blotting methods, and in other gene technologies like microarrays.

Probe design

A great advantage of hybridization for library screening is that it is extremely versatile.

In hybridization is **very stringent conditions** can be used

- so that only sequences identical to the probe are identified.
- This is necessary, for example, to identify genomic clones corresponding to a specific cDNA or to identify overlapping clones in a chromosome walk.

Alternatively, **less stringent conditions** can be used

- to identify both identical and related sequences.

- This is appropriate where a probe from one species is being used to isolate a homologous clone from another species.

Probes corresponding to a conserved functional domain of a gene may also cross-hybridize with several different clones in the same species at lower stringency, and this can be used to identify members of a gene family.

Hybridization thus has the potential to isolate any sequence from any library *if a probe is available*.

If a suitable DNA or RNA probe cannot be obtained from an existing cloned DNA,

- an alternative strategy is to make an oligonucleotide probe by chemical synthesis.
- This requires some knowledge of the amino acid sequence of the protein encoded by the target clone.

Chromosome walking

- The principle of chromosome walking is that overlapping clones will hybridize to each other, allowing them to be assembled into a contiguous sequence.
- This can be used to isolate genes whose function is unknown but whose genetic location is known, a technique known as ***positional cloning***.

To begin a chromosome walk, it is necessary to have in hand a genomic clone that is known to lie very close to the suspected location of the target gene.

For example - In humans,

- this could be a restriction fragment length polymorphism that has been genetically mapped to the same region.
- This clone is then used to screen a genomic library by hybridization, which should reveal any overlapping clones.
- These overlapping clones are then isolated, labeled and used in a second round of screening to identify further overlapping clones,

- and the process is repeated to build up a contiguous map.

Screening by PCR

The PCR is widely used to isolate specific DNA sequences from uncloned genomic DNA or cDNA, but it also a useful technique for library screening.

As a screening method,

- PCR has the same versatility as hybridization, and the same limitations.
- It is possible to identify any clone by PCR but only if there is sufficient information about its sequence to make suitable primers.

To isolate a specific clone, PCR is carried out with gene-specific primers that flank a unique sequence in the target.

- Instead of plating the library out on agar, as would be necessary for screening by hybridization, pools of clones are maintained in multiwell plates.
- Each well is screened by PCR and positive wells are identified.
- The clones in each positive well are then diluted into a series in a secondary set of plates and
- screened again.
- The process is repeated until wells carrying homogeneous clones corresponding to the gene of interest have been identified.

There are also several applications where the use of ***dege nerate primers*** is favourable.

- A degenerate primer is a mixture of primers, all of similar sequence but with variations at one or more positions.
- This is analogous to the use of degenerate oligonucleotides as hybridization probes, and the primers are synthesized in the same way.
- A common circumstance requiring the use of degenerate primers is when the primer sequences have to be deduced from amino acid sequences (Lee *et al.* 1988).

- Degenerate primers may also be employed to search for novel members of a known family of genes or to search for homologous genes between species.

Exploitation of structural or functional property of the clone

Screening expression libraries - (expression cloning)

If a DNA library is established using expression vectors, each individual clone can be expressed to yield a polypeptide.

Expression libraries are useful because they allow a range of alternative techniques to be employed, each of which **exploits some structural or functional property of the gene product**.

This can be important in cases where

- the DNA sequence of the target clone is completely unknown and
- there is no strategy available to design a suitable probe or set of primers.

Methods of screening are –

1. ***Immunological screening***
2. ***South-western and north-western screening***
3. ***Screening with alternative ligands***

Immunological screening

Immunological screening involves the use of antibodies that specifically recognize antigenic determinants on the polypeptide synthesized by a target clone.

- This is one of the most versatile expression cloning strategies, because it can be applied to any protein for which an antibody is available.
- The molecular target for recognition is generally an *epitope*, a short sequence of amino acids that folds into a particular three-dimensional conformation on the surface of the protein.

- Epitopes can fold independently of the rest of the protein and therefore often form even when the polypeptide chain is incomplete or when expressed as a fusion with another protein.
- Importantly, many epitopes can form under denaturing conditions, when the overall conformation of the protein is abnormal.

The first immunological screening techniques were developed in the late 1970s, when expression libraries were generally constructed using plasmid vectors.

- The method of **Broome and Gilbert (1978)** was widely used at the time.

This method exploited the fact that antibodies adsorb very strongly to certain types of plastic, such as polyvinyl, and that IgG antibodies can be readily labelled with ^{125}I by iodination *in vitro*.

1. As usual, transformed cells were plated out on Petri dishes and allowed to form colonies.
2. In order to release the antigen from positive clones, the colonies were lysed, e.g. using chloroform vapour or by spraying with an aerosol of virulent phage (a replica plate is required because this procedure kills the bacteria).
3. A sheet of polyvinyl that had been coated with the appropriate antibody was then applied to the surface of the plate, allowing antigen–antibody complexes to form.
4. The sheet was then removed and exposed to 125I-labelled IgG specific to a *different* determinant on the surface of the antigen (i.e. a determinant not involved in the initial binding of the antigen to the antibody-coated sheet (Fig. 6.1 4)).
5. The sheet was then washed and exposed to X-ray film.
6. The clones identified by this procedure could then be isolated from the replica plate.
7. _sandwich' technique is applicable only where two antibodies recognizing different determinants of the same protein are available.
8. However, if the protein is expressed as a fusion, antibodies that bind to each component of the fusion can be used, efficiently selecting for recombinant molecules.

- it is much more convenient to use bacteriophage-λ insertion vectors, because these have a higher capacity and the efficiency of *in vitro* packaging allows large numbers of recombinants to be prepared and screened.
- Immunological screening with phage-λ cDNA libraries was introduced by **Young and Davies** (1983) using the expression vector λgt11, which generates fusion proteins with β-galactosidase under the control of the *lac* promoter.

In the original technique,

screening was carried out using colonies of induced lysogenic bacteria, which required the production of replica plates, as above.

A simplification of the method is possible by directly screening plaques of recombinant phage.

In this procedure (Fig),

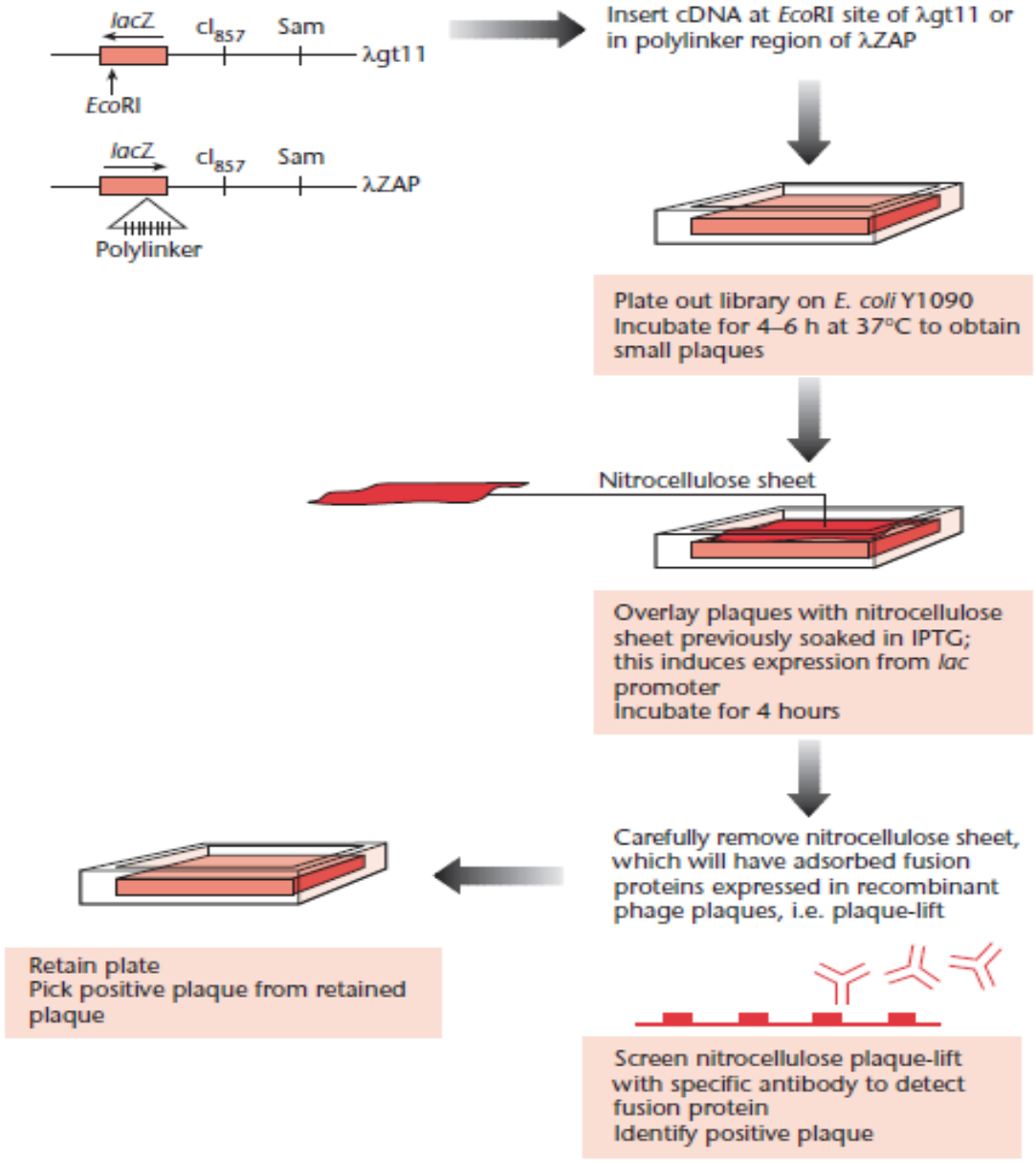
1. the library is plated out at moderately high density (up to 5 x 10^4 plaques/ cm2 plate), with *E. coli* strain Y1090 as the host.
2. This *E. coli* strain overproduces the *lac* repressor and ensures that no expression of cloned sequences (which may be deleterious to the host) takes place until the inducer isopropyl-β-ヒ-thiogalactoside (IPTG) is presented to the infected cells.
3. Y1090 is also deficient in the *lon* protease, hence increasing the stability of recombinant fusion proteins.
4. Fusion proteins expressed in plaques are absorbed on to a nitrocellulose membrane overlay, and this membrane is processed for antibody screening.
5. When a positive signal is identified on the membrane, the positive plaque can be picked from
6. the original agar plate (a replica is not necessary) and the recombinant phage can be isolated.

The original detection method using **iodinated antibodies** has been replaced by more convenient methods using non-isotopic labels, which are also more sensitive and have a lower background of nonspecific signal.

- Generally, these involve the use of unlabelled primary antibodies directed against the polypeptide of interest, which are in turn recognized by secondary antibodies carrying an enzymatic label.
- As well as eliminating the need for isotopes, such methods also incorporate an amplification step, since two or more secondary antibodies bind to the primary antibody.

Typically, the secondary antibody recognizes the species-specific constant region of the primary antibody and is conjugated to either horseradish peroxidase or alkaline phosphatase, each of which can in turn be detected using a simple colorimetric assay carried out directly on the nitrocellulose filter.

- Polyclonal antibodies, which recognize many different epitopes, provide a very sensitive probe for immunological screening, although they may also cross-react to proteins in the expression host.



- Monoclonal antibodies and cloned antibody fragments can also be used, although the sensitivity of such reagents is reduced because only a single epitope is recognized.

South-western and north-western screening

A closely related approach to immunochemical screening has been used for the screening and isolation of clones expressing sequence specific DNA-binding proteins.

Procedure

1. a plaque lift is carried out to transfer a print of the library on to nitrocellulose membranes.
2. the screening is carried out, without using an antibody, by incubating the membranes with a radiolabelled *double stranded* DNA oligonucleotide probe, containing the recognition sequence for the target DNA-binding protein.
3. This technique is called ***south-western* screening**, because it combines the principles of Southern and western blots.
4. It has been particularly successful in the isolation of clones expressing cDNA sequences corresponding to certain mammalian transcription factors.

Limitation of this technique is that,

- since individual plaques contain only single cDNA clones, transcription factors that function only in the form of heterodimers or as part of a multimeric complex do not recognize the DNA probe and the corresponding cDNAs cannot be isolated.
- the procedure can also be successful only in cases where the transcription factor remains functional when expressed as a fusion polypeptide.

More recently, a similar technique has been used to isolate sequence specific RNA-binding proteins, in this case using a single-stranded RNA probe.

By analogy to the above, this is termed ***north-western screening*** and has been successful in a number of cases

Screening with alternative ligands

- As well as DNA and RNA, a whole range of alternative _ligands' can be used to identify polypeptides that specifically bind certain molecules (for example, as an alternative to south-western screening).
- Such techniques are not widely used because they generally have a low sensitivity and their success depends on the preservation of the appropriate interacting domain of the protein when exposed on the surface of a nitrocellulose filter.
- the yeast two-hybrid system and its derivatives provide versatile assay formats for many specific types of protein–protein interaction, with the advantage that such interactions are tested in living cells, so the proteins involved are more likely to retain their functional interacting domains.

Basic techniques in screening recombinat DNA

Agarose gel electrophoresis

- The progress of the first experiments on cutting and joining of DNA molecules was monitored by velocity sedimentation in sucrose gradients. However, this has been entirely replaced by gel electrophoresis.
- Gel electrophoresis is not only used as an analytical method, it is routinely used preparatively for the purification of specific DNA fragments.
- The gel is composed of polyacrylamide or agarose.
  - Agarose is convenient for separating DNA fragments ranging in size from a few hundred base pairs to about 20 kb.
  - Polyacrylamide is preferred for smaller DNA fragments.
- The migration of the DNA molecules through the pores of the matrix must play an important role in molecular-weight separations since the electrophoretic mobility of DNA in free solution is independent of molecular weight.

- An agarose gel is a complex network of polymeric molecules whose average pore size depends on the buffer composition and the type and concentration of agarose used.
- DNA movement through the gel was originally thought to resemble the motion of a snake (reptation).
- DNA molecules display elastic behaviour by stretching in the direction of the applied field and then contracting into dense balls.
- The larger the pore size of the gel, the greater the ball of DNA which can pass through and hence the larger the molecules which can be separated.
- Once the globular volume of the DNA molecule exceeds the pore size, the DNA molecule can only pass through by reptation.
- This occurs with molecules about 20 kb in size and it is difficult to separate molecules larger than this without recourse to pulsed electrical fields.
- The bands of DNA in the gel are stained with the intercalating dye ethidium bromide and as little as 0.05 μg of DNA in one band can be detected as visible fluorescence when the gel is illuminated with ultraviolet light.
- In addition to resolving DNA fragments of different lengths, gel electrophoresis can be used to separate different molecular configurations of a DNA molecule.
- Gel electrophoresis can also be used for investigating protein–nucleic acid interactions in the so-called *gel retardation* or *band shift* assay.
- It is based on the observation that binding of a protein to DNA fragments usually leads to a reduction in electrophoretic mobility.
- The assay typically involves the addition of protein to linear double-stranded DNA fragments, separation of complex and naked DNA by gel electrophoresis and visualization.
- Marker DNA fragments of known sizes are run in a separate lane; this permits an accurate determination of the size of an unknown DNA molecule by interpolation. The gels are stained with the intercalcating dye ethidium bromide which, gives visible fluorescence on illumination of the gel with UV light; as little as 0.05 μg of DNA in one band can be detected by using this dye.

3.2. Nucleic acid blotting

Nucleic acid labelling and hybridization on membranes have formed the basis for a range of experimental techniques central to recent advances in our understanding of the organization and expression of the genetic material.

These techniques may be applied in the isolation and quantification of specific nucleic acid sequences and in the study of their organization, intracellular localization, expression and regulation.

A variety of specific applications includes the diagnosis of infectious and inherited disease.

Blotting describes the immobilization of sample nucleic acids on to a solid support, generally nylon or nitrocellulose membranes.

The blotted nucleic acids are then used as _targets' in subsequent hybridization experiments. The main blotting procedures are:

3.2.1. Southern Hybridization

The name of this technique is derived from the following:
(1) the name of its inventor, E.M. Southern, and

(2) the DNA-DNA hybridization that forms its basis. It is also called Southern blotting since the procedure for transfer of DNA from the gel to the nitrocellulose filter resembles blotting.

1. This technique has since been extended to the analysis of RNA (northern blotting) and proteins (western blotting); these names are only jargon terms, i.e., reverse of Southern being northern and so on, and do not reflect any functional or historical significance.
2. This approach is useful when few DNA fragments with considerable length differences are to be separated and studied. This approach also separates the closed circular (supercoiled), nicked (relaxed) and linear configurations of a single DNA molecule.

3. In many situations, it is critical to detect and identify DNA fragments in a sample that are complementary to a given DNA sequence, e.g., to demonstrate the presence of the gene in question in transgenics, to detect and stuty RFLP (restriction fragment length polymorphism), etc.

In Southern hybridization, a sample of DNA containing fragments of different sizes is subjected to electrophoresis using either polyacrylamide or agarose gel.

The DNA sample may either be subjected to mechanical shearing or to restriction endonuclease digestion in order to generate the fragments.

This is achieved by Southern hybridization in which the following steps are performed.

1. The restriction fragments of DNA present in agarose gel (after electrophoresis) are denatured into single-stranded form by alkali treatment.

2. They are then transferred onto a nitrocellulose filter membrane; this is done by placing the gel on top of a buffer saturated filter paper, then laying the nitrocellulose filter membrane on the top of gel, and finally placing some dry filter papers on top of this membrane. The buffer moves, due to capillary action, from the bottom filter paper through the gel carrying with it the denatured DNA present in the gel; the DNA becomes trapped in the nitrocellulose membrane as the buffer phases through it. This process is known as blotting and takes several hours to complete. The relative positions of the bands on the membrane remain the same as those in the gel and there is a minimal loss in their resolution (sharpness).

3. The nitrocellulose membrane is now removed from the blotting stack, and the DNA is permanently immobilized on the membrane by baking it at 80°C in vacuo.

4. Single-stranded DNA has a high affinity for nitrocellulose filter membrane. (Note that RNA lacks this affinity). Therefore, the baked membrane is treated, with a solution containing 0.2% each of Ficoll (an artificial polymer of sucrose), polyvinylpyrrolidon and

bovine serum albumin; this mixture is often supplemented with an irrelevant nucleic acid, e.g., t zRNA (pretreatment). This treatment prevents nonspecific binding of the radioactive probe (to be used in the next step) probably by attaching macromolecules to all the free binding sites on the membrane. Often the above mixture is included in the hybridization reaction itself.

5. The pretreated membrane is placed in a solution of radioactive, single-stranded DNA or an oligodeoxynucleotide (a DNA segment having few to several nucleotides) called probe. The name probe signifies the fact that this DNA molecule is used to detect and identify the DNA fragment in the gel/membrane that is complementary to the probe. The conditions during this step are chosen so that the probe hybridizes with the complementary DNA on the membrane to the greatest extent with a low nonspecific binding on the membrane; this step is known as hybridization reaction.

Usually, the initial hybridization reaction is carried out under conditions of relatively low stringency of hybridization to permit a high rate of hybridization; this is followed by a series of post hybridization washes of increasing stringency, i.e., higher temperature or, more commonly, lower ionic strength, with a view to eliminate the pairing of radioactive probe to related sequences and to allow only perfectly complementary pairing.

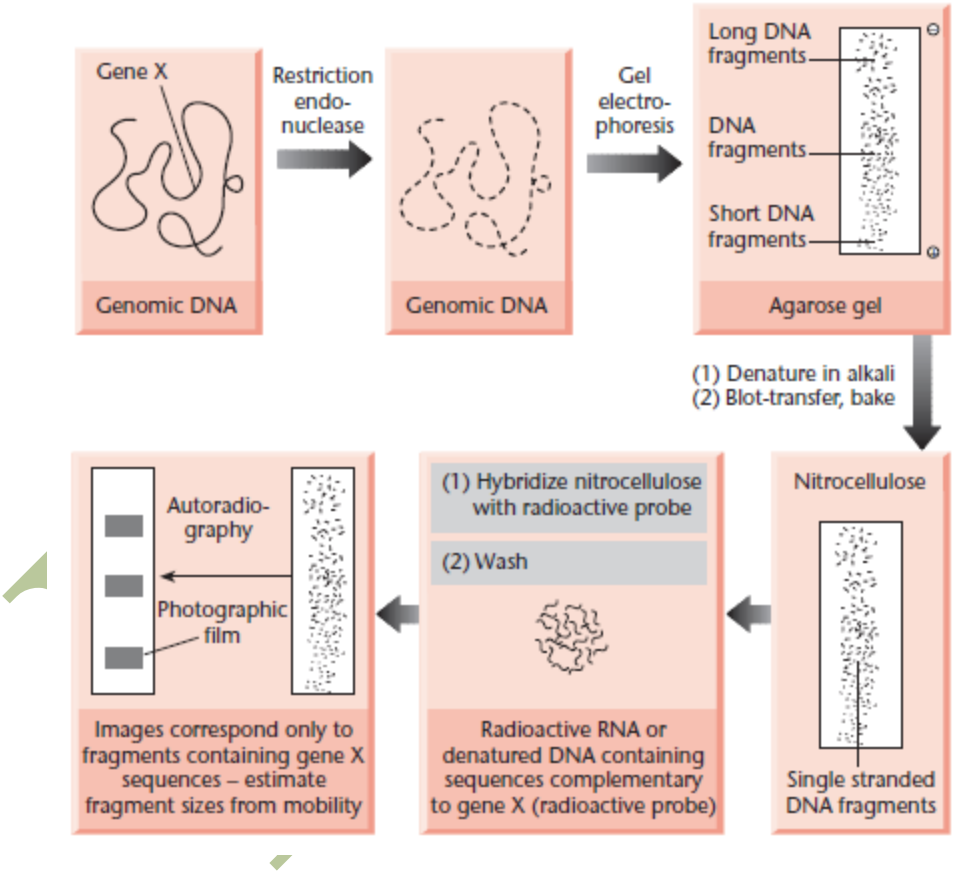
6. After the hybridization reaction, the membrane is washed to remove the unbound probes.

7. The membrane is now placed in close contact with an X-ray film and incubated for a desired period to allow images due to the radioactive probes to be formed on the film. The film is then developed to reveal distinct band(s) indicating positions in the gel of the DNA fragments that are complementary to the radioactive probe used in the study.

It should be kept in mind that electrophoresis of sheared or restricted DNA produces a smear in which the fragments are distributed in a continuum according to their size, and there are no distinct bands. The distinct bands are produced by the hybridization reaction of the selected

probe with one or few fragment sequences present in the gel.

The Southern blotting technique is extremely sensitive. It can be used to map the restriction sites around a single copy gene sequence in any genome (even of man). It is used for DNA fingerprinting, preparation of RFLP maps, detection and identification of the transferred genes in transgenic individuals, etc.



Recently some new membrane materials, e.g., nylon membranes, have been developed which have the following advantageous features:

(1) They are physically robust in comparison to nitrocellulose filter membranes,

(2) both DNA and RNA become cross-linked to them by a brief exposure to UV light, which

(3) saves the time needed for baking in vacuo in the case of nitrocellulose membranes, and

(4) the same membrane blot, e.g., a membrane onto which DNA/RNA has been transferred from a gel and cross-linked by UV exposure, can be used for search with more than one probe after removing the earlier probe by high temperature washing or some other denaturing procedure; in other words, the nylon membranes are reusable.

Northern Hybridization - In this technique,

RNAs are separated by gel electrophoresis, the RNA bands are transferred onto a suitable membrane, e.g., diazobenzyloxymethyl (DBM) paper or nylon membranes, and immobilized; the bands are hybridized with radioactive single stranded DNA probes, and the bands showing hybridization are detected by autoradiography.

Clearly, northern blotting/hybridization is simply an extension of the Southern blotting technique. The two techniques, however, show the following differences:

(1) in Southern hybridization, DNA is separated by gel electrophoresis, while in northern blotting RNAs are separated;

(2) as a result, in Southern hybridization DNA has to be denatured before blotting, while this step is not needed in northern hybridization;

(3) nitrocellulose membrane is generally not used for northern, while it is often used for Southern hybridization; and finally

(4) hybridization with the probe produces DNA: DNA hybrid molecules in Southern hybridization but RNA : DNA molecules in northern hybridization.

Initially, specially prepared paper (diazobenzyloxymethyl, DBM, paper prepared by diazotization of aminobenzyloxymethyl paper) was used for northern blotting since RNA did not bind to nitrocellulose membrane. RNA becomes covalently bound to DBM paper due to which these blot transfers are reusable.

DBM is also equally effective in binding to denatured DNA, and is more efficient than nitrocellulose in binding to small DNA fragments. Recently developed nylon membranes have superceded the use of DBM paper as they are robust, reusable and bind (by cross linking) to RNA on a brief exposure to UV light.

Northern hybridization is useful in the identification and separation of the RNA that is complementary to a specific DNA probe; this is a sensitive test for the detection of transcription of a DNA sequence that is used as probe.

Western Blotting - In western blotting,

proteins are electrophoresed in polyacrylamide gel, transferred onto a nitrocellulose or nylon membrane (to which they bind strongly), and the protein bands are detected by their specific interaction with antibodies, lectins or some other compounds. The various steps of this technique are briefly described below.

1. Protein bands are separated by polyacrylamide gel electrophoresis.
2. The protein bands are transferred onto a nitrocellulose or nylon membrane; initially this was achieved by a capillary movement of buffer similar to Southern blotting (capillary blotting), but nowadays it is usually done by electrophoresis (electrophoretic blotting).

Electrophoresis has been applied for the blotting step in Southern and northern hybridizations as well; in such 'cases buffer of low ionic strength (to avoid overheating during electrophoresis) and nylon membranes (since nucleic acids bind to nitrocellulose membrane only under conditions of

high ionic strength) are used. The electrophoretic blotting, both of proteins and nucleic acids is much faster and more efficient than capillary blotting.

3. The specific protein bands are identified in a variety of ways.

(i) Antibodies are the most commonly used as probes for detecting specific antigens.

(ii) Lectins are used as probes for the identification of glycoproteins.

- These probes may themselves be radioactive or a radioactive molecule may be tagged to them. Often the identification process is based on a 'sandwich' reaction.
- In such an approach, a species specific second antibody or protein A of Staphylococcus aureus (protein A binds to certain subclasses of IgG antibodies) or streptavidin (it binds to biotinylated antibodies) is used to bind to the antibodies bound to the protein bands.
- These second molecules may be labelled with radioactive, enzyme or fluorescent tags; a single preparation of these labelled molecules can be employed as a general detector for various probes.

PCR (Polymerase Chain Reaction)

- first proposed in the early **1970**s by **Ghobind Khorana** and his colleagues

- did not seem practicable at a time when genes had not yet been sequenced, synthesis of oligonucleotide primers and thermostable DNA polymerases had not been described.

After 15 years, **Kary Mullis and coworkers** at Cetus Corporation,

- Performed in vitro amplification of single-copy mammalian genes using the Klenow fragment of *Escherichia coli* DNA polymerase I.

- The use of a thermostable polymerase from *TlIermus aquaticus*

greatly increased the efficiency of PCR and opened the door to automation of the method.

By the end of the 1980s,

- cloning was no longer the only way to isolate genes:

- DNA sequencing had been revolutionized

- and **PCR had become a fundamental step of genetic and molecular analyses**.

Essential Components of Polymerase Chain Reactions

PCRs contain **seven** essential components:

Thermostable DNA polymerase

Pair of synthetic oligonucleotides

Deoxynucleoside triphosphates (dNTPs).

Divalent cations

Buffer

Monovalent cations

Template DNA

A thermostable DNA polymerase

To catalyze template~dependent synthesis of DNA.

A wide choice of enzymes is now available that vary in their fidelity, effIciency, and ability to synthesize large DNA products

For routine PCRs, Ta*q* polymerase (0.5-2.5 units per standard 25-50 µl reaction) remains the enzyme of choice.

The specific activity of most commercial preparations of *Taq* is ~ 80,000 units/mg of protein.

Standard PCRs contain 2 x 10^{12} to 10 X 10^{12} molecules of enzyme.

efficiency of primer extension with *Taq* polymerase is generally ~ 0. 7.

A pair of synthetic oligonucleotides

- **to prime DNA synthesis**.
- Careful design of primers is required to obtain the desired products in high yield, to suppress amplification of unwanted sequences, and to facilitate subsequent manipulation of the amplified product.
- Standard reactions contain non limiting amounts of primers, typically 0.1-0.5 µM of each primer (6 x 10^{12} to 3 X 10^{13} molecules).
- This quantity is enough for at least 30 cycles of amplification of a 1 -kb segment of DNA.
- Higher concentrations of primers favor mispriming, which may lead to nonspecific amplification.
- Oligonucleotide primers synthesized on an automated DNA synthesizer can generally be used in standard PCRs without further purification.

Deoxynucleoside triphosphates (dNTPs)

- Standard PCRs contain equimolar amounts of dATP, dTTP, dCTP, and dGTP.
- Concentrations of 200-250 f.1M of each dNTP are recommended for *Taq* polymerase in reactions containing 1.5 mM MgCl2.
- In a 50-Ill reaction, these amounts should allow synthesis of ~6-6.5Ilg of DNA, which should be suftlcient even for multiplex reactions in which eight or more primer pairs are used at the same time.
- High concentrations of dNTPs (>4 mM) are inhibitory, perhaps because of sequestering of Mg2+.

- However, a satisfactory amount of amplified product can be produced with dNTP concentrations as low as 20 11M 0.5-La pmole of an amplified fragment ·-1 kb in length,

Divalent cations.

- All thermostable DNA polymerases require free divalent cations – usually Mg^{2+} - for activity.
- Calcium ions are quite ineffective). Because dNTPs and oligonucleotides bind Mg2+, the molar concentration of the cation must exceed the molar concentration of phosphate groups contributed by dNTPs plus primers.
- It is therefore impossible to recommend a concentration of Mg2+ that is optimal in all circumstances.
- Although a concentration of 1.5 mM Mg2- is routinely used, increasing the concentration of Mg2+ to 4.5 mM or 6 mM has been reported to decrease nonspecific priming in some cases and to increase it in others.
- The optimal concentration of Mg2+ must therefore be determined empirically for each combination of primers and template.

Buffer to maintain pH

- Tris-Cl, adjusted to a pH between !U and 8.8 at room temperature, is included in standard PCRs at a concentration of 10 mM.
- Then incubated at 72°C (the temperature commonly used for the extension phase of PCR), the pH of the reaction mixture drops by more than a full unit, producing a buffer whose pH is -7.2.

Monovalent cations

- Standard PCR buffer contains 50 mM KCl and works well for amplification of segments Qf DNA >500 bp in length.
- Raising the KCI concentration to -70-100 Mm often improves the yield of shorter DNA segments.

Template DNA

- Template DNA containing target sequences can be added to PCR in single- or double-stranded form.
- Closed circular DNA templates are amplified slightly less efficiently than linear DNAs. Although the size of the template DNA is not critical, amplification of sequences embedded in high-molecular-weight DNA (> 10 kb) can be improved by digesting the template with a restriction enzyme that does not cleave within the target sequence.
- In the case of mammalian genomic DNA, up to 1.0 Ilg of DNA is utilized per reaction, an amount that contains ~3 x 105 copies of a single-copy autosomal gene.
- The typical amounts of yeast, bacterial, and plasmid DNAs used per reaction are 10 ng, 1 ng, and 1 pg, respectively.

Thermostable DNA Polymerases

- Thermostable DNA polymerases are isolated from two classes of organisms:
- the thermophilic and hyperthermophilic eubacteria Archaebacteria,
- *T. aquaticus*, an organism from the thermophilic Archaea family *Taq (T. aquaticus)* DNA polymerase, the first isolated and best understood of the thermostable DNA polymerases, remains the workhorse of PCR in most laboratories.

Programming Polymerase Chain Reactions

PCR is an iterative process, consisting of three elements:

- denaturation of the template by heat,
- annealing of the oligonucleotide primers to the single-stranded target seqllence(s), and
- extension of the annealed primers by a thermostable DNA polymerase.

Denaturation

- Double-stranded DNA templates denature at a temperature that is determined in part by their G+C content.

- The higher the proportion of G+C, the higher the temperature required to separate the strands of template DNA.
- The longer the DNA molecules, the greater the time required at the chosen denaturation temperature to separate the two strands completely.
- If the temperature for denaturation is too low or if the time is too short, only AT-rich regions of the template DNA will be denatured.
- When the temperature is reduced later in the PCR cycle, the template DNA will reanneal into a fully native condition.
- In PCRs catalyzed by *Taq* DNA polymerase, denaturation is carried out at 94-95°C, which is the highest temperature that the enzyme can endure for 30 or more cycles without sustaining excessive damage.
- In the first cycle of PCR, denaturation is sometimes carried out for 5 minutes to increase the probability that long molecules of template DNA are fully denatured.
- We recommend denaturation for 45 seconds at 94-95°C for routine amplification of linear DNA templateswhose contenls of G+C is 55% or less.
- Higher temperatures may be required to denature template and/or target DNAs that are rich in G+C (>55%).

Annealing of primers to template DNA

- The temperature used for the annealing step *(T*) is critical.
- If the annealing temperature is too high, the oligonucleotide primers anneal poorly, if at all, to the template and the yield of amplified DNA is very low.
- If the annealing temperature is too low, nonspecific annealing of primers may occur, resulting in the amplification of unwanted segments of DNA.
- Annealing is usually carried out 3-5°C lower than the calculated melting temperature at which the oligonucleotide primers dissociate from their templates.

- Many formulas exist to determine the theoretical melting temperature, but none of them are accurate for oligonucleotide primers of all lengths and sequences.
- It is best to optimize the annealing conditions by performing a series of trial PCRs at temperatures ranging from 20°C to 100°C below the lower of the melting temperatures calculated for the two oligonucleotide primers.
- Alternatively, the thermal cycler can be programmed to use progressively lower annealing temperatures in consecutive pairs or cycles ("touchdown" PCR).

Extension of oligonucleotide primers

- It is carried out at or near the optimal temperature for DNA synthesis catalyzed by the thermostable polymerase, which in the case of *Taq* DNA polymerase is 72-78°C.
- In the first two cycles, extension from one primer proceeds beyond the sequence complementary to the binding site of the other primer.
- In the next cycle, the first molecules are produced whose length is equal to the segment of DNA delimited by the binding sites of the primers.
- From the third cycle onward, this segment of DNA is amplified geometrically, whereas longer amplification products accumulate arithmetically.
- The polymerization rate of *Taq* polymerase is - 2000 nucleotides/minute at the optimal temperature (72-78^{o}C) and as a rule of thumb, extension is carried out for 1 minute for every 1000 bp of product.
- For the last cycle of PCR, many investigators use an extension time that is three times longer than in the previous cycles, to allow completion of all amplified products.

Number of cycles

- the number of cycles required for amplification depends on the number of copies of template DNA present at the beginning of the reaction and the efficiency or primer extension and amplification.

- Once established in the geometric phase, the reaction proceeds until one of the components becomes limiting.
- At this point, the yield of specific amplification products should be maximal, whereas nonspecific amplification products should be barely detectable, if at all.
- This is generally the case after -30 cycles in PCRs containing _ 10^5 copies of the target sequence and *Taq* DNA polymerase (efficiency ~O. 7).
- At least 25 cycles are required to achieve acceptable levels of amplification of single-copy target sequences in mammalian DNA templates.

Applications of PCR

1. Medical applications
2. Infectious disease applications
3. Forensic applications
4. Research applications
5. Other applications

Medical applications

PCR has been applied to a large number of medical procedures:

- The first application of PCR was for **genetic testing**, where a sample of DNA is analyzed for the presence of genetic disease mutations.
- Prospective parents can be tested for being genetic carriers, or their children might be tested for actually being affected by a disease.
- DNA samples for Prenatal testing can be obtained by amniocentesis, chorionic villus sampling, or even by the analysis of rare fetal cells circulating in the mother'sbloodstream.
- PCR analysis is also essential to Preimplantation genetic diagnosis, where individual cells of a developing embryo are tested for mutations.

- PCR can also be used as part of a sensitive test for **tissue typing**, vital to organ transplantation. As of 2008, there is even a proposal to replace the traditional antibody-based tests for blood type with PCR-based tests.
- Many forms of cancer involve alterations to **oncogenes**.

By using PCR-based tests to study these mutations, therapy regimens can sometimes be individually customized to a patient.

Infectious disease applications

Characterization and detection of infectious disease organisms have been revolutionized by PCR:

- The **Human Immunodeficiency Virus** (or **HIV**), responsible for **AIDS**, is a difficult target to find and eradicate.

The earliest tests for infection relied on the presence of antibodies to the virus circulating in the bloodstream.

- However, antibodies don't appear until many weeks after infection, maternal antibodies mask the infection of a newborn, and therapeutic agents to fight the infection don't affect the antibodies.
- PCR tests have been developed that can detect as little as one viral genome among the DNA of over 50,000 host cells.
- Infections can be detected earlier, donated blood can be screened directly for the virus, newborns can be immediately tested for infection, and the effects of antiviral treatments can be quantified.

- Some disease organisms, such as that for **Tuberculosis**, are difficult to sample from patients and slow to be grown in the laboratory. PCR-based tests have allowed detection of small numbers of disease organisms (both live or dead), in convenient samples.

Detailed genetic analysis can also be used to detect antibiotic resistance, allowing immediate and effective therapy.

The effects of therapy can also be immediately evaluated.

- The spread of a **disease organism** through populations of domestic or wild animals can be monitored by PCR testing.

In many cases, the appearance of new virulent sub-types can be detected and monitored. The sub-types of an organism that were responsible for earlier epidemics can also be determined by PCR analysis.

Forensic applications

The development of PCR-based genetic (or DNA) fingerprinting protocols has seen widespread application in forensics:

- In its most discriminating form, **Genetic fingerprinting** can uniquely discriminate any one person from the entire population of the world.

Minute samples of DNA can be isolated from a crime scene, and compared to that from suspects, or from a DNA database of earlier evidence or convicts.

Simpler versions of these tests are often used to rapidly rule out suspects during a criminal investigation.

Evidence from decades-old crimes can be tested, confirming or exonerating the people originally convicted.

- Less discriminating forms of DNA fingerprinting can help in **Parental testing**, where an individual is matched with their close relatives.

DNA from unidentified human remains can be tested, and compared with that from possible parents, siblings, or children.

Similar testing can be used to confirm the biological parents of an adopted (or kidnapped) child. The actual biological father of a newborn can also be confirmed (or ruled out).

Research applications

PCR has been applied to many areas of research in molecular genetics:

- PCR allows rapid production of short pieces of DNA, even when nothing more than the sequence of the two primers is known.

This ability of PCR augments many methods, such as generating **hybridization probes** for Southern or northern blot hybridization.

PCR supplies these techniques with large amounts of pure DNA, sometimes as a single strand, enabling analysis even from very small amounts of starting material.

- The task of **DNA sequencing** can also be assisted by PCR. Known segments of DNA can easily be produced from a patient with a genetic disease mutation.

Modifications to the amplification technique can extract segments from a completely unknown genome, or can generate just a single strand of an area of interest.

- PCR has numerous applications to the more traditional process of **DNA cloning**.

It can extract segments for insertion into a vector from a larger genome, which may be only available in small quantities.

Using a single set of 'vector primers', it can also analyze or extract fragments that have already been inserted into vectors.

Some alterations to the PCR protocol can **generate mutations** (general or site-directed) of an inserted fragment.

- **Sequence-tagged sites** is a process where PCR is used as an indicator that a particular segment of a genome is present in a particular clone.

The Human Genome Project found this application vital to mapping the cosmid clones they were sequencing, and to coordinating the results from different laboratories.

- An exciting application of PCR is the phylogenic analysis of DNA from **ancient sources**, such as that found in the recovered bones of Neanderthals, or from frozen tissues of Mammoths.

In some cases the highly degraded DNA from these sources might be reassembled during the early stages of amplification.

- A common application of PCR is the study of patterns of **gene expression**.

Tissues (or even individual cells) can be analyzed at different stages to see which genes have become active, or which have been switched off.

This application can also use Q-PCR to quantitate the actual levels of expression

- The ability of PCR to simultaneously amplify several loci from individual sperm has greatly enhanced the more traditional task of **genetic mapping** by studying chromosomal crossovers after meiosis.

Rare crossover events between very close loci have been directly observed by analyzing thousands of individual sperms.

Similarly, unusual deletions, insertions, translocations, or inversions can be analyzed, all without having to wait (or pay for) the long and laborious processes of fertilization, embryogenesis, etc.

Others applications

PCR is also important in answering basic scientific questions.

- In the field of evolutionary biology, PCR has been used to establish relationships among species.
- In anthropology, it has used to understand ancient human migration patterns.
- In archaeology, it has been used to help identify ancient human remains.

- Paleontologists have used PCR to amplify DNA from extinct insects preserved in amber for 20 million years.
- The Human Genome Project, which had a goal of determining the sequence of the 3 billion base pairs in the human genome, relied heavily on PCR.
- The genes responsible for a variety of human diseases have been identified using PCR.
- For example, a PCR technique called multiplex PCR identifies a mutation in a gene in boys suffering from Duchenne muscular dystrophy.
- PCR can also be used to search for DNA from foreign organisms such as viruses or bacteria.

DNA sequencing

Methods of sequencing:

Two different methods are now routinely used for determination of DNA sequences

1. Chemical degradation method
2. Chain termination method

Maxam and Gilbert's Chemical Degradation Method

- developed in the late 1970‘s, by Allan Maxam and Walter Gilbert
- first method to determine the sequence of a DNA molecule of up to 500 bp.
- based on chemical modification of DNA and subsequent cleavage at specific bases.

In this method, following steps are involved:

(i) Label the 3'ends of DNA with 32p.

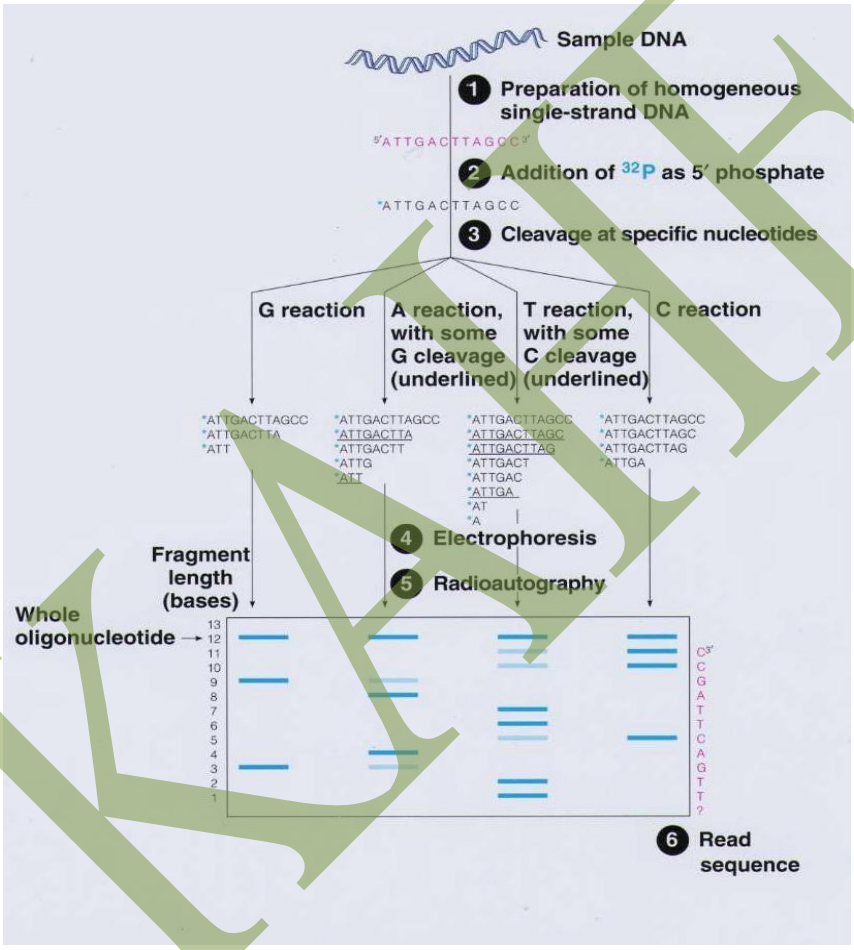
(ii) Separate two strands, both labelled at 3'ends.

(iii) Divide the mixture in four samples, each treated with a different reagent having the property of destroying either only G, or only C, or 'A and G' or 'T and C. The concentration of reagent is so adjusted that 50% of target base is destroyed, so that fragments of different sizes having 32p are produced.

(i) Electrophoreses each of the four samples in four different lanes of the gel.

(ii) Autoradiograph the gel and determine the sequence from positions of bands in four lanes.

vi) Read the sequence based on autoradiography.



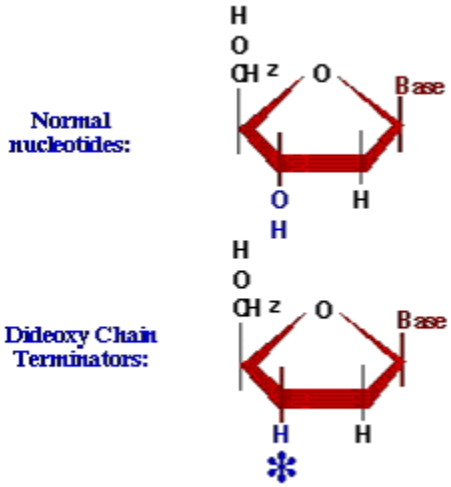
Disadvantages of this method:

This technology has the disadvantage of relying on toxic chemicals.

Chain termination method

- Developed by Fred Sanger
- Also called as Dideoxynucleotide Synthetic Method
- utilizing single stranded DNA as template for DNA synthesis, in which 2',3' dideoxynucleotides were incorporated leading to termination of DNA synthesis.
- These dideoxynucleotides are used as triphosphate and can be incorporated in a growing chain, but these dideoxynucleotides terminate synthesis, since they can not form a phosphodiester bond with next incoming deoxynucleotide triphosphate (dNTP).

Dideoxynucleotides:



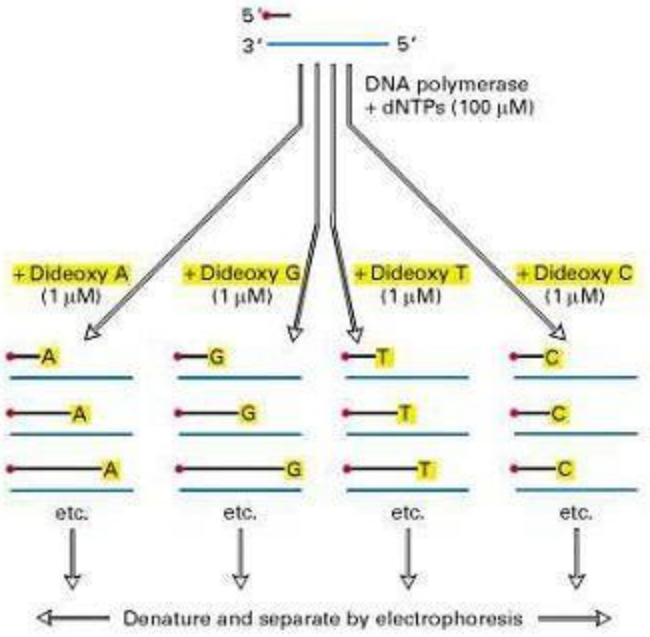
Following steps are involved in Sanger's dideoxy method for DNA sequencing:

(i) Four reaction tubes are set up, each containing single stranded DNA sample to be sequenced, all four dNTPs (radioactively labelled) and an enzyme for DNA synthesis (DNA polymerase I).

Each tube also contains a small amount (much smaller amount relative to four dNTPs) of one of the four ddNTP, so that four tubes have each a different ddNTP, bringing about termination at a specific base adenine (A), cytosine (C), guanine (G) and thymine (T).

ii) The fragments, generated by random incorporation of ddNTP leading to termination of reaction, are then separated by electrophoresis on a high resolution Polyacrylamide gel.

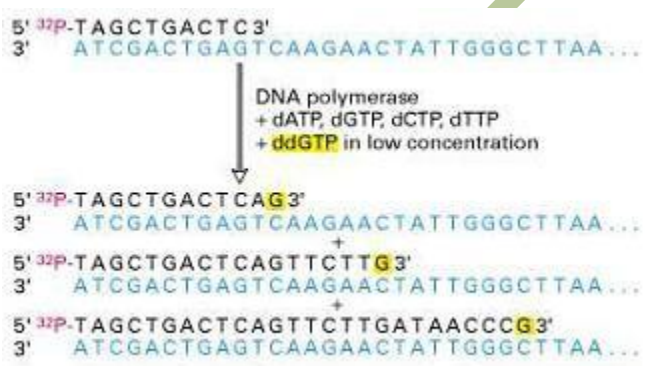
This is done for all the four reaction mixtures on adjoining lanes in the gel.



(iii) The gel is used for autoradiography so that the position of different bands in each lane can be visualized.

(iv) The bands on autoradiogram can be used for getting the DNA sequence.

Example



Now, variations of the above dideoxy method was developed as automatic sequencers.

In this automatic sequencer approach,

- a different fluorescent dye is tagged to the oligonucleotide primer in each of the four reaction tubes (blue for: A, red for C, etc.).
- The four reaction mixtures are pooled and electrophoreses together in a single Polyacrylamide of the tube.
- A high sensitivity fluorescence detector, placed near the bottom of the tube, measures the amount, of each fluorophore as a function of time.
- The sequence is determined from the temporal order of peaks corresponding to four different dyes.

The chain termination method of DNA sequencing can only be used for short strands (100 to 1000 base pairs), longer sequences must be subdivided into smaller fragments, and subsequently re-assembled to give the overall sequence.

Cloning and production of commercially important proteins

Cloning and production of proteins important to medicine and to industry –

particularly producing therapeutically important proteins for the treatment of numerous diseases e.g. cancer, allergies, autoimmune disease, neurological disorders, heart attacks, blood disorders, infections and wounds as well as the treatment of genetic diseases (**Gene therapy**).

Real Time PCR

Introduction to Real Time PCR

- As the name suggests, real time PCR is a technique used to monitor the progress of a PCR reaction in real time. At the same time, a relatively small amount of PCR product (DNA, cDNA or RNA) can be quantified. Real Time PCR is based on the detection of the fluorescence produced by a reporter molecule which increases, as the reaction proceeds.
- This occurs due to the accumulation of the PCR product with each cycle of amplification. These fluorescent reporter molecules include dyes that bind to the double-stranded DNA (i.e. SYBR® Green) or sequence specific probes (i.e. Molecular Beacons or TaqMan® Probes).
- Real time PCR facilitates the monitoring of the reaction as it progresses. One can start with minimal amounts of nucleic acid and quantify the end product accurately.
- Moreover, there is no need for the post PCR processing which saves the resources and the time. These advantages of the fluorescence based real time PCR technique have completely revolutionized the approach to PCR-based quantification of DNA and RNA. Real time PCR

assays are now easy to perform, have high sensitivity, more specificity, and provide scope for automation. Real time PCR is also referred to as real time RT PCR which has the additional cycle of reverse transcription that leads to formation of a DNA molecule from a RNA molecule. This is done because RNA is less stable as compared to DNA

Real Time PCR procedure

In a real time PCR protocol, a fluorescent reporter molecule is used to monitor the PCR as it progresses. The fluorescence emitted by the reporter molecule manifolds as the PCR product accumulates with each cycle of amplification. Based on the molecule used for the detection, the real time PCR techniques can be categorically placed under two heads:

1. Non-specific Detection using DNA Binding Dyes

2. Specific Detection Target Specific Probes

Non-specific Detection using DNA Binding Dyes

- In real time PCR, DNA binding dyes are used as fluorescent reporters to monitor the real time PCR reaction. The fluorescence of the reporter dye increases as the product accumulates with each successive cycle of amplification.
- By recording the amount of fluorescence emission at each cycle, it is possible to monitor the PCR reaction during exponential phase. If a graph is drawn between the log of the starting amount of template and the corresponding increase the fluorescence of the reporter dye fluorescence during real time PCR, a linear relationship is observed.

SYBR® Green is the most widely used double-strand DNA-specific dye reported for real time PCR. SYBR® Green binds to the minor groove of the DNA double helix. In the solution , the unbound dye exhibits very little fluorescence. This fluorescence is substantially enhanced when the dye is bound to double stranded DNA. SYBR® Green remains stable under PCR conditions and the

optical filter of the thermocycler can be affixed to harmonize the excitation and emission wavelengths. Ethidium bromide can also be used for detection but its carcinogenic nature renders its use restrictive.

Although these double-stranded DNA-binding dyes provide the simplest and cheapest option for real time PCR, the principal drawback to intercalation based detection of PCR product accumulation is that both specific and nonspecific products generate signal.

Specific Detection using Target Specific Probes

Specific detection of real time PCR is done with some oligonucleotide probes labeled with both a reporter fluorescent dye and a quencher dye. Probes based on different chemistries are available for real time detection, these include:

a. Molecular Beacons

b. TaqMan® Probes

c. FRET Hybridization Probes

d. Scorpion® Primers

Real Time PCR Applications Include

1 . Quantitative mRNA expression studies.

2 . DNA copy number measurements in genomic or viral DNAs.

3 . Allelic discrimination assays or SNP genotyping.

4 . Verification of microarray results.

5 . Drug therapy efficacy.

6 . DNA damage measurement.

Real Time PCR VS Traditional PCR

Real time PCR allows for the detection of PCR product during the early phases of the reaction. This ability of measuring the reaction kinetics in the early phases of PCR provide a distinct advantage over traditional PCR detection. Traditional methods use gel electrophoresis for the detection of PCR amplification in the final phase or at end-point of the PCR reaction.

Limitations of End-point PCR

In a PCR reaction as the reaction progresses, the reagents are being consumed as a result of amplification. Now the PCR product is no longer being doubled at each cycle due to this reagent constraint. This depletion will occur at different rates for each replicate. Thus, the samples begin to diverge in their quantities. This diminished amplification is the linear phase of the reaction. The plateau for each tube will differ due to the different reaction kinetics for each sample. It is in this phase where traditional PCR takes its measurement, also known as the end-point. This End-Point Detection has some problems such as low resolution, poor precision, low sensitivity and the need for post PCR processing. Also, the results of this detection are not expressed in numbers and there is no scope for automation.

Possible Questions:

1. Explain in detail about i) Linkers ii) Adapters and iii) Homopolymer tailing
2. Give a detailed note on Chromosome walking.
3. Discuss elaborately about sequence dependent screening for recombinants.
4. Discuss elaborately about sequence independent screening for recombinants
5. Give a detailed note on Southern blotting.
6. Explain the steps involved in Northern blotting.
7. Describe in detail about immunoblotting.
8. Give an elaborate note on immunological screening for recombinants.
9. Explain in detail about DNA hybridization techniques.
10. What are molecular probes? Explain in detail about probe labeling.
11. Discuss in detail about PCR and its applications.
12. What is DNA sequencing? Explain in detail about various methods of sequencing.

Unit III

SYLLABUS

Expression vectors, optimization of protein expression in heterologous systems, Fusion proteins, *In vitro* translation systems. Preparation and comparison of Genomic and cDNA library.

Expression vectors

- An expression vector, otherwise known as an expression construct, is usually a plasmid or virus designed for gene expression in cells.
- The vector is used to introduce a specific gene into a target cell, and can commandeer the cell's mechanism for protein synthesis to produce the protein encoded by the gene.
  - Expression vectors are the basic tools in biotechnology for the production of proteins. The vector is engineered to contain regulatory sequences that act as enhancer and promoter regions and lead to efficient transcription of the gene carried on the expression vector.
- The goal of a well-designed expression vector is the efficient production of protein, and this may be achieved by the production of significant amount of stable messenger RNA, which can then be translated into protein.
- The expression of a protein may be tightly controlled, and the protein is only produced in significant quantity when necessary through the use of an inducer, in some systems however the protein may be expressed constitutively.
- *Escherichia coli* is commonly used as the host for protein production, but other cell types may also be used. An example of the use of expression vector is the production of insulin, which is used for medical treatments of diabetes.
- Elements of expression vectors
  An expression vector has features that any vector may have, such as
  - An origin of replication
  - A selectable marker and

  A suitable site for the insertion of a gene such as the multiple cloning site.
- The cloned gene may be transferred from a specialized cloning vector to an expression vector, although it is possible to clone directly into an expression vector.
- The cloning process is normally performed in Escherichia coli, and vectors used for protein production in organisms other than E.coli may have, in addition to a suitable origin of

replication for its propagation in E. coli, elements that allow them to be maintained in another organism, and these vectors are called shuttle vectors.

Elements for expression

- An expression vector must have elements necessary for gene expression. These may include a promoter, the correct translation initiation sequence such as a ribosomal binding siteand start codon, a termination codon, and a transcription termination sequence.
- There are differences in the machinery for protein synthesis between prokaryotes and eukaryotes, therefore the expression vectors must have the elements for expression that is appropriate for the chosen host.
- For example, prokaryotes expression vectors would have a Shine-Dalgarno sequence at its translation initiation site for the binding of ribosomes, while eukaryotes expression vectors would contain the Kozak consensus sequence.
- The promoter initiates the transcription and is therefore the point of control for the expression of the cloned gene.
- The promoters used in expression vector are normally inducible, meaning that protein synthesis is only initiated when required by the introduction of an inducer such as IPTG. Gene expression however may also be constitutive (i.e. protein is constantly expressed) in some expression vectors. Low level of constitutive protein synthesis may occur even in expression vectors with tightly controlled promoters.

Protein tags

- After the expression of the gene product, it is usually necessary to purify the expressed protein; however, separating the protein of interest from the great majority of proteins of the host cell can be a protracted process.
- To make this purification process easier, a purification tag may be added to the cloned gene. This tag could be histidine (His) tag, other marker peptides, or a fusion partners such as glutathione S-transferase or maltose-binding protein.
- Some of these fusion partners may also help to increase the solubility of some expressed proteins. Other fusion proteins such as green fluorescent protein may act as a reporter gene for the identification of successful cloned genes.

Others

- The expression vector is transformed or transfected into the host cell for protein synthesis. Some expression vectors may have elements for transformation or the insertion of DNA into

the host chromosome, for example the *vir* genes for plant transformation, and integrase sites for chromosomal insertion.

- Some vectors may include targeting sequence that may target the expressed protein to a specific location such as the periplasmic space of bacteria.

Expression/Production systems

Prokaryotic expression systems

These are used to produce many enzymes and useful eukaryotic proteins/peptides

Most of the enzymes used in recombinant DNA technology, *e.g.* many restriction enzymes, DNA polymerases, DNA ligases, polynucleotide kinase and reverse transcriptases are now produced as recombinant proteins themselves.

In addition, simple eukaryotic proteins, such as protein hormones which are medically important (e.g. insulin, growth hormone, granulocyte colony-stimulating factor (G-CSF), can be produced in bacteria and in sufficient quantities for use as therapeutic agents.

Eukaryotic expression systems

may be necessary to produce some types of protein, such as proteins that require post-translational modifications or need to be folded properly for biological activity.

e.g. glycoproteins

Prokaryotes lack the enzymes that catalyse many of the post-translational modifications found on eukaryotic proteins. Proteins produced in prokaryotes may not be folded properly and/or can be insoluble - forming **inclusion bodies.** It may be very difficult to re-solubilise proteins found in inclusion bodies and 'restore' biological activity.

For many types of mammalian proteins, mammalian expression systems may be best.

- Study protein/gene function in a native environment
- Produce biologically active molecules
- Introduce new genes or alter existing ones in intact animals

Complex human proteins are produced by large-scale mammalian cell culture

- **Tissue plasminogen activator** (tPA), a protease that cleaves plasminogen, forming the active form plasmin. Plasmin degrades fibrin and is responsible for dissolving blood clots. Rapid administration of tPA to heart attack victims dissolves the life-threatening clots that lead to irreversible damage of heart muscle.
- **Factor VIII** is a protein required for normal clotting of blood and is administered to haemophiliacs who have a faulty Factor VIII gene.

Production of recombinant proteins

- One potential of recombinant DNA is the production of proteins.
- Many systems have been designed for the expression of recombinant proteins.
- Although *E. coli* are easy to grow in large quantities there are some problems associated with the expression of eukaryotic proteins in prokaryotes.
- Therefore, expression systems in eukaryotic systems have also been devised.

The following are the general considerations in the over production of recombinant proteins.

- strong promoter
- regulatable promoter
- gene dosage
- localization signals
- protease defective hosts
- fusion proteins

Ideally,

- the recombinant protein should be cloned in conjunction with a strong promoter.

Promoters are elements found on the 5'- end of genes and control their expression.

- A strong promoter results in a high level of gene expression.
- In addition to being strong, promoters should be regulatable so that transcription can be turned off and on.
- Recombinant proteins are sometimes toxic for the host cell and this toxicity can be minimized by controlling the expression.
- The use of high copy number plasmids also increases the production of recombinant proteins by increasing the gene dosage.
- it is often convenient to have recombinant proteins that are exported from the host cell. This can be accomplished by engineering localization signals within the recombinant protein so that they are directed to a particular cellular compartment.
- In addition, the recombinant protein can also be expressed as a fusion protein with another protein. This will sometimes stabilize the recombinant protein and/or assist in the purification and characterization of the recombinant protein.
- Proteolysis of recombinant proteins is often a major problem that is partially alleviated by cloning into protease deficient hosts.

Expression in bacterial cell

The general procedure for expression of cloned genes in *E. coli* involves

- the insertion of the coding region of interest into a vector, usually a plasmid, so the region is efficiently transcribed and translated.
- Since eukaryotic genes do not contain the proper signals for transcription initiation, ribosome recognition, translation initiation, and translation termination, these signals need to be supplied by the vector.

Common promoters used in bacterial expression vectors are:

λ PL, lac, tac and T7

Of these, the lac and tac promoters are the most widely used.

Control elements that regulate expression from these promoters are supplied by either the host or the vector.

The **λPL promoter** is controlled by a mutant cI repressor protein that is non-functional at 42°C.

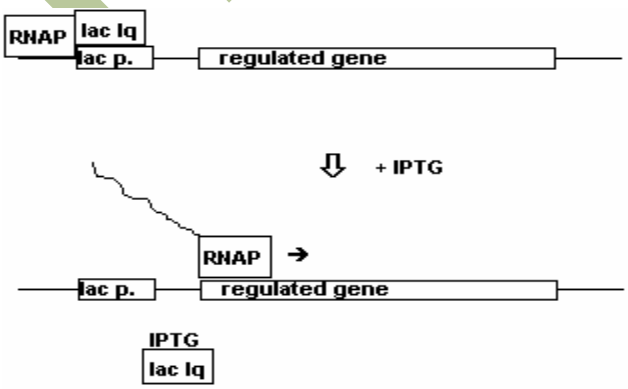
- At temperatures less that 42°C the repressor protein (cI) binds to the promoter and prevents expression.
- To induce expression the temperature is raised to that the repressor becomes nonfunctional and expression is now permitted.
- One problem with this system is that heat-shock proteins may also be induced.

The **T7 promoter** is from the T7 phage

- is only transcribed by the T7 RNA polymerase (T7 gene 1).
- The host cell must also contain the T7 gene 1 in order to express from this promoter.
- To regulate expression from this promoter it is necessary to be able to regulate the expression of the T7 gene 1.

The **lac** and **tac** promoters

- are controlled by the lac repressor.



- The lac repressor binds to the lac promoter and prevents RNA polymerase from transcribing the gene. Host strains with lacIq gene express the lac repressor at 10-fold higher concentrations than the normal
- lacI gene.
- IPTG (isopropyl-1-thio-β -D-galactopyanoside, an analog of lactose) binds to the repressor and prevents its interaction with the lac promoter and allowing RNA polymerase to transcribe the regulated gene.
- The tac promoter is a fusion of trp and lac promoters and is also regulated by IPTG.

The production of recombinant **fusion proteins** often stabilizes the expression of foreign proteins in *E. coli*.

Several plasmids that express recombinant proteins as fusion proteins have been developed (Table).

Fusion Protein	Affinity Matrix
Glutathione-Stransferase	glutathione agarose
Thioredoxin	phenylarsine oxide-agarose
Maltose Binding Protein	amyloseagarose
Six Histidine Residues	Ni-agarose

In addition to stabilizing the recombinant protein, the fusion partner is often exploited for affinity purification or for the analysis of the recombinant protein.

- Recombinant proteins expressed at high levels will sometimes form insoluble aggregates known a **inclusion bodies**.
- In some applications it is possible to take advantage of this phenomenon.
- For example, the inclusion bodies can be isolated by differential centrifugation and solubilized under denaturing conditions (eg., urea). It is sometimes possible to renature the protein and regain activity. In addition, fusions with *E. coli* thioredoxin can circumvent inclusion body formation.

Expression in eukaryotes

- Although expression of recombinant proteins in *E. coli* is usually fairly straight forward, it is often desirable or necessary to express cloned genes in eukaryotes.
- Eukaryotic expression systems are often needed to insure correct folding and disulfide-bond formation, post-translational modifications, and processing.
- Yeast, such as *Saccharomyces cerevisiae* and *Pichia* species, are useful hosts for the expression of recombinant proteins since they can be grown and manipulated like a bacteria.
- *Pichia* expression systems often allow for high level expression of recombinant proteins using a strong alcohol oxidase promoter which is induced by methanol.

Shuttle vectors contain both a bacterial origin of replication and the origin of replication of interest.

This allows for manipulations of the plasmids to be carried out in *E. coli* before transforming yeast or other eukaryotes.The two major strategies for expression of recombinant sequences in mammalian cells are

1) stable or transient expression of transfected DNA and

2) the use of viral expression vectors.

Mammalian and other eukaryotic cells are able to take up DNA.

Several different methods for introducing DNA to cells are available

- calcium phosphate
- DEAE-dextran
- electroporation
- liposomes
- protoplast fusion
- ballistics (gene gun)
- microinjection

Several different **eukaryotic viruses** can also be used as cloning vectors for the expression of recombinant proteins.

- Lytic viruses are good for transient expression whereas episomal viruses are better for constitutive expression.
- Retroviruses can become incorporated into the host cell genome and possibly lead to a stable transformation.

Baculoviruses are used for the production of recombinant eukaryotic proteins.

- *Autographica californica* is a nuclear polyhedrosis virus (AcNPV) that infects insects.
- Sf9 cells, derived from *Spodoptera frugiperda*, are readily grown in vitro and can be infected with baculovirus.

Insect cells

- Extremely high levels of recombinant proteins can be expressed and many eukaryotic post-translational modifications are correctly made in the host insect cells.
- This high level of expression is driven by a strong promoter for the polyhedrin gene.
- The polyhedron protein is expressed late in infection as the virus is killing the host cell and is needed for the dissemination of the virus in nature.

- However, the polyhedrin gene is unnecessary for viral growth in tissue culture and can be replaced with a gene of interest.

Fusion protein

- A fusion protein is a single protein that is produced by the expression of two formerly separate genes that have been combined into a single gene.
- A protein made from a fusion gene, which is created by joining parts of two different genes. Fusion genes may occur naturally in the body by transfer of DNA between chromosomes. For example, the BCR-ABL gene found in some types of leukemia is a fusion gene that makes the BCR-ABL fusion protein. Fusion genes and proteins can also be made in the laboratory by combining genes or parts of genes from the same or different organisms.
- Fusion proteins or chimeric are proteins created through the joining of two or more genes that originally coded for separate proteins.
- Translation of this fusion gene results in a single or multiple polypeptides with functional properties derived from each of the original proteins. Recombinant fusion proteins are created artificially by recombinant DNA technology for use in biological research or therapeutics.
- Chimeric or chimera usually designate hybrid proteins made of polypeptides having different functions or physico-chemical patterns.
- Chimeric mutant proteins occur naturally when a complex mutation, such as a chromosomal translocation, tandem duplication, or retrotransposition creates a novel coding sequence containing parts of the coding sequences from two different genes.
- Naturally occurring fusion proteins are commonly found in cancer cells, where they may function as oncoproteins. The bcr-abl fusion protein is a well-known example of an oncogenic fusion protein, and is considered to be the primary oncogenic driver of chronic myelogenous leukemia.

Natural occurrence

- Naturally occurring fusion genes are most commonly created when a chromosomal translocation replaces the terminal exons of one gene with intact exons from a second gene.
- This creates a single gene that can be transcribed, spliced, and translated to produce a functional fusion protein. Many important cancer-promoting oncogenes are fusion genes produced in this way.

Functions

- Some fusion proteins combine whole peptides and therefore contain all functional domains of the original proteins. However, other fusion proteins, especially those that occur naturally, combine only portions of coding sequences and therefore do not maintain the original functions of the parental genes that formed them.
- Many whole gene fusions are fully functional and can still act to replace the original peptides. Some, however, experience interactions between the two proteins that can modify their functions.
- Beyond these effects, some gene fusions may cause regulatory changes that alter when and where these genes act. For partial gene fusions, the shuffling of different active sites and binding domains have potential to result in new proteins with novel functions.

***In vitro* translation system**

- The *in vitro* synthesis of proteins in cell-free extracts is an important tool for molecular biologists and has a variety of applications, including the rapid identification of gene products (e.g., proteomics), localization of mutations through synthesis of truncated gene products, protein folding studies, and incorporation of modified or unnatural amino acids for functional studies.
- The use of in vitro translation systems can have advantages over *in vivo*gene expression when the over-expressed product is toxic to the host cell, when the product is insoluble or forms inclusion bodies, or when the protein undergoes rapid proteolytic degradation by intracellular proteases.
- In principle, it should be possible to prepare a cell-free extract for *in vitro* translation of mRNAs from any type of cells. In practice, only a few cell-free systems have been developed for *in vitro* protein synthesis.
- In general, these systems are derived from cells engaged in a high rate of protein synthesis.

Cell-Free Expression Systems

- The most frequently used cell-free translation systems consist of extracts from rabbit reticulocytes, wheat germ and Escherichia coli. All are prepared as crude extracts containing all the macromolecular components (70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination factors, etc.) required for translation of exogenous RNA.

- To ensure efficient translation, each extract must be supplemented with amino acids, energy sources (ATP, GTP), energy regenerating systems (creatine phosphate and creatine phosphokinase for eukaryotic systems, and phosphoenol pyruvate and pyruvate kinase for the E. *coli* lysate), and other co-factors (Mg2+, K+, etc.).

- There are two approaches to in vitro protein synthesis based on the starting genetic material: RNA or DNA. Standard translation systems, such as reticulocyte lysates and wheat germ extracts, use RNA as a template; whereas "coupled" and "linked" systems start with DNA templates, which are transcribed into RNA then translated. Each of these systems is discussed below.

Translation Systems

Rabbit Reticulocyte Lysate

Rabbit reticulocyte lysate is a highly efficient in vitro eukaryotic protein synthesis system used for translation of exogenous RNAs (either natural or generated in vitro).

- In vivo, reticulocytes are highly specialized cells primarily responsible for the synthesis of hemoglobin, which represents more than 90% of the protein made in the reticulocyte. These immature red cells have already lost their nuclei, but contain adequate mRNA, as well as complete translation machinery, for extensive globin synthesis.

- The endogenous globin mRNA can be eliminated by incubation with Ca2+-dependent micrococcal nuclease, which is later inactivated by chelation of the Ca2+ by EGTA. Ambion offers a nuclease-treated reticulocyte lysate.

- This type of lysate is the most widely used RNA-dependent cell-free system because of its low background and its efficient utilization of exogenous RNAs even at low concentrations (Figure). Exogenous proteins are synthesized at a rate close to that observed in intact reticulocyte cells.

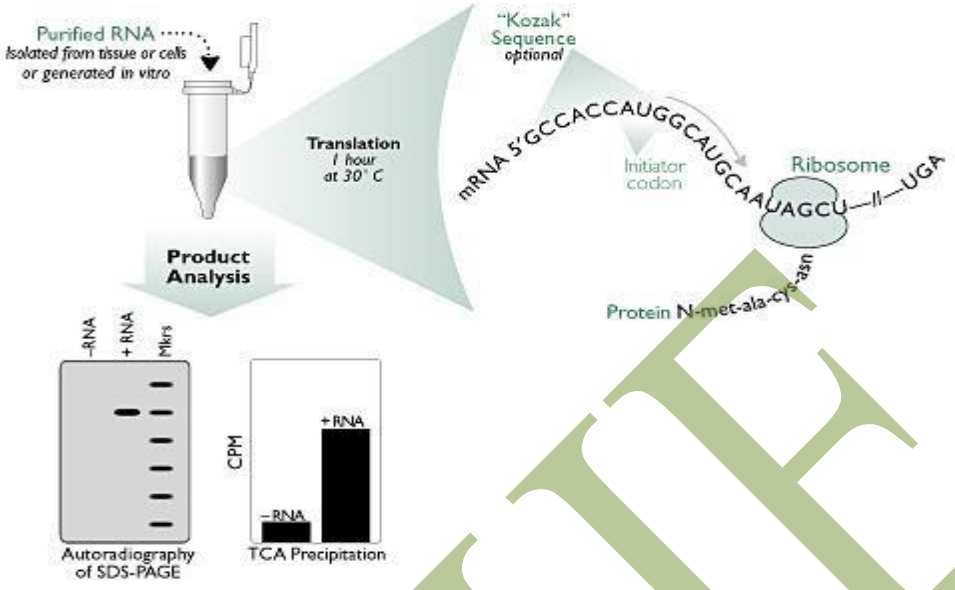


Figure : Standard in Vitro Translation Procedure Using Rabbit Reticulocyte Lysate or Wheat Germ Extract.

- Untreated reticulocyte lysate translates endogenous globin mRNA, exogenous RNAs, or both. This type of lysate is typically used for studying the translation machinery, e.g. studying the effects of inhibitors on globin translation.

- Both the untreated and treated rabbit reticulocyte lysates have low nuclease activity and are capable of synthesizing a large amount of full-length product. Both lysates are appropriate for the synthesis of larger proteins from either capped or uncapped RNAs (eukaryotic or viral).

Wheat Germ Extract

- Wheat germ extract is a convenient alternative to the rabbit reticulocyte lysate cell-free system. This extract has low background incorporation due to its low level of endogenous mRNA. Wheat germ lysate efficiently translates exogenous RNA from a variety of different organisms, from viruses and yeast to higher plants and mammals.

- The wheat germ extract is recommended for translation of RNA containing small fragments of double-stranded RNA or oxidized thiols, which are inhibitory to the rabbit reticulocyte lysate. Both retic and wheat germ extracts translate RNA isolated from cells and tissue or those generated by *in vitro* transcription.

- When using RNA synthesized *in vitro*, the presence of a 5' cap structure may enhance translational activity. Typically, translation by wheat germ extracts is more cap-dependent than translation by retic extracts. If capping of the RNA is impossible and the protein yield from an uncapped mRNA is low, the coding sequence can be subcloned into a prokaryotic vector and expressed directly from a DNA template in an *E. coli* cell-free system.

E. coli Cell-Free System

- E. *coli* cell-free systems consist of a crude extract that is rich in endogenous mRNA. The extract is incubated during preparation so that this endogenous mRNA is translated and subsequently degraded.

- Because the levels of endogenous mRNA in the prepared lysate is low, the exogenous product is easily identified. In comparison to eukaryotic systems, the *E. coli* extract has a relatively simple translational apparatus with less complicated control at the initiation level, allowing this system to be very efficient in protein synthesis.

- Bacterial extracts are often unsuitable for translation of RNA, because exogenous RNA is rapidly degraded by endogenous nucleases. There are some viral mRNAs (TMV, STNV, and MS2) that translate efficiently, because they are somewhat resistant to nuclease activity and contain stable secondary structure. However, *E. coli* extracts are ideal for coupled transcription:translation from DNA templates.

"Linked" And "Coupled" Transcription:Translation Systems

- In standard translation reactions, purified RNA is used as a template for translation. "Linked" and "coupled" systems, on the other hand, use DNA as a template. RNA is transcribed from the DNA and subsequently translated without any purification.

- Such systems typically combine a prokaryotic phage RNA polymerase and promoter (T7, T3, or SP6) with eukaryotic or prokaryotic extracts to synthesize proteins from exogenous DNA templates.

DNA templates for transcription:translation reactions may be cloned into plasmid vectors or generated by PCR (Primer Sequences for PCR-generated Translation Templates).

Linked Transcription: Translation

- The "linked" system is a two-step reaction, based on transcription with a bacteriophage polymerase followed by translation in the rabbit reticulocyte lysate or wheat germ lysate (Figure).

- Because the transcription and translation reactions are separate, each can be optimized to ensure that both are functioning at their full potential. Conversely, many commercially available eukaryotic coupled transcription:translation systems have compromised one or both reactions so that they can occur in a single tube. Thus, yield is sacrificed for convenience.

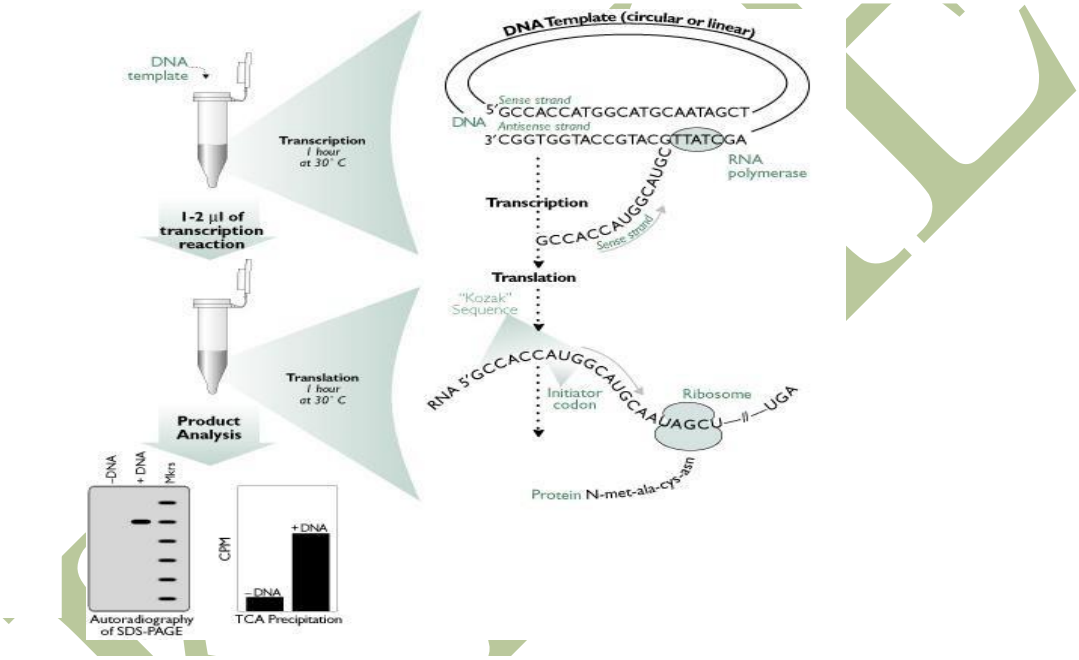


Figure : Linked in Vitro Transcription and Translation Procedure Using Rabbit Reticulocyte Lysate.

Coupled Transcription: Translation

- Unlike eukaryotic systems where transcription and translation occur sequentially, in E. coli, transcription and translation occur simultaneously within the cell. In vitro E. coli translation systems are thus performed the same way, coupled, in the same tube under the same reaction conditions (one-step reaction; Figure).
- During transcription, the 5' end of the RNA becomes available for ribosomal binding and undergoes translation while its 3' end is still being transcribed.

- This early binding of ribosomes to the RNA maintains transcript stability and promotes efficient translation. This bacterial translation system gives efficient expression of either prokaryotic or eukaryotic gene products in a short amount of time.
- For the highest protein yield and the best initiation fidelity, make sure the DNA template has a Shine-Dalgarno ribosome binding site upstream of the initiator codon. Capping of eukaryotic RNA is not required.
- Use of *E. coli* extract also eliminates cross-reactivity or other problems associated with endogenous proteins in eukaryotic lysates. Also, the E. coli S30 extract system allows expression from DNA vectors containing natural E. coli promoter sequences (such as lac or tac).

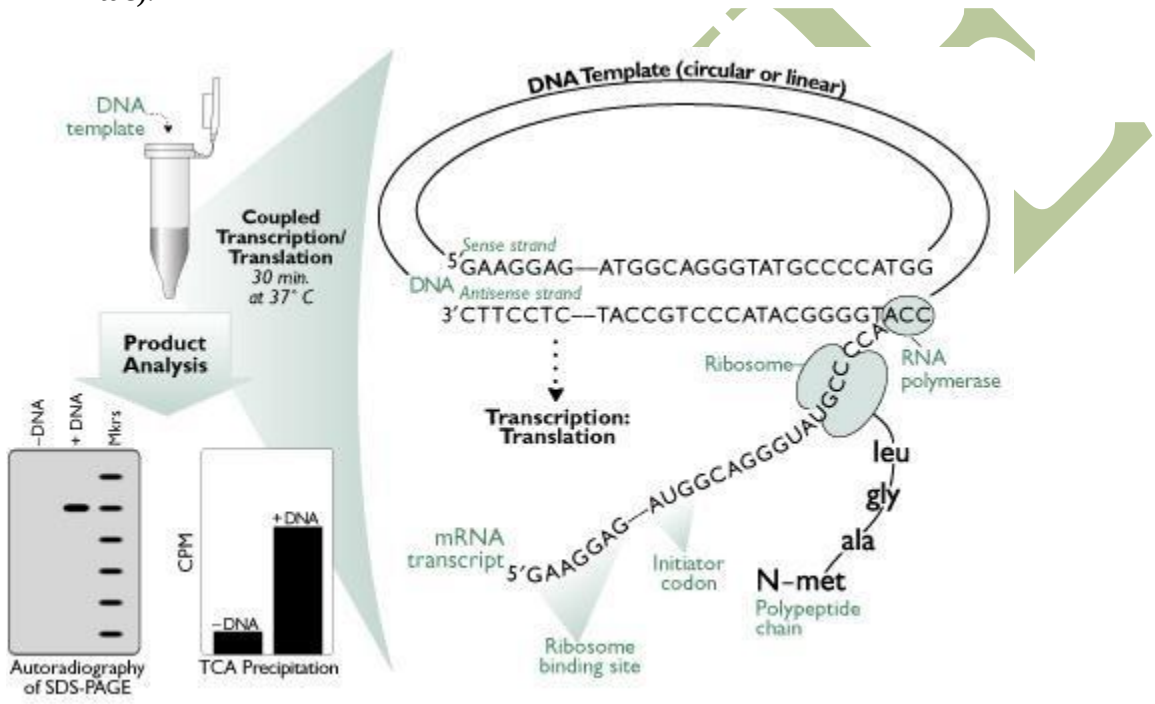


Figure : Coupled in Vitro Transcription:Translation Procedure Using E. coli Extract.

Important Elements for Translation

- There are some significant differences between prokaryotic and eukaryotic mRNA transcripts. Typically, eukaryotic mRNAs are characterized by two post-transcriptional modifications: a 5'-7 methyl-GTP cap and a 3' poly(A) tail.

- Both modifications contribute to the stability of the mRNA by preventing degradation. Additionally, the 5' cap structure enhances the translation of mRNA by helping to bind the eukaryotic ribosome and assuring recognition of the proper AUG initiator codon.

- This function may vary with the translation system and with the specific mRNA being synthesized. The consensus sequence 5'-GCCACCAUGG-3', also known as the "Kozak" sequence, is considered to be the strongest ribosomal binding signal in eukaryotic mRNA.

- For efficient translation initiation, the key elements are the G residue at the +1 position and the A residue at the -3 position. An mRNA that lacks the Kozak consensus sequence may be translated efficiently in eukaryotic cell-free systems if it possesses a moderately long 5'-untranslated region (UTR) that lacks stable secondary structure.

  In bacteria, the ribosome is guided to the AUG initiation site by a purine-rich region called the Shine-Dalgarno (SD) sequence. This sequence is complementary to the 3' end of the 16s rRNA in the 30S ribosomal subunit.

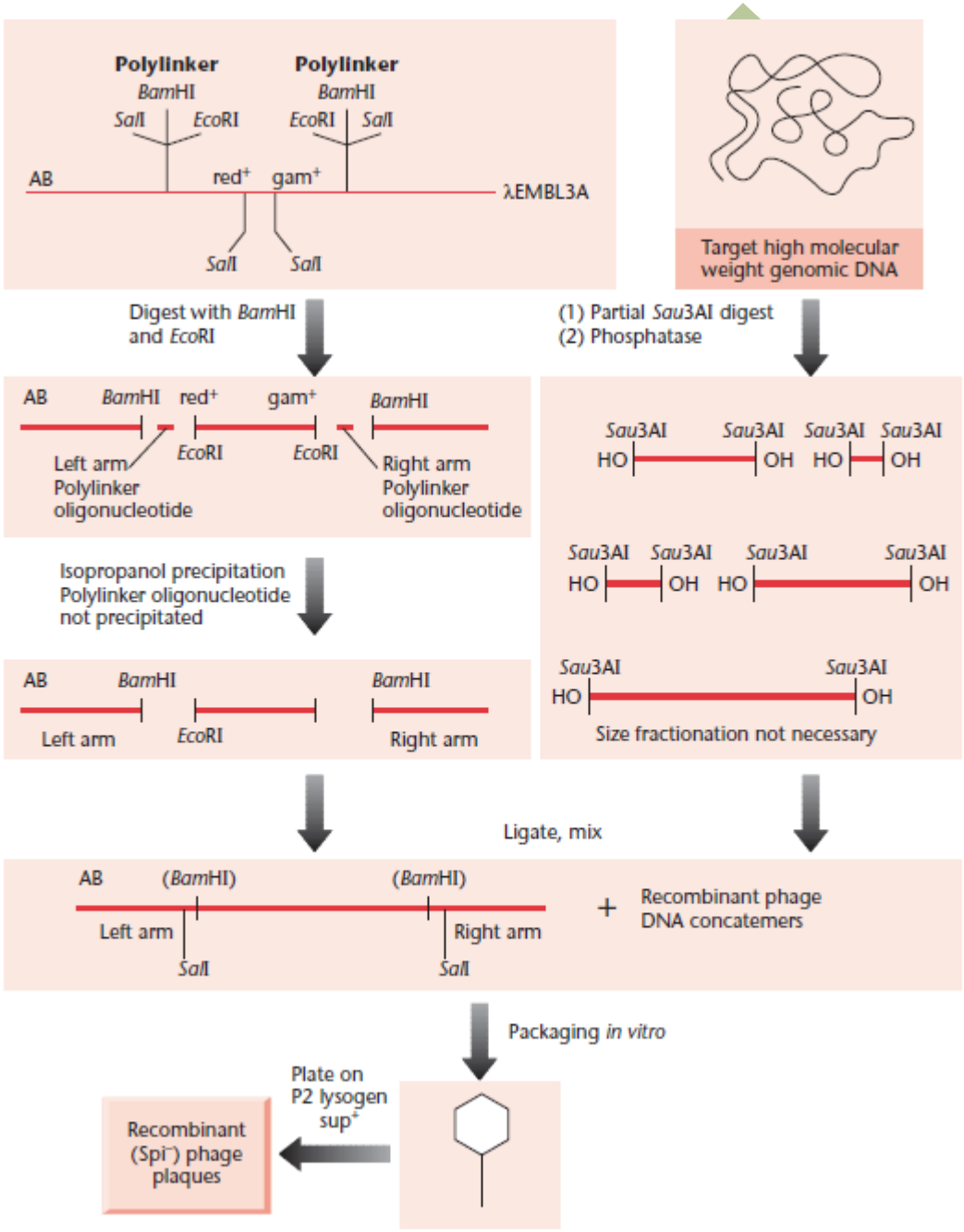
- Upstream from the initiation AUG codon, the SD region has the consensus sequence 5'-UAAGGAGGUGA-3'. Specific mRNAs vary considerably in the number of nucleotides that complement the anti-Shine-Dalgarno sequence of 16S rRNA, ranging from as few as two to nine or more.

- The position of the ribosome binding site (RBS) in relation to the AUG initiator is very important for efficiency of translation (usually from -6 to -10 relative to the A of the initiation site).

Construction of A Genomic Library

- For preparation of a genomic library, the total genomic DNA of an organism is extracted.
- The DNA is broken into fragments of appropriate size either by mechanical shearing (this generates blunt ended fragments), sonication, or by using a suitable restriction endonuclease for partial digestion of the DNA.
- Complete digestion is avoided since it generates fragments that are too heterogeneous in size.

- For partial digestion, restriction enzymes having 4-base (tetrameric) recognition sequences are employed in preference to those having 6-base (hexameric) target sites. This is because a given 4-base recognition site is expected to' occur every 44 (= 256) base pairs in a DNA molecule, while a 6 base target site would occur only after every 46 (=



4096) base pairs. (It is assumed here that the arrangement of the 4 bases in DNA molecules is random).

- Therefore, the fragments produced in partial digests with enzymes having 4 base recognition sites are more likely to be of appropriate size for cloning than those generated by enzymes having 6 base recognition sites.
- Single or mixed digestions with the enzymes AluI, HaeIII or Sau3A have been used for constructing genomic libraries.
- The use of restriction enzymes has the advantage that the same set of fragments are obtained from a DNA each time a specific enzyme is used, and many of the enzymes; produce cohesive ends.
- The partial digests of genomic DNA are subjected to agarose gel electrophoresis or sucrose gradient centrifugation for separation from the mixture of fragment of appropriate size.
- These fragments are then inserted into a suitable vector for cloning.
- This constitutes the shotgun approach to gene cloning. In principle, any vector can be used, but A vectors and cosmids have been the most commonly used since DNA inserts of upto 23-25 kb (kilobase pairs) can be cloned in these vectors. The vectors containing the inserts are cloned in a suitable bacterial host.

Genomic DNA libraries *Producing representative genomiclibraries in λ cloning vectors*

let us suppose that to clone a single-copy gene from the human genome.

How might this be achieved?

- simply digest total human DNA with a restriction endonuclease, such as *Eco*RI, insert the fragments into a suitable phage-λ vector and then attempt to isolate the desired clone.

How many recombinants have to screen in order to isolate the right one?

- Assuming *Eco*RI gives, on average, fragments of about 4 kb,

- and given that the size of the human haploid genome is 2.8 × 106 kb, we can see that over 7 × 10^5 independent recombinants must be prepared and screened in order to have a reasonable chance of including the desired sequence.

There are two problems with the above approach.

- First, the gene may be cut internally one or more times by *Eco*RI so that it is not obtained as a single fragment.
- These problems can be overcome by cloning *random* DNA fragments of a large size (for λ replacement vectors, approximately 20 kb).
- Since the DNA is randomly fragmented, there will be no systematic exclusion of any sequence.
- Furthermore, clones will overlap one another, allowing the sequence of very large genes to be assembled and giving an opportunity to 'walk' from one clone to an adjacent one.
- Because of the larger size of each cloned DNA fragment, fewer clones are required for a complete or nearly complete library.

How many clones are required?

Let *n* be the size of the genome relative to a single cloned fragment. Thus, for the human genome (2.8 × 10^6 kb) and an average cloned fragment size of 20 kb, $n = 1.4 \times 10^5$.

The number of independent recombinants required in the library must be greater than *n*, because sampling variation will lead to the inclusion of some sequences several times and the exclusion of other sequences in a library of just *n* recombinants.

Clarke and Carbon (1976) have derived a formula that relates the probability (*P*) of including any DNA sequence in a random library of *N* independent recombinants:

$$N = \frac{\ln(1-P)}{\ln\left(1-\frac{1}{n}\right)}$$

Therefore, to achieve a 95% probability ($P = 0.95$) of including any particular sequence in a random human genomic DNA library of 20 kb fragment size:

$$N = \frac{\ln(1 - 0.95)}{\ln\left(1 - \frac{1}{1.4 \times 10^5}\right)} = 4.2 \times 10^5$$

Notice that a considerably higher number of recombinants is required to achieve a 99% probability, for here $N = 6.5 \times 10^5$.

How can appropriately sized random fragments be produced?

Various methods are available.

- Random breakage by mechanical shearing is appropriate because the average fragment size can be controlled, but insertion of the resulting fragments into vectors requires additional steps.
- The more commonly used procedure involves restriction endonucleases. In the strategy devised by Maniatis *et al*. (1978)
- the target DNA is digested with a mixture of *two* restriction enzymes.
- These enzymes have tetranucleotide recognition sites, which therefore occur frequently in the target DNA and in a complete double-digest would produce fragments averaging less than 1 kb.
- However, only a partial restriction digest is carried out, and therefore the majority of the fragments are large (in the range 10–30 kb).
- Given that the chances of cutting at each of the available restriction sites are more or less equivalent, such a reaction effectively produces a random set of overlapping fragments.
- These can be size-fractionated, e.g. by gel electrophoresis, so as to give a random population of fragments.

Cloning of cDNA

cDNA Library

- A cDNA library is a population of bacterial transformants or phage lysates in which each mRNA isolated from an organism or tissue is represented as its cDNA insertion in a plasmid or a phage vector.
- The frequency of a specific cDNA in such a library would ordinarily depend on the frequency of the concerned mRNA in the tissue/organism in question.

Preparation of cDNA

- cDNA is the copy or complementary DNA produced by using mRNA (usually) as a template. In fact, any RNA molecule can be used to produce cDNA.
- DNA copy of an RNA molecule is produced by the enzyme reverse transcriptase (RNA dependent DNA polymerase; discovered by **Temin and Baltimore** in 1970) generally obtained from avian mycloblastosis virus (AMV).
- This enzyme performs similar reactions as DNA polymerase, and has an absolute requirement for a primer with a free 3' -OH.
- When eukaryotic mRNA is used as a template, a poly T oligonucleotide (more specifically, oligodeoxynucleotide) is conveniently used as the primer since these mRNAs have a poly-A tail at their 3' ends.
- But special tricks are required to utilize primers for other RNAs, e.g., prokaryotic mRNA, rRNA, RNA virus genomes, etc.

For example, a poly A tail may be added to 3' end of the RNA to make it analogous to eukaryotic mRNA (oligo-T is now used as primer); this reaction is catalyzed by the enzyme poly A polymerase.

The appropriate oligonucleotide primer (oligo- T for eukaryotic mRNA) is annealed with the mRNA; this primer will base-pair to the 3'-end of mRNA.

Reverse transcriptase extends the 3'-end of the primer using mRNA molecule as a template. This produces a RNA.DNA hybrid molecule, the DNA strand being the cDNA.

The RNA strand is digested either by RNase H or alkaline hydrolysis; this frees the single-stranded cDNA.

- Curiously, the 3'-end of this cDNA serves as its own primer and provides the free 3'-OH required for the synthesis of its complementary strand; therefore, a primer is not required for this step.
- The complementary strand of cDNA single strand is synthesized by either the reverse transcriptase itself or by E. coli DNA polymerase; this generates a hairpin loop in the cDNA. The hairpin loop is cleaved by a single strand specific nuclease to yield a regular DNA duplex.

Problems in cDNA Preparation

- Usually the double strand cDNA preparations are always a mixture of different kinds of molecules due to problems in copying of the RNA and also because even highly purified mRNAs are never absolutely pure.
- Physical and chemical methods are incapable of resolving these mixtures.
- Therefore, the cDNA mixture itself is used for cloning and the desired cDNA is identified and isolated in pure form from the appropriate bacterial clone.

Possible Questions:

1. List out the strategies and considerations in cloning.
2. Explain in detail about recombinant protein production in prokaryotes.
3. Explain in detail about recombinant protein production in Eukaryotes.
4. How genomic DNA library is constructed?
5. What are fusion proteins? Give its applications.
6. Explain in detail about protein synthesis.
7. Elaborate the steps involved in cDNA library construction.
8. What are *invitro* translation systems.
9. What are optimizations measures of prokaryotic system for protein production?

Unit IV

SYLLABUS

Primer extension and PCR based methods of site directed mutagenesis, Random mutagenesis, Gene shuffling, production of chimeric proteins, Protein engineering concepts and examples (any two).

Mutagenesis

Mutagenesis is the process of inducing mutations. Mutations may occur due to exposure to natural mutagens such as ultraviolet (UV) light, to industrial or environmental mutagens such as benzene or asbestos, or by deliberate mutagenesis for purposes of genetic research. For geneticists, the study of mutagenesis is important because mutants reveal the genetic mechanisms underlying heredity and gene expression. Mutations are also important for studying protein function: Often the importance of a protein cannot be characterized unless a mutant can be made in which that protein is absent.

Noninduced Mutagenic Agents

- Environmental agents can influence the mutation rate not only by increasing it, but also by decreasing it. For example, antioxidants, which are found commonly in fruits and vegetables, are thought by many to protect against mutagens that are generated by normal cellular respiration.
- In addition to protective agents, however, many plants also contain deleterious mutagens known as carcinogens . Many chemical mutagens exist both naturally in the environment and as a result of human activity. Benzo(a)pyrene, for example, is produced by any incomplete burning, whether of tobacco in a cigarette or of wood in forest fires.

Random mutagenesis

- To create proteins with new or improved properties, directed protein evolution is one of the most powerful strategies. Mutagenex provides mutant libraries with a high diversity and accuracy generated by an error-prone PCR and site-directed random mutagenesis (site saturation) at either single or multiple sites.
- Site-directed random mutagenesis is also called targeted random mutagenesis or site-saturation mutagenesis. Since most of conventional directed evolution strategies such as error-prone PCR and DNA shuffling generate large numbers of non-functional mutants rather than functional genes, these approaches inevitably require extensive high-throughput screening system, which request a hugh cost and long time.
- By targeting specific residues, most of codons remain as wild-type that can make mutants maintain natural structure of protein backbone despite of hyper diversity of functional region.

Therefore, the mutant library can contains high portion of functional mutants that are more likely to yield positive results.

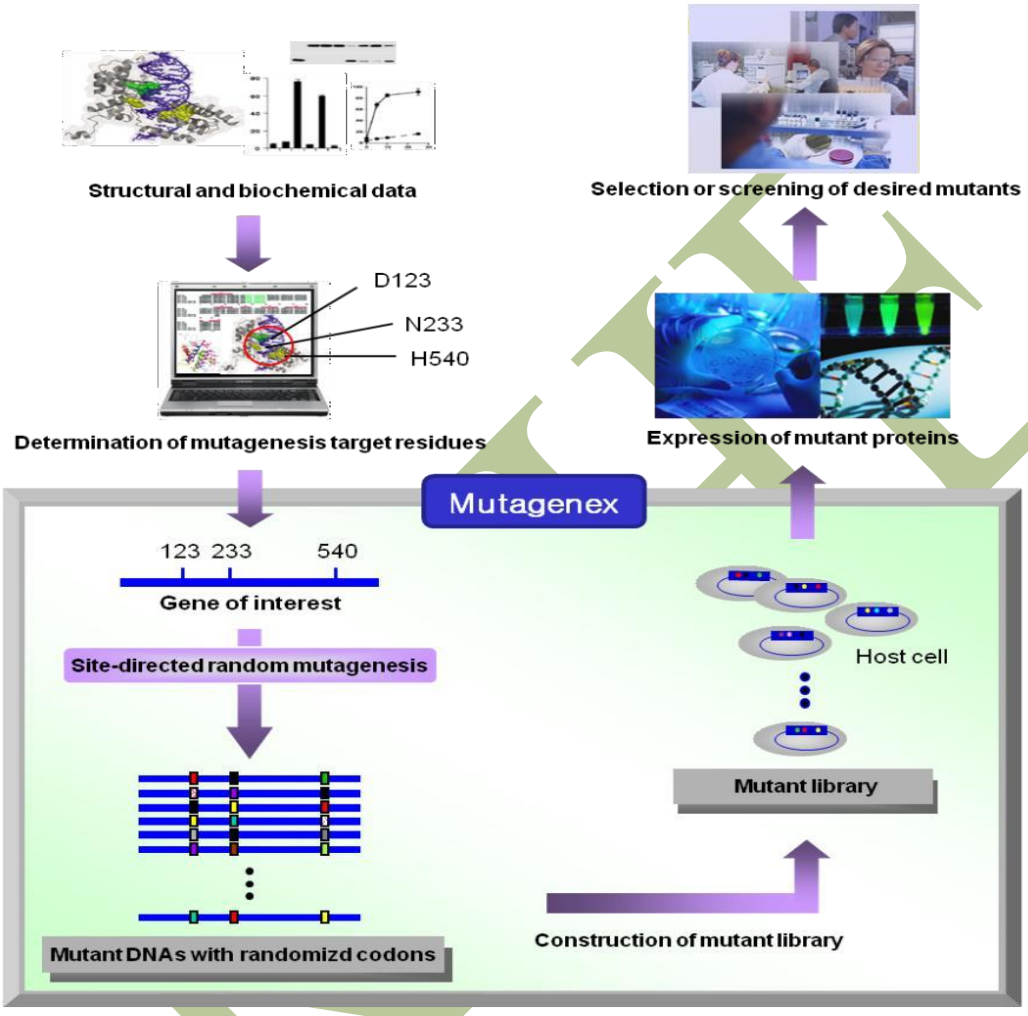


Figure : Random mutagenesis

Applications

Random mutagenesis has been successful for many directed evolution experiments. The applications of this method include

1. Enzyme
- Alteration of enontioselectivity
- Alteration substrate specificity
- New catalytic activity

- Stability at extreme temperature and pH

2. Antibody
- Enhanced binding activity and stability
- Decreased immunogenesity
- Increased serum half-life

3. Hormones
- Improved association rate or affinity
- Reduced affinity to increase recycling rate
- Increased serum half-life

4. Others
- Improvement of cytokines
- Alteration of GFP fluorescence property
- DNA or RNA sequence optimization

Approaches to Random Mutagenesis

- Random mutagenesis is an incredibly powerful tool for altering the properties of enzymes. Imagine, for example, you were studying a G-protein coupled receptor (GPCR) and wanted to create a temperature-sensitive version of the receptor or one that was activated by a different ligand than the wild-type. How could you do this?
- Firstly, you would clone the gene encoding the receptor, then randomly introduce mutations into the gene sequence to create a "library" containing thousands of versions of the gene. Each version (or "variant") of the gene in the library would contain different mutations and so encode receptors with slightly altered amino acid sequences giving them slightly different enzymatic properties than the wild-type.
- Next, you could transform the library into a strain where the receptor would be expressed and apply a high throughput screen to pick out variants in the library that have the properties you are looking for. Using a high throughput screen for GPCR activity you could pick out the variants from the library that were temperature-sensitive or were activated by different ligands.
- Sound easy? Well, of course it's not that easy. Creating a random mutant library that contains enough variants to give you a good chance of obtaining the altered enzyme you desire is a challenge in itself. There are many ways to create random mutant libraries, each with it's own pros and cons. Here are some of them:

1. Error-prone PCR.

- This approach uses a "sloppy" version of PCR, in which the polymerase has a fairly high error rate (up to 2%), to amplify the wild-type sequence.

- The PCR can be made error-prone in various ways including increasing the MgCl2 in the reaction, adding MnCl2 or using unequal concentrations of each nucleotide. Here is a good review of error prone PCR techiques and theory.
- After amplification, the library of mutant coding sequences must be cloned into a suitable plasmid.
- The drawback of this approach is that size of the library is limited by the efficiency of the cloning step. Although point mutations are the most common types of mutation in error prone PCR, deletions and frameshift mutations are also possible. There are a number of commercial error-prone PCR kits available, including those from Stratagene and Clontech.

2. Rolling circle error-prone PCR

- This is a variant of error-prone PCR in which wild-type sequence is first cloned into a plasmid, then the whole plasmid is amplified under error-prone conditions.
- This eliminates the ligation step that limits library size in conventional error-prone PCR but of course the amplification of the whole plasmid is less efficient than amplifying the coding sequence alone. More details can be found here.

3. Mutator strains

- In this approach the wild-type sequence is cloned into a plasmid and transformed into a mutator strain, such as Stratagene's XL1-Red. XL1-red is an E.coli strain whose deficiency in three of the primary DNA repair pathways (mutS, mutD and mutT) causes it to make errors during replicate of it's DNA, including the cloned plasmid.
- As a result each copy of the plasmid replicated in this strain has the potential to be different from the wild-type. One advantage of mutator strains is that a wide variety of mutations can be incorporated including substitutions, deletions and frame-shifts.
- The drawback with this method is that the strain becomes progressively sick as it accumulates more and more mutations in it's own genome so several steps of growth, plasmid isolation, transformation and re-growth are normally required to obtain a meaningful library.

4. Temporary mutator strains

- Temporary mutator strains can be built by over-expressing a mutator allele such as mutD5 (a dominant negative version of mutD) which limits the cell's ability to repair DNA lesions. By expressing mutD5 from an inducible promoter it is possible to allow the cells to cycle between mutagenic (mutD5 expression on) and normal (mutD5 expression off) periods of growth.
- The periods of normal growth allow the cells to recover from the mutagenesis, which allows these strains to grow for longer than conventional mutator strains.
- If a plasmid with a temperature-sensitive origin of replication is used, the mutagenic plasmid can easily be removed restore normal DNA repair, allowing the mutants to be grown up for

analysis/screening. An example of the construction and use of such a strain can be found. As far as I am aware there are no commercially available temporary mutator strains.

5. Insertion mutagenesis

- Finnzymes have a kit that uses a transposon-based system to randomly insert a 15-base pair sequence throughout a sequence of interest, be it an isolated insert or plasmid.
- This inserts 5 codons into the sequence, allowing any gene with an insertion to be expressed (i.e. no frame-shifts or stop codons are cause). Since the insertion is random, each copy of the sequence will have different insertions, thus creating a library.

6. Ethyl methanesulfonate (EMS)

- It is a chemical mutagen. EMS aklylates guanidine residues, causing them to be incorrectly copied during DNA replication. Since EMS directly chemically modifies DNA, EMS mutagenesis can be carried out either in vivo (i.e. whole-cell mutagenesis) or in vitro.
- An example of in vitro mutagenesis with EMS in which a PCR-amplified gene was subjected to reaction with EMS before being ligated into a plasmid and transformed can be found.

7. Nitrous acid

- Nitrous acid is another chemical mutagen.
- It acts by de-aminating adenine and cytosine residues (although other mechanisms are discussed) causing transversion point mutations (A/T to G/C and vice versa).
- An example of a study using nitrosoguanidine mutagenesis can be found.

8. DNA Shuffling

- It is a very powerful method in which members of a library (i.e. copies of same gene each with different types of mutation) are randomly shuffled.
- This is done by randomly digesting the library with DNAseI then randomly re-joining the fragments using self-priming PCR.
- Shuffling can be applied to libraries produced by any of the above method and allows the effects of different combinations of mutations to be tested.

Site directed mutagenesis

- Site-directed mutagenesis is an invaluable tool to modify genes and study the structural and functional properties of a protein, based on the structure, function, catalytic mechanism, and catalytic residues of enzymes.
- Site-directed mutagenesis includes single and combinational mutations. It is usually analyzed by bioinformatic methods.
- Single site-directed mutagenesis and multiple mutations have been used to expedite and simplify methods for mutagenesis.
- The properties of enzymes can be improved markedly by the combination of site-directed mutagenesis with other methods.

• For instance, the yield of maltose-binding protein-fused Hepl from recombinant E. coli was significantly improved (30.6% increase) by a thermostabilization strategy combining site-directed mutagenesis and calcium ion addition.

Methods for site-directed mutagenesis

• Site-directed mutagenesis is an in vitro method for creating a specific mutation in a known sequence. While often performed using PCR-based methods, the availability of custom-designed, synthetic, double-stranded DNA (dsDNA) fragments can drastically reduce the time and steps required to obtain the same sequence changes.

• several methods are described here- PCR-based methods for site-directed mutagenesis. Primers designed with mutations can introduce small sequence changes, and primer extension or inverse **PCR can be used to achieve longer mutant regions.**

• Using these site-directed mutagenesis techniques allows researchers to investigate the impact of sequence changes or screen a variety of mutants to determine the optimal sequence for addressing the question at hand. The IDT Mutagenesis Application Guide provides more details on these approaches.

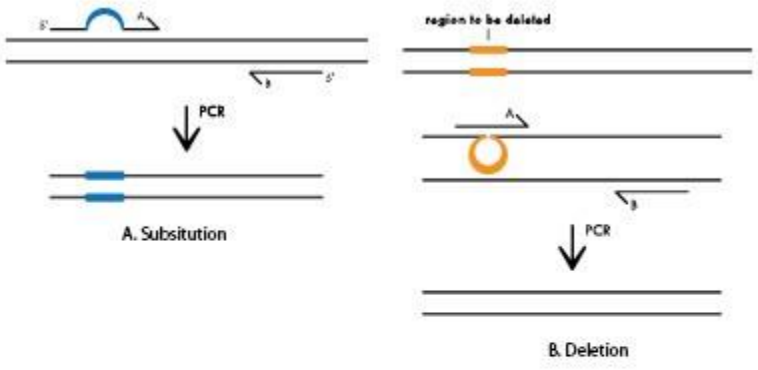
• Read our follow-up article, Site-directed mutagenesis—improvements to established methods, to learn how to use a simplified, alternative approach for generating similar mutagenesis designs quickly, with custom-designed, dsDNA fragments.

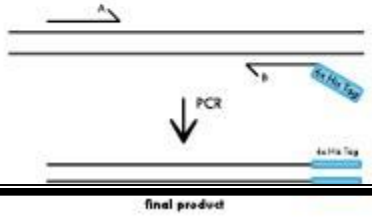
Traditional PCR

• When PCR is used for site-directed mutagenesis, the primers are designed to include the desired change, which could be base substitution, addition, or deletion.

• During PCR, the mutation is incorporated into the amplicon, replacing the original sequence.

• Mutations introduced by PCR can only be incorporated into regions of sequence complementary to the primers and not regions between the primers.





C. Terminal Substitution

Figure**: Site-directed mutagenesis by traditional PCR. Primers incorporating the desired base changes are used in PCR. As the primers are extended, the mutation is created in the resulting amplicon.**

Primer extension

- Site-directed mutagenesis by primer extension involves incorporating mutagenic primers in independent, nested PCRs before combining them in the final product.
- The reaction requires flanking primers (A and D) complementary to the ends of the target sequence, and two internal primers with complementary ends (B and C).
- These internal primers contain the desired mutation and will hybridize to the region to be altered. During the first round of PCR, the AB and CD fragments are created.
- These products are mixed for the second round of PCR using primers A and D. The complementary ends of the products hybridize in this second PCR to create the final product, AD, which contains the mutated internal sequence.
- Longer insertions can be incorporated by using especially long primers, such as IDT Ultramer™ oligonucleotides.
- To create a deletion, the internal primers, B and C, are positioned at either side of the region to be deleted to prevent it from being incorporated within fragments AB and CD from the first round of PCR.
- The complementary sequences at the ends of the these fragments, created by primers B and C, enable hybridization of AB to CD during the second round of PCR, and the final product with the desired deletion (AD) is created.

Inverse PCR

Inverse PCR enables amplification of a region of unknown sequence using primers oriented in the reverse direction. An adaptation of this method can be used to introduce mutations in previously cloned sequences. Using primers incorporating the desired change, an entire circular plasmid is amplified to delete, change or insert the desired sequence.

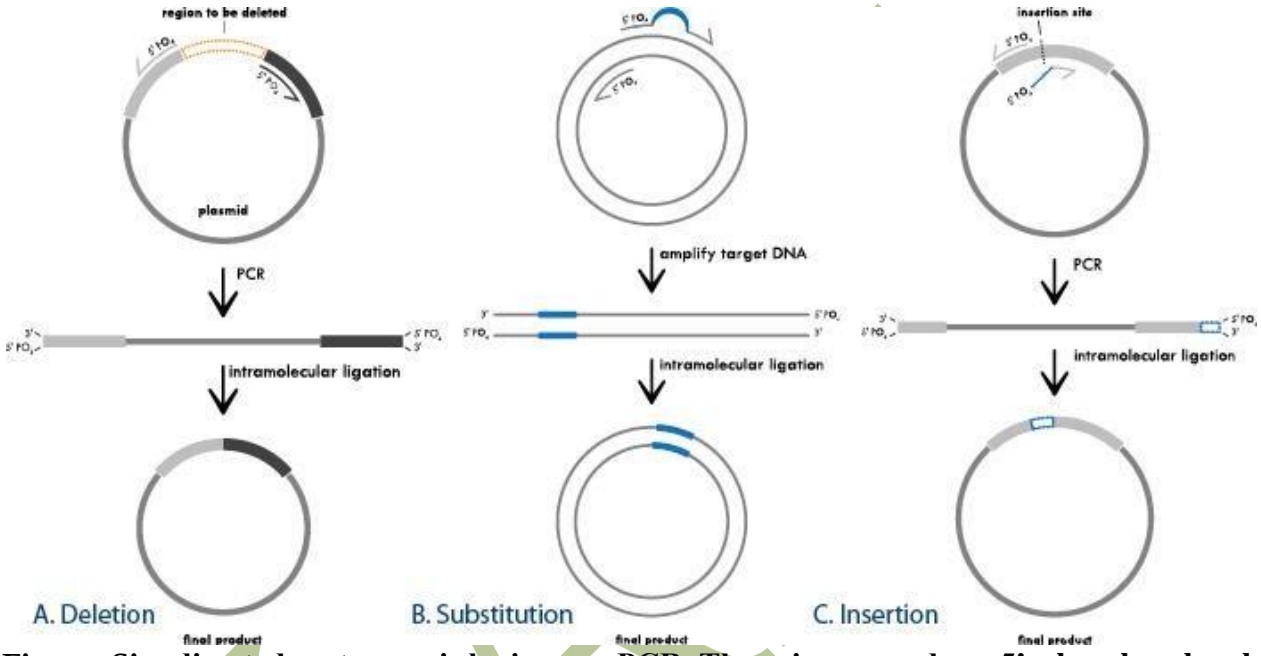


Figure: Site-directed mutagenesis by inverse PCR. The primers used are 5'-phosphorylated to allow ligation of the amplicon ends after PCR. A high fidelity DNA polymerase that creates blunt-ended products is used for the PCR to produce a linearized fragment with the desired mutation, which is then recircularized by intramolecular ligation. (A) Deletion: Primers that hybridize to regions on either side of the area to be deleted are used. (B) Substitution: One of the primers contains the desired mutation (blue bubble). (C) Insertion: The primers hybridize to regions on either side of the location of the desired insertion (black, dotted line). One primer contains the additional sequence that will be inserted (blue line).

Introduction Protein engineering

It is the design of new enzymes or proteins with new or desirable functions. It is based on the use of recombinant DNA technology to change amino acid sequences. The first papers on protein engineering date back to early 1980ies: in a review by Ulmer (1983), the prospects for protein engineering, such as X-ray crystallography, chemical DNA synthesis, computer modelling of

protein structure and folding were discussed and the combination of crystal structure and protein chemistry information with artificial gene synthesis was emphasized as a powerful approach to obtain proteins with desirable properties (Ulmer, 1983). In a later review in 1992, protein engineering was mentioned as a highly promising technique within the frame of biocatalyst engineering to improve enzyme stability and efficiency in low water systems (Gupta, 1992). Today, owing to the development in recombinant DNA technology and high-throughput screening techniques, protein engineering methods and applications are becoming increasingly important and widespread. In this Chapter, a chronological review of protein engineering methods and applications is provided.

Protein engineering methods

- Many different protein engineering methods are available today, owing to the rapid development in biological sciences, more specifically, recombinant DNA technology.
- These methods are chronologically reviewed in this section, and summarized in Table. The most classical method in protein engineering is the so-called "rational design" approach which involves "site-directed mutagenesis" of proteins (Arnold, 1993).
- Site-directed mutagenesis allows introduction of specific amino acids into a target gene. There are two common methods for site-directed mutagenesis. One is called the "overlap extension" method. This method involves two primer pairs, where one primer of each primer pair contains the mutant codon with a mismatched sequence.
- These four primers are used in the first polymerase chain reaction (PCR), where two PCRs take place, and two double-stranded DNA products are obtained. Upon denaturation and annealing of them, two heteroduplexes are formed, and each strand of the heteroduplex involves the desired mutagenic codon. DNA polymerase is then used to fill in the overlapping 3' and 5' ends of each heteroduplex and the second PCR takes place using the nonmutated primer set to amplify the mutagenic DNA. The other site-directed mutagenesis method is called "whole plasmid single round PCR". This method forms the basis of the commercial "QuikChange Site-Directed Mutagenesis Kit" from Stratagene.
- It requires two oligonucleotide primers with the desired mutation(s) which are complementary to the opposite strands of a double-stranded DNA plasmid template.
- Using DNA polymerase PCR takes place, and both strands of the template are replicated without displacing the primers and a mutated plasmid is obtained with breaks that do not overlap.
- DpnI methylase is then used for selective digestion to obtain a circular, nicked vector with the mutant gene. Upon transformation of the nicked vector into competent cells, the nick in

the DNA is repaired, and a circular, mutated plasmid is obtained (Antikainen & Martin, 2005).

Rational design

- It is an effective approach when the structure and mechanism of the protein of interest are well-known. In many cases of protein engineering, however, there is limited amount of information on the structure and mechanisms of the protein of interest.
- Thus, the use of "evolutionary methods" that involve "random mutagenesis and selection" for the desired protein properties was introduced as an alternative approach. Application of random mutagenesis could be an effective method, particularly when there is limited information on protein structure and mechanism. The only requirement here is the availability of a suitable selection scheme that favours the desired protein properties (Arnold, 1993).
- A simple and common technique for random mutagenesis is "saturation mutagenesis". It involves the replacement of a single amino acid within a protein with each of the natural amino acids, and provides all possible variations at that site. "Localized or region-specific random mutagenesis" is another technique which is a combination of rational and random approaches of protein engineering.
- It includes the simultaneous replacement of a few amino acid residues in a specific region, to obtain proteins with new specificities. This technique also makes use of overlap extension, and the whole-plasmid, single round PCR mutagenesis, as in the case of site-directed mutagenesis. However, the major difference here is that the codons for the selected amino acids are randomized, such that a mixture of 64 different forward and 64 different reverse primers are used, based on a statistical mixture of four bases and three nucleotides in a randomized codon (Antikainen & Martin, 2005).

Cell-free translation systems

- Described as important tools for protein engineering and production. They are an alternative to in vivo protein expression. When template DNA or mRNA is added to a reaction mixture, proteins are produced upon incubation in the absence of cells. PCR products can be used, and proteins are synthesized from cDNA rapidly.
- Cell-free translation systems are based on the ribosomal protein system of cells, which is provided as a cell extract from Escherichia coli etc. obtained as a supernatant upon centrifugation at 30'000 g.
- This supernatant contains necessary compounds for protein synthesis, such as ribosomes, t-RNAs, translation factors and aminoacyl-tRNA synthetases.

- Potential applications involve production of biologically active proteins, synthesis of membrane proteins for minimal cells, and artificial proteins.
- With further development, cellfree translation systems could be a strong alternative to in vivo protein expression, due to their high level of controllability and simplicity. The limitations of recombinant protein expression in living cells, such as protein degradation and aggregation will also be avoided (Shimizu et al., 2006).

Receptor-based QSAR methods

- They are also valuable for protein engineering studies. These methods are based on a computational combination of structure-activity relationship analysis and receptor structure-based design. They provide valuable pharmacological information on therapeutic targets.
- The Comparative Binding Energy (COMBINE) analysis, for example, probes bioactivity changes with respect to amino acid variations in a series of homologous protein receptors and with respect to conformational changes within a protein of interest (Lushington et al., 2007).
- The following table shows summary of various protein engineering techniques.

Method name	Reference(s)
Rational design	(Arnold, 1993)
Site-directed mutagenesis	(Arnold, 1993), (Antikainen & Martin, 2005)
Evolutionary methods/directed evolution	(Arnold, 1993)
Random mutagenesis	(Antikainen & Martin, 2005), (Wong *et al.*, 2006), (Jackson *et al.*, 2006), (Labrou, 2010)
DNA shuffling	(Antikainen & Martin, 2005), (Jackson *et al.*, 2006)
Molecular dynamics	(Anthonsen *et al.*, 1994)
Homology modeling	(Anthonsen *et al.*, 1994)
'MolCraft'*in vitro* protein evolution systems	(Shiba, 2004)
Computational methods (computational protein design)	(Jackson *et al.*, 2006), (Van der Sloot *et al.*, 2009), (Golynskiy & Seelig, 2010)
Receptor-based QSAR methods	(Lushington *et al.*, 2007)
NMR	(Anthonsen *et al.*, 1994)
X-ray crystallography	(Jackson *et al.*, 2006)
Peptidomimetics	(Venkatesan & Kim, 2002)
Phage display technology	(Antikainen & Martin, 2005), (Sidhu & Koide, 2007), (Chaput *et al.*, 2008)
Cell surface display technology	(Antikainen & Martin, 2005), (Gai & Wittrup, 2007), (Chaput *et al.*, 2008)
Flow cytometry / Cell sorting	(Mattanovich & Borth, 2006)
Cell-free translation systems	(Shimizu *et al.*, 2006)
Designed divergent evolution	(Yoshikuni & Keasling, 2007)
Stimulus-responsive peptide systems	(Chockalingam *et al.*, 2007)
Mechanical engineering of elastomeric proteins	(Li, 2008)
Engineering extracellular matrix variants	(Carson & Barker, 2009)
Traceless Staudinger ligation	(Tam & Raines, 2009)
De novo enzyme engineering	(Golynskiy & Seelig, 2010)
mRNA display	(Golynskiy & Seelig, 2010)

Possible Questions:

1. What is mutagenesis? Explain its types?
2. Give an detailed about gene shuffling.
3. What are chimeric proteins? Explain its production.
4. Elaborate random mutagenesis.
5. Explain about site directed mutagenesis.
6. What are the steps involved in production of chimeric proteins?
7. Explain about protein engineering concepts with one example.
8. Elaborate about PCR based methods of site directed mutagenesis.
9. What is gene shuffling and explain in detailed about chimeric proteins.
10. What is protein engineering and explain its concepts with two examples.

Unit V
SYLLABUS

In plants: use of *Agrobacterium tumefaciens* and *A. rhizogenes*, Ti plasmids, Strategies for gene transfer to plant cells, Direct DNA transfer to plants, Gene targeting in plants, Use of plant viruses as episomal expression vectors. In animals: Production and applications of transgenic mice, role of ES cells in gene targeting in mice, Therapeutic products - blood proteins, human hormones, immune modulators and vaccines (one example each). Ethical, legal and social issues.

Agrobacterium tumefacien

Agrobacterium tumefaciens, the cause of the economically important disease, crown gall, has also been studied for years because of its remarkable biology. The mechanism this bacterium uses to parasitize plant tissue involves the integration of some of its own DNA into the host genome resulting in unsightly tumors and changes in plant metabolism. *A. tumefaciens* prompted the first successful development of a biological control agent and is now used as a tool for engineering desired genes into plants.

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 anundifferentiated 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. 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 sequencewhich 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.

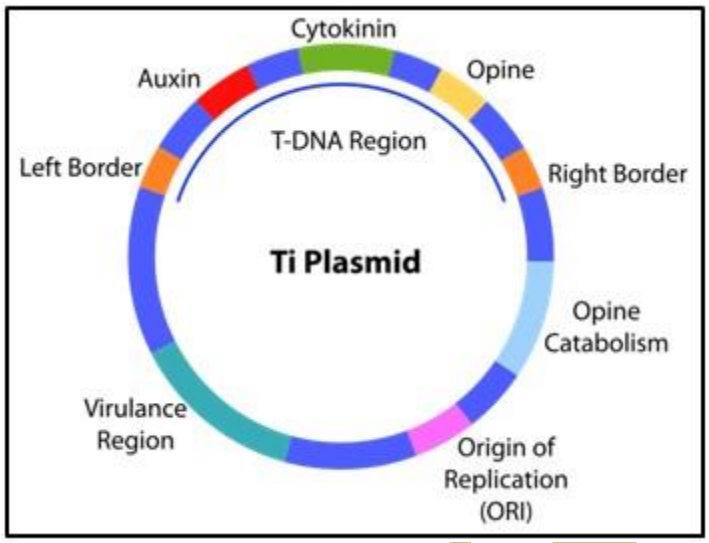


Figure: The structure of Ti Plasmid.

Organization of T-DNA

The transfer DNA (T-DNA) is the transferred DNA of the tumour inducing plasmid (pTi) of some*Agrobacterium* species of bacteria. T-DNA has both its side 24 kb direct repeat border sequence and contains the gene for tumor / hairy root induction and also for opines biosynthesis (Figure 25.2). pTi has three genes, two of these genes (*iaaM* and *iaaH*) encode enzymes which together convert tryptophane in to IAA (Indol-3-acetic acid) a type of auxin. If these two genes are deleted then shooty crown gall will produce. Therefore, the locus was earlier called „shooty locus' and the genes were designated as *tms* 1 (tumour with shoots) and *tms* 2. The third gene, *ipt*, encodes an enzyme which produces Zeatin-type cytokinin isopentenyl adenine. The deletion of *ipt*, causes rooty crown galls and the region was earlier designated as „rooty locus' and denoted by *tmr* (tumour having roots). In addition to these, another locus called *tml* and the deletion of which results in large tumours. Besides, T-DNA also contains genes involved in opine biosysnthesis which are located near the right border of T-DNA.

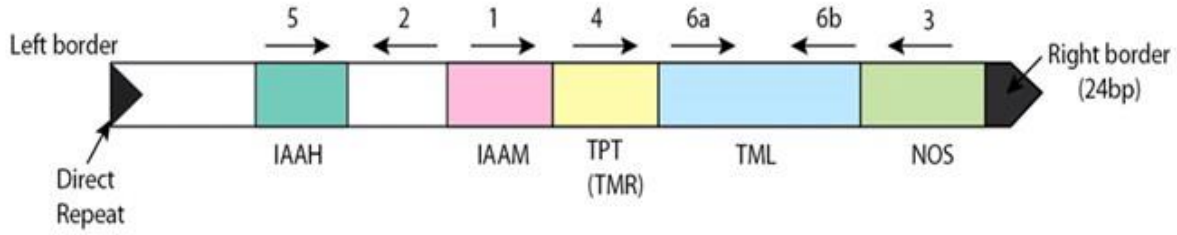


Figure 25.2: Nopaline type Ti plasmid T-DNA (Arrows indicating the direction of transcription and number indicates the transcriptional unit)

T-DNA transfer and integration

The steps involved in T-DNA transfer and integration in to the plant genome are explained in Figure 25.3.

Wounded plant cell releases phenolics substances and sugars (1); which are sensed by *vir* A, *vir* A activates *vir* G, *vir* G induces expression of *vir* gene of Ti-plasmid (2); *vir* gene produce all the *vir* - protein (3); *vir* D_1 and *vir* D_2 are involve in ssT-DNA production from Ti-plasmid and its export (4) and (5); the ssT-DNA (with associated *vir* D_1 and *vir* D_2) with *vir* E_2 are exported through transfer apparatus *vir* B (6); in plant cell, T-DNA coated with *vir* E_2 (7); various plant proteins influence the transfer of T-DNA + *vir* D_1 + *vir* D_2 + *vir* E_2 complex and integration of T-DNA to plant nuclear DNA(8). (LB= left border; RB= Right border; pTi = Ti plasmid, NPC = nuclear pore complex)

Agrobacterium rhizogenes, the causative agent of hairy root disease, is a soil dwelling gram negative bacterium capable of entering a plant through a wound and causing a proliferation of secondary roots. The mechanism of transformation is elaborated in below mentioned Figure. The biosynthetic capacity of the hairy root cultures is equivalent or sometimes more to the corresponding plant roots. Therefore, hairy root cultures have been developed as an alternate source for the production of root biomass and to obtain root derived compounds.

Establishment of hairy root cultures

For the production of hairy root cultures, the explant material is inoculated with a suspension of *A. rhizogenes* . The bacterial suspension is generated by growing bacteria in Yeast Mannitol Broth (YMB) medium for 2 days at 25°C under shaking conditions. Thereafter, pelleting by centrifugation (5 x 10 rpm; 20 min) and resuspending the bacteria in YMB medium to form a thick suspension (approx. 10^{10}viable bacteria/ml). Transformation may be inducesd in aseptic seedlings or surface sterilized detached leaves, leaf-discs, petioles, stem segments, from greenhouse

grown plants by scratching the leaf midrib or the stem of a plantlet with the needle of a hypodermic syringe containing a small (about 5-10 ul) droplet of thick bacterial suspension of *A. rhizogenes.*

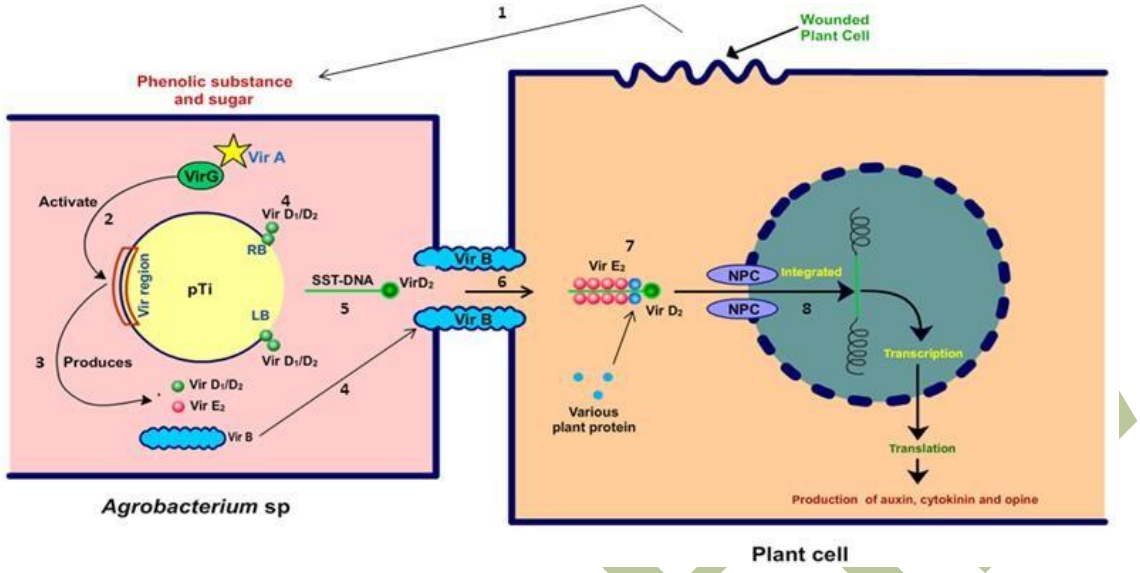


Figure : The *Agrobacterium* injects a plasmid (naked circular DNA) into the host cells

Wounded plant cell releases phenolic substances and sugar (1); which are sensed by *Vir* A, *Vir* A activates *Vir* G, *Vir* G induced for expression of *Vir* gene of Ri-plasmid (2); *Vir* gene produces all the *Vir*-protein (3); *Vir* D_1 and *Vir* D_2 are involved in ssT-DNA production from Ri-plasmid and its export (4) and (5); the ssT-DNA (associated with *Vir* D_1 and *Vir* D_2) with *Vir* E_2 are exported through transfer apparatus *Vir* B (6); in plant cell, T-DNA coated with *Vir* E_2 (7); various plant proteins influence the transfer of T-DNA + *Vir* D_1 + *Vir* D_2 + *Vir* E_2 complex and integration of T-DNA to plant nuclear DNA(8). (LB= left border; RB= Right border; pRi = Ri plasmid, NPC = nuclear pore complex)

Genetics of transformation

Ri plasmids contain one or two regions of T-DNA and a *Vir* (Virulence) region, all of which are necessary for tumorigenesis. The Ri plasmid is very similar to Ti plasmid except that their T-DNAs have homology only for auxin and opine synthesis sequences. The T-DNA of Ri plasmid lacks genes for cytokinin synthesis. The T-regions of Ti and Ri plasmids contain oncogenes that are expressed in the plants. Another type, present in Ri plasmids only, appears to impose a high hormone sensitivity on the infected tissue.

The T-DNA of Ri plasmids codes for at least three genes that each can induce root formation, and that together cause hairy root formation from plant tissue. Current results indicate that the products of these genes induce a potential for increased auxin sensitivity that is expressed when the transformed cells are subjected to a certain level of auxin. After this stage the transformed roots can be grown in culture without exogenous supply of hormones.

The Ri-plasmids are classified into two main classes according to the opines formed in transformed roots. First, agropine-type strains induce roots to synthesise agropine, mannopine and the related acids. Second, mannopine-type strains which induce roots to produce mannopine and the related acids. The agropine-type Ri-plasmids are very similar as a group and a quite distinct group from the mannopine-type plasmids. Perhaps the most studied Ri-plasmids are agropine-type strains, which are considered to be the most virulent and, therefore, more often used in the establishment of hairy root cultures.

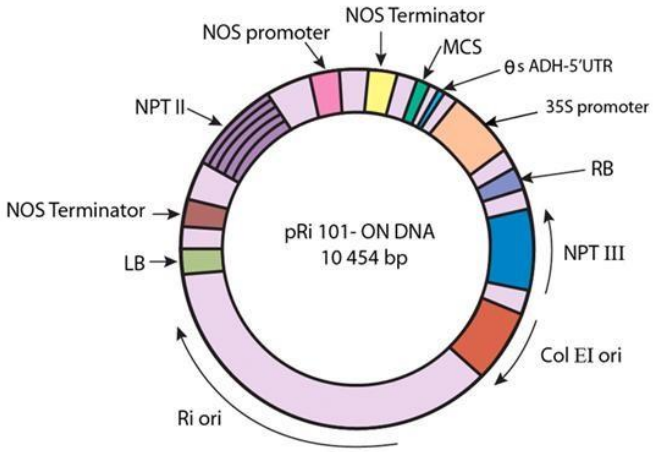


Figure : Structure of Ri-plasmid

The genes responsible for hairy root formation

The agropine-type Ri-plasmid consists of two separate T-DNA regions known as the TL-DNA and TR-DNA. Each of the T-DNA fragments is separated from each other by at least 15 kb of non-integrated plasmid DNA. These two fragments can be transferred separately during the infection procedure. The TR-DNA of the agropine type Ri-plasmid carries genes encoding auxin synthesis (*tms* 1 and *tms* 2) and agropine synthesis (*ags*).

The mannopine type Ri-plasmids contain only one T-DNA. TL-DNA region consists of four root locus (*rol*) genetic loci, *rol* A, *rol* B, *rol* C, and *rol* D, which affect hairy root induction. In

particular, *rol* B seems to be the most important in the differentiation process of transformed cells and also function as induction of hairy roots by hydrolyzing bound auxins leading to an increase in the intracellular levels of indole-3-acetic acid. Gene *rol* A involved in development of hairy root morphology, *rol* B is responsible for protruding stigmas and reduced length of stamens; *rol*C causes internode shortening and reduced apical dominance.

Factors influencing the transformation

Following factors influence the transformation process:

1. Virulence of *A. rhizogenes* strains
2. Medium
3. Age of the explant
4. Nature of the explant

Confirmation of transformation

Confirmation of transformation can be performed on the basis of following markers:

- Biochemical markers
- Opines
- Mannopines
- Genetic markers
- Southern hybridization
- Polymerase chain reaction

Screening of transformation

Screening of transformation can be performed by GUS assay, leaf callus assay, rooting and bleaching assays.

Properties of hairy roots

Hairy roots

Hairy roots have following properties

- high degree of lateral branching
- profusion of root hairs
- absence of geotropism
- they have high growth rates in culture, due to their extensive branching, resulting in the presence of many meristems.
- they do not require conditioning of the medium .

Hairy roots are genetically stable

Hairy roots are genetically stable consequently they exhibit biochemical stability that leads to stable and high-level production of secondary metabolites. Hairy root cultures apparently retain diploidy in all species so far studied. The stable production of hairy root cultures is dependent on the maintenance of organized states. The factors which promote disorganization and callus formation depress secondary metabolite production. The productivity of hairy root cultures is stable over many generations in contrast to disorganized cell cultures. This stability is reflected in both the growth rate and the level pattern of secondary metabolite production.

Application of hairy root cultures

Production of secondary metabolites

The hairy root system is stable and highly productive under hormone-free culture conditions. The fast growth, low doubling time, easy maintenance, and ability to synthesize a range of chemical compounds of hairy root cultures gives additional advantages as continuous sources for the production of plant secondary metabolites. Usually root cultures require an exogenous phytohormone supply and grow very gradually, resulting in the poor or insignificant synthesis of secondary metabolites.

Hairy roots are also a valuable source of photochemical that is useful as pharmaceuticals, cosmetics, and food additives. These roots synthesize more than a single metabolite; prove economical for commercial production purposes. Many medicinal plants have been transformed successfully by *A. rhizogenes* and the hairy roots induced show a relatively high productivity of secondary metabolites, which are important pharmaceutical products. Sevon has summarized the most important alkaloids produced by hairy roots, including *Atropa belladonna* L.,*Catharanthus trichophyllus* L., and *Datura candida* L.

Metabolic engineering offers new perspectives for improving the production of secondary metabolites by the over expression of single genes. This approach may lead to an increase of some enzymes involved in metabolism and, consequently, results in the accumulation of the target products.

This method utilizes the foreign genes that encode enzyme activities not normally present in a plant. This may cause the modification of plant metabolic pathways. Two direct repeats of a bacterial lysine decarboxylase gene, expressed in the hairy roots of *Nicotiana tabacum*, have markedly increased the production of cadaverine and anabasine (Feckeret al . 1993). The production of

anthraquinone and alizarin in hairy roots of *Rubia peregrina* L. was enhanced by the introduction of isochorismate synthase. *Catharanthus roseus* hairy roots harboring hamster 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) cDNA without the membrane-binding domain were found to produce more ajmalicine and catharanthine or serpentine and campesterol than the control.

Production of compounds not found in untransformed roots

Transformation may affect the metabolic pathway and produce new compounds that cannot be produced normally in untransformed roots. For example, the transformed hairy roots of *Scutellaria baicalensis* Georgi accumulated glucoside conjugates of flavonoids instead of the glucose conjugates accumulated in untransformed roots.

Changing composition of metabolites

Bavage et al. (1997) reported the expression of an *Antirrhinum* dihydroflavonol reductase gene which resulted in changes in condensed tannin structure and its accumulation in root cultures of *L. corniculatus*. The analysis of selected root culture lines indicated the alteration of monomer levels during growth and development without changes in composition.

Table: Pharmaceutical products produced using hairy root cultures

Plant species	Product
Bidens spp.	Polyacetylenes
Cinchona ledgeriana	Quinoline alkaloids
Datura spp.	Tropane
Cassia spp.	Anthraquinones
Echinacea purpurea	Alkaloids

GENE TARGETING IN PLANTS

- Targeting a single copy of a transgene into a pre-determined plant genomic location provides an efficient tool for securing long term stable expression.
- A gene targeting protocol may include: a) the ability to insert the transgene into regions of chromatin compatible with a desired developmental and tissue specific expression, and avoiding the knock-out of endogenous genes, b) the ability to control copy number, and/or c) subsequent removal of trans gene repeats, selection marker genes, and other undesired DNA sequences.

- Targeted transgenes should therefore provide a more efficient and informative means of securing and comparing the expression of various transgenic sequences than is available with current transgenic procedures.
- Applications of gene targeting may also be extended to the inactivation of undesired traits, modification of existing genes, and elucidation of gene function through the analysis of gene malfunction. As the sequencing projects are progressing in plants, the creation of null mutants by targeting of specific sequences would be a direct way of relating sequences to the genes, and subsequently genes to phenotypes.

Plant Viral vectors

- Exploitation of plant viruses as transformation vectors by massive infection may be harmful and even deleterious to the target plants. It is still however able to express and produce foreign proteins.
- Plant viruses must exhibit some of the exemplary features before they are considered as vectors. They should extend their broader host-range, spread of seed transmission and carry additional copies of gene of interest.
- Several viral vectors require suitable modification in order to accommodate extra nucleic acid and also aggressive in infection process. Although several groups of viruses have been identified, some moderate progresses have been made only in two groups. These two groups are Caulimo virus and Gemini virus, which have DNA genome as genetic material.

Cauliflower mosaic virus

- Cauliflower mosaic virus (CamV) belongs to the group caulimovirus, can be used as potential candidate to deliver foreign gene into the plant. It is perhaps the best studied viruses among plant virus, which infects several members belonging to Cruciferae family. Cauliflower mosaic virus contains circular double helical DNA as genetic material.

- As an infective agent, can cause disease in wide range of commercially important cultivated crops. Cauliflower mosaic DNA has been subjected to a wide range of manipulation. This was the only and first virus to be manipulated and used as a favourable choice for genetic engineering work. Elucidation of 8 kb CamV reveals that, it contains six major and two minor reading frames (Fig. 14.12).

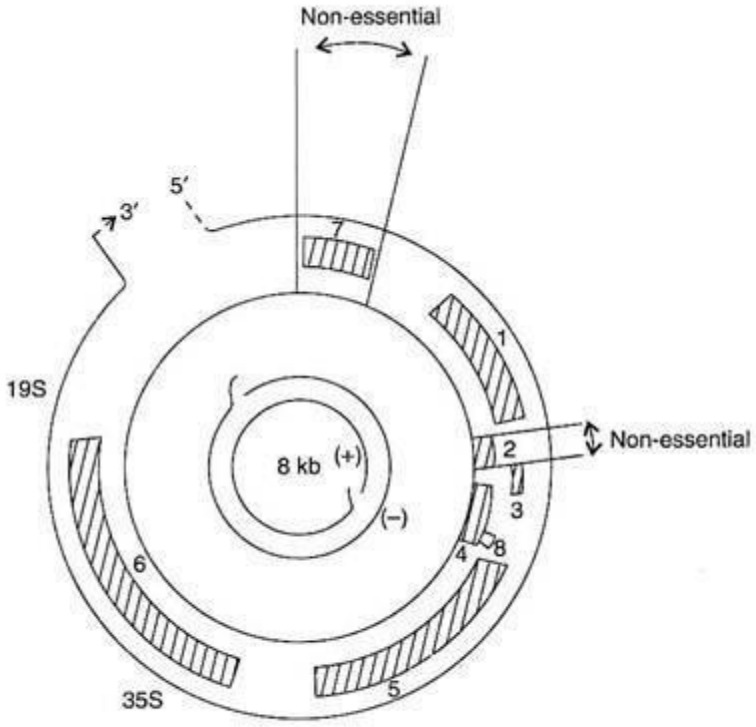


Fig. 14.12. Genetic map of cauliflower mosaic virus.

- Presence of ribonucleotide in DNA leads to the conclusion that CamV replication involves the synthesis of negative DNA strand by employing reverse transcriptase and followed by synthesis of positive DNA strand. Once the synthesis of double strand DNA completes, it is then packed into viral particles and continue the cycles of transcription and translation.

Gemini Virus

It is a DNA virus, known to infect wide range of economically important and agriculture crops of monocotyledonous and dicotyledonous plants. Several diseases such as maize streak virus and curly top virus are caused by Gemini virus. The genome is single Stranded Circular DNA and its replication takes place by DNA immediately.

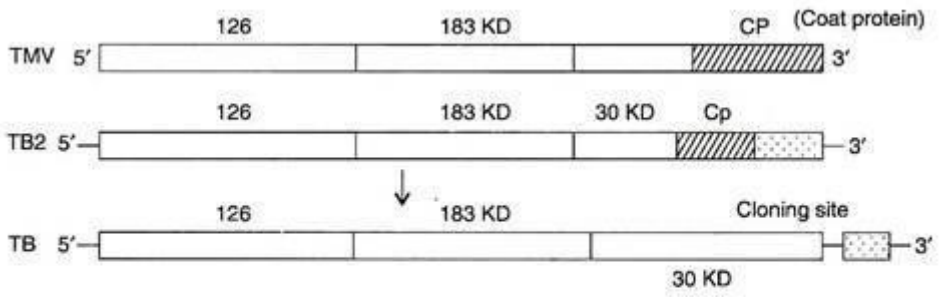
Tobacco Mosaic Virus Based Expression Vector:

Tobacco mosaic virus (TMV) is a RNA virus and shows several advantages by designing expression vector. TMV was the first virus to be purified and sequenced. As far as biohazard is considered TMV could be used as a comparatively safe recombinant virus in the field.

The coat protein of the TMV is one of the most accumulated proteins in plants reachable upto 10% of the dry weight of infected plant. Approximately, under ideal condition, 2000 kg tobacco protein can be produced per acre per year. Moreover, TMV can be purified in crystalline form in substantial quantity by simple methodology.

The single stranded RNA genetic material of TMV encloses 6300 nucleotides with four open reading frames (Fig. 14.13). The filamentous nature of the RNA virus is determined by the length of the viral nucleic acid. Both 183 kD read through protein and 126 kD coat proteins are translated from the 5′ end of the genomic RNA.

These two proteins form replicase complex. In addition to these two proteins, the 30-kD movement protein and 17.5-kD capsid proteins are translated at 3′ region of sub-genomic mRNA during replication.



14.13 Genomic organization of TMV vector (2 and 3)

TMV can be subjected to a wide range of manipulation by replacing the viral coat protein with a foreign protein, for example, replacement of coat protein with reporter gene chloramphenical acetyl transferase (CAT) resulted in a free-RNA virus that generated high CAT activity. Improvisation of TMV vector was achieved by placing CAT gene under the control of a coat protein sub-genomic promoter of TMV into the entire TMV genome.

This recombinant virus produced third sub-genomic mRNA and this was translated into CAT enzyme with high activity. Similarly, another TMV hybrid expression vector TB_2 was designed in which coat protein gene and neomycin phosphotransferase marker gene was placed under the control of sub-genomic promoters of TMV coat protein and ORSV, respectively.

This was referred as extra gene vector. TB_2 effectively produced the foreign protein without any major constraints. Another extra gene in TMV based vector, 4GD-PL, was developed from tomato green mosaic virus. The 4GD-PL vector was able to express foreign proteins systematically throughout plants.

All these studies demonstrated that proximity of genes to the 3′-untranslated region of the genome increases efficiency of their translation. Another improved TMV vector, 30B, was designed in which the start codon (AUG) of the capsid protein was mutated to AGA, and restriction cloning sites were engineered (40 nucleotide) to provide a full-size sub-genomic RNA promoter.

Possibility of satellite RNA to be used as vector has been considered. They vary in their size between 0.27 and 1.6 kb. They are not indispensable for virus replication. However, their functions can alter pathogenecity of virus.

Cow Pea Mosaic Virus Expression Vector:

Cow pea mosaic virus (CpmV) is also a RNA virus and infects species of legumes. There are two separate positive strand-RNA molecules present in the genetic material of CpmV. The number of nucleotides present in the RNA I and RNA II strand is 5889 and 3480, respectively. Although RNA I alone can replicate on its own but both RNAs are indispensable for infectivity.

The proteins involved in the replication of the virus are encoded by RNA I whereas movement proteins are encoded by RNA II. CpmV capside of both large (L) and small (S) coat protein of 30 copies each are in isohedral symmetry. The two capsid proteins are folded into three antiparallel β-barrel structures.

In the construction of CpmV expression vector, preference was given to the replacement of stable chimeras by insertion of foreign sequences rather than replacement for native residues.

Therefore, in the construction of viable and well refined CpmV vector, precise site of insertion of foreign sequence was given a prime choice by introducing foreign DNA sequence into βB-βC loop of the S protein for most chimeras foreign sequences inserts immediately upstream of proline 23 of the S protein. In view of propagating the chimeras, engineered pCP_2 and pCP_1 are linearised and inoculated on cow pea plants. (Fig. 14.14)

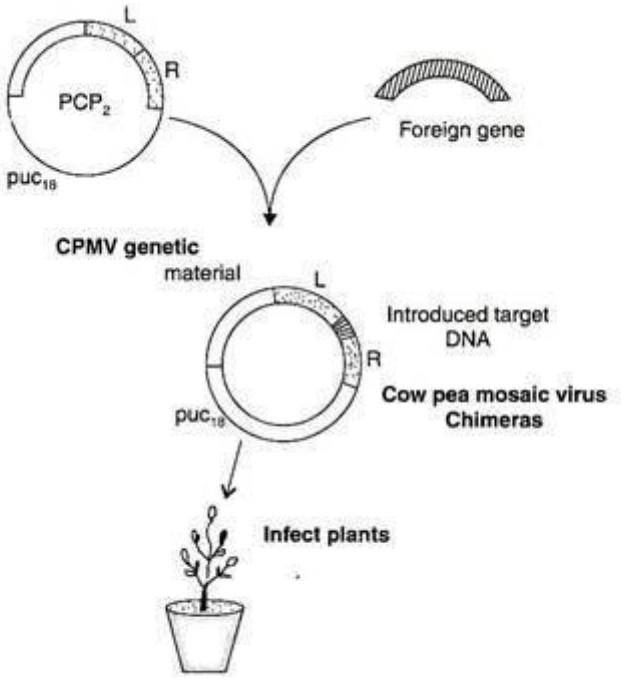


Fig. 14.14 CpmV expression vector construction and infection of plants

Practical Applications of DNA Technology

Medicine and the Pharmaceutical Industry

Diagnosis of diseases

- It is possible to clone entire DNA molecules or genes associated with various diseases (e.g., HIV DNA; genes associated with hemophilia, cystic fibrosis, Huntington‟s disease, and Duchenne muscular dystrophy)

- These can be cloned in vectors or by PCR to obtain large amounts, and then used to produce probes Probes can be used to detect infection (e.g., HIV) or the presence of a defective allele.
- Thus, it is possible to identify individuals carrying certain disease-related alleles even if they show no symptoms of the disease (e.g., heterozygous for cystic fibrosis)
- Even if a disease-causing allele has not been cloned and its precise locus is unknown, its presence can sometimes be detected by testing for restriction fragment length polymorphism (RFLP) markers that are very close to the gene in question

Human gene therapy

- For any genetic disorder traceable to a defective allele, it should theoretically be possible to replace or supplement the defective allele with a functional, normal allele using recombinant DNA techniques
- For gene therapy of somatic cells to be permanent, cells that receive the normal allele must be ones that multiply throughout the life of the patient
- Bone marrow stem cells are ideal candidates, because they continue to divide throughout life, and they give rise to cells of the blood and immune systems If, for example, the patient‟s cells lacked a particular normal gene, it might be possible to Insert the normal allele into a retrovirus
- Let the retrovirus infect bone marrow cells that have been removed from the patient and cultured
- Allow the retrovirus to insert its DNA, along with the normal allele, into the chromosomal DNA of the patient‟s bone marrow cells
- Inject the engineered cells into the patient

Pharmaceutical products

- By recombinant DNA technology discussed previously, human genes for various proteins have been engineered into expression vectors and then into bacterial host cells that produce and secrete the protein
- Examples include insulin, used to treat diabetes, and human growth hormone, used to treat hypopituitarism, which causes a form of dwarfism
- Another example is tissue plasminogen activator (TPA), which helps dissolve blood clots and reduces the risk of subsequent heart attacks if administered shortly after an initial attack
- It is also possible to construct desired molecules
- Genetically engineered proteins can block or mimic surface receptors on cell membranes; an example is a molecule designed to mimic a receptor protein that HIV binds to in entering white blood cells (if HIV binds to the drug molecules instead of those on the cell surface, it would fail to enter the blood cell)
- Recombinant DNA techniques can generate large amounts of proteins associated with the immune response against pathogens or be used to modify the genome of a pathogen to attenuate it, and thus lead to more specific and safer vaccines

Forensic, environmental, and agricultural applications

Forensic use of DNA technology

- DNA fingerprinting involving RFLP analysis by Southern blotting
- Due to previous work, many RFLP sequences are well know, and probes to them have been developed
- These select portions of the total DNA serve as markers when the probes are used to tag the corresponding DNA fragment bands and visualize them via autoradiography
- The pattern of RFLP bands is unique to each individual; therefore, one can compare DNA from a crime scene (or, perhaps, an item associated with a crime, such as clothing) to DNA from a victim and DNA from a suspect

- DNA fingerprinting involving satellite DNA analysis DNA fingerprinting is also done with satellite DNA consisting of tandemly repeated base sequences that vary in length from person to person
- The most useful are microsatellites, which are roughly 10 to 100 base pairs long, have repeating units of only a few base pairs, and are highly variable from person to person

For example, one person may have the sequence A-C-A repeated 10 times at one genome locus, 30 times at a second locus, and so on, while another person is likely to have different numbers of repeats at these loci

- Restriction fragments containing these simple tandem repeats (STRs) vary in size from one individual to another because of differences in STR lengths; therefore, they provide a unique pattern of markers for each individual
- Actually, it is possible for two individuals who are not identical twins to have identical DNA fingerprints, but the chances range from one in 100,000 to one in a billion, and that is only because in most cases relatively few markers are examined

Environmental use of DNA technology

- Considerable work is aimed at engineering organisms to cope with various environmental problems based on the ability of microorganisms to transform chemicals
- For example, the ability to extract heavy metals such as copper, lead, and nickel from the environment, which already exists in many bacteria, might lead to the development of ultra efficient genetically engineered microbes that could be used for mining minerals or cleaning up toxic mining wastes
- Likewise, genetic engineering is leading to the development of microbes that can degrade toxic compounds into nontoxic forms during the manufacturing process, at waste treatment facilities, at dump sites, and at sites of environmental disasters such as oil spills

Animal husbandry and "pharm" animals

- Farm animals are regularly treated with engineered products such as vaccines, antibodies, and hormones
- Transgenic organisms (organisms that contain genes from another species) have been developed, such as fish containing a foreign growth hormone gene, which makes them grow faster
- Transgenic animals can also be engineered to produce large amounts of an otherwise rare biological substance for medical use; this can be done such that the gene"s product is secreted in the animal"s milk.

Genetic engineering in plants

- It is possible to develop transgenic plants that contain genes for such things as resistance to herbicides, pathogens, pest insects, and freeze damage
- DNA containing the gene of interest is inserted into the T DNA area of the Ti plasmid from *Agrobacterium tumefaciens*
- The recombinant Ti plasmid is introduced into cultured plant cells where the T DNA carrying the gene of interest is inserted into the plant chromosomal DNA
- As the plant cell divides, each descendant receives a copy of the T DNA and the foreign gene
- Individual cells can then be induced to differentiate and regenerate an entire plant, all of the cells of which contain the foreign gene

Transgenic Mice

- It is designed to support investigators doing biology of aging research by creating mice that have been genetically altered by either inserting a new gene or removing a normal gene.

- This method has become one of the most exciting approaches of discovering the functions and interactions of genes in mammals. At the University Of Washington, Nathan Shock Center, this transgenic technology is used to develop new animal models for studying genetic mechanisms of the aging process.

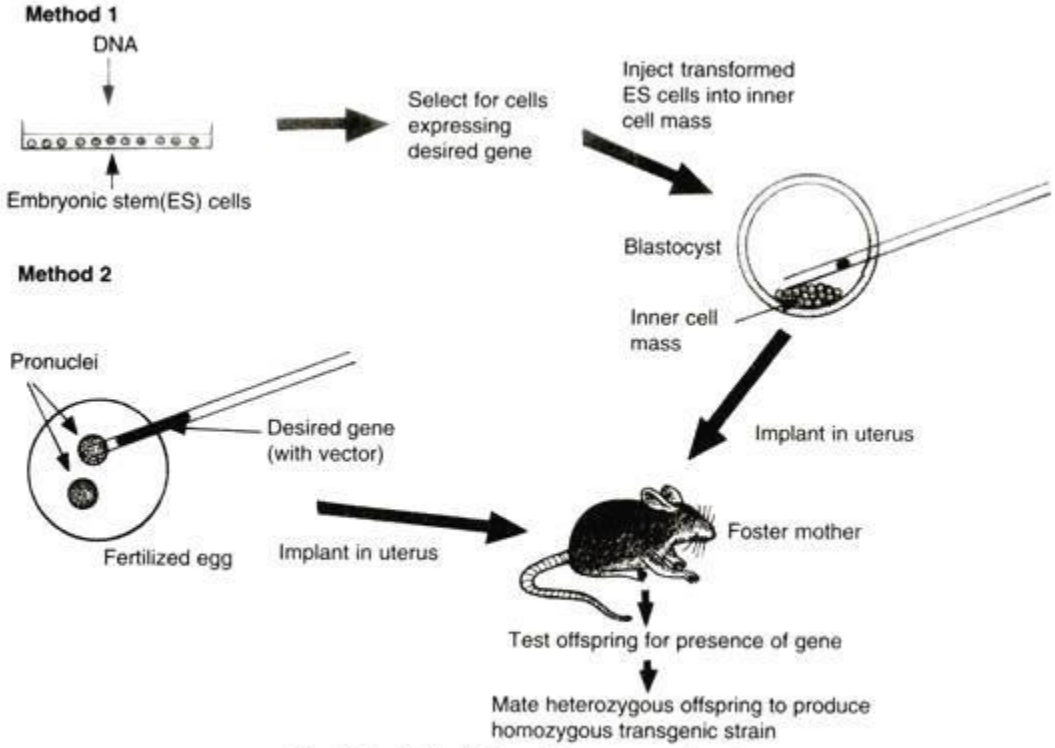


Fig. 18.4. Method of producing transgenic mice.

- During the previous year, transgenic mouse production has focused on constructs with enhanced defense against free radical injury in aging (e.g., catalase, superoxide dismutase, glutathione S-transferase), Werner Syndrome, adult onset diabetes, Alzheimer‟s disease, thrombospondin, and rheumatoid arthritis in aging. Almost 4000 embryos, mainly of the C57BL/6 ingred strain, have been transferred, 498 pups analyzed and at least 40 contained the integrated construct.

- In addition, this core concentrated an appreciable portion of effort into embryonal stem line (ES) methodologies for generation of knockouts and targeted ES transgenic. This included work to generate mouse models of Werner‟s Syndrome, models for study of presenillin genes related to Alzheimer‟s Disease and study of models of thrombospondin in aging. In the past year, a total of 396 embryos were transferred, and 79 pups were born, of which 37 were chimeric.
- The isolation of mammalian genes is of utmost importance to the biology and medicine of aging because of the contributions these studies can make to the understanding of physiology and development. Techniques for introducing foreign genes into the mouse germ line provide novel approaches for modeling human genetic and chronic degenerative diseases. Since the initial report in 1980 describing transgenic mice, methods for the direct microinjection of DNA into the pro-nuclei of fertilized embryos have become established. Foreign genes can be incorporated into somatic germ-line tissues, with expression of these elements in the progeny of founder mice.
- The creation of "transgenic" animals that make a specified gene product presents a spectrum of opportunities for basic studies in molecular pathogenesis and pre-clinical investigations applicable to a wide variety of medical problems of aging. An additional gene transfer technology developed in the 1980‟s involved the use of stem cells from the early embryo, so-called embryonic stem (ES) cells. The capacity of ES cells to undergo differentiation makes them useful for investigating the effects of genetic modifications of either the gain of function or loss of function.

Two methods of producing transgenic mice are widely used:

(1) Transforming embryonic stem cells (ES cells) growing in tissue culture with the desired DNA

(2) injecting the desired gene into the pro-nucleus of a fertilized mouse egg.

Method 1- The Embryonic Stem Cell Method:

Embryonic stem cells (ES cells) are harvested from the inner cell mass (ICM) of mouse blastocysts. They can be grown in culture and retain their full potential to produce all the cells of the mature animal, including its gametes.

1. Make your DNA-Using recombinant DNA methods, build molecules of DNA containing the structural gene you desire (e.g., the insulin gene), vector DNA to enable the molecules to be inserted into host DNA molecules, promoter and enhancer sequences to enable the gene to be expressed by host cells.

2. Transform ES cells in culture- Expose the cultured cells to the DNA so that some will incorporate it.

3. Select for successfully transformed cells.

4. Inject these cells into the inner cell mass (ICM) of mouse blastocysts.

5. Embryo transfer- Prepare a pseudo pregnant mouse (by mating a female mouse with a vasectomized male). The stimulus of mating elicits the hormonal changes needed to make her uterus receptive. Transfer the embryos into her uterus.

6. Test her offspring – Remove a small piece of tissue from the tail and examine its DNA for the desired gene. No more than 10-20% will have it, and they will be heterozygous for the gene.

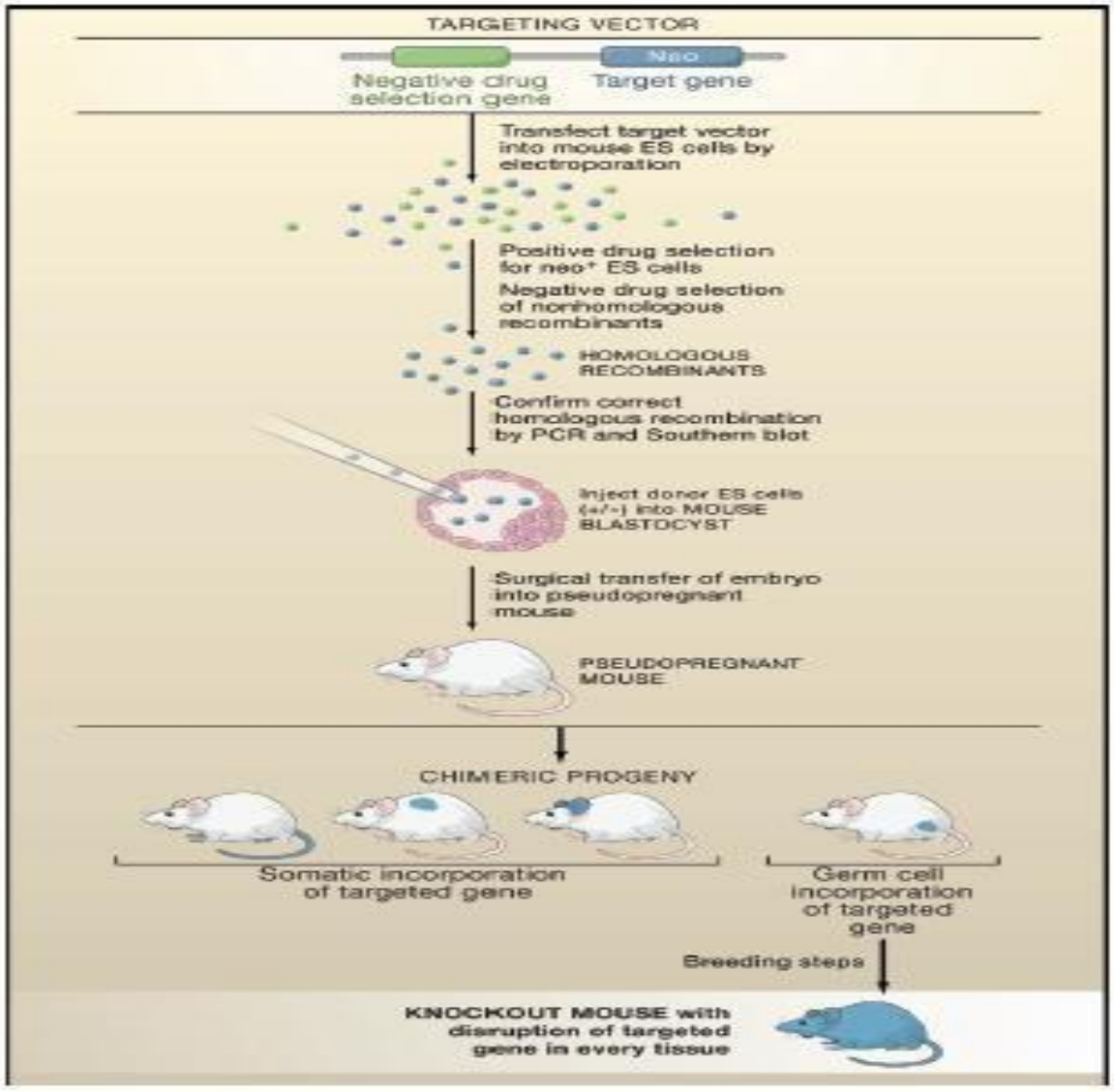
7. Establish a transgenic strain – Mate two heterozygous mice and screen their offspring for the 1:4 that will be homozygous for the transgene. Mating these will found the transgenic strain.

Method 2 -The Pro-nucleus Method:

1. DNA is prepared as in Method 1.

2. Transform fertilized eggs – Freshly fertilized eggs are harvested before the sperm head has become a pro-nucleus. The male pro-nucleus is injected with DNA. When the pro-nuclei have fused to form the diploid zygote nucleus, the zygote is allowed to divide by mitosis to form a 2-cell embryo. The embryos is implanted in a pseudo pregnant foster mother and preceded as in Method 1.

Role ES cells in gene targeting in Mice



- The development of transgenic technology, whereby genes (or mutations) can be stably introduced into the germline of experimental mammals, now allows investigators to create mice of virtually any genotype and to assess the consequences of these mutations in the context of a developing and intact mammal.
- In contrast to traditional "gain-of-function" mutations, typically created by microinjection of the gene of interest into the one-celled zygote, gene targeting via homologous

recombination in pluripotential embryonic stem cells allows one to modify precisely the gene of interest.

- The purpose of this review is to introduce the reader to the history of development of embryonic stem cell technology, the current methods employed to create "knock-out" mice, and the application of these methods to solve problems in biology.
- While the technology promises to provide enormous insight into mammalian development genetics, our desire is that this review will stimulate the application of gene targeting in embryonic stem cells to begin to unravel problems in complex regulatory pathways, specifically intermediary metabolism and physiology.

Production of plasma proteins

- During the past few years we have seen an unparalleled interest and growth in applied biology. The rapid progress in this area has been largely driven by laboratory advances in monoclonal antibody and recombinant DNA (r-DNA) techniques. Most of the early efforts to apply r-DNA technology to commercial product development were directed toward the production of small peptide hormones such as interferon, insulin and growth hormone.
- The only r-DNA produced therapeutic agent currently on the market is the human insulin preparation sold by Eli Lilly and Company. However, many observers believe that a human growth hormone preparation produced by Genentech could be available in 1985 and a number of other therapeutic agents such as tissue plasminogen activator are presently undergoing clinical trials.

Production of Growth hormone

HGH is a protein, and like all proteins, it's made from a chain of amino acid subunits. (In the case of HGH, the protein is roughly 190 amino acids long.) Before the invention of rDNA technology, HGH could only be produced laboriously by isolating it from pituitary gland tissue taken from human cadavers.

This process was inefficient, expensive and sometimes unsafe. For example, the resulting HGH product occasionally contained contaminants from cadaver tissues. Rarely, patients injected with HGH from cadavers developed Creutzfeld-Jakob disease, a very serious human version of mad cow disease. Infection is caused by proteins called prions. By eliminating the need for human tissue, rDNA technology avoids these and other potential contamination problems.

Isolation

Genes like the one for HGH contain coded instructions for protein production. Inside cells, this information is first re-coded from DNA, which provides long-term information storage, to a messenger RNA (mRNA) molecule, which provides specific instructions for HGH protein production.

Scientists begin by taking pituitary gland tissue and isolating the mRNA encoded by the HGH gene. Next, they used the mRNA as a template to create complementary DNA (cDNA). This DNA contains the coded instructions for making the HGH protein.

Transfer and Production

After scientists create the cDNA, they add it to a plasmid, a small loop of DNA taken from a bacterial cell. Next, they insert the plasmid into bacteria. When the bacteria are grown in culture, the cells use the transferred HGH gene to produce and isolate HGH much more quickly and with less effort and expense than was possible with human pituitary gland tissue. And, because the protein is produced by bacteria, contamination by components of cadaver tissue is not possible.

Recombinant Vaccine production

Vaccine generated using recombinant DNA technology is called recombinant vaccine. While there are various types of vaccines made possible by recombinant DNA technology, recombinant vaccines can be classified into two major categories.

DNA vaccines

Recombinant (protein subunit) vaccines

DNA vaccines

- These vaccines usually consist of synthetic DNA containing the gene that encodes the disease-agent protein. Usually, the plasmid DNA used as vaccine is propagated in bacteria such as E. coli and they are isolated and purified for injection.
- This "naked" DNA is usually injected intramuscularly or intradermally. The principle behind a DNA vaccine is that the antigen can be expressed directly by host cells in a way that simulates viral infection and invokes an immune response from the host.
- This is similar to GenScript's DNA Immunization Technology which is a powerful tool that aids in custom antibody production against membrane proteins, other problematic antigens, as well as for early DNA vaccine development studies.
- DNA immunization technique allows antigen production to occur in vivo, bypassing the need to produce and purify protein antigen in vitro. Click here to learn more about how GenScript's DNA immunization service can help you succeed in your next DNA vaccine development project. Schematic below illustrates concept of DNA vaccine.

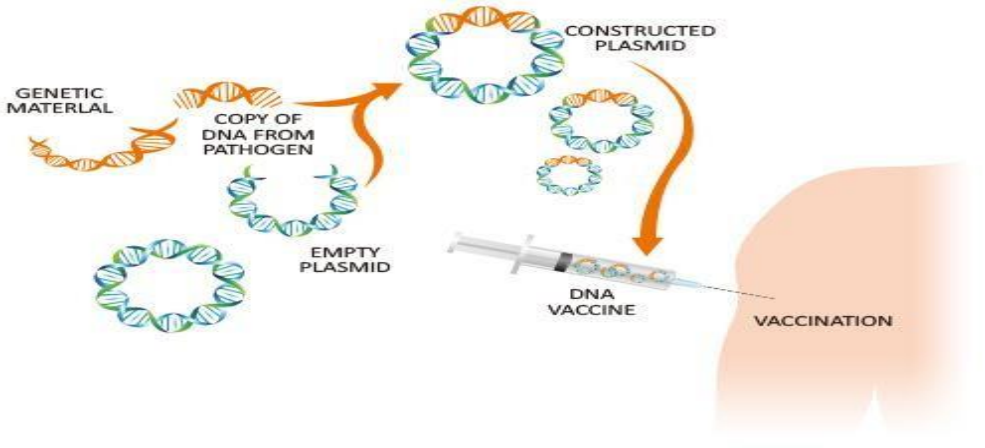


Figure : Principle of DNA vaccine

Recombinant (protein subunit) Vaccines

- These are subunit vaccines containing only a fraction of the pathogenic organism. Often time these are synthetic peptides that represent the protein component that induces an immune response.
- But they can also consist of protein subunits (antigens) expressed in a heterologous expression system (E. coli, yeast, insect etc.) using recombinant protein expression technologies.
- Most of the vaccines under investigation today are based on such purified recombinant proteins or subunits of antigens. One of the best examples of recombinant protein vaccine currently in use in humans is the vaccine against Hepatitis B Virus (HBV).

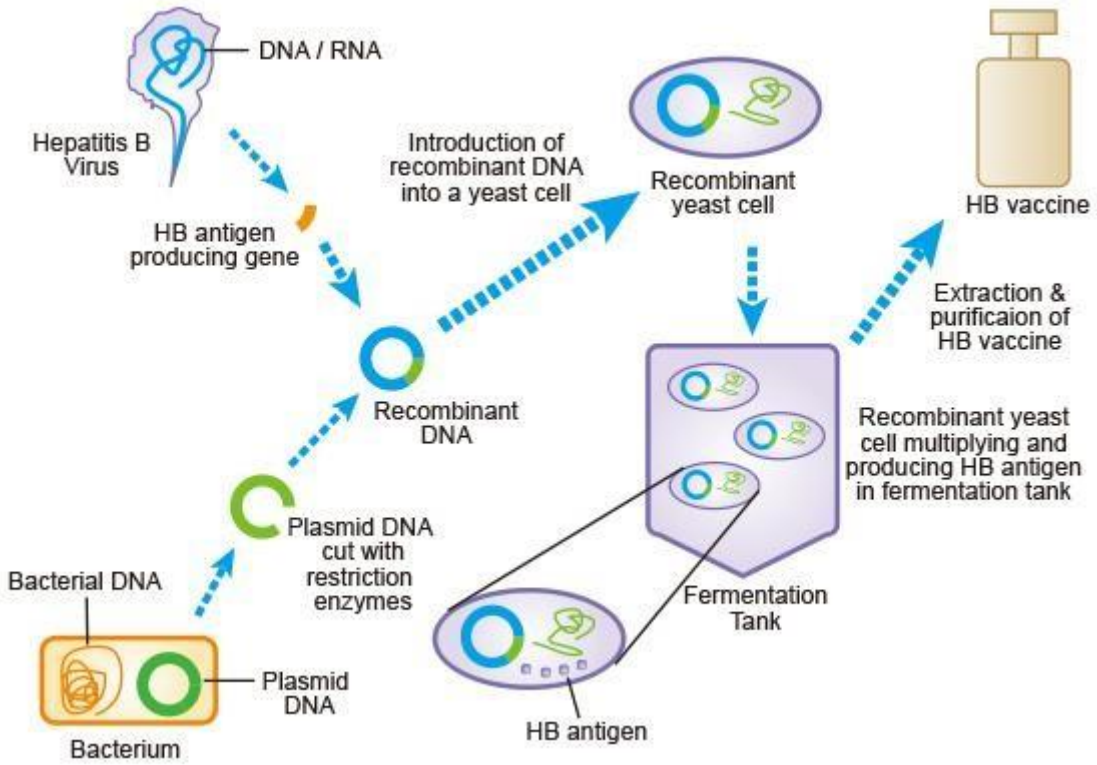


Figure : Recombinant Hepatitis B Vaccine Production Summary

Ethical issues involved in cloning

- The success rate in cloning is quite low.
- Even if we can increase the odds of success, problems can arise during the clone's development, both before and after pregnancy.
- What are the possible implications of cloning to society?
- All of us - researchers, policymakers and the public - have a responsibility to explore the potential effects of cloning technologies on our lives so that we can make informed decisions.

For each new application of cloning technologies, we must consider:

- What are the benefits?
- What are the risks?
- Whom will the technology help? Does it have the potential to hurt anyone?
- What does this mean for me? For my family? For others around me?
- Why might others not share my view?

1.1. Ethical, legal and social issues

There are several types of issues to consider as we think about cloning

- Ethical issues are those that ask us to consider the potential moral outcomes of cloning technologies
- Legal issues require researchers and the public to help policymakers decide whether and how cloning technologies should be regulated by the government.
- Social issues involve the impact of cloning technologies on society as a whole.

Pros of Cloning

- Cloning finds applications in genetic fingerprinting,
- amplification of DNA and alteration of the genetic makeup of organisms.
- It can be used to bring about desired changes in the genetic makeup of individuals thereby introducing psitive traits in them, as also for elimination of negative traits.
- Cloning can also be applied to plants to remove or alter defective genes, thereby making them resistant to diseases.
- Cloning may find applications in development of human organs, thus making human life safer. Here we look at some of the potential advantages of cloning.

Organ Replacement: If the vital organs of the human body can be cloned, they can serve as backup systems for human beings. Cloning body parts can serve as a lifesaver. When a body organ such as a kidney or heart fails to function, it may be possible to replace it with the cloned body organ.

Substitute for Natural Reproduction: Cloning in human beings can prove to be a solution to infertility. Cloning can serve as an option for producing children. With cloning, it would be possible to produce certain desired traits in human beings. We might be able to produce children with certain qualities.

Help in Genetic Research: Cloning technologies can prove helpful to researchers in genetics. They might be able to understand the composition of genes and the effects of genetic constituents on human traits, in a better manner. They will be able to alter genetic constituents in cloned human beings, thus simplifying their analysis of genes. Cloning may also help us combat a wide range of genetic diseases.

Obtain Specific Traits in Organisms: Cloning can make it possible for us to obtain customized organisms and harness them for the benefit of society. Cloning can serve as the best means to replicate animals that can be used for research purposes. Cloning can enable the genetic alteration of plants and animals. If positive changes can be brought about in living beings with the help of cloning, it will indeed be a boon to mankind.

Cons of Cloning

Like every coin has two sides, cloning has its flip side too.

- Though cloning may work wonders in genetics, it has potential disadvantages.
- Cloning, as you know, is copying or replicating biological traits in organisms.
- Thus it might reduce the diversity in nature.
- it is not clear whether we will be able to bring all the potential uses of cloning into reality.

Detrimental to Genetic Diversity: Cloning creates identical genes. It is a process of replicating a genetic constitution, thus hampering the diversity in genes. While lessening the diversity in genes, we weaken our ability of adaptation. Cloning is also detrimental to the beauty that lies in diversity.

Inivitation to Malpractices: While cloning allows man to tamper with genetics in human beings, it also makes deliberate reproduction of undesirable traits, a probability. Cloning of body organs might invite malpractices in society.

Will this Technology Reach the Common Man?: In cloning human organs and using them for transplant, or in cloning human beings themselves, technical and economic barriers will have to be considered. Will cloned organs be cost-effective? Will cloning techniques really reach the common man?

Man, a Man-made Being?: Moreover, cloning will put human and animal rights at stake. Will cloning fit into our ethical and moral principles? Cloning will make man just another man-made being. Won't it devalue mankind? Won't it demean the value of human life?

Possible questions

1. What are the steps involved in production of transgenic mice?
2. Give an detailed note about role of ES cells in gene targeting in mice.
3. What are the applications involved in transgenic mice?
4. Explain about the therauptic products of blood proteins and vaccines.
5. What are transgenic plants? Give its significance.
6. What is gene therapy? Explain in detail about its types.
7. Give ethical legal and social concern about cloning.
8. What is Bt gene? Give a detailed note on Bt cotton.
9. Give a detailed note on golden rice.
10. Explain in detail about flavr savr tomato.
11. What are GMOs. Give its advantages and disadvantages.

Reg. No.....

[08BTU22]

KARPAGAM ARTS AND SCIENCE COLLEGE

(AUTONOMOUS)

[AFFILIATED TO BHARATHIAR UNIVERSITY]

COIMBATORE – 641 021

(For the candidates admitted from 2008 onwards)

B.Sc. DEGREE EXAMINATION, NOVEMBER 2010

Fifth Semester

BIOTECHNOLOGY

RECOMBINANT DNA TECHNOLOGY

Time : 3 hours Maximum : 75 marks

PART – A (20 Marks) (30 Minutes)

(Question Nos. 1 to 20 Online Examinations)

PART B (5 X 7 = 35 Marks) (2 ½ Hours)

Answer any FIVE Questions

21. Explain the principle and applications of PCR.

22. Write short note on DNA polymerases.

23. Describe P ^{BR322}.

24. Briefly explain linkers and adaptors.

25. Explain shuttle vector.

26. Briefly describe the production of insulin in E.coli.

27. Comment on Gene therapy.

PART C (2 x 10 = 20 Marks)

Answer any TWO Questions

28. Explain Western blotting technique and its applications.

29. Write detailed notes on Plasmid and its types.

30. Illustrate Antisense technology with suitable example.

Reg. No.....

[09BTU601]

KARPAGAM UNIVERSITY

(Under Section 3 of UGC Act 1956)
COIMBATORE – 641 021
(For the candidates admitted from 2009 onwards)

B.Sc. DEGREE EXAMINATION, APRIL 2012

Sixth Semester

BIOTECHNOLOGY

RECOMBINANT DNA TECHNOLOGY

Time: 3 hours Maximum : 60 marks

PART – A (20X ½ = 10 Marks)
Answer ALL the Questions

1. The first restriction endonuclease was isolated by
   a. W. Arber b. H.O.Smith , K.W. Wilcox, T.J. Kelley c. Nathens d. S. Linn

2. The enzyme introduce or remove supercoils from covalently closed circular DNA are
   a. Nuclease b. Ligase c. Polymerase d. Topoisomerase

3. The type II restriction endonuclease requires the ions for cleavage
   a. Ca^{2+} b. Mg^{2+} c. Fe^{2+} d. Zn^{2+}

4. The first type II restriction endonuclease to be isolated was
   a. *EcoR* I b. *Hind* II c. *BamH* I d. *Alu* I

5. Vectors used for propagation of DNA inserts are called as
   a. Cloning vector b. Expression vector c. Insertional vector
   d. Replacement vector

6. Vectors used for expression of DNA inserts are called as
   a. Cloning vector b. Expression vector c. Insertional vector
   d. Replacement vector

7. The marker gene present in pBR 322
   a. Amp^+ and Tet^- b. Amp^+ c. Tet^+ d. Kan^+

8. The marker gene present in pUC vectors
   a. Amp^+ and Tet^- b. Amp^+ c. Tet^- d. Kan^+

9. The issues of cloning related to the moral outcome are called as
   a. Ethical issues b. Legal issues c. Social issues d. Non ethical issues

10. The issues of cloning related to the society as a whole are called as
    a. Ethical issues b. Legal issues c. Social issues d. Non ethical issues

11. First immunological screening for recombinant clone screening was developed by
    a. Broome and Gilbert b. W. Arber c. H.O.Smith d. S. Linn

12. Plaque lift screening method for recombinant clone was developed by
    a. Benton and Davis b. Grustein and Hogness c. W. Arber d. S. Linn

13. The method of detecting recombinant DNA molecule by hybridization is called as
    a. Southern blotting b. Northern blotting c. Electrophoresis d. Digestion

14. Southern blotting method was developed by
    a. E.M.Southern b. Grustein and Hogness c. W. Arber d. S. Linn

15. The matrix used for separating nucleic acid is called as
    a. Agarose b. Polyacrylamide c. Lipid d. Agar

16. The separation of nuclic acids and proteins by electrophoresis are based on the property of
    a. Size b. Charge c. Shape d. Form

17. PCR concept was first proposed by
    a. Ghobind Khorana b. E.M.Southern c. Grustein and Hogness d. W. Arber

18. The collection of microscopic DNA spots attached to a solid surface is called as
    a. Microarray b. Macroarray c. Nanoarray d. Serial array

19. The enzyme used in PCR is
    a. RNA polymerase b. *Taq* DNA polymerase c. RNase d. DNase

20. The divalent cations used in PCR is
    a. Mg^{2+} b. Ca^{2+} c. Cl^- d. I^-

PART B (5 X 4= 20 Marks)
Answer ALL the Questions

21. a. Briefly explain DNA polymerases and its function.
    Or
    b. Explain the importance of alkaline phosphatase in rDNA technology.

22. a. Describe artificial plasmid with suitable example.

Or

b. Comment on shuttle vector.

23. a. Mention the process of probe construction.

Or

b. Analyse the joining mechanism of DNA in relation with linkers and adaptors.

24. a. Briefly explain the technique behind flavrasavar tomato.

Or

b. Write note on insertional inactivation with example.

25. a. Explain the principle and applications of PCR.

Or

b. Differentiate RFLP and RAPD.

PART C (3 x 10 = 30 Marks)
Answer any THREE Questions

26. Discuss in detail nomenclature, types and restriction patterns of restriction endonuclease.

27. Elaborately explain plasmid and its types.

28. Explain the different steps involved in c DNA library construction and its applications.

29. Enumerate western Blotting technique.

30. Describe the principle and process involved in DNA finger printing.

Reg. No.....

[12BTU501]

KARPAGAM UNIVERSITY

(Under Section 3 of UGC Act 1956)
COIMBATORE – 641 021
(For the candidates admitted from 2012 onwards)

B.Sc. DEGREE EXAMINATION, NOVEMBER 2014

Fifth Semester

BIOTECHNOLOGY

RECOMBINANT DNA TECHNOLOGY

Time: 3 hours Maximum : 100 marks

PART – A (15 x 2 = 30 Marks)
Answer ALL the Questions

1. Define restriction endonucleases
2. What is the function of Taq Polymerase?
3. Define Linkers
4. What are the characters of an ideal vector?
5. Define cosmids
6. What is meant by cDNA library?
7. Define yeast artificial chromosome
8. Define probe
9. What is chromosome walking?
10. Define restriction mapping
11. Write the types of PCR
12. What is meant by expression vectors?
13. Define gene therapy
14. List out genetically modified plants
15. Bio ethics

PART B (5 X 14= 70 Marks)
Answer ALL the Questions

16. a. Briefly explain the isolation and purification of nucleic acids

Or

b. Write short notes on following.
i) DNA polymerases ii) Phosphatases iii) Ligases
iv) Reverse transcriptase

17. a. Write a short notes on following.
i) Bacterial artificial chromosome ii) Homopolymer tailing
iii) M_{13} phage based vectors

Or

b. Write in detailed account on solid phase synthesis of DNA and cDNA library.

18. a. Describe about probe preparation and chromosome walking.

Or

b. Briefly explain about colony and Plaque hybridization.

19. a. Describe about DNA sequencing and DNA finger printing

Or

b. Explain about the types of PCR.

20. **Compulsory : -**

Write notes on production of recombinant protein in bacteria and yeast.

Reg. No.....

[13BTU501]

KARPAGAM UNIVERSITY

Karpagam Academy of Higher Education
(Established Under Section 3 of UGC Act 1956)
COIMBATORE – 641 021
(For the candidates admitted from 2013 onwards)

B.Sc., DEGREE EXAMINATION, NOVEMBER 2015

Fifth Semester

BIOTECHNOLOGY

RECOMBINANT DNA TECHNOLOGY

Time: 3 hours Maximum : 60 marks

PART – A (20 x 1 = 20 Marks) (30 Minutes)
(Question Nos. 1 to 20 Online Examinations)

PART B (5 x 8 = 40 Marks) (2 ½ Hours)
Answer ALL the Questions

21. a. Briefly explain about the isolation of DNA

Or

b. Write in detail about the role of restriction enzymes, topo isomerase in gene cloning.

22. a. Account on plasmid vectors.

Or

b. Describe the construction of gene and cDNA library.

23. a. Explain about the sequence dependent screening strategies in rDNA technology.

Or

b. Write an account on preparation and labelling of probes

24. a. Explain DNA sequencing method.

Or

b. Explain the principle procedure and application of PCR

25. a. Illustrate the production of recombinant proteins.

Or

b. Account on following: a)Gene therapy b) transgenic plants

Reg. No.....

[14BTU501]

KARPAGAM UNIVERSITY

Karpagam Academy of Higher Education
(Established Under Section 3 of UGC Act 1956)
COIMBATORE – 641 021
(For the candidates admitted from 2014onwards)

B.Sc., DEGREE EXAMINATION, NOVEMBER 2016

Fifth Semester

BIOTECHNOLOGY

RECOMBINANT DNA TECHNOLOGY

Time: 3 hours Maximum : 60 marks

PART – A (20 x 1 = 20 Marks) (30 Minutes)
(Question Nos. 1 to 20 Online Examinations)

PART B (5 x 8 = 40 Marks) (2 ½ Hours)
Answer ALL the Questions

21. a. Write notes on nucleic acid isolation from bacteria.
Or
b. Briefly describe about restriction enzymes and its types.

22. a. Write detailed notes on cosmids.
Or
b. Briefly describe about cDNA libraries.

23. a. Write a notes on probe preparation.
Or
b. Briefly explain about southern blotting.

24. a. Explain in detail protein expression in heterologous systems.
Or
b. Explain about DNA finger printing.

25. a. Define gene therapy? Write detail notes on gene therapy.
Or
b. Write in detail notes on recombinant protein production from bacteria.
