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

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – I – SBB3102 – ADVANCED GENETIC ENGINEERING

Advanced molecular techniques

DNA fingerprinting

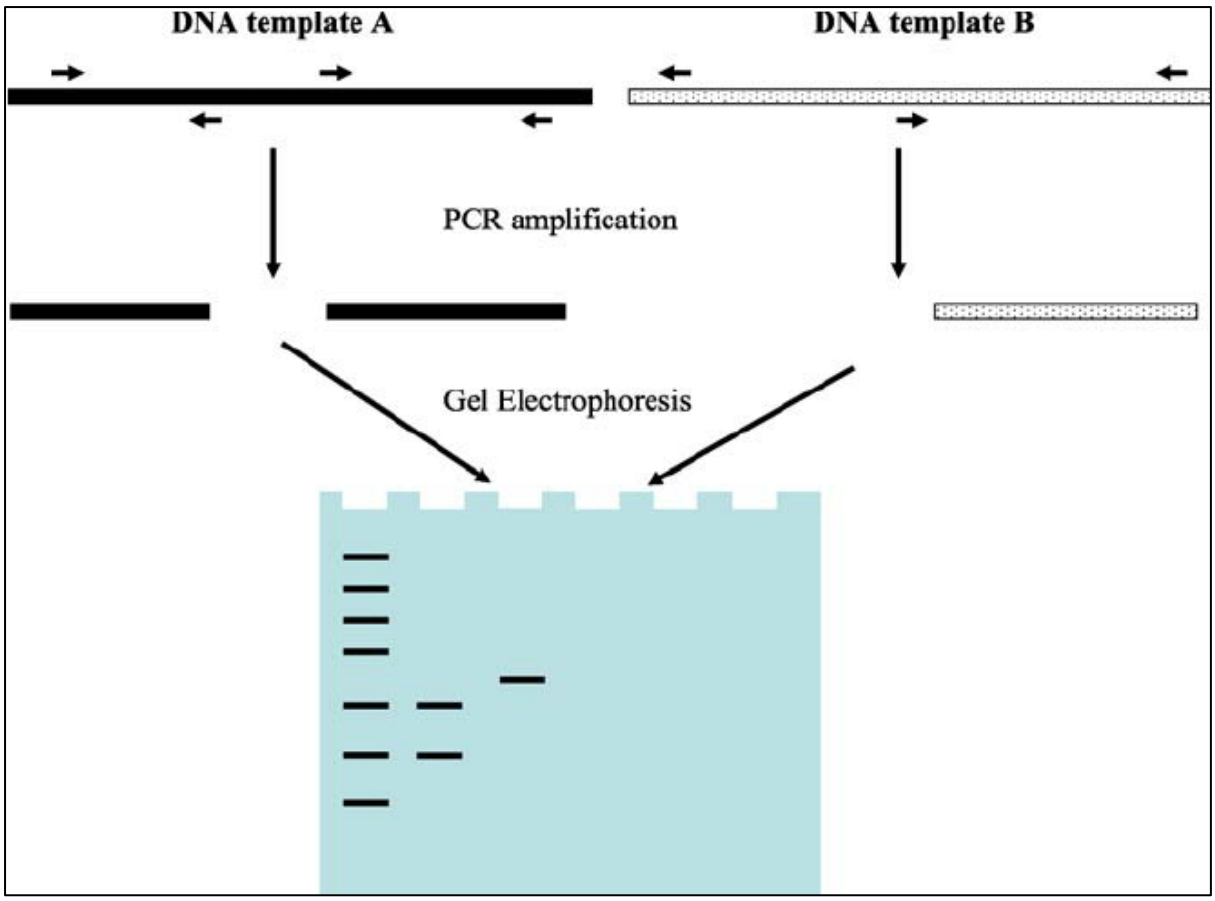
In the 1980's, the use of hypervariable regions of DNA called minisatellites was introduced. The molecular techniques associated with these regions, typically referred to as "DNA fingerprinting," have been a useful tool for pedigree analysis and in estimating the genetic variation in populations.

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

Random Amplified Polymorphic DNA (RAPD)

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

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



Limitations of RAPD

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

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

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

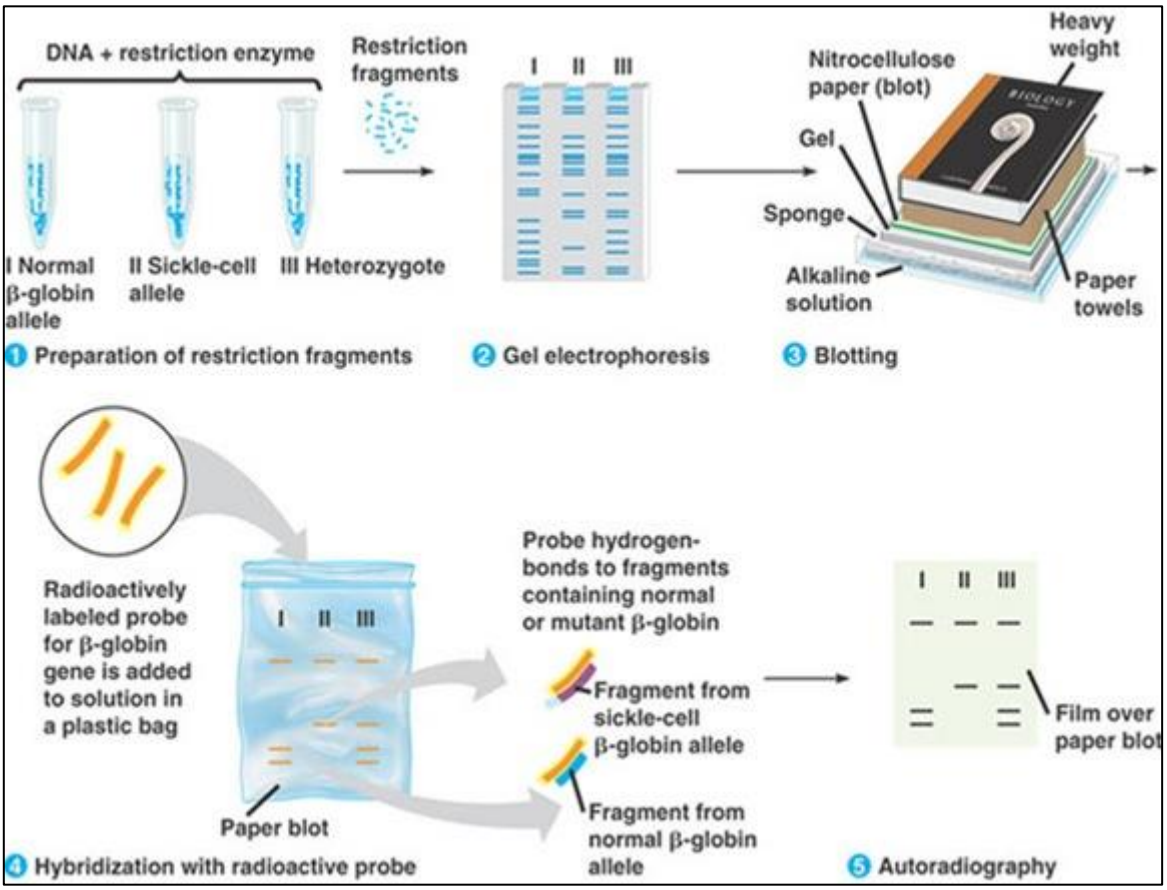
Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) is a difference in homologous DNA sequences that can be detected by the presence of fragments of different lengths after digestion of the DNA samples in question with specific restriction endonucleases. RFLP, as a molecular marker, is specific to a single clone/restriction enzyme combination. Most RFLP markers are co-dominant (both alleles in heterozygous sample will be detected) and highly locus-specific.

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

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

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



Genotyping

Developing RFLP probes

Total DNA is digested with a methylation-sensitive enzyme (for example, PstI), thereby enriching the library for single- or low-copy expressed sequences (PstI clones are based on the suggestion that expressed genes are not methylated).

The digested DNA is size-fractionated on a preparative agarose gel, and fragments ranging from 500 to 2000 bp are excised, eluted and cloned into a plasmid vector (for example, pUC18).

Digests of the plasmids are screened to check for inserts.

Southern blots of the inserts can be probed with total sheared DNA to select clones that hybridize to single- and low-copy sequences.

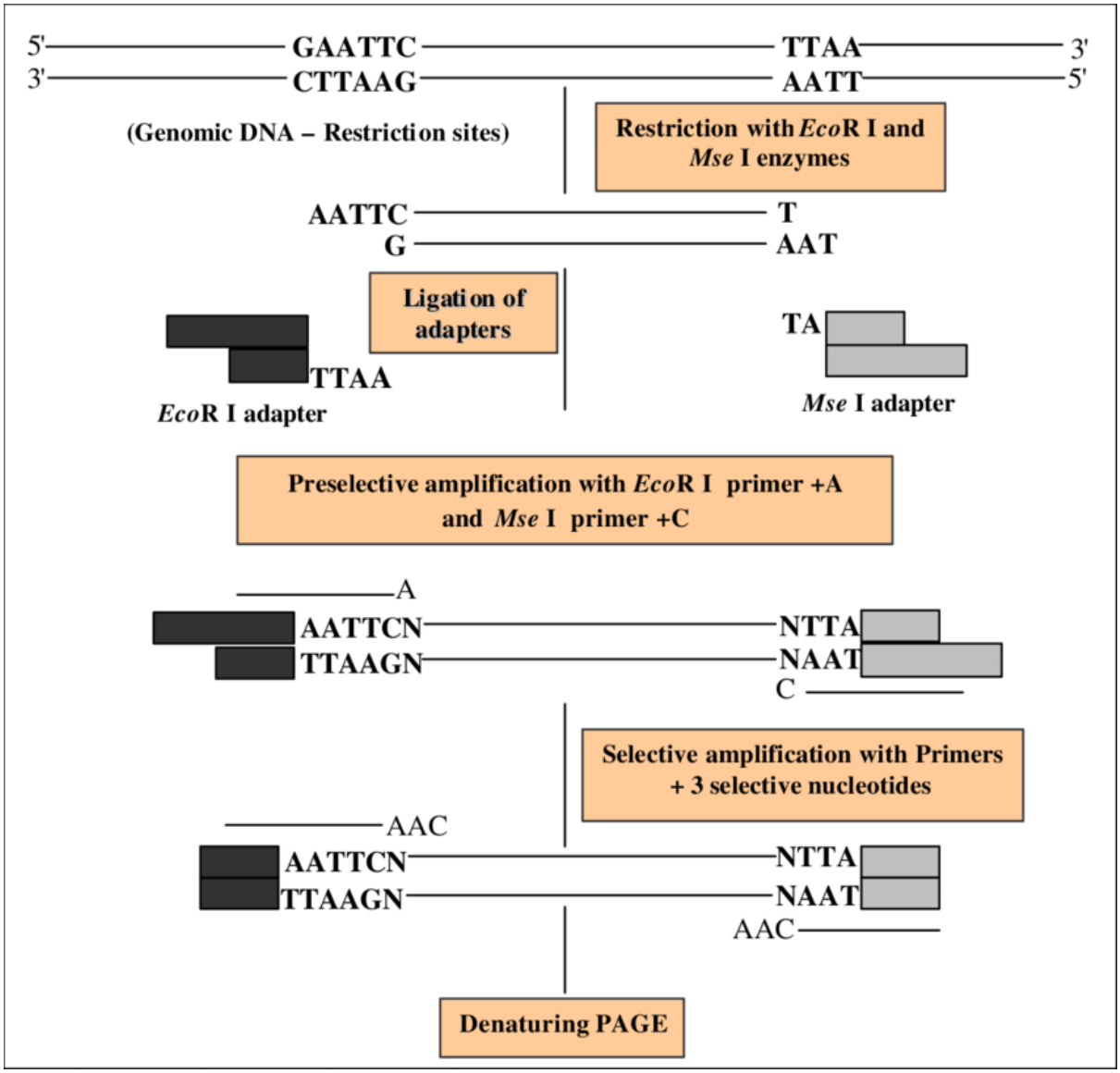
The probes are screened for RFLPs using genomic DNA of different genotypes digested with restriction endonucleases. Typically, in species with moderate to high polymorphism rates, two to four restriction endonucleases are used such as EcoRI, EcoRV, and HindIII. In species with low polymorphism rates, additional restriction endonucleases can be tested to increase the chance of finding polymorphism.

Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphisms (AFLPs) are differences in restriction fragment lengths caused by SNPs or INDELsthat create or abolish restriction endonuclease recognition sites. The AFLP technique is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA.

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

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



Weaknesses of AFLP

Proprietary technology is needed to score heterozygotes and ++ homozygotes. Otherwise, AFLP must be dominantly scored.

Developing locus-specific markers from individual fragments can be difficult.

Need to use different kits adapted to the size of the genome being analyzed.

Sequence-Tagged Sites (STS)

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

The STS concept was introduced by Olson et al (1989). In assessing the likely impact of the Polymerase Chain Reaction (PCR) on human genome research, they recognized that single-

copy DNA sequences of known map location could serve as markers for genetic and physical mapping of genes along the chromosome. The advantage of STSs over other mapping landmarks is that the means of testing for the presence of a particular STS can be completely described as information in a database: anyone who wishes to make copies of the marker would simply look up the STS in the database, synthesize the specified primers, and run the PCR under specified conditions to amplify the STS from genomic DNA.

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

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

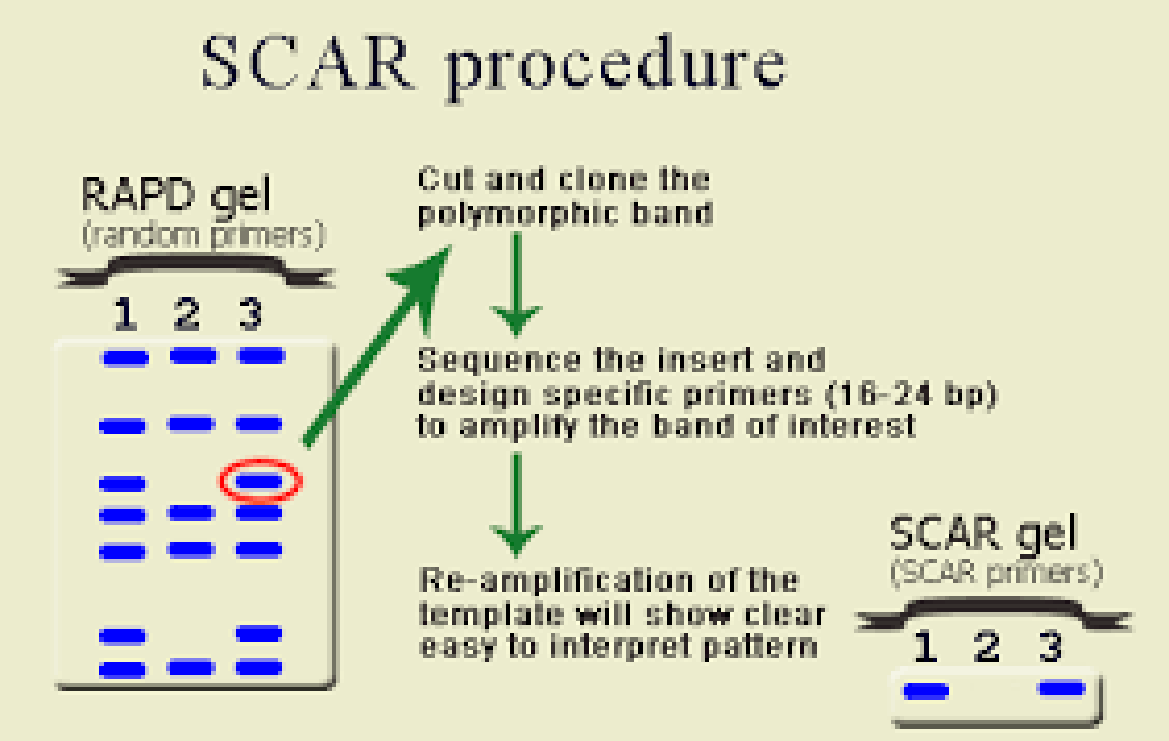
Thus, in broad sense, STS include such markers as microsatellites (SSRs, STMS or SSRPs), SCARs, CAPs, and ISSRs.

Microsatellites

Polymorphic loci present in nuclear DNA and organellar DNA that consist of repeating units of 1-10 base pairs, most typically, 2-3 bp in length, also called Simple Sequence Repeats (SSR), Sequence-Tagged Microsatellite Sites (STMS) or Simple Sequence Repeats Polymorphisms (SSRP). SSRs are highly variable and evenly distributed throughout the genome. This type of repeated DNA is common in eukaryotes. These polymorphisms are identified by constructing PCR primers for the DNA flanking the microsatellite region. The flanking regions tend to be conserved within the species, although sometimes they may also be conserved in higher taxonomic levels.

Sequence Characterized Amplified Region (SCAR)

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



Cleaved Amplified Polymorphic Sequences (CAPS)

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

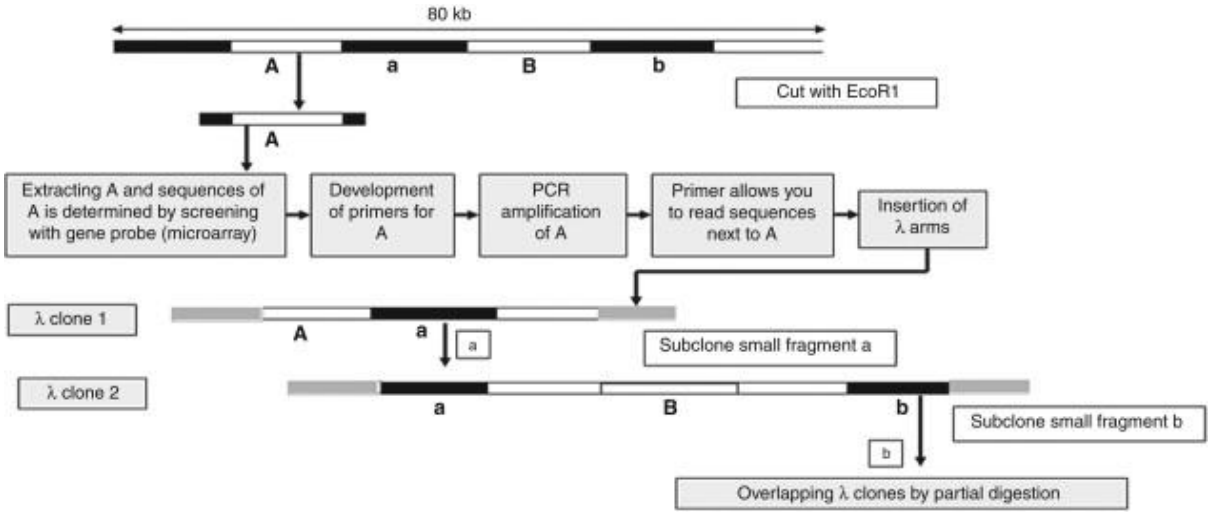
Inter-simple Sequence Repeats (ISSRs)

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

Chromosome walking

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

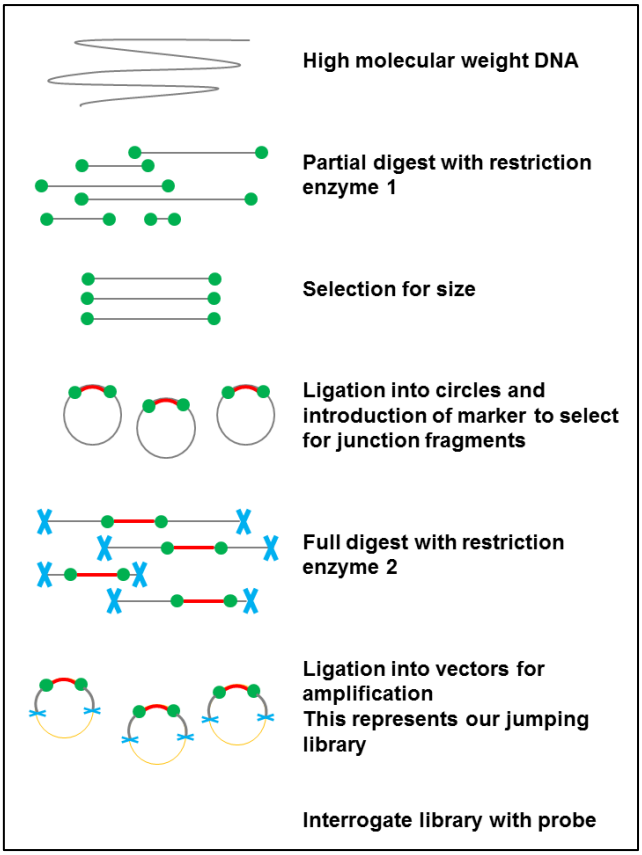
To locate a particular disease gene, the walking starts at the closest gene that has already been identified, known as a marker gene. Each successive gene in the sequence is tested repeatedly for what are known as overlap restrictions and mapped for their precise location in the sequence. Eventually,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



Chromosome jumping

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

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



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

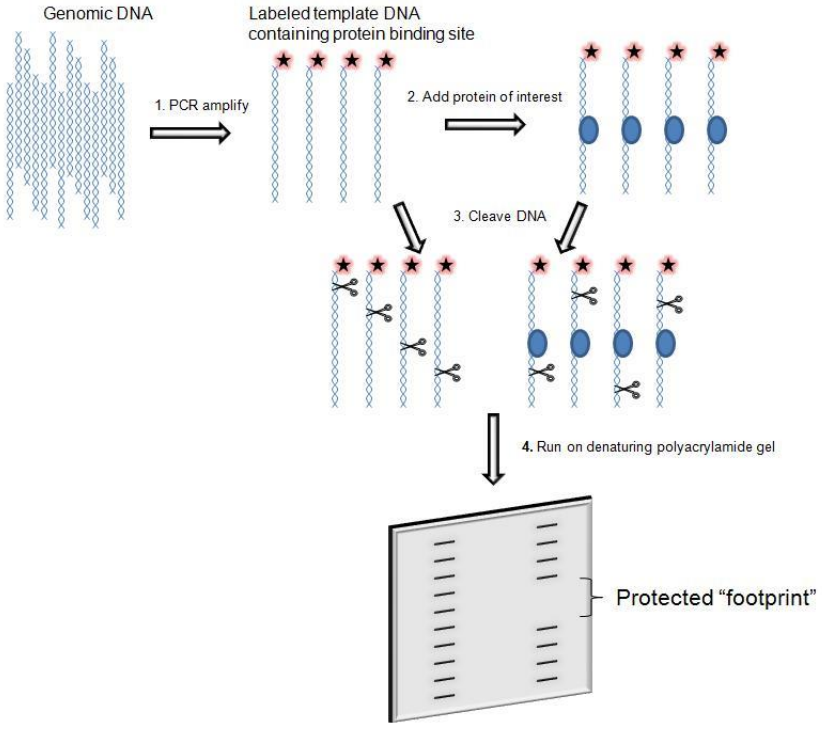
DNA footprinting

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

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

elucidate which proteins bind to these associated regions of DNA and unravel the complexities of transcriptional control. In 1978, David Galas and Albert Schmitz developed the DNA footprinting technique to study the binding specificity of the lac repressor protein. It was originally a modification of the Maxam- Gilbert chemical sequencing technique.

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



A variety of cleavage agents can be chosen.

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

Hydroxyl radicals are created from the Fenton reaction, which involves reducing Fe2+ with H2O2 to form free hydroxyl molecules. These hydroxyl molecules react with the DNA backbone, resulting in a break. Due to their small size, the resulting DNA footprint has high resolution. Unlike DNase I they have no sequence dependence and result in a much more evenly distributed ladder. The negative aspect of using hydroxyl radicals is that they are more time consuming to use, due to a slower reaction and digestion time.

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

CONSTRUCTION OF cDNA LIBRARY

In higher eukaryotes, gene expression is tissue-specific. Only certain cell types show moderate to high expression of a single gene or a group of genes. For example, the genes encoding globin proteins are expressed only in erythrocyte precursor cells, called reticulocytes. Using this information, a target gene can be cloned by isolating the mRNA from a specific tissue. The specific DNA sequences are synthesized as copies from mRNAs of a particular cell type, and cloned into bacteriophage vectors. cDNA (complementary DNA) is produced from a fully transcribed mRNA which contains only the expressed genes of an organism. Clones of such DNA copies of mRNAs are called cDNA clones.

A cDNA library is a combination of cloned cDNA fragments constituting some portion of the transcriptome of an organism which are inserted into a number of host cells. In eukaryotic cells, the mRNA is spliced before translation into protein. The DNA synthesized from the spliced mRNA doesn't have introns or non-coding regions of the gene. As a result, the protein under expression can be sequenced from the DNA which is the main advantage of cDNA cloning over genomic DNA cloning.

Construction of a cDNA Library

The construction of cDNA library involves following steps-

- Isolation of mRNA
- Synthesis of first and second strand of cDNA
- Incorporation of cDNA into a vector
- Cloning of cDNAs

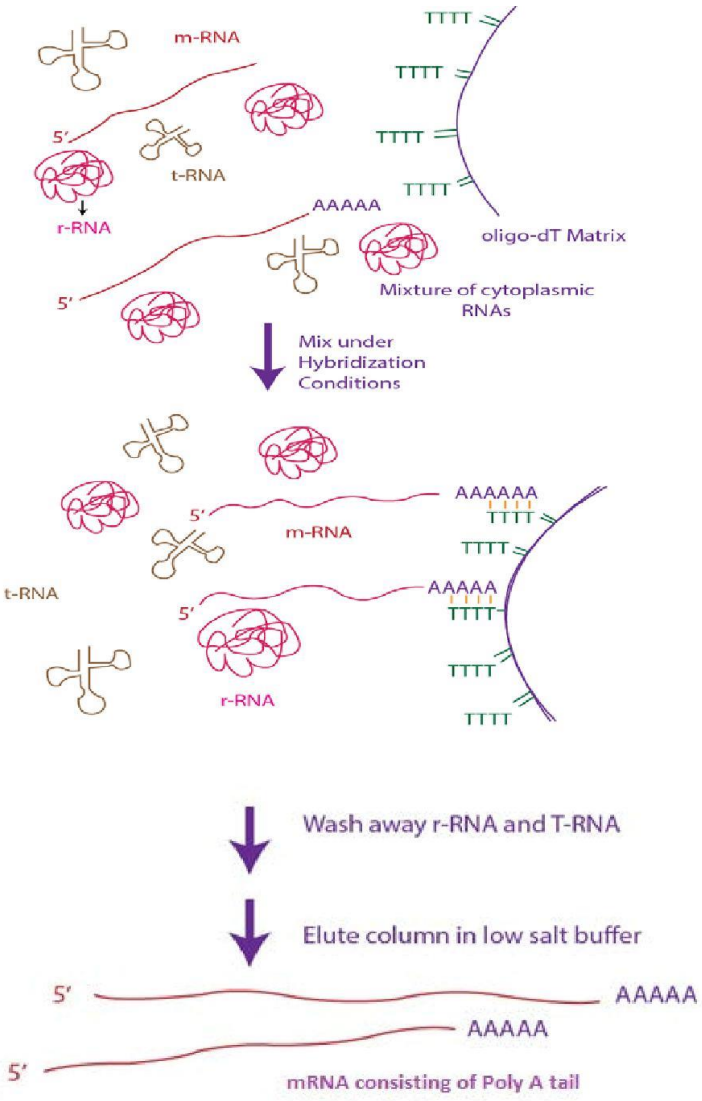
Isolation of mRNA

It involves the isolation of total mRNA from a cell type or tissue of interest. The amount of desired mRNA can be increased by following ways-

- Chromatographic purification of mRNA using oligo-dT column, which retains mRNA molecules, resulting in their enrichment.
- Spinning down mRNA by density gradient centrifugation.
- mRNA preparation from specialized cell types, e.g. developing seeds, chicken oviduct, erythrocytes, β cells of pancreas etc.

The 3′ ends of eukaryotic mRNA consist of a string of 50 – 250 adenylate residues (poly A Tail) which makes the separation easy from the much more prevalent rRNAs and tRNAs in a cell extract using a column containing oligo-dTs tagged onto its matrix.

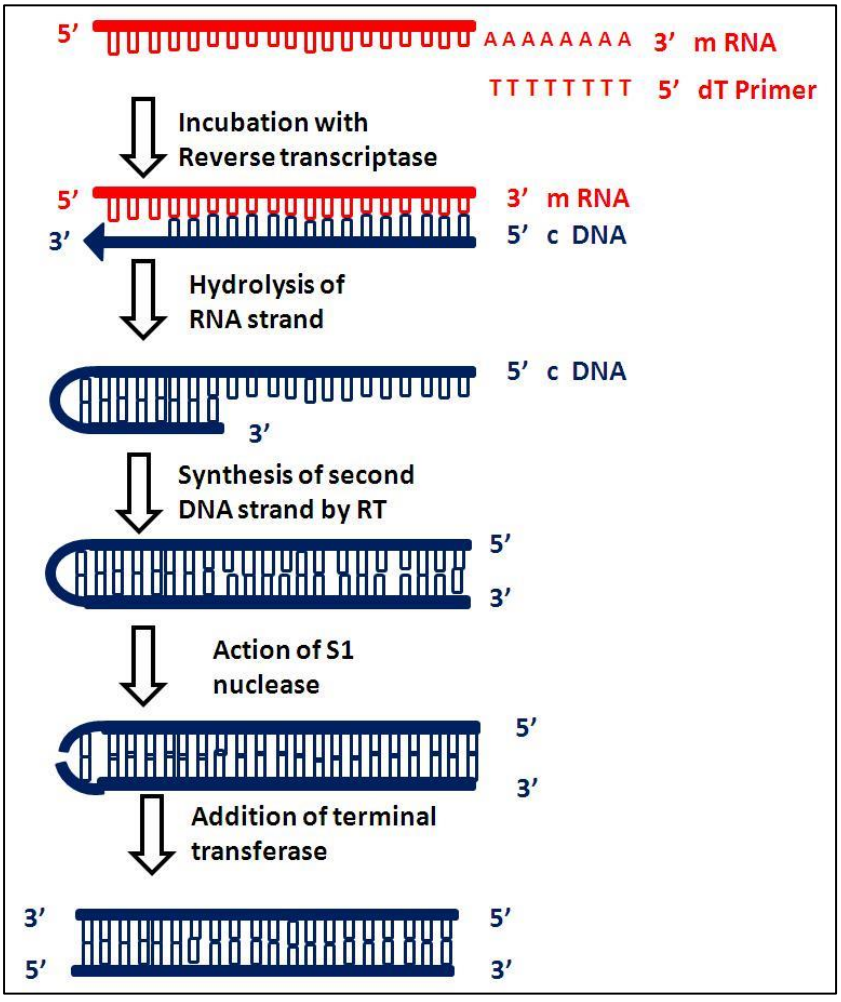
When a cell extract is passed through an oligo-dT column, the mRNAs bind to the column due to the complementary base-pairing between poly (A) tail and oligo-dT. Other RNAs (ribosomal RNAs and transfer RNAs) flow through as unbound fraction. The bound mRNAs can then be eluted using a low-salt buffer.



Synthesis of first and second strand of cDNA

- mRNA being single-stranded cannot be cloned as such and is not a substrate for DNA ligase. It is first converted into DNA before insertion into a suitable vector which can be achieved using reverse transcriptase (RNA-dependent DNA polymerase or RTase) obtained from avian myeloblastosis virus (AMV).
- A short oligo (dT) primer is annealed to the Poly (A) tail on the mRNA.

- Reverse transcriptase extends the 3´-end of the primer using mRNA molecule as a template producing a cDNA: mRNA hybrid.
- The mRNA from the cDNA: mRNA hybrid can be removed by RNase H or Alkaline hydrolysis to give a ss-cDNA molecule.
- No primer is required as the 3´end of this ss-cDNA serves as its own primer generating a short hairpin loop at this end. This free 3´-OH is required for the synthesis of its complementary strand.
- The single stranded (ss) cDNA is then converted into double stranded (ds) cDNA by either RTase or *E. coli* DNA polymerase.
- The ds-cDNA can be trimmed with S1 nuclease to obtain blunt–ended ds-cDNA molecule followed by addition of terminal transferase to tail the cDNA with C's and ligation into a vector.

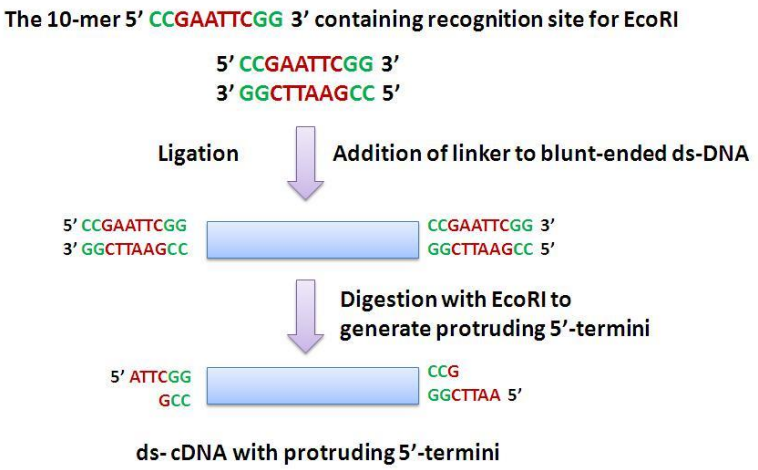


Incorporation of cDNA into a vector

The blunt-ended cDNA termini are modified in order to ligate into a vector to prepare ds-cDNA for cloning. Since blunt-end ligation is inefficient, short restriction-site linkers are first ligated to both ends.

Linker

It is a double-stranded DNA segment with a recognition site for a particular restriction enzyme. It is 10-12 base pairs long prepared by hybridizing chemically synthesized complementary oligonucleotides. The blunt ended ds-DNAs are ligated with the linkers by the DNA ligase from T4 Bacteriophage.



The resulting double-stranded cDNAs with linkers at both ends are treated with a restriction enzyme specific for the linker generating cDNA molecules with sticky ends. Problems arise, when cDNA itself has a site for the restriction enzyme cleaving the linkers. This can be overcome using an appropriate modification enzyme (methylase) to protect any internal recognition site from digestion which methylates specific bases within the restriction-site sequence, thereby, preventing the restriction enzyme binding.

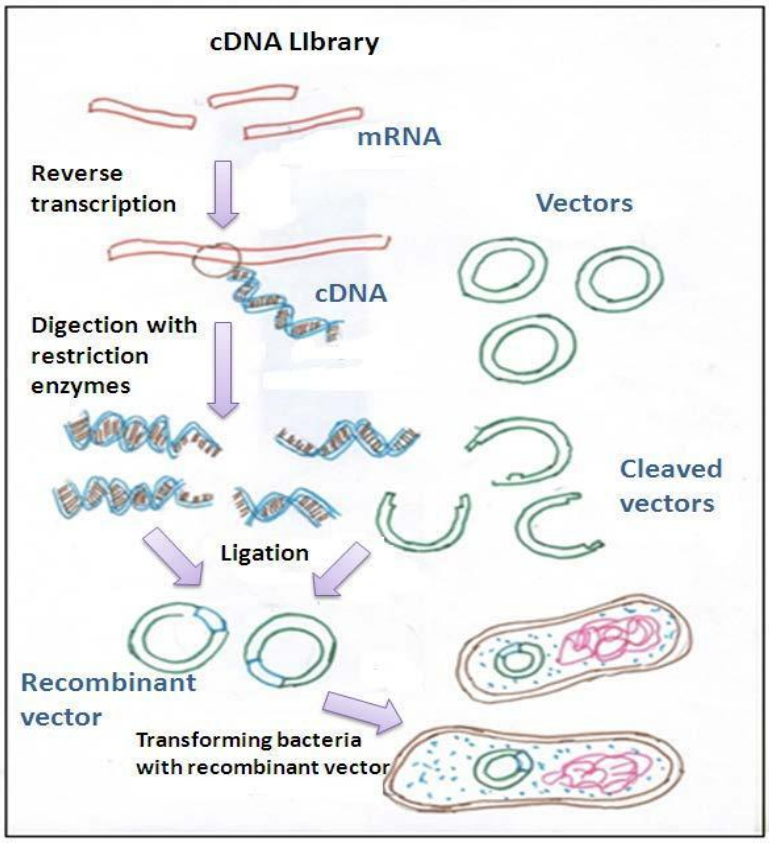
Ligation of the digested ds-cDNA into a vector is the final step in the construction of a cDNA library. The vectors (e.g. plasmid or bacteriophage) should be restricted with the same restriction enzyme used for linkers. The *E. coli* cells are transformed with the recombinant vector, producing a library of plasmid or λ clones. These clones contain cDNA corresponding to a particular mRNA.

Cloning of cDNAs

cDNAs are usually cloned in phage insertion vectors. Bacteriophage vectors offer the following advantageous over plasmid vectors,

- are more suitable when a large number of recombinants are required for cloning low-abundant mRNAs as recombinant phages are produced by *in vitro* packaging.
- can easily store and handle large numbers of phage clones, as compared to the bacterial colonies carrying plasmids.

Plasmid vectors are used extensively for cDNA cloning, particularly in the isolation of the desired cDNA sequence involving the screening of a relatively small number of clones.



Commonly used vectors for cDNA cloning and expression

Vector	Features
Lambda gt10, Lambda gt11	DNA inserts of 7.6 kb and 7.2 kb, respectively, inserted at a unique *EcoRI* cloning site; recombinant Lambda gt10 selected on the basis of plaque morphology ; Lambda gt11 has *E. coli LacZ* gene: *LacZ* and cDNA encoded protein is expressed as fusion protein.
Lambda ZAP series (phasmids)	Up to 10 kb DNA insert; therefore, most cDNAs can be cloned; polylinker has six cloning site; T3 and T7 RNA polymerase sites flank the polylinker so that riboprobes of both strands can be prepared; these features are contained in plasmid vector *pBluescript*, which is inserted into the phage genome; the plasmid containing cDNA recovered simply by co-infecting the bacteria with a helper f1 phage that helps excise from the phage genome.

Problems in cDNA preparation

- Large mRNA sequence results in inefficient synthesis of full- length cDNA. This cause problems during expression as it may not contain the entire coding sequence of the gene. This arises because of the poor processivity of RTase purified from avian myeloblastosis virus (AVM) or produced in *E.coli* from the gene of Moloney murine leukemia virus (MMLV).
- Use of S1 nuclease, the enzyme used to trim the ds cDNA, may remove some important 5´ sequences.

Strategies to overcome the limitations in cDNA preparation

Strategies that can be employed to overcome the above limitations are listed as follows-

- A specially designed *E. coli* vector can be used to avoid incomplete copying of the RNA.

- The use of single strand specific nuclease can be avoided by adding a poly-C tail to the 3´-end of the ss-cDNA produced by copying of the mRNA by the enzyme terminal deoxynucleotidyl transferase. Complementary oligonucleotide (Poly-G) is now used as a primer for the synthesis of complementary strand to yield ds-cDNA without a hairpin loop enhancing the full-length cDNA production.

Applications of cDNA libraries/cloning

- Discovery of novel genes.
- *in vitro* study of gene function by cloning full-length cDNA.
- Determination of alternative splicing in various cell types/tissues.
- They are commonly used for the removal of various non-coding regions from the library.
- Expression of eukaryotic genes in prokaryotes as they lack introns in their DNA and therefore do not have any enzymes to cut it out in transcription process. Gene expression required either for the detection of the clone or the polypeptide product may be the primary objective of cloning.
- To study the expression of mRNA.

Disadvantages of cDNA libraries

- cDNA libraries contain only the parts of genes found in mature mRNA. However, the sequences before and after the gene, for example, those involved in the regulation of gene expression, will not occur in a cDNA library.
- Construction of a cDNA library cannot be used for isolating the genes expressed at low levels as there will be very little mRNA for it in any cell type and may completely be out manoeuvred by the more abundant species.

Construction of genomic library

A genomic library is an organism specific collection of DNA covering the entire genome of an organism. It contains all DNA sequences such as expressed genes, non-expressed genes, exons and introns, promoter and terminator regions and intervening DNA sequences.

Construction of a genomic DNA library involves isolation, purification and fragmentation of genomic DNA followed by cloning of the fragmented DNA using suitable vectors. The eukaryotic cell nuclei are purified by digestion with protease and organic (phenol-chloroform) extraction. The derived genomic DNA is too large to incorporate into a vector and needs to be

broken up into desirable fragment sizes. Fragmentation of DNA can be achieved by physical method and enzymatic method. The library created contains representative copies of all DNA fragments present within the genome.

Mechanisms for cleaving DNA

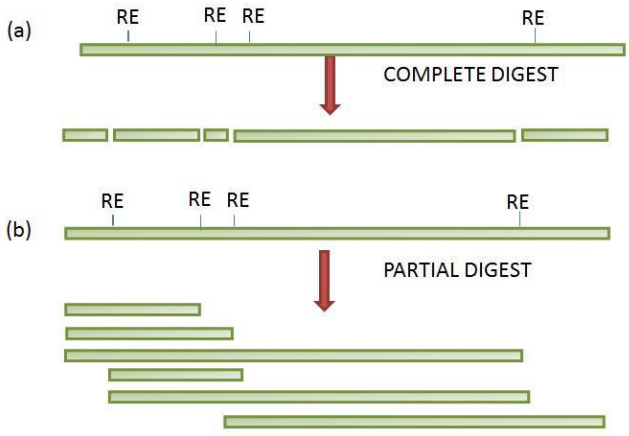
(a) Physical method

It involves mechanical shearing of genomic DNA using a narrow-gauge syringe needle or sonication to break up the DNA into suitable size fragments that can be cloned. Typically, an average DNA fragment size of about 20 kb is desirable for cloning into λ based vectors. DNA fragmentation is random which may result in variable sized DNA fragments. This method requires large quantities of DNA.

(b) Enzymatic method

- It involves use of restriction enzyme for the fragmentation of purified DNA.
- This method is limited by distribution probability of site prone to the action of restriction enzymes which will generate shorter DNA fragments than the desired size.
- If, a gene to be cloned contains multiple recognition sites for a particular restriction enzyme, the complete digestion will generate fragments that are generally too small to clone. As a consequence, the gene may not be represented within a library.
- To overcome this problem, partial digestion of the DNA molecule is usually carried out using known quantity of restriction enzyme to obtain fragments of ideal size.
- The two factors which govern the selection of the restriction enzymes are- type of ends (blunt or sticky) generated by the enzyme action and susceptibility of the enzyme to chemical modification of bases like methylation which can inhibit the enzyme activity.
- The fragments of desired size can be recovered by either agarose gel electrophoresis or sucrose gradient technique and ligated to suitable vectors.

Partial restriction digestion is achieved using restriction enzymes that produce blunt or sticky ends as described below-



i. Restriction enzymes generating blunt ends

The genomic DNA can be digested using restriction enzymes that generate blunt ends e.g. *HaeIII* and *AluI.*

```
HaeIII: 5'-GG|CC-3'    AluI: 5'-AG|CT-3'
        3'-CC|GG-5'          3'-TC|GA-5'
```

Blunt ends are converted into sticky ends prior to cloning. These blunt ended DNA fragments can be ligated to oligonucleotides that contain the recognition sequence for a restriction enzyme called linkers or possess an overhanging sticky end for cloning into particular restriction sites called adaptors.

Linkers

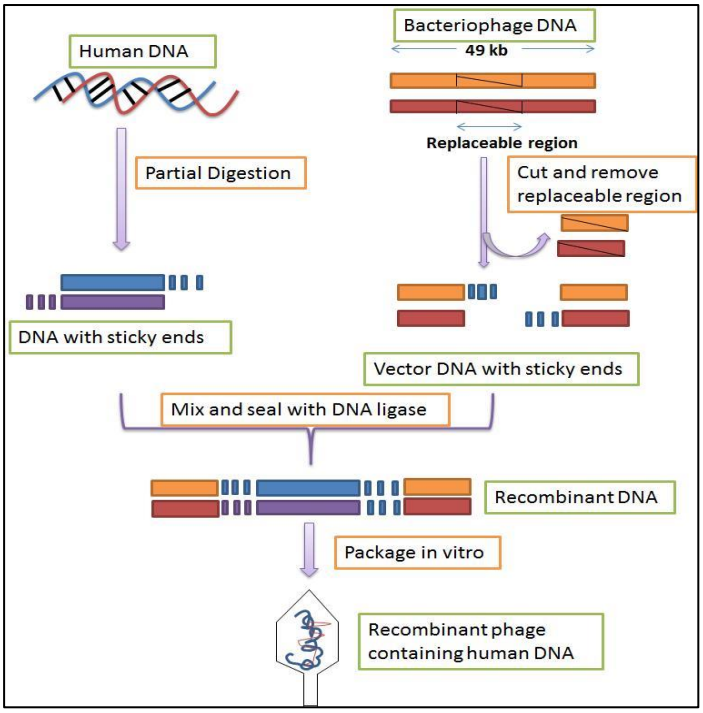
Linkers are short stretches of double stranded DNA of length 8-14 bp that have recognition site for restriction enzymes. Linkers are ligated to blunt end DNA by ligase enzyme. The linker ligation is more efficient as compared to blunt-end ligation of larger molecules because of the presence of high concentration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

Adapters

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

ii. Restriction enzymes that generate sticky ends

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



Cloning of genomic DNA

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

(1) λ replacement vectors

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

(2) High-capacity vectors

The high capacity cloning vectors used for the construction of genomic libraries are cosmids, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs) and yeast artificial chromosomes (YACs). They are designed to handle longer DNA inserts, much larger than for λ replacement vectors. So they require lower number of recombinantsto be screened for identification of a particular gene of interest.

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

Number of clones required for a library

The number of clones to be pooled depends upon the size of the genome *f* and average size of the cloned DNA.Let (*f*) be the fraction of the genome size compared to the average individual cloned fragment size, would represent the lowest possible number of clones that the library must contain.The minimum number of clones required can be calculated as-

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

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

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

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

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

Advantages of genomic libraries

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

Disadvantages of genomic library

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

Applications

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

Vector	Insert size	Features
λ phages	Up to 20-30 kb	Genome size-47 kb, efficient packaging system, replacement vectors usually employed, used to study individual genes.
Cosmids	Up to 40 kb	Contains *cos* site of λ phage to allow packaging, propagate in *E. coli* as plasmids, useful for sub-cloning of DNA inserts from YAC, BAC, PAC etc.
Fosmids	35-45 kb	Contains F plasmid origin of replication and λ*cos* site, low copy number, stable.
Bacterial artificial chromosomes (BAC)	Up to 300kb	Based on F- plasmid, relatively large and high capacity vectors.
P1 artificial chromosomes (PACs)	Up to 300 kb	Derived from DNA of P1 bacteriophage, combines the features of P1 and BACs, used to clone larger genes and in physical mapping, chromosome walking as well as shotgun sequencing of complex genomes.
Yeast artificial chromosomes (YAC)	Up to 2000kb	Allow identification of successful transformants (BAC clones are highly stable and highly efficient)

Comparison of Genomic and cDNA Libraries

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

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

PCR (Polymerase Chain Reaction)

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

Components of PCR

DNA template

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

DNA polymerase

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

Primers

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

Nucleotides (dNTPs or deoxynucleotide triphosphates)

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

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

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

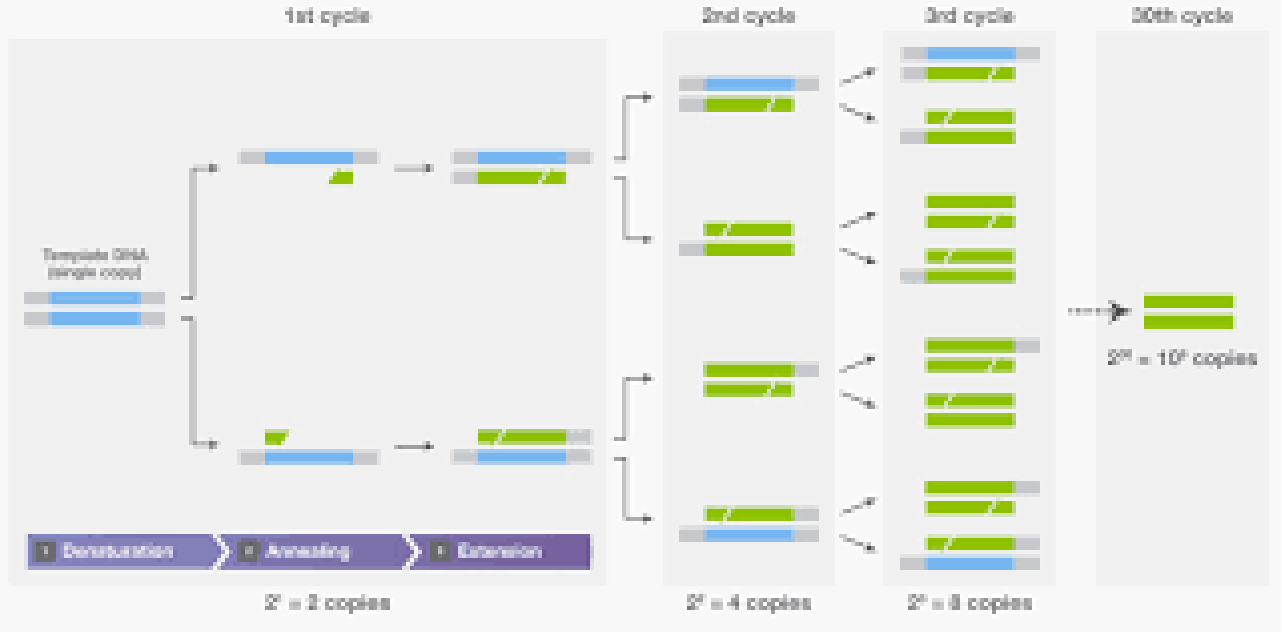
Typically, PCR consists of a series of 20–40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2–3 discrete temperature steps, usually three (Figure below). The cycling is often preceded by a single temperature step at a high temperature (>90 °C), and followed by one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

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

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

Extension/elongation step: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75–80 °C, and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the

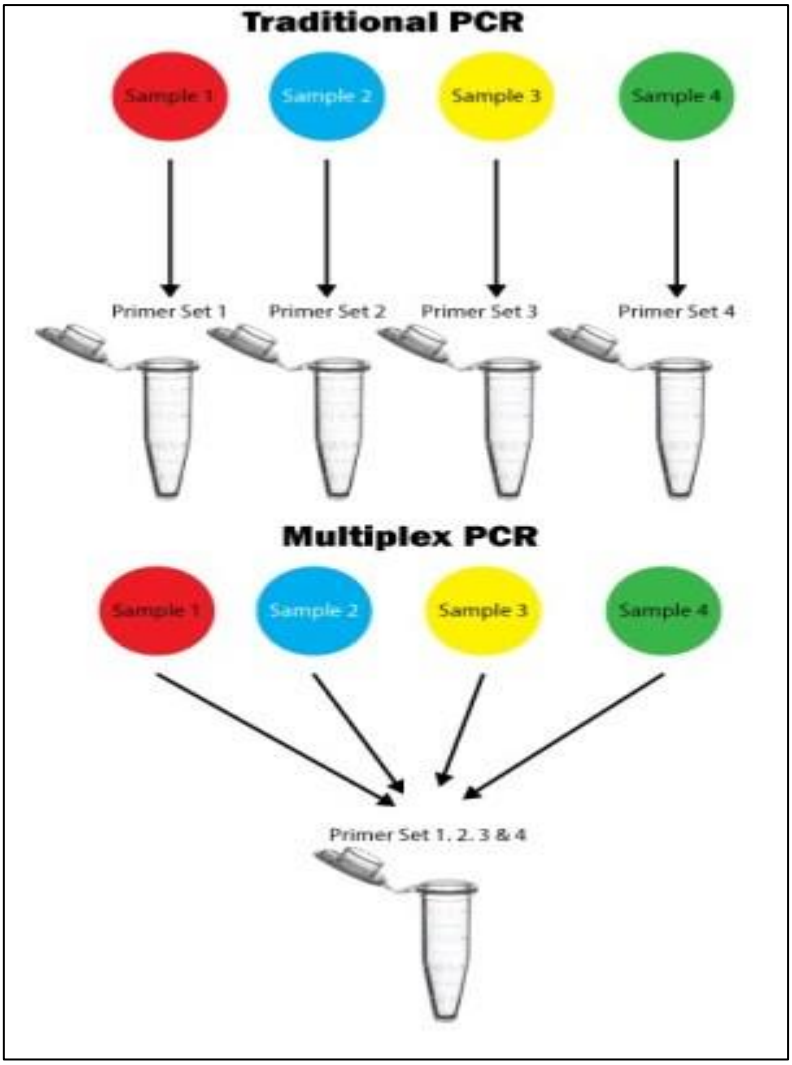
dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to amplify. As a rule-of-thumb, at its optimum temperature, the DNA polymerase polymerizes a thousand bases per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.



Multiplex PCR

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

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



Nested PCR

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

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

1. The target DNA undergoes the first run of polymerase chain reaction with the first set of primers, shown in green. The selection of alternative and similar primer binding sites gives a selection of products, only one containing the intended sequence.

2. The product from the first reaction undergoes a second run with the second set of primers, shown in red. It is very unlikely that any of the unwanted PCR products contain binding sites for both the new primers, ensuring the product from the second PCR has

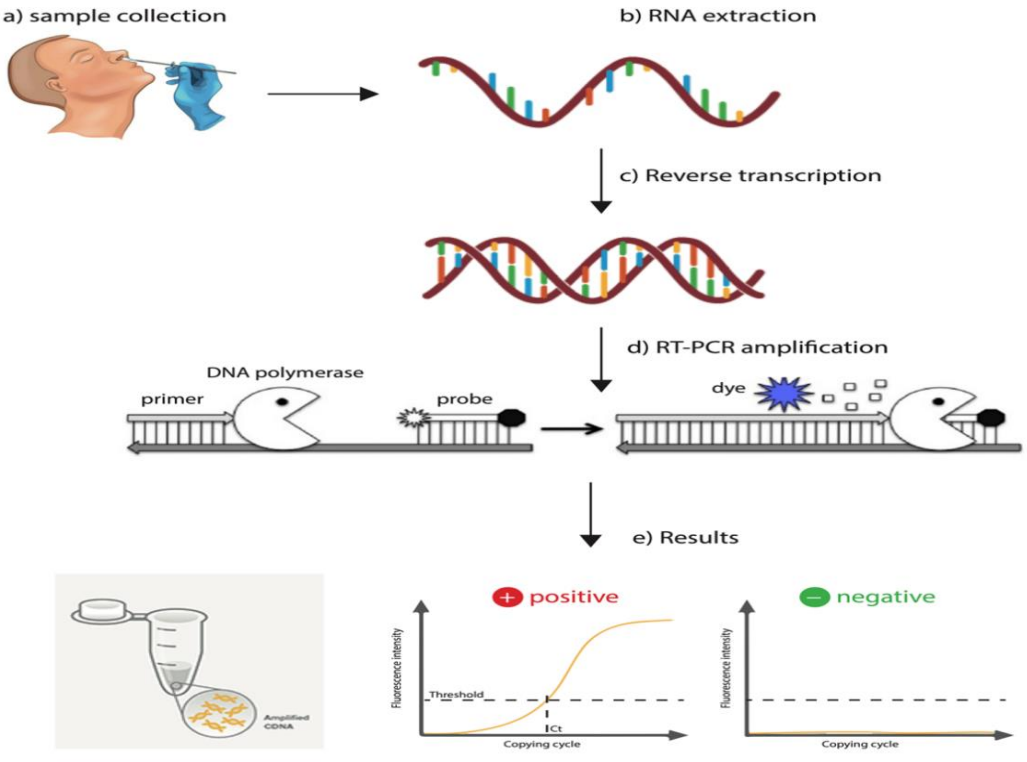
little contamination from unwanted products of primer dimers, hairpins, and alternative primer target sequences.

RT-PCR

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

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

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

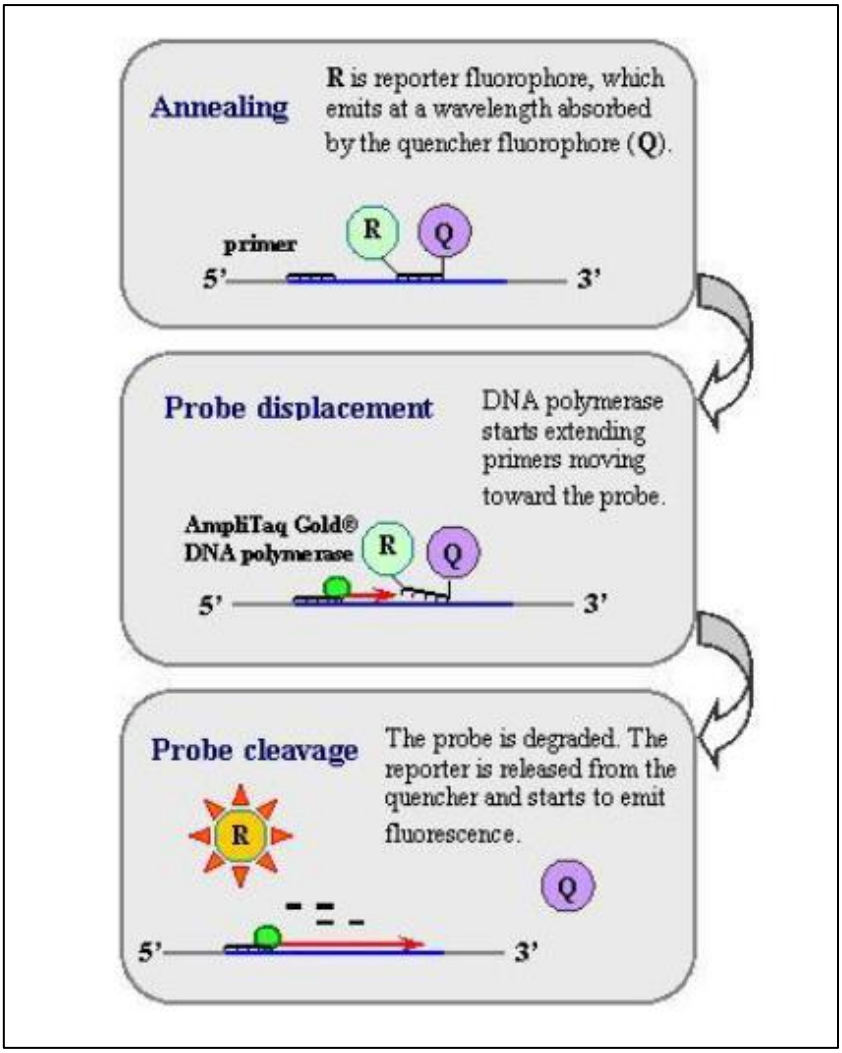


Real time PCR

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

TaqMan assay (named after *Taq* DNA polymerase) **was** one of the earliest methods introduced for real time PCR reaction monitoring and has been widely adopted for both the quantification of mRNAs and for detecting variation. The method exploits the 5' endonuclease activity of *Taq* DNA polymerase to cleave an oligonucleotide probe during PCR, thereby generating a detectable signal. The probes are fluorescently labeled at their 5' end and are non-extendable at their 3' end by chemical modification. Specificity is conferred at three levels: via two PCR primers and the probe.

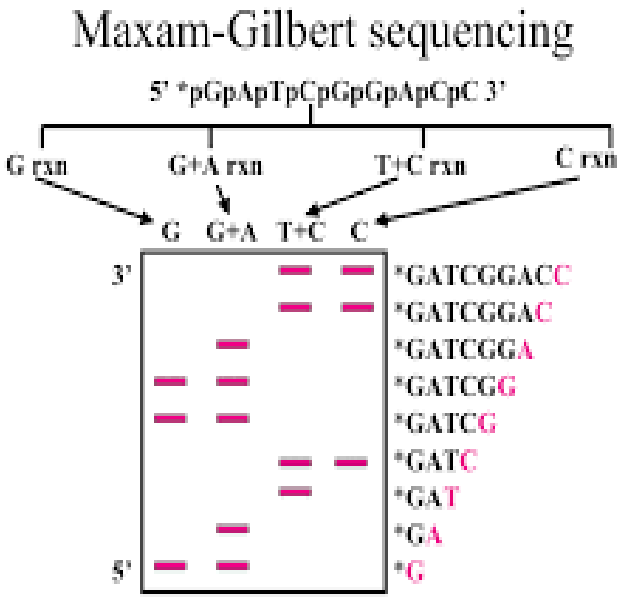
A DNA-binding dye binds to all double-stranded (ds) DNA in PCR, causing fluorescence of the dye. An increase in DNA product during PCR therefore leads to an increase in fluorescence intensity measured at each cycle. However, dsDNA dyes such as SYBR Green will bind to all dsDNA PCR products, including nonspecific PCR products (such as Primer dimer). This can potentially interfere with, or prevent, accurate monitoring of the intended target sequence.
In real-time PCR with dsDNA dyes the reaction is prepared as usual, with the addition of fluorescent dsDNA dye. Then the reaction is run in a real-time PCR instrument, and after each cycle, the intensity of fluorescence is measured with a detector; the dye only fluoresces when bound to the dsDNA (i.e., the PCR product). This method has the advantage of only needing a pair of primers to carry out the amplification, which keeps costs down; however, only one target sequence can be monitored in a tube.


Annealing
R is reporter fluorophore, which emits at a wavelength absorbed by the quencher fluorophore (Q).
primer
R
Q
5'
3'
Probe displacement
DNA polymerase starts extending primers moving toward the probe.
AmpliTaq Gold®
DNA polymerase
R
Q
5'
3'
Probe cleavage
The probe is degraded. The reporter is released from the quencher and starts to emit fluorescence.
R
Q
5'
3'

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

Maxam–Gilbert sequencing is a method of DNA sequencing developed by Allan Maxam and Walter Gilbert in 1976–1977. This method is based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified Maxam–Gilbert sequencing requires radioactive labelling at one 5′ end of the DNA fragment to be sequenced (typically by a kinase reaction using gamma-32P ATP) and purification of the DNA. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). For example, the purines(A+G) are depurinated using formic acid, the guanines (and to some extent the adenines) are methylated by dimethyl sulfate, and the pyrimidines (C+T) are hydrolysed using hydrazine. The addition of salt (sodium chloride) to the hydrazine reaction inhibits the reaction of thymine for the C-only reaction. The modified DNAs may then be cleaved by hot piperidine;(CH2)5NH at the position of the modified base. The concentration of the modifying chemicals is controlled to introduce on average one modification per DNA molecule. Thus a series of labelled fragments is generated, from the radiolabelled end to the first "cut" site in each molecule.

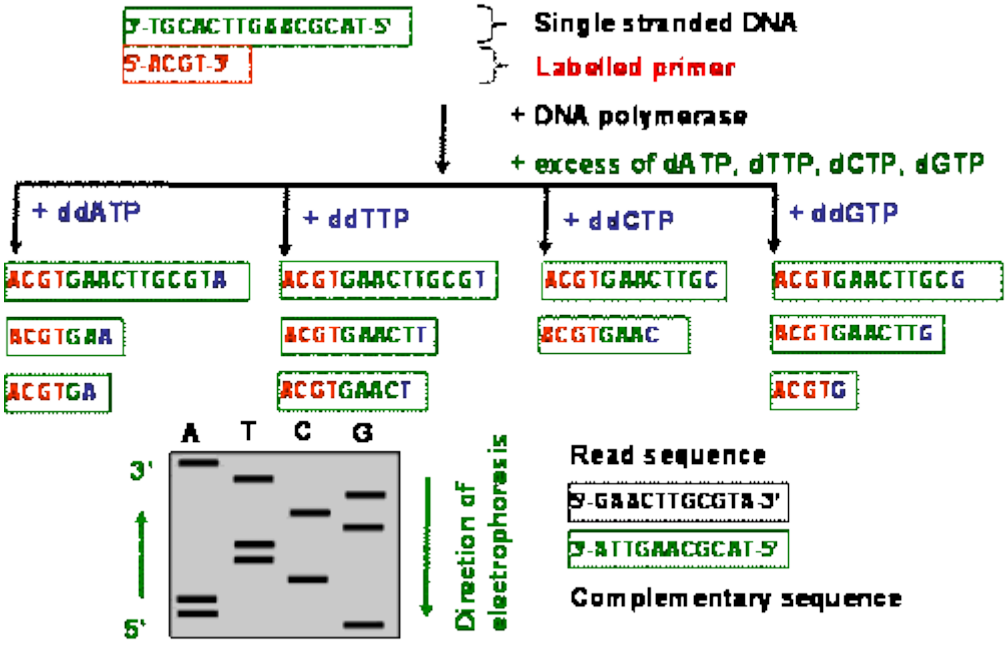
The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each showing the location of identical radiolabelled DNA molecules.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



Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication. Developed by Frederick Sanger and colleagues in 1977, it was the most widely used sequencing method for approximately 25 years. More recently, Sanger sequencing has been supplanted by "Next-Gen" sequencing methods, especially for large-scale, automated genome analyses. However, the Sanger method remains in wide use, for smaller-scale projects, validation of Next-Gen results and for obtaining especially long contiguous DNA sequence reads (>500 nucleotides). The classical chain-termination method requires a single-stranded DNA template, a DNA primer, a DNA polymerase, normal deoxynucleosidetriphosphates (dNTPs), and modified di-deoxynucleoside triphosphates (ddNTPs), the latter of which terminate DNA strand elongation. These chain-terminating nucleotides lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides, causing DNA polymerase to cease extension of DNA when a modified ddNTPs is incorporated. The ddNTPs may be radioactively or fluorescently labelled for detection in automated sequencing machines.

The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP), while the other added nucleotides are ordinary ones. Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat denatured and separated by size using gel electrophoresis. In the original publication of 1977, the formation

of base-paired loops of ssDNA was a cause of serious difficulty in resolving bands at some locations. This is frequently performed using a denaturing polyacrylamide-urea gel with each of the four reactions run in one of four individual lanes (lanes A, T, G, C). The DNA bands may then be visualized by autoradiography or UV light and the DNA sequence can be directly read off the X-ray film or gel image.



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

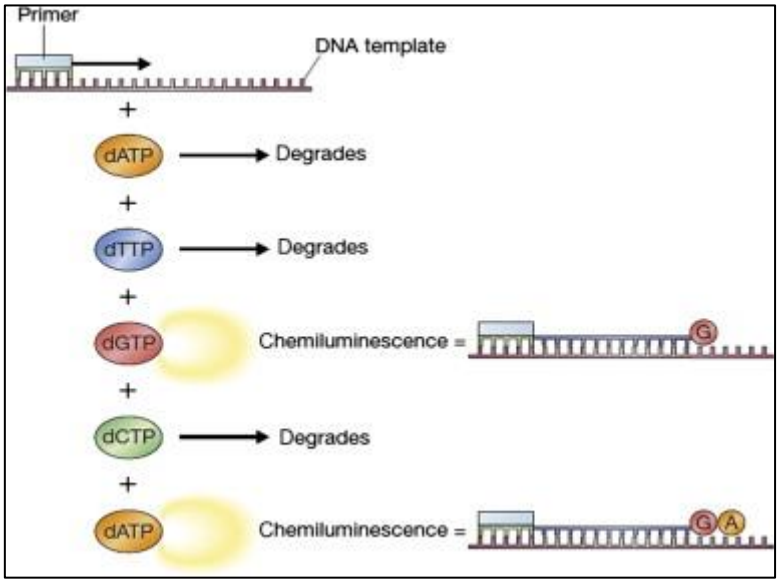
"Sequencing by synthesis" involves taking a single strand of the DNA to be sequenced and then synthesizing its complementary strand enzymatically. The pyrosequencing method is based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemiluminescent enzyme. Essentially,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

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

• The addition of one of the four deoxynucleoside triphosphates (dNTPs) (dATPαS, which is not a substrate for a luciferase, is added instead of dATP to avoid noise) initiates the second step. DNA polymerase incorporates the correct, complementary dNTPs onto the template. This incorporation releases pyrophosphate (PPi).

• ATP sulfurylase converts PPi to ATP in the presence of adenosine 5´ phosphosulfate. This ATP acts as a substrate for the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalysed reaction is detected by a camera and analysed in a pyro gram.

• Unincorporated nucleotides and ATP are degraded by the apyrase, and the reaction can restart with another nucleotide.



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

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – II – SBB3102 – ADVANCED GENETIC ENGINEERING

RECOMBINANT PROTEIN PRODUCTION

E. coli is a popular and well understood system for heterologous protein expression.

Expression options:

• **Direct expression.** *E. coli* cytoplasm is a reducing environment - difficult to ensure proper disulphide bonds formation.

• **Fusion expression.** Ensures good translation initiation. Can overcome insolubility and/or instability problems with small peptides. Has purification advantages based on affinity chromatography.

• **Secretion** a fusion alternative when proteins are fused to peptides or proteins targeted for secretion. Periplasm offers a more oxidising environment, where proteins tend to fold better. Major drawbacks: limited capacity for secretion (0.1-0.2% total cell protein compared to 10% produced intracellularly) and **inability for posttranslational modifications** of proteins.

Disadvantages

Insolubility of heterologous proteins produced in *E.coli* - one of the main problems

Inclusion bodies. Dense particles, containing precipitated proteins. Their formation depends on protein synthesis rate, growth conditions. Advantages: proteolysis resistant, big yield, relatively pure, easy to separate. Disadvantages: inactive product requires in vitro refolding and renaturation

Refolding of recombinant proteins

Solubilisation. High temperatures, detergents, high concentration of inorganic salts or organic solvents all used. The most commonly used organic solutes such as urea or guanidine-HCl often used in the presence of reducing agents (mercaptoethanol or DTT). Solubilised proteins purified by ion-exchange chromatography or other conventional methods, prior to refolding.

Refolding. If no S-S bonds present - remove denaturing agent to allow protein to fold correctly. If S-S bonds present - their formation can be accomplished: by air oxidation, catalysed by trace metal ions; by a mixture of reduced and oxidised thiol compounds - oxidised DTT, reduced DTT; GSSG/GSH; cystine and cysteine, cystamine and cysteamine.

Isolation and characterisation of correctly folded proteins.

 Biological activity.

 Purity. SDS-PAGE.

 Chromatography - reversed phase or ion-exchange.

 N-terminus determination by sequencing. Peptide mapping.

Yeast systems for heterologous expression: S*accharomyces cerevisiae*

Eukaryote, unicellular, GRAS (**G**enerally **R**egarded **A**s **S**afe), capable of performing posttranslational modifications. Excellent recombinant technology: vectors, markers, methods for transformation and gene manipulation, homologous recombination of cloned sequences by single cross over (insertion) and double cross over

Intracellular expression - higher protein yields, but more difficult extraction and purification. Additional potential problems with:

 co- and post-translational processing of proteins at N- and C-termini.

 proteolytic degradation

 addition of tags might result in aggregation and insolubility

Secretion

The yeast secretory pathway **is very similar to that in higher eukaryotes.** N-terminal signal sequences for co-translational translocation of screted proteins into the ER are removed by a signal peptidase. Examples of popular signal sequences used for secretion of heterologous proteins -these of Pho5, Suc2 and the a -factor.

Specific problems with secretion of heterologous proteins

Hyperglycosylation can inhibit reactivity with AB, or render proteins immunogenic (a problem for the production of therapeutic glycoproteins). The obvious solutions: glycosylation mutants (*mnn1, mnn9*) or elimination of potential sites for glycosylation. Alternatively use other yeast species like *P. pastoris*. The cell wall permeability can be a limiting factor. Some cell wall mutants have higher cell wall porosity and release, as a result, heterologous proteins better. Folding of secreted proteins in the ER and involves accessory proteins such as BiP (the product of *KAR2*), and PDI (protein disulphide isomerase). Overexpression of these genes has been beneficial in some cases.

Proteolytic processing could be limited by insufficient amounts of required processing enzymes, and in particular the products of *SEC11*, *KEX2*, *STE13* and *KEX1* in cases of multicopy expression of proteins. Again, might need to overexpress some of these genes.
Modification by N- linked (to asparagine) and O- linked (to serine/threonine) glycosylation.
Hyper glycosylation (outer chain extension) in the yeast Golgi is not typical of mammalian cells. Yeast proteins only modified by mannosylation (no other sugars).

Mammalian cell lines expression systems

• Two modes of expression - transient and stable.

• Cell lines used. Three cell types are dominant in transient expression: human embryonic kidney (HEK), COS and baby hamster kidney (BHK), whilst CHO (Chinese hamster ovary) cells are used predominantly for stable expression.

• Mammalian expression vectors. Eukaryotic origin of replication is from an animal virus: e.g. Simian virus 40 (SV40). Popular markers for selection are the bacterial gene Neor (encodes neomycin phosphotransferase), which confers resistance to G418 (Geneticin), and the gene, encoding dihydropholate reductase (*DHFR*). When *DHFR* is used, the recipient cells must have a defective *DHFR* gene, which makes them unable to grow in the presence of methotrexate (MTX), unlike transfected cells with a functiona l *DHFR* gene. Promoter sequences that drive expression of both marker and cloned heterologous gene, and the transcription termination (polyadenilation signals) are usually from animal viruses (human CMV, SV40, herpes simplex virus) or mammalian genes (bovine growth hormone, thymidine kinase).

Advantages:

• There are no examples of higher eukaryotic proteins, which could not be made in detectable levels, and in a form identical to the natural host (that includes all types of post-translational modifications).

Disadvantages:

• Cultures characterised by lower cell densities and lower growth rates. Maintenance and growing very expensive. Gene manipulations are very difficult. Mammalian cells might contain oncogenes or viral DNA, so recombinant protein products must be tested more extensively.

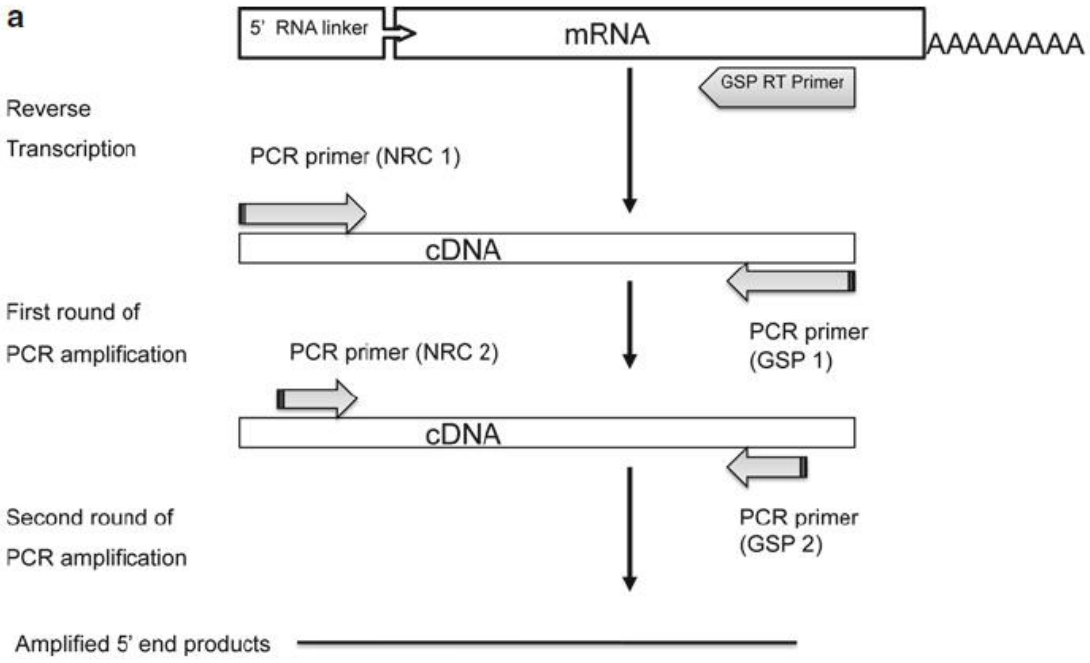
Rapid amplification of cDNA ends (**RACE**) is a technique used to obtain the full length sequence of an RNA transcript found within a cell. RACE results in the production of a cDNA copy of the RNA sequence of interest, produced through reverse transcription, followed by PCR amplification of the cDNA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

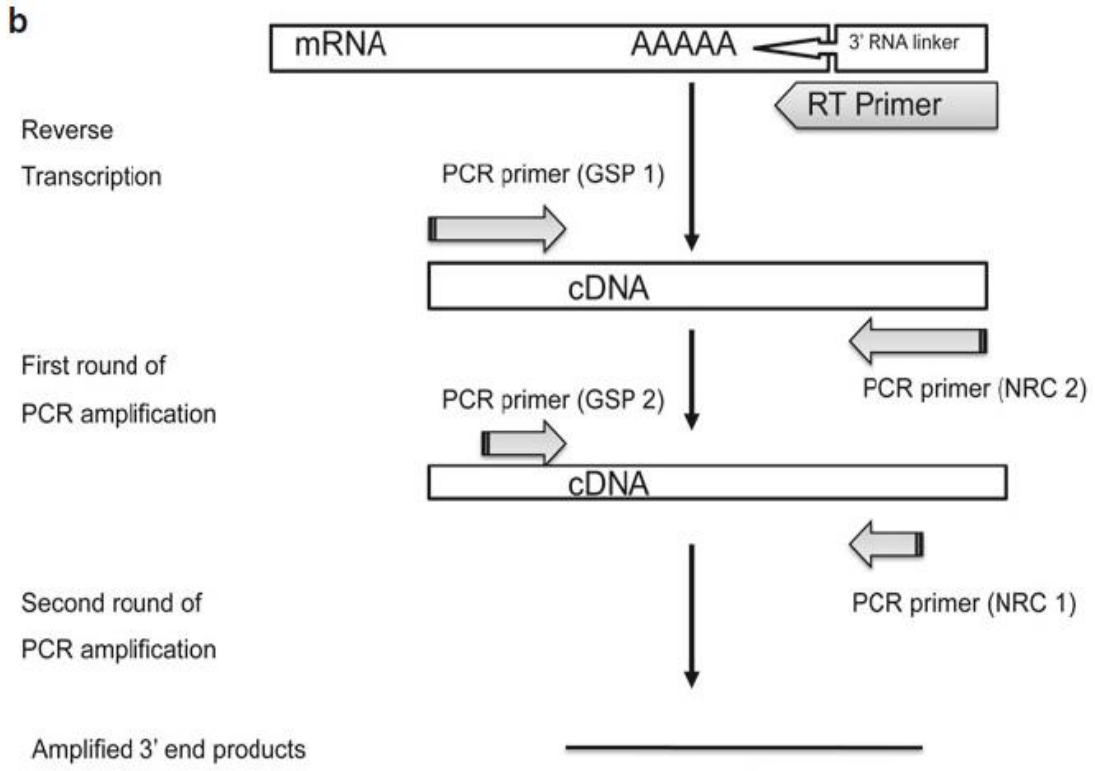
The first step in RACE is to use reverse transcription to produce a cDNA copy of a region of the RNA transcript. In this process, an unknown end portion of a transcript is copied using a known sequence from the center of the transcript. The copied region is bounded by the known sequence, at either the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

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

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

known region using a sense GSP, and an anti-sense primer complementary to the adaptor sequence.

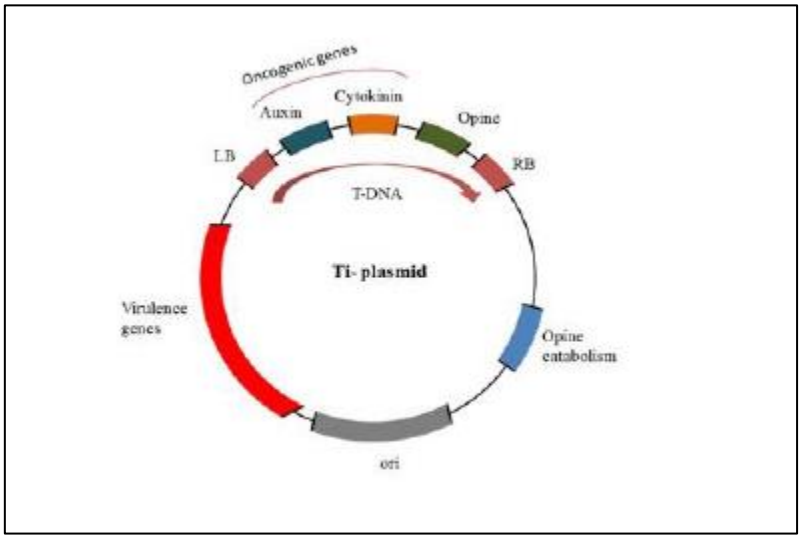




***Agrobacterium*-mediated transformation** - Ti-plasmid

The virulent strains of *A. tumefaciens* harbor large plasmids (140–235 kbp) known as tumor-inducing (Ti) plasmid involving elements like T-DNA, vir region, origin of replication, region enabling conjugative transfer and o-cat region (required for catabolism of opines).

Ti-plasmid of *Agrobacterium*



T-DNA It is a small, specific segment of the plasmid, about 24kb in size and found integrated in the plant nuclear DNA at random site. This DNA segment is flanked by right and left borders.

The functions of T-DNA genes are listed **Gene**	**Product**	**Function**
Ocs	Octopine synthase	Opine synthesis
Nos	Nopaline synthase	Opine synthesis
trns1 (iaaH, auxA)	Tryptophan-2-mono-oxygenase	Auxin synthesis
trns2 (iaaM, auxB)	Indoleacetamide hydrolase	Auxin synthesis
trnr (ipt, cyt)	Isopentyltransferase	Cytokinin synthesis
Frs	Fructopine synthase	Opine synthesis
Mas	Mannopine synthase	Opine synthesis
Ags	Agropine synthase	Opine synthesis

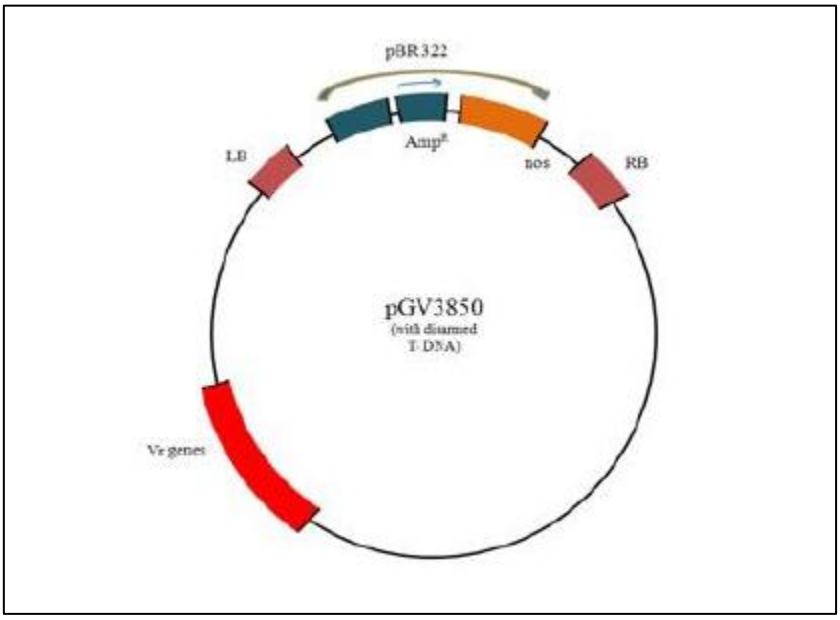
T- DNA: Border Sequences

T-regions are defined by direct repeats known as T-DNA border sequences (Right and Left Border i.e. RB and LB of 25 bp each). These are not transferred intact to the plant genome, but are involved in the transfer process. The RB is rather precise, but the LB can vary by about 100 nucleotides. Deletion of the RB repeat abolishes T-DNA transfer, but the LB seems to be non-essential. The LB repeat has little transfer activity alone.

Disarmed Ti-plasmid derivatives as plant vectors

Ti plasmid is a natural vector for genetically engineering plant cells due to its ability to transfer T-DNA from the bacterium to the plant genome. But wild-type Ti plasmids are not suitable as vectors due to the presence of oncogenes in T-DNA that cause tumor growth in the recipient plant cells. For efficient plant regeneration, vectors with disarmed T-DNA are used by making it non-oncogenic by deleting all of its oncogenes. foreign DNA is inserted between the RB and LB and then integrated into the plant genome without causing tumors.

The creation of disarmed T-DNA is an important step forward, but the absence of tumor formation makes it necessary to use an alternative method for the identification of transformed plant cells. Opine production using *pGV3850* was exploited as a screenable phenotype, and the *ocs* and *nos*genes are now widely used as screenable markers.



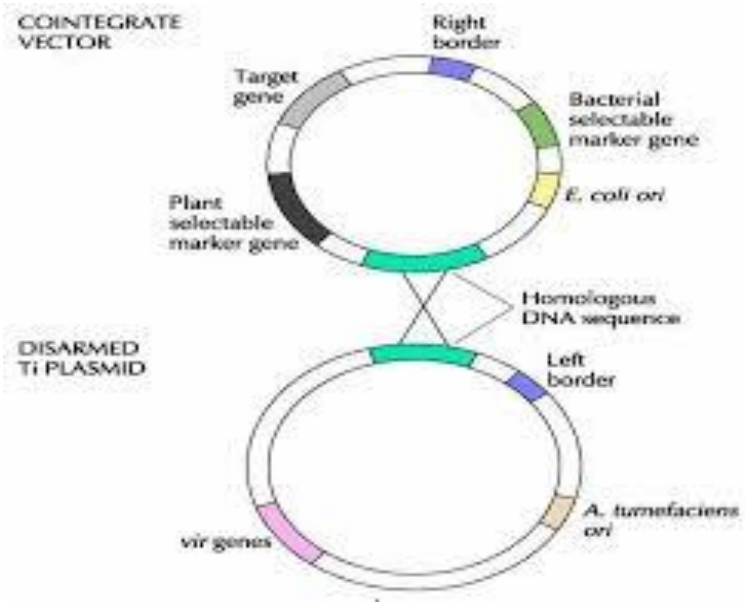
Drawbacks

Several drawbacks are associated with disarmed Ti- vector systems as discussed below;

- Necessity to carry out enzymatic assays on all potential transformants.
- Not convenient as experimental gene vectors due to large size.
- Difficulty in *in vitro* manipulation and
- Absence of unique restriction sites in the T-DNA.

Co- integrate vectors

Co-integrate vectors are the deletion derivatives of Ti-plasmids. The DNA to be introduced into the plant transformation vector is sub cloned in a conventional *Escherichia coli* plasmid vector for easy manipulation, producing a so-called *intermediate vector*. These vectors are incapable of replication in*A. tumefaciens* and also lack conjugation functions. Transfer is achieved using a 'triparental mating' in which three bacterial strains are mixed together: (i) An *E. coli* strain carrying a helper plasmid able to mobilize the intermediate vector in *trans*; (ii) *The E.coli* strain carrying the recombinant intermediate vector; (iii) *A.tumefaciens* carrying the Ti plasmid. Conjugation between the two *E. coli* strains transfers the helper plasmid to the carrier of the intermediate vector, which in turn is mobilized and transferred to the recipient *Agrobacterium.*Homologous recombination between the T-DNA sequences of the Ti plasmid and intermediate vector forms a large co- integrate plasmid resulting in the transfer of recombinant T-DNA to the plant genome.

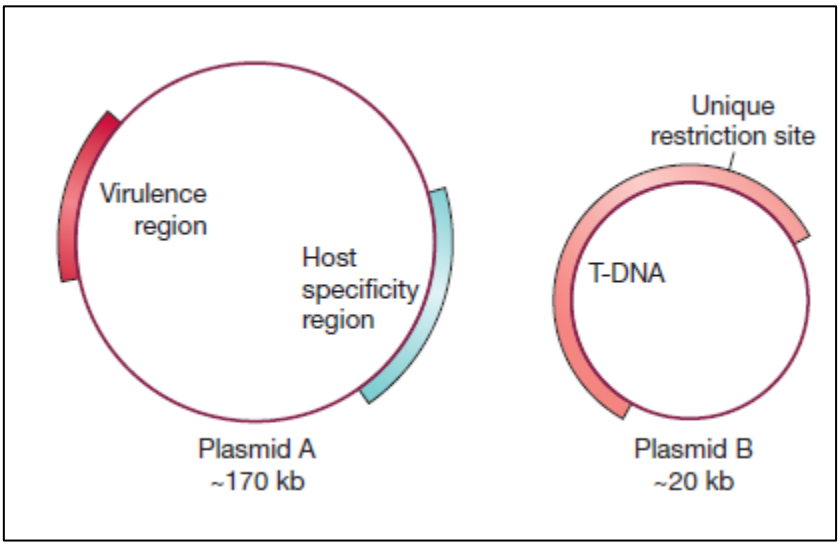


Binary vector

 Binary vector was developed by Hoekma *et al* (1983) and Bevan in (1984).

 It utilizes the trans- acting functions of the vir genes of the Ti-plasmid and can act on any T-DNA sequence present in the same cell.

 Binary vector contains transfer apparatus(the vir genes) and the disarmed T-DNA containing the transgene on separate plasmid



Advantages of Binary vector

 Small size due to the absence of border sequences needed to define T-DNA region and vir region.

 Ease of manipulation

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

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – III – SBB3102 – ADVANCED GENETIC ENGINEERING

Gene manipulation techniques

Site-directed mutagenesis

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

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

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

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

Kunkel's method

In 1987, Thomas Kunkel introduced a technique that reduces the need to select for the mutants.The DNA fragment to be mutated is inserted into a phagemid such as M13mp18/19 and is then transformed into an *E. coli* strain deficient in two enzymes, dUTPase (dut) and uracil deglycosidase (*ung*). Both enzymes are part of a DNA repair pathway that protects the bacterial chromosome from mutations by the spontaneous deamination of dCTP to dUTP. The dUTPase deficiency prevents the breakdown of dUTP, resulting in a high level of dUTP in the cell. The uracil deglycosidase deficiency prevents the removal of uracil from newly synthesized DNA. As the double-mutant *E. coli* replicates the phage DNA, its enzymatic machinery may, therefore, misincorporate dUTP instead of dTTP, resulting in single-strand DNA that contains some uracils

(ssUDNA). The ssUDNA is extracted from the bacteriophage that is released into the medium, and then used as template for mutagenesis. An oligonucleotide containing the desired mutation is used for primer extension. The heteroduplex DNA, that forms, consists of one parental non-mutated strand containing dUTP and a mutated strand containing dTTP. The DNA is then transformed into an E. coli strain carrying the wildtype *dut* and *ung* genes. Here, the uracil-containing parental DNA strand is degraded, so that nearly all of the resulting DNA consists of the mutated strand.

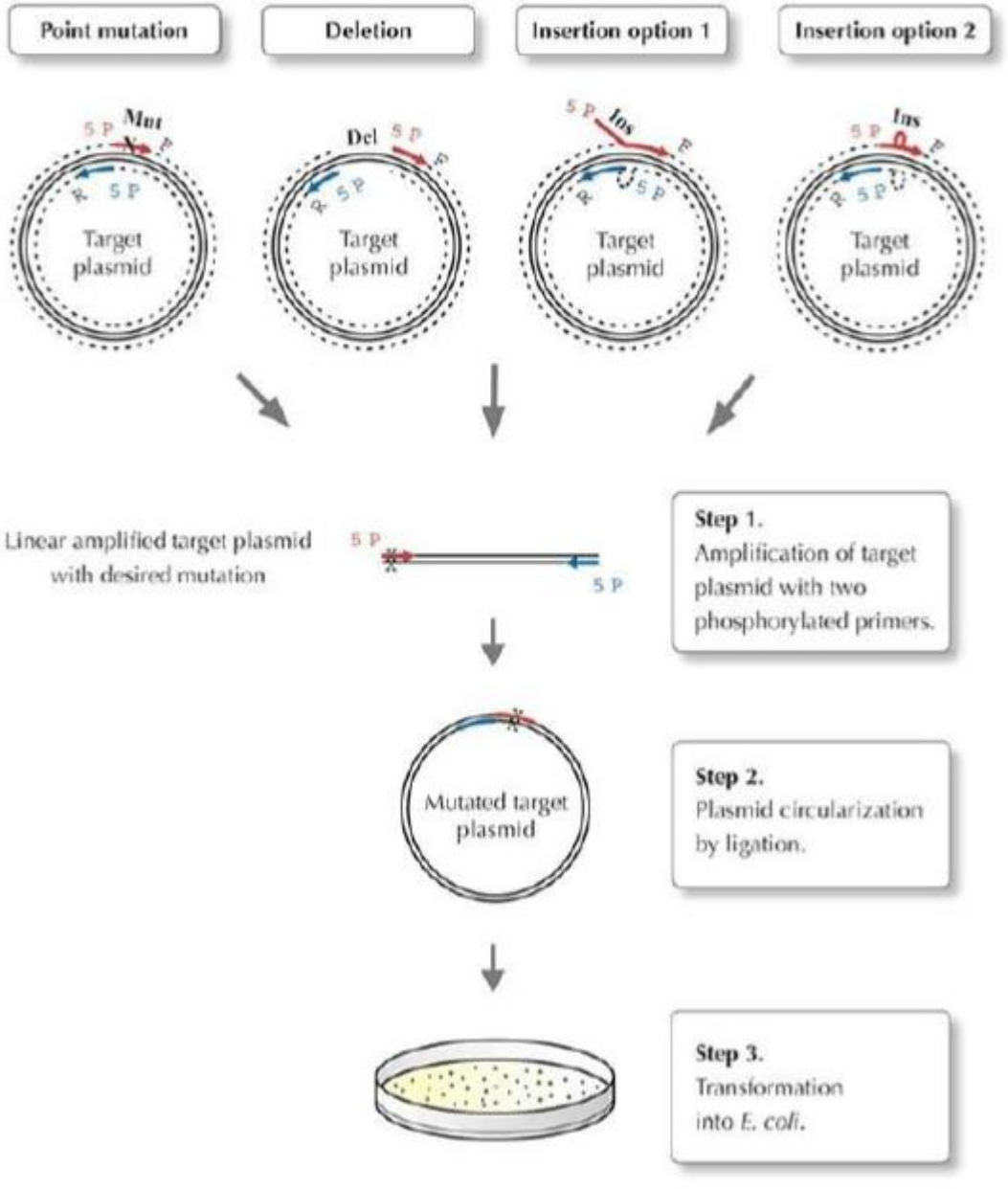
Cassette mutagenesis

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

PCR site-directed mutagenesis

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

from any convenient restriction site. These methods require multiple steps of PCR so that the final fragment to be ligated can contain the desired mutation.



GENE THERAPY

Introduction

Gene therapy is a novel treatment method which utilizes genes or short oligonucleotide sequences as therapeutic molecules, instead of conventional drug compounds. This technique is widely used to treat those defective genes which contribute to disease development. Gene therapy involves the introduction of one or more foreign genes into an organism to treat hereditary or acquired genetic defects. In gene therapy, DNA encoding a therapeutic protein is packaged within a "vector", which transports the DNA inside cells within the body. The disease is treated with minimal toxicity, by the expression of the inserted DNA by the cell machinery. In 1990 FDA for the first time approved a gene therapy experiment on ADA-SCID in the United States after the treatment of Ashanti DeSilva. After that, approximately 1700 clinical trials on patients have been performed with various techniques and genes for numerous diseases. Many diseases such as ADA-SCID, X-linked SCID, Leber's congenital amaurosis (a retinal disease), Parkinson's disease, multiple myeloma, chronic and acute lymphocytic leukemia, adrenoleukodystrophy have reported of successful clinical trials. But these are still not approved by FDA. Some other diseases on which gene therapy-based research is going on are Haemophilia, Tyrosinemia, Hyperbilirubinemia (Criglar-Nijjar Syndrom), Cystic Fibrosis and many other cancers. After 30 years of research and clinical trials, only one product called Glybera got approval in November 2012 which may be available in market in late 2013. It has the ability to cure lipoprotein lipase deficiency (LPLD) a very rare disease.

Types of gene therapy

There are several approaches for correcting faulty genes; the most common being the insertion of a normal gene into a specific location within the genome to replace a non-functional gene. Gene therapy is classified into the following two types:

1. Somatic gene therapy
2. Germ line gene therapy

Somatic Gene Therapy

In somatic gene therapy, the somatic cells of a patient are targeted for foreign gene transfer. In this case the effects caused by the foreign gene is restricted to the individual patient only, and not inherited by the patient's offspring or later generations.

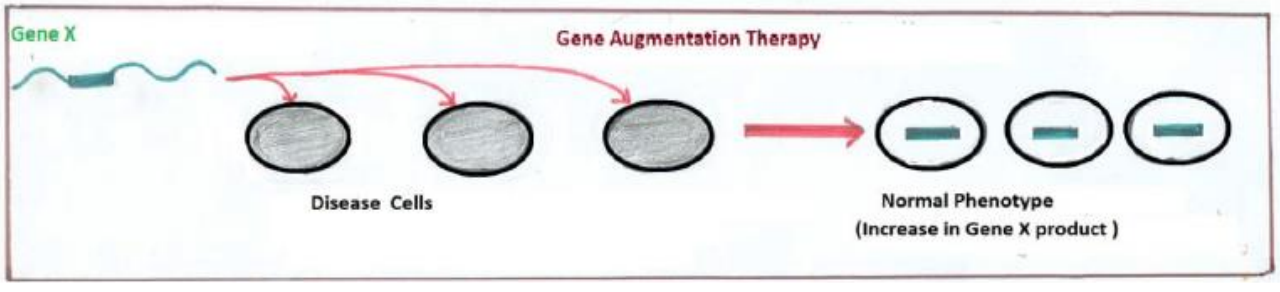
Germ Line Gene Therapy

Here, the functional genes, which are to be integrated into the genomes, are inserted in the germ cells, i.e., sperm or eggs. Targeting of germ cells makes the therapy heritable.

Gene Therapy Strategies

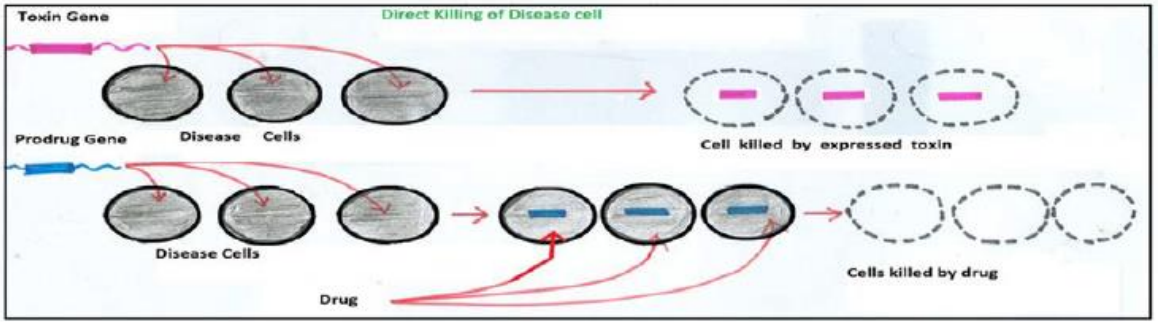
Gene Augmentation Therapy (GAT)

In GAT, simple addition of functional alleles is used to treat inherited disorders caused by genetic deficiency of a gene product, e.g. GAT has been applied to autosomal recessive disorders. Dominantly inherited disorders are much less amenable to GAT.

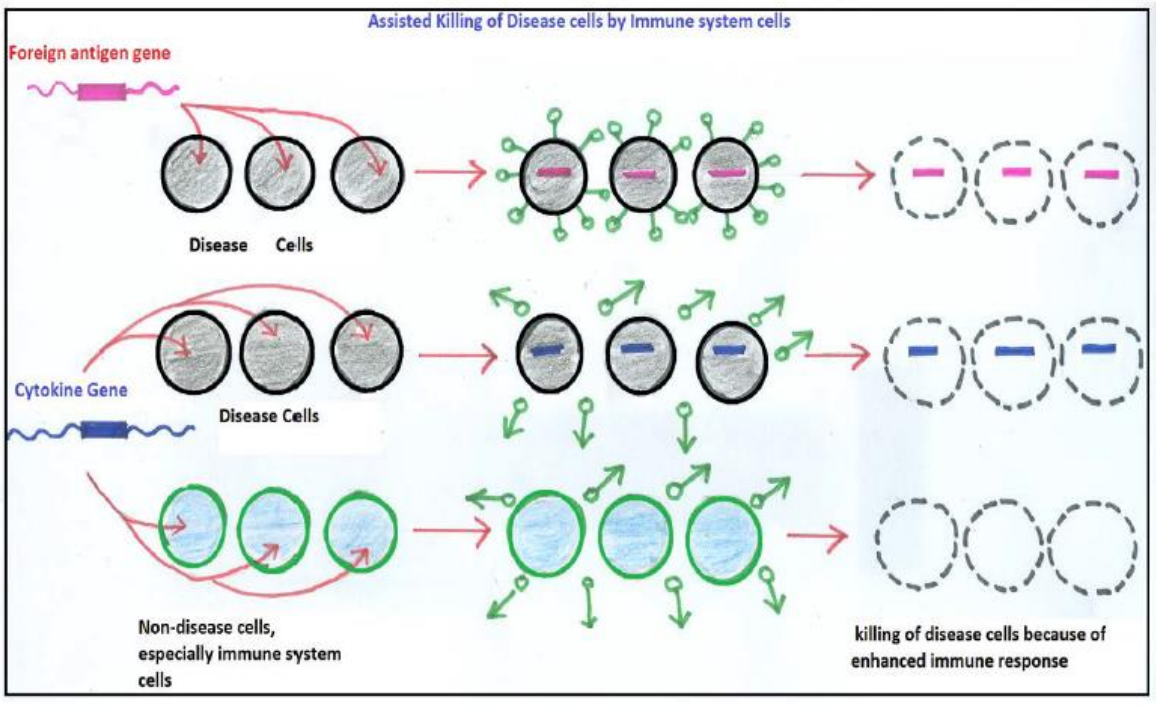


Targeted Killing of Specific Cells

It involves utilizing genes encoding toxic compounds (**suicide genes**), or **prodrugs** (reagents which confer sensitivity to subsequent treatment with a drug) to kill the transfected/ transformed cells. This general approach is popular in cancer gene therapies.

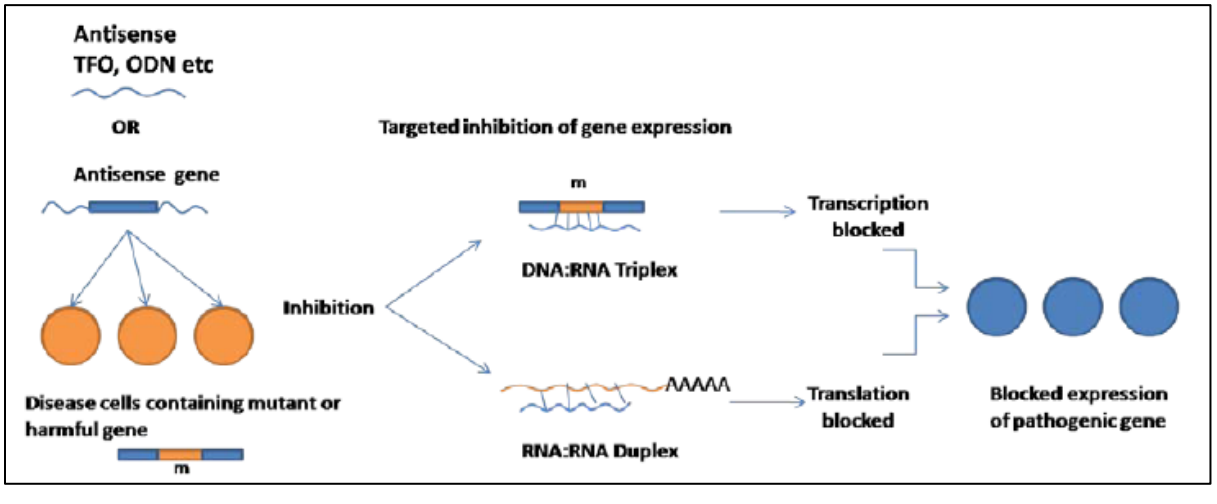


Thymidine kinase (TK) phosphorylates the introduced prodrug ganciclovir which is further phosphorylated by endokinases to form ganciclovir triphosphate, an competitive inhibitor of deoxyguanosine triphosphate. Ganciclovir triphosphate causes chain termination when incorporated into DNA.



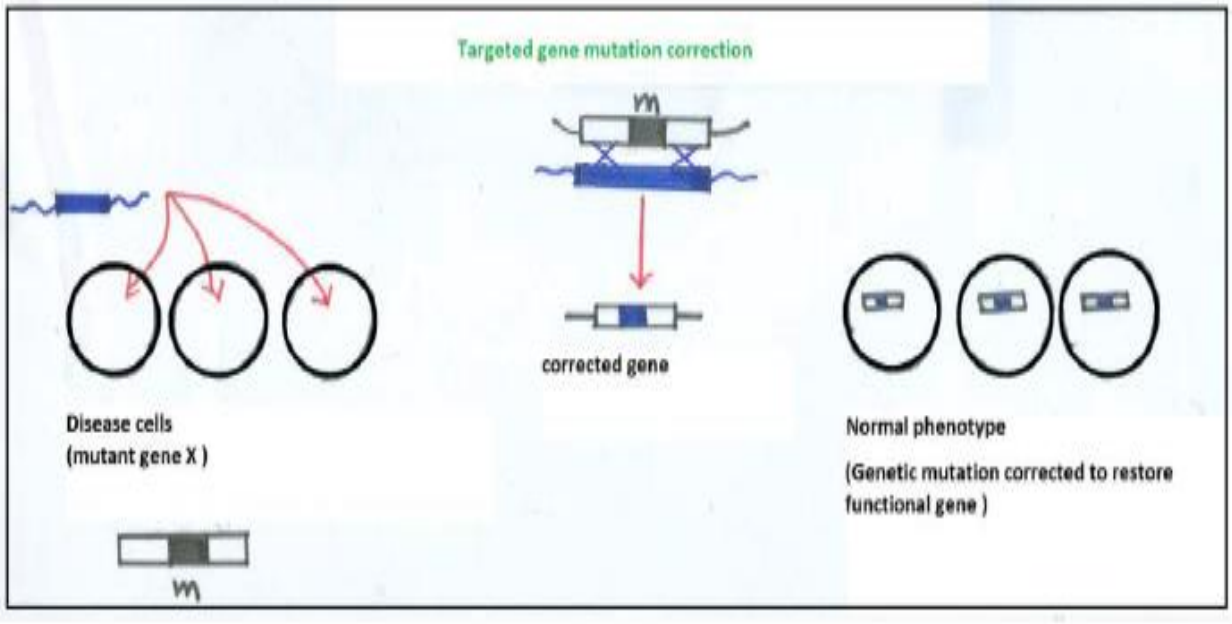
Targeted Inhibition of Gene Expression

This is to block the expression of any diseased gene or a new gene expressing a protein which is harmful for a cell. This is particularly suitable for treating infectious diseases and some cancers.



Targeted Gene Mutation Correction

It is used to correct a defective gene to restore its function which can be done at genetic level by homologous recombination or at mRNA level by using therapeutic ribozymes or therapeutic RNA editing.



Gene Therapy Approaches

Classical Gene Therapy

It involves therapeutic gene delivery and their optimum expression once inside the target cell. The foreign genes carry out following functions.

- Produce a product (protein) that the patient lacks;
- Produces toxin so that diseased cell is killed.
- Activate cells of the immune system so as to help in killing of diseased cells.

Non-classical gene therapy

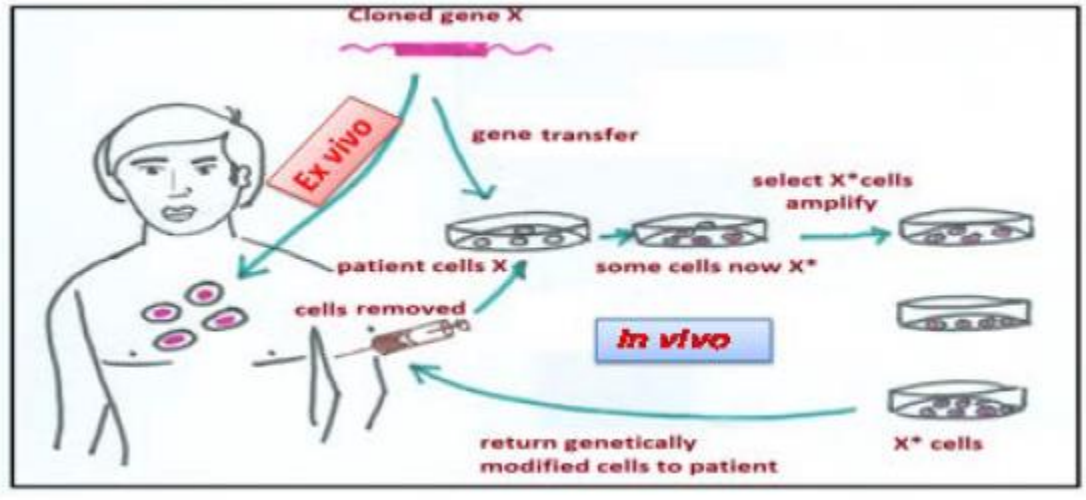
It involves the inhibition of expression of genes associated with the pathogenesis, or to correct a genetic defect and restore the normal gene expression.

Methods of gene therapy

There are mainly two approaches for the transfer of genes in gene therapy:

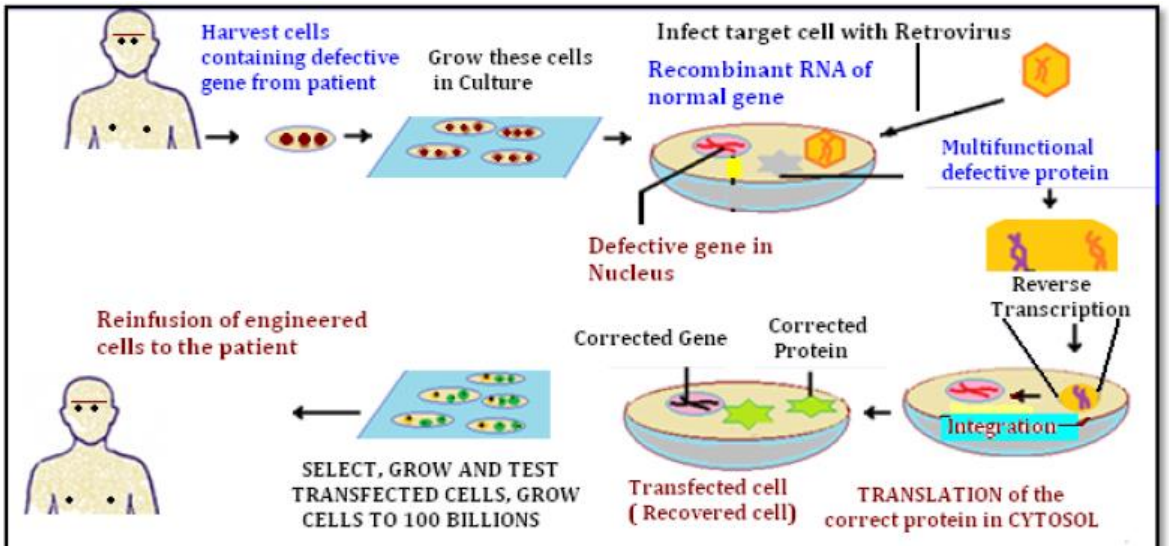
1**.** Transfer of genes into patient cells outside the body (*ex vivo gene therapy*)

2. Transfer of genes directly to cells inside the body (*in vivo*).



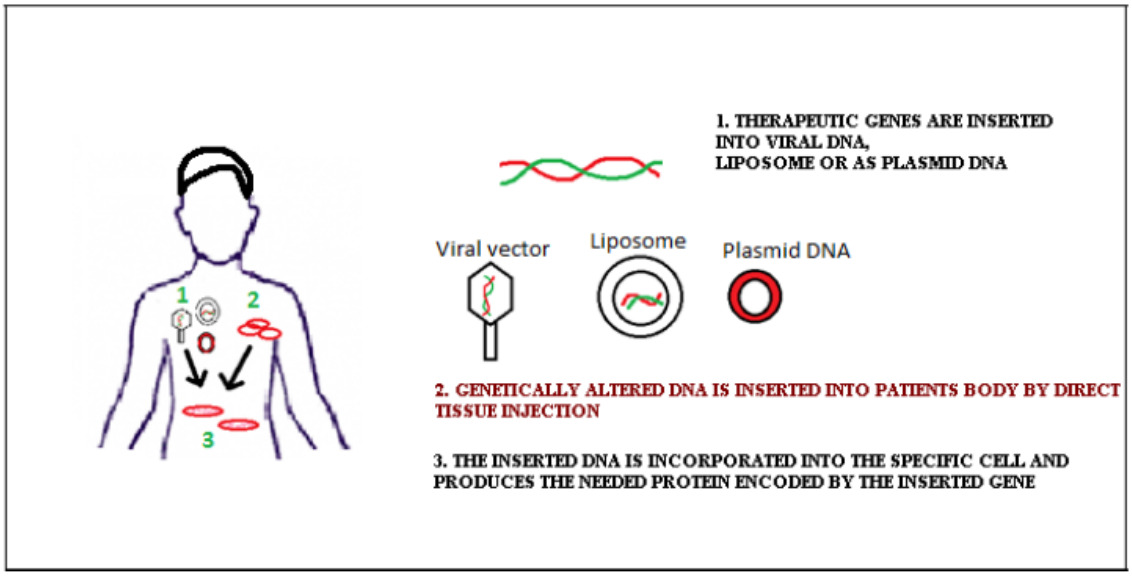
***Ex vivo* gene therapy**

- In this mode of gene therapy genes are transferred to the cells grown in culture, transformed cells are selected, multiplied and then introduced into the patient.
- The use of autologous cells avoids immune system rejection of the introduced cells.
- The cells are sourced initially from the patient to be treated and grown in culture before being reintroduced into the same individual.
- This approach can be applied to the tissues like hematopoietic cells and skin cells which can be removed from the body, genetically corrected outside the body and reintroduced into the patient body where they become engrafted and survive for a long period of time.



***In Vivo* Gene Therapy**

- *In vivo* method of gene transfer involves the transfer of cloned genes directly into the tissues of the patient.
- This is done in case of tissues whose individual cells cannot be cultured *in vitro* in sufficient numbers (like brain cells) and/or where re-implantation of the cultured cells in the patient is not efficient.
- Liposomes and certain viral vectors are employed for this purpose because of lack of any other mode of selection.
- In case of viral vectors such type of cultured cells were often used which have been infected with the recombinant retrovirus *in vitro* to produce modified viral vectors regularly. These cultured cells will be called as vector-producing cells (VPCs)). The VPCs transfer the gene to surrounding disease cells.
- The efficiency of gene transfer and expression determines the success of this approach, because of the lack of any way for selection and amplification of cells which take up and express the foreign gene.



Differences between In Vivo and Ex Vivo gene therapy Difference Between *in vivo* and *ex vivo* Gene Delivery Systems	
In vivo	***Ex vivo***
Less invasive	More invasive
Technically simple	Technically complex
Vectors introduced directly	No vectors introduced directly
Safety check not possible	Safety check possible
Decreased control over target cells	Close control possible

Target sites for Gene Therapy

Therapeutic genes have to be delivered to specific target sites for a specific type of disease. This table describes the list of such disease and their target sites for gene therapy.

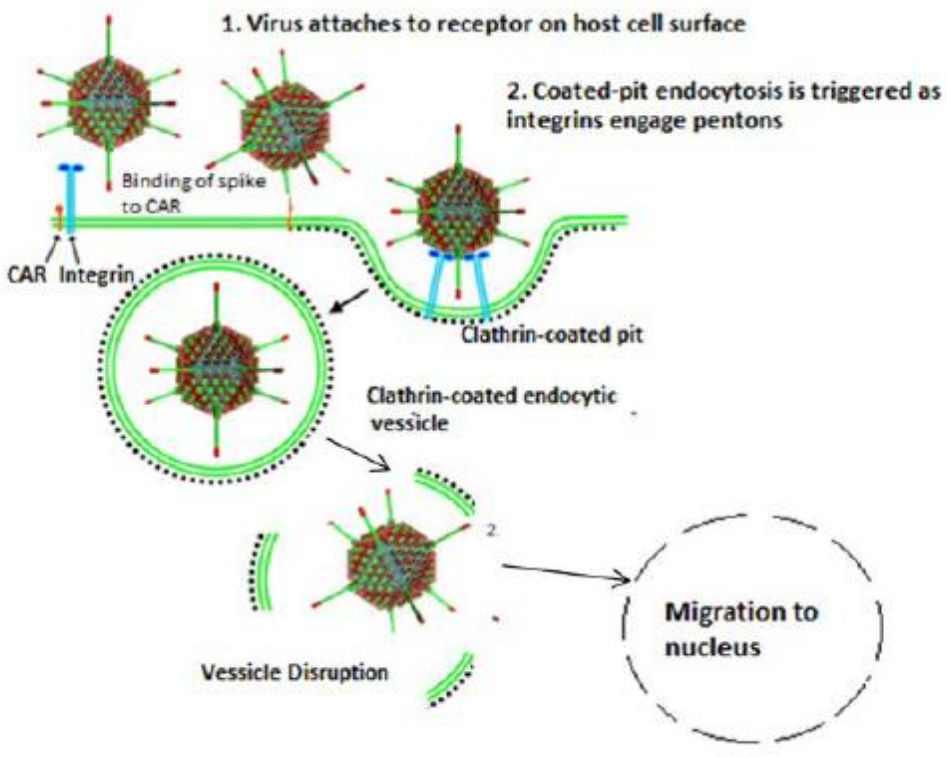
Target cells for gene transfer Target cells for gene transfer	
Disease	**Target Cells**
Cancer	Tumor cells, antigen presenting cells (APCs), blood progenitor cells, T cells, fibroblasts, muscle cells
Inherited monogenic disease	Lung epithelial cells, macrophages, T cells, blood progenitor cells, hepatocytes, muscle cells
Infectious disease	T cells, blood progenitor cells, antigen presenting cells (APCs), muscle cells
Cardiovascular disease	Endothelial cells, muscle cells
Rheumatoid arthiritis	Sinovial lining cells
Cubital tunnel Syndrome	Nerve cells

Viral vectors

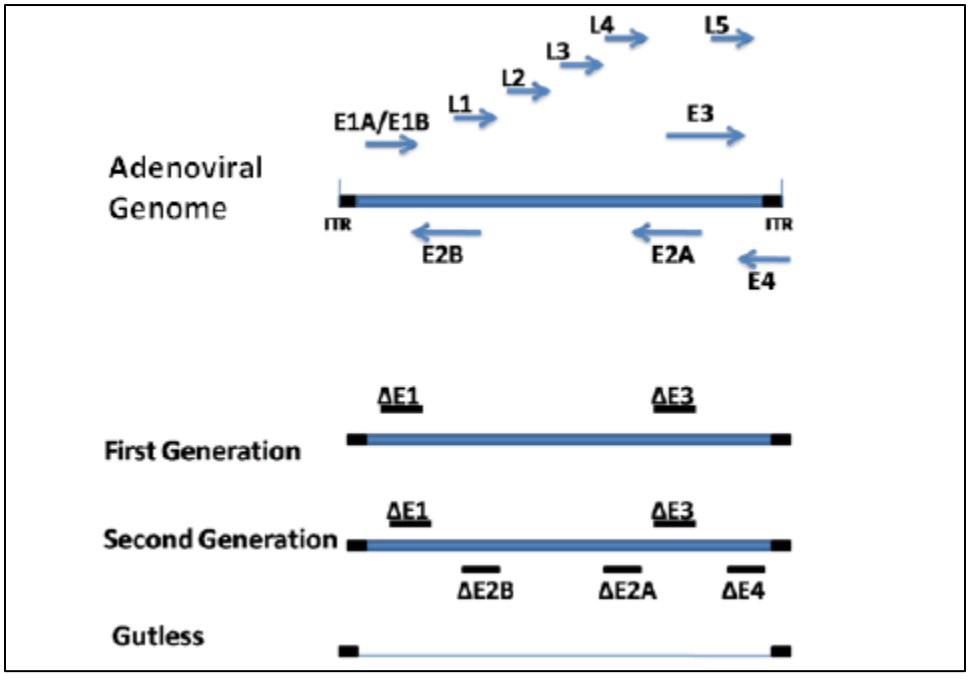
Retroviruses, adenoviruses and adeno-associated viruses (AAV) some commonly used viral vectors whereas some less commonly used viral vectors are derived from the Herpes simplex virus (HSV-1), the baculovirus etc.

Adenoviral vectors

Adenoviruses are large linear double-stranded DNA viruses that are commonly used for preparing gene transfer vectors. Adenovirus vectors are known to be the second most popular gene delivery vector for gene therapy of various diseases like cystic fibrosis and certain types of cancer. Figure shows how the adenoviruses enter cells by receptor-mediated endocytosis. A primary cellular receptor binds to viral fibre then the virus interacts with secondary receptors which are responsible for its internalization. Coxsackie and Adenovirus Receptor (CAR), Heparan sulphate glycosaminoglycans, sialic acid, CD46, CD80, CD86, alpha domain of MHC I are the primary receptors which are specific for specific strains of adenovirus. Integrins are the secondary receptors which helps in the internalization of viral particles. Some adenovirus directly interacts with integrins like in the case of fibre deficient Ad2 virions.



The adenoviral DNA has inverted terminal repeats (ITRs) and a terminal protein (TP) is attached covalently to 5' termini. The adenoviral genome is classified as early and late regions based on the proteins they express. Proteins encoded by early region (E1, E2, E3, E4) genes are involved in viral DNA replication, cell cycle modulation and defense system. The late region genes (L1, L2, L3, L4, L5) encodes the viral structural proteins. Three classes of adenoviral vectors namely first, second and third generation viral vectors are developed for gene therapy purpose.



First generation adenoviral Vectors

These vectors are constructed by replacing the E1/E3 expression cassette and inserting our candidate gene of 3-4kb size. E1 encodes proteins responsible for expressions of other viral genes required for viral growth. So cell lines that can provide E1 proteins *in trans* are required for the replication of the E1 deleted viral vectors.

Advantages:

• They are human viruses produced at very high titers in culture.

• They can infect a wide range of human cell types including non- dividing cells.

• They enter into cells by receptor mediated endocytosis with a very high transduction efficiency reaching upto 100% *in vitro*

• Their large size enables them to accept large inserts.

Disadvantages:

• Expression of foreign gene is for short period of time as they do not integrate into the chromosome.

• These vectors may generate immune response causing chronic inflammation.

Second generation adenoviral Vectors

These vectors have been developed to overcome these difficulties. Here of E1/E2 or E3/E4 expression cassettes are called deleted and replaced. The E1/E2 or E3/E4 proteins are required for

viral DNA replication. Similar to first generation viral vector, cell lines which can complement both E1and E2 or E3 and E4 are needed. It can carry DNA insert upto 10.5kb

Advantages:

• It has improved safety and increased transgene expression.

Disadvantages:

• These viral vectors are associated with immunological problems.

• Construction of these vectors is difficult.

Third generation adenoviral Vectors

These vectors are otherwise called as **gutless adenovirus.** These are also known as helper dependent adenovirus as they lack all the coding sequences and require helper virus which carries all the coding sequences. Helper virus for example AAV, or artificially disabled viruses provide the viral functions needed for successful infection like viral DNA replication, viral assembly and infection of new cells etc. The size of insert DNA can be 36kb and hence called as high capacity adenoviruses. They carry only 5' inverted terminal repeats (ITR) and 3' packaging signals (ψ).

Advantages:

• These are non-integrative and high-capacity vectors.

• It can be produced in high titer and the construction of these vector is easy.

• It shows longer stability and reduced immune response.

Disadvantages:

• Helper virus contamination contamination can cause diseases like conjunctivitis, pharyngitis, cold and respiratory disease.

Adeno- Associated Virus (AAV)

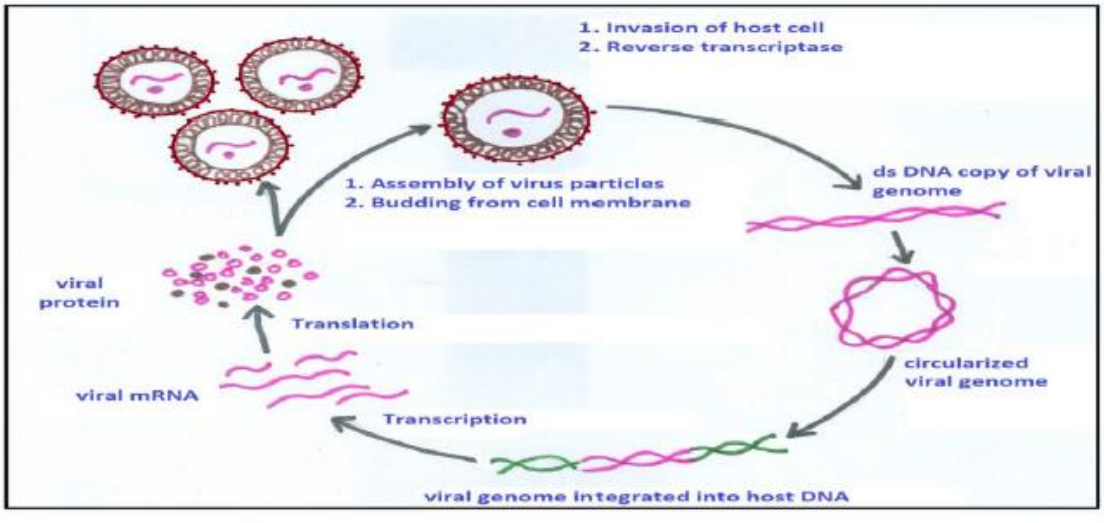
Adeno-associated viruses (AAVs) are a group of small, single-stranded DNA viruses which cannot usually undergo productive infection without co-infection by a helper virus, such as an adenovirus.

• The insert size for AAV is 4.5 kb, with the advantage of long-term gene expression as they integrate into chromosomal DNA.

• AAVs are highly safe as the recombinant adeno associated vectors contains only gene of interest and 96% viral genes are deleted.

Retroviral Vectors

Retroviruses are RNA viruses which possess a reverse transcriptase activity, enabling them to synthesize a complementary DNA. Following infection (transduction), retroviruses deliver a nucleoprotein complex (pre-integration complex) into the cytoplasm of infected cells. The viral RNA genome is reverse transcribed first and then integrates into a single site of the chromosome.



• **Tumor retroviruses**, example Moloney's murine leukemia virus (MoMuLV), is widely used for the generation of recombinant vectors. these are produced at low titers as all the viral genes are deleted.

• **Oncoretroviruses:** The cells that divide shortly after infection can only be transduced by oncoretrovirus. The preinitiation complex is excluded and their entry is restricted in to the nucleus as they can only enter when nuclear membrane dissolves during cell division the target cells for this viral vector is limited

• **Recombinant lentiviruses** are being developed that are non- pathogenic to humans and have the ability to transduce stationary cells.

Other Viral Vectors:

These include herpes simplex virus vectors and baculovirus.

Herpes simplex virus vectors: Herpes simplex virus-1 (HSV-1) is a 150 kb double-stranded DNA virus with a broad host range that can infect both dividing and non-dividing cells. the insert size is comparatively larger (>20kb) but have a disadvantage of short-term expression due to its inability to integrate into the host chromosome.

Baculovirus: They can take up very large genes and express them highly efficiently. They help in recombinant protein expression in insect cell. They can infect hepatocytes as an only mammalian cell type and the gene expresses under the control of either mammalian or viral promoter.

Simian Virus 40 Vectors (SV40): SV40 are icosahedral papovavirus with a circular double stranded DNA of 5.2kb size as genetic material. The genome encodes for early proteins viz; large T antigen (Tag) and small t antigen (tag) and late protein viz; a regulatory protein agnoprotein and three structural proteins (VP1, VP2, VP3). The Tag gene is removed as it is responsible for inducing immunogenicity in the recombinant SV40 vector. All the structural proteins except the major capsid protein VP1 is removed resulting in a genome of 0.5kb size which includes origin of replication (ori) and encapsidation sequence. Recombinant SV40 vectors allows expression of transduced gene

Non- viral vectors

It involves chemical and physical methods such as direct injection of naked plasmid DNA (particle bombardment), receptor-mediated endocytosis and gene transfer through liposomes, polymers, nano particles etc.

Some non viral methods

Direct injection/particle bombardment:

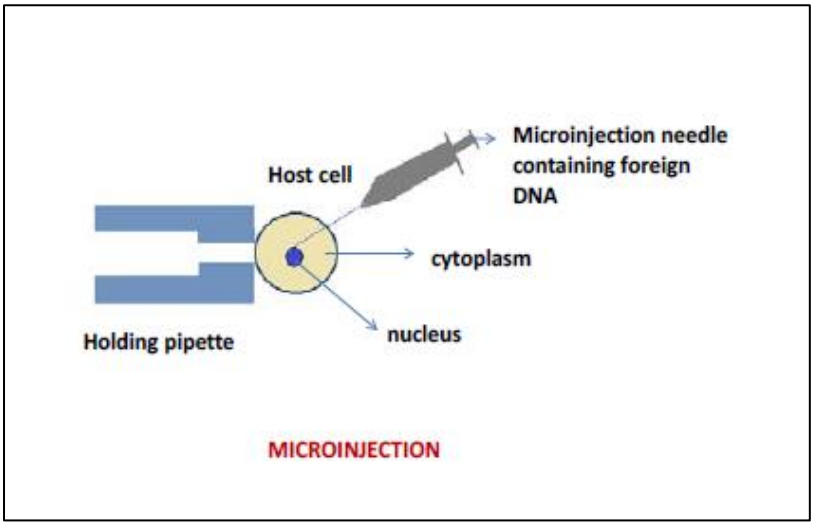
DNA can be injected parenterally which can be considered for Duchenne muscular dystrophy (DMD). An alternative approach uses particle bombardment (‘gene gun’) technique, in which DNA is coated on to metal micro particles and fired from a ballistic gun into cells/tissues. This technique is used to transfer the foreign DNA and its transient expression in mammalian cells *in vitro* and *in vivo* as well. It can cross the physical barriers like skin, muscle layer for which it is used for vaccination. Particle bombardment is used to deliver drugs, fluorescent dyes, antigenic proteins etc.

Advantage: Simple and comparatively safe.

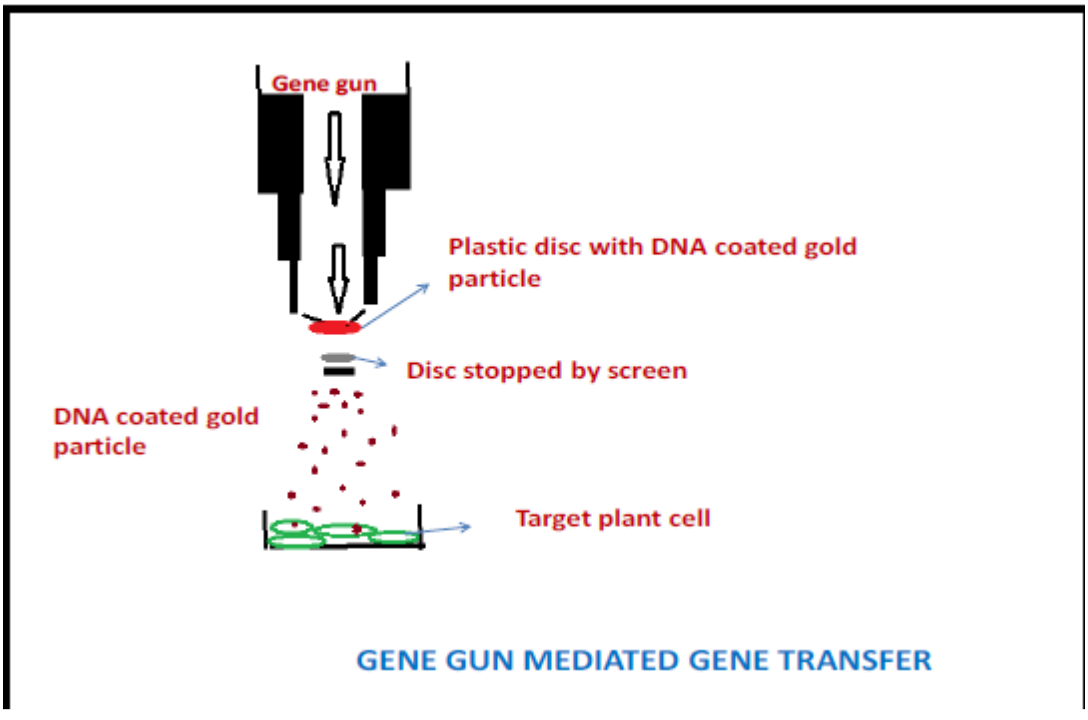
Disadvantage:

• Poor efficiency of gene transfer.

• A low level of stable integration of the injected DNA. Repeated injection may cause damage in the proliferating cells.



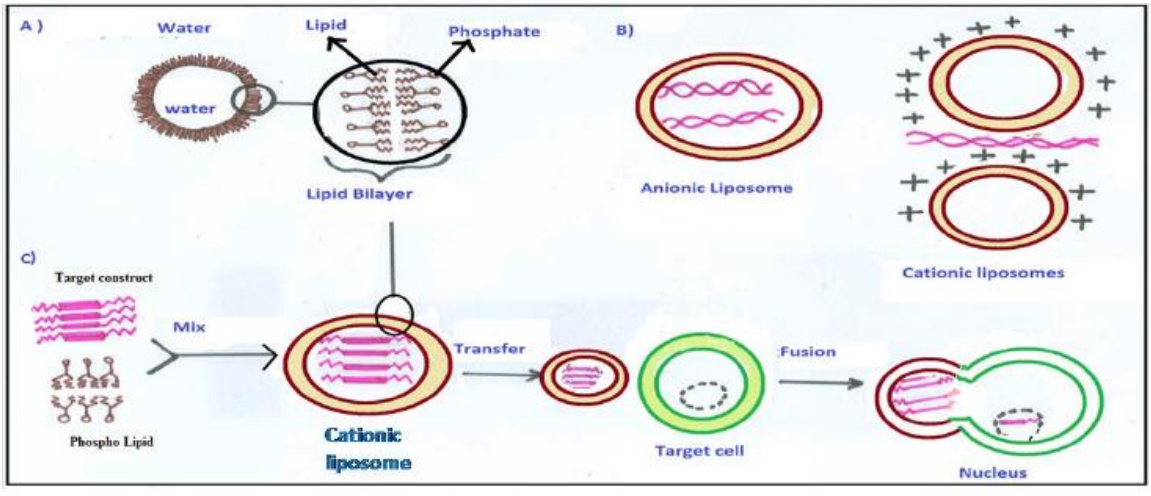
Microinjection involves the delivery of foreign DNA, by the help of glass micropipette into a living cell. The cell is held against a solid support or holding pipette and micro neeedle containing the desired DNA is inserted into the cell. The tip of the pipette used is about 0.5 to 5 micro meter diameter which resembles an injection needle. For this, glass micropipette is heated until the glass becomes somewhat liquefied and is quickly stretched to ressemble a injection needle. The delivery of foreign DNA is done under a powerful microscope (micromanupulator).



In **particle bombardment** method, the tungsten or gold particles (micro projectiles) are coated with the foreign DNA. Micro-projectile bombardment uses high-velocity metal particles to deliver biologically active DNA into the target cells. The macroprojectile is coated with the coated particles and is accelerated with air pressure and shot into plant the target tissue. A perforated plate is used, which allows the micro-projectiles to pass through to the cells on the other side of the plate and stops the macropojectile. Particle coated with the foreign gene releases the foreign gene when enters into the target cell and integrates into the chromosomal DNA. This technique is also used to transfer genes in mammalian cells. Mammalian cell lines like HEK 293, MCF7 showed gene expression when transfected with luciferase and green fluorescent genes and their gene expression was dependent on helium pressure, size and amount of gold particle and DNA load on each particle. Cell viability depends on helium pressure.

Liposomes Mediated

Liposomes are spherical vesicles which are made up of synthetic lipid bilayers which imitate the structure of biological membranes. DNA to be transferred is packaged into the liposome *in vitro* and transferred to to the targeted tissue. The lipid coating helps the DNA to survives *in vivo* and enters into the cell by endocytosis. Cationic liposomes, where the positive charge on liposomes is stabilized by binding of negatively charged DNA, are popular vehicles for gene transfer *in vivo.*



Advantage:

• The liposomes with the foreign DNA are easy to prepare.

• There is no restriction in the size of DNA that is to be transferred.

Disadvantage:

• Efficiency of gene transfer is low and transient expression of the foreign gene is obtained as they are not designed to integrate into the chromosomal DNA.

Electroporation

In electroporation, the external electric field is applied to the protoplast, which changes the electrical conductivity and the permeability of cell membrane; and thus the exogenous molecules found in the medium are taken up to either the cytoplasm (transient transfection) or into the nucleus (stable transfection).The efficiency of electroporation can be increased by giving the cell a heat shock, prior to the application of electric field or by using small quantity of PEG while doing electroporation. See lecture 3 of module 5 for detailed explanation.

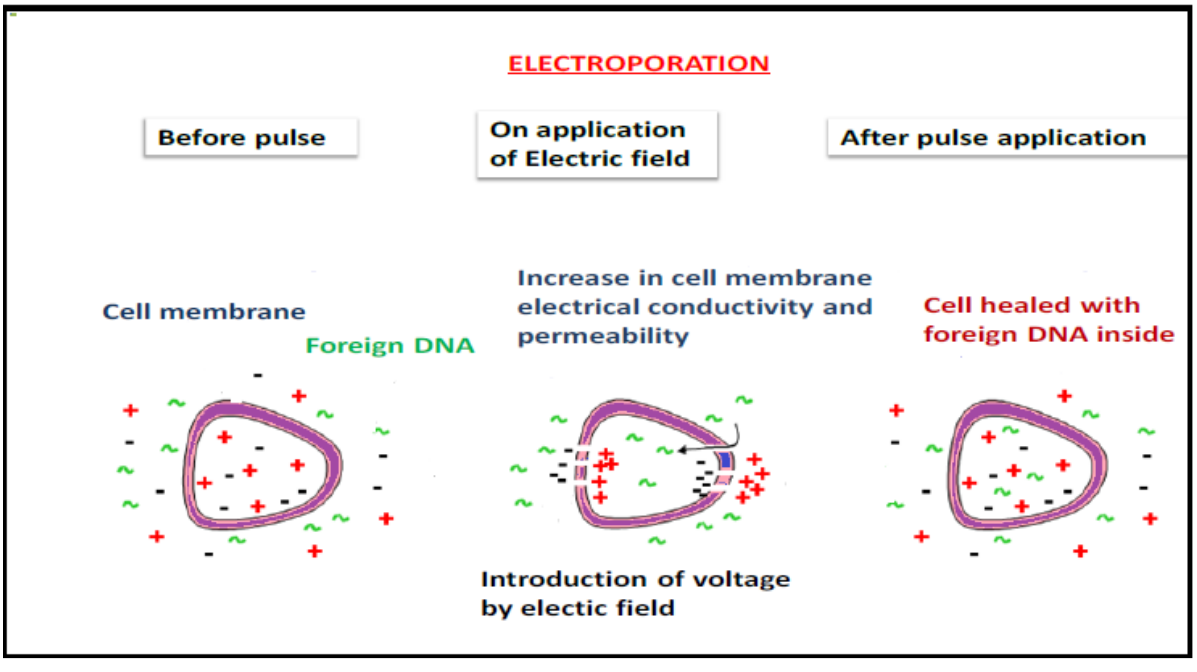
Advantage:

• By electroporation large numbers of cells can be processed at once, and thus the amount of time spent processing cells can be cut down.

Disadvantages:

• If the voltage applied is not calculated properly, the cells may damage.

• If electroporation does not occur in controlled environment, the potentially harmful substances can enter the cell or the impurities from solution may enter. This is because there is no way to control what enters the cell membrane.



Advantages of Gene Therapy

• Gene therapy can cure genetic diseases by addition of gene or by removal of gene or by replacing a mutated gene with corrected gene.

• Gene therapy can be used for cancer treatment to kill the cancerous cells.

• Gene expression can be controlled.

• Therapeutic protein is continuously produced inside the body which also reduces the cost of treatment in long term.

Gene Silencing

Gene silencing is a term of gene regulation used to describe the "switching off" of a gene by a mechanism without introducing any genetic modification. DNA is transcribed into mRNA but the mRNA is never translated into proteins. Gene regulation can be done either at the transcriptional level or at post-transcriptional level. In case of transcriptional level, it is done by inducing modification in histone protein, changing the environment for the binding of transcriptional machineries such as RNA polymerase, transcription factors, etc. However, in case of post-transcriptional level of gene regulations the transcribed mRNA by a particular gene is being blocked or destroyed. The post-transcriptional level of gene silencing is achieved by

1. Antisense Technology
2. RNA interference (RNAi)

Antisense Technology

Antisense technology talks about the production of complementary nucleic acid molecules against the mRNA molecule transcribed from the DNA in order to stop the translation into protein. These complementary molecules can be synthetically produced and delivered inside the cell to block the expression of diseased protein. It can be a short length of either RNA or DNA which commonly termed as Antisense oligonucleotides (AON). Here antisense refers the complementary nature of the synthetic molecule with respect to mRNA. When these AON inserted inside the cell it forms RNA duplex (i.e. double stranded RNA or RNA-DNA duplex). The formation of double stranded RNA inhibits gene expression at translation level as protein synthesis requires single stranded mRNA molecule as a template. This phenomenon is still not well understood but the current hypothesis about this is following-:

- blocking RNA splicing,
- accelerate the degradation process of the RNA and it also prevents the introns from splicing
- preventing the migration of mRNA from nucleolus to cytoplasm

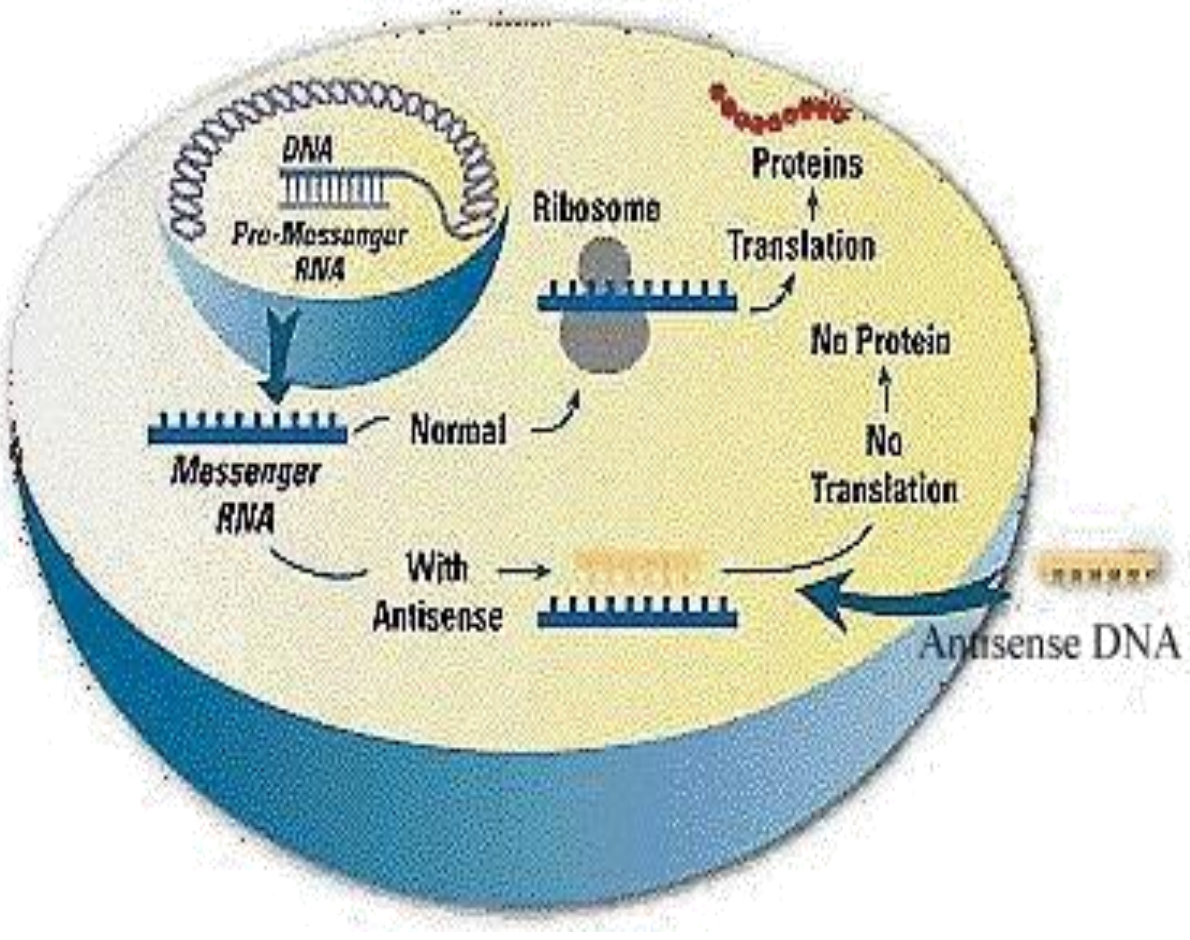
• stopping the translation of diseased protein, and

• If complementary DNA molecule is used there may be a formation of triplex in DNA template.

Mechanism of Antisense Technology

• The synthetic AON introduced inside the cell according to the gene of interest.

• If it is a DNA molecule it binds with the DNA inside the nucleus to form a triplex which inhibits the transcription and finally translation. Sometimes RNA-DNA heterodimer is also formed to stop the translation. In case of antisense RNA, it binds with mRNA to stop the translation.



Application of Antisense Technology

• In oncology antisense RNA has been used to inhibit many target proteins, such as growth factors receptors, growth factors, proteins responsible for invasion of cancerous cells and cell cycle proteins.

• If it is complementary to viral RNAs then may help in controlling various types of viral infections.

• Development of animal models for long-term normal blood pressure in hypertensive animals .

• In January 2013 a drug called mipomersen (trade name Kynamro) got approval from the FDA for curing homozygous hypercholesterolemia.

• Fomivirsen, an antiviral drug developed for the treatment of cytomegalovirus retinitis is basically an antisense oligonucleotide

RNA interference

RNAi has shown its importance in the analysis of gene functions and silencing of gene for therapeutic purpose. It was first reported by Andrew Fire and his team in the year of 1997 while studying the introduction of dsRNA into *C. elegans* for silencing a gene unc-22 gene. RNA interference (RNAi) is basically a post-transcriptional phenomenon which may be triggered by providing a double-stranded RNA (dsRNA) which is known as double RNA activation. Two types of small RNA molecules – microRNA (miRNA) and small interfering RNA (siRNA) plays a central role in RNA interference-based gene silencing. RNAi looks very similar to plant posttranscriptional gene-silencing (i.e.PTGS) and quelling in case of fungi.

Functions of RNAi are as follows:

• Immunity: the immune response to viruses and other foreign genetic material

(In case of plants)

• Down regulation of genes through mi RNA (micro RNA)

• Up-regulation of genes by using both siRNA and miRNA complementary to parts of a promoter.

Micro RNA (miRNA) and Small Interfering RNA (siRNA)

Both are considered as interfering RNA. Historically miRNA was discovered in 1993 by Ambros and his co-workers whereas siRNAs concept came in 1999 from another discovery in which a dsRNA showed its role in post-transcriptional gene silencing (PTGS) in plants by David Baulcombe's group. During PTGS at one stage there is a role of $\sim 20 - 25$ nt RNAs in silencing which was produced by the dsRNA. miRNAs is considered as regulators of cellular self-genes(i.e. endogenous genes), and siRNAs act as guards of foreign or invasive genes coming from viruses, transposons, and transgenes etc. which try to get integrated into host genome. The miRNA gene is always present in the host gnome which gets transcribed into primary-miRNA (pri-miRNA) first with the help of RNA polymerase II. This pri-miRNA is cleaved by an enzyme called Drosha which is a type of ribonuclease III enzyme. It liberates approximately 60 to 70 nt looped structure which is consider as precursor miRNA or pre- miRNA. This pre-miRNA is the transported with the help of Exportin 5 present in cytoplasm. Once the pre-miRNA is exported into cytoplasmic space another dsRNA specific enzyme called Dicer helps in duplexing with other miRNA. The unwinding of the duplexed miRNA is done by helicase. Now the both dsRNA-specific endonucleases enzymes (Drosha and Dicer) help to generate 2-nucleotide-long-3' overhangs near the cleavage site. After unwinding of the double stranded miRNA, the generation of target specific Guide strand and the passenger stand. Now the miRNA (i.e. Guide strand) is considered as mature miRNA which is then incorporated with the RNA-induced silencing complex (RISC). The target specific miRNA now binds with the mRNA and stops the translation. Finally, the gene is silenced with the help of miRNA and the cell undergo self-destruction pathway.

Mechanism of RNAi based Gene Silencing

A plasmid vector along with the target construct has to be delivered inside the diseased cell. This vector is able to transcribe a double stranded shRNA (short hairpin RNA). This shRNA is first processed into siRNA (small interfering RNA) and then siRNA inhibit the mRNA translation by sequence specific degradation process thereby silencing the gene. In first step of formation of siRNA the shRNA bind to a ribonuclease enzyme (similar to RNase III) and cleaved into 21 to 25 nucleotide siRNA. These siRNAs are complexed with RNA Interference Specificity Complex (RISC). RISC helps siRNA to find the mRNA complementary sequence and formation of the duplex.

The introduced plasmid has expressed a short hairpin RNA (shRNA). It requires Exportin 5 for the nuclear export.

2. Transactivating response (TAR) RNA-binding protein (TRBP) complexes with Dicer to form a dimer and then attaches to the shRNA.

3. Dicer generates 19-23 nucleotides siRNA and 2 nucleotides with 3' overhangs in one step only from the shRNA attached to the complex

4. Argonaute 2 (Ago 2) is a RNase which belongs to AGO subfamily and binds to the 3' overhang of siRNA in the RNA silencing complex resulting in unwinding of the dsRNA.,

5. The strand to which Ago 2 binds is called the guide strand and the other strand is known as "passenger strand". The latter is cleaved by Ago2.

6. Now the "passenger strand" becomes free to leave the complex.

7. The integrated "guide strand" is now known as the active RISC (RNA Interference Specificity Complex). The RISC that contains various other argonautes and also few argonaute-associated proteins.

8. The siRNA sequence remained in the complex (i.e. guide strand) help the RISC to find the mRNA and bind at the correct location.

9. Now the RISC bring the mRNA to a processing bodies (i.e. P- bodies or GW-bodies). It is a cytoplasmic focus where mRNA decay factors are in high concentration which leads to the mRNA degradation.

10. The mRNA in the P-body is now cleaved by Argonaute 2 and then degraded.

Application of RNAi

1. Tool for studying gene expression and regulations

2. Medical applications: Treatment of viral infections, cardiovascular diseases, cancer and metabolic disorders.

Examples:

− Treatment of age-related macular degeneration (AMD)

− RNAi is used to block production of VEGF (Vascular endothelial growth factor).

− Treatment of Hypercholesterolemia

− To block the production of LDL particles.

Advantage of RNAi

• It is target specific

• Very small amount of small dsRNA is sufficient for silencing gene expression

• A natural method of gene regulation

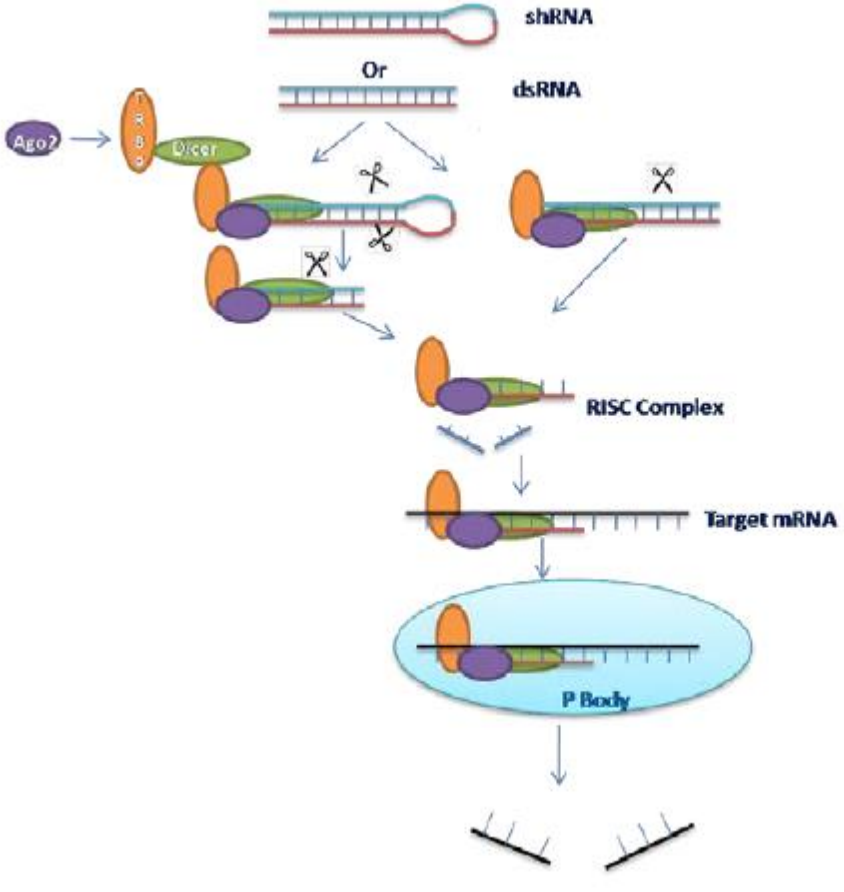
Limitations of RNAi

• Less is known regarding machinery and mechanism of RNA.

• Uses of siRNA to therapeutic purposes is now more concerns about the subject safety as there may be a chance of disturbances in natural regulation of the immune system.

• Problem of-target effects is not clear at present

• Not much is known regarding Dosage requirement, stabilization & synthesis of tailored/engineered mRNA.



miRNA versus siRNA

	Configuration Length	**Length**	**Complementary to target mRNA**	**Biogenesis**	**Action**	**Function**	**Clinical uses**
Micro RNA (miRNA)	Occur naturally in plants and animals Single stranded	19-25 nt	Not exact, and therefore, a single miRNA may target up to hundreds of mRNAs	Expressed by genes whose purpose is to make miRNAs, but they regulate genes (mRNAs) other than the ones that expressed them	Inhibit translation of mRNA	Regulators (inhibitors) of genes (mRNAs)	Possible therapeutic uses either as drug targets or as drug agents themselves. Expression levels of miRNAs can be used as potential diagnostic and biomarker tools

Short interfering RNA (siRNA)	Occur naturally in plants and lower animals. Whether or not they occur naturally in mammals is an unsettled question	Double stranded	21-22 nt	100% perfect match, and therefore siRNAs knock down specific genes, with minor off-target exceptions	Regulate the same genes that express them	Cleave mRNA	Act as gene-silencing guardians in plants and animals that do not have antibody-or cell-mediated immunity	siRNAs are valuable laboratory tools used in nearly every molecular biology laboratory to knock down genes. Several siRNAs are in clinical trials as possible therapeutic agents

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

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – IV – SBB3102 – ADVANCED GENETIC ENGINEERING

GENETICALLY MODIFIED ORGANISMS.

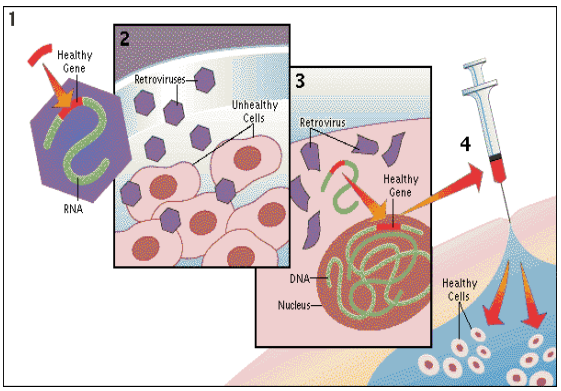
Gene therapy

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

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

Steps involved;

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



Non-viral Options

1. Direct introduction of therapeutic DNA
    a. But only with certain tissue
    b. Requires a lot of DNA
2. Creation of artificial lipid sphere with aqueous core, liposome
    a. Carries therapeutic DNA through membrane
3. Chemically linking DNA to molecule that will bind to special cell receptors

a. DNA is engulfed by cell membrane
b. Less effective ☹

4. Trying to introduce a 47th chromosome
    a. Exist alongside the 46 others
    b. Could carry a lot of information
    c. But how to get the big molecule through membranes?

Current Status

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

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

Problems with Gene Therapy

- Short Lived
    - Hard to rapidly integrate therapeutic DNA into genome and rapidly dividing nature of cells prevent gene therapy from long time
    - Would have to have multiple rounds of therapy
- Immune Response
    - new things introduced leads to immune response
    - increased response when a repeat offender enters
- Viral Vectors
    - patient could have toxic, immune, inflammatory response
    - also may cause disease once inside
- Multigene Disorders
    - Heart disease, high blood pressure, Alzheimer's, arthritis and diabetes are hard to treat because you need to introduce more than one gene
- May induce a tumor if integrated in a tumor suppressor gene because insertional mutagenesis

DNA vaccine

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

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

DNA vaccines are *third generation vaccines*, and are made up of a small, circular piece of bacterial DNA (called a plasmid) that has been genetically engineered to produce one or two specific proteins (antigens) from a micro-organism. The vaccine DNA is injected into the cells of the body, where the "inner machinery" of the host cells "reads" the DNA and converts it into pathogenic proteins. Because these proteins are recognised as foreign, they are processed by the host cells and displayed on their surface, to alert the immune system, which then triggers a range of immune responses. These DNA vaccines developed from “failed” gene therapy experiments. The first demonstration of a plasmid-induced immune response was when mice inoculated with a plasmid expressing human growth hormone elicited antibodies instead of altering growth.

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

Advantages and disadvantages of DNA vaccines

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

Plasmid vectors for use in vaccination

Vector design

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

Plasmids also include a strong polyadenylation/transcriptional termination signal, such as bovine growth hormone or rabbit beta-globulin polyadenylation sequences. Multicistronic

vectors are sometimes constructed to express more than one immunogen, or to express an immunogen and an immunostimulatory protein.

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

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

Vaccine insert design

Immunogens can be targeted to various cellular compartments in order to improve antibody or cytotoxic T-cell responses. Secreted or plasma membrane-bound antigens are more effective at inducing antibody responses than cytosolic antigens, while cytotoxic T-cell responses can be improved by targeting antigens for cytoplasmic degradation and subsequent entry into the major histocompatibility complex (MHC) class I pathway. This is usually accomplished by the addition of N-terminal ubiquitin signals

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

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

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

The method of delivery determines the dose of DNA required to raise an effective immune response. Saline injections require variable amounts of DNA, from 10 μg-1 mg, whereas gene gun deliveries require 100 to 1000 times less DNA than intramuscular saline injection to raise an effective immune response. Generally, 0.2 μg – 20 μg are required, although quantities as low as 16 ng have been reported. These quantities vary from species to species, with mice, for example, requiring approximately 10 times less DNA than primates. Saline injections require more DNA because the DNA is delivered to the extracellular spaces of the target tissue (normally muscle), where it has to overcome physical barriers (such as the basal lamina and large amounts of connective tissue, to mention a few) before it is taken up by the cells, while gene gun deliveries bombard DNA directly into the cells, resulting in less “wastage”.

Another approach to DNA vaccination is expression library immunization (ELI). Using this technique, potentially all the genes from a pathogen can be delivered at one time, which may be useful for pathogens which are difficult to attenuate or culture. ELI can be used to identify which of the pathogen’s genes induce a protective response. This has been tested with

Mycoplasma pulmonis, a murine lung pathogen with a relatively small genome, and it was found that even partial expression libraries can induce protection from subsequent challenge.

Summary of Plasmid DNA delivery methods

Method of Delivery		**Formulation of DNA**	**Target Tissue**	**Amount of DNA**
Parenteral	Injection (hypodermic needle)	Aqueous solution in saline	IM (skeletal); ID; (IV, subcutaneous and intraperitoneal with variable success)	Large amounts (approximately 100-200 μg)
	Gene Gun	DNA-coated gold beads	ED (abdominal skin); vaginal mucosa; surgically exposed muscle and other organs	Small amounts (as little as 16 ng)
	Pneumatic (Jet) Injection	Aqueous solution	ED	Very high (as much as 300 μg)
Topical application		Aqueous solution	Ocular; intravaginal	Small amounts (up to 100 μg)
Cytofectin-mediated		Liposomes (cationic); microspheres; recombinant adenovirus vectors; attenuated Shigella vector; aerosolised cationic lipid formulations	IM; IV (to transfect tissues systemically); intraperitoneal; oral immunization to the intestinal mucosa; nasal/lung mucosal membranes	variable

Advantages and disadvantages of commonly used DNA vaccine delivery methods

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

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

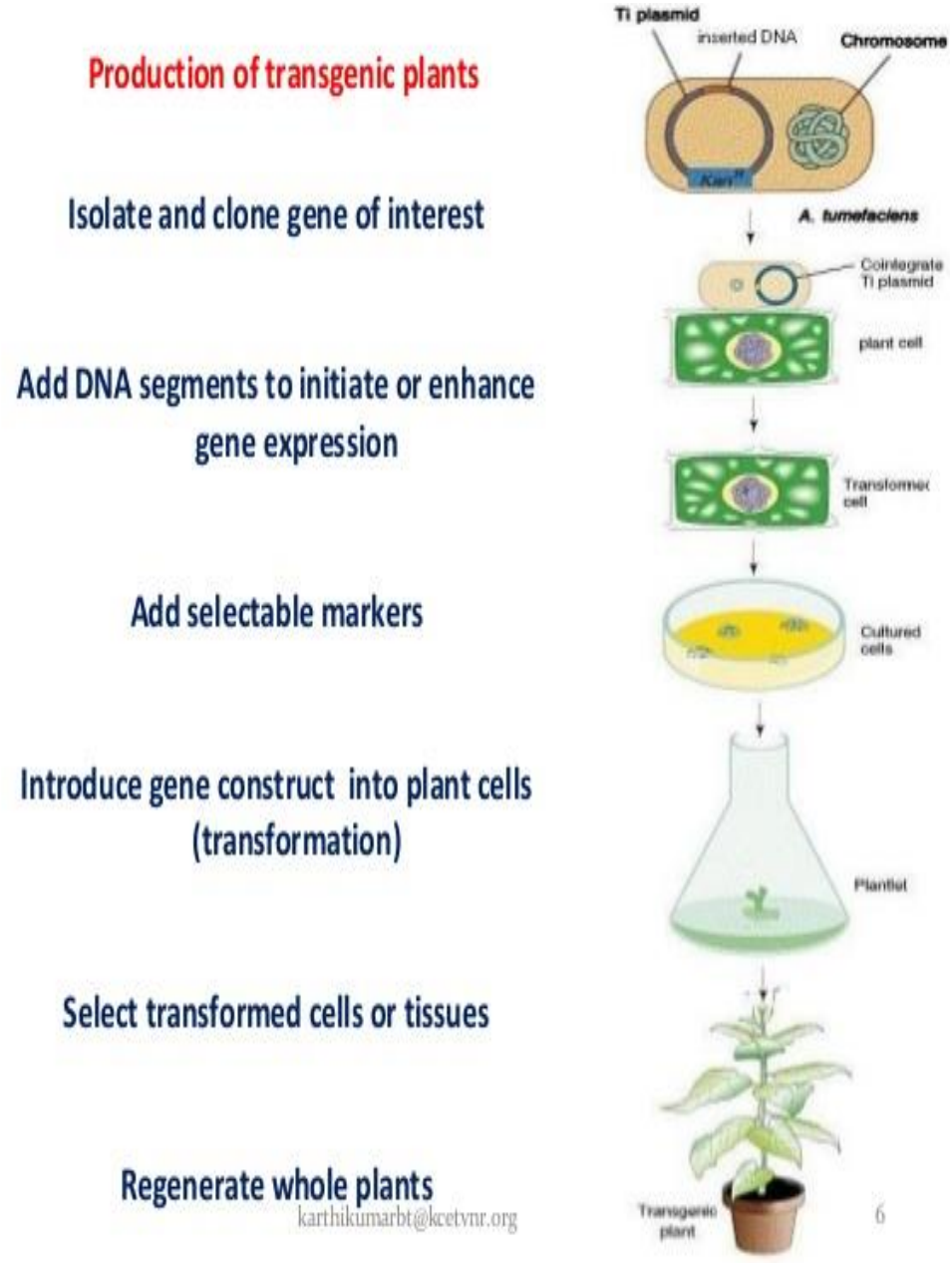
Transgenic plants

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

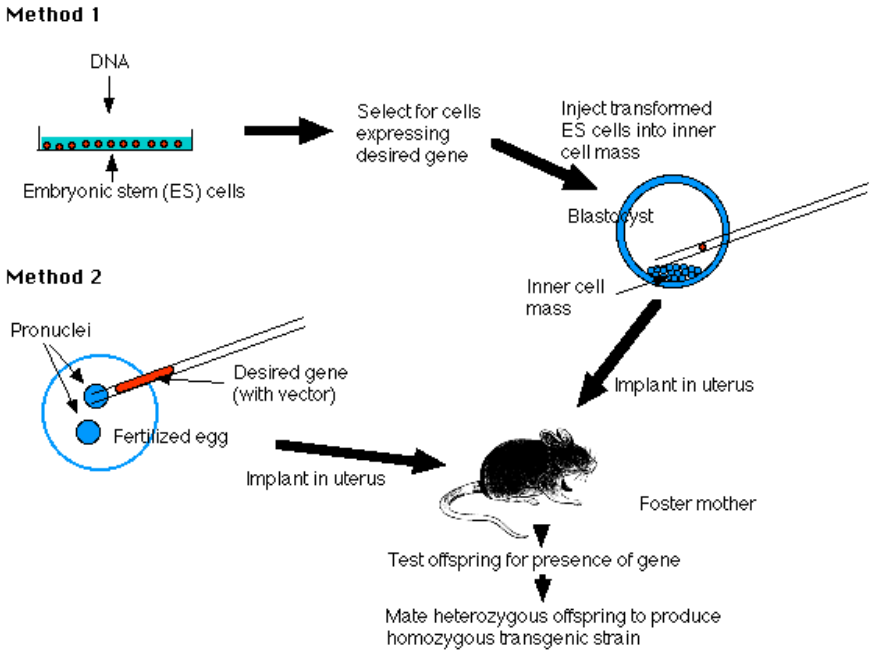
Transgenic plants have been developed for a variety of reasons: longer shelf life, disease resistance, herbicide resistance, pest resistance, and improved product quality. The first transgenic crop approved for sale in the US, in 1994, was the FlavrSavr tomato, which was intended to have a longer shelf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

Today there are more than 67.7 million hectares (677,000 km^2) of transgenic plants being grown throughout the world[1]. There are three general types of transgenic plants; those with genes to improve the quality of the product, those with genes to allow them to resist disease or herbivory (consumption by herbivores, usually insects), and plants with genes that allow them to be resistant to the effects of specific herbicides.

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



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



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

Production of Insulin:

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

two inter-chain and one intra-chain disulphide bridge. Insulin is synthesized as a single-chain precursor, pro-insulin, and produced by the proteolytic processing of pro-insulin in the pancreas .

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

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

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

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

Production of Growth Hormones:

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

Human growth hormone has versatile functions:

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

• Helps in fastening the production of DNA and RNA.

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

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

• Activates bone growth and skeletal development indirectly by producing intermediate factor IGF-1.

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

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

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

UNIT – V – SBB3102 – ADVANCED GENETIC ENGINEERING

Hazards and impact of GMOs

GUIDELINES

Definition of recombinant DNA: Recombinant deoxyribonucleic acid (rDNA) by definition involves *in vitro* introduction of different segments of DNA (one being the vector and the others normally unrelated DNA sequences) that are capable of replication in a host cell either autonomously or as an integral part of host's genome and maintenance of their continued propagation. This will include all types of cell fusion, microinjection of DNA or RNA or parts or all of chromosomes, genetic engineering including self-cloning and deletion as well as cell hybridization, transformation and other types of virus or pathogen introduction into unnatural hosts. The organisms involved may belong to these categories:

Intergeneric organisms

Well defined organisms with non-coding regulatory regions

Biological agents whose source of DNA is a pathogen

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

Classification of a pathogenic microorganisms

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

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

Characterisation of donor and recipient organisms

i) Characterisation of the modified organism

ii) Expression and properties of the gene product

Containment

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

Purpose of containment

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

Types of containment

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

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

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

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

i) Laboratory practice and technique**:**

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

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

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

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

These levels and the appropriate conditions are enumerated as follows:

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

Biosafety Level 2: These practices, safety equipment and facilities are applicable in clinical, diagnostic, teaching and other facilities in which work is done with the broad spectrum of indigenous moderate-risk agents present in the community and associated with human disease

of varying severity. Laboratory workers are required to have specific training in handling pathogenic agents and to be supervised by competent scientists. Accommodation and facilities including safety cabinets are prescribed, especially for handling large volume are high concentrations of agents when aerosols are likely to be created. Access to the laboratory is controlled.

Biosafety level 3: These practices, safety equipment and facilities are applicable to clinical, diagnostic, teaching research or production facilities in which work is done with indigenous or exotic agents where the potential for infection by aerosols is real and the disease may have serious or lethal consequences. Personnel are required to have specific training in work with these agents and to be supervised by scientists experienced in this kind of microbiology. Specially designed laboratories and precautions including the use of safety cabinets are prescribed and the access is strictly controlled.

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

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

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

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

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

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

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

Category II: Those requiring prior intimation of competent authority.

Experiments falling under containment levels II, III and IV.

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

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

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

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

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

(ii) Cloning of genes for vaccine production: e.g. Rinderpest and leprosy antigens. Rinderpest has been classified under Risk Group II in view of the common incidence of the disease in India, though it is listed under special category in the

Centres for Disease Control & National Institute of Health (CDC-NIH) system. Similarly, leprosy afflicts a large segment of population which calls for concerted programme to control the disease by vaccination and detection at early stages through immunodiagnostic tests. The containment should be decided by Review Committee on Genetic Manipulation (RCGM) on a case by case basis on experiment utilising DNA from non-defective genomes of organisms recognised as pathogen. In view of no demonstrated risk from handling free *M. laprae* antigens, inactivated whole cells as well as antigens can be assigned to Risk Group I. The details of the rDNA technology in development of vaccines for human and animal health giving containment conditions for observance of safeguards in large scale operations are given in Chapter V-B.

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

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

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

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

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

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

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

(x) Transgenosis in animal experiments : Transgenosis method is used to transform animal cells with foreign DNA by using viruses as vectors or by microinjection of DNA into eggs and pre-embryos. The expression of an inserted gene can be influenced both by the regulatory sequences associated with the gene and the sequences present at the site of integration of host genome. At present, there is no way to control where a gene is inserted into the chromosome of either an animal or

plant cell. Yet this site of insertion can affect not only the expression of the interested gene but also the regulation of the host cells- DNA e.g. by non-specific activation of cellular protooncogenes.

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

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

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

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

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

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

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

MECHANISM OF IMPLEMENTATION OF BIOSAFETY GUIDELINES

For implementation of the guidelines it is necessary to have an institutional mechanism to ensure the compliance of requisite safeguards at various levels. The guidelines prescribe specific actions that include establishing safety procedures for rDNA research, production and release to the environment and setting up containment conditions for certain experiments. The guidelines suggest compliance of the safeguards through voluntary as well as regulatory

approach. In this connection, it is proposed to have a mechanism of advisory and regulatory bodies to deal with the specific and discretionary actions on the following:

a. Self regulation and control in the form of guidelines on recombinant research activities; and
b. Regulation of large scale use of engineered organisms in production activity and release of organisms in environmental applications under statutory provisions.

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

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

Scope and functions of advisory committee and statutory body

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

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

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

2. Implementation Committees:

2.1 Institutional Biosafety Committee (IBSC)

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

Head of the Institution or nominee

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

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

(iii)One member nominated by DBT.

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

The biosafety functions and activity include the following:

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

IBSC will provide half yearly report on the ongoing projects to RCGM regarding the observance of the safety guidelines on accidents, risks and on deviations if any.

A computerised Central Registry for collation of periodic report on approved projects will be set up with RCGM to monitor compliance on safeguards as stipulated in the guidelines.

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

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

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

iii) Training of personnel on biosafety.

iv) Instituting health monitoring programme for laboratory personnel.

Complete medical check-up of personnel working in projects involving work with potentially dangerous microorganisms should be done prior to starting such

projects. Follow up medical checkups including pathological tests should be done periodically, at least annually for scientific workers involved in such projects. Their medical records should be accessible to the RCGM. It will provide half yearly reports on the ongoing projects to RCGM regarding the observance of the safety guidelines on accidents, risks and on deviations if any.

v) Adopting emergency plans.

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

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

The RCGM will have the functions:

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

i) To review the reports in all approved ongoing research projects involving high risk category and controlled field experiments, to ensure that safeguards are maintained as per guidelines.
ii) To recommended the type of containment facility and the special containment conditions to be followed for experimental trials and for certain experiments.
iii) To advise customs authorities on import of biologically active material, genetically engineered substances or products and on excisable items to Central Revenue and Excise.
iv) To assist Department of Industrial Development, Banks towards clearance of applications in setting up industries based on genetically engineered organisms.
v) To assist the Bureau of Indian Standards to evolve standards for biologics produced by rDNA technology.
vi) To advise on intellectual property rights with respect to rDNA technology on patents.

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

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

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

The functions include giving approval from environmental angle on:

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

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

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

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

The composition of the Committee would be as follows:

Chairman - Additional Secretary, Department of Environment

Co-Chairman - Expert Nominee of Secretary, DBT.

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

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

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

i) Monitoring of large scale release of engineered organisms/products into environment, oversee field applications and experimental field trials.

ii) To provide information/data inputs to RCGM upon surveillance of approved projects under industrial production, and in case of environmental releases with respect to safety, risks and accidents.

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

GENETICALLY MODIFIED ORGANISMS

For centuries, humans have altered plants and animals by selective reproduction (breeding, hybridizing). As a result, we have a wide range of domestic animals and plants grown for food and for a variety of non- food use (such as for fibers and decorative purposes and as a source of fuel). These efforts to adjust the characteristics of organisms in nature do not involve direct genetic modification by humans, but involve human actions working with existing natural processes for selection of traits. These traits are in the genes, so there are some differences in the genes of the original and modified versions of the plants and animals.

Direct genetic modification is a relatively new process based on a set of technologies that alter the genetic makeup of living organisms, including animals, plants, bacteria, or fungi by inserting genes rather than using cross-breeding and selection techniques. The purpose of the modification of the genes is to derive certain benefits. Genetic modification is accomplished by inserting one or more genes from one organism into a different organism (for example, from bacteria into a plant or from one species of plant into another). Combining genes from different organisms is known as recombinant DNA technology ("gene splicing"), and the resulting organism is said to be "genetically modified," "genetically engineered," or "transgenic." The end product we use may be part of the genetically modified organism itself (e.g., the beans of the soy plant) or something produced by the modified organism (for example, a drug produced by fermentation using modified bacteria or fungi).

SOME QUESTIONS; ETHICAL AND PRACTICAL MATTERS

The rapid introduction of these genetic engineering technologies has posed the serious question of whether we are rushing into an area of potential danger without giving it adequate thought. One can raise the "Jurassic Park" specter of messing with DNA and having the results come back to "bite you" just when you thought everything was going so well and that the few initial problems were resolved.

Some of the fears are generated by and kept active among groups who discuss the matter without having any significant training in biology and who may be relying on misconceptions and incorrect information. Science fiction writers and movie makers may help play into these fears as will authors of "non- fiction" works who sensationalize and misrepresent certain scientific concerns. Real problems with genetic engineering can then blend into the imagined scenarios, to make people become very agitated.

There have also been a number of technology scares that lead to a general concern about moving into new areas like genetic engineering. One can think of nuclear weapons and the threat of global destruction; nuclear power and the problems of Chernobyl and Three Mile

Island; the excessive use of the pesticide DDT and the potentially devastating outcomes for animals and humans that was averted by banning it; the contamination of water supplies with mercury and contamination of the soil with lead; the possibility that continued global warming and serious adverse consequences due to human activities, and so on. In these instances, drastic actions have been considered-and sometimes utilized- in order to prevent catastrophes, sometimes after a wake-up call from a limited disaster. In addition, there have been worries about human activities even from the non-technological end, such as decimating forests. So, some people simply want humans to back off from altering nature and leave things alone to the extent possible; that includes not altering the DNA of organisms.

On the other hand, most people like the numerous benefits of technology and simply want it to be used sensibly. Instead of protesting against technological innovations, they want to be assured that reasonable safeguards are in place. Some of the genetic modifications have raised virtually no objections, such as the use of bacteria to make insulin and the use of yeasts to make enzymes for cheese. These technologies are unlikely to stir up much controversy because the genetically modified organisms stay inside the factory and no problems have yet been detected. Some of the fears about genetically modified foods are not consistent with our knowledge of biology and toxicology. For example, eating a food that includes a protein (such as the one serving as a natural insecticide or protecting against herbicides) doesn't appear to pose any threat to humans (that protein already existed in nature and was present in small quant ities in some foods). The protein is not toxic to humans and is broken down, like other proteins, into amino acids that nourish the body. The gene is not going to change human genes. Nor is the gene within the food we eat going to mutate into a virus or other pathogen. Eating genetically modified soybeans will not have a direct adverse effect on the person eating them. However, the underlying issue of worrying about eating genetically modified foods is not entirely without certain merits, in that certain genetic modifications might affect humans. A

process was developed to make soybeans a richer source of nutrition by adding a gene from Brazil nuts. The purpose was to make the balance of amino acids in soybeans better for nourishing humans (something that was really not necessary). Soybean nutrition is compromised slightly by a relative deficiency in its methionine content; the Brazil nut gene for producing a methionine rich protein was introduced into the soybean genes. The problem is that some people are allergic to the protein produced by the gene from the Brazil nut (they are allergic to these nuts and this protein happens to be one of the allergenic substances). So, that GMO crop idea had to be abandoned (it had not been commercially introduced). Still, the fact

that scientists went down this particular path of potential product development shows that the technology can get into areas of trouble.

Another question that is raised is the extent to which the genetic modification actually provides a benefit. As an example, studies have suggested that some of the pest-resistant GM crops do not actually result in a significantly lower amount of pesticides being used on them. Thus, any potential risks of using the technology might not be balanced by sufficient bene fits. Also, nature can find a way around the genetic modifications. For example, with increasing sowing of the Roundup ready crops, this particular herbicide is being very extensively used; so weeds resistant to that pesticide are turning up.

Perhaps the biggest ethical problem is the one of the "slippery slope." Genetic engineering has definitely provided some benefits and also appears to have many more benefits to offer as the technology progresses. Companies and governments may rush into production one or more products of the new technologies that will turn out to be harmful, either to the environment or to humans directly. Consider, for example, a country where a large part of the population is starving (example: North Korea) and where researchers might find a way to vastly increase the yield of a crop or the nutritional benefits of a food. There would be a lot of pressure to move quickly to put this GM crop into commercial use, and to downplay any objections raised (as well as to consider that any problems that might arise could be resolved later). These genetically modified organisms are not always confined to the country where they are being used (particularly in the case where pollen is spread by the wind). Who knows what kind of ecological disaster might arise from failure to consider the unintended consequences? Similarly, when bacteria are used in batch cultures to produce proteins (as in the case of producing insulin), often the bacteria is one that is commonly found in nature (e.g., *E. coli*). If it escapes into the environment, could it then cause problems? Might these organisms be inadequately safe-guarded in some countries? Objections that do not involve the biology of genetic alterations might still be mentioned here in passing, such as the consideration of economics and society. The leading technologies and the ability to make use of them on a large scale is often dominated by countries with the greatest wealth or companies with patent protections. There is some concern that the utilization of the technology and the economic benefits may not be equitably shared. This is not unique to genetic modification, but because of the diversity of genetic work that can be accomplished with the state-of-the-art technology and production facilities, there is definitely a concentration of power in certain areas of the world.

The genetic engineering of animals has increased significantly in recent years, and the use of this technology brings with it ethical issues, some of which relate to animal welfare — defined by the World Organisation for Animal Health as "the state of the animal…how an animal is coping with the conditions in which it lives".

Several terms are used to describe genetically engineered animals: genetically modified, genetically altered, genetically manipulated, transgenic, and biotechnology-derived, amongst others. In the early stages of genetic engineering, the primary technology used was transgenesis, literally meaning the transfer of genetic material from one organism to another. However, with advances in the field, new technology emerged that did not necessarily require transgenesis: recent applications allow for the creation of genetically engineered animals via the deletion of genes, or the manipulation of genes already present. To reflect this progress and to include those animals that are not strictly transgenic, the umbrella term "genetically engineered" has been adopted into the guidelines developed by the Canadian Council on Animal Care (CCAC). For clarity, in the new *CCAC guidelines on: genetically-engineered animals used in science* (currently in preparation) the CCAC offers the following definition of a genetically engineered animal: "an animal that has had a change in its nuclear or mitochondrial DNA (addition, deletion, or substitution of some part of the animal's genetic material or insertion of foreign DNA) achieved through a deliberate human technological intervention." Those animals that have undergone induced mutations (for example, by chemicals or radiation — as distinct from spontaneous mutations that naturally occur in populations) and cloned animals are also considered to be genetically engineered due to the direct intervention and planning involved in creation of these animals.

Cloning is the replication of certain cell types from a "parent" cell, or the replication of a certain part of the cell or DNA to propagate a particular desirable genetic trait. There are 3 types of cloning: DNA cloning, therapeutic cloning, and reproductive cloning . For the purposes of this paper, the term "cloning" is used to refer to reproductive cloning, as this is the most likely to lead to animal welfare issues. Reproductive cloning is used if the intention is to generate an animal that has the same nuclear DNA as another currently, or previously existing animal. The process used to generate this type of cloned animal is called somatic cell nuclear transfer (SCNT).

During the development of the *CCAC guidelines on: genetically- engineered animals used in science,* some key ethical issues, including animal welfare concerns, were identified: 1) invasiveness of procedures; 2) large numbers of animals required; 3) unanticipated welfare concerns; and 4) how to establish ethical limits to genetic engineering (see Ethical issues of

genetic engineering).Genetic engineering technology has numerous applications involving companion, wild, and farm animals, and animal models used in scientific research. The majority of genetically engineered animals are still in the research phase, rather than actually in use for their intended applications, or commercially available.

Companion animals

By inserting genes from sea anemone and jellyfish, zebrafish have been genetically engineered to express fluorescent proteins — hence the commonly termed "GloFish." GloFish began to be marketed in the United States in 2003 as ornamental pet fish; however, their sale sparked controversial ethical debates in California — the only US state to prohibit the sale of GloFish as pets (5). In addition to the insertion of foreign genes, gene knock-out techniques are also being used to create designer companion animals. For example, in the creation of hypoallergenic cats some companies use genetic engineering techniques to remove the gene that codes for the major cat allergen Companion species have also been derived by cloning. The first cloned cat, "CC," was created in 2002. At the time, the ability to clone mammals was a coveted prize, and after just a few years scientists created the first cloned dog, "Snuppy".

With the exception of a couple of isolated cases, the genetically engineered pet industry is yet to move forward. However, it remains feasible that genetically engineered pets could become part of day-to-day life for practicing veterinarians, and there is evidence that clients have started to enquire about genetic engineering services, in particular the cloning of deceased pets .

Wild animals

The primary application of genetic engineering to wild species involves cloning. This technology could be applied to either extinct or endangered species; for example, there have been plans to clone the extinct thylacine and the woolly mammoth (5). Holt et al (8) point out that, "As many conservationists are still suspicious of reproductive technologies, it is unlikely that cloning techniques would be easily accepted. Individuals involved in field conservation often harbour suspicions that hi-tech approaches, backed by high profile publicity would divert funding away from their own efforts." However, cloning may prove to be an important tool to be used alongside other forms of assisted reproduction to help retain genetic diversity in small populations of endangered species.

Farm animals

Productivity of farm animal species can be increased using genetic engineering. Examples include transgenic pigs and sheep that have been genetically altered to express higher levels of growth hormone. Genetically engineered farm animals can be created to enhance food quality . For example, pigs have been genetically engineered to express the Δ12 fatty acid desaturase

gene (from spinach) for higher levels of omega-3, and goats have been genetically engineered to express human lysozyme in their milk. Such advances may add to the nutritional value of animal-based products.

Farm species may be genetically engineered to create disease-resistant animals. Specific examples include conferring immunity to offspring via antibody expression in the milk of the mother; disruption of the virus entry mechanism (which is applicable to diseases such as pseudorabies); resistance to prion diseases; parasite control (especially in sheep); and mastitis resistance (particularly in cattle).

Genetic engineering has also been applied with the aim of reducing agricultural pollution. The best-known example is the EnviropigTM; a pig that is genetically engineered to produce an enzyme that breaks down dietary phosphorus (phytase), thus limiting the amount of phosphorus released in its manure. Despite resistance to the commercialization of genetically engineered animals for food production, primarily due to lack of support from the public, a recent debate over genetically engineered AquAdvantageTM Atlantic salmon may result in these animals being introduced into commercial production. Effort has also been made to generate genetically engineered farm species such as cows, goats, and sheep that express medically important proteins in their milk. According to Dyck et al (12), "transgenic animal bioreactors represent a powerful tool to address the growing need for therapeutic recombinant proteins." In 2006, ATryn® became the first therapeutic protein produced by genetically engineered animals to be approved by the Food and Drug Administration (FDA) of the United States. This product is used as a prophylactic treatment for patients that have hereditary antithrombin deficiency and are undergoing surgical procedures.

Research animals

Biomedical applications of genetically engineered animals are numerous, and include understanding of gene function, modeling of human disease to either understand disease mechanisms or to aid drug development, and xenotransplantation.

Through the addition, removal, or alteration of genes, scientists can pinpoint what a gene does by observing the biological systems that are affected. While some genetic alterations have no obvious effect, others may produce different phenotypes that can be used by researchers to understand the function of the affected genes. Genetic engineering has enabled the creation of human disease models that were previously unavailable. Animal models of human disease are valuable resources for understanding how and why a particular disease develops, and what can be done to halt or reverse the process. As a result, efforts have focused on developing new genetically engineered animal models of conditions such as Alzheimer's disease, amyotrophic

lateral sclerosis (ALS), Parkinson's disease, and cancer. However, as Wells points out: "these [genetically engineered animal] models do not always accurately reflect the human condition, and care must be taken to understand the limitation of such models."

The use of genetically engineered animals has also become routine within the pharmaceutical industry, for drug discovery, drug development, and risk assessment. As discussed by Rudmann and Durham : "Transgenic and knock out mouse models are extremely useful in drug discovery, especially when defining potential therapeutic targets for modifying immune and inflammatory responses…Specific areas for which [genetically engineered animal models] may be useful are in screening for drug induced immunotoxicity, genotoxicity, and carcinogenicity, and in understanding toxicity related drug metabolizing enzyme systems."

Perhaps the most controversial use of genetically engineered animals in science is to develop the basic research on xenotrans-plantation — that is, the transplant of cells, tissues, or whole organs from animal donors into human recipients. In relation to organ transplants, scientists have developed a genetically engineered pig with the aim of reducing rejection of pig organs by human recipients. This particular application of genetic engineering is currently at the basic research stage, but it shows great promise in alleviating the long waiting lists for organ transplants, as the number of people needing transplants currently far outweighs the number of donated organs. However, as a direct result of public consultation, a moratorium is currently in place preventing pig organ transplantation from entering a clinical trial phase until the public is assured that the potential disease transfer from pigs to humans can be satisfactorily managed.

Ethical issues of genetic engineering

Ethical issues, including concerns for animal welfare, can arise at all stages in the generation and

life span of an individual genetically engineered animal. The following sections detail some of the issues that have arisen during the peer-driven guidelines development process and associated

impact analysis consultations carried out by the CCAC. The CCAC works to an accepted ethic of

animal use in science, which includes the principles of the Three Rs (Reduction of animal numbers, Refinement of practices and husbandry to minimize pain and distress, and Replacement

of animals with non-animal alternatives wherever possible). Together the Three Rs aim to minimize any pain and distress experienced by the animals used, and as such, they are

considered the principles of humane experimental technique. However, despite the steps taken to
minimize pain and distress, there is evidence of public concerns that go beyond the Three Rs and
animal welfare regarding the creation and use of genetically engineered animals.

Concerns for animal welfare

Invasiveness of procedures

The generation of a new genetically engineered line of animals often involves the sacrifice of some animals and surgical procedures (for example, vasectomy, surgical embryo transfer) on others. These procedures are not unique to genetically engineered animals, but they are typically required for their production.

During the creation of new genetically engineered animals (particularly mammalian species) oocyte and blastocyst donor females may be induced to superovulate via intraperitoneal or subcutaneous injection of hormones; genetically engineered embryos may be surgically implanted to female recipients; males may be surgically vasectomized under general anesthesia and then used to induce pseudopregnancy in female embryo recipients; and all offspring need to be genotyped, which is typically performed by taking tissue samples, sometimes using tail biopsies or ear notching. However, progress is being made to refine the genetic engineering techniques that are applied to mammals (mice in particular) so that less invasive methods are feasible. For example, typical genetic engineering procedures require surgery on the recipient female so that genetically engineered embryos can be implanted and can grow to full term; however, a technique called non-surgical embryo transfer (NSET) acts in a similar way to artificial insemination, and removes the need for invasive surgery. Other refinements include a method referred to as “deathless transgenesis,” which involves the introduction of DNA into the sperm cells of live males and removes the need to euthanize females in order to obtain germ line transmission of a genetic alteration; and the use of polymerase chain reaction (PCR) for genotyping, which requires less tissue than Southern Blot Analysis .

Large numbers of animals required

Many of the embryos that undergo genetic engineering procedures do not survive, and of those that do survive only a small proportion (between 1% to 30%) carry the genetic alteration of interest. This means that large numbers of animals are produced to obtain genetically engineered animals that are of scientific value, and this contradicts efforts to minimize animal use.

Other ethical issues

Genetic engineering also brings with it concerns over intellectual property, and patenting of created animals and/or the techniques used to create them. Preserving intellectual property can breed a culture of confidentiality within the scientific community, which in turn limits data and animal sharing. Such limits to data and animal sharing may create situations in which there is unnecessary duplication of genetically engineered animal lines, thereby challenging the principle of Reduction. Indeed, this was a concern that was identified in a recent workshop on the creation and use of genetically engineered animals in science .

It should be noted that no matter what the application of genetically engineered animals, there are restrictions on the methods of their disposal once they have been euthanized. The reason for this is to restrict the entry of genetically engineered animal carcasses into the natural ecosystem until the long-term effects and risks are better understood.

There are a number of ethical concerns over genetically modified (GM) foods and these have all affected public support of the products. The issues have also triggered controversy and regulations around GM foods and any company that produces these crops or products. Concerns range from the environment to risks to our food web or issues concerning disease, allergies and contamination.

Allergies And Disease

A key ethical concern about GM foods is their potential to trigger allergies or disease in humans. Given that a gene could be extracted from an allergenic organism and placed into another one that typically does not cause allergies, a person may unknowingly be exposed to an allergen. In turn, this could lead to an allergic reaction. There is also the fear that new allergies could occur from the mixing of genes from two organisms. Disease is a major health worry with regards to GM foods. Given that some of the crops modified are done so with DNA from a bacterium or virus, there is concern that a new disease may occur in humans who consume the GM food. With some GM crops having antibiotic resistant marker genes, there is also the worry that these genes could be passed on to microbes that cause disease and health problems in humans. With widespread antibiotic resistance currently already occurring, any new resistance could prove disastrous.

Damage To The Environment

Damage to the environment is another ethical fear with regards to GM crops. Unfortunately, the technology is still new enough that there is much we do not know about the effect of GM

crop production on the environment. Long-term studies take decades to complete and most studies of GM crop production involve short-term effects of the technology.

Another ethical issue around GM crops is our ability to contain them in a specific area. There are fears that if these crops do negatively impact the environment, they will spread in an out-of control fashion and we will not be able to stop their damaging effects. For instance, one type of sugar beet that had been engineered to be resistant to a specific herbicide ended up unintentionally having the genes to resist a different herbicide. When farmers went to eliminate the crop, they still found that a small percentage had survived.

Cross-Pollination

Cross-pollination is a challenge for any crop growth but it can typically be managed if c are is taken to use good growing practices. There is the possibility of genes from GM foods spreading to other plants and crops, which could create overzealous weeds that can't be contained at all.

Food Web And Risks

Risks to the food web are a very real ethical concern around GM technology. Any pesticide or herbicide from the crop could harm animals and other organisms in the environment. For example, GM sugar beets that were produced to be resistant to herbicides did successfully reduce weeds. However, Skylark birds that consume the seeds from this particular weed would now be required to find a new food source, thereby endangering their existence. An animal could also consume the GM crop itself, which means that if the crop has been engineered to produce a pesticide, the animal may become ill and die. In one North American study, caterpillars of the monarch butterfly were killed when they fed on pollen from GM corn crops.

Addressing Ethical Concerns For GM Foods

Unfortunately, the controversy and fears around GM foods and any company that produces these products still continue to persevere, although this could be viewed as a positive movement because it will challenge GM technology and help to make it safer and more regulated. In one public opinion poll, it was found that the more people read about GM foods, the more concerned they became about the technology.