MODULE 4- LECTURE 1

ISOLATION AND PURIFICATION OF NUCLEIC ACIDS (GENOMIC/PLASMID DNA AND RNA)

4-1.1. Introduction

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

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

4-1.2. Isolation and Purification of Genomic DNA

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

The method of isolation of genomic DNA from a bacterium comprises following steps (Figure 4-1.2.)-

1. Bacterial culture growth and harvest.
2. Cell wall rupture and cell extract preparation.

3. DNA Purification from the cell extract.
4. Concentration of DNA solution.

4-1.2.1. Growth and harvest of bacterial culture

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

4-1.2.2. Preparation of cell extract

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

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

Lysozyme

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

EDTA (Ethylene diamine tetra-acetic acid)

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

SDS (Sodium dodecyl sulphate)

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

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

4-1.2.3. Purification of DNA

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

4-1.2.3.1. Organic extraction and enzymatic digestion for the removal of contaminants

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

4-1.2.3.2. Using ion-exchange chromatography

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

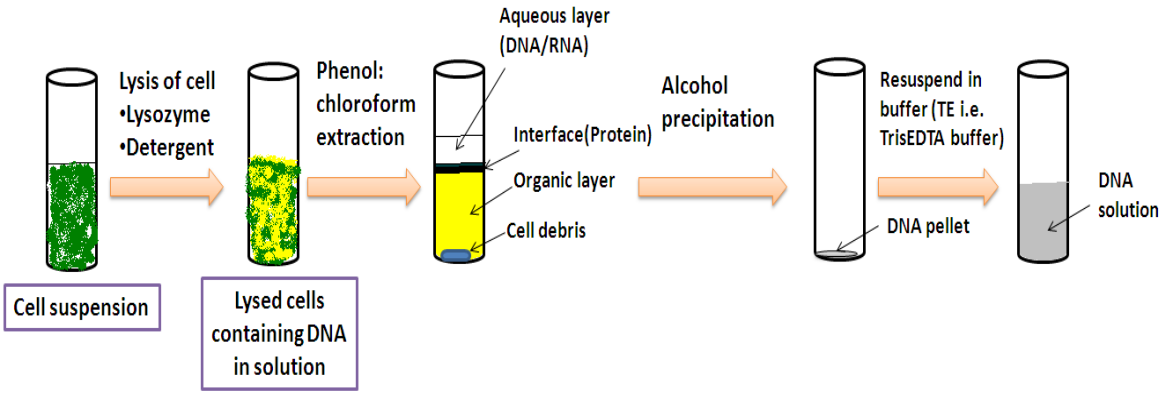


Figure 4-1.2. Preparation of genomic DNA

4-1.2.4. Concentration of DNA samples

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

Advantage

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

4-1.3. Isolation and Purification of Plasmid DNA

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

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

4-1.3.1. Growth of the bacterial cell

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

4-1.3.2. Harvest and lysis of bacteria

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

4-1.3.3. Purification of Plasmid DNA

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

4-1.3.3.1. Methods for separation of plasmid DNA

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

4-1.3.3.1.1. Separation based on size difference

- It involves lysis of cells with lysozyme and EDTA (function as described above in point **4-1.2.2.**) in the presence of sucrose (prevents the immediate bursting of cell).
- Cells with partially degraded cell walls are formed that retain an intact cytoplasmic membrane called as sphaeroplasts.
- Cell lysis is then induced by the addition of a non-ionic detergent (e.g. Triton X-100) or ionic detergents (e.g. SDS) causing chromosomal breakage.
- Bacterial chromosome attached to cell membrane, upon lysis gets removed with the cell debris.
- A cleared lysate consisting almost entirely of plasmid DNA is formed with very little breakage of the bacterial DNA. (Figure 4-1.3.3.1.1.).

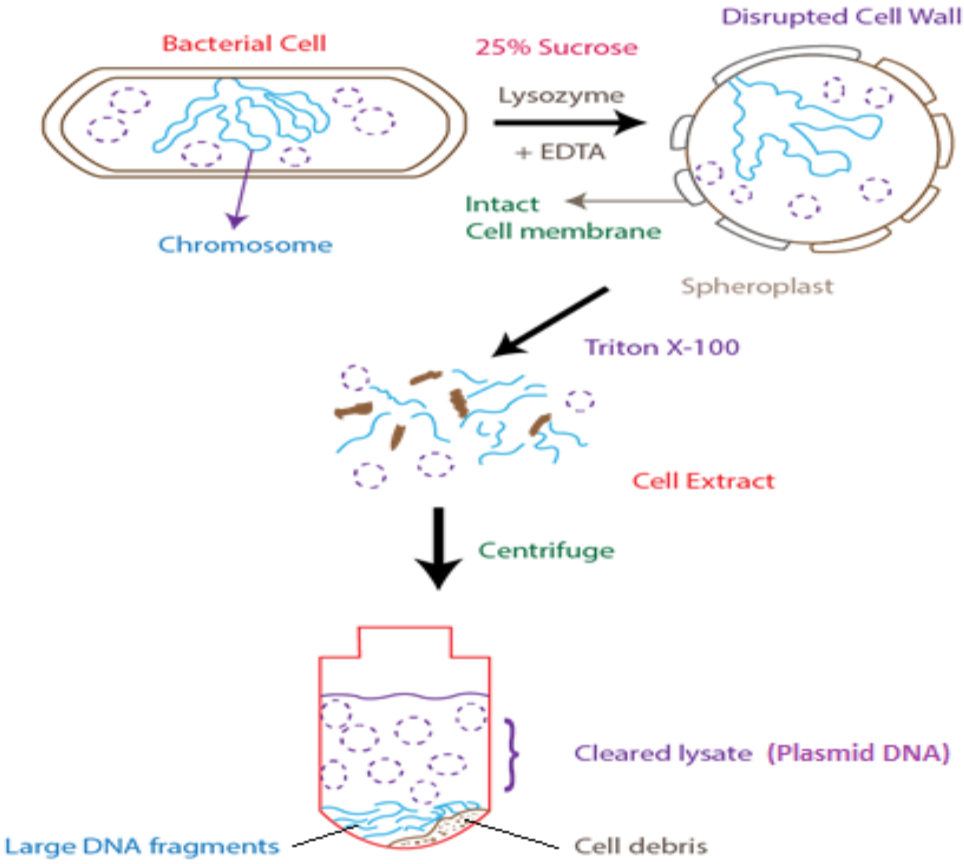


Figure 4-1.3.3.1.1. Separation of plasmid DNA on the basis of size.

4-1.3.3.1.2. Separation based on conformation

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

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

4-1.3.3.1.2(a). Alkaline denaturation method

- This method is based on maintaining a very narrow pH range for the denaturation of non-supercoiled DNA but not the supercoiled plasmid (Figure 4-1.3.3.1.2(a).).
- Addition of sodium hydroxide to cell extract or cleared lysate (pH12.0-12.5) results in disruption of the hydrogen bonds of non-supercoiled DNA molecules.
- As a result, the double helix unwinds and two polynucleotide chains separate.
- Further addition of acid causes the aggregation of these denatured bacterial DNA strands into a tangled mass which can be pelleted by centrifugation, leaving plasmid DNA in the supernatant.

Advantage

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

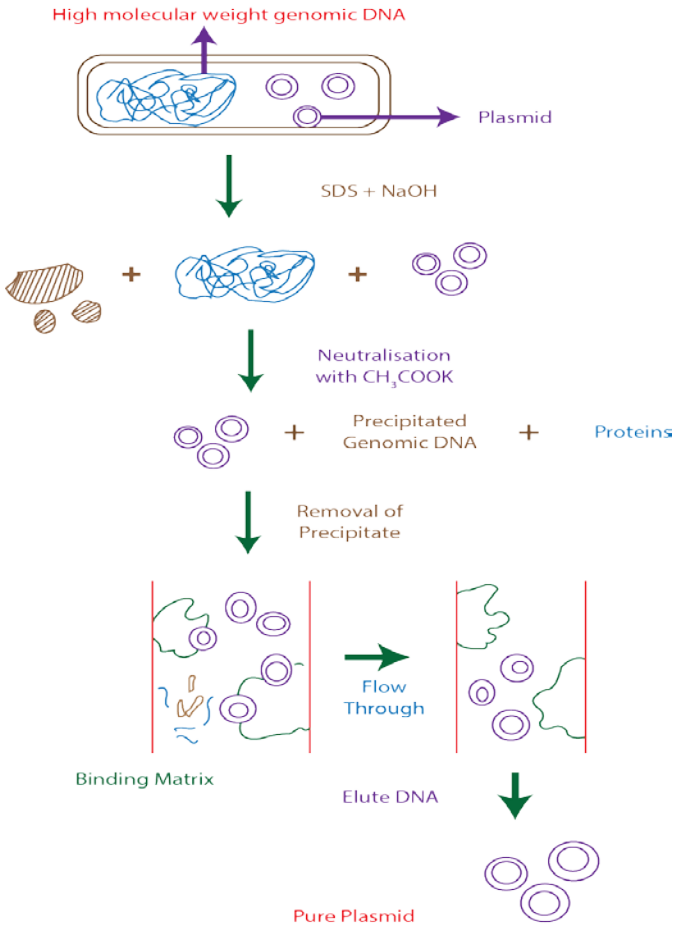


Figure 4-1.3.3.1.2(a). Separation of plasmid DNA by Alkaline denaturation method

4-1.3.3.1.2(b). Ethidium bromide-cesium chloride density gradient centrifugation

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

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

4-1.4. Isolation and Purification of RNA

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

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

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

4-1.4.1. Organic extraction method

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

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

4-1.4.2. Direct lysis methods

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

Advantages

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

Drawbacks

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

Bibliography

Brown TA. 2006. *Gene cloning and DNA analysis: an introduction.* 5th ed. Blackwell Scientific.

http://amrita.vlab.co.in/?sub=3&brch=186&sim=718&cnt=1

http://www.invitrogen.com/site/us/en/home/References/Ambion-Tech-Support/rnaisolation/general-articles/the-basics-rna-isolation.html

Reece RJ. 2003. *Analysis of Genes and Genomes.* John Wiley & Sons, U.K.

MODULE 4- LECTURE 2

QUANTIFICATION AND STORAGE OF NUCLEIC ACIDS

4-2.1. Quantification of nucleic acids

Quantification of nucleic acids is done to determine the average concentrations of DNA or RNA present in a mixture, as well as their purity. The accurate measurement is based on sensitivity, specificity and interference by contaminants. Various methods that can be employed to quantify the nucleic acid concentration are listed below,

(i) Spectrophotometric analysis
(ii) Nanodrop
(iii) Fluorescence based method
(iv) Fluorescence in situ hybridization (FISH)

4-2.1.1. Spectrophotometric analysis

It is a simple and accurate method to assess the concentration and purity of nucleic acids based on their absorption at different wavelengths (described in detail in Module 3-Lecture 1). According to Beer Lambert's Law, when light is passed through a substance of concentration c at a pathlength l (typically 1 cm), the absorbance is directly proportional to the concentration of the substance and the pathlength i.e.

$$\mathbf{A= \in \boldsymbol{cl}}$$

Where, A= absorbance of nucleic acids at a particular wavelength

$\in$=Molar extinction coefficient (M^{-1} cm^{-1}) or specific absorption coefficient (µg/ml) $^{-1}$ cm^{-} [1]that measures how strongly a substance absorbs light at a particular wavelength.

l = Pathlength of the spectrophotometer cuvette

c= concentration of a substance

The value of $\in$ for ss-DNA is 0.027; ds-DNA is 0.020 and ss-RNA is 0.025.

The concentration of nucleic acids can be calculated by the given equations-

DNA concentration (µg/ml) = OD_{260} x 100 (dilution factor) x 50 µg/ml

100

For ssDNA, c (pmol/µl) = A_{260}/10 x S

c(µg/ml)= A_{260}/0.027

For dsDNA, c (pmol/µl) = A_{260}/13.2 x S

c (µg/ml)=A_{260}/0.020

For ssRNA, c (µg/ml)=A_{260}/0.025

For **Oligonucleotide,** absorbance is significantly affected by composition of oligonucleotide bases.c (pmol/µl) = A260 x 100/1.5 N_A+ 0.71 N_C + 1.20 N_G + 0.84 N_T

where S denotes the size of DNA in Kb and N represents the number of bases A, G, C or T.

The value of ∈ for the DNA bases is listed in the **Table 4-2.1.1.**

Bases	**∈ (at 260 nm) (**M^{-1} $cm^{-1)}$
Adenine (A)	15,200
Cytosine (C)	7,050
Guanine (G)	12,010
Thymine (T)	8,400

The purity of the nucleic acids can be determined by A260/A280 ratio described in detail in Module 3-Lecture 1.

4-2.1.2. Nanodrop method

Unlike spectrophotometric method requiring 1-2 ml of the sample, this technique involves microvolume (1-2 µl) quantification of nucleic acid sample. The principle of NanoDrop method is described in detail in Module 3- Lecture 1. The nucleic acids having the concentration range from 2-15000 ng/µl can be assessed by this method. For this, 1-2µl of sample is loaded between the two optical surfaces and the software automatically calculates the concentration and purity of the nucleic acid and displays sample quality (purity) as a spectral output. Generally the pathlength is taken as 1 cm even though in

case of Hellma tray cell cuvette pathlength ranges from 0.2-1.0 mm during measurement cycle. Sensitivity measured by sample retention system of NanoDrop is increased by many folds compared to a standard spectrophotometer. Automatic pathlength adjustment facilitates the direct quantification of concentrated samples avoiding the need for sample dilutions.

4-2.1.3. Quantification using fluorescent dyes

This method is simple and more sensitive than spectrophotometric method which measures the fluorescence intensity of the dyes that fluoresce upon interaction with the nucleic acids. Various fluorescent dyes such as EtBr, Hoechst 33258, picogreen, DAPI can be used for quantification of nucleic acids described in detail in Module 3-Lecture 1. The sensitivity range of these dyes for the quantification is listed in Table 4-2.1.3. Quantification of nucleic acids separated by gel electrophoresis can be done by comparing the stained nucleic acids with stained standards of known concentration separated on the same gel. The dye: DNA complex shows greater fluorescence than the unbound dye by intercalating between the stacked base pairs. The fluorescence intensity of the band estimates the concentration/amount of DNA. The intensity of the stain is dependent at least in part on the base pair composition of the nucleic acid.

Nucleic acid	**UV Absorbance (A_{260})**	**Fluorescence methods**		
		Hoechst 33258	**EtBr**	**Picogreen**
DNA	1–50 µg/mL	0.01–15 µg/mL	0.1–10 µg/mL	0.025–1000 ng/mL
RNA	1–40 µg/mL	Not applicable	1–40 µg/mL	Minimal sensitivity
Ratio of signal (DNA/RNA)	0.8	400	2.2	>100

Table 4-2.1.3. Sensitivity range of quantification of nucleic acids by UV spectrophotometry and fluorescence based methods.

4-2.1.4. Fluorescence in situ hybridization (FISH)

In a cytometric system, quantification of fluorescence ISH signals and accurate estimation of fluorescence intensity of specific DNA sequences can be performed using an epi-fluorescence microscope with a multi-wavelength illuminator, fitted with a cooled charge couple device (CCD) camera. The principle of this method is described in detail in Module 3- Lecture 1. Quantitation by FISH was first applied as a basis for rudimentary cytogenetic assays. Detection using CCD camera enables the quantitative analysis of mRNA as well. The two factors affecting the fluorescence assay are reproducibility and signal irregularity as well as background noise which may vary from sample to sample and cell to cell. Various approaches have been developed to reduce background noise such as use of reducing agents e.g. sodium borohydride and pre-treatment by light irradiation, and image analysis methods. The former cannot be possible and effective always so the later are now employed for analysis. The output thus consist of a true signal and various noise components, the profile for each of them can be estimated and deleted by digital methods like independent component analysis. Several computational methods have been developed to overcome the differential intensity and color overlap caused by multi-color image analysis.

Further, FISH can be employed for interpretation using various approaches such as multi-color cytometry algotithms, dot-counting approaches, use of diagnostic probe sets etc.

4-2.2. Storage of Nucleic Acids

The purified DNA can be stored at -20°C or -70°C under slightly basic conditions (e.g., Tris- Cl, pH 8.0) as acidic conditions result in hydrolysis of DNA.

Diluted solutions of nucleic acids can be stored in aliquots and thawed once only. RNA preservation under frozen conditions is helpful.

Purified RNA can be stored at -20°C or -80°C in RNase-free solution such as-

- The RNA Storage Solution (1 mM sodium citrate, pH 6.4 ± 0.2): It is a buffer that delivers greater RNA stability than 0.1 mM EDTA or TE. The presence of sodium

citrate and low pH minimizes base hydrolysis of RNA. Sodium citrate acts both as a chelating and buffering agent.

- 0.1 mM EDTA
- TE Buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.0)

This RNase-free solution is compatible with all RNA applications including *in vitro* translation, reverse transcription, nuclease protection assays and northern analysis.

Bibliography

Armbrecht M, Gloe J, Goemann W. 2013. Determination of nucleic acid concentrationsusing fluorescent dyes in the Eppendorf BioSpectrometer® fluorescence. Application Note No. 271.

Arvey A, Hermann A, Hsia CC, Le E, Freund Y, McGinnis W. 2010. Minimizing off-target signals in RNA fluorescent *in situ* hybridization. *Nucleic Acids Research,* 38(10): e115.

Brown TA.2006. *Gene cloning and DNA analysis: an introduction.* 5th ed. Blackwell Scientific.

Desjardins P, Conklin D. 2010. NanoDrop Microvolume Quantitation of Nucleic Acids. *J. Vis Exp. 45: 2585.*

Gallagher SR, Wiley EA. 2008. *Current protocols essential laboratory techniques.* John Wiley & Sons, Inc.

http://amrita.vlab.co.in/?sub=3&brch=186&sim=718&cnt=1

http://www.invitrogen.com/site/us/en/home/References/Ambion-Tech-Support/rna-isolation/general-articles/the-basics-rna-isolation.html

Levsky JM, Singer RH. 2003. Fluorescence in situ hybridization: past, present and future. *Journal of Cell Science,* 116: 2833-2838.

Lu J, Tsourkas A. 2011. Quantification of miRNA abundance in single cells using locked nucleic acid-FISH and enzyme-labeled fluorescence. *Methods Mol Biol*, 680: 77-88.

Nederlof PM, van der Flier S, Verwoerd NP, Vrolijk J, Raap AK, Tanke HJ. 1992. Quantification of fluorescence in situ hybridization signals by image cytometry. Cytometry, 13(8): 846-52.

Reece RJ. 2003. Analysis of Genes and Genomes. *John Wiley & Sons* U.K.

MODULE 4- LECTURE 3

CONSTRUCTION OF cDNA LIBRARY

4-3.1. Introduction

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.

4-3.2. Construction of a cDNA Library

The construction of cDNA library involves following steps-

1. Isolation of mRNA
2. Synthesis of first and second strand of cDNA
3. Incorporation of cDNA into a vector
4. Cloning of cDNAs

4-3.2.1. 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.

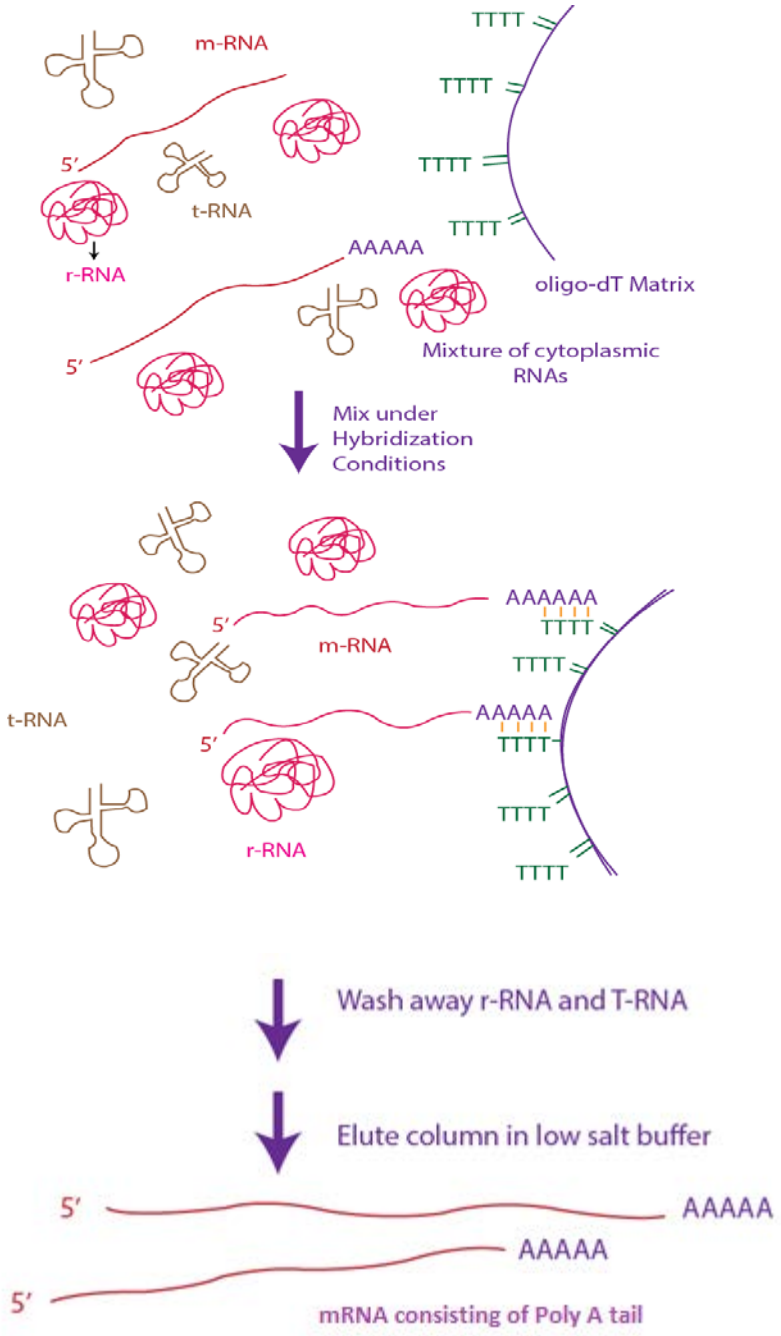


Figure 4-3.2.1. Isolation of mRNA using oligo-dT column chromatography.

4-3.2.2. 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.

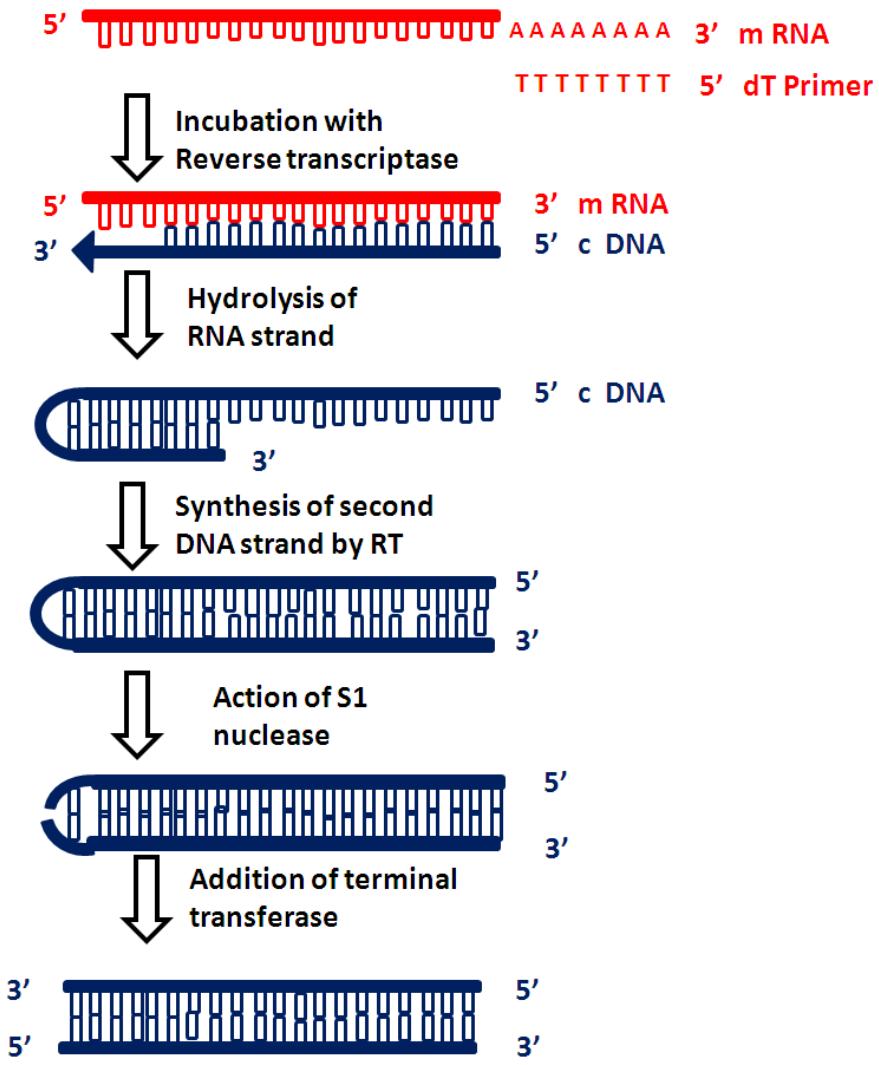


Figure 4-3.2.2. Synthesis of first and second strand of cDNA.

4-3.2.3. 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.

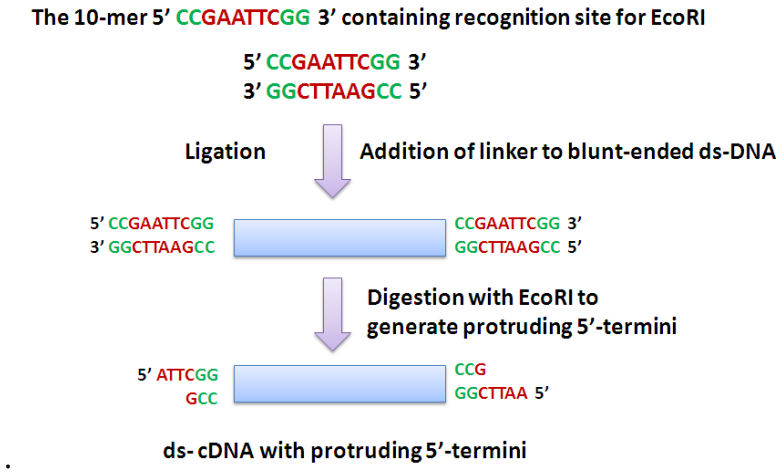


Figure 4-3.2.3.1. Modification of cDNA termini using linkers.

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.

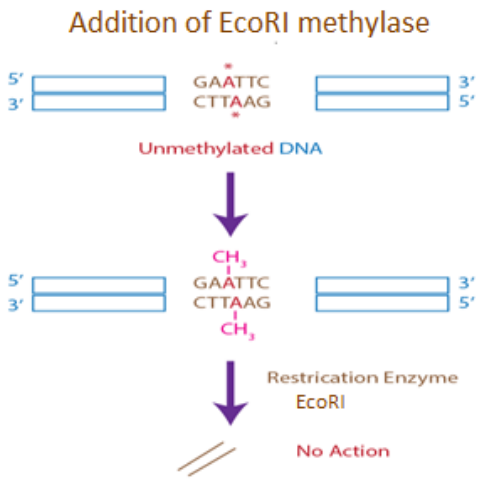


Figure 4-3.2.3.2. Action of methylases

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.

4-3.2.4. 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.

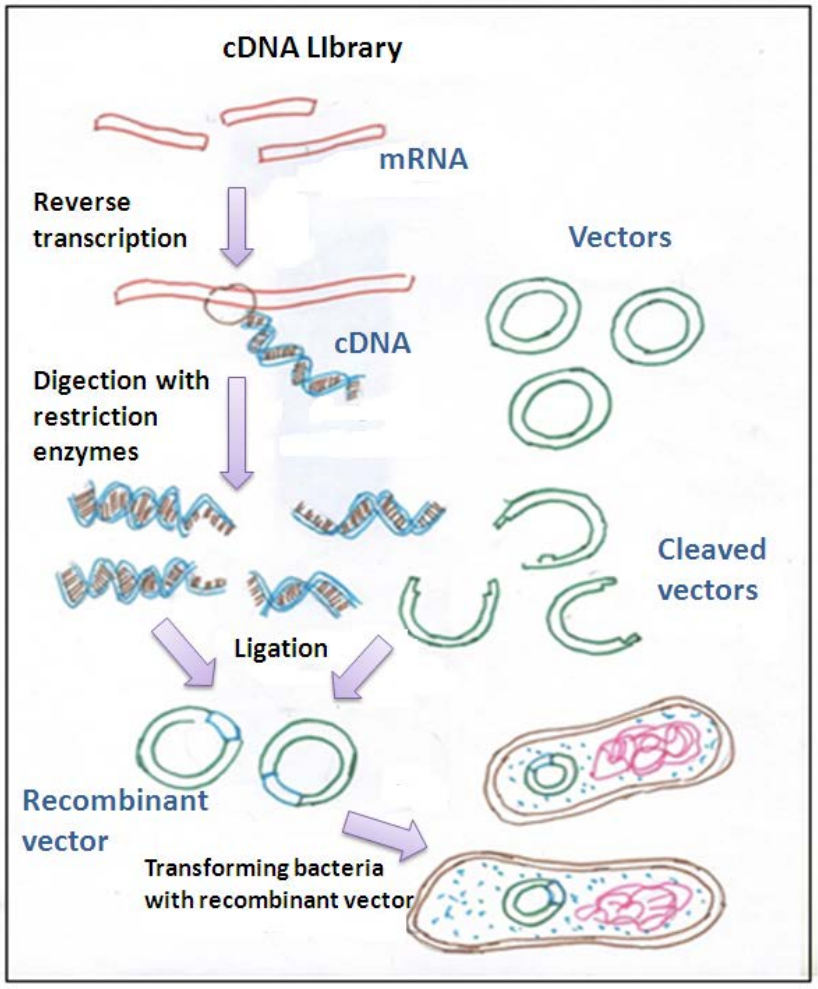


Figure 4-3.2.4. Cloning of cDNAs

4-3.3. 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.

4-3.4. 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.

4-3.5. 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.

4-3.6. 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.

4-3.7. 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.

Bibliography

Griffiths AJF, Miller JH, Suzuki DT *et al.* 2000. *An Introduction to Genetic Analysis.* 7th ed. W. H. Freeman. New York.

Lodge J. 2007. *Gene cloning: principles and applications.* Taylor and Francis Group.

Lodish H, Berk A, Zipursky SL *et al.* 2007. *Molecular Cell Biology.* 6th ed. W. H. Freeman.

Nicholl DST. 2008.*An Introduction to Genetic Engineering.* 3rd ed. Cambridge University Press.

Singh BD. 2008. *Biotechnology: Expanding Horizons.* 2nd ed. Kalyani Publishers.

MODULE 4-LECTURE 4

CONSTRUCTION OF GENOMIC LIBRARY

4-4.1. Introduction

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.

4-4.2. Construction of genomic library

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.

4-4.2.1. 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.

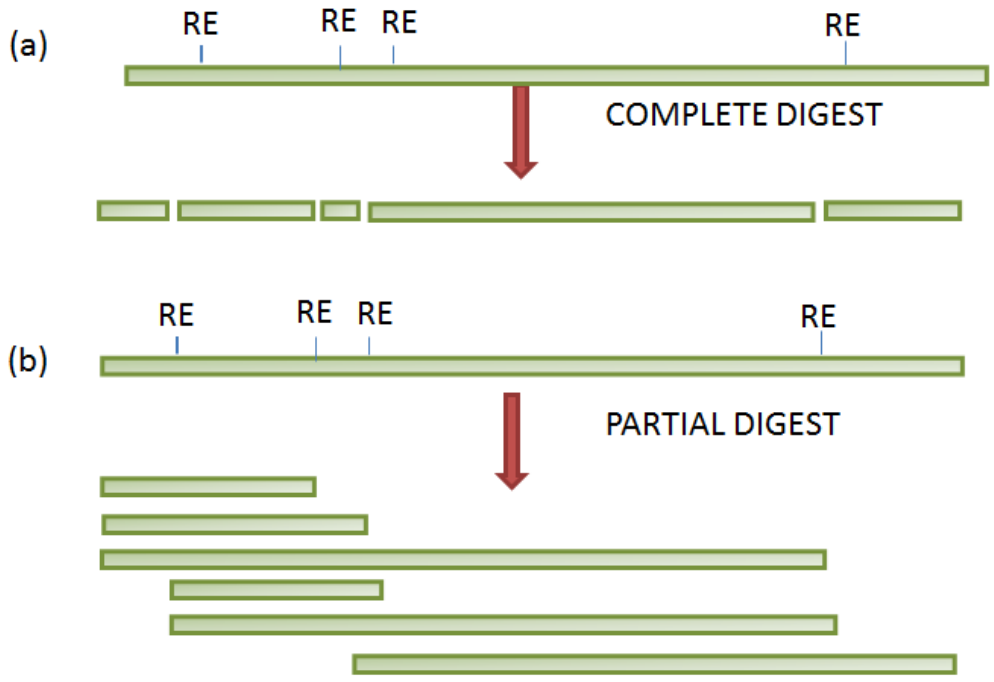


Figure4-4.2.1. The complete (a) and partial (b) digestion of a DNA fragment using restriction enzymes.

(Adapted from Reece RJ. 2000. Analysis of Genes and Genomes. John Wiley & Sons, U.K.)

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.

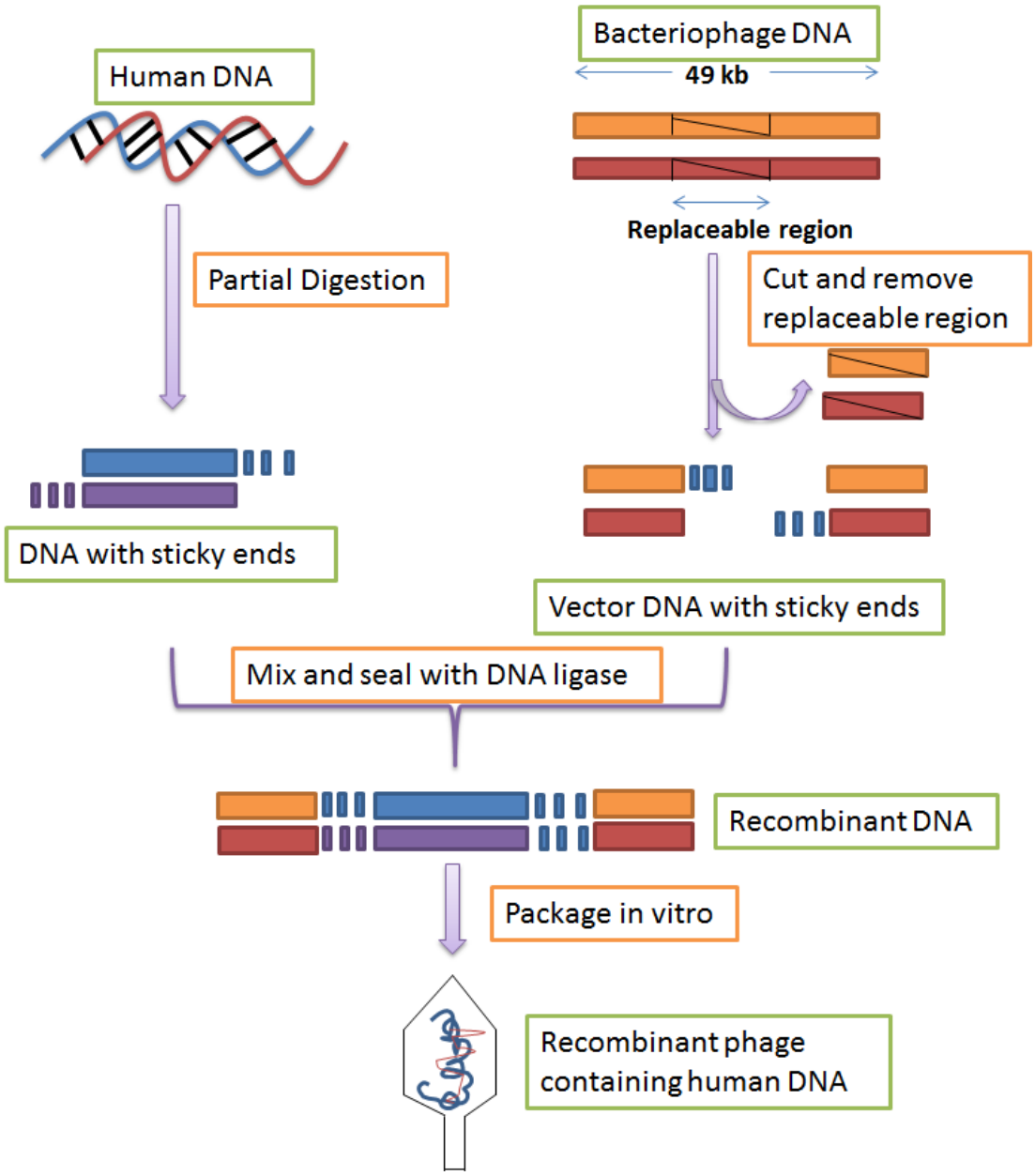


Figure 4-4.2. Construction of genomic library

4-4.2.2. 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.

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)

Table 4-4.2.2. Vectors used for cloning genomic libraries.

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.

4-4.2.3. 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-

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

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

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

4-4.3. Amplified library

- The primary library created is usually of a low titer and unstable. The stability and titer can be increased by amplification. For this, the phages or bacterial colonies are plated out several times and the resulting progenies are collected to form an amplified library.
- The amplified library can then be stored almost indefinitely due to longshelf-life of phages.
- It usually has a much larger volume than the primary library, and consequently may be screened several times.

- It is possible that the amplification process will result in the composition of the amplified library not truly reflecting the primary one.
- Certain DNA sequences may be relatively toxic to *E. coli* cells. As a consequence bacteria harboring such clones will grow more slowly than other bacteria harboring non-toxic DNA sequences. Such problematic DNA sequences present in the primary library may be lost or under-represented after the growth phase required to produce the amplified library.

4-4.4. Subgenomic library

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

4-4.5. 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.

4-4.6. 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.

4-4.7. 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.

4-4.8 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 4-4.8.

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

Table 4-4.8. Features of genomic and cDNA library

Bibliography

Brown TA. 2010. *Gene cloning and DNA analysis: An Introduction.* 6th ed. Wiley-Blackwell.

Lodge J. 2007. *Gene cloning: principles and applications.* Taylor & Francis Group.

Moore DD. 2001. Genomic DNA Libraries. *Current Protocols in Molecular Biology.* DOI: 10.1002/0471142727.mb0501s00

Primrose SB, Twyman RM, Old RW. 2001. *Principles of Gene Manipulation.* Oxford: Blackwell Scientific.

Reece RJ. 2000. *Analysis of Genes and Genomes.* John Wiley & Sons, U.K.

Singh BD. 2008. *Biotechnology: Expanding Horizons.* 2nd ed. Kalyani Publishers.

MODULE 4 LECTURE 5

SCREENING AND PRESERVATION OF DNA LIBRARIES

4-5.1. Introduction

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

- Detecting a particular DNA sequence and
- Gene expression.

4-5.2. Methods for screening based on detecting a DNA sequence

4-5.2.1. Screening by hybridization

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

The commonly used methods of hybridization are,

a) Colony hybridization
b) Plaque hybridization.

4-5.2.1(a). Colony hybridization

Colony hybridization, also known as replica plating, allows the screening of colonies plated at high density using radioactive DNA probes. This method can be used to screen plasmid or cosmid based libraries (Explained in detail in Module 3-Lecture 4 as point **3-4.4)**.

4-5.2.1(b). Plaque hybridization

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

- The nitrocellulose filter is applied to the upper surface of agar plates, making a direct contact between plaques and filter.
- The plaques contain phage particles, as well as a considerable amount of unpackaged recombinant DNA which bind to the filter.
- The DNA is denatured, fixed to the filter, hybridized with radioactive probes and assayed by autoradiography.

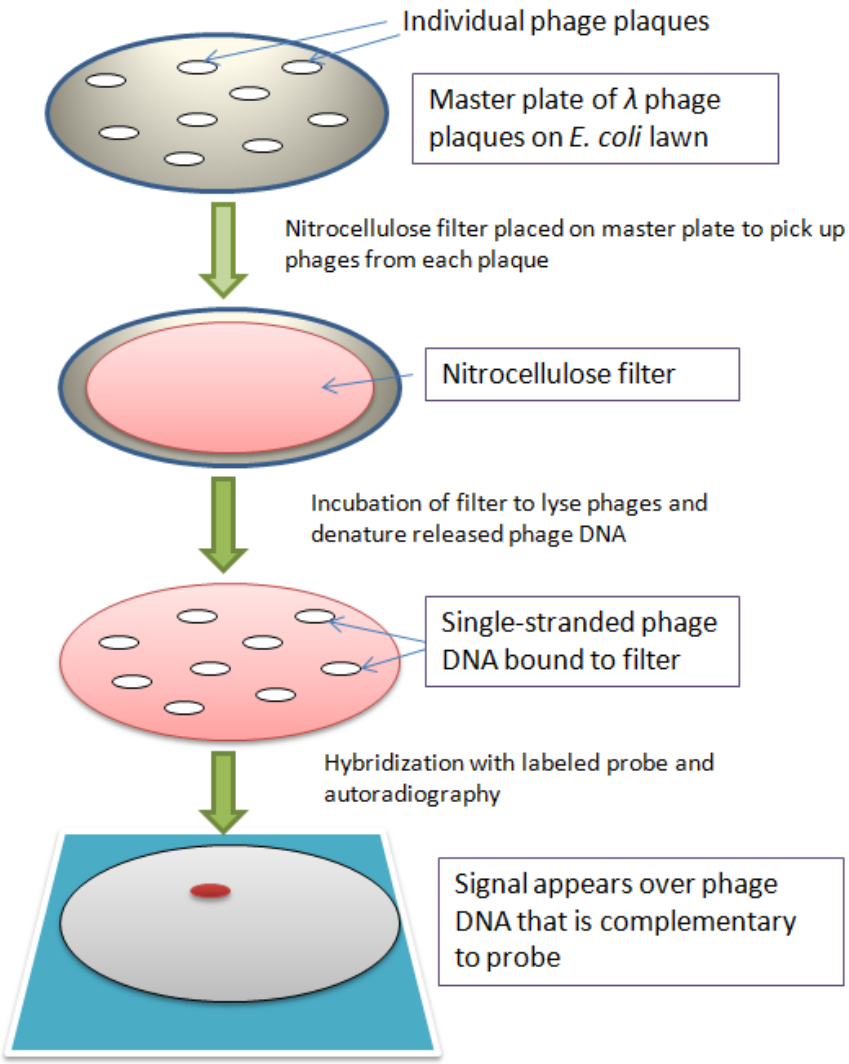
Advantages

- This method results in a 'cleaner' background and distinct signal (less background probe hybridization) for λ plaque screening due to less DNA transfer from the bacterial host to the nitrocellulose membrane while lifting plaques rather than bacterial colonies.

- Multiple screens can be performed from the same plate as plaques can be lifted several times.

- Screening can be performed at very high density by screening small plaques. High-density screening has the advantage that a large number of recombinant

clones can be screened for the presence of sequences homologous to the probe in a single experiment.



4-5.2.1(b). Schematic process for screening libraries by Plaque hybridization.

Probes used for hybridization

Cloned DNA fragments can be used as probes in hybridization reactions if a cDNA clone is available. DNA or synthetic oligonucleotide probes can be used for identification of a clone from a genomic library instead of RNA probes, for example, to study the regulatory sequences which are not part of the cDNA clone. A common method of labeling probes is the incorporation of a radioactive or other marker into the molecule. A number of alternative labeling methods are also available that involve an amplification process to detect the presence of small quantities of bound probe and avoid the use of radioactivity. These methods involve the incorporation of chemical labels such as digoxigenin or biotin into the probe which can be detected with a specific antibody or the ligand streptavidin, respectively.

4-5.2.2. Screening by PCR

PCR screening is employed for the identification of rare DNA sequences in complex mixtures of molecular clones by increasing the abundance of a particular sequence. It is possible to identify any clone by PCR only if there is available information about its sequence to design suitable primers.

Preparation of a library for screening by PCR can be done by following ways-

- The library can be plated as plaques or colonies on agar plates and individually inoculated into the wells of the multi-well plate. However it is a labor intensive process and can lead to bias in favor of larger colonies or plaques.
- The alternative method involves diluting the library. It involves plating out a small part of the original library (the packaging mix for a phage library, transformation for a plasmid library) and calculating the titer of the library. A larger sample is diluted to give a titer of 100 colonies per mL. Dispensing 100 μL into each well theoretically gives 10 clones in each well. These are then pooled and PCR reactions are carried out with gene-specific primers flanking a unique sequence in the target to identify the wells containing the clone of interest. This method is often used for screening commercially available libraries.

4-5.3. Screening methods based on gene expression

4-5.3.1. Immunological screening

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

Earlier immunoscreening methods employed radio-labeled primary antibodies to detect antibody binding to the nitrocellulose sheet (Figure 4-5.3.1(a).). It is now superseded by antibody sandwiches resulting in highly amplified signals. The secondary antibody recognizes the constant region of the primary antibody and is, additionally, conjugated to an easily assayable enzyme (*e.g.* horseradish peroxidase or alkaline phosphatase) which can be assayed using colorimetric change or emission of light using X-ray film (Figure 4-5.3.1(b).).

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

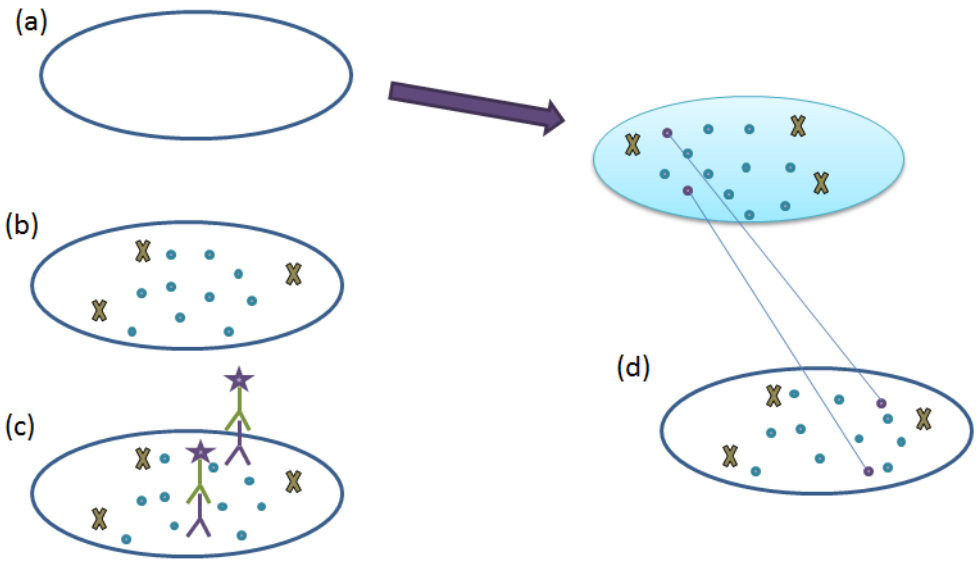


Figure 4-5.3.1(a). Schematic process of immunological screening (a) a nitrocellulose disk is placed onto the surface of an agar plate containing the phage library. Both agar plate and disk are marked so as to realign them later. (b) When the nitrocellulose disk is lifted off again, proteins released from the bacteria by phage lysis bind to the disk. (c) These proteins bind to specific antibody. (d) Plaques formed by bacteriophage that express the protein bound to the antibody will be detected by emission of light. The positive clones can be identified by realignment.

(Adapted from Lodge J. 2007.Gene cloning: principles and applications. Taylor & Francis Group)

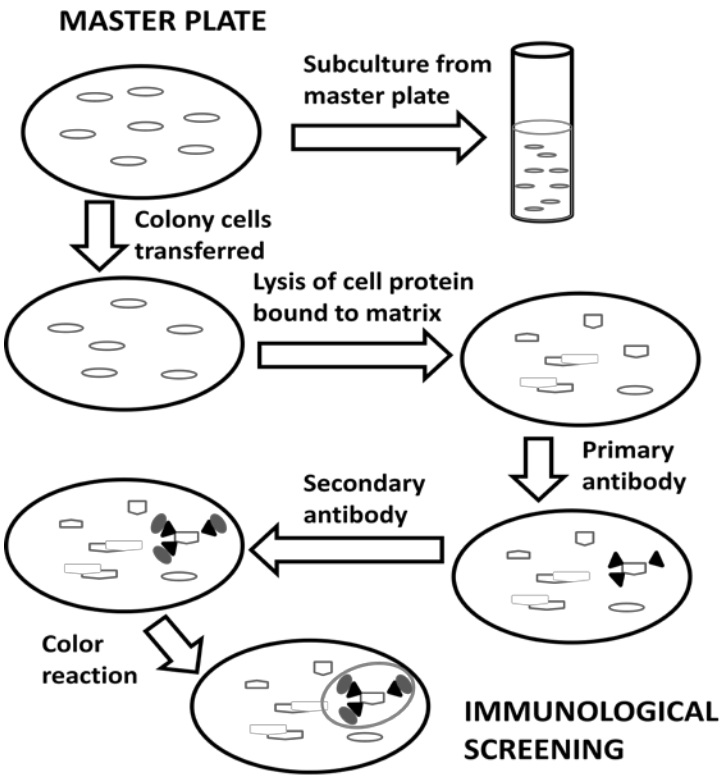


Figure 4-5.3.1(b). Schematic process of immunological screening using antibody sandwich.

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

4-5.3.2. Screening by functional complementation

Functional complementation is the process of compensating a missing function in a mutant cell by a particular DNA sequence for restoring the wild-type phenotype. If the mutant cells are non-viable, the cells carrying the clone of interest can be positively selected and isolated. It is a very powerful method of expression cloning and also useful for identification of genes from an organism having same role as that of defective gene in another organism. The selection and identification of positive clones is based on either the gain of function or a visible change in phenotype.

For example, the functional complementation in transgenic mice for the isolation of *Shaker-2* gene applied by Probst *et al* in1988 shown in Figure 4-5.3.2.

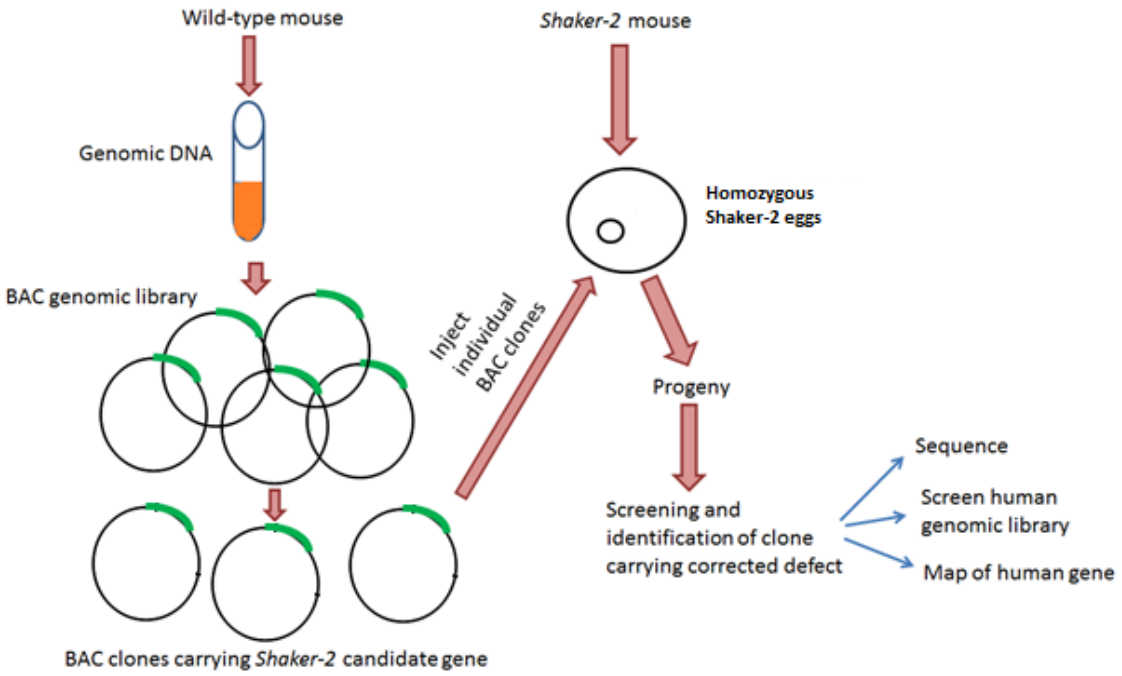


Figure 4-5.3.2. Functional complementation in transgenic mice for isolation of *Shaker-2* gene.

(Adapted from Primrose SB, Twyman RM. 2006. Principles of gene manipulation and genomics.7th ed. Blackwell Publishing.)

The *Shaker-2* mutation is due to the defective gene associated with human deafness disorder. The BAC clone from the wild type mice are prepared and injected into the eggs of Shaker-2 mutants. The resulting mice are then screened for the presence of wild type phenotype. Thus the BAC clone carrying the functional *Shaker-2* gene is identified which encodes a cytoskeletal myosin protein. This method can be used for screening human genomic libraries to identify equivalent human gene.

Drawbacks

- Presence of an assayable mutation within the host cell that can be compensated by the foreign gene expression which in most cases is not available. In addition, foreign genes may not fully compensate the mutations.

Applications

- This method can be used for the isolation of higher-eukaryotic genes (e.g. Drosophila topoisomerase II gene, a number of human RNA polymerase II transcription factors) from an organism.
- It can also be possible in transgenic animals and plants to clone a specific gene from its functional homologue.

Bibliography

Benton WD, Davis RW. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science,* 196: 180–182.

Benton WD, Davis RW. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. *Science,* 196 (4286): 180-182.

Campbell TN, Choy FYM. 2002. Approaches to Library Screening. *J. Mol. Microbiol. Biotechnol.* 4(6): 551-554.

Grunstein M, Hogness DS. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA*, 72: 3961–3965.

Lodge J. 2007. *Gene cloning: principles and applications.* Taylor & Francis Group.

Primrose SB, Twyman RM. 2006. *Principles of gene manipulation and genomics.* 7th ed. Blackwell Publishing.

Primrose SB, Tyman RM, Old RW. 2001. *Principle of Gene Manipulation.* 6th ed. Wiley-Blackwell.

Probst FJ, Fridell RA, Raphael Y et al. 1998. Correction of deafness in *shaker-2* mice by an unconventional myosin in a BAC transgene. *Science*, 280: 1444–1447.

Reece RJ. 2003. *Analysis of Genes and Genomes.* John Wiley & Sons, U.K.

Module 4-Lecture 6

DNA SEQUENCING AND CLONING STATEGIES

4-6.1 DNA SEQUENCING

Sequencing is the method of resolving the order of the nucleotide bases in a DNA molecule (genomic DNA, cDNA, or organellar DNA). It is a primary step in assessing regulatory sequences, coding and non-coding regions. For past few years, DNA sequencing has been a solid foundation for various research fields such as taxonomy, phylogeny, ecology and genetic studies.

There are two classical methods of sequencing described as below-

1. Sanger dideoxy sequencing
2. Maxam-Gilbert sequencing

4-6.1.1 Sanger dideoxy sequencing

The Sanger or chain termination method was first developed by Fred Sanger and colleagues in the mid-1970s. It involves *in vitro* synthesis of DNA on a single-stranded template by using a primer, a set of labelled deoxynucleotide triphosphates (dNTPs), and dideoxynucleotide triphosphates (ddNTPs). Each ddNTP is fluorescently-labeled and impedes chain elongation. It lacks hydroxyl group at the third position of the sugar component which is required for attachment of the next nucleotide.

The template DNA to be sequenced can be obtained by in vivo or in vitro cloning. It can be elongated by incorporation of deoxynucleotides at random positions. The ratio of dNTP todideoxyNTP should be such that an individual strand can be polymerized for a significant distance before addition of dideoxyNTP molecule. The reaction terminates at the position where the ddNTP rather than dNTP incorporates into the growing DNA chain. This process results in the generation of amplicons of different sizes each ending in dideoxyNTP.

The next step involves separation of these fragments which can be achieved by acrylamide gel electrophoresis to obtain a sequence of up to a few hundred bases. Four parallel reactions can be carried out in parallel, one for each base. The base sequence can then be read by autoradiography of the banding patterns.

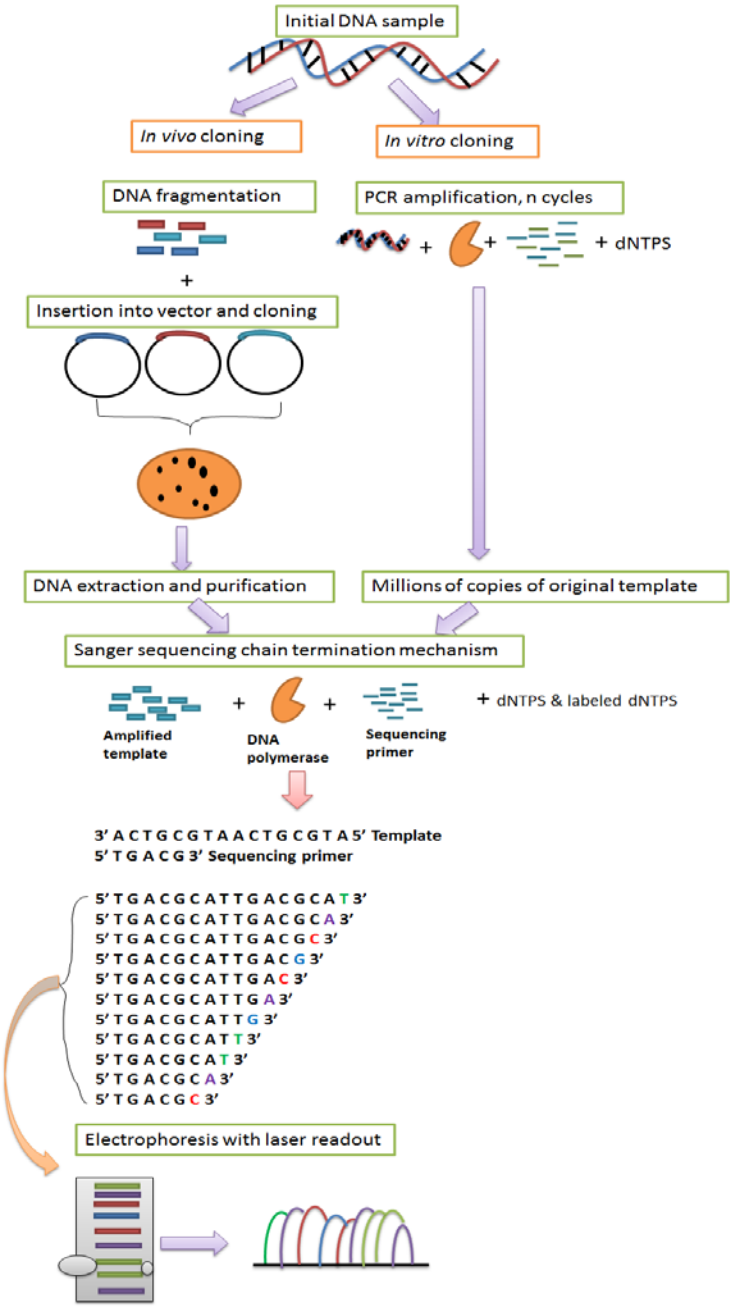


Figure4-6.1.1: Schematic representation of Sanger sequencing workflow.

(Adapted and modified from Hoy MA. 2013. DNA Sequencing and the Evolution of the "-Omics". Insect Molecular Genetics. 3rd Edn. Academic Press, San Diego, 251-305.)

4-6.1.2 Maxam-Gilbert Sequencing

This method was developed by Maxam and Gilbert in 1977 and is also known as chemical degradation method. In this method, double-stranded DNA to be sequenced is radioactively labeled at the 5′ end and undergoes selective fragmentation for the breakdown of specific base-pairing and dissociation of DNA. The resulting fragments are then loaded onto a polyacrylamide gel. Depending on the sizes of the fragments, the radioactively labelled fragments can be separated by electrophoresis, and the sequences are identified by autoradiography.

In contrast to Sanger sequencing, the samples to be sequenced need less complex preparation, but shorter reads (maximum 100 bp). This method itself is technically complex and utilizes hazardous chemicals.

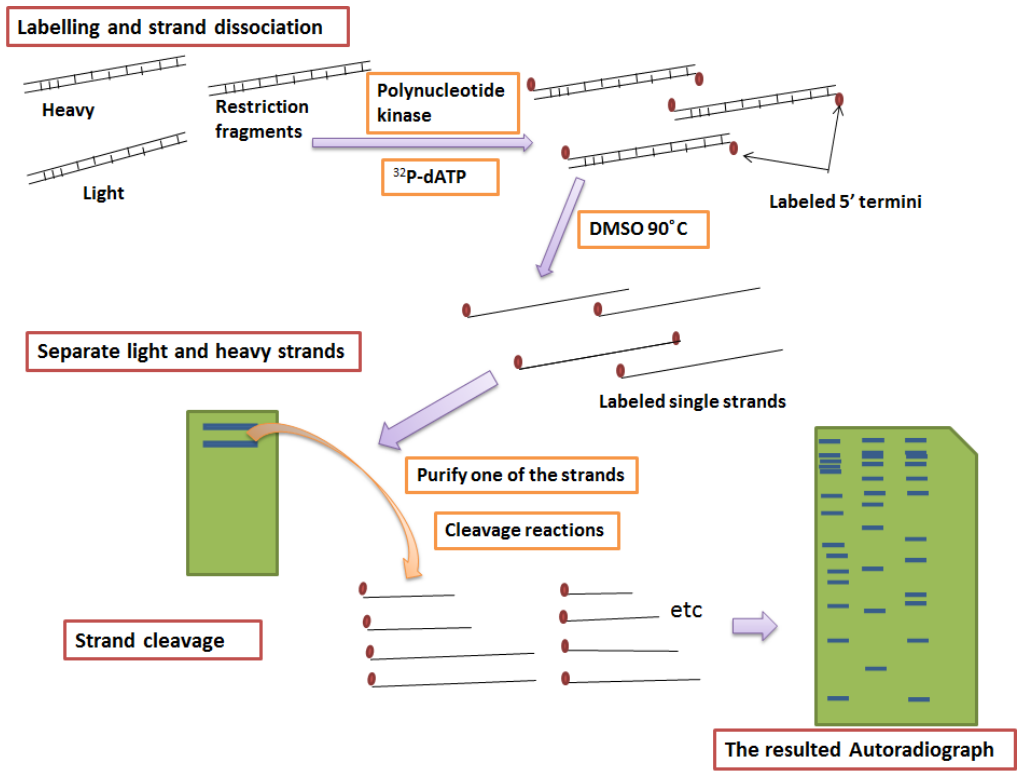


Figure 4-6.1.2 Schematic representation of Maxam-Gilbert sequencing workflow.

(Adapted and modified from Brown TA. 2006. Gene cloning and DNA analysis: an introduction.5th ed. Blackwell Scientific.)

4-6.2 CLONING STRATEGIES

The cloning of DNA fragment into a vector is routinely used method in recombinant DNA technology. Following are some strategies that are commonly used for cloning DNA fragment-

1. TA cloning
2. Gateway cloning

4-6.2.1 TA Cloning

It is a simple and more efficient cloning method without any requirement of restriction enzymes. It involves the use of terminal transferase activity of *Taq* polymerase which preferentially adds adenosine at 3’ end of the PCR amplified DNA molecule. To clone PCR products having 3’ overhangs, T-vector having 3’-T overhangs at both its ends can be used. This method is based on the complementary base pairing between the 3’-A overhangs of PCR product and 3’-T overhang of a vector molecule.

A single T-vector can be employed for easy and convenient cloning of any PCR amplified double-stranded DNA as well as blunt or sticky ended DNA molecules through certain minor modifications in TA cloning method. This technique, known as Universal TA cloning method, is useful in sub-cloning of DNA fragments when there are no compatible restriction sites available without compromising the cloning efficiency. Directional TA cloning can be possible by appropriate hemi-phosphorylation of the T-vectors and the inserts.

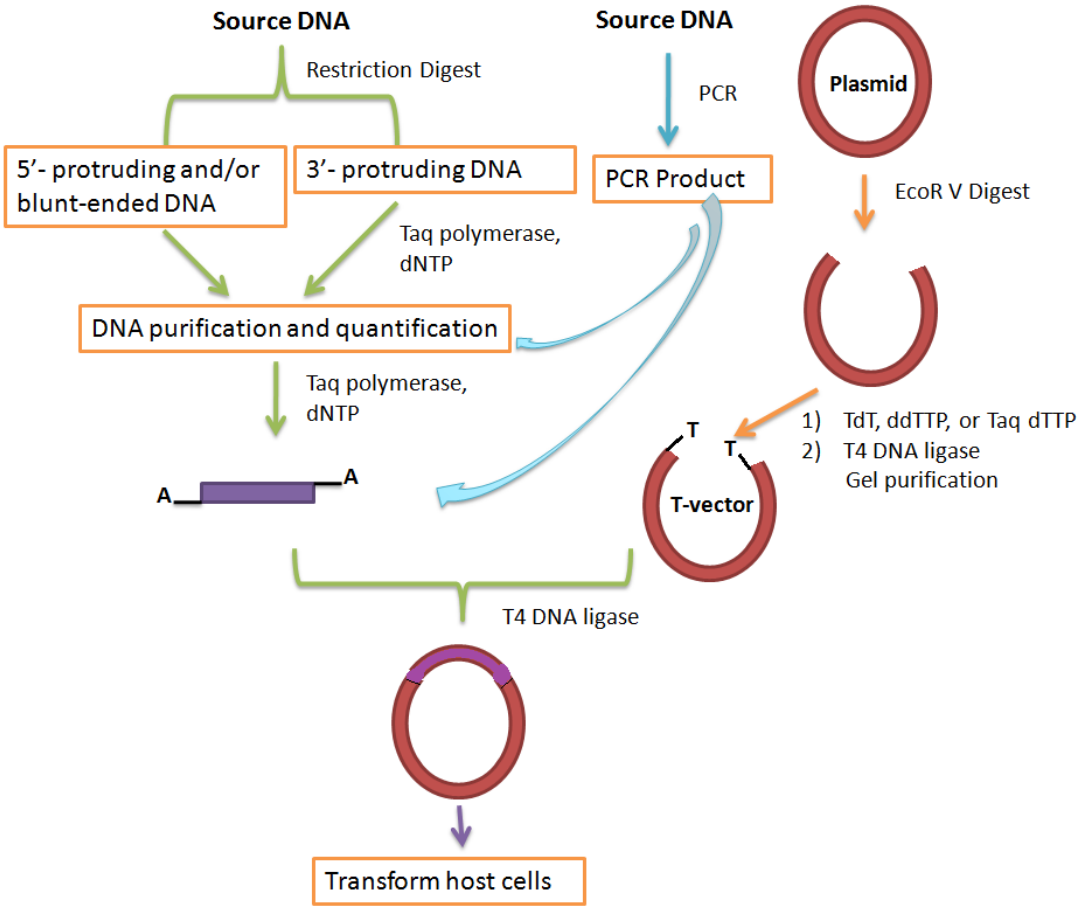


Figure 4-6.2.1(a) Schematic representation of the cloning procedures using the universal T-vectors

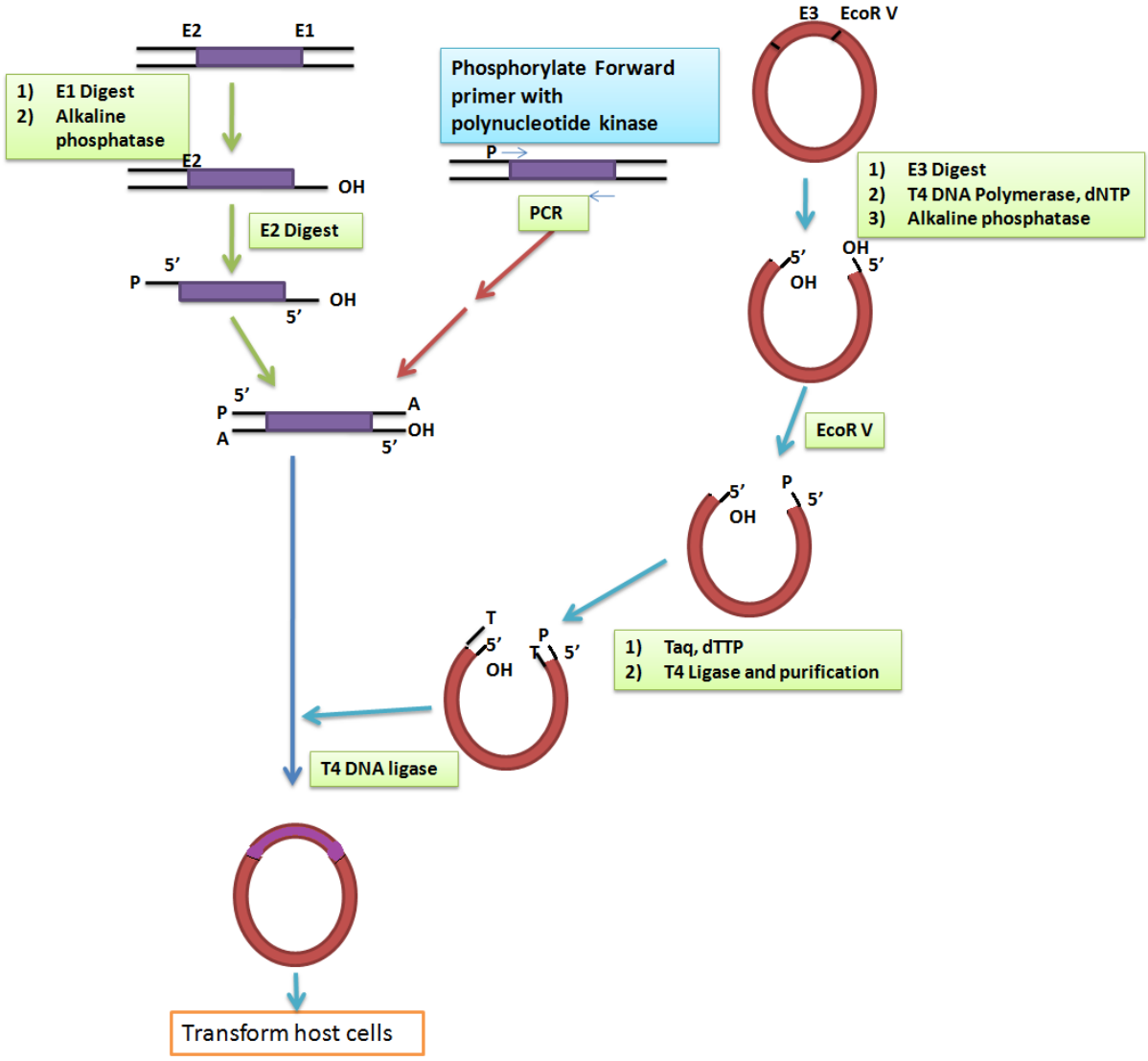


Figure 4-6.2.1(b) Schematic diagram demonstrating the strategies for directional TA cloning (E1, E2 and E3 - restriction enzymes, E3 - Generates blunt ends)

4-6.2.2 Gateway Cloning

The Gateway technology addresses the complications of efficiency, adaptability and compatibility of the traditional cloning procedures. It allows the cloning, combination and transfer of DNA fragments between various expression platforms in a high-throughput mode. It maintains the orientation and reading frame of the fragment/s of interest.

It is based on the site specific recombination method that mediates both integration and excision of bacteriophage λ. These events occur by recombination at specific attachment site on the phage DNA (*attp*) and bacterial chromosome (*attb*). The recombination begins when the phage encoded integrase and the bacterial integration host factor binds to the *attp* sequence and now this complex couples with the *attb* in the bacterial chromosome. Staggered DNA nicks are produced at the ends of *attp* and *attb* sequence followed by strand exchange between them. This results in the formation of a tiny heteroduplex joint at the point of exchange due to the presence of a short region (7 bp) of sequence homology in the two joined sequences. The holiday junction is then resolved and the prophage becomes flanked by two-hybrid *att* sites, called *attL* (for 'left') and *attR* (for 'right'). This initial reaction is referred as BP reaction. Although recombination reaction is reversible, different reaction condition prevail for each reaction ensuring that integration is not reversed by excision. Excision requires excisionase (Xis) and this reaction is known as LR reaction.

4-6.2.2.1 LR and BP Reactions

- The core of the Gateway is the *entry clone* which consists of a plasmid where the fragment of interest is flanked by *attL* sequences.
- Once the entry clone is obtained, the fragment of interest can be transferred to a secondary plasmid, the *destination vector* which is application-specific vector.
- This reaction is accomplished by mixing, *in vitro*, the entry clone (that usually carries *attL* sequences) with the destination vector of choice (that carries *attR* sequences) in the presence of a blend of the recombination proteins (Int, IHF and Xis), the LR Clonase™ enzyme mix.

- The product of this 'LR reaction' is the *expression clone*, which carries the DNA of interest and the donor vector, which is considered as a reaction byproduct.
- An aliquot of this reaction is used to transform competent *E. coli*.
- Two selection schemes are imposed to select the cells harboring the expression clone only.
  - First, the entry clone and the destination vector contain different antibiotic resistance genes (kanamycin and ampicillin in the as shown in Figure 4-6.2.2.1).
  - Second, the destination vector contains a counter selectable marker, the *ccdB* gene, which kills the cells by inducing gyrase-mediated double-stranded DNA breakage.
- Thus, transformants selected for ampicillin resistance contain only the expression clone.

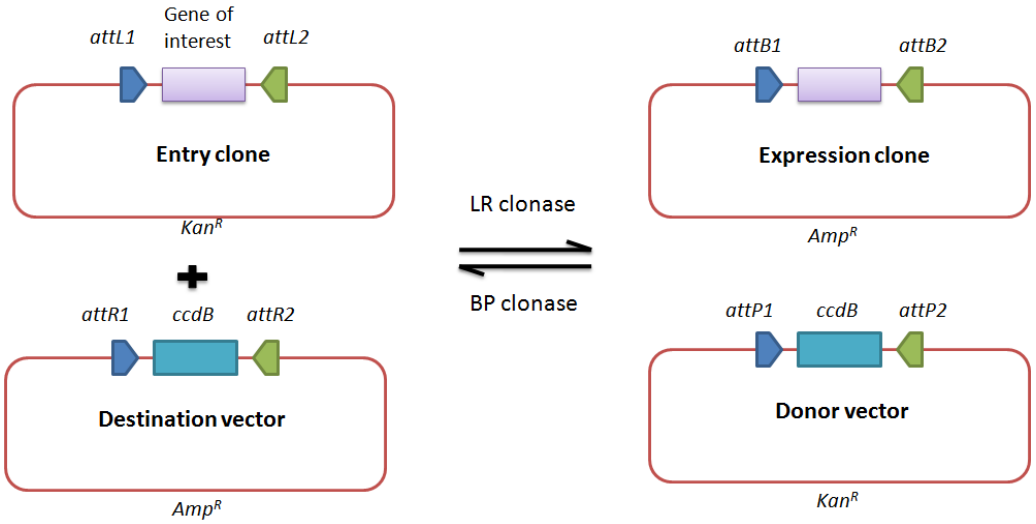


Figure 4-6.2.2.1 The Gateway reactions. The scheme shows the four types of plasmids and enzyme mixtures involved in Gateway cloning reactions. Red arrows represent the fragment of interest.

The Entry Clone

- The main door to enter a Gateway technology is the entry clone.
- There are 3 ways to generate this construct.
- First strategy is based on BP reaction.

- The second method is to ligate a PCR product into a TOPO- adapter entry clone. In this approach a single enzyme called vaccinia DNA topoisomerase having the ability to both cleave and rejoin DNA strands with extreme specificity at each step is used.
- Linearized vectors covalently attaches to the enzyme at 3' end to efficiently ligate the PCR product in a 5 minute reaction.
- The third method is the standard restriction cloning.

Destination Vectors

- A blunt-ended cassette comprising *attR1*-chloramphenicol resistance gene (CmR)-*ccdB-attR2* is inserted to any vector which results in a destination vector.
- Then the destination vector can be used in LR reaction to generate expression clone.
- Expression clones can also be generated by TOPO-cloning a PCR fragment into a TOPO-adapted expression vector.

Bibliography

- Brown TA. 2006. *Gene cloning and DNA analysis: an introduction.* 5th ed. Blackwell Scientific.
- Hoy MA. 2013. *DNA Sequencing and the Evolution of the "-Omics"*. Insect Molecular Genetics. 3rd Edn. Academic Press, San Diego, 251-305.
- Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, et al. 2013. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*, 110(1–2): 3-24.
- Ming-Yi Zhou and Celso E. Gomez-Sanchez. 2000. Universal TA Cloning. *Curr. IssuesMol. Biol*, 2(1): 1-7.
- Katzen F. 2007. Gateway® recombinational cloning: a biological operating system; Expert Opin. Drug Discov. 2(4):571-589
- Magnani E, Bartling L, Hake S. 2006. From Gateway to MultiSite Gateway in one recombination event. *BMC Molecular Biology*, 7:46.