Construction and Screening of Genomic and cDNA Libraries

Promila Sheoran
Ph.D. Biotechnology
GJU S&T Hisar

DNA (Gene) Libraries:

•A DNA library is a set of cloned fragments that collectively represent the genes of a particular organism. Particular genes can be isolated from DNA libraries, much as books can be obtained from conventional libraries.

•There are two general types of gene library: a genomic library, which consists of the total chromosomal DNA of an organism; and a cDNA library, which represents the mRNA from a cell or tissue at a specific point of time.

•The choice of the particular type of gene library depends on a number of factors, the most important being the final application of any DNA fragment derived from the library.

•If the ultimate aim understands the control of protein production for a particular gene or its architecture, then genomic libraries must be used.

•However, if the goal is the production of new or modified proteins, or the determination of tissue-specific expression of timing patterns, cDNA libraries are more appropriate.

•The main consideration in the construction of genomic or cDNA libraries is, therefore, the nucleic acid starting material.

•Since the genome of an organism is fixed, chromosomal DNA may be isolated from almost any cell type in order to prepare genomic DNA.

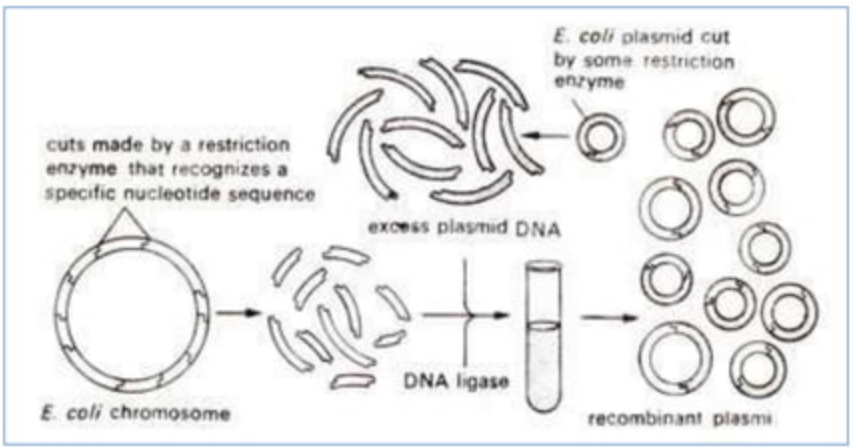
•In contrast, however, cDNA libraries represent only mRNA being produced from a specific cell type at a particular time in the cell's development. Thus, it is important to consider carefully the cell or tissue type from which the mRNA is to be deriver in the construction of cDNA libraries.

Constructing Gene Libraries:
Digesting Genomic DNA Molecules:

•After genomic DNA has been isolated and purified, it is digested with restriction endonucleases.

• These enzymes are the key to molecular cloning because of the specificity they have for particular DNA sequences.

•It is important to note that every copy of a given DNA molecule from a specific organism will give the same set of fragments when digested with a particular enzyme.



•Over 500 restriction enzymes, recognizing more than 200 different sites, have been characterized. The choice of which enzyme to use depends on a number of factors.

•For example, the recognition sequence of 6 bp will occur, on average, every 4096 (4^6) bases, assuming a random sequence of each of the four bases.

•This means that digesting genomic DNA with EcoRI, which recognizes the sequence 5′-GAATTC-3′, will produce fragments each of which is, on average, just over 4 kb. Enzymes with 8 bp recognition sequences produce much longer fragments.

•Therefore, very large genomes, such as human DNA, are usually digested with enzymes that produce long DNA fragments.

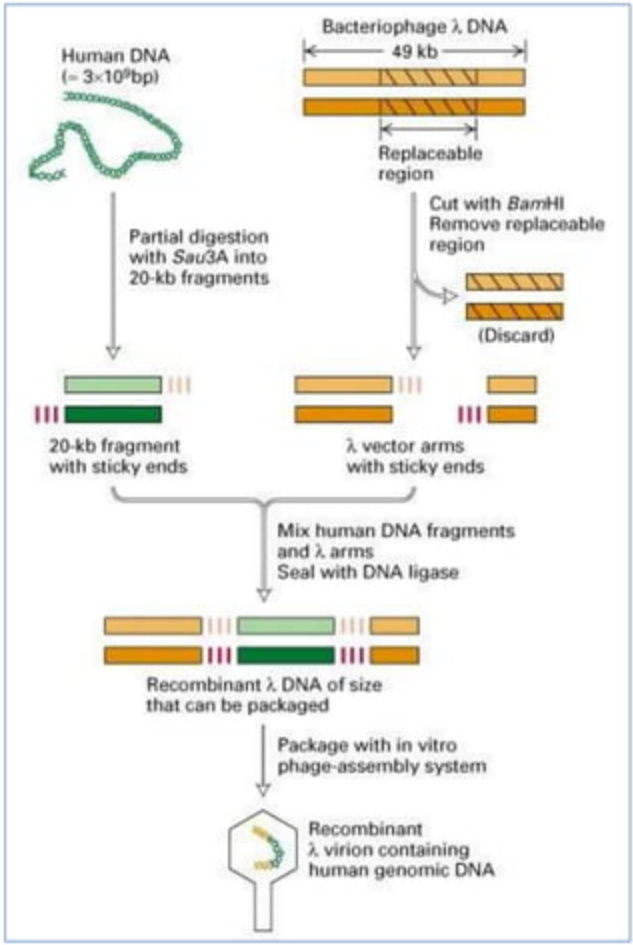
•This makes subsequent steps more manageable, since a smaller number of those fragments need to be cloned and subsequently analyzed.

Ligating DNA Molecules:

•The DNA products resulting from restriction digestion to form sticky ends may be joined to any other DNA fragments treated with the same restriction enzyme.

• Thus, when the two sets of fragments are mixed; base-pairing between sticky ends will result in the annealing of fragments that were derived from different starting DNA.

•All these pairing are transient, owing to the weakness of hydrogen bonding between the few bases in the sticky ends, but they can be stabilized by use of an enzyme, DNA ligase, in a process termed ligation.



•Each DNA fragment is inserted by ligation into vector DNA molecule, which allows the whole recombinant DNA to then be replicated indefinitely within microbial cells. In this way a DNA fragment can be cloned to provide sufficient material for further detailed analysis or for further manipulations.

•Thus, all of the DNA extracted from an organism and digested with a restriction enzyme will result in a collection of clones. This collection of clones is known as a gene library.

•Genomic libraries have been prepared from hundreds of different species. Many clones must be created to be confident that the genomic library contains the gene of interest.

•The probability, P, that some number of clones, N, contains a particular fragment representing a fraction, f, of the genome is
$P = 1 - (1 - f)^N$.
Thus, $N = \ln (1 - P) / \ln (1 - f)$.

•For example, if the library consists of 10 kbp fragments of the E. coli genome (4640 kbp total), over 2000 individual clones must be screened to have a 99% probability (P = 0.99) of finding a particular fragment.

•Since/ =10/4640 = 0.0022 and P – 0.99, N = 2093. For a 99% probability of finding a particular sequence within the 3×10^6kbp human genome, N would equal almost 1.4 million if the cloned fragments averaged 10 kbp in size.

•The need for cloning vectors capable of carrying very large DNA inserts becomes obvious from these numbers.

Screening DNA libraries to obtain gene of interest

•Library screening relies on information available about vector and genes of interest. In this, clones of the library are subjected to analysis to identify desired sequences. Different methods are used to identify clones harboring fragments of genes of interest in the library.

•A common method of screening plasmid-based genomic libraries is to carry out a colony hybridization experiment.

•In a typical experiment, host bacteria containing either a plasmid based or bacteriophage-based library are plated out on a petri dish and allowed to grow overnight to form colonies (or in the case of phage libraries, plaques).

•The probe and target DNA complementary sequences must be in a single stranded form if they are to hybridize with one another. Any DNA sequences complementary to probe DNA will be revealed by autoradiography of the nitrocellulose disc.

•Bacterial colonies (phage plaques) containing clones bearing target DNA are identified on the film and can be recovered from the master plate.

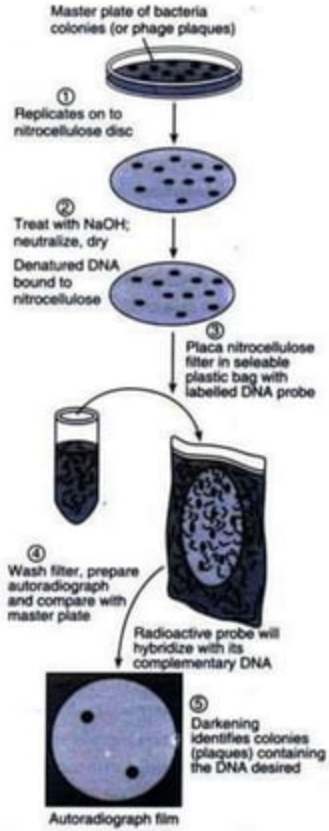


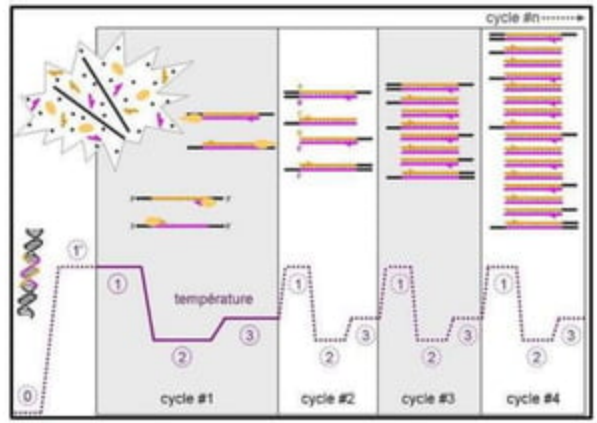
Fig. 4.10: Screening a genomic library by colony hybridization (or plaque hybridization). Host bacteria transformed with a plasmid-based genomic library or infected with a bacteriophage-based genomic library are plated on a petri plate and incubated overnight to allow bacterial colonies (or phage plaques) to form. A replica of the bacterial colonies (or plaques) is then obtained by overlaying the plate with a nitrocellulose disc. (1) Nitrocellulose strongly binds nucleic acids; single-stranded nucleic acids are bound more tightly than double-stranded nucleic acids. (Nylon membranes with similar nucleic acid – and protein-binding properties are also used.) Once the nitrocellulose disc has taken up an impression of the bacterial colonies (or plaques), it is removed and the petri plate is set aside and saved. The disc is treated with 2 M NaOH, neutralized, and dried. (2) NaOH both lyse any bacteria (or phage particles) and dissociates the DNA strands. When the disc is dried, the DNA strands become immobilized on the filter. The dried disc is placed in a sealable plastic bag, and a solution containing heat-denatured (single-stranded), labelled probe is added. (3) The bag is incubated to allow annealing of the probe DNA to any target DNA sequences that might be present on the nitrocellulose. The filter is then washed, dried, and placed on a piece of X-ray film to obtain an autoradiogram. (4) The position of any spots on the X-ray film reveals where the labelled probe has hybridized with target DNA. (5) The location of these spots can be used to recover the genomic clone from the bacteria (or plaques) on the original petri plate.

Polymerase Chain Reaction

•Colony hybridization is a laborious and time-consuming process and requires several rounds of plating and filter hybridization. It is also prone to false positives results.

•Polymerase chain reaction(PCR) has emerged as a robust technique in the area of molecular biology and is also efficient method for library screening. PCR allows isolation of DNA fragments from genomic DNA by selective amplification of a specific region of DNA.

•Although it is tedious to screen thousands of clones by but because of the specificity and sensitivity of PCR (if primer pair for gene of interest is available i.e. the sequence of the desirable fragment is known), it can be a useful way to identify clones of interest. Individual clones of the library can be subjected to PCR analysis.



cDNA Libraries:

•cDNAs are DNA molecules copied from mRNA templates. cDNA libraries are constructed by synthesizing cDNA from purified cellular mRNA.

•These libraries present an alternative strategy for gene isolation, especially eukaryotic genes.

•Because most eukaryotic mRNAs carry 3′-poly(A) tails, mRNA can be selectively isolated from preparations of total cellular RNA by oligo(dT)-cellulose chromatography.

•DNA copies of the purified mRNAs are synthesized by first annealing short oligo (dT) chains to the poly(A) tails.

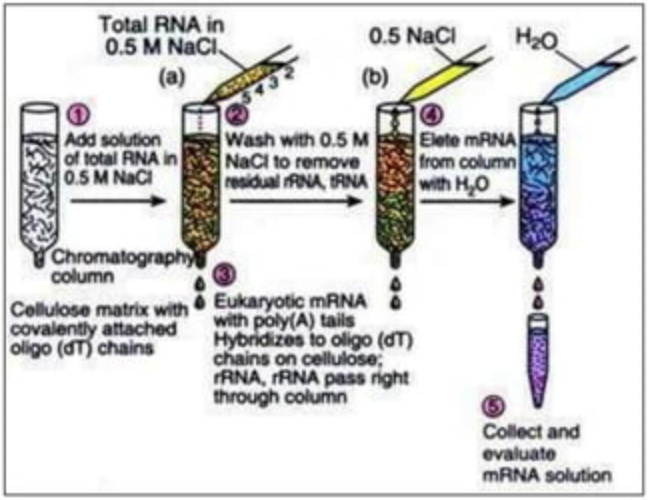


Fig. 4.12: Isolation of eukaryotic mRNA via oligo(dT)-cellulose chromatography. (a) In the presence of 0.5 M NaCl, the poly(A) tails of eukaryotic mRNA anneal with short oligo(dT) chains covalently attached to an insoluble chromatographic matrix such as cellulose. Other RNAs, such as rRNA (green), pass right through the chromatography column. (b) The column is washed with more 0.5 M NaCl to remove residual contaminants. (c) Then the poly(A) mRNA is recovered by washing the column with water because the base pairs formed between the poly(A) tails of the mRNA and the oligo(dT) chains are unstable in solutions of low ionic strength.

•These oligo(dT) chains serve as primers for reverse transcriptase-driven synthesis of DNA (Fig. 4.13). Reverse transcriptase is an enzyme that synthesizes a DNA strand, copying RNA as the template. DNA polymerase is then used to copy the DNA strand and form a double-stranded (duplex DNA) molecule.

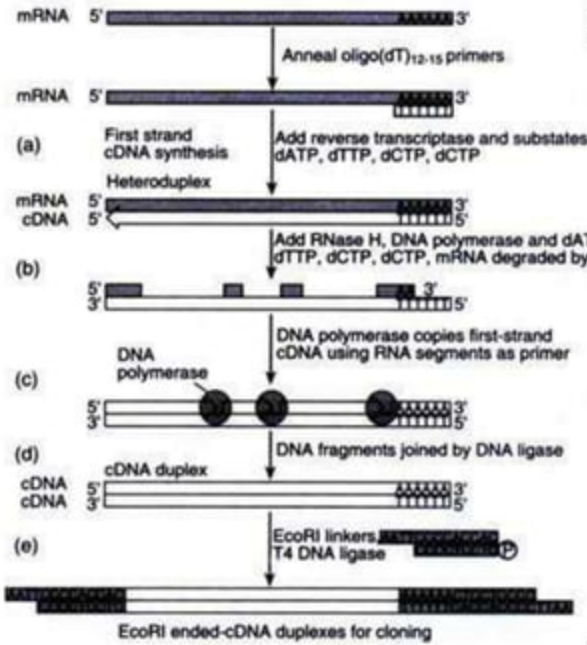


Fig. 4.13: Reverse transcriptase–driven synthesis of cDNA from oligo(dT) primers annealed to the poly(A) tails of purified eukaryotic mRNA. (a) Oligo(dT) chains serve as primers for synthesis of a DNA copy of the mRNA by reverse transcriptase. Following completion of first-strand cDNA synthesis by reverse transcriptase, RNase H and DNA polymerase are added (b). RNase H specifically digests RNA strands in DNA:RNA hybrid duplexes. DNA polymerase copies the first-strand cDNA, using as primers the residual RNA segments after RNase H has created nicks and gaps (c). DNA polymerase has a 5'→3' exonuclease activity that removes the residual RNA as it fills in with DNA. The nicks remaining in the second-strand DNA are sealed by DNA ligase (d), yielding duplex cDNA. *Eco*RI adapters with 5'-overhangs are then ligated onto the cDNA duplexes (e) using phage T4 DNA ligase to create *Eco*RI ended cDNA for insertion into a cloning vector.

•Lastly Linkers are added to the DNA duplexes rendered from the mRNA templates, and the cDNA is cloned into a suitable vector.

•Once a cDNA derived from a particular gene has been identified, the cDNA becomes an effective probe for screening genomic libraries for isolation of the gene itself.

•Because different cell types in eukaryotic organisms express selected subsets of genes, RNA preparations from cells or tissues in which genes of interest are selectively transcribed are enriched for the desired mRNAs.

• cDNA libraries prepared from such mRNA are representative of the pattern and extent of gene expression that uniquely define particular kinds of differentiated cells.

Thank You