Genomic DNA & cDNA Libraries

In order to study and map genes, researchers need to take potentially very large sections of DNA (such as a chromosome or whole genome), break them into smaller, manageable fragments, and clone these fragments to construct a DNA library.

Gene sequence are arranged in genome in a randon fashion and selecting or isolating a gene is a big task especially when the genomic sequences are not known. A small portion of genome is transcribed to give mRNA where as a major portion remained untranscribed. Hence, there are two ways to represent a genomic sequence information into the multiple small fragments in the form of a library: (1) Genomic library (2) cDNA library

Genomic library:

A genomic library is also known as clone bank or gene bank. It is a collection of DNA from a single organism, ideally though not necessarily containing its entire genomic DNA sequence. The DNA from the source organism of interest is divided into multiple fragments and packaged within cloning vectors such that each carries a portion of it. The vector DNA can then be inserted into host organisms - made by λ phage vectors, instead of plasmid vectors, for the following reasons.The entire human genome is about 3 x 109 bp long whereas a plasmid or λ phage vector may carry up to 20 kb fragment. This would require 1.5 x 105 recombinant plasmids or λ phages. When plating E. coli colonies on a 3" petri dish, the maximum number to allow isolation of individual colonies is about 200 colonies per dish. Thus, at least 700 petri dishes are required to construct a human genomic library. By contrast, as many as 5 x 104 λ phage plagues can be screened on a typical petri dish. This requires only 30 petri dishes to construct a human genomic library. Another advantage of the λ phage vector is that its transformation efficiency is about 1000 times higher than the plasmid vector.

Complementary DNA (cDNA):

cDNA is the reverse transcriptase product of mRNA and represents the coding sequence of all transcribed genes at the time of mRNA isolation

Genomic library and CDNA library are used in gene cloning to isolate different DNAs. The key difference between these two libraries is that genomic library contains DNA fragments that express the whole genome of an organism while in cDNA library, mRNA is taken from specific cells of an organism, and then cDNA is made from that mRNA in a reaction which is catalyzed by an enzyme.

Genomic Libraries	**cDNA Libraries**
1. From genomic DNA.	1. Reverse transcription of mRNA.
2. Frequency of hits independent of gene expression levels.	2. Frequency of hits dependent of gene expression levels.
3. May contain promoters and introns.	3. No promoters or introns.
4. Cannot express in heterologous system even if linked to suitable promoter.	4. Expression is feasible if linked to a suitable promoter.
5. Useful for genome analysis, map-based cloning, promoter studies, etc.	5. Useful for analysis of coding regions and gene functions

cDNA LIBRARY CONSTRUCTION

Historical Background of cDNA library synthesis

Gubler and Hoffman developed the original method for generating cDNA libraries by combining classical first-strand synthesis with RNase H DNA polymerase I mediated second-strand synthesis. This allows the conversion of first-strand to second-strand ds cDNA by RNA-primed

nick-translation without significant loss of sequence information (Gubler & Hoffman, 1983). Nowadays, modified Gubler and Hoffman's method is used.

Vectors used in the Construction of cDNA Library:

Both the bacterial and bacteriophage DNA are used as vectors in the construction of cDNA library.

The following table give a detailed information:

Table 6.2: Vectors used in the construction of cDNA library

Vectors	Insert size	Remarks
λ-phages	Up to 20-30kb (for replacement vectors) and 10-15kb (for insertion vectors)	• Maximum size of mRNA is about 8kb. Hence the capicity of DNA insert is not a major conern • Insertion vector system is usually employed • Useful for study of individual genes and their putative functions • Efficient packaging system, easy for gene transfer into *E. coli*, more representative than plasmid libraries, subcloning and subsequent DNA manipulation process are less convenient than plasmid systems
Bacterial plasmids	Up to 10-15kb	• Relatively easy to transform *E. coli* cells although may not be efficient as the λ-phage system for large scale gene transfer • Less representative than λ-phage libraries, subcloning and subsequent DNA manipulation processes are more convenient than the λ-pahge systems.

The cDNA Synthesis System

1. mRNA purification:

   Firstly, the mRNA is obtained and purified from the rest of the RNAs. Several methods exist for purifying RNA such as trizol extraction and column purification. Column purification is done by using oligomeric dT nucleotide coated resins where only the mRNA having the poly-A tail will bind. The rest of the RNAs are eluted out. The mRNA is eluted by using eluting buffer and some heat to separate the mRNA strands from oligo-dT.

2. cDNA synthesis:

   - first-strand DNA synthesis on the mRNA template (with a reverse transcriptase)

   Primers for "first strand" cDNA synthesis

   1) Oligo dT (binds polyA tails)
   2) Oligo dT with adaptors (restriction sites)
   3) Primers linked to a plasmid
   4) Random primers

   A short oligo (dT) primer is used. It anneals to the mRNA's poly (A) tail, allows reverse transcriptase to synthesize the cDNA (DNA-mRNA hybrid).

   cDNA Synthesis: The First Strand Reaction

   The classical conditions for the reverse transcription of RNA derive entirely from work with the Avian Myeloblastosis Viral (AMV) reverse transcriptase. This enzyme has been used extensively and its reaction parameters exhaustively documented. Available preparations of this enzyme are of variable quality, causing fluctuations in the quantity and quality of the first strand cDNA. The source of this variation is probably due to the level of contaminating RNase. As the RNase contamination becomes more severe, the first strand cDNAs will be shorter because the RNA template is degraded during first strand synthesis. In cases of extreme contamination, the yield of first strand cDNA will also decline as the size of the cDNA drops below the threshold of acid insolubility.

   An array of modifications and additions to the first strand reaction has been reported to improve the quality and yield of cDNA; some of these are attempts to reduce the effect of RNase contamination. The cDNA Synthesis System relies on a cloned enzyme, M-MLV reverse transcriptase. Cloned and overexpressed in *E. coli*, M-MLV reverse transcriptase is rigorously purified and does not require additional components to synthesize high quality first strand cDNA.

- removal of the RNA template

  RNase H degrades the mRNA strand, creating small fragments that serve as primers

- Second strand DNA synthesis using the first DNA strand as a template, (with a DNA-dependent DNA polymerase).

  DNA polymerase I makes new DNA fragments, DNA ligase connects them to make a complete chain.

  cDNA Synthesis: The Second Strand Reaction

  Many procedures for second strand synthesis have been developed. The main features of three commonly used methods are:

  1. Hairpin-Primed Synthesis

  2. Okayama and Berg Procedure

  3. Gubler and Hoffman Procedure

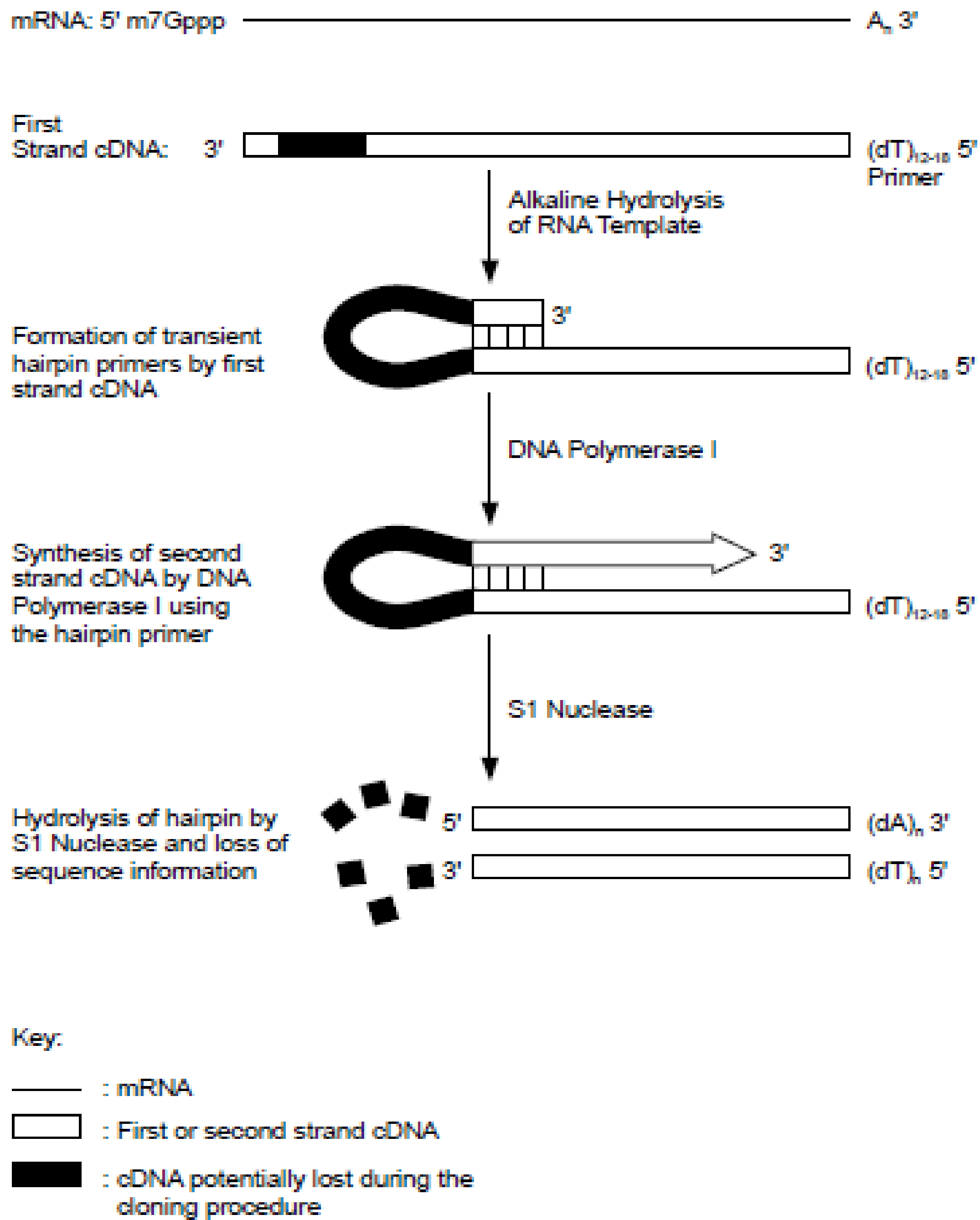

Figure 2. Hairpin-primed cDNA synthesis.

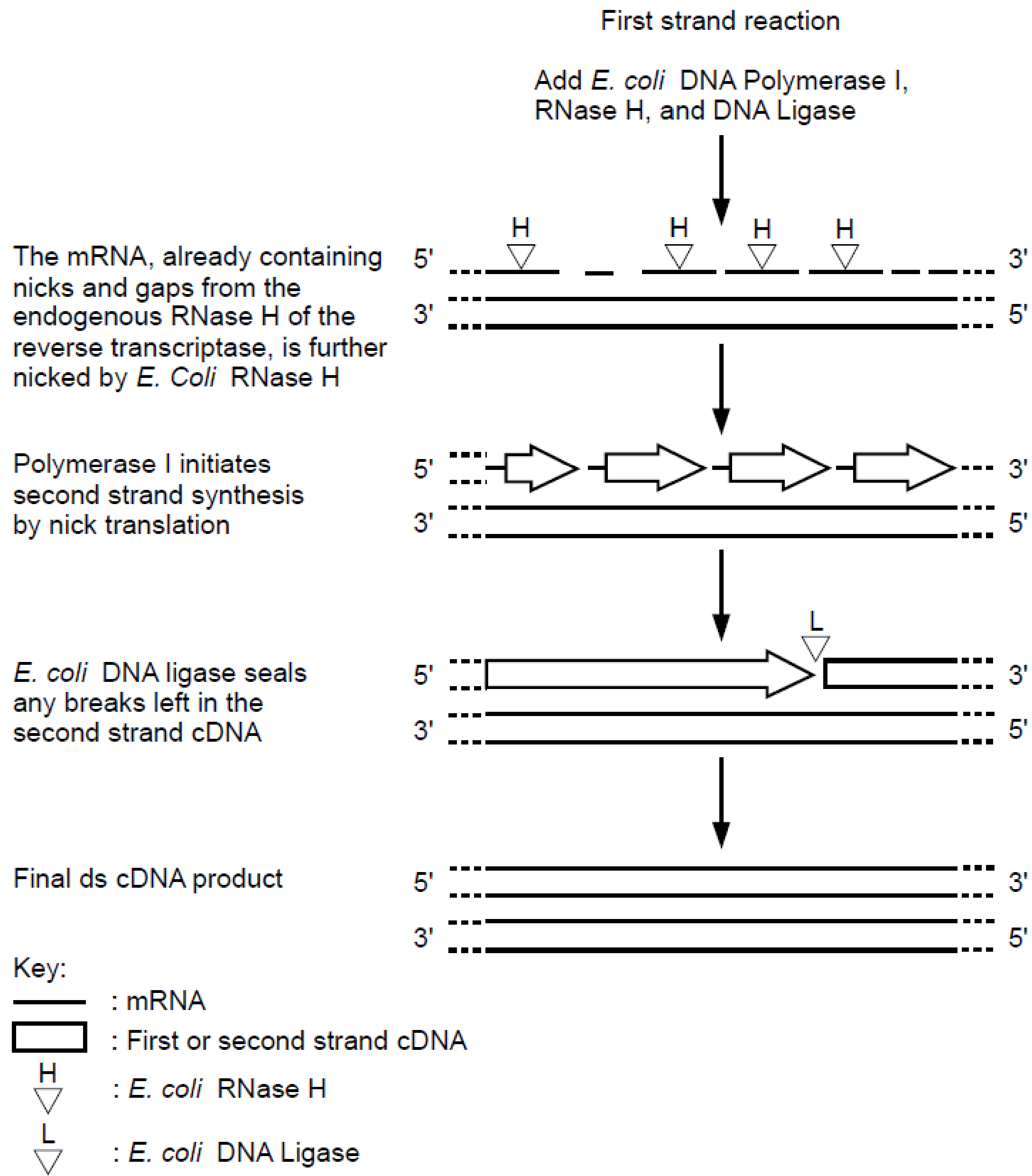

Figure 3. Second strand synthesis.

3. Preparation of dscDNA for cloning:

- Tailing with Terminal Transferase:

  One of the most popular methods for insertion of a cDNA into a plasmid is the addition of homopolymer tails to the DNA using terminal deoxynucleotidyl transferase. This method is fast and relatively simple. When dC tails of 20 to 30 nucleotides are added to the 3′ ends of the double-stranded cDNA, the cDNA can be annealed to a vector that has been linearized and tailed with dGTP. These annealed molecules do not need to be ligated and will transform *E. coli*. If the vector is linearized at a *Pst* I site and tailed with dGTP, the restriction endonuclease site is restored after repair *in vivo*. Thus, the cloned fragments can be excised with *Pst* I.

  The cDNA prepared by either of the methods can be tailed and cloned into a dG-tailed vector. The cDNA prepared in this manner will retain the greatest amount of sequence information originally encoded at the 5′ end of the message.

- Blunt-end Ligation:

  The cDNA can be cloned into either bacteriophage or plasmids by blunt-end ligation or the addition of molecular. If the RNase H is left out of the second strand reaction, the double stranded cDNA will contain residual RNA at one end. These molecules are very poor substrates for ligation, even if the ends are blunt, so the RNA must be removed by the combined action of E. coli RNase H and DNA polymerase I. The cDNA prepared should be treated with T4 DNA polymerase to render all of the termini perfectly blunt. After this treatment, the cDNA can be ligated using T4 DNA ligase into any plasmid that has been digested to produce blunt ends. This approach is simple but relatively inefficient and generally will result in fewer transformants per microgram of cDNA. Molecular linkers are a better alternative.

- Addition of Linkers:

  If it is important to maximize the number of clones obtained from a cDNA sample, the inefficient step of blunt-end ligation should be avoided by the addition of molecular linkers. Linker addition is generally performed when cloning into bacteriophage vectors but is equally applicable to plasmids. A disadvantage of linker addition lies in the number of enzymatic steps needed. The cDNA must first be methylated with the methylase specific for the linker that is to be added so that any internal restriction endonuclease sites resist cleavage. After repairing the termini with T4 DNA polymerase, linkers are added with T4 DNA ligase by blunt-end ligation.

In this case, the ligation is very efficient because the linkers are present at a very high concentration in the reaction. The DNA is then digested with the appropriate restriction endonuclease to expose the cohesive ends. The digestion products must be purified, usually by column chromatography, to separate the digested linkers from the cDNA before it can be ligated into a vector. Now that the cDNA has cohesive termini, it is ligated into vectors much more efficiently.

4. Cloning in vector:

   dscDNA prepared is then annealed in vectors for cloning. The vectors that can be used are:

   a. Plasmid:

      Up to 10-15 kb

      - Relatively easy to transform E. coli cells although may not be as efficient as the λ phage system for large scale gene transfer;
      - Less representative than λ phage libraries, subcloning and subsequent DNA manipulation processes are more convenient than the λ phage systems

   b. Bacteriophage:

      Maximum size for mRNA is about 8 kb, hence the capacity of DNA insert is not a major concern here;

      - Insertion vector system is usually employed;
      - Useful for study of individual genes and their putative functions
      - Efficient packaging system, easy for gene transfer into E. coli cells, more representative than plasmid libraries, subcloning and subsequent DNA manipulation processes are less convenient than plasmid systems

      - For replacement vectors: Up to 20-30 kb
      - For insertion vectors: 10-15 kb

5. Selection of cloned host. E.g. bacteria if bacterial plasmid is used for cloning.
6. Creation of stocks: stocks of the bacteria are created which are grown.
7. Sequencing to compile the cDNA library.

Top 3 Types of Specialized DNA Libraries :

1. Shelves:

Sometimes we know the size of the restriction fragment on which a particular gene is located. For example, this information may be acquired by probing a Southern blot of digested genomic DNA with a suitable sequence, such as an oligonucleotide probe, and measuring the size(s) of restriction fragment(s) that hybridize.

Once the size of the relevant restriction fragment is known, another digest of genomic DNA is then carried out with the same enzyme. The products are separated by electrophoresis, and DNA fragments of approximately that size are recovered from the gel.

They are then cloned into-a suitable vector. Because they are likely to be smaller than the random fragments used in making full genomic libraries, a plasmid vector is often suitable. The collection of recombinants generated is frequently called a shelf, as it is a subsection of a library.

2. Normalized Libraries:

A library made from cDNA that was prepared directly from mRNA will have a large number of members representing abundant RNAs and few representing the rare RNAs. If the cDNA we are looking for corresponds to an abundant RNA, that will increase the chances of finding an appropriate clone when the library is screened.

However, if the cDNA we are looking for corresponds to a rare RNA, then we will have to screen a large number of members of the library. It is possible to increase the representation of rare mRNAs in a library by a technique called normalization.

In the making of a normalized library a conventional cDNA library is first constructed using the RNaseH method for cDNA generation and cloning into a standard vector. The cDNA inserts are

then amplified using primers that flank the cloning site of the vector, melted by heating and then allowed to re-anneal.

Before re-annealing is complete, the DNA is passed down a column of hydroxyapatite, which binds more tightly to double-stranded nucleic acids than to single-stranded nucleic acids.

The single stranded material, therefore, passes through the column. This eluate is enriched for the less abundant sequences (as more abundant sequences will be more likely to have annealed in the time available). It can be enriched again by PCR if need be, and then cloned into a suitable vector as before.

3. Subtractive Libraries:

It is often very useful to make libraries that are enriched for sequences that are present in one sample but which are absent from another. These sequences might be present in the RNA from one tissue type but are absent from the RNA of another. Alternatively, they might be genomic DNA sequences that are present in a wild-type individual but are absent from a mutant that has a deletion in that region of the genome

Advantages of cDNA Library:

A cDNA library has two additional advantages. First, it is enriched with fragments from ac-tively transcribed genes. Second, introns do not interrupt the cloned sequences; introns would pose a problem when the goal is to pro-duce a eukaryotic protein in bacteria, because most bacteria have no means of removing the introns.

Disadvantages of cDNA Library:

The disadvantage of a cDNA library is that it contains only sequences that are present in mature mRNA. Introns and any other se-quences that are altered after transcription are not present; sequences, such as promoters and enhancers, that are not transcribed into RNA also are not present in a cDNA library.

It is also important to note that the cDNA library represents only those gene sequences ex-pressed in the tissue from which the RNA was isolated. Furthermore, the frequency of a par-ticular DNA sequence in a cDNA library depends on the abundance of the corresponding mRNA in the given tissue. In contrast, almost all genes are present at the same frequency in a genomic DNA library

Applications of cDNA Library :

1. It is used to express eukaryotic genes in prokaryotes. cDNA does not have introns and therefore can be expressed in prokaryotic cells.
2. It is useful in reverse genetics where the additional genomic information is of less use.
3. It is used for subsequently isolating the gene that codes for that mRNA.
4. It is used for analysis through bioinformatics.
    - The complete cDNA library of an organism gives the total of the proteins it can possibly express.
    - cDNA sequence gives the genetic relationship between organisms through the similarity of their cDNA
5. It is used in discovery of novel genes
6. It is used to obtain the sequences of genes for analysis, amplification, cloning, and expression
7. It is used for in vitro study of gene function
8. It is used in study of alternate splicing
9. It is useful in carcinogenic identification
10. It is used in commercial production of proteins and other biological products
11. It is used as a probe for DNA microarrays or as a clonal inserts into bacteria chromosomes
12. It is used for reproducing eukaryotic genomes, as the amount of information is reduced to remove the large numbers of non-coding regions from the library.