cDNA LIBRARY

In higher eukaryotes, many genes are transcribed into mRNA only in specialized cell types. For example, mRNAs encoding globin proteins are found only in erythrocyte precursor cells, called *reticulocytes.* Likewise, the mRNA encoding albumin, the major protein in serum, is produced only in liver cells where albumin is synthesized. The specific DNA sequences expressed as mRNAs in a particular cell type can be cloned by synthesizing DNA copies of the mRNAs isolated from that type of cell, and then cloning the DNA copies in plasmid or bacteriophage λ vectors. Just as a large collection of clones containing fragments of genomic DNA representing the entire genome of a species is called a ***genomic library,*** a large collection of cDNA copies of all the mRNAs in a cell type is called a ***cDNA library.***

DNA copy of mRNA are called ***complementary DNA (cDNA);*** clones of such DNA copies of mRNAs are called *cDNA clones.* In addition to representing only the sequences expressed as mRNAs in a particular cell type, cDNA clones lack the noncoding introns present in genomic DNA clones. Thus the amino acid sequence of a protein can be determined directly from the nucleotide sequence of its corresponding cDNA. Many genes in higher eukaryotes are too large to be included in a single λ clone because of their large introns. In contrast, all full-length cDNAs, containing the entire protein-coding sequence, can be included in a single λ clone.

A **cDNA library** is a collection of cloned cDNA (complementary DNA) fragments inserted into a collection of host cells, which together constitute some portion of the transcriptome of an organism.

cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism. Similarly, tissue specific cDNA libraries can be produced. In eukaryotic cells the mature mRNA is already spliced, hence the cDNA produced lacks introns and can be readily expressed in a bacterial cell. While information in cDNA libraries is a powerful and useful tool since gene products are easily identified, the libraries lack information about enhancers, introns, and other regulatory elements found in a genomic DNA library.

CONSTRUCTION OF cDNA LIBRARY

cDNA is created from a mature mRNA from an eukaryotic cell with the use of an enzyme known as reverse transcriptase. In eukaryotes, a poly-(A) tail (consisting of a long sequence of adenine nucleotides) distinguishes mRNA from tRNA and rRNA and can therefore be used as a primer site for reverse transcription.

Isolation of mRNAs:

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. The first step in preparing a cDNA library is to isolate the total mRNA from the cell type or tissue of interest. Nature has greatly simplified the isolation of eukaryotic mRNAs: the 3′ end of nearly all eukaryotic mRNAs consists of a string of 50–250 adenylate residues, the *poly(A) tail.* Because of their poly(A) tail, mRNAs can be easily separated from the much more prevalent rRNAs and tRNAs present in a cell extract by use of a column to which short strings of thymidylate (oligo-dTs) are linked to the matrix (oligomeric dT nuclotide coated resins Figure 1). When a cell extract is passed through an oligo-dT column, the mRNA poly(A) tails base-pair with the oligo-dTs, binding the mRNAs to the column. Since rRNAs, tRNAs, and other molecules do not bind to the column, they can be eluted out. The mRNA is eluted by using eluting buffer and some heat to separate the mRNA strands from oligo-dT. The bound mRNAs are recovered by elution with a low-salt buffer.

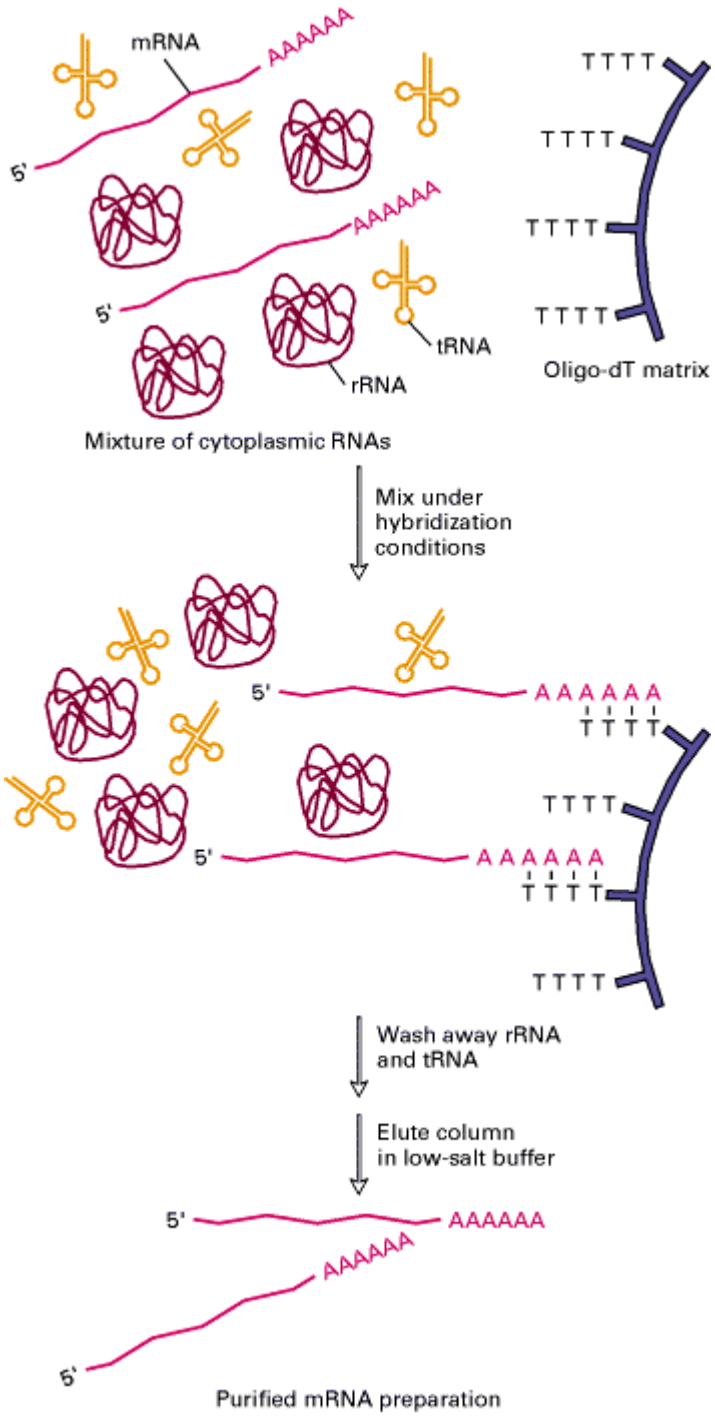

Fig. 1. Isolation of eukaryotic mRNA by oligo-dT column affinity chromatography.
Isolated cytoplasmic RNA consists mostly of ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs). The much less abundant mRNAs have 3′ poly(A) tails, which hybridize to oligo-dT covalently coupled to the column matrix. After hybridization, the rRNAs and tRNAs are washed out of the column; then the mRNAs are eluted with a low-salt buffer. The resulting purified mRNA preparation contains many different mRNA molecules encoding different proteins.

cDNA construction:

Once mRNA is purified, the following steps are performed to generate the cDNA;

- *oligo-dT* (a short sequece of deoxythymidine nucleotides) is attached as a complementary primer which binds to the poly-A tail providing a free 3'-OH.
- *Reverse transcription:* The 3′-OH end of the primer is extended by the enzyme **reverse transcriptase** to create the complementary DNA strand using the mRNA sequence as the template. Reverse transcriptase, purified from retroviruses, is used to synthesize a strand of DNA complementary to each mRNA molecule. This enzyme can polymerize deoxynucleoside triphosphates into a complementary DNA strand using an RNA molecule as template. Like other DNA polymerases, reverse transcriptase can add nucleotides only to the 3′ end of a preexisting primer base-paired to the template. Added free oligo-dT serves this function by hybridizing to the 3′ poly(A) tail of each mRNA template (Figure 2).

- *Removal of mRNA:* Now, the mRNA is removed by using a RNase enzyme leaving a single stranded cDNA (sscDNA).
- *Conversion of Single-Stranded cDNA to Double-Stranded cDNA*: This single stranded cDNA is converted into a double stranded DNA with the help of DNA polymerase I. However, for DNA polymerase to synthesize a complementary strand a free 3'-OH end is needed. **This is provided by the sscDNA itself by generating a *hair pin loop* at the 3' end by coiling on itself.** The polymerase extends the 3'-OH end.
- Later the loop at 3' end is cleaved by the scissoring action of the enzyme **S1 nuclease**.
- The enzyme terminal transferase may be used to add single stranded tails to this cDNA to facilitate their cloning in suitable vector.
- Restriction endonucleases and DNA ligase are then used to clone the sequences into bacterial plasmids.

Construction of cDNA library

The cloned bacteria are then selected, commonly through the use of antibiotic selection. Once selected, stocks of the bacteria are created which can later be grown and sequenced to compile the cDNA library.

USES OF CDNA LIBRARY

cDNA libraries are commonly used when reproducing eukaryotic genomes, as the amount of information is reduced to remove the large numbers of non-coding regions from the library. cDNA libraries are most useful in reverse genetics where the additional genomic information is of less use.

CDNA LIBRARY VS. GENOMIC DNA LIBRARY

As previously mentioned, a cDNA library lacks the non-coding and regulatory elements found in genomic DNA. Genomic DNA libraries provide much more detailed information about the organism, but are much more resource-intensive to generate and maintain.

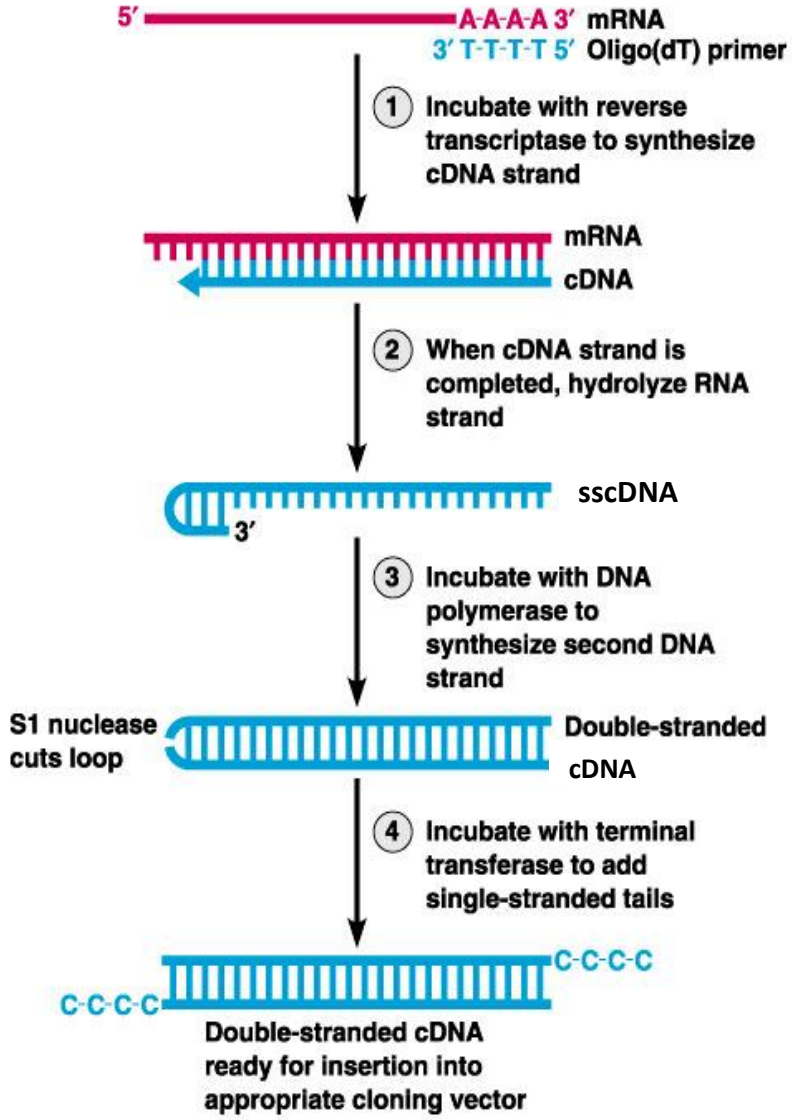

Fig. 2. Formation of cDNA from RNA

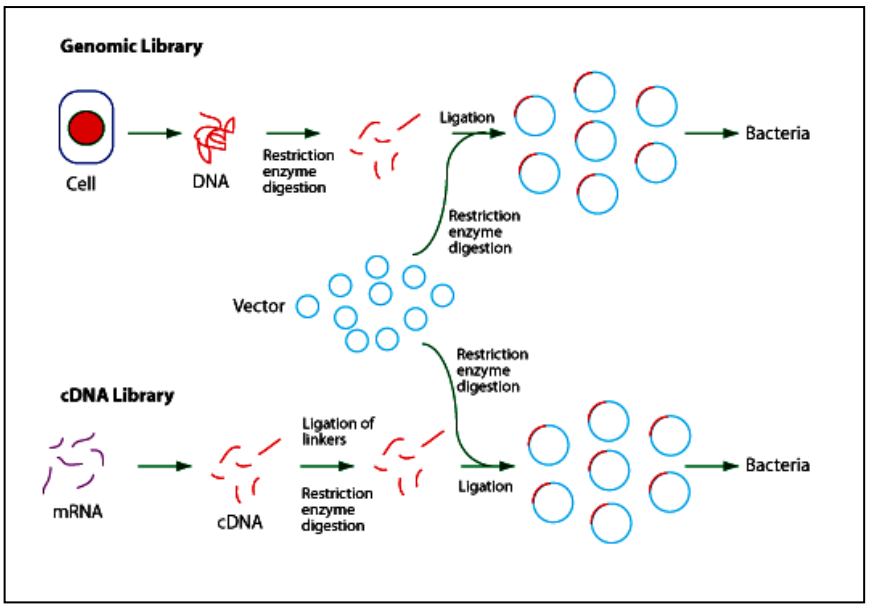

Fig. 3. Construction of Genomic and cDNA library