3.6: cDNA

cDNA clones are copies of mRNAs

Construction of cDNA clones involves the synthesis of complementary DNA from mRNA and then inserting a duplex copy of that into a cloning vector, followed by transformation of bacteria (Figure \(\PageIndex{1}\)).

![cDNA: first strand synthesis](https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes%2C_Nucleic_Acids%2C_Genomes_and_Chromosomes/3%3A_Isolating_and_Analyzing_Genes/3.06%3A_cDNA)

Product is complementary DNA, called **cDNA**. It is equivalent to the template strand of the duplex DNA.
Figure \ref{PageIndex1}:

\textit{Making cDNA clones}

a. \textbf{First strand synthesis}: First, one anneals an oligo dT primer onto the 3' polyA tail of a population of mRNAs. Then \textit{reverse transcriptase} will begin DNA synthesis at the primer, using dNTPs supplied in the reaction, and copy the mRNA into \textit{complementary DNA}, abbreviated \textit{cDNA}. The mRNA is degraded by the RNase H activity associated with reverse transcriptase and by subsequent treatment with alkali.

b. \textbf{Second strand synthesis}: For the primer to make the second strand of DNA (equivalent in sequence to the original mRNA), one can utilize a transient hairpin at the end of the cDNA. (The basis for its formation is not certain.) In other schemes, one generates a primer binding site and uses a primer directed to that site; one way to do this is by homopolymer tailing of the cDNA followed by use of a complementary primer. Random primers can also be used for second strand synthesis; although this precludes the generation of a full-length cDNA (i.e. a copy of the entire mRNA). However, it is rare to generate duplex copies of the entire mRNA by any means.

\textit{DNA polymerase} (e.g. Klenow polymerase) is used to \textit{synthesize the second strand}, complementary to the cDNA. The product is \textit{duplex cDNA}.

If the hairpin was used to prime second strand synthesis, it must be opened by a single-strand specific nuclease such as S1.

c. Insertion of the duplex cDNA into a cloning vector:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{cDNA_CloningDiagram.png}
\caption{Making cDNA clones}
\end{figure}
One method is to use terminal deoxynucleotidyl transferase to add a homopolymer such as poly-dC to the ends of the duplex cDNA and a complementary homopolymer such as poly-dG to the vector.

An alternative approach is to use linkers; these can be employed such that a linker carrying a cleavage site for one restriction endonuclease is on the 5’ end of the duplex cDNA and a linker carrying a cleavage site for a different restriction endonuclease is on the 3’ end. (In this context, 5’ and 3’ refer to the non-template, or “top” strand.) This allows “forced” cloning into the vector, and one has initial information about orientation, based on proximity to one cleavage site or the other.

The cDNA and vector are joined at the ends, using DNA ligase, to form recombinant cDNA plasmids (or phage).

d. The ligated cDNA plasmids are then transformed into E. coli. The resulting set of transformants is a library of cDNA clones.

Screening methods for cDNA clones

a. Brute force examination of individual cDNA plasmids.

If the mRNA is highly abundant in a given tissue, then many of the cDNA clones will be copies of that mRNA. One can examine DNA from individual clones and test for characteristic restriction cleavage patterns or a particular sequence. This was a common approach for screening cDNAs in the early days of recombinant DNA technology.

Starting in the mid-1990’s, cooperative efforts from corporations (such as Merck) and publicly funded genome centers (such as at Washington University) have generated the sequence of individual clones from large cDNA libraries from many tissues from human, mouse, and rat. Other consortia have sequenced cDNA libraries from other species. Each sequence is called an “expressed sequence tag” or EST. These are now a major source of partially or fully characterized cDNA clones. Hundreds of thousands of ESTs are available, and contain at part of the DNA sequence from many, if not most, human genes. The web site for NCBI (http://www.ncbi.nlm.nih.gov) is an excellent resource for examining the ESTs.

b. Hybridization with a gene-specific probe.

If the sequence of the desired cDNA is known, or if the sequence from homologs from related species is known, one can use synthetic oligonucleotides (or other source of the diagnostic sequence) as a radiolabeled hybridization probe to identify the cDNA of interest.

If the amino acid sequence has been determined for all or even just parts of the protein product of the gene of interest, then one can chemically synthesize oligonucleotides based on the genetic code for those amino acids. The oligonucleotides need to be at least 18 nucleotides or longer (so that they will anneal to specific sites in the genome), and because the genetic code is degenerate (more than one codon per amino acid; discussed in Part Two), they have to be degenerate as well. The oligonucleotides can be used directly as hybridization probes, although it is becoming more common to amplify the region between two oligonucleotides using the polymerase chain reaction, and to use that amplification product as a labeled probe.
The process of hybridization screening is illustrated schematically in Figure 3.16. The colonies of bacteria, each with a single cDNA plasmid, are transferred to a solid substrate (such as a nylon or nitrocellulose membrane), lysed, and the released DNA immobilized onto the membrane. Hybridization of this membrane (with the DNA attached) to a specific probe allows one to screen through thousands of colonies in a single experiment.

![Diagram of hybridization screening](https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes%2C_Nucleic_Acids%2C_Genomes_and_Chromosomes/3%3A_Isolating_and_Analyzing_Genes/3.06%3A_cDNA)

**Figure 3.2**: Hybridization Screening

**c. Express the cDNA, i.e. make the protein product encoded by the mRNA, and screen for that protein product** (Figure 3.3)). This is often in bacteria by constructing the clones in a vector that has an active E. coli promoter (for transcription) and efficient translation signals upstream from the site at which the cDNAs were inserted. The transformed bacterial cells will express the encoded protein, and one tries to identify it. One can also screen for expression in yeast, plant or mammalian cells. The expression vector has to contain gene-regulatory signals (such as promoters and enhancers, see Part Three) that allow expression of the desired gene in the appropriate cell.

![Diagram of screening for an expressed gene product](https://bio.libretexts.org/Bookshelves/Genetics/Book%3A_Working_with_Molecular_Genetics_(Hardison)/Unit_I%3A_Genes%2C_Nucleic_Acids%2C_Genomes_and_Chromosomes/3%3A_Isolating_and_Analyzing_Genes/3.06%3A_cDNA)

**Figure 3.3**: Screening for an Expressed Gene Product

1. One can use specific antiserato detect the desired colony expressing the gene of interest.
2. One can use a labeled ligand that will bind to the expressed cDNA on the cell surface. For example, cDNAs for receptors can be expressed in an appropriate cell (usually mammalian cells in culture) and identified by newly-acquired ability to bind a labeled hormone (such as growth hormone or erythropoietin)
3. by complementationof a known mutation in the host. E.g. a cDNA for the human homolog to yeast p34cdc2 was isolated by its ability to complement a yeast mutant that had lost the function of this key regulator of progress through the cell cycle.
4. Expression cloning can be done in mammalian cells, as long as one can screen or select for a new function generated by the expression. Use of this method to isolate the receptor for the glycoprotein hormone erythropoietin is illustrated in Figure 

\[ \text{Expression screening in eukaryotic cells.} \]

![Diagram](image)

\[ \text{Cell line that needs a cytokine (e.g. IL-3) to grow. Has no Epo receptor, will not grow in Epo.} \]

\[ \text{Expression library: cDNA inserts in a vector that will drive expression in eukaryotic cells.} \]

\[ \text{A transformed cell line that expresses the Epo receptor will now grow in Epo without IL-3. The plasmid with the Epo receptor cDNA can be isolated from this cell line.} \]

**Figure \(\PageIndex{4}\):** Expression screening in eukaryotic cells.

d. **Differential analysis:**

Often one is interested in finding all the genes (or their mRNAs) that are expressed uniquely in some differentiated or induced state of cells. Two classic examples are (i) identifying the genes whose products regulate the determination process that causes a multipotential mouse cell line (like 10T1/2 cells) to differentiate into muscle cells, and (ii) using the fact that the T-cell receptor is expressed only in T-lymphocytes, but not in their sister lineage B-lymphocytes, to help isolate cDNA clones for that mRNA. Both of these projects used subtractive hybridization to highly enrich for the cDNA clones of interest.

In this technique, the cDNA from the differentiating or induced cell of interest is hybridized to mRNA from a related cell line, but which has not undergone the key differentiation step. This allows one to remove mRNA-cDNA duplexes that contain the cDNAs for all the genes expressed in common between the two types of cells. The resulting single-stranded are enriched for the cDNAs that are involved in the process under study.

The subtractive hybridization scheme used in isolation of the muscle determination gene MyoD is illustrated in Figure \(\PageIndex{5}\).

A conceptually equivalent strategy, using PCR (see next section) rather than cDNA cloning, is differential display of PCR products from cells that differ by some process (e.g. differentiation, induction, growth arrest versus stimulation, etc.). In this technique, one uses several sets of PCR primers annealed to cDNA to mRNA from the two types of cells that are being compared. The sets of primers are empirically designed to allow many regions of cDNA to be amplified. The amplification products are resolved (or displayed) on polyacrylamide gels, and the products specific to the cell type of interest are isolated and used to screen through cDNA libraries. This technique is also called representational difference analysis.
Figure \(\PageIndex{5}\): Differential screening to find cDNAs of mRNAs expressed only in certain cell-types.

The advent of sequencing all or a very large number of genes from various organisms (e.g. *E. coli*, yeast, *Drosophila*, humans) has allowed the development of high-density microchip arrays of DNA from each gene. One can hybridize RNA from cells or tissues of interest, isolated under various metabolic conditions, to identify all (known) genes expressed. Even more useful are assays for genes whose expression *changes* during a shift in cell metabolism (cell cycle, heat shock, hormonal induction, etc.) or as a result of mutation of some other gene (e.g. a gene encoding a transcription factor of interest). This powerful new technology is being used more and more to examine global effects on gene expression.