Lecture 9 - Biochemistry of cDNA Synthesis (AMG text pp. 115-127)

September 18, 2001

Overview of cDNA synthesis reactions

One of the most useful applied molecular genetic methods is the cloning of complementary DNA (cDNA) sequences derived from the coding sequences of specific genes. The ability to "capture" mRNA as cDNA was made possible by the independent discovery of a viral RNA-dependent DNA polymerase called reverse transcriptase (RTase) by Howard Temin and David Baltimore. RTase uniquely synthesizes DNA products from primed RNA templates.

By synthesizing cDNA from purified mRNA, it makes it possible to focus only on the transcribed regions of the genome. RNA processing of primary transcripts creates the "business end" of genes. Craig Venter of Celera Genomics bet the company on this idea by initially only cloning and sequencing expressed genes using cDNA cloning.

Will cloning and sequencing only cDNA guarantee that you have identified all known genes of an organism? Explain.

What are randomly cloned cDNAs of unknown function referred to in the field of genomics?

Avian myelobastosis virus (AMV) reverse transcriptase was the first RTase specifically purified for use in first strand cDNA reactions. The active enzyme consists of two subunits that together encode the DNA polymerizing activity, and an RNase activity called RNase H which degrades RNA in RNA:DNA heteroduplexes.

The RTase of Moloney murine leukemia virus (MMLV) is a single polypeptide chain that encodes all the required RTase functions. The MMLV RTase has been cloned and re-engineered to have negligible levels of RNase H activity, without compromising its first strand cDNA polymerizing function.

Protein structure studies of MMLV RTase that reveal molecular interactions between the template DNA and the enzyme have been determined. The molecular structure of MMLV RTase protein can be viewed through the Protein Data Bank (PDB) Internet site.

The activity of HIV reverse transcriptase can be inhibited by nucleoside analogs.

What viral RTase do you think is one of the most studied enzymes in the last 10 years?

What would be the benefit of knowing the protein structure of this highly studied protein?

What is AZT and what is the molecular rationale for using it to treat human acquired immune deficiency syndrome?

What types of mutations in HIV reverse transcriptase would result in AZT resistance?

Why do you think HIV strains are able to mutate so quickly, what is it about their genome replication mode that promotes a high mutation rate?

The Four basic steps in constructing a cDNA library:

1. Purification of mRNA using chemical extraction and oligo-dT purification.
2. First strand cDNA synthesis using oligo-dT, random pdN6, or specific primers.
3. Second strand cDNA synthesis requires a priming event; done with RNaseH.
4. Repair of cDNA termini and ligation of adaptor oligos; clone into vector.

1. Purification of mRNA

Since RNA is single-stranded, any hydrolysis event that breaks the phosphate backbone will result in cleavage of the molecule into subfragments. Two factors contribute to the biochemical instability of RNA.

1. Endoribonucleases (RNases) are very stable enzymes that cannot be easily inactivated. In fact, human hands are a rich source of RNase and it is therefore necessary to wear clean latex gloves during RNA isolation procedures and to use RNase-free labware.

2. RNA is thermodynamically less stable than DNA because of the 2’ hydroxyl group on the ribose ring that promotes hydrophilic attack on the 5’-3’ phosphodiester bond to form a 2’-3’ cyclic phosphate. This cyclic phosphate intermediate is stabilized by Mg++, a component of many biochemical reactions.

It is critical that the RNA be isolated intact and pure so that it can function as a faithful template for first strand synthesis. One method is to purify mRNA from tissue culture cells using guanidinium thiocyanate and oligo dT cellulose. Total RNA remains in the aqueous phase under acidic pH conditions following phenol:chloroform extraction. After precipitation with isopropanol, the RNA solution is adjusted to high salt (0.5M NaCl) and loaded onto an oligo-dT cellulose column.

It has been observed that it is very difficult to isolate intact mRNA from some tissues such as rat pancreas. What makes some tissues more difficult to work with than others?

What can be done to try and minimize RNA degradation when working with "difficult" samples?

What is the recommended long term storage condition for highly purified mRNA?

How could you test if a mRNA sample is likely to be useful for constructing a cDNA library prior to committing time and effort to the actual library, i.e., what QC (quality control) procedure would you use?

2. First strand cDNA synthesis

Libraries made with oligo dT primers (a poly dT strand 12-18 nucleotides long) have the best chance of containing long cDNAs since the priming event is targeted to the 3' poly-A tail of the mRNA. Short random hexanucleotide primers can also be used to initiate cDNA synthesis. These nonspecific template primers bind to multiple sites along the mRNA and can provide a better distribution of coding and non-coding sequences. It is sometimes possible to use gene-specific primers to initiate first strand synthesis in cases where a portion of the 3' cDNA sequence is known.

3. Second strand cDNA synthesis

Second strand cDNA synthesis presents some challenges that aren't encountered during replication of the retroviral genome in infected cells. The big problem is how to prime second strand synthesis without creating DNA termini that block vector ligation.

In vivo retroviral replication involves a complicated series of priming events that are initiated by host cell tRNA molecules and require complementary sequence elements within the long terminal repeats (LTR) of the retroviral genome. Since cDNA copies of mRNA molecules do not contain these special priming sequences, an artificial priming event is required to initiate second strand synthesis.

Replication of retroviral RNA genomes in infected cells involves priming by cellular tRNA molecules that initiate the synthesis of a short stretch of cDNA that functions as a primer at the other end. Cool!

Several different priming strategies have been developed, one of which was based on the observation that RTase will occasionally create first strand 3' termini that serve as primers for second strand synthesis through the formation of "hairpin loops". However these hairpins only occur on <10% of the cDNA products and the looped structure needs to be cleaved by S1 nuclease in order to create a double-strand DNA terminus for vector ligation.

To improve the efficiency of second strand cDNA synthesis, researchers reexamined the retroviral RTase reaction and discovered that by adding small amounts of E. coli RNase H to the second strand synthesis reaction, they were able to find conditions that resulted in the production of short RNA primers. In the presence of E. coli DNA polymerase I, these RNA primers promote DNA synthesis at multiple sites along the cDNA template.

The "replacement synthesis" reaction for second strand cDNA synthesis using E. coli RNase H and DNA pol I. The RNase H activity creates short RNA fragments in the RNA:DNA heteroduplex that function as primers for second strand cDNA synthesis by E. coli DNA pol I. The addition of bacteriophage T4 DNA polymerase to the reaction creates blunt double-stranded DNA termini that are suitable substrates for adaptor ligation.

Why is it not a good idea to have RNase H present for the first strand reaction, but it is required for the second strand reaction in the replacement strategy?

Does the replacement strategy permit the synthesis of full-length cDNA such that the product includes the first nucleotide of the mRNA transcript? Explain.

4. Ligation of adaptor oligos and vector ligation

Modification of cDNA termini is the final step of cDNA synthesis that must be performed prior to vector ligation. The most straightforward, but least efficient, method is to ligate blunted-ended cDNA molecules to a vector that has been digested with an enzyme that produces blunt (flush) ends. Since blunt-end ligation by itself is inefficient, double-stranded oligonucleotide adaptor molecules are first ligated to the blunt-end cDNA termini. These special adaptor molecules convert the blunt-ended cDNA molecule to more conventional DNA fragments possessing cohesive ends.

EcoRI adaptors can be ligated onto blunted cDNA termini as a first step towards constructing a cDNA library in an EcoRI-digested lambda cloning vector. Most adaptors contain internal recognition sites for rare-cutting enzymes, such as NotI, to facilitate subcloning of insert fragments that contain internal EcoRI sites.

How is the blunt-ended ligation reaction using adaptor molecules and cDNA set-up to maximize the number of cDNA molecules that obtain covalently attached adaptors?

Assuming that the cDNA cloning vector contains the coding sequence for the alpha subunit of LacZ in the Eco RI cloning site, would blue-white screening distinguish between recombinants that contain cDNA inserts, and those that only have adaptor dimers? Explain.

If after isolating a cDNA clone from this library, you discover that the insert contains an internal Eco RI site, what two explanations could account for this? What would you need to do to distinguish between these possibilities?

How would you estimate the number of primary recombinants in your library and how would you amplify your library to create ampoules for sale on the open market?

Types of cDNA cloning vectors

The choice of using a plasmid vector or lambda phage vector for constructing a cDNA library depends on the screening method that will be used. Plasmid vectors are commonly used for cDNA expression libraries that will be screened by a functional assay. One of the most convenient lambda phage cDNA cloning vectors are those that incorporate an f1 bacteriophage origin of replication into the vector backbone to permit in vivo excision of insert cDNA. The lambda Zap cDNA cloning vectors from Stratagene Cloning Systems are examples of lambda phagemid vectors.

The primary advantage of lambda phagemid vectors is that they combine the high transfection efficiency of recombinant lambda phage vectors, with the rapid in vivo excision capability of M13-based phagemids. The linear lambda Zap Express cDNA

cloning vector can be converted in vivo to a circular eukaryotic expression plasmid by an excision process called phagemid rescue.

An example of a cDNA cloning vector based on the lambda Zap system is pBk-CMV which can be used to expressed cDNA inserts in transfected mammalian cells (CMV promoter is a eukaryotic viral promoter; chapter 7).

Screening cDNA libraries
cDNA libraries are commonly screened by either of two different approaches:

1. Isolation of clones based on cDNA sequence alone which reflects gene structure.

2. Isolation of clones based on a measureable function of the encoded protein.

Typically, the cloning vector determines which type of screen can be performed. If DNA sequence is the only criteria for identification then most any lambda vector is sufficient, however, if function of the encoded protein is required (or antigenicity), then the vector needs to contain transcriptional and translational control elements.

Screening a cDNA library with degenerate oligonucleotide probes
The amino acid sequence of polypeptide fragments of a highly purified protein can usually be determined by automated protein microsequencing based on the Edman degradation procedure or mass spectrometry. Peptide sequences can be used to predict a nucleic acid sequence based on triplet codons in the Genetic Code. Custom oligonucleotide synthesis can then be used to generate a nucleic acid probe for library screening.

Two polypeptide sequences from the same protein can be used to generate separate mixed oligonucleotide pools for use as probes for cDNA library screening.

By initially screening the cDNA library with one oligonucleotide pool from a specific region of the protein, and then re-screening all candidate clones with a second oligonucleotide pool, it is possible to identify a subset of double-positive clones that warrant further characterization.

How many different oligonucleotides are required to generate the NT probe based on the Genetic Code?

What is the rationale for using the two probes in successive screenings? Will any cDNAs be missed this way?

Does the oligonucleotide pool contain any probe molecules that are 100% complementary to the cDNA? Explain.

What is included in the hybridization solution to minimize differences in oligonucleotide Tm? Explain.