Sequencing cDNAs: An Introduction to DNA Sequence Analysis in the Undergraduate Molecular Genetics Course

Samuel Galewsky
Biology Department
Millikin University
1184 W. Main St.
Decatur, IL 62522
Email: sgalewsky@mail.millikin.edu

Abstract: DNA sequence analysis is a cornerstone of modern molecular genetics, yet it is difficult to provide students in the teaching lab with an experience that is meaningful and cost effective. At Millikin University in the upper division Molecular Genetics course, a series of lab exercises that culminates with a cDNA sequence that students can use for homology searches and open reading frame analysis has been implemented. The exercises begin with students picking a single colony from a Drosophila melanogaster embryo cDNA library. They purify the plasmid then analyze the insert through restriction digests and gel electrophoresis. In this semester’s example, three cDNAs were then randomly picked from the group and sent out for sequencing. The students had an ownership and interest in the sequence data because they represent an unknown gene and were selected by them.

Key Words: Drosophila melanogaster; cDNA; DNA sequencing; molecular biology; undergraduate teaching

Introduction
Molecular biology has become a fundamental tool in virtually all aspects of biology. From developmental genetics to taxonomy and environmental biology, the analysis of DNA sequence is a part of the array of techniques necessary for carrying out experiments that answer important questions. These days, a college course in molecular genetics has to bridge the gap between all of the historical and biological information and the huge amount of technical knowledge needed to work in the lab. Many of these techniques are relatively easy to carry out in the teaching lab and follow fairly simple “recipe” formats. Things like DNA isolations, gel electrophoresis and restriction digests have become exercises that are carried out in some high schools. A difficulty in the teaching lab setting occurs when more complicated ideas and techniques are incorporated. In this age of genome sequencing one of the most important concepts is becoming DNA sequence analysis. This requires the fusion of a wide range of techniques including all of the basic techniques mentioned above as well as computer analyses and homology searches. The importance of this area of molecular biology is apparent in the current literature and in the job market. As a way of exploring this critical concept, I developed a sequence of lab activities that allows students to isolate cDNA clones, characterize the inserts and have them sequenced by a remote facility. The resulting DNA sequence was recovered through the Internet and was analyzed for open reading frames and used to carry out homology searches. The resulting data provided a wide range of discussion opportunities. The students felt an ownership in the process and the kinds of difficulties, surprises and excitement that comes from doing research. The cost of DNA sequencing purified plasmids is well within the range of a lab budget.

Materials and Methods

cDNA Library
For this laboratory exercise, I chose a Drosophila cDNA library made with mRNA from 6-18 hr developing embryos (Brown and Kafatos, 1988). This library offers three advantages. First, the library is constructed in a plasmid that makes the DNA much simpler to isolate compared to a bacteriophage. Second, the plasmid is constructed in such a way that
the cDNA is directionally cloned and oriented in a specific way (Fig. 1). Third, the completion of the *Drosophila* genome sequence made the sequence analysis more detailed. Aliquots of an amplified stock of the library are available from the author; however, this stock should not be used for research purposes, as it has been amplified from an original stock and does not contain a representative set of cDNAs.

**Figure 1.** The cDNA cloning site of the pNB40 vector is shown. The insertion site is flanked by SP6 and T7 viral promoters. These allow for the use of standard DNA sequencing primers. The cDNA library was constructed in an orientation specific manner. The 5' end of the cDNA was cloned by using G-C tailing and the 3' end was cloned by using the poly-A tail. The EcoRI restriction sites were used to excise the insert for size analysis. The EcoRI restriction site flanking the SP6 promoter is approximately 300 bases from the beginning of the G-C tail.

**Plasmid Isolation**

Each student in the lab picked a bacteria colony from a prepared plate and streaked it out to recover a single isolated colony. The bacteria medium was standard Luria broth supplemented with 100mg/ml ampicillin (Sambrook, et al., 1989). Bacteria containing plasmid were grown in liquid culture overnight at 37 C overnight. The students isolated plasmid DNA using the alkaline lysis or boiling miniprep method according to standard protocols (Sambrook, et al., 1989). Isolated plasmid DNA was digested with EcoRI restriction enzyme. This liberated the cDNA insert along with a small fragment from the plasmid (Fig. 1). The cDNA insert was sized using 1% agarose gels run in 1x TBE (Sambrook, et al., 1989). We chose to prepare three samples for DNA sequence analysis. These plasmids were purified using the ClearCut miniprep plasmid DNA isolation kit (Stratagene Inc., La Jolla, CA ). This insured that the DNA would be of sufficient purity to allow for a good sequencing reaction and was recommended by the sequencing facility (Iowa State University Core Molecular Biology Facility). The DNA was quantified by gel electrophoresis and UV spectrophotometry.

**DNA Sequencing and Analysis**

Our cDNAs were sequenced by the Iowa State University Core Molecular Biology Facility. They use a PCR based cycle sequencing protocol that works effectively on double stranded DNA to easily provide up to 500 bases of quality sequence. By using the SP6 DNA sequencing primer we were able to unambiguously obtain sequence from the 5' end of the cDNA clone. The sequence facility uses the industry standard PCR cycle dideoxy-sequence protocols and the Applied Biosystems fluorescent DNA sequencing system. They post the data to their computer and we recovered them through the Internet as a data file that is read with ABI's Edit View™ program that is supplied free by the sequencing center. The sequencing center also supplies a straight text data file.

The cDNA sequences were analyzed, by the students for open reading frames, and used in homology searches. We used a set of sequence analysis tools on the web called “The Molecular Tool Kit” (http://arbl.cvmbs.colostate.edu/molkit/index.html) to determine open reading frames and to translate the cDNA into a putative protein sequence. The National Center for Biotechnology’s BLAST™ web site (http://www.NCBI.nlm.nih.gov/BLAST) was used to perform homology searches (Altschul et al., 1997).

**Results and Discussion**

Molecular Genetics at Millikin University is an upper division course that is taken by juniors and seniors who have already completed the Genetics course and the Molecular and Cell Biology course. The lab portion of the course covers a variety of techniques including basic nucleic acid manipulations and purifications. The lab experience described here
was carried out over a four-week period. Initially all twelve of the students in the course picked colonies from the cDNA library, isolated plasmid DNA and performed restriction digests. All of the digests were analyzed by gel electrophoresis. After this point, three clones were arbitrarily picked from the class for sequencing. The inserts ranged in size from 800 bases to 1.7 kilobases. The plasmid DNA from the three clones was purified using a spin column kit in order to prepare a DNA template that is optimal for double stranded sequencing. This step is recommended by the sequencing facility.

The reason for this experiment was to provide an unknown cDNA sequence that could be analyzed for open reading frames and used as a query sequence in a BLAST™ homology search. This demonstrated a variety of sequence analysis tools and served as an excellent starting point for discussions on molecular biology.

We recovered and sequenced three different DNA sequences. All of the templates provided quality data and gave at least 300 bases of sequence from the 5' end of the clone. Homology searches were very clear. One clone was a partial cDNA from the squid gene. A second clone was from the ribosomal protein L8 (rpL8) gene, while the third clone most likely represented an artifact of the library construction as the homology indicated that the sequence was the pGEM plasmid.

This lab exercise developed a number of excellent discussion points. The students could readily see the effects of sequence errors on open reading frame analysis. This was particularly apparent with the squid and rpL8 gene, since the complete sequences are in the GenBank™ database and any errors were seen through the homology search. The discovery of pGEM plasmid contamination in the library brought up a number of questions and possibilities for how this artifact could have been created. Most likely, there was plasmid contamination in a reagent during the cDNA synthesis or vector preparation.

In the next few years, the analysis of DNA sequence is going to become one of the most important jobs in molecular biology. This lab experience effectively introduces the students to techniques and skills required for sequence analysis. The students were able to prepare lab reports requiring interpretation of significant amounts of data. The complex, multi-step procedures provided an excellent series of exercises that brought DNA sequence analysis and its associated problems into the classroom.

**Literature Cited**

