Polymerase Chain Reaction Mutagenesis and Automated DNA Sequencing

Melanie L. Sims[1], Victor V. Tryon[1], Barry T. Nall[1], and Christy MacKinnon[2]

[1]University of Texas Health Science Center-San Antonio
[2]University of the Incarnate Word
4301 Broadway San Antonio, TX 78209
(210) 829-3147
mackinno@universe.uiwtx.edu
http://biology@universe.uiwtx.edu

Reprinted From: Sims, M. L., V. V. Tryon, B. T. Nall, and C. MacKinnon. 1999. Polymerase chain reaction mutagenesis and automated DNA sequencing. Pages 356-362, *in* Tested studies for laboratory teaching, Volume 20 (S. J. Karcher, Editor). Proceedings of the 20th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 399 pages.

- Copyright policy: http://www.zoo.utoronto.ca/able/volumes/copyright.htm

Although the laboratory exercises in ABLE proceedings volumes have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibility for risk. The Association for Biology Laboratory Education (ABLE) disclaims any liability with regards to safety in connection with the use of the exercises in its proceedings volumes.

© 1999 Melanie L. Sims, Victor V. Tryon, Barry T. Nall, and Christy MacKinnon

Introduction

This activity combines the Polymerase Chain Reaction, *in vitro* mutagenesis, and automated DNA sequencing. A model experimental system, a pEMBLYe30 plasmid containing the yeast cytochrome c gene, is utilized to analyze the relationship between genetic information and protein structure. This activity engages students in a pre-lab written activity that reviews the genetic code, types of amino acids, and DNA replication. Using the known sequence of the cytochrome c gene, students design an oligonucleotide mutagenic primer. The Megaprimer PCR protocol for *in vitro* mutagenesis is conducted by either the students or by the instructor as a demonstration. Concepts of DNA sequencing are discussed; examples of automated DNA sequencing chromatograms(and other support material)are available from the authors' web site. Additional investigations (suitable for student research projects) involving cloning and transformation of the plasmid vector also are described, as are *in vivo* assays of cytochrome c function, and spectral analysis of cytochrome c structure.

Mutagenesis is a process which introduces mutations, or permanent, heritable changes, in DNA. Mutagenic agents include physical (e.g., large temperature changes, ionizing radiation), chemical (e.g., aflatoxin B1, EMS), and genetic changes (e.g., transposable elements, recombination). Traditionally, mutations have been thought of as "bad," but they can produce valuable exceptions which allow specific genetic processes to be dissected.

Since Mueller's time, geneticists have been able to induce mutations, but these were more or less random. Creating specific mutations at precise locations in a given gene was a goal of geneticists and molecular biologists for some time. In the late 1970's, *in vitro* mutagenesis (or site-directed mutagenesis) was developed by Dr. Michael Smith and colleagues, and it has been widely utilized to study the role of specific amino acids in protein structure and function.

About this same time, Dr. Kary Mullis developed the Polymerase Chain Reaction (PCR). The PCR technique revolutionized genetics and biochemistry laboratories. Using PCR and the correct oligonucleotide primer sets, a single copy of DNA can be amplified into millions of copies. One of the beauties of the PCR protocol is that it is unnecessary to isolate and clone the sequence which is to be amplified, as the replicated strand of DNA in the genome is defined by the two primers used in the reaction.

For PCR, both of the strands of DNA can be used as templates, as long as the oligonucleotide primers are provided for each strand. Primers are chosen which flank the piece of DNA to be amplified (Figure 1). Each newly synthesized strand of DNA will start at the primer and extend beyond the position of that primer on the opposite strand. New primer binding sites are created on each new strand. The reaction mixture is then heated to separate both strands--the original template and the newly synthesized strand. These strands are then available for another cycle of replication.

The original *in vitro* mutagenesis protocols had steps which allowed for selection against the wild type (non-mutant) DNA strands to increase the chance of recovering mutant strands. These steps involved incorporation of uracil into the wild type template, or digestion with selected restriction enzymes (which would eliminate methylated wild type templates).

A novel mutagenesis protocol was developed by Drs. G. Sarkar and S.S. Sommer (Sarkar et al., 1990), using a "megaprimer" and PCR, which greatly increased the yield of mutant DNA strands. The protocol utilizes a pair of oligonucleotide primers (the 5' known primer and the 3' known primer), the internal (mutagenic) primer, and two rounds of PCR. The product of the first PCR reaction is the megaprimer, which contains the desired mutation (resulting from the internal primer), and the 5' known primer amplifies the beginning part of the gene. In order to synthesize the full-length gene, the second PCR reaction uses the megaprimer, and the 3' known primer.

Once the PCR is complete, how can we be sure that the resultant product has the correct DNA sequence? We must be able to analyze each nucleotide to determine the genetic sequence. In 1977, Drs. Allan Maxam and Walter Gilbert pioneered a powerful method of DNA sequence analysis--dideoxy DNA sequencing. Today, automated laser sequencers are used for large scale analysis of DNA. These automated sequencers use nucleotides which are "tagged" with four different types of dyes, or fluorochromes. When these dyes are irradiated by the laser, they emit light at different wavelengths. Each dideoxy reaction mixture is identified by a different dye label, so the mixtures are pooled and run in a single lane of a sequencing gel. As these labeled fragments migrate down the gel and pass in front of the laser beam, the light emitted by the fluorochromes is detected by a photomultiplier. The sequencer instrument thus "reads" the gel in real time during the electrophoresis of the gel. The emission peaks overlap so it is necessary to use a computer to analyze them, and convert the signals into a nucleotide sequence. Typically, a chromatogram of the sequence represents each nucleotide in a different color; the order of the peaks on the chromatogram corresponds to the fragments which are seen on a conventional (manual) sequencing gel.

Pre-Laboratory Activity

Read the following material and **complete the following questions which are in bold.**

The rules of folding linear polypeptide chains into three-dimensional structures are not known. However, we do know the rules for how DNA codes for proteins (collectively known as transcription and translation). The folding information must be contained in the amino acid sequence, and hence the DNA

sequence. An intense area of study in biochemistry and molecular genetics is the determination of the rules for folding, i.e, which amino acids are critical to the folding process of a polypeptide. This information will allow scientists to predict the protein product of any sequence, and will be a critical extension of the Human Genome Project to medical and pharmaceutical applications.

It has been hypothesized that certain amino acids, regard- less of the context of their sequence location, play a critical role in the folding of a polypeptide chain. One of the amino acids identified as influential in folding is proline. Due to its imide ring structure, proline is an unusual amino acid. The proline ring can be isomerized to a cis or trans state. It is thought that isomerization could be a rate-limiting step in the folding of a polypeptide chain (see Veeraragharan et al., 1995, and references therein).

If prolines are involved in the rate-limiting steps of folding, then substitution with a different amino acid (a mutated sequence) could cause changes in the folding rates of the mutated protein. These changes could help illustrate the effects of proline on the *in vivo* folding of the polypeptide chain.

Refer to the sequence of the yeast cytochrome c gene (Figure 2). Find all codons for proline. **Indicate the nucleotide numbers of each codon, and the amino acid number for each.**

Some of these prolines are referred to as "conserved" amino acids. This means that they are found in the genetic sequence for cytochrome c obtained from horses, humans, yeast, mice, and other eukaryotes. We know that in the yeast sequence, prolines at amino acid positions 30 and 71 are among these conserved prolines. These prolines must have essential roles in folding since variations at those positions result in nonfunctional proteins.

To test the essential role of conserved prolines, one could use an *in vitro* mutagenesis protocol and replace a conserved proline with an amino acid with a different side chain structure. **What is the most simple amino acid in terms of structure? Give examples of charged, polar, and nonpolar amino acids. Which type of amino acid is proline?**

Polar amino acid residues are more likely to be on the surface of a protein, and non-polar residues are more likely to be found in the hydrophobic core of a protein. **Where would you expect to find proline residues in a protein--on the surface, or in the hydrophobic core? Why?**

Site-directed mutagenesis can replace one amino acid with another by altering the codon. This can be accomplished by utilizing the DNA replication process, and synthetic DNA primers. DNA polymerase cannot initiate a nucleotide chain--it can only extend one. ***In vivo,* primase produces a product that DNA polymerase extends: what is that product?**

Synthetic DNA primers are known as oligonucleotides. Generally, they are about 18-20 nucleotides long, and can serve as *in vitro* or mutagenic primers. (They can be synthesized artificially at specialty labs and companies for less than $1 per base.) Oligonucleotides can be synthesized so that a codon has been changed from one amino acid to another (site-directed change). **Using the cytochrome c sequence, design a primer which is 18 codons long, and contains one codon change. The 5' end of your primer should start at least 3 codons away from the codon to be changed.** (This will minimize the polymerase "proof -reading" changes to wild type, but will still allow stable base pairing between primer and template). The changed codon should be altered to a different amino acid residue by changing only one nucleotide.

Assume your primer sequence has been synthesized, phosphorylated **(Why?)**, and is in a buffer solution suitable for use in the PCR protocol. **What are the minimum components you must have in a PCR reaction?**

Now, assuming you have conducted an *in vitro* synthesis reaction utilizing the oligonucleotide you just designed, **what percentage of the molecules resulting from an *in vitro* synthesis reaction will contain the wild type codon? What percentage of the molecules will contain the mutant codon?**

Laboratory Protocol for PCR Mutagenesis

The supplied template of yeast cytochrome c DNA was isolated from a pEMBLYe30 vector (Baldari et al,1985) using a Wizard Mini-Prep Kit from Promega. It may be wild type, or it may be a previously mutated sequence. The instructor will assign the codon to be mutated, and you will design a 18-mer mutagenic primer as practiced in the pre-lab activity.

The PCR Megaprimer protocol will be used to conduct the mutagenesis reaction. The reagents needed to perform the PCR are available in kit form. (Those used in this experiment are from the Strategene Pfu kit.) All oligonucleotide primers (mutagenic, 5' known, and 3' known primers) were synthesized by the Center for Advanced DNA Technologies, UTHSC-San Antonio, TX, using the Montgomery et al.(1980) published sequence of iso-2-cytochrome c.

Day One: The first PCR reaction will use the 5' known primer and the mutagenic primer. Using the instructions from the kit, a planning chart (Figure 3B) is constructed to facilitate the manipulations. Complete the planning chart before starting the PCR protocol.

If more than one PCR reaction is prepared, it is helpful to make a PCR "master mix" which contains all the essential chemicals, e.g., water, buffer, nucleotides, and polymerase. The amount of each reagent in the "master mix" is determined by multiplying the number of PCR times the amount for one reaction, plus 1, which compensates for pipetting errors (See Figure 3A for an example). The reagents are added in the order in which they are listed on the planning chart. The polymerase is always added last, just before you are ready to begin the PCR.

Label each PCR-size microfuge tube. Add 90 microliters of the master mix to each tube. Add the remaining components in each tube as per the planning chart. The final volume in each PCR tube will be 100 microliters. Mix and centrifuge each tube briefly. Overlay each reaction mixture with 50 microliters of PCR quality mineral oil. Place the tubes in the thermocycler and run on the preset program (determined by the instructor).

Day Two: Transfer a 10 microliter sample from the PCR to a clean microfuge tube. Add 2 microliters of tracking dye. Load your samples on a 2% agarose gel. Using the 5' known sequence, and the mutagenic primer sequence, predict the size of the amplified product that will appear in each sample lane. Stain and destain the gel as per the instructor's directions. Always wear gloves if ethidium bromide is used as the DNA stain!! It is a chemical mutagen!!

Prepare the second PCR using the same format as Day One (see Figure 3). Label each PCR-size microfuge tube. Add 90 microliters of the master mix to each tube, as well as the components listed on the planning chart. Mix and centrifuge each tube briefly. The final volume of the PCR mixture in each tube is 100 microliters. Overlay each reaction mixture with 50 microliters of PCR-quality mineral oil. Place the tubes in the thermocycler and run on the preset program.

Day Three: Examine the PCR product by gel electrophoresis using the same format as Day Two. Predict the size of the bands which will appear in each lane. Stain and destain as per the instructor's directions.

Extra Lab Days: As determined by your instructor, the PCR product may be sequenced. This can be accomplished manually or by automated sequencing. The PCR product is used as the template in the dideoxy sequencing reaction.) If the correct sequence is obtained (and if the Megaprimer protocol worked, you will obtain a mutant product, or no product at all!), you can plan a cloning protocol to insert the PCR product into a suitable plasmid vector. If expression of the mutant gene is to occur in yeast, then a vector which can express in both *E. coli* and yeast is necessary (pEMBLYe30 is a good choice). Expression in *E. coli* is necessary for the transformation protocol of the cloned gene, and subsequent plasmid isolation. The isolated plasmid then must be transformed into yeast (a more elaborate protocol than that used for *E. coli*) in order for in vivo folding of the cytochrome c protein to occur. *In vivo* function of the folded cytochrome c protein can be assayed using different strains of yeast that can use "fermentable" or "nonfermentable" carbon sources (i.e., carbon compounds which can be used in

glycolsis or enter the cellular respiration pathway after the glycolytic pathway, respectively). The two carbon sources commonly used are lactic acid and glucose.

Research projects may be designed in which the student selects which codon to replace, designs the mutagenic primer, and conducts the PCR protocols, sequencing, transformation, and *in vivo* assays. Simplified cytochrome c isolation protocols can be used to obtain protein samples, which can be analyzed by UV-VIS spectroscopy for the absorption spectrum profile of the protein structure.

Acknowledgments

Michael Pierce supplied the *E. coli* culture containing the pEMBLYe30 plasmid with the cytochrome c gene *Bam H*I fragment (with mutated codons H33N, H39K). Dr. Fred Ernani and Nancy Nicholls assisted with the PCR protocols, gel electrophoresis, and automated DNA sequencing. The research protocols were conducted by M. Sims in partial fulfillment of a MA in Biology.

References

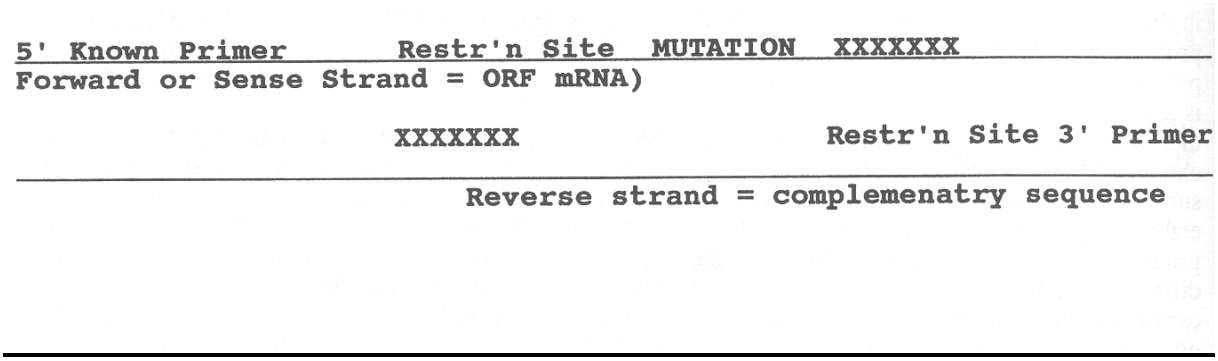
Baldari, C., and Cesareni, G. 1985. Plasmids pEMBLY: New single-stranded shuttle vectors for the recovery and analysis of yeast DNA sequences. Gene, 35:27-32.

Montgomery et al., Proceedings of the National Academy of Science USA., 77(1): 541-545 1980. Isolation and sequence of the gene for iso-2-cytochrome *c* in *Saccharomyces cerevisiae*. [The sequence shown corresponds to the open reading frame of mRNA].

Sarkar, G., and Sommer, S.S. 1990. The "megaprimer" method of site-directed mutagenesis. BioTechniques 8:404-407.

Veeraragharan, S., Rodriguez-Ghidarpour, S., MacKinnon, C., McGee, Wm. A.,Pierce, M. M., and Nall, B. T. 1995. Prolyl isomerase as a probe of stability of slow-folding intermediates. Biochemistry, 34:12892-12902.

Figure 1. Generalized PCR Reaction



Both the 3' Known Primer and the Mutant Oligo Primer will amplifiy the reverse strand; thus their sequences must be complementary to the forward sequence.

Figure 2. Cytochrome c Nucleotide Sequence Nucleotide sequence of the yeast iso-2-cytochrome c (Montgomery et al., 1980).

```
5'

ATG GCT AAA GAA AGT ACG GGA TTC AAA CCA GGC TCT GCA AAA
Met Ala Lys Glu Ser Thr Gly Phe Lys Pro Gly Ser Ala Lys

AAG GGT GCT ACG TTG TTT AAA ACG AGG TGT CAG CAG TGT CAT ACA
Lys Gly Ala Thr Leu Phe Lys Thr Arg Cys Gln Gln Cys His Thr

ATA GAA GAG GGT GGT CCT AAC AAA GTT GGA CCT AAT TTA CAT GGT
Ile Glu Glu Gly Gly Pro Asn Lys Val Gly Pro Asn Leu His Gly

ATT TTT GGT AGA CAT TCA GGT CAG GTA AAG GGT TAT TCT TAC ACA
Ile Phe Gly Arg His Ser Gly Gln Val Lys Gly Tyr Ser Tyr Thr

GAT GCA AAC ATC AAC AAG AAC GTC AAA TGG GAT GAG GAT AGT
Asp Ala Asn Ile Asn Lys Arg Val Lys Trp Asp Glu Asp Ser

ATG TCC GAG TAC TTG ACG AAC CCA AAG AAA TAT ATT CCT GGT ACC
Met Ser Glu Tyr Leu Thr Asn Pro Lys Lys Tyr Ile Pro Gly Thr

AAG ATG GCG TTT GCC GGG TTG AAG AAG GAA AAG GAC AGA AAC
Lys Met Ala Phe Ala Gly Leu Lys Lys Glu Lys Asp Arg Asn

GAT TTA ATT ACT TAT ATG ACA AAG GCT GCC AAA TAG
Asp Leu Ile Thr Tyr Met Thr Lys Ala Ala Lys Ter
```

Figure 3. Planning Charts for PCR

3A. Master Mix Planning Chart

Component	ul x Rxn + 1	Total
Water	62 x (3 + 1)	248
10 X Buffer	10 x (3 + 1)	40
ATP	4 x (3 + 1)	16
CTP	4 x (3 + 1)	16
GTP	4 x (3 + 1)	16
TTP	4 x (3 + 1)	16
Pfu Polymerase	2 x (3 + 1)	8

Figure 3B. Planning Chart First PCR Reaction

Component	Positive Control	Negative Control	Template + Primers
Water	0 ul	10.0 ul	4.0 ul
5' Known Primer	0 ul	0 ul	2.0 ul
Mutagenic Primer	0 ul	0 ul	2.0 ul
Template	0 ul	2.0 ul	2.0 ul
Control 5' Primer	5.0 ul	0 ul	0 ul
Control 3' Primer	5.0 ul	0 ul	0 ul
Control Template	2.0 ul	0 ul	0 ul
Master Mix	90 ul	90 ul	90 ul