**PCR and DNA Sequencing**

By: Aaron Coffey

**Abstract**

The objective of this experiment was to investigate whether pRSETB contained the GFP or RFP gene. A PCR followed by gel electrophoresis was performed alongside DNA sequencing and translation of the plasmid. Both tests concluded that pRSETB coded for GFP and not RFP.

**Introduction**

Four major techniques were used during this experiment to achieve the objectives and purify the product. These methods are PCR, gel electrophoresis, DNA sequencing, and PCR cleanup. All tests and inquiries were performed with regard to the nature of the plasmid pRSETB.

The plasmid was entered in a polymerase chain reaction, commonly shortened to PCR. PCR is widely used to exponentially amplify a sequence of nucleic acids *in vitro*, and it is also exploited in order to modify or amend the amplified sequence.\(^1\) This method originated from the work of Kary Mullis at Cetus Corporation in the early 1980s, and this account even goes as far as saying, “in the spring of 1983, Kary Mullis had the original idea for PCR while cruising in a Honda Civic on Highway 128 from San Francisco to Mendocino.”\(^2\)

Once the desired nucleotide sequence, its corresponding primers, individual nucleotides, and the polymerase enzymes have been added, PCR begins with denaturation. Denaturation, also termed “melting”, refers to the separation of the two strands in DNA held together by hydrogen bonds; this is accomplished by bringing the mixture up to an adequately high temperature. For DNA sequences that consist of more C-G pairs than A-T pairs, the melting
temperature is higher because there are three hydrogen bonds between cytosine and guanine and only two hydrogen bonds between adenine and thymine.

Following denaturation, the system is cooled to allow the primers to bind to the melted DNA. This step is called annealing. The cooling process is very delicate because if the DNA temperature is too low, the double-stranded structure will reform.

In this experiment the concoction was heated again to a temperature between the denaturing and annealing to maintain the denaturation of the DNA. For most polymerases temperatures this high would render them dysfunctional, but the enzyme Taq polymerase (now used prominently for PCR) functions optimally at 80°C.³ Taq polymerase is a DNA polymerase named after the extremophile *Thermus aquaticus* from which it was purified.³ This step was dubbed extension because it is when Taq polymerase elongates the primers by adding nucleotides one by one, effectually doubling the original quantity of the sequence.

These three steps (denaturation, annealing, and extension) are repeated several times, producing the exponential amplification Mullis reported.

Some of the product from the PCR was run through a gel via electrophoresis. “The basic concept of this process is relatively simple but still quite ingenious: ‘At pH near neutrality, DNA is negatively charged and migrates from cathode to anode [in an electric field] with a mobility dependent primarily on fragment size.”⁴ As a result, all pieces of DNA with the same length form a line in their lane of the gel. Often in the last column of the gel, a ‘ladder’ or scale sample is run alongside the subject samples; this ladder sample includes known premade fragments of DNA which can be used to estimate the other sample segment lengths by comparison. Before
running current through the gel, a running buffer is poured into the apparatus so that current will travel evenly throughout the gel and keep the DNA on-track.”

The pRSETB plasmid was sequenced using a 3730 DNA Analyzer. This process also involves PCR, but the new DNA strands are synthesized with fluorescently labeled nucleotides. The fragmented results of this reaction are drawn through capillary tubes by means of gel electrophoresis. An argon ion-laser beams perpendicularly through the tube and is detected by four optical sensors, each first filtered for light wavelengths of either 540nm, 560nm, 580nm, or 610nm. These are the wavelengths at which the labeled nucleotides fluoresce, thus indicating which base is passing at a given instant.

The remaining PCR product was cleaned with the Promega Wizard® SV Gel and PCR Clean-Up System so that further experiments with the pRSETB plasmid could be conducted. The kit, “is based on the ability of DNA to bind to silica membranes in the presence of chaotropic salts.” This system uses a binding solution to retain DNA in a column while impurities are removed with a washing solution. The desired DNA is then reclaimed by simply running water through the column.

PCR, gel electrophoresis, and DNA sequencing together made it possible to distinguish which gene was on pRSETB and discover said gene’s amino acid sequence.

**Methods**

Two samples were put in PCR. The first included 10µl sterile water, 5µl plasmid pRSETB (2ng/µl), 5µl each of forward GFP primer (5’ CGACTCACTATAGGG) and reverse GFP primer (5’ CCCAGCAGCTGTTACAAACTC) (40ng/µl ea.), and 25µl 2X PCR buffer (20mM Tris-HCl, pH 8.3, 100mM KCl, 3mM MgCl₂, 0.002% gelatin, 0.4mM dATP, 0.4mM
dCTP, 0.4mM dGTP, 0.4mM TTP, 0.06 units/microliter Taq DNA polymerase, inert red dye). The second sample was the same as the first with the exception of forward and reverse RFP primers, (5’ GCTTTGTGAGCGGATAAC) and (5’ CTCGGTTCTTTTCATACTGCTC), instead of GFP primers. Alan Conrad ran the samples in a thermal cycler with the following program: 94°C 5min., 25 cycles of 94°C 15sec. – 55°C 30sec. – 72°C 30sec., 72°C 5min.

Alan Conrad constructed two agarose gel (1% agarose in 0.5X TBE) electrophoresis apparatuses, and poured 0.5X TBE running buffer (44.5mM Tris base, 44.5mM boric acid, 1.0mM EDTA) into the main gel boxes.

Following the PCR, 10µl of each sample were loaded into their corresponding gels, and 10µl of a Promega 1kb DNA ladder (50ng/µl in 0.006% xylene cyanol FF, 0.006% bromophenol blue, 0.06% orange G, 2.5% Ficoll 400, 10mM Tris-HCl, pH 7.9, 10mM EDTA) were loaded into the last lane on each gel. The gel was electrophoresed at 100V for 1 hour then photographed with a UV shadowing device.

Unfortunately, the remaining 40µl of PCR product was prematurely washed instead of first being bound to the column. Had this mistake not been made, the 40µl of PCR product and 40µl Membrane Binding Solution would have been placed into Wizard® SV Minicolumns inside of microcentrifuge tubes and centrifuged at 13,000rpm for 1 min. The column then would have been centrifuged at 13,000rpm for 1min again after 700µl Membrane Wash Solution was added and again with only 500µl Membrane Wash Solution for 5 min. Finally the DNA would have been gathered with 50µl nuclease-free water added to the column and centrifuged at 13,000rpm for 1 min.
During the PCR cycles, a sample of 1µl pRSETB (250ng/µl), 2µl forward GFP primer, and 3µl sterile water was procured. This sample was taken to a DNA sequencing lab, in which the target GFP gene was sequenced using a 3730 DNA Analyzer.

**Results**

**GFP Test Gel**

![GFP Test Gel](image)

*Figure 1: Gel of the PCR results using GFP primers. The band in lane 1 indicates that pRSETB has the GFP gene.*

**RFP Test Gel**

![RFP Test Gel](image)

*Figure 2: Gel of the PCR results using RFP primers. The lack of a band in lane 1 indicates that pRSETB does not have the RFP gene.*

Above are the UV shadowing images of the resulting electrophoresis gels ran with the PCR products. These were used to determine whether pRSETB had the GFP or RFP gene.
pRSETB DNA Sequencing Feedback

The start codon is highlighted in blue, the His tag is highlighted in yellow, the reverse primer is highlighted in purple, and an estimated restriction cite for BamHI is highlighted in green.

Figure 3: Sequencing feedback from the 3730 DNA Analyzer for pRSETB with GFP primers. The start codon is highlighted in blue, the His tag is highlighted in yellow, the reverse primer is highlighted in purple, and an estimated restriction cite for BamHI is highlighted in green.

Translated Feedback of pRSETB DNA Sequencing

The total number of amino acids in this sequence (251) was also given with this translation and used to estimate the length of the GFP gene in base pairs (753b).
The melting temperatures of the four PCR primers can be calculated as follows:

**Equation 1:**

\[ T_m = 4^\circ C \times \left( \text{total \# \ of \ G \ and \ C} \right) + 2^\circ C \times \left( \text{total \# \ of \ A \ and \ T} \right) \]

\[ T_{m,GF} = (4^\circ C \times 6) + (2^\circ C \times 7) = 46^\circ C \]

\[ T_{m,GR} = (4^\circ C \times 11) + (2^\circ C \times 10) = 64^\circ C \]

\[ T_{m,RF} = (4^\circ C \times 9) + (2^\circ C \times 8) = 52^\circ C \]

\[ T_{m,RR} = (4^\circ C \times 16) + (2^\circ C \times 11) = 62^\circ C \]

\[ T_{m,avg} = \frac{T_{m,GF} + T_{m,GR} + T_{m,RF} + T_{m,RR}}{4} = 56^\circ C \]
Discussion

The annealing temperature was set about equal to the average of the primer melting temperatures, calculated in Equation 1 above. If the annealing temperature had been much higher, the primers would not have been able to bind to the plasmid because the hydrogen bonds could not be formed, and elongation would have been scarce if at all present. If the annealing temperature had been much lower than this, the denatured DNA would have rebound to itself and the primers would not have sites at which to bind, again hindering replication by Taq polymerase.

Plasmid A, found earlier to be pRSETB, was also found in this experiment to contain the gene which codes for green fluorescent protein or GFP.

The first indication that pRSTEB held DNA for GFP lies in Figure 1. The primary purpose of the gels in Figures 1 and 2 was to show whether the GFP or RFP primers would facilitate a successful PCR. The sample in which Taq polymerase was able to replicate the
plasmid would render DNA in quantities several orders of magnitude greater than that of the other sample. Although the band in lane 1 of Figure 1 is not the very bright, its existence, in contrast to the lack of a band in lane 1 of Figure 2, is enough to signify that pRSETB contains the GFP gene and not the RFP gene. However, this was all that could be concluded from the two gels because there was an apparent malfunction with the DNA ladder. With a working ladder, the size of the GFP gene could have also been estimated. Fortunately, not all of the eggs were put in the gel electrophoresis basket.

Figures 3 and 4 illustrate the results of the 3730 DNA Analyzer. The nucleotide sequence in Figure 3 was submitted to the ExPASy DNA Translator program⁹, which conveniently deciphered the series into its corresponding amino acid chain – Figure 4. In observing Figure 4, the Histidine tag (six consecutive histidine amino acids) and the correct translation frame were evident. This feature of the gene is highlighted in yellow in both Figure 3 and 4. Another conspicuous element (thanks to ExPASy⁹) of the sequences was the start codon – A-T-G or methionine as an amino acid. The start codon appears 9 nucleotides or 3 amino acids before the His tag and is highlighted blue in Figures 3 and 4. Two of the more elusive traits of the DNA sequence were the forward and reverse primer locations. The reverse primer setting is highlighted purple in Figure 3, but the forward primer locale could not be found. It is believed that this spot is contained within the numerous “N” residues at the start of the sequence. These “N”s represent the insufficient confidence the DNA analyzer had in indentifying a specific nucleotide at these points.

This condition is also believed to account for the interpretation of a singular “C” residue as being a “T”. The possible miscalculation is thought to have occurred towards the end of the
sequence (where such events are common) following the residues, “G-G-A-T-C-“. In the absence of other more likely candidates, this site had the best case for being one of the three known restriction enzymes’ sites. Figure 6 elaborates this belief: the confidence level of the analyzer in the “T” it had assigned, following the “G-G-A-T-C-“ set in the yellow box, is half its potential value. This possibility implicates this region as a BamHI restriction site, placing the GFP gene just before one of the two BamHI sites as demonstrated by Figure 5.

There was also a much larger incongruity between the translated DNA sequence and the known GFP amino acid sequence. It appears that 24 of the amino acids in the known sequence were not accounted for in the translation. What was even more baffling was that the missing residues belonged between residues that did match up with the known data. The only conceivable hypothesis is that this segment was lost at some point during the sequencing process.

Other than the missing fragment, the translation and known amino acid sequences are identical, proving DNA sequencing to be a more conclusive method in indentifying which gene was on pRSETB.

In light of the discovery that pRSETB had the GFP gene, it is plausible that when Eschericia coli bacteria are transformed by this plasmid, their colonies will be of a green hue.
References


