Thomas Richards and Mark Gustincic

June 7, 2016

BME:160

PCR into Restriction Digest Product Prediction Tool

**Introduction:**

The discovery of the polymerase chain reaction (PCR) evolved the study of experimental molecular biology. The reaction allows for the isolation and production of enormous amounts of a DNA sequence and is widely used by researchers to identify, sequence and clone nucleotide sequences. PCR utilizes DNA’s ability to reform complementary base pairs after denaturation to amplify a specific subsequence within a template sequence (1). Primers are designed to be complementary to sequences found in each DNA strand. During PCR, each primer anneals to its complementary DNA strand and DNA polymerase adds complementary bases to the new strand, replicating the DNA. Along with PCR, sequence fragmentation by use of restriction enzymes is a widely utilized molecular biology technique. Restriction enzymes cleave DNA sequences by recognizing and binding specific palindromic sequences within DNA. Once bound, the enzymes hydrolyze a phosphodiester bond in the DNA backbone, fragmenting the DNA (2). By using sequence data, researchers use these restriction enzymes to create DNA fragments that can be ligated to other DNA sequences or separated and analyzed by size using gel electrophoresis.

We designed a program using the most recent version of Python (Python 3.5.1) to be a tool that predicts the products and properties of PCR and restriction digest experiments. This program is meant to be run along an actual PCR or restriction digest experiment to help verify the accuracy of the results.

**Materials and Methods:**

The primary concept utilized in designing this program was calculating the sequence to be amplified by PCR. First, the program determines if and where each primer is present in the sequence or its complement strand. Next, it determines if the position of each primer facilitates amplification of the desired sequence by checking the position of each primer. If each position is satisfactory, DNA polymerase will replicate the sequence, producing two copies from the original sequence. The data for this is gathered in the method getInfo() and is handled by the method pcrProd() to determine the sequence of a PCR amplicon.

The program’s main() method handles the input for the program. Input can either be attained from a fasta file or input directly into the command line. Main() utilizes the FastAreader class to handle input from fasta files by reading through and gathering DNA sequences from a fasta file. The program parses though each genome present and prompts the user to perform PCR on each sequence individually.

In addition to modelling PCR, the method REfinder() in the program models restriction enzyme digestion and provides a restriction map of the amplified sequence containing labeled restriction enzyme cut sites, denoting their position in the sequence. This method utilizes a dictionary of restriction enzyme names and their specific cut sequences to locate the position of each cut site present in PCR product. Location of cut sites allows the program to output the fragments produced and their lengths for any and all restriction enzyme cut sites found in sequence.

Lastly, the methods protocol() and gelProtocol() are implemented as a series of dynamic print statements detailing the procedural steps involved in running PCR and gel electrophoresis on DNA. Each method takes input from the user to calculate specific volumes relative to the desired reaction volume of the user.

**Results:**

The program’s results are all printed to standard output in the terminal window. After prompting the user for the input sequence file/raw sequence and the primer sequences, the program asks the user whether the primers sequences are from the same strand. This is very important because it dictates what the program looks for when searching for the second primer in sequence. If the primer were given as a sequence on the same strand, then the actual target of that primer is its reversed complement sequence. If it were given as a sequence from the reverse strand, then only the reversed sequence is used. The first numerical data is outputted by the program as the primer lengths in base pairs, weights in kilo Daltons and GC contents of each primer.

If an amplicon is found within the template sequence, the program prompts the user if they want to print the full amplicon sequence or just its length, weight and GC concentration. Once the amplicon attributes are outputted, the program asks the user if they to see the restriction enzyme cut sites present in the amplicon. If the user returns yes, the program outputs the amplicon sequence containing all the restriction enzyme cut sites present in the sequence denoted by a “|” at the cut site. This is followed by the name of the restriction enzyme. In addition, the names of the all restriction enzymes present in sequence as well as the number of times the cut sequence is found in the sequence. The program asks the user for specific restriction enzymes to cut with and outputs the expected lengths of DNA fragments that would result from digestion with these specific restriction enzymes.

The last element of the program is a dynamic protocol for running PCR and gel electrophoresis using the inputted primers and template sequence. It prompts the user for the desired reaction volume and prints out a step by step procedure that includes the volumes and reagents needed for PCR and gel electrophoresis.

**Discussion:**

The most important goal for this program is to accurately predict the PCR products with a given sequence and primers and extended that to predict the results of a restriction digest performed on the PCR product. This was achieved by simply determining the position in the template sequence that each primer would anneal to and outputting the sequence contained between those annealing sites, assuming the expected PCR product is produced. The product is scanned for potential restriction enzyme cut sites using a dictionary containing restriction enzymes and the target sequence and the position of each cut site is stored.

From the user’s standpoint, this program automates the tedious calculations and analysis involved in PCR and restriction digest experiments to accurately predict their outcomes when performed. This program’s intended application is to be a tool to be utilized along side the experiments it models to assist the processing of sequence data attained from the experiment and verify product accuracy. The step by step procedure is designed to further assist the user in performing experiments by providing a dynamic set of instructions based on the reaction specifications determined by the user.

Biopython and a number of other packages could be utilized to improve this program by expanding the library of restriction enzymes present. ­­Currently, we are working to incorporate Biopython’s Restriction module to gain access to a more comprehensive data base of restriction enzymes in order to better identify restriction enzyme cut sites in any given sequence within our program. Additionally, we have been working on a method that determines sites on the template sequence with near complimentary sequences up to a certain threshold that could result in unexpected annealing and PCR products. The primer sequence is allowed a certain threshold of non conserved bases especially on its 5’ end but requires high specificity at the 3’ end to anneal the template strand and facilitate polymerase enzyme binding for amplification. Additionally, this can allow restriction cut sites present in the primers to be inserted into the amplicon sequence allowing it to easily be inserted into any genome.

**Acknowledgements:**

A huge thank you to Dr. Bernick for supplying us with the FastaReader class and to Robert Shelansky for making himself available when ever either of us had any questions.

**References:**

1. Garibyan L, Avashia N. Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *The Journal of investigative dermatology*. 2013;133(3):e6. doi:10.1038/jid.2013.1.

# Pingoud A, Jeltsch A. Structure and function of type II restriction endonucleases. Nucl. Acids Res. (2001) 29 (18):3705-3727. doi: 10.1093/nar/29.18.3705