Kimmie McGowen

Bio 131

IP: Sequencing Antibiotics

10 May 2016

Final Project Summary

a. Summarize the motivation of this project.

This project is based off of chapter 4 from the class textbook, which seeks to answer the question of how to sequence a peptide when it does not depend on a ribosome, i.e. when the peptide goes against the central dogma? I had never heard of non-ribosomal peptides before in the 6 years I have been taking biology courses, and it goes against the central dogma of biology, so that caught my interest and I decided to take on the question asked in chapter 4 for my independent project. Because the peptide is not translated from mRNA, traditional computational programs cannot determine the sequence of non-ribosomal peptides. This means that another approach must be used, which is what I developed for this project.

b. Describe any data formatting you did (where/how you got the data).

The question I asked in this project did not lend itself to analyzing a genome or anything requiring input data files. This is because a peptide is input into the linear and cyclic spectrum producing functions (to act as theoretical GC) and then the output is input into the sequencing function, which should output the same peptide you put into the GC mimicking function. Because of this input/output loop, peptides can be made up for testing this program. The test peptide I used came from chapter 4 of the book, which is used as an example.

c. Describe, at a high level, the steps of the program.

I answered the question by taking the end product peptide (where the amino acid sequence is unknown), theoretically using gas chromatography to break the peptide up into individual amino acids and amino acid clusters and determine the masses of each. The spectrum

from GC, along with the known masses of each amino acid, is used to synthesize the amino acid sequence of the peptide.

d. Describe your results and some discussion/interpretation/conclusions.

i. What did you discover?

Because this project isn't analyzing a data set to find an oriC or binding site motifs, etc., nothing was really discovered by creating a program that answers the target question.

ii. Were there any assumptions you made to get the results you did?

In order to create the GC mimicking function, the peptide was initially assumed to be linear, but the function was then modified to accommodate cyclic peptides. Because there are two amino acid pairs that share the same masses (I/L and K/Q), the program was written and implemented with I/L and K/Q being interchangeable. The way the program is now, when NQEL is input the output is NIEQ. If we assume that the difference between I/L is interchangeable and don't care which direction the peptide sequence reads in, the answer is assumed to match.

iii. How could you generalize your method to answer larger/broader questions?

While the book frames this question and program as being used for sequencing antibiotics, it can be used as it is written for any non-ribosomal peptides. These are not common so the program isn't particularly useful outside of this context, but it could also be used as a different way of sequencing any peptide if the linear versions of each function are used. This would only be useful if the peptide's GC spectrum is known and the sequence is not, which is not a likely scenario and other computational programs are more straightforward for a linear ribosomal peptide, but it is another way that this program can be applied.