

Tayla Isensee

Dr. Anna Ritz

Introduction to Computational Biology

7 May 2018

## Final Project Report

### **Motivation**

In this project, I wanted to implement a program that sequences cyclic peptides like antibiotics. Because of their unique structure and their absence from the coding genome of the organisms that synthesize them, sequencing cyclic peptides requires the use of mass spectrometry rather than analysis of DNA. This alternative sequencing technique really intrigued me, as did the biological importance of knowing the structure and sequence of antibiotics. Using mass spectrometry to sequence antibiotics was a chapter in our textbook, which also drew me to this project as the textbook and its companion website provided explanations, datasets, and pseudocode that helped further this project.

### **Data Formatting**

The datasets that I obtained came either from the textbook, *Bioinformatics Algorithms*, or from its companion website Rosalind. Most of the larger datasets came from Rosalind, while smaller and toy examples came from the textbook. The only manipulation of datasets was turning sets into lists so that they fit well into my program.

### **High-Level Steps**

My program can: generate both theoretical linear & cyclic spectrums of a given peptide and provide a possible mass sequence for a cyclic peptide, given ideal experimental data.

## I. Theoretical Spectra

- a. Linear Spectrum: takes peptide string and outputs a list of all the possible masses of the different slices of the peptide.
- b. Cyclospectrum: same thing as linear except the mass slices corresponding to the connected end and start are also added to the output list.

## II. Sequencing a Cyclopeptide

- a. Given an ideal experimental mass spectrum of a cyclic peptide, output its possible peptide sequence (converted also into a sequence of the component amino acids).
  - i. While no peptide yet matches the ideal cyclospectrum:
    - 1. E

bound algorithm structure, could be used to sequence other peptides and sequence-able elements of an organism that are not found in the genome.