BioMusic: Making Proteins Sing

Haylee Ham **Chelsea Ernhofer**

Abstract

1 Introduction

For this project, we explore protein sonification- that is, the translation of protein sequences to music. To achieve this goal, we investigate sonification algorithms and well as music theory in order to produce pieces of music that are both effective representations of sequences and melodiously pleasing. We then create a web application to demonstrate the process and share the generated music with users. We employ Javascript libraries to produce auditory and written music in browser.

2 Background

In the context of bioinformatics, sonification refers to the attempt to capture the information content of DNA and protein sequences auditorily. Although sequences are traditionally analyzed visually, using auditory methods to inspect DNA and proteins have benefits. Rather than learning about individual elements of sequences, sonification is useful when determining patterns and potential mutations across the entire DNA or protein strand. Sonification can also be used to aid visually impaired individuals, allowing them to hear the sequence when they are not able to see it.

There are six main algorithms currently used in the DNA sonification process. These algorithms are based on biological features and rules, such as the distinction between introns and exons in a DNA sequence. This ensures that music produced through sonification contains applicable information about the structure of the sequence in question. Most DNA sonification algorithms analyze nucleotides in groups of three to simulate the structure of codons; these algorithms generate reading frames and some even produce music with interlaced melodies to represent the three possible reading frames present in a DNA strand. Simpler algorithms use hash tables to map each of the four nucleotide bases to one of four musical notes or pairs of nucleotides to 16 possible musical notes.

3 Methods

4 Results

5 Discussion

6 References