Jade A. Minzlaff
BIOEN 537 Computational Systems Biology
November 13, 2023
Package Name: simulate_xna_signal.py

Component Specification

Required packages:

Pandas is used to read data from a .csv file containing a library of possible 4mers and their associated experimental nanopore signal into Python.

NumPy is used for numerical operations involved in signal processing and in the generation of synthetic noise. This uses data imported by Pandas.

MatPlotLib is used to generate the step function plot which is output by the package, dependent on numerical operations and data accomplished by Pandas and NumPy.

The required **user input** is a genetic sequence given as a string with no spaces or other symbols of the bases from the Artificially Expanded Genetic System (AEGIS): A,T,G,C, J, K, P, Sc, Sn, V, X, and Z.

The **generated output** is a step function of the simulated nanopore signal expected for the given user input, based on the experimental data from the imported library. The units of the y-axis of the plot are picoamperes (pA). The units of the x-axis are base position, where each line of the step function corresponds to a nucleotide in the sequence, in order of input.

Use case example: A user suspects that a mutation occurred in an oligo containing synthetic nucleotides and is seeking a visual reference with which to compare an experimental nanopore plot with the intended "correct" genetic sequence. They input the sequence to this package, which then uses the packages described above to generate the expected nanopore plot for the sequence, allowing the user to determine if their plot is trustworthy or if further investigation is required.