Group 2: CASP

Dedeepya Gudipati, Miles Keppler, Sam Carpenter

Proof of Concept

We’ve received the data files of five different sets of data from Dr. Jagodzinski. There are five sets of the data points, each with 20 files of the same protein. Each file has a different set coordinates for each atom. Although there is other information also included in the file, our approach will focus on just the coordinates of the amino acid sequence.The data sets will be visualized in Pymol and aligned to each other by rotating and the protein, so we have a general visual of the protein and how they compare to each other. Since each file has one protein sequence ( in total we have five protein datasets), we will put the file into BLAST and see which proteins best aligns to the query sequence. Since the structure of the protein determines its ability to We will pick a proteins above a specific threshold similarity (80%) and research the amino acids around the active site and how these help the function of the protein. The next step is to compare the selected sequences found in the homolog protein from the BLAST search to the same sequences in the proteins structures in the data sets.

The output we are expecting to get is a newly aligned protein that we will be able to perform calculations on. These calculations will show us the distance between the active sites of these new proteins and the ones we have gathered from BLAST. By comparing these we will see which of these proteins has the most similarities to already known proteins.