Preliminary Results.

After doing the preliminary analysis of our data, we developed a script that would read the amino acids and translate them into a single letter sequence. Then, we used BLAST to find homology proteins to our generated sequences to evaluate the 3D structure of our protein based on the structure of our query sequences. To get a precise account of the structure based on the sequence of the amino acid, we are going to use three different structures that score best in three different areas of the blast search, Max score, Query Length, and Identity (Figure 1). Using PyMol, we have measure the distance between the identical amino acid in the same sequence. Only amino acids in that are aligned to the query sequence will be used in the measurement. The amino acids were picked based on their importance in one of the selected protein structures, this information was gained through literature. This method replaces our previous method because we were not able to find a sequence of amino acids in literature at the active sites of the proteins and some of the proteins do not have active sites. We also picked the amino acids based on their presence in match with the query sequence of all three selected protein structures. Figure 2 is an example of three coordinate sets (serverxx\_TS1) with 3X1E’s amino acids. These scores will be put through another script that will output the scores to each of the 20 coordinate files based on how well the distances of the selected amino acid will match those of the selected pdb files (example in Figure 3). Future steps include ranking the data coordinate files based on how well they match the distances of the pdb files. We will then cross checking this method with one known structure of the peptide to see how effective it is.

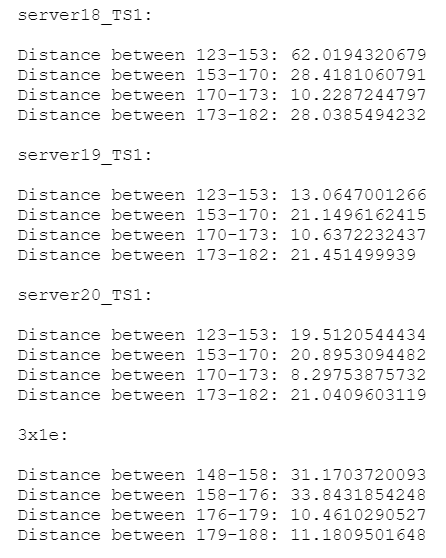


Figure 2: distances between amino acids

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Figure 1: PDB ID of Structures Selected to Investigate for Query Sequence 3D Structure | | | | |
| Sequence ID | Max Score | Query Length | Identity | Amino Acid |
| T0949 | 3X1E | 4HPO | 1SQB | Glutamic acid (GLU, E) |
| T0950 | 3O53 | 6EK8 | 4D82 | Glutamic acid (GLU, E) |
| T0951 | 5CBK | 3WO6 | 5DNU | Lysine (LYS, K) |
| T0953s1 | 2VCY | 2GMQ | 4EBB | Glycine (GLY, G) |
| T0953s2 | 3EEH | 6CN1 | 3JSA | Glycine (GLY, G) |