Abstract

Protein folding is the process by which a protein assumes its characteristic functional shape or native state. The primary structure of a protein consist of a chain of amino acids. Most proteins can carry out their biological functions only when folding has been completed, because three-dimensional shape of the proteins in the native state is critical to their function[3]. By further understanding the relationships between inter-residual contacts we can get a better understanding for the process of initial protein folding.

Methods

Starting with a 43 protein data set of proteins know to fold kinetically in a two state manner[1] (These proteins were chosen because of the proximity between the N and C terminal secondary structures). We began by calculating the geometric center of every amino acid on every protein. Geometric centers were separated into backbones (consisting of carbon, nitrogen, oxygen, and alpha carbon) and side chains consisting of everything else. This gave us a consensus single point for backbones ad another for side chains. Parsing is a process in which an input text is broken down into meaningful tokens and then stored in a data structure. In our case our input was pdb files and we extracted information such as location of helices and amino acid coordinates.

By using the distance formula we calculated distances between amino acid pairs that were contained inside helices. Those pairs that were within 6.5 angstroms of each other were plotted on to figure 2. Furthermore we wanted to know how prevalent each of the amino acids were in our data set. The distribution of amino acids is graphed onto figure 1.

Conclusion

- Further testing will be done on larger data sets and statistical and probability analysis should help us determine inter-residue contact energies. As expected amino acids that were more common in our data set tended to show up more frequently in figure 2.

References


2. Sanzo Miyazawa1 and Robert L. Jernigan (1996) “Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading”

3. Wikipidia http://en.wikipedia.org/wiki/Protein_folding