

Protein Fold Recognition Using Graph Theoretic Measures

Holly Johnsen | Advisors: Claus Wilke/Alpan Raval
Keck Graduate Institute, REU Program 2005



Protein structure prediction is one of the greatest challenges to biologists in the post-genomic era. Over 98% of the currently sequenced proteins have an unknown structure. However, only a limited number of 3D folds have actually been observed. Related proteins often adopt similar folds, and even non-homologous proteins often share similar structures. Thus, known protein structures can serve as templates for protein sequences of unknown structure. Prior studies have matched protein sequences to templates by comparing environmental properties.

We attempt to provide a new method of matching a protein sequence to a template through the use of graph theoretical properties. By representing a protein as a contact map, a representation of the contacts between residues in a protein, we can compute graph theoretical properties for each residue such as degree, closeness centrality, betweenness centrality, and clustering coefficient. We expect that these properties will allow discrimination between classes amino acids, which will be useful for aligning a protein sequence with a template structure.

In this project we have created the tools necessary to calculate these graph theoretic properties for template structures. We have conducted an exhaustive study of the effects of parameters such as distance threshold and minimum residue separation on the contact map and the resulting graph properties. We found the parameters which optimize discrimination between the amino acids for each graph property. We found that these properties correlated heavily with hydrophobicity scales. Since the centrality measures are global properties, we would not expect this. We believe that by including the artificial value of zero for centrality measures, we have skewed our results. Future work will attempt to achieve discrimination between amino acids that is not related to currently understood scales such as hydrophobicity.

