University of Utah
UNDERGRADUATE RESEARCH JOURNAL

STRUCTURAL PROFILING OF MICROPROTEINS: AIDING IN THE CLASSIFICATION AND IDENTIFICATION

Victoria C. Glomb

MicroProteins are small, single-domain proteins, usually less than 100 amino acids in length. Previous research has tried to classify the differences between a small protein and a MicroProtein. However, there are varied interpretations as to how micro-proteins should be identified leading to a poor definition of exactly what constitutes a MicroProtein. The goal of this research is to find if there are structural features that are characteristic of MicroProteins. With the definition of MicroProteins being unclear creating an exclusion and inclusion criteria was important. All the MicroProteins that were selected came from previous research papers about MicroProteins. Some of the identified MicroProteins were not of human origin. However, if found, a human homolog was used. The set of proteins that we selected were mostly human proteins, 3 non-human organisms. The amino acids of the pre-selected MicroProteins were extracted from SWISS/UNI PROT repository. These sequences were used as input to QUARK, in order to predict 3-D models of each protein. From the predicted tertiary structures, DSSP and bmi scripts were used to predict the secondary structures of these proteins. DSSP works on creating a secondary structure assignment based off of the location of the hydrogen bonds in the pdb. files collected from QUARK. The secondary assignments were used to determine the prevalence of alpha helixes, beta-sheets, and coils. We used BLAST used to ascertained that the proteins were of a single domain. Another program called STRING illustrated different gene connections of each MicroProtein. Although, due to the lack of literature on these proteins links were not found on all 48 proteins. RaptorX was used to find information on binding partners and structural properties. Comparisons of structural disorder, the types of solute binding, turns, coils, and helix percentages, gene connections and functions were used to create the structural profiles of the MicroProteins. In future work, STRING protein connections will be examined in SWISS/UNI PROT and related GO terms will be extracted for cluster analysis. Machine learning techniques also plan to be used to classify the rest of the protein features.