1. Ubuntu Programming to Simulate and Visualize Protein Structures

2. Kayla Nokeo and Dr. John M. Finke

3. **Abstract**

Proteins are folded via a process in which a polypeptide chain folds numerous times to produce a biologically active protein which, will present as a 3D structure. 3D structure proteins will be programmed using Ubuntu Virtual Box and put through a simulation to assess these behaviors and protein structure. This can aid in understanding biomolecule functions at the molecular level. This approach can be utilized to study structures from any organism. Further assessments and understanding of this can lead to advancements in medicine, giving the opportunity to help treat patients with a new approach. In this study, Chrymotrypsin Inhibitor 2 (CI2) was generated via programming in order to asses the degree in which simulation factors capture experimental values determined from protein folding. Two dominant conformational ensembles were prevalent and existed near the folding temperature, a folded and unfolded ensemble. These two states correlates with the transition state, a highly unstable conformation set with a structure between the folded and unfolded ensembles. To find the temperature that maximally populates the transition state, simulations were performed at varying temperatures. The optimal folding temperatures was indicated by an equal population of folded and unfolded states, there was frequent back and forth flipping between these states. Visualization software (VMD) was used to serve and aid in the observation of actual trajectory of CI2 as it flipped back and forth between the forementioned states, data was obtained to compare structural features of all three conformations (folded, unfolded, and optimal where these were equal). The data collected was separated into two graphs, displaying what amino acid residues were found to be in close proximity in each ensemble structurally. More contacts that were present were found to be more of the “native” protein structure in which it exists as. However, when this was compared with assessments of the transition state structure, the generated structure gave a +0.25 slope and +0.28 correlation coefficient. The results provided relative predicative value, however there is still room for improvement. This method can be applied to a plethora of different proteins found within organisms in order to predict their behavior under various conditions.