Background (3Qs)

1. Why are we answering this? What is the scientific question we are trying to answer?

2. What is currently known about this topic? Who has already worked on this?

3. What have you done?

Abstract

This project seeks to determine the more likely protein structure by minimizing the internal energy of proposed topologies. Internal energy is calculated by making approximations of Van der Waals forces, Coulomb forces, and solvation energy. The project is run in standard python 3 with preprocessed protein data files.

Background

Proteins are extremely important to help the human body function. Right now, we have a general idea of the general structures that a protein may have, specifically how many secondary and tertiary structures that a protein may have. Even for hundreds of proteins, many quaternary structures have been determined for a wide variety of proteins with functions ranging from regulation to structural purposes. However, given a polypeptide sequence, we would like to know the resulting structure. Structures include the widespread alpha helices and beta sheets in addition to more rare structures like the beta sheet hairpins.

Intermolecular forces such as Van der Waals, Coulomb, and solvation energy have a profound effect on protein folding. Gibbs free energy and spontaneity dictates that a molecule wants to reduce its internal energy. By predicting internal energy of multiple possibilities for 3D protein structures, one can identify the most probable folded structure by finding the structure that minimizes internal energy.

The approximation for Van der Waals energy is done using the equation for Lennard-Jones-Potential as shown in figure 1.

 Fig 1

According to wikipedia, the Lennard-Jones potential is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules. A form of this interatomic potential was first proposed in 1924 by John Lennard-Jones. ε is the depth of the potential well, σ is the distance at which the potential reaches its minimum.

, and r is the distance between the particles.

We would also like to determine the presence and location of misalignments and incorrect folding patterns that may have a small or huge impact on the protein affecting a person’s daily life. While using computer models have aided researchers with predicting protein models, algorithms built to analyze sequences always need improvement.

Currently, scientists in this field have been able to detect variations of the same protein to determine the most likely structure that the protein may have with very little variability. However, not much research has been done about finding accurate sequences from two different proteins and observing how they aligned with each other. We are not yet able to observe the variability from the sequence and the model compared to the protein in vivo. We plan to take a sequence that one may have to code for amino acids in a protein and accurately determine which model is more energetically favorable.

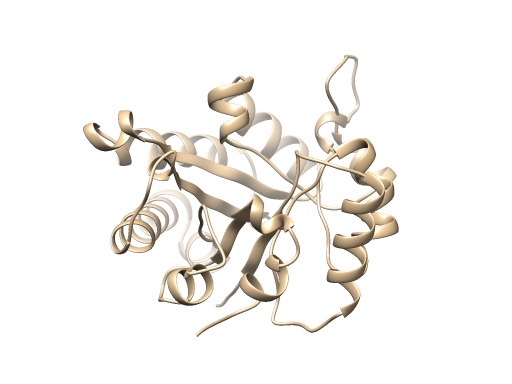
By predicting accurate models, researchers would be able to identify minute differences between distantly related models. We could use this data to observe how the structure of a protein changes due to mutations in the sequence. Once a library of models is made, professionals in the medical field can access this database and see which specific variations of protein they may need to target. Medical professionals may also use this data to determine if a person is able to react to certain virus or bacteria that may enter the immune system. They may also use this information to target enzymes that regulate or are a part of many metabolic pathways.

Methods

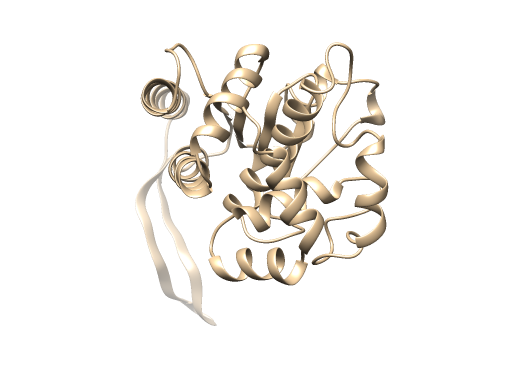
What we have done with the code: TBD

Results

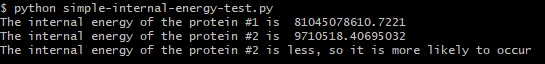
Protein 1:



Protein 2:



Given protein 1 and protein 2



The calculation indicates protein #2 is the more likely structure.

Some sources that I found right now (links below):

Toward the estimation of the absolute quality of individual proteinstructure models

Works Cited:

Breda A, Valadares NF, Norberto de Souza O, et al. Protein Structure, Modelling and Applications. 2006 May 1 [Updated 2007 Sep 14]. In: Gruber A, Durham AM, Huynh C, et al., editors. Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2008. Chapter A06. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6824/

Maiorov and Abagyan, 1998 V. Maiorov, R. Abagyan Energy strain in three-dimensional protein structures Fold. Des., 3 (1998), pp. 259-269 <https://www.sciencedirect.com/science/article/pii/S1359027898000376>

UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.