Protein Folding Prediction

Rufei Lu, Lauren Yarholar, Warren Yates and Miguel Bagajewicz

The search for an accurate and efficient protein conformation predicting method started around 1972 with a theoretical study on the ribonuclease folding. Surprisingly, physicist, chemists, and mathematicians have only been making very minute progress towards the empirical prediction algorithm of three-dimensional protein folding structures. The three most commonly accepted methods used in the past decade consist of homology prediction, folding recognition, and ab initio determination. Our research takes the ab inito approach using computational genetic algorithm (GA) search method. The GA is based on the basic concept of evolutionary natural selection for the "fittest" individual. To this technique we added an adaptation search with gene alteration. We have successfully developed a GA program that minimizes the potential energy of target sequence and generates the corresponding Cartesian coordinates for each atom. By using a three dimensional graphing software, Accelry's, we were able to visualize the predicted conformation and compare to natural conformation.

As mentioned above, our research takes the ab initio approach through the use of the genetic algorithm (GA). The GA starts with an initial population of protein conformations. The initial population size can vary depending on the preference of the user. After the initial population ahs been determined, a potential energy function is applied to the poplation. The potential energy function includes various energy terms that will help to determine the fitness of the molecule. Following the application of the potential energy function is the reproduction process which takes place with the occurrence of three genetic operators. Operators are rules that modify individuals and the population to include diversity to the process. Our GA contains four different operators: selection, crossover, mutation, and adaptation. The GA finally stops on the occurrence of one or two occasions. One is that there is a solution or the GA has proved the impossibility of the reproduction.

In conclusion, the GA based search and optimization method is a simple and efficient method for predicting short isolated native protein structures. Our GA has demonstrated its ability to determine the global minimal energy and generate natural conformations for unknown proteins. This result can be shown with the met-enkephaline predicted structure from our GA. However, the resolution and accuracy of GA depends largely upon the fitness function and the GA parameters optimization process. To further improve the prediction by the GA, more refined fitness function with improper torsion angle penalty, bond stretching, and bond angle bending should be used. Other possible alterations of the GA with the use of a various combination of operators may also improve the quality of the genetic algorithm.