

Undergraduate Summer Research Grant Program 2006  
Research Plan

**Research Topic:**

Development of a method that allows to calculate the tryptophan fluorescence for any given protein conformation. Finding the best application for this method to compare experimental data with data from protein folding simulations.

**Research Goals/Abstract:**

Protein motions play essential roles in biological processes. The function of a protein is determined by its three dimensional shape, its conformation. Many diseases are directly related to misfolded proteins, such as bovine spongiform encephalopathy (Mad Cow disease). Understanding the folding process may help to develop methods for preventing or correcting misfolding. In previous work methods were developed to simulate and analyze protein folding pathways. One successful method developed by our group is based on PRMs (Probabilistic Roadmap Methods), a computer science technique for simulating robot motion. In this method, a roadmap of conformations is created for a given protein. The roadmap is created in two main steps. The first step is sampling valid protein conformations, nodes, and storing them in the roadmap. The second step is finding connections between the conformations, edges. For proteins this is based on the free energy level of the conformations. After a roadmap is made, pathways between the unfolded conformations, and native, folded conformations of the protein can then be found and studied. The accuracy of the results can be evaluated using already existing experimental data such as secondary structure formation order.

While in previous evaluations the simulated results showed the same results as the experimental data, additional types of experimental data can be used to validate the protein folding simulation. For example data for the tryptophan fluorescence of proteins can be gathered. Tryptophan is one of the amino acids, or residues, that can exist in proteins. Since tryptophan is aromatic, when exposed to light, it adsorbs light and then emits light within a specific wavelength range. The intensity and wavelength of the fluorescence can be measured. The fluorescence increases with the number of native connections within the conformation in which tryptophan is involved. Data about the fluorescence can be gathered during the folding process. The fluorescence is highest for the native conformation and decreases with unfolding. This data is useful because large amounts of experimental data already exist and new measurements can be relatively easily done. I will develop a method that will calculate the tryptophan fluorescence of any given protein conformation. I will then explore the best method for to compare the experimental data with the data of the protein folding simulation. I hope that the tryptophan fluorescence data from protein folding simulations will match with the experimental data. The documentations will not only include a documentation about my code and results, but also a paper that can be included into a paper to be submitted to a technical conference or journal in this field.

## Tasks:

### I) Development of method

- reading Guang Song's thesis, which is the introduction to the already existing code, papers about tryptophan fluorescence and measurement methods, and others related to the topic.
- learning about tryptophan fluorescence measurements by visiting a biochemistry lab for such an experiment
- getting familiar with the code
- developing the pseudocode for the methods that are added to the program
- adding the code to the program
- running the program, which means creating roadmaps and calculating the fluorescence for the protein conformations within the roadmap

### II) Testing

- testing the accuracy of the new added methods
  - comparing to values with values that are calculated by hand
  - comparing the values with the RMSD value, which indicates how close a given conformation is to the native state. The tryptophan fluorescence supposed to be highest for the native state and lowest for the unfolded state. The fluorescence supposed to show correlations to the RMSD value.

### III) Finding best application

- finding the best suitable protein for to start with
- finding and testing the best application for the method
  - Extracting pathways from the roadmap (various methods) and evaluate tryptophan fluorescence by comparing experimental data with data from the simulation
  - Evaluating ensembles of pathways and finding the best way to compare the simulated folding process with existing experimental data

### IV) Documentation

- documentation of the code and results
- writing a paper that can be included into a paper to be submitted to a technical conference or journal in this field.

## Deliverables:

reading + looking at the code	1. week
integrating the new code, set up web page	2. week
Testing the code	3. week
Conference	4. week
Extracting pathways, evaluation ensembles evaluation	5.-7. week
documentation, paper, web page, poster	8. and 9. week