# A PROBABILISTIC ROAD MAP APPROACH TO STUDY PROTEIN FOLDING

## INTRODUCTION:

Protein Folding is the process by which a protein attains it's functional shape or conformation (structure with minimum energy/native state) from its constituent polypeptides (chain of amino acids residues) by folding in a particular pattern. A protein can follow different paths to attain its final conformation of which only a few are energetically feasible. Also, structurally similar proteins can exhibit different folding behaviour.

How can we discover this folding pathway for a given protein?

## A MOTION PLANNING APPROACH :

At first sight, robots and proteins seem to have little in common. But, they are very similar in their functionality which is based on their motions. Therefore, to study motions in these two different domains the same underlying algorithmic framework can be applied. Our approach towards Protein folding, based on this observation, shall be derived from Probabilistic Road Maps(PRMs), originally developed for robotic motion planning.

## MOTIVATION:

Why study Protein Folding?

Proteins play a vital role in various biochemical processes. Various diseases like Mad Cow disease and Alzeimer's disease are caused due to deviations (folding and aggregation) of protein structure synthesised in the body. This makes study of the procedure of formation of proteins a vital problem.

## METHODOLOGY:

Our PRM based method is based on the following procedure:

Firstly, sampling of nodes(where each node represents a conformation), which will be done on the basis of energies of the conformations. Those with lower energies will have higher probabilities of being retained and those with energies beyond a certain threshold will be rejected.

Secondly, connecting two nodes by an edge. The edge will be assigned a weight (more the energetic feasibility i.e. lower the energy, lesser will be the weight) based on the energies of the intermediate conformations along the edge.

Thus, an approximate map of the protein's energy landscape will be generated. In this map, all the shortest paths shall represent the most energetically paths and shall encode many feasible folding pathways.

Next, we will apply techniques of Map-based Monte Carlo simulation(MMC) and Map-based Master Equation(MME) calculation (derived from their traditional counterparts) on this reduced map. These methods are better than the traditional methods since they can be applied to much more complex molecules as they work on a reduced landscape model.

If time permits, we would like to further improve on our method by implementing non-linear dimensionality reduction.

REFERENCES:

1)A Motion Planning Approach to Studying Molecular Motions, Lydia Tapia, Shawna Thomas, Nancy M. Amato, Communications in Information and Systems, 10(1):53-68, 2010. Also, Technical Report, TR08-006, Parasol Laboratory, Department of Computer Science, Texas A&M University, Nov 2008.

2)Intelligent Motion Planning and Analysis with Probabilistic Roadmap Methods for the Study of Complex and High-Dimensional Motions, Lydia Tapia, Ph.D. Thesis, Parasol Laboratory, Department of Computer Science, Texas A&M University, College Station, Texas, Dec 2009.

DATA SOURCES:

Protein Secondary Structure Data Sets:

1)http://archive.ics.uci.edu/ml/datasets/Molecular+Biology%28Protein+Secondary+Structure%29

2)http://archive.ics.uci.edu/ml/datasets/Protein+Data