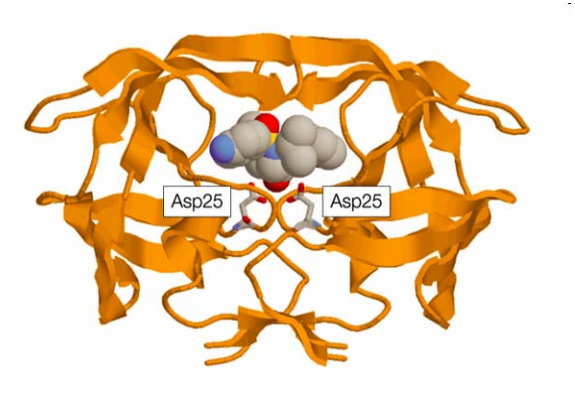
Sample Protein Functional

Peggy Yao, Jean-Claude Latombe
Biomedical Informatics, Compute Science
Stanford University

Conformations

Motivation (I)

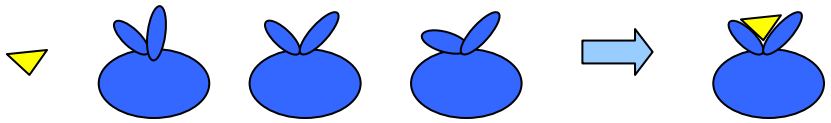


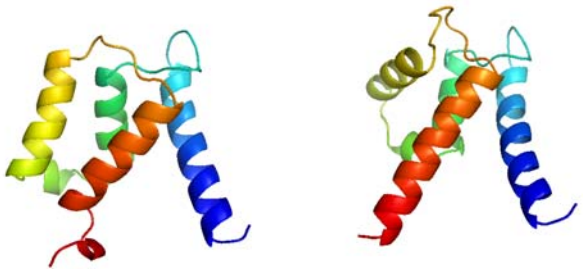
- Proteins
  - The major molecules that carry out our metabolic activities.
  - Function by interact with other molecules.
    - The interaction is largely determined by structures.
- Drugs
  - Small molecules that inhibit or facilitate the interactions between proteins and some specific molecules.
- Computer-aided drug design
  - Design drugs based on protein structural modeling.

Do we have good-enough structures? NO!

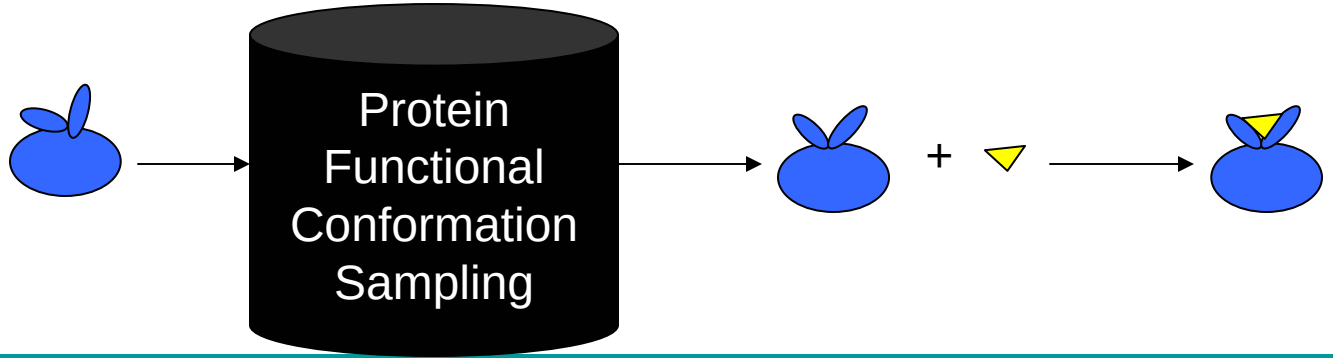
Motivation (II)

- Protein structures are dynamic
  - Conformation selection theory



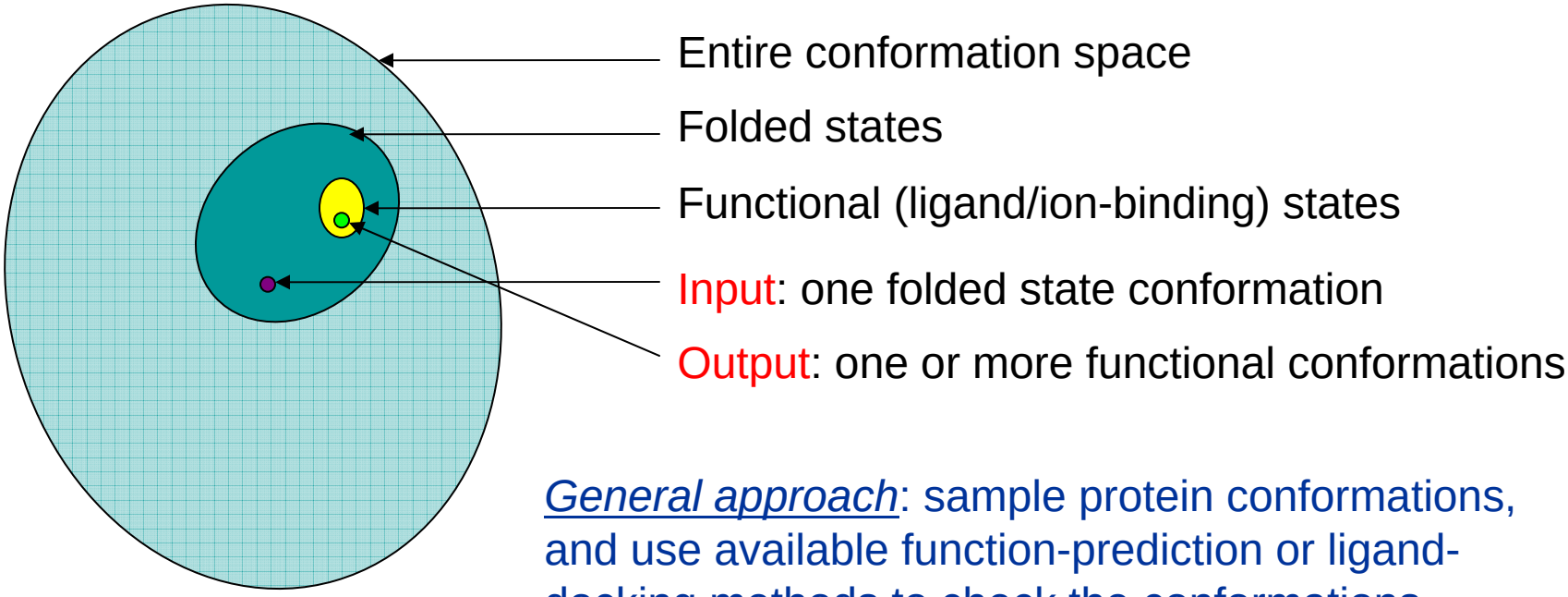


  - Experimental techniques can only provide very few conformations, which are not necessarily functional conformations.
- Need computational methods



Problem Definition

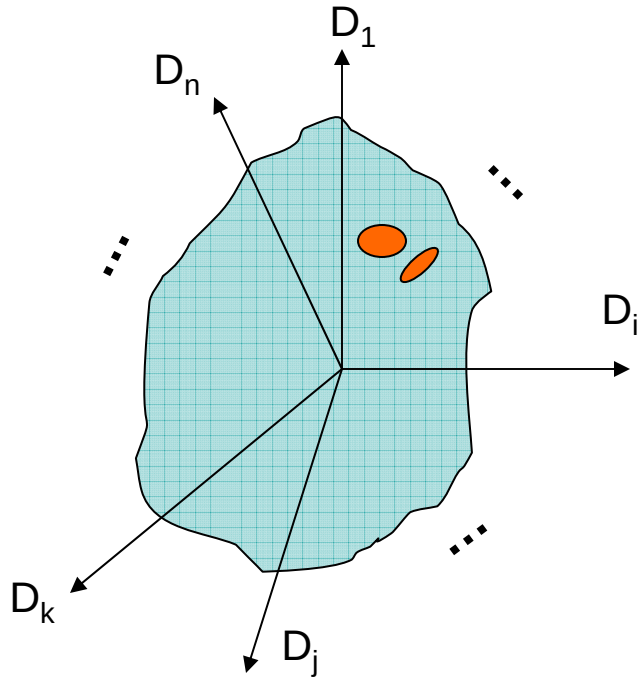
- Sampling protein functional conformations



General approach: sample protein conformations, and use available function-prediction or ligand-docking methods to check the conformations.

Challenges and Observations

- Lots of variable elements and constraints
  - Variable elements
    - Hundreds of atom positions, bond lengths, bond angles, dihedral angles, etc.
  - Constraints
    - Hundreds of bonds
    - No steric clash



- Observations
  - Not all variable elements are truly variable
    - For example, helices and sheets.

Research Framework

Protein Structure Model

Bond Constraints Model

Rigidity Analysis

Sampling Method

Applications: Protein-ligand Interaction Study

Aim 1:

To develop a good way to model bonds as constraints to facilitate efficient exploration of the conformation space.

Aim 2:

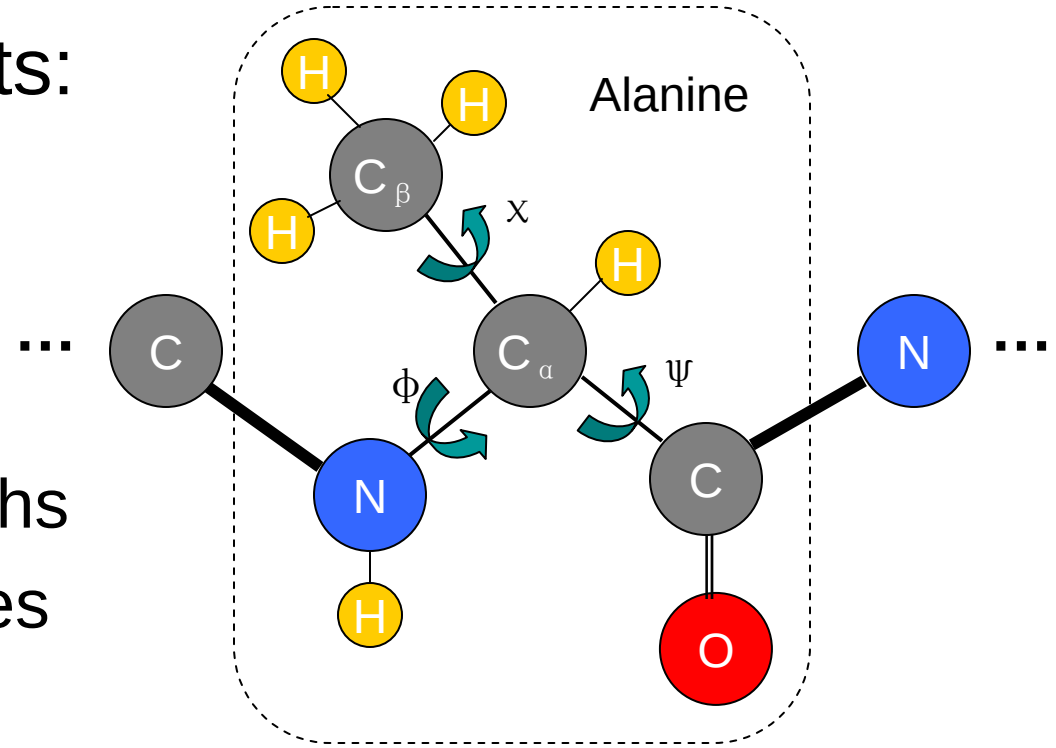
To develop a sampling approach to reach the functional states efficiently.

Aim 3:

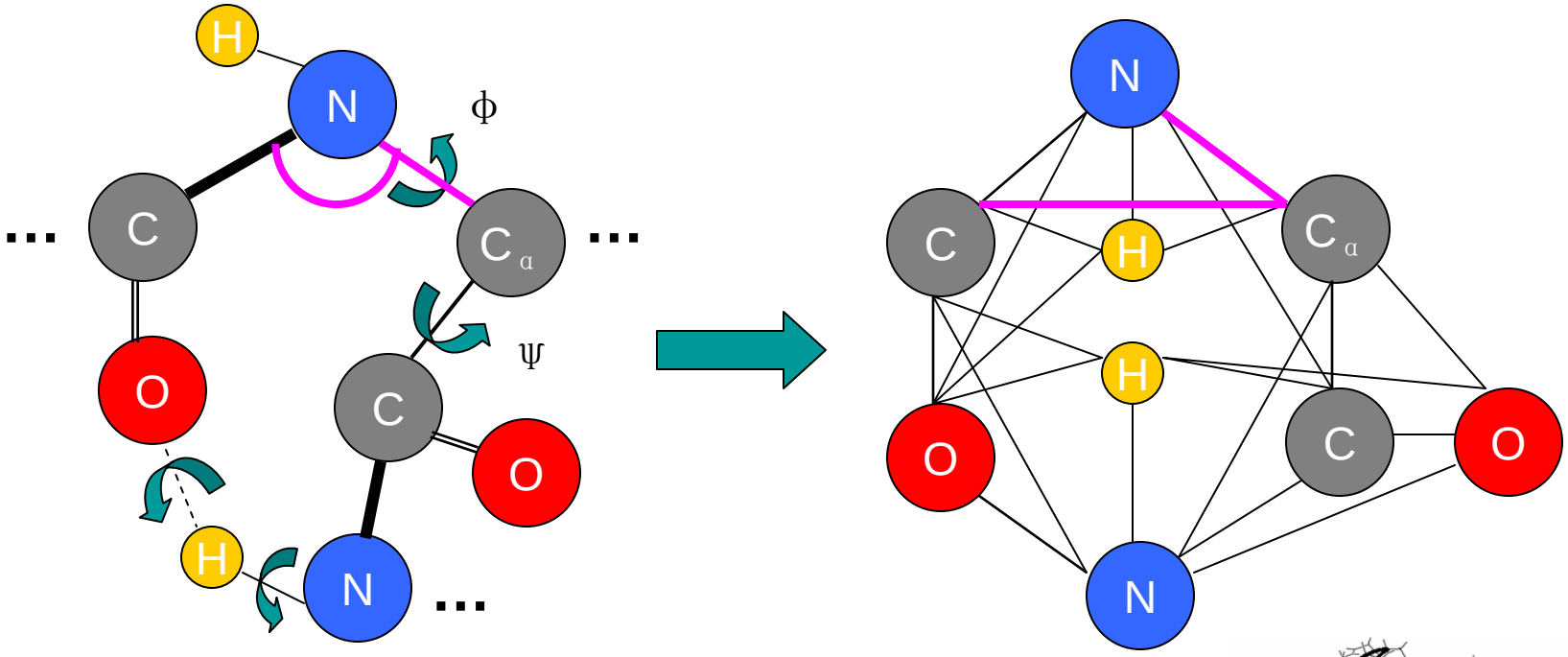
To apply the sampling method to function-prediction or ligand docking.

Linkage Model

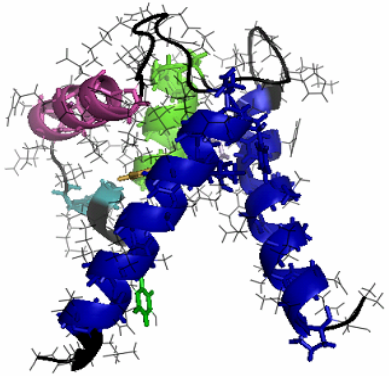
- Variable elements:
  - Dihedral angles
- Assumptions:
  - Fixed bond lengths
  - Fixed bond angles



Rigidity Analysis

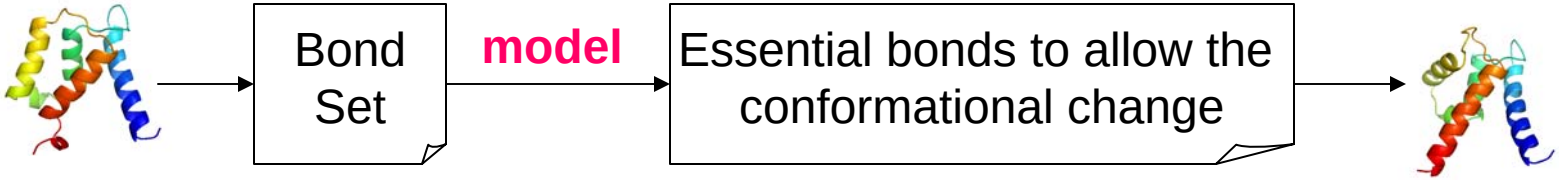


- Model bonds as *Distance Constraint Graph*.
- 3D Pebble Game
  - An algorithm to identify rigid regions, over-constrained regions, and collective motion regions.



Bond Constraint Model

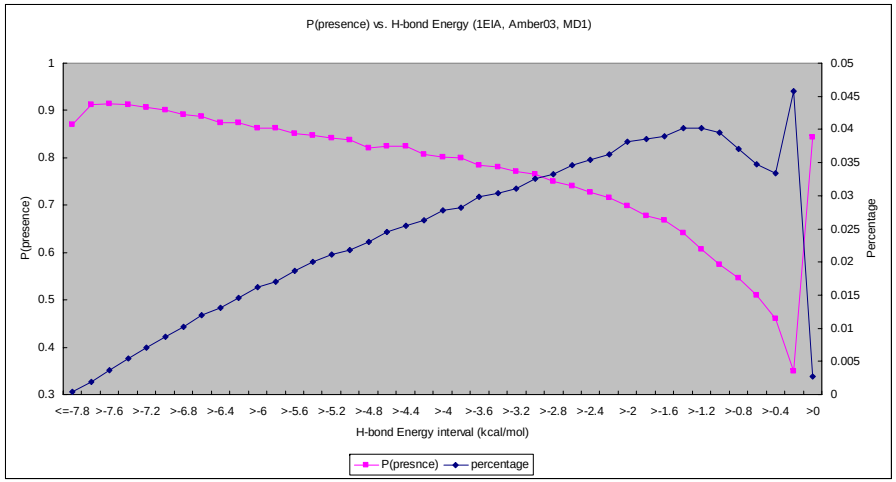
- *Question*: what are the bonds shall we model as distance constraints?

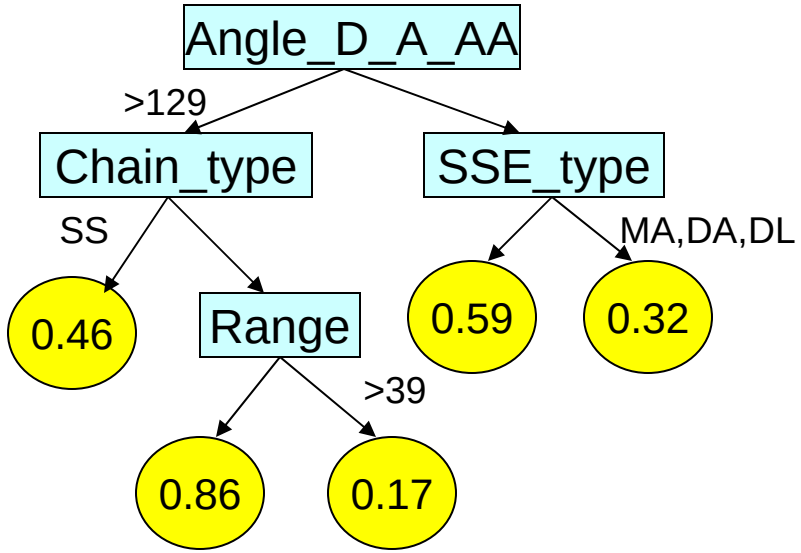


- Bond types
  - Covalent bonds
    - Strong and stable
  - Non-covalent bonds
    - Hydrogen bonds
    - Hydrophobic interactions
    - Many, weak, and dynamic

Hydrogen Bond Selection

- Learn H-bond stability from MD (Molecular Dynamics) simulation
  - Stability measurement: P(presence)
  - P(presence) vs. Energy
  - Decision-tree





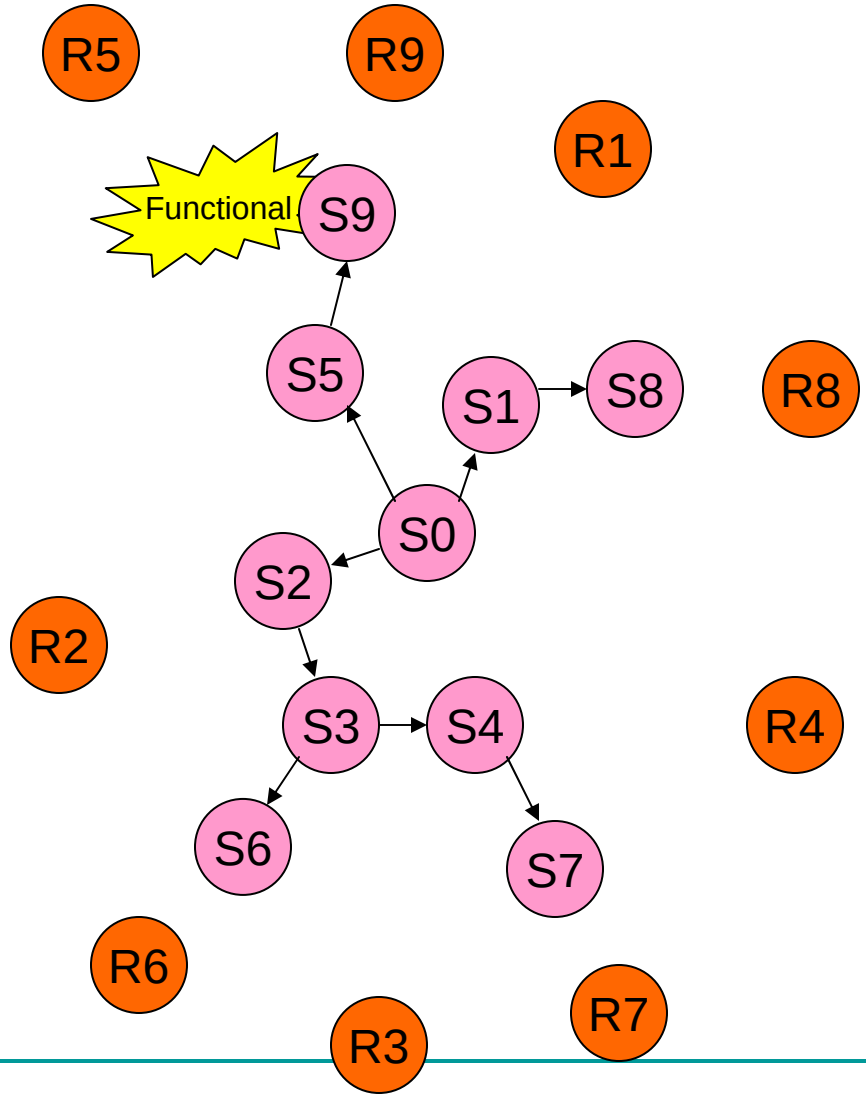
Conformation Sampling

- Goal:
  - Start from a folded state conformation, efficiently sample the valid conformation space until obtain a functional conformation.
  - Assume there exists a software which can recognize the conformation at the functional state.

Randomly-Guided Conformation Sampling Tree

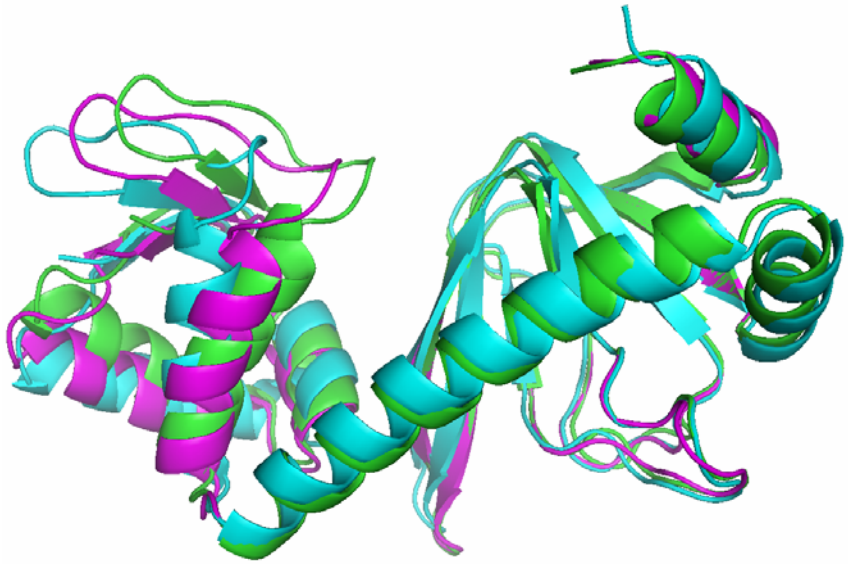
While (not reaching the goal):

1. Generate a random structure.
2. Find the node closest to the random structure, say node *i*.
3. Identify all H-bonds in node *i*.
4. Select a subset of H-bonds to be constraints, together with all covalent bonds.
5. Linear-interpolate *i* to the random structure for 100 steps while maintaining all rigid bodies.
6. Insert the new node into the tree as node *i+1*.



Preliminary Results

- Catabolite Gene Activator Protein (1G6N)
  - 200 amino acids => more than 800 total DOFs
  - Generated a tree with 100 nodes.



Green: initial conformation

Cyan: goal conformation

Magenta: best achieved