Prediction of Protein Folded Structure by Monte Carlo Simulation

Team 5: Alex Walsh, Jingqian Liu, Jinghong Liu

Proteins: The basics

- Proteins are complex macromolecules made of long chains of amino acids
- Proteins have several levels of structure, ranging from segments of residue chains coiling into α-helices, to multiple subunits combining into one large protein
- Folding is a complicated process that involves interaction between each AA and water molecules, hydrophobic/hydrophilic regions of other AAs, and sometimes other proteins
- This makes folding a computationally expensive problem

Protein Folding Models

Popular methods

- All-Atom MD with explicit solvents
  - Effective, but slow and limited to a small number of residues
- United-atom models
  - Levitt-Warshel, UNRES, KABS
- More recent developments use deep learning to simulate folding

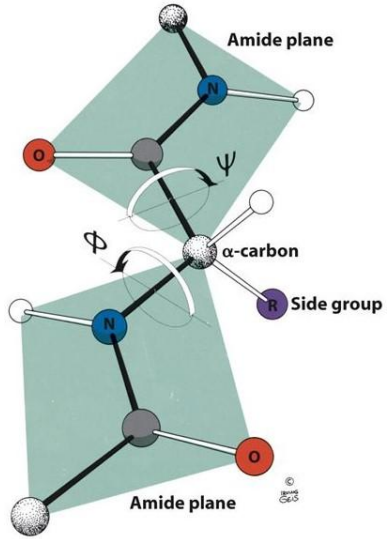
Our project

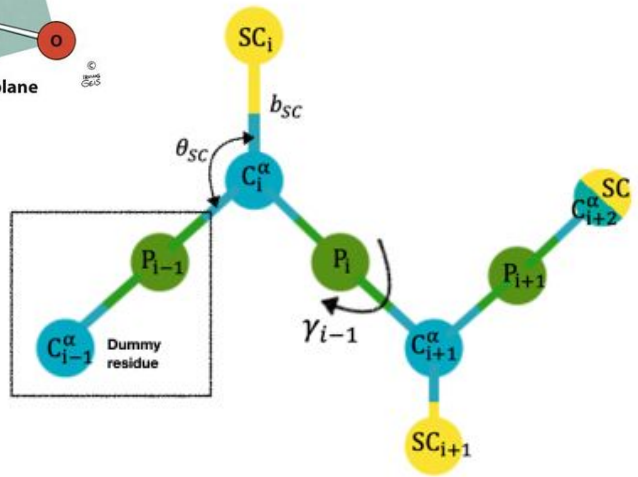
- Previously proposed method with united atom models
  - The Generalized Born model can simulate the effects of water implicitly
  - Levitt and Warshel model reduces each amino acid to two beads
  - Virtual Movement Monte-Carlo (VMMC) is a cluster algorithm that prevents low acceptance rates by moving neighbor particles with the seed
- To further improve the efficiency and feasibility we adopted **an united-atom model created by A. LIWO in 1993 (Precursor of UNRES)**

Protein Sci. 1993 Oct; 2(10): 1715–1731.

Geometry of the Model

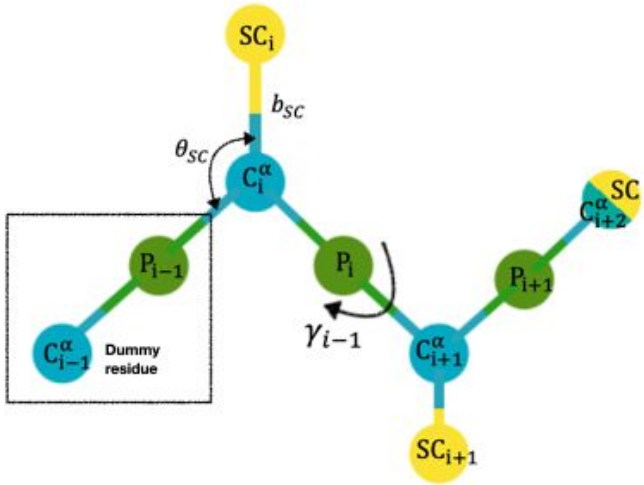
- Proteins are usually described by two angles, ϕ and ψ. This degree of freedom is reduced in the model
- By reducing the residues to three kinds of beads including C^{α} bead, peptide group bead and side chain bead.
- Fixed parameters:
  - b_{SC}, the distance between C^{α} and its side chain (SC)
  - θ_{SC}, the angle between SC and the peptide group of the previous residue (P)
  - C^{α}-C^{α} distance (3.8 Å)
  - Dihedral angle defined by SC(i)-C^{α}(i)-C^{α}(i-1)-C^{α}(i+1)
- Parameter can be changed
  - γ, the torsion angle, namely the dihedral angle defined by last two C^{α}s, C^{α} and SC. With a peptide with n C^{α} atoms, there are (n-3) γ, which means (n-3) degree of freedoms





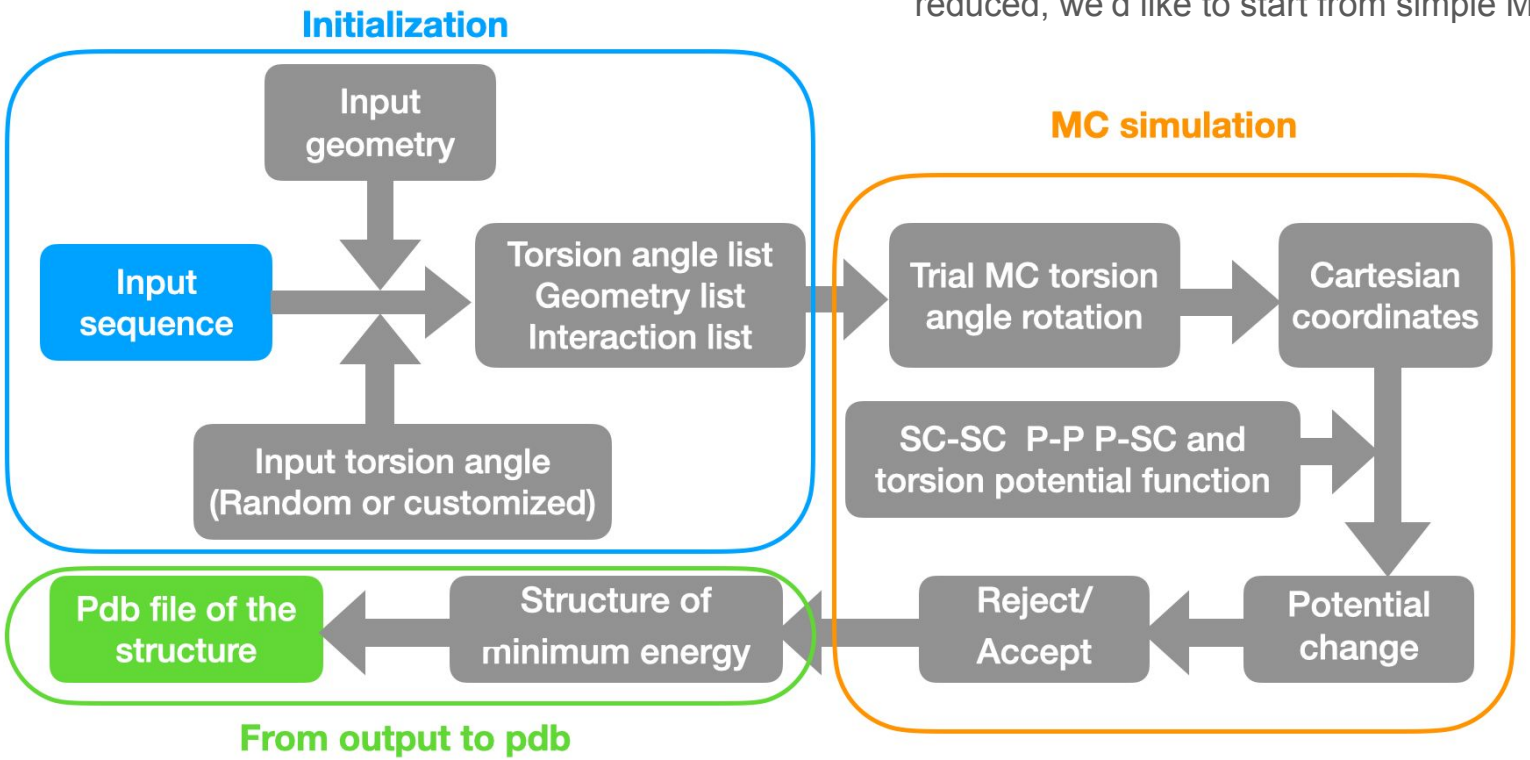
Geometry of the Model

- Dummy residues
  - Dummy beads are appended to the protein chain (if the end residue is not GLY) to add an additional degree of freedom
- Two lists for the model
  - Geometry list and interaction list
  - C^{α} s and dummy beads just included in the geometry list but not in interaction list



Summary of Algorithm

- Since the degree of freedoms is largely reduced, we'd like to start from simple MC

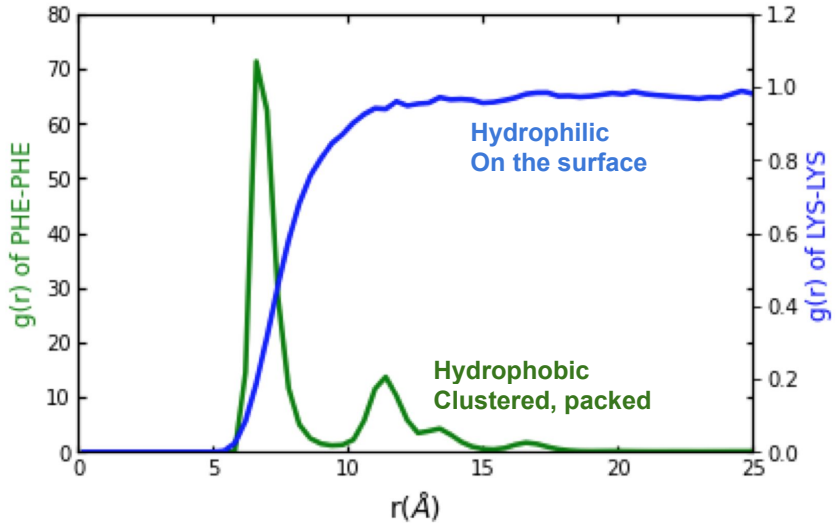


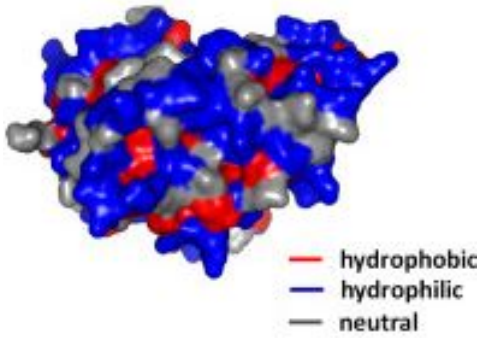
Force Field

- Side Chain-Side chain (SC-SC) interactions

$$U_{SC_i}U_{SC_j}\begin{cases}|\epsilon_{ij}|\left[\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12}-2\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6}\right] & \text{(hydrophobic)}\\ |\epsilon_{ij}|\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6} & \text{(hydrophilic)}\end{cases}$$

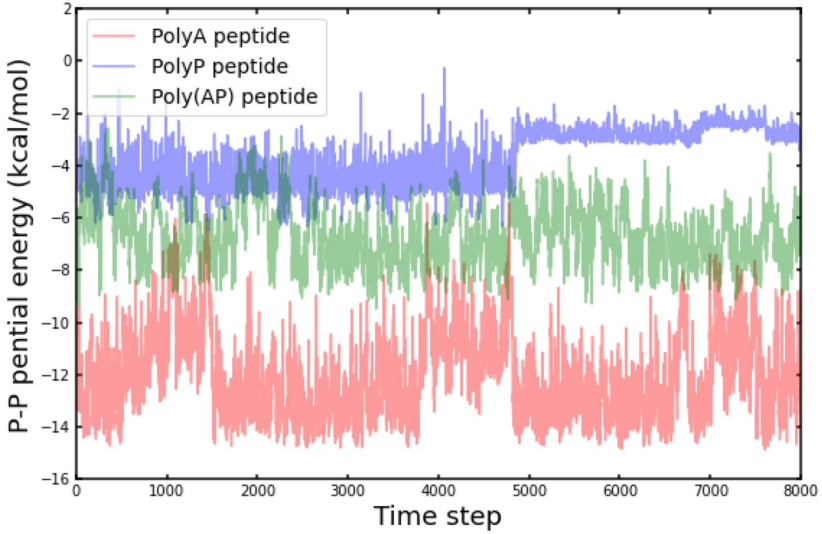
Validate the SC-SC interactions: 64 SC beads in a (80Å * 80 Å * 80 Å) PBC box





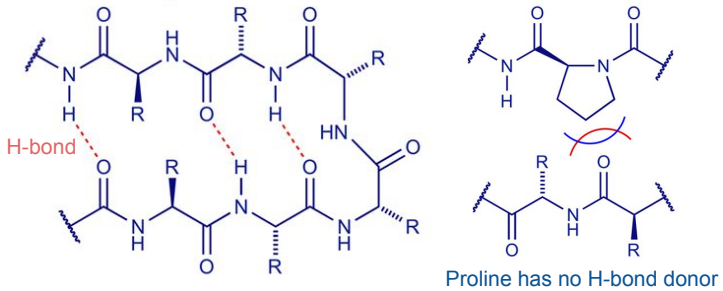
Force Field

- Peptide-Peptide (P-P)
  - Hydrogen bonds, electrostatic interactions
  - Important for forming secondary structures
  - Proline is special



$$U_{p_i p_j} = \frac{A_{p_i p_j}}{r_{ij}^3}(\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij}) - \frac{B_{p_i p_j}}{r_{ij}^6}[4 + (\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij})^2 - 3(\cos^2\beta_{ij} + \cos^2\gamma_{ij})] + \epsilon_{p_i p_j}\left[\left(\frac{r^0_{p_i p_j}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{p_i p_j}}{r_{ij}}\right)^{6}\right]$$

Type of residue pair	ϵ	r^0	A	B
ordinary-ordinary	0.305	4.51	3.73	1306
ordinary-proline	0.365	4.54	0.00	1129
proline-proline	0.574	4.48	5.13	335

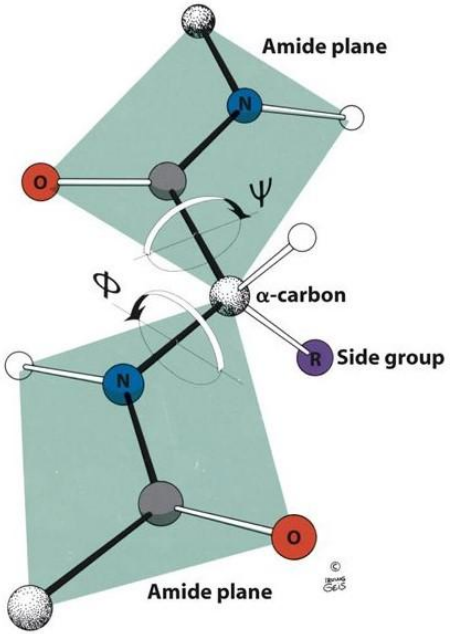


Force Field

- Side Chain-Peptide (SC-P)
  - Potential between SC and Ps which are not its neighbors
- Torsional Energy
  - Energy related with the torsion angle

$$U_{SC_i p_j} = \epsilon_{SCp} \left(\frac{r^0_{SCp}}{r_{ij}} \right)^6$$

$$U_{tor}(\gamma_i) = a_0 + \sum_{k=1}^{6} (a_k \cos k\gamma_i + b_k \sin k\gamma_i)$$

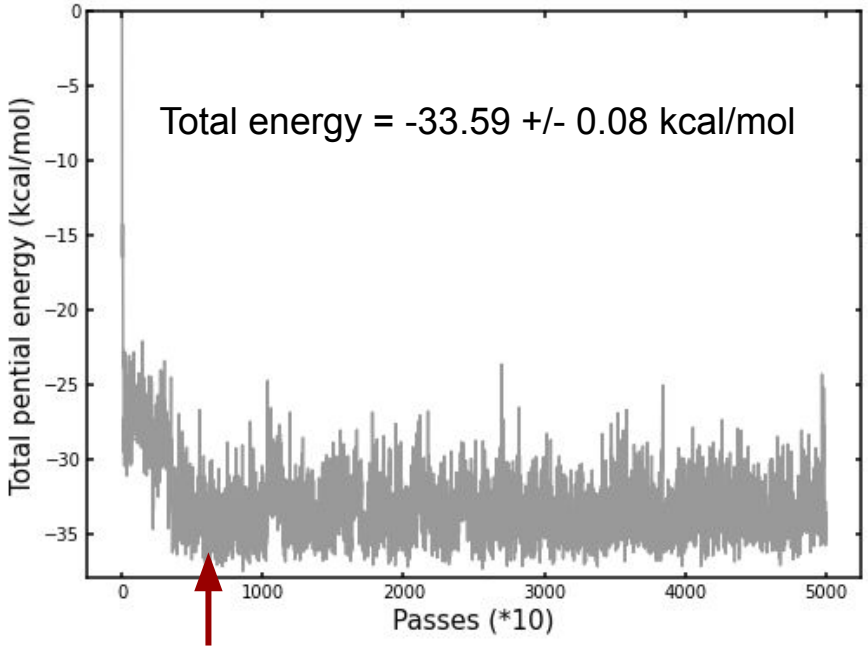


Settings for the Simulation

- Room temperature @ 298k
- Time steps (τ): 1 step = changing torsion angles in range of (-10°, 10°) (40-70% in different systems)
- Menu of moves: loop over all the γ in one pass
- No periodic boundary condition was used in structure prediction
    - Consider our model as 1 big molecule

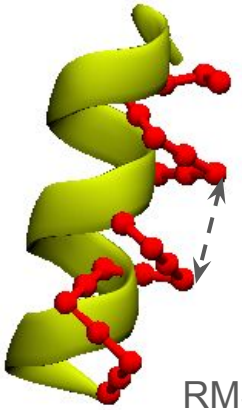
Structure prediction of α helix

- Protein: monomer of 6B17
- Sequence: TPRQARAARAAC
- Running 50000 passes takes ~44min



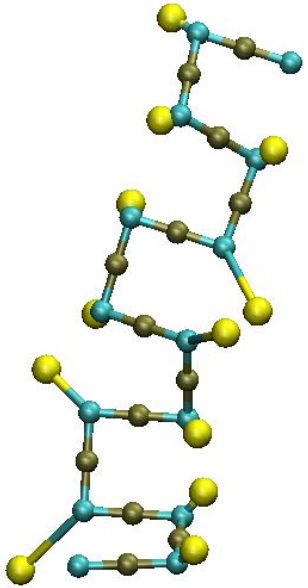
Structure prediction of α helix

- The distance between two turns in predicted structure: 4.9 Å (5.4Å in natural helix)
- RMSD (Root-mean-square deviation) between natural and predicted structure is 3.3 Å
- The top view of the predicted structure resemble the natural helix structure

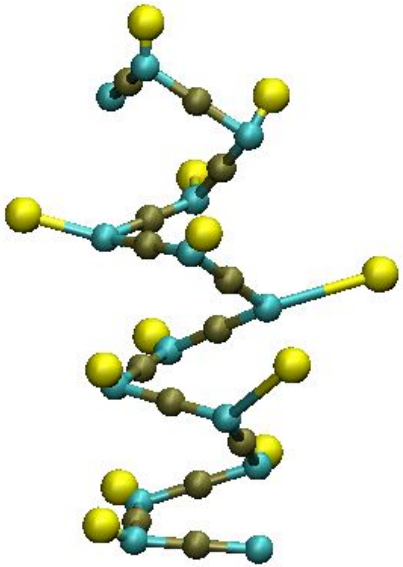


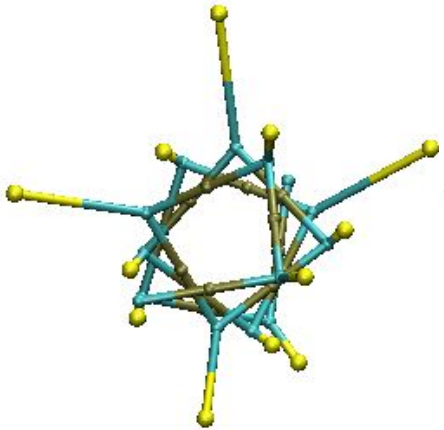
RMSD = 3.3 Å

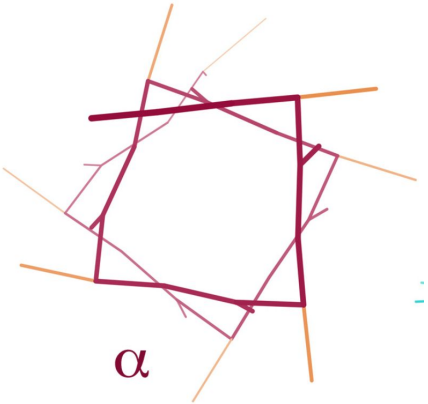
Input



Output



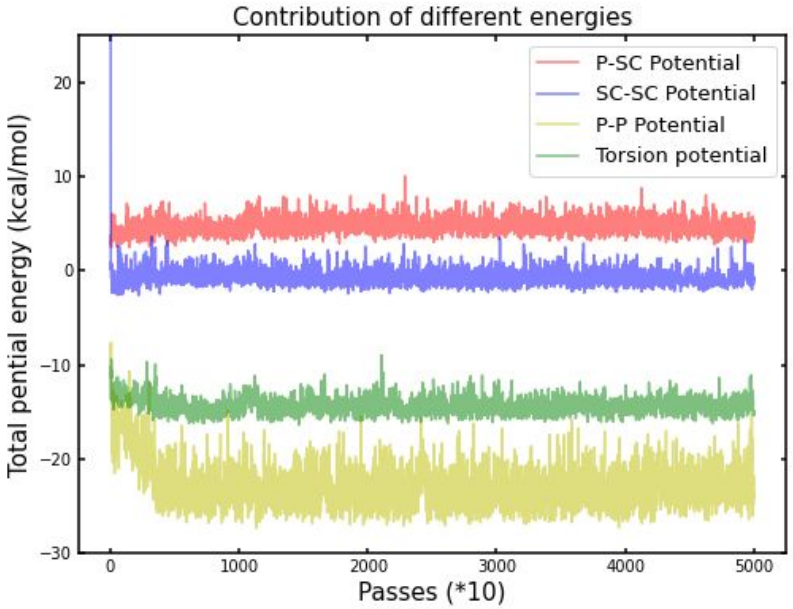




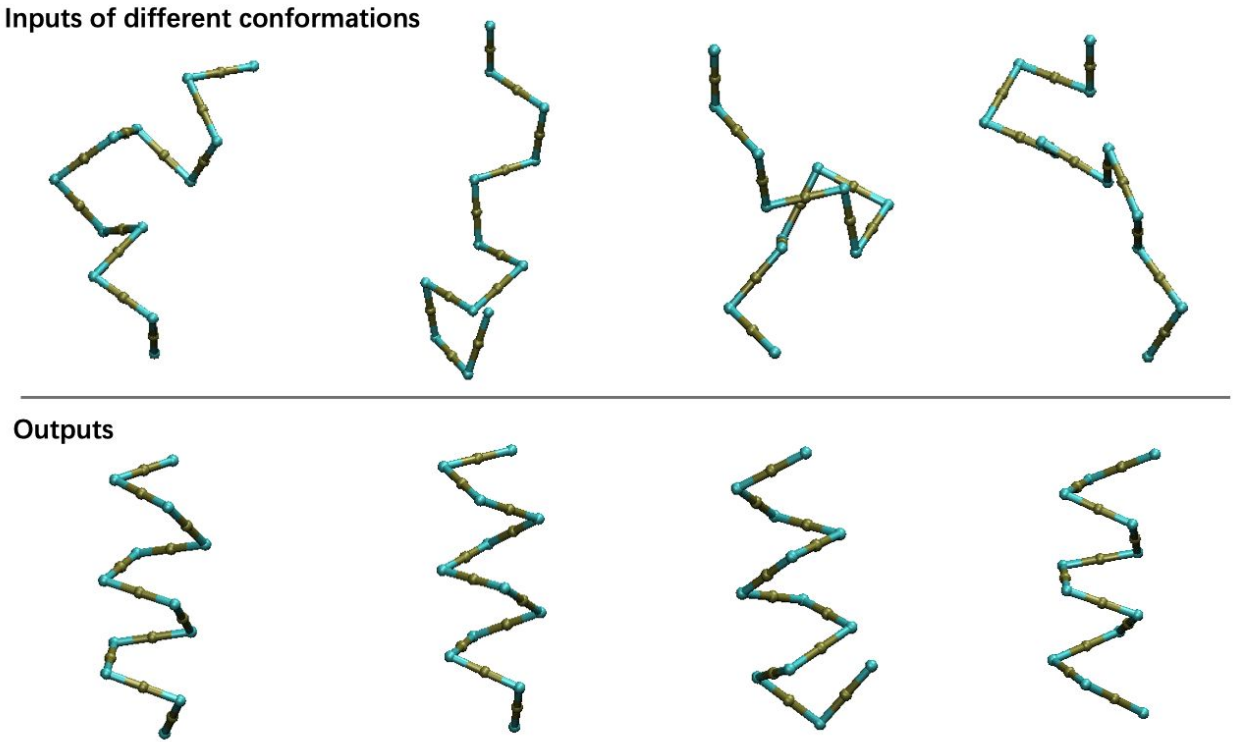
https://en.wikipedia.org/wiki/Alpha_helix

Structure prediction of α helix

- Energy profile: P-P potential contribute a lot to process of folding



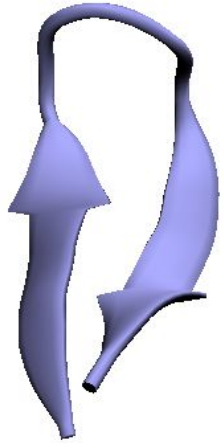
Structure prediction of α helix



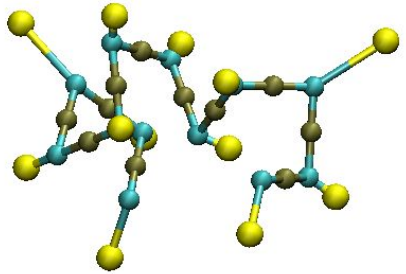
Structure prediction of β hairpin

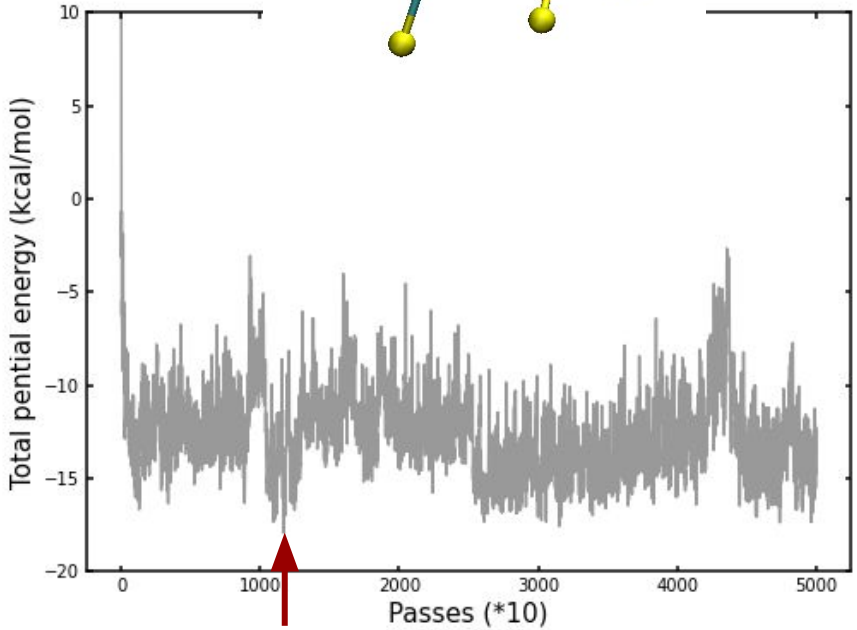
- Protein: 2EVQ
- Sequence: KTWNPATGKWTE

Natural structure

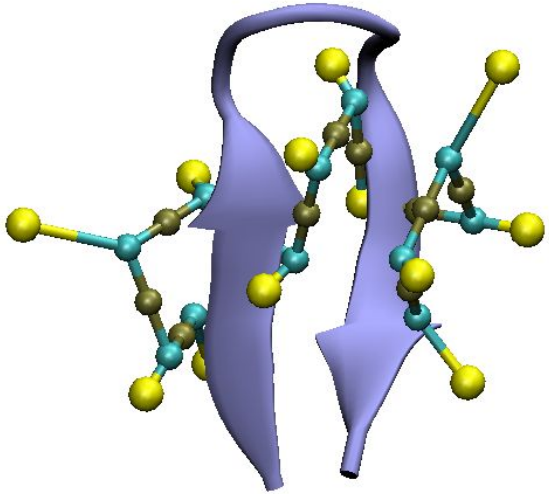


Predicted structure



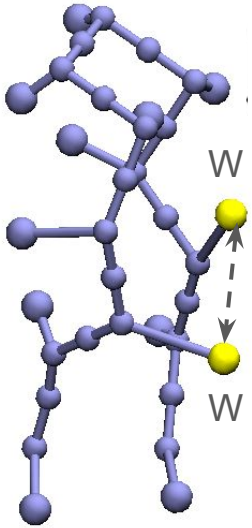


Structure prediction of β hairpin



- What cause the failure of prediction? Sampling method or force field or bugs?

- A test performed to figure out the issue: Put the natural structure under the force field



$$|\epsilon_{ij}| \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^{6} \right] \quad \text{(hydrophobic)}$$

5.0 Å< the van der Waals radii 7.2 Å

$$U_{SC_3SC_{10}} \approx 121\ kcal/mol$$

- This united-atom model may not be suitable for all the protein folding

Summary

- We made a tool to make protein structure prediction by implementing **A. LIWO model in 1993**
- We made evaluation of this force field by both pre-prediction simulation and also predicting simple α helix and β hairpin structures. α helix prediction shows consistency with natural structure while β hairpin prediction shows inconsistency with natural structure.
- Improvement can be done with **UNRES** model

Gitlab link: https://gitlab.engr.illinois.edu/physic466project/mc_proteinfolding

Questions?