Prediction of Protein Folded Structure by Monte Carlo Simulation

Team 5: Alex Walsh, Jingqian Liu, Jinghong Liu

Abstract

This project applies ensemble of an united-atom model of protein and Monte Carlo simulation to predict the protein structure. After preliminary validation of the force field, two test simulations were performed to evaluate the method's efficiency in predicting secondary structure of protein. In the tests, two protein sequences (PDB ID 6B17 and 2EVQ) whose natural structures are respectively α helix and β hairpin were investigated. The program was able to give decent prediction of the α helix conformation in our test simulation for 6B17. However, it could not provide a good prediction for structure of 2EVQ which has a natural structure of β hairpin. The discussion has been made to evaluate the tool based on these results.

1 Introduction

Proteins, polypeptide chains made of amino acids (residues), carry out the very basic functions for maintaining life and are the most crucial parts of any biological systems. The numerous conformations of proteins construct a great disparity of functions. These conformations are investigated by studying spontaneous folding mechanism of protein molecules, from the aspects of both intermolecular and intramolecular interactions. Depending on the temperature, the residue sequence on the polypeptide chain, the initial conformation of the chain, and the solvent, there would be varied outcomes of transitional and final conformations. Real proteins are extremely complicated. For example, 50 residues could bring about 200 degrees of freedom [**1**] during computation. The complexity of model requires modern protein studies aided by high performance computers and different algorithms.

To reduce the challenge of computational workload, people have brought up different models with reduced degrees of freedom to simulating protein behavior, especially protein folding, where each residue will be described by one or several beads instead of

all the atoms. In these united-atom models [**1**][**2**][**3**][**4**], only the most important degree of freedom will be remained and the potential function will be adapted to accurately describe interactions between beads. One of these united-atom models, developed by A. Liwo et al. [**5**], was adopted in this project, considering the limitation of accessible computational resource as well as the feasibility of implementation. It is the precursor of UNRES [**2**], one of the most popular untied-atom protein models. In this model, only three beads are included in one residue, which respectively present the C^{α}, the peptide group, and a side chain. Torsional angles between C^{α}s are then introduced and becoming the only variable parameter of the model. In this way, we will have less than n degrees of freedoms given a peptide with n residues, which largely reduces the computational effort, however, introduced risk of the misdescription of the natural peptide behavior. Therefore, by doing this project, we would like to realize a tiny tool for protein folded structure prediction, and evaluate the performance of the united-atom model by further analyzing produced data.

With these aims in mind, we performed pre-simulation validation for some of the potential functions. After that, two simulations were conducted to computationally predict the structure of sequence from PDB ID 6B17 [**6**] and 2EVQ [**7**], which proteins have well-documented structure of α helix and β hairpin. These two 12-residue proteins are selected for two reasons: to reduce computational time, and to prevent trapping in local minimum energy, which is possible when too many degree of freedoms are involved in the Monte Carlo (MC) simulation. As a result, the tool made decent prediction for 6B17 structure while failed on 2EVQ structure. A conjecture about the failure is made by further analysis on the potential inability of the force field when describing the interaction of two tryptophan (TRP) in a specific case.

2 Methods

2.1 United-atom Model for Polypeptide Chain

In the model developed by Liwo et al.[**2**], the polypeptide chain is simplified to a sequence of α-carbon atoms (C^{α}) with united side chains (SC) and peptide-group (P) centered between C^{α}. For the C^{α} beads. The distance of all the virtual bond $C^{\alpha} - C^{\alpha}$ are assigned as 3.8Å and the angle between the two virtual bonds $C^{\alpha}_{i-1} - C^{\alpha}_{i}$ and $C^{\alpha}_{i} - C^{\alpha}_{i+1}$ is assumed as 90°. The position of C^{α}_{i+1} will be decided by the position of C^{α}_{i-1}, C^{α}_{i}, SC_i and the ϕ_{SC} which are the fixed dihedral angles (decided by the SC_i type) of the beads $SC_i - C^{\alpha}_{i} - C^{\alpha}_{i-1} - C^{\alpha}_{i+1}$.

For a side chain bead, two degrees of freedom will be fixed, namely b_{SC} and θ_{SC} (Figure 1). The $C^{\alpha} - SC$ distance (b_{SC}) varies for each amino acid residue type. In

addition to b_{SC}, θ_{SC} is also set to corresponding values of different residue types. θ_{SC} defines the angle between $C^{\alpha}_{i-1} - C^{\alpha}_{i}$ and $C^{\alpha}_{i} - SC_i$. The side chain's movement is now depended on the only remaining degree of freedom – γ, a set of torsion angles between C^{α} atoms.

γ_i is defined as a dihedral angle between planes composed by beads $C^{\alpha}_{i-1} - C^{\alpha}_{i+1} - C^{\alpha}_{i+2} - SC_{i+2}$. Given a peptide with n residues, there will be (n-3) degrees of freedom which are defined by the virtual bond torsional angles γ_1 γ_2 γ_3 ... γ_{n-3} . If the N-end of C-end of the peptide is not a glycine, a dummy residue (Figure 1) will be add to allow the rotation of the side chains of what were previously the terminal residues.

Above we introduced how the geometry of the peptide chain will be presented. Besides a geometry list of beads, we will also have a list for interaction where all the dummy beads as well as the C^{α} beads are absent. Only beads of P and SC will be input to the potential while C^{α}s are used only for geometry presentation.

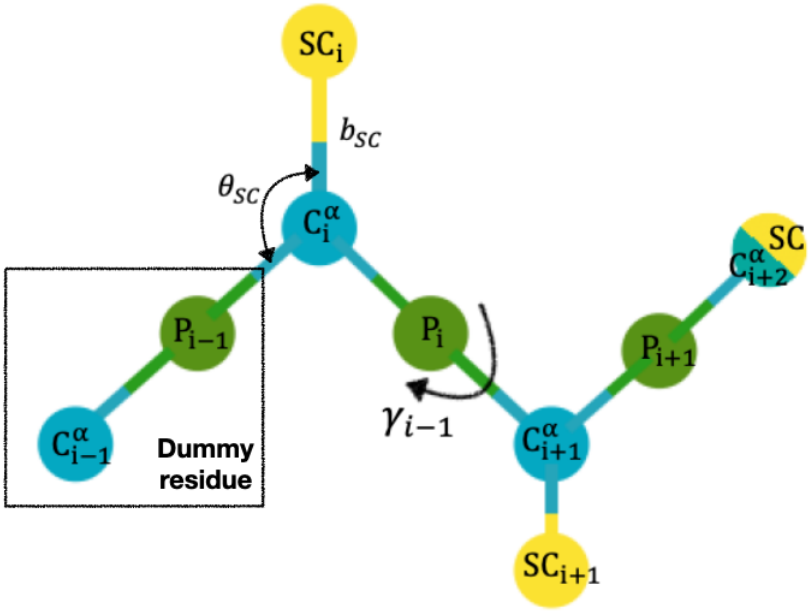

Figure 1: Geometry of a coarse-grained residue chain. One of the torsional angle, $C^{\alpha} - SC$ distance and angle between $C^{\alpha}_{i-1} - C^{\alpha}_{i}$ are respectively presented by γ_i,b_{SC} and θ_{SC}

2.2 United-atom Force Field

The chain energy function is the combination of four average energies:

$$U = \sum U_{SC_i} U_{SC_j} + \sum U_{SC_i p_j} + \sum U_{p_i p_j} + \sum U_{tor}(\gamma_i) \tag{1}$$

2.2.1 SC-SC

The first term $U_{SC_i}U_{SC_j}$ is the average energy of side-chain centroids (SC) interaction sites when residues i and j are in contact. These sites involve with hydrophobic/hydropholic interactions.

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

where ϵ_{ij} is the interresidue contact energy [**8**] and has different values for proline-proline, proline-nonproline, and other interaction (see more in Appendix). Also, r^0_{ij} is defined as:

$$r^0_{ij} = (r^0_i + r^0_j)/2 \tag{3}$$

with r^0 as the Van der Waals radii (Table A1 [**9**] in Appendix).

2.2.2 SC-p

When residues i and j are not in contact, an additional potential is required, simulating interactions between SC and peptide-bond centers (P), to prevent collapsing:

$$U_{SC_ip_j} = \epsilon_{SCp}\left(\frac{r^0_{SCp}}{r_{ij}}\right)^6 \tag{4}$$

with $\epsilon_{SCp} = 0.3$ kcal/mol and $r^0_{SCp} = 4$ Å.

2.2.3 p-p

The third term $U_{p_ip_j}$ calculates the average electrostatic interaction (including hydrogen bonding of the backbone) between the centers of the peptide group (p-p):

$$\begin{aligned}U_{p_ip_j} =& \frac{A_{p_ip_j}}{r^3_{ij}}(\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij})\\ &- \frac{B_{p_ip_j}}{r^6_{ij}}[4 + (\cos\alpha_{ij} - 3\cos\beta_{ij}\cos\gamma_{ij})^2 - 3(\cos^2\beta_{ij} + \cos^2\gamma_{ij})]\\ &+ \epsilon_{p_ip_j}\left[\left(\frac{r^0_{p_ip_j}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{p_ip_j}}{r_{ij}}\right)^{6}\right]\end{aligned} \tag{5}$$

where α_{ij} β_{ij} γ_{ij} are angles of the relative orientation of peptide groups (see details in Equation A3 of Appendix) and $A_{p_ip_j}$, $B_{p_ip_j}$, $\epsilon_{p_ip_j}$, $r^0_{p_ip_j}$ are constants (Table A2 in Appendix) which are varied for specific peptide groups (ordinary-ordinary, ordinary-proline, and proline-proline).

2.2.4 Torsional

The final term evaluates the torsional energy of local interactions by virtual bonds C^α-C^α:

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

where coefficients a_0 a_k b_k are constants γ_i are virtual-bond torsional angles (Table A4 in Appendix).

2.3 Workflow of Simulation

By implementing MC method for the introduced force field, we realize a tool in Python script. Figure 2 shows our designed workflow for this tool. First, a fasta file

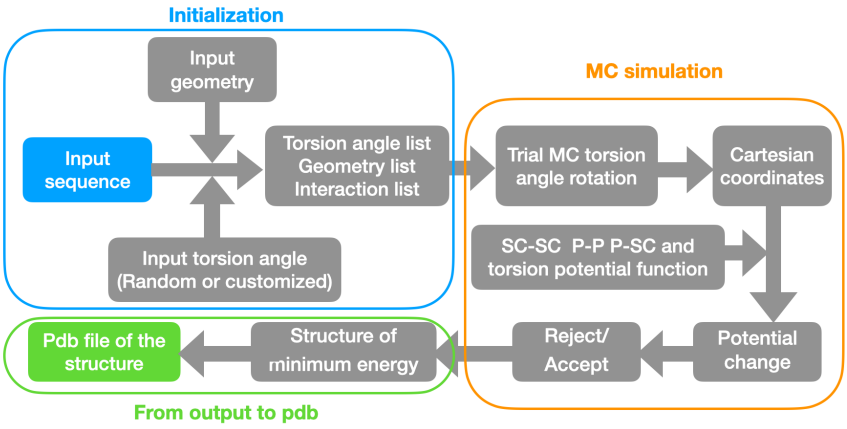

Figure 2: Flowchart of simulation for this project

will be read by the program along with all parameters files, such as geometry data of residues and constants for potential functions. From the input files, three lists, including geometry list (for all beads), interaction list (for only P and SC beads) as well as a torsional angle list are initialized. The former two lists will be decided by the input sequence while the torsional angle list will be generated either randomly or by a custom input. Then, the MC simulation will begin and loop through the torsional angle list, followed by combining the torsional angle list with all the other

input geometry parameters to generate Cartesian coordinates of all the beads. These coordinates will processed by four potential functions and see if the algorithm would accept this run. After enough sampling, the structure with the minimum energy will be generated as the output. Finally, one can convert the output file to a pdb version so that the visualization and comparison with natural structure can be achieved.

2.4 Program and Monte Carlo Settings

We have two kinds of system settings for our simulations. One is for pre-simulation force field validation for SC-SC potential with periodic boundary conditions (PBC). The other is for the actual protein folding simulation.

In the first one, 64 SC beads were placed in a $80\text{Å} \times 80\text{Å} \times 80\text{Å}$ periodic box. Here, Ewald summation can be ignored since both the hydrophobic and the hydrophilic interaction are short-range potentials. 21000 MC steps were run and the first 1000 steps, which considered as equilibration period, were removed. This simulation has been performed separately on phenylalanine (PHE) and lysine (LYS).

There second setting, where PBC was not applied, was used on the actual protein folding simulation and the validation for p-p potential validation. PBC is generally used to improve computational efficiency in a macroscopic system with many particles. In the actual runs, there is only one molecule – a simplified model of a protein, so there would be no need to apply PBC.

The simulation was set to initial at room temperature (298 K). It was started by arbitrarily setting a conformation or a customized conformation, which is virtual-bond torsional angles γ in our case. A Metropolis test is then conducted to accept or reject this conformation. For each step, one torsion angles will be rotated in a range of $(-10°, +10°)$, which yielded acceptance rate at 40 to 70%, depending on different systems. All the torsion angles will be looped over a pass. A total number of 50000 passes will be run through every structure prediction simulation.

3 Results

3.1 Force Field Validation

Before conducting actual simulations, several trial simulations were conducted to validate our force field functions. The SC-SC and p-p potential functions were selected because of their complicated forms and variable constants. The team decided

to examine them before implementation to reduce uncertainties.

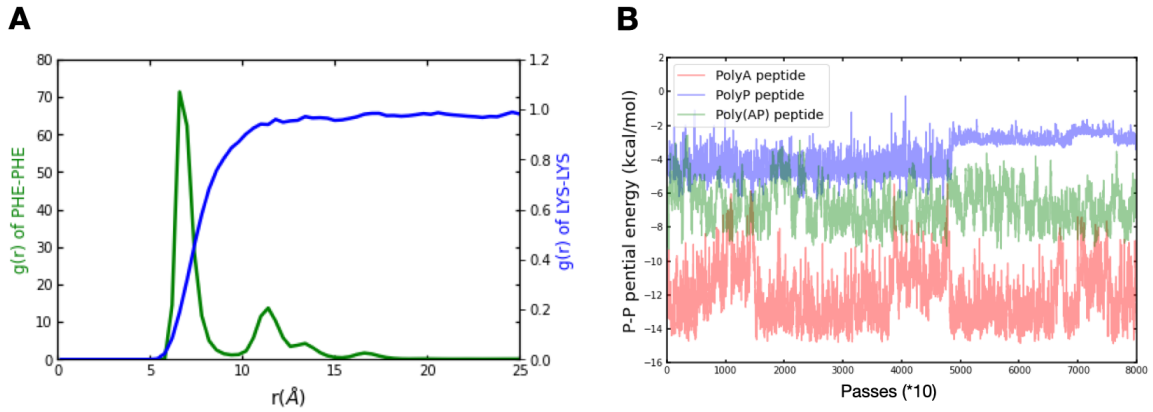

Figure 3: A) pair correlation functions for Phenylalanine and Lysine; B) p-p potential energies for three peptides (polyA, polyP and poly(AP)$_4$ (*10 on the x axis represents that we collected the energy data for every 10 passes)

The function that evaluates SC-SC potentials involves in two types of interactions: hydrophobic and hydrophilic. Runs were hence conducted to measure how well our algorithm could describe such scenario. The green line on Figure 3-A is the pair correlation function of PHE-PHE (only the SC beads), a hydrophobic amino acid, which shows two signature peaks. This form of curve indicates a hydrophobic property: the SC beads congregates in a water-soluble protein. The blue line is for LYS-LYS (also only the SC beads), a hydrophilic residue, which tend to be distributed outside of the phydrophobic cluster in a protein. Therefore, the blue curve intersects with the green one after first peak. These reasonable results brought confidence to the validation of force field functions.

The other function we validated was for computing energies of p-p interactions. This type of interaction is the major factor of constructing secondary structures in a protein by forming hydrogen bonds (H-bond). Proline (PRO) are special residues. They do not possess proton donors for the hydrogen bonds. This is also the reason we adopt different p-p interaction parameters for PRO (see Appendix Table A2). The team decided to examining the difference between the p-p interaction of peptide composed by only ordinary residues and those containing PRO. Three peptide sequences are designed for this aim. One is polyA peptide which contains eight alanine (ALA) residues. The other two are, respectively, polyP (poly PRO residues with length of eight) and poly(AP)$_4$ (with four dimers of A-P in peptide). Potential data were collected after running these sequence under the full force field. As Figure 4-B shows, for peptide polyA, largest amount of H-bond should be formed hence outputting a lower level of p-p potential ($\sim$ 12kcal/mol). The polyP peptide form least amount of H-bond, which can be confirmed by its highest potential ($\sim$ 4kcal/mol). Besides, poly(AP)$_4$ yields a p-p potential level between the other two lines, which means it generate H-bonds less than polyA but more than polyP. All of these results are consistent with

what were expected. Therefore, our p-p potential function was also inspected.

3.2 Prediction of α Helix

6B17 is a 12-residue protein (sequence: TPRQARAARAAC) consisting of a single alpha helix structure. The simulation on has been smooth and shown some promising results. The following figure describes an output of a minimum-energy conformation:

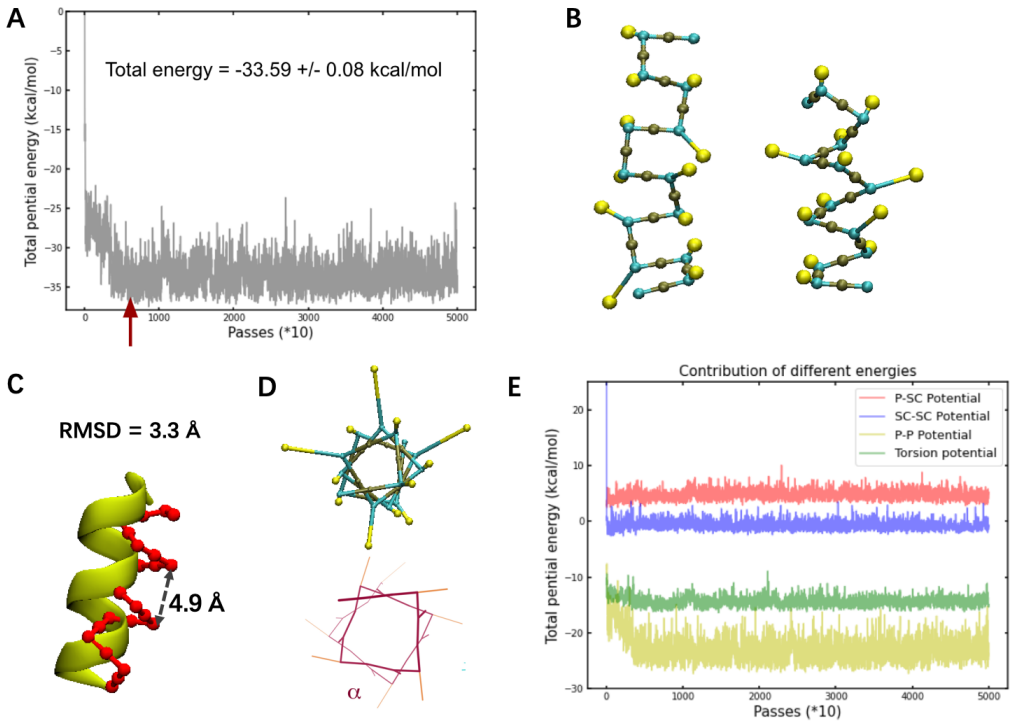

Figure 4: A) Total energy plots from the whole simulation, the red arrow indicates the minimum energy, where the structure is picked as an output of the prediction B) The conformations of input and output C) Final protein conformation from MC simulation (red) compared to known wild structure (yellow). RMSD is measured as 3.3Å and the distance between two turns is measure as 4.9Å (an averaged value from several distances) D) Top-down view comparing the simulated alpha helix (top) and a known alpha helix structure (from Wikipedia) E) Contribution of each potential to the total energy.

The simulation has run 50000 passes over 44 minutes. In Figure 4-A, the total energy of the whole process is given as -33.59 ± 0.08 kcal/mol (mean value $\pm$ estimated error). The protein conformation at the minimum energy was given as an output of folded structure prediction. As Figure 4-B show, the output successfully predicted the structure of the α helix. The predicted structure was compared with the original PDB coordinates, which gives a root-mean-square-deviation (RMSD) of 3.3Å. We also measure a parameter of the predicted α helix, which is the distance between two turns (pitch of the α helix) . This parameter is given as 4.9Å which is around

0.5Å difference from the literature-reported value 5.4Å [**10**] (Figure 4-C). The top-down view of the predicted structure also shares a similarity with the same view of natural helix structure, from which we can also conclude that the α helix structure has been successfully predicted (Figure 4-D). From Figure 4-E, we can see SC-SC potential contributes most to the result. However, this is thought to be a random contribution because of the poor initial conformation of the input. A gradual decrease of p-p potential, instead, is considered to provide the most valuable contribution to the whole process. This may suggest an equilibration stage where the H-bonds generate between the peptide group and finally assists the formation of α helix.

To further check if the output structure depends on the initial conformation, several initial conformations with different initial torsion angles inputs have been tried. As shown in the Figure 5, output results were similar – the program works stably for

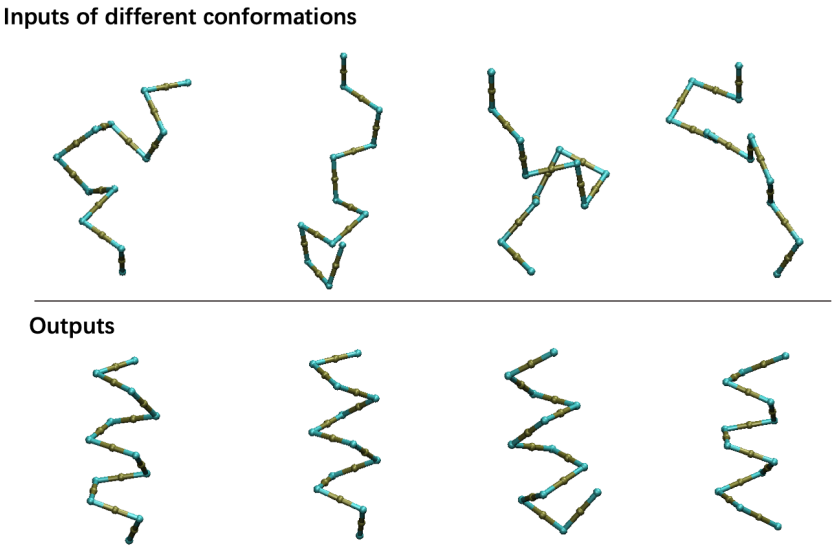

Figure 5: Multiple initial conformation vs. final output

varied input for this specific type of protein.

3.3 Prediction of β Hairpin

2EVQ is a 12-residue protein (sequence: KTWNPATGKWTE) consisting of a beta hairpin structure. The simulation was able output a final conformation but it does not match the beta hairpin structure of the protein (Figure 6-A) [**7**].

The team came up with two ideas: it was either an issue of the sampling method or the force field. An improper sampling method could trap the conformation in a local minimum-energy region instead of a global minimum-energy region. On the other

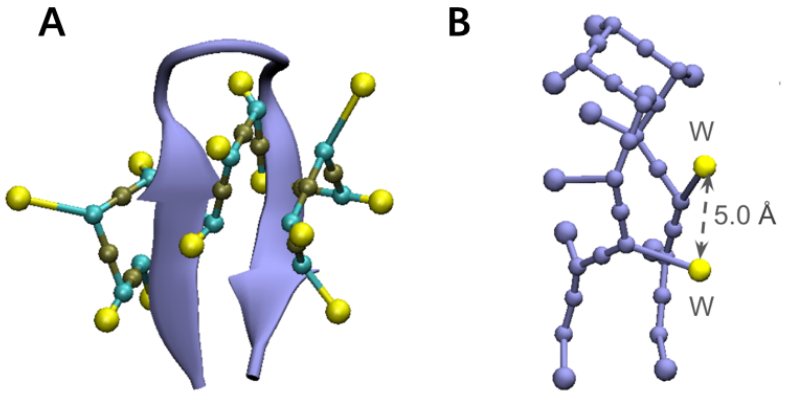

Figure 6: A) Direct comparison between the actual and simulated protein. The conformations do not match; B) Applying force field to natural structure. The distance between the TRP-TRP side chain is measured as 5.0Å

hand, a force field with deficiency can also gives completely different minimum-energy structure. A simple test, therefore, has been conducted to figure out which one might be the potential explanation. All the atoms in the natural structure are mapped to the list of beads and then an energy level corresponding to the natural structure under the force field is calculated. As a result, a high energy level is obtained. The most contribution (121 kcal/mol) comes from the SC-SC interaction between residue 3 and residue 10, which are two TRP residues. In natural structure, the distance between the two side chain is as small as 5.0Å, which is smaller than Van der Waals radii defined as 7.2Å in the force field. The 12-6 potential thereby becomes extremely high compared to the energy level that can be accepted in our MC simulation. (In our simulation the minimum energy is around -18 kcal/mol) This suggests the force field may not describe the β-hairpin structure on 2EVQ properly.

4 Discussion and Conclusion

In this project, the team implement MC algorithm to simulate the folding process of from a sequence to a folded structure. An united-atom model that simplified polypeptide chain was used instead of the all-atom model. The intermolecular interactions of residues were simulated by a force field, consist of hydrophobic/hydropholic interactions of side chains, a potential that prevents collapse of side chains on the peptides, average electrostatic interactions of peptides (mainly accounts for the hydrogen bonding) and energies involved with torsional angles of between C^{α}s.

Simulations were conducted on two protein sequence: 6B17 and 2EVQ, whose natural structures are α helix and β hairpin. For 6B17, the program was able to predict the secondary structure with acceptable error. However, our program did not

produce the accurate conformation for 2EVQ. We have narrowed down the factor to the force field – this algorithm failed to picture the hydrophobic interactions in this specific protein. As an united-atom model, the force field may not be able to explain some special scenarios but just an averaged interaction over all the conformations.These deficiency may also be noticed by the developers of the force field, given that the model is only proposed for the very beginning stage [**5**] of protein folding prediction. In Liwo's plan, they would eventually produce a result for a full-atom model. This model is also further revised into a more efficient one known as UNRES. [**2**]

Prediction of the final conformation was only one aspect of studying the protein folding mechanism. There are other approaches, such as Molecular Dynamic, which might even track the trajectories of atoms during the process and reveal even more secrets of this topic. In summary, this project, although was conducted in a simplified model under many assumptions, allowed us to gain insight of an astonishing way to predict protein folded structures.

References

[1] Levitt, M. and Warshel, A. *Computer simulation of protein folding.* Nature volume 253: 694–698 (1975)

[2] Liwo, A., Baranowski, M., Czaplewski, C., Gołaś, E., He, Y., Jagieła, D., . . . Zaborowski, B. *A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions.* Journal of Molecular Modeling, 20(8). (2014). doi:10.1007/s00894-014-2306-5

[3] Kolinski, A. *Protein modeling and structure prediction with a reduced representation.* Acta Biochimica Polonica, 51(2), 349-371. (2004). doi:10.18388/abp.2004_3575

[4] Kolinski, A., Jaroszewski, L, Rotkiewicz, P, and Skolnick, J *An Efficient Monte Carlo Model of Protein Chains. Modeling the Short-Range Correlations between Side Group Centers of Mass.*

[5] A. Liwo, M.R. Pincus, R.J. Wwawk, S. Rackovsky, and H.A. Scheraga. *Prediction of protein conformation on the basis of a search for compact structures: Test on avian pancreatic polypeptide.* Protein Science (1993). 2, 1715-1731

[6] Wu, H., Acharyya, A., Wu, Y., Liu, L., Jo, H., Gai, F., and DeGrado, W.F. *Design of a Short Thermally Stable alpha-Helix Embedded in a Macrocycle.*

[7] Andersen, N. H., Olsen, K. A., Fesinmeyer, R. M., Tan, X., Hudson, F. M., Eidenschink, L. A., amp; Farazi, S. R. *Minimization and Optimization of Designed*

β-Hairpin Folds. Journal of the American Chemical Society, 128(18), 6101-6110. (2006). doi:10.1021/ja054971w

[8] Miyazawa, S. and Jernigan, R.L . *Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation.* Macromolecules 18, 534-552 (1985).

[9] Levitt, M. *A simplified representation of protein conformations for rapid simulation of protein folding.* J. Mol. Biol. 104, 59-107.

[10] McCreath, S. B., amp; Delgoda, R 24.3.2.1 α Helix.in *Pharmacognosy: Fundamentals, applications and strategies.*. J. Mol. Biol. 104, 59-107. Amsterdam: Academic Press. (2016).

Appendix

Relative Equations

$$\begin{aligned}\epsilon_{\text{GlyGly}} &= 0.025\text{kcal/mol} \\ &= 0.6 \times (\varepsilon_{ij}^{MJ} - \varepsilon_{ij}^{0}) \quad \text{for } ij \neq \text{GlyGly}\end{aligned} \tag{A1}$$

where values of ε_{ij}^{MJ} are in Table A3.

$$\varepsilon_{ij}^{0} \begin{cases} \varepsilon_{\text{GlyGly}}^{MJ} & \text{if both } i \text{ and } j \text{ are not prolines} \\ \text{f}_{\text{Pro}}\varepsilon_{\text{GlyGly}}^{MJ} & \text{if one of the residues is proline} \\ \text{f}_{\text{ProPro}}\varepsilon_{\text{GlyGly}}^{MJ} & \text{if both residues are prolines} \end{cases} \tag{A2}$$

Angles α_{ij}, β_{ij}, and γ_{ij} are defined for the relative orientation of the two virtual bonds as:

$$\cos\alpha_{ij} = \text{v}_i \cdot \text{v}_j \quad \cos\beta_{ij} = \text{v}_i \cdot \text{e}_{rij} \quad \cos\gamma_{ij} = \text{v}_j \cdot \text{e}_{rij} \tag{A3}$$

where v_i and v_j are unit vectors pointing from C_i^α to C_{i+1}^α and C_j^α to C_{j+1}^α, e_{rij} is the unit vector pointing from p_i to p_j.

Parameter Values

Table A1: Residue Standard Geometries

Residue	b_{SC} (Å)	θ_{SC} (deg)	ϕ_{SC} (deg)	r^0 (Å)
CA	3.8	90.0	0.00	0.00
C	1.38	120.7	-148.5	5.0
M	2.34	120.5	-154.3	6.2
F	2.97	125.6	-154.2	6.8
I	1.83	125.2	-138.8	6.2
L	2.08	125.4	-152.1	6.3
V	1.49	127.7	-135.9	5.8
W	3.58	125.8	-154.2	7.2
Y	3.56	117.8	-163.4	6.9
A	0.77	125.3	-111.9	4.6
G	0.00	0.00	0.00	3.8
T	1.43	122.5	-129.8	5.6
S	1.28	122.7	-124.1	4.8
Q	2.58	125.3	-152.4	6.1
N	1.98	124.7	-150.6	5.7
E	2.63	124.9	-143.5	6.1
D	1.99	127.7	-141.7	5.6
H	2.76	124.3	-136.8	6.2
R	3.72	128.6	-150.7	6.8
K	2.94	128.9	-146.0	6.3
P	1.42	86.3	-123.3	5.6

Table A2: Values of constants in Equation 5

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

Table A3: Contact Energies in *RT* Units

	C	M	F	I	L	V	W	Y	A	G	T	S	Q	N	E	D	H	R	K	P
C	-5.44	-5.05	-5.63	-5.03	-5.03	-4.46	-4.76	-3.89	-3.38	-3.16	-2.88	-2.86	-2.73	-2.59	-2.08	-2.66	-3.63	-2.70	-1.54	-2.92
M	0.70	-6.06	-6.68	-6.33	-6.01	-5.52	-6.37	-4.92	-3.99	-3.75	-3.73	-3.55	-3.17	-3.50	-3.19	-2.90	-3.31	-3.49	-3.11	-4.11
F	0.52	-0.22	-6.85	-6.39	-6.26	-5.75	-6.02	-4.95	-4.36	-3.72	-3.76	-3.56	-3.30	-3.55	-3.51	-3.31	-4.61	-3.54	-2.83	-3.73
I	0.80	-0.18	0.14	-6.22	-6.17	-5.58	-5.64	-4.63	-4.41	-3.65	-3.74	-3.43	-3.22	-2.99	-3.23	-2.91	-3.76	-3.33	-2.70	-3.47
L	0.59	-0.09	0.06	-0.16	-5.79	-5.38	-5.50	-4.26	-3.96	-3.43	-3.43	-3.16	-3.09	-2.99	-2.91	-2.59	-3.84	-3.15	-2.63	-3.06
V	0.73	-0.02	0.14	0.00	-0.01	-4.94	-5.05	-4.05	-3.62	-3.06	-2.95	-2.79	-2.67	-2.36	-2.56	-2.25	-3.38	-2.78	-1.95	-2.96
W	0.67	-0.63	0.12	0.19	0.11	0.13	-5.42	-4.44	-3.93	-3.37	-3.31	-2.95	-3.16	-3.11	-2.94	-2.91	4.02	-3.56	-2.49	-3.66
Y	0.60	-0.12	0.25	0.25	0.41	0.19	0.04	-3.55	-2.85	-2.50	-2.48	-2.30	-2.53	-2.47	-2.42	-2.25	-3.33	-2.75	-2.01	-2.80
A	0.59	0.29	0.31	-0.05	0.19	0.10	0.03	0.18	-2.51	-2.15	-2.15	-1.89	-1.70	-1.44	-1.51	-1.57	-2.09	-1.50	-1.10	-1.81
G	0.64	0.37	0.79	0.55	0.56	0.50	0.43	0.36	0.19	-2.17	-2.03	-1.70	-1.54	-1.56	-1.22	-1.62	-1.94	-1.68	-0.84	-1.72
T	0.70	0.16	0.52	0.23	0.33	0.38	0.26	0.15	-0.04	-0.09	-1.72	-1.59	-1.59	-1.51	-1.45	-1.66	-2.31	-1.97	-1.02	-1.66
S	0.61	0.22	0.61	0.42	0.48	0.42	0.50	0.21	0.11	0.13	0.00	-1.48	-1.37	-1.31	-1.48	-1.46	-1.94	-1.22	-0.83	-1.35
Q	0.43	0.30	0.56	0.33	0.25	0.24	-0.01	-0.31	0.00	-0.01	-0.29	-0.18	-0.89	-1.36	-1.33	-1.26	-1.85	-1.85	-1.02	-1.73
N	0.93	0.33	0.67	0.91	0.70	0.91	0.39	0.10	0.61	0.32	0.14	0.23	-0.13	-1.59	-1.43	-1.33	-2.01	-1.41	-0.91	-1.43
E	1.23	0.43	0.50	0.47	0.58	0.49	0.36	-0.06	0.34	0.45	0.00	-0.15	-0.30	-0.04	-1.18	-1.23	-2.27	-2.07	-1.60	-1.40
D	0.54	0.61	0.59	0.68	0.79	0.71	0.28	0.01	0.16	-0.05	-0.32	-0.23	-0.33	-0.06	-0.161	-0.96	-2.14	-1.98	-1.32	-1.19
H	0.48	1.11	0.21	0.75	0.44	0.48	0.08	-0.17	0.55	0.53	-0.06	0.19	-0.02	0.18	-0.29	-0.26	-2.78	-2.12	-1.09	-2.17
R	0.71	0.23	0.58	0.48	0.43	0.38	-0.16	-0.28	0.44	0.10	-0.42	0.21	-0.72	0.07	-0.79	-0.80	-0.04	-1.39	-0.06	-1.85
K	1.11	-0.15	0.53	0.34	0.20	0.45	0.15	-0.31	0.08	0.18	-0.23	-0.15	-0.65	-0.19	-1.08	-0.90	0.24	0.57	0.13	-0.67
P	0.40	-0.49	0.29	0.23	0.42	0.10	-0.36	-0.43	0.03	-0.04	-0.21	-0.02	-0.69	-0.04	-0.22	-0.11	-0.19	-0.56	-0.15	-1.18

Table A4: Torsion Parameters

Residues	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	a_5	b_5	a_6	b_6
G-G	0.2358	0.0000	-0.0231	0.0000	0.1092	0.0000	0.0002	0.0000	-0.0660	0.0000	-0.0908	0.0000
A-G	-0.0664	0.1956	-0.3537	0.1232	-0.0296	0.0110	-0.1626	0.0854	-0.0928	0.0444	-0.0991	-0.0070
P-G	0.0067	0.1432	-0.3496	0.2297	0.2691	0.1251	-0.1589	-0.0090	-0.0608	0.0057	-0.0410	-0.0109
G-A	-0.0518	-0.6427	-0.4497	-0.1558	-0.1150	-0.0611	-0.2401	0.0307	-0.0272	-0.0048	-0.0379	-0.0435
A-A	0.5361	-0.4189	-0.0855	-0.3691	-0.1473	-0.2484	-0.2823	0.1008	-0.0952	0.1487	-0.0750	0.0532
P-A	0.5853	-0.5579	-0.1328	-0.1436	0.0579	-0.1318	-0.2751	0.0284	0.0261	0.0682	-0.0833	-0.1180
G-P	-0.3479	0.9326	-0.5485	-0.3216	0.9967	-0.0065	-0.2566	0.4761	-0.0892	-0.2608	0.3671	-0.0289
A-P	8.1793	5.7038	-0.8413	0.0401	3.3488	-3.2615	1.5991	0.5468	-1.2076	-0.6290	0.0894	-2.1917
P-P	3.2742	-1.2914	1.2540	-2.6647	1.0881	-1.9609	-0.3420	-0.6935	0.5555	-0.7774	0.1951	-0.2442