High Temperature Unfolding and Low Temperature Refolding Pathway of Chymotrypsin Inhibitor 2 Using Molecular Dynamics Simulation

N D Malau¹ and T Sumaryada¹²
¹Computational Biophysics and Molecular Modeling Research Group (CBMoRG), Department of Physics, Bogor Agricultural University, Jalan Meranti Kampus IPB Dramaga Bogor 16680, Indonesia
²Biopharmaca Research Center, Bogor Agricultural University, Jalan Taman Kencana No. 3, Bogor 16128, Indonesia

E-mail : tsumaryada@ipb.ac.id

Abstract. The mechanism that explains the unfolding/refolding process of the protein is still a major problem that has not been fully understood. In this paper we present our study on the unfolding and refolding pathway of Chymotrypsin Inhibitor 2 (CI2) protein through a molecular dynamics simulation technique. The high temperature unfolding simulation were performed at 500 K for 35 ns. While the low temperature refolding simulation performed at 200 K for 35 ns. The unfolding and refolding pathway of protein were analysed by looking at the dynamics of root mean squared deviation (RMSD) and secondary structure profiles. The signatures of unfolding were observed from significant increase of RMSD within the time span of 10 ns to 35 ns. For the refolding process, the initial structure was prepared from the structure of unfolding protein at t=15 ns and T=500 K. Analysis have shown that some of the secondary structures of CI2 protein that have been damaged at high temperature can be refolded back to its initial structure at low temperature simulation. Our results suggest that most of α-helix structure of CI2 protein can be refolded back to its initial state, while only half beta-sheet structure can be reformed.

1. Introduction

Unfolding and refolding of protein are rapid event which involve complex natural processes. The simplest unfolding pathway is two-states model, in which the denatured and native states are separated distinctly at transition point.

Chymotrypsin inhibitor 2 (CI2) was among the first protein that demonstrate the two-state folding mechanism, and it has since been the focus of a various number of experimental and theoretical studies [1,2,3,4,5]. CI2 protein consist of 64-residue with one α-helix and a three-stranded β-sheets. The main hydrophobic core of this protein is formed by the packing of the α-helix against the β-sheet [2]. The unfolding of CI2 was first characterized by simulation at high temperature of 498 K [3]. The use of artificially high temperatures simulation causes the unfolding process occurs more quickly without changing the unfolding pathway too much [4] and significantly reduce the usage of computational resources. Similar to unfolding, the protein refolding is also a
complex process which bring the unfolded protein back to its native state. In the experiment, protein refolding was performed at low temperatures in order to reduce protein aggregation [6].

In this paper we present our study on unfolding and refolding pathway of GB1 protein. The unfolding simulation were performed at 300 K, 400 K and 500 K, while the refolding process were done at 200 K. Both folding and refolding simulation were run for 35 ns.

![Figure 1. The initial structure of CI2 from Protein Data Bank (PDB).](image)

2. Materials and Methods
The crystal structure of Chymotrypsin inhibitor 2 (CI2) (pdb code : 1YPC) [8] that used in this simulation was taken from the Protein Data Bank (PDB) with 1.7 Å resolution (figure 1). This original file still contains of water molecules and ligand (glycerol) which need to be removed. Protein was prepared in a simulation box with the size of 60 x 60 x 60 Å³ filled with the solvent atoms, TIP3P [9]. To determine the potential energy of the crystal structure, CHARMM++ force field was used in this simulation. Neutralizing the total charge of the system was done by adding four Cl⁻ ion to water box. All those preparations were done by using Virtual Molecular Dynamics program (VMD) [10].

The next step is MD simulation which consists of several steps i.e minimization, heating, equilibration and production run. All MD steps were performed using NAMD v.2.9 simulator [11]. Molecules that have been solvated then must be minimized. The purpose of minimization is to avoid the van der Waals contacts that do not match (bad contact) and to minimize the steric effects that favor a high energy. Each simulation was started with a minimization for 100 ps to put the protein in its lowest energy. The second stage is heating and equilibration. Protein was heated from the temperature of 0 K up to 300 K, 400 K, 500 K for the unfolding process and 200 K for the refolding process, for 40 ps. Equilibration was done using Langevin protocol for 10 ps. Production run is the final stage of MD simulations where the constraints that have been applied to the equilibration process is removed so that protein now free to move. Production run performed for 35 ns for 300 K, 400 K and 500 K simulation. For the refolding process, production run were performed at 200 K for 35 ns as well.

The time step for MD simulations was set at 2.0 femto seconds (fs). The simulation was done by using periodic boundary condition (PBC) method to eliminate the effect of surface tension and to achieve density and pressure conditions were more uniform. Electrostatic energy system thoroughly calculated using the particle mesh ewald (PME) method, while the van der Waals interaction was calculated using the Lennard–Jones potential with 12 Å cutoff. All simulations was performed in single CPU powered by 3.4 GHz Intel® core i7 processor with 12 GB of RAM and using Ubuntu 12.04 Linux platform.

To analyze the unfolding and refolding pathway, VMD program was used to produce the output such as secondary structure and Root Mean Square Deviation (RMSD).
3. Result and Discussions

3.1 Unfolding

3.1.1 The Stability of the Secondary Structure
The stability of secondary structure during simulation plays an important role in maintaining the folded/unfolded state. The loss or damage of α-helix and β-sheet can be used as an indicator of unfolding or refolding event. The sequence of structural changes (or damages) of protein during the simulation will reveal the mechanism and the most responsible interaction that keep the protein stable.

![Graphs showing the percentage change in the composition of secondary structure during the simulation at 300 K (Black), 400 K (Blue) and 500 K (Red) for 35 ns for (a) beta-sheet structure. (b) α-helix structure. (c) Secondary structure changes in 500 K simulation for 50 ns; α-helix (Purple), β-sheet (Yellow) 3_10 helix (Blue), Turn (Green) and Coil (White).]

The composition of secondary structures, alpha-helix and beta-sheet, at temperature 300 K and 400 K has indicated a relatively stable and rigid structure. About 30% of the protein structure is in the form of beta-sheet and 18% α-helix. Whereas, at temperature of 500 K the composition of beta-sheet and α-helix were dramatically decreased. At this temperature beta-sheet is completely disappeared and
turned into coil and turn, while α-helix only slightly damaged. This found is similar with the results from other groups [1,2].

The composition of α-helix varies wildly at T=500 K indicates a very mobile and flexible structure. The α-helix structure at this temperature folds and refolds several times as also suggested by other [2]. The damage of secondary structure was initially occurred in β1 (beta-sheet number 1) at 10 ns, in β2 at 14 ns and followed by α-helix at 30 ns. Finally, the structure of β-sheet and α-helix will completely disappear at 35 ns.

3.1.2 Root Mean Square Deviation (RMSD)
Root Mean Square Deviation analysis is used to describe the change in secondary structure during the simulation. The RMSD change above 5.0 Å is commonly used as an indicator of significant conformational changes in protein structure.

Figure 3 shows the fluctuations RMSD values at various temperatures. At simulation temperature of 300 K and 400 K, there is no significant change in RMSD values. At temperature 500 K the RMSD is relatively constant until 9.0 ns and jumps drastically from 10 ns until 35 ns with several plateaus. Those plateaus (12 to 14 ns, 14 to 17 ns, 24 to 27 ns and 27 to 30 ns) might correspond to the dynamics of the tertiary structure exhibited by fluctuation of α-helix and the disruption of beta-sheet as the simulation progressed.

![Figure 3](image)

Figure 3. Root Mean Square Deviation (RMSD) as a function of simulation time (a) for temperature 300 K, 400 K, 500 K

3.2 Refolding
3.2.1 The Stability of the Secondary Structure
Studying the detailed pathway of refolding simulation from completely unfolded structure seems impossible and takes a very long time, but we can instead simulate the refolding event by starting from a transition state (TS) conformation [7]. Proteins were thought to fold not from a random coils or completely unfolded tertiary structure, but from a highly unfolded protein with some residual structures [12]. Based on that argument we choose CI2 unfolded structure at 15 ns as the initial structure for the refolding process. The low temperature of refolding was chosen to avoid aggregation of the protein as suggested by the experiment [6]. Similar to unfolding, the simulation of refolding will also be analysed from RMSD and secondary structure data.
Unfolding-refolding process as can be seen in figure 4 is represented by protein conformational changes during the simulation. Unfolding process begins with the change of beta-sheet structure into a coil and followed by the change of alpha-helix into a turn. In that process beta-sheet disrupted first and followed by alpha-helix. This unfolding simulation is in line with simulation from other groups [1,2]. After 35 ns all secondary structure was disappeared and transform into a completely random coil.

For refolding process, the rapid formation of alpha-helix precedes the formation of beta-sheet structure. At the end of refolding process, alpha-helix has been formed completely, while only some parts of beta-sheet structure reformed. Similar results have been reported by other researcher [2]. A Longer simulation time probably is needed here to obtain a completely folded structure of protein.

3.2.2 Root Mean Square Deviation (RMSD)
The initial structure for refolding process was taken from the unfolded structure at 15 ns with RMSD value about 5.1 Å. Figure 5 shows the fluctuation of RMSD values at various temperatures, at 300 K for the native state, 200 K for the refolding process, and 500 K for the unfolding process. For the native state, the average value of RMSD was 1.26 Å, which means that the protein was still in stable conditions. For the unfolding process, the RMSD value jumps drastically from 1.26 Å to 11.4 Å which indicates a transition from a folded to an unfolded structure. For the refolding process, the RMSD is slowly dropped from 5.1 Å to 3.06 Å, indicating a slow process of secondary structures reformation.
4. Conclusion

Analysis of molecular dynamics simulation have shown that CI2 protein that has been unfolded at high temperature can be refolded back to its native state at low temperature simulation. The unfolding event in this simulation is characterized by a complete disruption of beta-sheet and a fluctuation of α-helix structure which also found by others [1,2]. Using a transition state conformation which consist of a small portion of residual beta-sheet and α-helix structures (obtained from the unfolding process), the refolding simulation is able to reform the whole α-helix and some part of beta-sheet structure. A longer refolding simulation time would probably allow a complete description of folding process of CI2 protein.

References

[1] Li A and Daggett V 1994 Proc. Natl. Acad. Sci. USA. 91(22) 10430
[2] Kazmirski S L, Wong K, Freund S M V, Tan Y, Fersht A R and Daggett V 2001 Proc. Natl. Acad. Sci. USA. 98 (8) 4349
[3] Daggett V, Li A, Itzhaki L S, Otzen D E and Fersht AR 1996 J. Mol. Biol 257 (2) 430
[4] Day R, Bennion B J, Ham S and Dagget V 2002 J. Mol. Biol. 322 (1) 189
[5] Lindgren M, Westlund P O 2010 Biophysical Chemistry. 151 46
[6] Yamaguchi S, Yamamoto E, Mannen T and Nagamune T 2013 Biotechnol. J. 8 (1) 18
[7] Pande V S and Rokhsar A 1999 Proc. Natl. Acad. Sci. USA. 96 (16) 9062
[8] Harpaz Y, Elmasry N, Fersht A R and Henrick K 1994 Proc. Natl. Acad. Sci. USA. 91 (1) 311
[9] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 J. Chem. Phys. 79 926
[10] Humphrey W, Dalke A and Schulten K 1996 J.Molec.Graphics 14 33
[11] Phillips J C, Braun R, Wang W, Gunderman J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kalé L and Schulten K 2005 Journal of Computational Chemistry, 26 (16) 1781
[12] Pappu R V, Srinivasan R and Rose G D 2000 Proc. Natl. Acad. Sci. USA 97 12565