Elucidation of GB1 Protein Unfolding Mechanism via a Long-timescale Molecular Dynamics Simulation

T Sumaryada1, 2*, J Hati1, S T Wahyudi1, N D Malau1 and K N Sawitri1
1Computational Biophysics and Molecular Modelling 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. This study investigates the unfolding mechanism of 1GB1 protein at various simulation temperatures using a long-timescale molecular dynamics simulation. Analysis of structural parameters of molecular dynamics simulation have indicated that the unfolding process of GB1 protein has started at 95 ns for 475 K simulation, and at 745 ps for 500 K simulation. The unfolding process in this simulation exhibit the feature of hydrophobic core collapse model, in which the beta-hairpin destruction precedes the α-helix to coil transition. The unfolding was started with the increasing flexibility of the beta-sheets and hydrophobic core region, continued with beta-hairpins destruction, and ended with α-helix to coil and turn transition. The final structures of GB1 protein after unfolding, suggest an unfinished denaturation of protein as seen from the small remains of α-helix structure.

1. Introduction
The folding/unfolding mechanism of protein is still one of the hot topic in life sciences. The ability to predict the pathway of folding/unfolding event and mechanism of protein could lead to the breakthrough and revolution in biotechnology, especially in biomedical sciences and drug design engineering. Molecular dynamics (MD) simulation that basically rely on the first principal of physics laws can be used to simulate the folding/unfolding process of proteins and elucidate its mechanism. The typical time scale of folding and unfolding event in nature usually in microseconds or milliseconds time frame. To realistically mimic that event through a computer simulation, a computationally expensive process and time consuming effort must be done. This burden sometimes hinder scientists from using a full microseconds or milliseconds time frame simulation. However, the advance of computational algorithm, computer software and hardware nowadays has make long-timescale MD possible [1].

B1 domain of protein G (pdb code 1GB1) [2] is one of the small proteins which extensively studied to unveil the folding/unfolding mechanism of protein. The small number of residues (56 residues) and a highly symmetric structure has made 1GB1 protein as a as one of the favorite choices for scientists to study the folding/unfolding process and to unveil the most responsible interactions in maintaining protein in its native state. The high thermostability of beta hairpin fragment of GB1 protein have been well studied using MD simulation [3, 4-7]. The simulation of a full 56-residues of GB1 protein has also been done by some researchers [8-12]. Unlike beta-hairpin, simulation of a full
GB1 protein is more difficult and complex due to more secondary structures and molecular interactions involved.

Two models that commonly used in explaining the folding/unfolding mechanism of protein, the zipper model, and the hydrophobic core collapse model. In the zipper model [13] the folding process is dominated by hydrogen bond force [7, 14, 15], while in the hydrophobic collapse model the folding process is dominated by the hydrophobic force from the hydrophobic residues buried in the core of protein [5,17].

In this research, long-timescale (sub-microseconds) molecular dynamics simulations of GB1 protein have been conducted at three different temperatures, 450 K, 475 K, and 500 K for 0.10 μs to determine the pathway process of unfolding event and to reveal the unfolding mechanism of a GB1 protein. Some parameters that were analyzed including Root Mean Squared Deviation (RMSD), Root Means Squared Fluctuations (RMSF), secondary structure, and cartoon representation (snapshot) of the protein.

2. Materials and Methods
There are three important parts in this research. The first part is protein data preparation, the second is running the molecular dynamics simulation, and the last part is data analysis.

2.1. Preparation of protein file
The structure of B1 domain of protein G was taken from the Protein Data Bank with pdb code name 1GB1. This original file still contains of predicted coordinates of hydrogen atoms which have to be removed. Protein must be prepared in a simulation box filled with the solvent atoms, TIP3P water box [18] with the size of 80x80x80Å³. The sodium and chloride counterions were used to neutralize the protein during simulation. All those preparations were done by using Virtual Molecular Dynamics program (VMD) [19].

2.2. Molecular dynamics simulations
The MD simulations were performed using NAMD v.2.9 simulator [20] with CHARMM++ used as the force field. The Particle Mesh Ewald (PME) and Periodic Boundary conditions (PBC) methods were used to calculate the interactions between atoms. Simulation can be divided into two parts. The first part is minimizations, heating and equilibration. Minimizations were run in four steps and are aimed to minimize the energy of the protein. After minimizations, protein was heated from 0 K to the final temperatures (three different simulations with three different final temperatures i.e 450 K, 475 K and 500 K) then equilibrated at constant temperature using Langevin protocol. The second part is production run, where protein held at constant temperature for 100 ns. 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..

2.3. Data analysis
For the data analysis, the MD output (dcd type of files) were analyzed using VMD to produce the output such as RMSD, RMSF, secondary structures, and cartoon representation of protein structure. Some of the MD data need to be smoothed first using moving average method to be properly displayed in this paper. The data obtained after the unfolding event were cut and not included in the analysis.

3. Results and Discussions
The dynamics of protein structure and conformation can be observed through the RMSD plot. The RMSD change above 2.50 Å is commonly used as an indicator for significant conformational changes in protein structure, and RMSD above 8.0 Å indicates an unfolding event. For 450 K simulation, in general there was no significant change in RMSD occured, with the average RMSD lied within 1.0 Å to 2.0 Å range, except for some fluctuations at 33 ns and 65 ns (see figure 1(a)). For 475 K simulation,
a significant structural change occurred within 80.0 to 90.0 ns simulation time, and followed by a drastic change of RMSD at 95 ns (see figure 1(a)). This drastic change of RMSD up to 12 Å indicates a possible unfolding event happened at 475 K. For 500 K simulation the RMSD value jumps drastically from 1.0 Å to 9.5 Å at 745 ps which indicates the transition from a folded to an unfolded structure as seen in figure 1(b). This sharp transition is commonly expected from a typical two-state folding kinetics [13]. Notice that the result of RMSD at T=500 K is shown in the separate plot due to different time scaling.

![Figure 1](image1.png)

**Figure 1.** RMSD as a function of simulation time for (a) simulation temperature at 450 K and 475 K and (b) at 500 K.

The RMSF value shows the flexibility of each residue during the simulation. The RMSF of each residue is consistently increased as the temperature is raised. At 450 K the flexible residues were mostly found in coil (M1,E56) and turn (T11, A48) regions. At 475 K more flexibility were found in coil (M1, D22, G38, G41, E56) and in turn (T11, A48, K50) as seen from figure 2 (a). In contrast, when unfolding occurred at 500 K the flexibility were shifted to secondary structures, α-helix (E27, V29, Y33, D36), and beta-sheet (L5, I6, L7, G14, T16, E42) regions (figure 2(b)).

![Figure 2](image2.png)

**Figure 2.** RMSF as a function of simulation time for (a) simulation temperature at 450 K and 475 K and (b) at 500 K.

The increasing flexibility, or a rigidity weakening, of secondary structures indicates the beginning of unfolding process. We also found that, of five hydrophobic core residues in the beta-
hairpins region (L5, L7, G9, F52, V54), two of them (L5, L7) have shown a significant increasing of flexibility at transition point. This found emphasizes the domination of hydrophobic core collapse process and beta-hairpins destruction upon helix to coil and turn transformation. The above results are in agreement with findings from the other groups [5,10-12,17].

The secondary structures of protein at various temperatures were shown in figure 3. There is no significant change observed at temperature of 450 K and 475 K. At the end of 475 K simulation, about one third of α-helix transform into a turn structure, while the beta-sheets remain intact. The unfolding signature of GB1 protein occurred at 500 K as indicated by the disappearance of beta-sheets and some of α-helix structure which transformed into 310-helix, coil and turn structures. The cartoon representation of protein secondary structures (figure 4) suggest a complete destruction (collapse) of the hydrophobic core in the beta-sheet region at T=500 K.

![Figure 3. Evolution of secondary structures at various simulation temperatures.](image-url)
4. Conclusions
The long-time scale molecular dynamics simulation of GB1 protein has been performed at various temperatures. The unfolding signatures were found at transition point (T=500K and t=745 ps) as indicated by sudden changes in RMSD and secondary structure profiles. Analysis of MD simulation results has shown that the unfolding transition exhibited a two-state folding kinetics as suggested by other groups [13, 21-23]. Based on RMSD value, the indication of unfolding also found at 95 ns at 475 K, but the secondary structure analysis did not support the existence of this event. The high value of RMSD at 475 K likely correspond to a transformation from a compact globular into an open tertiary structure as can be seen from figure 4.

The structural parameters changes such as RMSF, secondary structures, and cartoon representation of protein clearly indicate an unfolding process of GB1 protein by showing the increasing flexibility of secondary structures and the disappearance of beta-sheet and some parts of α-helix structure at transition point. The fact that at 500 K there are six the most flexible residues in beta-sheet (including two hydrophobic core residues) compare to four in α-helix, indicates that beta-sheet destruction process initiate and dominate the overall unfolding process of GB1 protein. Secondary structures evolution and cartoon representation of protein emphasize a clear support for the hydrophobic core collapse model, as can be seen from the disappearance of all beta-sheets and a small part of α-helix structure.

5. Acknowledgment
The authors express their gratitude for the funding from the Directorate of Higher Education, Ministry of Education and Culture Republic of Indonesia through BOPTN research grant contract number 237/IT3.41.2/L2/SPK/2013 code 2013.089.521219.

6. References

[1] Klepeis J L, Lindorff-Larsen K, Dror R O and Shaw D E 2009 Curr. Op. Struct. Bio. 19 (2) 120
[2] Gronenborn A M, Fillipula D R, Essig N Z, Achari A, Whitlow M, Wingfield P T, Clore G M 1991 Science. 253 657
[3] Pande V S and Rokhsar D S 1999 Proc. Natl. Acad. Sci. USA. 96 9062
[4] Roccatano D, Amadei A, Di Nola A, Berendsen H J, 1999 Protein Sci. 8 (10) 2130
[5] Lee J and Shin S 2001 Biophys. J. 5 2507

Figure 4. Cartoon representation of protein final structures at various temperatures

450 K at 100 ns 475 K at 95 ns 500 K at 745 ps
[6] Klimov D K and Thirumalai D, 2000 Proc. Natl. Acad. Sci. USA. 97 (6) 2544
[7] Ma B and Nussinov R 2000 J. Mol. Biol. 296 (4) 1091
[8] Bui J M, Gsponer J, Vendruscolo M, Dobson C M 2009 Biophysics Journal. 97 2513
[9] Kouza M, Hansmann U H E 2012 J. Phys. Chem. B 116 6645
[10] Ceruso M A, Amadei A and Di Nola A 1991 Protein Science 8 147
[11] Sheinerman F B and Brooks C L 1998 J. Mol. Biol. 278 (2) 439
[12] Sheinerman F B and Brooks C L 1997 Proteins Struct. Funct. Gen. 29 (2) 193
[13] Muñoz V, Thompson P A, Hofrichter J and Eaton W A 1997 Nature 390 196
[14] Zhou R, Berne B J and Germain R 2001 Proc. Natl. Acad. Sci. 98 (26) 14931
[15] Zagrovic B, Sorin E J and Pande V 2001 J. Mol. Biol. 313 (1) 151
[16] Tanford C 1980 The hydrophobic Effect 2nd ed (New York :Wiley)
[17] Garcia A E, Sanbonmatsu K Y 2001 Proteins 42 (3) 345
[18] Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 J. Chem. Phys. 79 926
[19] Humphrey W, Dalke A and Schulten K 1996 J. Molec. Graphics 14 33
[20] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kalé L and Schulten K 2005 Journal of Computational Chemistry, 26 (16) 1781
[21] Blanco F J, Rivas G and Serano L 1994 Nat. Struct. Biol. 1 584
[22] Alexander P, Orban J and Bryan P Biochemistry 31(32) 7243
[23] Chung H S, Louis J M and Eaton W A 2010 Biophys. J. 98 696