The Investigation of the Folding Pathway of Trp-Cage Miniprotein Using Explicit Solvent Molecular Dynamics Simulation

Nurul Bahiyah Ahmad Khairudin, and Fatahiya Mohamed Tap

Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia

ABSTRACT

The objective of this study is to investigate the folding pathway of Trp-cage miniprotein. The structure and trajectories of this protein has been studied using Molecular Dynamics (MD) simulation. The simulation was run at 300K for 250ns. Clustering analysis was conducted to group the trajectories according to the RMSD value and six clusters were generated. From this, the best conformation was identified to best represent the Trp-cage miniprotein. The formation of the hydrogen bond that involved Gly11-Ser14 assisted the formation of 3^{10}-helix. In this study, it is strongly suggested that the hydrogen bond interactions determined the formation of secondary structures.

KEYWORDS: Molecular dynamics, protein folding, Trp-cage miniprotein.

1. INTRODUCTION

Protein folding is a process where proteins spontaneously fold into their unique conformations determined solely by the amino acid sequences. As previously stated, the function of a protein is determined by its three dimensional structure, which forms from many interactions among amino acid residues. In order for protein to function, it is important for the protein to fold into its correct or native tertiary structure. The process of protein folding can be investigated using molecular dynamics simulation at sufficiently small time interval [1]. Trp-cage miniprotein (PDB id: 1L2Y) was developed with the main purpose to study the pathway of protein folding [2]. Previous study that investigated the 1L2Y folding process was conducted by Chowdury and coworkers[3]. The amino acid sequence of this protein is NLYIQWLKDGGPSSGRPPPS. Based on the work done by Chowdhury *et al* as comparison to the ensemble of the NMR structure, the heavy atom RMSD of the heavy atom for the last 60 ns simulation was found to be 2Å, whereas the pair-wise RMSD of the 38 models in the NMR was in the range of 1.6Å to 2.8Å. So, this fact highlights that the simulation done by Chowdhury and colleagues was excellent because it was close to the experimental result. They also claimed that, the hydrophobic force is the main driving force for the protein folding process while salt bridge are important in providing additional stability to the native structure. The hydrophobic contact between Trp25 and Leu21 residues was found to occur in the early state, whereas the salt bridge was observed after completing the formation of the C terminus end of the α-helix in about 15 ns.

2. METHOD

The protein 1L2Y was simulated for 250 ns using an all atom classical simulation and slightly modified version of AMBER and the force field is FF99SB [4]. The system was heated up from 0K to 300 K. The recommended method for maintaining the temperature is Berendsen Thermostat [5]. The final stage was to run the production of MD with a constant temperature and pressure. Neutralisation process was employed by adding chloride ion (Cl⁻) to the system. The protein was immersed in the octahedron water box of TIP3P water model. Bond constraints were imposed on all bonds involving hydrogen atoms via SHAKE. The nonbonded interactions were treated using 10 Å cut-off.

3. DISCUSSION

The fully extended structure of 1L2Y was initially constructed at 300 K with RMSD of 15 Å. Then, the RMSD value decreased to 10 Å within 3 ns of MD simulation. The RMSD values fluctuated between 5 Å to

---

* **Corresponding Author:** Nurul Bahiyah Ahmad Khairudin, Tap Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia. r-bahiah@utm.my, fatahiyamohdtap@yahoo.com

10 Å from 3 ns to 20 ns. At the beginning of the folding process, there was only one hydrogen bond formed. However, the occupancy of the hydrogen bond formation was less than 50% and the distance was around 3.8 Å; which can be considered as unstable. The formation of a turn was noted at the beginning of the simulation at 3 ns together with the formation of hydrogen bond (Gly11-Ser14). However, this hydrogen bond was not stable and finally collapsed at 10 ns to form a 3_{10}-helix (Gly10-Ser13). The visual observation showed that the 3_{10}-helix was unstable but the turn was present throughout the whole simulation.

There were no secondary structure formations within 26 ns towards 47 ns. However, at 48 ns, an extended beta sheet and hydrogen bonds (Gly11-Ser14) with distance of 4.0 Å were found to form.   At this time, the percentage occupancy of hydrogen bond formation was more than 70% and the distance decreased to 3.0 Å.   The stability of hydrogen bond depended on its lifetime occupancy and the distance of bond.   The stability of hydrogen bond (Gly11-Ser14) increased when the simulation reached 91 ns, and at 93 ns, the 3_{10}-helix formation was noted again.   Besides the formation of hydrogen bonds, there was also formation of salt bridge (Asp9-Arg16). The salt bridge occupied towards the end of the simulation. In agreement with other findings, Anand and Hansmann [6] and Rovó and co-workers [7] also reported that there was a formation of salt bridge at residues Asp9-Arg16. During the simulation at 101 ns, the formation of the hydrogen bond Gly11-Ser14 assisted the formation of 3_{10}-helix. In this study, it is strongly showed that the hydrogen bond interactions determined the secondary structure formation.

In investigating the 1L2Y folding pathway, the trajectories had to be observed from the beginning of the simulation towards the end of the simulation as shown in Figure 1. Figure 1 a, b, c, and d show the conformation of the structure at 40 ns, 80 ns, 135 ns, and 145 ns, respectively. The trajectory at 80 ns showed the formation of secondary structure with stable values of RMSD. This observation suggested that the secondary structure formation led towards the lowest and stable RMSD value. This is because, the formation of secondary structure, is supported by the strong hydrogen bond interaction and salt bridge formation. However, from the previous study done by Wu *et al* [8] they claimed that salt bridge does not play a significant role in the protein folding. However, in this study, the salt bridge formation assisted in stabilizing the structure since the RMSD value was stable at the time where the salt bridge was formed.
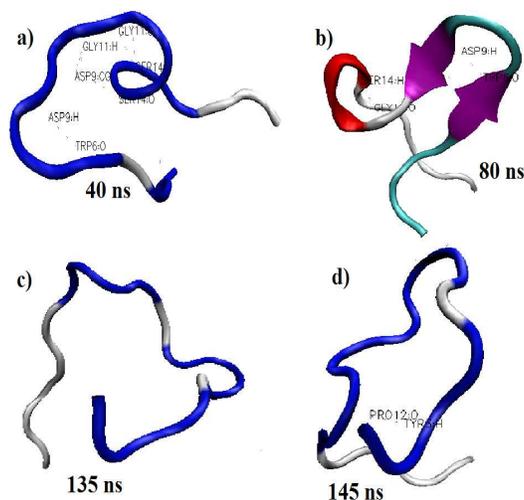
Figure 1:   The evolution of 1L2Y conformation throughout the 150 ns of simulation time. The conformation a, b, c, and d were extracted at 40 ns, 80 ns, 135 ns, and 145 ns, respectively. The red and purple ribbons represent 3_{10}-helix and beta-sheet, respectively.

These trajectories were further analysed in order to get the best structure. Therefore, clustering analysis was conducted to group the trajectories according to the RMSD value and six clusters were generated. The RMSD values of the centroid structure with reference to the NMR_{MD} structure, electrostatic energy, and van der Waals energy were obtained as shown in Table 1. Figure 2 shows the best structure represents each cluster. The SASA values were also considered in selecting the best cluster as presented in table 2. The lowest SASA

value reflects the most stable structure and more compact structure because the compact structure shows that the nonpolar amino acids aggregate together avoiding surrounding water. Figure 3 shows the surface presentation for the best structure from each cluster. Further inspection showed that cluster 1 could be classified as insignificant since it was formed at the early folding event. Besides that, the trajectory occupied the highest RMSD value (2.22 Å). Cluster 5 and 6 showed lower $RMSD_c$ than cluster 1, 2, 3, and 4; however the $RMSD_{c\text{-}NMRMD}$ values were high and they were extracted in the early stage of the folding process. Besides that, after superimposing cluster 5 and 6, both C-terminal and N-terminal of cluster 5 and 6 did not fit well with the NMR_{MD} structure.

Table 1:   Comparison of the best structural properties between each cluster

Cluster Number	Time (ns)	$RMSD_C$ (Å)	$RMSD_{C\text{-}best}$ (Å)	$RMSD_{best\text{-}NMRMD}$ (Å)	$RMSD_{C\text{-}NMRMD}$ (Å)	E_{eel}	E_{vdw}	G
1	29.39	2.22	1.24	5.14	4.79	-393.8	-24.1	-298.3
2	79.76	1.11	0.75	3.36	3.45	-466.0	-34.4	-321.7
3	144.08	1.29	0.87	5.24	4.98	-482.4	-19.9	-310.9
4	129.48	1.24	0.86	6.03	5.92	-458.4	-6.1	-283.8
5	34.71	1.05	0.74	5.42	5.32	-432.3	-16.4	-286.8
6	39.61	0.96	0.65	4.72	4.73	-455.1	-20.6	-295.4

$RMSD_C$ = RMSD between the best model and centroid structure; $RMSD_{C\text{-}best}$ = RMSD between the centroid structure and the best centroid structure; $RMSD_{best\text{-}NMRMD}$ = RMSD between the best structure and the NMR_{MD} structure; $RMSD_{C\text{-}NMRMD}$ = RMSD between the best-centroid structure and the NMR_{MD} structure.

Figure 2: The best structure from each cluster, a) cluster 1, b) cluster 2, c) cluster 3, d) cluster 4, e) cluster 5, f) cluster 6 at 29.39 ns ns, 79.76 ns, 144.08 ns, 129.48 ns, 34.71 ns and 39.61 ns respectively.

Table 2 The SASA value for the best structure from each cluster

Cluster Number	Time (ns)	SASA (Å²/kcal)
1	29.39	1897.14
2	79.76	1706.82
3	144.08	1723.67
4	129.48	1719.53
5	34.71	1873.21
6	39.61	1883.93

a) 29.39 ns   b) 79.76 ns

c) 144.08 ns   d) 129.48 ns

e) 34.71 ns   f) 39.61 ns
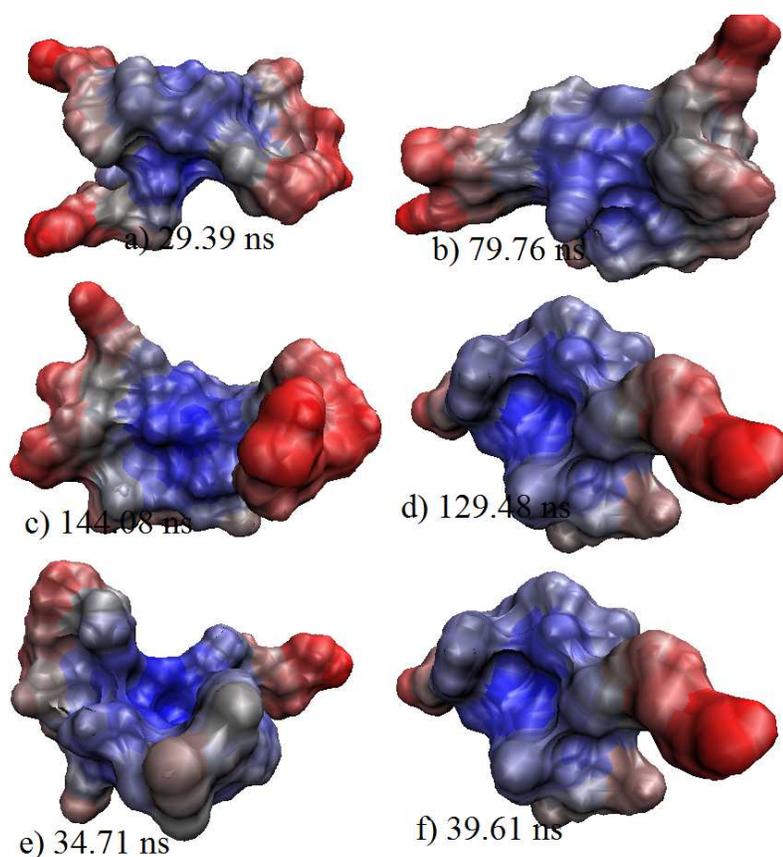
Figure 3: The surface presentation for the best structure from each clusters, a) cluster 1, b) cluster 2, c) cluster 3, d) cluster 4, e) cluster 5, f) cluster 6 at 29.39 ns ns, 79.76 ns, 144.08 ns, 129.48 ns, 34.71 ns and 39.61 ns respectively.

Besides looking at the RMSD values and time of simulation, the energy minimisation and SASA calculation also plays an important factor and needs to be considered in determining the best cluster.   In this study, the energy minimisation calculation was conducted on all the six clusters and the data were presented in Table 4.8.   In Table 4.9,   the electrostatic interaction energy could be highlighted as the important energy since the salt bridge helped in stabilising the structure [9].   Even though cluster 2 had the second lowest electrostatic interaction (466.0 kJ), it had the lowest total energy (321.7 kJ) and lowest SASA (1706.82 Å²/kcal).   Besides that, it also had the lowest $RMSD_{best\text{-}NMRMD}$ (3.36 Å), thus cluster 2   at 79.76 ns was identified as the best cluster.

4. **CONCLUSION**

The purpose of this study was to investigate the folding pathway of Trp-cage miniprotein using molecular dynamics simulation. The protein was chosen due to its small size. Clustering analysis was conducted to group the trajectories based on the RMSD values and six clusters were generated. From this, the best conformation was identified to represent the Trp-cage miniprotein. This study has suggested that the formation of secondary structure is supported by strong hydrogen bond interactions and salt bridge formations.

Acknowledgement

REFERENCES

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

[2] Neidigh, J. W., Fesinmeyer, R. M., & Andersen, N. H. (2002). Designing a 20-residue protein. *Nature Structural & Molecular Biology*, *9*(6), 425-430.

[3] Chowdhury, S., Lee, M. C., Xiong, G., & Duan, Y. (2003). Ab initio folding simulation of the Trp-cage mini-protein approaches NMR resolution. *Journal of molecular biology*, *327*(3), 711-717.

[4] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., ... & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of computational chemistry*, *26*(16), 1668-1688.

[5] Berendsen, H. J., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. R. H. J., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath.*The Journal of chemical physics*, *81*(8), 3684-3690.

[6] Anand, P., & Hansmann, U. H. (2011). Internal and environmental effects on folding and dimerisation of Alzheimer's β-amyloid peptide. *Molecular simulation*,*37*(06), 440-448.

[7] Rovó, P., Farkas, V., Hegyi, O., Szolomájer-Csikós, O., Tóth, G. K., & Perczel, A. (2011). Cooperativity network of Trp-cage miniproteins: probing salt-bridges. *Journal of Peptide Science*, *17*(9), 610-619.

[8] Wu, X., Yang, G., Zu, Y., Fu, Y., Zhou, L., & Yuan, X. (2012). Molecular dynamics characterisations of the Trp-cage folding mechanisms: in the absence and presence of water solvents.*Molecular Simulation*, *38*(2), 161-171.

[9] Williams, D. V., Byrne, A., Stewart, J., & Andersen, N. H. (2011). Optimal salt bridge for Trp-cage stabilization. *Biochemistry*, *50*(7), 1143-1152.