Electronic Supplementary Material (ESI) for RSC Advances.

Structures and Energies of Transition between Two Conformations of Alternate Frame Folding Calbindin-D_{9k} Protein: A Theoretical Study

Mingqiong Tong, Qing Wang, Yan Wang,* and Guangju Chen*

Key Laboratory of Theoretical and Computational Photochemistry, Ministry of Education, College of Chemistry, Beijing Normal University, Beijing 100875, China

Supporting Information

1. Conventional molecular dynamics simulation protocols

All CMD simulations for these built models were carried out using the AMBER 9 package[1] and ff03 all atom force field parameters[2-4]. The protocol for all CMD simulations is described as follows: (1) the systems were energetically minimized to remove unfavorable contacts. Four cycles of minimizations were performed with 2500 steps of each minimization and harmonic restraints on the AFF calbindin-D_{9k} protein from 100, 75, 50 to 25 kcal/(mol·Å^2), which means that the restraints were relaxed stepwise by 25 kcal/(mol·Å^2) per cycle. The fifth cycle consists of 5000 steps of unrestrained minimization before the heating process. The cutoff distance used for the non-bonded interactions was 10 Å. The SHAKE algorithm[5] was used to restrain the bonds containing hydrogen atoms; (2) each energy-minimized structure was heated over 120 ps from 0 to 300 K (with a temperature coupling of 0.2 ps), while the positions of AFF calbindin-D_{9k} protein were restrained with a small value of 25 kcal/(mol·Å^2). The constant volume was maintained during the processes; (3) the

unrestrained equilibration of 200 ps with constant pressure and temperature conditions was carried out for each system. The temperature and pressure were allowed to fluctuate around 300 K and 1 bar, respectively, with the corresponding coupling of 0.2 ps. For each simulation, an integration step of 2 fs was used; and (4) finally, conventional molecular dynamics (CMD) runs of 50 ns for the AFF-D_{9k}-N' and AFF-D_{9k}-N models of the AFF calbindin-D_{9k} protein in Ca^{2+}-free form were carried out, respectively, by following the same protocol.

2. MM-PBSA calculation for free energy

In MM-PBSA, the free energy analyses for the AFF-D_{9k}-N' and AFF-D_{9k}-N models were carried out from 2500 snapshots extracted by 4 ps intervals in the last 10 ns of each simulation. Free energy (G) of each state is estimated from molecular mechanical energy E_{MM}, solvation free energy G_{SOLV} and vibrational, rotational, and translational entropies S.

$$G = E_{MM} + G_{SOLV} - TS$$

$$E_{MM} = E_{int} + E_{vdw} + E_{ele}$$

$$G_{SOLV} = G_{pb/solv} + G_{np/solv}$$

Where T is the temperature; E_{int} is internal energy, i.e. the sum of bond, angle, and dihedral energies; E_{vdw} is van der Waals energy; E_{ele} is electrostatic energy; G_{SOLV} is the sum of electrostatic solvation free energy, $G_{pb/solv}$, and the non-polar salvation free energy, $G_{np/solv}$. The entropy S is estimated by a normal mode analysis of the harmonic vibrational frequencies, calculated using the Nmode module in Amber9 package.[6]

3. Targeted molecular dynamics simulation protocols

In the targeted molecular dynamics simulations,[7, 8] a time-dependent constraining force was applied onto the positions of backbone atoms of two AFF-D_{9k}-N' and AFF-D_{9k}-N models at each time-step to bias the trajectories toward each respective target structures. To obtain the appropriate small force constant k_1, four 8 ns independent TMD simulations were performed using different force constants of 1.0, 2.5, 5.0 and 7.5 kcal/(mol·Å^2), respectively. As shown in Figure S1 for the conversion of AFF-D_{9k}-N' to AFF-D_{9k}-N, although three out of the four TMD simulations reached a *RMSD* value of <2.0 Å extracting from the backbone atoms of two AFF-D_{9k}-N' and AFF-D_{9k}-N models, only two out of four simulations with the force constant of k_1 = 5.0 and 7.5 kcal/(mol·Å^2) reached a *RMSD* value of <2.0 Å from extracting the all atoms of two models. To achieve the satisfied transition process of the initial structure AFF-D_{9k}-N' to the target structure AFF-D_{9k}-N, k_1 = 5.0 kcal/(mol·Å^2) was chosen as the lowest harmonic force constant to apply onto all the backbone atoms of two models to bias the trajectories toward the target structure. All the other atoms in the two AFF-D_{9k}-N' and AFF-D_{9k}-N models were assumed to move freely and to rearrange at their convenience to let the protein reach its equilibrium under this constraint. The weighted RMSDs along the TMD trajectories from four, 8 ns independent simulations showed that each process could follow the same conformational transition pathway, which indicated that the external forces and initial velocities did not affect the transition pathway. Therefore, the TMD simulation with the 8ns simulation time followed by 22 ns of equilibration with an integration step of 2 ps was performed for this transition process.

4. Potential of mean force (PMF) simulations

In the umbrella sampling simulations[9, 10], the biasing force constant applied in different windows of umbrella-sampling ranged from 2.0 to 55.0 kcal/(mol·Å^2). For each umbrella-sampling window, the initial complex structure was selected from the trajectory of TMD simulations based on the reaction coordinate to be sampled for the window of interest. A total of 67 windows from the reaction coordinate of 23.1 Å to 10.3 Å with the increments of 0.2 Å were generated. The selected structure for each window was kept running for over 3 ns for production sampling. We obtained a smooth transition and good overlap between windows by using these conditions. The frequency for data collection was set to 2 fs, which was the same as that of the time step of umbrella-sampling MD. To test whether the PMF calculations reached convergence, we calculated the free energies for the conversion of AFF-D_{9k}-N' to AFF-D_{9k}-N by using different lengths of the MD trajectory.[11] As shown in Figure S2, the PMF for the conversion of AFF-D_{9k}-N' to AFF-D_{9k}-N was determined by using three different lengths of the MD trajectory for each window, i.e., 1 ns, 2 ns and 3 ns. There was no significant difference between the free energy profiles for this conversion determined on the two lengths of MD trajectory with the simulation time of 2 ns and 3 ns; the curve of the free energy profile corresponding to 2 ns almost perfectly overlaps with that corresponding to 3ns. These data suggest that 2 ns for each window of the umbrella sampling simulations should be sufficient for obtaining the converged result from the PMF calculations.

5. Calculations of interhelical angle, B-factor and correlation of atomic motions

In the calculations of interhelical angle, the program INTERHLX calculates the sign of the angle between two helices by following this convenient role: the two helices are taken to be positioned by helix I being in front of helix II. Helix I (from N

to C) is used to define first vertical vector. A second vertical vector is defined with its tail at the C-terminus of helix II. The angle between helices I and II is the rotation required to align the head of the second vector with the N-terminus of helix II. The vector is rotated in the direction that produces an angle of less than 180 degrees with the clockwise or counterclockwise rotation represented by positive or negative sign. This program can also provide other geometry based parameters such as interhelical distances.[12, 13]

In the present study, the B-factor (the temperature factor) values for the EF1 and EF2'/EF2 hands computed from the root-mean-square fluctuations (RMSFs). The RMSFs values of residues are a measure of fluctuations and flexibility of backbone Cα of protein over the trajectory broken down by residues in comparison to the average structures.[14] $RMSF_i$ of the Cα atom of each residue was calculated as follows:

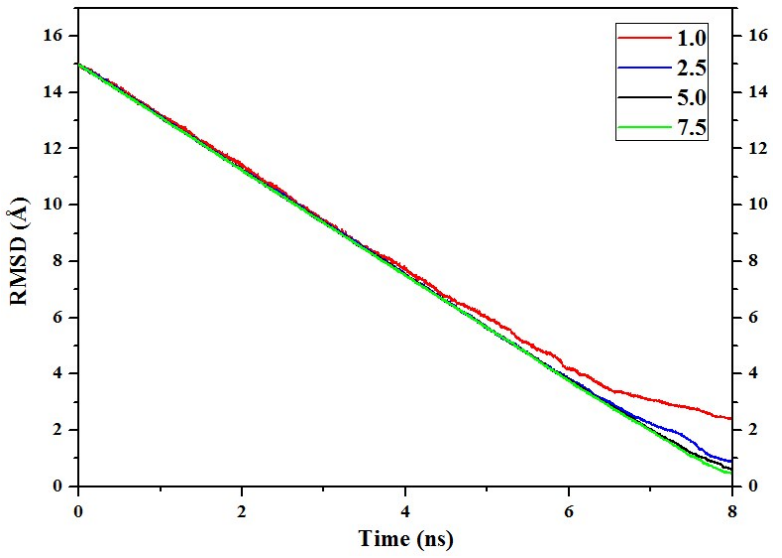
$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_i(t) - \langle r_i \rangle)^2}$$

Where T is the number of snapshots considered in the time trajectory, $r_i(t)$, the position of the Cα atom of residue i at time t, and $\langle r_i \rangle$, the time-averaged position of the Cα atom of residue i.

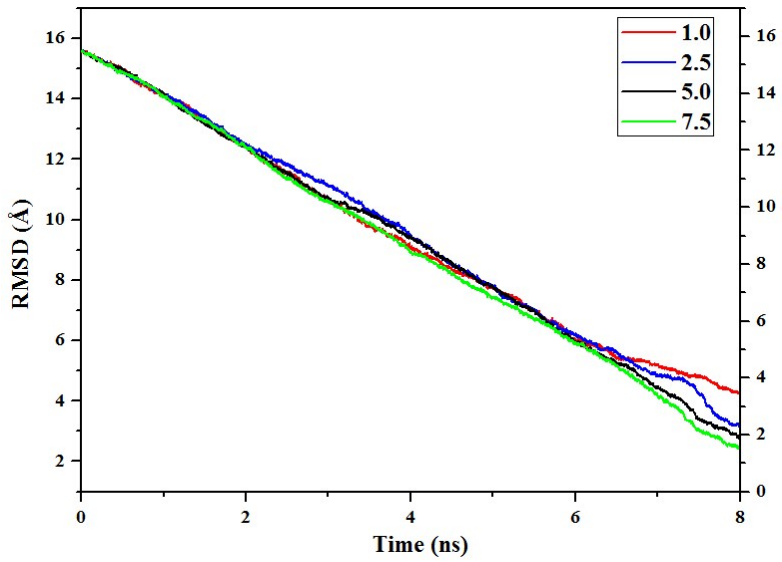
The correlation coefficients were averaged over the regions of protein, and the resulted cross-correlation coefficients are presented in the form of a two-dimensional graph.[14] These structure analyses in the present work were calculated by using the PTRAJ module of the AMBER 9 program.[1]

References

1. T. A. D. D.A. Case, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R., K. M. M. Luo, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S., A. R. Hayik, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H., L. Y. Gohlke, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. and W. S. R. Schafmeister, and P.A. Kollman, *AMBER 9*, University of California: San Francisco, CA, 2006.
2. M. C. Lee and Y. Duan, *Proteins*, 2004, **55**, 620-634.
3. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
4. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
5. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952-962.
6. K. J and C. DA, *Biopolymers*, 1990, **29**, 1409-1421.
7. J. Schlitter, M. Engels, P. Krüger, E. Jacoby and A. Wollmer, *Mol. Simulat.* , 1993, **10**, 291-308.
8. J. Schlitter, M. Engels and P. Krüger, *J. Mol. Graph. Model.*, 1994, **12**, 84-89.
9. G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* , 1977, **23**, 187-199.
10. C. Bartels and M. Karplus, *J. Comput. Chem.*, 1997, **18**, 1450-1462.
11. X. Huang, X. Zhao, F. Zheng and C.-G. Zhan, *J. Phys. Chem. B*, 2012, **116**, 3361-3368.
12. K. L. Yap, J. B. Ames, M. B. Swindells and M. Ikura, *Proteins*, 1999, **37**, 499-507.
13. J. B. A. Kyoko L. Yap, Mark B. Swindells, Mitsuhiko Ikura, *Methods Mol. Biol.* , 2002, **173**, 317-324.
14. S. K. Sadiq and G. De Fabritiis, *Proteins*, 2010, **78**, 2873-2885.



(a)



(b)

Figure S1. The root mean square deviation (RMSD) values of (a) all backbone atoms and (b) all atoms of the whole protein for the conversion of AFF-D_{9k}-N' to AFF-D_{9k}-N using varying force constants *k*, in kcal/(mol·Å^2), shown in the inset, respectively.

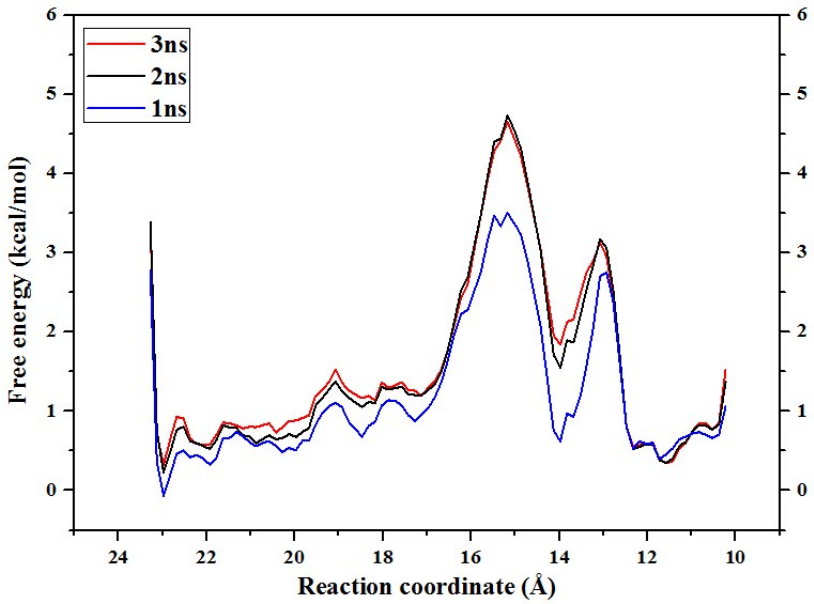
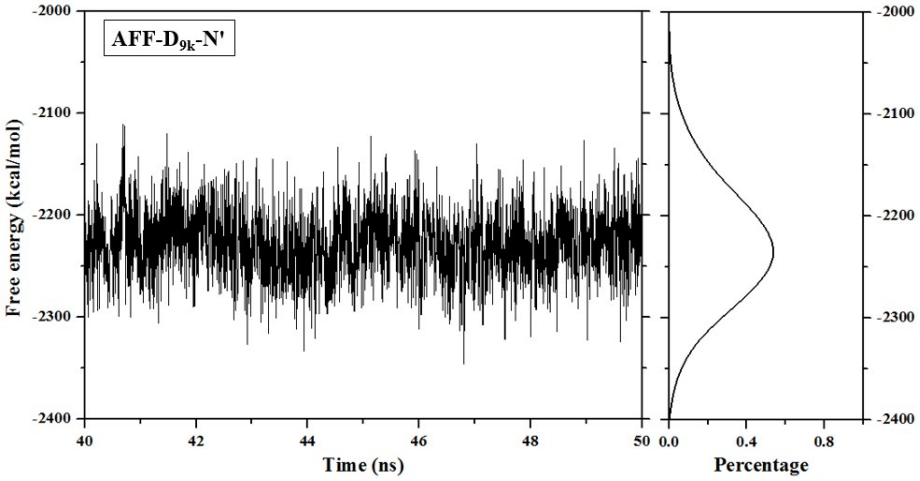
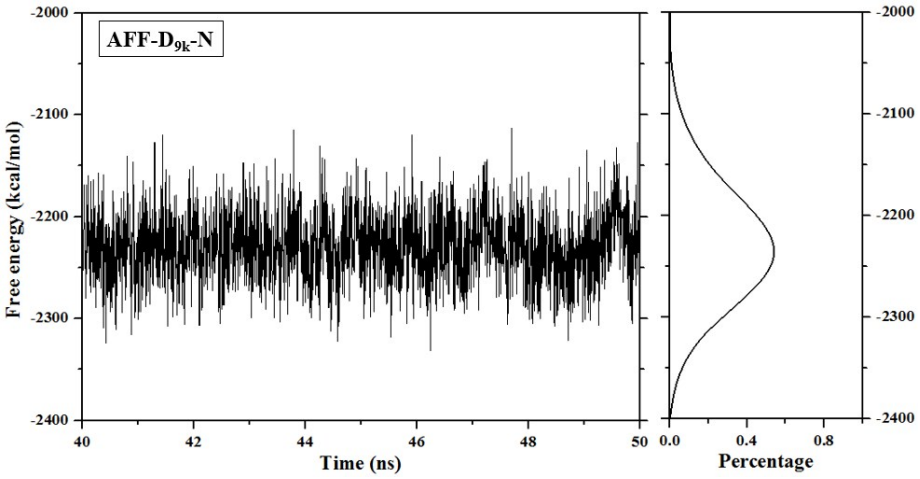


Figure S2. Calculated free energy profiles for the conversion of AFF-D_{9k}-N' to AFF-D_{9k}-N determined by using different lengths of MD trajectory from each window during the umbrella sampling simulations, *i.e.* the blue curve is determined by using 1 ns MD trajectory from each window, the black curve by using 2 ns MD trajectory, and the red curve by using 3 ns MD trajectory.



(a)



(b)

Figure S3. Conformational stability free energies (kcal/mol) and integrated distributions for (a) the AFF-D_{9k}-N' model; and (b) the AFF-D_{9k}-N model.

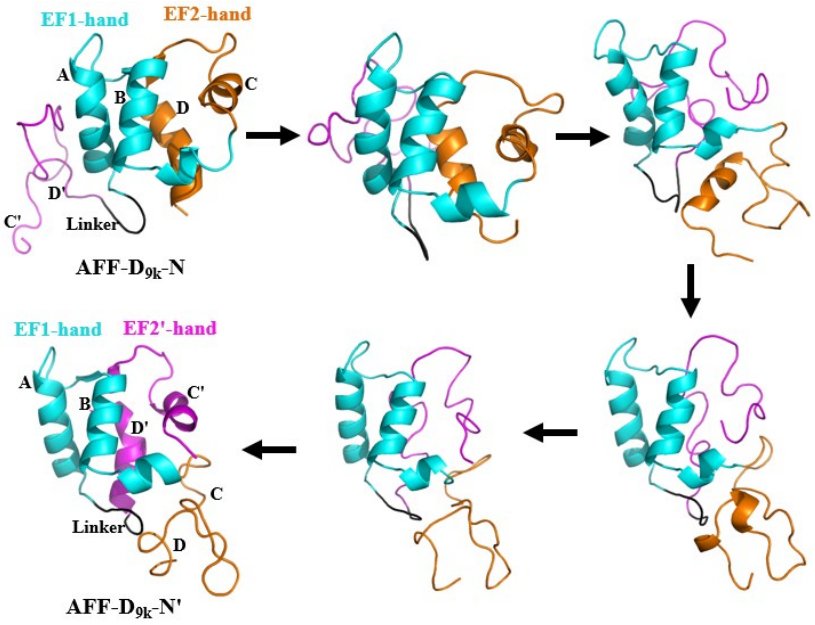


Figure S4. The three-dimensional structures of the conformational transition pathway from AFF-D_{9k}-N to AFF-D_{9k}-N' determined by the corresponding TMD simulation.