Solvent effect on the folding dynamics and structure of E6-associated protein characterized from Ab initio protein folding simulations

Xu, Zhijun; Lazim, Raudah; Sun, Tiedong; Mei, Ye; Zhang, Dawei

2012

Xu, Z., Lazim, R., Sun, T., Mei, Y., & Zhang, D. (2012). Solvent effect on the folding dynamics and structure of E6-associated protein characterized from ab initio protein folding simulations. *The Journal of Chemical Physics*, 136(13), 135102.

<https://hdl.handle.net/10356/95243>

<https://doi.org/10.1063/1.3698164>

© 2012 American Institute of Physics. This paper was published in *The Journal of Chemical Physics* and is made available as an electronic reprint (preprint) with permission of American Institute of Physics. The paper can be found at the following official DOI: [<http://dx.doi.org/10.1063/1.3698164>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 18:34:22 SGT

Solvent effect on the folding dynamics and structure of E6-associated protein characterized from ab initio protein folding simulations

Zhijun Xu, Raudah Lazim, Tiedong Sun, Ye Mei, and Dawei Zhang

Citation: *J. Chem. Phys.* **136**, 135102 (2012); doi: 10.1063/1.3698164

View online: <http://dx.doi.org/10.1063/1.3698164>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i13>

Published by the [American Institute of Physics](#).

Additional information on *J. Chem. Phys.*

Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT

Goodfellow
metals • ceramics • polymers • composites
70,000 products
450 different materials
small quantities fast
www.goodfellowusa.com

Solvent effect on the folding dynamics and structure of E6-associated protein characterized from *ab initio* protein folding simulations

Zhijun Xu,¹ Raudah Lazim,^{1,a)} Tiedong Sun,^{1,a)} Ye Mei,² and Dawei Zhang^{1,b)}

¹*Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore*

²*State Key Laboratory of Precision Spectroscopy and Department of Physics, Institute of Theoretical and Computational Science, East China Normal University, Shanghai 200062, China*

(Received 22 December 2011; accepted 12 March 2012; published online 2 April 2012)

Solvent effect on protein conformation and folding mechanism of E6-associated protein (E6ap) peptide are investigated using a recently developed charge update scheme termed as adaptive hydrogen bond-specific charge (AHBC). On the basis of the close agreement between the calculated helix contents from AHBC simulations and experimental results, we observed based on the presented simulations that the two ends of the peptide may simultaneously take part in the formation of the helical structure at the early stage of folding and finally merge to form a helix with lowest backbone RMSD of about 0.9 Å in 40% 2,2,2-trifluoroethanol solution. However, in pure water, the folding may start at the center of the peptide sequence instead of at the two opposite ends. The analysis of the free energy landscape indicates that the solvent may determine the folding clusters of E6ap, which subsequently leads to the different final folded structure. The current study demonstrates new insight to the role of solvent in the determination of protein structure and folding dynamics. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3698164>]

INTRODUCTION

Understanding the pathway and mechanism of protein folding is fundamental and of practical use in comprehending protein-related diseases and biological processes,^{1–5} such as protein degradation and protein translocation. Many factors such as solvent environment, chemical denaturant, and temperature have demonstrated critical effect on the structure and function of proteins. For example, the 2,2,2-trifluoroethanol (TFE) has been shown to stabilize/induce the formation of helical conformations in denatured proteins.^{6,7}

In the field of protein folding, molecular dynamics (MD) simulation is considered as a powerful tool for elucidating the folding dynamics,^{8–12} such as acquiring early folding events and ensembles during the self-assembly of proteins to their biological “native state” unbiased by native-state information. However, most folding studies, thus far, sample the complete phase space in order to find the native structure.¹³ Since protein folding in nature does not need to sample the entire phase space to approach the native state and virtual phase space, sampling is often computationally infeasible, and probing the folding pathway, guided by physical intramolecular and intermolecular forces, may provide a more convincing insight into the folding mechanism and is generally more practical. Since hydrogen bonds within proteins play an important role in maintaining secondary structures of proteins,¹⁴ electrostatic polarization in hydrogen bonding may also play an important role in folding dynamics and this rationale prompts us to develop a new charge scheme labeled as adaptive hydrogen

bond-specific charge (AHBC) in which the atomic charges of amino acids involved in the forming or breaking of hydrogen bonds during the folding simulation will be periodically updated by treating these amino acids with fragment quantum mechanical calculations.^{15–17} The successful application of AHBC scheme on the folding of 2KI9 peptide¹⁸ and polyalanine mutants¹⁹ has demonstrated that this on-the-fly charge fitting can ensure a relatively accurate electrostatic polarization effect for hydrogen bonds in immediate environment and thus better study protein folding dynamics.

In this work, we applied the AHBC scheme again to investigate the role of solvent in the folding of E6-associated protein (E6ap) peptide by carrying out *ab initio* folding simulations under two different solvent conditions. E6ap (Ref. 20) represents the E6 binding domain of the human cellular protein E6ap which functions as an ubiquitin-protein ligase.²¹ Thus, it mediates the human papillomavirus E6 protein-induced ubiquitination, which may provide insight into the interactions between E6 and E6ap.²⁰ Experimental research²² has shown that in 40% TFE, the NMR structure is composed of a well-defined α -helix at residues 4–13, while in pure aqueous solution, it has a helix content of about 20% based on circular dichroism (CD) measurements. It is known that in order to characterize the early folding events on nanosecond or microsecond timescales, approaches with fast time-resolution and high structure sensitivity such as time-resolved infrared (IR) and fluorescence spectrometers can be applied to probe the folding dynamics and protein structures along the folding/unfolding pathways. In the folding study of E6ap peptide, since only NMR structure in 40% TFE and CD-based helix content in water are available as references for comparison with our theoretical studies, this study aims

^{a)}R. Lazim and T. Sun contributed equally to this work.

^{b)}E-mail: zhangdw@ntu.edu.sg.

to provide insight for more understanding of the folding of E6ap at atomic level and the available experimental observations in different solvents offer a good chance to investigate the effect of solvent on the general principles of folding. Cui *et al.*²³ carried out Langevin dynamics (LD) simulations to explore the folding pathways of E6ap by starting from different denatured configurations obtained from LD simulations of E6ap performed at high temperatures. Through this study, leucine hydrophobic core was found to play an important role in the stability of the native structure. In the simulation presented here, the backbone hydrogen bonds of the peptide are checked every constant time interval and charge fitting of hydrogen bond donors and acceptors is performed at the level of B3LYP/6-31G*.

COMPUTATIONAL METHODS

In order to closely mimic the natural folding process, we started from the fully extended conformation of E6ap to conduct *ab initio* simulations with AHBC scheme. The linear structure was built by LEaP module of AMBER 9.0 (Ref. 24) using the E6ap sequence (IPESSELTLQELLGEERR). When performing AHBC, charge update will be carried out periodically based on density functional theory (DFT) charge-fitting calculation to account for the constant forming and breaking of main chain hydrogen bonds during the folding process. The basic procedures of DFT charge-fitting calculation can be described as follows: Poisson-Boltzmann equation is solved first to get the induced charges on protein-solvent interface. Then the DFT calculations are carried out with the induced charges taken as background charges to mimic the polarization effect from the solvent molecules. New atomic charges which are fitted to the electrostatic potential from DFT calculations will generate new induced charges. This self-consistent procedure will be applied until convergence is reached. The HB-plus program²⁵ is used to track hydrogen bonds formed during the simulation. Due to its heavy computational expense, periodic check of main chain hydrogen bonds and charge update are carried out every 10 ps. The solvation effect is represented by a generalized Born Igb5 model, which is widely used in the simulation of protein folding.^{26–29} The salt concentration is set to 0.1M and the external dielectric constant is 57.5 to mimic the 40% TFE solvent environment used by Baldwin and coworkers.³⁰ In the simulation, we use effective dielectric constant of 57.5 to describe 40% TFE solvent. This approximation is based on the works done by Goto, Scheraga, and Mark. The study by Yuji Goto *et al.*³¹ indicate that the dielectric constant of solvent is linearly related with TFE concentration. After that, Scheraga *et al.*³² carried out simulations by varying dielectric constant to mimic TFE/water mixed solvent and their results revealed that the key effect of TFE on α -helix stabilization may likely be due to the ability of TFE molecules aggregating around the surface of peptide and interacting weakly with nonpolar residues of peptides. This finding was confirmed by Mark's simulations³³ in explicit TFE/water solvent. All these studies make us believe that our treatment of TFE/water mixture based on Goto's work can mimic the key feature of TFE.

After the energy minimization, the system is heated up to 300 K in 100 ps, followed by 50 ns MD simulation with a time step of 2 fs. All bonds containing hydrogen atoms are fixed with SHAKE (Ref. 34) algorithm.

RESULTS AND DISCUSSIONS

First, we consider the electrostatic polarization effect on the E6ap protein folding dynamics in 40% TFE solution by comparing the MD simulations based on the standard AMBER force field to that performed with AHBC scheme. Both simulations based on AMBER and AHBC use generalized born (GB) solvation model and have the same starting velocity to make them identical except for the fact that the charges under AHBC scheme fluctuate. The available NMR structure in the TFE solution (PDB id 1EQX (Ref. 22)) was used as a standard structure for comparison with the simulation structure. It is observed that the folded protein approaches native state in less than 50 ns when AHBC scheme was implemented and this is shown by the stabilization of the backbone RMSD circa 1.0 Å after the 30 ns mark which is not observed when standard AMBER charges were used (Figure 1). Also, the RMSD of the protein relative to NMR structure experienced a rapid fall at the early folding stage till about 4.8 ns and the system continues to fluctuate until a further decrease of RMSD is observed at about 31 ns to form a stable structure with the lowest backbone RMSD of about 0.5 Å. Further analysis of the RMSD distributions during the entire simulation was done to give a direct comparison of the similarity between the simulated structures and native structure. As displayed in the inset of Figure 1, the most populated state under AMBER force field has a RMSD of 2.2 Å, while for the AHBC-incorporated simulations, folded structures with RMSD of 1.0 Å compared to NMR structure are well represented. In order to show the effect of AHBC on the charges, we picked out the charges in the last frame structure during the AHBC simulation and compared them with AMBER charges. As can be seen in Figure 2, after periodic charge update, partial atomic charges in the helix domain (residues 4–14) are generally more positive for main chain HB donor (@H) and more negative for main chain HB acceptor (@O) than AMBER charges (black dots). The charge difference clearly showed the changes introduced by polarization effect included in AHBC scheme. Therefore, given the importance of polarization effect in describing the secondary structures of proteins, all simulations hereafter in this study were carried out using the AHBC scheme. It allows us to delve into the effect of solvent in the folding of E6ap in different solvent environment based on the obtained high-resolution protein structures from the simulations performed with AHBC scheme.

In Figure 3, the temporal variation of backbone RMSD and helix content is presented for both simulation conditions which are in pure water and 40% TFE solution. For each condition, only two independent simulations are highlighted here for clarity. In 40% TFE solution, the two presented simulations reach the folded structure in approximately similar time frames but with some variations in the intermediate structures. For the simulation in pure water, the RMSD fluctuations

FIG. 1. Comparison of the RMSDs from two simulations based on the AHBC scheme (red) and AMBER force field (black) in 40% TFE solution. Distributions of the RMSD are given in the inset picture.

ranged from 3.0 to 5.0 Å after an abrupt decrease in RMSD to 4.0 Å at the start of the simulation. We also noted that the two independent simulations in pure water pursued diverse trajectories with obvious difference in the RMSD profiles. This is quite different from the explicit structures formed in TFE solution, which has similar RMSDs obtained from the two independent simulations. This indicates that in pure water, E6ap could not fold into a well-defined secondary structure. In order to quantitatively explore the folding dynamics of E6ap, the fractional helix content (percentage of residues present as helix) is also shown in Figure 3. Although helices start to form immediately for both solutions, the fractional helix content increased to almost 60% in just 2 ns in TFE solution and in pure water, it only increased to about 38%. As the simulation time increases, the number of residues present as helix continually increases to about 70% in TFE solution, and in the pure water, the helix content decreases and plateaued at about 24%

FIG. 2. The partial atomic charges of both main chain HB donor (@H) and acceptor (@O) in the last frame structure of the AHBC simulation are compared with AMBER charges (black dots).

after 5 ns, which is in good agreement with the CD experimental value (20%). Note that the NMR structure of E6ap in TFE solution with a well-defined central helix at residues 4–13 has a stable helix content of about 55%. The fact that our obtained helix content is higher than that of the NMR structure indicates that the two ends of the peptide also partook in the formation of the helical structure.

The above simulations replicate the tendency of E6ap to form a helical structure with different fractional helix

FIG. 3. RMSD (trajectories colored black and red in pure water and green and purple in 40% TFE) and fractional native helix content evolutions (red in pure water and green in 40% TFE) from the simulation based on the AHBC scheme in pure water and 40% TFE solution.

FIG. 4. Evolution of secondary structures obtained from DSSP analysis for E6ap in (a) 40% TFE solution and (b) pure water.

contents in different solutions as the helix content obtained from the simulations conducted is in good agreement with the experimental observations for both conditions. However, some detailed dynamical behaviors are still unknown, such as helix formation events occurring during the folding of E6ap and how development of the helical structures is affected by its solvent. Although only two independent simulations performed under each solvent condition are really not enough to give unbiased information regarding the early folding events such as nucleation and propagation, we can still obtain some indications from the presented simulations. For this purpose, the change in secondary structure of the peptide with simulation time was studied using the program DSSP for both TFE solution and pure water, as shown in Figure 4. In the presented simulation, we observed that at the beginning of the protein folding (about 1 ns), the formation of helical structures starts with the formation of two shorter helices at the two ends of the peptide which propagate toward the center of the peptide separately and merge within 4 ns to form a single helix, as shown in Figure 4 (top). During the simulation, a stable helix was formed at residues 4–15 in the TFE solution. A direct insight into the folding of E6ap from the snapshots obtained through cluster analysis can be further acquired. The cluster analysis was performed via K-means clustering, and the representative structures for the three most populated clusters are given in Figure 5. This further concurs the preference of E6ap to fold into a structure with high helix content in 40% TFE solution as opposed to in pure water.

In order to provide further insight on the evolution process of dynamically folding behavior of the protein, free energy contour maps were plotted to provide a direct comparison of the folding under different solvents, as shown in Figure 6. The RMSD and radius of gyration of the protein backbone are used as the reaction coordinates. Gener-

FIG. 5. Representative structures of conformations selected from the most populated clusters for two solutions: pure water (above) and 40% TFE solution (below). The most populated structure in 40% TFE solution is overlapped with the reference NMR structure (dark green).

ally, the contour maps for the two different solvent environments demonstrated completely different characteristics. As shown in Figure 4, two cluster regions are defined: I1 (RMSD = 3.8 Å, Rg = 5.8 Å) and I2 (RMSD = 2.7 Å, Rg = 7.0 Å). The representative structures at the defined points are also given on the right. It is observed that although E6ap has similar structural measurement for the backbone RMSD and radius of gyration at the region I1, the detailed structure of the protein is quite different in each solvent. In 40% TFE solution, the two ends of the peptides have formed two independent helical structures with a kink between them, which agrees with the secondary structure assignment and cluster analysis depicted in Figures 4 and 5, respectively. However, in pure water, only several residues in the central part of the protein folded into a helix. Similar behaviors could be observed for the cluster region I2. According to Figure 6(b), the lowest free energy well (RMSD = 0.9 Å, Rg = 6.0 Å) corresponds to the folded state of E6ap in 40% TFE solution, but in pure water, the lowest free energy is at the cluster region I1. This indicates different folding behaviors of E6ap occur in the two solvents.

As shown in Figure 6(c), to further corroborate the free energy landscape plotted (vide supra), a further analysis of the free energy based on backbone RMSD of the protein was conducted for E6ap in both 40% TFE solution and pure water. For the convenience of comparison, the free energy profiles have been positioned at the RMSD of 3.8 Å, corresponding to the cluster region I1. We observed that the protein needs to cross a free-energy barrier of about 1 kcal/mol from the cluster state I1 to the folded state in the TFE solution. But in pure water, the free-energy barrier is elevated to approximately 5 kcal/mol, thus preventing the peptide from approaching the folded state. This indicates that the solvent may determine the protein structures along the folding pathway of E6ap, which subsequently leads to the different final folded structures.

FIG. 6. Free energy landscapes of E6ap as a function of RMSD and radius of gyration in (a) pure water and (b) 40% TFE solution. Representative snapshots of the intermediate states are shown. (c) Free energy profile of E6ap from the simulation in pure water (magenta) and 40% TFE solution (blue) based on backbone-RMSD. The point marked by the red sphere is used to superpose the free energy.

CONCLUSION

With the newly developed charge update AHBC scheme, we performed *ab initio* protein folding simulations to investigate the solvent effect on the secondary structure and the folding of E6ap. The updated charge for the residues is obtained from the quantum mechanical calculation for a protein in solution at regular intervals based on the variation of main chain hydrogen bonds during the MD simulation. The obtained simulation results are in good agreement with experimental observations. The obtained structure in 40% TFE solution has a lowest backbone RMSD of about 0.9 Å with the available NMR structure as reference (PDB id 1EQX (Ref. 22)). In pure water, the calculated helix content (23%) is quite close to the CD measurements (about 20%). Based on the high agreement between the simulation results and experiments, we observed from the presented simulations that the folding of E6ap in TFE solution may begin at the two terminals of the peptide and progresses toward the center, while in pure water, only a few residues at the center of the peptide fold into a helix. Through the investigation of the free energy landscapes, cluster regions are observed at the same position within their respective landscapes but with different structural characteristics. This may hint that the folding of E6ap, which may be manipulated by the solvent, leads to the different final folded structures.

ACKNOWLEDGMENTS

D.W.Z. is supported in part by Nanyang Technological University start-up grant, and in part by Singapore AcRF Tier 1 Grant of M52110095. D.W.Z. thanks NTU HPC support and

resources. Y. M. is supported by the National Natural Science Foundation of China (Grant No. 21173082).

- ¹F. U. Hartl and M. Hayer-Hartl, *Nat. Struct. Mol. Biol.* **16**(6), 574–581 (2009).
- ²Y. W. Chen, F. Ding, H. F. Nie, A. W. Serohijos, S. Sharma, K. C. Wilcox, S. Y. Yin, and N. V. Dokholyan, *Arch. Biochem. Biophys.* **469**(1), 4–19 (2008).
- ³A. Naeem and N. A. Fazili, *Cell Biochem. Biophys.* **61**(2), 237–250 (2011).
- ⁴A. Relini, O. Cavalleri, R. Rolandi, and A. Gliozzi, *Chem. Phys. Lipids* **158**(1), 1–9 (2009).
- ⁵V. N. Uversky and A. K. Dunker, *Biochim. Biophys. Acta/Proteins Proteomics* **1804**(6), 1231–1264 (2010).
- ⁶S. Dave, S. Mahajan, V. Chandra, and P. Gupta, *Int. J. Biol. Macromol.* **49**(4), 536–542 (2011).
- ⁷F. Mehrnejad and N. Chaparzadeh, *J. Biomol. Struct. Dyn.* **26**(2), 255–262 (2008).
- ⁸K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *Science* **334**(6055), 517–520 (2011).
- ⁹S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Biophys. J.* **100**(9), L47–L49 (2011).
- ¹⁰H. A. Scheraga, M. Khalili, and A. Liwo, *Annu. Rev. Phys. Chem.* **58**(1), 57–83 (2007).
- ¹¹D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. B. Shan, and W. Wriggers, *Science* **330**(6002), 341–346 (2010).
- ¹²V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, *J. Am. Chem. Soc.* **132**(5), 1526–1528 (2010).
- ¹³K. Klenin, B. Strode, D. J. Wales, and W. Wenzel, *Biochim. Biophys. Acta/Proteins Proteomics* **1814**(8), 977–1000 (2011).
- ¹⁴J. U. Bowie, *Curr. Opin. Struct. Biol.* **21**(1), 42–49 (2011).
- ¹⁵D. W. Zhang and J. Z. H. Zhang, *J. Chem. Phys.* **119**(7), 3599–3605 (2003).
- ¹⁶A. M. Gao, D. W. Zhang, J. Z. H. Zhang, and Y. K. Zhang, *Chem. Phys. Lett.* **394**, 293–297 (2004).
- ¹⁷Y. Mei, D. W. Zhang, and J. Z. H. Zhang, *J. Phys. Chem. A* **109**, 2–5 (2005).
- ¹⁸L. L. Duan, Y. Mei, D. W. Zhang, Q. G. Zhang, and J. Z. H. Zhang, *J. Am. Chem. Soc.* **132**(32), 11159–11164 (2010).
- ¹⁹C. Y. Wei, D. Tung, Y. M. Yip, Y. Mei, and D. W. Zhang, *J. Chem. Phys.* **134**, 171101 (2011).

- ²⁰J. M. Huibregtse, M. Scheffner, and P. M. Howley, *Mol. Cell. Biol.* **13**(8), 4918–4927 (1993).
- ²¹A. L. Talis, J. M. Huibregtse, and P. M. Howley, *J. Biol. Chem.* **273**(11), 6439–6445 (1998).
- ²²X. Be, Y. Hong, J. Wei, E. J. Androphy, J. J. Chen, and J. D. Baleja, *Biochemistry* **40**(5), 1293–1299 (2001).
- ²³B. Cui, M. Y. Shen, and K. F. Freed, *Proc. Natl. Acad. Sci. U.S.A.* **100**(12), 7087–7092 (2003).
- ²⁴D. A. Case, T. A. Darden, T. E. Cheatham, III *et al.*, AMBER9, University of California, San Francisco, 2006.
- ²⁵I. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**(5), 777–793 (1994).
- ²⁶L. L. Duan, Y. Mei, Y. L. Li, Q. G. Zhang, D. W. Zhang, and J. Z. H. Zhang, *Sci. China, Ser. B: Chem.* **53**(1), 196–201 (2010).
- ²⁷S. Koulgi, U. Sonavane, and R. Joshi, *J. Mol. Graphics* **29**(3), 481–491 (2010).
- ²⁸M. S. Lee and M. A. Olson, *J. Chem. Theory Comput.* **6**(8), 2477–2487 (2010).
- ²⁹E. Lin and M. S. Shell, *J. Chem. Theory Comput.* **5**(8), 2062–2073 (2009).
- ³⁰C. A. Rohl, A. Chakrabarty, and R. L. Baldwin, *Protein Sci.* **5**(12), 2623–2637 (1996).
- ³¹D. P. Hong, M. Hoshino, R. Kuboi, and Y. Goto, *J. Am. Chem. Soc.* **121**, 8427 (1999).
- ³²J. A. Vila, D. R. Ripoll, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13075 (2000).
- ³³D. Roccatano, G. Colombo, M. Fioroni, and A. E. Mark, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12179 (2002).
- ³⁴V. Kräutler, W. F. van Gunsteren, and P. H. Hünenberger, *J. Comput. Chem.* **22**(5), 501–508 (2001).