Theor Chem Acc (2012) 131:1168
DOI 10.1007/s00214-012-1168-0

LETTER

Folding and thermodynamic studies of Trp-cage based on polarized force field

Ye Mei · Caiyi Wei · Yew Mun Yip · Chun Ying Ho · John Z. H. Zhang · Dawei Zhang

Received: 6 August 2011 / Accepted: 16 December 2011 / Published online: 2 March 2012

Abstract Two replica exchange molecular dynamics (REMD) simulations were carried out to study the thermodynamics of a 20-residue Trp-cage folding based on a newly developed polarized protein-specific charge (PPC). Starting from a fully extended conformation, Trp-cage native conformation was successfully sampled using REMD based on a 3-step PPC update. Next, the obtained Trp-cage folded conformation was then used to calculate the PPC in which another REMD was performed to explore the thermodynamic stability of Trp-cage. The theoretical melting temperature T_m of $\approx$325 K was found to be in close agreement with experimental melting temperature, T_m of 315 K. This indicates that the PPC was correctly predicting the temperature dependence. The current study provides a direct proof of how electrostatic polarization affects protein folding.

Keywords MFCC · Trp-cage · PPC · Charge update

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1168-0) contains supplementary material, which is available to authorized users.

Y. Mei · J. Z. H. Zhang
State Key Laboratory of Precision Spectroscopy and Department of Physics, Institute of Theoretical and Computational Science, East China Normal University, Shanghai, China

C. Wei · Y. M. Yip · C. Y. Ho · D. Zhang (✉)
Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
e-mail: zhangdw@ntu.edu.sg

J. Z. H. Zhang
Department of Chemistry, New York University, New York, NY 10003, USA

1 Introduction

Understanding how protein folds remains unsolved as one of the most fundamental problems in molecular biology despite of the enormous efforts carried out by various groups [1–3]. Solving such a problem not only clarifies the crucial features of protein sequences for the design of new well-folded functional proteins, but also assists in finding cures for diseases such as Alzheimer, Bovine *Spongiform Encephalopathy* (Mad Cow), Parkinson and many various cancers [4–7], which originate from improper protein folding. Although the study of protein folding via experimental techniques is an active area of research and a great deal of progress has been made to date, experimental probes have not yet been able to provide a comprehensive view in the folding dynamics and kinetics. Computer simulations can thus provide assistance to experimental efforts by providing valuable information at various levels of resolution on the details of protein folding [8–11].

Basically, the computational simulation of protein folding aims to answer three major questions. (1) What secondary structure will a protein sequence give rise to? (2) How does the native state of a protein stabilize thermodynamically? (3) How fast can a protein fold from the unfolded state to its native structure? There have been some notable successes in protein structure prediction reported by using molecular dynamics simulation [12–16]. In the previous few years, a 20-residue mini protein Trp-cage [17] has been an attractive target for computer simulation of protein folding due to its small size and fast-folding kinetics. As this protein can fold successfully under various force fields, it serves as a benchmark system for researchers to validate force fields and folding strategies.

Pitera and Swope [12] used replica exchange molecular dynamics (REMD) [18, 19] and the generalized Born

solvent model [20] to fold the Trp-cage and reproduced the stable folded structure that was <1 Å Cα RMSD from the NMR structure. Zhou [16] used the OPLSAA force field [21] with explicit SPC water model [22] to construct a folding-free energy landscape and proposed a two-step folding mechanism for the superfast folding process. García [15] performed REMD in explicit solvent and developed a stability diagram model for Trp-cage. He also studied the folding/unfolding thermodynamics by using ff99SB force field and the TIP3P water model and gave good agreement on the melting temperature [23]. Levy [24] combined REMD and TPT (transition pathway theory) to construct the folding pathways. Zacharias [25] applied standard temperature REMD and a biased potential REMD method to investigate the folding process and found the folding being favored by more van der Waals contributions. Zhang [26] also tested the different combinations of AMBER force fields and generalized Born models on Trp-cage folding/unfolding. All these works contribute much to our understanding of protein-folding process at atomic level. However, these simulation studies either give a much higher melting transition temperature T_m above 400 K, which deviates significantly from experimental 315 K or produce accurate T_m but incorrect enthalpy and entropy contributions to the thermodynamic stability [27]. Hence, there is a need to improve force field and folding strategies to obtain a more accurate prediction on protein thermostability.

Recent work done by Carlos Simmerling et al. [28] gave a very accurate T_m by reducing generalized Born(GB) radii of hydrogen atoms bound to charged nitrogen atoms in the folding simulation of Trp-cage. Even though changing the GB radius parameters proved to work for Trp-cage, such a method may not work for other proteins. Since the GB is derived from classical mechanics, the basis of reducing the radii which the simulation was started with will deviate at quantum scales due to the over-idealized approximation in classical mechanics. Thus, in this research, the folding and thermodynamics of Trp-cage was studied from another point of view, by using a newly developed atomic partial charge. Without tuning the GB radius parameters, AMBER charge was replaced by quantum-derived polarized protein-specific charge (PPC) while keeping other parameters in AMBER force field intact.

The standing point of this work is based on the fact that electrostatic interaction can play an important role in maintaining the thermodynamic stability of proteins [29–31]. Hence, a force field that models electrostatics by accounting for polarization effects may perform well for the prediction of protein stability. Current standard force fields for proteins, such as AMBER and CHARMM have demonstrated great successes in protein structure prediction; however, they have limitations in the ability to reproduce the thermodynamic stability of proteins due to the lack of explicit polarization effect. Specifically, in these force fields, the modeling of electrostatic interactions is based, in a homogeneous fashion, on the fixed atomic charges. These fixed charges developed from data on small amino acid fragment are not transferable across the different regions of protein and so must be determined case by case.

Unlike AMBER charges, PPC is derived from quantum mechanical calculation of protein in solution by using molecular fractionation with conjugate caps (MFCC) approach [32–36] incorporating the Poisson–Boltzmann (PB) equation [37, 38]. It contains the proper polarization effect, which should hence provide more reliable description in protein dynamics. Recent works also support the stand that using PPC can give a good description of the native structure [39, 40]. Thus, in this work, we employ PPC to include polarization effect to study the thermodynamic properties, that is, the population of folded state in various temperatures. The flow in the organization of the paper is as follows: Section 2, computational details for PPC calculation and REMD simulation are described. Section 3, one REMD simulation is carried out to search for the native conformation of Trp-cage based on a three-step PPC update. Section 4 gives a thermodynamic study of Trp-cage folding using another REMD simulation based on native conformation-derived PPC.

2 Computational methods

2.1 PPC charge calculation

Since the procedure to compute PPC for a given protein structure has already been reported elsewhere [41], a brief review of the method will be discussed here. The basic procedures in fitting the atomic charges of protein in our approach can be described as follows: First, gas phase quantum chemistry calculation of protein was performed with MFCC approach to obtain the initial electron density of the protein through fragment calculation for the given structure as described earlier [32, 35, 36]. The calculated electron density is then used to fit the atomic charges using the RESP procedure [42, 43]. The charge-fitting philosophy used here is the same as that used in the AMBER force field, and this guarantees that the PPC charge is consistent with other parameters of the AMBER force field. Solution of the PB equation is then carried out to obtain reaction field, from which discrete surface charges on the cavity surface are generated. The quantum chemistry calculation of protein fragment is performed again but now with the protein embedded in an external electrostatic potential produced by the induced solvent surface charges and other

fragments of the protein [36, 41]. The newly calculated protein atomic charges are used again to calculate the new solvent-induced charges, and such processes are repeated until convergence has reached. Since density functional theory (DFT) is expected to do a decent job of providing the electrostatic interaction of molecular systems, all the quantum calculations in this study were performed at the level of B3LYP/6-31G*.

2.2 REMD simulation to search Trp-cage native conformation

Since PPC charge is structure dependent, an effort was made to simulate the folding of Trp-cage without prior knowledge of its native structure in this work. A 3-step REMD simulation (S1, S2 and S3) with a periodic charge update was carried out first to search for the Trp-cage native conformation. The simulation started from a linear structure that was built by LEaP module of AMBER 9.0 using the Trp-cage amino acid sequence (NLYIQWLKDG GPSSGR-PPPS [17]). The extended conformation was then subjected to 1,000 steps of steepest descent and 3,000 steps of conjugate gradient minimization, followed by 80 ps equilibration in which the temperature was gently raised from 0 to 298 K in 30 ps and maintained at that temperature over 50 ps by using a Berendsen thermostat with a coupling time constant of 1 ps. The conformation generated after the equilibration step was used as both an input to calculate PPC and the starting structure for the first REMD step.

Replica exchange molecular dynamics (REMD) simulation was then carried out to search the Trp-cage structures, in which AMBER03 [44] force field parameters were applied, except for the atomic charges that were replaced by PPC. The temperatures of the 10 replicas are 253, 267, 282, 298, 315, 333, 352, 372, 393 and 417 K. Exchange was attempted every 10 ps, and the protein configurations were saved every 0.5 ps for later analysis. During the simulation, the PPC charges were updated every 7.5 ns of REMD simulation until a steady state was reached. Subsequent REMD steps made use of the lowest energy structure of the previous step as the starting structure for simulation and PPC calculation.

2.3 REMD simulation to study the folding thermodynamics of Trp-cage

Next, another REMD simulation was performed to study the folding/unfolding equilibrium thermodynamics of Trp-cage by employing the steady state-derived PPC obtained from the 3-step PPC update. The temperatures of the 16 replicas are 253, 267, 282, 298, 315, 333, 352, 372, 393, 417, 443, 472, 502, 535, 570 and 607 K. These temperatures give a uniform exchange acceptance ratio of about 12% under the converged PPC charge. A total of 2,500 exchange attempts were performed (corresponding to a 25 ns simulation for each replica). This process results in a total of 0.8 million configurations and an aggregate MD integration time of 0.4 μs.

All the simulations in this work use the generalized Born surface area (GBSA) model (modified Bondi radii, solvent dielectric constant 78.5, surface tension 0.005 kcal/mol $Å^2$) [20], mimicking solvent effects. This system is small enough such that all intrapeptide nonbonded interactions are included in the calculation. The time step is 2 fs, and the SHAKE algorithm [45] with relative geometrical tolerance of 10^{-5} is used to constrain all the bonds involving hydrogen atoms. To avoid occasional instability in dynamics runs, replica temperatures were maintained at an appropriate value by a combination of velocity reassignment from a Maxwell–Boltzmann distribution and Berendsen's temperature coupling method [46].

3 Results and discussions

3.1 Sample Trp-cage native conformation

Since partial atomic charge corresponding to a particular conformation fluctuates throughout the entire molecular dynamics simulation, charge re-distribution should be considered in simulations, especially protein folding, whereby electrostatic interaction plays an important role in maintaining protein stability. In order to study the possible influence of a change in protein conformation on atomic partial charge that affects folding processes, a 3-step PPC charge update method was proposed to simulate the folding process of Trp-cage, which is illustrated in Fig. 1. The results for the charge update method were then analyzed in terms of RMSD and potential energy (Fig. 2), and free energy landscapes (Fig. 1).

The simulation first begins from an extended linear Trp-cage structure (Fig. 1a), which is equilibrated using the AMBER program. It then rapidly collapses into a compact structure (Fig. 1b). This structure is used for both PPC calculation and initial structure for the first REMD simulation of 7.5 ns (Fig. 1, S1). In the RMSD graph of S1 (Fig. 2, S1, left), the graph is obtained by comparing the ensemble of structures obtained from S1 REMD simulation with reference to its initial structure (Fig. 1b). Cluster analysis is then conducted on the ensemble of structures generated where the ensemble with the lowest energy is selected to be the representative structure for the next REMD step. The same process continues for both REMD simulations S2 and S3 (Fig. 1) until the RMSD fluctuations have stabilized within 1 Å. (Figure 2, S3, left). The purpose of S3 is to confirm whether the folding has reached a

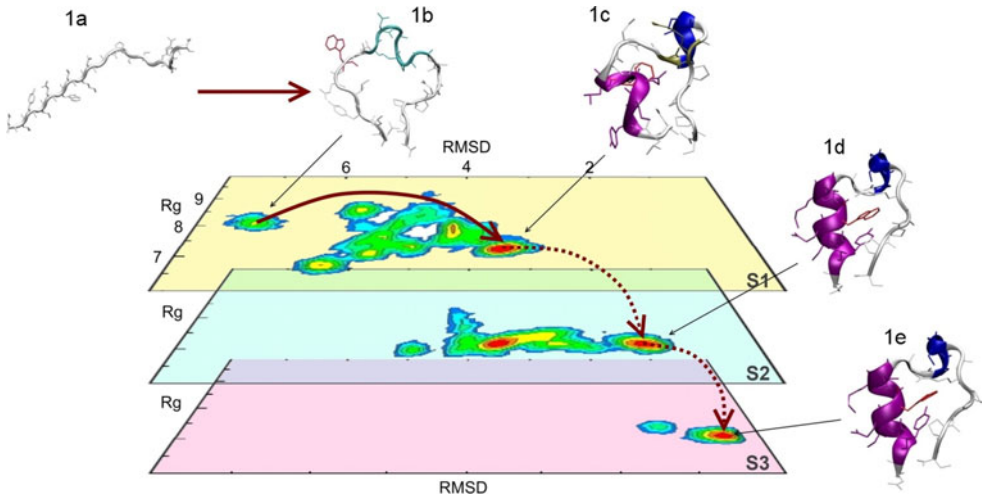

Fig. 1 Representative configurations of the Trp-cage during the 3-step REMD simulations: (*1a*) fully extended conformation; (*1b*) the conformation after 80 ps equilibration at 298 K, which is used as the starting structure for REMD S1; (*1c*) the conformation in the lowest potential energy cluster in REMD S1, which is used as the starting structure for REMD S2; (*1d*) the conformation in the lowest potential energy cluster in REMD S2, which is used as the starting structure for REMD S3; (*1e*) the conformation in the lowest potential energy cluster in REMD S3, which is <1 Å Cα REMD from the Trp-cage native structure; the folding-free energy landscape of the Trp-cage in each REMD is also shown. All of the data were extracted from 253 K replica

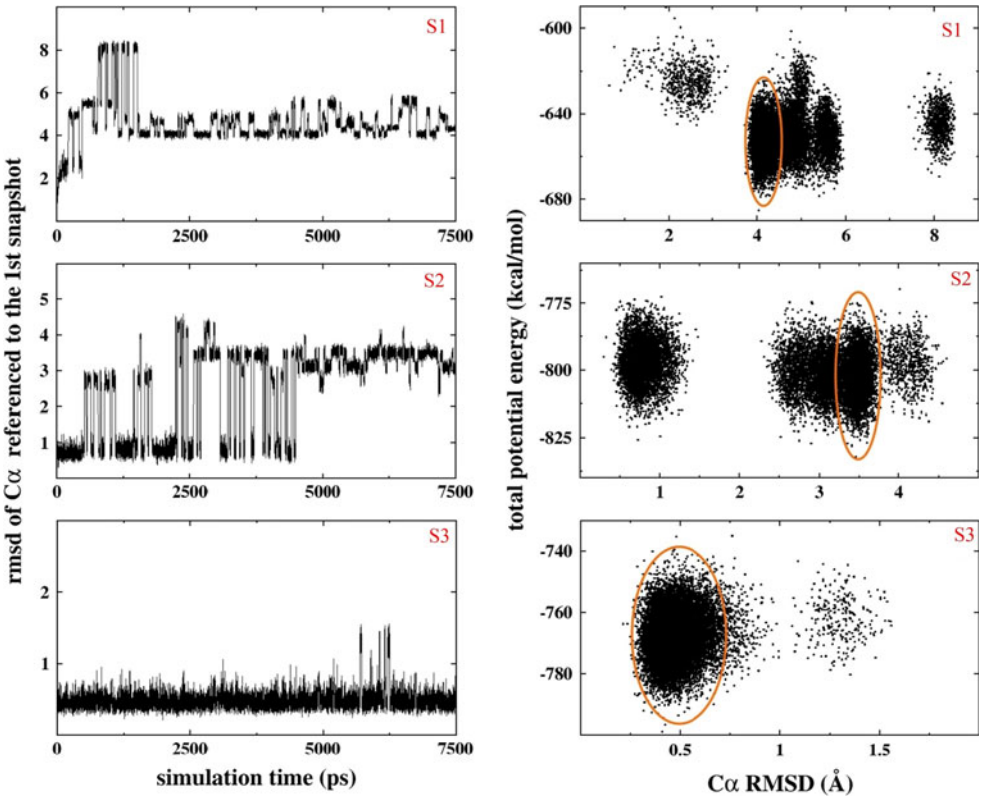

Fig. 2 *Left* RMSD of Cα referenced to the first structure in each REMD simulation. *Right* total potential energy versus Cα RMSD relative to the first structure. The cluster with the lowest potential energy in S1–S3 is circled

steady state and the charges have been converged. Under the charge derived from S2 which is a folded state close to the native structure, it is possible that the charges are biased a bit. However, in the REMD to study the folding thermodynamics of Trp-cage, a larger temperature range (253–607 K) is applied to minimize this happening. The low RMSD fluctuation together with the presence of one significant cluster indicates that the folding of protein has reached steady states and the comparison with the NMR structure affirmed that the native state has been obtained (<1.0 Å Cα RMSD). Thus, from such a 3-step PPC update method, we are able to sample conformations that are very

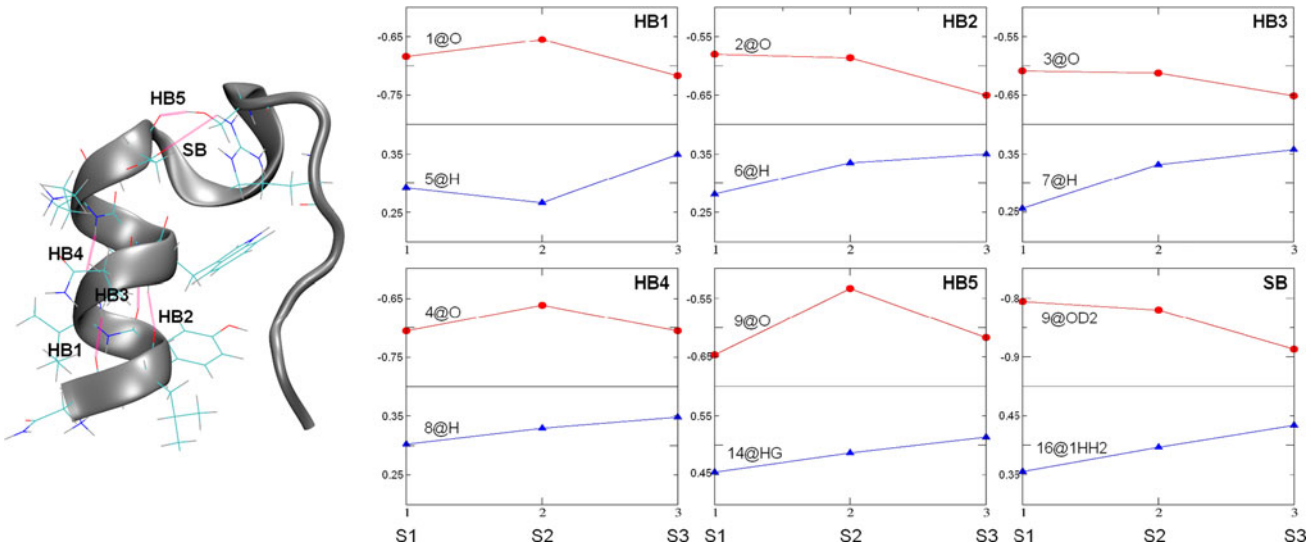

Fig. 3 Charge variation of the atoms involved in hydrogen bonds and salt bridge in three stages (S1–S3)

similar to the experimental folded state. The charge variation in the three stages is also presented in Fig. 3 to show the influence of the conformational change in PPC charges. For simplicity, only the charges of the atoms involved in hydrogen bond and salt bridge are shown. As we can see, with the re-evaluation of the PPC charge for three times, partial atomic charges are generally more positive for hydrogen bond donor and more negative for hydrogen bond acceptor. The same trend occurs for the salt bridge of D9-R16, which clearly shows the charge evolution of the salt bridge when the structure move progressively toward to the native. Although the charge re-evaluation for only 3 times may not be enough for other proteins during the whole-folding/unfolding simulation. This work, nevertheless, represents the first try during our development of on-the-fly charge update scheme in REMD simulation.

3.2 Thermodynamic study of Trp-cage based on PPC

Since replica exchange simulations under 3-step PPC update can sample conformations that are very similar to the experimental folded state, we can then study the thermodynamic principles of folding and stability of Trp-cage by performing a brand-new REMD simulation, which is simulated over a larger temperature range and in much wider conformational space to obtain a reliable thermodynamic properties. In the simulation, we simply replace the standard charges from the AMBER force field by structure 1e (shown in Fig. 1)-derived PPC while keeping the rest of force field parameters intact. During the whole simulation course, the PPC that is supposed to be accurate for structures in the vicinity of native state will be fixed and the temperature dependence of this charge will be explored.

To determine the fraction of folded structures, we use the Cα RMSD from the NMR structures as a measure to distinguish the folded and unfolded conformations. Specifically, if the RMSD of Cα atoms of a conformation is less than a given value to any of the 38 structures of the Trp-cage NMR ensemble, the structure is considered to be folded. The specific threshold value of RMSD is determined from the statistical distribution of RMSD shown in Fig. 4 left panel. As seen, distribution of the Cα RMSD under PPC charge exhibit a narrow peak I at ~1.7 Å, representing the folded configuration. Next, we use 2.5 Å in which a clear separation between folded and unfolded structures is visible as a threshold to estimate the Trp-cage population in each replica. The right panel in Fig. 4 shows the fraction of Trp-cage conformations that are within 2.5 Å Cα RMSD from any of the 38 structures of Trp-cage NMR ensemble as a function of temperature. The experimental unfolding molar fractions from NMR CSD and CD spectrums [17] are also shown in the *Inset* for comparison. It can be seen that the simulation under PPC decays similarly to that from the experiment, with a melting transition temperature around 325 K, which is in excellent agreement with the experimental 315 K. It is noted that the PPC charge set might not be suitable for high-temperature replicas; however, it will not damage the folding/unfolding fraction statistics too much because of the tight threshold of 2.5 Å. At high temperatures, Trp-cage will undergo a significant conformational change, which will be very different from the native structure no matter whether the polarization effect is considered or not.

Similarly, we have also run a REMD simulation using AMBER charge to determine its thermodynamic properties (shown in supporting information). However, with the use of AMBER charge, its melting transition temperature

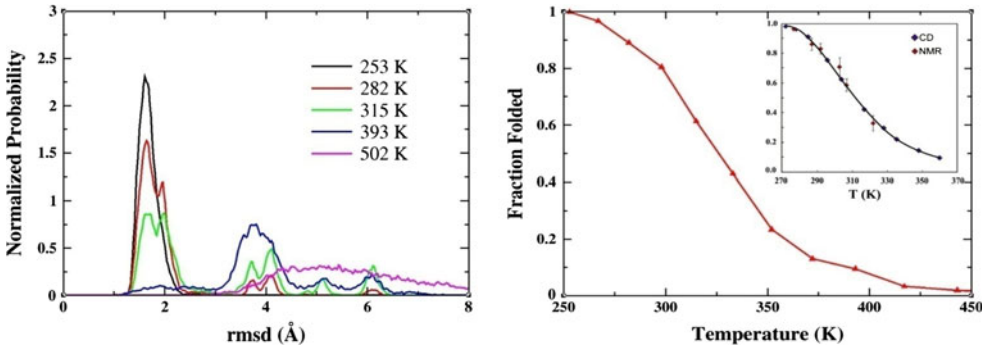

Fig. 4 Folding/unfolding equilibrium of the Trp-cage in the brand-new REMD simulation. *Left* Cα RMSD distributions for 5 selected temperatures under the structure 1e-derived PPC charge. *Right* the fraction of the Trp-cage population as a function of temperature under the PPC charge. The experimental results from CD and NMR CSD are shown in the *inner box* for comparison. The melting transition temperature in the simulation under the PPC is found to be ≈325 K, which is in excellent agreement with the experimental value of 315 K, indicating the good temperature dependence of the structure 1e-derived PPC charge

differs greatly from experimental value. Since we replaced AMBER charge by PPC while keeping all the other parameters in the force field, it can be concluded that the drastic difference in thermodynamic properties is due to the change in the charges involved. As mentioned earlier, the PPC solves the electronic response problem by combining quantum chemistry calculation of protein with the continuum solvation model. Since molecular polarizability is explicitly included in quantum mechanical calculation, electronic polarization information of the Trp-cage mini-protein will be embedded in the PPC. Therefore, it can give a better description of the polarized electrostatic state of the Trp-cage in both the native structure and the structures in its vicinity and may correctly describe the temperature dependence of the Trp-cage native structure, giving a melting temperature close to the experiment value. Since polarizable force field models such as AMOEBA can in principle incorporate polarization effect, our future work will target the comparison of the behavior of the PPC charges with the use of polarizable force field.

Here, we would like to compare our simulation results with another reported research [28] with the same simulation system. In that research, GB radii for hydrogen atoms bound to charged nitrogen atoms were tuned to make the simulated T_m agree with experimental value. We note that the scaling factor proposed in Ref. 23 may not be applicable in the folding dynamics of other proteins. In our scheme, without any ad hoc tuning of parameters, we simply replace AMBER charge by quantum-derived PPC while keeping the other parameters in the force field. Thus, the difference caused by this charge switch should be noteworthy.

Acknowledgments This work is supported by Nanyang Technological University Start-up Grant and Tier 1 grant and National Natural Science Foundation of China (Grants No. 20803034). Authors thank New York University and Nanyang Technological University for computational resources.

References

1. Daggett V, Fersht A (2003) Nat Rev Mol Cell Biol 4:497–502
2. Shakhnovich E (2006) Chem Rev 106:1559–1588
3. Thirumalai D, Hyeon C (2005) Biochemistry 44:4957–4970
4. Taubes G (1996) Science 271:1493–1495
5. Thomas PJ, Qu B-H, Pedersen PL (1995) Trends Biochem Sci 20:456–459
6. Hooper C (1991) J NIH Res 3:65–70
7. Rawls RL (1997) Chem Eng News 75:39–41
8. McCammon JA, Wolynes PG (2002) Curr Opin Struc Biol 12: 143–145
9. Zhou Y, Karplus M (1999) Nature 401:400–403
10. Brooks CL, Gruebele M, Onuchic JN, Wolynes PG (1998) Proc Natl Acad Sci 95:11037–11042
11. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Nature 420: 102–104
12. Pitera JW, Swope W (2003) Proc Natl Acad Sci 100:7587–7592
13. Zhou R, Berne BJ, Germain R (2001) Proc Natl Acad Sci 98: 14931–14936
14. Lei H, Wu C, Liu H, Duan Y (2006) Proc Natl Acad Sci 104: 4925–4930
15. Paschek D, Hempel S, García AE (2008) Proc Natl Acad Sci 105: 17754–17759
16. Zhou R (2003) Proc Natl Acad Sci 100:13280–13285
17. Neidigh JW, Fesinmeyer RM, Andersen NH (2002) Nat Struct Biol 9:425–430
18. Hukushima K, Nemoto KJ (1996) Phys Soc Jpn 65:1604–1608
19. Hansmann UHE (1997) Chem Phys Lett 281:140–150
20. Sitkoff D, Sharp KA, Honig BJ (1994) Phys Chem 98:1978–1988
21. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WLJ (2002) Phys Chem B 105:6474–6487
22. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces, vol 11. Reidel, Dordrecht, pp 331–342
23. Day R, Paschek D, Garcia AE (2010) Prot Struct Funct Bioinform 78:1889–1899
24. Zheng WH, Gallicchio E, Deng NJ, Andrec M, Levy RM (2011) J Phys Chem B 115:1512–1523
25. Kannan S, Zacharias M (2009) Int J Mol Sci 10:1121–1137
26. Duan LL, Mei Y, Li YL, Zhang QG, Zhang DW, Zhang JZH (2010) Sci China Chem 53:196–201
27. Freddolino PL, Harrison CB, Liu YX, Schulten K (2010) Nat Phys 6:751–758
28. Okur A, Wickstrom L, Layten M, Geney R, Song K, Hornak V, Simmerling C (2006) J Chem Theory Comput 2(2):420–443

29. Xiao L, Honig BJ (1999) Mol Biol 289:1435–1444
30. Giletto A, Pace CN (1999) Biochemistry 38:13379–13384
31. Roca M, Messer B, Warshel A (2007) FEBS Lett 581:2065–2071
32. Zhang DW, Zhang JZHJ (2003) Chem Phys 119:3599–3605
33. Gao AM, Zhang DW, Zhang JZH, Zhang YK (2004) Chem Phys Lett 394:293–297
34. Mei Y, He X, Xiang Y, Zhang DW, Zhang JZH (2005) Prot Struct Funct Bioinform 59:489–495
35. Mei Y, Zhang DW, Zhang JZHJ (2005) Phys Chem A 109:2–5
36. Mei Y, Ji CG, Zhang JZHJ (2006) Chem Phys 125:094906
37. Honig B, Sharp KA, Yang ASJ (1993) Phys Chem 97:1101–1109
38. Tannor DJ, Marten B, Murphy R, Friesner RA, Sitkoff D, Nicholls A, Honig B, Ringnalda M, Goddard WAJ (1994) Am Chem Soc 116:11875–11882
39. Lu Y, Mei Y, Zhang JZH, Zhang DW (2010) J Chem Phys 132: 131101
40. Wei CY, Mei Y, Zhang DW (2010) Chem Phys Lett 495:121–124
41. Ji CG, Mei Y, Zhang JZH (2008) Biophys J 95:1080–1088
42. Bayly CI, Cieplak P, Cornell W, Kollman PAJ (1993) Phys Chem 97:10269–10280
43. Cornell WD, Cieplak P, Bayly CI, Kollman PAJ (1993) Am Chem Soc 115:9620–9631
44. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PAJ (1995) Am Chem Soc 117:5179–5197
45. Ryckaert JP, Ciccotti G, Berendsen HJCJ (1977) Comput Phys 23:327–341
46. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JRJ (1984) Chem Phys 81:3684–3690