http://dx.doi.org/10.5012/bkcs.2012.33.3.828

Comparative Study of Implicit and Explicit Solvation Models for Probing Tryptophan Side Chain Packing in Proteins†

Changwon Yang and Youngshang Pak*

Department of Chemistry and Institute of Functional Materials, Pusan National University, Busan 609-735, Korea
**E-mail: ypak@pusan.ac.kr*
Received November 10, 2011, Accepted December 29, 2011

We performed replica exchange molecular dynamics (REMD) simulations of the tripzip2 peptide (beta-hairpin) using the GB implicit and TI3P explicit solvation models. By comparing the resulting free energy surfaces of these two solvation model, we found that the GB solvation model produced a distorted free energy map, but the explicit solvation model yielded a reasonable free energy landscape with a precise location of the native structure in its global free energy minimum state. Our result showed that in particular, the GB solvation model failed to describe the tryptophan packing of trpzip2, leading to a distorted free energy landscape. When the GB solvation model is replaced with the explicit solvation model, the distortion of free energy shape disappears with the native-like structure in the lowest free energy minimum state and the experimentally observed tryptophan packing is precisely recovered. This finding indicates that the main source of this problem is due to artifact of the GB solvation model. Therefore, further efforts to refine this model are needed for better predictions of various aromatic side chain packing forms in proteins.

Key Words : Replica exchange molecular dynamics simulation, Implicit solvation model, Tryptophan side chain packing, Protein folding

Introduction

Developments of efficient simulation strategies in conjunction with all-atom force fields have provided an important theoretical basis for understanding protein folding problems. The recent progress of effective simulation methods enables all-atom direct folding simulation of small proteins. In particular, predictions of native structures of small proteins are possible by computing free energy landscapes *via* direct folding approaches at all-atom level.[1] In such free energy based protein native structure predictions, however, the validity of empirical force fields is of critical importance to outcome of native structure predictions, since the force field is required to satisfy the condition that protein native structures should be located in the global free energy minimum state. Since proteins are mostly in aqueous environment, use of well-established solvation models is also crucial for protein folding studies. Although explicit solvation models offer more accurate representation of protein solvation environments, they are computationally demanding for direct folding simulations. Therefore, as an alternative to explicit solvation model, implicit solvation models can be a practical choice for long time folding simulations due to computational efficiency. Among many implicit solvation models, the Generalized Born (GB) implicit solvation model with surface area correction (SA)[2] has been one of the popular choices and used for a wide range of biomolecular simulations. Until now, we have endeavored to develop a transferable all-atom protein force field with the GBSA model, so that free energy based native structure predictions of diverse structural motifs become feasible at the all-atom resolution. In a previous study,[3] a modified version of the all-atom force field (param99MOD5/GBSA) was presented by training the backbone dihedral angle parameters of the param99 force field[4] and the GB intrinsic radius parameters in the GB implicit solvation model of Onufriev, Bashford, and Case [GB(OBC)],[5] such that each native-like structure of a training set containing α-helix, β-hairpin, and ββα folds was required to be in its global free energy minimum state. Due to this tuning process, the modified force field can predict secondary structure propensities in more balanced ways. As shown by several test simulation studies,[1] we demonstrated that this force field could differentiate native-like structures of an extended range of protein folds based solely on the first principle. Recently, during the process of the force field validation, despite successful applications of the aforementioned force field to various structural motifs, we found that this force field displayed some limitations for incorporating cross-strand aromatic side chain interactions in protein structures. The cross-strand aromatic interaction plays a role in governing structures and stability of some proteins, since it is associated with the preferential formation of hydrophobic clusters. In this work, as a benchmark system to probe aromatic side chain packing *via* the modified force field, we focus on tryptophan zipper (trpzip2). Trpzip2 is a 12-residue designed β-hairpin (PDB entry: 1LE1). Its hydrophobic core is maintained solely by two pairs of the Trp side chain interactions locking up the hairpin tightly (Trp residues 2 and 11 and Trp residues 4 and

†This paper is to commemorate Professor Kook Joe Shin's honourable retirement.

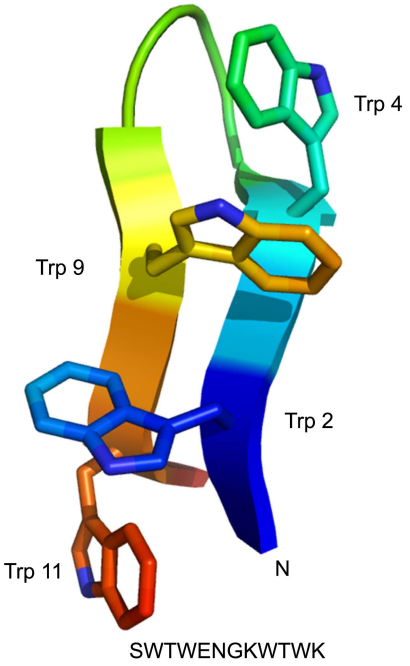

Figure 1. The NMR native structure of trpzip2 (PDB ID: 1LE1). The N-terminal is NH_3^+ and the C-terminal is NH_2. The two Trp pairs (4:9, 2:11) in the NMR structure interact in the edge to face (EtF) orientation. Here, "N" indicates the N-terminal.

9) (Figure 1). These Trp-Trp pair interactions led to an unusually high stability of this hairpin, as shown by a relatively high melting temperature (345 K).[6] Furthermore, the two Trp pairs adopt an interesting edge-to-face (EtF) geometry. Previously, using explicit-solvent MD simulations, Brooks and his coworker[7] found that the EtF packing in tripzip2 was caused by Trp side chain electrostatic interactions. Thus, in the framework of the GB implicit solvation model, it is of computational interest to see if param99MOD5/GBSA can predict the experimentally observed Trp side chain packing in trpzip2. In order to test our modified force field, we constructed a fully converged free energy map of trpzip2 by carrying out the replica exchange molecular dynamics (REMD) simulation[8] with param99MOD5/GBSA. Unfortunately, the REMD result of param99MOD5/GBSA showed that the native like structure of trpzip2 is not globally stable and its Trp packing is mostly a parallel-displaced (PD) one in contrast to the experimental observation. In order to find a source of this discrepancy, we also carried out the explicit-solvent REMD simulation using the same protein force field (param99MOD5) with an explicit solvent model (TIP3P). The explicit-solvent free energy surface appears to more realistic in that the native-like structure was found in the lowest free energy minimum basin with the correct Trp-Trp packing geometry reproduced. Therefore, the GB implicit solvation model may be responsible for the incorrect formation of the Trp side chain packing.

Method

In the present work, the paramMOD5 force field was used for the protein part in combination with the GB (OBC) implicit solvation model and the explicit solvent model (TIP3P).[9] The REMD simulation was employed for accelerating conformational searches of proteins on complex energy landscapes. This method involves a series of parallel independent MD simulations using several replicas of system. Each replica is assigned with a different temperature, so that the simulation covers the system with an extended range of temperatures. The key point of this method is a frequent trajectory exchange attempt for two adjacent replicas at temperatures T_i and $T_{i\pm1}$. This exchange event is accepted with a probability of $P=\min\{1,\exp[(\beta_i-\beta_{i\pm1})(U_i-U_{i\pm1})]\}$, where β_i is the inverse temperature ($1/k_BT$) and U_i is the total potential energy of the replica at T_i.

For implicit-solvent REMD simulations, a total of the 8 replicas were employed for a temperature range of 234-540 K. The acceptance probability of the exchange attempt is 36-38%. The cutoff distance of the non-bond interactions is 24 Å. This simulation started from a fully extended conformation. For explicit-solvent REMD simulations, use of a total of 44 replicas spans a temperature space of 284-530 K. Here, the protein molecule (tripzip2) was solvated by a total of 1581 TIP3P water molecules in a rhombic dodecahedron simulation box of which the image distance is 41.5 Å. For the charge balance, two chloride ions (Cl^-) were added to the simulation box. The cutoff distance of the van der Waals (VDW) interaction is 10 Å. The electrostatic interaction term was computed by the particle mesh-Ewald (PME) scheme with a cutoff length of 10.0 Å. This explicit-solvent simulation was run using the GROMACS program.[10] The velocity-Verlet integration with a time step of 2.0 fs was employed using the modified Berendsen thermostat by Bussi *et al.*[11] with a coupling constant of 0.1 ps. Compared to the original Berendsen thermostat,[12] this modified thermostat was improved for producing more accurate canonical distributions. Recently, it was found that for more reliable interpretation of simulation results, the REMD simulation should be combined with more improved thermostat leading to accurate canonical distributions.[13] The bond distances consisting of hydrogen atom were constrained by using RATTLE.[14] The replica exchange interval was 0.5 ps and each replica trajectory was saved at every 1.0 ps. The cutoff lengths of all non-bonding interactions and the GB solvation term were 24.0 Å. The acceptance rate for the exchange attempt is 30-40%. For the simulation using the explicit solvent, starting from the NMR native structure, a normal MD simulation was performed at 450 K for 10 ns. The resulting MD structures ranging from the native state to various unfold ones were randomly assigned to each replica for initiating the subsequent REMD simulation.

Result and Discussion

The REMD simulation with the GBSA model was run for a total of 500 ns for each replica. The first 200 ns data were discarded and the remaining set was used in the data analysis. The reason for such a large exclusion of the initial

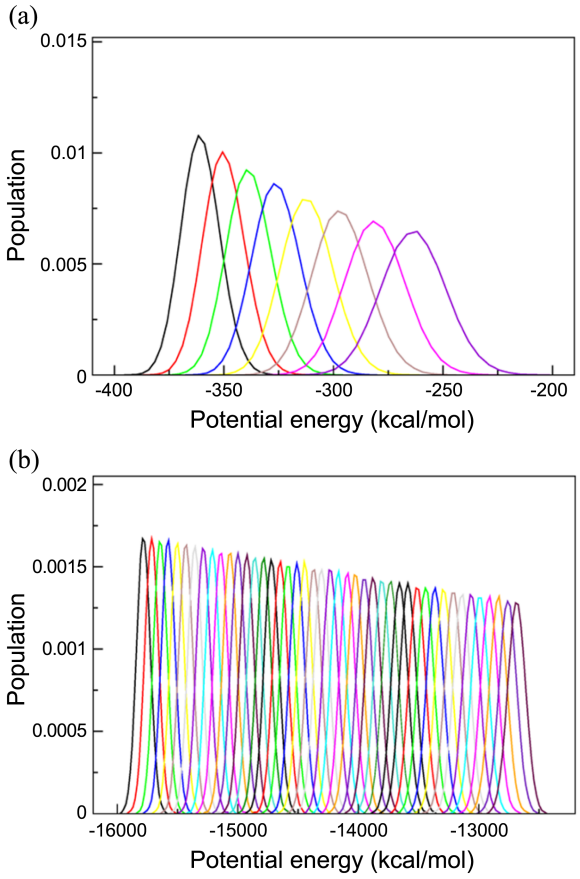

Figure 2. Potential energy distribution of each replica generated by REMD (a) the GBSA water model (b) the TIP3P water model.

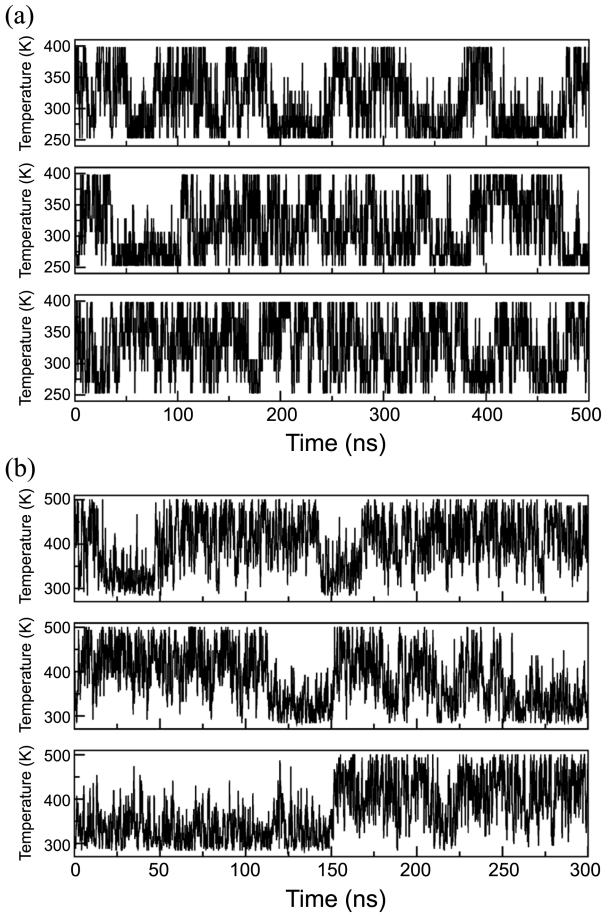

Figure 3. The exchange profile in temperature space for three different replicas at (a) T = 398.0 K (upper), 328 K (middle), and 307 K (lower) for the GBSA implicit solvation model (b) T = 500.8 K (upper), 372.3 K (middle), and 288.0 K (lower) for the TIP3P explicit solvation model. As an indication of good conformational mixing, each of selected replicas frequently visited from low to high temperature states. Thus, the replicas show a random walk in temperature space.

data is mainly due to a slow convergence of the simulation starting from a fully extended conformation. For the simulation using TIP3P solvation,[9] a total of 300 ns was run for each replica and the first 50 ns trajectory was excluded for the analysis. Figure 2 shows the potential energy distribution of each replica generated by REMD using the explicit and implicit solvation models. Each distribution is smooth and the distributions of any adjacent replicas are well overlapped, indicating that a reasonable conformation sampling is achieved. For monitoring the extent of sampling, we presented the trace indices of several replicas for both solvent models (Figure 3). Figure 3 clearly showed the replicas frequently visited from low to high temperature states and thereby sufficient conformational mixing in the temperature space occurred. In order to map the free energy surface from the REMD trajectory, two different sets of two-dimensional (2D) reaction coordinates: (*rmsd*, R_g) and (*rmsd*, L) were introduced. The *rmsd* is the root mean square deviation from the NMR native structure, R_g is the radius of gyration, and L is the sum of inner native hydrogen-bond distances from nitrogen to oxygen atom and distances between CD2 atoms of the three tryptophan residue pairs.[15] For the GBSA and TIP3P solvation models, the resulting free energy contour maps with (*rmsd*, R_g) and (*rmsd*, L) were given in Figure 4 and Figure 5, respectively. Unfortunately, as shown in Figures 4 and 5, the lowest free energy minimum structure from the implicit solvation somewhat substantially deviates from the native NMR structure. Instead, the most populated state was located in the partially fold region (*rmsd*, R_g) = (3.0, 6.0 Å) or (*rmsd*, L) = (2.0, 3.0 Å). In fact, a similarly distorted free energy profile was also reported in a previous simulation study using param99SB/GBSA.[16] In comparison with the implicit-solvent free energy surface, the TIP3P water model in conjunction with param99MOD5 produced qualitatively more realistic free energy representation of trpzip2 by correctly matching the lowest free energy predicted structure to the NMR native structure. Moreover, it is striking that the most populated structure was accurately predicted within a *rmsd* = 0.5 Å. In comparison, the present explicit water free energy map is similar to a recent simulation result using param99SB with explicit solvation.[17]

For the implicit and explicit solvation models employed in this work, the resulting lowest free energy minimum structures are shown in superimposition with the NMR native geometry in Figure 6. As shown in Figure 6(a), the Trp side chain core in the most populated conformer prefers the PD

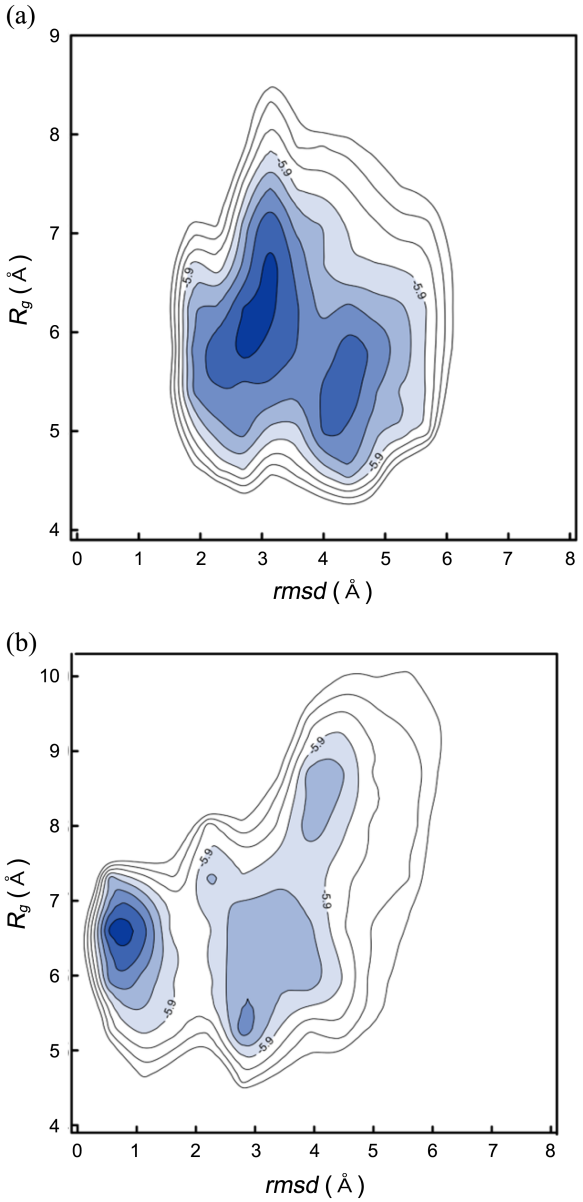

Figure 4. Free energy surfaces with (*rmsd*, R_g) using the GBSA model (a) and the TIP3P model (b). Herein, *rmsd* is the backbone root mean square deviation from the NMR native structure and R_g is the radius of gyration.

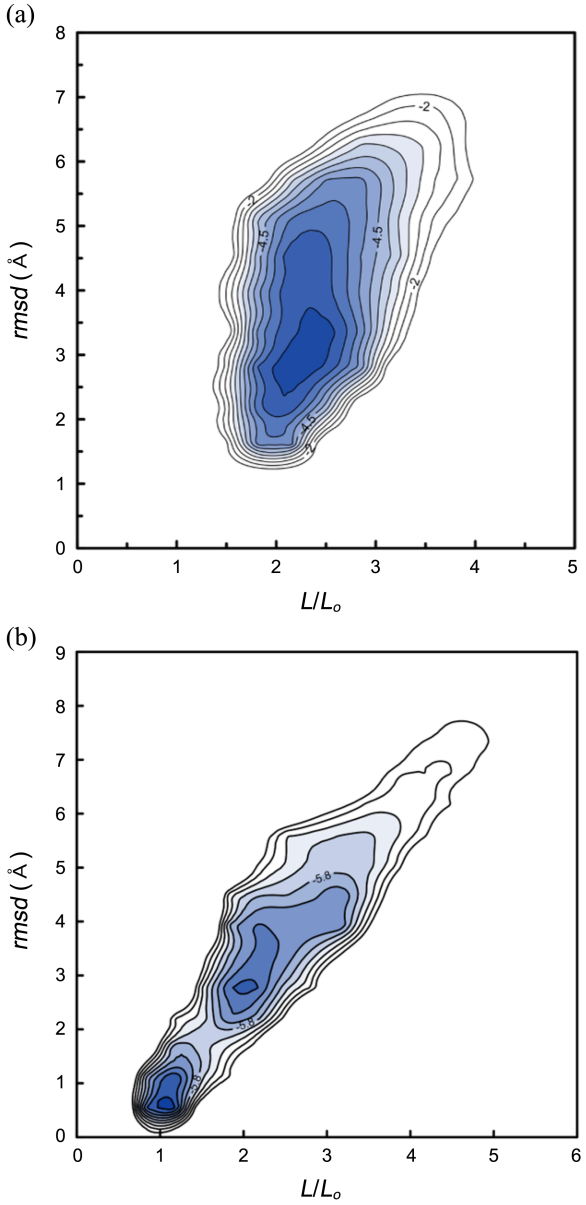

Figure 5. Free energy surfaces with (*rmsd*, *L*) using the GBSA model (a) and the TIP3P model (b). The *L* coordinate is defined to be the sum of inner native hydrogen-bond distances from nitrogen to oxygen atom and distances between CD2 atoms of the three tryptophan residue pairs.

packing to EtF one when the GB implicit solvation model was utilized. Therefore, the packing form of the Trp pairs was clearly misrepresented by the modified force field using the implicit solvation. Due to this mispacking, non-native conformer with a shifted turn was most stable. On the other hand, consistent with the NMR experiment, the global free energy minimum structure with the explicit solvation (Figure 6(b)) revealed that the two Trp side chain pairs (Trp residues 2 and 11 and Trp residues 4 and 9) of the lowest free energy predicted conformer interact in the EtF orientation. Making a note of the significantly improved performance of the explicit solvation model as opposed to the GBSA solvation model, the GB solvation model needs to be further refined for an improved prediction of the Trp side chain orientation in trpzip2. In particular, as a way to tune the GB solvation model toward better aromatic packing, relevant GB intrinsic radii in the Trp ring may be adjusted until the correct side chain packing geometry is recovered. The research on this issue is undergoing. Figure 7 is the denaturation profile from the TIP3P water simulation. Despite the improved free energy landscape and the correct Trp packing using the explicit solvation, the simulated melting profile is still in disagreement with the experiment with an underestimation of the folded state population. As a result, the predicted melting temperature is too much underestimated by 30 K. This underestimation of the melting temperature is in line with the observed small energy difference between

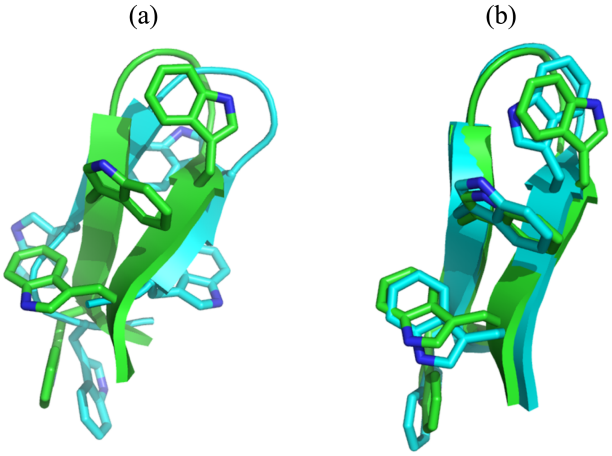

Figure 6. The lowest free energy predicted structure (cyan) superimposed with the NMR native one (green) from using the GBSA (a) and TIP3P (b) water models. With the TIP3P water model, the experimentally observed the EtF packing of the Trp residues was correctly predicted. However, with the GBSA water model, incorrect tryptophan packing of the PD form was observed.

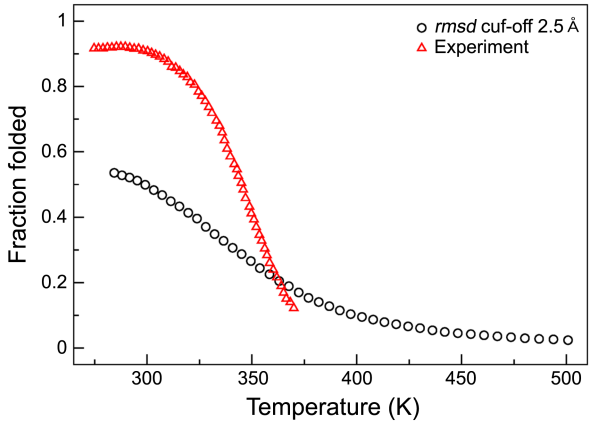

Figure 7. Thermal denaturation profile using REMD with param99MOD5/TIP3P. The folded fraction was obtained by using a criterion of *rmsd* < 2.5 Å.

the folded and unfolded states (< 2.0 k_BT). In order to reduce the gap between theory and experiment further, we expect that the param99MOD5 force field requires more adjustment together with the GB solvation model, which will be a future work.

The param99MOD5 force field was derived from param99 for more balanced treatments of α/β mixtures of proteins. This force field in conjunction with the GB implicit solvation model has been successfully applied to direct folding studies of diverse folds. Nevertheless, as shown in this work, use of param99MOD5 with explicit solvation seems to give more reliable results by circumventing the artifact of the implicit solvation model.

Acknowledgments. YP thanks the National Research Foundation of Korea (2010-0015929) for the financial support. The authors would like to acknowledge the support from the KISTI supercomputing center through the strategic support program for supercomputing application research (KSC-2010-C1-0036).

References

1. (a) Kim, E.; Jang, S.; Pak, Y. *J. Chem. Phys.* **2008**, *128*(17), 175014. (b) Kim, E.; Jang, S.; Pak, Y. *J. Chem. Phys.* **2009**, *131*(19), 195102. (c) Kim, E.; Jang, S.; Lim, M.; Pak, Y. *J. Phys. Chem. B* **2010**, *114*(22), 7686-7691.
2. Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129-152.
3. Kim, E.; Jang, S.; Pak, Y. *J. Chem. Phys.* **2007**, *127*(14), 145104.
4. Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*(12), 1049-1074.
5. Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*(2), 383-394.
6. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *P Natl. Acad. Sci. USA* **2001**, *98*(10), 5578-5583.
7. Guvench, O.; Brooks, C. L. *J. Am. Chem. Soc.* **2005**, *127*(13), 4668-4674.
8. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*(1-2), 141-151.
9. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*(2), 926-935.
10. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*(3), 435-447.
11. Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*(1), 014101.
12. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*(8), 3684-3690.
13. Rosta, E.; Buchete, N. V.; Hummer, G. *J. Chem. Theory Comput.* **2009**, *5*(5), 1393-1399.
14. Palmer, B. J. *J. Comp. Phys.* **2003**, *104*(2), 470-472.
15. Snow, C. D.; Qiu, L. L.; Du, D. G.; Gai, F.; Hagen, S. J.; Pande, V. S. *P Natl. Acad. Sci. USA* **2004**, *101*(12), 4077-4082.
16. Hayre, N. R.; Singh, R. R. P.; Cox, D. L. *J. Chem. Phys.* **2011**, *134*(3), 035103.
17. Zhang, C.; Ma, J. P. *J. Chem. Phys.* **2010**, *132*(24), 244101.