Using the generalized Born surface area model to fold proteins yields more effective sampling while qualitatively preserving the folding landscape

Peng Tao and Yi Xiao[*]
School of Physics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

(Received 17 April 2020; accepted 1 June 2020; published 22 June 2020)

Protein folding is a long-standing problem and has been widely investigated using molecular dynamics simulations with both explicit and implicit solvents. However, to what extent the folding mechanisms observed in two water models agree remains an open question. In this study, *ab initio* folding simulations of ten proteins with different topologies are performed in two combinations of force fields and water models (ff14SB+TIP3P and ff14SBonlysc+GB-Neck2). Interestingly, the latter combination not only folds more proteins but also provides a better balance of different secondary structures than the former in the same number of integration time steps. More importantly, the folding pathways found in the two types of simulations are conserved and they may only differ in their weights. Our results suggest that simulations with an implicit solvent may also be suitable for the investigation of the mechanism of protein folding.

DOI: 10.1103/PhysRevE.101.062417

I. INTRODUCTION

With the increase of computational power, molecular dynamics simulation has become more and more important in modern research, for example, it has been widely used in drug design [1–4], structural prediction and refinement [5–8], and folding mechanisms [9–15]. There are two commonly used models in atomistic molecular dynamics simulations, namely explicit-solvent molecular dynamics simulations (ESMDSs) and implicit-solvent molecular dynamics simulations (ISMDSs). In ESMDSs, water molecules are treated explicitly, which can provide high-resolution descriptions of stabilizations and dynamics of biomolecules. For example, using Anton, a special-purpose supercomputer, Shaw and coworkers fulfilled the reverse folding of a set of small proteins ranging from 10 to 80 amino acids [13]. However, such kinds of millisecond-scale simulations are still inaccessible to commonly used GPUs. Thus, in spite of its high resolution, the applications of ESMDSs are still limited on timescales in the range of tens of microseconds, which are still much shorter than the timescale of the folding process for most proteins. In such cases, a more reasonable choice is ISMDS because the calculations and sampling can be speeded up significantly. For example, since the water molecules are treated as a continuum media, the computational cost of ISMD is greatly reduced, resulting in the simulation speed increasing from 200–400 ns/day in explicit solvent to 800–1400 ns/day in implicit solvent for typical fast-folding proteins on a single Nvidia RTX 2080 GPU card. Besides, the energy landscape in implicit solvent models is considerably smoother than that in explicit water models, which makes the sampling in ISMDSs faster [16]. However, it is not well established whether the folding mechanisms supported by ESMDSs are in line with those of ISMDSs. To address this issue, we directly compare the folding mechanisms obtained by ESMDS and ISMDS for ten widely used model proteins with different secondary structures and topologies.

It should be pointed out that there have been hundreds of studies focusing on the effects of water on the stabilities and dynamics of proteins, but this work is different from theirs in the following aspects.

The force fields we applied are the newest Amber force fields, ff14SB and ff14SBonlysc [17]. Both of them are successors of ff99SB [18], one of the most widely used force fields, and contain the systematic-refitting side-chain dihedral parameters based on high-level quantum mechanics calculation. The difference between them is that, in addition to the side-chain parameters, small empirical-adjustment backbone dihedral parameters based on TIP3P simulations are further incorporated, resulting in ff14SB. This force field had been tested on a small set of peptides or proteins and shown to provide better secondary structure balance and fitting of NMR data than ff99SB [17]. Notably, a recent study [19] indicated that the fluctuation profiles of a triclinic lysozyme in ff14SB were closer to the experiment model when compared to CHARMM 36 [20], which strongly suggests high accuracy for ff14SB. However, ff14SB is mainly designed for explicit solvent since its backbone parameters were obtained from simulations in TIP3P explicit water. In contrast, it has been shown that the combination of ff14SBonlysc, excluding the backbone parameters of ff14SB, and a recently developed generalized Born surface area (GBSA) model GB-Neck2 [21] exhibits remarkable successes in describing the structure and dynamics for both proteins [22,23] and nucleic acids [24]. Thus, ff14SB and ff14SBonlysc reflect the accuracy of current nonpolarizable atomistic force fields for ESMDS and ISMDS, respectively.

In this study, we mainly concentrate on the similarity of folding mechanisms obtained from current best ESMDS and

[*] Corresponding author: yxiao@hust.edu.cn

ISMDS, but not on the effects of the solvent model on the protein folding mechanisms [25,26]. Thus, the force fields used in the different solvent models could be different. Here the folding mechanisms are defined by the folding pathways (the order of contact formation) and their weights. It is worth noting that, since the folding pathways defined here are not dependent on time, enhanced sampling simulations, such as replica exchange molecular simulations [27], could also be used to analyze the folding pathways of biomolecules, as suggested by several studies [28–30].

II. METHODS

A. Simulation details

1. *Initial system building*

The initial peptide structures were extended and built by TLEAP, a general program for preparing the input files for the MD simulations in Amber [31], with all backbone dihedrals at 180°. See Table S1 in the Supplemental Material [32] for the sequences of these peptides. The force fields we used in this study for ESMDS and ISMDS were ff14SB and ff14SBonlysc, respectively. For ESMDS, the peptides were solvated with TIP3P water molecules in a cuboidal box and the minimum distance between solute atoms and box boundary was 12 Å except for SH3 and CI2 (10 Å). After that, several Na^+ or Cl^- were added to neutralize the whole system, resulting in the size of final systems ranging from 7362 to 31 929 atoms.

2. *Energy optimization, heating, and equilibrium*

For ESMDS, we first minimized the energy of the solvent with 2000 steepest descent steps and subsequent 2000 conjugate gradient steps with 10 kcal/mol Å^2 harmonic restraints exerted on the peptides, then we performed another round of minimization on the whole system, with the same parameters as in the first step except that the restraints were removed. After energy optimization, 200 ps heating and another 200 ps equilibration were performed in NPT ensemble with a time step of 1 fs and restraints of 10 kcal/mol Å^2. The temperature increased gradually from 0 to 300 K in the first 100 ps and remained 300 K in the last 100 ps during the heating. The system equilibrated to 1 bar at 300 K during the equilibration.

For ISMDS, since there were no explicit solvent molecules and ions, only one round of minimization and heating was performed. The parameters of minimization were as in ESMDS. The heating from 0 to 300 K in 200 ps was divided into six minor stages and during each stage the temperature increased 50 K.

3. *Trajectory production*

For all simulations, the time step was set to 2 fs as all the bonds containing hydrogen were constrained by the SHAKE algorithm. For each protein, at least eight independent simulations were performed for both ESMDS and ISMDS, resulting in ~1.2-ms trajectories in total. See Table S2 in the Supplemental Material [32] for the simulation time and the folding and unfolding events for each protein. The temperature was kept around 300 K by Langevin dynamics and snapshots were saved every 0.1 ns. The nonbonded cutoffs were set to 9 and 999 Å in ESMDS and ISMDS, respectively. All simulations were performed using the PMEMD.CUDA [33,34] program on Nvidia GPUs (GTX 780, 980, 1080ti, and RTX 2080). Other parameters were set to their default values.

B. Trajectory analysis

Root-mean-square deviation (RMSD) calculations (C_α-RMSD in this study) excluded flexible termini or loops that were not well defined in the experimental structures, See Table S1 in the Supplemental Material [32] for the details. The secondary structures of trajectories were calculated by the define secondary structure of proteins (DSSP) method [35], which divides secondary structure types into seven classes, namely parallel β sheet, antiparallel β sheet, 3–10 helix, α helix, π helix, turn, and bend. To simplify the following analysis, the seven classes were transformed into three bigger classes: sheet (parallel and antiparallel β sheet), helix (3–10 helix, α helix, and π helix) and turn (turn and bend). The calculations of RMSD and secondary structures were implemented by cpptraj [36,37], a primary analysis tool in Amber.

It has been shown that the fraction of native contacts, Q, is a decent coordinate to measure the folding processes or the transition paths in atomistic simulations [38]. Thus, the Q values in this study were calculated based on mdtraj [39] using the same method described in Ref. [38]. To further characterize the folding mechanisms, we first chose two boundaries, $Q_\mathrm{f} = 0.8$ and $Q_\mathrm{u} = 0.2$, to define the folded states and unfolded states, respectively. And the folding and unfolding transitions were recorded only when the protein completely transformed between the folded and unfolded states. See Table S2 in the Supplemental Material [32] for the number of recorded folding and unfolding transitions. The Φ values were calculated using the Φ_2 approximation proposed by Best *et al.* [40], that is, the Φ value of residue i is

$$\Phi(i) = \frac{\sum_\mathrm{j} p(ij|\mathrm{TP})}{N_\mathrm{i}}, \tag{1}$$

where $p(ij|TP)$ is the probability of contact ij being formed on the transition pathways and $\mathrm{N_i}$ is the number of native contacts formed by residue i. In this approximation, only the native contacts are considered but the non-native contacts are ignored, and all native contacts are equally important during folding. It should be noted that the calculations of Φ values include either the folding or the unfolding transitions, as both of them have similar Φ-value profiles (see Figs. S8, S15, S23, S30, S37, S43, S51, S56 in the Supplemental Material [32] for these profiles). Since different folding pathways may have different Φ-value profiles or folding mechanisms, for specific proteins, the whole folding pathways were clustered by the root-mean-square difference between them using the hierarchical algorithm in SCIPY [41]. In practice, the cutoffs of clustering were 0.08, 0.2, 0.1, 0.15, and 0.15 for 1E0Q, HP35, GTT, NTL9, and protein B, respectively.

III. RESULTS

A. ISMDS can fold more proteins than ESMDS in the same number of integration time steps

Sampling is a key problem in molecular dynamics simulations; we cannot know the folding mechanisms unless the

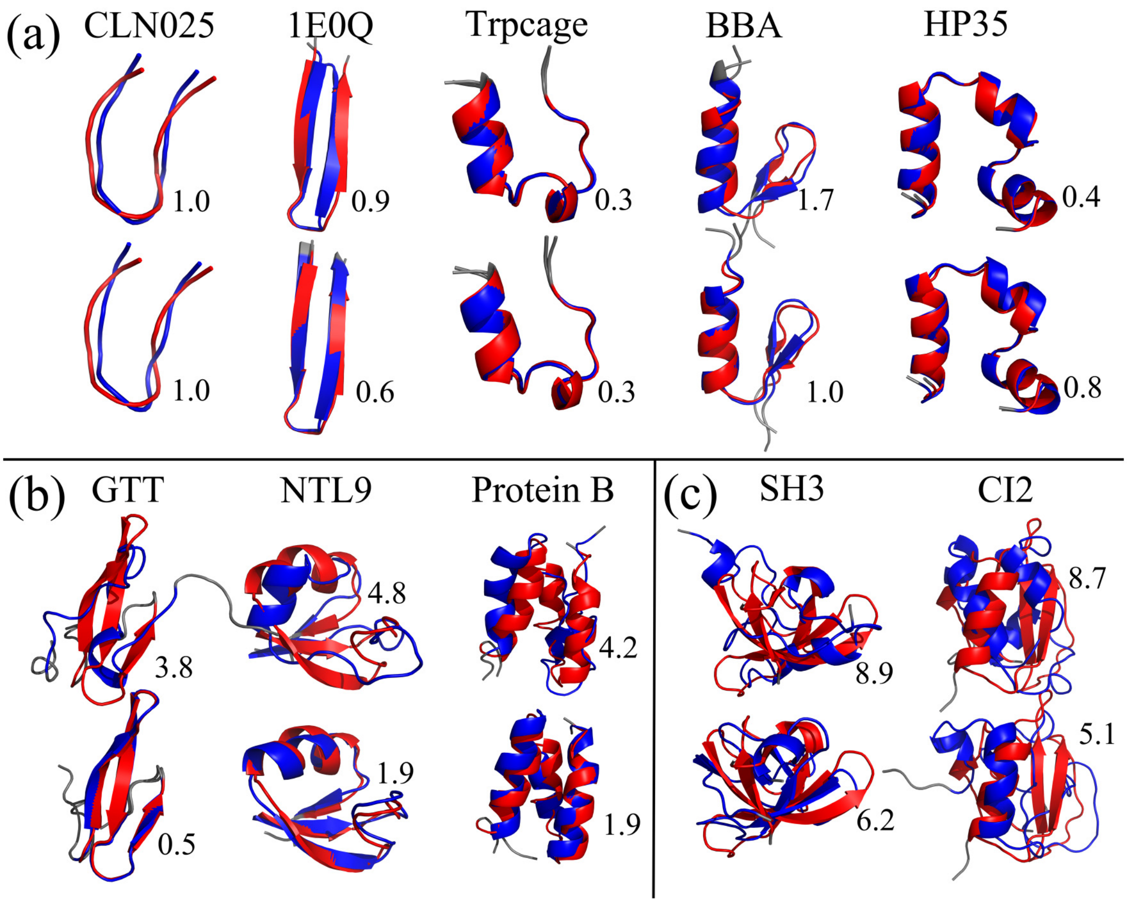

FIG. 1. The structures with the lowest RMSD in simulations (blue) are aligned to experimental structures (red). For each protein, the upper and lower structures are extracted from ESMDS and ISMDS, respectively. The RMSD values of the alignments are shown next to the structures. (a) five proteins that both ESMDS and ISMDS can find the native structure, (b) three proteins that only ISMDS can find the native structure, and (c) two proteins that both ISMDS and ESMDS cannot find the native structure. In this work, all cartoon structures are generated from PYMOL [42].

simulations can fold to the native state. To assess the sampling efficiency, we checked whether the simulations can fold to a native structure, defined here as a structure whose C_α-RMSD to the experimental structure is less than 2 Å. For each protein simulation, the structure with the lowest RMSD was extracted and then compared to the experimental structure. As shown in Fig. 1, ten proteins are divided into three classes. In the first class [CLN025, 1E0Q, Trpcage, BBA, and HP35, Fig. 1(a)] or third class [SH3 and CI2, Fig. 1(c)], both ESMDS and ISMDS can or cannot find the native structure. In contrast, in the second class [GTT, NTL9, and protein B, Fig. 1(b)], only ISMDS can find the native structure, suggesting a higher sampling efficiency in ISMDS since the number of integration time steps in ISMDS and ESMDS are identical for these proteins. Furthermore, since there are at least eight independent runs for each protein and simulation, we also analyzed the lowest RMSD in each independent run (see Fig. S1 in the Supplemental Material [32] for the distributions of these RMSD values). In ISMDS, the lowest RMSD values are significantly lower than that in ESMDS among seven of ten proteins except for the three smallest proteins, which also indicates that the sampling in ISMDS is faster.

Traditionally, one should further check whether the folded state is thermodynamically preferred. However, for most of these proteins, the sampling is still limited. Thus, such kind of conclusions could not be obtained based only on our current simulations.

B. ISMDS can provide a better balance of secondary structures than ESMDS within limited sampling

A secondary structure can provide local stable interactions that might be necessary for the formation of tertiary structures, thus it is vital for protein folding. As shown in Fig. 2, each pair of subplots (left and right were in ESMDS and ISMDS, respectively) represents the secondary structure populations during the simulations for each protein. We used the whole trajectories to calculate the populations of secondary structures, and there is no significant difference for using the whole trajectories or those near the end of the simulations since, in most cases, our simulations either reach the equilibrium or fall into local minima rapidly.

For all helical proteins (Trpcage, HP35, and protein B), there are lots of helical contents (red) but very little sheet structure (blue). Moreover, for Trpcage, the helical contents agree well with that in the native state in both ESMDS and ISMDS. In contrast, for HP35, the first and second helix could not be distinguished well in ISMDS, or there is no obvious

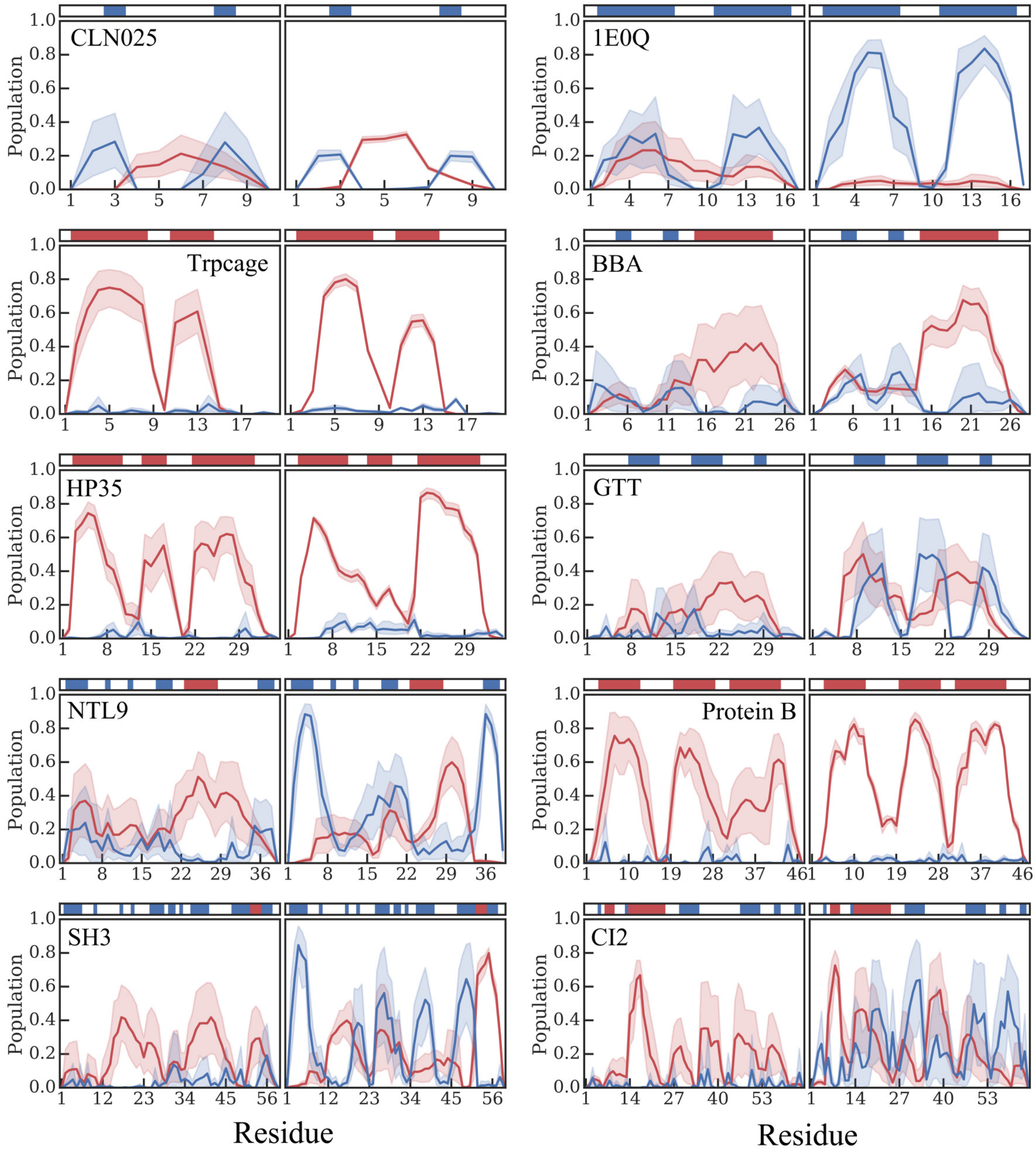

FIG. 2. Secondary structural populations for ten proteins during simulations. Each pair of subplots represents the population in ESMDS (left) and ISMDS (right), and helix and sheet are colored in red and blue, respectively. The native secondary structures are shown above each pair of subplots and the shadow of each curve implies the 95% confidence interval for independent runs.

peak for the second helix. This kind of disagreement could also be found in the last two helices of protein B in ESMDS. Together, these results suggest that both ESMDS and ISMDS have small defects in describing the helical properties.

For other sheet-containing proteins, there are also a large number of helical contents. This feature was shared by both ESMDS and ISMDS. However, in ESMDS, the populations of the sheet are significantly lower than those in ISMDS for these proteins except for CLN025. To give a more quantitative comparison, we calculated the area under each curve in Fig. 2; see Fig. S2 in the Supplemental Material [32] for the results. It can be seen that, for most proteins, the helical contents are comparative, but the sheet contents in ESMDS are significantly lower than that in ISMDS, suggesting the ISMDS can provide a more reasonable balance of secondary structures than ESMDS. It is noteworthy that, because of the limited sampling for most simulations, this propensity can only reflect the easiness of the formation of secondary structures, but not the accuracy of the two force fields, which will be discussed in the Discussion section.

C. Folding pathways are conserved in ESMDS and ISMDS but their weight may be different

To directly compare the folding mechanism between ESMDS and ISMDS, the transition paths were first extracted since the folding mechanisms are contained in the transitions between folded and unfolded states. The definitions

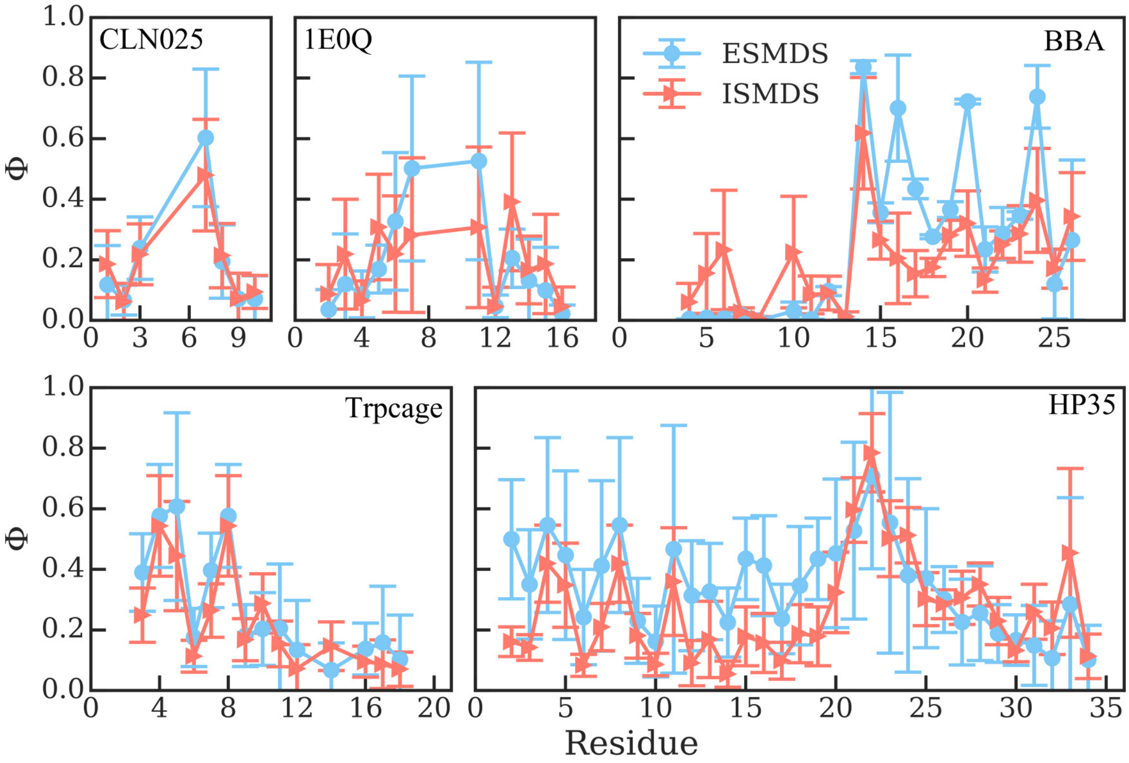

FIG. 3. Comparison of the Φ-value profiles obtained from ESMDS (tomato triangle) and ISMDS (light-blue circle) for five proteins. The error bars indicate the standard error of multiple folding pathways. The missing Φ values of some residues mean they do not form any contacts in the native state.

of transition paths are based on Q values, the fraction of native contacts. Then Φ value was used to characterize the transitions, a residue with higher Φ value is quanlitatively viewed as closer to the native structure. As shown in Fig. 3, the Φ-value profiles in ESMDS and ISMDS are quite similar, and the Pearson's correlation coefficient between two profiles for CLN025, 1E0Q, Trpcage, BBA, and HP35 are 0.98, 0.67, 0.93, 0.77, and 0.66, respectively. This result indicates that the folding mechanisms in ESMDS and ISMDS are highly conserved for CLN025 and Trpcage, but are somewhat different for the other three proteins.

To further seek the sources of these differences, we hypothesized that the weights of parallel folding pathways are altered as all these pathways might contribute to the calculated Φ value. To validate this hypothesis, the folding pathways were clustered based on the root-mean-square difference of the Φ-value profiles between different pathways for 1E0Q and HP35. BBA was excluded from this analysis because there is only one folding pathway in ESMDS. As shown in Fig. 4(a), there are two and three main folding pathways for 1E0Q in ESMDS and ISMDS, respectively. The main folding pathway is defined as the folding pathway whose proportion of all pathways is larger than 10% and number is larger than 1. In ESMDS, the two main folding pathways, PATH I and PATH II, are initialized by the formation of the loop at the bottom and the sheet at the top, respectively [Fig. 4(b)], and the weights of the two pathways are 60% and 20%, respectively. However, in ISMDS, in addition to PATH I and PATH II, there exists PATH III, which is initialized by the formation of the sheet in the middle, and the weights of the three pathways are 12%, 18%, and 27%, respectively. This result suggests that 1E0Q may have the same folding pathways in ESMDS and ISMDS, but the weights of these pathways are altered. Furthermore, the above result is further supported by the same analysis on HP35. As shown in Figs. 4(c) and 4(d), in ESMDS, HP35 has two main folding pathways, PATH I and PATH II, which are characterized by the early formation of last two and first two helices, respectively, and the weights of the two pathways are 50% and 25%, respectively. But in ISMDS, HP35 can only fold along PATH I, suggesting a dramatic change of the weights.

D. Folding mechanisms supported by ISMDS agree well with those found in previous studies

In our simulations, there are three proteins that ISMDS can sample their native state but not ESMDS, namely, GTT, NTL9 and protein B. Thus, we cannot directly compare the folding mechanisms observed in ISMDS with that in ESMDS. To examine whether or not the folding mechanisms are reasonable, we compared them with the folding mechanisms suggested by previous simulations or/and experiments.

For GTT WW domain [Figs. 5(a) and 5(b)], there are two folding pathways, PATH I and PATH II, which are characterized by the early formation of the first two and the last two sheets, respectively. The above two folding pathways are in line with the two folding pathways observed by Best *et al.* [40], who analyzed the long-time explicit solvent simulations performed by Shaw's group [13]. Furthermore, previous experiments have shown that the above two folding

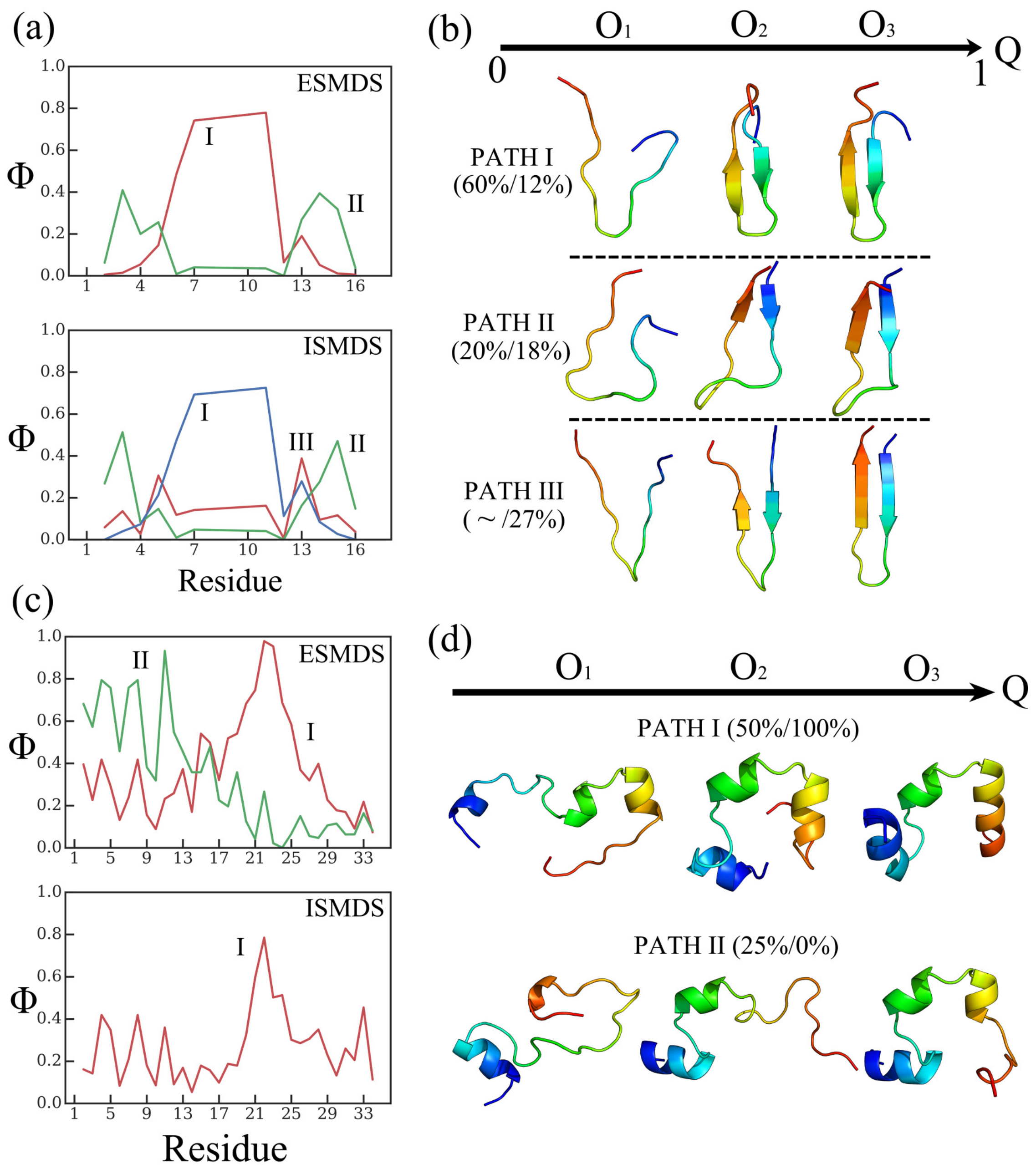

FIG. 4. The average Φ values of the different clusters of the folding pathways in both ESMDS and ISMDS for (a) 1E0Q and (c) HP35. The colors of the lines are red, green, and blue in order of decreasing cluster population. Representative structures (the color from blue to red correspond to the residue from *N* to *C* terminal) for each pathway at different intervals of *Q* values are shown for (b) 1E0Q and (d) HP35, where O_1, O_2, and O_3 indicate the *Q* values belong to [0.2, 0.4], [0.4, 0.6], and [0.6, 0.8], respectively. Besides, the two percentages in each bracket indicate the weight of the corresponding pathway in (before diagonal) ESMDS and ISMDS (after diagonal).

pathways indeed exist for the Pin1 WW domain and the first pathway PATH I is dominant [9,43], which agrees well with our finding (the weights of PATH I and PATH II are 64% and 36%, respectively) if we supposed the folding mechanisms are conserved enough for GTT and Pin1 WW domain.

For NTL9 [Figs. 5(a) and 5(c)], the K12M mutant of the *N*-terminal fragment of ribosomal protein L9 with a length of 39 residues, it has been shown that its folding time is about 700 μs at room temperature [44], thus we don't expect that there will be any folding event to be found. Surprisingly, we found one folding event in ISMDS in only 48-μs simulations (eight independent runs), and the corresponding folding pathway is also found in the explicit and implicit solvent simulations performed by Shaw's group [13] and Pande's group [12], respectively.

Protein B is an extremely stable protein [45] (melting temperature >373 K) and we used its mutant K5I/K39V to perform the simulations. It has been shown that the folding time of this mutant is extremely fast (~1 μs at 298 K) [46]. The coarse-grained simulations performed by Takada have

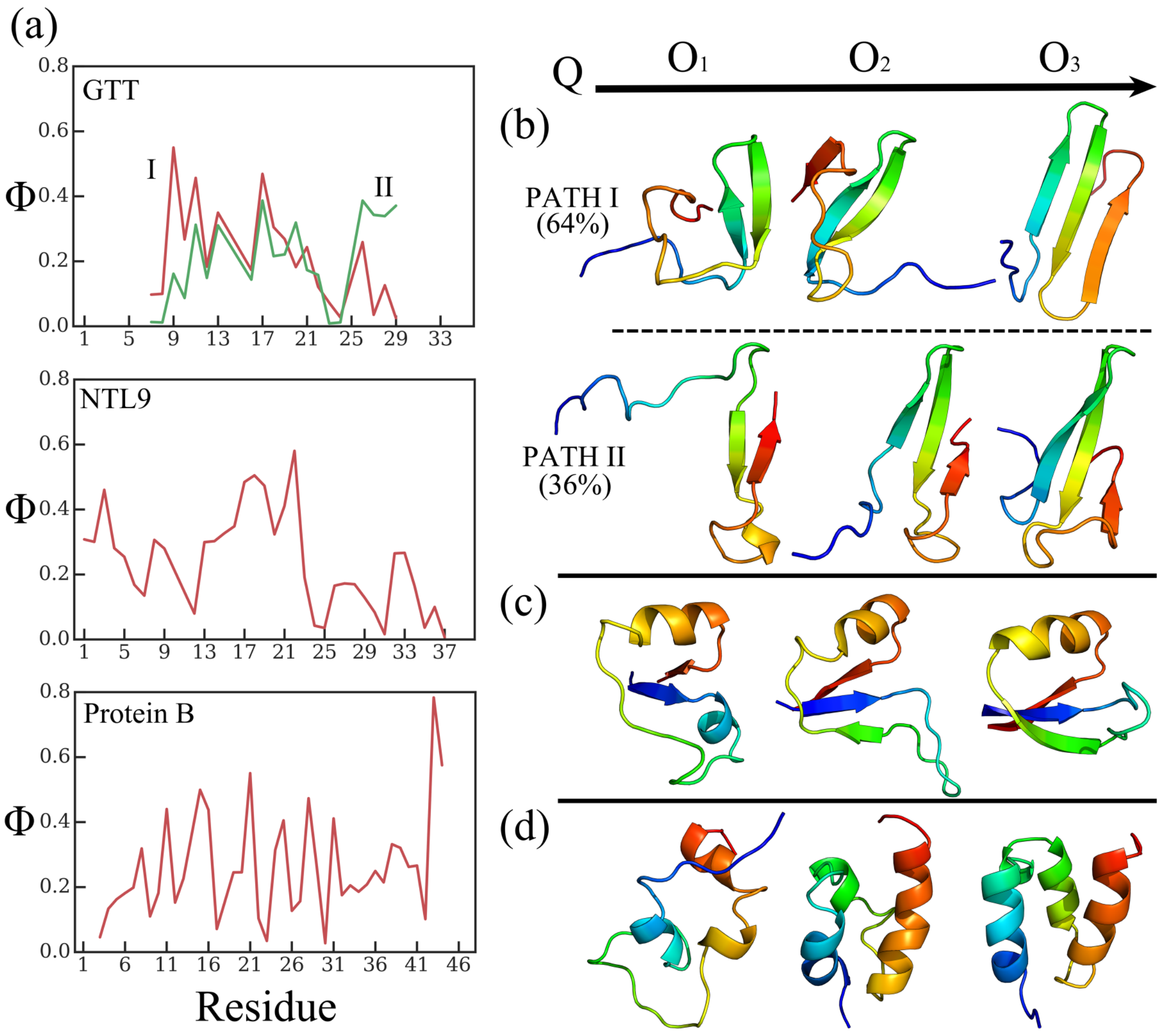

FIG. 5. (a) The average Φ values of the different clusters of the folding pathways in ISMDS for GTT, NTL9, and protein B. Representative structures for each pathway at different intervals of Q values are shown for (b) GTT, (c) NTL9, and (d) protein B, respectively. The meanings of O_1, O_2, O_3, and percentages are as in Fig. 4.

shown that the helices 1 and 3 are formed earlier and the helix 2 is unstable and cannot fold without that prefolded structure [47], which is in agreement with the folding mechanisms found in this work [Figs. 5(a) and 5(d)]. The same folding pathway can also be found in the simulations of Shaw's group though their folding pathways are more heterogeneous [13].

In summary, for all three proteins, GTT, NTL9, and protein B, although our ESMDSs have not reached the native states, the folding mechanisms suggested by ISMDSs provide an agreement at least in quantity, compared to previous simulations and experiments.

IV. DISCUSSION

It is well established that the sampling in ISMDSs is more efficient than that in ESMDSs [16], however, the applications of the former are limited by its inaccuracy on several aspects, e.g., salt bridges [48] and secondary structures [49]. Fortunately, the recent development of the generalized Born model [24] combined with the ff14SBonlysc force field allow us to run protein simulations under higher accuracy. In this study, using these parameters, eight out of ten proteins can fold into the native state, while only five of them can fold in the explicit solvent in the same number of integration time steps. In addition, such parameters also provide a better balance between different secondary structures under limited sampling, suggesting current implicit simulations are promising to investigate the dynamics of protein folding.

To quantitatively compare the folding mechanisms between ESMDSs and ISMDSs, a Φ-value analysis was performed. For the five smallest proteins, the Φ-value profiles in ESMDSs and ISMDSs are similar, especially for CLN025 and Trpcage, suggesting that the folding pathways are conserved. Further cluster analysis shows that the weights of these pathways may be different. This finding provides insight into ISMDS, namely that the protein folding landscapes are sensitive to the water model [50], but the folding pathways are not.

It is also noted that the folding pathways found in this work are consistent with previous studies. Particularly, for hairpin

1E0Q, the native contacts can form either from the termini to the loop or from the loop to the termini, corresponding to a "pincer" or a "zipper" mechanism, respectively [51]. Besides, the two major folding pathways of HP35 also are in line with that observed in both experiments [52,53] and simulations [54,55].

Through using these force fields and water models, our simulations can provide reasonable folding mechanisms. However, there is still room for further improvement. For ESMDSs, the populations of β sheet are significantly lower than that in ISMDSs for several proteins containing β sheet, suggesting the ff14SB force field may be too helical. For example, the mean folding time of GTT determined by experiment is 4.3 μs at 353 K [56], however, all eight independent simulations (10 μs each) at 300 K cannot generate substantial correct β sheet. To exclude the effects of temperature, another eight trajectories (6 μs each) were simulated at 350 K and the results are as that at 300 K (see Fig. S47 in the Supplemental Material [32] for the time courses of the secondary structure). It should be pointed out that the ff14SB force field is introduced to increase the stability of helix [17], but this increase may be too radical. For ISMDS, though the samplings are more efficient, however, the thermodynamical properties may be poorly defined. For example, the free energy landscape of HP35 has built from ISMDSs based on the potential of mean force (PMF) method, as the simulations have reached the equilibrium, but there is even no local minimum for the native state (see Fig. S35 in the Supplemental Material [32] for the one-dimensional free energy landscapes), which seriously contradicts to the previous knowledge. Indeed, the ISMDS is still not as mature as the ESMDS, thus several tricks could be introduced to fill the gaps, such as the constraints of dihedrals, as suggested by Nguyen and colleagues [22].

V. CONCLUSIONS

In this study, over 1.2-ms folding simulations of ten proteins with diverse topologies were performed to directly compare the folding mechanisms obtained from explicit and implicit solvents. Our results from Φ-value and cluster analysis indicate that the folding pathways observed in two solvent models are identical although their weights may be different. Furthermore, our simulations provide a large-scale benchmark for testing the accuracy of current Amber force fields, ff14SB and ff14SBonlysc, and several drawbacks have also been indicated.

ACKNOWLEDGMENT

This work was supported by the NSFC under Grant No. 11874162.

[1] H. Alonso, A. A. Bliznyuk, and J. E. Gready, Med. Res. Rev. **26**, 531 (2006).

[2] J. D. Durrant and J. A. McCammon, BMC Biol. **9**, 71 (2011).

[3] Y. B. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger, and D. E. Shaw, J. Am. Chem. Soc. **133**, 9181 (2011).

[4] X. Xu, M. Huang, and X. Zou, Biophys Rep. **4**, 1 (2018).

[5] S. Sharma, F. Ding, and N. V. Dokholyan, Bioinformatics **24**, 1951 (2008).

[6] A. Perez, J. A. Morrone, E. Brini, J. L. MacCallum, and K. A. Dill, Sci Adv. **2**, e1601274 (2016).

[7] A. Raval, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, Proteins **80**, 2071 (2012).

[8] A. Morriss-Andrews and J. E. Shea, Annu. Rev. Phys. Chem. **66**, 643 (2015).

[9] M. Jager, H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele, J. Mol. Biol. **311**, 373 (2001).

[10] H. Lei and Y. Duan, J. Mol. Biol. **370**, 196 (2007).

[11] P. L. Freddolino, F. Liu, M. Gruebele, and K. Schulten, Biophys. J. **94**, L75 (2008).

[12] V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, J. Am. Chem. Soc. **132**, 1526 (2010).

[13] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011).

[14] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Proc. Natl. Acad. Sci. USA **110**, 5915 (2013).

[15] E. C. Wang, P. Tao, J. Wang, and Y. Xiao, Phys. Chem. Chem. Phys. **21**, 18219 (2019).

[16] R. Anandakrishnan, A. Drozdetski, R. C. Walker, and A. V. Onufriev, Biophys. J. **108**, 1153 (2015).

[17] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**, 3696 (2015).

[18] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).

[19] P. A. Janowski, C. M. Liu, J. Deckman, and D. A. Case, Protein Sci. **25**, 87 (2016).

[20] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, J. Chem. Theory Comput. **8**, 3257 (2012).

[21] H. Nguyen, D. R. Roe, and C. Simmerling, J. Chem. Theory Comput. **9**, 2020 (2013).

[22] H. Nguyen, J. Maier, H. Huang, V. Perrone, and C. Simmerling, J. Am. Chem. Soc. **136**, 13959 (2014).

[23] Q. Shao and W. Zhu, Phys. Chem. Chem. Phys. **20**, 7206 (2018).

[24] H. Nguyen, A. Perez, S. Bermeo, and C. Simmerling, J. Chem. Theory Comput. **11**, 3714 (2015).

[25] R. Zhou, Proteins **53**, 148 (2003).

[26] C. Tan, L. Yang, and R. Luo, J. Phys. Chem. B **110**, 18680 (2006).

[27] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).

[28] J. Zhang, M. Qin, and W. Wang, Proteins **62**, 672 (2006).

[29] G. Zuo, W. Li, J. Zhang, J. Wang, and W. Wang, J. Phys. Chem. B **114**, 5835 (2010).

[30] X. Xue, W. Yongjun, and L. Zhihong, J. Theor. Biol. **365**, 265 (2015).

[31] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, Wires Comput. Mol. Sci. **3**, 198 (2013).

[32] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.101.062417 for tables with system details and simulations, plots of RMSD vs time, *Q* vs time, secondary structure vs time, RMSD histograms, *Q* histograms, and folding pathways.

[33] A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **8**, 1542 (2012).

[34] R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **9**, 3878 (2013).

[35] W. Kabsch and C. Sander, Biopolymers **22**, 2577 (1983).

[36] D. R. Roe and T. E. Cheatham, III, J. Chem. Theory Comput. **9**, 3084 (2013).

[37] D. R. Roe and T. E. Cheatham, J. Comput. Chem. **39**, 2110 (2018).

[38] R. B. Best, G. Hummer, and W. A. Eaton, Proc. Natl. Acad. Sci. USA **110**, 17874 (2013).

[39] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernandez, C. R. Schwantes, L. P. Wang, T. J. Lane, and V. S. Pande, Biophys. J. **109**, 1528 (2015).

[40] R. B. Best and G. Hummer, Proc. Natl. Acad. Sci. USA **113**, 3263 (2016).

[41] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J.van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, Nat. Methods **17**, 261 (2020).

[42] S. Yuan, H. S. Chan, S. Filipek, and H. Vogel, Structure **24**, 2041 (2016).

[43] H. Nguyen, M. Jager, A. Moretto, M. Gruebele, and J. W. Kelly, Proc. Natl. Acad. Sci. USA **100**, 3948 (2003).

[44] J.-C. Horng, V. Moroz, and D. P. Raleigh, J. Mol. Biol. **326**, 1261 (2003).

[45] M. U. Johansson, M. de Château, L. Björck, S. Forsén, T. Drakenberg, and M. Wikström, FEBS Lett. **374**, 257 (1995).

[46] T. Wang, Y. Zhu, and F. Gai, J. Phys. Chem. B **108**, 3694 (2004).

[47] S. Takada, Proteins **42**, 85 (2001).

[48] A. Okur, L. Wickstrom, and C. Simmerling, J. Chem. Theory Comput. **4**, 488 (2008).

[49] D. R. Roe, A. Okur, L. Wickstrom, V. Hornak, and C. Simmerling, J. Phys. Chem. B **111**, 1846 (2007).

[50] R. Anandakrishnan, S. Izadi, and A. V. Onufriev, J. Chem. Theory Comput. **15**, 625 (2019).

[51] R. B. Best and J. Mittal, Proc. Natl. Acad. Sci. USA **108**, 11087 (2011).

[52] L. Zhu, K. Ghosh, M. King, T. Cellmer, O. Bakajin, and L. J. Lapidus, J. Phys. Chem. B **115**, 12632 (2011).

[53] S. Nagarajan, S. Xiao, D. P. Raleigh, and R. B. Dyer, J. Phys. Chem. B **122**, 11640 (2018).

[54] R. Harada and A. Kitao, J. Chem. Theory Comput. **8**, 290 (2012).

[55] S. Piana, K. Lindorff-Larsen, and D. E. Shaw, Biophys. J. **100**, L47 (2011).

[56] S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele, and D. E. Shaw, J. Mol. Biol. **405**, 43 (2011).