Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 415–424
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Enhanced ab initio protein folding simulations in Poisson–Boltzmann molecular dynamics with self-guiding forces

Edward Z. Wen [a], Meng-Juei Hsieh [a], Peter A. Kollman [b,1], Ray Luo [a,*]

[a] *Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92697-3900, USA*
[b] *Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143, USA*

Abstract

We have investigated the sampling efficiency in molecular dynamics with the PB implicit solvent when self-guiding forces are added. Compared with a high-temperature dynamics simulation, the use of self-guiding forces in room-temperature dynamics is found to be rather efficient as measured by potential energy fluctuation, gyration radius fluctuation, backbone RMSD fluctuation, number of unique clusters, and distribution of low RMSD structures over simulation time. Based on the enhanced sampling method, we have performed ab initio folding simulations of two small proteins, ββα1 and villin headpiece. The preliminary data for the folding simulations is presented. It is found that ββα1 folding proceeds by initiation of the turn and the helix. The hydrophobic collapse seems to be lagging behind or at most concurrent with the formation of the helix. The hairpin stability is weaker than the helix in our simulations. Its role in the early folding events seems to be less important than the more stable helix. In contrast, villin headpiece folding proceeds first by hydrophobic collapse. The formation of helices is later than the collapse phase, different from the ββα1 folding.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Poisson–Boltzmann; Molecular dynamics; Self-guiding forces; Protein folding; BBA1; Villin headpiece

1. Introduction

Understanding the molecular mechanisms of protein folding at atomic detail has remained a major challenge. This is partly because there exists a huge gap between the time scale that can be probed in experiment and the time scale that can be simulated on computer with models of adequate atomic accuracy. This makes it difficult to study folding mechanisms with concerted efforts from both experiment and computation. In recent years, the gap has been gradually reduced [1,2], though mostly from the experimental side [3]. Computational simulations of folding kinetics have gradually been catching up. Among some notable studies in explicit solvents are ab initio folding simulation of turns [4–7], helices [8–10], β hairpins [11], and a three-helix bundle protein [12,13]. Due to the enormous demand of CPU time for folding simulations in explicit solvents, various implicit solvent models such as the distant-dependent dielectric screening and Generalized Born models have been applied in folding simulations [14–16]. Such ab initio folding simulations cover all-α-folds [17–20]; all-β-folds, including β-hairpin and three-stranded β sheet [17,21–27]; and α/β folds, including Trp-cage and BBA5 [20,26,28–31].

Efficient conformational sampling in ab initio folding simulations is important in order to cover the folding pathways and to identify the global minimum. Molecular dynamics is a natural method to study folding mechanisms, yet its difficulty in efficient conformational sampling has hindered progress in folding simulations. This has prompted many research efforts in developing methods to enhance sampling efficiency. Most previous efforts in enhancing sampling efficiency in ab initio folding simulations are based on the ideas of simulated annealing, diffusion equation method, locally enhanced sampling, replica exchange Monte Carlo, and self-guiding forces. Simulated annealing simulations employ higher temperatures to increase the frequency of energy barrier crossing events. They have been widely adopted in recent ab initio folding algorithms [32–35]. The diffusion equation method [36–39] has been shown to result in significant surface smoothing and greatly improved convergence of optimization algorithms. This approach modifies the energy function such that only a single minimum remains, which may correspond to the original global energy minimum. A promising approach in efficient sampling is the application of mean-field theory, which has been recently reviewed [40]. Of particular interest is the method of locally enhanced

* Corresponding author. Tel.: +1-949-824-9528; fax: +1-949-824-8551.
E-mail address: rluo@uci.edu (R. Luo).
[1] Deceased.

1093-3263/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2003.12.008

sampling [41–44]. This mean-field technique allows selective focusing of computational effort on a portion of the system, thus increasing the sampling of the region of interest. Recent improvements in sampling have relied on the idea of generalized ensembles such as multi-canonical ensembles and replica-exchange Monte Carlo [45]. Their applications in ab initio folding simulations have also been reported [46]. The self-guiding (force averaging) method reduces barriers through time and spatial averaging of non-bonded forces of a substructure of a molecule during dynamics simulations [47,48]. This results in "guiding forces" which are added to the instantaneous forces to enhance systematic motions [6,10,47–50]. In this study, we utilize self-guiding forces to enhance barrier crossing events. If explicit solvent is replaced with an implicit solvent, it is possible to increase the rate of barrier crossing by artificially reducing the friction constant of implicit solvent. This was demonstrated recently when an implicit solvent simulation converted a RNA tetra-loop from an incorrect to correct structure despite the inability of explicit solvent simulation to do so [51]. In a folding simulation of a 12-mer polyphenylacetylene helix, it was also shown that there exists a linear dependence of folding time on solvent viscosity for helix folding studied [52]. We turn off solvent friction forces to enhance sampling efficiency in the current study.

In this paper, we intend to gain a thorough understanding of the sampling efficiency of self-guiding forces. This will be helpful for future development of more robust sampling methods. This is achieved by comparing various benchmarks with respect to regular simulations at a higher temperature. Based on the analysis of the sampling efficiency, we further applies the self-guiding technique and our recently developed Poisson–Boltzmann implicit solvent model [53–55] to ab initio folding simulations for two proteins, ββα1 and villin headpiece to study their early folding events. To our knowledge, this is the first folding simulations with the Poisson–Boltzmann solvent model.

In the following, the method for simulations and data analysis is first described. This is followed by results and discussion of the theoretical native state, the comparative studies of sampling efficiency, and preliminary data for the folding simulations of ββα1 and villin headpiece. Finally, the concluding remarks are presented.

2. Methodology

In this section, a description of the proteins to be simulated is first described. This is followed by the refinement of native structures. Next are descriptions of our energy model and the enhanced sampling method for ab initio folding simulations. Finally, the simulation detail and data analysis are described.

2.1. Proteins to be studied

ββα1 (PDB id: 1HCW) is a 23-residue protein developed by the Imperiali group [56]. The sequence of ββα1 is Ace-YTVpSXTFSRSDELAKLLRLHAGNMe, with a non-natural amino acid X = 3-(1,10-phenanthrol-2-yl)-L-alanine (Fen) and a p = d-proline (D-pro). It contains all three types of secondary structures: β-strand (residues 2–7), type II' β-turn (residues 4–5), and α-helix (residues 12–20). In this designed protein, D-Pro4 is used to stabilize the type II′ β-turn, and Fen6 is used to stabilize the hydrophobic core between the hairpin and the helix. The hydrophobic core consists of Tyr1, Fen6, Phe8, Leu14, Leu 17, and Leu18.

Villin headpiece (PBD id: 1VII) is a 36-residue protein studied by McKnight et al. [57,58]. The sequence of villin headpiece is MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF with all natural amino acids. It is an all-α three-helix-bundle protein with a hydrophobic core consisting of residues Leu2, Phe7, Val10, Phe11, Phe18, Leu21, and Leu29.

2.2. Theoretical native states

Due to the uncertainty in NMR structures [59,60], we further refined the native structures by explicit water particle mesh Ewald (PME) simulations [61]. The refined native structures were used as references for later folding simulations. The refined native structures for both proteins were computed by taking a structural mean of a 2 ns equilibrium trajectory after a 1 ns relaxation. In our refinement simulations, the first frame in the NMR structure was used as the starting structure in each simulation. Counter ions (Cl^- or Na^+) were added to maintain system neutrality. The protein intra-molecular potential was computed by a standard AMBER force field (parm94) [62]. The force field parameters of Fen6 were taken from Wang et al. [22]. It should be pointed out that the χ_1 torsion term of Fen6 is directly taken from the standard amino acid Trp in this parameter set. This may result in the observed discrepancy between the PME and NMR structures for the native state (see below). The proteins were solvated with water boxes with a 12 Å buffer zone between the surfaces of the protein and the water box. The PME real-space Coulomb and van der Waals interactions used a cutoff of 9 Å. After an initial steepest descent minimization of 1000 steps, the systems were heated up to 300 K through 100 ps MD simulation, followed by 2.9 ns PME simulations at 300 K coupled to the Berendsen et al.'s heat bath [63] with a coupling constant of 0.2 ps. The simulation pressure was maintained at 1 bar by the Berendsen's pressure bath with a coupling constant of 0.2 ps. The SHAKE algorithm [64,65] was used to fix all bonds involving hydrogen atoms. A time step of 2 fs was used for the Verlet leap-frog integration method [66].

2.3. Energy model for folding simulations

In all folding simulations, protein intra-molecular interactions use the same force field as in the above PME simulations. The solvation model used is an implicit solvent based on finite-different Poisson–Boltzmann (FDPB) for the polar

solvation interaction [53,54]. Our previous studies show that the surface area (SA) model for the non-polar solvation interaction only contributes negligibly to protein stability and folding mechanisms [60,67] when the standard surface tension of 5 cal/mol-Å^2 [68] was used. Thus, this term was omitted as a reasonable approximation in the current simulations for efficiency. The solvent probe size was set to 1.4 Å for the molecular surface calculation in FDPB. The dielectric constants for the protein exterior and interior were set to 80 and 1, respectively. In the molecular dynamics simulation with FDPB (PBMD), it is possible to use the particle–particle and particle–mesh methods for Coulombic interactions to avoid the use of cutoff. The non-bonded van der Waals interactions for all proteins were also computed without cutoff since the proteins simulated are fairly small. Thus, all non-bonded interactions, including the solvation interaction, were computed with no cutoff. All non-bonded forces were computed at every time step.

2.4. Sampling method for folding simulations

As discussed in Section 1, the enhanced sampling in the current ab initio folding simulations utilizes both the friction-free nature of implicit PB solvents and the self-guiding (SG) forces. In a conventional MD simulation, the motion of atom i follows the Newtonian equation of motion:

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i \tag{1}$$

where m_i, $\ddot{\mathbf{r}}_i$ and $\mathbf{f}_i$ are, respectively, the mass, the displacement vector, and the force of atom i. In a MD simulation with self-guiding forces, a "guiding force" $\mathbf{g}_i$ of atom i, derived from its own simulation trajectory:

$$\mathbf{g}_i = \frac{1}{t_l}\int_{t-t_l}^{t} (\mathbf{f}_i + \lambda \mathbf{g}_i)\, \mathrm{d}\tau \tag{2}$$

is added upon atom i:

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i + \gamma \mathbf{g}_i \tag{3}$$

where λ is the guiding factor, and t_l the force running average window [47,48]. By increasing systematic motion of bonded substructures of a molecular system, self-guiding forces enhance barrier-crossing events to speed up conformational sampling [47]. In the present study, a guiding factor γ of 0.2 and local sampling time t_l of 0.2 ps for implicit solvents were used [48]. The relative sampling efficiency and sufficiency of SG/PBMD (at 300 K) was compared with high-temperature PBMD simulations (HT/PBMD) at 400 K. This is discussed below in Section 3.

2.5. Folding simulation detail

The initial structures were built with the InsightII modelling package as fully extended all-*trans* structures. After brief minimization, these initial structures were used for all ab initio folding simulations. Twenty independent folding simulations was initially started with a coarse grid spacing of 0.8 Å in FDPB for 20 ps due to the large size of the initial structure. Different random seeds were used to generate independent initial velocity assignments. All simulations were then restarted with a grid spacing of 0.5 Å in FDPB for the remaining of the 4 ns trajectories. The SHAKE algorithm [64,65] was used to fix all bonds. The time step was set to 2 fs. With 20 nodes of our PC cluster of Intel Pentium 4 2.4 GHz CPUs (512 MB, 1066 MHz RDRAM), 20 trajectories took about 4 days to finish.

2.6. Data analysis

The simulation snapshots were recorded every 1 ps in all folding simulations. The backbone RMSD values relative to the theoretical native structures were obtained using the PTRAJ module in the AMBER package [69]. The radius of gyration and the number of native contacts were evaluated using Multiscale Modeling Tools in Structural Biology (MMTSB) [70].

The clustering analysis was carried out with the "kclust" script in MMTSB tool set with a radius of 60° for "phipsi" mode. Due to the limitation of the clustering method, we used a potential energy "cutoff" to filter out high-energy structures sampled. A cutoff of 6 kcal/mol lower than the average potential energy of all trajectories was found to reduce the total number of snapshots to a manageable level in the cluster analysis.

The native-like snapshots were defined to be those with both high percentage native secondary structures (>70%) and low backbone RMSD values (<3.5 Å). The secondary structure analysis was performed by the DSSP package [71].

3. Results and discussion

In this section, we first present the data on the refinement of native structures in PME simulations. This is followed by an analysis of sampling efficiency when self-guiding forces are used in PBMD. Finally, folding pathways of both ββα1 and villin headpiece are discussed.

3.1. Theoretical native states

3.1.1. ββα1

As shown in Fig. 1(a), the potential energy reaches the equilibrium level and remains stable within the 3 ns PME simulation. Interestingly, from 400 to 600 ps, the hydrophobic core first unpacks and then repacks so that the Fen6 side-chain χ_1 angle is different from that in the NMR structure. This is reflected in the sudden increase in backbone RMSD during this period (Fig. 2(a)). The new Fen6 rotamer stays stable throughout the remaining of the trajectory. Partly due to this large conformational transition, the backbone deviation from the NMR structure is higher than that

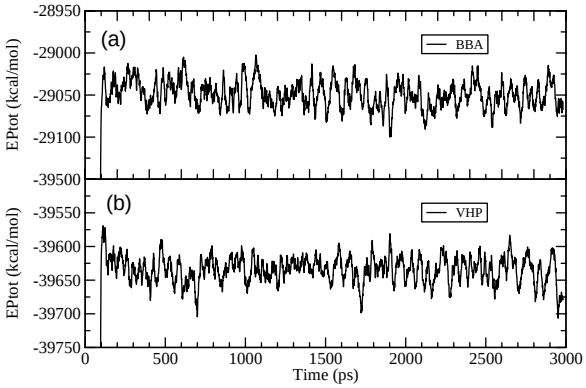

Fig. 1. PME refinement simulations of the native structures: potential energy fluctuation over time, running average with 20 ps window: (a) ββα1; (b) villin headpiece.

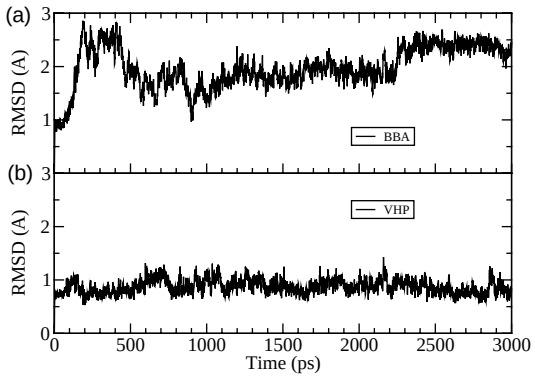

Fig. 2. PME refinement simulations of the native structures of backbone RMSD fluctuation over time: (a) ββα1; (b) villin headpiece.

of the villin headpiece (Fig. 2(a) and (b)). The discrepancy of the Fen6 side chain between PME and NMR structures, however, could imply either the NMR model is poorly determined, or the χ_1 torsion term in our force field needs improvement (see Section 2). Nevertheless, this is not the focus of this study on efficient sampling methods for ab initio folding simulations. After an initial relaxation of 1 ns, the remaining 2 ns equilibrium trajectory was used to compute an average structure as the theoretical native structure for later folding simulation analysis. The average structure maintains a nice hairpin and a good hydrophobic packing among the side chains of Fen6, Leu14, Leu17 and Leu18, although the RMSD between this structure and the initial NMR structure is 1.89 Å (Fig. 3(a)). Previously, Wang et al. also reported a 1 ns explicit water simulation of ββα1. Similar agreement to NMR structure was observed as here, though a cutoff distance was used for electrostatic interactions [22].

3.1.2. Villin headpiece

As shown in Figs. 1(b) and 2(b), the potential energy and backbone geometry quickly reaches equilibrium and remains quite stable within the 3 ns PME simulation trajectory. The backbone RMSD fluctuation with respect to the NMR structure is also lower than ββα1, mostly <1.0 Å. Furthermore, the hydrophobic core is well preserved and very stable in the simulation. After an initial 1 ns relaxation, the later 2 ns equilibrium simulation was used to compute an average structure. The average structure was used as a theoretical native structure for later folding simulation analysis. The RMSD between the average structure and the initial NMR structure is 0.94 Å. The agreement is similar to that of a simulated crystal structure (Fig. 3(b)). A 200 ns explicit water simulation on villin headpiece was previously reported by Duan and Kollman, though a cutoff distance was used for electro-static interactions in that simulation, resulting in a poorer agreement between simulation and NMR structures [12].

3.2. Enhanced sampling efficiency in folding simulations

The sampling efficiency in ab initio folding simulations was analyzed with ββα1 as a test case. The efficiency was first studied by monitoring the fluctuation of potential energy

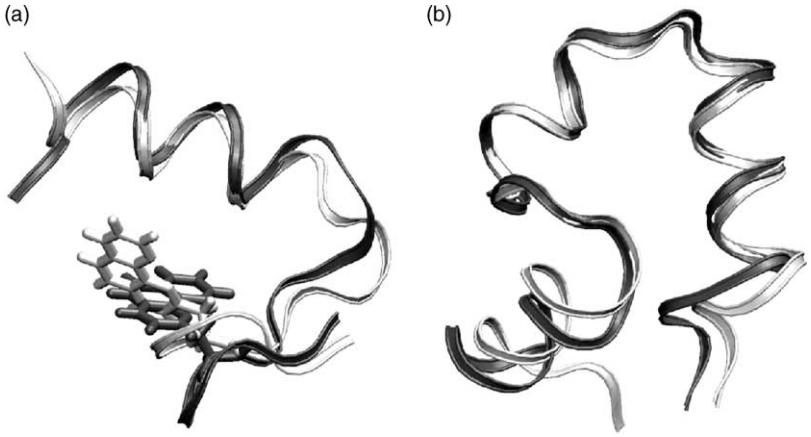

Fig. 3. PME refinement simulations of the native structures: superposition of the NMR and PME mean structures: (a) ββα1; (b) villin head-piece. Black ribbon is for the NMR structure, and gray ribbon is for the PME structure. The side chain of Fen6 is also shown for ββα1.

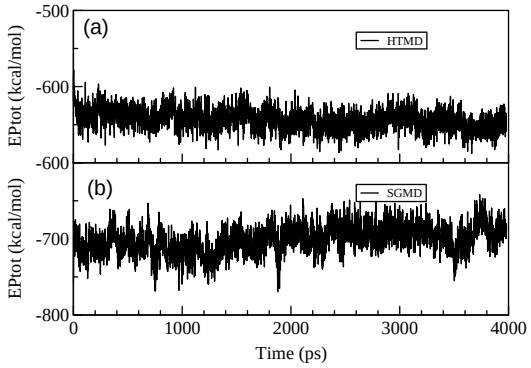

Fig. 4. Folding simulation with: (a) HT/PBMD for ββα1, potential energy vs. time; (b) SG/PBMD for ββα1, potential energy vs. time. Note that the potential energy (EPtot) in SG/PBMD fluctuates similarly with that in HT/PBMD.

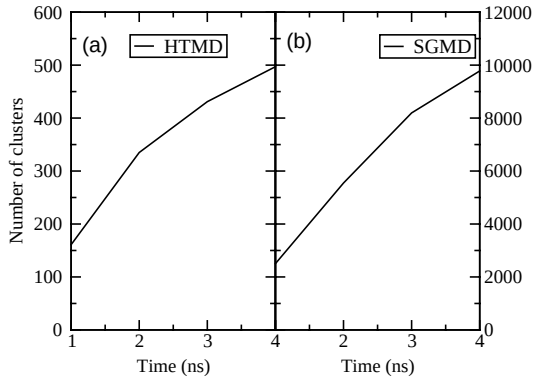

Fig. 6. Cumulative number of unique clusters from all: (a) 20 HT/PBMD folding simulations; (b) 20 SG/PBMD folding simulations.

during simulations (Fig. 4). In a representative SG/PBMD folding trajectory (Fig. 4(b)), the fluctuation of potential energy is much greater than that of room-temperature PBMD (RT/PBMD) simulation because of the "guiding forces". Although the average potential energy for SG/PBMD is very similar to that for RT/PBMD (−704.0 and −706.8 kcal/mol, respectively), the standard deviation (i.e., fluctuation) in potential energy is around 20 kcal/mol, which is almost twice as high as that for RT/PBMD (about 10 kcal/mol). Yet this fluctuation is similar to that of HT/PBMD, about 18 kcal/mol (Fig. 4(a)).

The sampling efficiency can also be demonstrated by the conformational fluctuation during a folding trajectory. The plots of gyration radius versus simulation time in the folding trajectories by both SG/PBMD and HT/PBMD are shown in Fig. 5. Apparently, the SG/PBMD folding simulation samples conformational space even more aggressively than HT/PBMD.

The sampling efficiency is further analyzed by comparing the unique conformations visited in all 20 folding trajectories

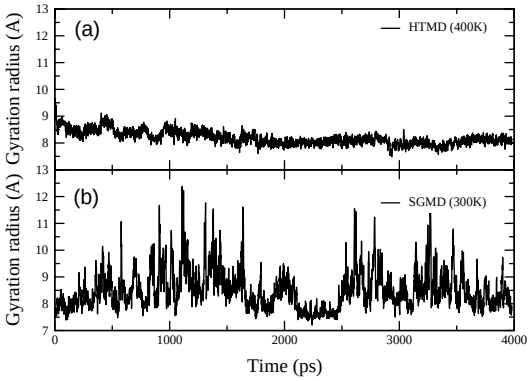

Fig. 5. Gyration radius vs. time in the same folding simulation with: (a) HT/PBMD as in Fig. 4(a) and (b); (b) SG/PBMD as in Fig. 4(b). Note that fluctuation of structures in SG/PBMD is much higher than that in HT/PBMD.

by either SG/PBMD or HT/PBMD. After filtering out high energy conformations in HT/PBMD, a total of 21,207 snapshots were collected for clustering analysis. Cluster analysis of these snap-shots gave a total of 497 clusters. After filtering out high energy conformations in SG/PBMD, a total of 24,899 snapshots were obtained, out of which 9778 clusters were found through the same clustering procedure. Moreover, after combining the 497 cluster centroids from HT/PBMD with those from SG/PBMD, no more new clusters were found than the 9778 clusters from SG/PBMD. This suggests that, by using the same simulation condition and the same simulation time, SG/PBMD is almost 20-fold as efficient as HT/PBMD in sampling protein conformational space.

Moreover, as shown in Fig. 6, which plots the number of cumulative unique conformations versus simulation time, the chance to find new unique conformations in HT/PBMD is declining, especially in the 4th ns—it is possible that after several more nanoseconds, HT/PBMD would soon reach a plateau where few new conformations would be sampled. This implies that more independent HT/PBMD simulations are needed to fully sample the conformational space. In contrast, after 4 ns simulations, SG/PBMD simulations still have the potential to sample larger conformational space with longer simulation time, indicating its higher tendency to cross energy barriers.

Finally, the sampling efficiency is investigated by the backbone RMSD fluctuation (Fig. 7) and low RMSD structures visited during folding simulations (Fig. 8). Indeed, with reference to the theoretical native structure, the lowest RMSD among all snapshots in HT/PBMD is 3.04 Å, compared to a much lower 1.74 Å in SG/PBMD simulations. During the 4 ns simulations, the lowest RMSD value for 15 out of all 20 trajectories is no larger than 3 Å, and the lowest RMSD for all 20 trajectories is 1.74 Å. A total of 174 snapshots are with RMSD below 3 Å. Fig. 8 shows distribution of low RMSD structures in all 20 SG/PBMD trajectories over simulation time. This indicates that our folding simulations are more sufficient than high-temperature simulations

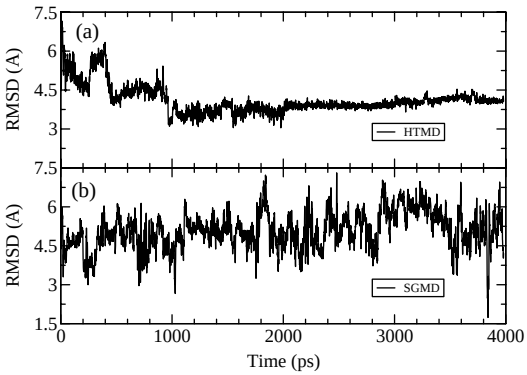

Fig. 7. Backbone RMSD fluctuation vs. time in the same folding simulation with: (a) HT/PBMD as in Fig. 4(a) and (b); (b) SG/PBMD as in Fig. 4(b).

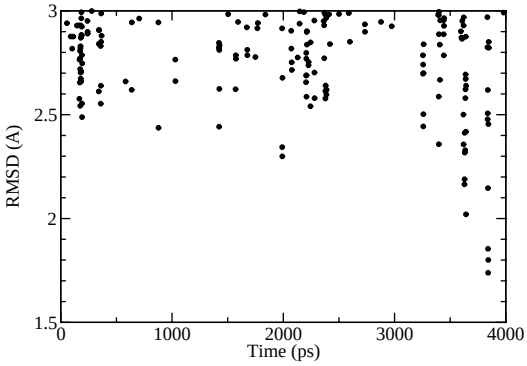

Fig. 8. Distribution of low RMSD structures (<3 Å) in all 20 SG/PBMD simulations over simulation time.

in sampling backbone conformations that are close to the native structure.

Based on the enhanced sampling of our simulation method, we performed folding simulations for ββα1 and villin headpiece. In the following, we present the preliminary data of these simulation trajectories.

3.3. Folding pathway of ββα1

Here, we present an observed folding pathway for ββα1 in the 20 folding simulation trajectories. Fig. 9(a) shows a superposition of the theoretical native state and the native-like structure at the end of the folding pathway. Interestingly, although structures with low RMSD values were observed since the early time in the folding simulation, it is only after 3 ns when snapshots with RMSD <2.0 Å were found, suggesting that observation of the native structure during the folding simulations may require much longer simulation time. Since the native state is not observed in any of the folding trajectories, we believe that the molecular events observed may be relevant to the early folding stage of the simulated protein. Fig. 10 shows the representative snapshots observed during one folding trajectory of ββα1. The folding trajectory shows that helix formation is much earlier than the hairpin formation and the hydrophobic collapse, though the type II' turn of the hairpin forms right after the beginning of the simulation, preceding the formation of helix. However, the formation of full hairpin is rare before the occurrence of the native-like structure in this trajectory, indicating that the hairpin may not be stable enough to populate during the early stage of folding. Instead, the helix not only forms early, but also populates much more than the hairpin, indicating its important role in the early folding stage. Compared with the villin headpiece below, it is interesting to note that the hydrophobic collapse is at most happening concurrently with the formation of helix, but not before the formation of helices as in villin headpiece (see below).

3.4. Folding pathways of villin headpiece

Fig. 11 presents a collection of representative snapshots in an early folding pathway of villin headpiece from the 20 folding trajectories simulated. Interestingly, villin headpiece demonstrates rather different folding mechanism from ββα1. For example, we observed a fairly low

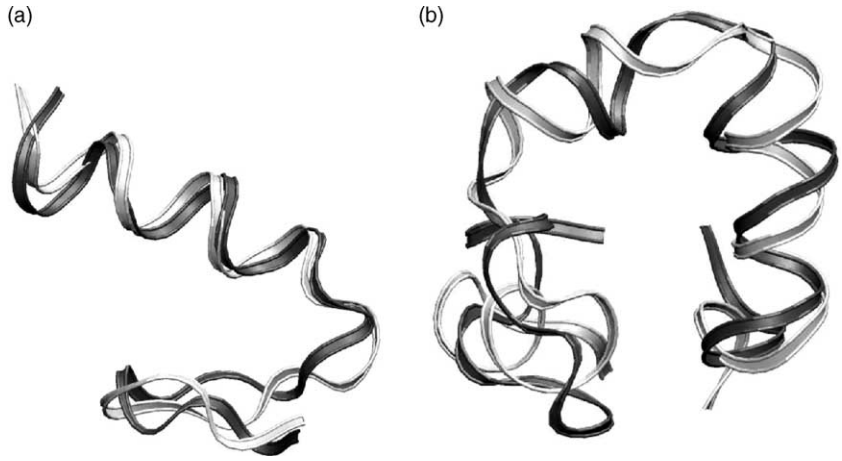

Fig. 9. Superposition of theoretical native states (gray) and native-like structures found in folding simulations (black): (a) ββα1; (b) villin headpiece.

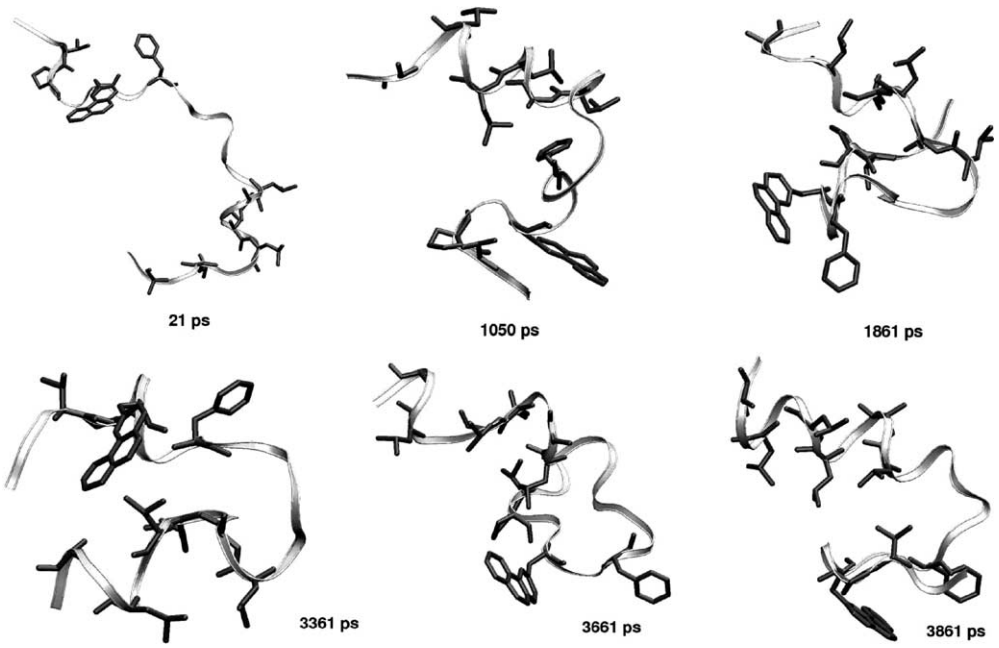

Fig. 10. Representative snapshots on a pathway to the native-like structure as in Fig. 9(a) for ββα1.

RMSD snapshot, 4.7 Å, at 38 ps. However, the native helix is <10% at this moment. This implies that hydrophobic collapse precedes the formation of secondary structures in villin headpiece folding. Subsequently, the folding process is dominated by large fluctuations of backbone RMSD and gyration radius. The percentage of native secondary structures also fluctuates, though the highest level is only 50%. Starting at 2.4 ns, a steady decline of backbone RMSD was observed (Fig. 12). This is also accompanied by a steady increase of native secondary structures. However, the gyration radius only fluctuates around 8.5–9.5 Å. This folding pattern is similar to that described in the folding theory utilizing the concept of molten globin state [72,73]: a folding process starts with an initial hydrophobic collapse and is followed by reorganization of secondary structures and tertiary contacts. The whole reorganization process takes over 1 ns simulation time. This is much longer than the time needed for initial collapse in our simulations, indicating the process may be the rating limitation step in folding. The folding mechanism observed in this enhanced sampling study also agrees with earlier studies of Duan and co-workers [12,13] and the Pande group [19]. A superposition of the theoretical

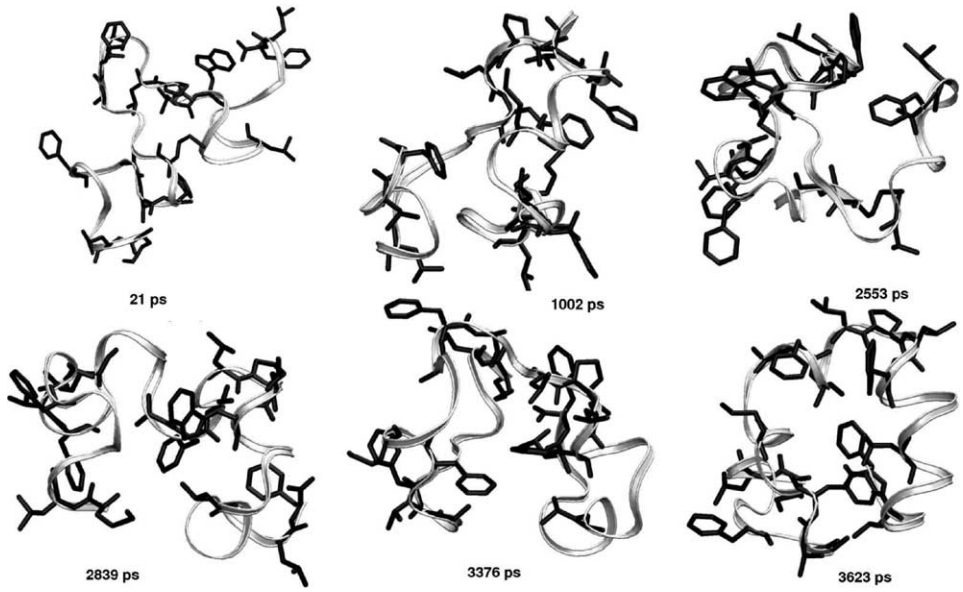

Fig. 11. Representative snapshots on a pathway to the native-like structure as in Fig. 9(b) for villin headpiece.

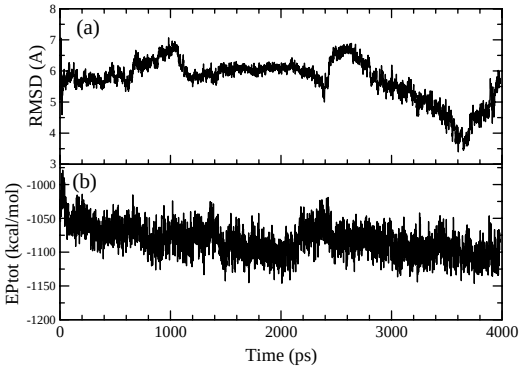

Fig. 12. Folding simulation trajectory for the pathway of villin headpiece as shown in Fig. 11: (a) backbone RMSD vs. time; (b) potential energy vs. time.

native structure and the native-like snapshot observed in the end of the folding pathway is shown in Fig. 9(b).

4. Conclusions

We examined the sampling efficiency in molecular dynamics with the PB implicit solvent when self-guiding forces are added. We found rather impressive efficiency as measured by potential energy fluctuation, gyration radius fluctuation, backbone RMSD fluctuation, number of unique clusters, and distribution of low RMSD structures over time compared with a high-temperature dynamics simulation. It is interesting to note that the fluctuation of potential energy in SG/PBMD simulations is similar to that of high-temperature PBMD simulations, and much higher than room-temperature PBMD simulations. The fluctuations of both gyration radius and backbone RMSD in SG/PBMD are also significantly higher than those of normal PBMD. This results in far more low-energy unique clusters identified in SG/PBMD simulations than normal PBMD simulations. The number of low RMSD structures identified in all folding simulations is also much higher than normal PBMD. The lowest backbone RMSD is also much lower than those from normal PBMD simulations.

Based on the enhanced sampling method, we performed ab initio folding simulations of two small proteins, ββα1 and villin headpiece. The preliminary data for the folding simulations were presented. We found that ββα1 folding proceeds by initiation of the turn and the helix. The hydrophobic collapse seems to be lagging behind or at most concurrent with the formation of the helix. The hairpin stability is weaker than the helix in our simulations. Its role in the early folding events seems to be less important than the more stable helix. In contrast, villin headpiece folding proceeds first by hydrophobic collapse. The formation of helices is later than the collapse phase, different from the ββα1 folding. We believe that this may well be due to the different helical stabilities between the two small proteins. Further studies of the stability of the helices will be interesting in understanding their roles in folding.

In this paper, we have primarily focused on enhancing sampling efficiency with the use of self-guiding forces in ab initio folding simulations. It is instructive to discuss its limitations for further improvements. One limitation is that native states are always unstable in equilibrium simulations at room temperature. This makes it difficult for a simulation trajectory to converge to the global free energy minimum. This is not a surprise because SGMD by design to push a conformation out of a local/global minimum. Several recent improvements in utilizing self-guiding forces in simulations have been proposed, including the uses of atomic momentum averaging, hybrid Monte Carlo, and/or Langevin dynamics in sampling trajectories [49,50]. These new improvements were shown to generate Boltzmann-like distributions while still preserving the accelerated conformational transition on model systems. The improvements in principle can resolve the problem of unstable native states given that the native states are global free energy minimum. Further analysis of these improved versions on peptides and proteins will be interesting to understand their performances.

The second limitation appears to be on the side chain sampling. Indeed, native side-chain packing is rarely observed in any trajectory, preventing us from discussing this very interesting phenomenon in folding processes. We believe that this is partly because side-chain sampling was enhanced only limitedly by self-guiding forces. The limited enhancement may well be due to the entropic nature of side-chain sampling. Indeed, self-guiding forces were developed to overcome enthalpic barriers, not entropic barriers. Another possible reason may be the distorted time-scales of the slow main-chain motions and fast side-chain motions since self-guiding forces enhance slow motions by design. It is likely that the slow motions are enhanced too much to interfere with the fast side chain motions. Further developments will be necessary to overcome this limitation.

References

[1] V. Daggett, A. Fersht, The present view of the mechanism of protein folding, Nat. Rev. Mol. Cell Biol. 4 (6) (2003) 497–502.
[2] V. Daggett, A.R. Fersht, Is there a unifying mechanism for protein folding? Trends Biochem. Sci. 28 (1) (2003) 18–25.
[3] W.A. Eaton, V. Munoz, S.J. Hagen, G.S. Jas, L.J. Lapidus, E.R. Henry, J. Hofrichter, Fast kinetics and mechanisms in protein folding, Annu. Rev. Biophys. Biomech. 29 (2000) 327–359.
[4] D. Mohanty, R. Elber, D. Thirumalai, D. Beglov, B. Roux, Kinetics of peptide folding: computer simulations of SYPFDV and peptide variants in water, J. Mol. Biol. 272 (3) (1997) 423–442.
[5] D. Mohanty, R. Elber, D. Thirumalai, Probing the role of local propensity in peptide turn formation, Int. J. Quantum. Chem. 80 (4/5) (2000) 1125–1128.
[6] X.W. Wu, S.M. Wang, Folding studies of a linear pentamer peptide adopting a reverse turn conformation in aqueous solution through molecular dynamics simulation, J. Phys. Chem. B. 104 (33) (2000) 8023–8034.

[7] G. Hummer, A.E. Garcia, S. Garde, Helix nucleation kinetics from molecular simulations in explicit solvent, Proteins 42 (1) (2001) 77–84.
[8] X. Daura, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, Reversible peptide folding in solution by molecular dynamics simulation, J. Mol. Biol. 280 (5) (1998) 925–932.
[9] X. Daura, I. Antes, W.F. van Gunsteren, W. Thiel, A.E. Mark, The effect of motional averaging on the calculation of nmr-derived structural properties, Proteins 36 (4) (1999) 542–555.
[10] X.W. Wu, S.M. Wang, Helix folding of an alanine-based peptide in explicit water, J. Phys. Chem. B. 105 (11) (2001) 2227–2235.
[11] X. Daura, K. Gademann, H. Schafer, B. Jaun, D. Seebach, W.F. van Gunsteren, The beta-peptide hairpin in solution: conformational study of a beta-hexapeptide in methanol by nmr spectroscopy and md simulation, J. Am. Chem. Soc. 123 (10) (2001) 2393–2404.
[12] Y. Duan, P.A. Kollman, Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, Science 282 (5389) (1998) 740–744.
[13] Y. Duan, L. Wang, P.A. Kollman, The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation, Proc. Natl. Acad. Sci. U.S.A. 95 (17) (1998) 9897–9902.
[14] T. Lazaridis, M. Karplus, Discrimination of the native from misfolded protein models with an energy function including implicit solvation, J. Mol. Biol. 288 (3) (1999) 477–487.
[15] P. Ferrara, J. Apostolakis, A. Caflisch, Evaluation of a fast implicit solvent model for molecular dynamics simulations, Proteins 46 (1) (2002) 24–33.
[16] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semi-analytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[17] P. Ferrara, A. Caflisch, Folding simulations of a three-stranded antiparallel beta-sheet peptide, Proc. Natl. Acad. Sci. U.S.A. 97 (20) (2000) 10780–10785.
[18] A. Hiltpold, P. Ferrara, J. Gsponer, A. Caflisch, Free energy surface of the helical peptide y(meara)(6), J. Phys. Chem. B. 104 (43) (2000) 10080–10086.
[19] B. Zagrovic, C.D. Snow, S. Khaliq, M.R. Shirts, V.S. Pande, Native-like mean structure in the unfolded ensemble of small proteins, J. Mol. Biol. 323 (1) (2002) 153–164.
[20] S. Chowdhury, M.C. Lee, G.M. Xiong, Y. Duan, Ab initio folding simulation of the trp-cage mini-protein approaches nmr resolution, J. Mol. Biol. 327 (3) (2003) 711–717.
[21] P. Ferrara, J. Apostolakis, A. Caflisch, Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations, J. Phys. Chem. B. 104 (20) (2000) 5000–5010.
[22] L. Wang, Y. Duan, R. Shortle, B. Imperiali, P.A. Kollman, Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures, Protein Sci. 8 (6) (1999) 1292–1304.
[23] H.W. Wang, S.S. Sung, Molecular dynamics simulations of three-strand beta-sheet folding, J. Am. Chem. Soc. 122 (9) (2000) 1999–2009.
[24] P. Ferrara, A. Caflisch, Native topology or specific interactions. What is more important for protein folding? J. Mol. Biol. 306 (4) (2001) 837–850.
[25] B. Zagrovic, E.J. Sorin, V. Pande, Beta-hairpin folding simulations in atomistic detail using an implicit solvent model, J. Mol. Biol. 313 (1) (2001) 151–169.
[26] S. Jang, S. Shin, Y. Pak, Molecular dynamics study of peptides in implicit water: ab initio folding of beta-hairpin, beta-sheet, and beta beta alpha-motif, J. Am. Chem. Soc. 124 (18) (2002) 4976–4977.
[27] A. Cavalli, P. Ferrara, A. Caflisch, Weak temperature dependence of the free energy surface and folding pathways of structured peptides, Proteins 47 (3) (2002) 305–314.
[28] C. Simmerling, B. Strockbine, A.E. Roitberg, All-atom structure prediction and folding simulations of a stable protein, J. Am. Chem. Soc. 124 (38) (2002) 11258–11259.
[29] C.D. Snow, N. Nguyen, V.S. Pande, M. Gruebele, Absolute comparison of simulated and experimental protein-folding dynamics, Nature 420 (6911) (2002) 102–106.
[30] C.D. Snow, B. Zagrovic, V.S. Pande, The trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations, J. Am. Chem. Soc. 124 (49) (2002) 14548–14549.
[31] J.W. Pitera, W. Swope, Understanding folding and design: replica-exchange simulations of "trp-cage" miniproteins, PNAS 100 (13) (2003) 7587–7592.
[32] R. Bonneau, J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C.E.M. Strauss, D. Baker, Rosetta in CASP4: progress in ab initio protein structure prediction, Proteins 45 (S5) (2001) 119–126.
[33] Y.X. Liu, D.L. Beveridge, Exploratory studies of ab initio protein structure prediction: Multiple copy simulated annealing, amber energy functions, and a generalized born/solvent accessibility solvation model, Proteins 46 (1) (2002) 128–146.
[34] R. Samudrala, Y. Xia, E. Huang, M. Levitt, Ab initio protein structure prediction using a combined hierarchical approach, Proteins 37 (S3) (1999) 194–198.
[35] D.J. Osguthorpe, Improved ab initio predictions with a simplified, flexible geometry model, Proteins 37 (S3) (1999) 186–193.
[36] R.J. Wawak, M.M. Wimmer, H.A. Scheraga, Application of the diffusion equation method of global optimization to water clusters, J. Phys. Chem. (US) 96 (12) (1992) 5138–5145.
[37] J. Kostrowicki, H.A. Scheraga, Application of the diffusion equation method for global optimization to oligopeptides, J. Phys. Chem. (US) 96 (18) (1992) 7442–7449.
[38] T. Huber, A.E. Torda, W.F. vanGunsteren, Structure optimization combining soft-core interaction functions, the diffusion equation method, and molecular dynamics, J. Phys. Chem. A 101 (33) (1997) 5926–5930.
[39] R.V. Pappu, G.R. Marshall, J.W. Ponder, A potential smoothing algorithm accurately predicts transmembrane helix packing, Nat. Struct. Biol. 6 (1) (1999) 50–55.
[40] P. Koehl, M. Delarue, Mean-field minimization methods for biological macromolecules, Curr. Opin. Struct. Biol. 6 (2) (1996) 222–226.
[41] A. Roitberg, R. Elber, Modeling side-chains in peptides and proteins—application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations, J. Chem. Phys. 95 (12) (1991) 9277–9287.
[42] C. Simmerling, T. Fox, P.A. Kollman, Use of locally enhanced sampling in free energy calculations: testing and application to the alpha→beta anomerization of glucose, J. Am. Chem. Soc. 120 (23) (1998) 5771–5782.
[43] C. Simmerling, J.L. Miller, P.A. Kollman, Combined locally enhanced sampling and particle mesh Ewald as a strategy to locate the experimental structure of a non-helical nucleic acid, J. Am. Chem. Soc. 120 (29) (1998) 7149–7155.
[44] C. Simmerling, M.R. Lee, A.R. Ortiz, A. Kolinski, J. Skolnick, P.A. Kollman, Combining MONSSTER and LES/PME to predict protein structure from amino acid sequence: application to the small protein CMTI-1, J. Am. Chem. Soc. 122 (35) (2000) 8392–8402.
[45] U.H.E. Hansmann, Y. Okamoto, New Monte Carlo algorithms for protein folding, Curr. Opin. Struct. Biol. 9 (2) (1999) 177–183.
[46] J. Skolnick, A. Kolinski, D. Kihara, M. Betancourt, P. Rotkiewicz, M. Boniecki, Ab initio protein structure prediction via a combination of threading, lattice folding, clustering, and structure refinement, Proteins 45 (S5) (2001) 149–156.
[47] X.W. Wu, S.M. Wang, Self-guided molecular dynamics simulation for efficient conformational search, J. Phys. Chem. B. 102 (37) (1998) 7238–7250.
[48] X.W. Wu, S.M. Wang, Enhanced systematic motion in molecular dynamics simulation, J. Chem. Phys. 110 (1999) 9401–9410.
[49] I. Andricioaei, A.R. Dinner, M. Karplus, Self-guided enhanced sampling methods for thermodynamic averages, J. Chem. Phys. 118 (3) (2003) 1074–1084.

[50] X.W. Wu, B.R. Brooks, Self-guided Langevin dynamics method, Chem. Phys. Lett. 381 (3-4) (2003) 512–518.
[51] D.J. Williams, K.B. Hall, Unrestrained stochastic dynamics simulations of the uucg tetraloop using an implicit solvation model, Biophys. J. 76 (6) (1999) 3192–3205.
[52] S. Elmer, V.S. Pande, A new twist on the helix-coil transition: a non-biological helix with protein-like intermediates and traps, J. Phys. Chem. B 105 (2) (2001) 482–485.
[53] R. Luo, L. David, M.K. Gilson, Accelerated Poisson–Boltzmann calculations for static and dynamic systems, J. Comput. Chem. 23 (2002) 1244–1253.
[54] J.Q. Lu, R. Luo, A Poisson–Boltzmann dynamics method with the non-periodic boundary condition, J. Chem. Phys. 119 (21) (2003) 11035–11047.
[55] E.Z. Wen, J.Q. Lu, W. Wang, R. Luo, Is the Poisson–Boltzmann surface area solvent sufficient for simulating dynamics of biomolecules? evidences from equilibrium and non-equilibrium simulations, in preparation.
[56] M.D. Struthers, R.P. Cheng, B. Imperiali, Economy in protein design: evolution of a metal-independent beta beta alpha motif based on the zinc finger domains, J. Am. Chem. Soc. 118 (13) (1996) 3073–3081.
[57] C.J. McKnight, D.S. Doering, P.T. Matsudaira, P.S. Kim, A thermostable 35-residue subdomain within villin headpiece, J. Mol. Biol. 260 (2) (1996) 126–134.
[58] C.J. McKnight, P.T. Matsudaira, P.S. Kim, Nmr structure of the 35-residue villin headpiece subdomain, Nat. Struct. Biol. 4 (3) (1997) 180–184.
[59] L.P. Lee, B. Tidor, Optimization of binding electrostatics: charge complementarity in the barnase–barstar protein complex, Protein Sci. 10 (2) (2001) 362–377.
[60] M.J. Hsieh, R. Luo, A physical scoring function based on the AMBER force field and the Poisson–Boltzmann implicit solvent for protein structure prediction, Proteins, submitted.
[61] T. Darden, D. York, L. Pedersen, Particle mesh Ewald—an $n\log(n)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (12) (1993) 10089–10092.
[62] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, J. Am. Chem. Soc. 117 (19) (1995) 5179–5197.
[63] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[64] W.F. van Gunsteren, H.J.C. Berendsen, Mol. Phys. 34 (1977) 1311.
[65] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327.
[66] L. Verlet, Computer experiments on classical fluids. Part i: Thermodynamical properties of Lennard–Jones molecules, Phys. Rev. 159 (1) (1967) 98–100.
[67] T.Z. Lwin, M.J. Hsieh, E.Z. Wen, R. Luo, Lessons learned from a comparison of enhanced samplings by self-guiding and replica-exchange simulated annealing, in preparation.
[68] D. Sitkoff, K.A. Sharp, B. Honig, Accurate calculation of hydration free energies using macroscopic solvation models, J. Phys. Chem. 98 (1994) 1978–1988.
[69] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force field for molecular mechanical simulation of nucleic acids and proteins, J. Am. Chem. Soc. 106 (1984) 765–784.
[70] M. Feig, J. Karanicolas, C.L. Brooks, Mmtsb tool set, MMTSB NIH Research Resource, The Scripps Research Institute.
[71] W. Kabsch, C. Sander, Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.
[72] K. Kuwajima, The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure, Proteins 6 (2) (1989) 87–103.
[73] M. Arai, P. Hamel, E. Kanaya, K. Inaka, K. Miki, M. Kikuchi, K. Kuwajima, Effect of an alternative disulfide bond on the structure, stability, and folding of human lysozyme, Biochemistry (US) 39 (12) (2000) 3472–3479.