Journal of Molecular Graphics and Modelling 72 (2017) 136–147

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Sampling conformational space of intrinsically disordered proteins in explicit solvent: Comparison between well-tempered ensemble approach and solute tempering method

Mengzhi Han[a,b], Ji Xu[a,*], Ying Ren[a,*]

[a] State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 6 July 2016
Received in revised form
18 December 2016
Accepted 23 December 2016
Available online 28 December 2016

Keywords:
Intrinsically disordered protein
Well-tempered ensemble
Replica exchange with solute tempering
Free energy surface
Molecular dynamics

ABSTRACT

Intrinsically disordered proteins (IDPs) are a class of proteins that expected to be largely unstructured under physiological conditions. Due to their heterogeneous nature, experimental characterization of IDP is challenging. Temperature replica exchange molecular dynamics (T-REMD) is a widely used enhanced sampling method to probe structural characteristics of these proteins. However, its application has been hindered due to its tremendous computational cost, especially when simulating large systems in explicit solvent. Two methods, parallel tempering well-tempered ensemble (PT-WTE) and replica exchange with solute tempering (REST), have been proposed to alleviate the computational expense of T-REMD. In this work, we select three different IDP systems to compare the sampling characteristics and efficiencies of the two methods Both the two methods could efficiently sample the conformational space of IDP and yield highly consistent results for all the three IDPs. The efficiencies of the two methods: are compatible, with about 5–6 times better than the plain T-REMD. Besides, the advantages and disadvantages of each method are also discussed. Specially, the PT-WTE method could provide temperature dependent data of the system which could not be achieved by REST, while the REST method could readily be used to a part of the system, which is quite efficient to simulate some biological processes.

1. Introduction

Intrinsically disordered proteins (IDPs) are a class of proteins that naturally exist in unstructured states and lack well-defined secondary or tertiary structures in solution [1–3]. Recent studies have shown that IDPs are abundant in eukaryotes [4] and play important roles in various physiological processes, especially in cellular signal transduction, translation regulation and transcription [5,6]. In addition to this, many IDPs are also strongly associated with human diseases [7,8]. Due to their heterogeneous nature, experimental characterization of IDP is challenging. Molecular dynamics (MD) simulations, with its high temporal-spatial resolutions that are not readily accessible to current experimental methods, constitute a valuable tool in investigating IDP systems [9–13]. However, the sampling ability of standard brute-force MD simulation is very limited, which usually only sample a local region of the conformational space. To alleviate this problem, various enhanced sampling methods have been proposed [14–17]. Among them, the temperature replica exchange molecular dynamics [18] (T-REMD) is one of the most commonly used methods.

In T-REMD, multiple replicas of a system are simulated at different temperatures simultaneously. Adjacent replicas are exchanged periodically using the Metropolis criterion, guaranteeing the canonical probability distribution at each temperature. The sampling of low temperature replicas is enhanced by travelling to high temperatures, in which the energy barriers are more easily overcome. Unlike collective variable (CV) based methods, T-REMD is easy to set up and does not require basic knowledge of the simulated system. For heterogeneous IDP systems, selecting a suitable set of CVs is not an easy task. Therefore, T-REMD is quite suitable for studying IDP systems. It has been successfully employed in the investigation of various IDPs [19–23]. However, a disadvantage of T-REMD is that the sampling efficiency relies on the fast diffusion in temperature space of each replica, which requires sufficient overlap between the energy distributions of neighboring replicas. Since the dependence of energy on system size, larger system sizes require closer temperature spacing, and therefore a greater number of replicas to cover a temperature range. When simulating in

* Corresponding authors.
E-mail addresses: xuji@ipe.ac.cn (J. Xu), yren@ipe.ac.cn (Y. Ren).

http://dx.doi.org/10.1016/j.jmgm.2016.12.014
1093-3263/

explicit solvent, the dramatic increase in the required number of replicas quickly makes the computational cost prohibitively large. There have been two methods which could alleviate this problem: one is the well-tempered ensemble (WTE) method [24]; the other is the replica exchange with solute tempering (REST) method [25,26].

The WTE method is originally from well-tempered metadynamics [27]. When the system potential energy is used as a biased CV in well-tempered metadynamics, one gets a biased ensemble termed WTE. In this ensemble, the average energy remains close to the canonical value, while its fluctuations are enhanced in a tunable way with the well-tempered metadynamics bias factor [24]. When combining with T-REMD, usually referred to as the PT-WTE method, the exchange probability between neighboring replicas is greatly increased, thus the number of replicas needed to cover a temperature range is greatly reduced [28]. The canonical averages of all properties of interest can be obtained by an elegant reweighing algorithm [29]. It has further been demonstrated that PT-WTE method could reduce the computational cost of T-REMD simulations with negligible effects on canonically averaged observables [30]. The PT-WTE method has been successfully used in the study of various protein systems [31–33], including IDP systems [34,35].

The REST method is originally from Hamiltonian replica exchange method [36,37] (HREM). In REST, the simulation system is first divided into two parts. For the simulation of protein in explicit solvent, the protein molecule is usually taken as one part (labeled as p) and the water molecules as the other part (labeled as w). Then the total interaction energy of the system (Hamiltonian of the system) can be decomposed into three components: the protein intramolecular energy, E_{pp}; the interaction energy between the protein and water, E_{pw}; and the interaction energy between water molecules, E_{ww}. Each replica of REST is run at the same temperature albeit on different deformed potential energy surfaces by properly scaling of these three components. The energy of the first replica, usually referred to as replica 0 or the 'cold' replica, is not scaled. As one climbs up the replica ladder, the energy term E_{pp} and E_{pw} are gradually scaled down. Since the ensemble probability only depends on $E/(k_BT)$, scaling down energy is equivalent to increasing temperature. So the protein is effectively heated up and sampled at high temperatures, thus increasing the barrier crossing, while the water remains cold in higher temperature replicas. These replicas are referred to as 'hot' replicas. The exchange between each replicas obeys the Metropolis criteria and detailed balance condition. In this way, the number of replicas required only scales as the protein degrees of freedom, which is relatively small compared to the whole system degrees of freedom. Thus a small number of replicas are sufficient to cover a temperature range and achieve good exchange probabilities, greatly reducing the computational cost [25,26]. It is important to note that the hot replicas in REST are sampled on deformed potential energy surfaces, which are not corresponding to physical high temperature simulations. Only replica 0 samples the physically meaningful canonical ensemble and other replicas are only used to accelerate sampling. The REST method has also been successfully utilized in many simulation works [38–41].

In this work, we select three different IDP systems to compare the sampling characteristics and abilities of the two enhanced sampling methods: the PT-WTE method and the REST method. The first two systems are unfolded model peptides of sequence EGAAXAAS, where X corresponding to G and W in this study. These peptides have been extensively investigated by means of NMR spectroscopy [42] and have been adopted in studying the accuracy of current force-fields in modeling protein disordered states [35]. The third system is a natural intrinsically disordered peptide, the extreme C-terminal domain (CTD) of tumor suppressor p53. The p53 protein has been extensively studied since it regulates many genes involved in cell cycle and apoptosis and is arguably one of the most important proteins in cancer [43]. The p53 CTD represents a widely accepted site for its regulation and performs various roles in cell [44,45]. Besides, it is one of a few IDPs that have been experimentally shown to be able to adopt different folded conformations upon binding to different proteins, including α-helix, β-strand and distinct loops [46]. Plenty of simulation works have been devoted to study the interaction of p53 CTD with its partners [47–50]. However, the conformational ensemble of this IDP has not been well characterized in explicit solvent. By simulating these three systems, we would like to investigate the relative efficiencies of each method compared to plain T-REMD, the accuracy and the ensemble generated by each method and the advantages and disadvantages of each method. Some practical differences between the two methods and some extensions are also discussed.

2. Methods

2.1. General setup

In this study, three different systems were selected to compare the two enhanced sampling methods, the PT-WTE method and the REST method, to simulate IDP systems in explicit solvent. The first two systems were unfolded model peptides with sequence EGAAGAASS (referred to as PepG) and EGAAWAASS (referred to as PepW), respectively. For each peptide, the initial conformation was created using the vmd [51] molefacture plugin with no secondary structure content. Both the N-termini and C-termini were not blocked and ionized (NH3+ and COO−, respectively), corresponding to the previous experimental structures. The protonation state of each amino acid was assigned according to the neutral biological conditions automatically by the pdb2gmx tool in Gromacs. The GLU residue in PepG and PepW was unprotonated. Each peptide was solvated in a dodecahedron periodic box containing about 1800 water molecules and certain number of ions. The third system was a 14-residue segment of p53 CTD (residues 375–388: QSTSRHKKLMFKTE). The initial conformation was obtained from a high temperature unfolding simulation starting from a crystal structure (PDB code: 1DT7 [52]). Neutral terminal patches ACE and CT3 were added to the N-termini and C-termini respectively to mimic its protein environment. The LYS381 residue in p53 CTD was protonated. The HIS380 was unprotonated in our simulation, which was assigned based on an optimal hydrogen bonding conformation. The peptide was also solvated in a dodecahedron box containing about 3000 water molecules and certain number of ions. Detail information of each system was summarized in Table S1 in Supporting information (SI).

All of the simulations were performed using GROMACS4 MD code [53] with PLUMED plugin [54,55] for PT-WTE and REST simulations. The Amber03w force field [56] and TIP4P/2005 model [57] were used for protein and water molecules, which have previously been shown to be able to accurately model IDPs [22,35,58,59]. The initial equilibrium procedures were as follows: each system was first energy-minimized using the steepest descent method. Then a 1-ns-long MD simulation was performed in the NVT ensemble for water relaxation with position constraints applied on protein heavy atoms. The temperature was kept at 298 K using the stochastic v-rescale thermostat [60]. After that, a 4-ns-long MD simulation was carried out in the NPT ensemble without any constraints and pressure and temperature kept to 1 atm and 298 K, respectively. The Parrinello-Rahman barostat [61,62] was utilized to enforce constant pressure. In all the simulations, an integration time step of 2 fs was used and all covalent bonds were constrained with the LINCS algorithm [63]. Particle Mesh Ewald method [64,65] was adopted for long-range electrostatics interactions, with a non-bonded cutoff equal to 1.2 nm. The final structures after equilibration were used for the following simulations.

2.2. Plain T-REMD simulation of PepG system

A plain T-REMD simulation of the PepG system was performed for reference and comparison. 50 replicas were simulated spanning the temperature range from 298 to 645 K. The replica temperatures were chosen according to the distribution proposed in reference [66], which could guarantee to achieve roughly uniform exchanges in explicit solvent. The plain T-REMD simulation was performed in NVT ensemble using the stochastic v-rescale thermostat to maintain the temperature. Each replica was simulated for 200 ns, resulting in a total simulation time of 10 μs. Exchange between neighboring replicas was attempted every 1 ps and the resulting acceptance probability was about 20–25% on average. The 298 K ensemble was analyzed with the initial 20 ns trajectory discarded as equilibration time.

2.3. PT-WTE simulations

The PT-WTE simulations were performed for all the three systems. In PT-WTE simulation, a metadynamics bias was constructed in the potential energy space to enhance energy fluctuations and reduce the number of replicas needed [24]. The WTE method is originally from the well-tempered metadynamics method. In metadynamics, a history dependent bias potential is introduced to some selected degrees of freedom which are known as CVs. If the CV is a d-dimensional vector, the metadynamics potential is given by:

$$V_G(S(s),t) = \omega \sum_{\substack{t' = \tau_G, 2\tau_G, \cdots \\ t' < t}} \exp\left(-\sum_{\alpha=1}^{d} \frac{(S_\alpha(x) - s_\alpha(t'))^2}{2\delta s_\alpha^2}\right)$$

where ω is the height of each Gaussian, τ_G is the size of the time interval between successive Gaussian depositions, and δs_α is the Gaussian width. In the well-tempered formalism, the bias deposition rate decreases over simulation time, which is implemented by rescaling the Gaussian height ω according to:

$$\omega = \omega_0 \tau_G e^{-\frac{V_G(S,t)}{k_B \Delta T}}$$

where ΔT is a temperature parameter and ω_0 is the initial deposition Gaussian height. In practice, the parameter ΔT is controlled by a bias factor γ defined as $\gamma = (T + \Delta T)/T$. When the system potential energy is selected as the biased CV in well-tempered metadynamics, one gets the WTE sampling method. In our simulations, the PT-WTE simulations were also performed in NVT ensemble using the stochastic v-rescale thermostat to maintain the temperature. For the PepG and PepW systems, 11 replicas were simulated spanning the temperature range of 298–645 K, with temperature distribution according to reference [66]. The Gaussian bias was deposited every picosecond with an initial height of 2.5 kJ/mol, a width of 200 kJ/mol and a bias factor of 40. Each replica was simulated for 200 ns with the initial 20 ns trajectory discarded as equilibration time. Exchange was attempted every 1 ps and the resulting acceptance probability was about 30%. For the p53 system, 12 replicas were simulated spanning the temperature range of 300–600 K. The Gaussian bias was also deposited every picosecond, with an initial height of 2.5 kJ/mol, a width of 400 kJ/mol and a bias factor of 40. In this case, each replica was simulated for 400 ns with the initial 50 ns trajectory discarded as equilibration time. Exchange was also attempted every 1 ps and the resulting acceptance probability was about 30%.

2.4. REST simulations

The REST simulations were also performed for all the three systems, which were also performed in NVT ensemble. In REST simulation, the total interaction energy of the system is decomposed into three parts: E_{pp}, E_{pw} and E_{ww}. All the replicas are run at the same temperature but on different potential energy surface. To be specific, the potential energy of replica m is scaled by:

$$E_m^{REST}(X) = \frac{\beta_m}{\beta_0} E_{pp}(X) + \sqrt{\frac{\beta_m}{\beta_0}} E_{pw}(X) + E_{ww}(X)$$

Here, X represents the configuration of the whole system, $\beta_m = 1/k_B T_m$ and T_0 is the temperature of replica 0. We refer to $\lambda = \beta_m/\beta_0$ as the scaling factor, which is a number smaller than 1. Therefore, enhanced sampling is achieved through scaling the intramolecular potential energy of the protein by a factor of λ, so that the barriers separating different conformations are lowered [26]. It should be noted that the potential for replica 0 reduces to the normal potential. In REST, the difference between each replica is the different λ used. However, to make comparison with other methods, we will use the "temperature" $T_m = T_0/\lambda$ to represent each replica, which means the effective temperature of the protein relative to the unscaled potential energy. There is no optimized temperature distribution method in REST, so we just used a geometric distribution. For the PepG and PepW systems, 9 replicas were simulated spanning the effective temperature range of 298–645 K. Each replica was simulated for 200 ns with the initial 20 ns trajectory discarded as equilibration time. For the p53 system, 12 replicas were simulated spanning the effective temperature range of 300–600 K. Each replica was simulated for 400 ns with the initial 50 ns trajectory discarded as equilibration time. In all cases, exchange was attempted every 1 ps and the resulting acceptance probability was about 30%. Additional information and parameters of each simulation were summarized in Table S1.

2.5. Analysis

All the analyses presented in this work were performed using the tools within GROMACS4, the PLUMED plugin and custom scripts. For PT-WTE simulations, the equilibrium probability distributions of all quantities were reconstructed using the reweighting scheme [29]. Secondary structure content was determined using the DSSP algorithm [67]. The clustering analyses were performed with g_cluster tool based on protein backbone atoms using the gromos method [68] with a 0.3 nm cutoff. The NMR $^3J_{HNHA}$-couplings constants for PepG and PepW systems were calculated using the Karplus equation [69] with Ubiquitin parameters [70,71]. The CA chemical shifts for PepG and PepW systems were calculated using Shiftx2 [72]. Free energy profiles along different CVs for p53 system were calculated using PLUMED plugin. Trajectory visualization and rendering of molecular representations was completed with VMD [51]. Statistical errors for all the quantities were estimated with the block-averaging method [73].

In order to further gauge the convergence, the Kullblach-Leibler (KL) divergence analysis [74] was utilized to systematically compare the conformational ensembles generated by each method. The KL analysis was completed with the MutInf code package [75]. For PepG system, the T-REMD ensemble was taken as reference ensemble and PT-WTE and REST ensembles as target ensemble. For PepW and p53 CTD system, which have not been simulated with T-REMD method, the PT-WTE ensemble was taken as reference ensemble and the REST ensemble as target ensemble. The simulation trajectories after equilibrium time were analyzed. The whole equilibrium ensembles were split into 6 blocks to make a bootstrapping statistical correction. The local KL divergences for each residue were

calculated in the backbone dihedral φ and ψ space. The average KL divergences of the target and reference ensemble and the null hypothesis reference divergences of the reference ensemble for each residue were shown in Tables S2–S4 for PepG, PepW, and p53 CTD system, respectively.

3. Results and discussion

3.1. PepG and PepW system

Two unfolded model peptide systems, PepG and PepW, were first simulated with both PT-WTE and REST methods. In addition, the PepG system was also simulated with plain T-REMD method for comparison. Each replica of the system was simulated for 200 ns in each method. The convergence of each simulation was accessed by a cluster analysis, which measured the number of clusters explored as a function of simulation time, as shown in Fig. S1. It could be demonstrated that the cluster numbers gradually tended to constant values and the simulations have tended to converge in their simulation time. The KL divergence analysis was also utilized to systematically compare the conformational ensembles generated by each simulation. As shown in Tables S2 and S3, the local KL divergences of each residue were all not significant, which further indicated that the simulations have gradually tended to converge and similar conformational ensembles were generated by different methods.

In order to characterize the conformational ensembles generated by each method, some structural properties of each peptide were first measured. Specifically, the distributions of the solvent accessible surface area (SASA), end-to-end distance (EtoE distance), the radius of gyration (R_g) and secondary structure content of each peptide were analyzed, as shown in Fig. 1. The number of backbone-backbone hydrogen bonds (H-bonds) was also calculated for each peptide using a switch function implemented in PLUMED. The ensemble average values of these properties were summarized in Table 1.

In general, the results obtained from each method showed strong agreement with each other for both PepG and PepW systems. For all the structural properties, both the ensemble average values and their shapes of distribution all corresponded well between each simulation method, indicating that the conformational ensembles generated by each method matched very well. The broad distributions of EtoE distance and R_g (Fig. 1B and C) demonstrated that the peptides were much flexible and could adopt both extended and compact conformations. The secondary structure content obtained from each method was also consistent well for both of the two systems. The conformations were mostly disordered with a high population of coil, bend and turn structures, a small helical structure and a negligible sheet structure, which is consistent with their unfolded nature. For the PepG system, the results of PT-WTE deviated a litter bigger from the plain T-REMD and REST results, but they all within their error ranges. For the PepW system, the results of PT-WTE and REST corresponded well with each other. Besides, these results were also corresponding well with a previous simulation work [35]. These indicated that the three methods, plain T-REMD, PT-WTE and REST, could all well characterize the conformational ensemble of PepG and PepW system.

The structural analysis showed similar results on the conformational ensembles generated by each method. To further compare these ensembles and determine whether they were corresponding to the physical one, several NMR observables were calculated and compared with available experimental data.

The $^3J_{HNHA}$-couplings, which are related to the φ backbone dihedral angle and assumed to correlate with the secondary structure population, were first calculated. The Root-Mean Square Deviation (RMSD) of simulated $^3J_{HNHA}$-couplings from the NMR couplings was used to quantify the agreement with experiments. As shown in Fig. 2A, the $^3J_{HNHA}$-couplings RMSD values calculated from different methods corresponded very well for both PepG and PepW. These RMSD values were slightly greater than the reported predictor error, consistent with previous simulation [35]. The CA chemical shifts were also calculated using Shiftx2 software. The average RMSD values from experimental data of each method for PepG and PepW were shown in Fig. 2B. An excellent agreement between different methods and between the calculated and experimental CA chemical shifts was observed. The average RMSDs were much lower than the method accuracy in all the cases. The residue level $^3J_{HNHA}$-couplings and CA chemical shifts deviations from experimental values were also calculated, as shown in Fig. S2 and S3 . The results from different methods were also consistent well. These analyses demonstrated that the plain T-REMD, PT-WTE and REST method could all well sample the conformational space of PepG and PepW and generate roughly the same results.

It should be noted that in REST simulation, the high temperature replicas are evolved on deformed potential energy surfaces. So they do not correspond to physical high temperature simulations and the true high temperature conformational ensembles could not be obtained. On the contrary, the high temperature ensemble could be obtained from plain T-REMD simulation or PT-WTE simulation by the reweighting algorithm. To investigate this property, structural properties and secondary structure content of highest temperature replica of the three methods for PepG system were compared, as shown in Fig. 3. As expected, the distributions of SASA, EtoE distance and R_g corresponded well between plain T-REMD and PT-WTE simulations. However the REST results deviated a bit bigger from plain T-REMD and PT-WTE results. This highest temperature ensemble of REST did not correspond to any physically meaningful ensemble. The only effect of high temperature replicas in REST is to enhance the sampling of the cold replica. Due to the high temperature, the peptide mostly adopted coil, bend and turn structures with negligible sheet and helical structures, as shown in Fig. 3D. Thus it is an advantage of PT-WTE method that it inherently provides temperature dependent data of the system. It could be utilized to investigate the folding thermodynamics of globular proteins or the temperature dependence of IDP properties, which could not be achieved with REST method.

To characterize the sampling efficiency of each method, a cluster analysis was performed on the PepG system, since it was simulated with all the three methods. The number of clusters explored by each method as a function of simulation time of all replicas, as well as simulation time per replica was shown in Fig. 4. The number of clusters sampled by each method was comparable as a function of simulation time per replica, with the PT-WTE and REST a little superior than plain T-REMD method. This indicated that the sampling efficiency for each replica of the three methods is comparable. However, 50, 11 and 9 replicas were used for plain T-REMD, PT-WTE and REST method, respectively. So the sampling efficiencies of PT-WTE and REST are much higher than plain T-REMD, with about 5 times more efficient, which is clearly indicated by the number of clusters explored as a function of simulation time of all replicas. The final number of clusters explored by plain T-REMD, PT-WTE and REST was 306, 305 and 310, respectively. Therefore, the REST method could explore slightly more clusters than the other two methods, although they were basically the same. This demonstrated that the conformational space of PepG was thoroughly sampled by all the three methods. Both PT-WTE and REST methods could greatly reduce the number of replicas needed compared to plain T-REMD, thus they have a considerably higher sampling efficiency and could greatly reduce the computational cost. The relative efficiency of PT-WTE and REST was more subtle and probably system dependent. In

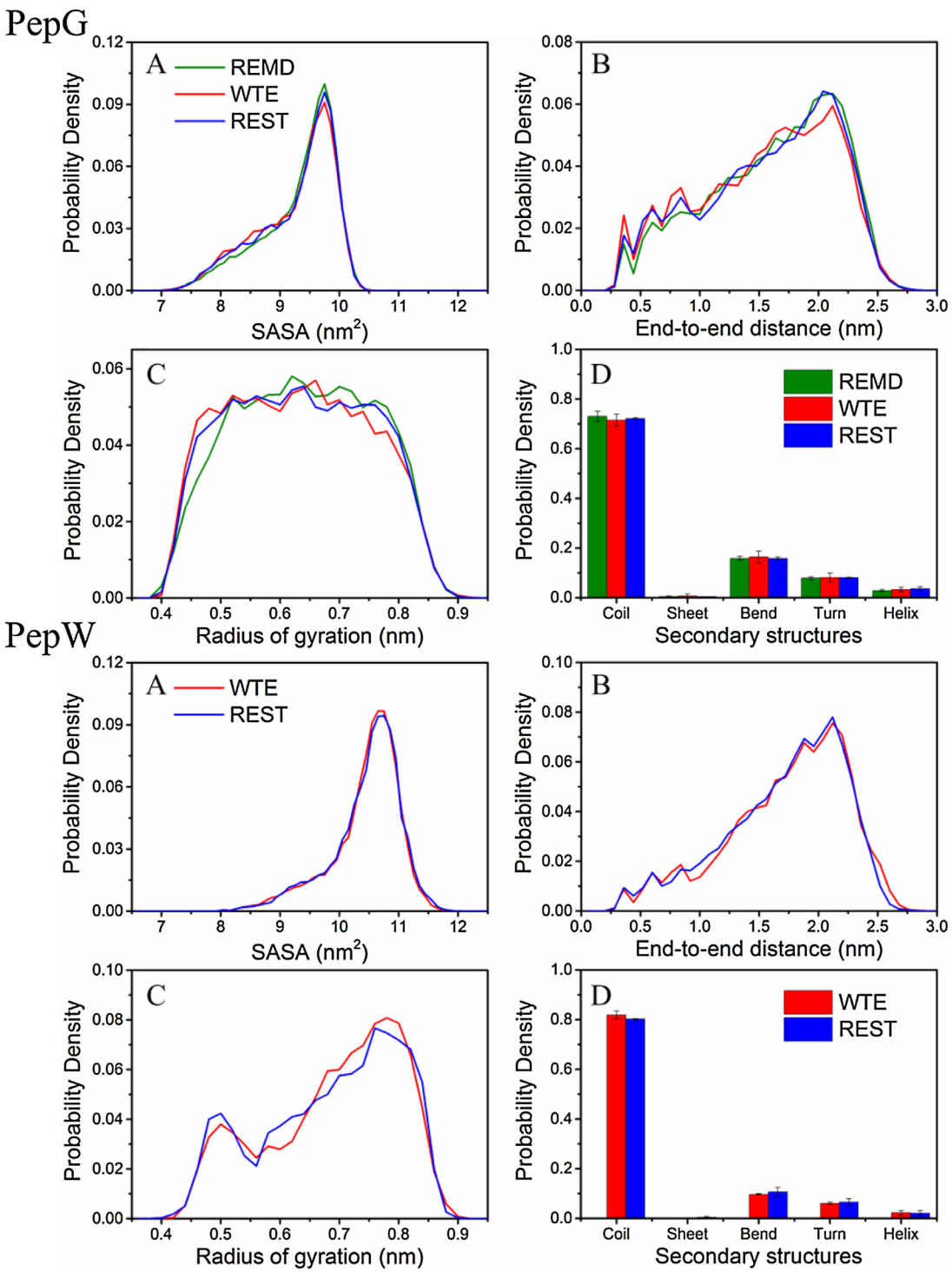

Fig. 1. Distributions of structural properties and secondary structure content generated by different methods for PepG (top) and PepW (bottom). (A) SASA. (B) EtoE distance. (C) R_g. (D) secondary structure content.

PT-WTE simulation, the energy fluctuation of each replica is dependent on the bias factor. In the PepG simulation, we used a bias factor of 40. Increasing this bias factor may further reduce the number of replicas needed. On the other hand, the exchange probability of REST and the number of replicas needed only depend on the protein degrees of freedom, thus they are system dependent. Adding more water molecules does not need more replicas in REST method. For IDP systems, which are often very flexible and could sample extended conformations, it often needs more water molecules to eliminate the boundary effect. So it is a potential advantage of REST method in IDP simulations.

The final number of clusters explored by each method for the PepG system was not the same, and the REST method explored the most number of clusters. The philosophy of REST method to enhance sampling is different from the PT-WTE or plain T-REMD method. To further investigate the conformational ensembles generated by each method, a cluster analysis was performed on the equilibrium ensembles (discard the first 20 ns trajectory) of

Table 1
Ensemble average values of the structural properties (SASA, EtoE distance, Rg and H-bonds) and secondary structure content generated by each method for PepG and PepW.

System		PepG			PepW	
Method		T-REMD	PT-WTE	REST	PT-WTE	REST
SASA (nm^2)		9.31	9.25	9.27	10.44	10.38
EtoE (nm)		1.62	1.55	1.58	1.81	1.72
Rg (nm)		0.64	0.63	0.63	0.69	0.66
H-bonds		1.28	1.33	1.33	1.58	1.62
Secondary structure	Coil	0.73	0.72	0.72	0.81	0.76
	Sheet	<0.01	<0.01	<0.01	<0.01	<0.01
	Bend	0.16	0.16	0.16	0.10	0.12
	Ture	0.08	0.08	0.08	0.06	0.08
	Helix	0.03	0.03	0.03	0.02	0.02

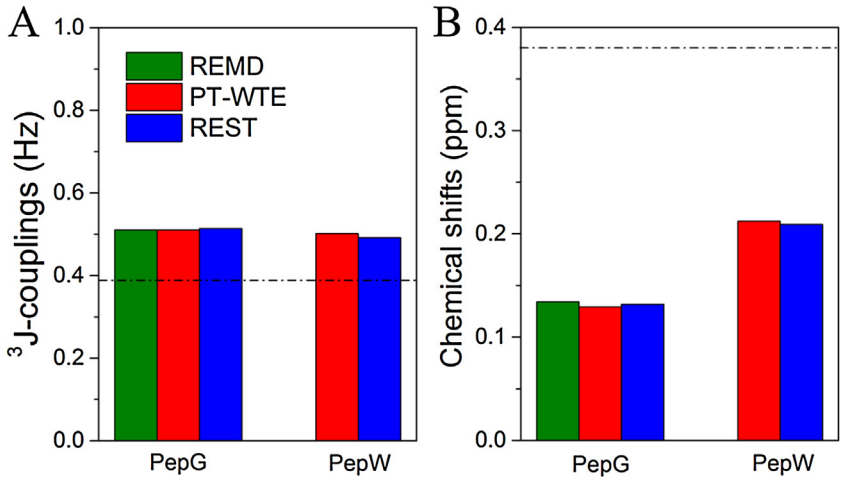

Fig. 2. RMSDs between simulated and experimental data for the $^3J_{HNHA}$-couplings (A) and CA chemical shifts (B) of each method for PepG and PepW. The horizontal dash dot lines indicate the typical predictor errors.

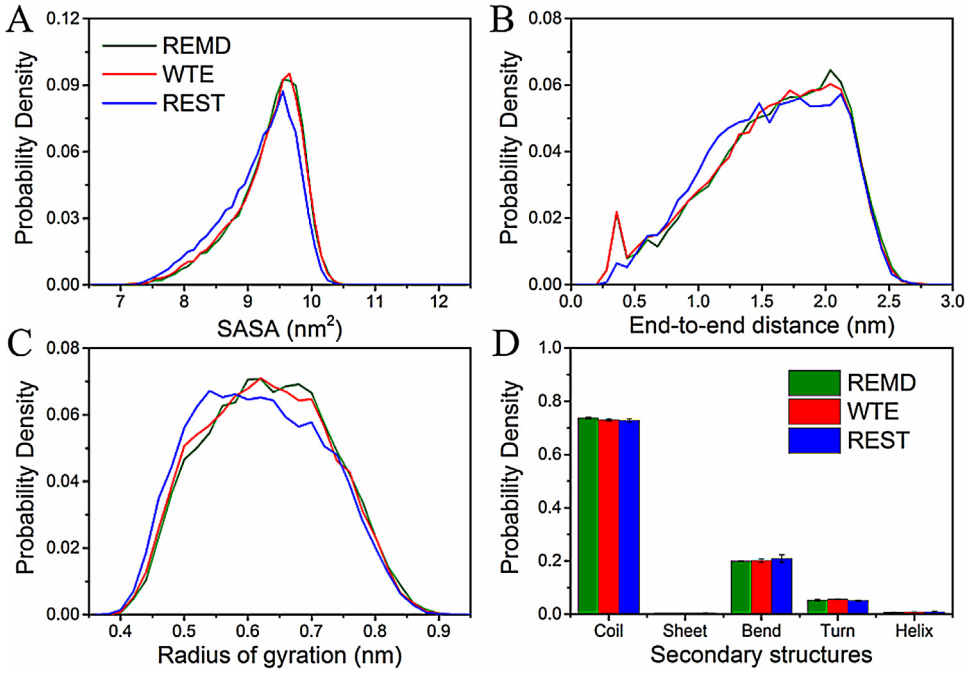

Fig. 3. Distributions of structural properties and secondary structure content of the highest temperature replica of each method for PepG system. (A) SASA. (B) EtoE distance. (C) R_g. (D) secondary structure content.

the three simulations. To verify the robustness of the clustering method, we utilized a block-averaging method to estimate the statistical errors for the occupancy of each cluster. The results were shown in Fig. S5 in SI. It clearly demonstrated that all the occupancies obtained were within their error ranges and the clustering method and parameters were chosen properly. The occupancies of

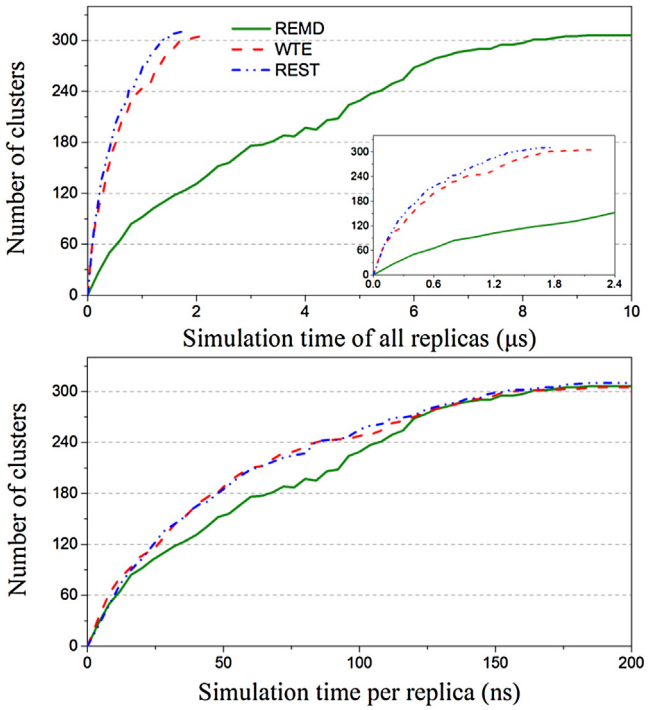

Fig. 4. Number of clusters explored as a function of simulation time of all replicas (top) and simulation time per replica (bottom) of each method for PepG system.

the top 20 clusters and representative structures of the top 6 clusters were shown in Fig. 5. It could be seen that the occupancies of top clusters were relatively small. The top 20 clusters only represented about 30% of the whole conformational space. This was consistent with the unfolded nature of PepG, which did not have dominant conformations in solution. The occupancies of the top 20 clusters sampled by each method corresponded very well with each other. In addition, the conformations of the top 6 clusters were also very similar, which mostly adopted extended structures. These all indicated that all the three simulations have tended to converge and the conformational ensembles generated by each method matched very well with each other, consistent with previous structural and NMR observable analyses. Even though the REST simulation explored slightly more clusters, these clusters had very small occupancies and located on high energy regions of the conformational space. Thus they contributed little to the ensemble average properties.

3.2. p53 system

To further compare the PT-WTE and REST methods, a natural intrinsically disordered peptide, 14-residue segment of p53 CTD, was selected and simulated with both the two methods. Although physiologically very important, the conformational ensemble of this peptide has not been well investigated in explicit solvent. It is usually difficult to characterize the ensemble of IDP since it mainly adopts random coil conformations or quickly converts between random coil and partially structured conformations. To compare the ensembles generated by PT-WTE and REST simulations, some CVs were first selected and the one dimensional free energy surfaces (FES) along these CVs were calculated. Specifically, we focused on the radius of gyration of C_α atoms (R_g), number of backbone hydrogen bonds (N—H bonds), coordination number of C_α-C_α contacts (C_α contacts), α-helix RMSD (α-RMSD), parallel β-sheet RMSD (Para-β-RMSD) and antiparallel β-sheet RMSD (Anti-β-RMSD). These CVs could efficiently measure the compactness or secondary structure content of proteins. The FESs along each

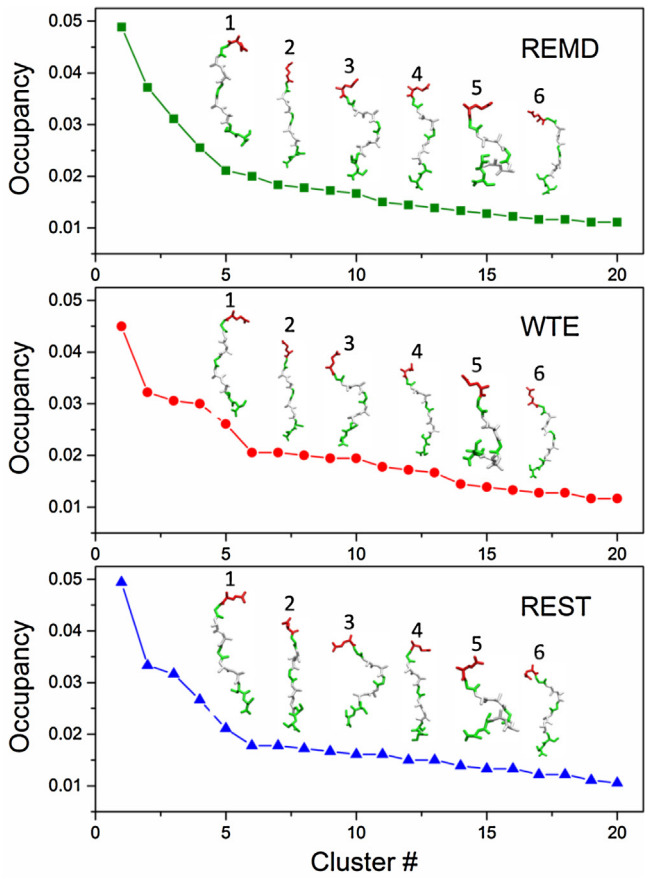

Fig. 5. Occupancies of the top 20 clusters and representative structures of top 6 clusters of the conformational ensembles generated by each method for PepG system.

CV were shown in Fig. 6. The convergences of the simulations were examined by monitoring the FESs as a function of simulation time, which were shown in Fig. S4. It could be demonstrated that the FESs along all the CVs converged rapidly after 160 ns simulation time and the simulations gradually tended to converge in 400 ns simulation time per replica. Besides, the KL divergence analysis was also utilized to compare the conformational ensembles generated by PT-WTE and REST. As PepG and PepW simulations, local KL divergences of each residue for p53 CTD were all not significant, indicating that the simulations have gradually tended to converge and generated similar conformational ensembles.

On the whole, the FES along different CVs obtained from PT-WTE and REST simulations corresponded very well with each other. The energy minima showed the same value on all the CVs except for the Cα contacts, which had a minimum on 80 for PT-WTE and 70 for REST. However, the FES along C_α contacts was quite flat. The energy differences over range 60–80 were usually within 2.5 kJ/mol, indicating that conformational ensemble generated by the two methods had little difference. The shapes of the FESs also showed similar behavior up to about 10 kJ/mol for all the CVs. The FES showed a slight discrepancy over range 0.3-1.2 for Para-β-RMSD and 0.6-2.0 for Anti-β-RMSD. From following analysis it could be seen that the peptide could hardly adopt β-sheet conformations. So larger deviations of FES along these two CVs were got since they were evaluated based on very few events. The sampling ability was also compatible for PT-WTE and REST methods. For some CVs, like R_g, N—H bonds, α-RMSD and Anti-β-RMSD, PT-WTE method could sample slight larger conformational space than REST, while for other CVs, Cα contacts and Para-β-RMSD, REST sampled larger conformational space. Judging from these one dimensional

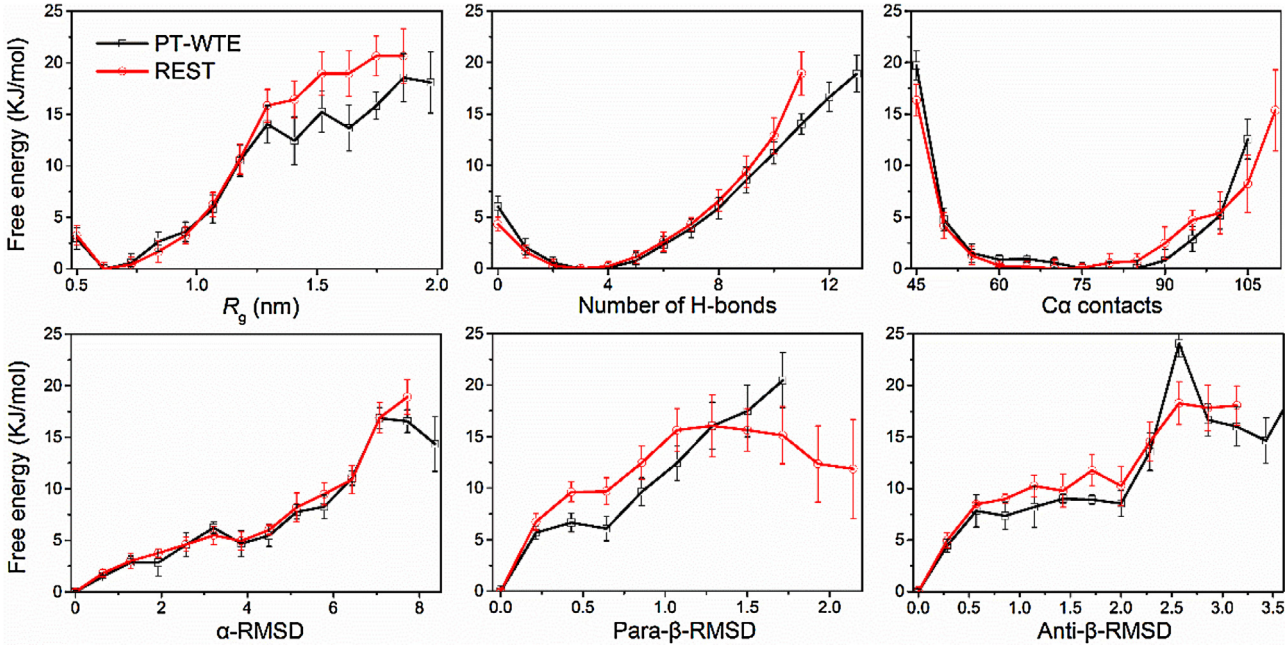

Fig. 6. Comparison of the FES projected on each different CVs obtained from PT-WTE and REST simulations. Error bars are calculated by the block-averaging method.

FESs, the peptide mainly presented compact conformations with a few secondary structures, which is consistent with its IDP natures.

Even though disordered, IDPs usually adopt transient residual secondary structures in solution. These residual secondary structures may of great importance for IDP physiological functions. Therefore, accurately estimating the secondary structure content of IDP is important to definitively evaluate its biophysical properties.

The average number of residues of p53 in each particular secondary structure and the per-residue secondary structure propensity was shown in Fig. 7. It could be seen that the results from both methods compare extremely well with each other for all structure types. Not only the average structure content, but also the per-residue secondary structure propensities were all corresponding very well between PT-WTE and REST. Consistent with the intrinsically disordered nature, a high preference of unstructured coil, bend and turn structures was observed. A moderate helical structures and a very low β-sheet like structures were also observed. The helical structures mainly focused on residue 7–9 (for α-helix) and residue 2–4 (for $3_{(10)}$-helix). Residue 5 and 12 had a high β-sheet propensity compared to other residues, although it is still very low compared to other structure types. The low β-sheet propensity also explained the big difference of the FES projected onto β secondary structure CVs (Fig. 6). Even though minor, these residual secondary structures may play important roles in IDP physiological functions. They may help to fine turn the IDP coupled folding and binding process. Both the PT-WTE and REST methods could accurately estimate the secondary structure content of p53 and generate rough the same results.

The p53 extreme C-terminus has been shown to be able to fold into different structures upon binding to different targets. Specifically, it could bind to S100B(ββ) (PDB ID: 1DT7), Sirtuin (PDB ID: 2H4J), Cyclin A2 (PDB ID: 1H26) or CBP (PDB ID: 1JSP) and fold into helix, partial helix, random coil or β-sheet like conformations, as shown in Fig. 8A. These folded conformations are quite different from each other. To examine whether any of these conformations could be sampled in the unbound state of p53, the distributions of C_α RMSD of the disordered ensembles to each of these folded conformations were calculated for both PT-WTE and REST simulations, as shown in Fig. 8B. As expected, the distributions obtained from PT-WTE and REST were almost the same, indicating the convergence and efficiency of the two methods. Unlike a previous implicit solvent simulation [47], minor peaks at small RMSD values were not observed, demonstrating that most of the conformations in the disordered ensemble were different from experimentally observed folded structures. Even though, it could still sample some conformations that are similar to some folded structures. For example, for 1H26 and 2H4J, there were significant parts of distributions smaller than 0.3 nm. However, for 1JSP, the distribution was almost all larger than 0.5 nm, indicating that the unbound conformations were quite different from 1JSP. It was consistent with the low β-sheet secondary structure of the unbound ensemble. In conclusion, the two methods generated almost the same distributions. Although minor, the unbound p53 could sample some conformations that resemble some experimentally observed folded structures.

Finally, a cluster analysis was performed based on the peptide backbone RMSD with a cutoff of 0.3 nm on both PT-WTE and REST ensembles to further compare the conformations generated by each method. The statistical errors for the occupancies of each cluster were also estimated by the block-averaging method and shown in Fig. S6 in SI. The results demonstrated that all the occupancies obtained were within their error ranges and the clustering method and parameters utilized were robust. A total of 83 and 95 clusters were found for PT-WTE and REST ensembles, respectively. The occupancies of the top 10 most populated clusters and representative structures of the top 5 most populated clusters were shown in Fig. 9. Unlike the PepG system, the conformational space of p53 CTD was not totally disordered and the top 10 clusters occupied about 53% of the whole conformational space. This is consistent with the IDP nature that it could rapid convert between random coil state and some partially folded states. Although the REST simulation generated a little more clusters than PT-WTE simulation, the occupancies of the top 10 clusters matched very well between the two methods. In addition, the representative structures of the top 5 clusters were also very similar. These demonstrated that the PT-WTE and REST simulations generated roughly the same ensemble. The conformations of the top clusters were largely coil structures,

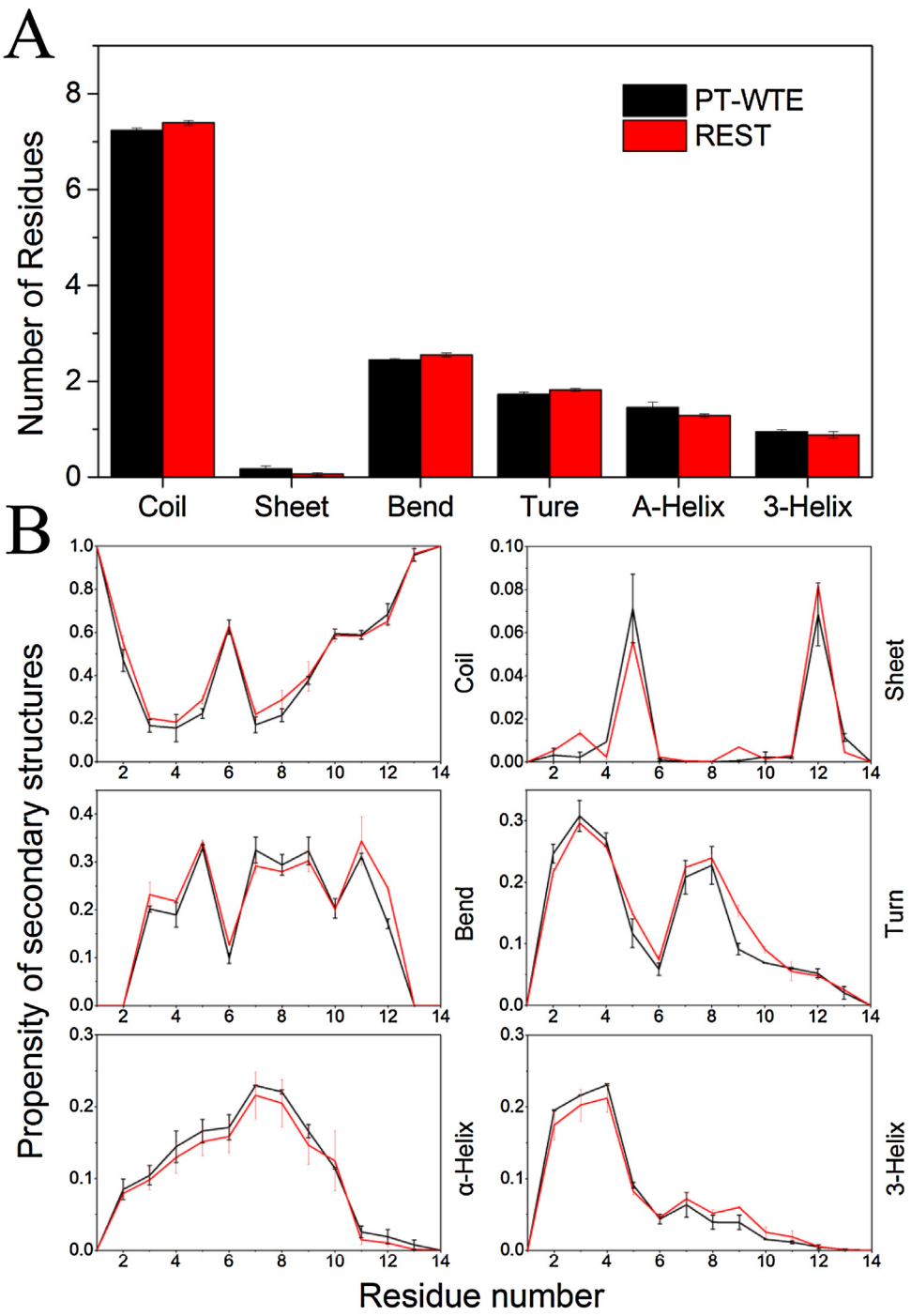

Fig. 7. (A) Average number of residues in each type of secondary structure determined from PT-WTE and REST simulations. (B) Per-residue secondary structure propensity determined from PT-WTE and REST simulations.

with some residual helical content, which is consistent with previous secondary structure analysis.

In summary, both the PT-WTE and REST methods could efficiently sample the conformational space of p53 and generate converged ensemble. The conformational ensembles generated by the two methods marched very well with each other. A similar number of replicas was used for PT-WTE and REST simulation, so the efficiency of the two methods was also compatible. It is worth noting that there is another advantage of the REST method compared to PT-WTE method. Although we did not employ in this work, the REST method could readily be used to only a part of the simulated system. This is a very useful feature in investigating some biological processes. For instance, in a previous work, the allosteric communication pathways in the KIX domain of CBP were simulated with the PT-WTE method [31]. To focuses on the sampling of the important part which takes major conformational changes and keep other parts un-perturbed, some restraints were employed to the non-important part to prevent their unfolding or conformational changes when travelling to high temperatures. These restraints should be fine turned to not perturb the system dynamics, which is not an easy task. However, if the REST method is utilized to simulate this process, these restraints are not required. One could only “heat” the important part to enhance its sampling and keep the other parts of the system un-perturbed. In this way,

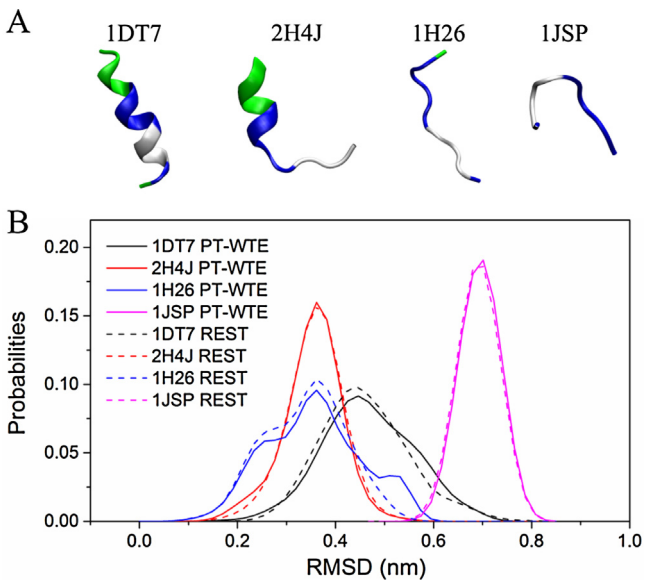

Fig. 8. (A) Cartoon representations of four folded conformations of p53 in complex with S100B(ββ) (1DT7), Sirtuin (2H4J), Cyclin A2 (1H26) and CBP (1JSP). (B) Probability distributions of Cα RMSD distances of the disordered ensembles obtained from the two methods to each of the folded conformations.

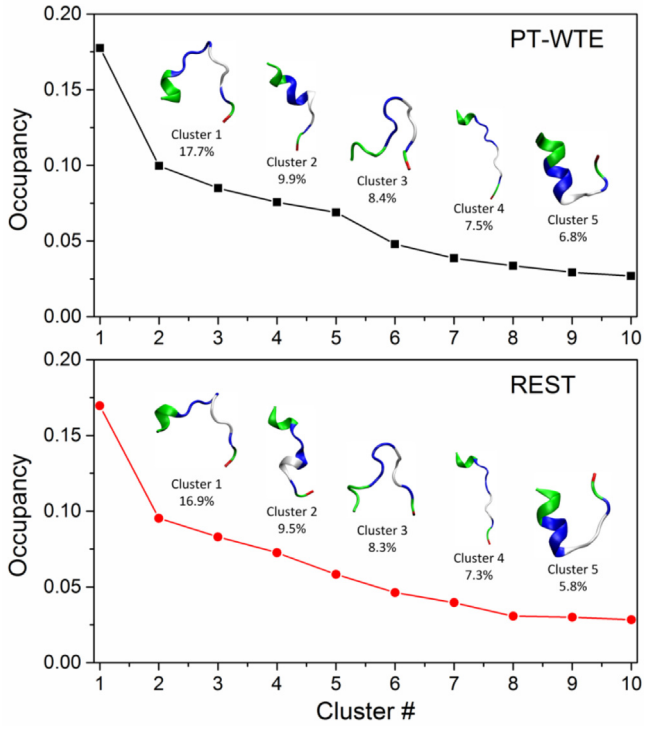

Fig. 9. Occupancies of top 10 clusters and representative structures of top 5 clusters of the conformational ensembles generated by PT-WTE and REST method.

the sampling of the important part is enhanced and the dynamics of the other parts are not changed. Another potential application of the REST method is to simulate the coupled folding and binding process of IDP to its partner. In this process, the major conformational changes occur in the IDP. The REST method could be only used to the IDP degrees of freedom and keep the dynamics of partner protein and water molecules unchanged.

4. Conclusions

Temperature replica exchange molecular dynamics is a popular and widely used enhanced sampling method. However, its application has been hindered due to its tremendous computational cost, especially for large systems in explicit solvent. Two methods, PT-WTE and REST, have been proposed to alleviate the computational expense of T-REMD. In this work, we selected three different IDP systems to investigate the sampling characteristics and efficiencies of the two methods. In general, both the two methods appear to be highly suitable for studying the structural properties of IDPs. Both methods yield highly consistent results based on various analyses. For PepG and PepW system, the structural properties and NMR observables are all consistent very well between the two methods, as well as with plain T-REMD simulation and previous work. For p53 system, both the one dimensional FESs projected on different CVs and the secondary structure content all correspond well between the two methods. Not only the average secondary structures of the entire peptide, but also the per-residue secondary structure propensities all agree well, indicating that the two methods generate roughly the same conformational ensemble for the p53 system.

Both PT-WTE and REST methods could greatly reduce the number of replicas needed when simulating in explicit solvent. The efficiencies of the two methods are compatible, with about 5–6 times better compared to plain T-REMD. In addition, the REST method needs slightly fewer replicas than the PT-WTE method for PepG and PepW systems. However, the relative efficiency of PT-WTE and REST is more subtle. The energy fluctuation depends on the bias factor in PT-WTE, so increasing the bias factor could further reduce the replicas needed to cover a certain temperature range. On the other hand, the number of replicas needed in REST only depends on the protein degrees of freedom. This is an advantage when simulating IDP systems, which could sample extended structures and may need large simulation box and more water molecules.

There is an advantage of PT-WTE method that it inherently provides temperature dependent data of the system. Such information could be utilized to investigate the folding thermodynamics of globular proteins or temperature dependence of IDP properties. However, in REST, the only effect of “high temperature” replicas is to enhance the sampling of the first replica, as they are evolved on deformed energy surfaces which do not correspond to any physical ensembles. On the other hand, an advantage of REST over PT-WTE is that it could be used to only a part of the system, which is quite efficient to simulate some biological processes like protein allosteric communications or the coupled folding and binding process of IDP to its partner.

Both PT-WTE and REST methods could readily combine with other CV based enhanced sampling methods, such as umbrella sampling and metadyanmics etc, to further enhance the sampling efficiency. There have been some works employed these combined methods [32,76]. In these methods, the major energy barriers along selected CVs are accounted for by CV based bias, while the minor hidden energy barriers which are not included in the selected CVs are well accounted for by the parallel tempering effect. Thus they are usually more efficient than the two individual methods. With these advanced enhanced sampling methods, more complicated conformational transition processes of proteins could be investigated in explicit solvent in future studies.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant Nos. 21103195 and 91434104, and

Ministry of Science and Technology of China under Grant No. COM2015A003.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2016.12.014.

References

[1] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, Intrinsically disordered protein, J. Mol. Graphics Modell. 19 (2001) 26–59.
[2] C.K. Fisher, C.M. Stultz, Constructing ensembles for intrinsically disordered proteins, Curr. Opin. Struct. Biol. 21 (2011) 426–431.
[3] R. van der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D.T. Jones, Classification of intrinsically disordered regions and proteins, Chem. Rev. 114 (2014) 6589–6631.
[4] R. Pancsa, P. Tompa, Structural disorder in eukaryotes, PLoS One 7 (2012) e34687.
[5] M. Fuxreiter, P. Tompa, I. Simon, V.N. Uversky, J.C. Hansen, F.J. Asturias, Malleable machines take shape in eukaryotic transcriptional regulation, Nat. Chem. Biol. 4 (2008) 728–737.
[6] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradović, A.K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, J. Mol. Biol. 323 (2002) 573–584.
[7] H.J. Dyson, P.E. Wright, Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell Biol. 6 (2005) 197–208.
[8] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annu. Rev. Biophys. 37 (2008) 215–246.
[9] J. Xu, M. Han, Y. Ren, J. Li, The principle of compromise in competition: exploring stability condition of protein folding, Sci. Bull. 60 (2015) 76–85.
[10] J. Xu, Y. Ren, J. Li, Multiscale simulations of protein folding: application to formation of secondary structures, J. Biomol. Struct. Dyn. 31 (2013) 779–787.
[11] M. Han, J. Xu, Y. Ren, J. Li, Simulations of flow induced structural transition of the β-switch region of glycoprotein Ibα, Biophys. Chem. 209 (2016) 9–20.
[12] S.-H. Chong, S. Ham, Assessing the influence of solvation models on structural characteristics of intrinsically disordered protein, Comp. Theor. Chem. 1017 (2013) 194–199.
[13] L. Shen, H.-F. Ji, Comparative study on the conformational stability of human and murine amyloid β peptide, Comp. Theor. Chem. 972 (2011) 44–47.
[14] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, J. Comput. Phys. 23 (1977) 187–199.
[15] A. Laio, M. Parrinello, Escaping free-energy minima, Proc. Natl. Acad. Sci. 99 (2002) 12562–12566.
[16] E. Darve, D. Rodríguez-Gómez, A. Pohorille, Adaptive biasing force method for scalar and vector free energy calculations, J. Chem. Phys. 128 (2008) 144120.
[17] S. Piana, A. Laio, A bias-exchange approach to protein folding, J. Phys. Chem. B 111 (2007) 4553–4559.
[18] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.
[19] N.G. Sgourakis, M. Merced-Serrano, C. Boutsidis, P. Drineas, Z. Du, C. Wang, A.E. Garcia, Atomic-level characterization of the ensemble of the Aβ (1–42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms, J. Mol. Biol. 405 (2011) 570–583.
[20] W. Zhang, D. Ganguly, J. Chen, Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins, PLoS Comput. Biol. 8 (2012) e1002353.
[21] M. Knott, R.B. Best, A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations, PLoS Comput. Biol. 8 (2012) e1002605.
[22] J. Mittal, T.H. Yoo, G. Georgiou, T.M. Truskett, Structural ensemble of an intrinsically disordered polypeptide, J. Phys. Chem. B 117 (2012) 118–124.
[23] C. Miller, G.I.H. Zerze, J. Mittal, Molecular simulations indicate marked differences in the structure of amylin mutants, correlated with known aggregation propensity, J. Phys. Chem. B 117 (2013) 16066–16075.
[24] M. Bonomi, M. Parrinello, Enhanced sampling in the well-tempered ensemble, Phys. Rev. Lett. 104 (2010) 190601.
[25] P. Liu, B. Kim, R.A. Friesner, B. Berne, Replica exchange with solute tempering: a method for sampling biological systems in explicit water, Proc. Natl. Acad. Sci. 102 (2005) 13749–13754.
[26] L. Wang, R.A. Friesner, B. Berne, Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2), J. Phys. Chem. B 115 (2011) 9431–9438.
[27] A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, Phys. Rev. Lett. 100 (2008) 020603.
[28] M. Deighan, M. Bonomi, J. Pfaendtner, Efficient simulation of explicitly solvated proteins in the well-tempered ensemble, J. Chem. Theory Comput. 8 (2012) 2189–2192.
[29] M. Bonomi, A. Barducci, M. Parrinello, Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics, J. Comput. Chem. 30 (2009) 1615–1621.
[30] A. Barducci, M. Bonomi, M. Parrinello, Linking well-tempered metadynamics simulations with experiments, Biophys. J. 98 (2010) L44–L46.
[31] F. Palazzesi, A. Barducci, M. Tollinger, M. Parrinello, The allosteric communication pathways in KIX domain of CBP, Proc. Natl. Acad. Sci. 110 (2013) 14237–14242.
[32] A. Barducci, M. Bonomi, M.K. Prakash, M. Parrinello, Free-energy landscape of protein oligomerization from atomistic simulations, Proc. Natl. Acad. Sci. 110 (2013) E4708–E4713.
[33] M. Deighan, J. Pfaendtner, Exhaustively sampling peptide adsorption with metadynamics, Langmuir 29 (2013) 7999–8009.
[34] P.R. Burney, N. White, J. Pfaendtner, Structural effects of methionine oxidation on isolated subdomains of human fibrin D and αC regions, PLoS One 9 (2014) e86981.
[35] F. Palazzesi, M.K. Prakash, M. Bonomi, A. Barducci, Accuracy of current all-atom force-fields in modeling protein disordered states, J. Chem. Theory Comput. 11 (2015) 2–7.
[36] H. Fukunishi, O. Watanabe, S. Takada, On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction, J. Chem. Phys. 116 (2002) 9058–9067.
[37] R. Affentranger, I. Tavernelli, E.E. Di Iorio, A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling, J. Chem. Theory Comput. 2 (2006) 217–228.
[38] L.B. Wright, T.R. Walsh, Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: insights from a quartz binding peptide, Phys. Chem. Chem. Phys. 15 (2013) 4715–4726.
[39] F. Musiani, E. Ippoliti, C. Micheletti, P. Carloni, S. Ciurli, Conformational fluctuations of UreG, an intrinsically disordered enzyme, Biochemistry 52 (2013) 2949–2954.
[40] A.H. Brown, P.M. Rodger, J.S. Evans, T.R. Walsh, Equilibrium conformational ensemble of the intrinsically disordered peptide n16N: linking subdomain structures and function in nacre, Biomacromolecules 15 (2014) 4467–4479.
[41] C.M. Miller, A.C. Brown, J. Mittal, Disorder in cholesterol-binding functionality of CRAC peptides: a molecular dynamics study, J. Phys. Chem. B 118 (2014) 13169–13174.
[42] S.A. Dames, R. Aregger, N. Vajpai, P. Bernado, M. Blackledge, S. Grzesiek, Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids, J. Am. Chem. Soc. 128 (2006) 13508–13514.
[43] A.J. Levine, C.A. Finlay, P.W. Hinds, P53 is a tumor suppressor gene, Cell 116 (2004) S67–S70.
[44] A.L. Kim, A.J. Raffo, P.W. Brandt-Rauf, M.R. Pincus, R. Monaco, P. Abarzua, R.L. Fine, Conformational and molecular basis for induction of apoptosis by a p53 C-terminal peptide in human cancer cells, J. Biol. Chem. 274 (1999) 34924–34931.
[45] K.L. Harms, X. Chen, The C terminus of p53 family proteins is a cell fate determinant, Mol. Cell. Biol. 25 (2005) 2014–2030.
[46] C.J. Oldfield, J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, A.K. Dunker, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, BMC Genom. 9 (2008) S1.
[47] J. Chen, Intrinsically disordered p53 extreme C-terminus binds to S100 B (ββ) through fly-casting, J. Am. Chem. Soc. 131 (2009) 2088–2089.
[48] W.J. Allen, D.G. Capelluto, C.V. Finkielstein, D.R. Bevan, Modeling the relationship between the p53 C-terminal domain and its binding partners using molecular dynamics, J. Phys. Chem. B 114 (2010) 13201–13213.
[49] C. McDowell, J. Chen, J. Chen, Potential conformational heterogeneity of p53 bound to S100B (ββ), J. Mol. Biol. 425 (2013) 999–1010.
[50] I. Staneva, Y. Huang, Z. Liu, S. Wallin, Binding of two intrinsically disordered peptides to a multi-specific protein: a combined Monte Carlo and molecular dynamics study, PLoS Comput. Biol. 8 (2012) e1002682.
[51] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[52] R.R. Rust, D.M. Baldisseri, D.J. Weber, Structure of the negative regulatory domain of p53 bound to S100B (ββ), Nat. Struct. Mol. Biol. 7 (2000) 570–574.
[53] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[54] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, M. Parrinello, PLUMED: a portable plugin for free-energy calculations with molecular dynamics, Comput. Phys. Commun. 180 (2009) 1961–1972.
[55] G.A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: new feathers for an old bird, Comput. Phys. Commun. 185 (2014) 604–613.
[56] R.B. Best, J. Mittal, Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse, J. Phys. Chem. B 114 (2010) 14916–14923.
[57] J.L. Abascal, C. Vega, A general purpose model for the condensed phases of water: TIP4P/2005, J. Chem. Phys. 123 (2005) 234505.
[58] G.H. Zerze, R.B. Best, J. Mittal, Modest influence of FRET chromophores on the properties of unfolded proteins, Biophys. J. 107 (2014) 1654–1660.

[59] G.H. Zerze, C.M. Miller, D. Granata, J. Mittal, Free energy surface of an intrinsically disordered protein: comparison between temperature replica exchange molecular dynamics and bias-exchange metadynamics, J. Chem. Theory Comput. 11 (2015) 2776–2782.
[60] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126 (2007) 014101.
[61] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190.
[62] S. Nosé, M. Klein, Constant pressure molecular dynamics for molecular systems, Mol. Phys. 50 (1983) 1055–1076.
[63] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[64] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N· log (N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[65] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[66] M.K. Prakash, A. Barducci, M. Parrinello, Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations, J. Chem. Theory Comput. 7 (2011) 2025–2027.
[67] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[68] X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, Peptide folding: when simulation meets experiment, Angew. Chem. Int. Ed. 38 (1999) 236–240.
[69] M. Karplus, Vicinal proton coupling in nuclear magnetic resonance, J. Am. Chem. Soc. 85 (1963) 2870–2871.
[70] A.C. Wang, A. Bax, Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations, J. Am. Chem. Soc. 118 (1996) 2483–2494.
[71] J.-S. Hu, A. Bax, Determination of ϕ and ξ1 angles in proteins from 13C-13C three-Bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? J. Am. Chem. Soc. 119 (1997) 6360–6368.
[72] B. Han, Y. Liu, S.W. Ginzinger, D.S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction, J. Biomol. NMR 50 (2011) 43–57.
[73] A. Grossfield, D.M. Zuckerman, Quantifying uncertainty and sampling quality in biomolecular simulations, Ann. Rep. Comput. Chem. 5 (2009) 23–48.
[74] C.L. McClendon, L. Hua, G. Barreiro, M.P. Jacobson, Comparing conformational ensembles using the Kullback–Leibler divergence expansion, J. Chem. Theory Comput. 8 (2012) 2115–2126.
[75] C.L. McClendon, G. Friedland, D.L. Mobley, H. Amirkhani, M.P. Jacobson, Quantifying correlations between allosteric sites in thermodynamic ensembles, J. Chem. Theory Comput. 5 (2009) 2486–2502.
[76] G. Bussi, F.L. Gervasio, A. Laio, M. Parrinello, Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics, J. Am. Chem. Soc. 128 (2006) 13435–13441.