JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

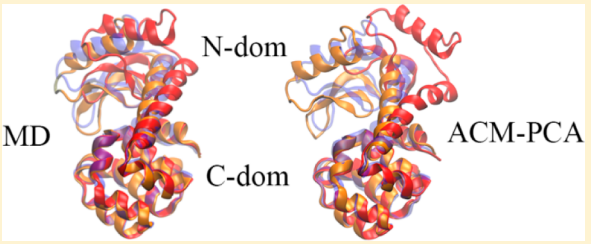
Simulating Large-Scale Conformational Changes of Proteins by Accelerating Collective Motions Obtained from Principal Component Analysis

Junhui Peng and Zhiyong Zhang*

Hefei National Laboratory for Physical Science at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

Supporting Information

ABSTRACT: Enhanced sampling methods remain of continuing interest over the past decades because they are able to explore conformational space of proteins much more extensively than conventional molecular dynamics (MD) simulations. In this paper, we report a new sampling method that utilizes a few collective modes obtained from principal component analysis (PCA) to guide the MD simulations. Two multidomain proteins, bacteriophage T4 lysozyme and human vinculin, are studied to test the method. By updating the PCA modes with a proper frequency, our method can sample large-amplitude conformational changes of the proteins much more efficiently than standard MD. Since those PCA modes are calculated from structural ensembles generated by all-atom simulations, the method may overcome an inherent limitation called "tip effect" that would possibly appear in those sampling techniques based on coarse-grained elastic network models. The algorithm proposed here is potentially very useful in developing tools for flexible fitting of protein structures integrating cryo-electron microscope or small-angle X-ray scattering data.



INTRODUCTION

Protein dynamics often plays an important role in its biological function.[1] For example, a multidomain protein consists of two or more well-folded domains connected by flexible linkers,[2,3] and those interdomain motions may facilitate conformational changes of the protein upon ligand binding.

Molecular dynamics (MD) simulation is one of the most popular computational techniques to investigate protein dynamics.[4−6] With more and more computer resources and advanced algorithms being available, MD simulation is becoming a powerful tool in the study of many biological processes.[7−9] However, MD simulations could be computationally very expensive that is an obstacle of their future applications. For a modest-size multidomain protein, a MD simulation at a time scale of microseconds (μs) is time-consuming, so one usually simulates only several hundreds of nanoseconds (ns). Since the time scale of large-amplitude domain motions is generally beyond ns, the limited-time MD simulation is not able to sample sufficient conformations and interpret protein motions properly.

Many efforts have been made to overcome the aforementioned limitation of MD, such as the enhanced sampling techniques.[10−13] The inefficient sampling in a MD simulation is due to not only the short simulation time[14] but also the frustrated shape of protein energy landscape.[15] In a MD simulation under physiological condition, the protein could be stuck in a few local states, whereas global conformational changes related to the protein's functional motions are rarely sampled. Many techniques have been developed to improve sampling efficiency in the MD simulations,[16−42] so the protein may jump out of the local minima in the energy landscape and transit globally between different conformational states.

A large number of sampling techniques[16,18,19,23,26,27,29−32,34,36,40−42] are based on collective coordinates of proteins determined by normal-mode analysis (NMA)[43−45] or principal component analysis (PCA).[46,47] Many studies have shown that functionally important motions of a protein occur along a few collective coordinates (also called collective modes), which have a dominant contribution to the protein dynamics.[48,49] This small number of collective modes can then serve as a basis set for enhanced sampling. In many NMA-based sampling methods,[26,27,34,40,42] a coarse-grained NMA called elastic network model (ENM[50]) is used because it can quickly compute collective modes from a single structure, which reflect the intrinsic motions of the protein.[51−53] However, Ma and co-workers have pointed out that ENM has an inherent limitation called "tip effect",[54] which may severely bias the sampling in those proteins containing tips. Although PCA does not suffer from the tip effect, there is remaining concern that the PCA modes calculated from a short simulation may not converge and have no predictive power for the protein dynamics at long time scale.[55,56] This issue needs to be addressed carefully in the PCA-based sampling methods.

In this paper, we present, to our knowledge, a new sampling technique that accelerates collective motions of a protein

Received: February 7, 2014

   dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

obtained from PCA (we denote it as ACM-PCA). Technically, a series of short simulations, each with several hundreds of ps, are carried out for the protein. After each run, PCA is performed on the trajectory, and we have found that these collective PCA modes can generally provide reliable directions of the domain motions. In the next simulation, motions along the collective modes are accelerated by coupling them to a thermostat with high temperature. Then, the above procedure repeats. That is to say, the PCA modes describing domain motions of the protein will be updated with a certain time interval during the ACM-PCA simulation, so the time evolution of the modes are likely captured, which makes ACM-PCA capable of simulating long time-scale and large length-scale conformational changes of the protein. Simulation results of two multidomain proteins, bacteriophage T4 lysozyme (T4L) and human vinculin, demonstrate that the sampling efficiency of the ACM-PCA method is significantly better than standard MD.

In the subsequent sections, after introducing PCA briefly, the ACM-PCA sampling method will be presented, followed by computational details of the simulations. The two multidomain proteins, T4L and vinculin, are used to illustrate the sampling efficiency of ACM-PCA compared to MD. After discussing key technical issues in detail, our method is compared with the ENM-based and other PCA-based sampling techniques. Concluding remarks are provided at the end.

THEORY AND METHODS

PCA. From an all-atom trajectory, we can construct an ensemble of coarse-grained (CG) protein structures. One CG site may represent an amino-acid residue or even a group of residues,[57] which locates at the position of the center-of-mass (COM). This CG procedure would significantly reduce the computational cost of PCA while the global collective modes are still preserved, since they are generally not related to the local structures of the protein. A standard PCA consists of the following steps: (1) All the CG conformations of the protein are superimposed to a reference structure, in order to remove overall translational and rotational motions of the system. (2) With the selected number of CG sites N, a covariance matrix of positional fluctuation is constructed. The elements of the matrix are $\langle (R_{I_h} - \langle R_{I_h} \rangle)(R_{J_k} - \langle R_{J_k} \rangle) \rangle$, where R_{I_h} is the Cartesian coordinate of h component of CG site I, and $\langle R_{I_h} \rangle$ means the ensemble average. (3) The covariance matrix is diagonalized to obtain eigenvectors Ψ_q (q is from 1 to $3N$) and eigenvalues λ_q (mean square fluctuations of the corresponding modes). Only a few eigenvectors (also called PCA modes) with the largest eigenvalues have a major contribution to collective motions of the protein (termed as essential modes), which will be used to guide the MD simulation.

ACM-PCA. According to atomic velocities $\mathbf{v}(t)$ at the simulation time t, we can compute the velocities of the CG sites $\mathbf{V}(t)$. If the first n_{ED} PCA modes with the largest eigenvalues are selected as the essential modes, the projection of $\mathbf{V}(t)$ onto these modes is given as

$$\mathbf{V}^{\mathrm{ED}}(t) = \sum_{q=1}^{n_{\mathrm{ED}}} (\mathbf{V}(t)\Psi_q)\Psi_q \tag{1}$$

In a constant-temperature MD simulation, a scaling factor is used to scale the atomic velocities in order to make the system temperature toward its target value T_0, by using algorithms such as the weak-coupling method,[58]

$$s = \left(1 + \frac{dt}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)\right)^{1/2} \tag{2}$$

where dt is the MD time step, τ_T is the temperature relaxation time, and $T(t)$ is the instantaneous temperature.

During an ACM-PCA simulation, in order to accelerate collective motions of the protein, the velocities along the essential PCA modes (eq 1) are coupled to a high temperature T_0^h (the corresponding scaling factor is s_h) while the remaining velocities are coupled to the ambient temperature T_0^n (the scaling factor is s_n). Therefore, the updated velocities are the combination of the above two components after different temperature scaling,

$$s_h\mathbf{V}^{\mathrm{ED}}(t) + s_n(\mathbf{v}(t) - \mathbf{V}^{\mathrm{ED}}(t)) \tag{3}$$

With the above strategy, the global conformational changes may be sampled extensively since the protein has a high kinetic energy to cross energy barriers and reach different conformational states. In the meanwhile, local structures of the protein remain stable because all of the high-frequency motions are still kept at the room temperature.

Work Flow of the ACM-PCA Method. Algorithmically, the ACM-PCA approach consists of the following steps:

(1) A short conventional MD simulation of the protein, such as 200 ps, is run first.
(2) A CG trajectory of the protein is constructed from the all-atom trajectory according to a predefined CG map, and PCA is then performed. Note that the reference structure for PCA should be the last frame of the trajectory that is also the starting structure of the next run.
(3) A short accelerated simulation, such as 200 ps, is carried out that is guided by the essential PCA modes generated from its preceding simulation.
(4) Update the PCA modes using the new trajectory, as in the step 2.
(5) Repeat the steps 3 and 4.

In summary, we perform a series of short simulations. PCA modes computed from one simulation are used to guide its next simulation. Therefore, only the first simulation is the conventional MD, and the following are all accelerated simulations.

COMPUTATIONAL DETAILS

T4L. T4L is a protein with 164 amino-acid residues, consisting of two domains connected by a long α-helix (Figure 1a). There is a deep cleft between its N-terminal and C-terminal domains, forming the active site for oligosaccharide binding. Many experimental structures of T4L and its variants have been determined, suggesting a hinge-bending domain motion that causes opening/closing of the active site.[59]

MD Simulation. A closed structure of T4L (PDB entry 2LZM)[60] with a resolution of 1.7 Å was taken to start a MD simulation, using the GROMACS-4.5.5 package[61] and the CHARMM27 force field.[62] The periodic boundary condition (PBC) with a cubic box type was used. The minimum distance between the solute and the box boundary was 1.2 nm. The box was filled with TIP3P water molecules.[63] The system with the protein and waters was energy-minimized by the steepest

dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

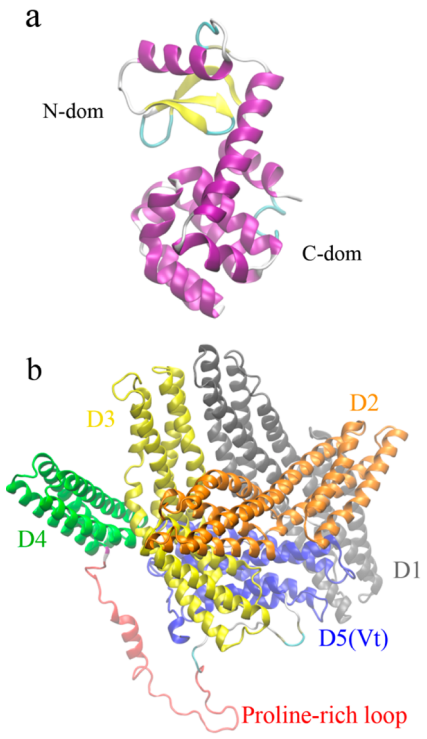


Figure 1. Simulation systems. (a) Bacteriophage T4 lysozyme (T4L), the N-terminal and C-terminal domains are labeled. (b) The full-length human vinculin, domains D1−D5 and the proline-rich region (PR loop) are labeled. D1−D4 are the vinculin head (Vh), and D5 is the tail (Vt). The missed PR loop in the crystal structure is built by homology modeling. All the protein structures in this figure and Figures 4 and 7 were created using VMD.[92]

descent method, until the maximum force was smaller than 1000 kJ mol^{-1} nm^{-1}. Eight Cl^- ions were added to compensate the net positive charges on the protein by replacing the same number of water molecules with the most favorable electrostatic potential. The final system was energy-minimized again using the steepest descent and then the conjugate gradient method, until the maximum force was smaller than 100 kJ mol^{-1} nm^{-1}. The simulation was conducted by using the leapfrog algorithm[64] with a 2 fs time-step. A 100 ps equilibration simulation with positional restraint was carried out, using a force constant of 1000 kJ mol^{-1} nm^{-2}. The initial atomic velocities were generated according to a Maxwell distribution at 300 K. The following production run was 200 ns.

The simulation was performed under the constant NPT condition. The three groups (protein, solvent, and ions) were coupled separately to a temperature bath of 300 K by using an velocity rescaling thermostat,[65] with a relaxation time of 0.1 ps. The pressure was kept at 1 bar with a relaxation time of 0.5 ps and the compressibility of 4.5×10^{-5} bar^{-1}. Covalent bonds in the protein were constrained using the P-LINCS algorithm.[66] Twin-range cutoff distances for the van der Waals interactions were chosen to be 0.9 and 1.4 nm, respectively, and the neighbor list was updated every 20 fs. The long-range electrostatic interactions were treated by the PME algorithm,[67] with a tolerance of 10^{-5} and an interpolation order of 4.

ACM-PCA. We have implemented the ACM-PCA method in the GROMACS-4.5.5 package. The parameters were largely the same as those in the standard MD simulation, except that collective motions of the protein defined by PCA were accelerated by coupling them to a high temperature bath. In PCA, each CG site represented the COM of a residue and all the 164 residues were used. The velocities projected onto the three largest-amplitude PCA modes were coupled to a high temperature of 800 K, while those along the other degrees of freedom of the protein were coupled to 300 K as in the MD simulation, by modifying the weak-coupling method.[58] The relaxation times were both set as 0.1 ps. The ACM-PCA simulation was comprised of 100 short simulations, each with 200 ps long, so the total simulation time was 20 ns.

Vinculin. Vinculin plays an important role in the maintenance and regulation of cell adhesion and migration.[68] The crystal structure of human vinculin has been determined with a resolution of 2.85 Å (PDB entry 1TR2), which adapts an autoinhibited conformation.[69] The full-length vinculin (1066 residues) consists of eight four-helix bundles, which are further organized into five domains (D1-D5, Figure 1b). D1 to D4 constitute the vinculin head (Vh) and D5 is the vinculin tail (Vt). D1 and D3 form a pair of pincers that hold Vt in its autoinhibited state. D4 and Vt are connected by a proline-rich region (PR loop) that is disordered and largely missing in the crystal structure.

MD Simulation. The crystallographic unit cell contains two vinculin monomers, and we took the chain A in PDB for the MD simulation. The missing proline-rich loop (residues 844−876) was predicted by homology modeling using SWISS-MODEL,[70] and those missing residues 1062−1066 in the C-terminus were placed with SwissPDBViewer.[71] The MD parameters were the same as those in the T4L simulation except for the following. The PBC with a dodecahedron box type was used. Eleven Na^+ ions were added to compensate the net negative charges on the protein. The final system was minimized until the maximum force was smaller than 200 kJ mol^{-1} nm^{-1}. The initial atomic velocities were generated according to a Maxwell distribution at 310 K. The production MD run was 100 ns. The three groups (protein, solvent, and ions) were coupled separately to a temperature bath of 310 K.

ACM-PCA. Parameters in the ACM-PCA simulation of the vinculin were essentially the same as those of T4L, except that the velocities projected onto the three largest-amplitude PCA modes were coupled to a high temperature of 1000 K, while those along the other degrees of freedom of the protein were coupled to 310 K. The simulation of the vinculin was comprised of 50 short simulations, each with 200 ps long, so the total simulation time was 10 ns.

RESULTS AND DISCUSSION

Comparison between ACM-PCA and MD. *T4L. Structure Properties.* For each residue of T4L, root-mean-square fluctuation (RMSF) of its COM was calculated from the different simulations, respectively. The results clearly indicate that the 20 ns ACM-PCA simulation (Figure 2, red solid line) has significantly larger RMSF values than the 200 ns MD simulation (Figure 2, black solid line). It should be noted that the two RMSF curves share a similar pattern with a high correlation coefficient of 0.82, which suggests that the large fluctuation observed in ACM-PCA should be reasonable.

Root mean square deviations (RMSD) of all the frames in a trajectory to its starting structure can indicate conformational changes of the protein. For the two simulations of T4L (MD and ACM-PCA), RMSD values were calculated, respectively, using the C_α atoms of residues 1−162. During the MD simulation, the protein remains in the closed state with most RMSD values around 1.0 Å (Figure 3a, black solid line). In the

dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

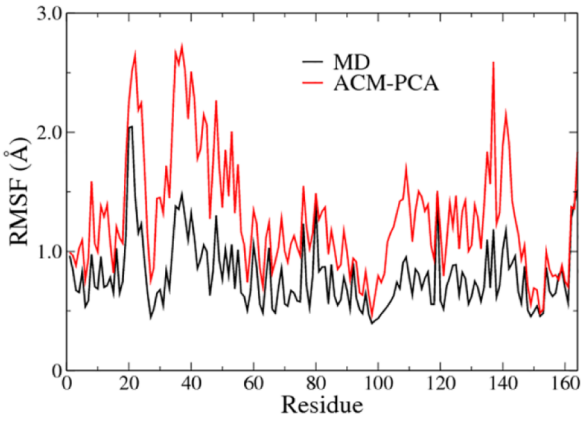


Figure 2. Root mean square fluctuations (RMSF) of the center-of-mass (COM) of residues in T4L, which were calculated from the MD (black solid line), and the ACM-PCA (red solid line) simulation, respectively.

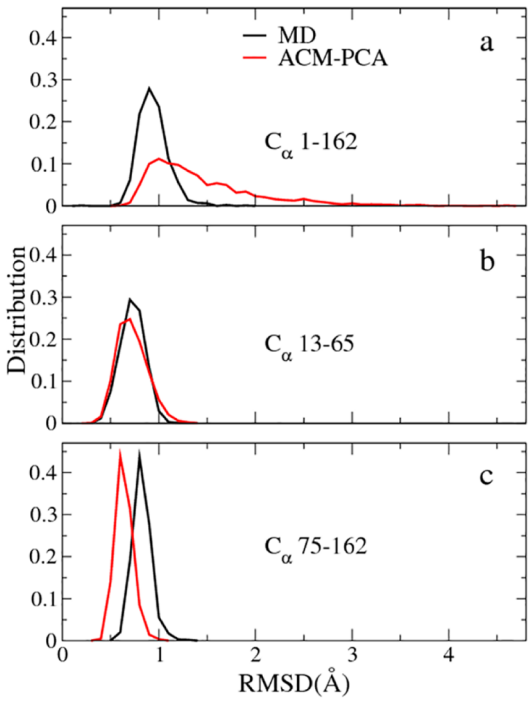


Figure 3. Distributions of root-mean-square deviations (RMSD) in the MD (black solid line) and the ACM-PCA (red solid line) simulation, respectively. The RMSD values to the starting structure were calculated using the C_α atoms of (a) residues 1−162, (b) residues 13−65 (the N-terminal domain), and (c) residues 75−162 (the C-terminal domain), respectively.

ACM-PCA simulation, the peak of RMSD distribution is also around 1.0 Å, but it does reach a much broader RMSD range than the MD simulation (Figure 3a, red solid line). There are clearly much larger conformational changes occurred in ACM-PCA than in MD. RMSD distribution of the N-terminal (C_α atoms of residues 13−65, Figure 3b) and the C-terminal domain (C_α atoms of residues 75−162, Figure 3c) was also computed, respectively. Contrary to the case of the whole protein (Figure 3a), the intradomain RMSD values in the ACM-PCA simulation are mostly smaller than 1.0 Å (Figure 3b and c, red solid line), which are essentially in the same range as those in the MD simulation (Figure 3b and c, black solid line). The above results strongly suggest that the large-amplitude conformational changes of T4L in the ACM-PCA simulation are ascribed to the interdomain motions while the internal structure of either domain remains intact.

The contents of major secondary structures in T4L were analyzed by DSSP.[72] In the MD ensemble, about 64% of the total 164 residues belong to α-helices and 8% of the residues belong to β-sheets. The corresponding percentages in the ACM-PCA ensemble are 64% and 7%, respectively. The results indicate that the secondary structures in the ACM-PCA simulation are as stable as those in MD, which are consistent to the intradomain RMSD values (Figure 3b and c).

Conformational Sampling of the Closed and Open States. Projecting a trajectory onto the subspace spanned by the essential PCA modes is a good way to visualize the conformational space sampled by the simulation. From various experimental structures of T4L and its mutants, 38 structures were selected that may represent possible conformations of the protein.[59,73,74] PCA was performed on this ensemble of structures using the C_α atoms of residues 1−162. The first mode describes an open/close motion between the N- and C-terminal domains with a dominant contribution (84%) to the total fluctuation in the protein, whereas the second mode represents a twist motion between the domains with a minor contribution of 6%. We projected the 38 experimental structures of T4L onto the 2D subspace defined by the above two PCA modes. It is clearly seen that there are two distinct clusters of structures along the open/close mode (Figure 4a, blue). The cluster on the right includes closed structures and the other one on the left consists of open structures. The trajectories of the MD and ACM-PCA simulations were all projected onto the 2D subspace. The MD simulation only samples a limited region on the right side of the plane (Figure 4a, black), which covers the cluster of closed structures but does not reach the cluster of open structures at all. On the contrary, the ACM-PCA simulation (Figure 4a, red) explores much larger areas in the subspace than MD, which almost covers the two clusters of experimental structures.

The ACM-PCA simulation of T4L consists of 100 continuous short runs, and their initial structures are connected sequentially by green lines in Figure 4a. Instead of expanding gradually from the closed to the open state in a targeted way, it appears that the protein samples in the essential subspace randomly and transits between the two states back and forth. The RMSD curve of the ACM-PCA simulation also supports this notion (Figure 4b). The results suggest that, instead of appearing gradually, those collective modes of T4L may present from the beginning of the simulation. To confirm this, we have checked the largest-amplitude PCA mode obtained from several short simulations, and they all nicely describe the hinge-bending domain motion of T4L (Supporting Information, Figure S1). It should be noted that this may not always be the case in other proteins, especially for very large conformational changes, in which some PCA modes might appear gradually during the simulation.

Free Energy Calculations. We used a semiempirical method called MM/GBSA (molecular mechanics generalized Born surface area)[75,76] to estimate absolute free energies of the conformations in different simulations. The theory and method of MM/GBSA were introduced in the Supporting Information. The free energies of the MD and ACM-PCA trajectories are both shown in Supporting Information, Figure S2. It appears that the free energies of ACM-PCA (red solid line) are a little

dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

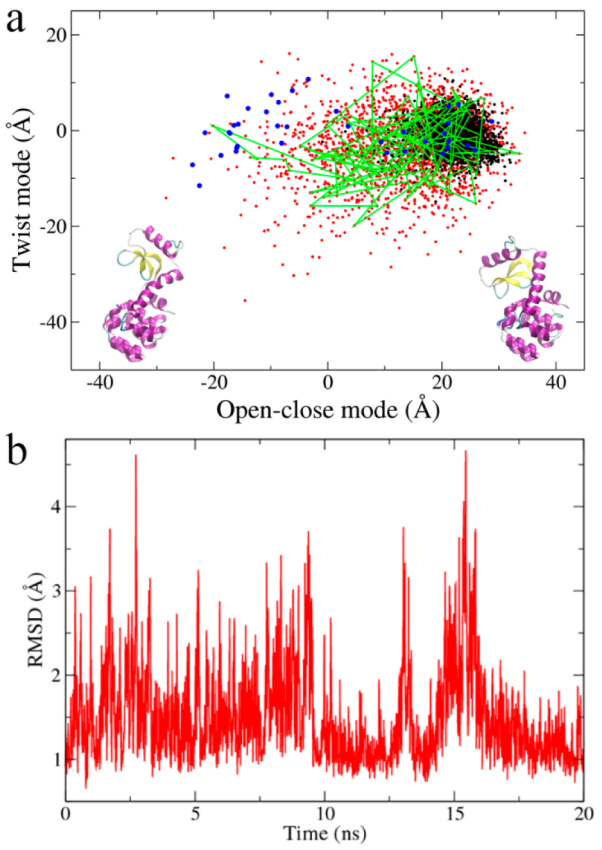


Figure 4. (a) Projections of T4L structures onto the essential subspace defined by the open-close and the twist modes. Each point on the plane represents a protein conformation. The experimental structures of T4L colored by blue form two clusters along the open-close mode, which are the closed state on the right and the open state on the left. The projections of MD (black) and ACM-PCA (red) indicate their sampled areas in the essential subspace, respectively. The starting structures of these short simulations in ACM-PCA are connected sequentially by green lines, in order to show the time evolution of conformational sampling. The inset shows the most closed and the most open conformations of T4L simulated by ACM-PCA. (b) The time evolution of RMSD in the ACM-PCA simulation, with respect to the starting structure. The RMSD values were calculated using the C_α atoms of residues 1−162.

larger than those of MD (black solid line), with the average free energy of the MD ensemble −22280 ± 148 kJ/mol and the value of the ACM-PCA ensemble −22200 ± 138 kJ/mol. These results are reasonable because the MD ensemble only contains the favorable closed conformations of the wild-type T4L whereas the ACM-PCA ensemble includes not only the closed state but also the open and intermediate states of the protein. However, it should be noted that the free energy differences between MD and ACM-PCA are marginal compared to the fluctuations of the free energies. In a recent work done by Harada and Kitao,[77] a small free energy difference of 0.8 k_BT (2.0 kJ/mol) between the open and closed state of T4L, and an energy barrier height 2.5 k_BT (6.2 kJ/mol) of the transition state with respect to the closed state, were obtained. Due to the large fluctuations of the free energies computed by MM/GBSA, one could hardly detect such minor changes along the free energy profiles (Supporting Information, Figure S2) although our results are qualitatively in agreement with theirs. Despite the above limitation, the MM/GBSA calculations do indicate that the conformations sampled by ACM-PCA have fairly reasonable free energies.

Taking the above comparison between ACM-PCA and MD together, the sampling in the ACM-PCA simulation is much more efficient than that in the standard MD simulation. Although the closed conformations of T4L are still the major state in the ACM-PCA simulation, the protein may transit between the closed and open states while the MD simulation cannot (Figure 4). The large conformational space explored by ACM-PCA comes from the fact that those PCA modes related to collective domain motions are coupled to the high temperature, which drives the system to cross the energy barrier between different states. The T4L conformations sampled by the ACM-PCA simulation should be physically reliable because (1) the intradomain structures (secondary structure elements) are well preserved as those in MD (Figure 3b and c), (2) they are within (or close to) the region defined by the 38 experimental structures (Figure 4a), and (3) their free energies are reasonable compared to those of the MD ensemble (Supporting Information, Figure S2). These results strongly support that the ACM-PCA method not only allows for adequate sampling of the conformational space but also restricts the protein conformations within reasonably low energy regions. In the ACM-PCA simulation, only very few degrees of freedom (PCA modes) are coupled to the high temperature while the dominant number of them are coupled to the room temperature, which makes ACM-PCA distinguish from a MD simulation under the high temperature. In the latter, the protein mainly samples the conformational space with high energies and its local structures would be broken.

Vinculin. Human vinculin is a multidomain protein that is 6−7 times larger than T4L, and it would be very time-consuming to simulate its domain motions by conventional MD simulations. As such, this study demonstrates how the ACM-PCA method can be applied to such a large system.

For each residue of the vinculin, RMSF of its COM was calculated from the MD and the ACM-PCA simulation, respectively. Similar to the case of T4L, the fluctuation of the vinculin in the 10 ns ACM-PCA simulation (Figure 5, red solid line) is larger than that in the 100 ns MD simulation (Figure 5, black solid line). In the mean time, the two RMSF curves have a highly similar pattern with the correlation coefficient of 0.91.

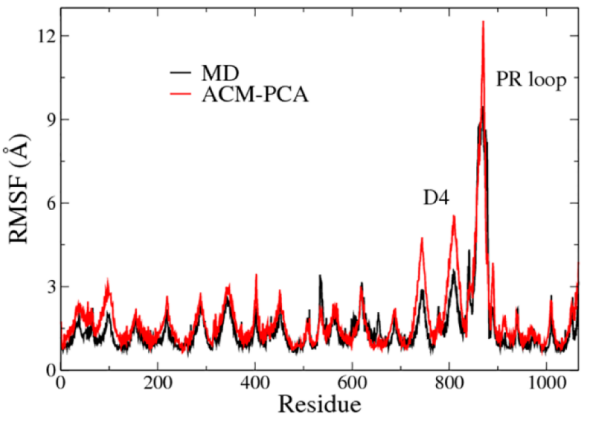


Figure 5. RMSF of the COM of residues in the vinculin, which were calculated from the MD (black solid line), and the ACM-PCA (red solid line) simulation, respectively. The PR loop and D4 are labeled.

dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The autoinhibited state of vinculin (Figure 1b) is a rather compact and stable structure, except for the PR loop. During the MD simulation, almost all the RMSD values of the C_α atoms in Vh+Vt are smaller than 3.0 Å (Figure 6a, black solid line), which support the above notion. However, in the ACM-PCA simulation (Figure 6a, red solid line), the conformations with large RMSD values are significantly more than those in the MD simulation. We then investigate the interdomain motions of Vh that is comprised by D1-D4. D1 and D3 form the pair of pincers that hold Vt tightly. D2, which consists of two four-helix bundles, sits on D1 and D3. The N-terminal bundle packs together with D3 with extensive contacts but the C-terminal bundle is less contacted. D4 is more or less isolated in Vh, suggesting that the major collective domain motions may occur between D4 and other domains.[78] RMSD distributions of the C_α atoms in D1+D3 are shown in Figure 6b, which indicate that structure of the pincers is very stable in both the simulations. The RMSD values of ACM-PCA are even a little smaller than those of MD because the simulation time of the former is only one-tenth of the latter. Figure 6c shows the RMSD distributions of D4 after superimposing all the structures by D1+D3, so these values indicate the relative movement between D4 and D1+D3. In the MD simulation, the RMSD values of D4 are centered on 2.5 Å (Figure 6c, black solid line). Few conformations have RMSD larger than 5 Å, and the largest one is 5.4 Å (Figure 7a). However, in the ACM-PCA simulation, quite some conformations have RMSD of D4 larger than 5.0 Å (Figure 6c, red solid line), and the largest one is 14.7 Å (Figure 7b), which are in agreement with the RMSF values of D4 (Figure 5, red solid line). The relative movement between D2 and D1+D3 was also measured. The RMSD values of D2 in ACM-PCA (Figure 6d, red solid line) are somewhat larger than those in MD (Figure 6d, black solid line), which show agreement with the contact analysis according to the crystal structure of vinculin. The results of RMSD distributions indicate that the ACM-PCA simulation does explore more domain motions (Figure 7b) than the MD simulation (Figure 7a), which are consistent to the interdomain movements proposed by other literature.[78]

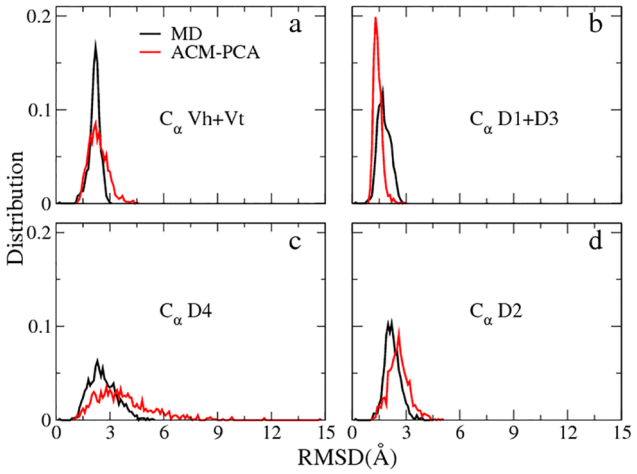


Figure 6. Distributions of RMSD in the MD (black solid line) and the ACM-PCA (red solid line) simulation, respectively. The RMSD values were calculated using the C_α atoms of (a) Vh+Vt, (b) D1+D3, (c) D4, and (d) D2, respectively. All the structures were superimposed to the starting structure using C_α atoms of D1+D3.

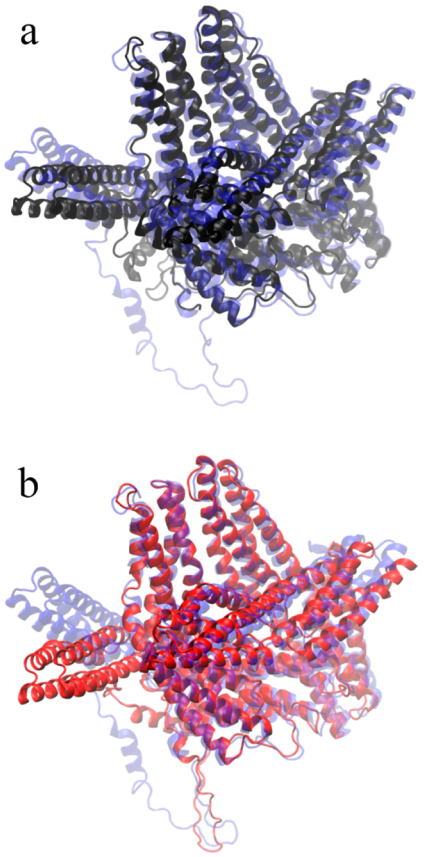


Figure 7. Domain motions of the vinculin in the (a) MD and (b) ACM-PCA simulation, respectively. In each simulation, the starting structure is colored by blue and the structure with the largest RMSD of D4 to D1+D3 is colored by black in MD, and red in ACM-PCA, respectively.

Technical Details of ACM-PCA. In the ACM-PCA method, several key issues need to be considered, such as the length of the simulation segment, the scheme of structure superposition, the number of PCA modes to be accelerated, and the high temperature coupled to the PCA modes. In this section, we discuss these issues by running various simulations of T4L.

Length of the Simulation Segment. The importance of this parameter comes from the fact that it determines the length of the trajectory used to perform PCA and the frequency of updating the PCA modes. It was said that the PCA modes calculated from a short MD trajectory may have a problem of convergence.[55,56] Amadei et al. have found that a simulation with the length of several hundreds of ps may provide a fairly reliable set of the essential PCA modes, at least within the time scale of ns.[79] In our initial test of ACM-PCA, we chose a segment length of 200 ps. The simulation results of T4L show significantly enhanced sampling of the conformational space (Figure 4). We have also tried other segment lengths, such as 20, 100, and 300 ps, respectively. The results of the ACM-PCA simulation using a 20 ps segment (Supporting Information, Figure S3a) indicate that the PCA modes from a too short trajectory may be noisy and lead to inefficient sampling in the essential subspace. Extensive domain motions of T4L are observed in these ACM-PCA simulations using the segment lengths ranging from 100 to 300 ps (Supporting Information, Figure S3b to d), which support the notion of Amadei et al.[79]

dx.doi.org/10.1021/ct5000988 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

One may notice that the ACM-PCA simulation using the 300 ps segment (Supporting Information, Figure S3d) samples a little smaller area in the essential subspace than that using the 200 ps segment (Supporting Information, Figure S3c), which is not strange since a longer segment length in ACM-PCA does not necessarily result in a broader sampling. Both the 200 and 300 ps trajectory can yield reliable PCA modes describing domain motion of T4L, so the small difference between the two ACM-PCA simulations may be ascribed to the randomness of the sampling. We have also found that a too long simulation segment (the extreme case would be just one single simulation) is not appropriate because the PCA modes need to be updated with a proper frequency during the simulation. This issue will be further discussed in another section when we compare ACM-PCA with other PCA-based sampling methods.

Scheme of Structure Superposition. Before PCA, one needs to superimpose all the conformations in the trajectory to a reference structure, in order to remove the translational and rotational motions of the system. There are several algorithms available,[80,81] and one may choose any part in the protein for structure superposition. In the ACM-PCA simulation of T4L, we performed the conventional least-squares superposition by using the COMs of all the 164 residues. The calculated PCA modes clearly demonstrate the collective domain motions (Supporting Information, Figure S1), which may explain why the ACM-PCA simulation can accelerate conformational changes between the closed and open states of the protein. We also tried superimposing all the conformations using the C-terminal domain (COMs of the residues 75−162) only, but performing PCA on the whole protein. The PCA modes do not describe the domain motions quite well because the C-terminal domain is almost static although the N-terminal domain has large movement (Supporting Information, Figure S4a). Therefore, the resulting ACM-PCA simulation is inefficient in sampling the essential subspace (Supporting Information, Figure S4b). Depending on what kind of motion to be accelerated, one has to choose the appropriate scheme of structure superposition.

Number of the PCA Modes to Be Accelerated. In the initial ACM-PCA simulation of T4L, the first three PCA modes were coupled to 800 K. We have also chosen the first two, four, and six PCA modes, and run ACM-PCA simulations, respectively. Generally speaking, larger conformational changes would be observed when more PCA modes are coupled to the high temperature (Supporting Information, Figure S5a). In order not to destroy local structures of the protein in ACM-PCA, one should only pick the first few PCA modes that describe global domain motions and exclude any modes reflecting local motions. On the other hand, these global modes should have a major contribution to the total fluctuation of the protein to ensure an adequate sampling in the essential subspace. Although PCA on the 38 experimental structures yields only two major modes (open-close and twist), the PCA modes from the simulated trajectories show some differences (Supporting Information, Figure S5b). Since more conformations, such as the intermediate states of T4L, are sampled in the simulation than the ensemble of experimental structures (Figure 4), some other collective modes may appear. We have found that coupling the first three PCA modes at the high temperature is a good choice, which contribute about 60%−80% of the fluctuation of T4L (Supporting Information, Figure S5b). A technical reason prevents us from using only one PCA mode for accelerated sampling because the temperature of the single mode may fluctuate wildly. Therefore, at some time steps, the temperature of this mode would be extremely high that may distort the protein structure, as we will discuss below.

High Temperature Coupled to the PCA Modes. Besides 800 K, we have tried other different temperatures to couple the first three essential PCA modes, and their sampling efficiency are shown in Supporting Information, Figure S6. Generally speaking, a higher temperature leads to a larger sampling area in the essential subspace. If the temperature is not high enough, such as 500 K, the open state of T4L would not be observed in the ACM-PCA simulation (Supporting Information, Figure S6a), such as the case in MD (Figure 4a, black). It should be noted that a too-high temperature would eventually distort the local structure of the protein since a leakage of energy may happen from the high-temperature degrees of freedom into the room-temperature ones.

Comparison with an ENM-Based Sampling Method. Besides PCA, collective modes calculated by NMA (or ENM) have also been used for enhanced sampling of protein's conformational space.[16,26,27,34,40,42] Typically ENM is an one site per residue CG model built from a single atomic structure of the protein,[50] and each residue is represented by its C_α atom or COM. Any pair of CG sites within a given cutoff distance is connected by a spring. With this simplified model, collective modes of the protein can be calculated rapidly. We have previously developed a sampling method that uses low-frequency ENM modes to accelerate collective motions of the protein.[27] In this section, we compare the sampling results between the ACM-PCA and this ENM-based method (denoted as ACM-ENM here).

For each simulation segment of T4L, we computed the root-mean-square inner products (RMSIP, eq S4 in Supporting Information) between the PCA modes from the short trajectory and the ENM modes from the final structure of the trajectory. The RMSIP values show that these PCA modes are rather similar to the ENM modes (Supporting Information, Figure S7a), which are in agreement with the results from Jernigan and co-workers.[82] Therefore, it is not surprising that both the ACM-PCA and the ACM-ENM simulation achieve an efficient sampling in the essential subspace (Supporting Information, Figure S7b). During the ACM-ENM simulation, about 64% and 8% of the residues belong to α-helices and β-sheets, respectively. The average free energy calculated by MM/GBSA is -22196 ± 140 kJ/mol. The results are quite close to those from the ACM-PCA simulation.

However, in the ACM-ENM simulation of vinculin, we encountered the "tip effect".[54] A tip in the protein, such as an isolated surface loop, has much less neighboring residues than other regions. In ENM, the tip region is loosely connected by only a few harmonic springs, which may result in an imbalance of elastic forces around the tip. Therefore, in some collective modes, the magnitudes of displacement in the tip region are abnormally large, which make the rest of the protein "static" since each mode is normalized. We have computed the RMSF of the COM of residues using the first three ENM modes (Eqn. S5 in the Supporting Information), and these values of the residues within the PR loop are abnormally much larger than those in the other regions of the protein (Supporting Information, Figure S8, black dash line), which indicate the existence of the tip effect.

In the ACM-ENM simulation, the PR loop fluctuates much more violently (Supporting Information, Figure S8, green solid line) than that in the MD simulation (Supporting Information,

Figure S8, black solid line). The RMSF values of the PR loop from the ACM-PCA simulation (Supporting Information, Figure S8, red solid line) are also larger than those in MD, but they are only about one-half of those in ACM-ENM. On the contrary, the domain movements in the ACM-PCA simulation are obviously larger than those in the ACM-ENM simulation (Supporting Information, Figure S9). The results indicate that the tip effect of the ENM modes have been brought into the ACM-ENM simulation, which may result in an oversampling to the tip while leaving the other regions less sampled. In order to tackle this problem, one could implement a modified ENM using a more complicated potential energy function[54] than the simple one used in the conventional ENM. As an alternative, ACM-PCA provides one of the solutions to eliminate the tip effect since the PCA modes are calculated from trajectories generated by all-atom force field. Therefore, the ACM-PCA methods may be applied to more proteins than the conventional ENM-based sampling method.

Comparison with Other PCA-Based Sampling Methods. There are quite some PCA-based sampling methods available,[18,19,23,29−32,36,41] which are mainly different at how to accelerate the collective motions defined by PCA. Many of these methods calculate the PCA modes only once and no longer update them during the simulation. In this work, we argue that it is important to update the PCA modes with a proper frequency due to the following reason. The PCA modes obtained from a short simulation may not be able to predict long time protein dynamics, especially when very large conformational changes occur. Therefore, it would be appropriate to redo PCA on the new generated trajectories using the current conformation of the protein as the new reference, in order to capture the time evolution of the PCA modes. As a test, we performed PCA on the 200 ns MD trajectory and then used the first three PCA modes to run a 20 ns ACM-PCA simulation without further updating. The conformational space sampled by this simulation is only a little larger than that by standard MD, and the open state of T4L is not observed (Supporting Information, Figure S10). It seems that updating the PCA modes on the fly, with a proper time interval, is really necessary for sampling the large-scale conformational changes efficiently.

CONCLUSION

This paper presents, to our knowledge, a new sampling technique that utilizes the PCA modes to accelerate conformational changes of proteins. The PCA-based method is superior to those ENM-based sampling methods when the tip effect exists in the protein. We have also addressed some important issues, such as determining the proper frequency of updating the PCA modes, which make our method an efficient tool in the study of large scale conformational changes of proteins.

It has been recognized that a protein's intrinsic dynamics plays an important role in its functional motions.[1,83] A large number of experimental structures of T4L and its mutants are available, suggesting that hinge-bending motions may be an intrinsic property of the enzyme, which would promote the substrates to enter and the products to leave the active site.[59,74] Site-directed spin labeling studies of T4L in solution also supported the notion.[73] Structural information on the vinculin is much less than that of T4L. However, comparison between the vinculin and α-catenin[84] has suggested that intrinsic domain motions of D4 is possible, which might facilitate the active of the vinculin upon binding of various ligands.[78] Although domain motions of the wild-type T4L and autoinhibited vinculin were observed in the MD simulations, they only occurred in small amplitude due to the short simulation times and the sampling problem. The ACM-PCA simulation would be well suited for investigating intrinsic dynamics of proteins under the physiological condition, since it not only increases the amplitude of conformational changes between different states but also restricts the sampling mainly within the regions containing fairly low-energy conformations.

Compared to standard MD, the extra computational cost of ACM-PCA are mainly from the calculations of PCA modes. For a modest-size protein such as T4L, PCA on the CG trajectories at the residue level can be done rather quickly, and the time cost is minor compared to the atomistic simulations. For a large-size protein such as vinculin or even larger systems, one can always reduce the CPU time by using a CG map with a resolution lower than one site per residue, as long as these collective motions to be accelerated are still preserved.[57]

It should be noted that the ACM-PCA method is a nonequilibrium simulation technique since the atomic velocities are decomposed into two parts and directly coupled to different temperature baths, respectively. Therefore, the ensemble sampled by the ACM-PCA simulation will deviate from the canonical Boltzmann distribution. This issue has recently been addressed by Liu and co-workers,[40] with the purpose of not only exploring the protein conformations extensively but also obtaining correct conformational distributions. Nevertheless, a potential use of the current ACM-PCA method is to develop a tool for flexible fitting of the protein structure into its target density map from electron microscope.[85−87] Another application is to combine the enhanced sampling technique and small-angle X-ray scattering (SAXS) data, to detect conformational states of proteins in solution.[88−91]

ASSOCIATED CONTENT

Supporting Information

Details of calculating the MM/GBSA free energy, the root-mean-square inner product between two sets of eigenvectors, and the root-mean-square fluctuation of residues from the ENM modes. Figures S1 to S10. This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*Tel: +86-551-63600854. Fax: +86-551-63600374. Email: zzyzhang@ustc.edu.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (grant 31270760), National Key Basic Research Program of China (grant 2013CB910203), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDB08030102), Anhui Natural Science Foundation (grant 1208085MC38), and the Fundamental Research Funds for the Central Universities (WK2070000017, WK2070000020).

REFERENCES

(1) Kern, D.; Zuiderweg, E. R. P. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748−757.

(2) Ekman, D.; Bjorklund, A. K.; Frey-Skött, J.; Elofsson, A. *J. Mol. Biol.* **2005**, *348*, 231−243.
(3) Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11079−11084.
(4) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(5) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589−1615.
(6) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H. F.; Shaw, D. E. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.
(7) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. B.; Wriggers, W. *Science* **2010**, *330*, 341−346.
(8) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(9) Arkhipov, A.; Shan, Y.; Das, R.; Endres, N. F.; Eastwood, M. P.; Wemmer, D. E.; Kuriyan, J.; Shaw, D. E. *Cell* **2013**, *152*, 557−569.
(10) Lei, H. X.; Duan, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 187−191.
(11) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134−139.
(12) Zuckerman, D. M. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(13) Mitsutake, A.; Mori, Y.; Okamoto, Y. *Methods Mol. Biol.* **2013**, *924*, 153−95.
(14) Clarage, J. B.; Romo, T.; Andrews, B. K.; Pettitt, B. M.; Phillips, G. N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3288−3292.
(15) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(16) Noguti, T.; Go, N. *Biopolymers* **1985**, *24*, 527−546.
(17) Schlitter, J.; Engels, M.; Krüger, P. *J. Mol. Graphics* **1994**, *12*, 84−89.
(18) Grubmüller, H. *Phys. Rev. E* **1995**, *52*, 2893−2906.
(19) Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Aalten, D. M. F.; Berendsen, H. J. C. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615−625.
(20) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817−824.
(21) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.
(22) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(23) Abseher, R.; Nilges, M. *Proteins* **2000**, *39*, 82−88.
(24) Jensen, M. Ø.; Park, S.; Tajkhorshid, E.; Schulten, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6731−6736.
(25) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(26) He, J. B.; Zhang, Z. Y.; Shi, Y. Y.; Liu, H. Y. *J. Chem. Phys.* **2003**, *119*, 4005−4017.
(27) Zhang, Z. Y.; Shi, Y. Y.; Liu, H. Y. *Biophys. J.* **2003**, *84*, 3583−3593.
(28) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(29) Chen, C. J.; Xiao, Y.; Zhang, L. S. *Biophys. J.* **2005**, *88*, 3276−3285.
(30) Lange, O. E.; Schafer, L. V.; Grubmüller, H. *J. Comput. Chem.* **2006**, *27*, 1693−1702.
(31) Kubitzki, M. B.; de Groot, B. L. *Biophys. J.* **2007**, *92*, 4262−4270.
(32) Spiwok, V.; Lipovova, P.; Kralova, B. *J. Phys. Chem. B* **2007**, *111*, 3073−3076.
(33) Gao, Y. Q. *J. Chem. Phys.* 2008, *128*.
(34) Isin, B.; Schulten, K.; Tajkhorshid, E.; Bahar, I. *Biophys. J.* **2008**, *95*, 789−803.
(35) Abrams, C. F.; Vanden-Eijnden, E. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 4961−4966.
(36) Tribello, G. A.; Ceriotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17509−17514.
(37) Zhang, C.; Ma, J. P. *J. Chem. Phys.* **2010**, *132*, 16.
(38) Markwick, P. R. L.; McCammon, J. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20053−20065.
(39) Dirks, R. M.; Xu, H. F.; Shaw, D. E. *J. Chem. Theory. Comput.* **2012**, *8*, 162−171.
(40) Hu, Y.; Hong, W.; Shi, Y.; Liu, H. *J. Chem. Theory. Comput.* **2012**, *8*, 3777−3792.
(41) Michielssens, S.; van Erp, T. S.; Kutzner, C.; Ceulemans, A.; de Groot, B. L. *J. Phys. Chem. B* **2012**, *116*, 8350−8354.
(42) Gur, M.; Madura, J. D.; Bahar, I. *Biophys. J.* **2013**, *105*, 1643−1652.
(43) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571−6575.
(44) Go, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696−3700.
(45) Levitt, M.; Sander, C.; Stern, P. S. *Int. J. Quantum Chem.* **1983**, 181−199.
(46) Garcia, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.
(47) Amadei, A.; Linnsen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412−425.
(48) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164−169.
(49) Berendsen, H. J. C.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165−169.
(50) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505−515.
(51) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586−592.
(52) Ma, J. P. *Structure* **2005**, *13*, 373−380.
(53) Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.
(54) Lu, M. Y.; Poon, B.; Ma, J. P. *J. Chem. Theory. Comput.* **2006**, *2*, 464−471.
(55) Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K. *J. Phys. Chem.* **1996**, *100*, 2567−2572.
(56) Grossfield, A.; Feller, S. E.; Pitman, M. C. *Proteins* **2007**, *67*, 31−40.
(57) Zhang, Z.; Lu, L.; Noid, W. G.; krishna, V.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 5073−5083.
(58) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(59) Zhang, X. J.; Wozniak, J. A.; Matthews, B. W. *J. Mol. Biol.* **1995**, *250*, 527−552.
(60) Weaver, L. H.; Matthews, B. W. *J. Mol. Biol.* **1987**, *193*, 189−199.
(61) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory. Comput.* **2008**, *4*, 435−447.
(62) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(63) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(64) Hockney, R. W.; Goel, S. P.; Eastwood, J. W. *J. Comput. Phys.* **1974**, *14*, 148−158.
(65) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* 2007, *126*.
(66) Hess, B. *J. Chem. Theory. Comput.* **2008**, *4*, 116−122.
(67) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(68) Carisey, A.; Ballestrem, C. *Eur. J. Cell. Biol.* **2011**, *90*, 157−163.
(69) Borgon, R. A.; Vonrhein, C.; Bricogne, G.; Bois, P. R. J.; Izard, T. *Structure* **2004**, *12*, 1189−1197.
(70) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res.* **2003**, *31*, 3381−3385.
(71) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714−2723.
(72) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(73) McHaourab, H. S.; Oh, K. J.; Fang, C. J.; Hubbell, W. L. *Biochemistry* **1997**, *36*, 307−316.
(74) de Groot, B. L.; Hayward, S.; van Aalten, D. M. F.; Amadei, A.; Berendsen, H. J. C. *Proteins* **1998**, *31*, 116−127.
(75) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712−3720.
(76) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383−394.
(77) Harada, R.; Kitao, A. *J. Chem. Phys.* **2013**, *139*, 035103.

(78) Bakolitsa, C.; Cohen, D. M.; Bankston, L. A.; Bobkov, A. A.; Cadwell, G. W.; Jennings, L.; Critchley, D. R.; Craig, S. W.; Liddington, R. C. *Nature* **2004**, *430*, 583−586.
(79) Amadei, A.; Ceruso, M. A.; Di Nola, A. *Proteins* **1999**, *36*, 419−424.
(80) Daidone, I.; Roccatano, D.; Hayward, S. *J. Mol. Biol.* **2004**, *339*, 515−525.
(81) Theobald, D. L.; Wuttke, D. S. *Bioinformatics* **2006**, *22*, 2171−2172.
(82) Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. *Structure* **2008**, *16*, 321−330.
(83) Bahar, I.; Chennubhotla, C.; Tobi, D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633−640.
(84) Yang, J.; Dokurno, P.; Tonks, N. K.; Barford, D. *EMBO J.* **2001**, *20*, 3645−3656.
(85) Tama, F.; Miyashita, O.; Brooks, C. L. *J. Mol. Biol.* **2004**, *337*, 985−999.
(86) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. *Structure* **2008**, *16*, 673−683.
(87) Zheng, W. S. *Biophys. J.* **2011**, *100*, 478−488.
(88) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656−5664.
(89) Pelikan, M.; Hura, G. L.; Hammel, M. *Gen. Physiol. Biophys.* **2009**, *28*, 174−189.
(90) Yang, S. C.; Blachowicz, L.; Makowski, L.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15757−15762.
(91) Różycki, B.; Kim, Y. C.; Hummer, G. *Structure* **2011**, *19*, 109−116.
(92) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.