JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

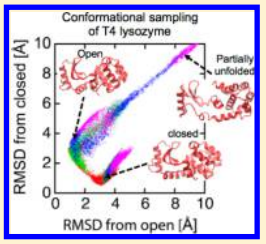
Nontargeted Parallel Cascade Selection Molecular Dynamics for Enhancing the Conformational Sampling of Proteins

Ryuhei Harada[†,‡] and Akio Kitao[*,§]

[†]Department of Physics, Graduate School of Pure and Applied Science, [‡]Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan

[§]Institute of Molecular and Cellular Bioscience, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

ABSTRACT: Nontargeted parallel cascade selection molecular dynamics (nt-PaCS-MD) is proposed as an efficient conformational sampling method to enhance the conformational transitions of proteins, which is an extension of the original targeted PaCS-MD (t-PaCS-MD). The original PaCS-MD comprises cycles of (i) selection of initial structures for multiple independent MD simulations toward a predetermined target and (ii) conformational sampling by the independent MDs. In nt-PaCS-MD, structures that significantly deviate from an average are regarded as candidates that have high potential to address other metastable states and are chosen as the initial structures in the selection. To select significantly deviated structures, we examine the root-mean-square deviation (RMSD) of snapshots generated from the average structure based on Gram−Schmidt orthogonalization. nt-PaCS-MD was applied to the folding of the mini-protein chignolin in implicit solvent and to the open−closed conformational transitions of T4 lysozyme (T4L) and glutamine binding protein (QBP) in explicit solvent. We show that nt-PaCS-MD can reach chignolin's native state and can also cause the open−closed transition of T4L and QBP on a nanosecond time scale, which are very efficient in terms of conformational sampling and comparable to that with t-PaCS-MD.



I. INTRODUCTION

Proteins often undergo anisotropic large-amplitude fluctuations when conducting biological functions such as protein folding, conformational transition, and molecular recognition.[1,2] A low number of collective atomic movements are typically related to large-amplitude fluctuations on a relatively slow time scale.[3−5] Molecular dynamics (MD) simulations with all-atom force fields can reproduce these slow large-amplitude fluctuations at atomic resolution and also provide fully time-dependent structural changes with femtosecond time resolution.[6] However, it is often a challenge to simulate anisotropic fluctuations relevant to biological functions because the characteristic time scale of functional motions can exceed feasible time scales simulated by standard MD simulations, even if significant computational resources are employed.[7] In conventional MD (CMD) simulations, Newton's equations of motion are integrated with a femtosecond-order time step, which requires 10^{12} time steps to reach milliseconds. In addition, biologically important conformational transitions tend to occur stochastically as rare events. Therefore, a long CMD simulation does not ensure that such rare events will be observed. More than a few efficient conformational sampling methods have been proposed to observe rare events relevant to biological functions.

Among these, one of the most relevant methods to the present work is targeted MD (TMD).[8] As a precondition to apply TMD, a target structure, e.g., a product with respect to a reactant, must be known a priori. In TMD, restraints are employed to reduce the root-mean-square deviation (RMSD) measured from the target structure. Another example is steered molecular dynamics (SMD),[9−11] in which a force bias is applied to pull part of a protein along a predetermined degree of freedom. These methods have been widely used to explore transition pathways, the binding/unbinding properties of ligands, and the folding/unfolding of proteins. In these methods, restraints or optimal force biases are imposed depending on the target systems, which then require adjustment of the parameter settings by trial and error.

In this study, we propose an effective and robust conformational sampling method to enhance conformational transitions of proteins without setting target structures and force biases, which is developed as an extension of parallel cascade selection molecular dynamics simulation (PaCS-MD).[12] The original PaCS-MD comprises cycles of (i) selection of initial structures for multiple independent MD simulations toward a predetermined target and (ii) conformational sampling by independent MD simulations. The initial structures of the MD simulations in each cycle are selected based on an appropriate quantity that characterizes conformational transitions so that the protein's conformation gradually approaches a target. The selection of snapshots enhances conformational changes by increasing the probability of capturing rare events that occur toward the target. Therefore, original PaCS-MD can be considered to be a targeted version of PaCS-MD (t-PaCS-MD). As an extension of PaCS-MD, we propose nontargeted PaCS-MD (nt-PaCS-MD), which does not require a specific target. In nt-PaCS-MD, we assume that large fluctuations in equilibrium can be often related to conformational transitions; large fluctuations could represent the directions of low-energy conformational changes, which can be related to the protein's response to perturbation, as suggested by

Received: July 30, 2015
Published: October 12, 2015

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

the fluctuation−dissipation theorem. These types of fluctuations are often related to biological functions.[13] The selection of snapshots that significantly deviate from the average structure in the PaCS-MD cycles enhances the probability of large conformational changes occurring in nt-PaCS-MD. We demonstrate the efficiency of nt-PaCS-MD for the folding of the mini-protein chignolin in implicit solvent and for the open−closed transition of T4 lysozyme (T4L) and glutamine binding protein (QBP) in explicit solvent, which were successfully simulated with computational costs of nanosecond-order MD.

II. MATERIAL AND METHODS

II.A. Original PaCS-MD. The original t-PaCS-MD generates conformational transition pathways that connect selected initial structures (reactant) to a target (product) based on parallel or distributed computing.[12] Therefore, a set of the reactant and product should be known a priori. Starting from a given reactant, independent multiple MD simulations (M, the number of MD simulations) with distinct initial conditions are performed as the first conformational sampling. t-PaCS-MD comprises cycles of (i) selections of M initial structures and (ii) conformational samplings with M independent MD simulations (Figure 1a). The initial structures for the next MD simulations are selected so that they gradually approach the product by ranking the MD snapshots in the previous cycle based on certain measurement quantities with respect to the product (Figure 1b). For example, the RMSD from the product was employed as the ranking measure (the smaller the RMSD, the better) in our previous study.[12] In each cycle, the top M snapshots are selected as the initial structures, and conformational sampling is conducted by restarting multiple independent MD simulations from the selected initial structures with regenerated initial velocities to reproduce the Maxwell−Boltzmann distribution. This cycle is subsequently repeated until highly ranked snapshots become sufficiently close to the product, e.g., the RMSD from the product is less than a threshold (we typically use 1 Å). A set of quasi-conformational transition pathways that connect the reactant and product is generated as concatenated PaCS-MD trajectories, which are referred to as reactive trajectories. Snapshots selected from the reactive trajectories are used to conduct multiple umbrella sampling to calculate the free energy change along the reactive trajectories.[12] The obtained pathways can also be refined by well-established path search methods such as the string method[14] or transition path sampling method.[15,16]

II.B. Nontargeted PaCS-MD. nt-PaCS-MD, an extension of PaCS-MD, also repeats the cycle of initial structure selection and conformational sampling with independent MD simulations, but the selection is conducted without particular targets. In nt-PaCS-MD, structures that significantly deviate from the average are regarded as candidates for the next cycle that have high potential to address other metastable states (Figure 1c). To select significantly deviated structures, RMSDs of generated snapshots from the average structure in each cycle were examined. In this work, the average structure used to select the initial structures of the ith cycle was calculated from 1000 randomly selected snapshots from past trajectories with a ratio of (the number of snapshots from the first to $(i-2)$th cycles)/(that from the $(i-1)$th cycle) = 9:1, except for the first cycle. In the first cycle, the trajectories generated by preliminary MD runs are employed to calculate the average structure.

After the average structure is determined, the initial structures for the next cycle are selected based on Gram−Schmidt orthogonalization (Figure 2). First, the most deviated snapshot

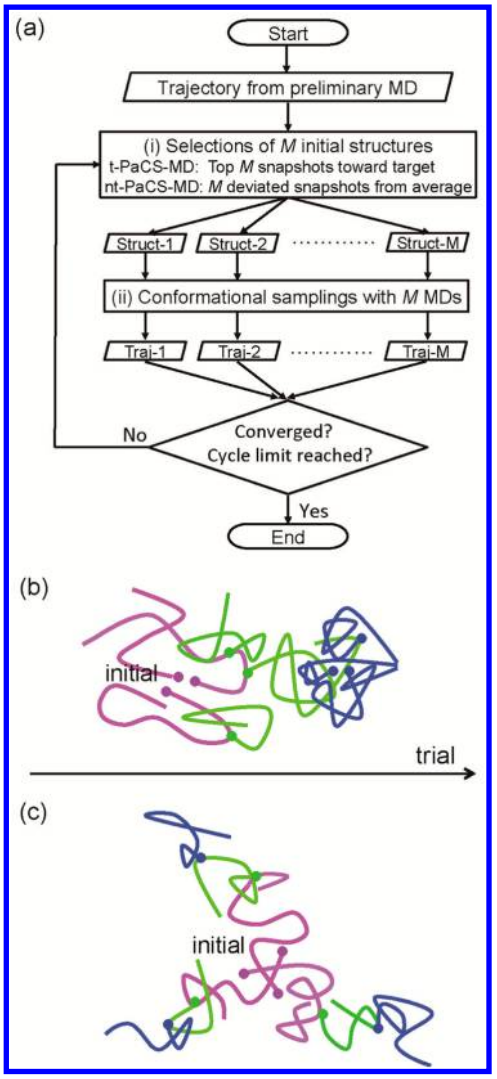


Figure 1. Procedure of PaCS-MD. (a) The flowchart of PaCS-MD. Conceptional trajectories generated by (b) t-PaCS-MD and (c) nt-PaCS-MD in the case of M = 3. The filled circles indicate the starting point of each trajectory. The trajectories are sequentially generated in the order magenta, green, and blue.

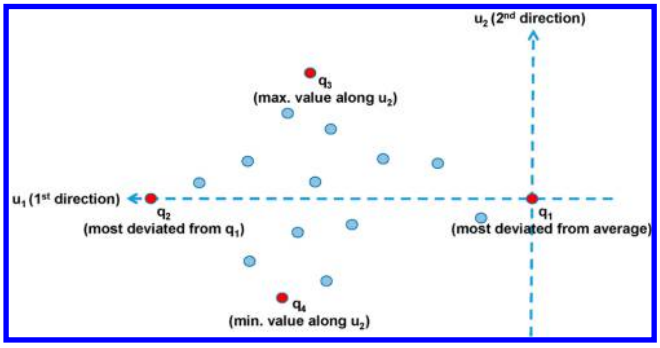


Figure 2. Selection of the initial structures in each cycle of nt-PaCS-MD using Gram−Schmidt orthogonalization.

from the average is defined as $\vec{q}_1$ and $\vec{q}_2$ is obtained as the most deviated snapshot from $\vec{q}_1$. The first axis, $\vec{u}_1$, is defined in a normalized form

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

Table 1. Summary of nt-PaCS-MD for Chignolin in Implicit Solvent

nt-PaCS-MD index	M	cycle	first cycle reached/cycle reached closest to the native	min/max C_α RMSD from the native structure [Å]	min/max hydrogen-bonding distances D3:N−G7:O [Å]	min/max hydrogen-bonding distances D3:N−T8:O [Å]
CH1	10	100	46/46	0.4/7.1	2.8/17.7	2.7/21.0
CH2	10	100	2/5	0.3/7.1	2.6/17.8	2.6/20.8
CH3	10	100	8/38	0.3/7.1	2.7/17.7	2.6/21.4
CH4	10	100	10/18	0.5/7.2	2.7/14.4	2.6/21.0
CH5	10	100	29/32	0.4/7.1	2.7/17.4	2.7/21.1
average			19.0/27.8	0.4/7.1	2.7/17.0	2.6/21.1

$$\vec{u}_1 = \frac{\vec{q}_2 - \vec{q}_1}{|\vec{q}_2 - \vec{q}_1|} \tag{1}$$

where $\vec{q}_1$ is employed as the origin. Each snapshot, $\vec{q}$, is orthogonalized based on the Gram−Schmidt orthogonalization with respect to $\vec{u}_1$ as

$$\vec{v}_2 = \frac{\Delta\vec{q} - (\vec{u}_1\cdot\Delta\vec{q})\vec{u}_1}{\|\Delta\vec{q} - (\vec{u}_1\cdot\Delta\vec{q})\vec{u}_1\|} \tag{2}$$

where $\Delta\vec{q} = \vec{q} - \vec{q}_1$ $\left(\vec{q} \neq \vec{q}_1, \vec{q} \neq \vec{q}_2\right)$. Herein, the absolute value of projection for $\Delta\vec{q}$ onto $\vec{v}_2$

$$|\vec{v}_2\cdot\Delta\vec{q}| \tag{3}$$

is considered. The third snapshot $\vec{q}_3$ is selected as that which gives the maximum value among the remaining snapshots

$$\max_{\Delta\vec{q}\neq 0, \Delta\vec{q}_2} |\vec{v}_2\cdot\Delta\vec{q}| \tag{4}$$

and the second axis $\vec{u}_2$ is defined as

$$\vec{u}_2 = \frac{\Delta\vec{q}_3 - (\vec{u}_1\cdot\Delta\vec{q}_3)\vec{u}_1}{\|\Delta\vec{q}_3 - (\vec{u}_1\cdot\Delta\vec{q}_3)\vec{u}_1\|} \tag{5}$$

where $\Delta\vec{q}_3 = \vec{q}_3 - \vec{q}_1$. $\vec{q}_4$ is selected as the snapshot that gives $\min_{\Delta\vec{q}\neq 0, \Delta\vec{q}_2, \Delta\vec{q}_3}\left(\vec{u}_2\cdot\Delta\vec{q}\right)$ from the remaining snapshots. The snapshot $\vec{q}_{2n-1}$ is generally defined to give $\max_{\Delta\vec{q}\neq 0, \Delta\vec{q}_2, \ldots, \Delta\vec{q}_{2n-2}} \left|\vec{v}_n\cdot\Delta\vec{q}\right|$ with

$$\vec{v}_n = \frac{\Delta\vec{q} - \sum_{i=1}^{n-1}(\vec{u}_i\cdot\Delta\vec{q})\vec{u}_i}{\|\Delta\vec{q} - \sum_{i=1}^{n-1}(\vec{u}_i\cdot\Delta\vec{q})\vec{u}_i\|} \tag{6}$$

The nth axis is defined as

$$\vec{u}_n = \frac{\Delta\vec{q}_{2n-1} - \sum_{i=1}^{n-1}(\vec{u}_i\cdot\Delta\vec{q}_{2n-1})\vec{u}_i}{\|\Delta\vec{q}_{2n-1} - \sum_{i=1}^{n-1}(\vec{u}_i\cdot\Delta\vec{q}_{2n-1})\vec{u}_i\|} \tag{7}$$

where $\Delta\vec{q}_{2n-1} = \Delta\vec{q}_{2n-1} - \vec{q}_1$. $\vec{q}_{2n}$ is selected as the snapshot that gives $\min_{\Delta\vec{q}\neq 0, \Delta\vec{q}_2, \ldots, \Delta\vec{q}_{2n-1}} \left(\vec{u}_n\cdot\Delta\vec{q}\right)$ from the remaining snapshots.

II.C. Computational Details. To demonstrate the efficiency of nt-PaCS-MD in conformational sampling, nt-PaCS-MD was applied to two systems: (i) folding of the mini-protein chignolin in implicit solvent and (ii) the open−closed conformational transition of T4L in explicit solvent. These systems were selected as targets because they have been well-investigated in the examination of sampling efficiency.[12,17−19]

For the folding study of chignolin (10 residues and 138 atoms), the generalized Born with surface area (GBSA) model was employed to include the solvent effect. The salt concentration was set to 0.2 M, and the surface tension was 0.005 kcal/mol/Å^2 (IGB = 5 in AMBER11).[20] MD simulations were performed using the SANDER module in AMBER11[20] with the AMBER ff99SB force field.[21] The simulation time step was 2 fs with bond constraints involving hydrogen atoms via SHAKE[22] and SETTLE.[23] A completely extended structure was first constructed based on the amino acid sequence (GYDPETGTWG) with the tLeaP module of AMBER11.[20] As preliminary runs of nt-PaCS-MD, ten 0.1 ns MD simulations were independently conducted with distinct initial velocities starting from the extended structure. As the initial structures of nt-PaCS-MD, the top 10 structures (M = 10) were selected with the Gram−Schmidt orthogonalization for the five axes (n = 5). The production of nt-PaCS-MD cycles, each of which comprised ten 0.1 ns MD simulations, was repeated 100 times. The total computational time is 100 ns (10 independent MDs × 0.1 ns × 100 cycles). Five distinct nt-PaCS-MDs (CH1−CH5) were conducted, as shown in Table 1.

For T4L, collective domain motion between the open and closed states was investigated. The goal of this application is to observe the conformational transitions from the open to closed states and vice versa. For the conformational transition from the open to closed states, the mutation of the I6M mutant structure (PDB ID 150L)[24] back to the wild type was considered as the reactant (starting structure). As the starting structure for the closed to open simulations, the wild-type closed structure (PDB ID 2LZM)[25] was selected. To model fully atomic systems in explicit solvent, these initial structures were solvated using the TIP3P water model.[26] Rectangular simulation boxes were constructed with a margin of at least 10 Å from the proteins to the periodic box boundaries. Each system contained 8876 water molecules and eight Cl^- ions to neutralize the system. Independent MD simulations of nt-PaCS-MD were performed with the PMEMD module of AMBER11[20] under a constant temperature condition at 300 K using a Langevin thermostat with a friction constant of 2.0 ps^{-1}. The AMBER ff99SB force field[21] was employed for T4L. Except for the initial pressure adjustment, MD simulations were performed with an NVT ensemble (T = 300 K). The electrostatic interactions were treated with the particle mesh Ewald method,[27] where the Lennard-Jones interaction and the real-space Ewald sum are smoothly switched to zero at 9 Å. The simulation time step was 2 fs with bond constraints involving hydrogen atoms. As preliminary runs for both directions (from the open to closed states and vice versa), 10 independent 0.1 ns MD simulations were started from the

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

Table 2. Summary of nt-PaCS-MDs for the Open−Closed Conformational Transition of T4L in Explicit solvent[a]

nt-PaCS-MD index	M	cycle	first cycle reached/cycle reached closest to the opposite	min/max C_α RMSD from the opposite structure [Å]	min/max values of PC1 [Å]	min/max values of PC2 [Å]
OP1	10	50	11/20	0.8/7.3	−1.7/5.3	−3.2/3.0
OP2	10	50	−/35	1.0/6.1	−2.4/4.2	−2.6/2.9
OP3	10	50	−/48	1.4/7.5	−2.2/5.4	−4.5/3.7
OP4	10	50	31/47	0.8/6.6	−2.6/4.4	−4.6/2.0
OP5	10	50	23/33	0.6/5.5	−2.2/3.9	−2.0/2.6
OP1−OP5 average			−/36.6	0.9/6.6	−2.2/4.6	−3.4/2.8
CL1	10	50	29/50	0.9/5.2	−3.0/3.4	−3.0/1.5
CL2	10	50	35/36	0.9/4.6	−2.4/4.7	−4.6/1.2
CL3	10	50	18/26	0.9/6.3	−2.4/3.2	−2.5/2.0
CL4	10	50	20/32	0.8/5.6	−2.3/2.5	−2.5/1.8
CL5	10	50	27/27	0.9/5.4	−2.4/2.8	−3.1/1.8
CL1−CL5 average			25.8/34.2	0.9/5.4	−2.5/3.3	−3.2/1.7
OP1L	10	200	11/20	0.8/9.8	−3.0/7.0	−5.1/5.6
CL1L	10	200	29/50	0.5/11.8	−3.1/3.9	−8.2/2.5
OP6	20	50	11/34	0.7/6.2	−3.5/4.4	−3.2/3.3
OP7	20	50	6/39	0.7/6.1	−2.7/4.2	−2.6/2.6
OP8	20	50	−/45	1.0/6.7	−2.5/4.0	−3.8/2.4
OP9	20	50	14/16	0.8/6.1	−2.4/4.2	−4.1/3.6
OP10	20	50	−/40	1.0/6.7	−1.8/4.7	−4.6/3.9
OP6−OP10 average			−/35	0.9/6.4	−2.6/4.3	−3.7/3.2

[a]OP1−OP10 represent the simulations from the open state, and CL1−CL5 indicate those from the closed states. The opposite structures were defined as average snapshots over 10 ns MD simulations started from the open/closed structures.

Table 3. Summary of nt-PaCS-MDs for the Open−Closed Conformational Transition of QBP in Explicit Solvent[a]

nt-PaCS-MD index	M	cycle	first cycle reached/cycle reached closest to the opposite	min/max C_α RMSD from the opposite structure [Å]	min/max values of PC1 [Å]	min/max values of PC2 [Å]
OPQ	10	200	121/144	1.7/11.1	−2.2/5.7	−3.5/6.8
CLQ	10	200	27/147	1.9/11.8	−4.4/6.0	−3.3/8.2

[a]OPQ and CLQ represent the simulations from the open and closed states, respectively. The opposite structures were defined as average snapshots over 10 ns MD simulations started from the open/closed structures.

open or closed structures (10 runs × 0.1 ns). The top five axes were considered to select the initial structures. The nt-PaCS-MD cycle was repeated 50 times (10 initial structures × 0.1 ns × 50 cycles, which corresponds to a totally computational cost of 50 ns). Five nt-PaCS-MD cycles starting from the open (OP1−OP5) and closed (CL1−CL5) structures were conducted, as shown in Table 2. OP1 and CL1 were further extended up to 200 cycles (OP1L and CL1L, respectively). Five nt-PaCS-MD cycles starting from the open structures with M = 20 were also performed (OP6−OP10).

The open−closed conformational transition of QBP was also investigated. QBP is one of the periplasmic binding proteins, consisting of 226 residues and 3535 atoms. Upon the binding of a glutamine to a cleft between two domains in the open form, large domain movement is induced toward the closed form.[28] The open apo (PDB ID 1GGG)[29] and closed holo (PDB ID 1WDN)[30] structures have already been determined by X-ray crystallography. Of note, the conformational transitions in the apo state (without the ligand) were investigated in this work. To model initial structures, the structures of the open (OPQ) and closed (CLQ) forms in the apo state were prepared from these X-ray structures; for the latter, the ligand was removed from the closed holo form. QBP was solvated with 9999 TIP3P water molecules and one Cl^- ion was added to neutralize the systems in a rectangular box, amounting to 33 533 atoms. The AMBER ff03 force field[31] was employed for QBP. The programs for MD simulations and parameters for temperature and pressure controls were the same with those used for T4L. After energy minimizations, short-time (300 ps) MD simulations were started from the modeled structures to adjust the volumes of the systems under the NPT ensemble (P = 1.0 atm, T = 300 K). Then, MD simulations were conducted with NVT at T = 300 K. As preliminary runs for both trials (from the open to close structures and vice versa), 10 independent 0.1 ns MD simulations were started from the initial structures (10 runs × 0.1 ns). The top five axes were considered to select 10 initial structures. The nt-PaCS-MD cycle was repeated 200 times (10 initial structures × 0.1 ns × 200 cycles; total computational cost of 200 ns). A set of nt-PaCS-MD was conducted as shown in Table 3.

III. RESULTS AND DISCUSSION

III.A. Folding of Chignolin in Implicit Solvent. To examine the convergence of conformational sampling in the case of chignolin, C_α root-mean-square fluctuation (RMSF) of the snapshots in each cycle from the average structure was monitored during the 100 cycles (Figure 3a). The error bars represent the standard deviation of the five distinct trials of nt-PaCS-MD started from different initial structures. The RMSF value rapidly increased during the first 20 cycles and gradually converged into a value (ca. 3.0 Å) at around the fortieth cycle.

A sampled conformation was determined as having reached the native state when the C_α RMSD from the native structure (the first nuclear magnetic resonance, NMR, structure) was less than 1.0 Å. To observe chignolin folding, conventional MD requires a microsecond-order simulation.[17,19,32] In contrast, nt-PaCS-MD first reached the native state with 19.0 cycles, on

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

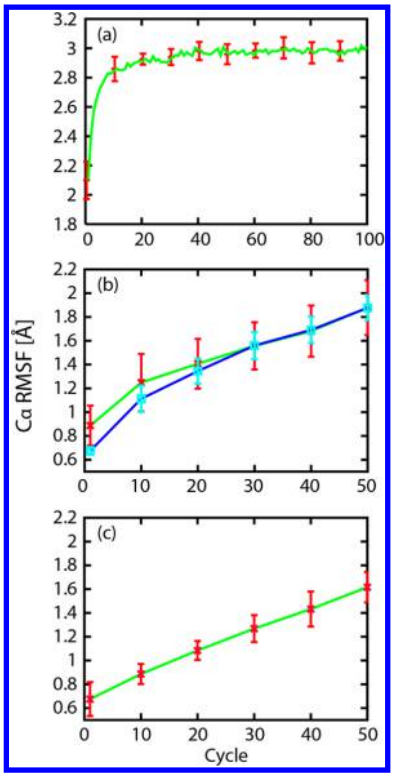


Figure 3. C_α root-mean-square fluctuations (RMSF) of the sampled snapshots in each cycle from the average. The average RMSF values (green) and standard deviations (error bars in red) over five distinct nt-PaCS-MDs are plotted. (a) 100 cycles for the folding of chignolin in implicit solvent (average over CH1−CH5) and (b, c) 50 cycles for the conformational transition of T4L started from the (b) open (average over OP1−OP5) and (c) closed states (CL1−CL5). The results of nt-PaCS-MD with 10 independent MDs are shown except for the result with 20 MDs (average over OP6−OP10) shown by average RMSF values (blue) and standard deviations (error bars in cyan) in (b).

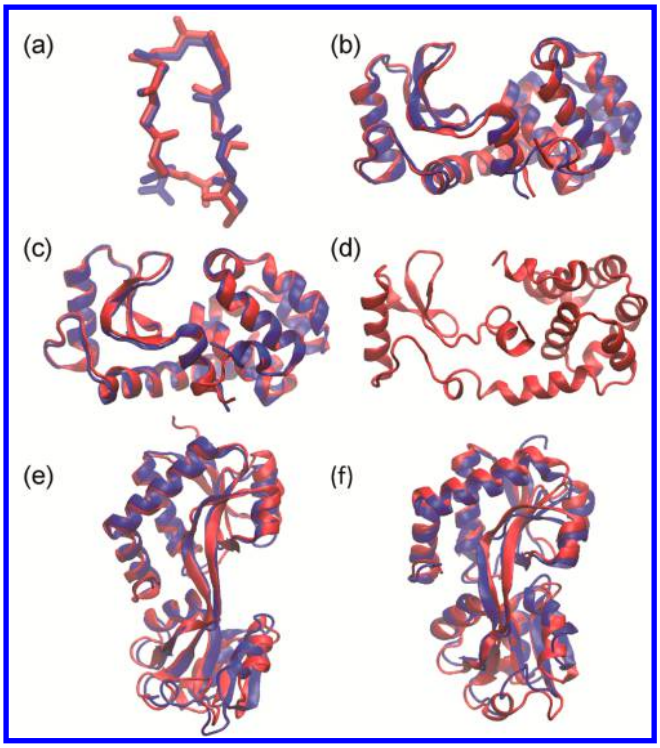


Figure 4. Superposition of the snapshots sampled by nt-PaCS-MD (red) with metastable structures in solution (blue). (a) First solution-NMR structure of chignolin (blue) and the snapshot closest to the NMR structure sampled by nt-PaCS-MD (red: RMSD 0.3 Å in CH2). (b) Average closed structure of T4L during 10 ns MD (blue) and the snapshot closest to the average (red: RMSD 0.6 Å in OP5). (c) Average open structure of T4L during 10 ns MD (blue) and the snapshot closest to the average (red: RMSD 0.8 Å in CL4). (d) Snapshot of CL1L at 17.4 ns (174 cycle) in the ninth MD. (e) Average closed structure of QBP during 10 ns MD (blue) and the snapshot closest to the average (red: RMSD 1.7 Å in OPQ). (f) Average open structure of QBP during 10 ns MD (blue) and the snapshot closest to the average (red: RMSD 1.9 Å in CLQ). Also, see Tables 1−3. This figure was created with VMD.[50]

average (Table 1), which is equivalent to 1.9 ns in the MD simulation time and 19.0 ns in total computational time. In the case of t-PaCS-MD, 19.4 cycles were required with the Langevin thermostat using the native structure as the target. Therefore, the computational costs required to reach the native state in nt- and t-PaCS-MDs are comparable. The superposition of the first NMR structure and the closest snapshot with the minimum C_α RMSD (0.3 Å in CH2) is shown in Figure 4a, which demonstrates that they are well-overlapped. The minimum C_α RMSD from the native structure and computational cost to sample it are listed in Table 1 for each trial.

The total computational cost of 19.0 ns MD (1.9 ns MD × 10) in nt-PaCS-MD is also compared to that with other methods. In replica exchange MD with hybrid Hamiltonian (REMhH),[33] the reversible folding of chignolin has been studied with a Poisson−Boltzmann model in implicit water with eight replicas, which required 20.8 ns MD to fold chignolin into the native state starting from the extended structure. This is equivalent to a total cost of 166.4 ns MD, which is computationally much more demanding than that with nt- and t-PaCS-MDs. In multicanonical MD (McMD) using the GB/SA model for 180 ns,[34] the native structure of chignolin was reached after 50 ns, which also required more computational cost than that with nt- and t-PaCS-MDs. In the transform and relax sampling (TRS) with explicit water,[18] folding into the native state was first observed at 0.2 and 1.2 ns in distinct TRS trials with single trajectories, which is more efficient than with PaCS-MDs. From the comparisons with the preceding studies, nt-PaCS-MD is shown to be a powerful conformational sampling method for observing conformational transitions efficiently.

The trajectories generated by nt-PaCS-MD were projected onto the subspace spanned by a set of hydrogen-bonding distances of the backbone that were defined in the preceding studies:[34,35] Asp3:N−Gly7:O and Asp3:N−Thr8:O (the red dots in Figure 5). These hydrogen-bond distances are essential to distinguish the misfolded (blue squares in Figure 5) and native states (green squares). In all trials, broad conformational spaces were sampled, and both the misfolded and native states were reached. High degrees of overlap between the distributions in Figure 5a−e (the first 50 cycles) and those in Figure 5f−j (all 100 cycles) also indicate the convergence of conformational sampling, which is consistent with the RMSD convergence shown in Figure 3a.

III.B. Open−Close Conformational Transitions of T4L in Explicit Solvent. To investigate the conformational space of T4L sampling with nt-PaCS-MD, a set of RMSD values from the open and closed structures was employed to project the nt-PaCS-MD trajectories. The red dots in Figure 6 show the projections of all of the trajectories during the 50 cycles of five independent nt-PaCS-MDs (OP1−OP5) started from distinct initial structures

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

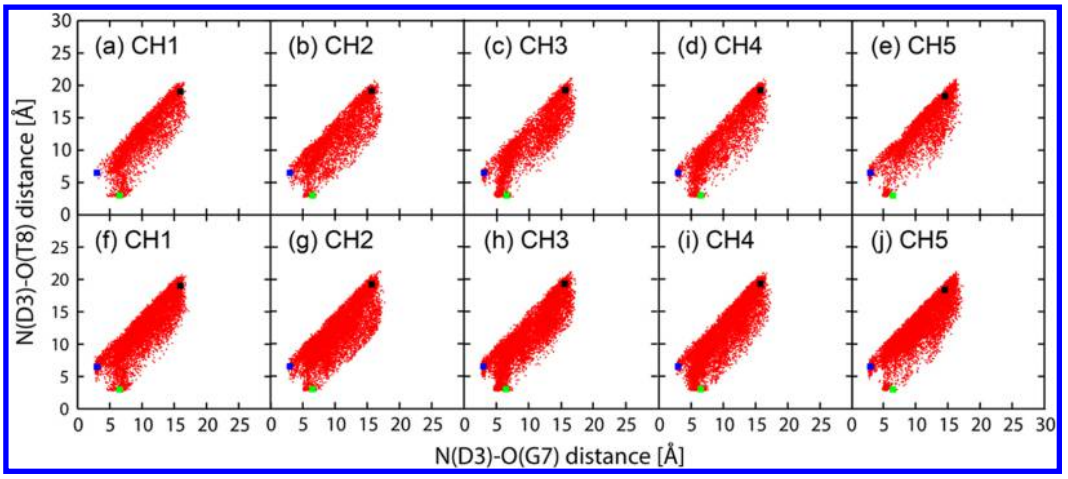


Figure 5. Projections of the trajectories generated by nt-PaCS-MD of chignolin onto the subspace spanned by two hydrogen-bond distances of the backbone of chignolin. (a–e) Trajectories of the five distinct trials CH1–CH5 during the first 50 cycles and (f–j) during 100 cycles. The squares in black, blue, and green represent the initial, misfolded, and native structures, respectively.

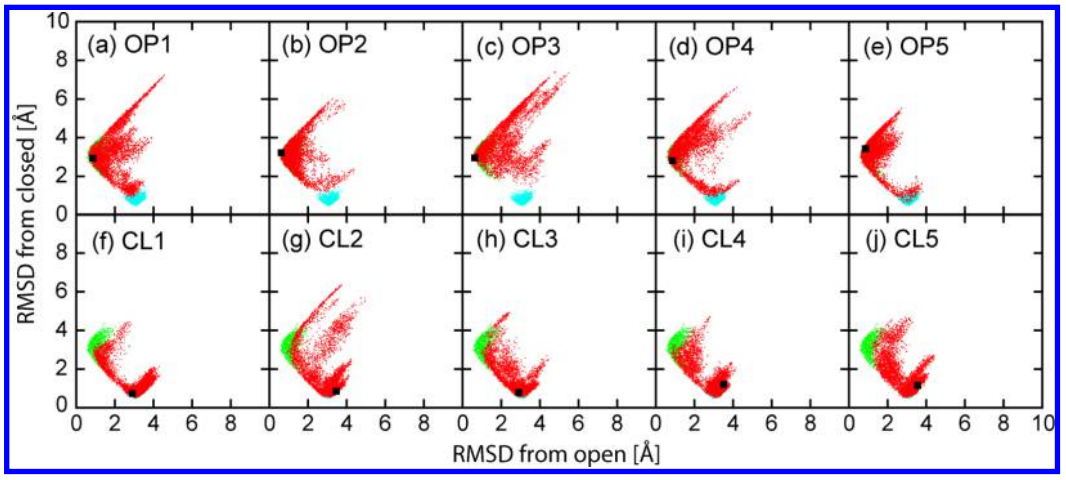


Figure 6. Projections of all of the trajectories generated by nt-PaCS-MD of T4L onto a subspace spanned by RMSDs from the open and closed structures during 50 cycles (red). (a–e) Trajectories started from the open (OP1–OP5 in Table 2) and (f–j) closed structures (CL1–CL5). Projections of 10 ns MD simulations starting from the open (green) and closed structures (cyan). The squares (black) correspond to the distinct initial structures of five independent nt-PaCS-MD.

around the open state (Figure 6a–e) and those started from the closed state (CL1–CL5; Figure 6f–j). Initial structures are shown by black squares. The green and blue dots represent snapshots from 10 ns canonical MD simulations at 300 K started from the open and closed structures, respectively. The results indicated that all of the nt-PaCS-MDs that were started from one of the states successfully reached around another state within 50 cycles without using information on the targets. The minimum C_α RMSDs from the opposite structure were within the range of 0.6–1.4 Å (Table 2), which shows that the sampled conformations approached very closely the opposite state in most cases (also see Figure 4b,c). More rigorously, if a conformation reached the native state with a C_α RMSD from the average structure of the opposite 10 ns MD snapshots <1.0 Å, then three out of five trials that started from the open state reached the closed state and all of the trials that started from the closed state successfully reached the open state. The minimum number of cycles required to reach another state were 11 (OP1) and 18 cycles (CL3), which correspond to simulation times of 1.1 and 1.8 ns and total simulation costs of 11 and 18 ns, respectively, i.e., high conformational sampling efficiency was achieved. Structural comparisons between the average open and closed structures and the closest sampled structures in Figure 4b,c indicate high similarity. In the case of t-PaCS-MD, the transitions from the open to closed structures and vice versa occurred with 7.6 and 11.6 cycles, on average, over five trials, respectively.[12] Therefore, the computational costs to sample the open–closed transitions of T4L with nt-PaCS-MD were slightly less efficient but still on the same order. The computational cost of nt-PaCS-MD is comparable to that of TRS[18] in which the open–closed transitions were observed several times during 20 ns in five distinct trials of TRS. Open–closed structural transitions take longer amounts of time with standard MD simulations, as seen in the example where a 450 ns MD simulation with the AMBER ff99 force field[36] stayed in the open state.[37] However, the structural transitions of T4L may be dependent on the force field because MD simulations with a modified version of GROMOS87 successfully sampled a transition from the open to closed states within 1 ns.[38]

The RMSFs of the snapshots in each cycle from the average structure shown in Figures 3b,c indicate that the sampling did not converge within 50 cycles. It is reasonable that the conformational space for this size of protein cannot be sampled on a nanosecond time scale. To examine the cycle dependence of the

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493–5502

sampled conformational space, OP1 and CL1 were further continued up to 200 cycles (Table 2). Figure 7 shows the

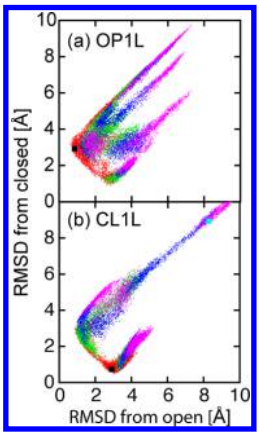


Figure 7. Projections of the (a) OP1L and (b) CL1L trajectories of T4L (200 cycles). The squares (black) indicate the initial structures. Each color indicates the results from different cycle number ranges: 1−50 (red), 51−100 (green), 101−150 (blue), and 151−200 cycles (magenta). The cyan square in (b) indicates the snapshot shown in Figure 4d.

projections of trajectories generated by nt-PaCS-MD during 200 cycles for (a) the open to closed and (b) closed to open structural transitions, which indicate that nt-PaCS-MD explored conformational space that is very far from the open and closed structures with more cycles. For example, Figure 4d show a representative snapshot that is very far from the open and closed structures that corresponds to the cyan dot in Figure 7b. The structure is partially unfolded around the hinge regions between two domains. From these projections, nt-PaCS-MD sampled not only structural transitions near the native state but also some states that may be relatively higher in energy.

Finally, the effects of the number of independent MD simulations in each cycle were examined. In addition, ten-dimensional basis sets (n = 10 and M = 20) were employed, and nt-PaCS-MD was conducted starting from the open T4L structures with 20 MDs. The generated trajectories were projected onto the two-dimensional space spanned by the first two principal components (PC1 and PC2) obtained from the mixed trajectories from the 10 ns conventional MD started from the open and closed states. Figure 8 shows the results of five trials of (Figure 8a−e) five-dimensional (10 MDs) and (Figure 8f−j) ten-dimensional (20 MDs) nt-PaCS-MD. The sampled conformational spaces between the five- and ten-dimensional cases were not significantly different, which indicates that increasing the number of dimensions for the Gram−Schmidt orthogonalization in eq 6 does not significantly affect the conformational sampling efficiency. This is confirmed by the RMSF plot comparison shown in Figure 3b. As a conclusion, the five-dimensional basis sets were sufficient to efficiently perform nt-PaCS-MD in this case.

Although T4L is much larger than chignolin in molecular size, why is the same value of M sufficient to produce efficient sampling? The results suggest that the optimum M values may not simply depend on the system size but also should depend more on the number of important degrees of freedom with which the process of interest is involved. T4L has a much larger number of degrees of freedom than chignolin; however, the open−closed transition of T4L occurs mainly in a subspace spanned by a small number of principal components, especially the first and second principal components. This can explain why chignolin and T4L require the same values of M in PaCS-MDs.

III.C. Open−Closed Conformational Transitions of QBP in Explicit Solvent. To examine the conformational space of QBP, a set of RMSDs from the average open and closed structures in the conventional MDs (Figure 9a,b) as well as the first principal components (Figure 9c,d) was employed as reaction coordinates to project the generated trajectories. In Figure 9, the green and cyan dots represent snapshots from the 10 ns canonical MD simulations started from the open and closed structures, respectively, and the red dots show the projections of all of the trajectories during 200 cycles for a set of nt-PaCS-MDs. The initial structures are indicated by the black squares. The nt-PaCS-MD trajectories starting from the open and closed structures reached close to the opposite states within 200 cycles successfully. The conformational areas sampled by nt-PaCS-MDs covered not only the areas around the open and closed structures and the areas between them but also broader

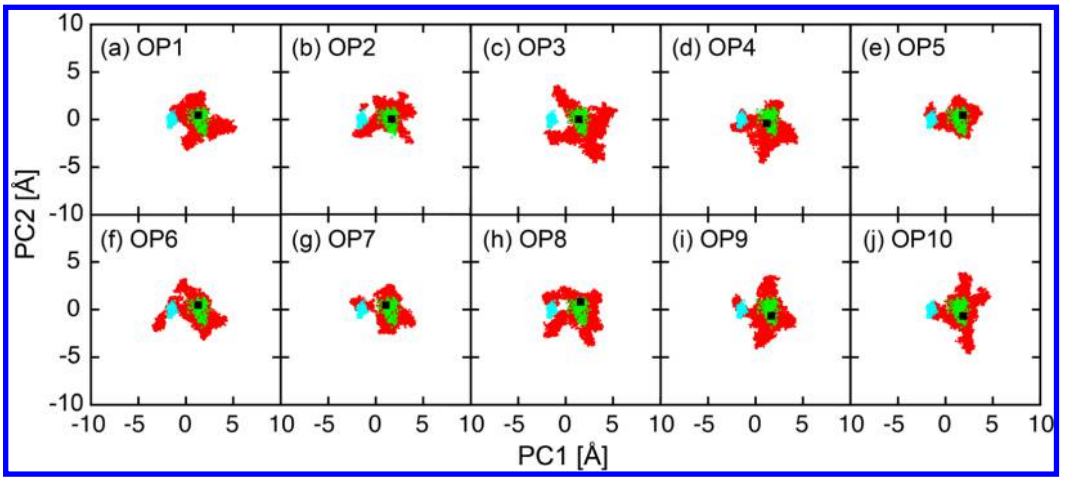


Figure 8. Projections of 50 cycles of trajectories generated by nt-PaCS-MD of T4L onto a subspace spanned by the first and second principal components (PC1 and PC2). (a−e) Five-dimensional basis (n = 5) with 10 independent MDs (OP1−OP5) and (f−j) ten-dimensional basis sets (n = 10) with 20 MDs (OP6−OP10). Projections of the 10 ns MD simulations started from the open (green) and closed (cyan) structures as well as the distinct initial structures for each trial (black).

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

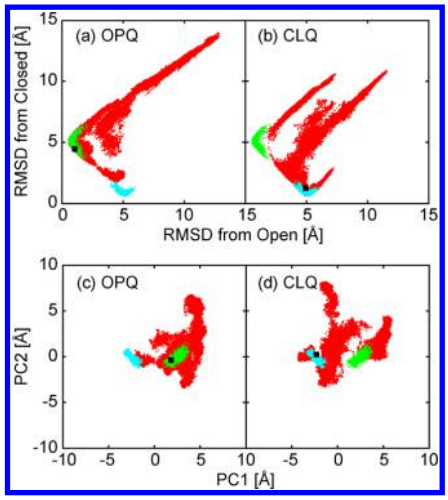


Figure 9. Projections of all of the trajectories generated by nt-PaCS-MD of QBP onto a subspace spanned by (a, b) RMSDs from the open and closed structures and by (c, d) the first and second principal components (PC1 and PC2) during 200 cycles (red). (a, c) Trajectories starting from the open (OPQ in Table 3) and (b, d) closed structures (CLQ). Projections of 10 ns MD simulations starting from the open (green) and closed structures (cyan) as well as the initial structures for each trial (black).

conformational areas that are very far from the initial states. This indicates the high conformational sampling efficiency of nt-PaCS-MD. The conformational sampling with TRS[18] also demonstrated a closed structure different from the closed structure with the ligand along the second principal component, which is consistent with the results of nt-PaCS-MD.

Figure 4e−f shows comparisons of the closest snapshots to the average open and closed structures, which also indicate high similarity. Table 3 shows the summary of the QBP simulations. The minimum C_α RMSDs from the opposite average structures were 1.7 and 1.9 Å for OPQ and CLQ, respectively. When C_α RMSDs from the opposite average structure reached within 2.0 Å, we judged the conformational transition to another state as having been completed. The conformational transition from the closed to open structures was observed with only 27 cycles (2.7 ns). In contrast, 121 cycles (12.1 ns) were required for the inverse transition from the open to closed structures. This difference might be related to the closed structure being relatively unstable in the apo state because it is stabilized by QBP−glutamine interactions. Therefore, observing the open to closed transition should be more difficult. The size of QBP (226 residues) is 140% larger than that of T4 lysozyme (164 residues), whereas the minimum number of cycles required to reach another state in QBP was much greater than this ratio. Additionally, it was also dependent on the directions of the movement. Judging from this result, sampling efficiency tends to depend on the softness or energy barrier of the conformational changes rather than the size of the system. Compared to other methods, TRS successfully captured open−closed conformational transitions a few times within 20 ns.[18] Therefore, the computational cost of nt-PaCS-MD (121 and 27 ns for one open to closed transition and vice versa, respectively) is slightly higher than that of TRS. The higher efficiency of TRS may be related the use of a transform force bias, which can strongly drive conformational changes along soft directions. Since PaCS-MD does not use a force bias, it can sometimes require more cycles to select steep uphill movements.

IV. CONCLUSIONS

In this study, an extension of the original targeted PaCS-MD (t-PaCS-MD)[12] was proposed as nt-PaCS-MD, which does not require specification of target quantities. In nt-PaCS-MD, snapshots that significantly deviate from an average are selected as the initial structures of the next cycle based on Gram−Schmidt orthogonalization. We have shown that nt-PaCS-MD can reach the native state of chignolin and can also cause the open−closed transition of T4L on a nanosecond time scale, which demonstrates very efficient conformational sampling that is comparable to that of t-PaCS-MD. As stated in a previous paper,[12] PaCS-MD is designed to find possible conformational pathways with relatively low computational cost and is expected to be combined with other more rigorous sampling methods, such as umbrella sampling,[39,40] transition path sampling,[15,16] the string method,[14] and the Markov state model (MSM).[41−43]

The sampling efficiency of PaCS-MD is achieved by selecting rare events that enhances the probability of their occurrence. The selection scheme for t- and nt-PaCS-MDs is not based on rigorous physical principles, which has both advantages and disadvantages. One of the major disadvantages is that PaCS-MD does not directly calculate free energy, as the relationships among the generated trajectories are not obvious. This is a disadvantage compared to a well-established method, for example, metadynamics.[44] However, this problem can be solved if the MSM[41−43] is applicable. For this purpose, each trajectory should be sufficiently long to be Markovian, and the number of trajectories should be statistically sufficient to estimate the Markov transition matrix, which will be reported elsewhere. Compared to metadynamics, which requires a set of reaction coordinates a priori to introduce Gaussian potentials, nt-PaCS-MD has an advantage in that reaction coordinates are not predefined. It is not necessarily trivial to specify a proper set of reaction coordinates to describe conformational transitions. One major advantage is that the selection can have variations depending on the purposes; nt-PaCS-MD can be considered one of these. Another advantage is that PaCS-MD can provide a fail-safe scheme for massive parallel and distributed computing. Each MD has no rigorous relationship with other MD simulations running in parallel; therefore, unexpected failure of some of the MD runs can be ignored. PaCS-MD is also suitable for massive parallel computing because it is composed of a series of multiple independent MD simulations. It should also be noted that no modification of MD programs is required for t- and nt-PaCS-MDs, but selecting initial structures and restarting the next cycle in the MD simulations can be handled with relatively simple scripts with straightforward implementation.

In nt-PaCS-MD, the conformational sampling in each cycle is restarted from the selected structures situated at marginal regions that are separated from stable states. Therefore, the selected structures are relatively close to transition states, and the MD simulations of the following cycles facilitate the transition into other metastable states with high probabilities, which tends to prevent confinement in a misfolded state. It might be nontrivial to judge the convergence of the conformational sampling because the selected structures are always situated at the edge of the calculated conformational distribution, which can gradually expand the sampled conformational space outward, even when the sampled conformation reaches high-energy space. Therefore, it is suggested that the nt-PaCS-MD cycle be stopped if the speed of the expansion becomes significantly slow.

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

If the target state is unknown, then a few solutions can be proposed to predict important conformational states. A possible solution is to investigate probability distributions of the generated structures by clustering methods. The obtained clusters can be candidates of important conformational states. Another solution is to calculate free energy landscapes with a combination of multiple umbrella samplings with the weighted histogram analysis method (WHAM)[45−47] starting from the snapshots generated by nt-PaCS-MD and to detect conformational states as free energy local minima. This procedure has been already conducted in our previous studies.[12,35,48,49] Conformational states that are detected can be candidates for further investigations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kitao@iam.u-tokyo.ac.jp.

Funding

This research was supported by a Japan Society for the Promotion of Science (JSPS) Research Fellowship for Young Scientists (no. 15J03797) and a Grant-in-Aid for Young Scientists (B) from JSPS (no. 26840056) to R.H., by a Grant-in-Aid in Innovative Areas (no. 25104002) and a Grant-in-Aid for Science Research B (no. 15H04357) to A.K. from JSPS and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and by Social and Scientific Priority Issues to be tackled using the Post K computer (achievement of a society that provides health and longevity, innovative drug discovery infrastructure through functional control of biomolecular systems) to A.K. from MEXT.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.H. would like to thank Drs. Kazuhiro Takemura, Yasuhiro Matsunaga, and Yu Yamamori for helpful discussions. The computations were partly performed using ACCC/RICC of RIKEN.

REFERENCES

(1) Frauenfelder, H.; Parak, F.; Young, R. D. Conformational Substates in Proteins. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 451−79.

(2) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science* **1991**, *254*, 1598−603.

(3) Kitao, A.; Hirata, F.; Go, N. The Effects of Solvent on the Conformation and the Collective Motions of Protein - Normal Mode Analysis and Molecular-Dynamics Simulations of Melittin in Water and in Vacuum. *Chem. Phys.* **1991**, *158*, 447−472.

(4) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. Effect of Solvent on Collective Motions in Globular Protein. *J. Mol. Biol.* **1993**, *234*, 1207−17.

(5) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412−25.

(6) Shea, J. E.; Brooks, C. L., III From Folding Theories to Folding Proteins: A Review and Assessment of Simulation Studies of Protein Folding and Unfolding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.

(7) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65.

(8) Schlitter, J.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. Targeted Molecular-Dynamics Simulation of Conformational Change - Application to the T[−]R Transition in Insulin. *Mol. Simul.* **1993**, *10*, 291−308.

(9) Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered Molecular Dynamics Investigations of Protein Function. *J. Mol. Graphics Modell.* **2001**, *19*, 13−25.

(10) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224−230.

(11) Leech, J.; Prins, J. F.; Hermans, J. Smd: Visual Steering of Molecular Dynamics for Protein Design. *IEEE Comput. Sci. Eng.* **1996**, *3*, 38−45.

(12) Harada, R.; Kitao, A. Parallel Cascade Selection Molecular Dynamics (Pacs-Md) to Generate Conformational Transition Pathway. *J. Chem. Phys.* **2013**, *139*, 035103.

(13) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein Structural Change Upon Ligand Binding: Linear Response Theory. *Phys. Rev. Lett.* **2005**, *94*, 078102.

(14) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.

(15) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition Path Sampling and the Calculation of Rate Constants. *J. Chem. Phys.* **1998**, *108*, 1964−1977.

(16) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv. Chem. Phys. Vol 123* **2002**, *123*, 1−78.

(17) Suenaga, A.; Narumi, T.; Futatsugi, N.; Yanai, R.; Ohno, Y.; Okimoto, N.; Taiji, M. Folding Dynamics of 10-Residue Beta-Hairpin Peptide Chignolin. *Chem. - Asian J.* **2007**, *2*, 591−598.

(18) Kitao, A. Transform and Relax Sampling for Highly Anisotropic Systems: Application to Protein Domain Motion and Folding. *J. Chem. Phys.* **2011**, *135*, 045101.

(19) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.

(20) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz,K. M.;Wang,B.;Hayik,S.;Roitberg,A.;Seabra,G.;Kolossvaŕy, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 11*; University of California: San Francisco, CA, 2010.

(21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.

(22) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(23) Miyamoto, S.; Kollman, P. A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(24) Zhang, X. J.; Matthews, B. W. Conservation of Solvent-Binding Sites in 10 Crystal Forms of T4-Lysozyme. *Protein Sci.* **1994**, *3*, 1031−1039.

(25) Weaver, L. H.; Matthews, B. W. Structure of Bacteriophage-T4 Lysozyme Refined at 1.7 a Resolution. *J. Mol. Biol.* **1987**, *193*, 189−199.

(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(27) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(28) Loeffler, H. H.; Kitao, A. Collective Dynamics of Periplasmic Glutamine Binding Protein Upon Domain Closure. *Biophys. J.* **2009**, *97*, 2541−2549.

(29) Hsiao, C. D.; Sun, Y. J.; Rose, J.; Wang, B. C. The Crystal Structure of Glutamine-Binding Protein from Escherichia Coli. *J. Mol. Biol.* **1996**, *262*, 225−242.

(30) Sun, Y. J.; Rose, J.; Wang, B. C.; Hsiao, C. D. The Structure of Glutamine-Binding Protein Complexed with Glutamine at 1.94

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502

Angstrom Resolution: Comparisons with Other Amino Acid Binding Proteins. *J. Mol. Biol.* **1998**, *278*, 219−229.
(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. A. Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(32) 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. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(33) Mu, Y. G.; Yang, Y.; Xu, W. X. Hybrid Hamiltonian Replica Exchange Molecular Dynamics Simulation Method Employing the Poisson-Boltzmann Model. *J. Chem. Phys.* **2007**, *127*, 084119.
(34) Satoh, D.; Shimizu, K.; Nakamura, S.; Terada, T. Folding Free-Energy Landscape of a 10-Residue Mini-Protein, Chignolin. *FEBS Lett.* **2006**, *580*, 3422−3426.
(35) Harada, R.; Kitao, A. Exploring the Folding Free Energy Landscape of a Beta-Hairpin Miniprotein, Chignolin, Using Multiscale Free Energy Landscape Calculation Method. *J. Phys. Chem. B* **2011**, *115*, 8806−8812.
(36) Wang, J. M.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (Resp) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(37) Sakuraba, S.; Joti, Y.; Kitao, A. Detecting Coupled Collective Motions in Protein by Independent Subspace Analysis. *J. Chem. Phys.* **2010**, *133*, 185102.
(38) de Groot, B. L.; Hayward, S.; van Aalten, D. M. F.; Amadei, A.; Berendsen, H. J. C. Domain Motions in Bacteriophage T4 Lysozyme: A Comparison between Molecular Dynamics and Crystallographic Data. *Proteins: Struct., Funct., Genet.* **1998**, *31*, 116−127.
(39) Torrie, G. M.; Valleau, J. P. Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578−581.
(40) Torrie, G. M.; Valleau, J. P. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(41) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
(42) Pan, A. C.; Roux, B. Building Markov State Models Along Pathways to Determine Free Energies and Rates of Transitions. *J. Chem. Phys.* **2008**, *129*, 064107.
(43) Prinz, J. H.; Keller, B.; Noe, F. Probing Molecular Kinetics with Markov Models: Metastable States, Transition Pathways and Spectroscopic Observables. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912−16927.
(44) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(45) Ferrenberg, A. M.; Swendsen, R. H. Optimized Monte-Carlo Data-Analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules 0.1. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(47) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(48) Harada, R.; Kitao, A. Multi-Scale Free Energy Landscape Calculation Method by Combination of Coarse-Grained and All-Atom Models. *Chem. Phys. Lett.* **2011**, *516*, 113−113.
(49) Harada, R.; Kitao, A. The Fast-Folding Mechanism of Villin Headpiece Subdomain Studied by Multiscale Distributed Computing. *J. Chem. Theory Comput.* **2012**, *8*, 290−299.
(50) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.jctc.5b00723
J. Chem. Theory Comput. 2015, 11, 5493−5502