

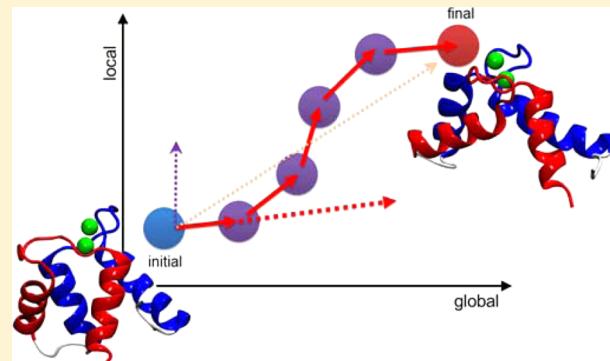
Linear Response Path Following: A Molecular Dynamics Method To Simulate Global Conformational Changes of Protein upon Ligand Binding

Koichi Tamura and Shigehiko Hayashi*

Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

Supporting Information

ABSTRACT: Molecular functions of proteins are often fulfilled by global conformational changes that couple with local events such as the binding of ligand molecules. High molecular complexity of proteins has, however, been an obstacle to obtain an atomistic view of the global conformational transitions, imposing a limitation on the mechanistic understanding of the functional processes. In this study, we developed a new method of molecular dynamics (MD) simulation called the linear response path following (LRPF) to simulate a protein's global conformational changes upon ligand binding. The method introduces a biasing force based on a linear response theory, which determines a local reaction coordinate in the configuration space that represents linear coupling between local events of ligand binding and global conformational changes and thus provides one with fully atomistic models undergoing large conformational changes without knowledge of a target structure. The overall transition process involving nonlinear conformational changes is simulated through iterative cycles consisting of a biased MD simulation with an updated linear response force and a following unbiased MD simulation for relaxation. We applied the method to the simulation of global conformational changes of the yeast calmodulin N-terminal domain and successfully searched out the end conformation. The atomistically detailed trajectories revealed a sequence of molecular events that properly lead to the global conformational changes and identified key steps of local–global coupling that induce the conformational transitions. The LRPF method provides one with a powerful means to model conformational changes of proteins such as motors and transporters where local–global coupling plays a pivotal role in their functional processes.



INTRODUCTION

Proteins respond to external stimuli and change their conformations in molecular functional processes. Binding of ligands to proteins is ubiquitously found for initiating protein functions involving large structural changes such as motors and transporters. Multiscale correlation of ligand binding, which is a relatively local event at a ligand binding site, with global protein conformational changes is therefore a key mechanism of the molecular functions. The protein conformational changes induced by ligand binding can be observed in an atomic detail through approaches of structural biology such as crystallographic diffraction and nuclear magnetic resonance (NMR) measurements. However, those approaches only resolve molecular structures in stable conformations in most cases and thus have difficulty in capturing short-lived intermediates during the conformational changes such as transition states which determine kinetics of conformational changes and thus directly represent the functional correlation. Despite the increasing availability of protein structures before and after the conformational changes upon the binding of ligands, little is known about a transition path connecting the two end states. Furthermore, in many cases, only one of the two end structures

is available due to experimental difficulties. One therefore needs to model the other end structure together with a transition path to examine the complete functional process.

Molecular dynamics (MD) simulations with atomistic force fields enable one to analyze the functional processes in molecular detail. MD simulations with a special purpose computer^{1,2} have shown that functional conformational changes involving spontaneous transitions via short-lived intermediates can be observed at atomic resolution in the course of a long-time trajectory on microto-millisecond time scale.^{3,4} However, MD simulations on commodity computational resources still suffer from difficulty in accessing to the time scale on which slow conformational changes of proteins in functional processes are fully simulated in an unbiased way.

To circumvent the limitation, several methodologies of biased MD simulations which accelerate conformational changes by applying biasing forces to atoms have been developed. The biased MD simulations can be classified into two approaches with nondirected^{5–12} and directed^{13–15} biases.

Received: February 9, 2015

In the MD simulations with nondirected biases, such as accelerated MD¹⁰ and metadynamics,¹² the waiting time for transitions between conformational substates is reduced by nondirected biasing forces that enhance escape from free energy basins of conformational substates. The methods therefore allow one to sample more conformational substates in the limited time of the simulations. However, nondirected search of very extensive configuration space of the system to identify conformational changes of overall functional processes still requires a longer time trajectory calculation; in general, simulation time to search out the final conformation in a nondirected manner is expected to be proportional to $r_{\text{RMSD}}^{\text{Nf}}$, where r_{RMSD} is change in root-mean-square deviation (RMSD) associated with the conformational changes, and N_f is the number of reaction coordinates that are involved in the conformational changes since the trajectory needs to evenly travel the entire energetically accessible region to find out the escaping path from the trapped free energy basin without any directed biases. Hence, the simulation time of the nondirected search required quickly increases when r_{RMSD} and N_f become large, and consequently, nondirected searches are less efficient to find the final conformation undergoing larger and more complex conformational changes. In order to circumvent the difficulty, nondirected simulations are often employed with path collective variables (PCVs) that reduce the reaction coordinates considered, i.e., N_f .¹⁶ Metadynamics¹² with PCVs have successfully been applied to simulation studies of large conformational changes of protein.^{17–19} However, in order to determine proper collective variables, details of both of the end structures of the conformational changes need to be known a priori.

The MD simulations with directed biases, however, provide means to simulate functional conformational changes more efficiently. In the simulations, forced conformational changes are biased along reaction coordinates introduced based on knowledge obtained experimentally. Figure 1 depicts a schematic view of the reaction coordinates introduced in representative methodologies with directed biases (see below). The knowledge-based directed biases reduce configuration space that needs to be searched, enabling one to induce conformational changes in shorter trajectory calculations. However, the predetermined restriction of configuration space search gives rise to a drawback; the simulations are not free from the possibility of inducing false conformational changes.

The most widely used method with a directed bias is targeted molecular dynamics (TMD). In a TMD simulation, conformational changes are forced in a direction toward a target structure of a functional process in the configuration space (see Figure 1). The TMD simulation therefore ensures that the conformational changes take place in a short trajectory calculation. However, obviously, the TMD method cannot be applied when one of the two end structures of the conformational changes is not known. Furthermore, TMD simulations suffer from artifacts of “large-scales-first” phenomenon²⁰ due to the unnaturally directed bias, although Gur et al. very recently developed a MD method in which a target conformation for a TMD simulation is sequentially updated to circumvent the problem of the conventional TMD.²¹

Steered molecular dynamics (SMD) is a method that utilizes more naturally directed biases. In the SMD simulations, directed biasing forces that represent relevant molecular events of functional processes are applied. The method has been

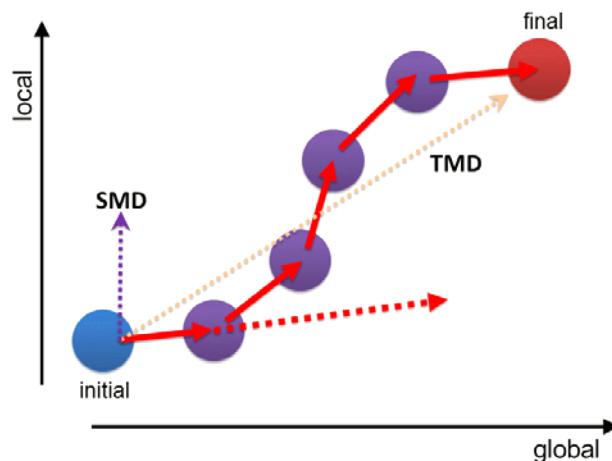


Figure 1. Schematic explanation of LRPF. The whole configuration space is described schematically. The horizontal (vertical) axis corresponds to the coordinate that represents global (local) motion. The two end states are represented as filled blue or red circles. Using the LRT displacement vector as a guiding force, the system is pulled to the adjacent substate (filled purple circles). Since there exists significant nonlinearity, the LRT vector does not point toward the final state (broken red arrows). Thus, the guiding force should be updated, and the procedure be iterated. The broken purple (orange) arrow represents the direction of the external force calculated in SMD (TMD) simulations.

successful to simulate ligand diffusion inside proteins and mechanical unfolding.²² In a simulation of global protein conformational changes induced by ligand binding, biasing forces that represent the ligand binding could be applied locally at the binding site of the protein without detailed knowledge of the final structure (Figure 1). However, weak coupling between the local ligand binding and global protein conformation changes requires a long time MD simulation with small biasing forces where slow response of global conformational changes to the local biasing forces needs to occur spontaneously. An attempt to accelerate the conformational changes by increasing biasing forces would lead to false unfolding at the local site.

Fortunately, the global conformational response to ligand binding, a key phenomenon in functional processes, can be predicted by a linear response theory (LRT) proposed by Ikeguchi et al.²³ LRT identifies coupling of the lowest order (Gaussian) fluctuation of protein in an equilibrium state with perturbation of ligand binding through very simple calculations. The LRT treatment has succeeded in reproducing experimentally observed global conformational changes of enzymes.²³ Note, however, that LRT is limited to the prediction of the protein response in a local region in the *configuration space*, even though predicted movement of atoms is globally distributed over an entire protein molecule in three-dimensional Cartesian space. Sole application of LRT therefore fails in predicting protein conformational changes that travel a long and nonlinear pathway in the configuration space (Figure 1).

In the present study, we have developed a biased MD method based on LRT, the linear response path following (LRPF) method, to simulate functional conformational changes of protein upon ligand binding without detailed knowledge of the final structure. In the LRPF method, a reaction coordinate in a local region in the configuration space is determined by LRT, and protein conformational transition between adjacent conformational substates in a local region in the configuration

space is efficiently induced by MD simulations with a directed biasing force along the local LRT reaction coordinate. The overall path of functional conformational changes is then obtained by iteratively performing the simulation cycles of the LRT conformational transition between adjacent conformational substates described above (Figure 1). The LRT treatment introduced in the biasing scheme overcomes the aforementioned difficulties in TMD and SMD by efficiently taking into account the multiscale coupling between local ligand binding and global conformational changes of protein with naturally directed biasing forces based on relevant functional events. The scheme of iterative update of conformational changes surmounts the local nature of the LRT prediction and enables one to identify the overall path of large, global, and complex conformational changes only with knowledge of local ligand binding sites.

We applied the LRPF method to the simulation of structural changes of yeast calmodulin N-terminal domain (CaMn) upon the binding of calcium ions, Ca^{2+} , as shown in Figure 2. CaMn

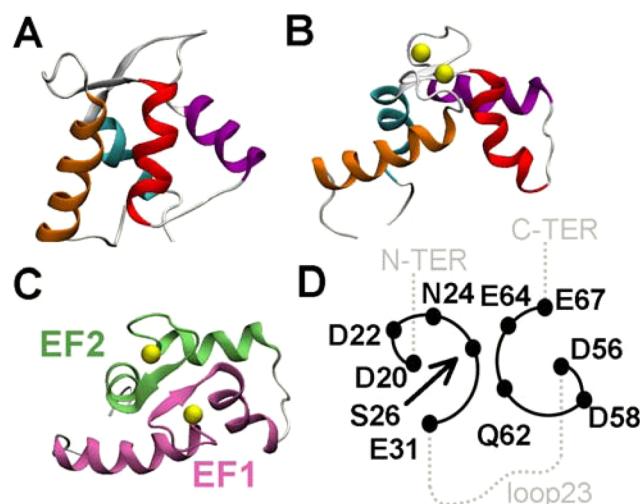


Figure 2. Structures of yeast CaMn. (A) CaMn in its closed apo state. (B) CaMn in its open holo state. Calcium ions are drawn as yellow spheres. (C) Two EF-hands are highlighted. They are linked by an antiparallel β -sheet. The middle loop of the EF-hand is a Ca^{2+} -binding loop, and protein's oxygen atoms at the loop capture Ca^{2+} atoms. (D) Schematic representation of CaMn. The protein backbone is drawn as a broken gray line, whereas calcium-binding loops are drawn as a solid black line. Each calcium-binding loop is composed of 12 residues (Asp20 to Glu31 in EF1 and Asp56 to Glu67 in EF2). The sites of selected residues are described as filled black circles.

is an appropriate system for a test of the method because of its reasonable size of the protein (77 residues) and simple structure of ligands (ions). The method succeeded in searching out the correct end conformation of the open form from the initial closed conformation without detailed knowledge of the final structure in multiple trials of MD simulations for submicro to microseconds. The transition paths were shown to significantly deviate from that obtained by TMD and uncover the complex nature of the conformational changes. The atomistic view of the conformational changes obtained by the LRPF simulations shed light on the molecular mechanism of the multiscale and nonlinear coupling between the local ligand binding and the global conformational changes of protein in the functional processes.

METHODOLOGICAL BACKGROUND

Overall Scheme. The LRPF method exploits the LRT by Ikeguchi et al.²³ to predict local conformational transitions involved in overall conformational changes of a functional process initiated by ligand binding. LRT showed that a protein's response to perturbation from ligands can be formulated as

$$\Delta\mathbf{r}_i \simeq \beta \sum_j C_{ij} \mathbf{G}_j \quad (1)$$

where $\Delta\mathbf{r}_i$ denotes a displacement vector of a protein atom i , β is the inverse temperature ($= 1/k_B T, k_B$ is the Boltzmann constant, and T is the absolute temperature), and C_{ij} is the variance–covariance matrix of the protein in the absence of the perturbation,

$$\mathbf{C} = \langle (\mathbf{r} - \mathbf{r}_{\text{ave}})(\mathbf{r} - \mathbf{r}_{\text{ave}})^T \rangle \quad (2)$$

where \mathbf{r} and $\mathbf{r}_{\text{ave}} = \langle \mathbf{r} \rangle$ denote protein coordinates and their averages, respectively, and $\langle \rangle$ indicates thermal average of the protein conformations without the perturbation. \mathbf{G}_j is the force exerted by the ligands on the protein atom j . Through the variance–covariance matrix, the LRT formula is capable of effectively taking into account coupling between the local perturbation of ligand binding and spatially distributed thermal fluctuation of protein, and consequently the protein response often appears to be global and collective in motion. However, the LRT relationship is valid only when the perturbation is weak and the protein exhibits Gaussian fluctuation in a linear response regime. Those conditions are satisfied only for structural changes in a local region around a stable conformation in the configuration space. As mentioned above, therefore, sole application of LRT for the prediction of conformational changes in functional processes, which often involve a nonlinear sequence of conformational transitions, is not appropriate. In fact, the overall structural changes of CaMn could not be described by LRT (see Figure S1C in Supporting Information). In the present LRPF scheme, therefore, LRT is utilized for predicting a local transition between neighboring conformational substates, and the overall conformational changes are simulated by a sequence of the local conformational transitions (Figure 1).

The LRT prediction is utilized to determine biasing forces that enhance local conformational transitions. We introduce the biasing force directed along a local reaction coordinate defined by the displacement vector of the LRT prediction, eq 1:

$$\mathbf{F}_{i,\text{LR}} = \alpha \Delta\mathbf{r}_i \quad (3)$$

where α is a scaling factor which determines the strength of the biasing force (see below). The total force acting on atom i , \mathbf{F}_i , in the biased MD simulation is given by a sum of the force originating from force field, $\mathbf{F}_{i,\text{forcefield}}$, and the biasing force along a local LRT reaction coordinate, $\mathbf{F}_{i,\text{LR}}$, which is called the LR force hereafter:

$$\mathbf{F}_i = \mathbf{F}_{i,\text{forcefield}} + \mathbf{F}_{i,\text{LR}} \quad (4)$$

As mentioned above, LRT can predict global protein response upon ligand binding that involves collective motion and thus exhibits slow relaxation. The LR force therefore stimulates the protein's slow collective motion coupled with ligand binding in the transition between conformational substates. In the present study, we applied the LR force only to backbone chain atoms of the protein, i.e., N and C atoms in

amides, and C_α atoms. Accordingly, we constructed the variance–covariance matrix and the perturbative force only for those atoms of the protein (see below). Although, in principle, the LR force can be evaluated for any set of protein atoms, it is less favorable to include atoms of side chains that exhibit nonlinear fluctuation in the evaluation of the LR force because of the linear response condition of the method.

The LRPF search of the conformational changes is constituted by iterative cycles of the conformational transition between adjacent conformational substates (Figure 1). The update of the local reaction coordinate with LRT introduces naturally directed biasing forces and overcomes difficulties in SMD and TMD as described above. Each of the iterative cycles consists of three sequential phases of MD simulations. In the first phase, an equilibrium simulation to obtain a LR force is carried out. For the evaluation of a LR force, the variance–covariance matrix, C_{ij} , and the perturbative force, \mathbf{G}_j , in the LRT formula, eq 1, are computed with conformational samples of protein obtained by the equilibrium MD simulation (see below). The LR force evaluated with eqs 1 and 3 is employed in a subsequent biased MD simulation in the second phase to induce conformational transition. In the last phase, the protein structure altered by the preceding biased simulation is relaxed and equilibrated in a conformational substate by a nonbiased MD simulation, which is prerequisite for the equilibrium MD simulation in the first phase of the next iteration.

Perturbative Force. For the perturbative force, \mathbf{G}_i , in eq 1, we employed a mean force acting from ligands, $\langle \mathbf{G}_i \rangle$, evaluated with conformational samples taken from an equilibrium MD simulation in the first phase. The perturbative force was derived from the force field and included contributions of Coulombic and Lennard-Jones interactions between ligand Ca^{2+} ions and the protein atoms. The force vectors at the protein atoms point along the directions between the protein atoms and the positions of Ca^{2+} ions. Since we employ the LR force acting on the peptide chain atoms as described above, the force components acting on atoms attached to the peptide chain atoms are gathered and added to the corresponding peptide chain atoms as follows:

$$\bar{\mathbf{G}}_i = \langle \mathbf{G}_{i,\text{self}} \rangle + \sum_{j \in I(i)} \langle \mathbf{G}_j \rangle \quad (5)$$

$\bar{\mathbf{G}}_i$ is the contracted force acting on the backbone chain atom i , and $\mathbf{G}_{i,\text{self}}$ is a perturbative force directly acting on atom i . The second term in the right-hand side sums up force components of atoms attached to the backbone chain atom, i.e., atoms of side chains and H_α atom for C_ω H and O atoms for N and C in an amide group, respectively. Interactions between atoms other than the backbone chain atoms and ligands are therefore effectively included.

Variance–Covariance Matrix. The variance–covariance matrix, C_{ij} , of the backbone chain atoms used in eq 1 is computed with conformational samples obtained from an equilibrium MD simulation in the first phase of the LRPF scheme. In the calculation of the covariance matrix, translational and rotational motions of the protein were removed from the trajectory. It is evident that statistical convergence of the variance–covariance matrix requires sufficient conformational samples. However, it was found that convergence behavior of the variance–covariance matrix remains unstable by straightforward statistical averaging with conformational samples of a longer MD trajectory. The instability arises from characteristic dynamics of protein such as intermittent and nonlinear

conformational fluctuation appearing in a longer MD trajectory, which violate the linear response (Gaussian fluctuation) condition in which the variance–covariance matrix is physically relevant. To circumvent the difficulty, we introduced a sliding window function which limits the statistical average to be taken with statistical samples in a narrow time window and thus diminishes long-time correlation of intermittent and nonlinear fluctuation. In the present study, we employed a Hanning window function of sample size N ,

$$w_l^k = 0.5 - 0.5 \cos \frac{2\pi(k-l)}{N-1} \quad (6)$$

where k and l are indices of conformational samples in a time series of the trajectory. The weighted statistical average in each window is calculated as

$$\mathbf{r}_{i,\text{ave}} = \frac{\sum_{k=0}^{N-1} w_l^k \mathbf{r}_k}{\sum_{k=0}^{N-1} w_l^k} \quad (7)$$

where $\mathbf{r}_{i,\text{ave}}$ is the weighted average in the l^{th} window, and \mathbf{r}_k is the k^{th} conformational sample of the trajectory. A variance–covariance matrix of each window was then constructed as

$$\mathbf{C}^i = \langle (\mathbf{r} - \mathbf{r}_{i,\text{ave}})(\mathbf{r} - \mathbf{r}_{i,\text{ave}})^T \rangle_i \quad (8)$$

where $\langle \cdot \rangle_i$ denotes the average taken over the i^{th} window. Finally, the total covariance matrix is obtained as the mean value of \mathbf{C}^i 's, i.e.,

$$\mathbf{C}^{\text{tot}} = \frac{1}{M} \sum_{i=1}^M \mathbf{C}^i \quad (9)$$

where M is the number of samples in the trajectory.

Biasing Force. Together with the perturbative force, eq 5, and the variance–covariance matrix, eq 9, described above, the LR biasing force, eq 3 based on eq 1, is evaluated as

$$\mathbf{F}_{i,\text{LR}} = \alpha \beta \sum_j C_{ij}^{\text{tot}} \bar{\mathbf{G}}_j \quad (10)$$

Now, we determine the scaling factor, α , which controls the strength of the biasing force. For that purpose, we introduced a rough estimate of work applied by the biasing force:

$$W = \sum_i (|\mathbf{F}_{i,\text{LR}}| \times d) \quad (11)$$

where d represents RMSD of the backbone chain atoms undergoing forced displacement in the biased MD simulation with the LR force. In the current study, α is set to a value in the biased MD simulation to give $W = 10\text{--}30$ kcal/mol for $d = 3 \text{ \AA}$. We estimate a relative magnitude of the biasing force with W for $d = 3 \text{ \AA}$ hereafter.

Biased MD Simulation. After obtaining the LR force in the first phase of the LRPF scheme, a biased MD simulation with the LR force is carried out in the second phase. The magnitude of the biasing force was gradually raised by $W = 1$ kcal/mol every 200 ps. When W grows enough, the biasing force is kept constant, and the biased MD simulation is continued until RMSD reaches a plateau. Then, the biasing force is gradually decreased by $W = 1$ kcal/mol every 200 ps until the biasing force vanishes. After the biased MD simulation in the second phase, a nonbiased MD simulation is performed for relaxation and equilibration in the third phase. A cycle of the LRPF iteration ends when RMSD reaches a plateau in the third phase,

and the first phase of the next cycle then begins. Computational details of the MD simulations are described in Materials and Methods.

RESULTS

Conformational Changes Observed in Experiments.

Before describing simulation results, we briefly summarize the conformational changes of CaMn observed in experiments. In its ligand unbound form (apo-CaMn, PDB id: 1F54²⁴), CaMn takes a closed structure (Figure 2A). The protein binds two ligand Ca^{2+} at its characteristic binding sites in two EF-hands, EF1 and EF2, respectively,²⁵ and undergoes considerable structural changes (RMSD = 5.1 Å) to an open conformation where its hydrophobic patch is exposed (Figure 2B, PDB id: 1F55²⁴). The EF-hand (Figure 2C) in CaMn is a helix-loop-helix motif, and the middle loop is a Ca^{2+} -binding loop which consists of 12 amino acids (Asp20 to Glu31 in EF1 and Asp56 to Glu67 in EF2). The closed-open conformational transition involves a characteristic movement of the side chain of His61, which is situated inside of the Ca^{2+} -binding loop in the closed conformation, and flips away and moves to the outside of the protein in the formation of the open one.

Overview of MD Simulations. In the present study, we performed 17 MD simulations summarized in Table 1. All

Table 1. Summary of Simulations

simulation	for what	duration (ns)
Eq-prep	preparatory equilibration of closed holo-CaMn	120
Eq-long	long time equilibration of closed holo-CaMn	3,880
Eq-open	equilibration of open holo-CaMn	300
LRPF1	biased MD	572
LRPF2	biased MD	1,254
LRPF3	biased MD	2,173
LRPF4	biased MD	1,650
TMD	targeted molecular dynamics simulation	500
SMD	steered molecular dynamics simulation	720
aMD1	accelerated molecular dynamics simulation	100
aMD2	accelerated molecular dynamics simulation	100
aMD3	accelerated molecular dynamics simulation	100
aMD4	accelerated molecular dynamics simulation	200
aMD5	accelerated molecular dynamics simulation	500
aMD5_t2	accelerated molecular dynamics simulation	500
aMD5_t3	accelerated molecular dynamics simulation	500
aMD5_t4	accelerated molecular dynamics simulation	500
		total 13.669 μ s

calculations were carried out with bound Ca^{2+} ions. First, after a preparatory equilibrium MD simulation (120 ns, Eq-prep), we carried out nonbiased MD simulations for both of the closed and open conformations (Eq-long and Eq-open, respectively). We examined spontaneous conformational changes from the closed conformation in Eq-long simulation for 3.88 μ s. The nonbiased MD simulations also defined structures of the initial (closed) and final (open) states of the conformational transition in the use of the force field adopted in the present study. From different initial structures of the closed conformation sampled in the preparatory equilibrium MD simulation, four independent LRPF biased MD simulations (LRPF1–4) for search of the conformational transitions to the open conformation are carried out. The multiple trials of the conformational transition search confirm the stability of the

search scheme. We also performed other biased MD simulations, SMD,^{14,15} TMD,¹³ and accelerated MD (aMD),¹⁰ for the same conformational transition to evaluate the performance of the present method. Computational details of the MD simulations are given in Materials and Methods.

Preparatory Equilibrium MD Simulation of Closed Holo-CaMn.

The preparatory equilibrium MD simulation (Eq-prep) of a closed holo-CaMn system, which was modeled from a NMR structure of closed apo-CaMn²⁴ by manually adding Ca^{2+} ions, was performed (Figure 3). During equilibrium simulation for 120 ns, the protein was perturbed by the Ca^{2+} ions and changed the structure of Ca^{2+} -binding loops (Figures 3A and B). Consequently, the Ca^{2+} ions are coordinated by the protein atoms in the loops; the coordination numbers of the protein atoms to Ca^{2+} are 5 for EF1 and 2 or 3 for EF2 (see Materials and Methods for the computational definition of the coordination numbers). Although the protein backbone was also perturbed by the addition of Ca^{2+} , resulting in the significant conformational change (Figure 4A), the structure is still in a closed conformation; the hydrophobic patch remains embedded inside the protein.

Long-Time Equilibrium MD Simulation of Closed Holo-CaMn.

Starting from the last conformation of the preparatory simulation described above, a relatively long (3.88 μ s) equilibrium MD simulation (Eq-long) was performed (Figure S2 in Supporting Information). Significant conformational changes of Ca^{2+} -binding loops and reorganization of the coordination sphere of Ca^{2+} were observed, although the overall protein conformation was kept closed.

First, Asp56 bound to Ca^{2+} in a monodentate manner in EF2 at ~2,045 ns (Figures 3 and S2F and S3 in Supporting Information). Subsequently, the loop conformation around His61 changed, and the side chain of His61 was flipped away. Accompanying the structural changes in EF2, helix 3 (the helices are numbered from the N-terminus) underwent a slight tilt as seen in Figure 4A. The abrupt RMSD changes in Figures 4C and S2C in Supporting Information correspond to the conformational changes of EF2 and helix 3. Then, Asp58 bound to Ca^{2+} at ~2,260 ns (Figure S3 in Supporting Information). After that, the coordination sphere of Ca^{2+} in EF2 remains stable until the end of the simulation, giving the coordination number of 5 or 6 (Figure S2F in Supporting Information). As a consequence of those events in Eq-long simulation, the conformation of EF2 became similar to that observed in the NMR holo structure (Figure 3). However, the complete coordination of the EF2 loop observed in the NMR holo structure is not achieved yet; Gln62 did not participate in the coordination during Eq-long simulation (Figures 3C and S3 in Supporting Information) and needs to bind to Ca^{2+} to complete the coordination sphere (Figure 3E).

A significant conformational transition also happened in EF1 at ~3 μ s (Figures 3 and S2E in Supporting Information). The transition was initiated by dissociation of Ser26 from Ca^{2+} in EF1 (Figure S4 in Supporting Information), which was soon followed by a coordination switch of Asp20 from a monodentate coordination to a bidentate one (Figure S4 in Supporting Information), retaining the coordination number in EF1. Then, the N-terminal part of the EF1 Ca^{2+} -binding loop underwent rather complicated structural changes, resulting in the binding of Asn23 to Ca^{2+} at ~3,095 ns, binding of Glu31 in a bidentate manner at around 3,110 ns, and dissociation of Asn23 from Ca^{2+} at ~3,175 ns (Figure S4 in Supporting Information). Conformation of the binding loop and the

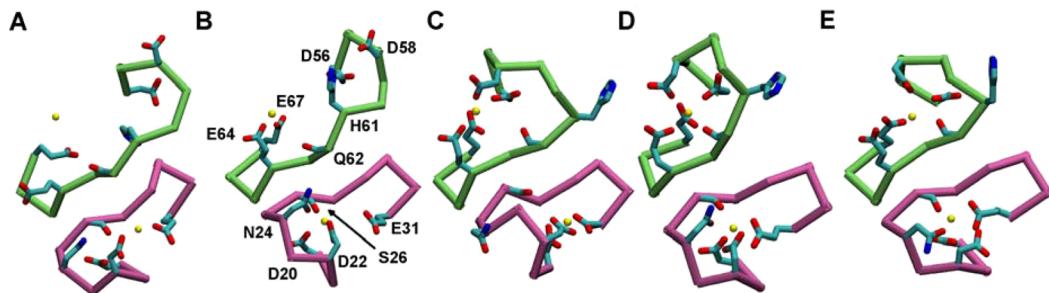


Figure 3. Structural changes in Ca^{2+} -binding loops. (A) A NMR structure of closed apo-CaMn (PDB id: 1F54). (B–D) The last snapshots of Eq-prep (B), Eq-long (C), and Eq-open (D). (E) A NMR structure of open holo-CaMn (PDB id: 1F55). In A, calcium ions were manually added to the NMR structure of closed apo-CaMn. Only calcium-binding loops and relevant residues are shown. Hydrogen atoms are not shown for clarity. Protein backbones drawn in purple and green are EF1 and EF2, respectively. Carbon, oxygen, and nitrogen atoms in the side chains as well as the main chain carbonyl groups of Ser26 and Gln62 are colored in cyan, red, and blue, respectively. Calcium ions are drawn as yellow spheres.

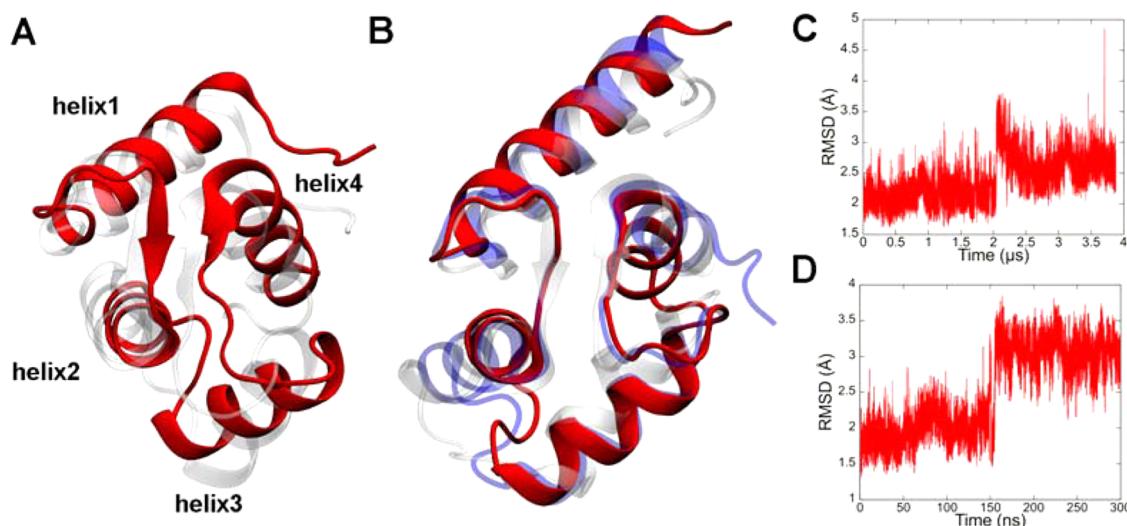


Figure 4. Structural changes in the equilibrium simulations. (A) The average structure of Eq-prep (red) superimposed to the NMR apo-CaMn structure (transparent white). (B) The average structure of the partially open state (red) superimposed to that of the open state (transparent blue) and the NMR open structure (transparent white). (C) RMSD change in Eq-long with respect to the NMR closed apo structure. (D) RMSD change in Eq-open with respect to the NMR open structure.

coordination sphere after the conformational changes were stable until the end of the Eq-long simulation. In the last phase, the final coordination number of Ca^{2+} in EF1 fluctuated between 5 and 6 (Figure S2E in Supporting Information). Similar to EF2 described above, the conformation of EF1 approached that observed in the NMR holo structure, although the Ca^{2+} coordination sphere is yet to be completed (Figure 3).

Despite the conformational rearrangements in EF1 and EF2 loops and the slight tilt of helix 3, the overall α helix bundle structure consisting of helices 1–4 were almost unchanged. The solvent accessible surface area (SASA) of hydrophobic residues of the protein did not increase during the simulation (Figure SSF in Supporting Information), indicating that the hydrophobic region between those helices is embedded in the protein and that the α helix bundle conformation stayed in the closed form.

Equilibrium MD Simulation of Open Holo-CaMn

Starting from the NMR solution structure of open holo-CaMn (PDB id: 1F55²⁴), an equilibrium MD simulation for 300 ns was performed (Eq-open). It was observed that the experimentally captured open holo structure underwent spontaneous transition to a partially open state at ~ 150 ns (Figure 4D). In the partially open state, the distance between

helices 2 and 4 shrinks, while helices 1 and 3 keep the open conformation, and the hydrophobic patch is still exposed to bulk water (Figure 4B). Although the SASA of hydrophobic residues decreased upon the transition to the partially open state, the SASA of the partially open state is still larger than that of the closed form in Eq-long (Figure S5 in Supporting Information). The transition from the open conformation to the partially open one accompanies significant movement of the hydrophobic side chain of Phe19, which has been suggested to play a role in the closed-to-open conformational changes.^{26–28} The coordination sphere of Ca^{2+} also exhibits slight differences from one in the NMR structure (Figures 3D and E). In EF1, Asp22 and Asn24 altered their orientation and gave the coordination number of 6 or 7. In EF2, Asn60 dissociated from Ca^{2+} , and alternatively, Glu64 bound to the ion, resulting in the coordination number of 6. The deviation from the NMR structure may be attributed to the limitation of the force field employed.

We therefore defined the partially open state, which was observed to be stable in the Eq-open simulation by the use of the force field adopted in the present study, as the final state of the closed-to-open transition simulated by the same force field. Even in the partially open state, the four helices of CaMn,

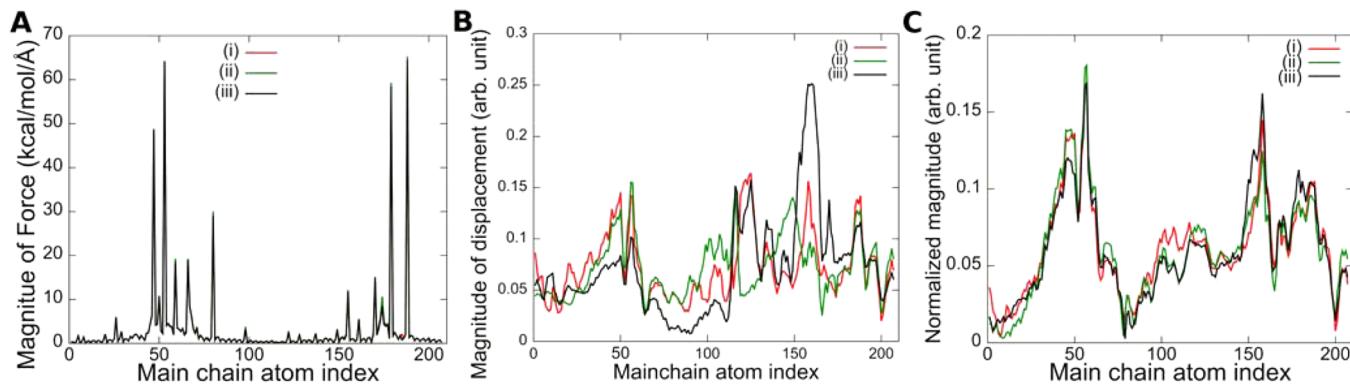


Figure 5. Convergence of the contracted mean force and LR biasing forces. (A) Magnitude of the contracted mean force. (B) Normalized magnitude of the LR biasing force without windowing and (C) with windowing. The following sampling periods in Eq-prep were used for the calculations: period (i), 90–100 ns; period (ii), 100–110 ns; period (iii), 110–120 ns.

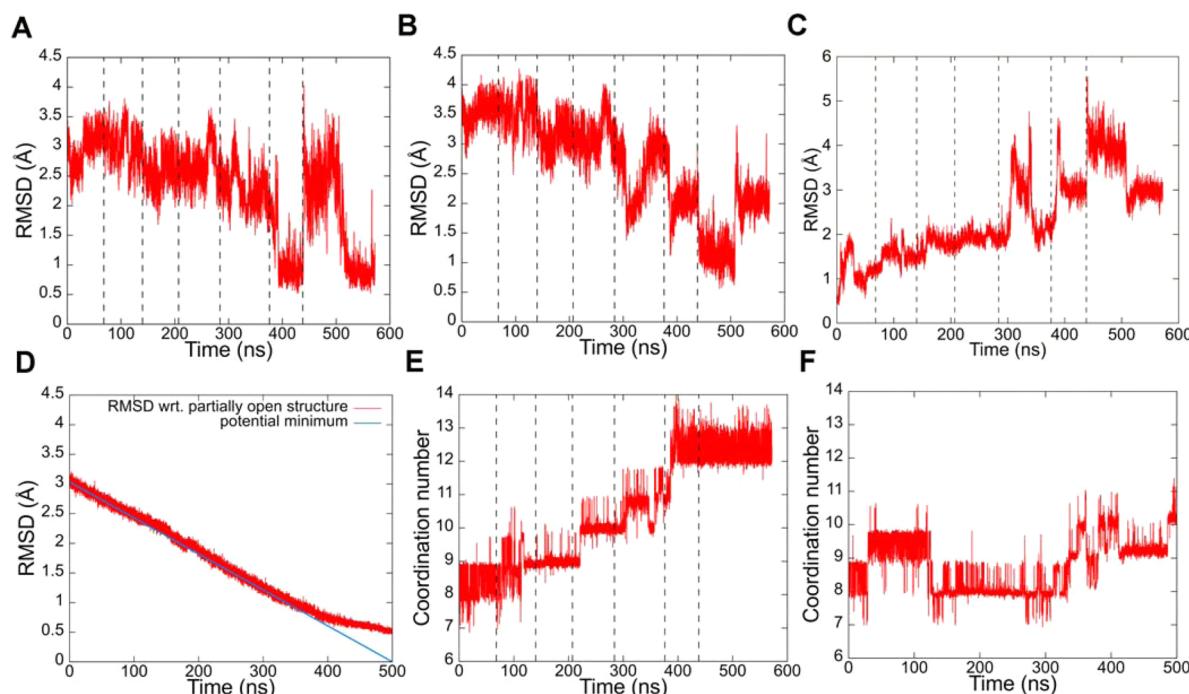


Figure 6. Temporal changes of RMSD and the coordination number of Ca^{2+} in LRPF1 and TMD simulations. (A) RMSD in LRPF1 with respect to the averaged partially open structure obtained from the Eq-open simulation. (B) RMSD in LRPF1 with respect to the averaged open structure obtained from the Eq-open simulation. (C) RMSD in LRPF1 with respect to the averaged closed structure obtained from the Eq-prep simulation. (D) RMSD in TMD with respect to the averaged partially open structure. The minimum of the guiding potential is also shown by a blue line. (E) Total coordination number of Ca^{2+} in LRPF1. (F) Total coordination number in TMD. Vertical dashed lines in the figures indicate the beginnings of the cycles of LRPF iterations.

especially helices 1 and 3, already form a significantly open conformation, which was not observed in the long-time nonbiased MD simulation (Eq-long). Since the purpose of the present study is development of the LRPF method, detailed examination of the transition between the open conformation and the partially open one is beyond the scope of the present study. It is noteworthy, nevertheless, that the fully open conformation of the four helices was successfully found during the LRPF search as well, although it relaxed spontaneously to the partially open conformation in the relaxation phase of the LRPF scheme (see below).

Perturbative and Biasing Forces of LRPF Simulations.

The LRPF simulation scheme includes construction of the LR biasing force, eq 10, obtained with the variance–covariance

matrix and the perturbative force. As described in Methodological Background, the statistical averaging with a sliding window function for the calculation of the variance–covariance matrix, eq 6–9, was introduced in the present study to circumvent the difficulty of its unstable statistical convergence. In order to assess the statistical convergence by the present approach, we calculated the contracted force, eq 5, and the LR biasing force, eq 10, with three independent samples from mutually nonoverlapping sampling periods (100,000 conformations from 10 ns each) of the last part of the Eq-prep trajectory (Figure 5). The statistical convergence of the contracted mean force is already sufficient with the 10 ns samples (Figure 5A). However, the LR biasing forces obtained without the windowing largely deviate from one another (Figure 5B),

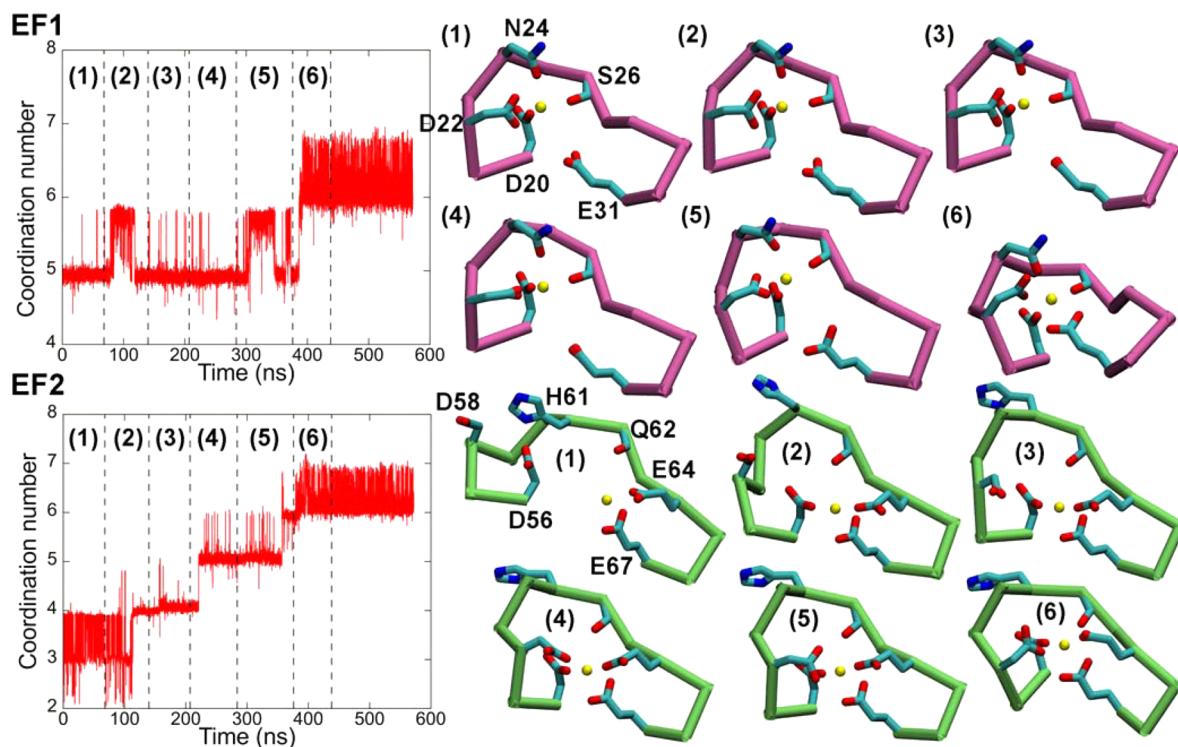


Figure 7. Conformational changes in Ca^{2+} -binding loops in LRPF1. Upper and lower panels show the conformational changes in the Ca^{2+} -binding loop of EF1 and EF2, respectively, in LRPF1. The figures on the left show the coordination numbers in EF-hands. Vertical dashed lines indicate the LRPF cycle as described in Figure 6. The average structures of the LRPF cycles 1–6 are depicted in the right panels. Each of the average structures is obtained with samples in the last 10 ns of each LRPF cycle. The color code is the same as that in Figure 3.

indicating the statistical instability of the variance–covariance matrix. The windowing with the 100 ps window width (i.e., 1,000 samples, see eq 6), however, successfully reduces the serious instability and drastically improves the convergence behavior of the LR biasing force (Figure 5C). It was also found that widening of the window width to 200 and 400 ps gives rise to worse convergence (Figure S1 in Supporting Information), indicating that the instability of the statistical sampling originates from slow and intermittent nonlinear fluctuation included more in the wider windows. In the present study, we adopted the averaging with the 10 ns sampling and the 100 ps window function for the calculation of the LR biasing forces throughout the entire LRPF search.

One may consider that, on the contrary, it is more beneficial to capture larger conformational fluctuations by including nonlinear fluctuations with a wider window to induce large conformational changes more efficiently. However, from a theoretical point of view, to express nonlinear fluctuations with second moments of fluctuation, i.e., linear fluctuations, of variance–covariance matrix is obviously erroneous. The second moments of fluctuation are largely overestimated when nonlinear fluctuation which involves much larger displacements of atoms is expressed as the linear fluctuation. Such erroneous behavior is observed for the biasing force (iii) shown in Figure 5B; the biasing force in the local region of the atom indices 150–170 is largely increased compared with other biasing forces in the different time regions and the converged ones. The overestimation of the biasing force is very unfavorable as it can lead to the unfolding of protein due to the overshooting of induced conformational changes. We therefore decided not to use the theoretically erroneous variance–covariance matrix to keep the stability of the method. Examination of the use of the

unconverged variance–covariance matrix, which could seriously damage the stability of the method, is beyond the scope of the present study and is left for future studies.

Figure 5A shows that the perturbative forces are localized on several groups of the protein. The localized forces originate from large Coulombic interactions of Ca^{2+} ions with the negatively charged moieties in the binding site of EF-hands and thus point toward Ca^{2+} ions. The perturbative forces therefore appropriately represent coordinations of the negatively charged moieties in EF-hands to Ca^{2+} ions in the Ca^{2+} binding process. However, the LR biasing forces shown in Figure 5C are widely distributed throughout the protein with broad peaks at EF-hands, indicating that the local–global coupling is taken into account in the biased simulations of the LRPF method.

LRPF Search of the Closed-to-Open Transition. Four independent LRPF simulations (LRPF1, LRPF2, LRPF3, and LRPF4) were performed. Each LRPF simulation started from a different conformation of the closed state sampled in Eq-prep (see Supporting Information). The final open conformations were successfully found in the transition searches of LRPF1 and LRPF2. In LRPF3 and LRPF4, conformational transitions were not completed in the simulation times due to traps in relatively stable conformational substates where key events were absent or took place in a different order.

In Figure 6, temporal changes of RMSDs and the coordination numbers of Ca^{2+} atoms during the successful closed-to-open transition in LRPF1 are shown. Detailed descriptions of the conformational changes are summarized in Table S1 (Supporting Information). The successful transition in LRPF2 is essentially the same as that in LRPF1 (see Supporting Information). As seen in Figure 6A, following a gradual decrease of RMSD to the conformation of the partially

open state, which is defined as the final state (see above), until the fifth cycle, the protein underwent large conformational change to the partially open state at 390 ns ($\text{RMSD} \approx 1 \text{ \AA}$). The LRPF search also found the fully open state during ~ 440 ns to ~ 510 ns (Figure 6B), although it spontaneously relaxed to the partially open state as observed in the Eq-open simulation (see above). In those final states, SASAs of hydrophobic residues increased to the values obtained for the open holo-CaMn seen above (Figure S5B in Supporting Information).

The closed-to-open transition simulated in the LRPF1 search accompanied the formation of the proper Ca^{2+} coordination structure. As seen in Figure 6E, the coordination number of Ca^{2+} gradually increases even in the closed conformation until 390 ns. During this period, the coordinations of EF-loops to Ca^{2+} were almost completed, except for those of Glu31 in EF1 and Gln62 in EF2, respectively. The developments of the coordination spheres in EF-loops are depicted in Figure 7 and Figures S3 and S4 in Supporting Information. The coordinations in EF1 except for Glu31 were already established in the Eq-prep simulation. In EF2, the coordination of Asp56 accompanied by a characteristic flip of His61 first occurred in ~ 200 ns, and then Asp58 participated in the coordination sphere (Figure S3 in Supporting Information). Finally, the remaining coordinations of Glu31 in EF1 and Gln62 in EF2 took place at the same moment of the opening conformational change at 390 ns, leading to the formation of the same coordination conformations in EF-loops as those observed in the Eq-open simulation (Figures 3D and 7).

In LRPF3 and LRPF4 simulations, the transition path searches were trapped at the conformational substates other than those appearing in the successful paths in LRPF1 and LRPF2, and consequently the closed-to-open conformational changes were not observed during the simulations for 1,873 ns and 1,350 ns, respectively, longer than those of LRPF1 and LRPF2 (Figures S6 and S7 in Supporting Information). In LRPF3, the backbone carbonyl group of Ser26 in EF1 dissociates from Ca^{2+} , and instead, Glu31 in EF1 established a stable coordination to Ca^{2+} in the closed conformation (Figures S4 and S8 in Supporting Information). Such rearrangement of the coordinations is not observed in the successful conformational changes in LRPF1 and LRPF2. Since the closed-to-open conformational transition is coupled with the coordination of Glu31 that completes the coordination sphere as described above, the rearrangement of the coordinations in EF1 by which the Glu31 binds to Ca^{2+} in the closed conformation before the other coordination is established loses the coupling between the completion of the coordination sphere and the closed-to-open conformational changes. Rearrangement of coordinations to Ca^{2+} in EF2 was also observed. Because of the coordination of the backbone carbonyl group of Gln62 in an early stage of the coordination sphere formation which is absent in LRPF1 and LRPF2 (Table S2 and Figure S3 in Supporting Information), coordination of Asp56 to Ca^{2+} took place without the characteristic flip of His61 observed in LRPF1 and LRPF2 (Figure S8 in Supporting Information). The lack of the flip of His61 seemingly hampered further development of the coordination sphere in EF2. LRPF4 simulation also involved the rearrangement of the coordinations in EF1 preventing the closed-to-open conformational change described above (Figures S4 and S9 in Supporting Information), although the EF2 loop formed the same coordination structure as that found in the closed conformation

before the transition to the open form in LRPF1 and LRPF2 (Figures S3 and S9 in Supporting Information). Those conformational substates in the last periods of LRPF3 and LRPF4 were stable during additional equilibrium MD simulations for 300 ns after the LRPF searches of LRPF3 and LRPF4 (Figures S6 and S7 in Supporting Information).

It should be noted that the conformational substates in LRPF3 and LRPF4 are neither unfolded states nor false final open states but can be regarded as intermediate states of the closed-to-open transition. The overall α helix bundle structures in the closed form are kept during the simulations. RMSDs (Figures S6 and S7 in Supporting Information) and SASAs (Figure S5 in Supporting Information) of LRPF3 and LRPF4 trajectories are comparable with those of Eq-long unbiased simulation where the closed form is maintained. The structural changes that characterize the conformational substates are local rearrangement of Ca^{2+} coordination from Ser26 to Glu31 in EF1 and flip of the side-chain of His61 in EF2 one as described above (Figures S8 and S9 in Supporting Information). Figure 8 depicts a schematic view of the

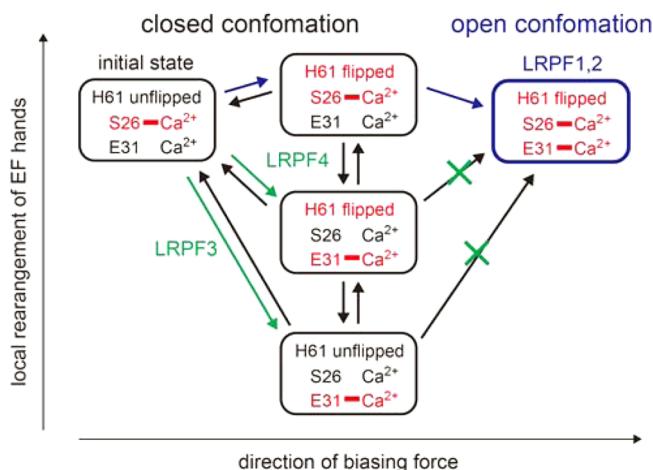


Figure 8. Schematic diagram of the conformational landscape observed in LRPF simulations.

conformational landscape of the transitions. Before the transition to the open form, several conformational substates exist in the closed form. Since those conformational changes are local, i.e., the conformational substates locate near one another in the configuration space, transitions between those substates are possible, and thus, the trajectory of the closed-to-open transition can travel among those intermediates.

TMD Simulation. The conformational change of protein was also simulated by the TMD method. We employed the last snapshot structure of Eq-prep and the average partially open structure as the starting structure and the target one of TMD simulation, respectively. The simulation time of TMD was set to 500 ns. The simulation time is comparable to that of LRPF1 which successfully searched out the target structure. Figure 6D shows time evolution of RMSD with respect to the target structure. Because of the biasing potential, eq 12, the minimum of which is moving linearly from the initial structure to the target one, the conformational changes constantly evolved along the minimum until ~ 300 ns. The linear feature of the RMSD change presents a contrast to that of LRPF1 and indicates a “large-scale-first” behavior of the conformational changes. Figure 6F shows the time evolution of the

coordination numbers of the protein atoms to Ca^{2+} in TMD. Unlike the successful LRPF trajectories, the side chains of the protein began to bind to Ca^{2+} ions at ~ 350 ns after the opening conformational changes have already developed to some extent. However, despite the opening conformational changes exerted by the biasing potential, some of the side chains failed to bind to Ca^{2+} in the simulation, and the coordination number did not reach the target value of 12 for the complete coordinations. Derivation from the linear decrease of RMSD in the last 200 ns shown in Figure 6D also represents the failure of the complete coordinations. Details of the coordination structures are presented in Supporting Information.

SMD Simulation of the Closed Holo-CaMn. We performed SMD simulations from the starting structures taken from the LRPF1 trajectory. We did not observe the opening global conformational changes induced in the simulation time, and even the simulations were started with conformations where the coordinations to Ca^{2+} ions in EF-hands were already developed to some extent. Moreover, despite the enhanced interactions between the protein and ligands, proper coordinations to Ca^{2+} were not observed. The details are found in Supporting Information.

aMD Simulation. We performed aMD simulations launched from pre-equilibrated snapshots of the closed conformation. Eight aMD simulations were carried out to adjust parameters employed for the acceleration of conformational changes and to assess reproducibility (see Supporting Information for details of the simulations). First, in four simulations with parameters determined based on those utilized in a previous study,²⁹ collapses of the secondary and tertiary structures of the protein were observed within 100–200 ns (Figures S21–S25, Supporting Information). The result indicates that acceleration with those parameters is too aggressive and thus induces false conformational changes. The parameters were then tuned to reduce the acceleration. Finally, we carried out four other simulations for 500 ns each starting at different pre-equilibrated snapshots but with the same parameters to assess reproducibility (Figures S26–30, Supporting Information). It should again be noted that, in the simulation time of 500 ns, LRPF1 succeeded in finding out the final open conformation. In the case of aMD simulations, however, we observed either collapses of the protein structure or the closed conformation maintained throughout the simulations, and consequently that none of the simulations led to the final structure in the open or partially open conformation in 500 ns. The observation of the coexistence of the collapsed conformation and the retained closed one indicates that the acceleration is still too strong, and thus longer simulations with more reduced acceleration are necessary to eliminate the false conformational changes.

■ DISCUSSION

The MD simulations demonstrated that the LRPF method developed in the present study is capable of simulating at atomic resolution functionally relevant global conformational changes of protein only with prior knowledge of local ligand binding sites, which is often available by biochemical experiments. For the purpose of modeling of the conformational changes without detailed knowledge of the final structure, it is crucial that the simulation does not produce any false positives, i.e., final structures different from the true one, since, obviously, one cannot identify the true final structure out of various final structures without detailed knowledge of the true final

structure. Although, in the present simulations of the closed-to-open transition of CaMn, two trials (LRPF3 and LRPF4) provided structures different from the other two trials (LRPF1 and LRPF2), one can selectively recognize the open forms found in LRPF1 and LRPF2 as the final states of successful transitions without prior knowledge of the final structure. First, one can inspect the structures resulting from the trials and see that the open form undergoes large conformational changes at the α helix bundle, exposing the hydrophobic patch. Given that the function of this protein involves molecular recognition of α helices upon Ca^{2+} binding, the open form can be a good candidate for the final state of the functional transition. Next, the structures of the open form found in the two trials (LRPF1 and LRPF2) are essentially the same, indicating that the transition to the open form is reproducible by the simulations. Finally, the structures obtained in the other two trials (LRPF3 and LRPF4) which did not lead to the open form are neither unfolded ones nor other open ones but in the closed forms where the overall α helix bundle structures are almost maintained as in the initial state. Given again the function of the protein and structural comparison with the open form obtained by the other trials, one can readily recognize that the structures are of intermediate states before the transition to the open form. Thus, the open form in the two trials can be selected as the true positive of the final state of the functional transition since the structures in the other two trials are not false positives. The key feature here is the high stability of the present method; no false positives, i.e., unfolded structures or other open ones, were obtained in the multiple trials.

In the LRPF simulations, the trajectories reached at the final structure in the open conformation in $\sim 1 \mu\text{s}$, whereas no global conformational changes was observed in the unbiased MD simulation for more than $3.8 \mu\text{s}$. Experimentally, a time constant of the conformational change of the N-lobe part in the entire CaM protein upon ligand binding was measured to be $\sim 20 \text{ ms}$.³⁰ Although the time constant for the entire CaM protein may not directly represent that for the truncated CaMn (N-lobe only) one studied in the present study, it is expected that the LRPF method can accelerate the slow process of the global conformational changes by several orders of magnitude.

The method succeeded in reproducing the global conformational changes without detailed knowledge of the final protein structure, which is often difficult with other methodologies. The difficulty arises from complex coupling between local and global events in functional processes which makes it difficult to determine *a priori* appropriate reaction coordinates in biasing simulations for accelerating conformational changes. As shown in the SMD simulation for CaMn in the present study, sole enhancement of local interaction at the ligand binding sites failed to induce the global conformational changes presumably due to weak local–global coupling and slow fluctuation–relaxation of the global motion. This indicates that explicit biasing along reaction coordinates that represent global conformational changes is necessary for accelerating the process. However, obviously, intuitive determination of reaction coordinates for global motion of many atoms is not trivial. The use of the LRT in the LRPF method overcomes the difficulty through evaluation of the explicit local–global coupling, allowing one to determine reasonable reaction coordinates for global conformational changes.

The idea of LRT was also utilized to explore structural changes with random forces for the perturbation in previous studies.^{11,31,32} Because of the robustness of LR displacements,

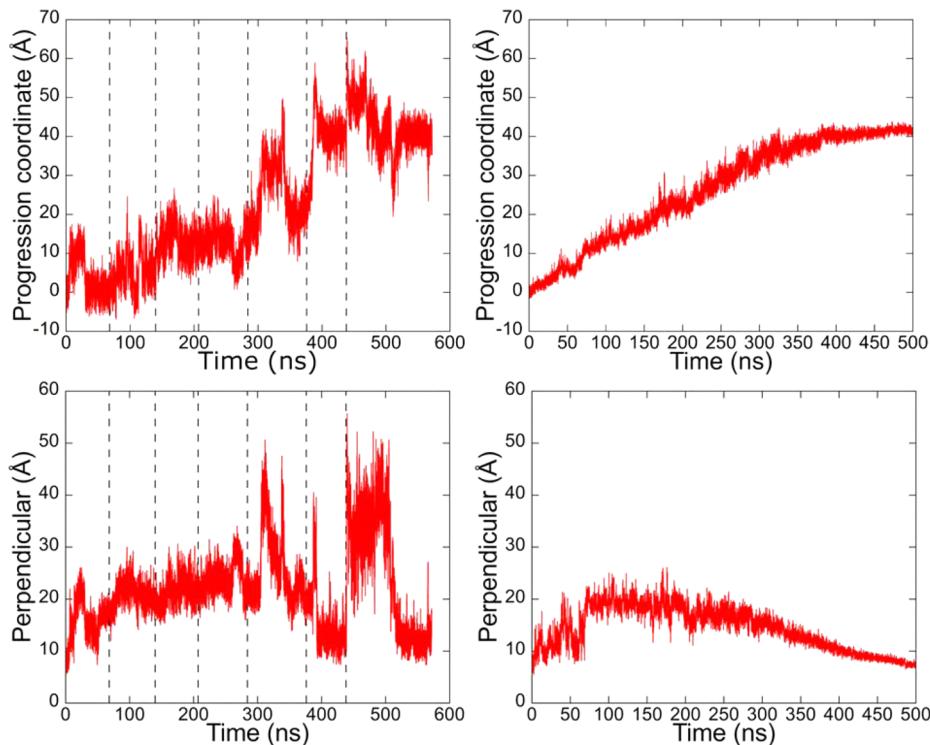


Figure 9. Time evolution of the conformational changes along progression and perpendicular coordinates in LRPF1 (left panels) and TMD (right panels). Vertical dashed lines are the same as those in Figure 6.

eq 1, against the perturbative forces,²³ functional conformational changes can be searched by randomly applied perturbative forces. Unlike the present approach, those methods do not require prior knowledge of ligand binding. However, the use of random forces for perturbation is not very efficient for simulations of large scale conformational changes. Note that such large scale conformational changes and their relaxation are very slow. In fact, a typical simulation time of each iteration in the LRPF procedure is tens of nanoseconds. As described below, the simulation time in the present study could still be too short to sufficiently sample the conformational changes. It is therefore desirable to persistently employ the perturbative force that provides the biasing force inducing the functional conformational changes most efficiently. The present choice of the perturbative force acting on the ligand binding site is appropriate for the purpose. However, the use of random forces for the perturbation frequently gives unnecessary biasing forces and therefore is expected to require much longer MD simulation time to induce the functional conformational changes.

We also tested the aMD method,¹⁰ a nondirected biasing method, for the simulation of the conformational changes of CaMn. aMD has been shown to be a powerful method to simulate conformational changes of proteins.^{29,33–41} However, in the case of the conformational changes of CaMn, the true final structure was not found by aMD in simulation time more than that of LRPF1 which successfully found the final open conformation. The low efficiency of aMD in the present case can be well explained by the simple idea described above that simulation time to search out proper conformational changes in a nondirected manner can quickly increase when the conformational changes become large and complex. It should be noted that the closed-to-open conformational changes of CaMn are large, i.e., RMSD change is more than 2 Å, and

involve very complex coupling between ligand bindings at two sites and the large opening motion of the four helix bundle proceeding through a strongly winding pathway in the configuration space (Figure 6 and 10). For simulations of large and complex conformational changes, therefore, the well-directed biasing scheme of LRPF can be advantageous over nondirected biasing ones to induce the proper conformational changes.

Reaction coordinates representing global conformational changes along which the conformational changes are induced by directed biasing forces can be determined when the final structure is known a priori. Moradi and Tajkhorshid succeeded in identifying conformational transition pathways of a bacterial ABC transporter MsbA with guiding forces that describe global conformational changes between the two end structures observed in X-ray crystallographic studies.⁴² A more systematic and commonly used way to simulate global conformational changes when the final structure is a priori known is the TMD approach. However, the TMD method employs biasing forces toward a direction pointing to the target structure anywhere in the configuration space, which are physically less relevant and thus can induce physically less realistic conformational changes such as the “large-scales first” phenomenon.²⁰ In fact, the closed-to-open conformational transition observed in the TMD simulation showed global movement of the protein backbone before local rearrangement of coordination spheres to Ca²⁺ ions at the binding sites as described above and presents a marked contrast to that in the LRPF simulation as seen in Figure 6.

In order to characterize the protein conformational changes during the transition more clearly, the instantaneous protein coordinate in the configuration space is projected onto the following two coordinates. The first one is the progression coordinate which is a linear line interpolating between the two end structures and thus represents progress of the conforma-

tional changes. The other one is the perpendicular coordinate which is a normal distance from the linear line of the progression coordinate and thus shows deviation from it (see Materials and Methods for details). Figure 9 shows the time evolution of the closed-to-open conformational changes represented in the two coordinates. In the TMD simulation, the conformational changes linearly evolved in time and were restricted around the linearly interpolating line of the progression coordinate in the configuration space. This behavior is a consequence of the biasing forces constantly directed to the target structure. However, the conformational changes in the LRPF simulation exhibit large deviation from the linearly interpolating line. The deviation became especially large when the opening motion occurs in 300–400 ns. Figure 10 depicts paths of the conformational changes in the space of

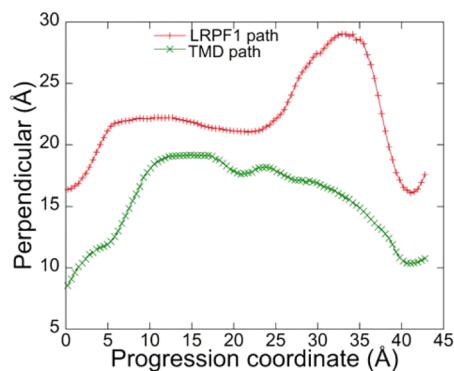


Figure 10. Comparison of TMD and LRPF paths of the conformational changes in the space of progression and perpendicular coordinates.

those two coordinates. Compared with the TMD path, the LRPF one largely deviates from the linearly interpolating line and exhibits a large bent in the configuration space. Note that the trajectory of the TMD simulation failed to reach the experimentally observed target structure even though it utilized detailed information on the target structure to strongly restrict conformational changes around the linearly interpolating line to the target structure. The observation indicates that the proper conformational changes are characterized by the winding path in the configuration space so that the biasing forces of TMD which prefer a linear path directed to the target structure did not lead to convergence to the proper path. However, the frequent update of the direction of the biasing forces with LRT in the iterative procedure of the LRPF simulation permits the trajectory to follow the complex winding pathway of the conformational changes. It should be noted that the pathway found by the LRPF method is still an approximated one, and it will need to be refined by postprocess such as path sampling⁴³ to obtain a more accurate pathway. Nevertheless, the advantageous feature of the present method over TMD can be exploited to obtain a better initial guess for simulations of transition path by path sampling methods and for free energy calculations associated with the conformational transition process. Although the TMD is generally more computationally efficient to produce an initial guess of transition path, path sampling simulation with an initial path largely deviating from the correct pathway increases the possibility that the refinement of path fails due to the trap at a local minimum of the path. It is also noted that the refinement of the path by path sampling requires huge computational resources as the method uses a

number of replicas along the path. It is therefore considered that the path obtained by the LRPF based on the physically more reasonable scheme can be a better choice as the initial path for time-consuming path sampling in terms of both computational accuracy and efficiency than that by TMD was forced to proceed along the physically less relevant reaction coordinate.

The successful LRPF paths began with local conformational changes represented by a stepwise increase of the Ca^{2+} coordination number in EF2 (Figures 7 and S10 in Supporting Information) accompanying a gradual decrease of RMSD to the final state (Figures 6A and S11B in Supporting Information) before the sudden and relatively large RMSD changes, which are indicative of the transitions to the partially open state, observed when the coordination spheres were completed by the bindings of Glu31 in EF1 and Gln62 in EF2. The local conformational changes in EF loops for Ca^{2+} coordinations before the transition to the open form was also observed in the unbiased Eq-long simulation (Figure 3C and Figures S3 and S4 in Supporting Information). The characteristic tilt of helix 3 accompanying the coordination of Asp56 to Ca^{2+} in Eq-long described above was also seen in the LRPF simulations. Small but abrupt changes of RMSDs at 130 ns for LRPF1 (Figures 6A and C) and at 500 ns for LRPF2 (Figures S11B and S11D in Supporting Information) correspond to the tilt of helix 3. Thus, the unbiased simulation of Eq-long and LRPF simulations share the characteristic conformational changes before the transition to the open form, indicating that the successful LRPF pathways do not include serious artifacts introduced by the biasing force. It is also noted that the conformational changes for Ca^{2+} coordinations before the transition to the open form in LRPF simulations took place in a much earlier time range (<1 μs) than those in Eq-long (>2 μs) (Figures S3 and S4 in Supporting Information). The negatively charged groups in EF-hands were enhanced to approach the Ca^{2+} ions by the biasing forces represented by broad peaks at EF-hands in Figure 5. However, the flip of His61 was unlikely to be directly induced by the biasing force as the perturbative force acting on His61 is much smaller than those acting on the negatively charged groups. However, since the flip of His61 needs to occur when coordinations of Asp56 and Asp58 to a Ca^{2+} ion are established, the enhanced approach of those negatively charged groups to a Ca^{2+} ion may stimulate the flip of His61. These results indicate that although the LR biasing forces possess a global character, the present scheme does not always force global motions to occur but also enhances flexible local conformational changes prerequisite for proper local–global coupling.

The multiple trials of the LRPF simulations showed that a strict order of events even in the formation of the Ca^{2+} coordination spheres at the binding sites is necessary for efficient conformational transition. The order of events found in LRPF1 is probably the one that might give the most efficient pathway, as indicated by its relatively short simulation time to observe the closed-to-open transition (~400 ns). In LRPF2, however, Ca^{2+} in EF1 transiently dissociated from Ser26 in EF1, which was not observed in LRPF1 (Figure S4 in Supporting Information), although it is eventually coordinated by Ser26 again. The transition to the partially open state was seemingly hampered by transient dissociation because Ca^{2+} in EF1 was not held in the proper position for the coordination of Glu31, which is coupled with the closed-to-open conformational transition. The transient rearrangement at EF1 might

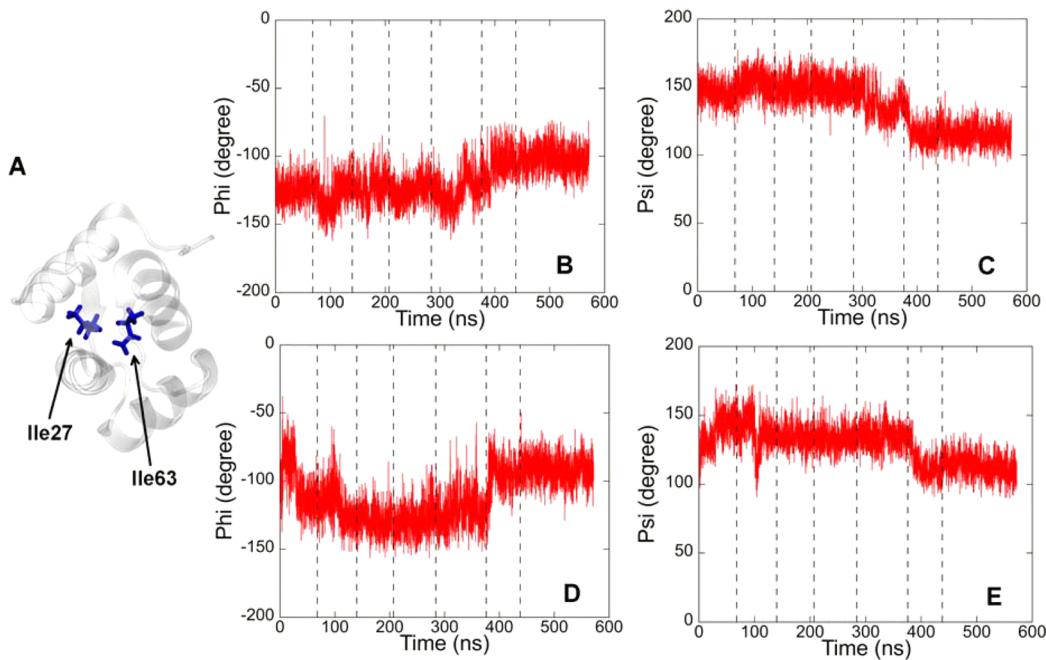


Figure 11. Changes in dihedral angles in LRPF1. (A) Positions of Ile27 and Ile63 are highlighted. The protein backbone is drawn in a transparent white ribbon representation, whereas Ile is shown in blue stick form. (B,C) Changes in phi and psi angles of Ile27. (D,E) Changes in phi and psi angles of Ile63. The vertical dashed lines are the same as those in Figure 6.

explain the relatively long simulation time (~ 950 ns) required for the overall conformational changes as compared with LRPF1. The order of events observed in LRPF3 and LRPF4 (Tables S3 and S4 and Figures S3 and S4 in Supporting Information) were quite different from those observed in LRPF1 and LRPF2 and the closed-to-open transition was not observed in the simulation time of $\sim 2 \mu\text{s}$ as described above. The observation indicates that there exist complex branches of pathways to conformational substates representing the rearrangement of the Ca^{2+} coordination which do not connect to the path of the closed-to-open transition (Figure 8). It should be noted that transient branching to the intermediates by local conformational rearrangement is the physically natural phenomenon in conformational transition processes and hence can be found in straightforward unbiased simulations as well as any kind of biased ones. The failure to observe the transition to the open form in LRPF3 and LRPF4 trajectories is therefore not due to artifacts of the LRPF methodology but to the limited simulation time which is still not long enough to simulate the local conformational rearrangement. Such conformational substates of local rearrangement impose a limitation of the present scheme of the LRPF method. Note that, although the linear biasing force is applied to introduce the conformational changes in the present method, nonlinear conformational changes accompanying the conformational transitions such as the rotations of main/side-chains and hydration changes of water molecules need to occur spontaneously during MD simulations for tens of nanoseconds in each cycle of the LRPF procedure. Unfortunately, the LR biasing forces are in principle not very efficient to enhance local nonlinear conformational changes, although the local–global coupling may contribute to the stimulation of local conformational changes to some extent. The inhibitory substates of the Ca^{2+} coordination conformations may also appear due to artifacts of the force field employed. Piana et al. recently showed that the default force field parameters (CHARMM22) overestimate the association

constant of guanidinium acetate,⁴⁴ indicating overpolarization of carboxylate groups and thus providing an explanation for the strong binding of Glu31 to Ca^{2+} , instead of those of Asn24 and Ser26, commonly observed in the unsuccessful simulations of LRPF3, LRPF4, and Eq-long (Figures 3C, S8, and S9 in Supporting Information). A longer LRPF simulation extending conformational sampling of the local transitions or a combination of a method for enhanced sampling of local conformational substates with the LRPF method together with the use of improved force field parameters would be expected to overcome the difficulty.

In the successful simulations of LRPF1 and LRPF2, the global opening conformational transitions are coupled with the coordinations of Glu31 in EF1 and Gln62 in EF2 to Ca^{2+} ions, respectively, as described above. The local–global coupling is consistent with the two-step ligand binding mechanism proposed by Grabarek based on a crystallographic study for holo-CaM with the N-terminal domain kept closed by a disulfide bond introduced between helix2 and helix4.⁴⁵ In the proposed mechanism, the binding of Ca^{2+} ions to N-terminal halves of EF1 and EF2 loops first takes place, and then coordinations of glutamates at 12th positions in both of EF1 and EF2 loops to Ca^{2+} ions accompanying rotation of the exiting helices (helix2 and 4) occur. The glutamate at the 12th position in EF1 is Glu31 and thus the local–global coupling in EF1 agrees with our observation. However, the residue which participates in the local global coupling in EF2 found in our simulation, Gln62, does not locate at the 12th position, showing asymmetric coupling between those two binding loops. Grabarek also suggested that isoleucine residues at the middle positions of the two binding loops undergo rotation of the main-chain dihedral angles upon the rotation of the exiting helices and thus act as hinges for the opening transition. In the present simulations of LRPF1 and LRPF2, the dihedral angles of the corresponding isoleucine residues, Ile27 and Ile63 in EF1 and EF2, were also found to rotate (Figures 11 and S12 in

Supporting Information) at the moments of the coordinations of Glu31 in EF1 and Gln62 in EF2 to Ca^{2+} ions (Figures 7 and S10 in Supporting Information) coupled with the closed-to-open transition (Figures 6A and S11B in Supporting Information).

It has also been suggested that Phe19 (or its counterpart in the C-terminal domain, Phe92) of CaM plays a critical role in maintaining its function.^{26,27} Its substitution to alanine significantly decreases CaM's ability to bind and activate target proteins. A study employing a hydrophobic probe showed that the F92A mutation of the human CaM C-terminal domain resulted in a decrease in the solvent exposed hydrophobic surface, suggesting that the conformational transition from the closed structure to the open one is hindered by the mutation.²⁶

In the present simulations of Eq-open, LRPF1, and LRPF2, an atomistic mechanism of the role of Phe19 was revealed. In Eq-open, Phe19 was trapped in a pocket formed by Leu32, Val35, and Met36 (termed "LEU-MET-VAL pocket"²⁸) when the structure was in the open conformation in the first stage of Eq-open simulation starting from the NMR structure. Judging from the native contact analysis, however, the side-chain of Phe19 did not fit well to the pocket (Figure S13A in Supporting Information). Consequently, the protein underwent transition from the open conformation to the partially open one, during which Phe19 exited from the pocket (Figures 12

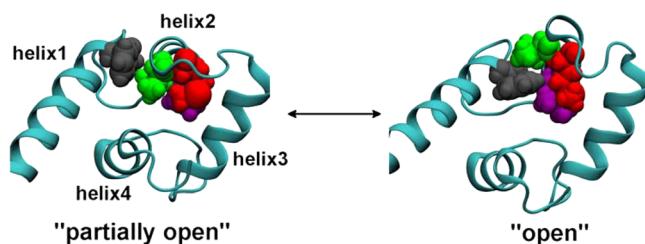


Figure 12. Movement of Phe19 upon the conformational transition between the partially open state and the open one. Protein backbones are drawn in a ribbon representation, while residues involved in the interaction of Phe19 are drawn in a vdW representation (Phe19, black; Leu32, purple; Val35, green; and Met36, red).

and S13A in Supporting Information). Conversely, Phe19 was inserted into the pocket upon the transition from the partially open conformation to the open one in LRPF1 and LRPF2 (Figures S13B and C in Supporting Information). The open state trapped by the insertion of Phe19 into the pocket lasted for tens of nanoseconds, followed by spontaneous expulsion of Phe19 and transition to the partially open state. Those observations clearly indicate that the transition between the open conformation and the partially open one is coupled with the insertion of Phe19 into the pocket. The transition from the partially open conformation to the open one is also correlated with increase in SASA of hydrophobic residues in the protein (Figure S5 in Supporting Information). The insertion of Phe19 into LEU-MET-VAL pocket is therefore suggested to play a role in stabilizing the open conformation with more exposed hydrophobic surface and thus in stimulating binding of a target protein. Transient insertion of Phe19 with the LEU-MET-VAL pocket was also observed in a simulation of open-to-closed conformational changes upon the removal of Ca^{2+} ions with the activation–relaxation technique by Dupuis et al.,²⁸ although, unlike the observation in the present study, the insertion

appeared during the closing process and thus was interpreted to contribute to an efficient closure.

Tripathi et al. developed coarse-grained variational models of CaMn and CaMc (C-terminal domain of CaM) which allow one to obtain an analytical expression of mean square fluctuations of individual residues.⁴⁶ Their models suggest that, in CaMc, the loop connecting helix2 and helix3 (loop23, Figure 2D) undergoes conformational cracking when the closed-to-open transition takes place, whereas any cracking, or local unfolding, is not involved in the transition of CaMn. The lack of cracking in CaMn is attributed to higher fluctuation that can release local stresses more efficiently. In accord with their results, we did not find any indication of cracking in LRPF trajectories for CaMn.

Finally, the applicability of the LRPF approach to other larger systems with more complex molecular ligands and possible limitations are discussed. First, limitation of the MD simulation time imposed by an increase of the system size would cause problems due to insufficient search and relaxation in the present scheme, although the limitation is commonly applied to unbiased MD simulations as well as most of the simulations with biased schemes. However, the MD simulation time for larger systems can be extended relatively easily by parallel computation. In this case, the system size does not matter for the present method as the overhead of the LRPF part is very small. In general, the present method is expected to be even more applicable to some larger systems where large structural changes are realized by collective accumulation of smaller displacements of atoms in larger domains, which is a better condition for the linear response treatment.

However, a major limitation of the method would appear when one treats highly nonlinear conformational changes that are hardly traced by linear response forces as explained for the local rearrangements of LRPF3 and LRPF4. Protein recognition of intact calmodulin is an example of such systems. In this case, one needs to treat protein–protein interaction with large nonlinear conformational changes, and thus, it could be very difficult to define proper perturbative forces and to describe the large scale movement of the protein which completely change its shape. Proteins where the linear response couplings are weak were also found by precise classification of the linear response couplings for available protein structures in PDB.⁴⁷ Nevertheless, the simplicity of the present iterative scheme with the well-defined LRT enables one to estimate the applicability of the method relatively easily.

The LRPF method can also treat any kind of ligand with some limitations. First, it is noteworthy that the biasing force is insensitive to changes of the perturbative force at the ligand site as shown by Ikeguchi et al.²³ Thus, approximate perturbative force without details of the ligand structure can be utilized to induce large conformational changes by the biasing force. However, as found in the present study, large scale conformational transition can also be coupled with selective formation of local interaction at the binding site. In this case, one must describe the ligand binding structure more precisely, and thus, treating larger ligands would introduce further complexity in the simulation.

Because of uncertainty coming from the complex nature of protein conformational changes upon ligand bindings, multiple trials of the LRPF search are necessary. One may obtain many false positives at first attempt. However, the high stability of the method demonstrated in the present study allows one to tune the parameters of the LRPF procedure, such as choice of the

perturbative force, strength of the biasing force, and the simulation time, to minimize false positives, and thus to identify true positives or at least better candidates for them which can be verified with biochemical experimental evidence.

■ CONCLUDING REMARKS

We have developed a method, LRPF, to accelerate conformational transitions of proteins upon ligand binding. The method introduces a biasing force based on linear response theory, which determines a local reaction coordinate in the configuration space that represents linear coupling between local events, i.e., ligand bindings, and global conformational changes, and thus provides one with fully atomistic models undergoing large conformational changes without knowledge of a target structure. Through iterative cycles consisting of a biased MD simulation with an updated linear response force and a following unbiased MD simulation for relaxation, nonlinear conformational changes in the overall transition process are described. The method was applied to elucidate the conformational transition pathway of yeast CaMn upon the binding of Ca^{2+} ions. Two out of four LRPF simulations successfully searched out the end structure of the conformational transition within the simulation time of $<1.5\ \mu\text{s}$. The entire pathway was found to be characterized by a sequence of three phases: (i) formation of the Ca^{2+} coordination structures in EF-hands other than two coordination groups at Glu31 and Gln62; (ii) binding of Glu31 and Gln62 to ligand ions, which couples with the transition to the partially open state; (iii) insertion of Phe19 to the “LEU-MET-VAL” pocket accompanying structural changes to the open state. Crucial importance of the proper formation of the Ca^{2+} coordination structures in EF-hands in phase i preceding the global conformational transition to the partially open state in phase ii is suggested since a lack of the proper coordination structures due to being stuck at different conformational substates in phase i did not lead to the transition to the partially open state in phase ii within the simulation time in the two unsuccessful searches. Since local–global coupling is a ubiquitous feature in functional processes of proteins such as motors and transporters, the LRPF method explicitly taking into account the local–global coupling is expected to be widely applicable to model the functional conformational changes of proteins.

■ MATERIALS AND METHODS

Eq-prep and Eq-long Simulations. We employed a solution NMR structure of apo-CaMn (PDB id: 1F54, model 1²⁴) for modeling of the closed conformation. His61 was assumed to be positively charged, considering that two acidic residues (Asp56 and Asp58) are located in its vicinity. We manually added a calcium ion to each EF-hand to set up the initial state of the closed–open conformational changes. The entire protein was immersed in a rectangular box of size $\sim 60^3\ \text{\AA}^3$. The atomic coordinates of the simulation box filled with water molecules were set up with the LEaP module of AMBER9.⁴⁸ The system was neutralized by adding five sodium ions randomly with the Autoionize plugin of VMD.⁴⁹

The CHARMM22⁵⁰ force field, together with CMAP correction,⁵¹ was used for the protein. The Lennard-Jones parameters for Na^+ and Ca^{2+} ions were taken from references 52 and 53, respectively. TIP3P^{50,54} parameters were employed for water molecules. Nonbonded interactions were smoothly cut off at 12 Å, with the switching function beginning at 10 Å.

Long range electrostatic interactions were calculated with the particle mesh Ewald method.⁵⁵ The grid size was set to 64³. Hereafter, the same parameter set was used unless otherwise stated.

The system was first energetically minimized with the conjugate gradient method. All MD simulations were carried out under the constant NPT condition with $P = 1\ \text{bar}$ and $T = 298\ \text{K}$. The integration time step was set to 2 fs for all MD simulations except for the heating phase where a time step of 1 fs was used. The energetically minimized system corresponding to 0 K was gradually heated by reassigning velocities at every step. The temperature of the system was linearly elevated to 298 K in 298 ps. After the heating phase, further equilibration at 298 K lasted for 102 ps. To further equilibrate water molecules around the protein, atoms except for water molecules were restrained at their initial positions by harmonic potentials, and a restrained simulation was carried out for 2 ns. The force constants of restraining potentials were gradually decreased from 10 to 0.0625 kcal mol⁻¹ Å⁻² in the course of the simulation. Langevin dynamics with a damping coefficient of 2 ps⁻¹ was used to keep the temperature constant. Pressure was maintained with Berendsen's method,⁵⁶ using a relaxation time of 100 fs. The internal degrees of freedom of water molecules were fixed with SETTLE.⁵⁷ Bonds including a hydrogen were constrained by the RATTLE⁵⁸ algorithm. After the restrained simulation, an equilibrium simulation without restraint was conducted for 120 ns (Eq-prep). All calculations mentioned were done with NAMD.⁵⁹ Eq-long was started from the last snapshot of Eq-prep and lasted for 3,880 ns, using AMBER12⁶⁰ with the same force field. The calculation was performed on a GPU with SPFP mode. We utilized the same force field and the simulation protocols (i.e., Langevin dynamics and Berendsen pressure coupling, etc.), except for the switching function (AMBER12 does not support this feature).

Eq-open Simulation. Open holo-CaMn was modeled with a solution NMR structure reported by Ishida et al.²⁴ (PDB id: 1F55, model 1). The same protocols as those for Eq-prep were employed for minimization, heating, and equilibration to prepare the equilibrated system. Total simulation time was $\sim 300\ \text{ns}$. Calculations were done with NAMD.

TMD Simulation. In TMD, a biasing potential is added to the force field, “pulling” the system to a target conformation in a predefined simulation time (500 ns in this study). The biasing potential was defined as

$$V(\mathbf{r}^N(t)) = \frac{1}{2}k[\text{RMS}(t) - \rho(t)]^2 \quad (12)$$

In the formula, $\mathbf{r}^N(t) = (\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_N(t))$ denotes N atoms' coordinates at time t , k is the force constant, and N is the number of atoms on which the potential acts. RMS(t) is defined as,

$$\text{RMS}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i(t) - \mathbf{r}_{i,\text{target}})^2} \quad (13)$$

where $r_{i,\text{target}}$ represents the coordinate of the i^{th} atom in the target structure. $\rho(t)$ determines the minimum value of the potential at time t , and it decreases linearly from $\rho(0) = \text{RMS}(0)$ to $\rho(\text{last}) = 0$. The external potential was applied to the main chain atoms only. Accordingly, in the case of CaMn, which has $77 - 8 = 69$ residues (excluding the first and the last four residues), $N = 207$. The force constant k was set to 207 kcal mol⁻¹ Å⁻². The simulation time was 500 ns, and $\rho(t)$

decreases linearly during the simulation from $\rho(0) = \text{RMS}(0) = 3$ to $\rho(\text{last}) = 0$ ($0.06 \text{ \AA}/10 \text{ ns}$). Thus, the instantaneous RMSD with respect to the target conformation (eq 13) is expected to decrease from 3 to 0 \AA in 500 ns.

SMD Simulations. In SMD simulations, \bar{G}_i in eq 5 was added to the force derived from the force field. Details of the simulations are described in Supporting Information.

Trajectory Analysis. RMSD was calculated with respect to the main chain atoms unless otherwise noted. To evaluate progress of structural changes, each trajectory was projected onto the axis parallel to the line connecting the two end states (Figure S14 in Supporting Information),

$$\hat{e}_{\text{diff}} = \frac{1}{|\mathbf{r}_{\text{final}} - \mathbf{r}_{\text{initial}}|} (\mathbf{r}_{\text{final}} - \mathbf{r}_{\text{initial}}) \quad (14)$$

Here, $\mathbf{r}_{\text{initial}}$ and $\mathbf{r}_{\text{final}}$ are the structures corresponding to the closed and open CaMn, respectively, which were aligned by mutual RMSD fit. The former is calculated as an average structure over the last 10 ns part of Eq-prep trajectory and the latter as an average over the last 20 ns part of Eq-open trajectory, i.e., $\mathbf{r}_{\text{final}}$ represents the partially open state. Each trajectory $\mathbf{r}(t)$ aligned by RMSD fit to $\mathbf{r}_{\text{initial}}$ was projected as

$$P(t) = \hat{e}_{\text{diff}} \cdot (\mathbf{r}(t) - \mathbf{r}_{\text{initial}}) \quad (15)$$

where $P(t)$ is called the progression coordinate ($P(0) \approx 0$, $P(\text{"partially open"}) \approx 43$). A component of $\mathbf{r}(t) - \mathbf{r}_{\text{initial}}$ perpendicular to \hat{e}_{diff} was also calculated as

$$\mathbf{r}^{\perp}(t) = \mathbf{r}(t) - \mathbf{r}_{\text{initial}} - P(t)\hat{e}_{\text{diff}} \quad (16)$$

where $\mathbf{r}^{\perp}(t)$ is called the perpendicular coordinate (see Figure S14 in Supporting Information). $P(t)$ vs $|\mathbf{r}^{\perp}(t)|$ plots in Figure 10 were constructed as follows. The entire region of the projection coordinate was divided in 100 bins, and the average value of $|\mathbf{r}^{\perp}(t)|$ in each bin was calculated and plotted. The analyses above were carried out with respect to the main chain atoms only.

The coordination number of protein's oxygen atoms to the ligand ones (Ca^{2+} in CaMn) was calculated with a smoothed step function as⁶¹

$$C(t) = \sum_{i=1}^{N_{\text{cal}}} \sum_{j=1}^{N_0} \frac{1 - (d_{ij}(t)/d_0)^{10}}{1 - (d_{ij}(t)/d_0)^{20}} \quad (17)$$

where N_{cal} is the number of Ca^{2+} ions, N_0 is the number of protein's oxygen atoms, and d_{ij} is the distance between particles i and j . The threshold value d_0 was set to 3.1 \AA .

In the native contact analysis of Phe19 (Figure S13 in Supporting Information), residues A and B are considered to contact with each other if the distance between the center of mass of the residues is less than 7 \AA . Then,

$$\text{contact index} = \begin{cases} 1 & \text{if Phe 19 contacts Leu 32, Val 35, and Met 36} \\ & \text{simultaneously} \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

However, the contact index of a residue in EF-hands to Ca^{2+} is the number of oxygen atoms that bind to Ca^{2+} ions (Figures S3 and S4 in Supporting Information). The same criterion of 3.1 \AA was applied to judge the contact.

Solvent accessible surface area was calculated with VMD. The probe radius was set to 1.4 \AA . Other parameters were set to default values.

ASSOCIATED CONTENT

Supporting Information

Simulation protocols and detailed description of the results of LRPF2, LRPF3, LRPF4, TMD, SMD, and aMD. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00120.

AUTHOR INFORMATION

Corresponding Author

*Tel: +81-75-753-4006. Fax: +81-75-753-4000. E-mail: hayashig@kuchem.kyoto-u.ac.jp.

Funding

This work was financially supported by grants from the Japanese Ministry of Education, Culture, Sports, Science, and Technology (25104004 and 25291034), the Programme for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry, and Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries and Food Industry.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Some computations were performed at the Research Center for Computational Science, Okazaki, Japan. Molecular figures were prepared with VMD.⁴⁹

REFERENCES

- Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis*; ACM: Portland, OR, 2009; pp 1–11.
- 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.; Wriggers, W. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330*, 341–346.
- Jensen, M. Ø.; Jogini, V.; Borhani, D. W.; Leffler, A. E.; Dror, R. O.; Shaw, D. E. Mechanism of voltage gating in potassium channels. *Science* **2012**, *336*, 229–233.
- Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transition to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 7270–7275.
- Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **1995**, *52*, 2893–2906.
- Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. C. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615–625.
- Wu, X.; Wang, S. Enhancing systematic motion in molecular dynamics simulation. *J. Chem. Phys.* **1999**, *110*, 9401–9410.
- Wu, X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **2003**, *381*, 512–518.
- Yang, L.; Gao, Y. Q. An approximate method in using molecular mechanics simulations to study slow protein conformational changes. *J. Phys. Chem. B* **2007**, *111*, 2969–2975.
- Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
- Kitao, A. Transform and relax sampling for highly anisotropic systems: Application to protein domain motion and folding. *J. Chem. Phys.* **2011**, *135*, 045101.

- (12) Laio, A.; Parrinello, M. Escaping free energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (13) Schlitter, J.; Engels, M.; Krüger, 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.
- (14) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997–999.
- (15) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **1997**, *72*, 1568–1581.
- (16) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.
- (17) Berteotti, A.; Cavalli, A.; Branduardi, D.; Gervasio, F. L.; Recanatini, M.; Parrinello, M. Protein conformational transitions: The closure mechanism of a kinase explored by atomistic simulations. *J. Am. Chem. Soc.* **2009**, *131*, 244–250.
- (18) Prakash, M. K.; Barducci, A.; Parrinello, M. Probing the mechanism of pH-induced large-scale conformational changes in Dengue virus envelope protein using atomistic simulations. *Biophys. J.* **2010**, *99*, 588–594.
- (19) Formoso, E.; Limongelli, V.; Parrinello, M. Energetics and structural characterization of the large-scale functional motion of adenylyl kinase. *Sci. Rep.* **2015**, *5*, 8425.
- (20) Ovchinnikov, V.; Karplus, M. Analysis and elimination of a bias in targeted molecular dynamics simulations of conformational transitions: Application to calmodulin. *J. Phys. Chem. B* **2012**, *116*, 8584–8603.
- (21) Gur, M.; Madura, J. D.; Bahar, I. Global transitions of proteins explored by a multiscale hybrid methodology: Application to adenylyl kinase. *Biophys. J.* **2013**, *105*, 1643–1652.
- (22) Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.
- (23) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein structural change upon ligand binding: Linear response theory. *Phys. Rev. Lett.* **2005**, *94*, 078102.
- (24) Ishida, H.; Takahashi, K.; Nakashima, K.; Kumaki, Y.; Nakata, M.; Hikichi, K.; Yazawa, M. Solution structures of the N-terminal domain of yeast calmodulin: Ca²⁺-dependent conformational change and its functional implication. *Biochemistry* **2000**, *39*, 13660–13668.
- (25) Kretsinger, R. H.; Nockolds, C. E. Carp muscle calcium-binding protein. *J. Biol. Chem.* **1973**, *248*, 3313–3326.
- (26) Meyer, D. F.; Mabuchi, Y.; Grabarek, Z. The role of Phe-92 in the Ca²⁺-induced conformational transition in the C-terminal domain of calmodulin. *J. Biol. Chem.* **1996**, *271*, 11284–11290.
- (27) Okano, H.; Cyert, M. S.; Ohya, Y. Importance of phenylalanine residues of yeast calmodulin for target binding and activation. *J. Biol. Chem.* **1998**, *273*, 26375–26382.
- (28) Dupuis, L.; Mousseau, N. Understanding the EF-hand closing pathway using non-biased interatomic potentials. *J. Chem. Phys.* **2012**, *136*, 035101.
- (29) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997–3002.
- (30) Park, H. Y.; Kim, S. A.; Korlach, J.; Rhoades, E.; Kwok, L. W.; Zipfel, W. R.; Waxham, M. N.; Webb, W. W.; Pollack, L. Conformational changes of calmodulin upon Ca²⁺ binding studied with a microfluidic mixer. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 542–547.
- (31) Atilgan, C.; Gerek, Z. N.; Ozkan, S. B.; Atilgan, A. R. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* **2010**, *99*, 933–943.
- (32) Atilgan, A. R.; Aykut, A. O.; Atilgan, C. Subtle pH differences trigger single residue motions for moderating conformations of calmodulin. *J. Chem. Phys.* **2011**, *135*, 155102.
- (33) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. Ras conformational switching: Simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics. *PLoS Comput. Biol.* **2009**, *5*, e1000325.
- (34) Bucher, D.; Grant, B. J.; Markwick, P. R.; McCammon, J. A. Accessing a hidden conformation of the maltose binding protein using accelerated molecular dynamics. *PLoS Comput. Biol.* **2011**, *7*, e1002034.
- (35) de Oliveira, C. A. F.; Grant, B. J.; Zhou, M.; McCammon, J. A. Large-scale conformational changes of *Trypanosoma cruzi* proline racemase predicted by accelerated molecular dynamics simulation. *PLoS Comput. Biol.* **2011**, *7*, e1002178.
- (36) Lindert, S.; Kekenes-Huskey, P. M.; Huber, G.; Pierce, L.; McCammon, J. A. Dynamics and calcium association to the N-terminal regulatory domain of human cardiac troponin C: A multiscale computational study. *J. Phys. Chem. B* **2012**, *116*, 8449–8459.
- (37) Wereszczynski, J.; McCammon, J. A. Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 7759–7764.
- (38) Blachly, P. G.; de Oliveira, C. A. F.; Williams, S. L.; McCammon, J. A. Utilizing a dynamical description of IspH to aid in the development of novel antimicrobial drugs. *PLoS Comput. Biol.* **2013**, *9*, e1003395.
- (39) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and dynamic network of the M2 muscarinic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10982–10987.
- (40) Singh, R.; Ahlawat, N.; Murarka, R. K. Activation of corticotropin-releasing factor 1 receptor: Insights from molecular dynamics simulations. *J. Phys. Chem. B* **2015**, *119*, 2806–2817.
- (41) Miao, Y.; Caliman, A. D.; McCammon, J. A. Allosteric effects of sodium ion binding on activation of the M3 muscarinic G-protein-coupled receptor. *Biophys. J.* **2015**, *108*, 1796–1806.
- (42) Moradi, M.; Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 18916–18921.
- (43) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (44) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47–L49.
- (45) Grabarek, Z. Structure of a trapped intermediate of calmodulin: Calcium regulation of EF-hand proteins from a new perspective. *J. Mol. Biol.* **2005**, *346*, 1351–1366.
- (46) Tripathi, S.; Portman, J. J. Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2104–2109.
- (47) Amemiya, T.; Koike, R.; Fuchigami, S.; Ikeguchi, M.; Kidera, A. Classification and annotation of the relationship between protein structural change and ligand binding. *J. Mol. Biol.* **2011**, *408*, 568–584.
- (48) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. AMBER 9; University of California: San Francisco, CA, 2006.
- (49) Humphrey, W.; Dalke, A.; Schulter, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (50) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; 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.; Mattox, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

- (51) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (52) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.
- (53) Marchand, S.; Roux, B. Molecular dynamics study of calbindin D9k in the apo and singly and doubly calcium-loaded states. *Proteins* **1998**, *33*, 265–284.
- (54) 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.
- (55) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $\text{N} \log(\text{N})$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (57) 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.
- (58) Andersen, H. C. Rattle: A “velocity” version of the Shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.
- (59) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (60) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Saguì, C.; Babin, V.; Luchko, S.; Gusarov, A.; Kovalenko, A.; Kollman, P. A. AMBER 12; University of California: San Francisco, CA, 2012.
- (61) Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Phys. Rev. Lett.* **2003**, *90*, 238302.