

Analysis and Elimination of a Bias in Targeted Molecular Dynamics Simulations of Conformational Transitions: Application to Calmodulin

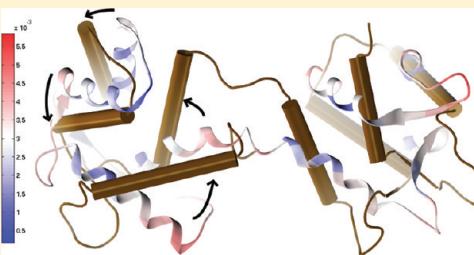
Victor Ovchinnikov*,† and Martin Karplus*,†,‡

*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

‡Laboratoire de Chimie Biophysique, ISIS, Université de Strasbourg, 67000 Strasbourg, France

 Supporting Information

ABSTRACT: The popular targeted molecular dynamics (TMD) method for generating transition paths in complex biomolecular systems is revisited. In a typical TMD transition path, the large-scale changes occur early and the small-scale changes tend to occur later. As a result, the order of events in the computed paths depends on the direction in which the simulations are performed. To identify the origin of this bias, and to propose a method in which the bias is absent, variants of TMD in the restraint formulation are introduced and applied to the complex open ↔ closed transition in the protein calmodulin. Due to the global best-fit rotation that is typically part of the TMD method, the simulated system is guided implicitly along the lowest-frequency normal modes, until the large spatial scales associated with these modes are near the target conformation. The remaining portion of the transition is described progressively by higher-frequency modes, which correspond to smaller-scale rearrangements. A straightforward modification of TMD that avoids the global best-fit rotation is the locally restrained TMD (LRTMD) method, in which the biasing potential is constructed from a number of TMD potentials, each acting on a small connected portion of the protein sequence. With a uniform distribution of these elements, transition paths that lack the length-scale bias are obtained. Trajectories generated by steered MD in dihedral angle space (DSMD), a method that avoids best-fit rotations altogether, also lack the length-scale bias. To examine the importance of the paths generated by TMD, LRTMD, and DSMD in the actual transition, we use the finite-temperature string method to compute the free energy profile associated with a transition tube around a path generated by each algorithm. The free energy barriers associated with the paths are comparable, suggesting that transitions can occur along each route with similar probabilities. This result indicates that a broad ensemble of paths needs to be calculated to obtain a full description of conformational changes in biomolecules. The breadth of the contributing ensemble suggests that energetic barriers for conformational transitions in proteins are offset by entropic contributions that arise from a large number of possible paths.



1. INTRODUCTION

Many proteins of interest populate multiple conformational states, whose relative stability is modulated by ligands, other biomolecules, and the ambient solution environment. Transitions between different conformations of such proteins in response to changes in the environment (e.g., binding of a ligand, interactions with other proteins or DNA) are an essential element in many cellular processes, such as enzyme catalysis,¹ DNA replication,^{2,3} refolding of misfolded proteins,⁴ ATP synthesis and organelle transport by molecular motors,^{5,6} and muscle contraction.⁷ Molecular dynamics (MD) simulations are a useful tool for the study of biomolecules with known structures that can be performed at atomic spatial and femtosecond temporal resolution, which are typically not accessible in experiments. In principle, conformational transitions can be observed in a sufficiently long MD simulation trajectory. Present-day computer resources, however, limit the duration of typical MD simulations to tens or hundreds of nanoseconds. Since many transitions occur on the scale of

microseconds and longer, they are beyond the reach of conventional MD. Many techniques have been developed to extend the conformational sampling beyond that achievable by conventional MD, e.g., umbrella sampling,^{8,9} metadynamics,¹⁰ adaptive biasing force,¹¹ adiabatic MD,¹² temperature accelerated MD,¹³ and orthogonal space random walk.¹⁴ These methods are well-suited for the exploration of phase space but not for computing physically meaningful transition paths between known conformations.

The finite-temperature string method^{15,16} can be used to compute transition paths, as well as the associated free energy profiles and transition rates. However, the string method requires an initial path that must be generated independently,

Special Issue: Macromolecular Systems Understood through Multi-scale and Enhanced Sampling Techniques

Received: December 31, 2011

Revised: March 12, 2012

Published: March 13, 2012

e.g., using zero-temperature methods,^{17–20} or by the methods discussed in the present study. Specialized path-based collective variables have also been devised for use with umbrella sampling or metadynamics to improve upon an initial path, in the spirit of the string method.²¹ Transition path sampling (TPS)^{22,23} can improve upon an initial path connecting two structures that correspond to metastable states (although, in principle, only one intermediate configuration is required to start TPS), by performing a “random walk” in trajectory space. However, because TPS relies on the generation of unbiased trajectories by MD, it can be very costly to apply to large biomolecules in which transitions are dominated by diffusive motions.

A simple method for finding transition paths that does not require the specification of collective variables is targeted molecular dynamics (TMD).^{24,25} In the original TMD implementation,²⁴ the simulation system is initialized from one structure and evolves according to a standard MD potential augmented by an evolving holonomic constraint, which gradually steers the simulation structure toward a specified (target) structure. The constraint guarantees that the root-mean-square difference (RMSD) between the simulation structure and the target structure, typically computed using all atoms in the molecule, is reduced to near zero within the prescribed simulation time. In the standard TMD method, the presence of an evolving constraint implies that the simulation system does not sample an equilibrium ensemble, and therefore TMD cannot be used to compute equilibrium properties, such as free energy profiles along the transition. In addition, because of the constraint, the system may be forced to overcome very high energy barriers.²⁶ Despite these limitations, TMD has been found to generate plausible transition paths, and has been used in studies of allosteric and a variety of transitions in large proteins.^{4,27–30} Transition paths obtained with TMD have also been employed to initialize TPS,^{31,32} umbrella sampling,³³ and the string method.³⁴

Several related methods have been introduced since the original TMD publication. In restricted perturbation (RP) TMD,²⁶ the evolution of the holonomic constraint is modified such that the simulation system is not allowed to undergo a constrained displacement toward the target structure that exceeds a prescribed threshold value. The instantaneous value of the constraint is calculated from the actual displacement. The main feature of RPTMD is that the evolution of the constraint slows down in the regions of large free energy gradients, allowing the simulation trajectories to lengthen naturally and explore more conformational space in such regions. RPTMD also does not restrict spontaneous evolution of the simulated structure toward the target.²⁶

In an alternative formulation of TMD, which we refer to as restrained (R)TMD to distinguish it from the original method, the holonomic constraint is replaced by a restraint potential based on the RMSD between the simulation and target structures.³⁵ The RTMD method is simpler to implement than TMD because it does not require modification of the MD integrator (see section 2.1). A recent study found that TMD and RTMD generate similar transition paths.³⁶ The difference between the simulation and target structures was measured in terms of the mean-squared interatom distance³⁷ (MSID) defined as

$$\zeta(t) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j>i}^N (d(t)_{ij} - d_{ij}^0)^2 \quad (1)$$

where $d(t)_{ij}$ and d_{ij}^0 are the distances between atoms i and j in the simulation and target structures (at time t), respectively, and N is the number of atoms to which restraint forces are applied. To our knowledge, a systematic comparison between TMD and RTMD using the RMSD as the progress coordinate has not been made. We note that the RTMD method is conceptually similar to steered molecular dynamics (SMD),³⁸ in that both methods employ a time-dependent restraint potential to guide the simulation from one state to the other. The difference is that, in the original applications of SMD to the unbinding of ligands (e.g., unbinding of the avidin–biotin complex^{39,40}), the restraint potential was a function of the linear distance between groups of atoms. When the biasing coordinate is the best-fit RMSD, it seems more appropriate to use the term RTMD to emphasize the similarity with TMD. We will use the term SMD for steered dynamics in dihedral angles (see below).

A variant of SMD is biased (B)MD.³⁷ The main difference between BMD and other restraint methods (SMD or RTMD) is that in BMD the biasing potential is “one-sided”; i.e., forces are applied only if the value of the progress coordinate (e.g., MSID) computed from the simulation structure is ahead of the corresponding target value. The target value of the progress coordinate is updated only if it is behind the value computed from the structure. BMD is similar in concept to RP-TMD in that the target value of the progress coordinate evolves more slowly along path segments that correspond to high free-energy gradients as a function of the progress coordinate, and in that spontaneous evolution toward the target structure is not restricted. In the original BMD method,³⁷ the restraint potential is defined on the basis of the MSID. BMD was used to simulate forced unfolding of β -sandwich and α -helical proteins,^{37,41} and the conformational change in the F₁-ATPase.²⁷ A hybrid method, which combines the use of general progress variables with the one-sided application of the BMD restraint potential was used to generate initial trajectories for use with TPS.³²

The TMD class of methods for modeling conformational transitions has been widely applied to large biomolecules due to its algorithmic simplicity, its ability to generate plausible paths within a prescribed simulation time (typically of the order of a nanosecond^{27,29,30,36,42,43}), and the absence of a requirement to specify a reaction coordinate. A particular bias that appears to be shared by all existing TMD algorithms is the tendency of large-scale motions to precede small-scale ones, which we call below the “large-scales-first” feature. Although this feature of TMD has been identified previously,^{26,29,36,44} to our knowledge, neither the causes nor a possible modification to remove the bias have been investigated. Addressing this question, however, is of paramount importance in the modeling of conformational transitions that involve the interplay between small-scale events and large-scale events. Examples are allosteric transitions in proteins that involve the motions of smaller loops due to, e.g., the binding of a ligand, coupled to motions of larger subdomains, such as those occurring in myosins or ATPases upon nucleotide binding.^{30,45,46} In the modeling of such transitions by TMD, large-amplitude motions will usually occur before small-amplitude ones, which could disagree with the actual sequence of transition events (e.g., those that would be observed in unbiased dynamical trajectories). A cursory examination of such transition paths might, for example, rule out the induced fit mechanism in favor of a population shift mechanism.^{47,48} In the latter, ligand binding effectively “traps”

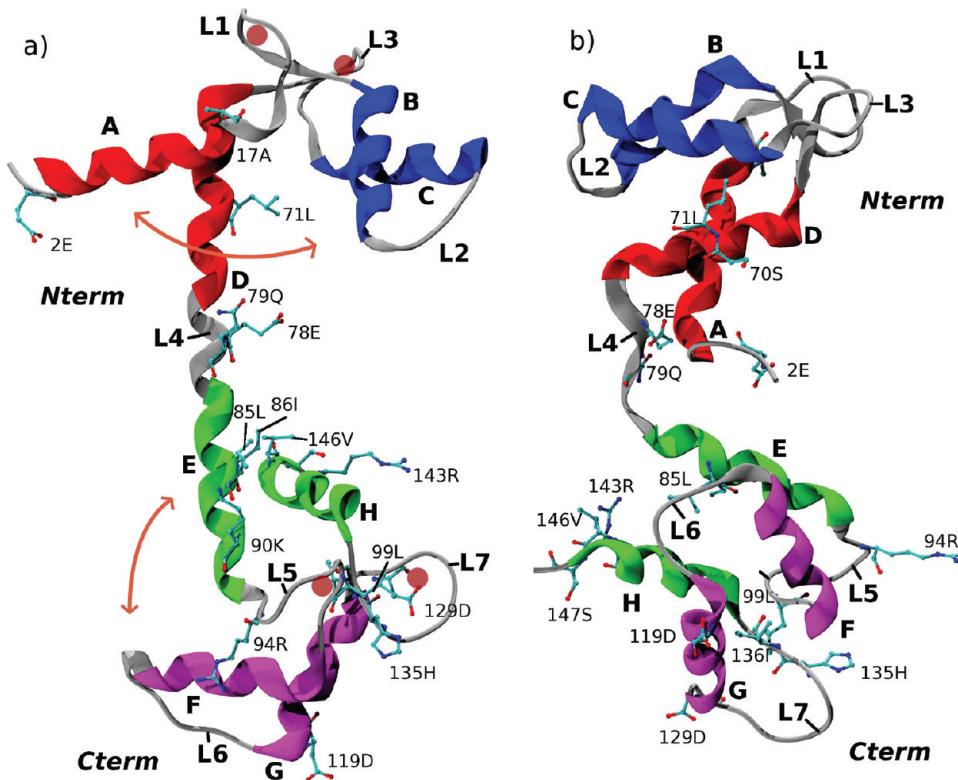


Figure 1. Calmodulin in (a) open form (1EXR) and (b) closed form (1DMO). Helix–loop–helix (HLH) motifs A-L1-B and C-L3-D comprise two EF-hands in the N-terminal domain (*Nterm*), and HLH motifs E-L5-F and G-L7-H comprise two EF-hands in the C-terminal domain (*Cterm*). Loops L1, L3, L5, and L7 form binding sites for Ca²⁺ ions, which were removed prior to simulation (the location of the ions within the loops is indicated by light red circles). Loop L4 forms the linker between the N-terminal and C-terminal domains. This loop is α -helical in the open structure and unwound in the closed structure. Residues that were mutated in the *Xenopus laevis* sequence are shown in ball-and-stick representation (see text). Residues that comprise the various structural elements are listed in Table S1 in the Supporting Information. The heavy-atom RMSDs between the *Nterm* and *Cterm* domains in the two structures are 5.6 and 5.5 Å, respectively. Images were generated with VMD.⁹⁴

the target molecule in a particular conformation that is observed only after certain large-scale motions, shifting the distribution of populations in favor of the ligand-bound conformations.

To investigate the source of the “large-scales-first” feature, and to propose a method in which the feature is absent, we test three biasing methods, two of which are straightforward extensions of the RTMD method, and the other involves biasing in internal coordinates. We then compare the transition paths generated by each method with those obtained from standard TMD. Variants of TMD and RTMD have previously been tested on small systems such as the alanine dipeptide.^{26,35} The simplicity of such systems implies that paths connecting multiple conformations are few and relatively simple, and will generally lack the broad range of transition length scales that are desirable for the present study. Similarly, although the overall rigor \rightarrow postrigor conformational transition studied by RTMD³⁰ is quite complicated, the domain to which biasing forces were applied was localized and small (\sim 300 atoms), with a relatively simple conformational change. Huang et al.³⁶ used the catalytic domain of the Src kinase protein Lyn as a test system. However, despite the moderately large size of the Src kinase domains (ca. 300 res.), the conformational transition can be described using only a few parameters, such as the conformation of a 20-residue loop A, the opening of a “lid” domain, and an approximately rigid motion of the α C helix.³⁶

As a test case to compare the various transition paths, we chose the Ca²⁺-binding protein calmodulin (CaM), which

undergoes a complex closed \leftrightarrow open conformational transition.⁴⁹ The CaM family and its homologues have been studied extensively because of their ubiquity in signal transduction and regulation of many cellular processes (see refs 49–55 and references therein). CaM proteins are known to bind to more than 300 target proteins⁵⁶ in response to changes in the Ca²⁺ concentration. The broad substrate specificity is made possible by the structure of CaM, in which two globular N-terminal and C-terminal domains are connected by an \sim 28-residue linker (see Figure 1). Each globular domain is composed of two helix-turn-helix motifs (EF-hands),⁵¹ which undergo a closed \rightarrow open transition upon the binding of Ca²⁺ ions, partially exposing hydrophobic target-binding pockets. There is strong evidence that the binding of Ca²⁺ ions to EF-hand loops (see Figure 1) is coupled to the opening of the EF-hand helices⁵⁴ and that Ca²⁺-binding occurs in a cooperative fashion.^{49,54,55} In addition to the open and closed EF-hand structures shown in Figure 1, partially open states have been observed in solution experiments^{57,58} and in MD simulations.^{59,60} The interdomain linker is known to be flexible in solution,^{49,57,58,60,61} and, together with the hydrophobic pockets, provides a deformable target-binding interface that permits high-affinity binding to a variety of protein targets.^{49,57,62} The present state of understanding of the mechanism of CaM function suggests that the closed \leftrightarrow open transition involves a delicate interplay between different spatial scales, making CaM a realistic system for testing transition path simulation methods. The Ca²⁺-bound and the

Ca^{2+} -free *apo* structures of CaM are very different, with the heavy-atom RMSD of $\sim 13 \text{ \AA}$ between the two structures (see also Figure 1). The large RMSD is expected to illustrate clearly the differences between the biasing methods employed in this study. The only TMD simulation of a CaM homologue known to us investigates the closed \rightarrow open transition in the Ca^{2+} -binding protein recoverin.⁴³ Although the order of events in the transition is not discussed from the viewpoint of large vs small spatial scales, it is evident from the discussion of the transition path that large-scale rearrangements precede smaller-scale ones. The transition was shown to begin with a reorientation of the C-terminal and N-terminal domains, followed by rearrangements of α -helices, ending with the motion of a myristoyl moiety.⁴³ The precedence of large-scale rearrangements is qualitatively consistent with the TMD and RTMD simulations of CaM reported here in section 3.2.

2. METHODS

As mentioned in the Introduction, for conformational transitions in which large-scale motion occurs in response to small-scale rearrangements, such as the binding of a small ligand to a large protein, followed by a rearrangement of protein subdomains, transition paths generated by TMD are likely to show an incorrect order of events. The methods described in this section were used to investigate the origins of the large scale bias, and to provide possible approaches that eliminate the bias, while preserving the modest computational requirements of TMD.

2.1. Scaled-Force TMD Method. To investigate the extent to which the difference in the biasing force magnitudes *per se* favor the large-scale motions, here we introduce a scaled-force (SF) TMD method, which limits the magnitude of the targeting force. The algorithm is based on the restrained formulation of TMD,^{24,35} in which a restraint potential, rather than a holonomic constraint, drives the system to the target configuration. It was shown recently³⁶ that the standard TMD and restrained RTMD generate similar transition paths when the MSID progress variable is employed. The present results show that this is also the case if the RMSD progress variable is used (see section 3.2).

Let $\hat{r}^* = \{r_1^*, \dots, r_N^*\}$ be a set of N coordinate triplets of the atoms to which targeting forces are applied, and $w_i = m_i / \sum_{j=1}^N m_j$, where m_i are the atomic masses. Defining center-of-mass (COM) coordinates by $r_i = r_i^* - \sum_{j=1}^N w_j r_j^*$, we apply the following harmonic potential to the simulated structure:

$$\begin{aligned} U_{\text{RTMD}}(\hat{r}, \hat{r}_T; t) &= \frac{k_{\text{RTMD}}}{2} \\ &\times (\rho(\hat{r}, A(\hat{r}, \hat{r}_T)\hat{r}_T) - \rho_0(t))^2 \end{aligned} \quad (2)$$

In eq 2, k_{RTMD} is the force constant,

$$\rho(\hat{x}, \hat{y}) = [\sum_{j=1}^N w_j \|x_j - y_j\|^2]^{1/2} \quad (3)$$

is the RMSD function, and \hat{r} and \hat{r}_T represent the current and target atomic coordinates, respectively. $\rho_0(t)$ is the desired equilibrium value of ρ at time t , which in an RTMD simulation is reduced gradually from an initial value, e.g., the RMSD between the initial and target structures, to a target value, e.g., $0\text{--}1 \text{ \AA}$. $A(\hat{r}, \hat{r}_T)$ is a rotation matrix, chosen such that $\rho(\hat{r}, A(\hat{r}, \hat{r}_T)\hat{r}_T)$ is minimum, i.e., $A(\hat{r}, \hat{r}_T)\hat{r}_T$ is a “best fit” of \hat{r}_T onto \hat{r} ^{63–66} (see Appendix A for an alternative definition). Henceforth, we

omit the subscripts on U_{RTMD} and k_{RTMD} if the meaning is clear from the context. Explicit time dependence of U and ρ_0 is also omitted, and $A(\hat{r}, \hat{r}_T)$ is written simply as A . Restraint forces are obtained from eq 2:

$$F_{x_m}^* = -k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) w_m (x_m - A_x r_{T,m}) \quad (4)$$

where $r_{T,j}$ denotes the target coordinate triplet $(x_{T,j}, y_{T,j}, z_{T,j})$, A_x is the first row of A , and the equation is written for the x -component of the force. Similar expressions hold for the y - and z -components. The simplicity of eq 4 comes from the fact that the derivatives of the rotation matrix A make no contribution to the force, as was claimed in ref 67. In Appendix A, we give a simple proof of this claim, and discuss a special case for which derivatives of A must be included (which are calculated in the Supporting Information).

Equation 4 shows that in RTMD (as in TMD^{24–26,36}) the force acting on an atom is proportional to the difference between its current and target positions, so that RTMD is likely to favor the occurrence of large-scale motions before small-scale ones. To remove the bias arising in the RTMD equations, we limit the maximum force that can be applied in eq 4:

$$F_{x_m}^* = -k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) w_m \alpha_m (x_m - A_x r_{T,m}) \quad (5)$$

$$\alpha_m = \min \left\{ 1, \frac{\Delta r_{\max}}{\|r_m - A r_{T,m}\|} \right\} \quad (6)$$

Equations 5 and 6 ensure that the forces on all atoms whose current and target positions differ by more than Δr_{\max} will have the same magnitude. Therefore, large-scale motion will not be biased by the magnitude differences of the RTMD forces to occur first. A feature of the modified RTMD equations is that the force in eq 5 can no longer be written as the gradient of a potential. This means that the simulations will not follow Hamiltonian dynamics. In particular, free-energy umbrella sampling simulations⁸ that use an RMSD potential⁶⁸ cannot be used. This drawback, which is present also in constrained TMD, is not important for the present purpose of rapid generation of transition paths between conformations of a molecule. We note that, after the simulation structure is sufficiently close to the target structure (i.e., $\|r_i - A_i r_{T,i}\| < \Delta r_{\max} \forall i \in \{1, \dots, N\}$), eq 5 becomes identical to eq 4.

In addition, the simulated molecule will now experience a net force and torque. This does not occur in the original RTMD formulation because rotation/translation invariance is implicit in eq 2 (see Appendix A). One possible way to remove net force and torque is to apply Eckart constraints.^{35,69,70} This approach, however, requires modifying the MD integrator, which is difficult to do in some MD packages optimized for speed.⁷¹ Since applying external forces is typically easy, we instead compute a minimum-norm force correction \hat{f} to the forces in eq 5 using standard tools in numerical methods,^{72,73} such that the corrected forces $\hat{F}^* + \hat{f}$ sum up to zero and generate no net torque on the system. For the SFTMD calculations performed here, the RMS magnitude of \hat{f} is 5–10% of that of \hat{F}^* , indicating that the forces are not changed significantly by the correction. Details of the construction of \hat{f} are provided as Supporting Information.

2.2. Locally Restrained TMD Method. The locally restrained TMD method (LRTMD) is one of two methods

developed to test the effect of the global best-fit alignment between the simulation and target structures that is used as part of TMD methods (including SFTMD). Although the LRTMD method is similar in implementation to the standard RTMD, it is based on local regions of the protein molecule, each comprised of a small group of residues. The biasing forces are derived from the restraint potential U_{LRTMD} , which is built up from a sum of conventional RTMD potentials acting on subsets of atoms that are, generally, much smaller in size than the molecule itself. The potential has the form

$$U_{\text{LRTMD}}(\hat{r}, \hat{r}_T; t) = \frac{k_{\text{LRTMD}}}{2} \times \left(\underbrace{\left[\sum_{n=1}^M \rho(\hat{r}^n, A^n \hat{r}_T^n)^2 \right]^{1/2}}_{\delta(\hat{r}, \hat{r}_T)} - \delta_0(t) \right)^2 \quad (7)$$

For brevity, we have used A^n to denote $A(\hat{r}^n, \hat{r}_T^n)$, and the coordinate sets \hat{r}^n and \hat{r}_T^n correspond to different (possibly overlapping) subsets of the structure. Note that the best-fit rotation matrix A^n is recomputed for each of the M RMSDs above. Comparing eqs 7 and 2, one sees that the RMSD of eq 2 is replaced by a generalized distance δ , in which N Cartesian interatomic distances are replaced by M RMSDs between groups of atoms. The “hierarchical” potential in eq 7 is expected to enforce the global conformational transition *via* a superposition of a number of local small-scale transitions. The fact that the potential is constructed from local (small-scale), and not global (large-scale), transitions suggests that small-scale rearrangements will occur before large-scale ones. That this is the case is shown in section 3.4. In analogy to $\rho_0(t)$ in section 2.1, $\delta_0(t)$ is the prescribed equilibrium value of the potential at time t , which is reduced during a simulation from the initial value (computed such that the potential in eq 7 vanishes) and the final value (which we set to 0 Å). We note that the forces arising from the potential are easily calculated from eq 4 and the chain rule. To speed up force calculation in the cases for which the number of component RMSDs M is large (e.g., >50), the contributions to the forces arising from different RMSD components are evaluated on different processors, and subsequently combined using message-passing calls (MPI).

2.3. Steered Dynamics in Dihedral Angles. In addition to targeting dynamics in Cartesian coordinates described in the previous subsections, we investigated steered dynamics with a restraint applied to a set of dihedral angles (DSMD). In this approach, the following restraint potential was used

$$U_{\text{SMD}}(\hat{r}, \hat{r}_T; t) = \frac{k_{\text{SMD}}}{2} \times \left(\underbrace{\left[\frac{1}{N} \sum_{j=1}^N |\psi_j(\hat{r}) - \psi_j(\hat{r}_T)|^2 \right]^{1/2}}_{\epsilon(\hat{r}, \hat{r}_T)} - \epsilon_0(t) \right)^2, \quad (8)$$

in which ψ_j are the N dihedral angles included in the potential and $\epsilon_0(t)$ is reduced gradually from an initial value (computed such that the potential vanishes initially) to the final value (which we set to zero radians). To remove the ambiguity in the value of the difference $\psi_j(\hat{r}) - \psi_j(\hat{r}_T)$, which arises due to the periodicity of torsional angles, we use the value that has the smallest magnitude in the range $(-\pi, \pi]$. The dihedral angles that are included in the potential of eq 8 are discussed in section 2.6.

2.4. Calculation of Free Energy Profiles. The finite-temperature string method (FTS)^{15,16,74} was used to calculate the free energy (FE) profiles associated with transition paths obtained using RTMD, LRTMD, and DSMD.

A numerical implementation of the FTS algorithm has been described in ref 16. The essential difference in its use in the present calculations is that the transition tube is kept centered on the initial path, and is assigned a prescribed width of ~1 Å. The two constraints allow one to compare different paths directly on the basis of the associated free energies, since the transition tubes remain restricted to the vicinity of the corresponding paths. A drawback of the second constraint is the requirement to specify a value for the tube width *a priori*, which may lead to the neglect of important regions of conformational space, effectively overestimating the free energy at corresponding locations along the path. Although such errors may lead to the overestimation in the free energy along a single given path, they are expected to have a smaller effect on the comparison of different paths,³⁴ which is the subject of the present study. The use of this constraint is important for the present calculations so that the conformational space within the tube is small enough to be sampled in relatively short 5 ns MD simulations (see below). The FTS calculation simulation protocol is as follows. Trajectories of 1.5 ns in length were generated by RTMD, LRTMD, or DSMD, and were sampled in 1 ps intervals. The resulting set of 1500 structures was used to generate a sequence of 128 equispaced structures (string of replicas) using linear interpolation in Cartesian coordinates. To remove bad contacts generated by the interpolation procedure, while maintaining the equal spacing between adjacent replicas, 500 000 iterations of steepest descent minimization combined with a linear interpolation/reparametrization procedure²⁰ were performed. Reparametrization was done after every 20 steps of minimization. The resulting string was used to initialize 128 MD simulations, which were carried out subject to the constraint (here implemented using a restraint potential) that the corresponding MD replicas are only allowed to move in hyperplanes perpendicular to the initial path. In addition, as described above, each MD replica was restrained to stay within ~1 Å of the initial string by using a one-sided restraint potential centered on 0.75 Å. Temperature was maintained by use of Langevin dynamics, as described for the biased simulations (see below). Restraint forces acting on the hyperplanes were averaged for 5 ns, and subsequently integrated to yield the free energy profiles.¹⁶ The FTS simulations required about 24 h for each MD replica running on 32 2.1 AMD MagnyCours CPU cores.

2.5. Preparation of Structures. The X-ray crystal structure of Ca²⁺-bound CaM from *Paramecium tetraurelia* at 1.0 Å resolution⁵⁰ was obtained from the protein data bank (entry 1EXR). No X-ray structures of monomeric Ca²⁺-free apoCaM have been solved. Schumacher et al.⁵³ reported X-ray structures of a domain-swapped⁷⁵ dimeric apoCaM. As a consequence of domain-swapping in these structures, the C-terminal lobes of the two monomers are “intertwined”, such that separated monomers do not represent plausible monomeric solution structures. Therefore, we used the first structure from *Xenopus laevis* obtained by NMR⁴⁹ (PDB entry 1DMO). Because the CaM sequences from the two organisms differ at 18 residue locations, the following mutations were made to the 1DMO structure: D2E, S17A, T70S, M71L, D78E, T79Q, I85L, R86I, R90K, K94R, Y99L, E119D, N129D, Q135H, V136I, Q143R, T146V, and A147S (the first and last characters

correspond to residue codes in PDB files 1DMO and 1EXR, respectively). Missing coordinates for atoms in the mutated residues were added using internal coordinate tables in CHARMM.^{76,77} The mutated residues occur mostly on the surface in the C-terminal domain (see Figure 1).

Ca^{2+} ions were deleted from the simulation structures. Because the conformational state of CaM is known to depend on the Ca^{2+} concentration,^{49,55} and because the binding of Ca^{2+} to CaM is cooperative,⁵⁴ reproducing faithfully the dynamics of the open (1EXR) conformer as well as transitions between the open and closed forms would probably require including Ca^{2+} ions in the simulation. However, the absence of Ca^{2+} should not affect the relative performance of different TMD methods used in the present study. For both conformers, all histidines were singly protonated on the δ -nitrogen. The all-atom CHARMM22 topology and parameter files with the CMAP correction were used.⁷⁸ The FACTS solvation model was employed to approximate the effect of solvent.⁷⁹ Further details on the preparation of simulation structures and on the equilibration in the presence of restraints are provided in the Supporting Information.

The structures were equilibrated in the absence of restraints at 300 K for 1 ns using the Langevin dynamics thermostat with the friction coefficient γ set to 1 ps^{-1} . During the equilibration, two types of conformational changes were observed, which resulted in the deviation of the final equilibration structures from the initial structures by $\sim 4.5 \text{ \AA}$. These changes are quantified in terms of RMSD in Table 1. In the open structure,

friction that is typically absent in implicit solvent models (see, e.g., ref 34 for a discussion). A similar, albeit less pronounced closure was also observed for the *Cterm* domain. The second type of conformational change observed involves the relative motion of the *Nterm* and *Cterm* domains about the linker loop L4 (see Figure 1b). This motion was observed mainly for the closed structure, in which L4 is unwound, and is consistent with the flexible behavior of solution structures of *apo*CaM.^{49,80} To test whether EF-hand closure in the *Nterm* domain observed in the simulations was due to the removal of Ca^{2+} , we repeated the equilibration simulation in explicit solvent but with the Ca^{2+} ions bound to the structure. The EF-hands remained open for the entire 20 ns of the simulation trajectory, suggesting strongly that removal of Ca^{2+} is the cause of the EF-hand collapse (see also Table 1). We note that these results appear to be in conflict with the conclusions of Vigil et al.,⁵⁹ who proposed that the solution structure of the *Nterm* domain of calmodulin exists predominantly in the closed conformation but visits the open conformation frequently. The conclusions were based largely on the interpretation of $\sim 2.5 \text{ ns}$ MD simulations with X-ray scattering data. The radius of gyration in the MD simulations of the *Nterm* domain decreased from the initial value of $\sim 13 \text{ \AA}$ to $11.7\text{--}12 \text{ \AA}$, which the authors interpreted as EF-hand closure. In the present explicit water simulations, the radius of gyration of *Nterm* decreased over several nanoseconds from 12.9 \AA to $\sim 12.5 \text{ \AA}$, and remained at this value throughout the simulation. In comparison, the radius of gyration of *Nterm* in the closed structure is $\sim 11.5 \text{ \AA}$. In the simulations performed here in the absence of Ca^{2+} , the radius of gyration of *Nterm* decreased to $\sim 11.9 \text{ \AA}$. Thus, the present explicit water simulations support a Ca^{2+} -loaded solution structure of *Nterm* that is more open than closed. This structure appears similar to the solution structure of Ca^{2+} -loaded CaM derived from NMR by Chou et al.,⁵⁷ whose measurements ruled out a model in which the closed *apo* conformation is partially populated in the Ca^{2+} -bound state.

The relatively large RMSDs between the initial and the equilibrated structures reported in Table 1 appear consistent with the conclusions in refs 49 and 50 for the Ca^{2+} -free and Ca^{2+} -bound states, respectively, that each static structure represents a large number of conformations in solution. Further discussion is provided as Supporting Information.

2.6. Biased MD Simulations. The simulations performed in this study are summarized in Table 2. All of the biased simulations described below were initiated from the equilibrated structures. As the corresponding target structure, we chose to use the starting energy-minimized X-ray or NMR structure, as was done previously in TMD simulations of GroEL.⁴ With the use of the energy-minimized structures, all of the biasing methods except DSMD were able to generate transition paths with final simulation structures within $\sim 2 \text{ \AA}$ of the target structures. The DSMD simulations converged to the target structures only when then equilibrated target structures were used. The reasons for this behavior are discussed in section 3.5. Using equilibrated (rather than energy-minimized) structures in TMD or RTMD simulations did not produce significant differences in the final RMSD to the target structures, the sequence of events, or in the increase in temperature when the thermostat was switched off. The temperature was kept at 300 K using the Langevin dynamics thermostat with $\gamma = 1 \text{ ps}^{-1}$. This value of γ was sufficiently high to keep the temperature of the system at 300 K during the entire simulation, despite the fact that nonconservative work was performed on the system by the biasing forces. In the

Table 1. Conformational Change in the CaM Structures after Equilibration without Restraints as Quantified by RMSD Quoted in \AA (See Text)

domain	RMSD from X-ray 1EXR (open)	RMSD from NMR 1DMO (closed)
Open Structure		
heavy atoms	4.5	12.4
<i>Nterm</i>	4.5	3.5
<i>Cterm</i>	3	4.5
heavy atoms ^a	4.5	13
<i>Nterm</i> ^a	4.5	3.5
<i>Cterm</i> ^a	2.2	5.4
heavy atoms ^b	4	13
<i>Nterm</i> ^b	3	4.7
<i>Cterm</i> ^b	2.2	5.4
Closed Structure		
heavy atoms	13.3	5
<i>Nterm</i>	5.5	2.7
<i>Cterm</i>	5.4	2.2

^aEquilibration calculation was performed in explicit solvent.

^bEquilibration calculation was performed in explicit solvent and with four Ca^{2+} ions bound to CaM.

from which Ca^{2+} were deleted prior to simulation (see above), the EF hands in the *Nterm* domain (see Figure 1) undergo partial closure, which is consistent with the NMR study of Zhang et al.,⁴⁹ who found that the removal of Ca^{2+} leads to the closure of EF-hand motifs within CaM. As qualitative validation, we repeated the equilibration simulation of the open structure in explicit solvent, and observed a similar EF-hand closure. In the explicit water equilibration simulation, this conformational change occurred after 8 ns of simulation, consistent with the fact that explicit solvation introduces

Table 2. Summary of MD Simulations^a

simulation (no. of traj.)	$\Delta\rho_0/\Delta t$ (Å) $\Delta\delta_0/\Delta t^d$ (Å) $\Delta\varepsilon/\Delta t^e$ (rad)	k_{RTMD}^d k_{SMD}^e	Δr_{\max} (Å) cluster size ^d torsions ^e P_f (Å)	$\rho(t_{\max})$ (Å) O → C/C → O	t_{\max} (ns) O → C/C → O
RTMD(1)	-0.00001	500	n.a.	1.03/0.93	1.5/1.5
RTMD(10)	-0.00001	1000	n.a.	0.7/0.66	1.5/1.5
RTMD(1)	-0.000005	1000	n.a.	0.7/0.67	3.0/3.0
RTMD(1)	-0.00001	1500	n.a.	0.6/0.53	1.5/1.5
TMD ^c (10)	-0.00001	n.a.	n.a.	1.0/1.0	1.14/1.23
TMD ^{b,c} (1)	-0.00001	n.a.	n.a.	1.0/1.0	1.14/1.23
SFTMD(10)	-0.00001	1000	1	1.25/1.26	1.5/1.5
SFTMD ^b (1)	-0.00001	1000	1	1.0/n.p.	1.5/n.p.
SFTMD(1)	-0.00001	1000	5	0.7/n.p.	1.5/n.p.
SFTMD(1)	-0.00001	1000	3	0.7/n.p.	1.5/n.p.
SFTMD(1)	-0.00001	2000	1	0.94/1.05	1.5/1.5
LRTMD(10)	-0.00001	1000	6	0.8/0.5	1.5/1.35
LRTMD ^b (1)	-0.00001	1000	6	0.45/0.95	1.5/1.5
LRTMD ^b (1)	-0.00001	200	6	0.7/n.p.	1.5/n.p.
LRTMD(1)	-0.00001	1000	10	0.45/0.5	1.5/1.5
DSMD(1)	-6.7×10^{-7}	20000	Φ, Ψ	7.3/7.7	1.5/1.5
DSMD(1)	-9.4×10^{-7}	20000	$\Phi, \Psi, \chi_1, \chi_2$	4.7/6.25	1.5/1.5
DSMD(1)	-9.6×10^{-7}	40000	$\Phi, \Psi, \chi_1-\chi_5$	3.8/n.p.	1.5/n.p.
DSMDeq(1)	-9.4×10^{-7}	20000	$\Phi, \Psi, \chi_1, \chi_2$	4.4/5.6	1.5/1.5
DSMDeq(1)	-9.6×10^{-7}	40000	$\Phi, \Psi, \chi_1-\chi_5$	1.1/1.5	1.5/1.5
DSMDeq ^b (1)	-9.6×10^{-7}	40000	$\Phi, \Psi, \chi_1-\chi_5$	5.0/1.9	1.5/1.5
BMD ^c (1)	-0.000333	1000	n.a.	2.5/2.9	6.0/1.5
BMD ^{b,c} (1)	-0.000333	3000	n.a.	1.85/2.11	1.5/1.5
BMD ^c (1)	-0.000333	5000	n.a.	1.80/n.p.	1.5/1.5
RPTMD ^c (1)	n.a.	n.a.	0.0002	4.8/n.p.	1.5/n.p.
RPTMD ^c (1)	n.a.	n.a.	0.001	3.0/n.p.	3.0/n.p.
RPTMD ^{b,c} (1)	n.a.	n.a.	0.0005	4.0/n.p.	1.5/n.p.
RPTMD ^{b,c} (1)	n.a.	n.a.	0.001	3.4/n.p.	1.5/n.p.
RPTMD ^{b,c} (1)	n.a.	n.a.	0.003	2.1/n.p.	1.5/n.p.

^aTMD constraints and RTMD forces were applied only to the heavy atoms (1165 atoms). The number of corresponding simulations of a given type (performed identically but with different random seeds) is given in parentheses next to the simulation type. $\rho(t_{\max})$ is averaged over identical simulations. The force constants k_{RTMD} and k_{LRTMD} are given in units of kcal/mol/Å² (applicable to RTMD, SFTMD, and LRTMD simulations). The force constant k_{SMD} is given in units of kcal/mol/rad² (applicable to DSMD simulations). For brevity, "O" and "C" denote the open (1EXR) and closed (1DMO) structures of CaM. The RMSD between the heavy atoms in the open and closed structures is ≈ 13 Å. t_{\max} corresponds to simulation duration. $\rho(t_{\max})$ is the RMSD between the simulation and target structures at the end of simulation. $\Delta\rho_0/\Delta t$, $\Delta\delta_0/\Delta t$, and $\Delta\varepsilon_0/\Delta t$ represent the change per simulation timestep in the target values of the variables ρ , δ , and ε , respectively (see Methods). Cluster size is given in residues (see Methods); n.a. indicates that the parameter is not applicable to the type of simulation; n.p. indicates that the simulation was not performed. In DSMDeq simulations, the equilibrated X-ray and NMR structures were used as the target structures. ^bSimulation was performed with the thermostat turned off (see text). ^cSimulation results are shown in the Supporting Information. ^dOnly for LRTMD simulations. ^eOnly for DSMD simulations. ^fOnly for RPTMD simulations (defined in the Supporting Information).

absence of the thermostat, the temperature at the end of the simulations increased by 60–120 K depending on the simulations (see section 3). Additional details specific to each biasing method are given below.

Constrained TMD Simulations. The target structure was rotated to obtain a minimum-RMSD fit to the simulation structure at each simulation step. The maximum allowed RMSD between the simulation and target structures was decreased by 1×10^{-5} Å per time step (1 fs). With the RMSD value of ≈ 13 Å between the two energy-minimized structures and the prescribed final value of 1 Å, the duration of each TMD simulation was ≈ 1 200 000 steps (1.2 ns). Ten TMD simulations were performed in each direction (i.e., open → closed and closed → open) using different random seeds to initialize the thermostat.

Restrained TMD and Scaled-Force TMD Simulations. Structures were initialized as described above for TMD simulations, and the same thermostat was used. The force

constant k_{RTMD} was in the range 500–2000 kcal/mol/Å², which was sufficient to reduce the RMSD between the simulation structure and the target structure to ≈ 1 Å (see Figure 2a and Table 2). Using higher values for the force constant results in a smaller value for the RMSD but does not change the computed paths qualitatively (e.g., the "large-scales-first" order persists regardless of the value of the force constant). The maximum allowed RMSD between the simulation and target structures was decreased by 1×10^{-5} Å per time step, and the duration of each simulation was 1.5 ns (in all but one RTMD simulation, in which $\Delta\rho_0 = 5 \times 10^{-6}$ Å, and $t_{\max} = 3$ ns; see Table 2). The maximum allowed RMSD between the simulation and target structures was held fixed after it reached zero (after ≈ 1.3 ns of simulation). The value of Δr_{\max} in eq 6 was set to 1, 3, and 5 Å in different SFTMD simulations. For $\Delta r_{\max} = 1$ Å, which corresponds to the greatest force reduction compared to regular RTMD, 10 simulations with different random number seeds were performed in each direction.

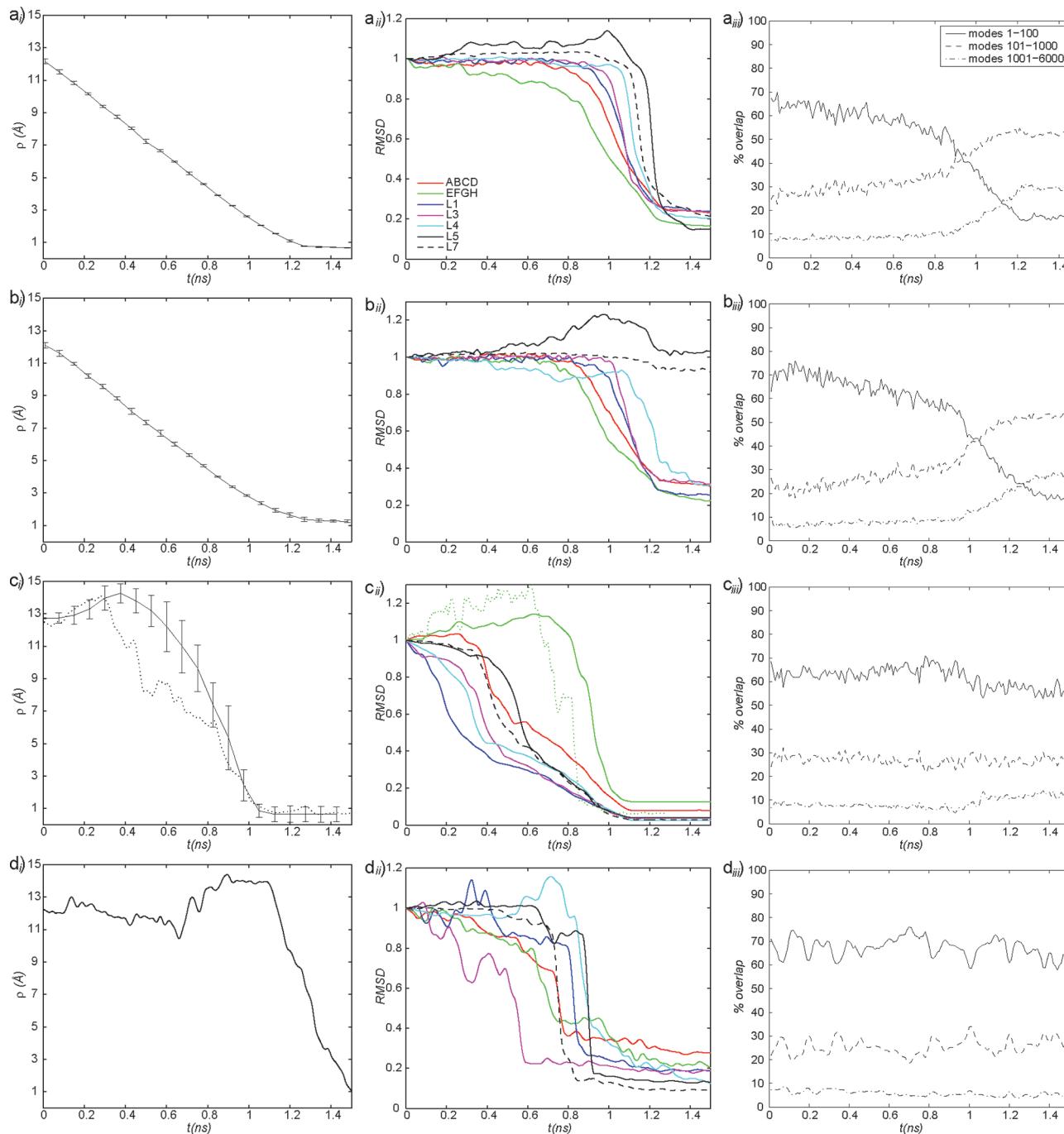


Figure 2. (a_i, b_i, c_i, d_i) Evolution of the heavy-atom RMSD between the simulation and target structures (ρ) in (a_i) RTMD simulations, (b_i) SFTMD simulations, (c_i) LRTMD simulations, and (d_i) DSMD simulations. Error bars represent standard deviations computed from all trajectories in the corresponding ensemble (results of one DSMD simulation are shown); solid line, O → C; dotted line, O → C, $k_{\text{LRTMD}} = 200 \text{ kcal/mol}/\text{\AA}^2$ (i only). (a_{ii}, b_{ii}, c_{ii}, d_{ii}) Evolution of the normalized heavy atom RMSD computed for different CaM subdomains (see Figure 1). Curve labels are provided in a_{ii}. The dotted green curve in c_{ii} corresponds to the portion of one trajectory that was chosen for subsequent free energy analysis. (a_{ii}) RTMD O → C; (b_{ii}) SFTMD O → C; (c_{ii}) LRTMD O → C; (d_{ii}) DSMD O → C. (a_{iii}, b_{iii}, c_{iii}, d_{iii}) Evolution of the overlap coefficients between the normal mode vectors (NMV) and trajectory displacements in (a_{iii}) RTMD simulations, (b_{iii}) SFTMD simulations, (c_{iii}) LRTMD simulations, and (d_{iii}) DSMD simulations. All curves correspond to the O → C direction. Curve labels are provided in a_{iii}. NMV were computed using the open structure. The modes corresponding to rigid body motions are not included (i.e., numbering starts from the first mode that has a positive frequency).

Locally Restrained TMD Simulations. Each of the M sets of atoms from which the coordinates \hat{r}^n , for $n \in \{1, \dots, M\}$ were computed, was composed of all non-hydrogen atoms in a cluster of six adjacent residues. The clusters were chosen to be rather small to demonstrate clearly the existence of a broad ensemble of paths, in which small-scale motions do not follow

but can even precede the large-scale ones. Each cluster except the first overlaps with the preceding cluster by about 50%, such that, e.g., cluster 1 corresponds to residues 1–6 and cluster 2 corresponds to residues 4–9, etc. For the 148-residue sequence of CaM studied here, this scheme results in $M = 49$ clusters, with cluster 49 spanning residues 145–148. The use of

overlapping clusters is not required but ensures that the global target conformation is attained at the end of the simulation. Each cluster is composed of approximately 45 atoms. The effect of increasing the cluster size to 10 residues is discussed in section 3.4.

Dihedral SMD Simulations. Three sets of protein torsional angles were used. The first set contains all of the Φ and Ψ backbone torsion angles (294 in total), the second set also contains all of the χ_1 and χ_2 side chain torsion angles (489 in total), and the third set contains additional torsional angles sufficient to specify the conformation of the entire residue (590 in total; the restraints do not involve hydrogen atoms). For example, for arginine, seven dihedral angles are included, Φ , Ψ , and χ_1 – χ_5 ; for phenylalanine, only angles up to χ_2 are included because of the rigidity of the aromatic ring. Adding to the restraint set the ω dihedral angles, which specify the orientation of the peptide bonds (typically trans), did not make a significant difference in the results. A comprehensive list of the dihedral angles in the third set is provided in Table S2 of the Supporting Information. The RMS sum of the dihedral angle difference ε is steered linearly from its initial value (≈ 0.8 for set 1 and ≈ 1.1 for sets 2 and 3) to zero over 1 200 000 simulation steps; the restraints are held at the final value for 300 additional steps, for a total of 1.5 ns of simulation at 1 fs per step. Initially, target structures for the DSMD simulations were taken to be the corresponding energy-minimized X-ray (1EXR) and NMR (1DMO) structures. Because the final structures from the corresponding simulations were quite far from the target structures, the DSMD simulations were repeated using equilibrated target structures, which resulted in a markedly closer approach to the target structures. The reasons for this behavior are discussed in section 3.5.

2.7. Normal Mode Analysis. Calculation of Normal Modes. Energy-minimized structures from section 2.5 were subjected to additional minimization in the absence of restraints until the difference in the potential energy at two consecutive steps was less than 10^{-7} . This required $\sim 30\,000$ iterations with the ABNR minimizer in CHARMM. Normal mode vectors (NMV) were computed for each structure with CHARMM, as described in ref 81, using a second-order central finite-difference approximation for second derivatives (the solvation model FACTS, which was employed in the NMA, currently does not provide analytical formulas for the second derivatives). The first six NM represent rigid body motion and were discarded. (The frequencies corresponding to the rigid body modes were scattered around zero with the mean deviation from zero of 0.066 cm^{-1} .) The remaining 6810 modes were scaled to unit norm and stored for visualization and projection analysis (described below). The lowest non-rigid-body NM frequency was $\approx 1\text{ cm}^{-1}$.

Normal Mode Projection Analysis. To quantify the extent of the large-scale displacement preference in the simulations, we project the displacements obtained from simulation trajectories onto the NMV computed for the open and closed conformations of CaM. This approach, which we call normal-mode projection analysis (NMPA), provides a quantitative measure of the dominant scales of motion. Since large-amplitude motions of biomolecules typically correspond to a low-frequency region of the normal mode spectrum^{45,82} and the smallest-amplitude modes (such as bond and angle vibrations) correspond to the highest-frequency region, one can infer the size of the dominant scale involved in the trajectory displacement by determining the frequency of the

normal modes with which the trajectory displacement vector has the highest overlap.

To reduce computation time, trajectories obtained from TMD and RTMD calculations were sampled in 10 ps time increments. To remove rigid-body motion, the trajectory frames were aligned such that any two consecutive structures have minimum mass-weighted RMSD. A normalized displacement vector was calculated between each pair of consecutive trajectory frames, and projected onto the NMV. The normalization was such that the squares of the projections (also called overlap or involvement coefficients^{82,83}) add up to unity. We also tried computing the overlap coefficients after realigning all frames to minimize the mass-weighted RMSD to the corresponding energy-minimized structures; the results were qualitatively similar to those obtained using the first alignment method described.

3. RESULTS

3.1. Overview. The form of the TMD equations (section 2.1) suggests that the “large-scales-first” phenomenon may be caused by the linear dependence of the atomic force magnitude on the length of the displacement of a given simulation atom from its position in the target structure (see eq 4).

This possibility was investigated by comparing trajectories generated using the standard TMD and RTMD methods with those generated using a scaled-force RTMD method (section 2.1), in which magnitudes of the forces acting on the simulation atoms are capped at a maximum value. By employing a projection of trajectory displacements onto normal mode vectors (see section 2.7), we find that TMD and RTMD trajectories tend to follow low-frequency normal modes during the initial $\sim 60\%$ of the transition simulation. In the final portion of the trajectories, most of the motion occurs along higher-frequency modes, with the contribution of such modes increasing rapidly as the target structure is approached. During this stage, a large amount of nonreversible work is dissipated, which accounts for a large temperature increase in the simulations in which the thermostat was switched off. The dissipated work appears to arise from forcing the system to overcome barriers associated with the rearrangements of small loops. Simulation trajectories obtained using the SFTMD method revealed that the “large-scales-first” feature was still present (see below). This result indicated that the direction of the applied forces has a larger influence on the evolution of TMD trajectories than the relative magnitudes of the forces. Two new biasing methods are used to generate additional transition paths. In both methods, the direction of the force vector is not related to global orientation. The first method, LRTMD, is a straightforward generalization of the original RTMD method, in which the bias potential corresponds to an RMS sum of conventional RTMD potentials based on small localized residue clusters in the protein (section 2.2). In the second method, called dihedral (D)SMD, steering forces are applied to a prescribed set of dihedral angles (see section 2.3). The “large-scales-first” feature is absent from paths produced by both methods. A direct examination of the corresponding trajectories shows that the transitions in the small scales often (although not always) precede transitions in the large scales. An additional feature of the trajectory ensemble generated by the LRTMD method is that thermostatted trajectories initiated with different random seeds are substantially more diverse than those generated by the standard TMD/RTMD methods. The LRTMD path ensemble (used here to refer to a collection of

trajectories rather a thermodynamic ensemble) is therefore broader in comparison to TMD/RTMD ensembles. It is likely that these trajectories correspond to a larger number of physical transition paths. We use the finite-temperature string method^{15,16,74} (see section 2.4) to calculate the free energy profiles along a narrow transition tube associated with each path. The free energy barriers along the paths are found to be similar, suggesting that the transitions proceed *via* paths with similar probabilities.

3.2. RTMD Simulations. First, we discuss ensembles of transition paths generated with standard TMD and RTMD methods and quantify the “large-scales-first” feature of the paths. We focus on twenty 1.5 ns transition path trajectories that were generated with the RTMD method in the open → closed direction, as described in Methods and summarized in Table 2. Results are shown in the main text only for the RTMD simulations because trajectories generated using TMD show very similar behavior. Additional results are given in the Supporting Information. The evolution of the RMSD function $\rho(t) \equiv \rho(\hat{r}(t), A\hat{r}_T)$ (defined in eq 3) is shown in Figure 2a_i for the RTMD simulations. The decrease in ρ is linear, in accordance with the driving protocol outlined in section 2.6. The small error bars computed from the trajectory ensemble indicate that the actual RMSD ρ remains close to the target value ρ_0 throughout the simulations. To examine the order of events in the transitions, seven elements in the CaM structure were selected (see Figures 1 and 2a_{ii}), and the evolution of the normalized best-fit RMSDs between the atoms in each element was computed. The best-fit involved removal of the rigid-body motion, as required to determine whether a given element has reached its target conformation in the sense of RMSD. The normalization ensures that all curves start at unity. The transition of a given element appears as a decrease in the RMSD. Two of the elements are essentially the *Nterm* and *Cterm* domains of CaM excluding loops, namely, helices A, B, C, and D (*Nterm*) and helices E, F, G, and H (*Cterm*) (the structure of CaM is shown in Figure 1 where the structural elements are indicated; they are defined in the Supporting Information). The remaining five elements are loops L1, L3–L5, and L7. Loops L2 and L6 are excluded to make the plots clearer. Figure 2a_{ii} indicates that the larger elements ABCD and EFGH undergo transition before the remaining loops. It is instructive to examine the simulation structure at $t \approx 1$ ns for one of the simulations taken from the open → closed RTMD ensemble (Figure 3). At this time, the simulation structure is within ≈ 3 Å RMSD of the target structure (relative to the starting value of 12.4 Å). Consistent with the plots in Figure 2a_i, Figure 3 confirms that the overall conformation of the simulation structure is close to the closed target structure. The insets in Figure 3 show, however, that the conformation of all of the loops is very different from those in the target structure. The above results collectively illustrate what we call the ‘large-scales-first’ feature of RTMD and TMD simulations, namely, the precedence of large-scale conformational motions (e.g., *Nterm* and *Cterm*) over small-scale ones (e.g., L1–L7). A more precise analysis can be obtained by projecting the RTMD trajectory displacements onto the normal mode vectors (NMV) calculated from the energy-minimized end point structures, as described in section 2.7. The time evolution of the overlap coefficients for the open → closed transition is shown in Figure 2a_{iii}. The corresponding plot in the reverse direction is very similar and is not shown. Figure 2c shows that ~50–70% of the motion observed in the first 1 ns of the RTMD simulations can

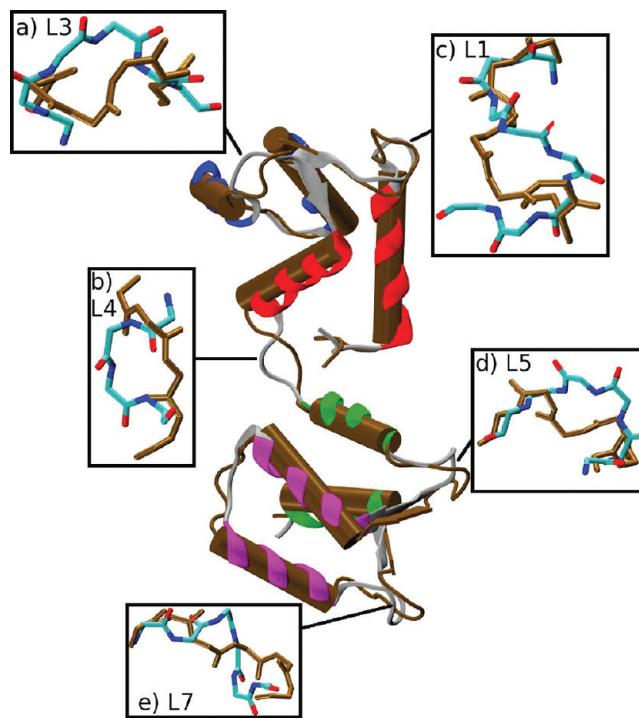


Figure 3. Superposition of an instantaneous simulation structure from the open → closed RTMD ensemble at $t = 1$ ns (colors) onto the target structure (brown). The regions separating the helices in the simulation structure are shown as gray ribbons. [The β -strand of EF-hand 2 (helices C and D in Figure 1) partially blocks the view of helix B, which remains intact in both structures.] The insets provide an expanded overlay of loops L1–L7 (L2 and L6 are omitted).

be captured by the 100 lowest-frequency NMV computed from the open conformation (out of 6810 vectors in total); 20–30% of the motion is accounted for by modes 101–1000, and the remaining 10% is captured only by the higher-frequency modes 1001–6000.

In the portion of the simulations from 1 to 1.5 ns, as the simulated structure approaches the target, the low-frequency NMV from the starting structure provide a progressively worse description of the dynamics (~18% near the end). If the NMV computed from the target structure are used for comparison (see the Supporting Information), the improvement in the overlap is modest (~26% near the end), which indicates that low-frequency modes from neither NMV set describe well the motions in the late stages of TMD simulations. The scalar products between the normalized NMV derived from the two structures are shown in the Supporting Information. The overlap between the 10 lowest-frequency NMV is significantly lower (0.3–0.6) for the CaM conformations than observed for, e.g., myosin⁴⁶ (~0.8) because the two CaM conformations are less similar than the two myosin conformations (whose motor domains differ by ~3.5 Å).

Projection of the displacements generated in the portion of the RTMD trajectories from 1 to 1.5 ns onto the remaining 6710 higher-frequency NMV computed from either conformation results in a progressively better overlap as the simulated structure approaches the target. Thus, the evolution of overlap coefficients in the projection analysis provides a quantitative measure of scale preference in transition paths computed by RTMD. Additional insight into the “large-scales-first” feature of the transition is obtained by monitoring the kinetic temperature

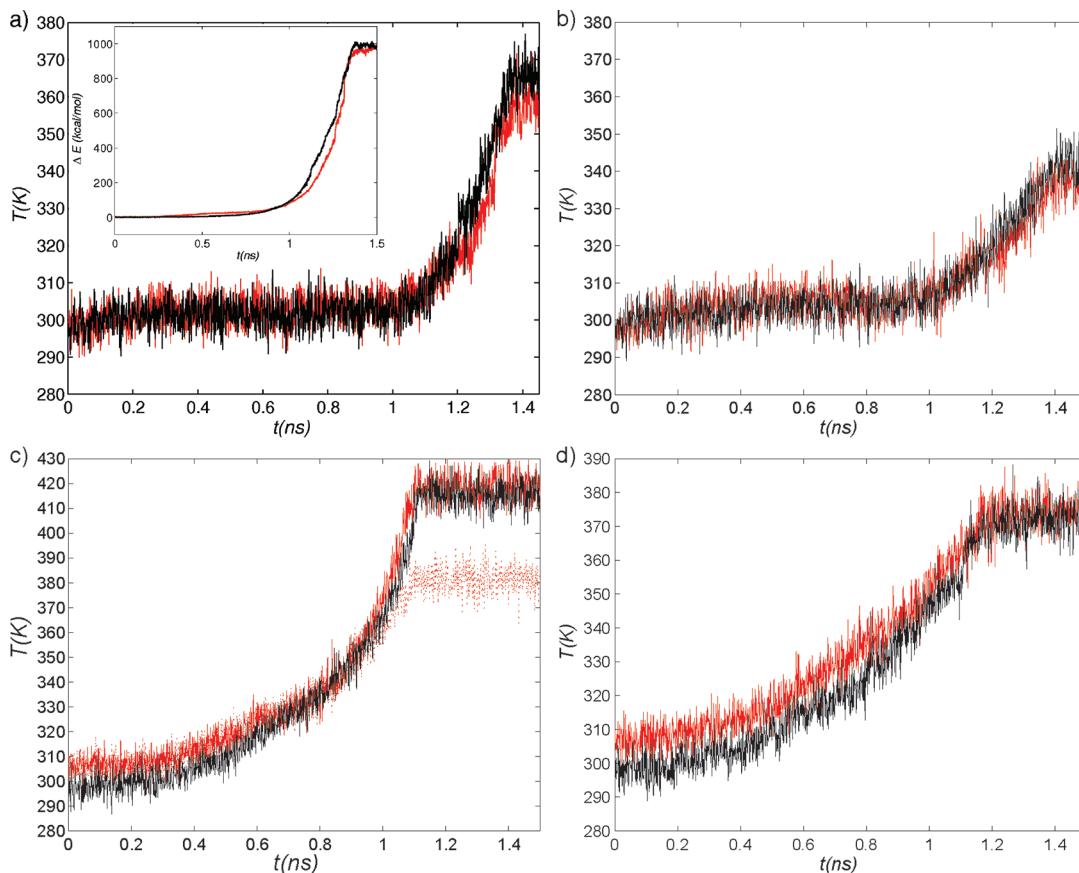


Figure 4. Temperature evolution in the absence of thermostating in (a) RTMD, (b) SFTMD, (c) LRTMD, and (d) DSMD simulations. Red, solid: $O \rightarrow C$; black, solid: $C \rightarrow O$; red, dotted: $O \rightarrow C$, $k_{\text{LRTMD}} = 200 \text{ kcal/mol}/\text{\AA}^2$ (c only). The inset in part a shows the difference between the total energy of the simulation and the initial total energy. The energy is averaged over 0.1 ns intervals to yield smoother plots.

in the RTMD simulations performed in the absence of a thermostat (see Table 2). Figure 4a shows that the temperature begins to increase rapidly around 1 ns into the simulation, beginning at $\sim 300 \text{ K}$ and reaching $\sim 360 \text{ K}$ by the end of the simulation. The inset of Figure 4a shows that the increase of the total energy parallels closely the temperature increase, and reaches $\simeq 1000 \text{ kcal/mol}$ at the end of the simulation (the kinetic and potential energies contribute approximately 500 kcal/mol each to the total increase). The onset of the temperature increase coincides with the increase in the overlap between the trajectory displacements and the higher-frequency NMV (Figure 2a_{iii}), suggesting that the temperature increases due to motions along high-frequency modes. In normal mode analysis, the potential energy in the vicinity of a local minimum is approximated to second order as

$$\begin{aligned} E(\hat{r}_0 + \Delta\hat{r}) &= \frac{1}{2}\Delta\hat{r}^t \nabla^2 E(\hat{r}_0)\Delta\hat{r} \\ &= \frac{1}{2}\Delta\hat{r}^t V^t N^2 V \Delta\hat{r} \end{aligned} \quad (9)$$

where V is the eigenvector matrix of the Hessian ($\nabla^2 E$) evaluated at \hat{r}_0 (i.e., V contains the normal modes), N is the diagonal frequency matrix, the atom coordinates \hat{r} have been scaled by mass ($r_i \rightarrow (m_i)^{1/2}r_i$), and $(\cdot)^t$ denotes matrix transpose. Equation 9 implies that a displacement of $\Delta\hat{r}$ along a given mode requires more energy for high frequency modes than for low-frequency modes. The analysis shows that transition paths generated by RTMD are expected to cross

the lowest barriers first and the highest barriers last. Qualitatively, by following normal modes in the order of increasing frequency, the transitions in RTMD occur along paths that encounter the least local resistance. In the next sections, we employ the variants of the RTMD method described in Methods with the aim of removing the “large-scales-first” feature characterized in this section. The above analysis applies equally to the constrained TMD algorithm (see Figure S3 in the Supporting Information).

3.3. Scaled-Force TMD Simulations. As discussed in section 2.1, the magnitude of the biasing force acting on an atom in RTMD (or TMD) is proportional to the length of the displacement vector from the current atomic position to the target position. This relationship leads to a distribution of the forces illustrated in Figure 5a, in which the initial (open) simulation structure is colored according to the total magnitude of the force acting on the corresponding residue. The force magnitude correlates with the approximate distance to the target (closed) structure. To determine whether such a force distribution can, by itself, cause the observed “large-scales-first” pattern of conformational change, we performed 20 simulations using the SFTMD method, in which the force cutoff was $\Delta r_{\text{max}} = 1 \text{ \AA}$. Since this value is more than 10 times smaller than the RMSD between the initial and the target structure ($\simeq 13 \text{ \AA}$), the level of force scaling will be large during most of the simulation, until the RMSD between the simulation target structures approaches Δr_{max} . (The high level of force scaling was chosen for a clearer comparison between the SFTMD and conventional RTMD trajectories.) The evolution of the average

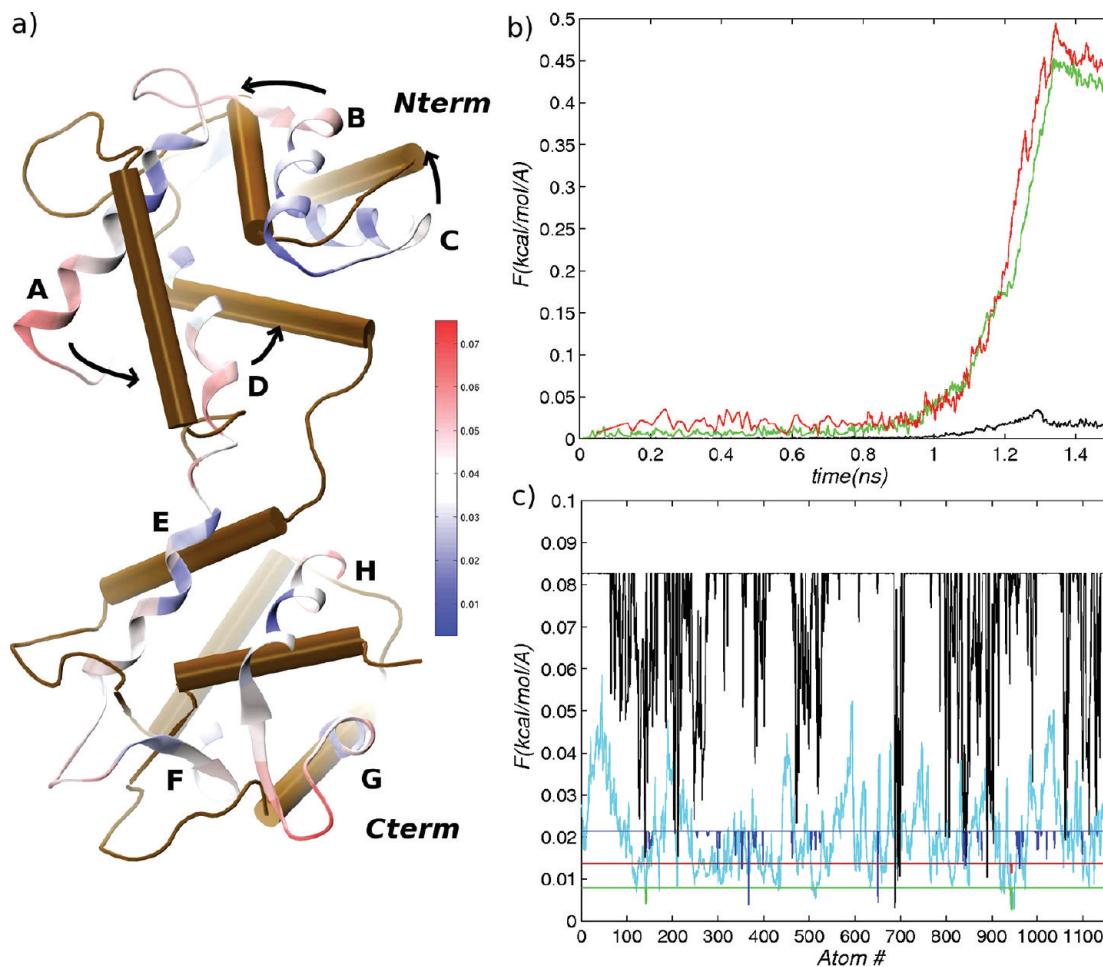


Figure 5. (a) Superposition of an instantaneous simulation structure taken from the open \rightarrow closed RTMD ensemble at $t = 500$ ps onto the target structure. The simulation structure is colored by the magnitude of the biasing force, and the target structure is in brown. The direction of the conformational change in the N-terminal EF-hands is indicated by black arrows. (b) Evolution of the average force magnitude acting on the simulation atoms. red: RTMD simulation, O \rightarrow C, $k_{\text{RTMD}} = 1000$ kcal/mol; green: SFTMD simulation, O \rightarrow C, $k_{\text{RTMD}} = 1000$ kcal/mol; black: SFTMD simulation (force correction; see section 2.1). (c) Profiles of instantaneous biasing force acting on the simulation atoms. red: SFTMD at $t = 100$ ps; green: SFTMD at $t = 500$; blue: SFTMD at $t = 1000$ ps; black: SFTMD at $t = 1100$ ps; cyan: RTMD at $t = 500$ ps. SFTMD simulations were performed with $\Delta r_{\max} = 1 \text{ \AA}$ (see text).

magnitude of the force acting on the simulation atoms is given in Figure 5b. As expected from the discussion, the average forces in SFTMD are generally lower than those in RTMD due to the force scaling. The atomic distributions of the biasing force magnitudes in the SFTMD method are shown in Figure 5c at $t = 100, 500, 1000$, and 1100 ps. For comparison, the distribution of the biasing forces in the RTMD method at $t = 500$ ps is also shown. During the first 500 ps of the simulations, the atomic forces are approximately equal in magnitude for all atoms. As the simulated structure approaches the target and the RMSD target ρ_0 approaches Δr_{\max} (e.g., $t = 1100$ ps), a nonuniform distribution of force magnitudes is observed, as expected from the discussion in section 2.1. Figure 2d shows that the final simulation structure approaches the target to within $\approx 1.25 \text{ \AA}$, averaged over the 10 trajectories in each direction (see also Table 2). This value is slightly higher than the one obtained using standard RTMD ($\approx 0.7 \text{ \AA}$). In addition, the standard deviations of the final RMSD are higher for the SFTMD case.

The higher mean value and standard deviations are caused by the generally smaller forces in the SFTMD method as a result of force-limiting, despite the use of the same force constant to

drive the transition (see Table 2). Although a closer approach to the target structure can be achieved by increasing the force constant, the present focus is on the “large-scales-first” phenomenon, which is still present in the SFTMD trajectories. Figure 2b_{ii} is similar to Figure 2a_{ii}, showing that the smaller loops undergo a transition after the larger structures ABCD and EFGH. In addition, because the forces on loop L5 are relatively small, due to the force-limiting, L5 generally does not reach the target conformation (although the overall RMSD to the target structure is around 1 \AA). Figure 2b_{iii} confirms that the transitions proceed initially along low-frequency modes, and end along high-frequency modes, as was observed for the TMD/RTMD simulations. Evolution of the temperature in an SFTMD simulation performed without the thermostat (Figure 4b) is similar to that obtained for the RTMD simulations, and implies that the simulation system moves along high-frequency/high-energy modes at the end of the simulation. The overall increase in the temperature (from 300 to 340 K) is somewhat less than that for the corresponding RTMD simulation (360 K) because of the smaller force magnitudes relative to RTMD (see above).

The similarities in the trajectory analysis results obtained for the RTMD and SFTMD simulations show that reducing the magnitudes of the forces applied to simulation atoms, as done in the SFTMD method, is not, by itself, sufficient to produce a significant difference in the trajectory ensembles. As mentioned in the Introduction, the methods have in common the global best-fit alignment procedure,⁶⁴ which effectively determines the direction (as well as the magnitude) of the applied biasing forces. In the next section, we discuss ensembles of trajectories generated with the locally restrained TMD method, LRTMD (section 2.2), which does not use a global alignment. As a consequence, the early large scale motions are absent from these trajectories.

3.4. Locally Restricted TMD Simulations. In the LRTMD simulation method, the single global RMSD potential of eq 2 is replaced by an RMS sum of a set of local potentials, as described in section 2.2. Correspondingly, the global best-fit alignment is replaced by a set of alignments, each based on a subset of the structure. The evolution of the RMSD function ρ (computed *a posteriori* using a global best-fit alignment) shown in Figure 2c_i for the LRTMD trajectory ensembles indicates that the final simulation structures are within ~ 0.6 Å of the corresponding target structures. In contrast to the results of the previous simulations, the profiles of ρ are not generally monotonic; e.g., the average ρ for the open \rightarrow closed transition increases in the first 400 ps of the transition. The RMSD function ρ computed for the LRTMD trajectory ensemble is much higher in this region than that for the TMD or the RTMD ensembles. This behavior is expected, since in RTMD (and TMD²⁴) the bias potential acts to restrain $\rho(t)$ to a monotonically decreasing ρ_0 but in LRTMD the simulation value of $\rho(t)$ is determined indirectly by enforcing a different restraint potential. The high standard deviations in Figure 2c_i suggest that the corresponding trajectory ensemble has a higher variability than that obtained in TMD and RTMD. For a qualitative measure of the variability in the trajectory ensembles computed in this study, we define the average RMSD between all trajectory frames computed at a particular time t :

$$\bar{\rho}(t) = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j>i}^N \rho(\hat{r}_i(t), A\hat{r}_j(t)) \quad (10)$$

where N is the number of trajectories in the ensemble and $A(\hat{r}_i(t), \hat{r}_j(t))$ is a best-fit rotation matrix. Profiles of $\bar{\rho}(t)$ for the ensembles generated with TMD, RTMD, and LRTMD are compared in Figure 6. Consistently with the standard deviations in Figure 2c_v, the LRTMD ensemble exhibits the highest variability over the first nanosecond; the variability is much smaller in the final 0.5 ns portion of the trajectories because the simulation structures are essentially converged to the target structures (to within ~ 1 Å) after 1 ns. This suggests that the ensemble of unbiased transition paths may be more meaningful than those found with standard TMD/RTMD simulations; i.e., the conformational transition can occur *via* a larger number of paths. The high variability also suggests that the conformational transitions in the small domains that define the LRTMD potential (eq 7) are somewhat independent, and need not follow a particular order. Figure 2c_{ii} demonstrates that the precedence of large scales in the transition is no longer present. With the choice of six-residue clusters, transitions in the loops L1–L7 are seen to precede those in the EF-hand pairs ABCD and EFGH. A simulation structure extracted from

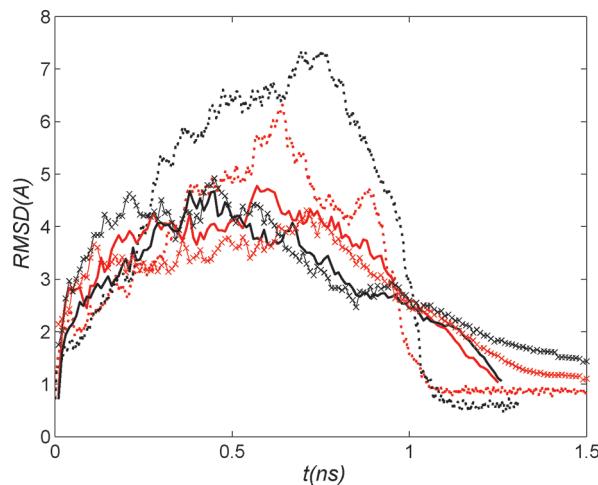


Figure 6. Average RMSD between corresponding instantaneous structures computed for the TMD, RTMD, and LRTMD trajectory ensembles (see text). red solid line, TMD ($O \rightarrow C$); black solid line, TMD ($C \rightarrow O$); red \times , RTMD ($O \rightarrow C$); black \times , RTMD ($C \rightarrow O$); red dotted line, LRTMD ($O \rightarrow C$); black dotted line, LRTMD ($C \rightarrow O$).

the open \rightarrow closed LRTMD trajectory ensemble at $t = 600$ ps is shown in Figure 7. Although the global overlap with the target

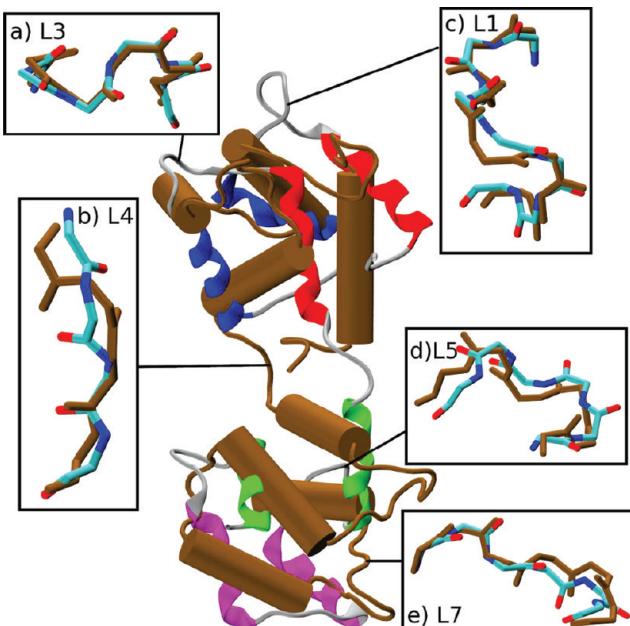


Figure 7. Superposition of an instantaneous simulation structure from the open \rightarrow closed LRTMD ensemble at $t = 600$ ps (colors) onto the target structure (brown). The insets provide an expanded overlay of loops L1–L7 (L2 and L6 are omitted).

structure is quite poor (~ 11 Å), the individual loops fit to the target structures well (RMSD of ~ 1 Å).

The evolution of the overlap coefficients between the LRTMD trajectory displacements and the NMV are shown in Figure 2c_{iii}. It is clear that, during all stages of the transition, the trajectory motions are described well by the low frequency NMV. In particular, at the end of the conformational transition (1.5 ns), the 100 lowest-frequency NMV computed from the energy-minimized closed (target) structure account for >70%

of the trajectory motion. This behavior is different from that seen in the RTMD/TMD ensembles for the end of the transition, in which the overlap between the lowest-frequency motions falls to 20–30%, and the trajectory motions follow predominantly the remaining high-frequency NMV (Figure 2a_{iii} and b_{iii}).

To investigate the effect of the cluster size on the scale precedence in the transition, two LRTMD simulations were performed with the cluster size increased to 10 adjacent residues per cluster (see Table 2). The evolution of the RMSD between the simulation and target structures is provided in Figure S4 in the Supporting Information; it is qualitatively similar to that described above for the LRTMD simulations with six-residue clusters. The overlap coefficients between the trajectory displacements and the NMV, shown in Figure 8,

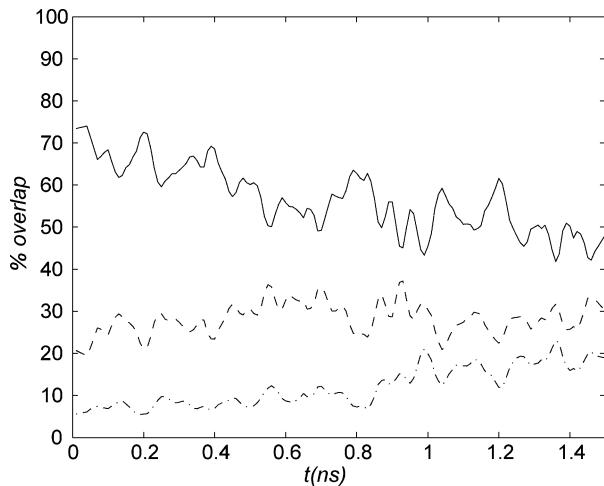


Figure 8. Evolution of the overlap coefficients between the NMV and trajectory displacements in the O → C LRTMD simulation with 10-residue clusters (see text). The labeling of traces is the same as in Figure 2c_{iii}.

suggest a slight precedence of larger scales over smaller scales. Larger cluster sizes or more complex cluster composition were not tried. However, in view of the results, we expect that larger clusters will favor the precedence of large-scale motions, and smaller clusters will favor small-scale motions. This clearly raises a concern regarding the most physically meaningful sequence of motions (see section 3.6).

It was shown in sections 3.2 and 3.3 that the increase in the high-frequency motions at the end of the trajectories is accompanied by a rather abrupt temperature increase (in the simulations with the thermostat switched off). Two LRTMD simulations using six-residue clusters were performed in the absence of a thermostat (Table 2). The time evolution of the temperature is shown in Figure 4c. In contrast to the RTMD case, the temperature increase is gradual, with a progressively higher slope, and begins near the start of simulation. This behavior appears to be consistent with the explanation that the temperature increase arises due to the crossing of barriers associated with small-scale motions, which, in the LRTMD cases, take place concomitantly with the large-scale motions.

In the RTMD simulations without a thermostat, the temperature rises to ~360 K, and in the corresponding LRTMD simulations, the final values are ~415 K. These differences appear to be due to the use of different biasing potentials in RTMD and LRTMD (eq 2 vs eq 7) with the same

forcing constant ($k = 1000 \text{ kcal/mol}/\text{\AA}^2$). Since the biasing potentials are qualitatively different, using the same force constant does not necessarily produce equivalent biasing conditions. As a test, we performed an additional LRTMD simulation in the open → closed direction with the force constant set to 200 kcal/mol/ \AA^2 (Table 2). The final global best-fit RMSD between the simulation and target structures was ~0.7 Å, similar to the main LRTMD case (see Figure 2c), but the final temperature was substantially lower, at ~380 K (see Figure 4c). Decreasing the force constant to $k = 40 \text{ kcal/mol}/\text{\AA}^2$ results in an even smaller temperature increase (~345 K) but at the expense of increasing the RMSD between the final simulation and target structures to ~5 Å. The results imply that it is difficult to decide based on the evolution of the temperature in the absence of a thermostat which biasing method results in smaller nonconservative work performed on the simulated system. Rather, the temperature evolution in all of the methods underscores the possibility that the transition paths are far from minimum free energy paths. Other methods, e.g., annealing in trajectory space³² or the string method,^{15,16,74} may be used to equilibrate biased trajectories, although for large biomolecular systems such methods are expensive computationally. Biased MD simulations performed in this study (see Table 2 and supporting text and Figure S6 in the Supporting Information) suggest that obtaining unbiased trajectories using trajectory annealing³² would require extremely long simulation trajectories.

Each free energy calculation using the string method reported here required a hundred thousand CPU core hours to obtain converged results (see section 3.6).

3.5. Dihedral SMD Simulations. In the DSMD method, the biasing potential is based on the RMS difference between chosen dihedral angles in the simulation and target structures, and does not involve best-fitting. Three sets of dihedral angles were chosen for the simulations. The first set contained only the backbone angles Φ and Ψ , the second set also contained the χ_1 and χ_2 angles, and the third set contained the angles $\chi_1-\chi_5$ in addition to the backbone dihedral angles, which is sufficient to specify the conformation of all residues (see section 2.6 and Table S2 in the Supporting Information). For the initial DSMD simulations, the target structures were the energy-minimized X-ray and NMR structures. As shown in Table 2, however, the RMSD between the final simulation structures and the target structures could not be reduced below ~4 Å; an improvement from 7.3 to 4.7 Å was observed for the open → closed simulation when dihedral set 2 was used instead of set 1, and a further improvement to 3.8 Å resulted from using set 3. In the second set of DSMD simulations, equilibrated X-ray and NMR structures were used as the target structures. When dihedral set 3 was used along with the equilibrated target structures, the RMSD between the final simulation and the target structures decreased to 1.1 and 1.5 Å for the open → closed and closed → open directions, respectively. (However, when the thermostat was switched off to monitor the increase in the temperature, the RMSD was again higher; see Table 2.) The reasons for the dependence of the final RMSD on the choice of target structure are explored at the end of this section.

Parts d_i–d_{iii} of Figure 2 correspond to the DSMD simulations performed with dihedral set 3 and equilibrated target structures. The simulation behavior is qualitatively more similar to the LRTMD simulations than the RTMD simulations. The projection of trajectory displacements in

Figure 2d_{iii} shows that the DSMD simulations do not have a length scale bias. (These plots are similar to Figures 2c_{iii} and 8, which correspond to the LRTMD simulations.) Figure 2d_{ii} shows that changes in small loops are approximately in tune with the large-scale changes. The increase in the simulation temperature with the thermostat turned off is similar to that in the LRTMD simulations (Figure 4c and d), occurring gradually in the simulation. As in Figure 2c_i (LRTMD case), the evolution of the RMSD function ρ is not necessarily monotonic (Figure 2d_i). A significant difference between the DSMD and the other simulations is in the behavior of the RMSD function in the open → closed trajectory. Although the target value of the biasing potential ε_0 reaches zero at $t = 1.2$ ns (see Methods), the RMSD between the simulation and the target structures at this time is ≈ 9 Å. This behavior can be understood by considering four simulation snapshots corresponding to 1.2, 1.3, 1.4, and 1.5 ns, which are shown in Figure 9. The individual

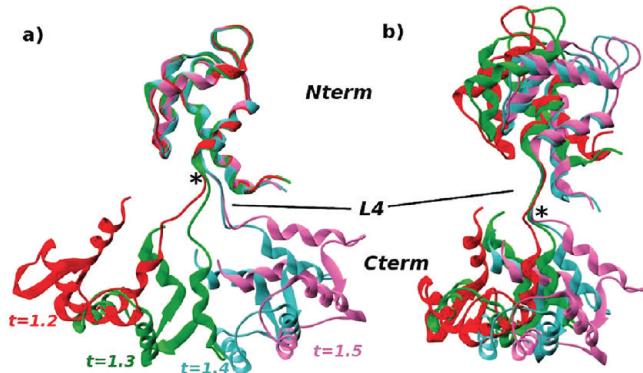


Figure 9. Four simulation structures from the open → closed transition in CaM enforced by DSMD. The structures correspond to $t = 1.2, 1.3, 1.4$, and 1.5 ns. The superposition is performed using the *Nterm* subdomain in part a and using the L4 loop in part b.

subdomains are very close to the corresponding target positions by $t = 1.2$ ns (e.g., *Nterm* in Figure 9a and linker loop L4 in Figure 9b), but the global alignment changes considerably between $t = 1.2$ ns and $t = 1.5$ ns. In particular, Figure 9 suggests that the structural differences are caused by small changes in the conformation of L4 (marked by asterisks in Figure 9). Although L4 is, essentially, in the target state (see also Figure 2d_{ii}), small conformational differences near the termini of L4 (res. 77 and 81; see Table S1 in the Supporting Information for subdomain definitions) are amplified into large differences in the relative positions of *Nterm* and *Cterm*, which are responsible for the large values of the RMSD (Figure 2d_i) at $t = 1.2$ ns. An additional 300 ps of equilibration in the presence of the fixed restraint (from $t = 1.2$ ns to $t = 1.5$ ns) results in slow relative motion of the *Nterm* and *Cterm* subdomains, which reduces the Cartesian RMSD to 1.1 Å by $t = 1.5$ ns. We examined the evolution of the dihedral angles in the period from $t = 1.2$ ns to $t = 1.5$ ns and found that the backbone dihedral angles involving loop L4 were farther from their target values than most of the dihedrals in the forcing set, even though their RMSD sum $\epsilon(\hat{r}, \hat{r}_T)$ in eq 8 did not change in the trajectory period. The findings indicate that, for highly flexible domains such as loop L4 in CaM, applying restraints in internal coordinates may not be an efficient way to enforce Cartesian positions. For such cases, the LRTMD method, which retains the locality of the restraint inherent in DSMD but employs Cartesian coordinates, appears to be a better-suited alternative.

The causes of the higher RMSD between the final simulation and target structures that were observed with the energy-minimized target structures are discussed in the Supporting Information.

3.6. Free Energy Profiles along Transition Paths. It was demonstrated in the preceding sections that the use of all biasing methods in the absence of a thermostat leads to a large increase in the temperature by the end of the simulations, indicating that a large amount of nonequilibrium work is performed on the simulation system. It was also shown in section 3.4 (see Figure 4c) that the temperature increase is sensitive to the value of the force constant used in the biased dynamics. Since the values for the force constants cannot be compared directly for different biasing methods, the magnitude of the temperature increase in a simulation is not an accurate indicator of the quality of generated paths, as discussed above. To estimate the relative thermodynamic importance of the transition paths generated using the biasing methods presented in this study, we used the finite-temperature string method,^{15,16,74} as described in section 2.4.

The FE profiles corresponding to transition paths generated by RTMD, LRTMD, and DSMD are compared in Figure 10.

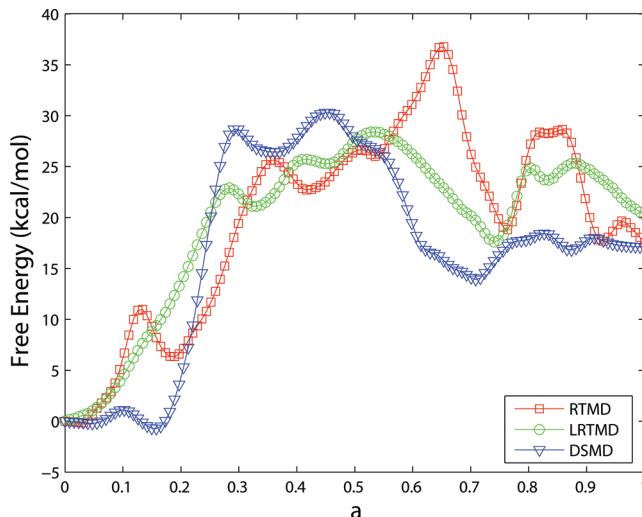


Figure 10. Free energy profiles associated with a narrow transition tube corresponding to three initial paths (see text). The open and closed states correspond to $\alpha = 0$ and $\alpha = 1$, respectively.

Although the FE calculations are not sufficiently long to obtain error estimates directly, we found that the FE difference between the end point states varies by less than 5 kcal/mol, depending on the transition path used. We therefore take this value as a rough measure of uncertainty in the FE values.

First, we note that the end point $\alpha = 1$, which corresponds to the equilibrated closed structure, is higher in FE by ≈ 18 kcal/mol than the end point $\alpha = 0$, which corresponds to the equilibrated open structure after the removal of Ca²⁺ ions. While there are no experimental data concerning the free energy difference between the Ca²⁺-free solution structure and the Ca²⁺-free X-ray structure, which could be used to validate the calculations, studies of CaM by solution NMR observe that the molecule samples a very large ensemble of conformations,^{49,84,85} implying that the conformation observed in the open X-ray crystal structure⁵⁰ should be accessible at room temperature. The rather large value of the FE difference computed here is in conflict with this possibility. The large

calculated FE difference could be due to the low resolution NMR model of the closed structure, in which the atomic coordinates are not as optimal as in the high (1.0 Å) resolution of the open X-ray structure. An additional factor is the mutation of the 18 residues in the closed structure that was necessary to maintain the same residue sequence for both structures (see section 2.5).

We note that the temperature increase in the unthermostatted biased trajectories (Figure 4) is similar in both steering directions, suggesting that the above FE difference is small in comparison to the nonequilibrium work generated by the steering protocols. Figure 10 also shows that the free energy barriers associated with the three paths are similar within the uncertainty estimate (≈ 5 kcal/mol). This suggests that the paths make comparable contributions to the ensemble of all accessible transition paths, implying that conformational transitions can occur *via* a variety of different paths. As mentioned above, the relatively high barriers in Figure 10 (≈ 30 kcal/mol, starting from the open structure) may be overestimates caused by the constraint of a narrow transition tube. We also note that the generation of unbiased trajectories using TPS for processes that have multiple free energy barriers (such as those in Figure 10) has been achieved only by splitting the transition paths into disjoint segments, with each segment containing a single barrier.³¹

4. CONCLUDING DISCUSSION

A major reason for the popularity of targeted molecular dynamics (TMD) and its variants is that it can produce conformational transition paths within a very short simulation time. However, in all existing TMD algorithms, large-scale motions tend to precede small-scale motions.^{26,29,36,44} This length scale bias is a concern because it implies that the order of transition events observed in a simulation depends on the direction in which the simulation is performed; i.e., standard TMD and restrained TMD (RTMD) methods are inherently hysteretic. It seems difficult to reconcile this property with the generally accepted belief that most conformational transitions occur near thermodynamic equilibrium. In particular, the dominant conformational transition pathways are expected to be microscopically reversible. This argument provides the motivation for developing transition modeling methods in which the length scale bias is absent or diminished.

The present results indicate that the differences in the numerical techniques by which the biasing is implemented in TMD or RTMD do not produce significant differences in the order of transition events (a similar conclusion was reached by Huang et al.³⁶). Using the scaled-force TMD (SFTMD) method, in which the magnitude of the maximum force applied on an atom is independent of its displacement vector from the current to the target position, we established that differences in the magnitude of the forces applied to simulation atoms are, by themselves, not sufficient to cause “large-scales-first” order of transition events. The similarities in the trajectory ensembles produced with TMD, RTMD, and SFTMD (Figure 2) can be explained by a step that is performed identically in the three TMD variants, i.e., the global best-fit alignment procedure. (In the present calculations, the alignment is repeated at every simulation step. Although it can be performed once in a larger number of MD steps, typically to achieve a modest increase in speed, our tests indicate that, when the best-fitting is performed once in every 10 iterations, the trajectories are not significantly changed. In addition, the resulting dynamics will not be

Hamiltonian, since the forces can no longer be expressed as the gradient of a potential.) The effects of the best-fitting can be understood by referring to a well-known property of normal mode vectors (NMVs). For two conformations of a protein that are superposed using a best-fit alignment, the displacement vector pointing from one conformation to the other is often well approximated by a small number of low-frequency NMVs.^{45,46,82,83} For example, ref 45 reports that about 50% of the conformational change in the recovery stroke of myosin II can be represented by 10 NMV vectors. Further, the higher-frequency NMVs have a progressively smaller overlap with the displacement vector as the frequency increases.⁴⁵ Since in the TMD (as well as in SFTMD) the force vectors are aligned with the displacement vector (see Methods), the best-fitting biases the force vectors to lie in a subspace spanned by a relatively small number of low-frequency normal modes. As a result, trajectories produced by all three TMD variants tend to follow low-frequency normal modes until the associated large scale structures are in their corresponding target conformations (which occurs around 1 ns in the present simulations), and the remainder of the conformational change is dominated by high-frequency modes (see Figure 2a_{iii} and b_{iii}).

Two variants of biased dynamics, in which the direction of the force vectors is not directly related to the global best-fitting between simulation and target structures, were introduced into CHARMM. The locally restrained TMD (LRTMD) is similar in implementation to the standard RTMD method in that it is formulated in Cartesian coordinates, yet it differs conceptually in that the biasing forces are derived from enforcing directly a series of local conformational changes, with the global transition resulting from the combined effect of the local changes. In particular, the global best-fit alignment is replaced by best-fit alignments based on the corresponding local structures. The results of section 3.4 show that the early large-scale displacements are absent in trajectories generated by this method. The breadth of the path ensemble generated with the LRTMD method indicates that the local clusters of the structure, to some extent, evolve independently. A broad ensemble also suggests that energetic barriers for conformational transitions in large proteins are offset by entropic contributions that arise from a large number of possible paths. The lowering of free energy barriers by entropic contributions from a broad transition state ensemble (TSE) has been predicted in coarse-grained models of protein folding⁸⁶ and allosteric transitions.⁸⁷ We found a rather broad TSE in a recent study of a conformational transition in myosin VI, which was performed using the string method.³⁴ A broad TSE was also observed by Junker and Rief⁸⁸ in their experiments of calmodulin unfolding by mechanical force.

The fact that the LRTMD potential is constructed from “local” contributions is not in contradiction with the importance of low-frequency normal modes in conformational transitions. First, the low-frequency content of the trajectory displacements in the LRTMD simulations is essentially as high as that in standard TMD/RTMD trajectories. In fact, after ~ 1 ns, TMD/RTMD trajectory displacements have progressively smaller overlap with low-frequency modes than those obtained with LRTMD (see Figure 2a_{iii}). In addition, the importance of low-frequency motions in describing conformational transitions does not imply a particular order of events. Equipartition of energy (EoE) in the space of NMV predicts that the mass-weighted mean-square displacement of the system along the *i*th normal mode scales as $k_B T / \nu_i^2$, where ν_i is the corresponding

frequency, implying that equilibrium excursions along low-frequency modes are higher in amplitude than those along high-frequency modes. Qualitatively, motions along low-frequency modes are expected to encounter less resistance than motions along high-frequency modes, but this result, *per se*, does not give information about the order of events in the transition.

In the other variant, the dihedral SMD method (DSMD), forces were applied to sets of dihedral angles. When equilibrated structures were used as the target structures, the simulation trajectories converged to the target structures to within 1–1.5 Å RMSD. As in the LRTMD case, DSMD paths did not exhibit the “large-scales-first” feature. In particular, the overlap coefficients obtained by projecting the trajectory displacements onto the normal mode vectors are approximately constant along the trajectories, which was also consistent with a direct examination of the trajectories.

For completeness, in the Supporting Information, we compare the results obtained using biased molecular dynamics (BMD)^{32,37} with the heavy-atom RMSD as the progress variable and restricted perturbation (RP) TMD.²⁶ These methods are conceptually similar in that both aim to reduce the accumulation of dissipative work by restricting the target value of the potential $\rho_0(t)$ to a vicinity of the actual RMSD value $\rho(\hat{r}, \hat{r}_T)$. With the RMSD function in eq 3 as the progress variable, both methods produce trajectories which retain the “large-scales-first” feature. This finding is consistent with the results obtained with the other TMD variants, which indicate that this bias is less dependent on the differences between the constraint vs restraint implementation than on the direction of the applied force.

As discussed in the Results, the transition paths generated with LRTMD and DSMD are qualitatively different from those generated with TMD or RTMD. In paths obtained with TMD/RTMD, the transition begins with the relative reorientation of the *Nterm* and *Cterm* domains, followed by the closing of the EF-hands in these domains, and terminating with the rearrangement of the loops L1–L7. In contrast, both LRTMD and DSMD produce paths in which the loop rearrangements either precede or occur simultaneously with EF-hand closure, which generally precedes the relative reorientation of the *Nterm* and *Cterm* domains. An outstanding question is which of the transition paths obtained in this study are most physically meaningful, i.e., which ones would be observed with a high probability in an unbiased ensemble of transition paths? Answering this question directly would require the generation of an unbiased ensemble of paths, which is computationally infeasible for large biomolecular systems. To obtain an approximate answer, we used the finite-temperature string method to calculate the free energy profiles along a narrow transition tube associated with each path. The free energy barriers along the paths obtained from RTMD, LRTMD, and DSMD were found to be similar, suggesting that they were meaningful, and that transitions should proceed *via* the paths with similar probabilities. However, the use of a narrow transition tube, which was required to obtain convergence, may lead to overestimates of the free energy at corresponding locations along the path.

To obtain more information about the physical significance of the paths obtained in this study, we consider some experimental data for calmodulin. As mentioned in section 2.5, the removal of Ca²⁺ ions alters the dynamics of the open structure, and is likely to impact the character of the transition

paths generated by all of the methods. Despite this caveat, paths generated with the LRTMD method for the open \leftrightarrow closed transition in CaM appear physically meaningful. In particular, the fact that the Ca²⁺-binding loops approach the target conformations before the global structure is consistent with the experimental observations that calcium binding drives conformational change in the EF-hands.^{49,54,89} The fact that Ca²⁺ concentration affects the folding rate of CaM but has no effect on the rate of unfolding⁸⁹ supports an induced-fit, rather than a pre-equilibration mechanism of CaM folding.⁹⁰ In contrast, the association of CaM with target domains is thought to proceed *via* the pre-equilibration mechanism.^{84,89,90} The relatively high mean values and variability of the global RMSD function $\rho(t)$ in the LRTMD ensemble at $t \approx 600$ ps (Figure 2c_i), after the smaller-scale structure in the Ca²⁺-binding loops is, essentially, in the target conformation (see insets in Figure 7), probably reflects the variable relative orientation of the *Nterm* and *Cterm* domains, as observed for ensembles of *apo*CaM structures in solution.^{49,57,61,80} The variability of transition paths obtained with the various methods appears to be consistent with a broad transition state ensemble for CaM folding observed by Junker and Rief.⁸⁸ For further validation of the present paths, however, additional simulations are necessary, e.g., those in which the conformational change in the EF-hand helices is not explicitly forced but emerges only as a consequence of rearrangements in the Ca²⁺-binding loops. Moreover, properly accounting for the presence of Ca²⁺ ions is likely to be essential for capturing the actual dynamics of loop transitions.

APPENDIX A

To derive eq 4 from eq 2, we have

$$\begin{aligned}
 F_{x_m^*} &= -\frac{\partial U}{\partial x_m^*} \\
 &= -\sum_{i=1}^N \frac{\partial U}{\partial x_i} \frac{\partial x_i}{\partial x_m^*} \\
 &= -k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) \sum_{i=1}^N \sum_{j=1}^N w_j (r_j - Ar_{T,j})^t \\
 &\quad \times \left(\frac{\partial r_j}{\partial x_i} - \frac{\partial A}{\partial x_i} r_{T,j} \right) (\delta_{im} - w_m) \\
 &= -k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) \left[w_m (x_m - A_x r_{T,m}) \right. \\
 &\quad \left. - w_m \sum_{i=1}^N w_i (x_i - A_x r_{T,i}) \right. \\
 &\quad \left. - \sum_{j=1}^N w_j (r_j - Ar_{T,j})^t \frac{\partial A}{\partial x_m} r_{T,j} \right. \\
 &\quad \left. + w_m \sum_{j=1}^N w_j (r_j - Ar_{T,j})^t \left(\sum_{i=1}^N \frac{\partial A}{\partial x_i} \right) r_{T,j} \right] \quad (A1)
 \end{aligned}$$

where A is the (best-fit) rotation matrix for which the sum $\sum_{n=1}^N w_n \|r_n - Ar_{T,n}\|^2$ is minimized, $(\cdot)^t$ denotes matrix transpose, A_x is the first row of A , δ_{im} is the Kronecker delta

tensor, and we used $\partial x_i / \partial x_m^* = \delta_{im} - w_m$. The last three terms in eq A1 are identically zero. The second term in eq A1 is zero because the centers of mass (COMs) of \hat{r} and \hat{r}_T are zero by construction. To see that the third term is zero, we use A' to denote $\partial A / \partial x_m$ and differentiate the identity $AA^t = I$ for rotation matrices, to obtain $A'A^t = -AA^t = -(A'A^t)^t$, which shows that $A'A^t$ is skew-symmetric. Using $A'A^t$ to represent an infinitesimal rotation through an angle ϕ^{91} and the fact that A corresponds to the optimal rotation, we have

$$\begin{aligned} & \sum_{j=1}^N w_j \|r_j - (Ar_{T,j} + \phi(A'A^t)Ar_{T,j})\|^2 \\ & \sum_{j=1}^N w_j \|r_j - Ar_{T,j}\|^2 \end{aligned} \quad (\text{A2})$$

In the limit $\phi \rightarrow 0$, eq A2 can be written as

$$\begin{aligned} 0 &= \frac{d}{d\phi} \left[\sum_{j=1}^N w_j \|r_j - (I + \phi A'A^t)Ar_{T,j}\|^2 \right]_{\phi=0} \\ &= \sum_{j=1}^N w_j (r_j - Ar_{T,j})^t A' r_{T,j} \end{aligned} \quad (\text{A3})$$

as required. We note that the last term can be written as $\sum_{j=1}^N w_j r_j^t A' r_{T,j}$, because A is a pure rotation and therefore $2(Ar_{T,j})^t A' r_{T,j} = (\|Ar_{T,j}\|^2)' = (\|r_{T,j}\|^2)' = 0$. We note also that, if we use an arbitrary vector u to represent the infinitesimal rotation,⁹¹ eq A3 becomes

$$\begin{aligned} 0 &= \frac{d}{d\phi} \left[\sum_{j=1}^N w_j \|r_j - (I + \phi u \times)Ar_{T,j}\|^2 \right]_{\phi=0} \\ &= - \sum_{j=1}^N w_j (r_j - Ar_{T,j})^t u \times Ar_{T,j} \\ &= u^t \sum_{j=1}^N w_j r_j \times Ar_{T,j} \end{aligned} \quad (\text{A4})$$

Since u is arbitrary, we must have $\sum_{j=1}^N w_j r_j \times Ar_{T,j} = 0$, which is equivalent to the second Eckart condition.⁶⁹ (The first condition holds because the COMs of \hat{r} and \hat{r}_T are zero.) The last term in eq A1 is zero because $\sum_{i=1}^N \partial A / \partial x_i$ is zero (shown in the Supporting Information).

The need to calculate contributions to the force from terms such as $w_j r_j^t A' r_{T,j}$ arises if it is desirable to perform best-fitting using one subset of the structure, and to enforce the RMSD restraint using a different subset. Such an approach may be useful if certain subdomains of the initial and target structures of a molecule are in similar conformations, while other subdomains are in different conformations. In this case, one can perform best-fit alignment using only the similar subdomains ("orientation" set), and apply targeting forces to the dissimilar subdomains ("forcing" set). In the case that the orientation set is different from the forcing set, eq A1 is slightly modified. We let o and f denote orientation and forcing atoms,

respectively, and introduce two different sets of weights, w^o and w^f , which are used to compute the rotation matrix and the RMSD, i.e., A is chosen from the group of rotation matrices to minimize the sum

$$\sum_{j=1}^{N_o} w_j^o \|\hat{r}_{o,j} - B\hat{r}_{T,o,j}\|^2 : B^t B = I, \det B = 1$$

$$\rho(\hat{r}, A\hat{r}_T) = \left[\sum_{j=1}^{N_f} w_j^f \|\hat{r}_{f,j} - A\hat{r}_{T,f,j}\|^2 \right]^{1/2} \quad (\text{A5})$$

where $\hat{r} = \hat{r}_* - \sum_{j=1}^{N_o} w_j^o r_j^*$ and N_o and N_f are the number of orientation and forcing atoms, respectively. For convenience, we write two separate equations for the force on the orientation and forcing atoms:

$$\begin{aligned} F_{x_{o,m}}^{o*} &= k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) \left[w_m^o \sum_{j=1}^{N_f} w_j^f (x_{f,j} - A_x r_{T,f,j}) \right. \\ &\quad \left. + \sum_{j=1}^{N_f} w_j^f r_{f,j}^t \frac{\partial A_x}{\partial x_{o,m}} r_{T,f,j} \right] \end{aligned} \quad (\text{A6})$$

$$F_{x_{f,m}}^{f*} = -k \left(1 - \frac{\rho_0}{\rho(\hat{r}, A\hat{r}_T)} \right) w_m^f (x_{f,m} - A_x r_{T,f,m}) \quad (\text{A7})$$

Note that overlapping orientation and forcing sets may also be defined. In the case that an atom belongs to both sets, eqs A6 and A7 represent two contributions of the total RTMD force acting on the atom. (In the special case in which the forcing set contains the orientation set, and the orientation weights are proportional to the corresponding forcing weights, i.e., $w_j^o = \alpha w_j^f$ for some α , the sums in eq A6 may exclude all indices j if the corresponding atoms belong also to the orientation set, since their contribution to the force will vanish in view of eqs A1–A3.) Equation A7 is identical to eq 4; the additional forces in eq A6 act on the orientation atoms to ensure that no net force or torque is applied to the simulation system. If the orientation and forcing sets are not identical, the Eckart conditions will not hold. The Eckart conditions are used to define a coordinate frame for a molecule such that the rigid-body degrees of freedom (DOF) are approximately separable from the internal DOF.^{69,92,93} Such a coordinate frame does not play a direct role in the RTMD method, and thus the Eckart conditions need not be satisfied.

ASSOCIATED CONTENT

Supporting Information

Additional details concerning preparation of simulation structures, analysis of the simulation data, and miscellaneous derivations are provided in Supporting Text sections S1–S5, Tables S1 and S2, and Figures S1–S6. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ovchinnv@georgetown.edu (V.O.); marci@tammy.harvard.edu (M.K.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Kwangho Nam for many helpful discussions. The work was supported by the Human Frontiers Science Program and by the National Institutes of Health. Supercomputing resources were provided by the National Energy Resource Supercomputing Centers. V.O. acknowledges financial support under the NRSA fellowship 1F32GM083422-01.

REFERENCES

- (1) Benkovic, S.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196–1202.
- (2) Johnson, K. *Annu. Rev. Biochem.* **1993**, *62*, 685–713.
- (3) Joyce, C.; Steitz, T. *Annu. Rev. Biochem.* **1994**, *63*, 777–822.
- (4) Ma, J.; Siegler, P.; Xu, Z.; Karplus, M. *J. Mol. Biol.* **2000**, *302*, 303–313.
- (5) Karplus, M.; Gao, Y. *Curr. Opin. Struct. Biol.* **2004**, *14*, 250–259.
- (6) Houdusse, A.; Sweeney, H. *Curr. Opin. Struct. Biol.* **2001**, *11*, 182–194.
- (7) Geeves, M.; Holmes, K. *Annu. Rev. Biochem.* **1999**, *68*, 687–727.
- (8) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (9) Bartels, C.; Schaefer, M.; Karplus, M. *J. Chem. Phys.* **1999**, *111*, 8048.
- (10) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
- (11) Hénin, J.; Chipot, C. *J. Chem. Phys.* **2005**, *123*, 244906.
- (12) Rosso, L.; Tuckerman, M. *Mol. Simul.* **2002**, *28*, 91.
- (13) Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168–175.
- (14) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.
- (15) E, W.; Ren, W.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2005**, *413*, 242–247.
- (16) Ren, W.; Vanden-Eijnden, E.; Maragakis, P.; E, W. *J. Chem. Phys.* **2005**, *123*, 134109.
- (17) Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375–380.
- (18) Fisher, S.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 252–261.
- (19) Jónsson, G.; Jacobsen, K. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B., Ciccotti, G., Coker, D., Eds.; World Scientific: Singapore, 1998; pp 385–404.
- (20) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.
- (21) Branduardi, D.; Gervasio, F.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.
- (22) Bolhuis, P.; Chandler, D.; Dellago, C.; Geissler, P. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
- (23) Dellago, C.; Bolhuis, P.; Geissler, P. *Adv. Chem. Phys.* **2002**, *123*, 1.
- (24) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.
- (25) Schlitter, J.; Engels, M.; Krüger, P. *J. Mol. Graphics* **1994**, *12*, 84–89.
- (26) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2005**, *122*, 114903.
- (27) Ma, J.; Flynn, T.; Cui, Q.; Leslie, A.; Walker, J.; Karplus, M. *Structure* **2002**, *10*, 921–931.
- (28) Madhusoodanan, M.; Lazaridis, T. *Biophys. J.* **2003**, *84*, 1926–1939.
- (29) Yu, H.; Ma, L.; Yang, Y.; Cui, Q. *PLoS Comput. Biol.* **2007**, *3*, 0214–0230.
- (30) Ovchinnikov, V.; Trout, B.; Karplus, M. *J. Mol. Biol.* **2010**, *395*, 815–833.
- (31) Radhakrishnan, R.; Schlick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5970–5975.
- (32) Hu, J.; Ma, A.; Dinner, A. *J. Chem. Phys.* **2006**, *125*, 114101.
- (33) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106.
- (34) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. *J. Chem. Phys.* **2011**, *134*, 085103.
- (35) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099–2108.
- (36) Huang, H.; Ozkirimli, E.; Post, C. *J. Chem. Theory Comput.* **2009**, *5*, 1304–1314.
- (37) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441–459.
- (38) Izrailewitz, B.; Baudry, J.; Kosztin, J. G. D.; Schulten, K. *J. Mol. Graphics Modell.* **2001**, *19*, 13–25.
- (39) Grubmüller, H.; Heyman, B.; Tavan, P. *Science* **1996**, *271*, 997–999.
- (40) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568–1581.
- (41) Paci, E.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6521–6526.
- (42) van der Vaart, A.; Ma, J.; Karplus, M. *Biophys. J.* **2004**, *87*, 562–573.
- (43) Li, J.-L.; Geng, C.-Y.; Bu, Y.; Huang, X.-R.; Sun, C.-C. *J. Comput. Chem.* **2009**, *30*, 1135–1145.
- (44) Golosov, A. A.; Warren, J. J.; Beese, L. S.; Karplus, M. *Structure* **2010**, *18*, 83–93.
- (45) Li, G.; Cui, Q. *Biophys. J.* **2004**, *86*, 743–763.
- (46) Cecchini, M.; Houdusse, A.; Karplus, M. *PLoS Comput. Biol.* **2008**, *4*, e1000129.
- (47) Cui, Q.; Karplus, M. *Protein Sci.* **2008**, *17*, 1295–1307.
- (48) Tsai, C.-J.; del Sol, A.; Nussinov, R. *Mol. BioSyst.* **2009**, *5*, 205–216.
- (49) Zhang, M.; Tanaka, T.; Ikura, M. *Nat. Struct. Biol.* **1995**, *2*, 758–767.
- (50) Wilson, M.; Brunger, A. *J. Mol. Biol.* **2000**, *301*, 1237–1265.
- (51) Hoeflich, K.; Ikura, M. *Cell* **2002**, *108*, 739–742.
- (52) Johnson, C. K. *Biochemistry* **2006**, *45*, 14233–14246.
- (53) Schumacher, M.; Crum, M.; Miller, M. *Structure* **2004**, *12*, 849–860.
- (54) Grabarek, Z. *J. Mol. Biol.* **2005**, *346*, 1351–1366.
- (55) Valeev, N.; Bates, D.; Heslop-Harrison, P.; Postlethwaite, I.; Kotov, N. *BMC Syst. Biol.* **2008**, *2*, 48.
- (56) Shen, X.; Valencia, C.; Szostak, J.; Dong, B.; Liu, R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5969–5974.
- (57) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990–997.
- (58) Wyttenbach, T.; Grabenauer, M.; Thalassinos, K.; Scrivens, J. H.; Bowers, M. T. *J. Phys. Chem. B* **2010**, *114*, 437–447.
- (59) Vigil, D.; Gallagher, S.; Trewella, J.; Garcia, A. *Biophys. J.* **2001**, *80*, 2082–2092.
- (60) Chen, Y.-G.; Hummer, G. *J. Am. Chem. Soc.* **2007**, *129*, 2414–2415.
- (61) Bertini, I.; Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zaddor, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6841–6846.
- (62) Frederick, K.; Marlow, M.; Valentine, K.; Wand, A. *Nature* **2007**, *448*, 325–330.
- (63) Farrell, J. L.; Stuelpnagel, J. C.; Wessner, R. H.; Velman, J. R.; Brook, J. E. *SIAM Rev.* **1966**, *8*, 384–386.
- (64) Kabsch, W. *Acta Crystallogr.* **1976**, *A32*, 922–923.
- (65) Horn, B. *J. Opt. Soc. Am. A* **1987**, *4*, 629–642.
- (66) Coutsias, E.; Seok, C.; Dill, K. *J. Comput. Chem.* **2004**, *25*, 1849–1857.
- (67) Schlitter, J.; Swegat, W.; Mülders, T. *J. Mol. Model.* **2001**, *7*, 171–177.
- (68) Banavali, N.; Roux, B. *Structure* **2005**, *13*, 1715–1723.
- (69) Eckart, C. *Phys. Rev.* **1935**, *47*, 552–558.
- (70) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580–5601.
- (71) Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (72) Kaneko, I.; Plemons, R. *Int. J. Numer. Methods Eng.* **1984**, *20*, 983–998.
- (73) Cline, R.; Plemons, R. *SIAM Rev.* **1976**, *18*, 92–106.
- (74) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
- (75) Bennett, M.; Choe, S.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *9*, 3127–3131.

- (76) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (77) Brooks, B.; Brooks, C. III; Mackerell, A. Jr.; Nilsson, L.; Petrella, R.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (78) MacKerell, A. Jr.; Feig, M.; Brooks, C. III. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (79) Haberthür, U.; Caflisch, A. *J. Comput. Chem.* **2007**, *29*, 701–715.
- (80) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C.; Bax, A. *Nat. Struct. Biol.* **1995**, *2*, 768–776.
- (81) Brooks, B.; Janežič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522–1542.
- (82) Ma, J.; Karplus, M. *J. Mol. Biol.* **1997**, *274*, 114–131.
- (83) Tama, F.; Sanejouand, Y. *Protein Eng.* **2001**, *14*, 1–6.
- (84) Swindells, M.; Ikura, M. *Nat. Struct. Biol.* **1996**, *3*, 501–504.
- (85) Evenas, J.; Forsen, S.; Malmendal, A.; Akke, M. *J. Mol. Biol.* **1999**, *289*, 603–617.
- (86) Du, R.; Pande, V.; Grosberg, A.; Tanaka, T.; Shakhnovich, E. *J. Chem. Phys.* **1998**, *109*, 334–350.
- (87) Itoh, K.; Sasai, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 7775–7780.
- (88) Junker, J.; Rief, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 3306–3309.
- (89) Junker, P.; Ziegler, F.; Rief, M. *Science* **2009**, *323*, 633–637.
- (90) Best, R.; Hummer, G. *Science* **2009**, *323*, 593–594.
- (91) Altmann, S. *Rotations, Quaternions, and Double Groups*; Dover Publications: New York, 2005.
- (92) Sayvets, A. *J. Chem. Phys.* **1939**, *7*, 383–389.
- (93) Ferigle, S.; Weber, A. *Am. J. Phys.* **1953**, *21*, 102–107.
- (94) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.